



**CATÓLICA
LISBON**
BUSINESS & ECONOMICS

**FROM DATA TO INNOVATION:
EMPOWERING SUSTAINABLE NEW PRODUCT
DEVELOPMENT WITH AI AND DATA SCIENCE
- INSIGHTS FROM PORTUGUESE INDUSTRIES**

Ricardo David Francisco

Dissertação apresentada para cumprimento parcial dos requisitos para a obtenção do grau de Mestre em Gestão Aplicada (Applied Management) na Universidade Católica Portuguesa.

Setembro de 2023

Dissertação escrita sob a supervisão do Professor Doutor Paulo Cardoso do Amaral.

ABSTRACT

TITLE: From Data to Innovation: Empowering Sustainable New Product Development with AI and Data Science - Insights from Portuguese Industries

AUTHOR: Ricardo David Francisco

Unlocking the potential of AI and Data Science (AI/DS) in New Product Development (NPD) stands as a promising avenue for innovation acceleration. According to Gomes (2022), despite recognizing the potential of these technologies for better decision-making, many Portuguese firms lack the resources, hindering competitiveness and innovation.

This dissertation examines this paradox, aiming to reveal the AI/DS potential for 'Innovation Analytics' in Portugal's NPD landscape. Through a mixed-methods approach involving surveys, semi-structured interviews, and a practical application, this study investigates the impact of AI/DS adoption on sustainable NPD.

Key contributions of this research include a proposed conceptual framework harmonizing both AI/DS and functional departments with Cooper's Stage-Gate model (potentially acting as a roadmap for AI/DS implementation for NPD), insights from AI/DS adoption and a practical examples for rapid innovation using AI tools.

The study's insights highlight correlations between AI/DS interest and innovation in NPD, supported by real-world cases. Additionally, a practical application employing AI tools like DALL·E and ChatGPT demonstrates these technologies transformative potential for rapid innovation, improving market research efficiency and product concept visualization.

While some Portuguese industry sectors exhibited varied AI/DS adoption profiles, the study found that these sectoral differences were not significant.

The insights offered provide guidance for venturing into the realm of AI/DS-powered sustainable innovation.

Keywords: Sustainable New Product Development; Artificial Intelligence and Data Science in NPD, Innovation Analytics, Stage-Gate, Portuguese Industries

RESUMO

TÍTULO: Dos Dados à Inovação: Como potenciar o Desenvolvimento Sustentável de Novos Produtos com IA e Ciência de Dados - Insights da Indústria Portuguesa

AUTOR: Ricardo David Francisco

A inclusão da Inteligência Artificial e Ciência de Dados (IA/CD) no Desenvolvimento de Novos Produtos (DNP) surge como potencial acelerador da inovação. Segundo Gomes (2022) apesar de considerarem estas tecnologias para uma melhor tomada de decisão, muitas empresas portuguesas carecem de recursos, limitando a competitividade e inovação.

Pretende-se examinar o potencial da IA/CD no contexto do DNP sustentável em Portugal. Utilizando uma abordagem com métodos mistos: inquéritos, entrevistas semiestruturadas e uma aplicação prática, investigou-se o impacto da IA/CD no DNP.

As contribuições-chave englobam uma proposta de *framework* conceptual alinhando IA/CD com departamentos funcionais, baseada no modelo *Stage-Gate* de Cooper (servindo como guia de implementação), insights das diversas indústrias na adoção da IA/CD e um exemplo prático para a inovação sustentável.

Foram descobertas correlações entre o interesse em IA/CD e a inovação no DNP. Adicionalmente, um guia prático, utilizando ferramentas de IA como DALL·E e ChatGPT, evidencia o potencial transformador destas tecnologias para inovação célere, melhorando a pesquisa de mercado e a visualização de conceitos de produtos.

Apesar de alguns setores da indústria portuguesa apresentarem perfis variados de adoção de IA/CD, concluiu-se que essas diferenças setoriais não são significativas.

Reconhecendo limitações, esta dissertação fornece orientações no domínio da inovação sustentável impulsionada pela IA/CD.

Palavras-chave: Desenvolvimento Sustentável de Novos Produtos; Inteligência Artificial e Ciência de Dados no DNP, Innovation Analytics, Stage-Gate, Indústrias Portuguesas

ACKNOWLEDGMENTS

I would like to express my deep gratitude to the following people and institutions, whose steadfast support, love, and encouragement played a key role in the conclusion of this dissertation:

To my family - my foundation, my haven and constant source of inspiration. Thank you, mother, father, and grandmother, for your boundless love, endless sacrifices, and unwavering faith in my abilities. Your support has been the wind that supports my wings throughout this journey.

To my beloved wife, my traveling companion – your patience and unwavering support have been my guiding light. Your presence in my life is the most precious gift, and I am deeply grateful for the countless words of encouragement, true and dedicated love. Together we are one in search of great achievements.

To my wife's family - thank you for your continued support and firm belief that “all will be well”. Your kindness has made this journey even more enriching.

To my esteemed advisor, Professor Paulo Cardoso do Amaral - his guidance, experience, and vote of confidence in me were invaluable for my academic and personal growth. In fact, your belief in my potential has been a driving force behind my academic and professional success.

To my cherished company, Nimble, and to Marco Vidal, for allowing me to have space and time for the writing of this dissertation and for the undeniable support until its completion.

To the people representing the companies that participated in this study – your willingness to collaborate and share their insights has enriched the depth and quality of my research. Your contributions were indispensable to achieve the objectives of this study.

To Católica Lisbon School of Business & Economics - for providing me with the resources, the environment, and the opportunities to pursue higher education and research. I am grateful for the platform they have given me to develop my skills and contribute to the academic community.

To the people who dream of achieving their goals and thrive every day for a better world:

“Standing on the shoulders of giants is how progress is made.” - *Artificial Intelligence*

TABLE OF CONTENTS

ABSTRACT	i
RESUMO	iii
ACKNOWLEDGMENTS.....	v
TABLE OF CONTENTS	vii
LIST OF FIGURES.....	xiii
LIST OF TABLES	xv
LIST OF ABBREVIATIONS	xvii
1. INTRODUCTION.....	1
1.1. Chapter Introduction.....	1
1.2. Background and Context	1
1.2.1. Interplay of Innovation, Dynamic Capabilities, and Technological Adoption ...	1
1.2.2. Innovating New Product Development through Dynamic Capabilities.....	2
1.2.3. The Role of Dynamic Capabilities and Innovation on Firm Performance.....	2
1.2.4. AI, Data Science, and NPD in Portugal	2
1.3. Problem Statement.....	3
1.4. Research Question	3
1.5. Conceptual Framework.....	3
1.6. Methodological Approach	4
1.7. Structure.....	4
1.8. Chapter Summary	5
2. LITERATURE REVIEW.....	7
2.1. Chapter Introduction.....	7
2.2. Theoretical Foundations of Innovation and Dynamic Capabilities.....	7
2.2.1. The Commercial Value of Innovation.....	7
2.2.2. The Nature of Dynamic Capabilities.....	7
2.2.3. The Interplay between Dynamic Capabilities and Innovation	8

2.2.4.	The Role of Innovation in Firm Performance and Competitiveness.....	8
2.3.	Data Science and AI: A New Paradigm for Innovation	8
2.3.1.	Understanding Data Science: Concepts and Techniques	8
2.3.2.	Exploring Artificial Intelligence: Capabilities and Constraints	9
2.3.3.	AI and Data Science as Catalysts for Dynamic Capabilities and Innovation	10
2.4.	Advancing Innovation through Analytic Capabilities	12
2.4.1.	The Potential of Innovation Analytics in the Innovation Process.....	12
2.4.2.	The Transformative Impact of AI and Data Science on Dynamic Capabilities and Firm Performance.....	12
2.5.	The Role of AI and Data Science in New Product Development	13
2.5.1.	An Overview of the Stage-Gate NPD Process and Its Key Stages	13
2.5.2.	Organizational Structures: Influence and Impact on NPD.....	15
2.5.3.	Stage-Gate NPD: From Functional Structures to AI-Enabled Practices.....	16
2.5.4.	Unified Approach: AI, Data Science, Firm Structure, and NPD	17
Management		17
Research and Development.....		18
Marketing and Sales		18
Operations		19
Finance		19
2.6.	Understanding the Adoption of AI and Data Science	21
2.6.1.	Industry Adoption and Application of AI and Data Science Technologies.....	21
2.6.2.	AI and Data Science in the Portuguese Market.....	22
2.7.	Barriers and Strategies for AI and Data Science Implementation.....	22
2.7.1.	Key Factors Influencing the Adoption of AI and Data Science in NPD.....	22
2.7.2.	Overcoming Obstacles in AI and Data Science Implementation.....	23
2.8.	Chapter Summary	23
3.	METHODOLOGY AND DATA COLLECTION.....	25

3.1.	Chapter Introduction.....	25
3.2.	Development of Conceptual Framework.....	25
3.2.1.	Evolving the Stage-Gate Framework: Assumptions and Developments	26
3.2.2.	Developing the Conceptual framework: Harmonizing an Established Model with Innovative Technologies	27
3.2.3.	AI and Data Science Enhanced Roles Across Departments: A Conceptual Framework.	30
3.3.	Hypothesis	31
3.3.1.	Hypothesis for the Survey (H1)	32
3.3.2.	Hypothesis for the Semi-structured Interviews (H2)	32
3.4.	Research Design	32
3.4.1.	Preliminary Data Collection.....	33
Data Collection Instruments.....	33	
Pilot Study.....	33	
Sampling Strategy	34	
Sample Description	34	
Pre-Assessment Survey (Quantitative Data Collection)	34	
Semi-Structured Interviews (Qualitative Data Collection)	35	
Ethical Considerations.....	37	
Potential Limitations and Biases	37	
3.4.2.	Practical Application	37
Web Scrapping for Data Mining	38	
Sentiment Analysis.....	38	
AI Product Generation	38	
Automated marketing promotion	39	
Ethical Considerations.....	39	
Potential Limitations and Biases	39	
3.1.	Chapter Summary	39

4.	ANALYSIS AND RESULTS	41
4.1.	Chapter Introduction.....	41
4.2.	Data Presentation.....	41
4.2.1.	Survey Data: Respondents	41
4.2.2.	Survey Data: Company Profiles.....	42
4.2.3.	Interview Data: Thematic Insights	42
4.3.	Statistical Analysis	43
4.3.1.	Descriptive Statistics	43
	Survey Data: Respondents' Statistics	43
	Survey Data: Company Profiles' Statistics.....	45
	Interview Data: Thematic Statistics	45
4.3.2.	Inferential Statistics.....	47
	Fisher's Exact Test.....	47
	Correlation Analysis.....	47
	Kruskal-Wallis Test.....	48
4.3.3.	Multivariate Statistics.....	49
	Cluster Analysis	49
4.4.	Results of Statistical Analysis	51
4.4.1.	Hypothesis H1 – AI Integration & Innovation Rates in Portuguese NPD.....	51
	H1 Evidence from the Results.....	51
	Conclusion for H1	51
4.4.2.	Hypothesis H2 - AI in NPD: Sectoral Variance.....	51
	H2 Evidence from the Results.....	52
	Conclusion for H2	52
4.1.	Survey Evaluation and Robustness Checks.....	52
4.1.1.	Survey Evaluation	52
4.1.2.	Robustness Check	53

4.2.	Applied Analysis	54
4.2.1.	Web Scraping for Data Mining	54
4.2.2.	Sentiment Analysis.....	55
4.2.3.	AI Product Generation	56
4.2.4.	Automated Marketing Promotion.....	58
4.3.	Results of Applied Analysis	59
	Applied Analysis - H1 Evidence from the Results	59
	Applied Analysis - Conclusion for H1	59
4.4.	Discussion of Results.....	59
4.5.	Chapter Summary	60
5.	CONCLUSION	61
5.1.	Preliminary Considerations	61
5.2.	Insights from Literature Review	62
5.3.	Methodological Framework and Data Acquisition	62
5.4.	Interpretation of Findings	63
5.1.	Limitations and Future Research	64
	BIBLIOGRAPHY	1
	APPENDIX I - Data Collection	A
	APPENDIX II - Datasets.....	K
	APPENDIX III - Data Analysis	U

LIST OF FIGURES

Figure 1: Key disciplines of the data scientist (Jones, 2013).....	8
Figure 2: Author’s interaction with ChatGPT (Model: 4, Version: May 3rd) in 17/05/2023..	10
Figure 3: Organizations have increased their AI capabilities, from an average of 1.9 capabilities per firm in 2018 to 3.8 capabilities in 2022 (McKinsey, 2022).....	11
Figure 4: AI is likely to have the biggest impact in knowledge work, especially decision-making and collaboration, areas previously considered less automatable. (McKinsey, 2023).....	11
Figure 5: The Agile-Stage-Gate model with Agile project management built into the stages (Cooper, 2021).	13
Figure 6: Example of Amazon’s functional structure in the 2000s (Jones, 2013).....	15
Figure 7: AI Recommender Output for Marketing Initiatives outlining KPIs (BCG, 2023)...	19
Figure 8: Ai adoption within functional departments (McKinsey, 2020).....	20
Figure 9: AI leaders worldwide and their GDP ties across regions from 2009-2018 (European Commission, 2018)	22
Figure 10: Proposed Conceptual Framework.....	25
Figure 11:1) Cooper's Stage-Gate model forms the conceptual framework's foundation. 2) Framework emphasizes core departments to enhance NPD in diverse organizations.	28
Figure 12:3) AI and Data Science as means on improved automation and decision-making. 4) Framework embeds AI and Data Science within departments, boosting performance.....	29
Figure 13:5) Functional departments, enhanced by AI and Data Science, integrated into NPD. 6) Final framework enhances NPD through automation, data-driven decisions, adaptability, and experimentation.	29
Figure 14: Word Cloud of the Semi Structured Interviews	36
Figure 15: Strong Correlations Found Between AI Interest, Familiarity, and Product Adoption.	48
Figure 16: KModes Algorithm Groups Data: Elbow method (left) identified 4 Clusters, and the t-SNE technique for 2D representation (right).....	49

Figure 17: Most Influential Features for Clustering.....	49
Figure 18: Cluster Rankings Reflect Varied AI and DS Adoption Across Groups.	50
Figure 19: Survey Results: 40% Perfect Score, 7% Lowest; AI Ratings Vary.....	53
Figure 20: Good Highlights: Product Quality Emphasizes Durability and Reliability Trends (ChatGPT).	55
Figure 21: Downsides: quality control issues and misleading information in products (ChatGPT).	56
Figure 22: DALL·E 2 transforms ChatGPT's good feedback into a sleek, user-friendly water bottle.....	57
Figure 23: DALL·E 2 envisions a flawed bottle, influenced by bad descriptors.....	57
Figure 24: ChatGPT and DALL·E unite, crafting eco-friendly Instagram post for MAMEIDO.	58
Figure 25: Survey Guide	A
Figure 26: Interview Guide	J
Figure 27: Complete Fisher’s Exact Test Code in R Language.....	U
Figure 28: Complete Correlations Code in Python Language	W
Figure 29: Complete Kruskal Wallis Code in Python Language	Y
Figure 30: Complete Clustering Code in Python Language	BB

LIST OF TABLES

Table 1: Senior managers favour Agile-Stage-Gate over traditional methods for innovation projects due to improved flexibility and efficiency (Edwards et al., 2019).....	15
Table 2: Survey Dataset - Exploring Demographics, Professions, and AI Familiarity	41
Table 3: Survey Dataset - Exploring Companies' Characteristics and AI Adoption Trends ...	42
Table 4: Dataset of Interviews - Revealing Strategies, Challenges, and AI/Data Science Perspectives.....	43
Table 5: Respondents Snapshot: A Balanced Gender Mix, Dominance of 25-34 Age Group, and Popular Roles in Marketing and AI/Data Science	44
Table 6: Companies' Profiles: Varied Ages, Sizes, Industries; Mixed AI Adoption & Disclosures.	44
Table 7: Corporate Strategies: Data-Driven Shift, AI Adoption, and Innovation Emphasized.	46
Table 8: Fisher's Test Unveils Data Relationships: Purple for Significant Links, Yellow for Important Insights.	47
Table 9: Kruskal-Wallis Test Reveals No Significant Median Differences in AI Interests. ...	48
Table 10: Chatbot Survey: High Quality Ratings, Strong AI User Satisfaction.....	52
Table 11: Amazon Review Analysis: Octoparse Scraping, ChatGPT Analysis, UK Dominance, September Peak.....	54
Table 12: Complete Survey Dataset.....	K
Table 13: Complete Interview Dataset.....	L
Table 14: Complete Survey Quality Dataset.....	T

LIST OF ABBREVIATIONS

Abbreviation	Meaning
A.I.	Artificial Intelligence
ICT	communication technology
BDA	Big Data Analytics
I.T.	Information Technologies
R&D	Research and Development
NPD	New Product Development
R&D	Research and Development

1. INTRODUCTION

1.1. Chapter Introduction

This dissertation explores how 'Innovation Analytics' impacts the New Product Development (NPD) in Portuguese firms. 'Innovation Analytics' as defined by Kakatkar et al. — the amalgamation of AI, Data Science, and innovation process (Kakatkar et al., 2020) — impacts firms. As AI usage doubled from 1.9 capabilities per company in 2018 to 3.8 capabilities per company in 2022 (McKinsey, 2022), hurdles still exist (European Commission, 2018).

The ongoing digital transformation, seen worldwide, has profound implications for industries globally (European Commission, 2018). In terms of deploying artificial intelligence for strategic decision-making, Portuguese companies encounter barriers, as it remains a relatively novel area of exploration in their context (Gomes, 2022).

This research probes the capabilities and barriers associated with the deployment of Innovation Analytics within the Portuguese market. Its primary aim is to offer insights that could potentially facilitate the enhancement of dynamic capabilities and organizational performance through the efficacious implementation of AI and Data Science in NPD.

1.2. Background and Context

1.2.1. *Interplay of Innovation, Dynamic Capabilities, and Technological Adoption*

Innovation, considered as “the successful combination of hardware, software, and orgware (the various components of the innovation system)” (Hekkert et al., 2007), helps firms to augment their product portfolio and stay competitive (Hoonsopon & Ruenrom, 2012), which, in turn, this competitive advantage aligns with the theory of dynamic capabilities.

As stated by Barreto (2010), dynamic capabilities aid firms to outperform competitors by systematically addressing challenges, capitalizing on opportunities, and making “timely, strategic, market-oriented decisions”. Firms with robust dynamic capabilities often foster innovation and excel in the market competition beyond traditional strategies (Alves et al., 2017).

The integration of Data Science and AI amplifies these dynamics, contributing to 'Innovation Analytics' (Kakatkar et al., 2020). According to Joshi et al. (2010) and Verganti et al. (2020), such advanced techniques further accelerate innovation by improving the speed and accuracy of knowledge retrieval and directing processes efficiently.

1.2.2. Innovating New Product Development through Dynamic Capabilities

According to (Bhuiyan, 2011) effective NPD, dedicated to the customer's voice, prioritizes gathering extensive ideas, primarily from customers, to facilitate the design and development of successful new products. This focus on clients' decisions is aligned with the NPD process and it is particularly important in today's rapidly changing consumer landscape (Chaochotechuang et al., 2015) allowing firms to make “timely, strategic, market-oriented decisions” (Barreto, 2010).

One important framework for NPD is Cooper's Stage-Gate model (Cooper, 1990, 2016, 2021) incorporating ideation, concept development, product design, testing, and product launching.

1.2.3. The Role of Dynamic Capabilities and Innovation on Firm Performance

The study by Alves et al. (2017) suggests that a firm's innovative performance is significantly influenced by its management capability within the dynamic capabilities framework. The ability of a firm to integrate, reconfigure, and evolve competencies in response to changing environments fosters innovation and subsequently, competitive advantage (Barreto, 2010; Teece, 2018).

Besides, competitive advantage derived from innovation, particularly product differentiation, is seen as an important element enhancing firm performance (Gunday et al., 2011; Porter, 1985).

This study thus explores the synergy between innovation, dynamic capabilities, and both data science and AI, in fostering competitive advantage and successful NPD within the Portuguese context.

1.2.4. AI, Data Science, and NPD in Portugal

According to (Gomes, 2022) the adoption of AI and Data Science in Portugal is still emerging. While, in his study (Gomes, 2022), only one firm reportedly used AI for strategic

decisions, the prevalence of operational IT usage was apparent. The increasing investments in computational resources signal promising growth potential (Almeida, 2022).

1.3. Problem Statement

Data Science and AI offer the promise of accelerating innovation by streamlining knowledge retrieval and process efficiency (Alves et al., 2017), potentially enhancing product portfolios (Hoonsopon & Ruenrom, 2012) and competitiveness through dynamic capabilities (Gunday et al., 2011; Porter, 1985).

However, Gomes (2022) highlights the limited incorporation of these technologies in Portugal's industries, potentially hindering effective decision-making central on successful NPD.

This discordance with the imperative of precise NPD decision-making raises questions about the attainment of complete dynamic capabilities. Thus, this study seeks to unravel the paradox of untapped AI and Data Science potential for 'Innovation Analytics' as defined by Kakatkar et al. (2020) in Portugal's NPD landscape. It investigates the impact of AI and Data Science on NPD, aiming to uncover integration barriers and reveal their latent potential in fostering sustainable innovation.

1.4. Research Question

This dissertation aims to illustrate the current state, successful instances, and challenges associated with the utilization of AI and Data Science in Portugal's New Product Development process.

1.5. Conceptual Framework

This study proposes a conceptual framework that integrates AI, Data Science, and functional organizational structures into Cooper's Stage-Gate model (Cooper, 1990, 2016, 2021). Furthermore, this proposed conceptual framework embodies principles of innovation, dynamic capabilities, and technological integration aiming to leverage the transformative potential of AI and Data Science within NPD.

1.6. Methodological Approach

To verify if the actions proposed in the conceptual framework are already being adopted by Portuguese companies or, if not, to understand the barriers in place, this study uses a survey and semi-structured interviews. Additionally, to test the practicality of the proposed framework, it is also employed a practical application. This triangulated approach (surveys, semi-structured interviews, and practical application) offers light to both quantitative and qualitative data and provides a controlled environment for experimentation.

1.7. Structure

The dissertation's structure unfolds as follows:

Chapter 1: INTRODUCTION offers an overview, delineates the problem statement, and enumerates research objectives.

Chapter 2: LITERATURE REVIEW analyses existing literature, exploring the fusion of "Innovation," "ICT," and "Dynamic Capabilities" in NPD phases within the "Innovation Analytics" domain. The section highlights case studies exemplifying innovation analytics, scrutinizing AI and Data Science's transformative influence on dynamic capabilities and firm performance, their role in new product development, and their industry-specific adoption.

Chapter 3: METHODOLOGY AND DATA COLLECTION devises a conceptual framework aligned with the central research question, probing AI and Data Science integration in diverse Portuguese sectors to drive sustainable innovation in NPD. It outlines the comprehensive research model, expounding on the methodology encompassing initial data collection, analysis, and practical implementation. The chapter expands on the theoretical constructs shaping the conceptual framework and delineates the derived hypotheses.

Chapter 4: ANALYSIS AND RESULTS examines hypotheses stemming from the methodology, presenting an in-depth analysis of semi-structured interview data, evaluating an AI-facilitated "Consumer Insights Report", and deliberating outcomes that affirm or challenge the hypotheses.

Chapter 5: CONCLUSION AND FUTURE WORK encapsulates discussed themes, identifies dissertation contributions and limitations, and charts prospects for future research.

1.8. Chapter Summary

This chapter highlighted the transformative role of 'Innovation Analytics' on firms, particularly within the emerging Portuguese market. The following study is aimed to unravel the potential and barriers of these technologies, their impact on dynamic capabilities and firm performance, as well as the factors influencing their integration in the NPD.

2. LITERATURE REVIEW

2.1. Chapter Introduction

In this literature review, key concepts of innovation and dynamic capabilities are explored, laying the foundation for the role of AI and Data Science as a driver of sustainable innovation within the Portuguese market. Furthermore, factors affecting the adoption of these technologies in the NPD process are discussed, setting the stage for the proposed conceptual framework.

2.2. Theoretical Foundations of Innovation and Dynamic Capabilities

This section elucidates key concepts including the commercial value of innovation, the essence of dynamic capabilities, their interplay, and their collective impact on firm performance and competitiveness.

2.2.1. The Commercial Value of Innovation

Innovation can be defined as "the process of translating an idea or invention into a good or service that creates value or for which customers will pay" (Schumpeter, 1934). Open innovation, on the other hand, refers to "a distributed innovation process based on purposively managed knowledge flows across organizational boundaries" (Chesbrough, 2006).

2.2.2. The Nature of Dynamic Capabilities

Dynamic capabilities refer to a firm's proficiency in integrating, reconfiguring, and developing competencies to tackle evolving environments, enabling adaptation, innovation, and competitive advantage (Barreto, 2010; Eisenhardt & Martin, 2000; Teece et al., 1997). These capabilities, instrumental in the digital transformation era, drive competitiveness (Teece, 2018). Important firms like Apple and Google exemplify this through their evolving endeavours (McGrath, 2013). McKinsey's (2017) report substantiates that prioritizing such capabilities aligns with high growth and success. This view also aligns with a Boston Consulting Group's (2018) finding, which affirms outperformance by innovative firms in growth, profitability, and shareholder returns.

2.2.3. The Interplay between Dynamic Capabilities and Innovation

Dynamic capabilities drive firm's innovation, allowing it to “integrate, build, and reconfigure competences” (Teece et al., 1997). Also, these capabilities enable firms to effectively sense and seize opportunities, adapt to changing environments (Helfat & Winter, 2011), and transition successfully, as IBM moved from hardware to IT services (Pisano, 2017). Ultimately, dynamic capabilities positively impact a firm's innovation, leading to competitive advantage and long-term success (Zahra et al., 2006).

2.2.4. The Role of Innovation in Firm Performance and Competitiveness

Innovation relates with firm performance, generating unique value (Tidd & Bessant, 2009). This link, substantiated by numerous studies, influences sales growth, profitability, and market value (Hult et al., 2004; Rosenbusch et al., 2011). Furthermore, actively adopting open innovation also enhances innovation performance (Laursen & Salter, 2006). PwC’s (2018) report noted that the top Global Innovation 1000 firms surpassed peers in revenue growth and shareholder returns.

2.3. Data Science and AI: A New Paradigm for Innovation

This study combines: stage-gate theory for NPD, AI and Data Science technologies, and firm’s functional organization, mapping their interplay and contact points.

2.3.1. Understanding Data Science: Concepts and Techniques

Data Science can drive knowledge creation from data by offering analytical tools for comprehending phenomena (Provost & Fawcett, 2013). It involves coding, data analysis, and storytelling to aid decision-making (Davenport & Patil, 2012). Applied to quantitative and/or qualitative methods, Data Science aims to predict outcomes and optimize solutions (Waller & Fawcett, 2013).

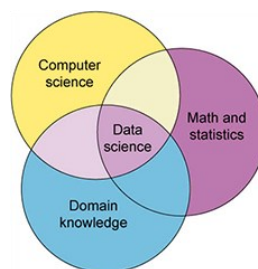


Figure 1: Key disciplines of the data scientist (Jones, 2013)

Efficient data scientists (Figure 1, p 8) merge analytical skills with domain knowledge to translate data into business insights (Waller & Fawcett, 2013). Their relevance spread through multiple distinct departments like marketing and finance, fostering data-driven decisions (Provost & Fawcett, 2013).

In summary, data science provides a methodology for leveraging knowledge from data (Provost & Fawcett, 2013).

2.3.2. Exploring Artificial Intelligence: Capabilities and Constraints


Artificial Intelligence (AI) is a computer science subfield, building intelligent machines that emulate human behaviour, often surpassing biological methods (McCarthy, 2007 as cited in (Manning, 2020). Intelligence is defined as “goal achievement computational power” (Mccarthy, 2007).

AI improves data science with advanced algorithms for machine learning and task execution (Provost & Fawcett, 2013). While data science fosters knowledge creation, AI automates tasks such as speech recognition, natural language understanding, and computer vision (Mccarthy, 2007).

ChatGPT, one of several AIs, exemplifies the use of data science principles in its training and operations, while its AI demonstrates the ability to comprehend and generate text (Brandl & Ellis, 2023; Nerdynav, 2023). Nonetheless, at the time of this series, ChatGPT reveals AI's limitations (Figure 2, p 10), such as lack of real-time internet data access, event knowledge limited until 2021, and occasional nonsensical responses (Brandl & Ellis, 2023; Nerdynav, 2023).

SU find 5 published scientific papers related to: data science or big data or artificial intelligence and innovation. The papers you find must always mention innovation. Provide the url for the papers



 As an AI developed by OpenAI, I can't browse the internet in real-time to provide URLs of specific papers. However, I can simulate the response based on my understanding and suggest the type of papers that might exist on these topics. Here are hypothetical examples of such papers:

1. "Innovation in Data Science: A Comprehensive Review of Emerging Techniques and Technologies"
2. "The Role of Big Data in Fostering Innovation in Business Processes"
3. "Artificial Intelligence and Innovation: An Interdisciplinary Perspective"
4. "Impact of Big Data Analytics on Innovation: Evidence from Start-Up Ecosystem"
5. "Innovation Through Artificial Intelligence: A Data-Driven Approach to Business"

Remember, these are simulated examples and may not correspond to real papers. You would need to search for these or similar topics in the databases you've mentioned to find actual papers.



Figure 2: Author's interaction with ChatGPT (Model: 4, Version: May 3rd) in 17/05/2023

Despite, AI's understanding, interpretation, and data-deriving capacity it still demands data scientists' expertise (Waller & Fawcett, 2013). AI's complex problem-solving relies on data scientists, as experts equipped with domain knowledge, to oversee and improve results (Mccarthy, 2007).

2.3.3. *AI and Data Science as Catalysts for Dynamic Capabilities and Innovation*

Teece (2007) describes dynamic capabilities as enabling firms to perform exceptionally through sensing, seizing, and transforming abilities. BDA, encompassing AI and data science, supports these capabilities by offering actionable insights (Mikalef et al., 2018; Wiener et al., 2020). Evidence shows BDA positively influences dynamic capabilities, which are essential in fast-changing contexts (Mikalef et al., 2020; Wamba et al., 2017).

These technologies stimulate innovation and enable dynamic capabilities (Lehrer et al., 2018; Dremel et al., 2020 as cited in (Mikalef et al., 2021). BDA's management, people, and technological dimensions collectively enhance firm performance (Davenport, 2018; Davenport & Patil, 2012); Barton & Court, 2012 as cited in (Wamba et al., 2017).

Research underlines BDA's impact (Figure 3, p 11) on firm performance and dynamic capabilities, emphasizing its role in business outcomes (Wamba et al., 2017). BDA's value (Figure 4 p 11) extends beyond IT investment, influencing organizational capabilities and decision-making (Gao & Sarwar, 2022).

Responses show an increasing number of AI capabilities embedded in organizations over the past five years.

Average number of AI capabilities that respondents' organizations have embedded within at least one function or business unit¹

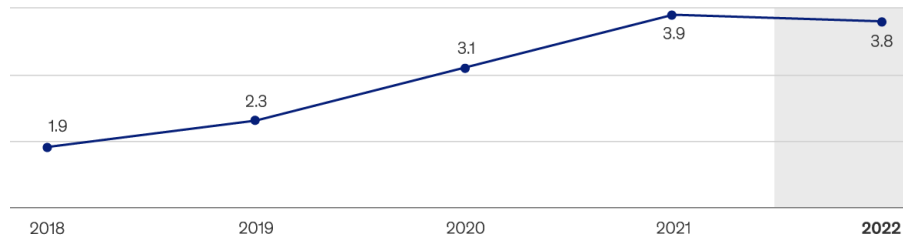


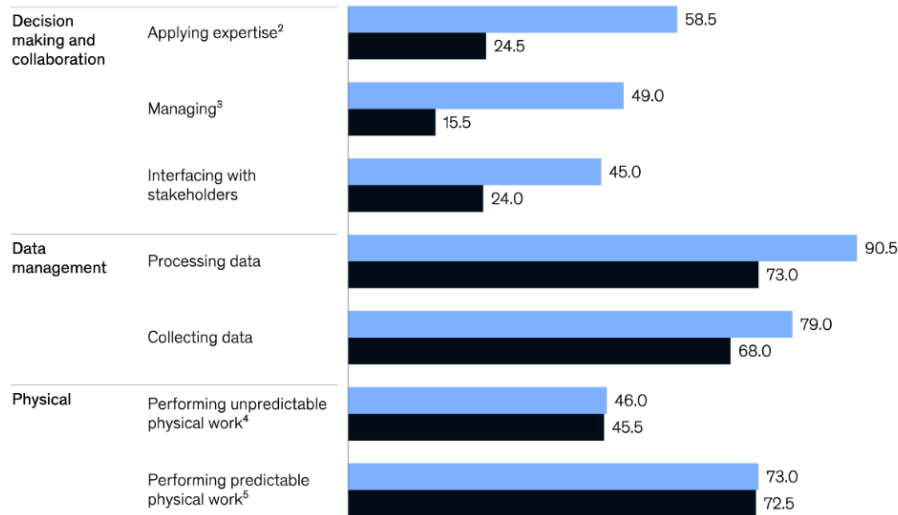
Figure 3: Organizations have increased their AI capabilities, from an average of 1.9 capabilities per firm in 2018 to 3.8 capabilities in 2022 (McKinsey, 2022)

Generative AI could have the biggest impact on collaboration and the application of expertise, activities that previously had a lower potential for automation.

Overall technical automation potential, comparison in midpoint scenarios, % in 2023

■ With generative AI
 ■ Without generative AI¹

Activity groups



Note: Figures may not sum, because of rounding.
¹Previous assessment of work automation before the rise of generative AI.
²Applying expertise to decision making, planning, and creative tasks.
³Managing and developing people.
⁴Performing physical activities and operating machinery in unpredictable environments.
⁵Performing physical activities and operating machinery in predictable environments.
 Source: McKinsey Global Institute analysis

Figure 4: AI is likely to have the biggest impact in knowledge work, especially decision-making and collaboration, areas previously considered less automatable. (McKinsey, 2023)

AI enriches big data's capabilities, providing situational understanding and problem-solving (Cigref, 2018; S. Gupta, Kar, Baabdullah, & Al-Khowaiter, 2018, as cited in (Tiguint & Hossari, 2020). However, efficient integration and development of these technologies require skilled management and data scientists (Tiguint & Hossari, 2020).

2.4. Advancing Innovation through Analytic Capabilities

Innovation Analytics involves the use of computer-enabled, data-derived insights within the innovation process (Kakatkar et al., 2020). It employs qualitative and quantitative methods to provide insights, identify trends, and assess innovation initiatives (Kakatkar et al., 2020). This technique allows companies to evaluate progress, identify areas for improvement, and make data-informed decisions enhancing innovation capabilities (Pisano, 2015). It assists in better resource allocation and project prioritization, improving the chances of successful innovation (Pisano, 2015). Consequently, companies can comprehend factors influencing innovation performance and devise strategies promoting an innovation culture (Hidalgo & Albors, 2008).

2.4.1. The Potential of Innovation Analytics in the Innovation Process

Data science and AI may enhance dynamic capabilities by aiding organizations to sense, seize, and transform opportunities (Conboy et al., 2020; McAfee & Brynjolfsson, 2012; Mikalef et al., 2019a, 2020). Through comprehensive data analysis (Grossi et al., 2021), companies understand customer needs and market dynamics better, thus increasing performance (Bhatti et al., 2022; Mikalef et al., 2021). Incorporating data science (Ransbotham, 2017) and AI (Mikalef et al., 2019b) into operations may foster innovation, thereby escalating competitiveness and ensuring longevity (De Luca et al., 2021; Ghasemaghahi & Calic, 2020; Suoniemi et al., 2020). AI-powered tools aid in collaboration and knowledge sharing, amplifying innovation processes' effectiveness (Füller et al., 2022). Furthermore, data science and AI support companies in informed decision-making, optimizing resource allocation, and bolstering their innovation ecosystem (Davenport & Patil, 2012; Davenport et al., 2018; Joloudari et al., 2022; Morariu et al., 2020).

2.4.2. The Transformative Impact of AI and Data Science on Dynamic Capabilities and Firm Performance

Intelligent mechanisms may reveal themselves essential for effective knowledge retrieval and structuring, thus enabling organizations to leverage their intellectual assets and drive innovation (Marina, 2007; Zia & Shafiq, 2016). Tools such as machine learning, natural language processing, and semantic technologies expedite the extraction and analysis of relevant knowledge (Dessi et al., 2021; Gambhir & Gupta, 2017; Khurana et al., 2023). This enhances knowledge management, fosters collaboration, and eases knowledge sharing (Alavi & Leidner,

2001; Jarrahi et al., 2023). Improved decision-making, efficiency in innovation processes, and better firm performance are outcomes of effective knowledge structuring (Anshari et al., 2023; Liao et al., 2010; Wang & Wang, 2012). With the rapid growth of available information, adopting these mechanisms is important for organizations to remain competitive (Deloitte, 2018; UNESCO, 2021).

2.5. The Role of AI and Data Science in New Product Development

By exploring the transformative impact of AI and Data Science on NPD, this study investigates how these technologies advance traditional Stage-Gate NPD models (Cooper, 1990, 2006, 2008, 2011, 2014, 2016, 2019, 2021). Furthermore, it aims to reimagine product development within corporate structures and highlights the potential of AI and Data Science to address Stage-Gate models' limitations.

2.5.1. An Overview of the Stage-Gate NPD Process and Its Key Stages

Robert G. Cooper, a key author in NPD, developed the Stage-Gate process, a sequential model for managing new products (Cooper, 1990, 2006, 2008, 2011, 2014, 2016, 2019, 2021).

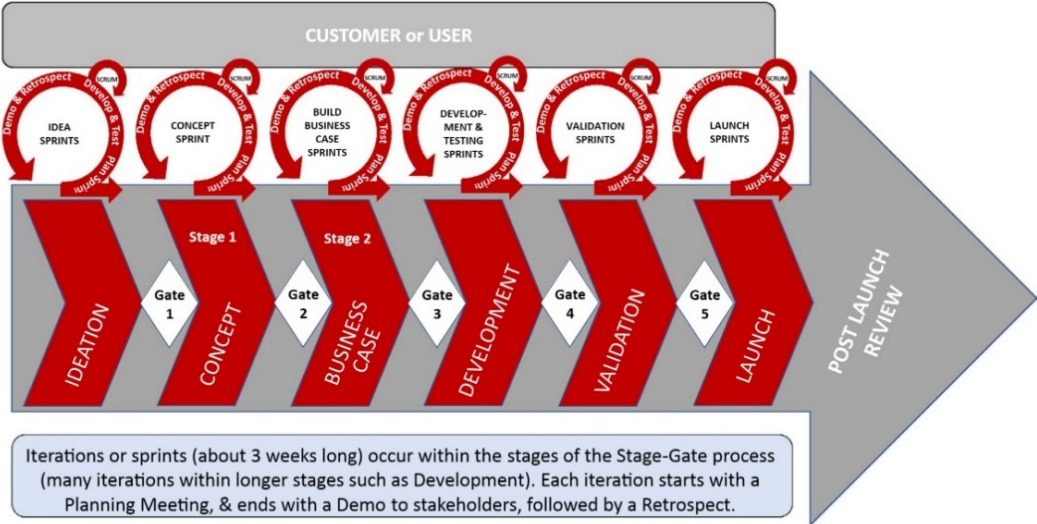


Figure 5: The Agile-Stage-Gate model with Agile project management built into the stages (Cooper, 2021).

Its model involves five stages (Ideation, Concept, Business Case, Development, Validation, Launch) with decision points to evaluate new product ideas (Cooper, 1990, 2006, 2021; Eppinger & Ulrich, 2011; Kahn, 2006):

- Idea Generation: This phase is the birthplace of the new product concept, serving as the process's inception, leading to Gate 1 (Cooper, 1990).

- Gate 1, "Initial Screen" and Stage 1, "Preliminary Assessment": Initial project commitment is evaluated based on strategic fit, feasibility, magnitude, and market attractiveness. Subsequent assessment focuses on technical and market viability (Cooper, 1990).
- Gate 2, "Second Screen" and Stage 2, "Definition": These steps require reassessment of the project utilizing insights from Stage 1. After approval, the project's appeal is solidified before committing to major investments (Cooper, 1990).
- Gate 3, "Decision on Business Case" and Stage 3, "Development": This phases, the last termination point before heavy costs, gives way to product creation, testing, and strategic planning (Cooper, 1990).
- Gate 4, "Post-Development Review" and Stage 4, "Validation": Gate 4 scrutinizes project progress and ongoing product attractiveness. Validation then assesses product, process, customer feedback, and economic feasibility (Cooper, 1990).
- Gate 5, "Pre-Commercialization Decision" and Stage 5, "Commercialization": The final termination point, Gate 5, precedes full commercialization. Stage 5 executes marketing and operational plans (Cooper, 1990).
- Post-Implementation Review: Post-implementation, the project is evaluated and concluded, becoming a standard component of the organization's portfolio (Cooper, (Cooper, 1990).

Conversely, the Stage-Gate models have faced criticism for inhibiting experimentation (Becker, 2006; Lenfle and Loch, 2010; Cooper, 2014; Sætre & Brun, 2013; Sethi and Iqbal, 2008; Salerno et al., 2015, as cited in (Cooper, 2014). In addition, critics urged a more flexible approach (Ajamian and Koen, 2002; Kreiner, 1995, as cited in (Cooper, 2014).

In response, Cooper (2016) proposed a hybrid Agile-Stage-Gate model (Figure 5, p 13), blending the structure of Stage-Gate with Agile's adaptability (Karlstrom and Runeson, 2005, 2006, as cited in (Cooper, 2016). This blending enhances team communication, customer engagement, developer motivation, and resource optimization (Cooper, 2013, as cited in (Cooper, 2016).

An evaluation (Table 1, p 15) by Edwards et al. (2019) acknowledged its effectiveness in SMEs, leading to faster market reach, better success rates, and improved product process.

Performance metric	Company A	Company B	Company C
Development time	-20%	-20%	-20%
Development cost	Same	+10%	Same
Overall innovation process	Better	Better	Same
Estimated probability of success	Higher	Higher	Higher

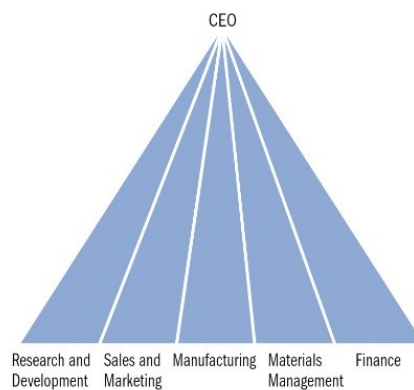
Table 1: Senior managers favour Agile-Stage-Gate over traditional methods for innovation projects due to improved flexibility and efficiency (Edwards et al., 2019)

Cooper (2021) suggested the inclusion of digital tools in the Agile-Stage-Gate system to match pandemic-driven innovation. He denotes that AI accelerates technical solution selection, evident in pharmaceuticals (Rotman, 2018; Mishcon & Robinson, 2019, as cited in (Cooper, 2021)). In addition, machine learning expedites product design, reducing time-to-market (Waltz, 2020, as cited in (Cooper, 2021)).

2.5.2. Organizational Structures: Influence and Impact on NPD

Organizations employ structures like functional, product, or matrix to accommodate growth and diversification (Ahmady et al., 2016; Duncan, 1979; Jones, 2013). A functional structure, grouping employees by skills, is a common choice (Duncan, 1979; Jones, 2013). This dissertation focuses on functional structures (Figure 6, p 15) defined by departmental specialization (Jones, 2013).

A. This format shows that each function has its own hierarchy



B. This format shows the position of each function within the organization's hierarchy

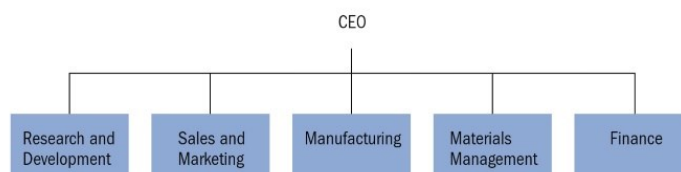


Figure 6: Example of Amazon's functional structure in the 2000s (Jones, 2013)

Functional structures enhance role efficiency and effectiveness, promoting clear communication, eliminating redundancy, and accelerating problem-solving, aligning with the sequential nature of the stage-gate NPD model (Jones, 2013):

a) **Management** oversees planning, organization, control, and leadership (Jones, 2013; Lumen Learning, 2018).

b) **R&D** focuses on innovation and competitive advantage (Jones, 2013; Lumen Learning, 2018).

c) **Operations** manage the supply chain and meet customer needs (Jones, 2013; Lumen Learning, 2018).

d) **Marketing/Sales** is responsible for product design, promotion, pricing strategy, and revenue generation (Jones, 2013; Lumen Learning, 2018).

e) **Finance** conducts resource planning, procurement, and decision-making (Jones, 2013; Lumen Learning, 2018).

Evidence is latent on key decision-making roles such as CEO for Management, COO for Operations, CFO for Finance, CRO for Marketing/Sales, and CTO for R&D (Jones, 2013).

2.5.3. Stage-Gate NPD: From Functional Structures to AI-Enabled Practices

By bringing together stage-gate theory (Cooper, 1990, 2016, 2021) and functional organizational structures (Ahmady et al., 2016; Duncan, 1979; Jones, 2013), can provide individual departments with critical roles:

a) **Management**: Steers the process towards strategic objectives (Jones, 2013; Lumen Learning, 2018). According to Cooper (1990, 2016, 2021) these actions are pivotal where resources are committed.

b) **R&D**: Handles the technical aspects of product development (Jones, 2013; Lumen Learning, 2018). According to Cooper (1990, 2016, 2021) these are tasks relevant in testing feasibility, product development, and patents (Cooper, 1990, 2016, 2021).

c) **Marketing/Sales**: They decipher market needs and consumer preferences (Jones, 2013; Lumen Learning, 2018). In harmony with Cooper (1990, 2016, 2021) this is relevant for carrying out market assessments, concept testing, market research, and implementing marketing plans (Cooper, 1990, 2016, 2021).

d) **Operations:** Controls the supply chain, including procurement and logistics (Jones, 2013; Lumen Learning, 2018). Cooper (1990, 2016, 2021) stresses its importance in executing production plans, debugging processes, and implementing operational plans.

e) **Finance:** Oversees strategic planning, funds procurement, debt management, and providing decision-making accounting data (Jones, 2013; Lumen Learning, 2018). Cooper's model (1990, 2016, 2021) assigns finance, from early calculations, to advanced analyses at later stages.

These roles are often typical but may fluctuate based on an organization's structure and strategy (Ahmady et al., 2016; Duncan, 1979; Jones, 2013).

2.5.4. Unified Approach: AI, Data Science, Firm Structure, and NPD

In the previous sections, it is highlighted the importance of Cooper's Stage-Gate process for NPD. Furthermore, it is also portrayed how functional structures enable workforce specialization (Duncan, 1979; Jones, 2013). Finally, it is also explained how Cooper's Stage-Gate can be put to practice in a functional organized firm, thus benefiting from this specialized workforce.

In this section, it is explained how AI and Data Science are being applied in functional structures. Furthermore, it is depicted how AI and Data Science practices are in line with Cooper's Stage-Gate model, which, according to the author should mitigate issues such as bureaucracy and excessive time consumption (Salerno et al., 2015; Becker, 2006; Lenfle and Loch, 2010; Sethi and Iqbal, 2008, as cited in (Cooper, 2016).

Management

Data Science influences managerial decision-making through predictive modelling and optimization, informing strategic planning and resource allocation (Coussement & Benoit, 2021; Power, 2016). Data-driven decision-making is becoming increasingly prevalent (McKinsey, 2020).

Artificial Intelligence complements data science by leveraging predictive analytics to forecast trends and improve decision accuracy (Araujo et al., 2020; Pomerol, 1997).

Within Stage-Gate NPD, these technologies aid in predicting market trends and discerning customer needs (McKinsey, 2020). Furthermore, AI assesses strategic alignment and

market attractiveness (Kabaivanov & Markovska, 2021). By anticipating market trends (Vachhani et al., 2020) and evaluating product feasibility (Mak & Pichika, 2019), predictive analytics promote informed decisions.

Research and Development

Data Science's statistical methods drive experimental analysis and product performance forecasts, enhancing R&D efficiency and risk assessment in product development (Kim & Lee, 2019; Sharma et al., 2020).

AI enriches product development by encouraging concept generation and supplementing human creativity, also accelerating commercialization (Damioli et al., 2021; Johnson et al., 2022; Luusua & Ylipulli, 2020). Machine learning, an AI component, enhances product development by evaluating quality and optimizing manufacturing processes (Johnson et al., 2022; D. B. Larson & Boland, 2019).

Within the Stage-Gate NPD model, machine learning may assess product quality, automate manufacturing, performance monitoring, and trend prediction (Johnson et al., 2022). Furthermore, AI aids in prototype creation (Johnson et al., 2022), while Data Science improves decision-making in R&D, cutting costs and time (Pacheco Gutierrez et al., 2023).

Marketing and Sales

Data science thrives for optimized sales and marketing operations, driving customer engagement using customer data for market segmentation and targeted marketing initiatives (Rosário et al., 2021; José Ramón Saura, 2020; Jose Ramon Saura et al., 2023).

AI harmonizes marketing (Figure 7, p 19) and sales data, influencing financial and risk management while revolutionizing market analysis and strategy implementation (Huang & Rust, 2021; Rosário et al., 2021). AI also refines customer data analysis, aiding prospect scoring and sales funnel optimization, and harnessing user-generated content for product development (Paschen et al., 2019).

When integrated with the Stage-Gate NPD process, data science aids in market segmentation (Huang & Rust, 2021; Rosário et al., 2021) and AI's sentiment analysis aid in real-time pricing adjustments (Huang & Rust, 2021). These technologies also enhance customer engagement, refine validation procedures, and guide strategic decision-making in future projects (Paschen et al., 2019; Rosário et al., 2021; Jose Ramon Saura et al., 2023).

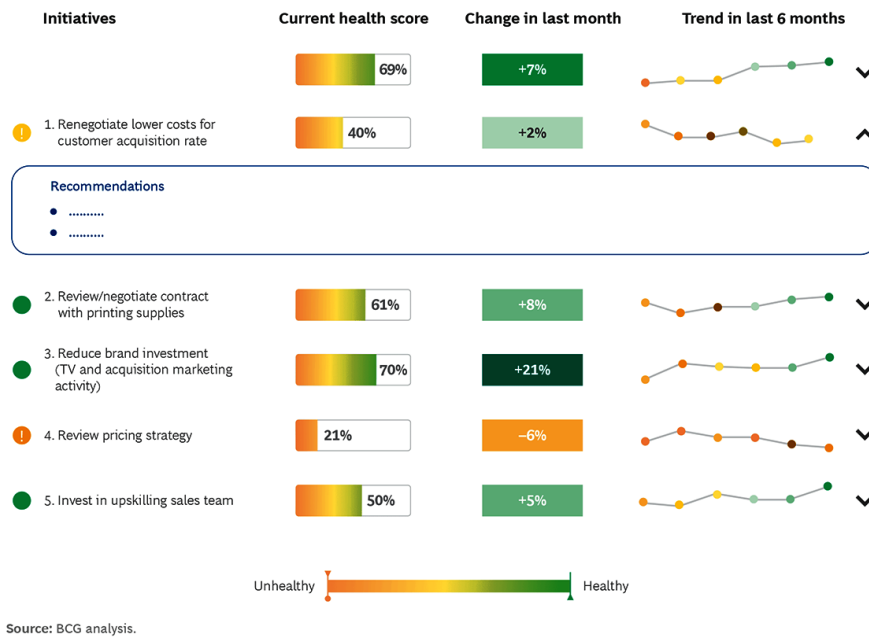


Figure 7: AI Recommender Output for Marketing Initiatives outlining KPIs (BCG, 2023)

Operations

Data Science increases operational efficiency through predictive modelling and statistical analysis (Chung et al., 2020; Russo et al., 2015). Data Science enables precise machinery downtime prediction, improved logistics, and enhanced production schedules (Chung et al., 2020; Lamba & Prakash, 2018).

AI increases operational efficiency by enhancing production, inventory control, and demand prediction, utilizing machine learning for trend analysis and cost-cutting (De Giovanni, 2019; Niranjana et al., 2021).

Data Science and AI in NPD can jointly optimize manufacturing efficacy and product quality (Lamba & Prakash, 2018). Furthermore, AI refines inventory and demand forecasting and enables high-quality validations via simulations (Huang & Rust, 2021; Niranjana et al., 2021).

Finance

Data science techniques facilitate strategic decisions in organizations by predicting customer behaviour, including churn rates (Leung et al., 2021; Pisoni et al., 2021).

AI refines financial services, improving analysis, and has significant roles from risk management to customer behaviour analysis, dynamic pricing models in e-commerce thus

enhancing efficiency and customer satisfaction (Cao et al., 2021; Fernandez, 2019; Hendershott et al., 2021; Yin & Han, 2021).

Data Science and AI together improve financial actions (Cao et al., 2021). Predictive modelling, machine learning data science in NPD aids feasibility assessment, provides insights, and forecasts financial performance (Cao et al., 2021; Hendershott et al., 2021; Leung et al., 2021; Mosavi et al., 2020). Furthermore, they enhance operational plans for successful launches (Cao et al., 2021).

AI adoption is highest within the product- or service-development and service-operations functions.

AI use cases most commonly adopted within each business function, % of respondents



Figure 8: Ai adoption within functional departments (McKinsey, 2020)

In conclusion, these technologies (Figure 8, p 20) improve market trend prediction, concept generation, customer engagement, and operational efficiency (Chung et al., 2020;

Johnson et al., 2022; Rosário et al., 2021). Furthermore, they aid in financial performance forecasting and risk management (Cao et al., 2021; Hendershott et al., 2021).

2.6. Understanding the Adoption of AI and Data Science

In the epoch of emerging technologies, understanding the adoption of AI and Data Science across industries becomes vital (McKinsey, 2018). This section provides an overview of AI and data science adoption in diverse sectors and explores specific impacts and challenges in the Portuguese market building upon the work of Gomes (2022).

2.6.1. *Industry Adoption and Application of AI and Data Science Technologies*

McKinsey's study (McKinsey, 2018) identified AI and data science's transformative role in industries via NPD processes, boosting efficiency and innovation.

Lemonade (insurance), for instance, eliminated human brokers in favour of an AI bot “Maya” to calculate policies and another AI bot "Jim" to streamline claims, thus enhancing customer service (Bort, 2016; Lemonade, 2017). Mayo Clinic (healthcare) utilized Google Cloud for AI-based image analysis to boost patient outcomes (Mayo Clinic, 2019). Both P&G (Consumer goods) and Walmart (Retail) applied AI in NPD processes to optimize various areas, including product development and supply chain (P&G, 2020, 2021; Walmart, 2016). Netflix (Entertainment) increased viewer engagement through AI personalized recommendations (Gomez-Uribe & Hunt, 2016; Verganti et al., 2020). Tesla (Automotive) and NVIDIA (Electronics) incorporated AI into NPD processes to develop advanced automotive features and autonomous vehicles (Bansal et al., 2018; NVIDIA, 2017, 2020; Verganti et al., 2020). Stitch Fix (Clothing) utilized AI and data science in NPD and sales expansion (Marr, 2018; Stitch Fix, 2018).

2.6.2. AI and Data Science in the Portuguese Market

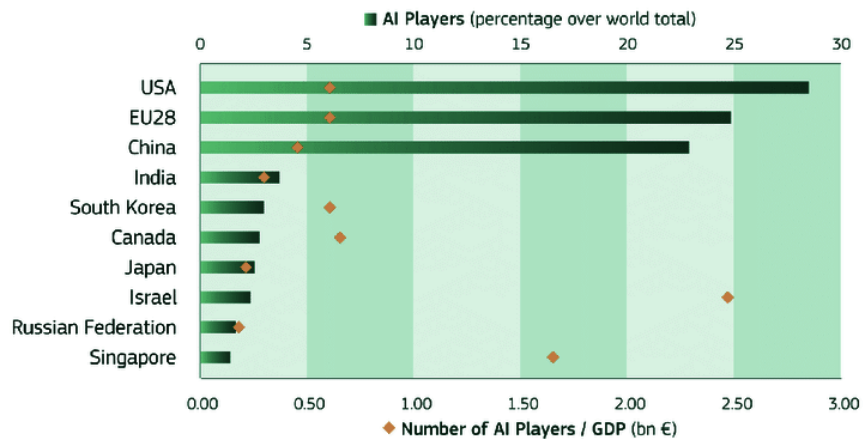


Figure 9: AI leaders worldwide and their GDP ties across regions from 2009-2018 (European Commission, 2018)

The European Commission (2018) emphasizes the need for region-specific AI and Data Science impact studies (Figure 9, p 22), such as in Portugal, for understanding diverse adoption contexts.

In the case of Portugal, Gomes (2022) reported only one firm using AI for strategic decision-making but prevalent operational use of I.T. along with human decision.

2.7. Barriers and Strategies for AI and Data Science Implementation

This section explores the intricate factors shaping the adoption of AI and Data Science in NPD, and subsequent strategies to overcome the identified barriers, underpinned by legal, ethical, economic, and data perspectives.

2.7.1. Key Factors Influencing the Adoption of AI and Data Science in NPD

Adopting AI and Data Science in NPD requires a comprehensive approach, encompassing ethical, legal, economic, cybersecurity, and data considerations (European Commission, 2018). Successfully leveraging big data from multiple sources necessitates transparent AI design and an understanding of data analytics (Del Vecchio et al., 2018; European Commission, 2018; Mikalef et al., 2019b).

Integration complexity increases in open innovation strategies due to the need to maintain data reliability and authenticity (Del Vecchio et al., 2018). Thus, possessing technically proficient human resources becomes essential (Gao & Sarwar, 2022).

Societal acceptance and clear regulatory frameworks are critical (European Commission, 2018). Managing privacy is critical to maintaining brand reputation and trust, as misuse of data mining can harm the image of the company (Del Vecchio et al., 2018).

Moreover, AI's implications on growth, employment, and wages, and the potential for power concentration among a few internet economy actors add to the economic concerns (European Commission, 2018). Also, aligning Big Data use with innovation strategies is especially demanding for large corporations (Del Vecchio et al., 2018).

2.7.2. Overcoming Obstacles in AI and Data Science Implementation

Ethical, legal, and educational challenges are major obstacles to AI implementation, requiring balanced regulation and enhanced learning strategies (European Commission, 2018). Regional economic impacts need monitoring to enable appropriate (European Commission, 2018). Addressing data volume and quality issues entails strategies for data acquisition, including partnerships and third-party data sets (McKinsey, 2018). Techniques like reinforcement learning reduce data labelling efforts, and bias can be addressed through representative training data (McKinsey, 2018).

Building data analytics capabilities combines recruitment, learning culture, and embedding big data decision-making (Mikalef et al., 2019a). Effective AI utilization requires innovation reframing and infrastructure development (Verganti et al., 2020).

Recognizing data scientists' role is key, and their abilities can be augmented via effective communication and external partnerships (Del Vecchio et al., 2018; Gao & Sarwar, 2022).

2.8. Chapter Summary

This literature chapter underscores the significant potential of AI and data science in driving innovation, particularly in the new product development process. It emphasizes the critical role of dynamic capabilities and the significant value of open and sustainable innovation. The barriers and strategies towards successful implementation in varying industries, particularly within the Portuguese market, are also highlighted.

3. METHODOLOGY AND DATA COLLECTION

3.1. Chapter Introduction

This chapter delineates the methodology adopted seeking to answer the central research question:

This dissertation aims to illustrate the current state, successful instances, and challenges associated with the utilization of AI and Data Science in Portugal's New Product Development process.

Expanding upon the literature chapter's findings, a conceptual framework is proposed. It clarifies AI, Data Science, and stage-gate NPD's interaction in functional organizations, suggesting a tech-oriented NPD sustainable process.

3.2. Development of Conceptual Framework

This dissertation proposes a conceptual framework (Figure 10, p 25), which incorporates AI and Data Science into an established stage-gate model, aiming to achieve a sustainable NPD process while considering firms' organizational layouts.

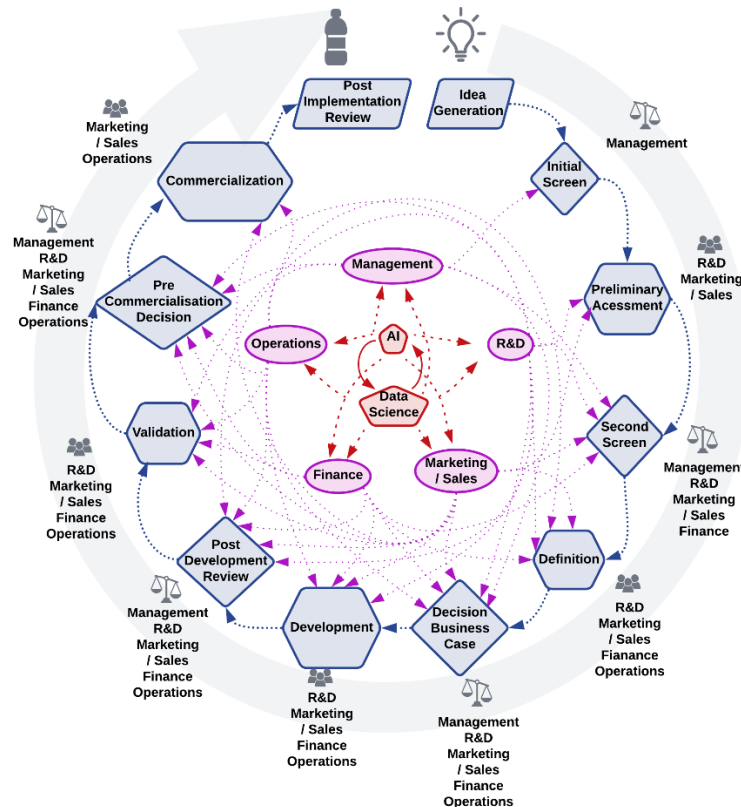


Figure 10: Proposed Conceptual Framework

As said, the foundational base of this conceptual framework is Cooper's Stage-Gate model (1990, 2016, 2021) to which modifications, informed by the literature chapter, are added; namely ICT technologies and company departments. The literature chapter underscored limitations of Cooper's Stage-Gate model (1990, 2016, 2021) such as bureaucracy and innovation blocking (Cooper, 2014). Hence the goal of this conceptual framework is to create a product development roadmap that harnesses the capabilities of AI and Data Science combined with specialized knowledge produced by functional departments.

3.2.1. Evolving the Stage-Gate Framework: Assumptions and Developments

This section outlines the goals of the proposed conceptual model, which builds upon Cooper's models (1990, 2016, 2021), while acknowledging certain assumptions and limitations inherent in this approach:

a) Limitations:

- 1) **Strategic Alignment (Assumption):** The integration of AI and Data Science into New Product Development (NPD) is proposed to address bottlenecks such as bureaucracy and time delays, as identified by Becker and Lenfle & Loch (2006; 2010; as cited in (Cooper, 2014)). This dissertation assumes that the capabilities offered by AI and Data Science can mitigate these issues, though further empirical validation is needed.
- 2) **Model Improvement (Assumption):** Incorporation of AI and Data Science is aimed at addressing criticisms of rigidity in the Stage-Gate model (Becker, 2006; Lenfle & Loch, 2010 ; as cited in (Cooper, 2014)). The assumption is that this integration can indeed enhance the model's adaptability, but empirical confirmation is necessary.
- 3) **Innovation Encouragement (Assumption):** The proposed structure seeks to promote new approaches to NPD, addressing innovation challenges raised by Lenfle & Loch (2010; as cited in (Cooper, 2014)). It is assumed that leveraging these technologies can stimulate innovation by streamlining processes and fostering a culture of experimentation, but this also requires further investigation.

b) New Goals:

- 1) **Data-Centricity:** Emphasizing data throughout NPD ensures informed decision-making via predictive modelling (Chung et al., 2020); Lenfle & Loch, 2010 as cited in (Cooper, 2014; Russo et al., 2015).

- 2) **Flexibility:** The framework seeks to align with various workflows, from conventional to Agile.
- 3) **Efficiency Enhancement:** AI assists in prototype development and trend forecasting (Johnson et al., 2022).

While this proposed conceptual framework draws heavily from Cooper's models, it recognizes that not every work in the field could be exhaustively considered in the period of this dissertation. Furthermore, this framework positions AI and Data Science not as simple additions to Cooper's models, but rather as essential instruments leveraged by a specialized workforce operating within various functional departments.

3.2.2. Developing the Conceptual framework: Harmonizing an Established Model with Innovative Technologies

This section outlines the steps involved in the conceptual framework's development:

- 1) **Selecting Cooper's Stage-Gate Model as the Foundation:** The conceptual framework's is based on Cooper's Stage-Gate model (1990, 2016, 2021). This conceptual model's structured and proven methodology forms the basis of the conceptual framework (Figure 11, p 28, 1).
- 2) **Recognizing Core Functional Departments:** The conceptual framework focuses on core functional departments common in various organizational structures. These include Management, R&D, Sales & Marketing, Finance, and Operations (Jones, 2013). By emphasizing these departments, the conceptual framework enhances its practical value by catering to a broad spectrum of organizational structures while adding specialized personnel (Jones, 2013) to carry out the diverse tasks of the NPD (Figure 11, p 28, 2)

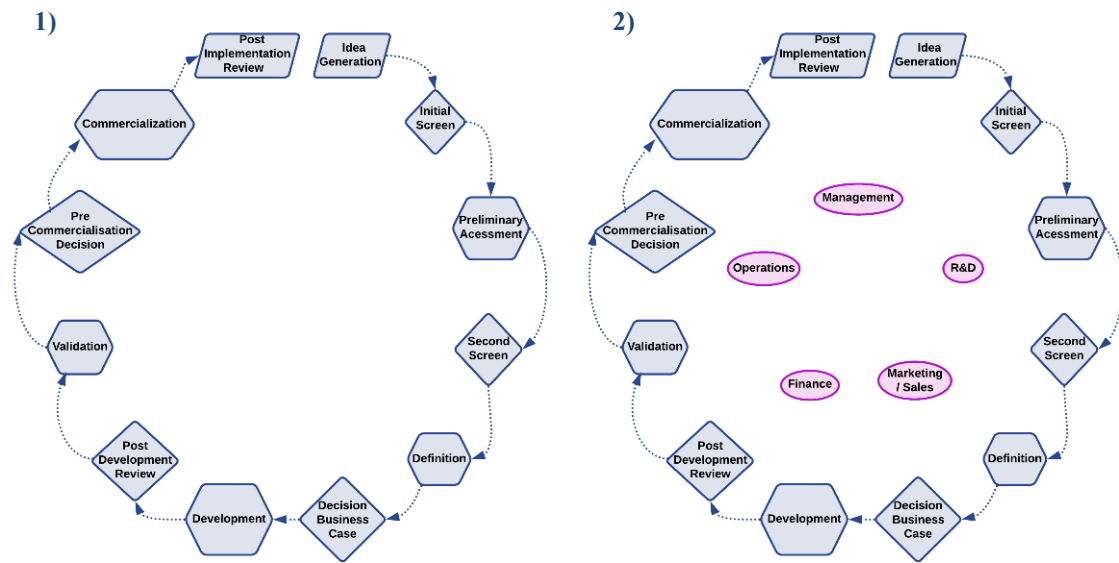


Figure 11:1) Cooper's Stage-Gate model forms the conceptual framework's foundation.
 2) Framework emphasizes core departments to enhance NPD in diverse organizations.

- 3) **Introducing AI and Data Science:** the literature review covered the benefits and potential drawbacks associated with these technologies. Incorporating these technologies, which interact and mutually leverage one another, enhances automation and decision-making capabilities (Tigunt & Hossari, 2020), while also influencing the NPD processes (Figure 12, p 29, 3).
- 4) **Fostering AI and Data Science in Functional Departments:** The conceptual framework converges technology and organizational departments by incorporating AI and Data Science within each functional department. This integration allows each department to benefit from the predictive capabilities and decision-making advantages of AI and Data Science, thereby enriching their performance (Cao et al., 2021; Hendershott et al., 2021; Johnson et al., 2022; McKinsey, 2020; Rosário et al., 2021) (Figure 12, p 29, 4).

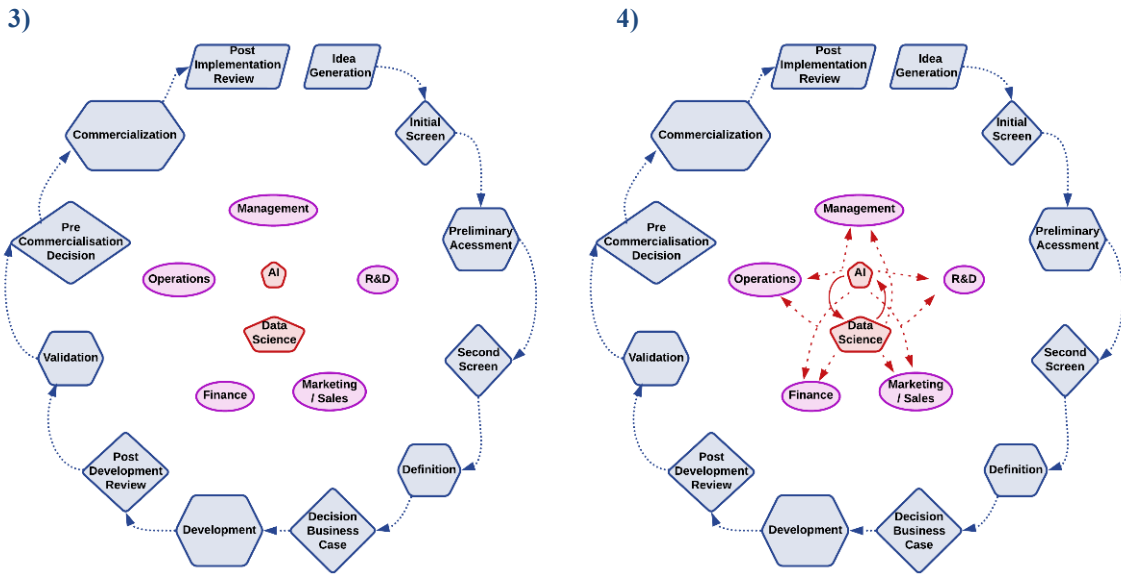


Figure 12:3) AI and Data Science as means on improved automation and decision-making.
 4) Framework embeds AI and Data Science within departments, boosting performance.

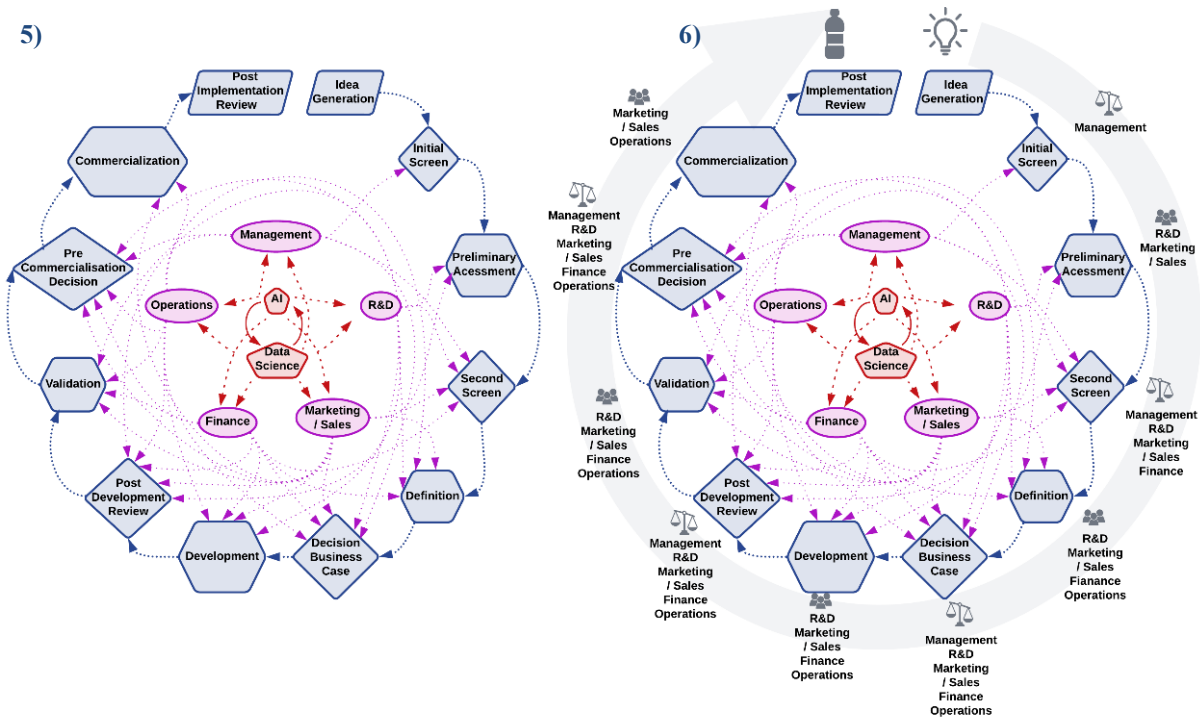


Figure 13:5) Functional departments, enhanced by AI and Data Science, integrated into NPD.
 6) Final framework enhances NPD through automation, data-driven decisions, adaptability, and experimentation.

- 5) **Integrating Functional Departments into the Stage-Gate NPD:** this concluding step incorporates the functional departments with their specialized personnel (Jones, 2013), now empowered with AI and Data Science, into the Stage-Gate model. This integration aims for alignment of the technology-augmented departments with each stage of the NPD process, thus forming a conceptual framework for NPD (Figure 13, p 29, 5).

The developed conceptual framework (Figure 13, p 29, 6) offers a tool for NPD, merging Cooper's structured approach with the power of AI and Data Science within functional departments. Again, the aim of this model is to provide a step-by-step guide to nurture sustainable innovation by promoting efficiency, enabling informed decision-making, and enhancing the NPD overall process.

3.2.3. AI and Data Science Enhanced Roles Across Departments: A Conceptual Framework.

This section examines the role of AI and Data Science in the distinct functional departments within the proposed conceptual framework.

Idea Generation: The NPD begins with a novel product concept (please consider the Practical Application in page 37 as a proof of concept to find new product ideas based on review insights) that can emerge from any department, and it's then forwarded to Gate 1, Initial Screening (Cooper, 1990).

Management: Integrating AI and Data Science into management practices aids in predictive modelling and strategic decision-making, informing strategic planning and resource allocation (Coussement & Benoit, 2021; Power, 2016). AI enhances these capabilities by employing predictive analytics to forecast trends and improve decision accuracy (Araujo et al., 2020; Pomerol, 1997). Thus, AI and data science insights enable better-informed decision-making, boosting innovation and efficiency across NPD stages 1 (First Screen) to 5 (Pre-Commercialization Decision).

Research and Development: Data science drives experimental analysis and product performance forecasts (Sharma et al., 2020; Kim & Lee, 2019), while AI supplements human creativity in concept generation and accelerates commercialization (Damioli et al., 2021; Johnson et al., 2022; Luusua & Ylipulli, 2020). These technologies combined expectedly will enhance R&D innovation, from gate 1 (Initial Screen) to 5 (Pre-Commercialization Decision).

Sales and Marketing: Here, data science optimizes operations by analysing customer data for market segmentation and targeted marketing initiatives (Rosário et al., 2021; José Ramón Saura, 2020; 2023). AI, in turn, revamps these data-driven marketing strategies, revolutionizing market analysis and strategy implementation (Huang & Rust, 2021; Rosário et al., 2021). It is expected that AI and data science synergy enhances sales, marketing, in all steps of the NPD.

Finance: Financial management can benefit from data science techniques that predict customer behaviour, including churn rates (Leung et al., 2021; Pisoni et al., 2021). AI refines these capabilities further, improving analysis and influencing areas from risk management to customer behaviour analysis, dynamic pricing models in e-commerce, enhancing efficiency, and customer satisfaction (Cao et al., 2021; Fernandez, 2019; Hendershott et al., 2021; Yin & Han, 2021) from gates 2 (Second Screen) to 5 (Pre-Commercialization Decision).

Operations: Data science increases efficiency through predictive modelling and statistical analysis (Chung et al., 2020; Russo et al., 2015). AI enhances these benefits, optimizing processes like production, inventory control, and demand prediction (De Giovanni, 2019; Niranjana et al., 2021). The operations department, empowered by AI and Data Science, can improve efficiency from stage 2 (Definition) to stage 5 (Commercialization).

Post-Implementation Review: After commercialization, the project ends, the team splits and the product joins the firm's regular lineup and lastly, sales performance is assessed (Cooper, 1990). Data Science can be used to access actual revenues, costs, and profits and then compare with previous forecasts (Pisoni et al., 2021). AI can be used in a post-audit to evaluate the project's pros and cons, scan product reviews and find ways to improve future projects (Cao et al., 2021) (please consider the Practical Application in page 37 as a proof of concept to scan product reviews). This review signifies the project's conclusion (Cooper, 1990).

3.3. Hypothesis

The literature review, focusing on the research question, "This dissertation aims to illustrate the current state, successful instances, and challenges associated with the use of AI and Data Science in Portugal's New Product Development process", identified potential enhancements to Cooper's Stage Gate that may arise by incorporating AI and Data Science benefits within functional organizational structures.

The development of this conceptual framework underscored the need to investigate AI

and Data Science's application in Portuguese NPD. Consequently, this framework inspired the formulation of two hypotheses, H1 and H2. These hypotheses examine AI and Data Science's role in addition to potential advantages and barriers across various Portuguese industries.

3.3.1. Hypothesis for the Survey (H1)

H1: Among the surveyed companies in Portugal, those who report a higher level of AI and Data Science adoption in their NPD process will indicate a higher level of perceived innovation and efficiency.

H1 suggests that integrating AI and Data Science into NPD could enhance innovation and efficiency. Validating this hypothesis would support the proposed conceptual model, signifying the potential benefits of such integration in NPD. A pre-interview survey collects initial data, shaping the interview questions. The interview addresses AI and Data Science adoption levels and perceived value.

3.3.2. Hypothesis for the Semi-structured Interviews (H2)

H2: The adoption and impact of AI and Data Science in the NPD process will vary across sectors, with some sectors benefiting more than others due to various factors, including potential barriers.

H2 examines AI and Data Science adoption variability across sectors. Verifying this hypothesis would suggest the model's applicability varies by sector hence the need for it to be flexible to accommodate sector specific practices. Through semi-structured interviews, the questions assess AI and Data Science integration, benefits, challenges, and sector comparisons, providing insight into the hypothesis and framework's sectoral relevance.

3.4. Research Design

The research design intends to provide an answer to the intricacies of the research question by integrating quantitative and qualitative techniques.

As posited by Heigham and Croker (2009), understanding the multifaceted dimensions of today's world, the author's perspective suggests that embracing flexible data gathering strategies is often beneficial. Considering this viewpoint, this study adopts a mixed-methods framework, allowing both quantitative measures and qualitative insights.

The investigation initiated with a pre-interview survey. Designed to capture a quantitative overview of AI and Data Science incorporation in Portuguese enterprises, this survey aimed to gather information for hypothesis H1 and developing subsequent semi-structured questions.

Following this, semi-structured discussions were carried. Adhering to established practices in qualitative data, these dialogues enabled participants to share their distinct experiences, thus enriching the study's depth (Galletta, 2013; Oates, 2006; Whiting, 2008). Such conversations provided further insight into the role of AI and Data Science within NPD processes, addressing both hypotheses H1 and H2.

This study wraps up with an empirical application, assessing the integration of AI and Data Science tools in the NPD framework. Drawing from quasi-experimental methodologies (Galletta, 2013; Runeson & Höst, 2009; Whiting, 2008), this phase aspires to offer evidence of the efficacy of these tools.

3.4.1. Preliminary Data Collection

This section discusses the initial phase research that involves both quantitative and qualitative data collection and its analysis.

Data Collection Instruments

The pre-assessment survey was conducted using Collect.chat and the semi-structured interviews were carried out via Google Meet. The chatbot-based survey aimed to simulate conversation and maintain participant engagement, along with testing the viability of chatbot roles in NPD. The interviews were driven by a PowerPoint guide. Each interview, lasted around 30 minutes and was recorded, transcribed, and then translated ensuring data reliability as suggested by Galletta, (2013), M. L. Larson (1991), Temple & Young (2004) and Whiting (2008). Automated transcriptions were executed using Google Cloud's algorithms to assess the feasibility of integrating AI into the NPD process in accordance with the approach advocated by the proposed Conceptual Framework. Translations were performed with Office 365 and human re-checked.

Pilot Study

In this research, a pilot study assessed the efficacy of data collection tools, including surveys and interview protocols (van Teijlingen & Hundley, 2002). Four respondents' interactions with the chatbot-mediated survey led to modifications in the final survey clarity (Collins, 2003). Subsequently, the pilot study with other three-member sample assisted in refining interview protocols. The pilot aimed at enhancing the primary study's reliability and bolstering the research's integrity (Collins, 2003; Stockemer, 2019).

Sampling Strategy

The study used a combined sampling method, integrating both convenience and purposive elements (Stockemer, 2019). Convenience sampling allowed data collection from accessible sources, though possibly lacking complete representativeness (van Teijlingen & Hundley, 2002). Purposive sampling aimed to include diverse respondents (Stockemer, 2019). Varied roles and industries were targeted for insights into AI and data science use. Potential biases like self-selection might have influenced sample diversity (van Teijlingen & Hundley, 2002). Thus, the results from this methodology required cautious interpretation.

Sample Description

For the study, 10 companies were analysed, predominantly operating within the Information Technologies sector (30%). The other sectors, including Construction, Health, Market Studies, Media, Transportation, and Wholesale Trade, each constituted 10% representation. Concerning company size, 40% are classified as small, housing 1-50 employees, while very large entities account for 30%. Company age varies, with 30% having operated for over 50 years and up to 100 years.

Turning to the respondents representing these companies, there was an even gender distribution, with males and females each constituting 50%. The majority (60%) fall within the age bracket of 25-34 years. Professionally, roles such as Marketing Manager and AI/Data Science Specialist are most prevalent at 20%, with other roles like Administrator, Assistant Director, and CEO each representing 10%. On average, participants have been affiliated with their current companies for 3.6 years.

Pre-Assessment Survey (Quantitative Data Collection)

The research employed an AI like chatbot from Collect.chat, exemplifying AI's instrumental role in data collection, resonating with the proposed conceptual framework emphasis on AI and Data Science for Marketing and R&D. The survey, in APPENDIX I, p A, divided into ten sections, showcased a blend of question types, including dichotomous, categorical, interval, ordinal, and open-ended, ensuring precision and analytical simplicity.

The "Introduction and Authorization" section established the voluntary nature of participation and elucidated the survey's aims. Subsequently, the "Participant's Identification" section captured respondents' emails, pivotal for future interactions and data cross-referencing.

The "Demographic Information" section, utilizing categorical and interval questions, was designed for segmentation and statistical analysis.

The "Professional & Company Details" section, focusing on Hypothesis H1, collected data on respondents' roles and tenure. The "Industry Specifics" section, pivotal for Hypothesis H2, discerned the industry sector, setting the stage for sector-wise comparisons. The "AI and Data Science in the Organization" section probed into AI familiarity and adoption, shedding light on both Hypotheses H1 and H2.

The "Company's Product and Market Share" section, leveraging open-ended questions, provided a window into the company's offerings and market stance, indirectly bolstering Hypothesis H1. The "Feedback" section, though not directly tied to any hypothesis, was instrumental for survey validation.

The "Interview Scheduling" section streamlined interview setups, and the final section conveyed appreciation, nurturing a positive respondent relationship. Data was automatically archived in a CSV dataset by the chatbot. Further enrichment of this dataset introduced additional columns, enabling a more profound analytical dive and insight extraction.

Semi-Structured Interviews (Qualitative Data Collection)

The methodology employed a semi-structured interview design to delve deeply into the intricacies of AI and Data Science integration in Portugal's NPD process. This approach, chosen for its adaptability, aimed to unearth unforeseen insights, enriching the understanding of the research problem. The interview questions, in APPENDIX I, p J, were meticulously crafted, reflecting the research hypotheses, H1 and H2. They probed organizational decision-making, strategies, AI adoption, and industry comparisons. Informed by survey results, this method aligns with the sequential explanatory strategy in mixed-methods research.

Key questions explored:

- a) **Data Sources for Decision-Making:** This sought to understand the variety and nature of data sources used, shedding light on potential AI applications.
- b) **New Product Development Process:** This aimed to discern the organization's approach to product ideation and development, pinpointing potential stages for AI integration.
- c) **Customer Engagement Strategies:** This investigated customer targeting techniques,

illuminating AI's potential role in enhancing engagement.

- d) **Inventory Management and Demand Forecasting:** This probed inventory optimization methods, highlighting AI's potential in enhancing efficiency.
- e) **Customer Satisfaction and Industry Trends:** This gauged strategies for staying updated on market shifts, revealing AI's potential in market intelligence.
- f) **Current Use of AI and Data Science:** This assessed the extent of AI adoption, directly testing the research hypotheses. Depending on responses, further inquiries delved into practical applications, benefits, challenges, or barriers to adoption.

Post-interview data was transcribed to Excel, with answers segmented into columns. Subsequent columns were created to capture emerging themes, facilitating a nuanced analysis, and cross-referencing with survey data.



Figure 14: Word Cloud of the Semi Structured Interviews

Ethical Considerations

The study followed ethical guidelines, prioritizing participant rights as recommended by McBeath (2020). Consent was obtained, and the research's objectives were clearly outlined. Interviews were recorded, ensuring confidentiality of identifiable information. Anonymity was paramount; all data was anonymized, with minimal personal details retained. Also aligning with McBeath (2020), participants were informed about data handling and storage.

Potential Limitations and Biases

This study recognizes potential limitations and biases. It faces potential self-selection bias, where those choosing to participate might have varied AI familiarity (Stockemer, 2019). The risk of social desirability bias was also present, with participants possibly responding favourably (Stockemer, 2019). Participant abandonment during survey was addressed by incorporating encouraging survey statements. Finally, abandonment from survey to the interview was addressed using the chatbot ability to send interview reminder emails. Still, 4 survey respondents did not followed to the interview.

The study employed a four-point Likert scale in its survey to assess respondent familiarity with AI and data science. This choice aimed to simplify respondent decision-making and reduce potential decision overload bias as described by (Haynes, 2009). It also intended to prevent neutral responses, enhancing accuracy (Haynes, 2009). Nevertheless, this choice might have increased the risk of Type I errors as stated by Jacoby & Matell (1971).

3.4.2. Practical Application

The main goal of this experiment is to highlight how AI and Data Science can be beneficial in the NPD process. The research aims to demonstrate how simple techniques and tools can automate the collection of data and the creation of knowledge, assisting in better decision-making in NPD. Additionally, it addressed potential challenges in this approach.

In the experiment, AI and Data Science tools were used to study different stages of a product's development, especially focusing on R&D and marketing. A product from Amazon, which had room for enhancements (4.3 stars considering 4,654 ratings having 17% ratings bellow 4 stars), was chosen. The customer feedback was researcher then gathered using Octoparse. This feedback was analysed for key sentiments with the help of ChatGPT. Following this, OpenAI's DALL-E was used to visualize potential product improvements. Finally, insights

from this analysis were transformed into an advertisement, which was shared on Instagram to see its alignment with customer preferences.

Web Scrapping for Data Mining

In the study, web scraping, specifically via Octoparse, was employed to extract data from Amazon.co.uk. Parsing converted text-based in html code into structured output, while extraction stored the data in CSV. The focus was on the "MAMEIDO 1l Water Bottle Rich Black" product. Upon inputting the product [URL](#), Octoparse detected webpage data, facilitating a data extraction workflow. Data fields were customized, validated, and data exported to Excel.

Sentiment Analysis

In the study, sentiment analysis, a primary technique of Natural Language Processing (NLP), was employed to understand sentiments within customer reviews of the "MAMEIDO 1l Water Bottle Rich Black". OpenAI's ChatGPT model 4 was used because it is ready to use without need to train any model for the analysis. The sentiment analysis followed the steps:

- 1) ChatGPT created a list of pros and cons from amazon raw reviews. Using structured prompts, the model discerned implicit meanings from reviews.
- 2) From the feedback the model generated lists of the product's benefits and drawbacks.
- 3) ChatGPT synthesized positive reviews (4-5 stars) into a coherent summary.
- 4) Similarly, negative reviews (1-3 stars) were summarized.
- 5) Two summarizing sentences, highlighting the product's advantages and disadvantages, were produced. These were later used with DALL-E to create visual representations of customer perspectives.

AI Product Generation

The AI model DALL-E, developed by OpenAI, was employed due to its proficiency in generating images from text prompts. The model, based on a GPT-3 variant, operates on a subscription basis, offering virtual representation capabilities (*Dall-e*, 2022). For the research, DALL-E rendered images from both positive and negative product prompts. These digital mock-ups (DMUs) hold relevance in product development by promoting efficient design reviews and decision-making, eliminating the need for physical prototypes (Riascos et al., 2015). Verlinden et al. (Verlinden et al., 2003) noted that virtual prototypes merge various

design aspects, facilitating analysis and communication. The design's strategic role in the NPD process integrates user perspectives effectively (Veryzer & Borja de Mozota, 2005).

Automated marketing promotion

In the study, an automated method was developed to generate content for an Instagram post, encompassing both textual and visual elements. The process started with text generation, using the AI model, ChatGPT, merging the brand's tone with positive sentiments derived from prior analysis to create the content copy. Subsequently, the AI model, DALL-E, was tasked to create an advertising image using a specific prompt. This approach aimed at demonstrating the feasibility of integrating AI in automated marketing, emphasizing efficient resource use and effective product feature communication.

Ethical Considerations

The methodology utilized for the study involved data mining from Amazon, not complying with its Terms of Service. The intent was to demonstrate the methodology's potential without seeking commercial gain from Amazon's data. In practical applications, data would have been sourced legally, perhaps from customer feedback on Amazon or a company's platform.

Potential Limitations and Biases

Data was web-scraped only from Amazon, limiting its scope, and potentially introducing biases like fake reviews. Analysis using ChatGPT may miss nuanced insights. Furthermore, summarizing reviews can omit details, while ChatGPT's multilingual understanding also varies. DALL-E's image generation depends on its training, possibly affecting accuracy.

3.1. Chapter Summary

In this chapter, the research design for addressing the study's central question was presented. The methodology employed a combination of quantitative and qualitative methods, beginning with a pre-interview survey on AI and Data Science's role in Portuguese enterprises, followed by semi-structured interviews to explore their impact on NPD. Subsequently, a practical application evaluated the feasibility of using AI and Data Science tools in NPD.

4. ANALYSIS AND RESULTS

4.1. Chapter Introduction

In this chapter, the analysis and results of the research are presented. Recalling the methodology, a mixed-methods approach was employed, beginning with a pre-interview survey followed by semi-structured interviews. Two hypotheses were posited: H1, suggesting that higher AI and Data Science adoption in NPD processes among Portuguese companies correlates with increased perceived innovation and efficiency; and H2, proposing that AI and Data Science adoption in NPD varies across sectors. This chapter will systematically test these hypotheses.

4.2. Data Presentation

4.2.1. Survey Data: Respondents

Respondents Survey Dataset

	Gender	Age	Job_Role	Years_at_Company	AI_DS_Familiarity
A	Female	25-34 years	Marketing Manager	1	1 - Slightly familiar
B	Male	35-44 years	Sales Manager	16	1 - Slightly familiar
C	Male	45-54 years	Administrator	1	2 - Moderately familiar
D	Female	25-34 years	Assistant Director	5	0 - I don't know
E	Female	35-44 years	Project Manager	2	1 - Slightly familiar
F	Female	25-34 years	AI and/or Data Science Specialist	1	3 - Very familiar
G	Male	25-34 years	Marketing Manager	1	1 - Slightly familiar
H	Female	25-34 years	Customer Service	1	1 - Slightly familiar
I	Male	25-34 years	CEO	3	3 - Very familiar
J	Male	45-54 years	AI and/or Data Science Specialist	5	3 - Very familiar

Table 2: Survey Dataset - Exploring Demographics, Professions, and AI Familiarity

The dataset (Table 2, p 41; complete in APPENDICES K) captured survey responses regarding the views of respondents. It encompassed four main questions, spanning from question `2 to 5`. These questions delved into aspects like gender, age, job role, years at the company, and familiarity with AI and Data Science. The dataset provided a comprehensive insight into the demographics and professional backgrounds of the respondents, as well as their exposure to AI and Data Science techniques.

4.2.2. Survey Data: Company Profiles

Company Survey Dataset - Part 1

	Company_Age	Company_Age_Bin	Company_Size	Company_Size_Bin	Industry_Sector
A	Over 50 years, up to 100 years	Established	Large (251-1000 employees)	Large	Health
B	Over 100 years	Legacy	Very large (over 1000 employees)	Very Large	Construction
C	6 to 10 years	New	Small (1-50 employees)	Small	Media
D	Over 50 years, up to 100 years	Established	Medium (51-250 employees)	Medium	Wholesale Trade
E	1 to 5 years	New	Small (1-50 employees)	Small	Market Studies
F	11 to 20 years	New	Very large (over 1000 employees)	Very Large	Information Technologies
G	6 to 10 years	New	Small (1-50 employees)	Small	Information Technologies
H	21 to 50 years	Established	Medium (51-250 employees)	Medium	Health
I	1 to 5 years	New	Small (1-50 employees)	Small	Information Technologies
J	Over 50 years, up to 100 years	Established	Very large (over 1000 employees)	Very Large	Transportation

Company Survey Dataset - Part 2

	AI_DS_Interest	AI_DS_In_ProductDev	AI_DS_Other_Activities	AI_DS_Other_Functions	Main_Product_Service
A	2 - Moderate interest	1 - Low adoption	1 - Low adoption	-	Logistics
B	0 - No Interest	0 - No adoption	0 - No adoption	-	Construction
C	2 - Moderate interest	1 - Low adoption	0 - No adoption	-	Media & Advertising
D	1 - Low Interest	1 - Low adoption	1 - Low adoption	-	Industrial Equipment & HVAC
E	1 - Low Interest	1 - Low adoption	2 - Moderate adoption	-	Marketing Services
F	3 - High Interest	2 - Moderate adoption	1 - Low adoption	-	Unknown
G	1 - Low Interest	1 - Low adoption	1 - Low adoption	CRM and content marketing	Security Services
H	0 - No Interest	0 - No adoption	0 - No adoption	-	Healthcare
I	3 - High Interest	2 - Moderate adoption	1 - Low adoption	-	Restaurant Tech
J	2 - Moderate interest	1 - Low adoption	1 - Low adoption	Operational Management	Transportation Services

Table 3: Survey Dataset - Exploring Companies' Characteristics and AI Adoption Trends

The dataset (Table 3, p 42; complete in APPENDICES K) captured survey responses from questions `5 to 9` about the companies. It encompassed several questions, spanning various aspects of the companies such as their age, size, industry sector, and their interest and adoption levels in AI and Data Science. The dataset provided a comprehensive insight into the characteristics and AI adoption trends of the surveyed companies.

4.2.3. Interview Data: Thematic Insights

This dataset (Table 4, p 43; complete in APPENDICES L) is derived from a thematic analysis of semi-structured interviews with 10 companies. The themes were encoded as categories, capturing various aspects like data sources, product development processes, and AI adoption trends. It offers a detailed view into companies' strategies, challenges, and their stance on AI and Data Science.

Interview Dataset - Part 1

	Data_Sources_Description_Cat	New_Product_Development_Process	Customer_Targeting_Strategies_Cat	Inventory_Optimization_Strategies_Cat
A	Brainstorming and Market Reading	Intuition	Multi-Channel Marketing	Automated Systems
B	Brainstorming and Market Reading	Market Opinions	Multi-Channel Marketing	Demand-Driven
C	ERP Systems	Brainstorming	Customer Interaction	Unknown Methods
D	ERP Systems	Market Opinions	Brand and Event Marketing	Demand-Driven
E	Market Research	Market Research	Demographic Specialization	Client/Project-Based Forecasting
F	Customer Feedback	Proof-of-Concept	Bragging and Customer Feedback	Unknown Methods
G	CRM and Analytics	Intuition	Lead Generation and Account-Based Marketing	Client/Project-Based Forecasting
H	Pharmacy Data System	Planning Structure	Multi-Channel Marketing	Automated Systems
I	CRM and Analytics	Community Engagement	Customer Interaction	Client/Project-Based Forecasting
J	Brainstorming and Market Reading	Customer Feedback	Competitive Pricing and Collaboration	Demand-Driven

Interview Dataset - Part 2

	Customer_Satisfaction_Insights_Cat	AI_DS_Use	AI_DS_Use AI_DS_Examples	AI_DS_Difficulties_Cat
A	Surveys	Active AI & Data Science User	AI in Marketing	Trust & Reliability Issues
B	Surveys	AI & Data Science Non-User	Not Used	Not Used
C	Surveys	AI & Data Science Non-User	Not Used	Not Used
D	Customer Interaction	Traditional ERP User	Not Used	Not Used
E	Sales and Performance Monitoring	Partial AI & Data Science User	Not Used	Not Used
F	Constant Communication	AI for Marketing & CRM User	AI in Marketing	Resource & Skill Deficit
G	Customer Interaction	AI & Data Science Non-User	AI in Marketing	Innovation Resistance
H	Sales and Performance Monitoring	Active AI & Data Science User	Automation in Inventory	Resource & Skill Deficit
I	Constant Communication	Data Science for Forecasting & Results Analysis User	Data Science in E-commerce	Trust & Reliability Issues
J	Predictive Analytics	Active AI & Data Science User	Predictive Modelling	Trust & Reliability Issues

Interview Dataset - Part 3

	Open_To_AI_DS	Barriers_To_AI_DS_Cat	Competitor_AI_DS_Use	Innovation_Importance_Cat
A	Already Use	Already Use	Unaware of AI/DS Use	Very Important
B	Medium-Term Consideration	Cultural Resistance	Unaware of AI/DS Use	Medium Important
C	Open to AI & DS	Technological and Cost Challenges	Unaware of AI/DS Use	Not Important
D	Open to AI & DS	Technological and Cost Challenges	Unaware of AI/DS Use	Very Important
E	Open to AI & DS	Customer Resistance	Specific AI/DS Application	Very Important
F	Already Use	Already Use	Specific Applications	Very Important
G	Already Use	Already Use	Specific Applications	Very Important
H	Already Use	Already Use	Indirectly Suggest AI/DS Use	Very Important
I	Already Use	Already Use	Have No Competitors	Not Important
J	Already Use	Already Use	Specific AI/DS Application	Very Important

Table 4: Dataset of Interviews - Revealing Strategies, Challenges, and AI/Data Science Perspectives

4.3. Statistical Analysis

4.3.1. Descriptive Statistics

Survey Data: Respondents' Statistics

In the survey dataset, gender distribution was balanced, with males and females each at 50%. Most respondents (60%) were aged 25-34, and 20% each in the 35-44 and 45-54 groups. Professionally, "Marketing Manager" and "AI/Data Science Specialist" roles were most common at 20%. Other roles like "CEO" accounted for 10%. Respondents averaged 3.6 years at their companies, with a 1.5-year median. Regarding AI familiarity, 50% were slightly acquainted, 30% very, 10% moderately, and 10% unfamiliar (Table 5,p 44).

Descriptive Statistics of the Respondents Survey Dataset

	Gender	Age	Job_Role	Years_at_Company	AI_DS_Familiarity
description	The gender of the respondent.	The age of the respondent.	The respondent's role within their company.	The number of years the respondent has been working at their company.	The respondent's familiarity with AI and data science techniques.
count	10	10	10	10.0	10
unique	2	3	8	-	4
top	Female	25-34 years	Marketing Manager	-	1 - Slightly familiar
freq	5	6	2	-	5
mean	-	-	-	3.6	-
std	-	-	-	4.6475800154489	-
min	-	-	-	1.0	-
25%	-	-	-	1.0	-
50%	-	-	-	1.5	-
75%	-	-	-	4.5	-
max	-	-	-	16.0	-
Frequencies	Female 50.00% Male 50.00%	25-34 years 60.00% 35-44 years 20.00% 45-54 years 20.00%	AI and/or Data Science Specialist 20.00% Administrator 10.00% Assistant Director 10.00% CEO 10.00% Customer Service 10.00% Marketing Manager 20.00% Project Manager 10.00% Sales Manager 10.00%	1 50.00% 2 10.00% 3 10.00% 5 20.00% 16 10.00%	0 - I don't know 10.00% 1 - Slightly familiar 50.00% 2 - Moderately familiar 10.00% 3 - Very familiar 30.00%

Table 5: Respondents Snapshot: A Balanced Gender Mix, Dominance of 25-34 Age Group, and Popular Roles in Marketing and AI/Data Science

Descriptive Statistics of the Company Survey - Part 1

	Company_Age	Company_Age_Bin	Company_Size	Company_Size_Bin	Industry_Sector
description	The age of the respondent's company or its parent group.	The age of the respondent's company or its parent group in three bins.	The size of the respondent's company or its parent group.	The size of the respondent's company or its parent group in bins.	The industry sector of the respondent's company.
count	10	10	10	10	10
unique	6	3	4	4	7
top	Over 50 years, up to 100 years	New	Small (1-50 employees)	Small	Information Technologies
freq	3	5	4	4	3
Frequencies	1 to 5 years 20.00% 11 to 20 years 10.00% 21 to 50 years 10.00% 6 to 10 years 20.00% Over 100 years 10.00% Over 50 years, up to 100 years 30.00%	Established 40.00% Legacy 10.00% New 50.00%	Large (251-1000 employees) 10.00% Medium (51-250 employees) 20.00% Small (1-50 employees) 40.00% Very large (over 1000 employees) 30.00%	Large 10.00% Medium 20.00% Small 40.00% Very Large 30.00%	Construction 10.00% Health 20.00% Information Technologies 30.00% Market Studies 10.00% Media 10.00% Transportation 10.00% Wholesale Trade 10.00%

Descriptive Statistics of the Company Survey - Part 2

	AI_DS_Interest	AI_DS_In_ProductDev	AI_DS_Other_Activities	AI_DS_Other_Functions	Main_Product_Service
description	The company's level of interest in adopting AI and data science.	The level of AI and data science adoption in the company's product development process.	The level of AI and data science adoption in other activities within the company.	Other functions in which the company uses AI and/or data science.	The main product or service offered by the company.
count	10	10	10	10	10
unique	4	3	3	3	10
top	2 - Moderate interest	1 - Low adoption	1 - Low adoption	-	Logistics
freq	3	6	6	8	1
Frequencies	0 - No Interest 20.00% 1 - Low Interest 30.00% 2 - Moderate interest 30.00% 3 - High Interest 20.00%	0 - No adoption 20.00% 1 - Low adoption 60.00% 2 - Moderate adoption 20.00%	0 - No adoption 30.00% 1 - Low adoption 60.00% 2 - Moderate adoption 10.00%	- 80.00% CRM and content marketing 10.00% Operational Management 10.00%	Construction 10.00% Healthcare 10.00% Industrial Equipment & HVAC 10.00% Logistics 10.00% Marketing Services 10.00% Media & Advertising 10.00% Restaurant Tech 10.00% Security Services 10.00% Transportation Services 10.00% Unknown 10.00%

Table 6: Companies' Profiles: Varied Ages, Sizes, Industries; Mixed AI Adoption & Disclosures.

Survey Data: Company Profiles' Statistics

The distribution of the companies revealed that 30% were established for over 50 years, up to 100 years, while 50% were relatively new companies with 6 to 10 years of history. Company sizes varied, with 40% being small (1-50 employees) and 30% being very large (over 1000 employees). The most prevalent industry sector was Information Technologies, at 30%. In terms of AI and Data Science interest, 30% showed moderate interest, and 60% had low adoption in various aspects. The majority did not disclose their market share.

Interview Data: Thematic Statistics

The analysis of a dataset comprising 10 companies provided insights into their operational strategies and AI adoption. 'Brainstorming and Market Reading' was the prime method under 'Data_Sources_Description_Cat' at 30%. Both 'Intuition' and 'Market Opinions' led the 'New_Product_Development_Process' with 20% each. 'Customer_Targeting_Strategies_Cat' highlighted a 20% preference for 'Customer Interaction'.

'Inventory_Optimization_Strategies_Cat' was governed by 'Client/Project-Based Forecasting' and 'Demand-Driven' (30% each). The most favoured customer insight method was 'Surveys' (30%). In the AI realm, 30% were 'Active AI & Data Science Users'. However, a 60% majority already used AI, signifying adoption barriers. Emphasizing innovation's role, 70% deemed it 'Very Important'. The dominant company 'Strategy' was 'Data-Driven' (40%). This data underscores a shifting corporate inclination towards data-driven and AI-integrated strategies.

Descriptive Interview Dataset - Part 1

	Data_Sources_Description_Cat	New_Product_Development_Process	Customer_Targeting_Strategies_Cat	Inventory_Optimization_Strategies_Cat
description	Categorical representation of the types of data sources the company uses.	Description of the company's process for identifying, evaluating, and developing new product ideas, including any research and development methods used.	Categorical representation of the strategies the company uses to target customers.	Categorical representation of the strategies the company uses for inventory optimization.
count	10	10	10	10
unique	6	8	7	4
top	Brainstorming and Market Reading	Intuition	Multi-Channel Marketing	Demand-Driven
freq	3	2	3	3
mean	-	-	-	-
std	-	-	-	-
min	-	-	-	-
25%	-	-	-	-
50%	-	-	-	-
75%	-	-	-	-
max	-	-	-	-
Frequencies	Brainstorming and Market Reading 30.00% CRM and Analytics 20.00% Customer Feedback 10.00% ERP Systems 20.00% Market Research 10.00% Pharmacy Data System 10.00%	Brainstorming 10.00% Community Engagement 10.00% Customer Feedback 10.00% Intuition 20.00% Market Opinions 20.00% Market Research 10.00% Planning Structure 10.00% Proof-of-Concept 10.00%	Bragging and Customer Feedback 10.00% Brand and Event Marketing 10.00% Competitive Pricing and Collaboration 10.00% Customer Interaction 20.00% Demographic Specialization 10.00% Lead Generation and Account-Based Marketing 10.00% Multi-Channel Marketing 30.00%	Automated Systems 20.00% Client/Project-Based Forecasting 30.00% Demand-Driven 30.00% Unknown Methods 20.00%

Descriptive Interview Dataset - Part 2

	Customer_Satisfaction_Insights_Cat	AI_DS_Use	AI_DS_Use_AI_DS_Examples	AI_DS_Difficulties_Cat
description	Categorical representation of the methods the company uses to gain customer satisfaction insights.	Whether the company is currently using any AI or data science techniques in its decision-making processes, product development, marketing or selling strategies.	Examples of how these AI or data science techniques have been implemented and their impact on the company (if the answer to AI_DS_Use was "yes").	Categorical representation of the difficulties the company faces in implementing AI and data science techniques (if the answer to AI_DS_Use was "yes").
count	10	10	10	10
unique	5	6	5	4
top	Surveys	Active AI & Data Science User	Not Used	Not Used
freq	3	3	4	4
mean	-	-	-	-
std	-	-	-	-
min	-	-	-	-
25%	-	-	-	-
50%	-	-	-	-
75%	-	-	-	-
max	-	-	-	-
Frequencies	Constant Communication 20.00% Customer Interaction 20.00% Predictive Analytics 10.00% Sales and Performance Monitoring 20.00% Surveys 30.00%	AI & Data Science Non-User 30.00% AI for Marketing & CRM User 10.00% Active AI & Data Science User 30.00% Data Science for Forecasting & Results Analysis User 10.00% Partial AI & Data Science User 10.00% Traditional ERP User 10.00%	AI in Marketing 30.00% Automation in Inventory 10.00% Data Science in E-commerce 10.00% Not Used 40.00% Predictive Modeling 10.00%	Innovation Resistance 10.00% Not Used 40.00% Resource & Skill Deficit 20.00% Trust & Reliability Issues 30.00%

Descriptive Interview Dataset - Part 3

	Open_To_AI_DS	Barriers_To_AI_DS_Cat	Competitor_AI_DS_Use	Innovation_Importance_Cat
description	Whether the company is open to exploring the use of AI and data science in the future (if the answer to AI_DS_Use was "no").	Categorical representation of the barriers the company faces in implementing AI and data science.	Whether the other main market players use any AI or data science techniques.	Categorical representation of the importance of innovation to the company.
count	10	10	10	10
unique	3	4	5	3
top	Already Use	Already Use	Unaware of AI/DS Use	Very Important
freq	6	6	4	7
mean	-	-	-	-
std	-	-	-	-
min	-	-	-	-
25%	-	-	-	-
50%	-	-	-	-
75%	-	-	-	-
max	-	-	-	-
Frequencies	Already Use 60.00% Medium-Term Consideration 10.00% Open to AI & DS 30.00%	Already Use 60.00% Cultural Resistance 10.00% Customer Resistance 10.00% Technological and Cost Challenges 20.00%	Have No Competitors 10.00% Indirectly Suggest AI/DS Use 10.00% Specific AI/DS Application 20.00% Specific Applications 20.00% Unaware of AI/DS Use 40.00%	Medium Important 10.00% Not Important 20.00% Very Important 70.00%

Table 7: Corporate Strategies: Data-Driven Shift, AI Adoption, and Innovation Emphasized.

4.3.2. Inferential Statistics

Fisher's Exact Test

Dimension	Test Pair	P Value
01_respondent	Age vs Industry_Sector	0.0397
	AI_DS_Familiarity vs Job_Role	0.0714
	Gender vs Industry_Sector	0.5238
	Age vs AI_DS_Familiarity	0.5714
	Age vs Job_Role	0.6349
02_company	AI_DS_In_ProductDev vs AI_DS_Interest	0.0016
	AI_DS_Other_Activities vs Strategy	0.0333
	AI_DS_In_ProductDev vs AI_DS_Other_Activities	0.2619
	AI_DS_In_ProductDev vs Strategy	0.2825
	AI_DS_Interest vs Strategy	0.3867
	AI_DS_Other_Functions vs Market_Share	0.5333
03_sector	AI_DS_Interest vs Market_Share	0.55
	AI_DS_Other_Activities vs Industry_Sector	0.175
	AI_DS_Familiarity vs Industry_Sector	0.2464
	Industry_Sector vs Market_Share	0.2917
	Industry_Sector vs Strategy	0.5746
04_survey	AI_DS_Other_Functions vs Industry_Sector	0.8
	AI_DS_In_ProductDev vs Industry_Sector	0.8952
	AI_DS_Familiarity vs Survey_Quality	0.0905
	AI_Experience_Evaluation vs Survey_Quality	0.1238
	Job_Role vs Survey_Quality	0.3968
	Strategy vs Survey_Quality	0.5276
	AI_Experience_Evaluation vs Job_Role	0.5302
Age vs Survey_Quality	0.581	
	AI_DS_Interest vs AI_Experience_Evaluation	0.7714

Table 8: Fisher's Test Unveils Data Relationships: Purple for Significant Links, Yellow for Important Insights.

Fisher's Exact Test analysed associations between categorical variables from a survey dataset. Designed for small sample sizes, this test accurately determines probability distribution without data distribution assumptions. It assessed pairs of categorical columns, identifying significant links between 'Age' and 'Industry_Sector' (p=0.0397) regarding the respondents themselves, and between 'AI_DS_In_ProductDev' and 'AI_DS_Interest' (p=0.0016) as well as 'AI_DS_Other_Activities' and 'Strategy' (p=0.0333) when considering the industry aspects. These results hint at relationships between a company's interest in AI and Data Science and their implementation in product development and other functions. The analysis of industry sectors showed no significant results.

Correlation Analysis

In the analysis, relationships between numerical variables were evaluated using Kendall's and Spearman's tests. Significant correlations emerged: 'AI_DS_Familiarity_Numerical' and 'AI_DS_Interest_Numerical' had a coefficient of 0.764 (p-value: 0.010), while between

'AI_DS_Interest_Numerical' and 'AI_DS_In_ProductDev_Numerical', it was 0.896 (p-value: 0.000). This suggests that increased company interest in AI and Data Science correlates with the respondents AI familiarity. Similarly increased company interest in AI and Data Science correlates with a company's intent to adopt AI in product development.

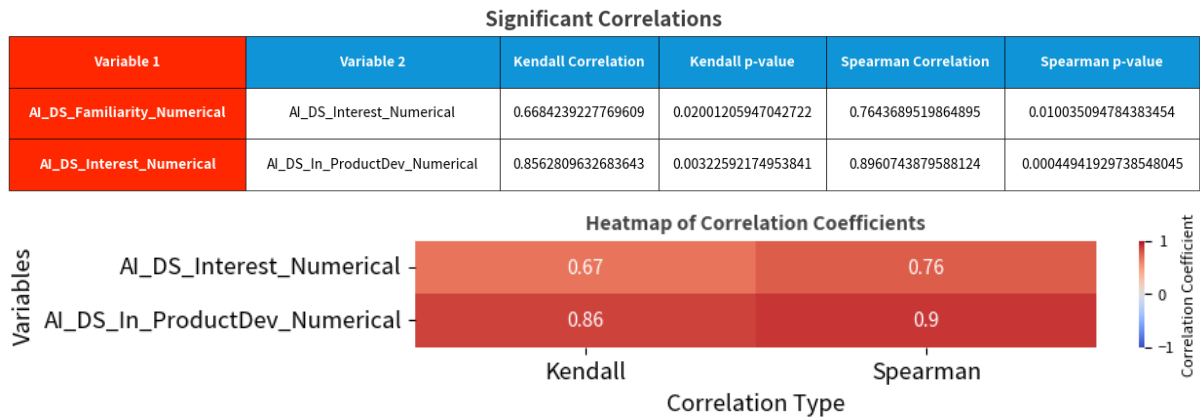


Figure 15: Strong Correlations Found Between AI Interest, Familiarity, and Product Adoption.

Kruskal-Wallis Test

To assess the differences across 'Industry_Sector' for 'AI_DS_Interest_Numerical' and 'AI_DS_In_ProductDev_Numerical', the Kruskal-Wallis H Test, since it does not assume normality, it was applied to this sample of 10 respondents. No significant differences emerged with p-values of 0.5565 and 0.7898. After sector consolidation, results were consistent. Bootstrapping with 1000 iterations yielded p-values of 0.1399 and 0.2820 for each variable, exceeding the 0.05 threshold. In conclusion, no significant median variations were detected for either variable, and the test's power was below the 80% benchmark, perhaps due to sample size or minimal effect.

Kruskal-Wallis Test Results

	AI_DS_Interest_Numerical	AI_DS_In_ProductDev_Numerical
Statistic	4.901612903225811	3.149999999999999
p-value	0.5564926535839894	0.7897985033388103

Kruskal-Wallis Fewer Categories Test Results

	AI_DS_Interest_Numerical	AI_DS_In_ProductDev_Numerical
Statistic	4.901612903225811	3.149999999999999
p-value	0.5564926535839894	0.7897985033388103

Kruskal-Wallis Bootstrap Test Results

	AI_DS_Interest_Numerical	AI_DS_In_ProductDev_Numerical
Statistic	4.901612903225811	3.149999999999999
p-value	0.16474724441983324	0.30852678628684643

Table 9: Kruskal-Wallis Test Reveals No Significant Median Differences in AI Interests.

4.3.3. Multivariate Statistics

Cluster Analysis

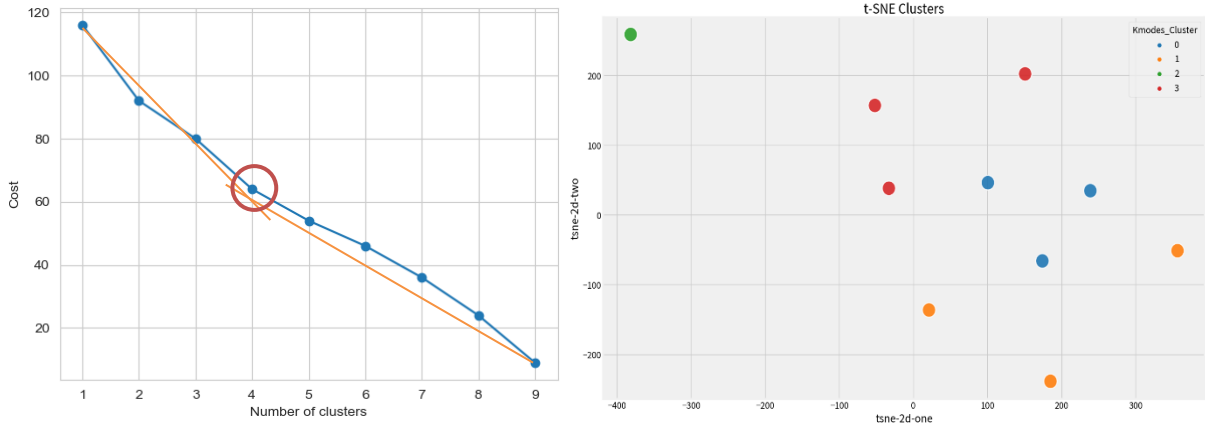


Figure 16: KModes Algorithm Groups Data: Elbow method (left) identified 4 Clusters, and the t-SNE technique for 2D representation (right).

The study employed the 'KModes' algorithm, a variant of unsupervised machine learning tailored for categorical data, to group data points by similarities. The selection of the optimal seed ensured stable and reproducible clusters, with the best seed identified by the least negative cost. The elbow method indicated 4 clusters. Subsequent analysis used the t-SNE technique for 2D representation and a decision tree to detect influential features.

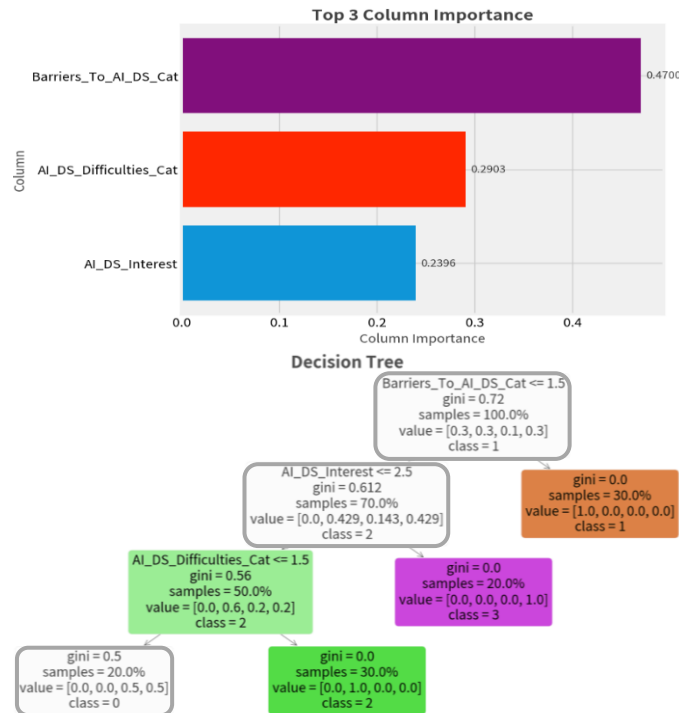


Figure 17: Most Influential Features for Clustering

Four clusters were discerned:

Established Traditionalists, predominantly legacy companies, existed for over a century and often boasted over 1000 employees. Low adoption rates of AI and DS, indicating resistance to these technologies.

Open-Minded Traditionalists, with a history of over 50 years, exhibited varied AI and DS adoption rates, suggesting a cautiously receptive stance.

Diverse Adopters, with varied ages and sizes, displayed keen interest in AI and DS but faced barriers to full integration.

Emerging Tech Enthusiasts, newer companies, showed high enthusiasm for AI and DS but encountered resource constraints and trust challenges.

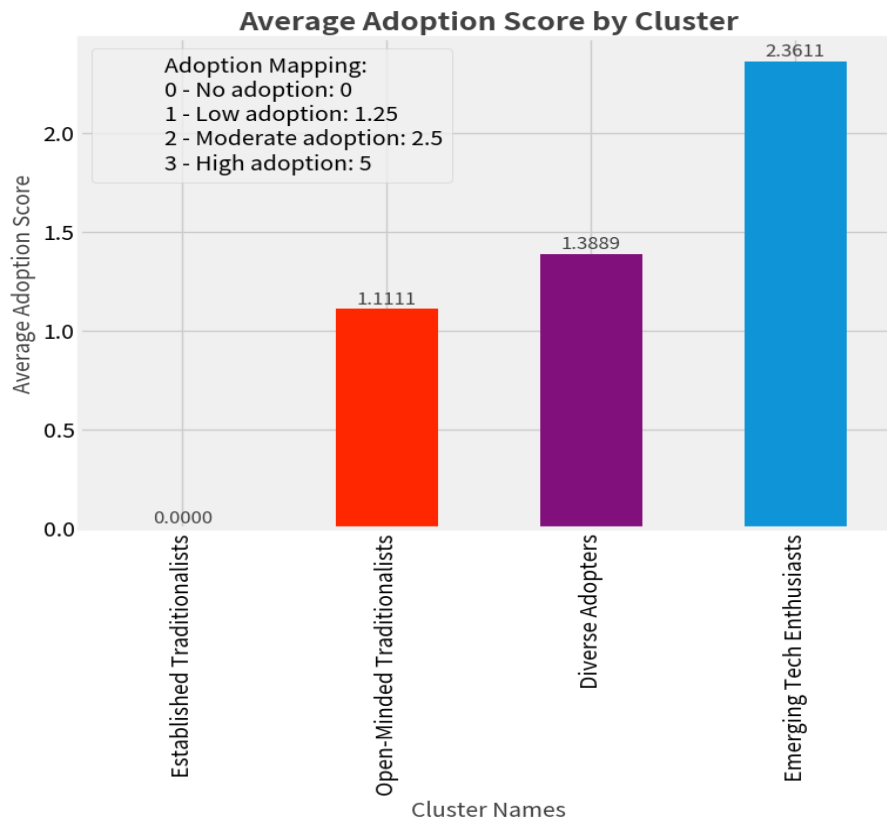


Figure 18: Cluster Rankings Reflect Varied AI and DS Adoption Across Groups.

Clusters were ranked by their average AI and DS adoption scores, and the resulting values (0 to 5) were mapped from “0-No Adoption” to “3 – High Adoption” for better context. Once again, the “Established Traditionalist” showed “0 - No Adoption”, while the maximum was “2 – Moderate Adoption” for “Emerging Tech Enthusiasts”. Neither of these clusters showed “3 – High Adoption”.

4.4. Results of Statistical Analysis

To address the two hypotheses based on the results provided:

4.4.1. Hypothesis H1 – AI Integration & Innovation Rates in Portuguese NPD

H1: Among the surveyed companies in Portugal, those who report a higher level of AI and Data Science adoption in their NPD process will indicate a higher level of perceived innovation and efficiency.

H1 Evidence from the Results

- 1) A significant correlation was identified between 'AI_DS_Interest_Numerical' and 'AI_DS_In_ProductDev_Numerical', with a coefficient of 0.896 (p-value: 0.000). This indicates that companies with a growing interest in AI and Data Science tend to implement these technologies in product development.
- 2) In the thematic statistics, 70% of the surveyed companies consider innovation as 'Very Important'. Additionally, 30% of these companies are categorized as 'Active AI & Data Science Users', while another 60% are already in the process of AI adoption.
- 3) The Fisher's Exact Test also showed a significant association between 'AI_DS_In_ProductDev' and 'AI_DS_Interest'. This underlines the close relationship between a company's interest in AI and Data Science and its practical application.
- 4) The dominant company strategy being 'Data-Driven' (40%) further emphasizes the importance these companies place on data-centric methods.

Conclusion for H1

There is strong evidence supporting a positive correlation between the level of interest in AI and Data Science and its adoption in the NPD process. However, while many companies lean towards a data-driven strategy and consider innovation paramount, the exact magnitude of innovation and efficiency resulting from AI/DS integration in NPD remains ambiguous.

4.4.2. Hypothesis H2 - AI in NPD: Sectoral Variance

H2: The adoption and impact of AI and Data Science in the NPD process will vary across sectors, with some sectors benefiting more than others due to various factors, including potential barriers.

H2 Evidence from the Results

- 1) The cluster analysis identified four distinct company profiles based on their AI and DS adoption. Notably, 'Established Traditionalists' portrayed resistance to the adoption of these technologies. In contrast, 'Emerging Tech Enthusiasts' demonstrated high enthusiasm but faced challenges like resource limitations and trust issues.
- 2) The Kruskal-Wallis H Test applied on 'AI_DS_Interest_Numerical' and 'AI_DS_In_ProductDev_Numerical' across different 'Industry_Sectors' revealed no significant median differences. This suggests that the rate of AI and DS adoption does not differ markedly across industries, at least in the sampled group.

Conclusion for H2

While there are discernible patterns in AI and DS adoption levels among different company profiles, the sectoral differences are not markedly supported by the results. Although the Kruskal-Wallis Test was inconclusive, potentially due to sample size limitations, the cluster analysis provided nuanced insights into company-specific adoption tendencies.

4.1. Survey Evaluation and Robustness Checks

4.1.1. Survey Evaluation

Descriptive Survey Quality Dataset

	Survey_Quality	AI_Experience_Evaluation	Survey_Comments
description	The respondent's perception of the quality of the survey questions.	The respondent's evaluation of their experience with the AI used in the survey.	The respondent's comments about their experience of the survey.
count	10.0	10	10
unique	-	3	10
top	-	Good	It's more interactive and less boring
freq	-	5	1
mean	8.9	-	-
std	1.1005049346146119	-	-
min	7.0	-	-
25%	8.0	-	-
50%	9.0	-	-
75%	10.0	-	-
max	10.0	-	-
Frequencies	7 10.00% 8 30.00% 9 20.00% 10 40.00%	Excellent 30.00% Good 50.00% Regular 20.00%	A little longer 10.00% Different. And therefore attracts attention 10.00% Greater speed and efficiency 10.00% I couldn't understand where they applied AI here? It seemed to me a very straightforward form 10.00% It was more interactive and interesting to answer the questions 10.00% It's more interactive and less boring 10.00% More humanized but with lower perception of survey completion percentage 10.00% More interactive and captivating 10.00% Some questions lacked context for scenarios where companies are international. 10.00% The fact that it is a chat makes the survey more "humanized" 10.00%

Table 10: Chatbot Survey: High Quality Ratings, Strong AI User Satisfaction.

The chatbot survey (Table 10, p 52; complete in APPENDICES T), gauging respondents' perceptions of its quality and their AI experience, received an average quality rating of 8.9 out of 10. Most respondents (40%) gave it a perfect score, while the lowest was 7. In terms of AI experience, half (50%) rated it as "Good", followed by 30% who found it "Excellent" (Figure 19, p 53).

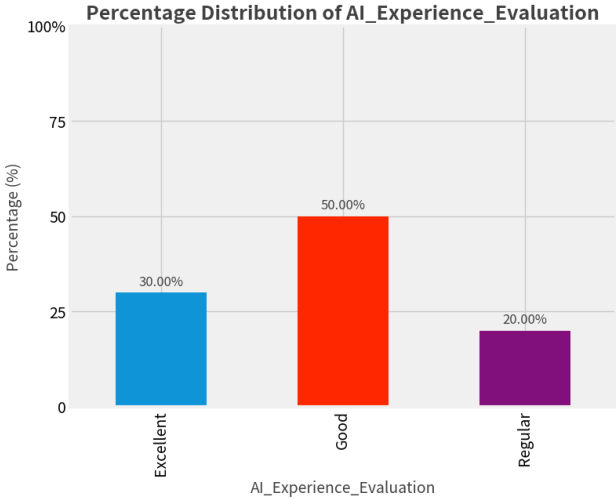


Figure 19: Survey Results: 40% Perfect Score, 7% Lowest; AI Ratings Vary

Comments highlighted the AI-driven survey's interactivity and novelty, with one remarking on its "humanized" feel. However, some found it longer, and one respondent couldn't discern the AI's role. Overall, the chatbot survey was largely well-received. The chatbot survey's high-quality rating and positive feedback support AI's potential in R&D and Market Research for NPD (in line with the proposed conceptual model), emphasizing interactivity, novelty, and a "humanized" approach, thus reducing the chance of participant abandonment.

4.1.2. Robustness Check

To ensure the robustness of the data collection, both Mann-Whitney U and Wilcoxon Signed-Rank Tests were employed:

The Mann-Whitney U test revealed significant differences in survey evaluations based on domain knowledge. Specifically, pairs ('Survey_Quality', 'Years_at_Company') and ('Survey_Quality', 'AI_DS_Familiarity_Numerical') had corrected p-values of 0.0086 and 0.0007 respectively, indicating that survey evaluations weren't attributed randomly by respondents.

The Wilcoxon Signed-Rank Test for repeated measures assessed consistency between

survey and interview responses. With p-values greater than 0.05 for comparisons like 'AI_DS_Interest_Binary' (survey variable) and 'AI_DS_Use_Binary' (interview variable) ($p=0.083$), the test failed to reject the null hypothesis, suggesting consistent responses across surveys and interviews.

These findings support the reliability and validity of the data collection.

4.2. Applied Analysis

4.2.1. Web Scraping for Data Mining

The study used Octoparse to web-scrape Amazon product reviews, which were then processed in a Jupyter notebook. ChatGPT was utilized for automated text analysis. The dataset had columns like Review, Rating, and Country, among others. Out of 741 unique reviews, "Perfekt" was the most frequent comment. The average rating stood at 3.75, with "Good" sentiment at 64.52%. Most reviews came from the UK (92.31%). Reviews were primarily submitted on Saturdays and Tuesdays, peaking in September, April, and May (Table 11, p 54).

Descriptive Statistics of the Amazon Reviews

	Rating	Feeling	Country	Article_Name
description	Numerical rating given by the reviewer, ranging from 1 to 5.	Categorical data indicating the sentiment of the review (e.g., Good, Bad).	Country from which the review was posted.	Name of the article or product being reviewed.
Number of non-missing values	747.0	747	26	732
Number of unique values	-	2	2	9
Most common value	-	Good	United Kingdom	Flamingo Pink
Frequency of most common value	-	482	24	155
mean	3.749665327978581	-	-	-
std	1.5542846670357033	-	-	-
min	1.0	-	-	-
25%	2.0	-	-	-
50%	5.0	-	-	-
75%	5.0	-	-	-
max	5.0	-	-	-
Frequencies	1 15.93% 2 10.58% 3 8.97% 4 11.65% 5 52.88%	Bad 35.48% Good 64.52%	Netherlands 7.69% United Kingdom 92.31%	Carbon Grey 18.44% Emerald Green Gold 2.60% Flamingo Pink 21.17% Frosted White Gold 13.11% Ivory Beige Gold 13.39% Rich Black 6.56% Rosé Quartz 2.60% Sapphire Blue 1.78% Urban Taupe Gold 20.36%

Table 11: Amazon Review Analysis: Octoparse Scraping, ChatGPT Analysis, UK Dominance, September Peak.

4.2.2. *Sentiment Analysis*



Figure 20: Good Highlights: Product Quality Emphasizes Durability and Reliability Trends (ChatGPT).

ChatGPT analysed reviews from two datasets: `good_df` and `bad_df`. In `good_df` (Figure 20 ,p 55), themes like "Product Quality" emphasized durability and reliability, while "Aesthetics" focused on design and appearance. In contrast, `bad_df` revealed concerns like "Product Defects," indicating quality control issues, and themes like "Misleading Information" and "Complexity." Sample reviews highlighted praises for design in `good_df` and product defects in `bad_df`. Keywords for `good_df` included "Quality" and "Design," while `bad_df` featured "Defect" and "Misleading."

Bad Reviews:

1. "Schön und hält dicht, aber die Beschichtung des Deckels löst sich in dem Geschirrspüler leider und auch der Kleber, der das Gewinde im Deckel hält löst sich." - *Translation: Nice and tight, but unfortunately the coating of the lid comes off in the dishwasher and the glue that holds the thread in the lid also comes off.*
2. "Il materiale non è dei migliori. Sicuramente è comoda e leggera per portare con sé sempre 1l di acqua, ma dopo poco all'interno si sono create macchie (nonostante i lavaggi). In più all'interno del tappo c'è la guarnizione di gomma che non è fissata e ogni volta viene via. Parecchie volte mi ha allagato la borsa! Non la consiglio" - *Translation: The material is not the best. It is certainly comfortable and light to always carry 1l of water with you, but after a short time stains were created inside (despite washing). In addition, inside the cap there is a rubber gasket that is not fixed and comes off every time. Several times it flooded my bag! I don't recommend it.*
3. "Wie in einigen Rezensionen bereits erwähnt, löst sich der Gummiring sehr leicht, wodurch die Flasche undicht wird. Somit ist sie leider nicht zu nutzen. Ich werde sie zurückschicken müssen. Schade!" - *Translation: As already mentioned in some reviews, the rubber ring comes off very easily, making the bottle leak. So unfortunately it cannot be used. I will have to send it back. Too bad!*
4. "Nichts" - *Translation: Nothing.*
5. "Die Flasche war nach kurzer Zeit undicht und ich kann mir eine neue kaufen" - *Translation: The bottle was leaking after a short time and I have to buy a new one.*

Figure 21: Downsides: quality control issues and misleading information in products (ChatGPT).

4.2.3. AI Product Generation

In the study, DALL·E 2 was employed to visualize product mock-ups based on feedback from ChatGPT's analysis of good_df and bad_df reviews. For the positive mock-up, ChatGPT crafted a description emphasizing features like "sleek and modern design" and "comfortable grip," leading DALL·E 2 to generate an image of a contemporary, user-friendly water bottle. Conversely, for the negative mock-up, terms like "flimsy structure" and "visible defects" were used, guiding DALL·E 2 to visualize a bottle with evident imperfections and a low appearance.



Figure 22: DALL·E 2 transforms ChatGPT's good feedback into a sleek, user-friendly water bottle.



Figure 23: DALL·E 2 envisions a flawed bottle, influenced by bad descriptors.

4.2.4. Automated Marketing Promotion

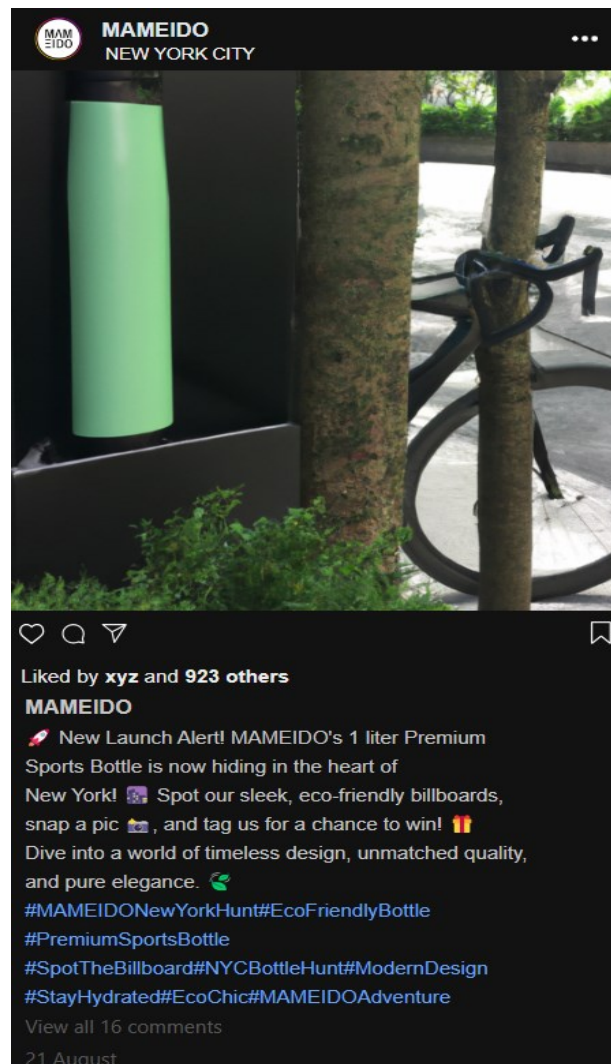


Figure 24: ChatGPT and DALL·E unite, crafting eco-friendly Instagram post for MAMEIDO.

Finally, in the study, ChatGPT and DALL·E were combined to create an automated Instagram post for MAMEIDO's product launch. ChatGPT crafted an SEO-optimized text highlighting the product's eco-friendliness and a New York billboard hunt. Simultaneously, DALL·E produced an image aligning with the brand's ethos. The result was a unified Instagram post, showcasing AI's potential in digital marketing.

4.3. Results of Applied Analysis

Recalling H1:

H1: Among the surveyed companies in Portugal, those who report a higher level of AI and Data Science adoption in their NPD process will indicate a higher level of perceived innovation and efficiency.

Applied Analysis - H1 Evidence from the Results

- 1) The research highlighted AI tools' efficiency in market research, achieving results in an hour compared to traditional methods that took weeks. This indicates a potential boost in efficiency for companies integrating AI.
- 2) DALL·E's proficiency in generating diverse product mock-ups illustrates AI's innovative role in product development, supporting the hypothesis of AI fostering innovation.
- 3) The swift adaptation and visualization of product concepts, contrasting with conventional R&D, accentuate AI tools' innovative and efficient nature.
- 4) The collaboration between DALL·E and ChatGPT emphasized the merging of linguistic and visual AI capabilities, indicating a holistic approach to innovative product development.

Applied Analysis - Conclusion for H1

The findings offer compelling evidence for the hypothesis that AI and Data Science integration in NPD can enhance innovation and efficiency. The demonstrated abilities of AI tools, especially DALL·E and ChatGPT, emphasize AI's transformative role in product development. However, the study's constraints, including dataset limitations and potential nuances missed by ChatGPT, should be factored into the interpretation of these results.

4.4. Discussion of Results

The results for H1 indicate a strong correlation between interest in AI and Data Science and its application in the NPD process among Portuguese companies. While many firms prioritize a data-driven strategy, the exact impact of AI on perceived innovation remains slightly unclear. For H2, while distinct company profiles emerged regarding AI adoption, sectoral differences weren't significant. This might be due to sample size limitations. Additionally, the

practical application showcased that AI tools, like DALL·E and ChatGPT have a transformative potential in NPD, emphasizing efficiency and innovation. However, dataset constraints and potential nuances missed by AI tools warrant consideration.

4.5. Chapter Summary

This chapter presented a comprehensive analysis of the research data, focusing on the role of AI and Data Science in NPD among Portuguese companies. Descriptive statistics, inferential statistics, cluster analysis were used to test the hypothesis and a practical application showcased the applicability of these tools in NPD.

5. CONCLUSION

5.1. Preliminary Considerations

This dissertation delves deeply into 'Innovation Analytics', a term introduced by Kakatkar et al. (2020), which synergizes AI, Data Science, and the innovation process. Schumpeter (1934) defines innovation as the metamorphosis of an idea into a valuable offering that garners consumer interest, a process that inherently correlates with enhanced firm performance and unique value creation (Tidd & Bessant, 2009).

Central to this is the concept of dynamic capabilities. As elucidated by Teece et al. (1997), these capabilities empower firms to combine, evolve, and reshape competencies, facilitating the effective sensing of opportunities and adaptation to shifting landscapes. Helfat & Winter (2011) further emphasize their role in seizing opportunities. Alves et al. (2017) highlight that firms with dynamic capabilities often command innovation, outpacing rivals. Zahra et al. (2006) affirm that these capabilities amplify innovation, fostering competitive advantage and enduring success. Furthermore, McKinsey's (2022) report underscores the increasing AI capabilities, noting a leap from 1.9 capabilities per company in 2018 to 3.8 in 2022.

However, the European Commission (2018) accentuates the nuanced challenges accompanying these advancements, spanning ethical, legal, economic, cybersecurity, and data dimensions.

Zooming into Portugal, Gomes (2022) offers an intricate portrayal of Portuguese companies, highlighting hurdles in leveraging AI for strategic decisions while also emphasizing the embryonic stage of AI adoption in these firms.

These findings within the Portuguese industry emerged as a paradox: despite their potential to accelerate innovation (Zahra et al., 2006) and empower dynamic capabilities (Alves et al., 2017), the constrained integration of AI and Data Science, as highlighted by Gomes (2022) in Portugal's industries, appeared as an inhibitor of sustainable NPD.

This observation, of a potential underutilization of AI and Data Science in Portugal's industrial landscape, potentially curbing innovation, formed the foundation of this dissertation's problem statement. Consequently, the study sought to uncover the untapped potential of AI and Data Science within the domain of 'Innovation Analytics' (Kakatkar et al., 2020) in Portugal's NPD landscape.

Emerging from the problem statement, the main research question revolved around uncovering the present condition, instances of achievement, and obstacles linked to the integration of AI and Data Science for NPD within Portuguese Industries.

5.2. Insights from Literature Review

The literature review delved into innovation, dynamic capabilities, AI, and data science's roles to achieve sustainable innovation. The review explained innovation's commercial value, dynamic capabilities' adaptive essence, and their synergistic impact on firm performance. Innovation, AI, and data science's roles were underscored, with examples such as Lemonade and Mayo Clinic.

In the context of the Portuguese market, adoption and impact were explored considering Gomes' (2022) study. In his report Gomes (2022) found only one firm using AI for strategic decision-making but prevalent operational use of I.T. along with human decision.

Analysing the actual NPD process, the integration of AI and data science with the Stage-Gate model and functional structures was examined. This assessment revealed potential contributions of these technologies on decision-making, R&D, marketing, operations, and finance.

5.3. Methodological Framework and Data Acquisition

The main research question guided the literature review. Within the typical NPD process (Cooper's stage-gate), the literature analysis revealed a gap in integrating AI, Data Science, and functional organizational structures.

Consequently, a key contribution of this dissertation emerged: a conceptual framework that harmonizes AI, Data Science, and functional organizational structures with Cooper's Stage-Gate model. This innovation-grounded framework, enriched by principles of dynamic capabilities and technological integration, aims to serve as a roadmap for companies aspiring to leverage AI and Data Science's transformative potential within NPD.

Likewise, the creation of the proposed conceptual framework imposed a deeper understanding of the current state, success narratives, and challenges associated with AI and Data Science utilization in Portugal's NPD process.

From the proposed conceptual framework two hypotheses were considered.

Indeed, Hypothesis H1 stated that among surveyed companies in Portugal, those with higher AI and Data Science adoption in their NPD process would indicate greater levels of perceived innovation and efficiency. This hypothesis aimed to validate the potential benefits of integrating AI and Data Science into NPD, using survey data to shape subsequent interviews.

Whereas hypothesis H2 suggested that the adoption and impact of AI and Data Science in the NPD process would vary across sectors due to factors including potential barriers, leading to differential benefits. This second hypothesis aimed to explore sector-specific differences in AI and Data Science integration, benefits, challenges, and comparisons through semi-structured interviews.

Seeking to provide an answer to these hypotheses, a mixed-methods methodological approach was designed. For this reason, the methodology comprised surveys, semi-structured interviews, and a practical application, offering a triangulated perspective that illuminated both quantitative and qualitative aspects of the research. Regarding the order of the data collection, the investigation started with a pre-interview survey, followed by in-depth semi-structured interviews, culminating in an empirical application test, where each method provided insights for the following one.

The study employed a mixed sampling approach - convenience and purposive - combining accessible data collection with diverse respondent inclusion for insights into AI and data science use, acknowledging potential biases and the need for careful result interpretation.

The study examined 10 companies, mainly in Information Technologies (30%), and other sectors like Construction, Health, and more (each 10%). Company size ranged from small to very large, and their age varied. Among company respondents, gender distribution was even (50% males, 50% females), primarily aged 25-34 (60%). Predominant roles were Marketing Manager and AI/Data Science Specialist (20% each), while Administrator, Assistant Director, and CEO each represented 10%. Average company affiliation duration was 3.6 years.

5.4. Interpretation of Findings

Inferential statistics, particularly Fisher's Exact Test and Correlation Analysis, revealed significant associations and correlations. Fisher's exact test revealed significant relationships ($p=0.0016$) between 'AI_DS_In_ProductDev' and 'AI_DS_Interest,' and ($p=0.0333$) between 'AI_DS_Other_Activities' and 'Strategy,' suggesting links between AI/DS interest and implementation in product development and functions, with no sector-specific significance. In

addition, a notable correlation was identified between a company's interest in AI and Data Science and its application in product development, with a coefficient of 0.896 (p-value: 0.000).

Hence, for Hypothesis H1, the evidence, derived from both the survey and the categorically encoded semi structured interviews, suggests that higher AI and Data Science adoption in NPD processes among Portuguese companies correlates with increased perceived innovation and efficiency.

Regarding Hypothesis H2, cluster analysis, derived from interview data, identified four distinct company profiles with AI and DS adoption ranging from “Established Traditionalists” (no adoption) to “Emerging Tech Enthusiasts” (moderate adoption). However, the Kruskal-Wallis H Test applied across different 'Industry_Sectors' revealed no significant differences in AI and DS adoption rates.

And so, for Hypothesis H2, the evidence suggests that while there are discernible patterns in AI and DS adoption levels among different company profiles, sectoral differences are not markedly supported by the results. Thus, Hypothesis H2, proposing that AI and Data Science adoption in NPD varies across sectors, was not conclusively verified.

Finally, practical applications of AI tools, specifically DALL·E and ChatGPT, were showcased, emphasizing their transformative potential in NPD. The research highlighted the efficiency of these AI tools in market research, achieving results in an hour compared to traditional methods that in general take weeks, therefore supporting the proposed framework.

5.1. Limitations and Future Research

The study acknowledges potential limitations and areas for future exploration. Self-selection bias might influence results due to varying participant familiarity with AI. Social desirability bias could impact participants' responses. While efforts were made to mitigate survey and interview abandonment, some respondents didn't transition to interview.

In the practical application, data scope was restricted to Amazon, possibly introducing biases. ChatGPT's and DALL·E's insights might lack nuance, and their limitations should be balanced with human expertise.

Future research could address these limitations by increasing sample sizes, widening data sources, increasing sample size, refining AI tools, and exploring diverse industry contexts for a comprehensive understanding.

BIBLIOGRAPHY

- Ahmady, G. A., Mehrpour, M., & Nikooravesh, A. (2016). Organizational Structure. *Procedia - Social and Behavioral Sciences*, 230, 455–462.
- Alavi, M., & Leidner, D. E. (2001). Review: Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues. *The Mississippi Quarterly*, 25(1), 107–136.
- Almeida, R. de J. de. (2022). *The impact of intelligent systems on management control of 21st century organizations* (R. J. L. da Costa & R. A. H. Gonçalves, Eds.) [ISCTE Business School]. https://repositorio.iscte-iul.pt/bitstream/10071/25600/1/master_rita_jesus_almeida.pdf
- Alves, A. C., Barbieux, D., Reichert, F. M., Tello-Gamarra, J., & Zawislak, P. A. (2017). Innovation and dynamic capabilities of the firm: Defining an assessment model. *Revista de Administração de Empresas*, 57(3), 232–244.
- Anshari, M., Syafrudin, M., Tan, A., Fitriyani, N. L., & Alas, Y. (2023). Optimisation of Knowledge Management (KM) with Machine Learning (ML) Enabled. *Information. An International Interdisciplinary Journal*, 14(1), 35.
- Araujo, T., Helberger, N., Kruijkemeier, S., & de Vreese, C. H. (2020). In AI we trust? Perceptions about automated decision-making by artificial intelligence. *AI & Society*, 35(3), 611–623.
- Bansal, M., Krizhevsky, A., & Ogale, A. (2018). ChauffeurNet: Learning to Drive by Imitating the Best and Synthesizing the Worst. In *arXiv [cs.RO]*. arXiv. <http://arxiv.org/abs/1812.03079>
- Barreto, I. (2010). Dynamic Capabilities: A Review of Past Research and an Agenda for the Future. *Journal of Management*, 36(1), 256–280.

- BCG. (2023). *AI Is Revolutionizing How Companies Manage Transformations*. Boston Consulting Group. <https://www.bcg.com/publications/2023/artificial-intelligence-strategies-for-leading-business-transformation>
- Bhatti, S. H., Ahmed, A., Ferraris, A., Hirwani Wan Hussain, W. M., & Wamba, S. F. (2022). Big data analytics capabilities and MSME innovation and performance: A double mediation model of digital platform and network capabilities. *Annals of Operations Research*. <https://doi.org/10.1007/s10479-022-05002-w>
- Bhuiyan, N. (2011). A Framework for successful new product development. *Journal of Industrial Engineering and Management*, 4(4), 746–770.
- Bort, J. (2016, October 10). How Lemonade’s founders raised a massive seed round just by talking. *Business Insider*. <https://www.businessinsider.com/lemonade-daniel-schreiber-13-million-just-by-talking-2016-10>
- Brandl, R., & Ellis, C. (2023). *Survey: ChatGPT and AI Content Can people tell the difference?* tooltester.com. <https://www.tooltester.com/en/blog/chatgpt-statistics/>
- Brynjolfsson, E., Rock, D., & Syverson, C. (2017). *Artificial Intelligence and the Modern Productivity Paradox: A Clash of Expectations and Statistics*. <https://papers.ssrn.com/abstract=3065841>
- Cao, L., Yang, Q., & Yu, P. S. (2021). Data science and AI in FinTech: an overview. *International Journal of Data Science and Analytics*, 12(2), 81–99.
- Chaochotechuan, P., Daneshgar, F., & Sindakis, S. (2015). Innovation Strategies of New Product Development (NPD). *International Journal of Knowledge and Systems Science*, 6(2), 57–75.
- Chesbrough, H. W. (2006). *Open Business Models: How To Thrive In The New Innovation Landscape* (Vol. 17). Wiley.

- Chung, S.-H., Ma, H.-L., Hansen, M., & Choi, T.-M. (2020). Data science and analytics in aviation. *Transportation Research Part E: Logistics and Transportation Review*, *134*, 101837.
- Collins, D. (2003). Pretesting survey instruments: an overview of cognitive methods. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation*, *12*(3), 229–238.
- Conboy, K., Mikalef, P., Dennehy, D., & Krogstie, J. (2020). Using business analytics to enhance dynamic capabilities in operations research: A case analysis and research agenda. *European Journal of Operational Research*, *281*(3), 656–672.
- Cooper, R. G. (1990). Stage-gate systems: A new tool for managing new products. *Business Horizons*, *33*(3), 44–54.
- Cooper, R. G. (2006). The seven principles of the latest Stage-Gate method add up to a streamlined, new-product idea-to-launch process. *Stage-Gate International*.
https://www.five-is.com/wp-content/uploads/2013/12/Cooper_2006_Formula_for_Success.pdf
- Cooper, R. G. (2008). Perspective: The Stage-Gate® Idea-to-Launch Process—Update, What’s New, and NexGen Systems*. *Journal of Product Innovation Management*, *25*(3), 213–232.
- Cooper, R. G. (2011). *Winning at new products: Creating value through innovation*. Basic Books.
- Cooper, R. G. (2014). What’s Next?: After Stage-Gate. *Research-Technology Management*, *57*(1), 20–31.
- Cooper, R. G. (2016). Agile–Stage-Gate Hybrids. *Research-Technology Management*, *59*(1), 21–29.

- Cooper, R. G. (2019). The drivers of success in new-product development. *Industrial Marketing Management*, 76, 36–47.
- Cooper, R. G. (2021). Accelerating innovation: Some lessons from the pandemic. *The Journal of Product Innovation Management*, 38(2), 221–232.
- Coussement, K., & Benoit, D. F. (2021). Interpretable data science for decision making. *Decision Support Systems*, 150, 113664.
- Dall·e. (2022, March 25). <https://openai.com/dall-e-2>
- Damioli, G., Van Roy, V., & Vertesy, D. (2021). The impact of artificial intelligence on labor productivity. *Eurasian Business Review*, 11(1), 1–25.
- Davenport, T. H. (2018). *The AI Advantage: How to Put the Artificial Intelligence Revolution to Work*. MIT Press.
- Davenport, T. H., & Patil, D. J. (2012). Data scientist: the sexiest job of the 21st century. *Harvard Business Review*, 90(10), 70–76, 128.
- Davenport, T. H., Ronanki, R., & Others. (2018). Artificial intelligence for the real world. *Harvard Business Review*, 96(1), 108–116.
- De Giovanni, P. (2019). Digital Supply Chain through Dynamic Inventory and Smart Contracts. *Science in China, Series A: Mathematics*, 7(12), 1235.
- De Luca, L. M., Herhausen, D., Troilo, G., & Rossi, A. (2021). How and when do big data investments pay off? The role of marketing affordances and service innovation. *Journal of the Academy of Marketing Science*, 49(4), 790–810.
- Del Vecchio, P., Di Minin, A., Petruzzelli, A. M., Panniello, U., & Pirri, S. (2018). Big data for open innovation in SMEs and large corporations: Trends, opportunities, and challenges. *Creativity and Innovation Management*, 27(1), 6–22.

- Deloitte. (2018). *Knowledge Management & Big Data: Making Smart Enterprise a Reality*.
 Deloitte. <https://www2.deloitte.com/content/dam/Deloitte/in/Documents/technology-media-telecommunications/in-tmt-knowledge-management-and-big-data-noexp.pdf>
- Dessi, D., Osborne, F., Reforgiato Recupero, D., Buscaldi, D., & Motta, E. (2021). Generating knowledge graphs by employing Natural Language Processing and Machine Learning techniques within the scholarly domain. *Future Generations Computer Systems: FGCS*, *116*, 253–264.
- Duncan, R. (1979). What is the right organization structure? Decision tree analysis provides the answer. *Organizational Dynamics*, *7*(3), 59–80.
- Edwards, K., Cooper, R. G., Vedsmand, T., & Nardelli, G. (2019). Evaluating the Agile-Stage-Gate Hybrid Model: Experiences From Three SME Manufacturing Firms. *International Journal of Innovation and Technology Management*, *16*(08), 1950048.
- Eisenhardt, K. M., & Martin, J. A. (2000). Dynamic capabilities: what are they? *Strategic Management Journal*, *21*(10–11), 1105–1121.
- Eppinger, S., & Ulrich, K. (2011). *Product Design and Development*. McGraw-Hill Education.
- European Commission. (2018). *Artificial intelligence: An European perspective*. European Commission. Joint Research Centre. <https://doi.org/10.2760/936974>
- Fernandez, A. (2019). *Artificial Intelligence in Financial Services*.
<https://doi.org/10.2139/ssrn.3366846>
- Füller, J., Hutter, K., Wahl, J., Bilgram, V., & Tekic, Z. (2022). How AI revolutionizes innovation management – Perceptions and implementation preferences of AI-based innovators. *Technol. Forecast. Soc. Change*, *178*, 121598.
- Galletta, A. (2013). Mastering the semi-structured interview and beyond: From research design to analysis and publication. *Qualitative Studies in Psychology*, *245*.
<https://doi.org/10.18574/nyu/9780814732939.001.0001>

- Gambhir, M., & Gupta, V. (2017). Recent automatic text summarization techniques: a survey. *Artificial Intelligence Review*, 47(1), 1–66.
- Gao, J., & Sarwar, Z. (2022). How do firms create business value and dynamic capabilities by leveraging big data analytics management capability? *Information Technology & Management*, 1–22.
- Ghasemaghaei, M., & Calic, G. (2020). Assessing the impact of big data on firm innovation performance: Big data is not always better data. *Journal of Business Research*, 108, 147–162.
- Gomes, A. I. V. (2022). *Artificial Intelligence empowerment in managerial decision-making* (D. J. Crispim, Ed.) [Escola de Economia e Gestão da Universidade do Minho]. <https://repositorium.sdum.uminho.pt/handle/1822/81745>
- Gomez-Uribe, C. A., & Hunt, N. (2016). The Netflix Recommender System: Algorithms, Business Value, and Innovation. *ACM Trans. Manage. Inf. Syst.*, 6(4), 1–19.
- Greenberg, E., Hirt, M., & Smit, S. (2017). *The global forces inspiring a new narrative of progress*. McKinsey & Company. <https://www.mckinsey.com/capabilities/strategy-and-corporate-finance/our-insights/the-global-forces-inspiring-a-new-narrative-of-progress>
- Grossi, V., Giannotti, F., Pedreschi, D., Manghi, P., Pagano, P., & Assante, M. (2021). Data science: a game changer for science and innovation. *International Journal of Data Science and Analytics*, 11(4), 263–278.
- Gunday, G., Ulusoy, G., Kilic, K., & Alpkan, L. (2011). Effects of innovation types on firm performance. *International Journal of Production Economics*, 133(2), 662–676.
- Gupta, M., & George, J. F. (2016). Toward the development of a big data analytics capability. *Information & Management*, 53(8), 1049–1064.

- Haynes, G. A. (2009). Testing the boundaries of the choice overload phenomenon: The effect of number of options and time pressure on decision difficulty and satisfaction. *Psychology & Marketing, 26*(3), 204–212.
- Heigham, J., & Croker, R. (2009). *Qualitative Research in Applied Linguistics: A Practical Introduction*. Springer.
- Hekkert, M. P., Suurs, R. A. A., Negro, S. O., Kuhlmann, S., & Smits, R. E. H. M. (2007). Functions of innovation systems: A new approach for analysing technological change. *Technological Forecasting and Social Change, 74*(4), 413–432.
- Helfat, C. E., & Winter, S. G. (2011). Untangling Dynamic and Operational Capabilities: Strategy for the (N)Ever-Changing World. *Strategic Management Journal, 32*(11), 1243–1250.
- Hendershott, T., Zhang, X. (michael), Zhao, J. L., & Zheng, Z. (eric). (2021). FinTech as a Game Changer: Overview of Research Frontiers. *Information Systems Research, 32*(1), 1–17.
- Hidalgo, A., & Albors, J. (2008). Innovation management techniques and tools: a review from theory and practice. *R and D Management, 38*(2), 113–127.
- Hoonsopon, D., & Ruenrom, G. (2012). The Impact of Organizational Capabilities on the Development of Radical and Incremental Product Innovation and Product Innovation Performance. *Journal of Managerial Issues, 24*(3), 250–276.
- Huang, M.-H., & Rust, R. T. (2021). A strategic framework for artificial intelligence in marketing. *Journal of the Academy of Marketing Science, 49*(1), 30–50.
- Hult, G. T. M., Hurley, R. F., & Knight, G. A. (2004). Innovativeness: Its antecedents and impact on business performance. *Industrial Marketing Management, 33*(5), 429–438.
- Jacoby, J., & Matell, M. S. (1971). Three-Point Likert Scales Are Good Enough. *JMR, Journal of Marketing Research, 8*(4), 495–500.

- Jarrahi, M. H., Askay, D., Eshraghi, A., & Smith, P. (2023). Artificial intelligence and knowledge management: A partnership between human and AI. *Business Horizons*, 66(1), 87–99.
- Jaruzelski, B., Chwalik, R., & Goehle, B. (2018). *The Global Innovation 1000 study*. PwC. <https://www.strategyand.pwc.com/gx/en/insights/innovation1000.html>
- Johnson, P. C., Laurell, C., Ots, M., & Sandström, C. (2022). Digital innovation and the effects of artificial intelligence on firms' research and development – Automation or augmentation, exploration or exploitation? *Technological Forecasting and Social Change*, 179, 121636.
- Joloudari, J. H., Mojrian, S., Saadatfar, H., Nodehi, I., Fazl, F., Shirkharkolaie, S. K., Alizadehsani, R., Dipu Kabir, H. M., Tan, R.-S., & Rajendra Acharya, U. (2022). The state-of-the-art review on resource allocation problem using artificial intelligence methods on various computing paradigms. In *arXiv [cs.AI]*. arXiv. <http://arxiv.org/abs/2203.12315>
- Jones, G. R. (2013). *Organizational Theory, Design and Change, Global Edition*. Pearson Education.
- Joshi, K. D., Chi, L., Datta, A., & Han, S. (2010). Changing the Competitive Landscape: Continuous Innovation Through IT-Enabled Knowledge Capabilities. In *Information Systems Research* (Vol. 21, Issue 3, pp. 472–495). <https://doi.org/10.1287/isre.1100.0298>
- Kabaivanov, S., & Markovska, V. (2021). Artificial intelligence in real estate market analysis. *THERMOPHYSICAL BASIS OF ENERGY TECHNOLOGIES (TBET 2020)*. THERMOPHYSICAL BASIS OF ENERGY TECHNOLOGIES (TBET 2020), Tomsk, Russia. <https://doi.org/10.1063/5.0041806>
- Kahn, K. B. (2006). *New Product Forecasting: An Applied Approach*. M.E. Sharpe.

- Kakatkar, C., Bilgram, V., & Füller, J. (2020). Innovation analytics: Leveraging artificial intelligence in the innovation process. *Business Horizons*, 63(2), 171–181.
- Khurana, D., Koli, A., Khatter, K., & Singh, S. (2023). Natural language processing: state of the art, current trends and challenges. *Multimedia Tools and Applications*, 82(3), 3713–3744.
- Kim, J.-H., & Lee, Y.-G. (2019). Change of Data-Driven Drug Design Trends Through Patent Analysis. *Processes*, 7(8), 492.
- Lamba, K., & Prakash, S. S. (2018). Modeling big data enablers for operations and supply chain management. *The International Journal of Logistics Management*, 29(2), 629–658.
- Larson, D. B., & Boland, G. W. (2019). Imaging Quality Control in the Era of Artificial Intelligence. *Journal of the American College of Radiology: JACR*, 16(9 Pt B), 1259–1266.
- Larson, M. L. (1991). *Translation: Theory and Practice, Tension and Interdependence*. John Benjamins Publishing.
- Laursen, K., & Salter, A. (2006). Open for innovation: the role of openness in explaining innovation performance among U.K. manufacturing firms. *Strategic Management Journal*, 27(2), 131–150.
- Lemonade. (2017). *The Empathetic Bot*. <https://www.lemonade.com/blog/the-empathetic-bot/>
- Leung, C. K., Pazdor, A. G. M., & Souza, J. (2021). Explainable Artificial Intelligence for Data Science on Customer Churn. *2021 IEEE 8th International Conference on Data Science and Advanced Analytics (DSAA)*, 1–10.
- Liao, C., Wang, H.-Y., Chuang, S.-H., Shih, M.-L., & Liu, C.-C. (2010). Enhancing knowledge management for R&D innovation and firm performance: An integrative view. *African Journal of Business Management*, 4(14), 3026.

- Lumen Learning. (2018, December). *Functional Areas of Business*.
<https://courses.lumenlearning.com/wm-introductiontobusiness/chapter/reading-functional-areas-of-business/>
- Luusua, A., & Ylipulli, J. (2020). Artificial Intelligence and Risk in Design. *Proceedings of the 2020 ACM Designing Interactive Systems Conference*, 1235–1244.
- Mak, K.-K., & Pichika, M. R. (2019). Artificial intelligence in drug development: present status and future prospects. *Drug Discovery Today*, 24(3), 773–780.
- Manning, C. (2020). *Artificial Intelligence Definitions*. Stanford University.
<https://hai.stanford.edu/sites/default/files/2020-09/AI-Definitions-HAI.pdf>
- Marina, du P. (2007). The role of knowledge management in innovation. *Journal of Knowledge Management*, 11(4), 20–29.
- Marr, B. (2018, May 25). Stitch Fix: The Amazing Use Case Of Using Artificial Intelligence In Fashion Retail. *Forbes Magazine*.
<https://www.forbes.com/sites/bernardmarr/2018/05/25/stitch-fix-the-amazing-use-case-of-using-artificial-intelligence-in-fashion-retail/>
- Mayo Clinic. (2019, September 10). *Mayo Clinic selects Google as strategic partner for health care innovation, cloud computing*.
<https://newsnetwork.mayoclinic.org/discussion/mayo-clinic-selects-google-as-strategic-partner-for-health-care-innovation-cloud-computing/>
- McAfee, A., & Brynjolfsson, E. (2012). Big data: the management revolution. *Harvard Business Review*, 90(10), 60–66, 68, 128.
- McBeath, A. (2020). Doing Quantitative Research with a Survey. In S. Bager-Charleson & A. McBeath (Eds.), *Enjoying Research in Counselling and Psychotherapy: Qualitative, Quantitative and Mixed Methods Research* (pp. 175–193). Springer International Publishing.

- Mccarthy, J. (2007). *What is artificial intelligence?*
<https://www.diochnos.com/about/McCarthyWhatisAI.pdf>
- McGrath, R. G. (2013). *The End of Competitive Advantage: How to Keep Your Strategy Moving as Fast as Your Business*. Harvard Business Review Press.
- McKinsey. (2018). *Notes from the AI frontier: Applications and value of deep learning*. McKinsey Global Institute. <https://www.mckinsey.com/featured-insights/artificial-intelligence/notes-from-the-ai-frontier-applications-and-value-of-deep-learning>
- McKinsey. (2020). *Global survey: The state of AI in 2020*. McKinsey. <https://www.mckinsey.com/capabilities/quantumblack/our-insights/global-survey-the-state-of-ai-in-2020>
- McKinsey. (2022). *The state of AI in 2022—and a half decade in review*. McKinsey. <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2022-and-a-half-decade-in-review>
- McKinsey. (2023). *The economic potential of generative AI: The next productivity frontier*. McKinsey. <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-AI-the-next-productivity-frontier>
- Mikalef, P., Boura, M., Lekakos, G., & Krogstie, J. (2019a). Big data analytics capabilities and innovation: The mediating role of dynamic capabilities and moderating effect of the environment. *British Journal of Management*, 30(2), 272–298.
- Mikalef, P., Boura, M., Lekakos, G., & Krogstie, J. (2019b). Big data analytics and firm performance: Findings from a mixed-method approach. *Journal of Business Research*, 98, 261–276.
- Mikalef, P., Krogstie, J., Pappas, I. O., & Pavlou, P. (2020). Exploring the relationship between big data analytics capability and competitive performance: The mediating roles of dynamic and operational capabilities. *Information & Management*, 57(2), 103169.

- Mikalef, P., Pappas, I. O., Krogstie, J., & Giannakos, M. (2018). Big data analytics capabilities: a systematic literature review and research agenda. *Information Systems and E-Business Management*, 16(3), 547–578.
- Mikalef, P., van de Wetering, R., & Krogstie, J. (2021). Building dynamic capabilities by leveraging big data analytics: The role of organizational inertia. *Information & Management*, 58(6), 103412.
- Morariu, C., Morariu, O., Răileanu, S., & Borangiu, T. (2020). Machine learning for predictive scheduling and resource allocation in large scale manufacturing systems. *Computers in Industry*, 120, 103244.
- Mosavi, A., Faghan, Y., Ghamisi, P., Duan, P., Ardabili, S. F., Salwana, E., & Band, S. S. (2020). Comprehensive Review of Deep Reinforcement Learning Methods and Applications in Economics. *Science in China, Series A: Mathematics*, 8(10), 1640.
- Nerdynav. (2023, April). *91 Important ChatGPT Statistics & User Numbers In April 2023 (GPT-4, Plugins Update) - Nerdy Nav*. <https://nerdynav.com/chatgpt-statistics/>
- Niranjan, K., Narayana, K. S., & Rao, M. V. A. L. N. (2021). Role of Artificial Intelligence in Logistics and Supply Chain. *2021 International Conference on Computer Communication and Informatics (ICCCI)*, 1–3.
- NVIDIA. (2017, March 30). *NVIDIA DRIVE Solutions*. <https://developer.nvidia.com/drive>
- NVIDIA. (2020). *NVIDIA Announces Financial Results for Second Quarter Fiscal 2021*. NVIDIA. <https://nvidianews.nvidia.com/news/nvidia-announces-financial-results-for-second-quarter-fiscal-2021>
- Oates, B. J. (2006). *Researching Information Systems and Computing*. SAGE.
- Pacheco Gutierrez, D., Folkmann, L. M., Tribukait, H., & Roch, L. M. (2023). How to accelerate R&D and optimize experiment planning with machine learning and data science. *Chimia*, 77(1/2), 7.

- Paschen, J., Kietzmann, J., & Christian, K. T. (2019). Artificial intelligence (AI) and its implications for market knowledge in B2B marketing. *Journal of Business & Industrial Marketing*, 34(7), 1410–1419.
- P&G. (2020). *Oral-B iO™ Unveiled at Consumer Electronics Show, Marking the Brand's Most Innovative Power Toothbrush to Date*. <https://news.pg.com/news-releases/news-details/2020/Oral-B-iO-Unveiled-at-Consumer-Electronics-Show-Marking-the-Brands-Most-Innovative-Power-Toothbrush-to-Date/default.aspx>
- P&G. (2021). *P&G Announces Fiscal Year 2021 Second Quarter Results*. <https://news.pg.com/news-releases/news-details/2021/PG-Announces-Fiscal-Year-2021-Second-Quarter-Results/default.aspx>
- Pisano, G. P. (2015). You need an innovation strategy. *Harvard Business Review*, 93(6), 44–54.
- Pisano, G. P. (2017). Toward a prescriptive theory of dynamic capabilities: connecting strategic choice, learning, and competition. *Industrial and Corporate Change*, 26(5), 747–762.
- Pisoni, G., Molnár, B., & Tarcsi, Á. (2021). Data Science for Finance: Best-Suited Methods and Enterprise Architectures. *Applied System Innovation*, 4(3), 69.
- Pomerol, J.-C. (1997). Artificial intelligence and human decision making. *European Journal of Operational Research*, 99(1), 3–25.
- Porter, M. E. (1985). *Competitive advantage : creating and sustaining superior performance* [Xviii, 557 p. : ill. ; 25 cm.]. Free Press ; Collier Macmillan.
- Power, D. J. (2016). Data science: supporting decision-making. *Journal of Decision Systems*, 25(4), 345–356.
- Provost, F., & Fawcett, T. (2013). Data Science and its Relationship to Big Data and Data-Driven Decision Making. *Big Data*, 1(1), 51–59.

- Ransbotham, S. K. (2017). Analytics as a Source of Business Innovation. In *MIT Sloan Management Review; Cambridge volume* (Vol. 58, Issue 3, p. 0).
<https://search.proquest.com/openview/0f389efc750dac8a49eed0aaa00b1f28/1?pq-origsite=gscholar&cbl=26142>
- Riascos, R., Levy, L., Stjepandić, J., & Fröhlich, A. (2015). Digital Mock-up. In J. Stjepandić, N. Wognum, & W. J. C. Verhagen (Eds.), *Concurrent Engineering in the 21st Century: Foundations, Developments and Challenges* (pp. 355–388). Springer International Publishing.
- Ringel, M., Manly, J., & Möller, C. (2018). *The Most Innovative Companies 2018*. Boston Consulting Group. <https://www.bcg.com/publications/collections/most-innovative-companies-2018>
- Rosário, A., Moniz, L. B., & Cruz, R. (2021). Data Science Applied to Marketing: A Literature Review. *Journal of Information Science and Engineering*, 37, 1067–1081.
- Rosenbusch, N., Brinckmann, J., & Bausch, A. (2011). Is innovation always beneficial? A meta-analysis of the relationship between innovation and performance in SMEs. *Journal of Business Venturing*, 26(4), 441–457.
- Runeson, P., & Höst, M. (2009). Guidelines for conducting and reporting case study research in software engineering. *Empirical Software Engineering*, 14(2), 131–164.
- Russo, I., Confente, I., & Antonio, B. (2015, September 3). Using big Data in the Supply Chain Context: Opportunities and Challenges. *The 16th European Conference on Knowledge Management ECKM 2015*. <http://dx.doi.org/>
- Saura, José Ramón. (2020). Using Data Sciences in Digital Marketing: Framework, methods, and performance metrics. *Journal of Innovation & Knowledge*, 6(2).
<https://doi.org/10.1016/j.jik.2020.08.001>

- Saura, Jose Ramon, Palacios-Marqués, D., & Ribeiro-Soriano, D. (2023). Digital marketing in SMEs via data-driven strategies: Reviewing the current state of research. *Journal of Small Business Management*, 61(3), 1278–1313.
- Schumpeter, J. A. (1934). *The Theory of Economic Development: An Inquiry into Profits, Capital, Credit, Interest, and the Business Cycle*.
<https://papers.ssrn.com/abstract=1496199>
- Sharma, A. K., Goel, N., Rajput, J., & Bilal, M. (2020). An Intelligent Model For Predicting the Sales of a Product. *2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, 341–345.
- Stitch Fix. (2018). *Stitch Fix Company Fact Sheet About Stitch Fix*. Stitch Fix Company .
https://newsroom.stitchfix.com/wp-content/uploads/2016/09/StitchFix_FactSheet-15.pdf
- Stockemer, D. (2019). Constructing a Survey. In D. Stockemer (Ed.), *Quantitative Methods for the Social Sciences: A Practical Introduction with Examples in SPSS and Stata* (pp. 37–56). Springer International Publishing.
- Suoniemi, S., Meyer-Waarden, L., Munzel, A., Zablah, A. R., & Straub, D. (2020). Big data and firm performance: The roles of market-directed capabilities and business strategy. *Information & Management*, 57(7), 103365.
- Teece, D. J. (2007). Explicating dynamic capabilities: the nature and microfoundations of (sustainable) enterprise performance. *Strategic Management Journal*, 28(13), 1319–1350.
- Teece, D. J. (2018). Dynamic capabilities as (workable) management systems theory. *Journal of Management & Organization*, 24(3), 359–368.
- Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic Capabilities and Strategic Management. *Strategic Management Journal*, 18(7), 509–533.

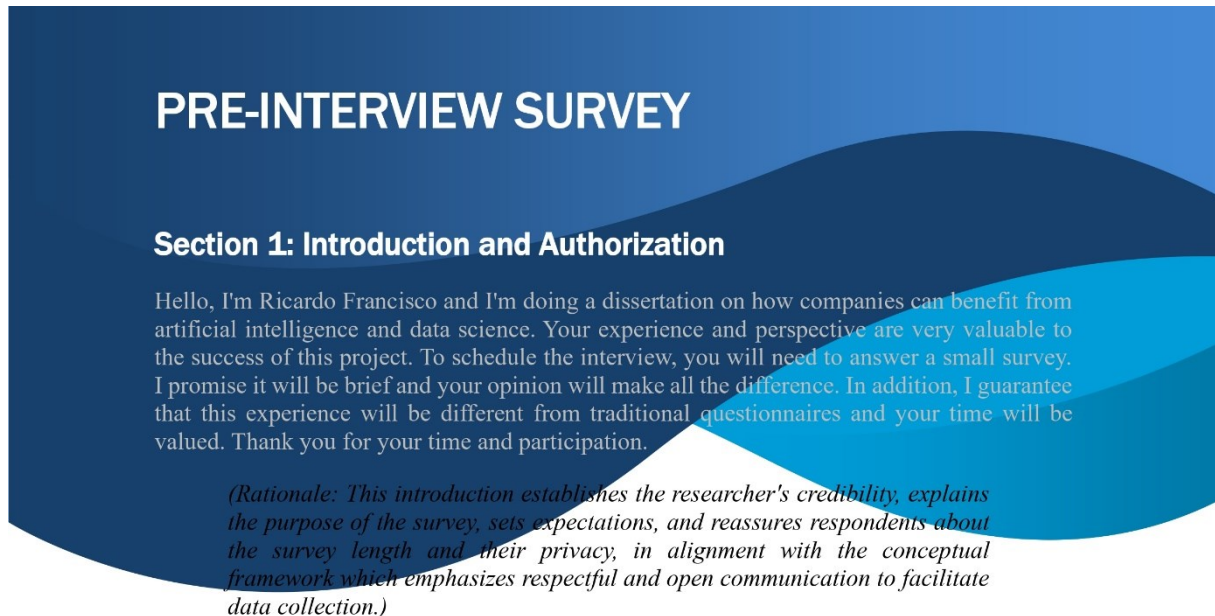
- Temple, B., & Young, A. (2004). Qualitative Research and Translation Dilemmas. *Qualitative Research: QR*, 4(2), 161–178.
- Tidd, J., & Bessant, J. (2009). *Managing innovation : integrating technological, market and organizational change*. John Wiley.
- Tiguint, B., & Hossari, H. (2020). *Big Data Analytics and Artificial Intelligence: A Meta-Dynamic Capability Perspective*. <https://doi.org/10.2139/ssrn.3709077>
- UNESCO. (2021). *Understanding the Impact of Artificial Intelligence on Skills Development. Education 2030*. UNESCO. <https://eric.ed.gov/?id=ED612439>
- Vachhani, H., Obiadat, M. S., Thakkar, A., Shah, V., Sojitra, R., Bhatia, J., & Tanwar, S. (2020). Machine Learning Based Stock Market Analysis: A Short Survey. *Innovative Data Communication Technologies and Application*, 12–26.
- van Teijlingen, E., & Hundley, V. (2002). The importance of pilot studies. *Nursing Standard: Official Newspaper of the Royal College of Nursing*, 16(40), 33–36.
- Verganti, R., Vendraminelli, L., & Iansiti, M. (2020). Innovation and design in the age of artificial intelligence. *The Journal of Product Innovation Management*, 37(3), 212–227.
- Verlinden, J. C., de Smit, A., Peeters, A. W. J., & van Gelderen, M. H. (2003). Development of a flexible augmented prototyping system. *Journal of WSCG*. <https://dspace5.zcu.cz/handle/11025/1627>
- Veryzer, R. W., & Borja de Mozota, B. (2005). The impact of user-oriented design on new product development: An examination of fundamental relationships. *The Journal of Product Innovation Management*, 22(2), 128–143.
- Waller, M. A., & Fawcett, S. E. (2013). Data science, predictive analytics, and big data: A revolution that will transform supply chain design and management. *Journal of Business Logistics*, 34(2), 77–84.

- Walmart. (2016, September 19). *Walmart Completes Acquisition of Jet.com, Inc.* Corporate - US. <https://corporate.walmart.com/newsroom/2016/09/19/walmart-completes-acquisition-of-jet-com-inc>
- Wamba, S. F., Gunasekaran, A., Akter, S., Ren, S. J.-F., Dubey, R., & Childe, S. J. (2017). Big data analytics and firm performance: Effects of dynamic capabilities. *Journal of Business Research*, 70, 356–365.
- Wang, Z., & Wang, N. (2012). Knowledge sharing, innovation and firm performance. *Expert Systems with Applications*, 39(10), 8899–8908.
- Whiting, L. S. (2008). Semi-structured interviews: guidance for novice researchers. *Nursing Standard: Official Newspaper of the Royal College of Nursing*, 22(23), 35–40.
- Wiener, M., Saunders, C., & Marabelli, M. (2020). Big-data business models: A critical literature review and multiperspective research framework. *Journal of Information Technology Impact*, 35(1), 66–91.
- Yin, C., & Han, J. (2021). Dynamic pricing model of E-commerce platforms based on deep reinforcement learning. *Computer Modeling in Engineering & Sciences: CMES*, 127(1), 291–307.
- Zahra, S. A., Sapienza, H. J., & Davidsson, P. (2006). Entrepreneurship and dynamic capabilities: A review, model and research agenda. *The Journal of Management Studies*, 43(4), 917–955.
- Zia, S., & Shafiq, M. (2016). Innovation and knowledge management: A literature review and research framework. *Journal of Quality and Technology Management*, 13(1). https://www.researchgate.net/profile/Muhammad-Shafiq-32/publication/335541380_INNOVATION_AND_KNOWLEDGE_MANAGEMENT_A_LITERATURE_REVIEW_AND_RESEARCH_FRAMEWORK/links/5d6c4bb09

2851c8538858452/INNOVATION-AND-KNOWLEDGE-MANAGEMENT-A-
LITERATURE-REVIEW-AND-RESEARCH-FRAMEWORK.pdf

APPENDIX I - DATA COLLECTION

Survey Questions



PRE-INTERVIEW SURVEY

Section 1: Introduction and Authorization

Hello, I'm Ricardo Francisco and I'm doing a dissertation on how companies can benefit from artificial intelligence and data science. Your experience and perspective are very valuable to the success of this project. To schedule the interview, you will need to answer a small survey. I promise it will be brief and your opinion will make all the difference. In addition, I guarantee that this experience will be different from traditional questionnaires and your time will be valued. Thank you for your time and participation.

(Rationale: This introduction establishes the researcher's credibility, explains the purpose of the survey, sets expectations, and reassures respondents about the survey length and their privacy, in alignment with the conceptual framework which emphasizes respectful and open communication to facilitate data collection.)

Question I1

Can we start scheduling the interview? Your contribution is very important to me.

(Question Rationale: This question seeks to confirm the participant's willingness to proceed with the interview and participation in the survey respecting ethical considerations in research.)

Answer Options: "Yes", "No"

(Response Type and Rationale: Dichotomous type question which provides clear yes or no options, enabling a quick understanding of the participant's willingness to proceed.)

Section 2: Participant's Identification

Question 01

What is your email?

(Question Rationale: This question collects the participant's email address, ensuring future correspondence and serving as an identifier for data analysis, hence allowing to cross data against the semi-structured interviewers.)

Answer Options: Open-ended response.

(Response Type and Rationale: This question requires a direct input from the respondent.)

Figure 25: Survey Guide

Section 3: Demographic Information

Question 02

What is your gender?

(This question collects demographic information for segmenting and understanding the responses. It is critical for investigating potential gender differences in AI and Data Science adoption, relevant to Hypothesis H1.)

Answer Options: "Male", "Female" or "Non-Binary"

Response Type and Rationale: Pre-defined categories provide a simple way to quantify the responses and allow for easy comparison and segmentation of responses.

Question 03

What is the number of trips you have made around the sun? (i.e., what is your age?)

(Question Rationale: This question collects demographic data in a humorous, non-intrusive way, encouraging response and reducing non-response bias. Age may correlate with perspectives on AI and Data Science, linking to Hypothesis H1.)

Answer Options: "25-34 years", "35-44 years", "45-54 years", "55-64 years", "65 years or more"

(Response Type and Rationale: This question uses an interval type response, which allows for demographic segmentation and identification of patterns or trends based on age groups.)

Encouraging Statement 1

Age is just a number, but it's very important for my research!

(Rationale: Injecting humour and respect into the interaction can contribute to a more enjoyable and engaging experience for the respondent. This can increase response rates by creating a positive, engaging atmosphere which can reduce the likelihood of non-response bias. This approach recognizes the importance of the respondent's experience and seeks to make the process of survey-taking less burdensome and more enjoyable.)

Section 4: Professional & Company Details

Question 04

What is your role within the company?

(Question Rationale: This question aims to understand the respondent's role within the company to assess their perspective and potentially segment responses based on roles. Different roles may interface differently with AI and

Data Science, relevant for the investigation under Hypothesis H1. Furthermore, this acceptance of roles justifies the assumption of functional firm organization which, in turn, forms the core of the Conceptual Framework)

Answer Options: "CEO; Administrator; Project Manager; Human Resources Manager, Financial Manager; Marketing Manager, Operations Manager, Sales Manager, IT Manager; Analyst; AI and/or Data Science Specialist; Customer Service; R&D; Other (please specify)"

(Response Type and Rationale: Pre-defined categories allow for comparison of responses between different roles. An open-ended option ("Other") is included to ensure all possible responses are captured.)

Encouraging Statement 3

Thank you very much! Understanding the spectrum of companies participating in the study will allow me to gain valuable insights about the impact of AI and data science across different sectors and sizes. Your contribution to science is very much appreciated.

(Rationale: This statement is designed to motivate the respondent and reinforce the value of their participation. It can enhance respondent engagement and satisfaction, potentially reducing non-response bias in this intrusive question. This strategy aligns with the ethical considerations of conducting research, fostering a positive respondent experience.)

Question 05

How many years have you been with the company?

(Question Rationale: This question aims to understand the tenure of the respondent within the company. The responses can help correlate the impact of tenure on the perception and use of AI and Data Science in the organization. The length of tenure could influence employees' exposure to, and understanding of, these technologies, relevant to Hypothesis H1.)

Answer Type: Open-ended response.

(Response Type and Rationale: The answer is expected as an integer to allow for calculation of metrics such as averages.)

Question 06

How old is the company or the parent group?

(Question Rationale: This question seeks to understand the company's age, potentially impacting its adoption and use of AI and data science. Older companies may have different attitudes and capacities for technological innovation compared to newer ones, relevant to Hypothesis H2.)

Answer Options: "Less than 1 year; 1 to 5 years; 6 to 10 years; 11 to 20 years; 21 to 50 years; More than 50 years, up to 100 years; More than 100 years"

(Response Type and Rationale: Pre-defined categories enable the grouping of companies into different bins, simplifying the comparison and analysis process.)

Question 07

What is the size of the company or the parent group?

(Question Rationale: This question aims to understand the size of the organization, which could influence its resources and capabilities to adopt and use AI and data science, which aligns with the conceptual framework's consideration of organizational resources.)

Answer Options: “Small (1-50 employees); Medium (51-250 employees); Large (251-1000 employees); Very large (more than 1000 employees)”

(Response Type and Rationale: The pre-defined categories allow for comparison of responses across different company sizes.)

Section 5: Industry Specifics

Question 08

What is the industry sector?

(Question Rationale: This question aims to understand the industry context of the respondent's organization. Different industries may have different levels of AI and data science adoption. This information is key to test Hypothesis 2, which states that the adoption and impact of AI and data science in the NPD process can vary across sectors. The data gathered will contribute to a nuanced understanding of the interaction between the industry sector and AI adoption, revealing how context can shape technological implementation and use. This question is pivotal to validate the context-sensitive aspect of the conceptual framework.)

Answer Options: “Agriculture; Construction; Education; Energy; Financial; Health; Tourism; Information Technology; Manufacturing; Retail; Transport; Logistics; Insurance; Other (please specify)”

(Response Type and Rationale: Pre-defined categories ensure responses can be easily compared across different industry sectors. An open-ended option ("Other") is included to ensure all possible responses are captured.)

Encouraging Statement 4

I appreciate your willingness to participate in this investigation and contribute to a greater understanding of how AI and data science could affect different industries.

(Rationale: This statement provides supportive feedback to the participant, emphasizing the importance of their contribution. It's intended to maintain participant engagement and motivation, supporting a positive survey-taking experience. It underscores the importance of understanding the effects of AI and data science across different industries.)

Section 6: AI and Data Science in the Organization

Question 09

How familiar are you with AI and/or data science techniques?

(Question Rationale: This question seeks to understand the respondent's familiarity with AI and data science, which could impact their responses and perceptions about the use of these technologies in their organization. It is essential to test Hypothesis 1, which suggests that higher levels of AI and data science adoption in the NPD process will indicate a higher level of perceived innovation and efficiency. The respondent's familiarity with these technologies can provide insights into the organization's capacity and readiness to implement them, hence influencing the level of adoption.)

Answer Options: “0 – Unfamiliar; 1 - Slightly familiar; 2 - Moderately familiar; 3 - Very familiar”

(Response Type and Rationale: The ordinal scale captures different degrees of familiarity, allowing for a nuanced understanding of the respondent's level of knowledge and experience with AI and data science.)

Fun fact

A study published in the International Journal of Human Resource Management found that employees who worked with AI reported higher job satisfaction, higher levels of focus, and better work-life balance.

(Rationale: Offering a fun fact can serve to break up the monotony of the survey-taking process, potentially increasing respondent engagement and retention. The fact chosen is also related to the survey's topic, subtly reinforcing the relevance and importance of the respondent's contribution. This approach can help to reduce respondent fatigue and maintain interest throughout the survey, thus reducing “non-response”, “abandonment”, or “dissent” biases.)

Question 10

How interested is your company in adopting AI and/or Data Science?

(Rationale: This question aims to gauge the company's interest level in AI and/or Data Science, offering insights into its readiness and intention to adopt these technologies. The responses can help identify potential barriers that might hinder this adoption, which contributes to understanding how prepared companies are for integrating such innovations. Understanding this interest level can also shed light on the company's strategic priorities and openness to technological advancement, both of which are key for effectively utilizing AI and Data Science.)

Answer Options: “0 - No Interest, 1 - Low Interest, 2 - Moderate Interest, 3 - High Interest”

(Response Type and Rationale: The ordinal scale captures different degrees of interest, allowing for a nuanced understanding of the respondent's company's stance towards AI and data science.)

Question 11

What is the level of AI and/or Data Science adoption in your product/service development process?

(Rationale: This question seeks to understand the extent to which AI and Data Science have been adopted in the company's product or service development process. However, it goes beyond merely assessing the level of adoption. The responses to this question can reveal patterns in how these technologies are utilized within the organization's development process. Such information is crucial to validate the hypothesis about AI and Data Science's impact on innovation and efficiency. Through these insights, we can identify trends and correlations between AI and Data Science adoption and innovative practices within various business sectors.)

Answer Options: “0 - No Adoption, 1 - Low Adoption, 2 - Moderate Adoption, 3 - High Adoption”

(Response Type and Rationale: The ordinal scale allows for capturing varying levels of adoption, providing a nuanced understanding of how extensively the respondent's company has integrated AI and data science into its product/service development processes.)

Question 11.01 (Follow-up to Question 11)

If the company uses AI and/or Data Science in product/service development, please specify what types of products.

(Rationale: The rationale for this question is to further delve into the specifics of AI and data science use in new product development. If a company is leveraging these technologies in their product/service development, understanding the nature of the products involved could offer critical insights. This information could potentially reveal correlations between product characteristics and the extent of AI and data science integration. The responses can contribute to an enriched analysis, shedding light on the applicability and benefits of these technologies across various product types).

Answer Options: “Open-ended”

(Response Type and Rationale: it requires a direct input from the respondent, allowing for the capture of specific, detailed information.

Question 12

What is the level of AI and/or Data Science adoption in other activities within your company (excluding product development)?

(Rationale: Like question 11, this question is designed to explore the degree and patterns of AI and Data Science adoption across the company's other functions. This data can provide a broader view of how these technologies are integrated within different areas of the organization, revealing the versatility of AI and Data Science applications. By uncovering the extent of adoption in non-product development functions, we can better understand the role of AI and Data Science in enhancing overall organizational efficiency and innovation, thereby providing a more comprehensive perspective for validating the research hypotheses.)

Answer Options: "0 - No Adoption, 1 - Low Adoption, 2 - Moderate Adoption, 3 - High Adoption"

(Response Type and Rationale Similar to Questions 10 and 11, the ordinal scale allows for capturing varying levels of adoption.)

Question 12.01 (Follow-up to Question 12)

If the company uses AI and/or Data Science in other functions, please specify which ones.

(Rationale: The intent of this question is to gather further specifics if the respondent indicates that their company uses AI and data science in areas other than new product development. The additional detail can provide valuable insights into the scope and versatility of AI and data science applications within the company. It can reveal areas where these technologies are particularly beneficial, further validating their broad relevance and versatility in a business context. The responses could also inform strategies for expanding AI and data science adoption in different organizational functions.)

Answer Type: Open-ended response.

(Response Type and Rationale: This question requires a direct input from the respondent, providing detailed insights into the different areas where AI and data science are applied.)

Section 7: Company's Product and Market Share

Question 13

What is the company's main product or service? You can choose "I'd rather not disclose".

(Question Rationale: This question seeks to understand the key offerings of the company. The responses can help relate the company's products or services to their use of AI and data science. Understanding the product or service context is vital to explore Hypothesis 3, which suggests that AI and data science can expedite innovation in companies, fostering dynamic capabilities and product refinement. By gathering data about the company's main products or services, it is possible to analyse how AI and data science are leveraged for product innovation and development.)

Answer Type: Open-ended response.

(Response Type and Rationale: This question requires a direct input from the respondent, providing specific insights into the company's offerings. The open-ended nature of this question allows for capturing a wide variety of responses.)

Question 14

What is the company's national market share?

(Question Rationale: This question seeks to understand the company's position and competitiveness in the market. This can be an indicator of the company's resources, capabilities, and potentially their adoption of AI and data science.

The company's market share can reflect its success in the market, which can be influenced by its level of AI and data science adoption. This information will contribute to the investigation of Hypothesis 1 by providing insights into the correlation between AI and data science adoption and company success.)

Answer Options: "Do not know; Less than 10%; 10-25%; 26-50%; 51-75%; More than 75%"

(Response Type and Rationale: The pre-defined categories allow for comparison of responses across different market shares.)

Section 8: Feedback

Question F1

How would you evaluate the survey questions, were they clear and concise?

(Question Rationale: This question seeks feedback on the survey to understand if the questions were easily understood by the respondents. This feedback can be valuable for refining future surveys and ensuring the reliability and validity of the data collected. It is also important to assess the efficacy of using an AI-driven chatbot for survey administration, contributing to the practical application aspect of Hypothesis 3, which suggests that AI can enhance various business functions, including marketing and customer engagement.)

Answer Options: On a scale from 1 to 10, with 1 representing "Not at all clear" and 10 representing "Very clear and concise."

(Response Type and Rationale: This question uses a numerical (interval) scale, allowing for more nuanced responses. By providing a range from 1 to 10, the survey can capture a wider spectrum of views regarding the clarity of the survey questions. This can yield more detailed and useful feedback for the survey design.)

Question F2

How would you rate the quality of this survey on a scale of 1 to 5 stars?

(Rationale: This question is designed to gauge the overall quality of the survey from the respondent's perspective. This feedback can be instrumental in identifying areas of improvement and enhancing the survey design for future iterations. Furthermore, by rating the survey on a scale, respondents can express their level of satisfaction in a quantified manner, enabling easier comparison and analysis of feedback. This can also contribute to the practical implications of the research, providing evidence on the perceived quality of AI-driven surveys.)

Answer Options: "1 - Very poor; 2 - Poor; 3 - Fair; 4 - Good; 5 - Excellent"

(Response Type and Rationale: This question uses a Likert scale to capture the participant's rating of the survey quality, enabling a quantified measure of user satisfaction.)

Question F3

Please leave a comment about this experience compared to traditional surveys.

(Rationale: This open-ended feedback question provides respondents with an opportunity to share their personal experience and views about the survey process. This could include comments on the survey's format, length, content, or the use of an AI chatbot for administration. This feedback can provide rich, qualitative data that complements the quantitative feedback from Question B. It can help identify unforeseen issues or benefits of the AI-driven survey, further contributing to the research's practical implications.)

Answer Type: Open-ended response.

(Response Type and Rationale: This question requires a direct input from the respondent, allowing for the capture of detailed, personalized feedback.)

Section 9: Interview Scheduling

Question S1

What is the most convenient time for your interview until May 7, 2023?

(Question Rationale: The purpose of this question is to schedule the follow-up interview at a time convenient for the respondent. By offering a calendar response format, respondents can easily indicate their availability, facilitating the scheduling process. This respects the respondent's time and preferences, contributing to a positive research experience. Additionally, this question showcases the potential of AI-driven surveys for automating aspects of the research process, aligning with the practical implications of the research.)

Answer Type: Calendar response.

(Response Type and Rationale: This question requires the respondent to assign an hour and date in a calendar, ensuring that the interview can be scheduled at a mutually convenient time.)

Section 10: Concluding Remarks

Thank you so much for participating in this survey! Your input is invaluable for our understanding of the intersection of AI, data science, and corporate practices. Please remember that all your responses will be kept strictly confidential and used solely for the purposes of this research.

We deeply appreciate your contribution and time spent completing this survey. We understand that your time is precious, and we aim to make the best use of the data collected. The findings of this study will help businesses and individuals better comprehend the changing landscape of technology adoption in the corporate sector, contributing to more informed decision-making processes in the future.

Interview Questions



Script

- 1.** Can you describe the data sources your company uses in its decision-making process to develop and prioritize new ideas? (4 minutes)
- 2.** Can you describe your company's process for identifying, evaluating and developing new product ideas, including any research and development methods used? (4 minutes)
- 3.** What strategies does your company use to reach and captivate customers? (3 minutes)
- 4.** How does your company optimize stock management and/or forecast demand? (3 minutes)
- 5.** How does your company stay informed about customer satisfaction and any changes in industry trends and/or customer preferences? (4 minutes)
- 6.** Does your company use any AI or data science techniques in its decision-making, product development, marketing or point-of-sale strategies? (3 minutes)

Considering your answer to question 6:

If you answered "yes":

- 7y.** Can you give examples of how these techniques have been implemented and their impact on the company? (3 minutes)
- 8y.** What are the main difficulties or problems? (4 minutes)

If you answered "no":

- 7n.** Is your company open to exploring the use of AI and data science in the future? (3 minutes)
- 8n.** What are the main barriers to implementing data science or AI in the enterprise? (4 minutes)
- 9.** Do the other major *players* in the market use any AI or data science techniques? (2 minutes)

Your and company data will always remain anonymous!

Thank you for contributing to science!

Figure 26: Interview Guide

APPENDIX II - DATASETS

Survey Dataset

Company	Company_Age	Company_Age_Bin	Company_Size	Company_Size	Company_Size_Bin	Industry_Sector	AI_DS_Interest	AI_DS_In_ProductDev	AI_DS_Other_Activiti	AI_DS_Other_Functions	Main_Product_Service	Market_Share
A	Over 50 years, up to 100 years	Established	Large (251-1000 employees)	Large	Health	2 - Moderate interest	1 - Low adoption	1 - Low adoption	-	-	Logistics	10-25%
B	Over 100 years	Legacy	Very large (over 1000 employees)	Very Large	Construction	0 - No Interest	0 - No adoption	0 - No adoption	-	-	Construction	Less than 10%
C	6 to 10 years	New	Small (1-50 employees)	Small	Media	2 - Moderate interest	1 - Low adoption	0 - No adoption	-	-	Media & Advertising	Can't say
D	Over 50 years, up to 100 years	Established	Medium (51-250 employees)	Medium	Wholesale Trade	1 - Low interest	1 - Low adoption	1 - Low adoption	-	-	Industrial Equipment & HVAC	Can't say
E	1 to 5 years	New	Small (1-50 employees)	Small	Market Studies	1 - Low interest	1 - Low adoption	2 - Moderate adoption	-	-	Marketing Services	Can't say
F	11 to 20 years	New	Very large (over 1000 employees)	Very Large	Information Technologies	3 - High interest	2 - Moderate adoption	1 - Low adoption	-	-	Unknown	Can't say
G	6 to 10 years	New	Small (1-50 employees)	Small	Information Technologies	1 - Low interest	1 - Low adoption	1 - Low adoption	CRM and content marketing	-	Security Services	Can't say
H	21 to 50 years	Established	Medium (51-250 employees)	Medium	Health	0 - No Interest	0 - No adoption	0 - No adoption	-	-	Healthcare	Can't say
I	1 to 5 years	New	Small (1-50 employees)	Small	Information Technologies	3 - High interest	2 - Moderate adoption	1 - Low adoption	-	-	Restaurant Tech	Can't say
J	Over 50 years, up to 100 years	Established	Very large (over 1000 employees)	Very Large	Transportation	2 - Moderate interest	1 - Low adoption	1 - Low adoption	Operational Management	-	Transportation Services	51-75%

Table 12: Complete Survey Dataset

Interview Dataset

Company	Data Sources Description	Data Sources Description Cat	New Product Development Process Cat	New Product Development Process	Customer Targeting Strategies	Customer Targeting Strategies Cat	Inventory Optimization Strategies	Inventory Optimization Strategies Cat	
A	<ul style="list-style-type: none"> The company's marketing department relies on brainstorming and self-initiative as sources of data for developing and prioritizing new ideas. Ideas are evaluated by marketing managers and the Chief Marketing Officer (CMO) for final approval. Once approved, the process becomes agile, and changes have been well accepted. <p>Brainstorming and Market Reading</p>	Brainstorming and Market Reading	<ul style="list-style-type: none"> The interviewees are not familiar with the company's process for identifying, evaluating, and developing new product ideas, indicating that their department is not involved in this aspect. The company's process for identifying and developing new product ideas is based on empirical readings and opinions gathered from other market players. While the company does not specifically identify the methods used, it does indicate that there is some market benchmarking in the specification and definition of new ideas. The company lacks a formal process for identifying, evaluating, and developing new product ideas. The process involves a network of commercials and agents who brainstorm and design new products or variations of existing ones. The interviewees did not mention specific research and development methods used for products or services. The company primarily operates as a retailer, importing products and providing associated services, rather than developing new products. The company does have a service component where they offer solutions based on the products they sell, such as programming and conditioning components and sensors. The decision-making for developing new services comes from both PHC and Exec. <p>Market Opinions</p>	<ul style="list-style-type: none"> The company's marketing department uses an annual strategic marketing plan to derive marketing campaigns aligned with their goals. They engage customers through various channels, including email, social media, direct sales interactions, call centres, and focus groups. The company's strategies to target and engage customers are based on market insights that meet customer needs and desires. While the company does not heavily use statistical readings of customer needs, it relies on a large sample of customers and former customers to understand their current needs. The company does not specifically identify the methods used to maintain a low profile on social networks and media. <p>Multi-Channel Marketing</p>	<ul style="list-style-type: none"> The company optimizes inventory management by adopting a first-in-first-out strategy due to dealing with products with shelf life. They use a robot in the warehouse to manage products and prioritize those with closer expiration dates. Forecasting demand is done using SAP and statistical analysis to understand customer buying patterns, allowing them to anticipate and optimize stock. <p>Automated Systems</p>	<ul style="list-style-type: none"> The company optimizes inventory management and forecasts demand by adapting the speed of sales to the available product. They use stock accelerators like marketing investments and adjusts the product's price to control the speed of sales. <p>Demand-Driven</p>	<ul style="list-style-type: none"> Inventory management and demand forecasting are not relevant to the industry in which the company operates (media industry). Since it is a media industry, there is no stock management or need to predict demand for products. <p>Unknown Methods</p>	<ul style="list-style-type: none"> The company does not have a specific model for predicting stock levels due to the high volatility of the market. They rely on a listing of the current stock and use an estimated approach to manage their inventory. <p>Demand-Driven</p>	<ul style="list-style-type: none"> The company optimizes inventory management and predicts demand through fixed clients who pay a fee, allowing them to anticipate a certain monthly demand. For more specific cases, a project manager is assigned to manage the inventory, ensuring timely delivery and exploring the need to hire additional resources or renegotiate deadlines. <p>Client/Project-Based Forecasting</p>
B	<ul style="list-style-type: none"> The company primarily relies on empirical readings made by the company's administrators to understand what is happening in the market and to generate new ideas. It follows a top-down approach. Marketing managers play a crucial role in idea generation, there is also some consideration of the company's internal ideas and feedback from their sales department. The company primarily relies on internal sources for decision-making related to new products. The addition of knowledge from new employees is also taken into consideration for decision-making. The company does not have significant data sources, and the data used is mostly from an accounting and financial standpoint. The company primarily operates as a service provider, offering solutions regarding sales, billing orders, and payments. For service-related decisions and new service development, the company uses Excel in addition to PHC due to the variability in product requirements and associated complexities. The company does not have a formal process for developing new products and they have begun to leverage the full potential of PHC for this purpose. <p>ERP Systems</p>	Brainstorming	<ul style="list-style-type: none"> The company primarily operates as a retailer, importing products and providing associated services, rather than developing new products. The company does have a service component where they offer solutions based on the products they sell, such as programming and conditioning components and sensors. The decision-making for developing new services comes from both PHC and Exec. <p>Market Opinions</p>	<ul style="list-style-type: none"> The company's strategy to target and engage customers is based on market insights and knowledge of Generation Z in Portugal. When approaching potential customers, the company presents the research they have conducted to demonstrate their deep understanding of their target audience. This specialized knowledge is used to deliver tailored and finely tuned campaigns, focusing on visual, copy, and audience behaviour. <p>Brand and Event Marketing</p>	<ul style="list-style-type: none"> The company's strategy to target and engage customers is based on market insights and knowledge of Generation Z in Portugal. When approaching potential customers, the company presents the research they have conducted to demonstrate their deep understanding of their target audience. This specialized knowledge is used to deliver tailored and finely tuned campaigns, focusing on visual, copy, and audience behaviour. <p>Demographic Specialization</p>	<ul style="list-style-type: none"> The company optimizes inventory management and predicts demand through fixed clients who pay a fee, allowing them to anticipate a certain monthly demand. For more specific cases, a project manager is assigned to manage the inventory, ensuring timely delivery and exploring the need to hire additional resources or renegotiate deadlines. <p>Client/Project-Based Forecasting</p>	<ul style="list-style-type: none"> The company optimizes inventory management and forecasts demand by adapting the speed of sales to the available product. They use stock accelerators like marketing investments and adjusts the product's price to control the speed of sales. <p>Demand-Driven</p>	<ul style="list-style-type: none"> The company does not have a specific model for predicting stock levels due to the high volatility of the market. They rely on a listing of the current stock and use an estimated approach to manage their inventory. <p>Demand-Driven</p>	<ul style="list-style-type: none"> The company optimizes inventory management and predicts demand through fixed clients who pay a fee, allowing them to anticipate a certain monthly demand. For more specific cases, a project manager is assigned to manage the inventory, ensuring timely delivery and exploring the need to hire additional resources or renegotiate deadlines. <p>Client/Project-Based Forecasting</p>
C	<ul style="list-style-type: none"> The company's decision-making process for developing and prioritizing new ideas heavily relies on customer feedback and requirements. Working closely with customers are conducted to understand their pain points and product usage scenarios. Customer requirements are used to extract ideas and prioritize features for inclusion in new products. The company places a strong emphasis on understanding customer needs and using that information as a foundation for new product development. <p>Customer Feedback</p>	Market Research	<ul style="list-style-type: none"> The company follows an expeditious proof-of-concept philosophy to identify, evaluate, and develop new product ideas. When a potential need or market opportunity is identified, a team of product engineers is assigned to develop a proof of concept. The proof of concept is tested with potential customers during an early access phase. If the proof of concept shows promise, a more complete contract is created, and deadlines are established for the product's development. The company iterates on product ideas rapidly and iterates on product ideas before committing fully to their development. <p>Proof-of-Concept</p>	<ul style="list-style-type: none"> The company's strategy to target and engage customers is based on market insights and knowledge of Generation Z in Portugal. When approaching potential customers, the company presents the research they have conducted to demonstrate their deep understanding of their target audience. This specialized knowledge is used to deliver tailored and finely tuned campaigns, focusing on visual, copy, and audience behaviour. <p>Demographic Specialization</p>	<ul style="list-style-type: none"> The company optimizes inventory management and predicts demand through fixed clients who pay a fee, allowing them to anticipate a certain monthly demand. For more specific cases, a project manager is assigned to manage the inventory, ensuring timely delivery and exploring the need to hire additional resources or renegotiate deadlines. <p>Client/Project-Based Forecasting</p>	<ul style="list-style-type: none"> The company optimizes inventory management and forecasts demand by adapting the speed of sales to the available product. They use stock accelerators like marketing investments and adjusts the product's price to control the speed of sales. <p>Demand-Driven</p>	<ul style="list-style-type: none"> The company does not have a specific model for predicting stock levels due to the high volatility of the market. They rely on a listing of the current stock and use an estimated approach to manage their inventory. <p>Demand-Driven</p>	<ul style="list-style-type: none"> The company optimizes inventory management and predicts demand through fixed clients who pay a fee, allowing them to anticipate a certain monthly demand. For more specific cases, a project manager is assigned to manage the inventory, ensuring timely delivery and exploring the need to hire additional resources or renegotiate deadlines. <p>Client/Project-Based Forecasting</p>	
D	<ul style="list-style-type: none"> The company's decision-making process for developing and prioritizing new ideas heavily relies on customer feedback and requirements. Working closely with customers are conducted to understand their pain points and product usage scenarios. Customer requirements are used to extract ideas and prioritize features for inclusion in new products. The company places a strong emphasis on understanding customer needs and using that information as a foundation for new product development. <p>Customer Feedback</p>	Market Research	<ul style="list-style-type: none"> The company follows an expeditious proof-of-concept philosophy to identify, evaluate, and develop new product ideas. When a potential need or market opportunity is identified, a team of product engineers is assigned to develop a proof of concept. The proof of concept is tested with potential customers during an early access phase. If the proof of concept shows promise, a more complete contract is created, and deadlines are established for the product's development. The company iterates on product ideas rapidly and iterates on product ideas before committing fully to their development. <p>Proof-of-Concept</p>	<ul style="list-style-type: none"> The company's strategy to target and engage customers is based on market insights and knowledge of Generation Z in Portugal. When approaching potential customers, the company presents the research they have conducted to demonstrate their deep understanding of their target audience. This specialized knowledge is used to deliver tailored and finely tuned campaigns, focusing on visual, copy, and audience behaviour. <p>Demographic Specialization</p>	<ul style="list-style-type: none"> The company optimizes inventory management and predicts demand through fixed clients who pay a fee, allowing them to anticipate a certain monthly demand. For more specific cases, a project manager is assigned to manage the inventory, ensuring timely delivery and exploring the need to hire additional resources or renegotiate deadlines. <p>Client/Project-Based Forecasting</p>	<ul style="list-style-type: none"> The company optimizes inventory management and forecasts demand by adapting the speed of sales to the available product. They use stock accelerators like marketing investments and adjusts the product's price to control the speed of sales. <p>Demand-Driven</p>	<ul style="list-style-type: none"> The company does not have a specific model for predicting stock levels due to the high volatility of the market. They rely on a listing of the current stock and use an estimated approach to manage their inventory. <p>Demand-Driven</p>	<ul style="list-style-type: none"> The company optimizes inventory management and predicts demand through fixed clients who pay a fee, allowing them to anticipate a certain monthly demand. For more specific cases, a project manager is assigned to manage the inventory, ensuring timely delivery and exploring the need to hire additional resources or renegotiate deadlines. <p>Client/Project-Based Forecasting</p>	
E	<ul style="list-style-type: none"> The company's decision-making process for developing and prioritizing new ideas heavily relies on customer feedback and requirements. Working closely with customers are conducted to understand their pain points and product usage scenarios. Customer requirements are used to extract ideas and prioritize features for inclusion in new products. The company places a strong emphasis on understanding customer needs and using that information as a foundation for new product development. <p>Customer Feedback</p>	Market Research	<ul style="list-style-type: none"> The company follows an expeditious proof-of-concept philosophy to identify, evaluate, and develop new product ideas. When a potential need or market opportunity is identified, a team of product engineers is assigned to develop a proof of concept. The proof of concept is tested with potential customers during an early access phase. If the proof of concept shows promise, a more complete contract is created, and deadlines are established for the product's development. The company iterates on product ideas rapidly and iterates on product ideas before committing fully to their development. <p>Proof-of-Concept</p>	<ul style="list-style-type: none"> The company's strategy to target and engage customers is based on market insights and knowledge of Generation Z in Portugal. When approaching potential customers, the company presents the research they have conducted to demonstrate their deep understanding of their target audience. This specialized knowledge is used to deliver tailored and finely tuned campaigns, focusing on visual, copy, and audience behaviour. <p>Demographic Specialization</p>	<ul style="list-style-type: none"> The company optimizes inventory management and predicts demand through fixed clients who pay a fee, allowing them to anticipate a certain monthly demand. For more specific cases, a project manager is assigned to manage the inventory, ensuring timely delivery and exploring the need to hire additional resources or renegotiate deadlines. <p>Client/Project-Based Forecasting</p>	<ul style="list-style-type: none"> The company optimizes inventory management and forecasts demand by adapting the speed of sales to the available product. They use stock accelerators like marketing investments and adjusts the product's price to control the speed of sales. <p>Demand-Driven</p>	<ul style="list-style-type: none"> The company does not have a specific model for predicting stock levels due to the high volatility of the market. They rely on a listing of the current stock and use an estimated approach to manage their inventory. <p>Demand-Driven</p>	<ul style="list-style-type: none"> The company optimizes inventory management and predicts demand through fixed clients who pay a fee, allowing them to anticipate a certain monthly demand. For more specific cases, a project manager is assigned to manage the inventory, ensuring timely delivery and exploring the need to hire additional resources or renegotiate deadlines. <p>Client/Project-Based Forecasting</p>	
F	<ul style="list-style-type: none"> The company's decision-making process for developing and prioritizing new ideas heavily relies on customer feedback and requirements. Working closely with customers are conducted to understand their pain points and product usage scenarios. Customer requirements are used to extract ideas and prioritize features for inclusion in new products. The company places a strong emphasis on understanding customer needs and using that information as a foundation for new product development. <p>Customer Feedback</p>	Market Research	<ul style="list-style-type: none"> The company follows an expeditious proof-of-concept philosophy to identify, evaluate, and develop new product ideas. When a potential need or market opportunity is identified, a team of product engineers is assigned to develop a proof of concept. The proof of concept is tested with potential customers during an early access phase. If the proof of concept shows promise, a more complete contract is created, and deadlines are established for the product's development. The company iterates on product ideas rapidly and iterates on product ideas before committing fully to their development. <p>Proof-of-Concept</p>	<ul style="list-style-type: none"> The company's strategy to target and engage customers is based on market insights and knowledge of Generation Z in Portugal. When approaching potential customers, the company presents the research they have conducted to demonstrate their deep understanding of their target audience. This specialized knowledge is used to deliver tailored and finely tuned campaigns, focusing on visual, copy, and audience behaviour. <p>Demographic Specialization</p>	<ul style="list-style-type: none"> The company optimizes inventory management and predicts demand through fixed clients who pay a fee, allowing them to anticipate a certain monthly demand. For more specific cases, a project manager is assigned to manage the inventory, ensuring timely delivery and exploring the need to hire additional resources or renegotiate deadlines. <p>Client/Project-Based Forecasting</p>	<ul style="list-style-type: none"> The company optimizes inventory management and forecasts demand by adapting the speed of sales to the available product. They use stock accelerators like marketing investments and adjusts the product's price to control the speed of sales. <p>Demand-Driven</p>	<ul style="list-style-type: none"> The company does not have a specific model for predicting stock levels due to the high volatility of the market. They rely on a listing of the current stock and use an estimated approach to manage their inventory. <p>Demand-Driven</p>	<ul style="list-style-type: none"> The company optimizes inventory management and predicts demand through fixed clients who pay a fee, allowing them to anticipate a certain monthly demand. For more specific cases, a project manager is assigned to manage the inventory, ensuring timely delivery and exploring the need to hire additional resources or renegotiate deadlines. <p>Client/Project-Based Forecasting</p>	

Table 13: Complete Interview Dataset

G	<ul style="list-style-type: none"> The company primarily uses CRM and Google Analytics data to make decisions. They also plan to integrate data from LinkedIn and their website after a new CRM is implemented. Besides these sources, they also incorporate data from partners and studies conducted in the industry, like the hotel sector. 	CRM and Analytics	<ul style="list-style-type: none"> The process of identifying, evaluating, and developing new product ideas is currently done spontaneously by individuals, without a structured approach. The company acknowledges the need for a structured approach and plans to address this issue with the new CRM implementation. 	Intuition	<ul style="list-style-type: none"> The company's primary strategies for engaging customers are Lead Generation and account-based marketing. They currently do not have an active customer loyalty process in place but are working towards building stronger relationships through contracted maintenance services. 	Lead Generation and Account-Based Marketing	<ul style="list-style-type: none"> The company minimizes its stock by analysing supplier inventories and predicting delivery times. They use a forecasting method based on previous month's data and historical sales data from Sage, their billing system. Sales data from CRM and Sage are used to generate two types of reports - the billing report and the actual sales report. 	Client/Project-Based Forecasting
H	<ul style="list-style-type: none"> The company, a group of community pharmacies, primarily relies on its own system called 'Sifarma 2000' developed by Glint to generate reports on sales, inventory, and scores data related to daily pharmacy activities and customer information. Sifarma 2000 allows them to segment customers based on various categories such as age, purchase history, buying trends, or pathology to develop communication actions and prioritize ideas. 	Pharmacy Data System	<ul style="list-style-type: none"> The company does not focus on new product development, as it is a community pharmacy. Instead, it engages in actions within the community, such as disease screenings, using the Sifarma 2000 database to support these activities. 	Planning Structure	<ul style="list-style-type: none"> The company's strategy for customer targeting is two-fold: it aims to captivate younger customers through online platforms and social networks, while also serving the older customer base, who typically seek the pharmacy for medication needs. 	Multi-Channel Marketing	<ul style="list-style-type: none"> The company optimizes inventory management using the Sifarma 2000 system, which calculates average consumption of products per month and predicts future needs. However, human perception is also important for adjusting averages, especially during periods of increased demand or for seasonal products. 	Automated Systems
I	<ul style="list-style-type: none"> The company's decision-making process for coming up and prioritizing new ideas is heavily reliant on the data generated by the CRM. They utilize a CRM system to collect and store customer information with multiple data points. The collected data is then analysed to identify correlations and gain insights into their operations. Data-driven decision-making enables the company to make informed choices about product development and strategic decisions. 	CRM and Analytics	<ul style="list-style-type: none"> The company has a structured process for identifying, evaluating, and developing new product ideas. They have two distinct teams: "Discovery" and "Delivery". The Discovery team conducts research and validates ideas, while the Delivery team implements the validated ones. Customer feedback and success stories are vital sources of information for generating new ideas and improving existing functionalities. The Product Manager plays a crucial role in following up on customer feedback and translating it into actionable development stories. 	Community Engagement	<ul style="list-style-type: none"> The company uses Cold Calling as their primary strategy to reach and engage customers. Since they offer unique services, there is no need to convince potential customers to switch from other suppliers. 	Customer Interaction	<ul style="list-style-type: none"> As a market leader with no significant competition, team allocation management is not a major concern for the company. Resource allocation is mainly focused on optimizing the management of teams when dealing with large, specific customer needs. 	Client/Project-Based Forecasting
J	<ul style="list-style-type: none"> The company uses specific areas of Business Intelligence that consolidate information from various departments, such as sales, marketing, and departments for decision-making. However, the information is dispersed and not directly used to generate new innovations or ideas. Innovations in the company typically arise from situations where decision-makers request analysis of specific information or processes that may not be functioning optimally. 	Brainstorming and Market Reading	<ul style="list-style-type: none"> The company operates in a stable sector and focuses on improving user experiences, particularly in mobile applications so that consumers can manage their travelling information. The emphasis is not on being at the cutting edge of research and development but rather on enhancing existing products and services. 	Customer Feedback	<ul style="list-style-type: none"> The company employs strategies focused on competitive pricing and service quality to reach and engage customers. Marketing campaigns are designed in collaboration with partners, and the company uses marketing automation tools, including Google, for analysis and segmentation. 	Competitive Pricing and Collaboration	<ul style="list-style-type: none"> The company practices just-in-time stock management, especially for maintenance parts. For demand forecasting, the company uses a bucket system with different price tiers, but this approach can lead to profitability challenges during peak seasons. 	Demand-Driven

Customer_Satisfaction_Insights	Customer_Satisfaction_Insights_Cat	AI_DS_Use_Cat	AI_DS_Use	AI_DS_Use_AI_DS_Examples	AI_DS_Examples	AI_DS_Difficulties	AI_DS_Difficulties_Cat	Open_To_AI_DS
<ul style="list-style-type: none"> The company stays informed about customer satisfaction through two annual satisfaction studies, one conducted by an external entity, and another conducted internally with a specific focus on their company. Focus groups are also conducted to gather feedback and improve services efficiently. 	Surveys	Active AI & Data Science User	Active AI & Data Science User	<ul style="list-style-type: none"> The use of ChatGPT in the marketing department has helped with document creation, generating action ideas, and writing communication texts. 	AI in Marketing	<ul style="list-style-type: none"> The main difficulty with ChatGPT is the trustworthiness of its answers, especially since it lacks a personal stamp and may not fully align with the company's tone or customer interactions. 	Trust & Reliability Issues	question not applicable
<ul style="list-style-type: none"> The company stays informed about customer satisfaction and industry trends through satisfaction surveys and direct interaction with customers. Surveys are conducted almost daily through their sales agents, providing valuable insights into customer satisfaction. 	Surveys	AI & Data Science Non-User	AI & Data Science Non-User	question not applicable	Not Used	question not applicable	Not Used	<ul style="list-style-type: none"> The company is not open to exploring the use of Artificial Intelligence or data science in the near future, but it may consider it in the medium term. It wants to observe how AI and data science is used in other industries and gather evidence of increased productivity before adopting it.
<ul style="list-style-type: none"> The interviewee doesn't have complete information on how the company tracks customer satisfaction, but they mention that the company is attentive to customers and partners to gauge satisfaction. Other industries are apparent and clear due to the nature of the media industry. 	Surveys	AI & Data Science Non-User	AI & Data Science Non-User	question not applicable	Not Used	question not applicable	Not Used	<ul style="list-style-type: none"> The company is open to exploring the use of AI and data science in the future.
<ul style="list-style-type: none"> The company uses annual questionnaires and QR codes on shipping boxes to assess customer satisfaction, though these methods have had low adherence. The company also gets feedback from the quality system evaluations that the customers have themselves. They are planning to implement surveys at the end of provided services for immediate feedback. 	Customer Interaction	Traditional ERP User	Traditional ERP User	question not applicable	Not Used	question not applicable	Not Used	<ul style="list-style-type: none"> The company is open to use AI and data science in the future. They are currently working on cleaning and organizing their information in databases and training employees to correctly input data into the system.
<ul style="list-style-type: none"> The company measures customer satisfaction by requesting feedback from clients when presenting proposals or campaigns. They also evaluate the success of campaigns through surveys, which helps increase engagement among the targeted audience. Winning awards for certain campaigns, serves as an indicator of success and satisfaction. The company stays informed about customer satisfaction through various channels, including customer support teams that actively gather feedback from customers. Word-of-mouth is used by announcing acquired customers to attract new customers and demonstrate its credibility in the market. They are planning to launch new industry trends and competitors' products to stay informed about changes in customer preferences and market demands. The subject expresses concern about relying too much on customer feedback alone, as this approach may lead to a lack of product identity and potential issues with product attractiveness to other companies in the market. There is a need for a balance between customer-driven features and maintaining a strong product identity. 	Sales and Performance Monitoring	Partial AI & Data Science User	Partial AI & Data Science User	question not applicable	Not Used	question not applicable	Not Used	<ul style="list-style-type: none"> The company is open to exploring the use of AI and data science in the future. They believe it will be inevitable given the strong interest in technology among their target audience.
<ul style="list-style-type: none"> The subject's specific developed product did not use AI or data science techniques. The subject mentions the existence of other products within the company that may have utilized data science models for customer service purposes. Data science or AI is not used in the company for decision-making, only to aggregate value to reliable products. 	Constant Communication	AI for Marketing & CRM User	AI for Marketing & CRM User	<ul style="list-style-type: none"> The subject provides an example of a prediction model that the company initially disregarded but later implemented in customer service, leading to significant improvements in key indicators and customer satisfaction. The successful implementation of the prediction model demonstrates the value of data science when given proper space and time to work on solving specific business problems. 	AI in Marketing	<ul style="list-style-type: none"> The main difficulties in implementing data science or AI techniques include a lack of consideration for data science from the beginning of product development. The company's strong engineering culture may not fully understand the complexities and time required for data science work, leading to challenges in estimating effort and resources. Data-related difficulties are also present, such as data preparation and potential changes in data structure by engineering teams, which can affect the performance of data science models. There is a need for a stronger data culture within the company to foster the effective integration of data science solutions. 	Resource & Skill Deficit	question not applicable

<ul style="list-style-type: none"> The company currently only measures customer dissatisfaction when customers report it. They are implementing a new CRM which will allow them to automatically send questionnaires to customers after each invoice, measure overall customer satisfaction, and predict customer sentiments through an Artificial Intelligent module called Zia. 	Customer Interaction	<ul style="list-style-type: none"> The marketing department uses AI in the form of Opovox is chosen for content creation and scheduling as well as SEO. The company will use AI for decision-making processes with the new CRM, particularly in lead scoring. The company plans to implement an internal AI-based chatbot for technical problem-solving, which will have long-term impacts on their operations. 	AI & Data Science Non-User	<ul style="list-style-type: none"> The use of AI is currently limited to the marketing department and is being tested for optimizing time and decision making in an employment way. The use of AI reduced the time of scheduling marketing campaigns from 4 hours to 5 minutes. The impact of AI in training technicians is expected to be seen in the medium to long term, around 1 to 2 years. 	AI in Marketing	<ul style="list-style-type: none"> Resistance to innovation and acceptance, particularly among older employees. The process of democratizing technologies, especially when decision makers are from older age groups makes it harder to convince management about the positive impact of AI despite clear improvements in efficiency. Concerns about the potential misuse of rapidly evolving AI technology for negative purposes. The validity of information and potential bias in AI generated outputs. 	Innovation Resistance question not applicable
<ul style="list-style-type: none"> The company faces challenges in measuring customer satisfaction, especially for prescription medications, where the customer does not have much say in the choice. For nonprescription products, satisfaction is often gathered through word of mouth or via feedback during repurchases. 	Sales and Performance Monitoring	<ul style="list-style-type: none"> The company is not actively using AI or data science techniques, but it has implemented robots in two of its pharmacies for stock management and organization, which might involve AI applications. 	Active AI & Data Science User	<ul style="list-style-type: none"> The use of AI robots in two pharmacies for stock management has significantly improved the efficiency of inventory management. The robots collect sales data, reducing human errors and providing better traceability. However, it's not clear if the robots have AI. Data from the Sigma 2000 system is utilized for a various marketing campaigns. For instance, the company can extract information from the system to identify the target audience based on specific purchase histories, ensuring that marketing messages are sent to relevant customers. 	Automation in Inventory	<ul style="list-style-type: none"> Finding employees who are skilled in entering and utilizing the data for stock management is challenging. Many pharmacies still use traditional manual storing methods, so finding individuals experienced with robotic systems can be difficult. The complexity of tracking medication validity manually is a time-consuming task, particularly with thousands of different medications with varying shelf lives. The use of robots helps with this issue, but adoption and adaptation to the new system might pose initial challenges. 	Resource & Skill Deficit question not applicable
<ul style="list-style-type: none"> The company stays informed about customer satisfaction and industry/customer preferences through constant communication with customers. Product teams, including the CEO, product manager, and designer, maintain direct contact with customers to understand their needs and concerns. By being in touch with customers regularly, the company can quickly respond to feedback and make necessary improvements to their product and services. 	Constant Communication	<ul style="list-style-type: none"> While the company does not use Artificial Intelligence for decision-making, product development, or marketing, they employ data science techniques for certain aspects of their operations. Data science techniques are primarily used for forecasting sales, churn, and repeat analysis. The company emphasizes the importance of data engineering and organizing data streams to ensure accurate and real-time results. Additionally, the company's product features incorporate AI, which automatically analyzes that assist their customers in making decisions. 	Data Science for Forecasting & Results Analysis User	<ul style="list-style-type: none"> The company's product focuses on e-commerce analysts, considering three key parameters: the number of visitors, conversion rate, and average purchase value. Data science techniques are used to understand the challenges faced by their customers' (retailers) and to optimize their customers' operations. While specific algorithms are not mentioned, the company aims to enhance the accuracy of values and real-time data. 	Data Science in E-commerce	<ul style="list-style-type: none"> Ensuring data consistency and speed for decision-making is one of the main challenges faced by the company. They must rely on data engineering and organization to ensure the quality of the data streams they analyze. Building customer trust is crucial for the success of the company. They must convey reliability and accuracy in their data-driven approaches to gain customer confidence. Hiring qualified talent for data science and AI in pharmacies is a challenge, although the company has a dedicated team with 20 people dedicated to these tasks. 	Trust & Reliability Issues question not applicable
<ul style="list-style-type: none"> The company stays informed about customer satisfaction and industry trends through statistics and data available in the regulated sector. They also develop prediction models, including those for extreme disruptions like the COVID-19 pandemic, no show models, and fleet delay predictions. Quality data and prediction models are both considered essential for staying ahead in a highly competitive market. 	Predictive Analytics	<ul style="list-style-type: none"> The company uses AI and data science techniques in decision-making, including for sales and treasury budgeting. Fleet optimization is another area where more robust models are employed, indicating a proactive approach to resource management. The application of data science and AI in marketing and point-of-sale strategies remains uncertain, with potential usage by partners. 	Active AI & Data Science User	<ul style="list-style-type: none"> Data science and machine learning have been widely implemented in pricing and informed decision-making. Models for sales and treasury budgeting have resulted in a more proactive business plan and an optimizer fleet. The company also utilizes chatbots for customer communication, showcasing the use of AI-driven customer service solutions. 	Predictive Modelling	<ul style="list-style-type: none"> A major challenge in adopting data science and machine learning is the company's traditional sector focus, where technological innovations may be difficult to demonstrate. The use of Black Box type machine learning models has led to difficulties in justifying their results, which could impact their acceptance and wider implementation. 	Trust & Reliability Issues question not applicable

Open to AI & DS Use	Barriers to AI & DS Use	Competitor AI & DS Use	Innovation Importance	Innovation Importance	Functional Departments Presence	AI & DS Use	AI & DS Use Binary
<p>Already Use</p> <p>question not applicable</p> <ul style="list-style-type: none"> The main barrier to implementing data science or AI in the company is its culture, which avoids pursuing the adoption of new technologies. The subject prefers a wait-and-see approach to understand new technologies are implemented and avoid potential growing pains. 	<p>Already Use</p> <ul style="list-style-type: none"> The main barrier to implementing data science or AI in the company are technical complexity and cost. The interviewees does not mention any reason of resistance to adopting these technologies. 	<p>Barriers to AI & DS Use</p> <p>Already Use</p> <ul style="list-style-type: none"> The subject does not have concrete data about other major market players using AI or data science techniques, but they do so, only after their interviewees might be using these technologies in small areas, possibly in marketing. 	<p>Very Important</p> <ul style="list-style-type: none"> Yes, innovation is essential for this company as evident from their process of developing and prioritizing new ideas through brainstorming and self-initiative. The company's openness to using AI, data science techniques, and robots in the warehouse also reflects their willingness to adopt innovative approaches. 	<p>Very Important</p> <ul style="list-style-type: none"> Yes, there is evidence of functional departments in the company. The interviewees clearly states that they can only speak about the data sources and decision-making process in the marketing department. Additionally, they mention that their department is not involved in new product or service data, suggesting that other departments handle this aspect. 	<p>AI & DS Use</p> <ul style="list-style-type: none"> Yes, the company does use data science and AI techniques. The marketing department has employed ChatGPT to support various tasks, including document structuring, generating action ideas, and writing communications for internal and external purposes. They also mentioned selling through data analysis to clients. 	<p>AI & DS Use Binary</p> <p>1</p>	
<p>Medium-Term Consideration</p> <ul style="list-style-type: none"> The main barriers to implementing data science or AI in the company are technical complexity and cost. The interviewees does not mention any reason of resistance to adopting these technologies. 	<p>Barriers to AI & DS Use</p> <p>Already Use</p> <ul style="list-style-type: none"> The subject admits not knowing if other competing players are using AI or data science techniques in their processes. The subject mentions that some companies may claim to use such techniques, but it is unclear how competency they are implemented in their processes. The subject also notes that some communication systems are becoming common across various products as a popular application of AI in the market. 	<p>Medium Important</p> <ul style="list-style-type: none"> The company relies on empirical reading, market benchmarking, and customer feedback to generate new ideas. However, they do not have a formal process but they do so, only after their competitors, so they avoid being innovative. 	<p>Medium Important</p> <ul style="list-style-type: none"> Yes, there is evidence of functional departments in the company. The interviewees clearly states that they can only speak about the data sources and decision-making process in the marketing department. Additionally, they mention that their department is not involved in new product or service data, suggesting that other departments handle this aspect. 	<p>Medium Important</p> <ul style="list-style-type: none"> Yes, there is evidence of functional departments in the company. The interviewees clearly states that they can only speak about the data sources and decision-making process in the marketing department. Additionally, they mention that their department is not involved in new product or service data, suggesting that other departments handle this aspect. 	<p>AI & DS Use</p> <ul style="list-style-type: none"> No, the company does not currently use data science or AI techniques in its decision-making processes, product development, marketing, or product development, marketing, or point-of-sale strategies. 	<p>0</p>	
<p>Open to AI & DS</p> <ul style="list-style-type: none"> The main barriers to implementing data science or AI in the company are customer resistance to change and the need to demonstrate the benefits and gains of adopting these technologies gradually. Some clients are resistant to move away from traditional processes that they believe have worked for them. There is also a lack of readiness among customers to embrace new technologies, which the company aims to address by showing tangible results and positive returns on investments. 	<p>Challenges</p> <p>Technological and Cost</p> <ul style="list-style-type: none"> The interviewees is not certain but believes that other major players in the market may also use AI or data science techniques. 	<p>Not Important</p> <ul style="list-style-type: none"> The company's lack of a formal process for generating and evaluating new product ideas suggests a more ad-hoc or well-structured focus. 	<p>Not Important</p> <ul style="list-style-type: none"> The interviewees does not provide explicit evidence of functional departments within the company. However, the mention of communications, product development, marketing, or sales suggests a structured approach to product idea generation. 	<p>Not Important</p> <ul style="list-style-type: none"> The interviewees does not provide explicit evidence of functional departments within the company. However, the mention of communications, product development, marketing, or sales suggests a structured approach to product idea generation. 	<p>AI & DS Use</p> <ul style="list-style-type: none"> The company does not currently use data science or AI techniques in its decision-making processes, product development, marketing, or product development, marketing, or point-of-sale strategies. 	<p>0</p>	
<p>Open to AI & DS</p> <ul style="list-style-type: none"> The main barriers to implementing data science or AI in the company are customer resistance to change and the need to demonstrate the benefits and gains of adopting these technologies gradually. Some clients are resistant to move away from traditional processes that they believe have worked for them. There is also a lack of readiness among customers to embrace new technologies, which the company aims to address by showing tangible results and positive returns on investments. 	<p>Challenges</p> <p>Customer Resistance</p> <ul style="list-style-type: none"> The subject admits not knowing if other competing players are using AI or data science techniques in their processes. The subject mentions that some companies may claim to use such techniques, but it is unclear how competency they are implemented in their processes. The subject also notes that some communication systems are becoming common across various products as a popular application of AI in the market. 	<p>Very Important</p> <ul style="list-style-type: none"> Yes, innovation is important for this company as they heavily rely on customer feedback and requirements to develop and promote new products. They also mention their willingness to explore and test new ideas and functionalities to address specific needs and stay competitive in the market. 	<p>Very Important</p> <ul style="list-style-type: none"> Yes, there is evidence of functional departments within the company. The subject mentions that they have a structured approach to product development, including the existence of specific product development teams. They also mention departments such as marketing and sales. 	<p>Very Important</p> <ul style="list-style-type: none"> Yes, there is evidence of functional departments within the company. The subject mentions that they have a structured approach to product development, including the existence of specific product development teams. They also mention departments such as marketing and sales. 	<p>AI & DS Use</p> <ul style="list-style-type: none"> The company is not currently using AI or data science techniques in its decision-making or product development processes. However, they express openness to exploring these technologies in the future, particularly in market analysis and data analysis. 	<p>1</p>	

Already Use	question not applicable	Already Use	<ul style="list-style-type: none"> The subject does not have explicit data about the AI and data science usage of other major players in the market. They mention that some of their partners use AI or data science techniques. However, the interviewees mentioned that pharmaceutical laboratories, both in medicines and cosmetics, are larger companies that are very attentive to these technologies. It can be inferred that other major market players might indeed be using data science or AI techniques in their operations. 	Specific Applications	<ul style="list-style-type: none"> Innovation is crucial for this company as evident from their constant pursuit of new technologies like AI, data analytics, and CRM for decision-making, customer engagement, and inventory management. Their focus on implementing a new CRM and utilizing an AI-based chatbot demonstrates their commitment to progress. They recognize the significance of optimizing resources and decision-making through technology, including AI. 	Very Important	<ul style="list-style-type: none"> The company clearly has evidence of functional departments using AI for content creation and SEO, a sales department utilizing sales data for forecasting, and a technical staff involved in implementing an internal chatbot for training technicians, which suggests the presence of an operational department where technicians work. 	<ul style="list-style-type: none"> Yes, the company uses AI and data science. They employ AI for content creation and SEO in the marketing department. Additionally, they plan to utilize AI and data science in the new CRM for decision-making, forecasting, customer sentiment analysis, lead scoring, and internal knowledge base creation. 	1
Already Use	question not applicable	Already Use	<ul style="list-style-type: none"> The interviewees did not provide a direct answer to whether other major players use AI or data science techniques. However, the interviewees mentioned that pharmaceutical laboratories, both in medicines and cosmetics, are larger companies that are very attentive to these technologies. It can be inferred that other major market players might indeed be using data science or AI techniques in their operations. 	Indirectly Suggest AI/DS Use	<ul style="list-style-type: none"> Yes, innovation seems important for this company, as it has adopted the use of robots for stock management in two of its pharmacies. It also mentions the development of the marketing department, including the use of online surveys and social networks to reach a younger audience. 	Very Important	<ul style="list-style-type: none"> Yes, there is evidence of functional departments in the company. The interviewees mention the marketing department, which is being greatly developed and has grown through online stores and social networks. Additionally, there is a department focused on actions in the community, such as screenings, which suggests the presence of specialized functional areas. 	<ul style="list-style-type: none"> The company currently uses robots in two pharmacies for stock management and organization. These robots might involve data science or AI applications, as they handle information on medication batches, their fill, and track sales data for traceability. 	0
Already Use	question not applicable	Already Use	<ul style="list-style-type: none"> The company currently does not have a dedicated AI department, but they are exploring this option. However, they anticipate potential competitive advantages in the future. 	Have No Commitments	<ul style="list-style-type: none"> Yes, innovation is essential for the company's decision-making and growth. They are currently focused on customer engagement and their products and services. Yes, innovation is crucial for the company's growth and improvement. The company focuses on enhancing user experiences and continuously seeks ways to improve its products and services, particularly in mobile applications. 	Not Important	<ul style="list-style-type: none"> Yes, the company has a development department, a sales department, and a CEO. 	<ul style="list-style-type: none"> Yes, the company employs data science and analytics in its product development phase. They also use feedback artificial intelligence algorithms in their product to support customer decision-making. 	1
Already Use	question not applicable	Already Use	<ul style="list-style-type: none"> Other major players in the market may not extensively use artificial intelligence. However, data science and machine learning are being employed by manufacturers to identify parts with higher failure rates and adjust production accordingly. This indicates that while AI adoption may not be widespread across the industry, specific applications are being utilized for optimization purposes. 	Specific AI/DS Application	<ul style="list-style-type: none"> Yes, innovation is crucial for the company's decision-making and growth. They are currently focused on customer engagement and their products and services. Yes, innovation is crucial for the company's growth and improvement. The company focuses on enhancing user experiences and continuously seeks ways to improve its products and services, particularly in mobile applications. 	Very Important	<ul style="list-style-type: none"> Yes, there is evidence of functional departments within the company. The company uses specific areas of Business Intelligence that consolidate information from different sectors and departments for decision-making. Robustness metrics are used for optimizing their management. Potential use in marketing and point-of-sale strategies, although uncertain. 	<ul style="list-style-type: none"> Yes, the company uses data science and AI techniques in several areas: <ul style="list-style-type: none"> Decision-making: Sales and treasury budgets Forecasting: Sales and treasury budgets Optimization: Robustness metrics are used for optimizing their management. Potential use in marketing and point-of-sale strategies, although uncertain. 	1

AI_DS_Advantages	AI_DS_Challenges	Final_Insights	Strategy	Highlight_Curiosity
<ul style="list-style-type: none"> Data science and AI techniques, such as ChatGPT, have revolutionized the marketing department in document structuring and generating ideas for marketing actions, improving efficiency in decision-making processes. The company's ability to self-analyzed data analysts to clients indicates an additional revenue stream and demonstrates the value of their data-driven approach. 	<ul style="list-style-type: none"> The interviewees expressed concerns about the trustworthiness of ChatGPT's answers, highlighting the challenge of fully relying on AI-generated content for certain tasks like writing texts that require a personal touch. The main challenge lies in the company's culture, which avoids being a pioneer in adopting new technologies and prefers to wait for evidence of increased productivity and successful implementation in other sectors before exploring the use of AI. 	<ul style="list-style-type: none"> The company demonstrates a data-driven and innovative approach, with a strong focus on using data to inform decision-making, support marketing strategies, and optimize inventory management. While they have embraced AI and data science techniques, there are reservations about certain aspects, particularly regarding the personalized touch in communications. The company's decision-making and idea generation processes heavily rely on empirical readings and opinions from top-level management and other market players. While the company emphasizes meeting customer needs and desires, it does not heavily rely on statistical data to understand customer preferences. The company is currently not keen to adopting AI or data science but may consider it in the medium term once evidence of increased productivity is available. 	<p>Data-Driven</p> <p>Personal interaction Focused</p>	<ul style="list-style-type: none"> It's interesting to note that the company has a colleague who has mastered AI technology, programs, and assists the sales department. This suggests that there are individuals within the company who possess specialized knowledge and skills in AI, driving its implementation in various departments. The company's cautious approach to adopting new technologies and its emphasis on avoiding pioneering can impact its competitive advantage in the market. While waiting for evidence of increased productivity can be sensible, it also poses the risk of falling behind competitors who are early adopters of innovative technologies. This approach raises interesting questions about the company's ability to adopt and stay competitive in a rapidly evolving business landscape.
<ul style="list-style-type: none"> The main advantage of data science or AI for this company could be the potential to gain deeper insights into market trends and customer preferences, leading to more targeted product development and marketing strategies. 	<ul style="list-style-type: none"> The main challenge would be dealing with technical complexity and cost associated with implementing data science or AI technologies. 	<ul style="list-style-type: none"> The company appears to be more focused on personal interactions with customers, emphasizing face-to-face strategies and direct engagement. There is a potential opportunity for the company to leverage data science and AI in the future to improve its decision-making processes and product offerings. 	<p>Personal interaction Focused</p> <p>Innovative</p>	<ul style="list-style-type: none"> One interesting point to note is that the interviewee mentioned that the company is in a niche industry and doesn't deal with stock management or demand forecasting. This suggests a unique business model compared to traditional product-based companies.
<ul style="list-style-type: none"> The main advantage could be enhanced decision-making, product development, and customer targeting through data-driven insights. The main advantages would be more sophisticated analytics and forecasting, better decision-making, and potential improvements in customer satisfaction tracking. 	<ul style="list-style-type: none"> The challenges include lack of computer literacy among employees, lack of understanding about AI and data science among management, and difficulty in hiring data science professionals due to high salaries. 	<ul style="list-style-type: none"> The company shows a strong willingness to innovate and improve, despite being a traditional company in an old industry. There is potential for AI and data science to play a significant role in their operations, but the company will need to overcome several challenges, particularly in terms of training and hiring. Before these technologies can be fully integrated. The company's focus on market research and understanding customer needs is a key strength of their approach to product development and marketing. They believe that AI and data science will play a significant role in their industry in the future, and they are open to exploring these technologies to stay ahead in a rapidly evolving market. 	<p>Personal interaction Focused</p> <p>Innovative</p>	<ul style="list-style-type: none"> Despite the digital nature of their business, the company still relies on Excel for some of their operations. This is due to the high variability in their product and service offerings and an older workforce (750 years old), which makes migrating these processes to their ERP system challenging.
<ul style="list-style-type: none"> The main advantage of data science or AI for this company could be the potential to gain deeper insights into market trends and customer preferences, leading to more targeted product development and marketing strategies. 	<ul style="list-style-type: none"> However, challenges include a lack of consideration for data science from the beginning of product development, a strong reliance on traditional methods, and a need to understand the complexities of data science work, and data-related difficulties such as data preparation and changes in data structure affecting the models. 	<ul style="list-style-type: none"> The company values customer feedback and requirements as a primary source of data for new product development. There is room for improvement in incorporating data science and AI techniques more proactively to enhance products and services. A multidisciplinary approach, involving both data science and engineering teams, could address some of the challenges in implementing data science solutions. 	<p>Customer-Feedback Focused</p> <p>Innovative</p>	<ul style="list-style-type: none"> It's interesting to note that the company has implemented Conversational Z, an advanced form of AI and data science techniques, emphasizing the need to use technologies that themselves consume and understand as part of the target audience. This forward push for innovation could drive the company towards integrating new technologies in their market research and product development processes.
<ul style="list-style-type: none"> The main advantage of data science or AI for this company could be the potential to gain deeper insights into market trends and customer preferences, leading to more targeted product development and marketing strategies. 		<ul style="list-style-type: none"> The company values customer feedback and requirements as a primary source of data for new product development. There is room for improvement in incorporating data science and AI techniques more proactively to enhance products and services. A multidisciplinary approach, involving both data science and engineering teams, could address some of the challenges in implementing data science solutions. 	<p>Customer-Feedback Focused</p> <p>Innovative</p>	<ul style="list-style-type: none"> A curious observation is that the subject's product, named "Predictive," did not actually have any predictive capabilities. The data science was only used to generate reports and insights. This highlights the importance of having a clear understanding of what data science entails and the potential it can bring to a product's development and performance.

<ul style="list-style-type: none"> AI and data science offer significant advantages for this company, including time and decision-making optimization, technical problem-solving, and understanding customer sentiments. Leveraging OpenAI's ChatGPT can streamline content creation in marketing, while AI techniques can enhance SEO and website development. Additionally, chat-based questionnaires and Zoom's AI aid in measuring customer satisfaction. 	<ul style="list-style-type: none"> However, challenges such as resistance to change, data security concerns, and potential biases in AI-generated outputs need to be addressed. 	<ul style="list-style-type: none"> The main advantage of using data science or AI for this company is seen in the implementation of robots in pharmacies for stock management. These robots save time, reduce human error, and provide better traceability of medication (e.g., 	<ul style="list-style-type: none"> The challenges include ensuring data consistency and real-time speed, gaining customer trust in data-driven approaches, and finding qualified talent for data science and AI initiatives. 	<ul style="list-style-type: none"> Predictive models aid in decision-making, leading to more informed and optimized business plans and resource allocation. Data science and AI techniques help in predicting delays, optimizing fleet management, and improving sales and treasury budgeting. 	<ul style="list-style-type: none"> The company's primary data source for decision-making is its own system, Sigma 2000, which provides real-time inventory, sales, and customer behavior data. The data-driven approach contributes to efficient customer segmentation and decision-making processes. The use of robots for stock management showcases the company's willingness to embrace technological innovations, improving inventory management efficiency and reducing human errors. The marketing department is actively developing strategies to reach younger customers through online platforms and social networks, indicating a commitment to engaging with a broader audience. While the company has adopted data science and AI techniques for stock management, there is potential for further exploration and implementation in other areas to enhance overall efficiency and business growth. 	<p>Data Driven</p>	<ul style="list-style-type: none"> The company plans to use AI (Zia) for sentiment analysis of customer emails, improving responses. Despite efficiency gains like faster content creation, convincing management about AI's positive impact remains a challenge. Organizational resistance may stem from decision-makers not fully understanding AI's benefits.
<ul style="list-style-type: none"> The main challenge mentioned is the lack of knowledge or awareness about the potential use of AI or data science in other areas of the company, such as decision-making, product development, or marketing. 	<ul style="list-style-type: none"> The company's approach revolves around using data science and data-driven decision-making to enhance their products and customer experiences. They emphasize the importance of accurate and real-time data, which enables them to lead the market in their niche. 	<ul style="list-style-type: none"> The company's emphasis on data-driven decision making and optimization showcases its commitment to leveraging technology for better performance. Innovation is valued for enhancing user experiences and staying competitive in the market. Challenges in integrating data science and AI into the traditional sector highlight the need for a strong business-technical framework. 	<ul style="list-style-type: none"> The subject mentions ChatGPT as a potential game-changer that raises curiosity about the broader impact and applications of AI language models in various industries. He suggests that AI advancements like ChatGPT have the potential to revolutionize traditional sectors, even when the full understanding of underlying patterns might be challenging. 	<ul style="list-style-type: none"> A point of curiosity is the company's claim of being the first in the market to use AI in decision-making, despite being in the same space. This implies a unique advantage or offering that sets them apart in the market. 	<p>Data Driven</p>	<ul style="list-style-type: none"> It's interesting to note that while the company has successfully adopted robots for stock management and utilizes the Sigma 2000 system for customer segmentation, yet there's a knowledge gap about the broader potential of data science and AI. 	
<ul style="list-style-type: none"> The company operates in a traditional sector, making it challenging to demonstrate the potential of data science and machine learning. Concerns about using Black Box-type machine learning models and the inability to justify their results may hinder wider adoption. 	<ul style="list-style-type: none"> The company's approach revolves around using data science and data-driven decision-making to enhance their products and customer experiences. They emphasize the importance of accurate and real-time data, which enables them to lead the market in their niche. 	<ul style="list-style-type: none"> The company's emphasis on data-driven decision making and optimization showcases its commitment to leveraging technology for better performance. Innovation is valued for enhancing user experiences and staying competitive in the market. Challenges in integrating data science and AI into the traditional sector highlight the need for a strong business-technical framework. 	<ul style="list-style-type: none"> The subject mentions ChatGPT as a potential game-changer that raises curiosity about the broader impact and applications of AI language models in various industries. He suggests that AI advancements like ChatGPT have the potential to revolutionize traditional sectors, even when the full understanding of underlying patterns might be challenging. 	<ul style="list-style-type: none"> A point of curiosity is the company's claim of being the first in the market to use AI in decision-making, despite being in the same space. This implies a unique advantage or offering that sets them apart in the market. 	<p>Data Driven</p>	<ul style="list-style-type: none"> It's interesting to note that while the company has successfully adopted robots for stock management and utilizes the Sigma 2000 system for customer segmentation, yet there's a knowledge gap about the broader potential of data science and AI. 	

Survey Quality Dataset

Company	Survey_Quality	AI_Experience_Evaluation	Survey_Comments
A	10	Excellent	It's more interactive and less boring
B	10	Excellent	The fact that it is a chat makes the survey more "humanized"
C	10	Excellent	A little longer
D	7	Good	Different. And therefore attracts attention
E	10	Regular	It was more interactive and interesting to answer the questions
F	9	Good	Some questions lacked context for scenarios where companies are international.
G	8	Good	More humanized but with lower perception of survey completion percentage
H	8	Good	More interactive and captivating
I	8	Regular	I couldn't understand where they applied AI here? It seemed to me a very straightforward form
J	9	Good	Greater speed and efficiency

Table 14: Complete Survey Quality Dataset

APPENDIX III - DATA ANALYSIS

Fisher's Exact Test Code

```
1 #!/usr/bin/env python
2 # coding: utf-8
3
4 ### Step 1: Load the Dataset
5
6 # In[ ]:
7
8
9 # Load necessary libraries
10 install.packages('googlesheets4', quietly = TRUE)
11 library(googlesheets4)
12 install.packages('ggplot2', quietly = TRUE)
13 library(ggplot2)
14 install.packages('gridExtra', quiet = TRUE)
15 library(gridExtra)
16 library(grid)
17
18 # In[ ]:
19
20
21 # Set the Google Sheet URL
22 sheet_id <- "13GJpuDCgEsfhFNWZuQySuqyAFjoVxHqC_ockamL3oj0"
23 sheet_name <- "Dataset_Full_Clean"
24 url <- paste0("https://docs.google.com/spreadsheets/d/", sheet_id, "/gviz/tq?tqx=out:csv&sheet=", sheet_name)
25
26 # Load the dataset
27 df <- read.csv(url, stringsAsFactors = FALSE)
28
29 # Display the first few rows of the dataset
30 head(df)
31
32 # In[ ]:
33
34
35 # List columns using names()
36 print(names(df))
37
38 # In[ ]:
39
40
41 # List of wildcard strings
42 wildcards <- c("Company", "Survey_Comments", "Company_Age_Bin", "Company_Size_Bin", "*_Numerical", "Data_Sources_Description",
43 "New_Product_Development_Process", "Customer_Targeting_Strategies", "Innovation_Importance", "Inventory_Optimization_Strategies",
44 "Customer_Satisfaction_Insights", "AI_DS_Use", "Open_To_AI_DS", "Cluster", "AI_DS_Difficulties", "Barriers_To_AI_DS", "Functional_Departments_Presence",
45 "AI_DS_Advantages", "AI_DS_Challenges", "Final_Insights", "Highlight_Curiosity")
46
47 # Exclude columns based on wildcard strings
48 exclude_cols <- unlist(lapply(wildcards, function(w) {
49   grep(w, names(df))
50 }
51 ))
52 df <- df[, -exclude_cols]
53 print(names(df))
54 df
55
56 ### Step 2: Inspect Unique Values in Each Column
57
58 # In[ ]:
59
60
61 # Remove rows with missing values from the original dataframe 'df'
62 df_clean <- na.omit(df)
63
64 # Compute the number of unique values for the cleaned dataframe 'df_clean'
65 unique_counts_clean <- sapply(df_clean, function(col) length(unique(col)))
66
67 # Display the number of unique values for each column in 'df_clean'
68 data.frame(Column = names(df_clean), Unique_Values_Count = unique_counts_clean)
69
70 ### Step 3: Perform Fisher's Exact Tests
71
72 # In[ ]:
73
74
75 # Identify categorical columns based on unique value counts for the cleaned dataframe
76 categorical_columns <- names(df_clean)[unique_counts_clean <= 10]
77
78 # Function to perform Fisher's Exact Test for a pair of columns
79 fisher_test_pair <- function(col1, col2) {
80   table_data <- table(df_clean[[col1]], df_clean[[col2]])
81   test_result <- fisher.test(table_data)
82   return(test_result$p.value)
83 }
84
85 # Perform Fisher's Exact Test for each pair of categorical columns
86 results_list <- list()
87 for (i in 1:length(categorical_columns)) {
88   for (j in 1:length(categorical_columns)) {
89     if (i != j) {
90       p_value <- fisher_test_pair(categorical_columns[i], categorical_columns[j])
91       results_list[[paste(categorical_columns[i], 'vs', categorical_columns[j])] ] <- p_value
92     }
93   }
94 }
95
96 # Convert results to a dataframe
97 results_df <- data.frame(Column_Pair = names(results_list), P_Value = unlist(results_list))
98
99 # Remove repeated pairs
100 results_df$Column_Pair <- sapply(strsplit(results_df$Column_Pair, ' vs '), function(x) {
101   paste(sort(x), collapse = ' vs ')
102 })
```

Figure 27: Complete Fisher's Exact Test Code in R Language

```

99 })
100 results_df <- unique(results_df)
101
102 results_df
103
104 # In[ ]:
105
106
107 # Identify categorical columns based on unique value counts for the cleaned dataframe
108 categorical_columns <- names(df_clean)[unique_counts_clean <= 10]
109
110 # Function to perform Fisher's Exact Test for a pair of columns
111 fisher_test_pair <- function(col1, col2) {
112   table_data <- table(df_clean[[col1]], df_clean[[col2]])
113   test_result <- fisher.test(table_data)
114   return(test_result$p.value)
115 }
116
117 # Perform Fisher's Exact Test for each pair of categorical columns
118 results_list <- list()
119 for (i in 1:length(categorical_columns)) {
120   for (j in 1:length(categorical_columns)) {
121     if (i != j) {
122       p_value <- fisher_test_pair(categorical_columns[i], categorical_columns[j])
123       results_list[[paste(categorical_columns[i], 'vs', categorical_columns[j])] ] <- p_value
124     }
125   }
126 }
127
128 # Convert results to a dataframe
129 results_df <- data.frame(Column_Pair = names(results_list), P_Value = unlist(results_list))
130
131 # Remove repeated pairs
132 results_df$Column_Pair <- sapply(strsplit(results_df$Column_Pair, ' vs '), function(x) {
133   paste(sort(x), collapse = ' vs ')
134 })
135 results_df <- unique(results_df)
136
137 results_df
138
139 ### Step 4: Visualize Fisher's Exact Test Results
140
141 # In[ ]:
142
143
144 # Filter the results to keep only those with p-values less than 0.05
145 significant_results <- results_df[results_df$P_Value < 0.05, ]
146
147 # Modify the filter function to exclude results based on wildcard strings
148 exclude_wildcard <- function(data, exclude_patterns) {
149   excluded_data <- data[!grepl(paste(exclude_patterns, collapse = '|'), data$Column_Pair), ]
150   return(excluded_data)
151 }
152
153 # Filter results to exclude those pairs that match the wildcard patterns 'cluster' or '*_Numerical'
154 excluded_wildcard_results <- exclude_wildcard(significant_results, c('.*_Numerical'))
155
156 # Display the excluded results based on wildcard patterns
157 excluded_wildcard_results
158
159 # **NOTE:** The statistic values for the Fisher's Exact Test results are NULL. This is expected behavior for Fisher's Exact Test in R, as the test does not
160 # compute a test statistic in the same way that, for example, a chi-squared test does. Instead, it directly computes the p-value.
161
162 # In[ ]:
163
164 # Define colors based on p-value ranges
165 color_assign <- function(p_value) {
166   if (p_value >= 0.035) {
167     return('#0F95D7')
168   } else if (p_value >= 0.015) {
169     return('#FF2700')
170   } else {
171     return('#810F7C')
172   }
173 }
174
175 # Assign colors to the dataframe based on p-value ranges
176 excluded_wildcard_results$Color <- sapply(excluded_wildcard_results$P_Value, color_assign)
177
178 # Plotting the bar chart
179 library(ggplot2)
180
181 # Adjusting the width of the plot
182 options(repr.plot.width=10, repr.plot.height=8)
183
184 # Plotting the bar chart with values on the x-axis sorted by ascending order of P-value
185 ggplot(excluded_wildcard_results, aes(x = reorder(Column_Pair, -P_Value), y = P_Value, fill = Color)) +
186   geom_bar(stat = 'identity') +
187   coord_flip() +
188   ylim(0, 0.06) +
189   labs(title = 'P-Values from Fisher's Exact Tests',
190        x = 'Column Pairs',
191        y = 'P-Value') +
192   scale_fill_identity() +
193   theme_minimal() +
194   theme(text = element_text(color = '#414141', size = 12, face = 'bold'),
195         plot.title = element_text(size = 18, face = 'bold', hjust = 0),
196         axis.title.x = element_text(size = 12),
197         axis.title.y = element_text(size = 12),
198         legend.position = 'none')
199
200

```

Correlations Code

```
1 #!/usr/bin/env python
2 # coding: utf-8
3
4 # In[ ]:
5
6
7 # Import necessary libraries
8 import pandas as pd
9 import itertools
10 import seaborn as sns
11 import matplotlib.pyplot as plt
12 from matplotlib.lines import Line2D
13 from scipy.stats import pearsonr, kendalltau, spearmanr
14
15 # Load the dataset from the provided Google Sheets link
16 sheet_id = "13G3puDCgEfShfNWZuQySuqyAFjoVxHqC_ockamL3oj0"
17 sheet_name = "Sheet1"
18 url = f"https://docs.google.com/spreadsheets/d/{sheet_id}/gviz/tq?tqx=out:csv&sheet={sheet_name}"
19
20 df = pd.read_csv(url,
21                 # Set first column as rownames in data frame
22                 index_col=0
23                 )
24
25 # Display the first few rows of the dataset
26 print(df.dtypes)
27 df.head(5)
28
29 # In[ ]:
30
31
32 # Filter columns with dtype 'int64'
33 df = df.select_dtypes(include=['int64'])
34 # Display the first few rows of the dataset
35 print(df.dtypes)
36 df
37
38 # In[ ]:
39
40
41 # List of wildcard strings
42 wildcards = ["Cluster_Found", "AI_DS_Use_Binary"]
43
44 # Exclude columns based on wildcard strings
45 exclude_cols = df.filter(regex='|'.join(wildcards)).columns
46 df = df.drop(columns=exclude_cols)
47 print(df.columns)
48 df
49
50 # In[ ]:
51
52
53 # Select only numeric columns
54 numeric_columns = df.select_dtypes(include=['float64', 'int64']).columns
55
56 # Create all possible pairs of numeric columns
57 numeric_pairs = list(itertools.combinations(numeric_columns, 2))
58
59 # Perform correlation tests for each pair
60 correlation_data = []
61 for pair in numeric_pairs:
62     # Remove missing values
63     pair_data = df[list(pair)].dropna()
64
65     # Calculate Kendall correlation
66     kendall_corr, kendall_p = kendalltau(pair_data[pair[0]], pair_data[pair[1]])
67
68     # Calculate Spearman correlation
69     spearman_corr, spearman_p = spearmanr(pair_data[pair[0]], pair_data[pair[1]])
70
71     # Store the results
72     correlation_data.append({
73         'Variable 1': pair[0],
74         'Variable 2': pair[1],
75         'Kendall Correlation': kendall_corr,
76         'Kendall p-value': kendall_p,
77         'Spearman Correlation': spearman_corr,
78         'Spearman p-value': spearman_p
79     })
80
81 # Convert the results to a DataFrame
82 correlation_df = pd.DataFrame(correlation_data)
83
84 # Print the results
85 correlation_df
86
87 # In[ ]:
88
89
90 # Check if any of the tests were significant at the 0.05 level
91 significant_correlations = correlation_df[
92     (correlation_df['Kendall p-value'] < 0.05) |
```

Figure 28: Complete Correlations Code in Python Language

```

93 |         (correlation_df['Spearman p-value'] < 0.05)
94 |     ]
95 |
96 | significant_correlations
97 |
98 | # Create a Summary
99 | summary = []
100 |
101 | for _, row in significant_correlations.iterrows():
102 |     summary.append(f"For the variables '{row['Variable 1']}' and '{row['Variable 2']}':")
103 |     summary.append(f" - Kendall Correlation: {row['Kendall Correlation']:.3f} (p-value: {row['Kendall p-value']:.3f})")
104 |     summary.append(f" - Spearman Correlation: {row['Spearman Correlation']:.3f} (p-value: {row['Spearman p-value']:.3f})")
105 |     summary.append("")
106 |
107 | text_summary = "\n".join(summary)
108 | print(text_summary)
109 |
110 | import matplotlib.pyplot as plt
111 | import numpy as np
112 |
113 | # Data for the table
114 | columns = ['Variable 1', 'Variable 2', 'Kendall Correlation', 'Kendall p-value', 'Spearman Correlation', 'Spearman p-value']
115 | cell_text = significant_correlations.values.tolist()
116 |
117 | # Plotting
118 | fig, ax = plt.subplots(figsize=(10, 2))
119 | ax.axis('off')
120 |
121 | # Creating the table
122 | table = ax.table(cellText=cell_text, colLabels=columns, cellLoc='center', loc='center')
123 | table.auto_set_font_size(False)
124 | table.set_fontsize(10)
125 | table.auto_set_column_width(col=list(range(len(columns))))
126 |
127 | # Variables to control row heights
128 | header_height = 0.12
129 | first_row_height = 0.12
130 | last_row_height = 0.12
131 | other_rows_height = 0.12
132 |
133 | # Adjusting row heights and styles
134 | cellDict = table.get_cellDict()
135 | for i in range(0, len(cell_text) + 1):
136 |     for j in range(0, len(columns)):
137 |         if i == 0:
138 |             cellDict[(i, j)].set_height(header_height)
139 |             cellDict[(i, j)].set_facecolor('#0F95D7')
140 |             cellDict[(i, j)].set_text_props(color='white', weight='semibold')
141 |         elif i == 1:
142 |             cellDict[(i, j)].set_height(first_row_height)
143 |         elif i == len(cell_text):
144 |             cellDict[(i, j)].set_height(last_row_height)
145 |             cellDict[(i, j)].set_facecolor('white')
146 |         else:
147 |             cellDict[(i, j)].set_height(other_rows_height)
148 |             if i % 2 == 0:
149 |                 cellDict[(i, j)].set_facecolor('#FAFAFA')
150 |             if j == 0:
151 |                 cellDict[(i, j)].set_facecolor('#FF2700')
152 |                 cellDict[(i, j)].set_text_props(color='white', weight='semibold')
153 |
154 | table.scale(1, 2.7)
155 |
156 | # Title
157 | ax.set_title('Significant Correlations', fontsize=15, fontweight='semibold', color='#414141')
158 |
159 | # Set the facecolor of the figure and axes to white
160 | fig.set_facecolor('white')
161 |
162 | plt.show()
163 |
164 | # Defining the correlation_matrix using the significant correlations data
165 | correlation_matrix = significant_correlations[['Kendall Correlation', 'Spearman Correlation']].values
166 | correlation_types = ['Kendall', 'Spearman']
167 |
168 |
169 | # Adjusting the size of the legend and axis text
170 | legend_fontsize = 8
171 | axis_fontsize = 12
172 |
173 | # Plotting the heatmap with adjusted text sizes
174 | plt.figure(figsize=(8, 1), facecolor='white')
175 | heatmap = sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', vmin=-1, vmax=1, center=0,
176 |                       yticklabels=significant_correlations['Variable 2'].values,
177 |                       xticklabels=correlation_types, cbar_kws={'label': 'Correlation Coefficient'})
178 | cbar = heatmap.collections[0].colorbar
179 | cbar.set_label('Correlation Coefficient', size=legend_fontsize)
180 | cbar.ax.tick_params(labelsize=legend_fontsize)
181 | plt.title('Heatmap of Correlation Coefficients', fontsize=10, fontweight='semibold', color='#414141')
182 | plt.ylabel('Variables', fontsize=axis_fontsize)
183 | plt.xlabel('Correlation Type', fontsize=axis_fontsize)
184 | plt.xticks(rotation=0, fontsize=axis_fontsize)
185 | plt.yticks(rotation=0, fontsize=axis_fontsize)
186 | plt.show()
187 |

```

Kruskal Wallis Code

```
0
1 # Import necessary libraries
2 import pandas as pd
3
4 # Load the dataset from the provided Google Sheets link
5 sheet_id = "13GJpuDCgEFShfNWZuQySuqyAFjovxHqC_ockaml3oj0"
6 sheet_name = "No_coment_text"
7 url = f"https://docs.google.com/spreadsheets/d/{sheet_id}/gviz/tq?tqx=out:csv&sheet={sheet_name}"
8
9 df = pd.read_csv(url, index_col=0)
10
11 # Check the frequencies of the 'Industry_Sector' column
12 industry_sector_frequencies = df['Industry_Sector'].value_counts()
13 industry_sector_frequencies
14
15 # In[ ]:
16
17 # Displaying the data types of the columns to identify continuous or ordinal columns
18 df.dtypes
19
20 # In[ ]:
21
22 from scipy.stats import kruskal
23
24 # Kruskal-Wallis H Test for AI_DS_Interest_Numerical across different Industry_Sector groups
25 groups_interest = [df['AI_DS_Interest_Numerical'][df['Industry_Sector'] == sector] for sector in df['Industry_Sector'].unique()]
26 stat_interest, p_interest = kruskal(*groups_interest)
27
28 # Kruskal-Wallis H Test for AI_DS_In_ProductDev_Numerical across different Industry_Sector groups
29 groups_productdev = [df['AI_DS_In_ProductDev_Numerical'][df['Industry_Sector'] == sector] for sector in df['Industry_Sector'].unique()]
30 stat_productdev, p_productdev = kruskal(*groups_productdev)
31
32 # Plotting the results in a table
33 import matplotlib.pyplot as plt
34 import numpy as np
35
36 # Data for the table
37 columns = ['', 'AI_DS_Interest_Numerical', 'AI_DS_In_ProductDev_Numerical']
38 rows = ['Statistic', 'p-value']
39 cell_text = [['Statistic', stat_interest, stat_productdev], ['p-value', p_interest, p_productdev]]
40
41 # Creating a dataframe to display the results
42 import pandas as pd
43 data = {'': ['Statistic', 'p-value'],
44        'AI_DS_Interest_Numerical': [stat_interest, p_interest],
45        'AI_DS_In_ProductDev_Numerical': [stat_productdev, p_productdev]}
46 results_df = pd.DataFrame(data)
47 print(results_df)
48 print()
49
50 # Plotting
51 fig, ax = plt.subplots(figsize=(6, 1))
52 ax.axis('off')
53
54 # Creating the table
55 table = ax.table(cellText=cell_text, colLabels=columns, cellLoc='center', loc='center')
56 table.auto_set_font_size(False)
57 table.set_fontsize(10)
58 table.auto_set_column_width(col=list(range(len(columns))))
59
60 # Variables to control row heights
61 header_height = 0.12
62 first_row_height = 0.12
63 last_row_height = 0.12
64 other_rows_height = 0.12
65
66 # Adjusting row heights and styles
67 cellDict = table.get_celld()
68 for i in range(0, len(cell_text) + 1):
69     for j in range(0, len(columns)):
70         if i == 0:
71             cellDict[(i, j)].set_height(header_height)
72             cellDict[(i, j)].set_facecolor('#0F95D7')
73             cellDict[(i, j)].set_text_props(color='white', weight='semibold')
74         elif i == 1:
75             cellDict[(i, j)].set_height(first_row_height)
76         elif i == len(cell_text):
77             cellDict[(i, j)].set_height(last_row_height)
78             cellDict[(i, j)].set_facecolor('white')
79         else:
80             cellDict[(i, j)].set_height(other_rows_height)
81             if i % 2 == 0:
82                 cellDict[(i, j)].set_facecolor('#FAFAFA')
83             if j == 0:
84                 cellDict[(i, j)].set_facecolor('#FF2700')
85                 cellDict[(i, j)].set_text_props(color='white', weight='semibold')
86
87 table.scale(1, 3)
```

Figure 29: Complete Kruskal Wallis Code in Python Language

```

96
97 # Title
98 ax.set_title('Kruskal-Wallis Test Results', fontsize=15, fontweight='semibold', color='#414141')
99
100 # Set the facecolor of the figure and axes to white
101 fig.set_facecolor('white')
102
103 plt.show()
104
105 # In[ ]:
106
107
108 # Combining multiple categories into fewer ones
109 df['Industry_Sector'] = df['Industry_Sector'].replace(['Agriculture', 'Construction', 'Manufacturing', 'Wholesale', 'Retail'], 'Primary/Secondary
Industry')
110 df['Industry_Sector'] = df['Industry_Sector'].replace(['Transportation', 'Hospitality', 'Entertainment', 'Real Estate'], 'Tertiary Industry A')
111 df['Industry_Sector'] = df['Industry_Sector'].replace(['IT', 'Finance', 'Education', 'Healthcare'], 'Tertiary Industry B')
112
113 # Kruskal-Wallis H Test for AI_DS_Interest_Numerical across different Industry_Sector groups
114 groups_interest = [df['AI_DS_Interest_Numerical'][df['Industry_Sector'] == sector] for sector in df['Industry_Sector'].unique()]
115 stat_interest, p_interest = kruskal(*groups_interest)
116
117 # Kruskal-Wallis H Test for AI_DS_In_ProductDev_Numerical across different Industry_Sector groups
118 groups_productdev = [df['AI_DS_In_ProductDev_Numerical'][df['Industry_Sector'] == sector] for sector in df['Industry_Sector'].unique()]
119 stat_productdev, p_productdev = kruskal(*groups_productdev)
120
121 # Creating a dataframe to display the results
122 import pandas as pd
123 data = {'': ['Statistic', 'p-value'],
124        'AI_DS_Interest_Numerical': [stat_interest, p_interest],
125        'AI_DS_In_ProductDev_Numerical': [stat_productdev, p_productdev]}
126 results_df = pd.DataFrame(data)
127 print(results_df)
128 print()
129
130 # Plotting the results in a table
131 import matplotlib.pyplot as plt
132 import numpy as np
133
134 # Data for the table
135 columns = ['', 'AI_DS_Interest_Numerical', 'AI_DS_In_ProductDev_Numerical']
136 rows = ['Statistic', 'p-value']
137 cell_text = [['Statistic', stat_interest, stat_productdev], ['p-value', p_interest, p_productdev]]
138
139 # Plotting
140 fig, ax = plt.subplots(figsize=(6, 1))
141 ax.axis('off')
142
143 # Creating the table
144 table = ax.table(cellText=cell_text, colLabels=columns, cellloc='center', loc='center')
145 table.auto_set_font_size(False)
146 table.set_fontsize(10)
147 table.auto_set_column_width(col=list(range(len(columns))))
148
149 # Variables to control row heights
150 header_height = 0.12
151 first_row_height = 0.12
152 last_row_height = 0.12
153 other_rows_height = 0.12
154
155 # Adjusting row heights and styles
156 cellDict = table.get_celld()
157 for i in range(0, len(cell_text) + 1):
158     for j in range(0, len(columns)):
159         if i == 0:
160             cellDict[(i, j)].set_height(header_height)
161             cellDict[(i, j)].set_facecolor('#0F95D7')
162             cellDict[(i, j)].set_text_props(color='white', weight='semibold')
163         elif i == 1:
164             cellDict[(i, j)].set_height(first_row_height)
165         elif i == len(cell_text):
166             cellDict[(i, j)].set_height(last_row_height)
167             cellDict[(i, j)].set_facecolor('white')
168         else:
169             cellDict[(i, j)].set_height(other_rows_height)
170             if i % 2 == 0:
171                 cellDict[(i, j)].set_facecolor('#FAFAFA')
172             if j == 0:
173                 cellDict[(i, j)].set_facecolor('#FF2700')
174                 cellDict[(i, j)].set_text_props(color='white', weight='semibold')
175
176 table.scale(1, 3)
177
178 # Title
179 ax.set_title('Kruskal-Wallis Fewer Categories Test Results', fontsize=15, fontweight='semibold', color='#414141')
180
181 # Set the facecolor of the figure and axes to white
182 fig.set_facecolor('white')
183
184 plt.show()
185
186 # In[ ]:
187
188
189 # Bootstrap resampling for Kruskal-Wallis test
190 def bootstrap_resample_kruskal(data, column, n_iterations=1000, valid_p_values_required=1000):
191     p_values = []
192     while len(p_values) < valid_p_values_required:
193         resampled_data = data.sample(n=len(data), replace=True)
194         groups = [resampled_data[resampled_data['Industry_Sector'] == sector][column] for sector in resampled_data['Industry_Sector'].unique()]

```

```

195
196     # Check if any group has all identical values or if there's only one group
197     if any(group.nunique() == 1 for group in groups) or len(groups) < 2:
198         continue
199
200     _, p = kruskal(*groups)
201     p_values.append(p)
202
203     # Optional: break the loop if it's running too long
204     if len(p_values) > n_iterations:
205         break
206
207     return np.mean(p_values)
208
209 # Bootstrap resampling for AI_DS_Interest_Numerical and AI_DS_In_ProductDev_Numerical
210 p_value_interest_resampled = bootstrap_resample_kruskal(df, 'AI_DS_Interest_Numerical')
211 p_value_productdev_resampled = bootstrap_resample_kruskal(df, 'AI_DS_In_ProductDev_Numerical')
212
213 # Creating a dataframe to display the results
214 import pandas as pd
215 data = {'': ['Statistic', 'p-value'],
216         'AI_DS_Interest_Numerical': [stat_interest, p_value_interest_resampled],
217         'AI_DS_In_ProductDev_Numerical': [stat_productdev, p_value_productdev_resampled]}
218 results_df = pd.DataFrame(data)
219 print(results_df)
220 print()
221
222 # Plotting the results in a table
223 import matplotlib.pyplot as plt
224 import numpy as np
225
226 # Data for the table
227 columns = ['', 'AI_DS_Interest_Numerical', 'AI_DS_In_ProductDev_Numerical']
228 rows = ['Statistic', 'p-value']
229 cell_text = [['Statistic', stat_interest, stat_productdev], ['p-value', p_value_interest_resampled, p_value_productdev_resampled]]
230
231 # Plotting
232 fig, ax = plt.subplots(figsize=(6, 1))
233 ax.axis('off')
234
235 # Creating the table
236 table = ax.table(cellText=cell_text, colLabels=columns, cellLoc='center', loc='center')
237 table.auto_set_font_size(False)
238 table.set_fontsize(10)
239 table.auto_set_column_width(col=list(range(len(columns))))
240
241 # Variables to control row heights
242 header_height = 0.12
243 first_row_height = 0.12
244 last_row_height = 0.12
245 other_rows_height = 0.12
246
247 # Adjusting row heights and styles
248 cellDict = table.get_celld()
249 for i in range(0, len(cell_text) + 1):
250     for j in range(0, len(columns)):
251         if i == 0:
252             cellDict[(i, j)].set_height(header_height)
253             cellDict[(i, j)].set_facecolor('#0F95D7')
254             cellDict[(i, j)].set_text_props(color='white', weight='semibold')
255         elif i == 1:
256             cellDict[(i, j)].set_height(first_row_height)
257         elif i == len(cell_text):
258             cellDict[(i, j)].set_height(last_row_height)
259             cellDict[(i, j)].set_facecolor('white')
260         else:
261             cellDict[(i, j)].set_height(other_rows_height)
262             if i % 2 == 0:
263                 cellDict[(i, j)].set_facecolor('#FAFAFA')
264             if j == 0:
265                 cellDict[(i, j)].set_facecolor('#FF2700')
266                 cellDict[(i, j)].set_text_props(color='white', weight='semibold')
267
268 table.scale(1, 3)
269
270 # Title
271 ax.set_title('Kruskal-Wallis Bootstrap Test Results', fontsize=15, fontweight='semibold', color='#414141')
272
273 # Set the facecolor of the figure and axes to white
274 fig.set_facecolor('white')
275
276 plt.show()
277

```

Clustering Code

```
13
14 # Installing the kmodes library which will be used for clustering categorical data.
15 !pip install -q kmodes
16
17 # Importing necessary libraries for data manipulation and clustering
18 import pandas as pd
19 import numpy as np
20 from kmodes.kmodes import KModes
21 from sklearn.manifold import TSNE
22 import matplotlib.pyplot as plt
23 import seaborn as sns
24 from collections import Counter
25
26 # Loading the dataset from a Google Sheet
27 #https://docs.google.com/spreadsheets/d/1YHU17WbGhNGuiejktWbyJWc7waP23pHCGfsu-mcbYd4/edit#gid=522748653
28 sheet_id = "1YHU17WbGhNGuiejktWbyJWc7waP23pHCGfsu-mcbYd4"
29 sheet_name = "Dataset_Full_Clean"
30 url = f"https://docs.google.com/spreadsheets/d/{sheet_id}/gviz/tq?tqx=out:csv&sheet={sheet_name}"
31
32 df = pd.read_csv(url, index_col=0)
33 original_df=df.copy()
34
35 # Displaying the first few rows of the dataset to get an overview
36 df.head()
37
38 ### Clustering Process
39 # 1. We use the `KModes` algorithm to cluster the categorical data. This algorithm is suitable for clustering categorical data.
40 # 2. The optimal number of clusters is chosen based on the "elbow" method, where we look for a point in the plot where the rate of decrease of the cost slows down.
41 #
42
43 # In[ ]:
44
45
46 # Finding 4 clusters using KModes
47 km = KModes(n_clusters=4, init='Huang', n_init=5, verbose=1)
48 clusters = km.fit_predict(df)
49 df['KModes_Cluster'] = clusters
50
51 # Analyze the clusters
52 cluster_analysis = df.groupby('KModes_Cluster').agg(lambda x: x.value_counts().index[0])
53
54 # Display the cluster analysis
55 cluster_analysis
56
57 ### Visualization of Clusters
58 # 1. To visualize the clusters in a 2D space, we use the t-SNE (t-distributed Stochastic Neighbor Embedding) technique. This technique reduces the dimensionality of the data while preserving the relationships between data points.
59 # 2. We plot the data points in the 2D space and color them based on their cluster assignments.
60 #
61
62 # In[ ]:
63
64
65 # Convert the categorical variables to numerical for t-SNE
66 df_encoded = df.apply(lambda x: pd.factorize(x)[0])
67
68 # Initialize t-SNE with perplexity=5
69 tsne = TSNE(n_components=2, verbose=1, perplexity=5, n_iter=300)
70
71 # Perform t-SNE
72 tsne_results = tsne.fit_transform(df_encoded)
73
74 # Add t-SNE results to the dataframe
75 df['tsne-2d-one'] = tsne_results[:, 0]
76 df['tsne-2d-two'] = tsne_results[:, 1]
77
78 # Display the first few rows of the dataset with t-SNE results
79 df.head()
80
81 ### Cluster Analysis
82 # 1. We analyze each cluster to understand its characteristics and the common traits of the companies within each cluster.
83 # 2. This analysis helps in understanding the different groups of companies based on their responses and provides insights into their practices and challenges.
84 #
85
86 # In[ ]:
87
88
89 # Set the FiveThirtyEight style for visualization
90 plt.style.use('fivethirtyeight')
91
92 # Plot the clusters
93 plt.figure(figsize=(16,10), facecolor='white')
```

Figure 30: Complete Clustering Code in Python Language

```

94 sns.scatterplot(
95     x='tsne-2d-one', y='tsne-2d-two',
96     hue='Kmodes_Cluster',
97     palette=sns.color_palette('tab10', 4),
98     data=df,
99     legend='full',
100    alpha=0.9,
101    s=500
102 )
103 plt.title('t-SNE Clusters')
104 plt.show()
105
106 # In[ ]:
107
108
109 from sklearn.tree import DecisionTreeClassifier, plot_tree
110 from sklearn.preprocessing import OrdinalEncoder
111 import matplotlib.pyplot as plt
112 import pandas as pd
113
114 # Assuming df is your dataframe with the Kmodes_Cluster column
115 # Merge the new dataframe with the original dataframe using a left join
116 df = original_df.merge(df[['Kmodes_Cluster']], left_index=True, right_index=True, how='left')
117
118 # Ordinal encode the categorical variables
119 encoder = OrdinalEncoder()
120 X_ordinal = encoder.fit_transform(df.drop('Kmodes_Cluster', axis=1))
121 y = df['Kmodes_Cluster']
122
123 # Fit a decision tree
124 clf = DecisionTreeClassifier(max_depth=3) # Limiting depth for visualization purposes
125 clf.fit(X_ordinal, y)
126
127 # Visualize the decision tree
128 plt.figure(figsize=(20, 10), facecolor='white')
129 plot_tree(clf, feature_names=df.drop('Kmodes_Cluster', axis=1).columns, class_names=y.unique().astype(str), filled=True, rounded=True, proportion=True)
130 plt.title('Decision Tree', fontsize=28, fontweight='semibold', color='#414141')
131 plt.show()
132
133 # Get feature importances
134 importances = clf.feature_importances_
135
136 # Get top 5 columns based on importance
137 sorted_indices = importances.argsort()[::-1][:-1]
138 top_columns = df.drop('Kmodes_Cluster', axis=1).columns[sorted_indices]
139 top_importances = importances[sorted_indices]
140
141 # Plot feature importance for the top 3 columns
142 colors = ['#0F95D7', '#FF2700', '#810F7C']
143 plt.figure(figsize=(8, 6), facecolor='white')
144 bars = plt.barh(top_columns[::-1], top_importances[::-1], color=colors[:len(top_columns)]) # Reverse the order for descending
145
146 # Add number labels to bars
147 for bar in bars:
148     plt.text(bar.get_width() + 0.005, bar.get_y() + bar.get_height()/2,
149             f'{bar.get_width():.4f}',
150             va='center', ha='left', color='#414141', fontsize=12)
151
152 plt.xlabel('Column Importance', fontsize=14, color='#414141')
153 plt.ylabel('Column', fontsize=14, color='#414141')
154 plt.title('Top 3 Column Importance', fontsize=18, fontweight='semibold', color='#414141')
155 plt.show()
156
157
158 # In[ ]:
159
160
161 df_backup = df.copy()
162
163 # Quantify AI and Data Science attributes for the specified columns
164 interest_mapping = {
165     "0 - No Interest": 0,
166     "1 - Low Interest": 1.25,
167     "2 - Moderate interest": 2.5,
168     "3 - High Interest": 5
169 }
170
171 adoption_mapping = {
172     "0 - No adoption": 0,
173     "1 - Low adoption": 1.25,
174     "2 - Moderate adoption": 2.5,
175     "3 - High adoption": 5
176 }
177
178 # Map the values
179 df['AI_DS_Interest_Score'] = df['AI_DS_Interest'].map(interest_mapping)
180 df['AI_DS_In_ProductDev_Score'] = df['AI_DS_In_ProductDev'].map(adoption_mapping)
181 df['AI_DS_Other_Activities_Score'] = df['AI_DS_Other_Activities'].map(adoption_mapping)
182
183 # Compute the AI and Data Science adoption score for each row based on the specified columns
184 df['AI_DS_Adoption_Score'] = df[['AI_DS_Interest_Score', 'AI_DS_In_ProductDev_Score', 'AI_DS_Other_Activities_Score']].mean(axis=1)
185
186 # Compute the average adoption score for each cluster
187 average_adoption_by_cluster = df.groupby('Kmodes_Cluster')['AI_DS_Adoption_Score'].mean()
188
189 # Rank the clusters based on the average adoption score
190 # The 'rank' method will rank the scores, with 1 being the highest rank
191 cluster_rankings = average_adoption_by_cluster.rank(ascending=False).astype(int)
192
193 # Map the rank to each row in the dataframe based on its cluster

```

```

194 df['AI_DS_Adoption_Rank'] = df['Kmodes_Cluster'].map(cluster_rankings)
195
196 # Group by 'Kmodes_Cluster' and compute the mean for 'AI_DS_Adoption_Score'
197 grouped_df = df.groupby('Kmodes_Cluster').agg({
198     'AI_DS_Adoption_Score': 'mean',
199     'AI_DS_Adoption_Rank': 'first' # Since rank is the same for all rows in a cluster, we can just take the first value
200 }).reset_index()
201
202 # Group by 'Kmodes_Cluster' and compute the mean for 'AI_DS_Adoption_Score'
203 grouped_df = df.groupby('Kmodes_Cluster').agg({
204     'AI_DS_Adoption_Score': 'mean',
205     'AI_DS_Adoption_Rank': 'first' # Since rank is the same for all rows in a cluster, we can just take the first value
206 }).reset_index()
207
208 # Rename the column
209 grouped_df = grouped_df.rename(columns={'AI_DS_Adoption_Score': 'AI_DS_Avg_Adoption_Score'})
210
211 # Sort by 'AI_DS_Adoption_Score' in descending order
212 grouped_df = grouped_df.sort_values(by='AI_DS_Avg_Adoption_Score', ascending=False)
213
214 # Display the sorted grouped dataframe
215 print(grouped_df[['Kmodes_Cluster', 'AI_DS_Avg_Adoption_Score', 'AI_DS_Adoption_Rank']])
216
217 # final dataset
218 df = df_backup.reset_index().merge(grouped_df[["Kmodes_Cluster", 'AI_DS_Avg_Adoption_Score']], on="Kmodes_Cluster"
219 ).sort_values(by="Company").set_index("Company")
220 df
221 # In[ ]:
222
223
224 #For each cluster, we'll calculate the percentage occurrence of each category in the columns
225 def calculate_percentage_newline(column_data):
226     total = len(column_data)
227     counts = Counter(column_data)
228     percentages = {key: f'{{(value/total)*100:.1f}}%' for key, value in counts.items()}
229     return '\n'.join([f'{{key}}: {{value}}' for key, value in percentages.items()])
230
231 # Group by 'Kmodes_Cluster' and apply the percentage calculation function to all columns
232 grouped_all_columns = df.reset_index().groupby('Kmodes_Cluster').agg({col: calculate_percentage_newline for col in df.reset_index().columns if col !=
233 'Kmodes_Cluster'})
234
235 grouped_all_columns
236 # In[ ]:
237
238
239 # Refining the cluster names based on their characteristics
240 # Note: These names are based on the analysis of the data and the provided documentation.
241 refined_cluster_names = {
242     0: 'Diverse Adopters',
243     1: 'Open-Minded Traditionalists',
244     2: 'Established Traditionalists',
245     3: 'Emerging Tech Enthusiasts'
246 }
247
248 # Update the cluster names in the dataset
249 df['Kmodes_Cluster_Names'] = df['Kmodes_Cluster'].map(refined_cluster_names)
250
251 # Display the first few rows of the dataset with the refined cluster names
252 df[['Company_Age', 'Company_Size', 'AI_DS_Interest', 'AI_DS_In_ProductDev', 'Industry_Sector', 'Kmodes_Cluster', 'Kmodes_Cluster_Names']].head(20)
253
254 # In[ ]:
255
256
257 # Setting size of our plot
258 fig, ax = plt.subplots(figsize=(8, 6), facecolor='white')
259
260 # Grouping by Kmodes_Cluster_Names and calculating the mean of AI_DS_Avg_Adoption_Score
261 cluster_scores = df.groupby('Kmodes_Cluster_Names')['AI_DS_Avg_Adoption_Score'].mean()
262
263 # Sorting the clusters for better visualization (optional)
264 cluster_scores = cluster_scores.sort_values()
265
266 # Plotting the bar chart with four distinct colors
267 bars = cluster_scores.plot(kind='bar', ax=ax, color=['#FFD700', '#FF2700', '#810F7C', '#0F95D7'])
268
269 # Title and labels
270 ax.set_title('Average Adoption Score by Cluster', fontsize=18, fontweight='semibold', color='#414141')
271 ax.set_xlabel('Cluster Names', fontsize=14, color='#414141')
272 ax.set_ylabel('Average Adoption Score', fontsize=14, color='#414141')
273
274 # Add number labels to bars
275 for bar in bars.patches:
276     ax.text(bar.get_x() + bar.get_width()/2, bar.get_height() + 0.005,
277            f'{{bar.get_height():.4f}}',
278            va='bottom', ha='center', color='#414141', fontsize=12)
279
280 # Create a legend to explain the context of Average Adoption Score
281 import matplotlib.patches as mpatches
282 # Create a multiline legend label
283 adoption_label = 'Adoption Mapping:\n' + '\n'.join([f'{{k}}: {{v}}' for k, v in adoption_mapping.items()])
284 adoption_patch = mpatches.Patch(color='none', label=adoption_label)
285 ax.legend(handles=[adoption_patch], loc='upper left')
286
287
288 # Displaying the plot
289 plt.show()
290
291

```