



UNIVERSIDADE CATÓLICA PORTUGUESA

Gestão e Análise de Informação de Múltiplas Fontes de Dados

por

Rafael Oliveira da Silva

Católica Porto Business School
fevereiro de 2020



UNIVERSIDADE CATÓLICA PORTUGUESA

Gestão e Análise de Informação de Múltiplas Fontes de Dados

Trabalho Final na modalidade de Dissertação
apresentado à Universidade Católica Portuguesa para obtenção do grau de
mestre em Gestão com a especialidade em Business Analytics

por

Rafael Oliveira da Silva

sob orientação de
Professor Doutor António Andrade

Católica Porto Business School
fevereiro de 2020

Agradecimentos

Ao longo da elaboração desta dissertação, pude contar com o apoio de várias pessoas. Deste modo, gostaria de agradecer principalmente à minha família por estarem sempre presentes e apoiarem-me constantemente durante o meu percurso de vida.

Aos meus amigos que tentaram sempre ajudar e motivar quando estava mais em baixo ou desmotivado, com a sua preocupação e amizade sempre incondicional.

Ao meu professor e orientador António Andrade, pela paciência e por estar sempre disponível.

Resumo

O *Business Intelligence* era tradicionalmente uma especialidade explorada fundamentalmente por profissionais da gestão de dados de TI, matemáticos e estatísticos. A tendência é agora, e cada vez mais, a descentralização do acesso e da análise de dados em toda a organização, havendo cada vez mais empresas a reconhecer o *Business Intelligence* como uma componente ao serviço da estratégica.

Este trabalho visa testar uma abordagem de *Business Intelligence* no desenvolvimento de um sistema de *analytics* de auxílio à gestão e análise de informação de múltiplas fontes de dados no meio empresarial. De forma a desenvolver o sistema de *analytics*, este projeto utilizou como caso de estudo a informação retirada de duas fontes distintas de dados de uma empresa que fabricava e vendia bicicletas, peças e acessórios. O design metodológico da investigação foi suportado pela abordagem CRISP-DM, tradicional na construção de um sistema de *analytics* que, neste caso, recorreu especialmente ao Power BI. No processo de preparação das duas bases de dados estruturadas foi necessária a utilização das ferramentas SQL Server Management, DAX e Excel. Como recurso de apoio ao *Business Intelligence*, foi concebida uma *dashboard* interativa com várias páginas, de forma a simular o potencial de gestão do Power BI nas diversas vertentes da empresa, tendo sido exploradas técnicas descritivas e prescritivas para a análise dos dados.

Atendendo às diversas vertentes exploradas, ficou comprovado o grande potencial do Power BI como ferramenta de suporte ao *Business Intelligence*, quer na perspetiva económica quer na perspetiva *analytics*.

Palavras-chave: *Business Intelligence*, Power BI, *Dashboard* e CRISP-DM

Abstract

Business Intelligence was traditionally a specialty explored primarily by professionals in IT data management, mathematicians and statisticians. The trend is now, and increasingly, the decentralization of access and data analysis across the organization, with more and more companies recognizing Business Intelligence as a component at the service of strategy.

This work aims to test a Business Intelligence approach in the development of an analytics system to aid the management and analysis of information from multiple data sources in the business environment. In order to develop the analytics system, this project used as a case study the information taken from two different sources of data from a company that manufactured and sold bicycles, parts and accessories. The methodological design of the investigation was supported by the CRISP-DM approach, traditional in the construction of an analytics system that, in this case, resorted especially to Power BI. In the process of preparing the two structured databases, it was necessary to use the tools SQL Server Management, DAX and Excel. As a resource to support Business Intelligence, an interactive dashboard with several pages was designed in order to simulate the management potential of Power BI in the different aspects of the company, and descriptive and prescriptive techniques for data analysis were explored.

Attending to the various aspects explored, the great potential of Power BI as a tool to support Business Intelligence was proven, both from an economic perspective and from an analytics perspective.

Keywords: Business Intelligence, Power BI, Dashboard and CRISP-DM

Índice

Agradecimentos	v
Resumo	vii
Abstract	ix
Índice	xi
Índice de Figuras.....	xv
1. Introdução.....	17
1.1. Enquadramento e Motivação	17
1.2. Objetivos do Estudo.....	18
1.3. Estrutura da Dissertação	18
2. A Tecnologia e a Gestão.....	20
2.1. Sistemas de Gestão Empresarial	20
2.1.1. Conceitos Elementares	20
2.1.2. Função dos Sistemas de Informação	22
2.1.3. Enterprise Resources Planning (ERP).....	36
2.2. Sistemas de Gestão e a Empresa em Contexto.....	41
2.2.1. Back-office e Front-office	41
2.2.2. Conexões com o Estado	41
2.2.3. Cloud e o Software as a Service.....	42
2.3. Big Data	44
3. Visualização Interativa de Dados.....	49
3.1. Tipologia dos Dados	49
3.1.1. Dados Estruturados.....	49
3.1.2. Dados Não Estruturados	50
3.1.3. Big Data e Small Data para ter Smart Data	50
3.2. Algoritmos de Tratamento de Dados.....	52
3.3. Visualização	54
3.4. Ferramentas e Tecnologia	56
3.4.1. Tipo utilizadores	56
3.4.2. Inteligência Artificial.....	58
3.4.3. Internet of Things	60

3.4.4. Elementos de Avaliação.....	60
3.4.5. Ferramentas de Análise	65
4. Metodologia.....	70
4.1. Questões e Finalidade.....	70
4.2. Paradigma Metodológico.....	70
4.3. Procedimentos de Investigação.....	72
4.3.1. Seleção do Objeto de Estudo.....	72
4.3.2. Tratamento	72
5. Construção de uma Dashboard em Power BI	75
5.1. Compreensão do Negócio.....	75
5.1.1. Power BI	75
5.1.2. Dashboard.....	76
5.1.3. Sistemas Semelhantes	76
5.2. Compreensão dos Dados	77
5.3. Preparação dos Dados	79
5.3.1. Structured Query Language	79
5.3.2. Excel.....	80
5.3.3. Data Analysis Expressions	81
5.4. Modelação	82
5.4.1. Recursos Humanos.....	82
5.4.2. Produção	83
5.4.3. Produtos.....	84
5.4.4. Compras.....	85
5.4.5. Vendas.....	86
5.5. Avaliação	89
5.6. Implementação	89
6. Conclusão.....	90
6.1. Conclusão dos Resultados Obtidos	90
6.2. Limitações e Trabalho Futuro	93
Referências Bibliográficas.....	95
Anexo.....	101

Apêndices	102
-----------------	-----

Índice de Figuras

Figura 1 - Atividades da Função de Sistemas de Informação (Varajão, 2002)....	22
Figura 2 - Atividades do Planeamento de Sistemas de Informação (Varajão, 2002).	24
Figura 3 - Atividades do Desenvolvimento de Sistemas de Informação (Varajão, 2002).	28
Figura 4 - Atividades da Exploração de Sistemas de Informação (Varajão, 2002).	32
Figura 5 - Anatomia de um Sistema Corporativo (Davenport, 1998).....	38
Figura 6 - Modelos de serviços da cloud (Abd Elmonem et al., 2016).....	44
Figura 7 - Hierarquia de aprendizagem (Gama et al., 2012)	53
Figura 8 - Magic Quadrant for Analytics and Business Intelligence Platforms (Howson et al., 2019)	65
Figura 9 - Metodologia CRISP-DM (Chapman et al., 2000)	73
Figura 10 - Medida para calcular a idade dos colaboradores	81
Figura 11 - Medida para indicar o destino dos produtos comprados pela empresa	82
Figura 12 - Modelo preditivo ARIMA em R Script	88
Figura 13 - Previsão ARIMA das vendas	88
Figura 12 - Diagrama do AdventureWorks (Microsoft, 2005)	101
Figura 13 - View das Compras	102
Figura 19 – Página dos Recursos Humanos na Dashboard.....	103
Figura 20 - Página da Produção na Dashboard.....	104
Figura 21 - Página dos Produtos na Dashboard.....	105
Figura 22 - Página das Compras na Dashboard.....	106
Figura 23 - Página das Vendas na Dashboard.....	107

Capítulo 1

Introdução

1. Introdução

1.1. Enquadramento e Motivação

O mercado empresarial é cada vez mais agressivo e vinga quem utiliza estratégias que permitem reunir não apenas o máximo de informação possível, mas também informação com maior qualidade, ou seja, a informação certa no momento e no local certo.

Devido aos avanços nas tecnologias de informação, existe a necessidade de transformar os dados previamente armazenados em informação, e esta em conhecimento útil, de modo a poder suportar a tomada de decisões. É nessa ótica que surge o *Business Intelligence* (BI), o qual permite agregar um vasto conjunto de tecnologias. Em particular, as técnicas de Data Mining que, a partir de dados em bruto, pretendem extrair padrões e tendências que possam ser utilizadas na melhoria da tomada de decisões de uma organização.

Com esta dissertação pretende-se desenvolver um estudo sobre a forma como o Power BI (software da Microsoft lançado em 2015), pode complementar o BI ao permitir retirar com facilidade um maior proveito da informação de uma base de dados, usando as suas ferramentas de visualização de dados, análise e recursos de conexão de dados (ex: Dynamics CRM, Salesforce, Excel, Oracle, etc).

1.2. Objetivos do Estudo

Este estudo tem como objetivo central simular um sistema que teste a capacidade de gestão e de análise de dados oriundos de múltiplas fontes de dados, através do Power BI.

Para atingir este objetivo, definiram-se os seguintes objetivos específicos:

- Conceber e desenvolver um sistema de *analytics* de apoio à gestão em diversos departamentos de uma empresa, suportado em várias fontes de dados;
- Explorar o potencial do Power BI como ferramenta de apoio ao BI, através da implementação de *dashboards* interativas;
- Explorar técnicas descritivas e prescritivas na análise dos dados.

1.3. Estrutura da Dissertação

A estrutura da dissertação encontra-se organizada ao longo de seis capítulos, cujo primeiro capítulo é a presente introdução. No capítulo 2 é efetuada uma revisão à literatura teórica, referente à área dos Sistemas de Gestão Empresarial, começando por introduzir alguns conceitos elementares, seguindo-se os Sistemas de Informação, onde são referidos as suas funções, tipos e contexto em que normalmente atuam, e a *Big Data*. Segue-se o capítulo 3, no qual é abordada a visualização interativa de dados, começando pelos tipos de dados existentes, seguindo-se os algoritmos de tratamento, a sua visualização, e por último, as ferramentas e tecnologia, com referência ao tipo de utilizadores, à inteligência artificial, elementos de avaliação das ferramentas e à sua análise.

Após esta exposição, sentiu-se necessidade de criar um capítulo relativo à metodologia aplicada (capítulo 4) com o planeamento necessário, de forma a levar a cabo as etapas seguintes, as ferramentas utilizadas, as técnicas e métricas que se pretendiam utilizar para o desenrolar da aplicação do presente caso,

tendo sido adotado o método de trabalho denominado CRISP-DM (Cross-industry standard process for data Mining), pois esta é uma metodologia especificamente desenhada para processos de mineração de dados.

No capítulo 5 é desenvolvida a *dashboard* referenciada nos objetivos, utilizando o Power BI, de forma a assim melhor compreender o potencial do BI.

Por último, no capítulo 6 esta dissertação é finalizada com a conclusão do trabalho.

Capítulo 2

A Tecnologia e a Gestão

2. A Tecnologia e a Gestão

2.1. Sistemas de Gestão Empresarial

2.1.1. Conceitos Elementares

Para enquadrar o presente estudo optamos por uma abordagem sintética e simples, porém abrangente, ao âmbito dos sistemas de informação, em que se insere, no contexto empresarial. Alguns conceitos são de seguida identificados e definidos para facilitar a compreensão do leitor ao longo da dissertação. Os conceitos essenciais para tal compreensão são os dados, a informação, o conhecimento, os sistemas, os sistemas de informação e as tecnologias de informação.

Os dados são factos ou eventos, imagens ou sons que podem ser pertinentes para o desempenhar de uma tarefa, no entanto por si só não é possível obter uma compreensão de determinados factos ou situações (Rascão, 2004).

Já informação é um conjunto de dados que quando obtida de forma e a tempo adequado, melhora o conhecimento de quem a recebe ficando ela mais habilitada a desenvolver determinada atividade ou a tomar determinada decisão (Galliers, 1987a).

Por outro lado, o conhecimento “é a capacidade de uma pessoa relacionar estruturas complexas de informação para um novo contexto. Novos contextos implicam mudança-ação, dinamismo. O conhecimento não pode ser partilhado,

embora a técnica e os componentes da informação possam ser partilhados” (Grenier & Metes, 1991).

Um Sistema é um grupo de elementos inter-relacionados ou de elementos interagindo formando um todo (O’Brien & Marakas, 1990).

O sistema de informação tem dois subsistemas: controlo e tomada de decisão (Zorrinho, 1991). Um sistema de informação pode ser definido como um conjunto de componentes inter-relacionados que recolhe (ou recuperam), processam, armazenam e distribuem informações para dar suporte à tomada de decisão e ao controlo numa organização. Além de apoiar a tomada de decisões, a coordenação e o controlo, os sistemas de informação também podem ajudar gestores e colaboradores a analisar problemas, visualizar assuntos complexos e criar novos produtos (School, 2002).

A tecnologia da informação é a aplicação de computadores e telecomunicações na recolha, processamento, armazenamento e disseminação de voz, gráficos, texto, imagem e informações numéricas (Flowers, 1988).

A gestão da informação pode ser entendida como a gestão eficaz de todos os recursos de informação relevantes para a organização, tanto de recursos gerados internamente como os produzidos externamente e fazendo apelo, sempre que necessário, à tecnologia de informação (Wilson, 1989).

Por ultimo, é essencial explicar o que é o BI, sendo que este é um termo abrangente que agrupa arquiteturas, ferramentas, bases de dados, aplicações e metodologias (Raisinghani, 2004). Um dos principais objetivos do BI é o facto de possibilitar o acesso e a manipulação de forma interativa (por vezes em tempo real) de dados, e permitir aos gestores e analistas o acesso a análises mais controladas e detalhadas. O processo de BI, em geral, baseia-se na transformação de dados em informação, posteriormente em decisões e por fim em ações (Turban, Delen, Sharda, & King, 2007).

2.1.2. Função dos Sistemas de Informação

Sendo a informação um recurso fundamental na gestão de qualquer organização, ao promover a otimização do sistema de informação, a Função dos Sistemas de Informação (FSI) assume um papel decisivo para o seu sucesso.

Podemos idealizar a FSI através de quatro atividades principais: o Planeamento de Sistemas de Informação (PSI), o Desenvolvimento de Sistemas de Informação (DSI), a Exploração de Sistemas de Informação (ESI) e a Gestão de Sistemas de Informação (GSI) (Varajão, 2002).



Figura 1 - Atividades da Função de Sistemas de Informação (Varajão, 2002).

Estas atividades da FSI estão relacionadas entre si e são interdependentes (figura 1). Elas funcionam num processo contínuo e interativo, podendo estabelecer-se uma ordem das atividades, na qual o SI é pensado (PSI), produzido (DSI) e finalmente utilizado (ESI) (Varajão, 2002).

Ao longo do tempo, os SI precisam de se adaptar de forma a poderem suportar as necessidades da organização. Assim, a FSI deve ser encarada como contínua, mas também com carácter cíclico. Em cada geração do sistema, as suas atividades

alimentam-se mutuamente, mantendo entre si fortes ligações. O PSI permite identificar os sistemas necessários, seguindo-se o DSI que será responsável pelo seu desenvolvimento. Cabe à ESI assegurar a correta utilização dos sistemas. A GSI gere todas as atividades (Varajão, 2002).

2.1.2.1. Planeamento de Sistemas de Informação

O Planeamento de Sistemas de Informação (PSI) pode ser entendido como a fase em que uma organização decide qual o futuro que pretende para o seu SI, de que modo as TI o deverão suportar e qual a melhor forma de concretizar esse processo (Amaral, 1994). O objetivo será garantir o desenvolvimento e utilização dos SI de forma a obter o máximo retorno da sua operação (J. M. Ward & Griffiths, 1996)

Ou seja, o PSI corresponde fundamentalmente a um método de integração das diversas vertentes do SI com o modelo de planeamento da organização, permitindo desta forma ligar diretamente este modelo à gestão operacional do DSI, com recurso à aquisição de TI, e ao desenvolvimento, exploração e manutenção de aplicações, entre outros (Galliers, 1987b). Desta forma, também atua como um processo que permite integrar e intervir na formulação das estratégias organizacionais.

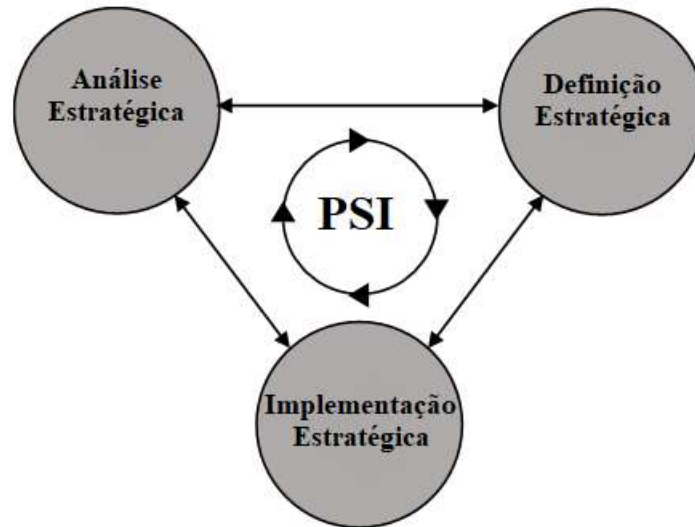


Figura 2 - Atividades do Planejamento de Sistemas de Informação (Varajão, 2002).

No PSI, todas as atividades envolvidas têm um relacionamento entre si, dependendo umas das outras e podendo suceder em simultâneo. Trata-se de um processo analítico e criativo (Bunn, Bartlett, & Mclean, 1989) que não constitui uma atividade isolada no tempo, sendo antes um processo contínuo (Robson, 1994; J. M. Ward & Griffiths, 1996). Com o decorrer do tempo, as políticas e planos que dele resultam, deverão evoluir de forma a acompanhar a implementação e evolução dos próprios sistemas (CCTA, 1993)

Pode-se analisar o PSI através de três atividades principais (fig.2), em que cada uma procura obter resposta para uma certa questão (Varajão, 2002):

- **Análise Estratégica** → *Qual é a situação atual?*
- **Definição Estratégica** → *O que se pretende para o futuro?*
- **Implementação Estratégica** → *O que fazer para o conseguir?*

Através do PSI deverá ser possível obter uma visão global do SI da organização, incluindo ao mesmo tempo todos os elementos que permitam a sua operacionalização (desenvolvimento) (Carvalho & Amaral, 1993). Deverá identificar a arquitetura e o ambiente da organização, desenvolver uma

perspetiva para o futuro e estipular as regras orientadoras dos processos e de utilização das TI (Varajão, 2002). Ou seja, em tempo de “transformação digital” é mais clara a importância das tecnologias nessa transformação, ficando a parecer neste ponto claro que a tecnologia faz inevitavelmente parte da estratégia quando durante anos procurou “alinhar” com a estratégia, sendo esta tratada como algo acessório para otimização de processos internos.

Análise Estratégica

O planeamento de sistemas que permitam suportar verdadeiramente uma organização, exige um grande conhecimento do negócio, das estratégias e objetivos, do ambiente competitivo e do meio externo que a influencia. Portanto, é fundamental uma boa perceção do ambiente interno e externo, entendendo as restrições, pressões e fatores que moldam a sua atividade (Ricardo, 1995).

Na base de todo o processo de PSI está a Análise Estratégica. O seu objetivo é identificar e caracterizar da melhor forma a realidade da organização e entender as suas necessidades de SI, não só através de uma profunda análise do negócio e ambiente da organização, mas também da revisão dos atuais sistemas e respetivas matérias, tendo em vista a identificação de problemas e oportunidades para o seu desenvolvimento e exploração. Além de uma visão precisa da atual situação, o resultado da Análise Estratégica deverá também espelhar as pretensões e direções estratégicas da organização e do respetivo SI (Varajão, 2002).

Definição Estratégica

A Análise Estratégica permite preparar o trabalho para a fase de Definição Estratégica, uma vez que possibilita uma boa perceção tanto da atual como da futura situação da organização e do seu SI, dos objetivos que se deseja alcançar,

das suas estratégias e da sua posição em relação a outros ambientes (Varajão, 2002).

Ponderando o futuro pretendido para o negócio de acordo com o futuro previsível do meio envolvente e das oportunidades e ameaças encontradas, a Definição Estratégica deverá determinar o futuro desejado para o SI, bem como este deverá ser suportado pelas TI, isto é, qual o papel desejado para o SI na estrutura e atividade da organização (Amaral, 1994).

Dependendo das dificuldades, oportunidades e cenários alternativos previamente identificados na Análise Estratégica e tendo por base o conhecimento dos diferentes ambientes, a Definição Estratégica deverá permitir definir um conjunto de grandes opções para o futuro do SI. No contexto estratégico da organização, tais opções devem ser analisadas de modo a avaliar os seus méritos relativos, de forma a selecionar as potencialmente mais vantajosas e viáveis. Assim, a finalidade da Definição Estratégica é desenvolver as estratégias que melhor possam beneficiar a organização, devendo estas fornecer um bom mecanismo defensivo contra previsíveis ameaças futuras e a possibilidade de explorar as oportunidades potenciais (Varajão, 2002).

Implementação Estratégica

Depois de concluída a definição de estratégias, é preciso transpô-las para planos de ação que sejam viáveis (Laudon & Laudon, 1994), de forma a que os recursos sejam devidamente conseguidos e alocados, que seja estabelecida uma organização adequada e que as pessoas sejam estimuladas de maneira a poderem ser atingidos os objetivos inerentes a essas estratégias (J. M. Ward & Griffiths, 1996).

Os planos para a implementação são habitualmente desenvolvidos em cada ano numa perspetiva anual e estabelecem como e quando os objetivos da estratégia deverão ser conseguidos. Permitem definir e orçamentar as ações a

desenvolver a curto prazo e os meios necessários, que habitualmente inclui cronogramas para a compra de equipamentos e suportes lógicos, cronogramas para o desenvolvimento de sistemas, planeamentos técnicos, planos de recrutamento e alocação de recursos humanos, de reorganização interna, de suporte aos utilizadores, entre outros (Varajão, 2002).

À medida que vão sendo implementadas, as estratégias deverão ser constantemente avaliadas, supervisionadas e controladas no modo como conseguem sustentar a organização no encaço dos seus objetivos (J. M. Ward & Griffiths, 1996). Para a que ao longo do tempo o desenvolvimento de SI possa efetivamente suportar a evolução das exigências de informação da organização, as estratégias necessitam de responder a mudanças, tanto da organização como tecnológicas, devendo por isso ser acompanhadas, revistas e ajustadas (Bunn et al., 1989) de acordo com as necessidades do negócio.

As estratégias e planos de SI têm constantemente de se manter consistentes com a visão e planos de negócio, por isso devem ser revistos em sintonia com as revisões da estratégia e visão de negócio (J. M. Ward & Griffiths, 1996). O controlo da implementação das estratégias permite assegurar que os SI são desenvolvidos e explorados de forma eficaz e mantidos em concordância com alterações das circunstâncias, e se houver desvios nos planos, deve assegurar que sejam efetuadas as devidas ações corretivas (Bunn et al., 1989). Permite também averiguar se os objetivos e os benefícios esperados são alcançados ou não.

Os requisitos iniciais dos sistemas podem ser alterados devido a possíveis mudanças, como por exemplo, a alteração das condições internas ou da natureza do ambiente de negócio e os avanços tecnológicos. Por esse motivo, todos os sistemas devem ser permanentemente acompanhados, de modo a garantir que satisfazem as necessidades da organização e a identificar quando devem ser preservados, aperfeiçoados ou substituídos (McKeown & Leitch, 1993).

2.1.2.2. Desenvolvimento de Sistemas de Informação

A definição do futuro ambicionado para o SI é da responsabilidade do PSI. Um sistema que seja desenvolvido sem o adequado planeamento, irá provavelmente conduzir a uma grande insatisfação dos seus utilizadores e acabará por cair em desuso (Kendall & Kendall, 1992). Por outro lado, a falta de um processo de PSI poderá favorecer o desenvolvimento de “ilhas tecnológicas” com todos os inevitáveis inconvenientes, tais como, incompatibilidades entre tecnologias utilizadas, problemas de integração dos diversos tipos de SI, desperdício de recursos, redundâncias, confusão no desenvolvimento e na utilização, incerteza sobre o aproveitamento de oportunidades, etc. (Amaral, 1994). Desta forma, o PSI assume-se como um precursor indispensável do DSI, uma vez que é responsável pela definição de planos, políticas e normas que conduzirão todas as suas atividades (Varajão, 2002).

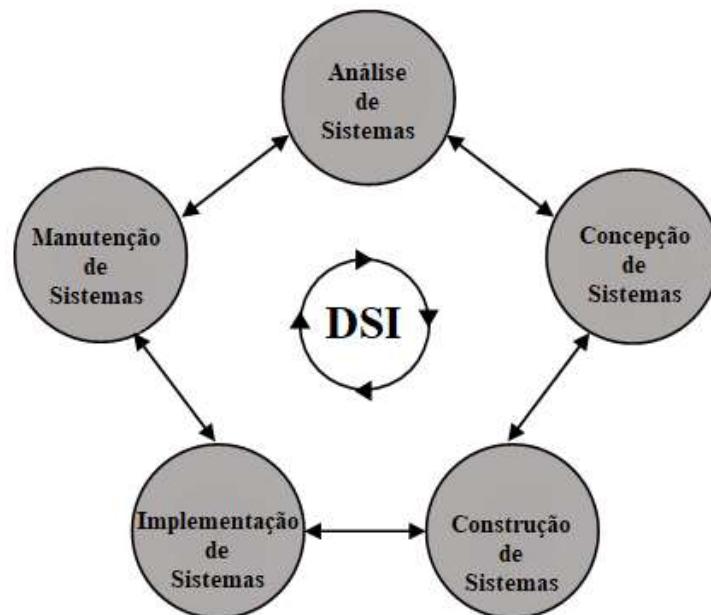


Figura 3 - Atividades do Desenvolvimento de Sistemas de Informação (Varajão, 2002).

Essencialmente, o DSI é caracterizado por ser um processo de mudança que busca a melhoria do desempenho de um (sub)sistema de informação (Carvalho

& Amaral, 1993), e refere-se a todas as atividades envolvidas na produção de SI (Laudon & Laudon, 1994) que suportem devidamente a organização, não apenas no apoio aos seus processos, como também na criação de vantagens competitivas.

Análise de Sistemas

A Análise de Sistemas envolve o estudo ponderado dos requisitos de informação da organização e dos utilizadores finais, das atividades, dos recursos e dos sistemas existentes, de forma a detalhar a natureza dos sistemas propostos pelo PSI (Varajão, 2002).

Trata-se de uma das atividades mais complexas do DSI e na qual habitualmente se colocam dificuldades. Da definição completa e precisa de requisitos, dependerá qualitativamente o sucesso do DSI. No caso de tal não ser conseguido, poderá verificar-se o desenvolvimento de um sistema que não sirva os interesses da organização, pelo menos em toda a sua extensão (Martin, DeHayes, Hoffer, & Perkins, 1994).

Construção de Sistemas

A Construção de Sistemas é uma atividade fundamentalmente técnica (Barnatt, 1996), na qual as especificações dos sistemas previamente definidas, são convertidas em software e hardware (Dickson & Wetherbe, 1985; McKeown & Leitch, 1993) que irão desempenhar as desejadas funções de suporte às necessidades de informação da organização. É também nesta atividade que se desenvolve a documentação que permite explicitar o sistema a programadores e utilizadores (Alter, 1996).

Assim, esta atividade encarrega-se da aquisição e/ou desenvolvimento de todo o software e hardware necessário (Lucas Jr., 1992), da documentação de suporte (O'Brien, 1993) e da integração de todas as componentes num SI funcional

(McKeown & Leitch, 1993), tal como foi definido nas especificações do sistema (Kendall & Kendall, 1992).

Implementação de Sistemas

Um novo sistema pode ser implementado após a sua construção (Barnatt, 1996). Este é o processo de tornar o sistema operacional na organização (Alter, 1996), após o qual poderá começar a ser utilizado (Lucas Jr., 1992; McKeown & Leitch, 1993). Esta atividade envolve diversas atividades, desde a preparação das instalações, da instalação de equipamentos e software, de conversão dos sistemas existentes para os novos sistemas, à formação de utilizadores e de especialistas que serão os responsáveis pela sua exploração (Varajão, 2002).

É importante estar consciente de que um sistema, apesar de bem concebido e desenvolvido, poderá falhar se não for devidamente implementado (O'Brien, 1993). Uma vez que um novo sistema habitualmente conduz à alteração dos processos existentes, é fundamental uma formação adequada dos recursos humanos, para que estes o possam entender e utilizar corretamente (Sager, 1990).

Manutenção de Sistemas

A Manutenção é o trabalho desenvolvido para melhorar e corrigir os sistemas após a sua implementação (Alter, 1996) e envolve habitualmente recursos significativos (Kendall & Kendall, 1992). Assim, apesar de a implementação de um novo sistema estar terminada, não quer dizer que o seu desenvolvimento esteja concluído (Barnatt, 1996), pois existe uma permanente necessidade o manter e desenvolver.

Muito provavelmente, ao longo do tempo o sistema irá necessitar da incorporação de novas tecnologias, de alterações devido à evolução dos requisitos de informação, de correção de eventuais erros do software, etc. (Varajão, 2002).

À medida que o sistema é usado, os utilizadores vão conhecendo-o melhor e desenvolvendo novas ideias para o aperfeiçoar. Apesar de bem-sucedido, um sistema estará sempre inacabado, e a sua longevidade dependerá da aptidão da FSI em mantê-lo em harmonia com as mudanças de requisitos a que a organização está frequentemente sujeita (Varajão, 2002).

2.1.2.3. Exploração de Sistemas de Informação

O PSI e o DSI produzem fundamentalmente uma adequada combinação de SI/TI que suportam a organização. Além disso, promovem desenvolvimento da capacidade da própria organização para a uma eficiente e eficaz utilização. Mas os sistemas não são estáticos, nem funcionam perpetuamente por si só, antes pelo contrário, mesmo depois de completar o seu desenvolvimento, necessitam continuamente de supervisão, suporte e manutenção (Varajão, 2002).

A Exploração de Sistemas de Informação (ESI) é a atividade que permite assegurar o bom funcionamento do SI. Além de ser responsável pela operação dos sistemas existentes, também assume um papel decisivo na definição de futuras estratégias, restringindo ou facilitando o PSI e o DSI. De referir que, a seja qual for a qualidade do planeamento ou do desenvolvimento de um sistema, o sucesso deste dependerá sempre da qualidade da sua exploração (Varajão, 2002).



Figura 4 - Atividades da Exploração de Sistemas de Informação (Varajão, 2002).

Na ESI existem três atividades focadas nos diferentes recursos de SI: Operação dos Sistemas, Administração de Recursos Humanos e Administração de Tecnologias de Informação, uma atividade centrada na Aquisição de Serviços e Recursos, e finalmente, Atividades de Carácter Diversificado, na qual se podem enquadrar todas as atividades que ocorram pontualmente no âmbito da FSI, como por exemplo, da Investigação e Desenvolvimento (Varajão, 2002).

Contrariamente ao PSI e DSI, dada a natureza da ESI, não faz sentido atribuir a esta uma sequência lógica de atividades. Enquanto que no planeamento e no desenvolvimento de sistemas existem diversas atividades que podem ser consideradas fases desde o pensar dos sistemas até à sua concretização, no caso da ESI, as atividades estão naturalmente inter-relacionadas, mas ocorrem sempre simultânea e continuamente, como ilustra a figura 4 (Varajão, 2002).

Operação do Sistema

Para assegurar o funcionamento e utilização nas melhores condições, um novo sistema depois de instalado precisa de ser acompanhado permanentemente, de

modo a resolver eventuais problemas que possam ocorrer e para fornecer suporte à organização na sua exploração. Algumas das atividades fundamentais para garantir uma eficaz operação dos sistemas são: a segurança e controlo da informação; a administração de dados; os procedimentos e o suporte à organização (Varajão, 2002).

Administração das Tecnologias de Informação

A Administração de TI deverá garantir um bom funcionamento das TI, assegurando a resolução atempada de eventuais problemas que possam surgir durante a sua exploração.

Esta atividade deverá de ser desenvolvida em colaboração com o DSI, uma vez que após a Implementação de Sistemas, que é responsável pela aquisição e instalação das TI, cabe à Administração de TI a manutenção e controlo das mesmas. Assim, estão a cargo desta atividade várias tarefas: instalação e configuração de equipamentos e de sistemas, planeamento e afinação da capacidade dos sistemas, recolha de equipamentos para reparação ou substituição, instalação e administração das redes informáticas, controlo de versões instaladas, controlo de contratos de manutenção das TI, etc (Varajão, 2002).

Administração de Recursos Humanos

A competência da organização para adquirir, desenvolver e preservar recursos humanos qualificados e aptos será fundamental para o retorno do seu SI a longo prazo (J. Ward, 1995), e o seu sucesso dependerá em primeiro lugar da qualidade dos recursos humanos responsáveis pelo respetivo planeamento, desenvolvimento, exploração e gestão. Por esse motivo, esta é uma atividade que assume uma importância determinante na FSI, devendo assegurar que a mesma

está devidamente preparada para lidar com as exigências existentes e para encarar as futuras necessidades técnicas (Varajão, 2002).

Habitualmente, esta atividade é desenvolvida em colaboração com a função de recursos humanos da organização, mantendo esta última a seu cargo as questões de carácter laboral, tais como a celebração de contratos, processos de despedimento, relações com sindicatos, etc. (Varajão, 2002).

Aquisição de Serviços e Recursos

Uma vez que as grandes iniciativas de aquisição deverão ser enquadradas no PSI, DSI e Atividades de Carácter Diversificado, esta atividade apenas considera a aquisição corrente de recursos e serviços, envolvendo diversas ações tais como a pesquisa no mercado de soluções, a procura de novas tecnologias de hardware e de software, o planeamento e aquisição de necessidades de equipamento, software e economato, o controlo das existências de forma a assegurar stocks mínimos, e as tarefas de elaboração de propostas, de análise e verificação de encomendas a fornecedores (Varajão, 2002).

Atividades de Carácter Diversificado

Além das diversas atividades oportunamente analisadas, existem ainda outras intimamente relacionadas com diferentes aspetos burocráticos, com as práticas e com a própria cultura da organização. Exemplo disso são o correio eletrónico, a gestão de correspondência (interna e externa), os contactos telefónicos, a definição e normalização de procedimentos, a gestão da qualidade, a difusão interna de informação, os meios de transporte e logística, ou os eventos organizados pela própria organização (Varajão, 2002).

Pontualmente, podem também surgir atividades desenvolvidas com o intuito de explorar eventuais oportunidades de SI/TI ou para resolver uma determinada situação. Tratando-se de uma função essencial nas atuais organizações, a

Investigação e Desenvolvimento em SI poderá tipicamente constituir um destes projetos, embora caso se trate de uma organização com um contínuo investimento em Investigação e Desenvolvimento, esta deva ser definida como uma atividade da ESI (Varajão, 2002).

2.1.2.4. Gestão de Sistemas de Informação

Em última instância a GSI tem a seu cargo a gestão de toda a FSI, uma vez que é responsável pelas atividades que permitem garantir a existência de um SI adequado às necessidades de informação numa organização. Tais atividades são:

- O Planeamento - diz respeito ao desenvolvimento de programas de ação e a definição de políticas, de objetivos e estratégias, abrangendo a perceção e análise de oportunidades e problemas. De facto, o planeamento das atividades de PSI, DSI e ESI é da competência da GSI, devendo esta desempenhar um papel especialmente ativo e direto no PSI, uma vez que o PSI é entendido fundamentalmente como uma tarefa de gestão que trata da integração das questões relacionados com SI no processo de planeamento da organização (Varajão, 2002).
- A Estruturação - constitui a atribuição de tarefas a indivíduos e grupos através da delegação de autoridade e responsabilidade. Basicamente, trata da forma como as atividades devem ser organizadas e os recursos utilizados (Varajão, 2002).
- A Direção - consiste na liderança da organização e que é exercida através da comunicação e motivação das pessoas. Compete à GSI a coordenação e fiscalização do trabalho, motivar e recompensar ou sancionar, resolver eventuais conflitos, etc. Trata-se de um dos papéis fundamentais do gestor ao definir um modelo de liderança e motivação que permita conciliar os objetivos organizacionais com os objetivos individuais (Varajão, 2002).

- O Controlo - corresponde à observação e medida do desempenho da organização e das atividades do ambiente, de forma a ajustar os planos e atividades se tal for necessário. Em termos práticos consiste em vários tipos de controlo: do cumprimento dos planos estabelecidos, das equipas de trabalho, do orçamento, de custos, de faturação e fornecimento de serviços ou de recursos, etc. (Varajão, 2002).

2.1.3. Enterprise Resources Planning (ERP)

Uma empresa integrada é um pré-requisito para todos os estágios evolutivos subsequentes da mesma, ou seja, uma empresa recetiva e centrada no cliente, renovadora, colaborativa e bem informada (Kale, 2016).

Os sistemas ERP vieram permitir a integração de unidades de negócios heterogéneas e dispare, funções e processos para coordenar, cooperar e colaborar no alinhamento das operações de negócios da empresa com sua estratégia corporativa (Kale, 2016).

O desenvolvimento de melhores tecnologias cliente/servidor permitiram a implementação de novos sistemas, tendo o conceito de ERP surgido no final da década de 80. Mas nos últimos 30 anos estes sistemas evoluíram consideravelmente, como resultado de melhorias contínuas na gestão de negócios e no desenvolvimento de novas tecnologias de informação (Kale, 2016).

Consistindo em vários conjuntos integrados de módulos de software que partilham dados e fornecem conectividade, os sistemas ERP vieram permitir reduzir muito o custo de manutenção de sistemas. O fornecedor do sistema assume a responsabilidade de aperfeiçoar as funcionalidades, fornecendo atualizações técnicas e incorporando as mais recentes regulamentações específicas do país e requisitos legais (Kale, 2016).

Após o registo dos dados no sistema, estes ficam disponíveis para todas as áreas da empresa, permitindo que a informação necessária sobre os processos da

empresa seja representada de forma consistente e atualizada em todas as áreas de negócios e em qualquer momento. Desta forma, o ERP fornece os dados operacionais mais atuais, corretos, consistentes e completos que podem ser disponibilizados no *data warehouse* corporativo para análise e criação de relatórios (Kale, 2016).

Atualmente, o ERP constitui um "*backbone*" corporativo multifuncional que integra e automatiza muitos processos internos de negócios e sistemas de informação dentro das funções de vendas e distribuição, produção, logística, contabilidade e RH de uma empresa (Kale, 2016).

2.1.3.1. Perspetiva de Davenport

A implementação de sistemas corporativos nem sempre é um processo simples e direto que aumenta automaticamente a eficiência. Na realidade, as empresas encontram vários problemas ao tentar instalar sistemas corporativos, que podem acabar por se transformar num pesadelo. Parte da culpa pelas falhas dos sistemas corporativos está nos desafios técnicos da implementação. Esses sistemas são caros e complexos, exigindo grandes investimentos em dinheiro, tempo e conhecimento para instalação. No entanto, a principal razão pela qual os sistemas corporativos falham, está relacionada com os negócios: as empresas falham em reconciliar os imperativos tecnológicos dos sistemas corporativos com as necessidades comerciais da própria empresa (Davenport, 1998).

No centro de um sistema corporativo, há uma única base de dados abrangente. Quando novas informações são inseridas num único local, as informações relacionadas são atualizadas automaticamente. Apesar da sua capacidade de integração, o sistema corporativo não é flexível. É uma solução genérica que nem sempre se encaixa nas características individuais de cada empresa (Davenport, 1998).

Segundo Davenport, um sistema corporativo é ilustrado da seguinte forma (figura 5):

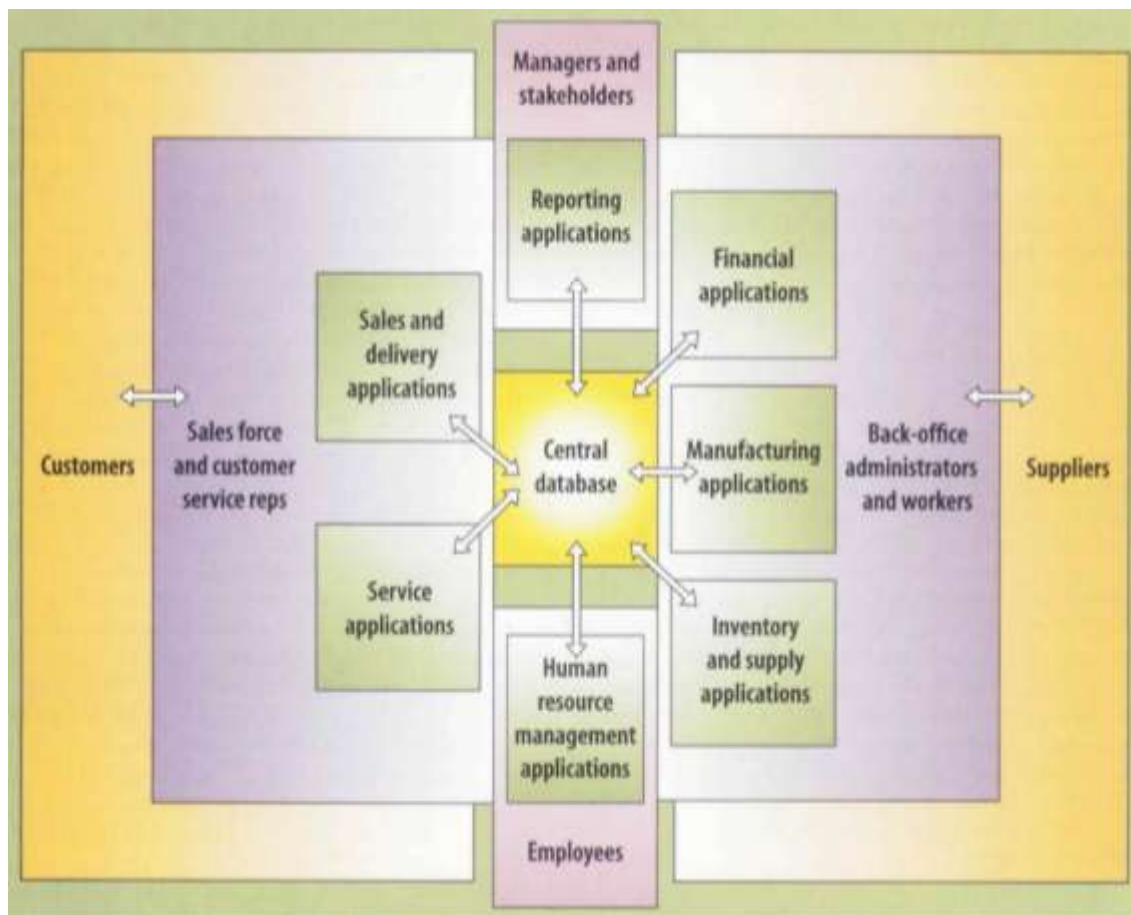


Figura 5 - Anatomia de um Sistema Corporativo (Davenport, 1998).

Devido à sua natureza, um sistema corporativo impõe a sua própria lógica na estratégia, organização e cultura de uma empresa. Ele empurra a empresa para a integração total, mesmo quando um certo grau de segregação de unidades de negócios pode ser do seu melhor interesse (Davenport, 1998).

É possível desenvolver algum grau de personalização num sistema corporativo. Por exemplo, como os sistemas são modulares, as empresas podem instalar apenas os módulos mais adequados aos seus negócios. No entanto, a complexidade dos sistemas inviabiliza grandes alterações. Como resultado, a maioria das empresas que instala sistemas corporativos, precisa de adaptar ou

mesmo refazer completamente os seus processos para os ajustar aos requisitos do sistema. (Davenport, 1998).

Além das implicações estratégicas, os sistemas corporativos também têm um impacto direto e paradoxal na organização e na cultura de uma empresa. Ao fornecer acesso universal e em tempo real a dados operacionais e financeiros, eles permitem às empresas otimizar as suas estruturas de gestão e criar organizações mais democráticas e mais horizontais. Por outro lado, elas também envolvem centralização do controle sobre informações e padronização dos processos, que são qualidades mais consistentes com organizações hierarquizadas com culturas uniformes (Davenport, 1998).

Finalmente, como um sistema corporativo tem implicações profundas nos negócios, a gestão de topo deve estar envolvida no processo de tomada de decisão. O risco de deixar a implementação de sistemas corporativos para os tecnólogos é muito perigoso. A gestão deve controlar o desenvolvimento e a implementação do sistema corporativo de forma a garantir que o sistema se enquadra na empresa, ou corre o risco de ficar sob o controlo de um sistema genérico que força a empresa a mudar para se adequar ao sistema corporativo (Davenport, 1998).

2.1.3.2. Customer Relationship Management (CRM)

Customer Relationship Management (CRM) é uma abordagem integral (holística) para identificar, atrair e reter clientes. O CRM lida com a criação de uma empresa centrada no cliente. Isso envolve dois aspetos principais: foco no cliente e capacidade de resposta ao cliente. Todas as atividades devem, eventualmente, agregar valor ao cliente, refletidas na sua disposição de pagar pelos produtos e/ou serviços; os elementos de adição sem valor devem ser removidos rapidamente no *Internet-time*, porque os clientes têm várias outras

opções. Isso implica concentrar todas as estratégias, planos e ações no cliente, em vez do foco tradicional nos produtos e/ou serviços (Kale, 2016).

Uma estratégia de CRM eficaz visa alcançar o seguinte (Kale, 2016):

- Conquistar continuamente novos clientes;
- Obter compreensão (insights) e gerir a intimidade do cliente;
- Manter os clientes lucrativos e acabar com os clientes não lucrativos;
- Estabelecer relacionamentos de longo prazo com os clientes atuais
- Aumentar os gastos e lucros do cliente através de *cross-selling* e *upselling* (venda cruzada e adicionais).

2.1.3.3. Supply-chain Management (SCM)

A Gestão da Cadeia de Fornecedores (SCM) pode ser definida como a gestão de processos e atividades intra e interorganizacionais com o objetivo de atender aos requisitos do cliente, fornecendo bens e serviços desde o ponto de origem até ao ponto de consumo na hora certa e no mais alto nível de qualidade, ao menor custo total (Kale, 2016).

Para realizar uma cadeia de fornecimento, além da empresa focal, inúmeras empresas diferentes poderão estar envolvidas desde o ponto de origem das matérias-primas até o ponto de consumo pelos consumidores finais; estas incluem fornecedores de matérias-primas/componentes, fabricantes, grossistas/distribuidores, retalhistas e clientes. Diferentes clientes podem precisar de diferentes produtos e serviços diferenciados, e um conjunto diferente de empresas pode estar envolvido para entregar a diferentes clientes. Mesmo para o mesmo portfólio de produtos, diferentes clientes podem exigir um conjunto diferente de serviços de valor agregado (Kale, 2016).

O último elemento-chave na definição de SCM é o foco simultâneo na minimização de custos, redução de tempo e otimização de qualidade (Kale, 2016).

2.2. Sistemas de Gestão e a Empresa em Contexto

2.2.1. Back-office e Front-office

No passado, os tradicionais sistemas ERP eram chamados de *back-office* uma vez que os clientes e público em geral, não estavam diretamente envolvidos nas respetivas atividades e processos. Normalmente, esses ERP suportavam várias funções tais como a produção, compras, vendas, gestão de stocks, contabilidade, finanças, gestão de recursos humanos ou logística. Em geral, os tradicionais sistemas de ERP tinham por objetivo uma maior eficiência e eficácia, mas esta em menor grau (McGaughey & Gunasekaran, 2007).

Já os sistemas ERP contemporâneos foram projetados para otimizar e integrar processos operacionais e fluxos de informações dentro de uma empresa para promover sinergia (Nikolopoulos, Metaxiotis, Lekatis, & Assimakopoulos, 2003) e maior eficácia organizacional. Muitos novos sistemas ERP foram além do *back-office* para dar suporte a processos e atividades de *front-office*, sendo que tal aconteceu, pois com o passar dos anos e com o aumento da importância do *E-commerce* e da globalização dos negócios, os sistemas ERP evoluíram, começando a focar-se mais nas aplicações de *front-office*, pois estas interagem com diversos aspetos externos dos negócios, tais como clientes, fornecedores, parceiros e muito mais. Os principais *players* como Infor, Oracle, PeopleSoft e SAP incorporaram Advanced Planning and Scheduling (APS), Sales Force Automation (SFA), Customer Relationship Management (CRM), Supply Chain Management (SCM) (McGaughey & Gunasekaran, 2007).

2.2.2. Conexões com o Estado

Na “ligação” ao Estado, os ERP permitem preparar certos ficheiros de acordo com as normas governamentais para o pagamento de impostos e guias de

transportes para a Autoridade Tributária e Aduaneira (AT) (AT - Autoridade Tributária e Aduaneira, n.d.).

Em Portugal, todas as empresas que tenham atividade comercial, industrial ou agrícola têm que enviar mensalmente à AT, a informação relativa à sua faturação. Para fazer esta comunicação, é utilizado um formato de ficheiro específico chamado SAF-T(PT) (Standard Audit File for Tax Purposes - Portuguese version), que se trata de um documento digital em formato XML (AT - Autoridade Tributária e Aduaneira, n.d.).

Em qualquer momento, o software permite facilmente reunir e exportar neste formato, a necessária informação fiscal e contabilística de uma empresa, relativa a um determinado período de tempo, o que permite satisfazer os requisitos de fornecimento de informação aos serviços de inspeção, acionistas, auditores internos ou externos e revisores de contas (AT - Autoridade Tributária e Aduaneira, n.d.).

Por sua vez, o software da AT está preparado para importar este tipo de documentos e disponibilizar da melhor forma toda a informação para os inspetores e auditores tributários, para a análise dos registos contabilísticos. A adoção deste formato permite simplificar procedimentos e evitar que elementos da AT necessitem de se especializar em vários sistemas (AT - Autoridade Tributária e Aduaneira, n.d.).

2.2.3. Cloud e o Software as a Service

Os sistemas ERP modernos são desenvolvidos para utilização na internet, tendo vindo a aperfeiçoar os recursos de comércio eletrónico e a capacidade de integração e colaboração com os fornecedores, parceiros, portais de clientes, bem como o acompanhamento aperfeiçoado das matérias-primas recebidas e produtos finais enviados, de modo obter uma maior visibilidade e controlo, tanto dentro como fora da empresa. A implementação destes sistemas é um processo

que se torna mais oneroso à medida que aumenta com o tamanho da empresa. O custo de um projeto de ERP é estimado por muitas empresas apenas pelo custo da licença do software, mas existem vários outros custos a serem consideradas no orçamento de implementação do sistema, tais como o hardware, os serviços de implementação, ou taxas de manutenção e de formação (Abd Elmonem, Nasr, & Geith, 2016).

Os sistemas ERP tradicionais podem ser classificados em duas categorias: ERP local e ERP hospedado (Klaus, Rosemann, & Gable, 2000). No caso de um sistema ERP local, este é carregado e executado na infraestrutura corporativa, utilizando os seus servidores, redes, plataformas, computadores, etc. A empresa executa, opera e gere o sistema ERP de acordo com o modelo de licença de software. O custo de execução, de operação e de manutenção estão a cargo da empresa, bem como da recuperação de desastres. Um ERP hospedado pode ser definido como um serviço vendido a um indivíduo ou organização por um fornecedor que hospeda os servidores físicos que executam esse serviço num outro local. O serviço é habitualmente oferecido através de uma ligação direta de rede que pode eventualmente ser feita via internet (Swart, 2011).

A computação na *cloud* é um ambiente de computação que fornece disponibilidade, escalabilidade e flexibilidade dos recursos do computador em diferentes níveis de abstração, com baixo custo de operação. Pode ser definida como uma forma de computação para fornecer serviços de computação enquanto vantagem para atender às necessidades diárias da comunidade geral de negócios. A computação na *cloud* diz respeito a aplicações, hardware e software fornecidos como serviços via internet (Buyya, Vecchiola, & Selvi, 2013). Os serviços de computação na *cloud* são fornecidos em três modelos (Xu, 2012):

- Software como Serviço (SaaS) tem como alvo o utilizador final ou empresa. Trata-se da entrega de uma aplicação de software pela internet a vários utilizadores. Os *Cloud* ERP pertencem a esta categoria.

- Plataforma como serviço (PaaS) é a entrega de *middleware* que contém ferramentas, serviços e plataformas direcionadas aos desenvolvedores de software, de modo a permitir que eles criem aplicativos SaaS.
- Infraestrutura como um Serviço (IaaS) é a entrega de hardware e software avançados de computação direcionados aos administradores. A empresa paga de acordo com as suas necessidades e atualiza a utilização na medida do crescimento do negócio.

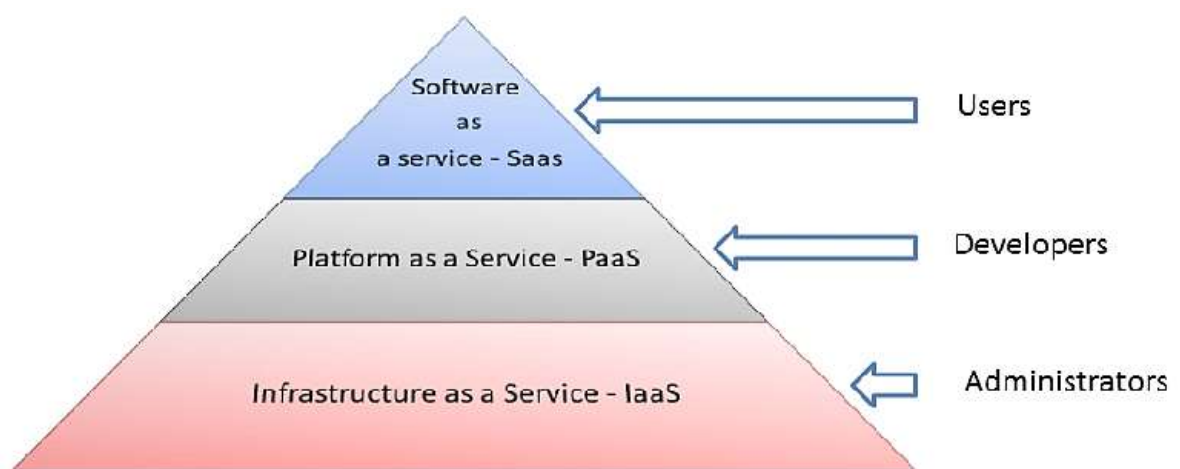


Figura 6 - Modelos de serviços da cloud (Abd Elmonem et al., 2016)

As soluções de *Cloud ERP* são fornecidas através do software como um modelo de serviço. Muitos dos sistemas ERP oferecidos no mercado são sistemas *Cloud ERP* (Scavo, Longwell, & Newton, 2012). Estes sistemas devem ser acessíveis via internet através do browser do utilizador, sem ser necessário instalar ou configurar o seu sistema. Um dos mais famosos *Cloud ERP* do mercado de software é o SAP Business ByDesign (Abd Elmonem et al., 2016).

2.3. Big Data

O *big data* refere-se a grandes conjuntos de dados complexos, estruturados e não estruturados, nos quais as técnicas de processamento tradicionais e/ou

algoritmos não conseguem operar. Ele visa revelar padrões ocultos e levar a uma evolução de um paradigma científico orientado a modelos para um paradigma científico orientado a dados (Taylor-Sakyi, 2016).

Frequentemente, o *big data* é representado por 7 V's, sendo estes o Volume, a Velocidade, a Variedade, a Veracidade, a Validade, a Volatilidade e o Valor, sendo que o desejável no *big data* é o Valor (Khan, Uddin, & Gupta, 2014).

2.3.1. Volume

O volume do *big data* refere-se ao tamanho dos dados que vão sendo criados a partir de todas as fontes, incluindo texto, áudio, vídeo, redes sociais, estudos de pesquisa, dados médicos, imagens espaciais, relatórios de crime, previsão do tempo, desastres naturais, etc (Khan et al., 2014). Os dados que entram para os sistemas de *big data* podem ser conversas de redes sociais, registos do servidor web, sensores do fluxo de tráfego, imagens de satélite, transmissão de fluxos de áudio, transações, música de rock, o conteúdo de páginas web, digitalizações de documentos do governo, trilhos de GPS, telemetria de automóveis, dados do mercado financeiro, etc. No entanto, esse volume de dados é desorganizado e desconhecido (Dumbill, 2012).

2.3.2. Velocidade

Ao contrário do que se poderia pensar, a velocidade a que são gerados dados usando os smartphones e a *World Wide Web* (WWW), está muito longe de estar controlada. Esta elevada velocidade é diretamente responsável pelo elevado volume de dados. Com uma tal elevada velocidade de dados a chegar, as empresas têm de estar preparadas com tecnologia e mecanismos das próprias bases de dados para os processar consoante vão necessitando deles. Há momentos em que simplesmente não é possível esperar por um relatório para executar um trabalho. Isso implica que é da maior importância a velocidade do

ciclo de feedback por que os dados passam desde a sua entrada até à decisão. Portanto, não é apenas importante a velocidade dos dados recebidos, mas também a rápida transmissão dos mesmos para um grande suporte de armazenamento, para processamento e análise posteriores (Khan et al., 2014).

2.3.3. Variedade

Como é sabido, os dados surgem em várias formas, como áudio, vídeo, texto e imagens, introduzindo uma enorme complexidade ao seu tratamento. Por esse motivo, não se pode mais chamar de base de dados relacional. É um grande desafio estabelecer ou construir um sistema para que essas misturas de dados possam ser integradas nele diretamente. A variedade de dados afeta diretamente a integridade dos mesmos, ou seja, uma maior variedade de dados implica que mais erros estes conterão (Khan et al., 2014).

2.3.4. Veracidade

Por Veracidade, quer-se dizer a autenticidade dos dados. Por outras palavras, quão certo se está sobre esses dados? Antes deste V ser adicionado ao universo de *big data*, foi assumido na comunidade científica e de pesquisa que os dados recebidos são limpos e precisos. Tal suposição foi realmente seguida por muitos *data warehouses* tradicionais. Lida-se com dados não estruturados e em grande quantidade que podem ser provenientes de publicações no Facebook, tweets, publicações no LinkedIn, etc. Confia-se em tudo o que se vê lá? Apesar de se aproveitarem essas publicações de dados enquanto se lê e contribui, mas poderemos não perceber que não se pode confiar nelas para as vendas e negócios ou decisões críticas (Khan et al., 2014).

2.3.5. Validade

A Validade dos dados pode parecer similar à veracidade dos dados. No entanto, não se trata do mesmo conceito, mas sim de algo similar. Por validade, quer-se dizer a exatidão e precisão dos dados de acordo com o uso pretendido. Por outras palavras, os dados podem não ter problema algum de veracidade, mas apesar disso, não serem válidos se não forem entendidos. Criticamente falando, o mesmo conjunto de dados pode ser válido para um aplicativo ou uso e, simultaneamente, ser inválido para uma outra aplicação ou uso. Mesmo assim, está-se a lidar com dados cujo relacionamento pode não ser facilmente definido ou em fases iniciais, mas é muito importante verificar com alguma profundidade a relação entre os elementos dos dados com os quais se está a trabalhar, para na medida do possível, os validar em relação ao uso pretendido, (Khan et al., 2014).

2.3.6. Volatilidade

Quando se fala na volatilidade da *big data*, pode-se lembrar facilmente da política de retenção de dados estruturados que se implementam a cada dia nos negócios. Quando o período de retenção dos dados expirar, estes podem ser destruídos. Por exemplo: uma empresa de comércio eletrónico online pode não querer manter o histórico de uma compra de um cliente para além do respetivo período de garantia. *Big data* não é exceção a essa regra e política no mundo real do armazenamento de dados. Esta questão é mais ampliada no mundo do *big data* e não é tão fácil de gerir como no mundo dos dados tradicional. O período de retenção da *big data* pode ser excessivo e o armazenamento e a segurança podem tornar-se caros para serem implementados. Na verdade, a volatilidade torna-se significativa devido ao volume, variedade e velocidade dos dados (Khan et al., 2014).

2.3.7. Valor

Trata-se este V como especial por um motivo. Ao contrário dos outros V's da *big data* que foram referidos anteriormente, este V é o resultado desejado do processamento de *big data*.

Aqui deve-se procurar o verdadeiro valor dos dados que são usados para trabalhar. Por outras palavras, o valor dos dados deve exceder tanto o seu custo como a propriedade ou a gestão. Deve-se prestar atenção ao investimento de armazenamento de dados (Khan et al., 2014). O armazenamento pode ser rentável e relativamente mais barato no momento da compra, mas o subinvestimento pode prejudicar dados altamente valiosos. Por exemplo, armazenar dados de ensaios clínicos para novas drogas em produtos de armazenamento baratos e não confiáveis pode permitir economizar dinheiro hoje, mas pode também colocar os dados em risco amanhã (Hurwitz, Nugent, Halper, & Kaufman, 2013). O valor dos dados depende muito do mecanismo de governança também. Isto é, da forma como se definem as políticas e estruturas que acabarão por trazer equilíbrio entre recompensa e risco dos dados (Hurwitz et al., 2013). Ao mesmo tempo, essas políticas e estruturas, se não forem cuidadosamente definidas e implementadas, podem restringir as empresas a extrair o real valor dos dados. Ou seja, isso tornará os dados subvalorizados. Outro ponto importante que é muitas vezes ignorado, é que o verdadeiro valor está nos olhos do cliente de dados corporativos. Outro facto digno de nota é que alguns dados no momento de recolha podem não ter o mesmo valor para a taxa de risco, mas pode desenvolver-se ao longo do tempo (Khan et al., 2014).

Capítulo 3

Visualização Interativa de Dados

3. Visualização Interativa de Dados

3.1. Tipologia dos Dados

3.1.1. Dados Estruturados

O termo dados estruturados refere-se geralmente a dados com comprimento e formato definidos. Exemplos de dados estruturados incluem números, datas e grupos de palavras e números chamados sequências de caracteres (por exemplo, nome, endereço de um cliente etc.). A maioria dos especialistas concordam que esse tipo de dados representa cerca de 20% dos dados existentes, sendo que estes correspondem a atributos de entidades do mundo real que representamos no mundo da informação. Dados estruturados são os dados com os quais será mais provável estar-se acostumado. Geralmente são armazenados em bases de dados. É possível consultá-los usando uma linguagem como a linguagem de consulta estruturada (SQL - Structured Query Language) (Hurwitz et al., 2013).

A empresa poderá estar já a reunir dados estruturados de fontes "tradicionais". Estes podem incluir os dados do CRM, dados operacionais do ERP e dados financeiros. Geralmente em empresas de grande dimensão, esses elementos de dados são integrados num *data warehouse* para análise (Hurwitz et al., 2013), enquanto nas empresas de menor dimensão optam por tecnologias como os cubos OLAP (Online Analytics Processing). Em outras de pequena dimensão, apenas os dados das transações ficam guardados e apenas desses se fazem *backups*.

Embora isso possa parecer normal, na realidade os dados estruturados estão a assumir uma nova função no mundo da *big data*. A evolução da tecnologia fornece novas fontes de dados estruturados, sendo estes produzidos normalmente em tempo real e em grandes volumes. As fontes de dados são normalmente divididas em duas categorias (Hurwitz et al., 2013):

- Dados gerados por computadores ou máquinas: os dados gerados por máquinas geralmente referem-se a dados criados por uma máquina sem intervenção humana, ou seja, *internet of things*.
- Dados gerados por humanos: são dados que as pessoas, em interação com os computadores, fornecem.

3.1.2. Dados Não Estruturados

Os dados não estruturados são dados que não seguem um formato específico. Se 20% dos dados disponíveis para as empresas são estruturados, os outros 80% não são estruturados. Assim, dados não estruturados são realmente a maioria dos dados que se encontra. No entanto, até recentemente a tecnologia não tinha grandes formas de lhes dar utilidade, exceto armazená-los ou analisá-los manualmente (Hurwitz et al., 2013).

Dados não estruturados estão por todo o lado. De facto, a maioria das pessoas e organizações orientam as suas vidas à volta de dados não estruturados. Tal como nos dados estruturados, os dados não estruturados são gerados por máquinas ou por humanos (Hurwitz et al., 2013), como por exemplo o email e as plataformas sociais.

3.1.3. Big Data e Small Data para ter Smart Data

Tal como foi referido antes, a *big data* é composto por 7 Vs (volume, velocidade, variedade, veracidade, validade, volatilidade e valor), já a *small data* é um conjunto de dados que contém atributos muito específicos. Esta é usada para

determinar estados e condições atuais ou podem ser gerados através da análise de conjuntos de dados maiores (Kavis, 2015).

Por outro lado, a *smart data* filtra o ruído e retém os dados valiosos, que podem ser efetivamente usados pela empresa para resolver problemas de negócios. A análise qualitativa dos dados permite não apenas tornar-se orientada por dados, mas também cria oportunidades para se orientar criativamente, eliminando o ruído permitindo assim uma abordagem mais lógica (Creamfinance, 2016).

Um dos grandes problemas que muitas empresas enfrentam é: quantidade versus qualidade. Os dados precisam ser rigorosamente analisados e avaliados quanto à sua veracidade e funcionalidade. Qual é a variação, facilidade e capacidade de extração? Está incorporado em um conjunto de outras informações irrelevantes? A recolha e o uso da *big data* é significativo apenas quando usados para otimizar e automatizar soluções e resolver problemas. É necessário mudar o foco de apenas recolher grandes quantidades de todos os dados possíveis para contextualizar o que é recolhido, dentro de uma área específica. Devido a este motivo a *big data* tende a ficar mais inútil e prova ser até mesmo dispendioso como um todo (Creamfinance, 2016).

Smart data, por outro lado, torna-se mais inteligente. Facto é que os dados que foram classificados e estruturados adequadamente podem ser utilizados por muito tempo após a data de validade dos dados típicos. Possibilita às empresas a capacidade de ir aos seus arquivos e identificar tendências, procurar anomalias e delinear padrões para o futuro. Essa capacidade só é possível se os dados forem abordados de maneira inteligente e, obviamente, deve haver intenção e visão ao recolher, agrupar e utilizá-los (Creamfinance, 2016).

3.2. Algoritmos de Tratamento de Dados

Os ERP atualmente possibilitam a emissão de relatórios sobre o que aconteceu ou está a acontecer no momento exato e combina as informações de várias áreas, como as finanças, as vendas e a produção, ou seja, é um meio excelente para as empresas recolherem dados para analisar e obter mais conhecimento. Mas para tal acontecer é necessário a Extração de Conhecimento de Dados (ECD), tal vai permite a análise dos dados e destes extrair ainda mais conhecimento para a empresa.

A ECD tem como base um princípio de inferência chamado de indução através da programação de computadores de forma a aprender com a experiência passada, sendo que a partir de uma determinada quantidade de exemplos, permite alcançar conclusões genéricas. Partindo de conjuntos de dados representativos de instâncias do problema que se pretende resolver, os algoritmos de ECD aprendem a induzir uma função, ou hipótese, que possibilita a resolução do problema (Gama, Carvalho, Faceli, Lorena, & Oliveira, 2012).

Mas muitos conjuntos de dados apresentam inconsistências, contêm redundâncias, dados em falta, ou apresentam algum tipo de deficiência, sendo por isso muito importante que os algoritmos de ECD estejam devidamente preparados para lidar com estes problemas e possuam a robustez necessária de forma reduzir o seu efeito em todo o processo de indução de hipóteses (Gama et al., 2012).

Como estes problemas podem prejudicar o todo processo de indução, são frequentemente utilizadas técnicas de pré-processamento de modo a identificar e minimizar a sua ocorrência (Gama et al., 2012).

Recentemente, a área de ECD tem conhecido um grande desenvolvimento graças à enorme capacidade dos atuais recursos computacionais e pelo desenvolvimento de software cada vez mais eficiente. Assim, a sua utilização está

cada vez mais diversificada, existindo já diversas aplicações com soluções para problemas reais que foram bem-sucedidas em áreas como a condução autónoma de veículos, o reconhecimento de voz, o diagnóstico de doenças ou em jogos (Gama et al., 2012).

Já o modelo de aprendizagem adotado por um algoritmo de ECD para lidar com uma tarefa, pode ser classificado como sendo preditivo ou descritivo (Gama et al., 2012).

Nas tarefas de previsão são usados dados de treino para se encontrar uma função que permita prever um rótulo que caracterize um novo exemplo, tendo por base os valores dos respetivos atributos de entrada. Por isso, é essencial que cada objeto do conjunto de treino possua atributos de entrada e de saída. Estes algoritmos seguem o padrão da aprendizagem supervisionada (preditiva) (Gama et al., 2012).

Por outro lado, no caso das tarefas de descrição pretende-se explorar (descrever) um conjunto de dados. Como neste caso o atributo de saída é ignorado, diz-se que estes algoritmos seguem o padrão de aprendizagem não supervisionada (descritiva) (Gama et al., 2012).

Através da figura 6 é possível ver a hierarquia de aprendizagem, de acordo com os tipos de tarefas de aprendizagem (Gama et al., 2012).



Figura 7 - Hierarquia de aprendizagem (Gama et al., 2012)

No início há a aprendizagem indutiva, processo pelo qual são realizadas as generalizações a partir dos dados. Depois, esta ramifica-se em dois tipos de aprendizagem: a supervisionada (preditivo) e a não supervisionada (descritivo) (Gama et al., 2012).

As tarefas preditivas diferenciam-se pelo tipo dos rótulos dos dados (Gama et al., 2012):

- Contínuas: no caso de regressão;
- Discretas: no caso de classificação.

Já nas tarefas descritivas são genericamente divididas em (Gama et al., 2012):

- Agrupamento: os dados são agrupados de acordo com sua semelhança;
- Sumarização: tem como objetivo encontrar uma descrição simples e compacta de um conjunto de dados;
- Associação: consiste em encontrar padrões frequentes de associações entre os atributos de um conjunto de dados.

Na tarefa de aprendizagem por reforço, que se encontra fora do enquadramento das anteriores tarefas, a política é reforçar (recompensar) as ações positivas, punindo as ações negativas. Por exemplo, numa tarefa para encontrar a melhor trajetória entre dois pontos, a passagem por caminhos promissores é recompensada pelo algoritmo de aprendizagem, sendo punida a passagem pelos caminhos que são pouco promissores (Gama et al., 2012).

3.3. Visualização

A visualização de dados é o processo que disponibiliza os dados num formato de imagem ou gráfico de forma a facilitar a sua interpretação. Atualmente, vive-se na era da *big data*, onde os dados são descritos como matéria-prima para os negócios. O volume de dados usados em negócios, indústrias, organizações de pesquisa e desenvolvimento tecnológico é enorme e cresce rapidamente todos os

dias. Quanto mais dados se recolhe e analisa, mais eficiente será possível a tomada de decisões críticas de negócios. No entanto, com o enorme crescimento de dados, tornou-se mais difícil para as empresas extrair informações cruciais dos dados disponíveis. É aí que a importância da visualização de dados se torna evidente. A visualização de dados ajuda as pessoas a entender a importância dos dados, resumindo e apresentando uma enorme quantidade de dados num formato simples e fácil de entender, de modo a comunicar as informações de forma clara e eficaz (Embarak, 2018).

As pessoas entendem melhor os dados através de imagens, em vez de ver números em linhas e colunas. Conseqüentemente, se os dados forem apresentados num formato gráfico, as pessoas terão efetivamente mais facilidade de encontrar conexões e levantar questões importantes (Embarak, 2018).

A visualização de dados ajuda a empresa a alcançar vários objetivos (Embarak, 2018):

- Convertendo os dados de negócios em gráficos interativos para interpretação dinâmica para atender às metas dos negócios;
- Transformação de dados em *dashboards* interativos visualmente apelativos, provenientes de várias fontes de dados para melhor atender aos negócios com essas informações;
- Criação de *dashboards* mais atraentes e informativos de várias representações gráficas de dados;
- Tomar decisões apropriadas, aprofundando os dados e encontrando novas ideias;
- Descobrir padrões, tendências e correlações nos dados em análise para determinar onde se devem melhorar os processos operacionais e, assim, expandir os seus negócios;
- Fornecer uma imagem mais completa dos dados em análise;

- Organizar e apresentar dados maciços intuitivamente para apresentar importantes conclusões dos dados;
- Tomar melhores decisões, mais rápidas e melhor informadas através da visualização de dados.

3.4. Ferramentas e Tecnologia

3.4.1. Tipo utilizadores

Os utilizadores podem ser repartidos entre três categorias diferentes: utilizadores casuais, utilizadores avançados e super utilizadores

Utilizadores casuais visualizam e interagem com aplicações analíticas, relatórios e *dashboards* (Eckerson, 2016).

Eles podem ser divididos em dois grupos (Eckerson, 2016):

- Os espectadores que usam relatórios, mas não interagem com eles. Muitos recebem relatórios estáticos através do email, podendo solicitar mais;
- Os navegadores que interagem com relatórios e painéis de forma a aumentar sua compreensão dos dados. Entre outras coisas, eles aprofundam, filtram, classificam, dinamizam, agrupam, visualizam, calculam e anotam dados. Normalmente salvam *snapshots* de relatórios que foram modificados, mas nunca editam o modelo de dados subjacente (Eckerson, 2016).

Já os utilizadores avançados são analistas de negócios (por exemplo, finanças, vendas, operações, marketing) e cientistas de dados com vários graus de conhecimento em dados e análises (Eckerson, 2016).

Existem várias categorias de utilizadores avançados (Eckerson, 2016):

- Os analistas iniciados que adaptam relatórios e *dashboard* existentes, usando o modelo de dados subjacente para criar novas métricas, dimensões,

hierarquias e visualizações. Estes analistas arrastam objetos do modelo ou da biblioteca de dados para criar pesquisas *ad hoc* e montar novas visualizações dos dados existentes na forma de relatórios e painéis *ad hoc*;

- Os analistas intermédios vão mais além ligando-se a novas fontes de dados - arquivos locais, aplicações corporativas ou fontes de dados externas - com as quais estão intimamente familiarizados. Eles usam os dados para criar novos relatórios a partir do zero ou ampliar os relatórios existentes, misturando dados novos e existentes num novo modelo de dados;
- Analistas avançados são cientistas de dados que sabem programar em SQL, Java, Python ou outras linguagens e possuem um conhecimento profundo dos dados originais. Eles podem criar relatórios e análises personalizados.

Os super utilizadores são os principais componentes da infraestrutura de suporte do ERP. Além de estarem na linha de frente com os utilizadores finais, eles também servem como elos de ligação entre as empresas e as organizações de TI para se concentrar em aspetos mais técnicos do suporte ao sistema (Kamaleshn, 2012).

O super utilizador é um especialista local pronto para ajudar os colegas de trabalho no uso de equipamentos específicos. Esse papel deve ser desempenhado por um indivíduo motivado e orientado a serviços, interessado no uso bem-sucedido do equipamento, agora e no futuro. Esse indivíduo deve ser um solucionador de problemas, disposto a ouvir as preocupações e questões dos colegas de equipa, conhecedor dos recursos disponíveis e que esteja disposto a levar os problemas adiante às entidades adequadas (Kamaleshn, 2012).

A função do super utilizador inclui (Kamaleshn, 2012):

- Funcionar como uma pessoa de suporte baseado numa unidade "socorrista" de apoio ao seu departamento;
- Frequentar aulas e atualizações, e desenvolver anualmente a sua própria aptidão;

- Atuar como recurso da unidade e solucionador de problemas;
- Fornecer *feedback* aos funcionários relativamente a atualizações necessárias;
- Assumir a liderança na formação da unidade e na verificação de competências, tanto anualmente como para novas contratações;
- A participação em testes ou projetos-piloto do departamento, focado no equipamento envolvido, conforme aplicável.

3.4.2. Inteligência Artificial

Inteligência Artificial (IA), consiste na capacidade de um computador digital ou um robô controlado por computador para executar tarefas normalmente associadas com seres inteligentes. O termo é frequentemente aplicado em projetos de desenvolvimento de sistemas dotados dos processos intelectuais característicos dos seres humanos, como a capacidade de raciocinar, descobrir significado, generalizar ou aprender com as experiências passadas. Desde o desenvolvimento do computador digital nos anos 40, foi demonstrado que os computadores podem ser programados para realizar tarefas muito complexas - como por exemplo, descobrir provas de teoremas matemáticos ou jogar xadrez - com grande proficiência. Ainda assim, apesar dos contínuos avanços na velocidade de processamento do computador e na capacidade de memória, ainda não existem programas que possam corresponder à flexibilidade humana em domínios mais amplos ou em tarefas que exigem muito conhecimento quotidiano. Por outro lado, alguns programas atingiram os níveis de desempenho de especialistas humanos e profissionais na execução de determinadas tarefas específicas, de modo que a IA nesse sentido limitado é encontrada em aplicações tão diversas quanto diagnóstico médico, mecanismos de busca por computador e reconhecimento de voz ou caligrafia (Copeland, 2003).

O Excel é uma das ferramentas mais utilizadas para calcular e analisar dados, sendo o Ignite uma função que permite a utilização de quatro recursos de IA que tornam o Excel ainda mais útil (Spataro, 2018):

- **Ideas** - trata-se de um serviço de informações com IA que ajuda a aproveitar todo o potencial do Office. Este recurso ajuda na criação de documentos, apresentações e folhas de cálculo profissionais em menos tempo, recorrendo a sugestões proativas, adaptadas à tarefa em questão. Por exemplo, o Ideas ajuda a identificar tendências, padrões e outliers num conjunto de dados no Excel.
- **Tipos de dados** - através dos tipos de dados, o Excel pode reconhecer conceitos do mundo real, começando com os *stocks* e a geografia. Utilizando IA, este recurso permite transformar um único texto numa entidade interativa que contém camadas de informações valiosas.
- **Inserir dados através de uma imagem** - a informação analógica contida numa foto de uma tabela de dados impressos, através de reconhecimento de imagem pode ser facilmente convertida numa folha de cálculo completamente editável no Excel.
- **Matrizes dinâmicas** - Usando matrizes dinâmicas, qualquer fórmula que retorne uma matriz de valores será "derramada" (*spill*) facilmente nas células desocupadas vizinhas, facilitando a obtenção de uma matriz de valores retornada e o trabalho numa única célula. Desta forma, em vez de escrever muitas fórmulas complexas para resolver um problema de várias células, basta escrever uma fórmula simples para obter uma matriz de valores retornados.

3.4.3. Internet of Things

A Internet of Things (IoT) pode ser entendida como uma infraestrutura global para a sociedade da informação que possibilita serviços avançados através da interconexão (física e virtual) entre “coisas”, baseados em tecnologias de informação e comunicação interoperáveis, existentes e em evolução. Estas coisas são objetos do mundo físico (coisas físicas) ou da informação (mundo virtual) capazes de serem identificados e integrados nas redes de comunicação. As coisas físicas são capazes de serem sentidas, manipuladas e conectadas. Exemplos disso incluem o ambiente circundante, robôs industriais, bens e equipamento elétrico. Coisas virtuais são capazes de serem armazenadas, processadas e acedidas, como por exemplo: conteúdo multimédia e aplicações de software (International Telecommunication Union, 2012).

Através da exploração dos recursos de identificação, captura de dados, processamento e comunicação, a IoT permite fazer o uso total de "coisas" para oferecer serviços a todos os tipos de aplicações, garantindo ao mesmo tempo os requisitos de segurança e privacidade (International Telecommunication Union, 2012).

Espera-se um grande crescimento da integração da IoT nas principais tecnologias, como as relacionadas com a comunicação avançada máquina a máquina, redes autónomas, *data mining* e tomadas de decisão, proteção de segurança e privacidade, computação em *cloud*, tecnologias para deteção e manipulação avançadas (International Telecommunication Union, 2012).

3.4.4. Elementos de Avaliação

Segundo a Gartner em 2019, os elementos de avaliação são compostos por um total de 5 formas de uso 15 recursos críticos de uma plataforma de análise e BI,

sendo as 5 formas de uso são as seguintes (Howson, Richardson, Sallam, & Kronz, 2019):

1. Aprovisionamento ágil e centralizado do BI: suporta um *workflow* ágil e habilitado para TI, desde os dados à entrega centralizada e conteúdo analítico gerido, usando os recursos independentes de gestão de dados da plataforma.
2. Análise descentralizada: suporta um workflow de dados para análise de *self-service*, e inclui análises para unidades de negócios e utilizadores individuais.
3. Descoberta de dados governados: suporta um workflow de os dados para análise de *self-service* ao SOR (System of Record), conteúdo gerido por TI com governança, reutilização e promoção do conteúdo gerado pelo utilizador para dados certificados e conteúdo de análises.
4. Análise integrada de OEM (Original Equipment Manufacturer) ou incorporada: suporta um *workflow* de dados para conteúdo de BI incorporado em um processo ou aplicação.
5. Implantação de extranet: suporta um *workflow* semelhante ao provisionamento ágil e centralizado do BI para o cliente externo ou, no setor público, o acesso dos cidadãos ao conteúdo analítico.

Por outro lado, os 15 recursos críticos são os seguintes:

A infraestrutura (Howson et al., 2019):

1. **A administração, a segurança e a arquitetura da plataforma de BI:** recursos que permitem a segurança da plataforma, administração de utilizadores, auditoria do acesso e utilização da plataforma e alta disponibilidade e recuperação de desastres.

2. **Cloud BI:** plataforma como serviço e aplicação analítica como serviço para criar, implantar e gerir análises e aplicativos analíticos na *cloud*, com base em dados na *cloud* e no local.
3. **Conectividade de fonte de dados e ingestão:** recursos que permitem aos utilizadores se conectar a dados estruturados e não estruturados contidos em vários tipos de plataforma de armazenamento, tanto no local como na *cloud*.

Gestão de dados (Howson et al., 2019):

4. **Gestão de metadados:** Ferramentas que permitem aos utilizadores aproveitarem um modelo semântico e metadados comuns. Eles devem fornecer uma maneira robusta e centralizada para que os administradores pesquisem, capturem, armazenem, reutilizem e publiquem objetos de metadados, tais como dimensões, hierarquias, medidas, KPI's (Key Performance Indicators, além de relatar objetos de layout, parâmetros etc. Os administradores devem ter a capacidade de promover um *mashup* de dados e metadados definidos pelo utilizador comercial para os metadados do SOR.
5. **Opções de armazenamento e carregamento de dados:** recursos da plataforma para aceder, integrar, transformar e carregar dados num mecanismo de desempenho independente, com a capacidade de indexar dados, gerir carregamentos de dados e agendar atualizações.
6. **Preparação de dados:** "*Drag and drop*" combinação de dados adicionados pelo utilizador de diferentes fontes e a criação de modelos analíticos, como medidas, conjuntos, grupos e hierarquias definidas pelo utilizador. As capacidades avançadas incluem a machine learning com descoberta automática de semântica, junções inteligentes, perfil inteligente, geração hierárquica, linhagem de dados e combinação de dados em diversas fontes de dados, incluindo dados multi-estruturados.

7. **Escalabilidade e complexidade do modelo de dados:** o grau em que o mecanismo na memória ou a arquitetura na base de dados lida com grandes volumes de dados, modelos de dados complexos, otimização de desempenho e grandes implantações de utilizadores.

Análise e Criação de Conteúdo (Howson et al., 2019):

8. **Análise avançada para cientistas de dados comuns:** permite que os utilizadores acessem facilmente a recursos avançados de análise independentes da plataforma, através de opções orientadas por menus ou pela importação e integração de modelos desenvolvidos externamente.

9. **Dashboards analíticos:** a capacidade de criar *dashboards* e conteúdos altamente interativos com exploração visual e análises avançadas e geoespaciais incorporadas para consumo de terceiros.

10. **Exploração visual interativa:** permite a exploração de dados por meio de uma variedade de opções de visualização que vão além das tabelas básicas de gráficos circular, barra e linha, para incluir mapas de calor e de árvore, mapas geográficos, gráficos de dispersão e outros recursos visuais especiais. Esses recursos permitem que os utilizadores analisem e manipulem dados, interagindo diretamente com uma representação visual destes, exibindo-os como percentagens, posições e grupos.

11. **Descoberta aumentada de dados:** encontra, visualiza e narra automaticamente descobertas importantes, como correlações, exceções, clusters, links e previsões de dados relevantes para os utilizadores, sem a necessidade de criar modelos ou escrever algoritmos. Os utilizadores exploram dados por meio de visualizações, narração gerada em linguagem natural, pesquisa e tecnologias de consulta de linguagem natural (NLQ).

12. **Exploração móvel e autoria:** permite que as organizações desenvolvam e entreguem conteúdo para dispositivos móveis num modo de publicação

e/ou interativo e aproveita os recursos nativos dos dispositivos móveis, como o touchscreen, a câmara e o reconhecimento de local.

Partilha de resultados (Howson et al., 2019):

13. **Incorporação de conteúdo analítico:** recursos, incluindo um kit de desenvolvedor de software com APIs (Application Programming Interface) e suporte a padrões abertos para criar e modificar conteúdo analítico, visualizações e aplicativos, incorporando-os a um processo de negócios e/ou aplicativo ou portal. Esses recursos podem residir fora do aplicativo, reutilizando a infraestrutura analítica, mas devem ser acedidos de maneira fácil e perfeita a partir do aplicativo, sem forçar os utilizadores a alternar entre sistemas. Os recursos para integrar análise e BI com a arquitetura do aplicativo permitirão que os utilizadores escolham em que parte do processo de negócios a análise deve ser incorporada.

14. **Publicar, partilhar e colaborar no conteúdo analítico:** recursos que permitem aos utilizadores publicar, implantar e operacionalizar conteúdo analítico através de vários tipos de *output* e métodos de distribuição, com suporte para pesquisa, programação e alertas de conteúdo. Esses recursos permitem que os utilizadores partilhem, discutam e rastreiem informações, análises, conteúdo analítico e decisões por meio de tópicos de discussão, conversa e anotações.

Plataforma geral (Howson et al., 2019):

15. **Facilidade de uso, apelo visual e integração do *workflow*:** A facilidade de administrar e implantar a plataforma, criar conteúdo, consumir e interagir com o conteúdo, bem como o grau em que o produto é visualmente apelativo. Essa capacidade também considera o grau em que as capacidades são oferecidas num único produto e *workflow* contínuo, ou em vários produtos com pouca integração.

3.4.5. Ferramentas de Análise

O Quadrante Mágico (figura 6) concentra-se em produtos que cumprem com os critérios da Gartner para uma moderna plataforma de análise e BI, sendo que destes destacam-se o IBM, Qlik, Tableau e Power BI (Microsoft).



Figura 8 - Magic Quadrant for Analytics and Business Intelligence Platforms (Howson et al., 2019)

IBM

A IBM oferece o Cognos Analytics, uma moderna plataforma de análise e BI, aprimorada recentemente com capacidades aumentadas de análise que existiam anteriormente apenas no IBM Watson Analytics. Como o Cognos Analytics também é uma atualização de versões anteriores do Cognos, ele traz relatórios

formatados no estilo de produção para o Modo 1 (concentra-se na previsibilidade e tem como objetivo a estabilidade), juntamente com exploração com uma base visual e agilidade para análises e BI do Modo 2 (é exploratório, envolvendo experimentação para resolver novos problemas e otimizado para áreas de incerteza).

Dos clientes de referência da IBM, mais da metade (59%) utiliza-o para um ágil provisionamento de BI centralizado. O próximo uso mais comum é para análises descentralizadas (49%).

A IBM é um *niche player*, pois permanece interessado principalmente para a base instalada do Cognos, e a inovação abrandou enquanto a IBM se concentrou em criar recursos aprimorados no Cognos Analytics.

Forças:

- Relatórios combinados e análises aumentadas;
- Visão do produto;
- Presença global.

Cuidados:

- Produto imaturo com lacunas básicas;
- Perda de impulso e percepção como inovador;
- Operações.

Qlik

A Qlik fornece uma descoberta controlada de dados, análises ágeis e BI através do seu produto líder, o Qlik Sense, construído na plataforma Qlik Analytics. A plataforma pode ser usada para criar aplicativos personalizados via um extenso conjunto de APIs para suportar casos de utilização de análise incorporada. O produto original da Qlik, Qlik View, responde por grande parte da base instalada da empresa, mas agora o Qlik Sense gera mais de 67% da sua receita de licenças. De acordo com os clientes de referência, o Qlik é usado principalmente para casos

de utilização de análise descentralizada (70%) e casos de provisionamento ágil de BI centralizado (66%).

A posição da Qlik no quadrante Leaders é parcialmente atribuível ao seu forte roteiro de produtos, alcance geográfico e entendimento do mercado. Mas as suas pontuações de capacidade de resposta ao mercado são inferiores às de outros líderes.

Forças:

- Recursos e extensibilidade do produto;
- Experiência do cliente;
- Visão do produto.

Cuidados:

- Workflow do produto;
- Experiência de migração;
- Momento mais baixo.

Tableau

O Tableau oferece uma experiência de exploração intuitiva, interativa e baseada em visual que permite que os utilizadores comerciais acedam, preparem, analisem e apresentem resultados nos seus dados, sem habilidades técnicas ou codificação. A oferta do Tableau é implantada principalmente no local, como um aplicativo de desktop independente ou integrado com um servidor para partilhar conteúdo. Tableau Online é um SaaS baseado numa *cloud*.

Os clientes de referência do Tableau relatam que o usam principalmente para análises descentralizadas (70%) e provisionamento ágil de BI centralizado (51%).

O Tableau é um líder, graças à popularidade do seu produto, altos índices de satisfação do cliente e um forte *roadmap*.

Forças:

- Fácil exploração visual e manipulação de dados;

- Clientes como fãs;
- Momento.

Cuidados:

- Lacunas do produto;
- Declínio do suporte;
- Experiência em vendas, contratação e custo.

Power BI

A Microsoft oferece preparação de dados, descoberta de dados com base visual, *dashboards* e análises aumentadas através de um único produto, o Power BI. Está disponível como uma opção SaaS em execução na *cloud* do Microsoft Azure ou como uma opção local no Power BI Report Server. O Power BI Desktop pode ser usado como uma ferramenta de análise pessoal gratuita e independente. Também é necessário quando utilizadores avançados desenvolvem *mashups* de dados complexos que envolvem fontes de dados locais.

Embora o Power BI tenha começado a ser usado principalmente como uma ferramenta *self-service* de análise e BI, os clientes de referência da Microsoft usam-no agora principalmente para análises descentralizadas (61%) e provisionamento ágil e centralizado de BI (54%).

A Microsoft é um líder. Possui um *roadmap* de produtos abrangente e visionário, com o objetivo de globalizar e democratizar o Power BI para todos os casos de uso de análise. Ao mesmo tempo, continua a demonstrar forte aceitação e adoção global do Power BI, com altos níveis de satisfação do cliente.

Forças:

- Titular de baixo preço com experiência de vendas positiva;
- Facilidade de uso para tipos complexos de análise;
- Visão abrangente do produto.

Cuidados:

- Diferenças no serviço local e na *cloud* apenas com a *cloud* do Azure;
- Integração do Modo 1 e Modo 2;
- Múltiplos produtos.

Capítulo 4

Metodologia

4. Metodologia

4.1. Questões e Finalidade

A finalidade deste trabalho de investigação é a construção de um sistema que teste a capacidade de gestão e de análise de dados oriundos de múltiplas fontes de dados, através do Power BI. Por esse motivo, definiram-se os seguintes objetivos específicos:

- Conceber e desenvolver um sistema de *analytics* de apoio à gestão em diversos departamentos de uma empresa, suportado em várias fontes de dados;
- Explorar o potencial do Power BI como ferramenta de apoio ao BI, através da implementação de *dashboards* interativas;
- Explorar técnicas descritivas e prescritivas na análise dos dados.

Cumprindo estes objetivos teremos uma perceção clara do potencial deste recurso e dos diferentes níveis de conhecimento técnico necessário para dele tirar partido.

4.2. Paradigma Metodológico

O paradigma de investigação na área dos sistemas de informação, que se cruza com o comportamento humano e organizacional, é frequentemente o Design Science Research (Hevner, March, Park, & Ram, 2004). Trata-se de um quadro conceptual que enquadra a busca de um guião esclarecido para definir ideias,

práticas, técnicas, produtos, análises, implementação ou a gestão de sistemas de informação de forma inovadora ou, pelo menos, mais eficiente.

Neste enquadramento de procura de um guia claro para realizar uma análise específica de software de manipulação analítica de dados, identificamos as melhores técnicas para o efeito: CRISP-DM, SEMMA e KDD (Shafique & Qaiser, 2014).

São métodos adotados em processos de data mining que identificam etapas precisas para o bom êxito deste complexo exercício.

O Sample, Explore, Modify, Model e Assess (SEMMA) foi criada pelo SAS Institute e tem semelhanças com CRISP-DM focando-se na criação do modelo de dados sem considerar objetivos de negócio. O Knowledge Discovery in Databases (KDD) é um modelo mais antigo, focado exclusivamente na análise dos dados.

Para o desenvolvimento deste trabalho foi adotado o método baseado no standard para a implementação de projetos de *data mining*: o CRISP-DM. Este é um método de *data mining* que define uma abordagem que tem em linha de conta os objetivos de negócio para além do modelo de dados, ao contrário do semelhante SEMMA (Chapman et al., 2000).

A utilização do método CRISP-DM permite obter várias vantagens, nomeadamente: torna a implementação de projetos de *data mining* mais rápida, mais simples, mais barata, e mais fácil de gerir; é independente da indústria e da ferramenta de mineração de dados. A CRISP-DM é também independente da indústria e da ferramenta usada para a implementação, ou seja, esta pode ser usada independentemente do negócio em causa (saúde, comércio, retalho, etc.), bem como fazendo uso de qualquer ferramenta de *data mining* (Chapman et al., 2000).

O CRISP-DM permite a todos os profissionais desta área ou ligados a ela, do mais inexperiente até a um especialista, seguir um mapa, ou desenho técnico, que

permite conduzir todo o projeto de *data mining* de forma simples, eficiente e com sucesso alinhados aos objetivos do negócio (Chapman et al., 2000).

4.3. Procedimentos de Investigação

4.3.1. Seleção do Objeto de Estudo

O objeto de estudo é a informação proveniente de duas fontes de dados distintas de uma empresa do mercado das bicicletas chamada Adventure Works Cycles.

A base de dados principal é a AdventureWorks que se trata de uma amostra fornecida originalmente pela Microsoft para o Microsoft SQL Server 2008 ao 2014. Esta base de dados suporta cenários de processamento padrão de transações onde se incluem a produção, vendas, compras, gestão de produtos e recursos humanos.

Nas organizações pretende-se, frequentemente, combinar fontes de dados transacionais com outras fontes, nomeadamente relativas às das plataformas sociais e de tráfego no website. Assim, para avaliar o grau de facilidade com que o Power BI relaciona tabelas de múltiplas fontes, é utilizada uma outra base de dados desenvolvida pelo próprio, a “Bicicletas – Rating”. Esta simula uma avaliação efetuada por clientes, no site da empresa, aos seus diversos produtos, seguindo uma estrutura semelhante ao que se pode encontrar em diversas empresas globais de venda online.

4.3.2. Tratamento

Como foi referido anteriormente o método escolhido para este estudo é a CRISP-DM. Este método tem uma natureza cíclica, e tal como se pode observar na figura 9, é composto por seis etapas que têm dependências entre si. A

possibilidade de a execução de uma etapa encadear novas questões, mais focadas e específicas podem originar um regresso à etapa anterior.

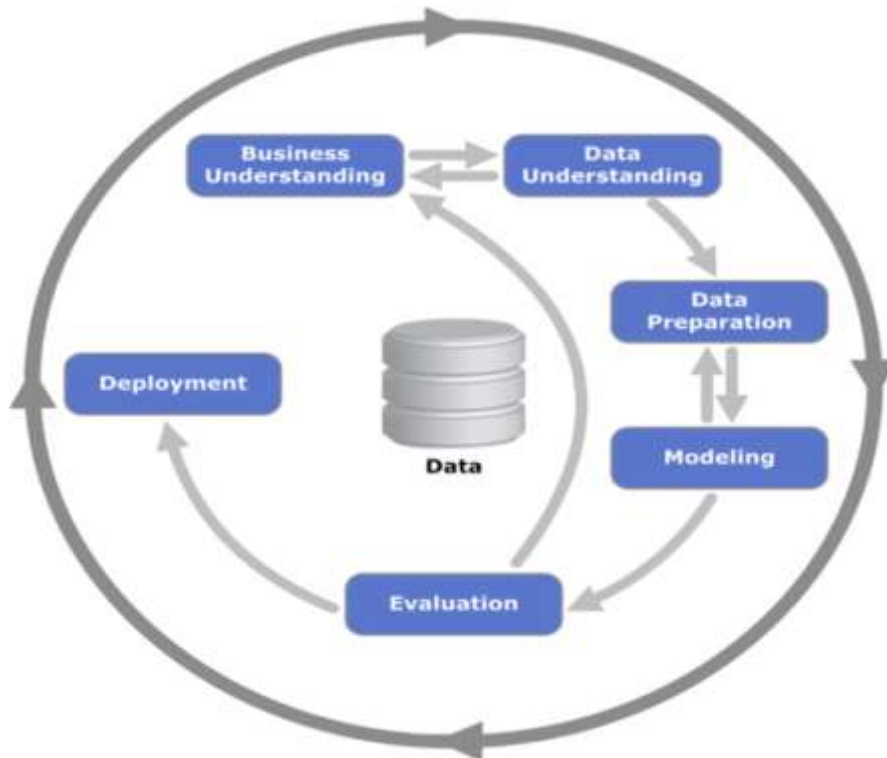


Figura 9 - Metodologia CRISP-DM (Chapman et al., 2000)

As seis etapas que compõem o CRISP-DM são as seguintes (Chapman et al., 2000):

- **Compreensão do Negócio:** Nesta etapa é efetuada a análise dos objetivos e dos requisitos funcionais, técnicos e temporais na perspetiva do negócio. Para além disso, é ainda efetuado o enquadramento desses objetivos e restrições na formulação do problema de *data mining*, bem como a elaboração de uma estratégia preliminar para alcançar os objetivos propostos.
- **Compreensão dos Dados:** Nesta segunda etapa é efetuada a recolha e análise dos dados para verificar se estes correspondem aos requerimentos. Nessa verificação é possível serem identificados problemas que obriguem a retornar à fase anterior, compreensão do negócio, e rever o plano.

- **Preparação dos Dados:** Nesta etapa é aplicado um conjunto de atividades com a finalidade de construção do conjunto de dados para que possam ser objeto de análise pelas ferramentas de modelação. É efetuada a seleção de casos e campos que se pretendem analisar. Para além disso, se necessário, deve efetuar-se também limpeza dos dados e algumas transformações.
- **Modelação:** Nesta etapa é onde se selecionam as técnicas de modelagem a aplicar.
- **Avaliação:** Nesta etapa é efetuada a avaliação e revisão das atividades realizadas na construção do modelo e verificação da sua contribuição para o alcance dos objetivos do negócio.
- **Implementação:** Atividades que conduzem à organização do conhecimento e à sua disponibilização, como relatórios e outros tipos de documentos. Por vezes os resultados obtidos poderão dar lugar a um novo projeto de *data mining*.

Com a adoção possível deste método procuramos explorar uma situação próxima do real e assim realizar os testes e a avaliação com uma precisão mais objetiva.

Capítulo 5

Construção de uma Dashboard em Power BI

5. Construção de uma Dashboard em Power BI

5.1. Compreensão do Negócio

De forma a atingir o objetivo central da investigação nesta dissertação, que pretende simular um sistema que teste a capacidade de gestão e de análise de dados oriundos de múltiplas fontes de dados, através do Power BI, elaboram-se os seguintes objetivos específicos:

- Conceber e desenvolver um sistema de *analytics* de apoio à gestão em diversos departamentos de uma empresa, suportado em várias fontes de dados;
- Explorar o potencial do Power BI como ferramenta de apoio ao BI, através da implementação de *dashboards* interativas;
- Explorar técnicas descritivas e prescritivas na análise dos dados.

5.1.1. Power BI

O Power BI da Microsoft é uma ferramenta poderosa que permite ajudar os utilizadores a recolher e interpretar os dados, de modo a facilitar a monitorização e a tomada de decisões. Esta é uma solução de análise de informação que possibilita a visualização de dados e partilhar informações na organização através da utilização de *dashboards* e relatórios dinâmicos, os quais podem ser incorporados numa aplicação ou site. Esta ferramenta oferece uma interface simples que permite aos utilizadores criar as suas próprias *dashboards* e relatórios.

A solução está disponível de várias formas, das quais se destacam as seguintes: Power BI Services, que se trata de uma implantação de SaaS, e Power BI Desktop que se trata de uma versão de instalação local. Estas são soluções robustas de software que podem ser usadas para conectar múltiplas fontes de dados, visualizar padrões e descobrir *insights*. Nesta investigação foi utilizado o Power BI Desktop, pois o Power BI Services é mais limitado em termos de modelação.

5.1.2. Dashboard

A *dashboard* é uma ferramenta interativa de gestão de informações que, de forma visual e personalizada, permite rastrear, analisar e exibir KPIs, métricas e pontos de dados importantes para monitorizar a saúde de um negócio, um departamento ou um processo específico. Nos bastidores, uma *dashboard* encontra-se conectada aos ficheiros, anexos, serviços e API's, mas na superfície exibe todos esses dados na forma de tabelas, gráficos de linhas, gráficos de barras e medidores. Esta ferramenta permite implementar num único local, uma forma eficiente de análise de várias fontes de dados, facilitando a monitorização e análise do desempenho desde um simples processo até uma empresa. A monitorização em tempo real reduz as horas de análise e a longa linha de comunicação que anteriormente desafiavam os negócios.

5.1.3. Sistemas Semelhantes

Como já foi antes referido, há dois grandes sistemas bastante semelhantes ao Power BI:

- O Tableau é uma ferramenta de BI especializada em visualizações avançadas de dados que facilitam a exploração e análise de dados de várias fontes. Este software permite a consulta de bases de dados relacionais, bases de dados em *cloud*, cubos e tabelas de OLAPs, para gerar uma variedade de tipos de visualizações. Usando a linguagem de consulta VizQL, o Tableau

converte ações de arrastar e soltar dos utilizadores, em consultas de dados que são expressas visualmente. Os utilizadores do Tableau podem criar e partilhar *dashboards* que descrevem relatórios, padrões e densidade de dados para disseminação eficiente de informações.

- O QlikView é uma plataforma de descoberta de dados e *insight* de clientes da Qlik. Foi desenvolvido para oferecer dados de autoatendimento que podem ajudar a direcionar decisões e gerar ROI (Return on Investment) significativo para utilizadores de qualquer nível de habilidade técnica. São oferecidos vários recursos, incluindo *dashboards* pré-elaborados, ou seja, *dashboards* pré-configurados para análise e interpretação dos dados. Foi desenvolvido para ser acessível, escalável e adaptável. Este software permite utilizar dados de diversas fontes, como grandes fluxos de dados, dados baseados em ficheiros, dados locais ou na *cloud*. O QlikView é conhecido pelas suas associações de dados e função de relacionamento, mantendo automaticamente os dados no respetivo contexto.

5.2. Compreensão dos Dados

Na segunda etapa do CRISP-DM, de forma a atingir o objetivo foram utilizadas duas fontes distintas de dados estruturados, de uma empresa do mercado das bicicletas chamada Adventure Works Cycles. Trata-se de uma multinacional que fabricava e vendia bicicletas, peças e acessórios (por exemplo capacetes), para os mercados da América do Norte, Europa e Ásia entre 2011 e 2014.

A base de dados principal é a AdventureWorks que se trata de uma amostra fornecida originalmente pela Microsoft para o Microsoft SQL Server 2008 ao 2014. Encontra-se segmentada em seis áreas, sendo possível observar cada uma delas através do diagrama da base de dados (ver anexo 1):

- *Sales* (vendas): o módulo das vendas contém informações sobre carrinho de compras, pedidos de vendas, ofertas especiais e vendedores;
- *Human Resources* (recursos humanos): informação sobre os colaboradores;
- *Person* (gestão de contactos): consiste nos nomes e endereços de clientes, fornecedores e colaboradores;
- *Purchasing* (compras): fornecedores aos quais as peças e produtos são comprados;
- *Production* (produção e gestão dos produtos):
 - Produção: produtos fabricados e vendidos pela empresa;
 - Gestão dos produtos: gestão de inventários dos produtos.
- Dbo: administração do sistema (não utilizada por não ser relevante para o trabalho).

De forma cumprir o objetivo central desta dissertação, foi desenvolvida pelo próprio uma base de dados secundária, a “Bicicletas – *Rating*”, que pretende constituir uma avaliação hipotética efetuada por clientes no site da Adventure Works Cycles, aos seus diversos produtos. Uma vez que esta base de dados secundária provém de um site, a sua fonte de dados é distinta da base de dados principal, cujos dados foram recolhidos e organizados através de um ERP.

A “Bicicletas – *Rating*” é composta por cinco campos, tal como se pode ver no apêndice 1. São as seguintes:

- *ProductID*: é a chave comum identificadora de cada produto no sistema da empresa, permitindo que o Power BI estabeleça automaticamente uma ligação com a base de dados AdventureWorks;
- *Product*: o nome do produto que o cliente avaliou;
- *Date*: data em que foi feita a avaliação do produto pelo cliente;
- *Reviews Rating*: avaliação feita pelo cliente a um produto, podendo esta variar entre 1 (mau) e 5 (excelente);
- *Username*: nome com que o cliente fez a avaliação no site da empresa.

5.3. Preparação dos Dados

Nesta etapa é decidido quais os dados a ser utilizados na construção da *dashboard* no Power BI. Este permite aceder diretamente e de forma mais rápida à base de dados AdventureWorks no SQL Server, o que possibilitaria a visualização interativa dos dados em tempo real.

No entanto, esta base de dados não se encontra a ser atualizada em tempo real e contém centenas de milhares de linhas e centenas de colunas, ou seja, uma quantidade excessiva de dados em relação às necessidades para os objetivos desta investigação, possuindo ainda missing values e diversos valores codificados. Por estes motivos, optou-se pela utilização dos programas SQL Server Management e Excel.

5.3.1. Structured Query Language

De forma a facilitar o desenvolvimento do conjunto de dados de modo a poderem ser objeto de análise pelas ferramentas de modelação, foi efetuada uma seleção de várias tabelas que permitiram criar cinco *views* no SQL Server Management, ou seja, pretende-se transformar o *big data* em *small data*. Uma *view* funciona como uma tabela virtual que disponibiliza os dados de uma ou mais tabelas, apesar de não os conter. Trata-se apenas de um conjunto de consultas aplicadas a uma ou mais tabelas armazenadas na base de dados e que constitui um objeto.

Uma vez que se trata de uma base de dados muito grande e não sendo necessária uma tal quantidade de dados, optou-se pela utilização de *views* na construção da *dashboard*, tendo sido apenas escolhidas algumas tabelas de cada área, de forma a permitir explorar com o Power BI alguns dos principais departamentos da empresa. Estas *views* foram denominadas e compostas pelas seguintes tabelas, tal como pode ser observado no exemplo dado no apêndice 1:

- **Recursos Humanos:** HumanResources.EmployeeDepartmentHistory; HumanResources.Employee; HumanResources.Department; HumanResources.Shift; HumanResources.EmployeePayHistory e Person.Person;
- **Produção:** Production.WorkOrder; Production.Product; Production.ScrapReason; Production.WorkOrderRouting; Production.Location; Production.TransactionHistory e Production.ProductCostHistory;
- **Produtos:** Production.Product; Production.ProductSubcategory; Production.ProductCategory e Production.ProductModel;
- **Compras:** Purchasing.ProductVendor; Purchasing.PurchaseOrderHeader; Purchasing.ShipMethod; Purchasing.PurchaseOrderDetail; Production.Product; Purchasing.Vendor e Person.BusinessEntity; HumanResources.vEmployee;
- **Vendas:** Production.Product; Sales.vStoreWithContacts; Sales.vStoreWithAddresses; Sales.vStoreWithDemographics; Sales.vStoreWithAddresses.BusinessEntityID; Sales.Customer; Sales.SalesTerritoryHistory; Sales.SalesTerritory; Sales.SalesOrderHeader e Sales.SalesOrderDetail.

5.3.2. Excel

No Excel, os dados foram expurgados, ou seja, foram eliminados os *missing values* (valores em falta) existentes nos campos das *views* criadas no SQL Server Management, tendo sido também atribuídos valores adequados nos campos que se encontravam codificados.

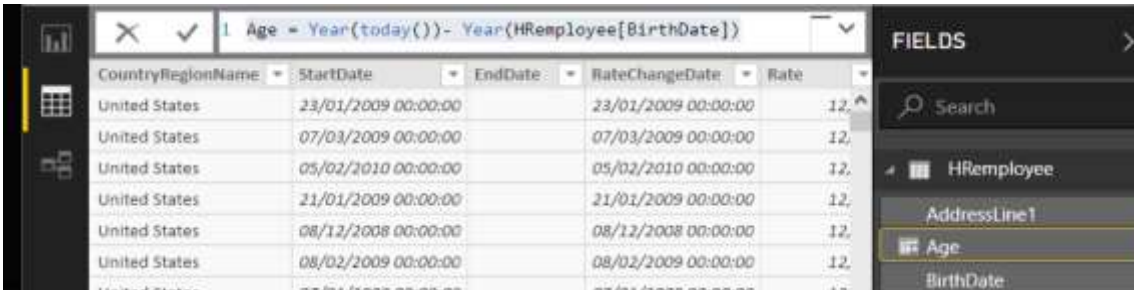
Esta preparação dos dados foi realizada em Excel por se tratar de uma solução mais simples e uma forma de experimentar a sua versatilidade para um utilizador mais comum. Contudo, este tratamento dos dados poderia ter sido

concretizado diretamente no Power BI recorrendo à linguagem Data Analysis Expressions (DAX).

5.3.3. Data Analysis Expressions

O DAX é uma linguagem de fórmula e *query* desenvolvida pela Microsoft para manipular dados em várias plataformas, tais como a SSAS (SQL Server Analysis Services), o Power BI e o Power Pivot no Excel. Esta linguagem inclui funções, operadores e valores para executar cálculos e consultas avançadas sobre dados em tabelas e colunas relacionadas em modelos de dados tabulares.

Foi necessário utilizar o DAX no Power BI para criar duas medidas. A primeira foi necessária para calcular a idade dos colaboradores da empresa, pois apenas estava disponível a respetiva data de nascença. De forma a obter-se a idade, foi utilizada a fórmula em DAX ($\text{Age} = \text{Year}(\text{today}()) - \text{Year}(\text{HREmployee}[\text{BirthDate}])$) que se pode observar na figura 10.



The screenshot shows the Power BI interface. At the top, the DAX formula bar contains the measure: $\text{Age} = \text{Year}(\text{today}()) - \text{Year}(\text{HREmployee}[\text{BirthDate}])$. Below the formula bar is a table with the following data:

Country/RegionName	StartDate	EndDate	RateChangeDate	Rate
United States	23/01/2009 00:00:00		23/01/2009 00:00:00	12
United States	07/03/2009 00:00:00		07/03/2009 00:00:00	12
United States	05/02/2010 00:00:00		05/02/2010 00:00:00	12
United States	21/01/2009 00:00:00		21/01/2009 00:00:00	12
United States	08/12/2008 00:00:00		08/12/2008 00:00:00	12
United States	08/02/2009 00:00:00		08/02/2009 00:00:00	12
United States	07/01/2008 00:00:00		07/01/2008 00:00:00	13

On the right side, the 'FIELDS' pane is visible, showing a search bar and a list of fields including 'HREmployee', 'AddressLine1', 'Age', and 'BirthDate'. The 'Age' field is currently selected.

Figura 10 - Medida para calcular a idade dos colaboradores

A segunda medida foi necessária para criar um campo que indica se certos produtos comprados pela empresa se destinam a revenda ou como componente para fabrico de outros produtos, tendo sido utilizada a fórmula em DAX ($\text{Product Purpose} = \text{IF}(\text{CONTAINS}(\text{Products}; \text{Products}[\text{Make Flag}]; \text{"Purchased"}); \text{"Re-selling"}; \text{"Components"})$) que se pode observar na figura 11.

BusinessEntityID	ProductID	OrderDate	Employee	JobTitle	Product	OrderQty	ReceivedQty	RejectedQty	Stocked
253	375	24/01/2012 00:00:00	Linda P Meisner	Buyer	Hex Nut 5	3	3	0	0
253	376	24/01/2012 00:00:00	Linda P Meisner	Buyer	Hex Nut 6	3	3	0	0
253	377	24/01/2012 00:00:00	Linda P Meisner	Buyer	Hex Nut 16	3	3	0	0
253	378	24/01/2012 00:00:00	Linda P Meisner	Buyer	Hex Nut 17	3	3	0	0

Figura 11 - Medida para indicar o destino dos produtos comprados pela empresa

5.4. Modelação

Na quarta etapa do método CRISP-DM é finalmente construída a *dashboard* com cinco páginas (Recursos Humanos, Produção, Produtos, Compras e Vendas), ou seja, uma página por cada *view* criada a partir da AdventureWorks. Na página dos Produtos é também utilizada a base de dados da “Bicicleta – Rating”, de forma a aceder à avaliação dos produtos comprados que foi registada no site. Além disso, foi também concebido em R um modelo preditivo na página das Vendas.

5.4.1. Recursos Humanos

Na *dashboard*, o objetivo da página dos Recursos Humanos é permitir ao utilizador analisar e disponibilizar a informação dos seus colaboradores de forma mais organizada e acessível.

Assim, tal como se pode observar no apêndice 2, esta página da *dashboard* é composta por:

- Cinco filtros que permitem segmentar os colaboradores por sexo, estado civil, idade, função na empresa e data de admissão;
- A tabela – “*Employee Detail Table*” – apresenta em maior detalhe os seguintes dados de cada colaborador:
 - nome;
 - departamento;
 - trabalho;
 - sexo;

- estado Civil;
- total de horas de baixa médica;
- total de horas de férias;
- número de telemóvel;
- email;
- data de admissão.
- Gráfico de barras – “*Employee Departements*” – onde é exibido o número de colaboradores por departamento;
- Gráfico de *donut* – “*Sick Leave Hours by Job*” – onde se pode constatar a percentagem de cada função na empresa relativamente ao total do número de horas de baixa médica;
- Mapa que permite saber onde vivem os colaboradores.

5.4.2. Produção

Esta página tem como objetivo permitir à empresa analisar vários aspetos da sua produção e obter a respetiva informação de forma organizada.

Assim, tal como se pode observar no apêndice 3, esta página da *dashboard* é composta por:

- Quatro filtros que permitem segmentar a produção pelo produto, pelo ano em que foi produzido, pela cor e pelo estilo do produto.
- Uma tabela – “*Product Details*” – onde se pode observar com maior detalhe o que foi produzido, através dos seguintes campos:
 - nome do produto;
 - ano, mês e dia em que foi produzido;
 - onde foi produzido;
 - linha de montagem;
 - custo de produção;
 - quantidade.

- Gráficos de barras – “*Product Cost*” e “*Product Quantity*” – onde são exibidos o custo total e a quantidade de cada produto;
- Gráfico de linhas – “*Sales*” – onde se pode observar as vendas ao longo do tempo;
- Gráfico de *donut* – “*Product Average Price*” – onde se pode observar a média do preço de cada produto desde que começou a ser produzido;
- Gráfico de previsão da produção para o ano de 2015, desenvolvido com o *Marketplace Visuals* existente nas *Visualizations* no Power BI.

5.4.2.1. Marketplace Visuals

O *Marketplace Visuals* é um sítio onde os membros da comunidade e a Microsoft contribuem com seus recursos visuais do Power BI para benefício público e os publicar no mercado AppSource. Desta forma é possível fazer o *download* desses recursos visuais e adicioná-los aos relatórios do Power BI. A Microsoft testa e aprova esses recursos visuais em termos de funcionalidade e qualidade. O AppSource é um local onde se podem encontrar aplicações, suplementos e extensões para os softwares da Microsoft, permitindo conectar milhões de utilizadores de produtos como Office 365, Azure, Dynamics 365 e Power BI, a soluções que os ajudam a realizar trabalhos com mais eficiência, rapidez e produtividade.

5.4.3. Produtos

Esta página na *dashboard* foi criada para os Produtos de forma a permitir à empresa analisar, ter a informação organizada e mais acessível dos seus produtos vendidos.

Tal como se pode observar no apêndice 4, esta página da *dashboard* é composta por:

- Quatro filtros que permitem segmentar os produtos vendidos de acordo com a origem (produzido ou comprado), classe (qualidade), subcategoria e categoria do produto;
- Uma tabela - "*Product Details*" - apresenta em maior detalhe os dados dos produtos da empresa:
 - nome do produto;
 - linha de montagem;
 - categoria;
 - subcategoria;
 - classe (qualidade);
 - origem (produzido ou comprado);
 - peso de cada produto nas vendas aparecendo em percentagem;
 - média das avaliações feitas pelos clientes aos produtos no site da empresa.
- Gráfico de *donut* - "*Subcategory Sales*" - onde se pode constatar a percentagem que cada subcategoria dos produtos em relação ao total de vendas;
- Dois indicadores - "*Sales Total*" e "*Quantity Sold*" - onde é possível observar o total das vendas e a quantidade de produtos vendida;
- Gráfico de Cluster - "*Subcategory Cluster*" - onde os produtos foram segmentados em clusters.

5.4.4. Compras

Esta página na *dashboard* foi criada para as Compras e tem como função permitir à empresa analisar, ter a informação organizada e mais acessível dos produtos que comprou.

Assim, tal como se pode observar no apêndice 5, esta página da *dashboard* é composta por:

- Seis filtros que permitem segmentar as compras feitas pela empresa pela data de compra, pelo produto, pelo colaborador que fez a compra em nome da empresa, pelo vendedor, pela classificação do crédito do vendedor e pelo estado da relação com vendedor (se continua ou não a fazer negócio com o vendedor);
- Uma tabela - "*Purchasing Details*" - apresenta em maior detalhe os dados dos produtos que a empresa comprou:
 - nome do produto;
 - colaborador que fez a compra;
 - total pago pelo produto;
 - preço unitário do produto;
 - quantidade encomendada;
 - quantidade rejeitada;
 - quantidade recebida;
 - quantidade em stock;
 - data da encomenda;
 - estado da relação com vendedor;
 - destino dos produtos.
- Gráficos de barras - "*Purchased from Sellers*" e "*Total Purchased of each Product*" - o primeiro gráfico permite observar quanto é que foi gasto nas compras feitas aos vendedores, enquanto o segundo gráfico mostra o total gasto em cada produto adquirido;
- Gráfico de *donut* - "*Paid for each Product*" - onde é exibido em percentagem o peso dos gastos da empresa na aquisição de cada produto.

5.4.5. Vendas

Nesta página foram colocadas as Vendas com o objetivo de permitir à empresa analisar as suas vendas e ter a informação destas organizada.

Tal como se pode observar no apêndice 6, esta página da *dashboard* é composta por:

- Quatro filtros que permitem segmentar as vendas pelo país onde foi vendido, vendedor do produto, ano e quadrimestre em que foi vendido;
- Uma tabela – “*Sales Geografic Distribution*” – onde se pode observar as vendas consoante a distribuição geográfica de forma mais detalhada, através dos seguintes campos:
 - província;
 - cidade;
 - loja;
 - vendas.
- Gráfico de barras – “*Products Total Sales*” – onde é possível observar o total de vendas de cada produto;
- Gráfico de linhas – “*Sales*” – onde se pode verificar como é que foram as vendas ao longo do tempo;
- Dois indicadores – “*Sales Total*” e “*Quantity Sold*” – onde é possível observar o total das vendas e a quantidade de produtos vendida;
- Gráfico ARIMA, construído através do R, onde se pode observar a previsão de vendas para os próximos três anos.

5.4.5.1. Linguagem R

De acordo com o que já foi referido anteriormente na página de *dashboard* das Vendas, foi construído um gráfico de previsão usando o modelo ARIMA com recurso a R no Power BI. Tal foi possível através de uma ligação entre o Power BI e o R Studio, como se pode observar na figura 12

```

1 # Input load. Please do not change #
2 `dataset` = read.csv('C:/Users/Rafael/REditorWrapper_f02db0a0-6278-4434-ac91
-cdb8e1f12677/input_df_1a1451c6-bfde-4b49-a559-da850ed92f06.csv', check.names =
FALSE, encoding = "UTF-8", blank.lines.skip = FALSE);
3 # Original Script. Please update your script content here and once completed copy
below section back to the original editing window #
4 library(forecast)
5 is.na(dataset)
6 dataset$Year <- as.factor(dataset$Year)
7 dataset<- ts(c(t(dataset$Month)),start = c(2011,02), end = c(2014,12),frequency
=12)
8 ForecastSales<- forecast(auto.arima(dataset))
9 ForecastSales
10 plot(ForecastSales, main="Time Forecast ARIMA")
11

```

Figura 12 - Modelo preditivo ARIMA em R Script

O R Studio é um programa que utiliza R Script, o qual permite executar um *script* constituído por uma série comandos R num arquivo de texto sem formatação, para executar os comandos de uma só vez e economizar muito tempo. O R é uma linguagem de programação e ambiente de software livre, para computação estatística e gráficos. Esta linguagem é amplamente utilizada entre estatísticos e *data miners* para o desenvolvimento de software estatístico e análise de dados.

Como foi antes referido, foi utilizado uma análise preditiva temporal utilizando o modelo de auto regressão ARIMA com o objetivo de prever os resultados dos próximos três anos da empresa, tal como se pode observar na figura 13.

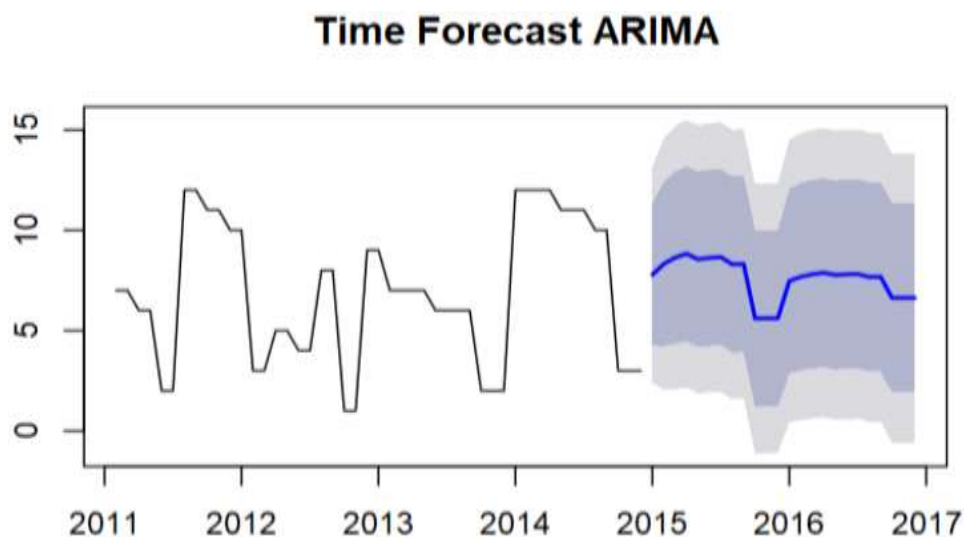


Figura 13 - Previsão ARIMA das vendas

5.5. Avaliação

O desenvolvimento desta *dashboard* foi conseguido graças à utilização de várias técnicas e recursos. Na etapa da preparação de dados foram utilizadas duas fontes de dados distintas, tendo sido desenvolvido na etapa seguinte um sistema de *analytics* de apoio à gestão recorrendo a uma *dashboard* interativa composta por diversas páginas. Nas diversas vertentes de análise dos dados que esta *dashboard* possibilita, são utilizadas várias técnicas descritivas e prescritivas.

Atendendo a todos estes fatores, o potencial do Power BI ficou comprovado como ferramenta de apoio ao BI.

5.6. Implementação

A última etapa do método CRISP-DM não se enquadra nesta investigação uma vez que esta tem como base uma dissertação, não havendo assim seguimento para uma fase de implementação.

Capítulo 6

Conclusão

6. Conclusão

6.1. Conclusão dos Resultados Obtidos

Sendo o objetivo central deste trabalho a simulação de um sistema para testar a capacidade de gestão e de análise de dados provenientes de fontes de dados diversas utilizando o Power BI, podem-se destacar duas conclusões principais, uma na perspetiva económica e a outra na perspetiva *analytics*.

Na perspetiva económica o foco está em demonstrar que o sistema de *analytics* de apoio à gestão em diversos departamentos de uma empresa, utilizando *dashboards* interativas implementadas em Power BI, está alinhado com o objetivo principal deste sistema, ou seja, possibilitar a criação de condições para obter um impacto positivo na organização. Tal impacto pode assumir vários aspetos, como por exemplo o aumento do rendimento ou a melhoria do desempenho interno dos colaboradores na empresa. Estas melhorias são possíveis graças ao sistema de *analytics* que faculta os meios necessários para os utilizadores analisarem e disponibilizarem a informação da empresa de forma mais organizada e acessível.

É possível constatar através da análise da página dos recursos humanos da *dashboard* (ver apêndice 2), que o total de horas de baixas médicas é semelhante ao total de horas de férias, e também que existem vários colaboradores com mais horas de baixa médica do que de férias. Estes factos levantam questões que são importantes para compreender as razões de eventuais anomalias e corrigi-las caso seja possível. Por exemplo, será importante determinar o número baixas que são devidas a lesão no trabalho.

Outras questões podem ser levantadas ao observar a página dos Produtos da *dashboard*. Trata-se da avaliação dos produtos feita pelos clientes, existindo vários produtos sem avaliação ou com uma avaliação média muito baixa. Tais factos indiciam alguns problemas:

- no caso dos produtos sem avaliação, a empresa deveria tentar arranjar uma forma de se ligar melhor com os clientes de modo a incentivá-los a visitarem e deixarem a sua avaliação no site da empresa;
- a questão dos produtos com uma avaliação média baixa é bastante importante, pois indicia a existência de várias avaliações más em produtos da empresa. Assim, esta deveria averiguar o motivo destas más avaliações de forma a corrigir eventuais problemas nos seus produtos.

A *dashboard* desenvolvida durante esta investigação com recurso ao Power BI mostrou que estará mais vocacionada para uma utilização predominantemente analítica. O Power BI será assim, um recurso mais adequado num contexto empresarial de *back-office* do que em *front-office*.

Por outro lado, na perspetiva *analytics* pode-se concluir que o Power BI consegue satisfazer grande parte das necessidades dos utilizadores de BI, mostrando ser uma ferramenta muito potente e versátil para a criação de *dashboards* avançadas.

Ao permitir a ligação a uma diversidade enorme de fontes de dados, o Power BI facilita a tarefa de um utilizador avançado no seu trabalho de preparação dos dados. Assim, ele pode utilizar uma grande diversidade de fontes de dados tradicionais tais como o SQL Server, Oracle, SAP, MySQL, Azure, Access, Excel ou Txt/CSV, mas também dados provenientes de IoT ou mesmo de SAFT-T, entre outras. Conectando-se ao IoT através dos serviços do Azure, pode desenvolver trabalhos de monitorização e análise, controle de inventário ou deteção de anomalias em tempo real de dispositivos, bem como a sua manutenção preditiva. Poderá também utilizar um ficheiro SAFT-T recorrendo a um conector para o

Power BI, obtendo desse modo acesso à informação dos clientes, para criar relatórios de vendas, definir alertas e notificações, obter balancetes ou rácios financeiros.

O Power BI integra várias funcionalidades de IA que facilitam a preparação de dados, como o reconhecimento de imagem e a análise de texto. Além disso, estão disponíveis outras funcionalidades de IA tais como: o Quick Insights, linguagem natural (Q&A), determinar e analisar KPIs, e ajuda à construção de modelos de Machine Learning, que permitem encontrar respostas e identificar padrões, auxiliando a exploração e análise dos dados.

No desenvolvimento da sua *dashboard*, o utilizador avançado tem ao dispor diversos tipos de recursos visuais, e se necessário, poderá ainda recorrer ao *Marketplace Visuals* que disponibiliza uma grande variedade de recursos. Durante a construção da *dashboard*, a customização desses recursos é habitualmente uma tarefa acessível. Todo este processo permite a criação de *dashboards* complexas que disponibilizam não só um elevado nível de interação ao usar variados tipos de gráficos, mas permitem também aprofundar, filtrar, classificar, dinamizar, agrupar e integrar vários tipos de modelos avançados, recorrendo à utilização de ferramentas como o R Studio ou o Python.

Após o desenvolvimento das *dashboards*, os utilizadores avançados necessitam frequentemente de partilhar o seu trabalho com utilizadores casuais, tipicamente quadros médios e altos da empresa, que podem facilmente usar as várias componentes visuais para explorar e analisar os dados da empresa de modo a obter conhecimento que ajude nas tomadas de decisão, na análise de padrões e deteção de anomalias, tal como foi apresentado em exemplos na perspetiva económica.

No caso da versão Power BI Desktop, esta partilha está restringida apenas à utilização da versão *mobile* (Power BI Mobile), podendo no entanto estender essa partilha à internet, recorrendo ao Power BI Services. Para usufruir de todas as

capacidades de partilha do Power BI, os profissionais ligados à área da BI, tipicamente utilizadores avançados, utilizam versões mais completas disponibilizadas através de SaaS. Estas versões facultam diversas formas de partilhar: partilha *peer-to-peer*, internet e vários produtos da Microsoft (Teams, SharePoint/SharePointOnline, Power Apps e Power BI Mobile).

Em suma, com a finalização desta investigação, foi possível concluir que o sistema de *analytics* de apoio à gestão utilizando *dashboards* interativas, está alinhado com o seu objetivo central. Relativamente ao Power BI, verificou-se que se trata de uma ferramenta altamente interativa e versátil, permitindo conectar-se com facilidade a múltiplas fontes de dados e utilizando *dashboards* interativas, explorar diferentes técnicas descritivas e prescritivas na análise de dados. Serão estes os principais motivos pelos quais o Power BI é atualmente considerado a melhor ferramenta de apoio ao BI no mercado.

6.2. Limitações e Trabalho Futuro

Chegada a conclusão do presente trabalho, importa referir as limitações que condicionaram o mesmo, bem como indicar alguns caminhos que poderiam permitir uma nova investigação nesta matéria.

Infelizmente, não foi possível desenvolver este trabalho com a colaboração concreta de uma empresa, não sendo assim possível contar com o valioso conhecimento do respetivo funcionamento interno, objetivos e estratégias de gestão. Desta forma, o desenvolvimento desta *dashboard* com recurso ao Power BI, foi condicionado pela dificuldade adicional em obter conhecimento de uma realidade específica, de modo a garantir o melhor impacto no funcionamento global de uma tal organização.

Relativamente a trabalho futuro, sugerem-se três ideias:

- Implementar um projeto deste tipo em ambiente real, de modo a verificar o seu impacto e as mais-valias na organização;
- Desenvolver uma *dashboard* para uma empresa de venda de serviços e uma outra de venda de bens, de forma a poder comparar os resultados;
- Aplicar a *dashboard* na área da educação como ferramenta de apoio ao ensino.

Referências Bibliografias

- Abd Elmonem, M. A., Nasr, E. S., & Geith, M. H. (2016). Benefits and challenges of cloud ERP systems – A systematic literature review. *Future Computing and Informatics Journal*, 1(1–2), 1–9. <https://doi.org/10.1016/j.fcij.2017.03.003>
- Alter, S. (1996). *Information systems: a management perspective* (2nd ed.). Benjamin/Cummings.
- Amaral, L. A. M. do. (1994). *PRAXIS Um Referencial para o Planeamento de Sistemas de Informação*. Universidade do Minho.
- AT - Autoridade Tributária e Aduaneira. (n.d.). SAF-T PT (Standard Audit File for Tax purposes) - Versão Portuguesa. Retrieved December 2, 2019, from http://info.portaldasfinancas.gov.pt/pt/apoio_contribuinte/SAFT_PT/Paginas/news-saf-t-pt.aspx
- Barnatt, C. (1996). *Management Strategy and Information Technology: Text and Readings*. Thomson.
- Bunn, G., Bartlett, C., & Mclean, D. (1989). *Strategic Planning for Information Systems: Ensuring that the business benefits*. John Wiley & Sons.
- Buyya, R., Vecchiola, C., & Selvi, S. T. (2013). *Mastering Cloud Computing*. In *Morgan Kaufmann* (1st ed.). <https://doi.org/10.1016/b978-0-12-411454-8.00001-2>
- Carvalho, J. A., & Amaral, L. A. M. (1993). Matriz de Actividades: um enquadramento conceptual para as Actividades de Planeamento e Desenvolvimento de Sistemas de Informação. *Sistemas de Informação*, 37–48.
- CCTA. (1993). *SSADM Version 4: Reference Manual*. NCC.
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). *CRISP-DM 1.0: Step-by-step Data Mining Guide*. Retrieved from <https://www.semanticscholar.org/paper/CRISP-DM-1.0%3A-Step-by->

- step-data-mining-guide-Chapman-Clinton/54bad20bbc7938991bf34f86dde0babfbd2d5a72
- Copeland, B. J. (2003). Artificial intelligence. <https://doi.org/10.1070/RM1988v043n01ABEH001547>
- Creamfinance. (2016). Big Data VS Smart Data - What is what? Retrieved November 13, 2019, from <https://www.creamfinance.com/big-data-smart-data/>
- Davenport, T. H. (1998). Putting the Enterprise into the Enterprise System. *Harvard Business Review*, 121–132. Retrieved from http://www8.informatik.umu.se/digitalAssets/1/1404_Davenport.pdf
- Dickson, G. W., & Wetherbe, J. C. (1985). *The Management of Information Systems*. McGraw-Hill.
- Dumbill, E. (2012). Volume, Velocity, Variety: What You Need to Know About Big Data. Retrieved Jan.
- Eckerson, W. W. (2016). *Governed Data Discovery - Balancing Flexibility and Standards*. Retrieved from http://go.timextender.com/hubfs/downloads/Analyst_Reports/Eckerson_Group_Govened_Data_Discovery_3-1-16.pdf
- Embarak, D. O. (2018). Data Analysis and Visualization Using Python. In *Data Analysis and Visualization Using Python*. <https://doi.org/10.1007/978-1-4842-4109-7>
- Flowers, S. (1988). *Success in information processing*. John Murray Pubs Ltd.
- Galliers, R. (1987a). *Information analysis: selected readings*. Addison-Wesley.
- Galliers, R. (1987b). Information Systems Planning in the United Kingdom and Australia - a comparison of current practice. *Oxford University Press*.
- Gama, J., Carvalho, A. P. de L., Faceli, K., Lorena, A. C., & Oliveira, M. (2012). *Extração de Conhecimento de Dados* (3th ed.). Retrieved from http://www.silabo.pt/Conteudos/8117_PDF.pdf

- Grenier, R., & Metes, G. (1991). *Enterprise Networking: Working Together Apart* (1st ed.). Bedford: Digital Press.
- Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). DESIGN SCIENCE IN INFORMATION SYSTEMS RESEARCH. In *Design Science in IS Research MIS Quarterly* (Vol. 28). Retrieved from https://wise.vub.ac.be/sites/default/files/thesis_info/design_science.pdf
- Howson, A. C., Richardson, J., Sallam, R., & Kronz, A. (2019). *Magic Quadrant for Analytics and Business Intelligence Platforms*. Retrieved from <https://www.gartner.com/doc/reprints?id=1-68720FP&ct=190213&st=sb>
- Hurwitz, J. S., Nugent, A., Halper, F., & Kaufman, M. (2013). *Big Data for Dummies*. Retrieved from [https://eecs.wsu.edu/~yinghui/mat/courses/fall2015/resources/Big data for dummies.pdf](https://eecs.wsu.edu/~yinghui/mat/courses/fall2015/resources/Big%20data%20for%20dummies.pdf)
- International Telecommunication Union. (2012). Overview of the Internet of things. *Series Y: Global Information Infrastructure, Internet Protocol Aspects and next-Generation Networks - Frameworks and Functional Architecture Models*, 22. Retrieved from https://www.itu.int/rec/dologin_pub.asp?lang=e&id=T-REC-Y.2060-201206-I!!PDF-E&type=items
- Kale, V. (2016). *Enhancing Enterprise Intelligence*. Taylor & Francis Group.
- Kamaleshn. (2012). Super User Role Description | Competence (Human Resources) | Emergence. Retrieved September 18, 2019, from <https://www.scribd.com/document/77988101/Super-User-Role-Description>
- Kavis, M. (2015). Forget Big Data -- Small Data Is Driving. *Forbes*. Retrieved from <http://www.forbes.com/sites/mikekavis/2015/02/25/forget-big-data-small-data-is-driving-the-internet-of-things/#3de078be661b>
- Kendall, K. E., & Kendall, J. E. (1992). *Systems Analysis and Design* (2nd ed.). Prentice Hall.
- Khan, M. A., Uddin, M. F., & Gupta, N. (2014). *Seven V's of Big Data understanding Big Data to extract value*. Retrieved from

- <http://www.asee.org/documents/zones/zone1/2014/Professional/PDFs/113.pdf>
- Klaus, H., Rosemann, M., & Gable, G. G. (2000). What is ERP? *Information Systems Frontiers*, 2(2), 141–162. <https://doi.org/10.1023/A:1026543906354>
- Laudon, J. P., & Laudon, K. C. (1994). *Management Information Systems: Organization and Technology* (3rd ed.). MacMillan.
- Lucas Jr., H. C. (1992). *The Analysis, Design and Implementation of Information Systems* (4th, Ed.). McGraw-Hill.
- Martin, E. W., DeHayes, D. W., Hoffer, J. A., & Perkins, W. C. (1994). *Managing Information Technology: What Managers Need to Know* (2nd ed.). MacMillan.
- McGaughey, R. E., & Gunasekaran, A. (2007). Enterprise resource planning (ERP): Past, present and future. In A. Gunasekaran (Ed.), *Selected Readings on Strategic Information Systems* (pp. 23–35). <https://doi.org/10.4018/978-1-60566-090-5.ch023>
- McKeown, P. G., & Leitch, R. A. (1993). *Management Information Systems: Managing With Computers*. Fort Worth, Texas: The Dryden Press.
- Microsoft. (2005). AdventureWorks OLTP Database Diagram. Retrieved from <https://www.microsoft.com/en-us/download/details.aspx?id=10331>
- Nikolopoulos, K., Metaxiotis, K., Lekatis, N., & Assimakopoulos, V. (2003, April). Integrating industrial maintenance strategy into ERP. *Industrial Management and Data Systems*, Vol. 103, pp. 184–191. <https://doi.org/10.1108/02635570310465661>
- O'Brien, J. A., & Marakas, G. M. (1990). *Management Information Systems*. McGraw-Hill Education.
- O'Brien, J. A. (1993). Management Information Systems: A Managerial End User Perspective. In *Irwin*.
- Raisinghani, M. (2004). Business intelligence in the Digital Economy: Opportunities, Limitations and Risks. In *Idea group Inc* (Vol. 1). Retrieved

- from
<https://trove.nla.gov.au/work/29948296?selectedversion=NBD25055617>
- Rascão, J. P. (2004). *Sistemas de informação para as organizações : a informação chave para a tomada de decisão* (2nd ed.). Sílabo.
- Ricardo, A. M. A. (1995). *Um Contributo para a Descrição e Avaliação de Técnicas no Planeamento de Sistemas de Informação*. Universidade do Minho.
- Robson, W. (1994). *Strategic Management and Information Systems: An Integrated Approach*. London: Pitman.
- Sager, M. (1990). *Managing Advanced Information Systems*. Prentice Hall.
- Scavo, F., Longwell, J., & Newton, B. (2012). Choosing Between Cloud and Hosted ERP, and Why It Matters. *Computer Economics Report*, 34(8), 1. Retrieved from <https://www.computereconomics.com/custom.cfm?name=postPaymentGateway.cfm&id=1728>
- School, G. C. (2002). *Management Information Systems* (1st ed.). Singapore.
- Shafique, U., & Qaiser, H. (2014). A Comparative Study of Data Mining Process Models (KDD, CRISP-DM and SEMMA). *International Journal of Innovation and Scientific Research*, 12(1), 217–222. Retrieved from https://www.researchgate.net/publication/268770881_A_Comparative_Study_of_Data_Mining_Process_Models_KDD_CRISP-DM_and_SEMMA
- Spataro, J. (2018). Bringing AI to Excel—4 new features announced today at Ignite. Retrieved November 30, 2019, from <https://www.microsoft.com/en-us/microsoft-365/blog/2018/09/24/bringing-ai-to-excel-4-new-features-announced-today-at-ignite/>
- Swart, O. (2011). Cloud vs. Hosted Services, what's the difference? Retrieved December 2, 2019, from IT News Africa website: <https://www.itnewsafrika.com/2011/04/cloud-vs-hosted-services/>
- Taylor-Sakyi, K. (2016). *Big Data: Understanding Big Data*.

- Turban, E., Delen, D., Sharda, R., & King, D. (2007). *Business Intelligence: A Managerial Approach* (1st ed.). Pearson Education Limited.
- Varajão, J. (2002). *Função de Sistemas de Informação: Contributos para a melhoria do sucesso da adopção de tecnologias de informação e desenvolvimento de sistemas de informação nas organizações*. Universidade do Minho.
- Ward, J. (1995). *Principles of Information Systems Management*. Londra, RU: Thomson Learning.
- Ward, J. M., & Griffiths, P. M. (1996). *Strategic Planning for Information Systems* (2nd ed.). John Wiley & Sons, Inc.
- Wilson, T. (1989). Towards an information management curriculum. *Journal of Information Science*, 15(4-5), 203-209.
<https://doi.org/10.1177/016555158901500403>
- Xu, X. (2012). From cloud computing to cloud manufacturing. *Robotics and Computer-Integrated Manufacturing*, 28(1), 75-86.
<https://doi.org/10.1016/j.rcim.2011.07.002>
- Zorrinho, J. C. das D. (1991). *Gestão da informação* (1st ed.). Editorial Presença.

Anexo

Anexo 1

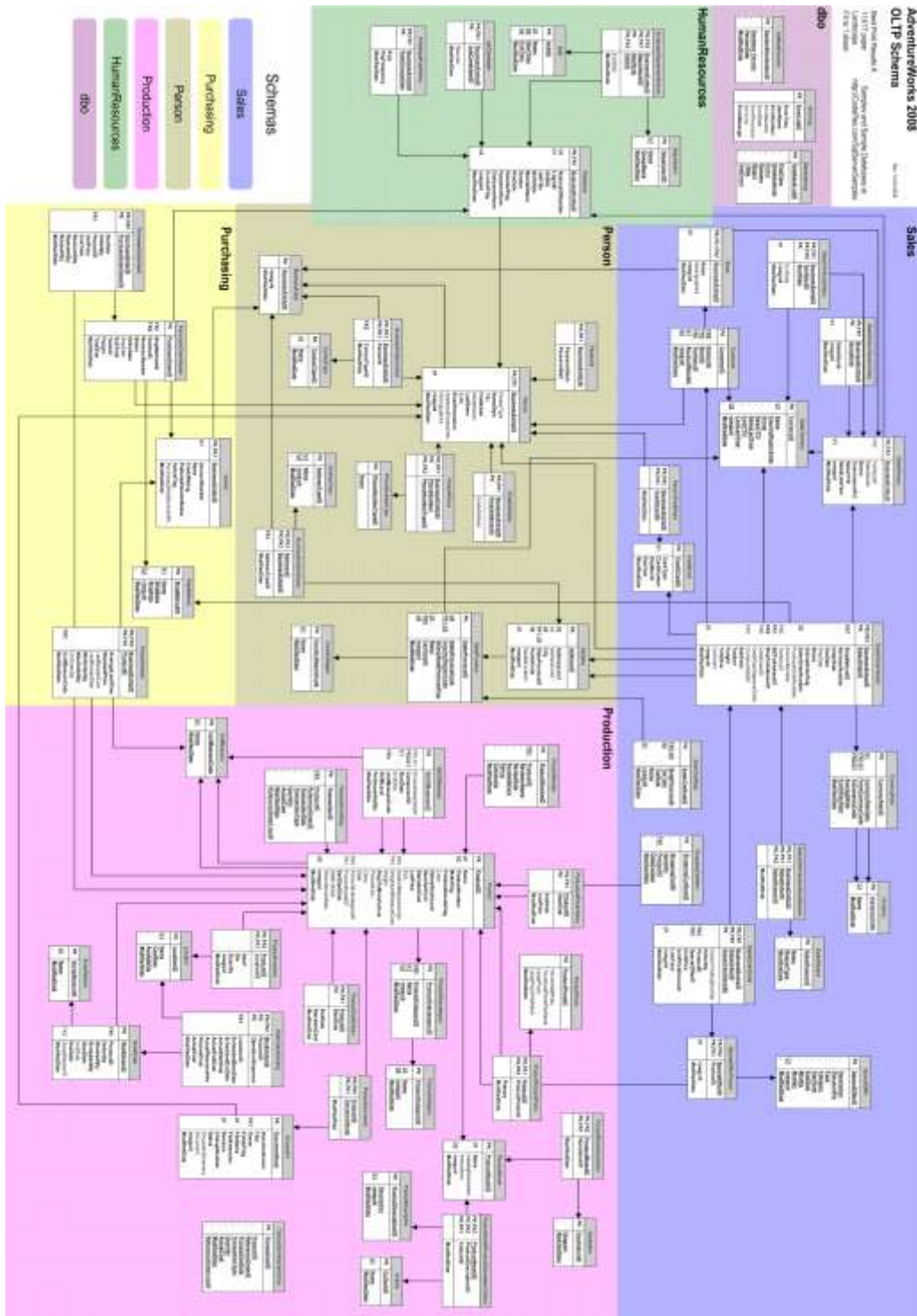


Figura 14 - Diagrama do AdventureWorks (Microsoft, 2005)

Apêndices

Apêndice 1

The screenshot displays a Microsoft Access interface. At the top, a query design grid shows the following tables and fields:

- Supplier (P...)**: Name, ShipDate, ShipSize, Longuid, ModifiedDate
- PurchaseOrderHeader (P...)**: (All Columns), PurchaseOrderID, RevisionNumber, Status, EmployeeID, VendorID, ShipMethod, OrderDate, ShipDate, SubTotal, Taxamt, Freight, TotalDue
- Employee (Humaniza...)**: (All Columns), BusinessEntityID, Title, FirstName, MiddleName, LastName
- Vendor (Partners...)**: AccountNumber, Name, CreditRating, PreferredVendorS, ActiveFlag, PurchasingWorks
- ProductVendor (P...)**: (All Columns), Product, BusinessEntityID, AverageLeadTime, StandardCost, LastReceiptCost, LastReceiptDate, LastRevisedQty, MinOrderQty, MaxOrderQty, OnOrderQty
- Product (Production)**: ProductID

The design grid also includes a 'Column' section with the following fields:

- BusinessEntityID
- ProductID

Below the design grid, the SQL text for the query is visible:

```
FROM PurchasingProductVendor INNER JOIN
PurchasingPurchaseOrderHeader INNER JOIN
PurchasingShipMethod ON PurchasingPurchaseOrderHeader.ShipMethodID = PurchasingShipMethod.ShipMethodID ON PurchasingPurchaseOrderDetail ON PurchasingPurchaseOrderHeader.PurchaseOrderID = PurchasingPurchaseOrderDetail.PurchaseOrderID INNER JOIN
ProductionProduct ON PurchasingPurchaseOrderDetail.ProductID = ProductionProduct.ProductID ON PurchasingProductVendor.ProductID = ProductionProduct.ProductID INNER JOIN
BusinessEntity ON PurchasingPurchaseOrderHeader.VendorID = PurchasingVendor.BusinessEntityID AND PurchasingProductVendor.BusinessEntityID = PurchasingVendor.BusinessEntityID INNER JOIN
BusinessEntity ON PurchasingPurchaseOrderHeader.BusinessEntityID = BusinessEntity.BusinessEntityID INNER JOIN
```

The data table at the bottom shows the following columns and values:

BusinessEntityID	ProductID	OrderDate	FirstName	MiddleName	LastName	JobTitle	Product	OrderQty	ReceivedQty	RejectionQty	StandardQty	SafetyStock	ReorderPt...	St...
256	1	2011-04-16	Elsa		Hoggen	Buyer	Adjustable - 4	3,00	0,00	1,00	1000	750	0,1	
254	339	2011-04-16	Fabrizio		Ogata	Buyer	Thin-Lam HL 3	3,00	0,00	1,00	1000	750	0,1	
254	340	2011-04-16	Fabrizio		Ogata	Buyer	Thin-Lam HL 3	3,00	0,00	1,00	1000	750	0,1	
257	530	2011-04-16	Ede		Kujala	Buyer	Seat Post	500,00	0,00	500,00	500	375	0,1	

Figura 15 - View das Compras

Apêndice 3

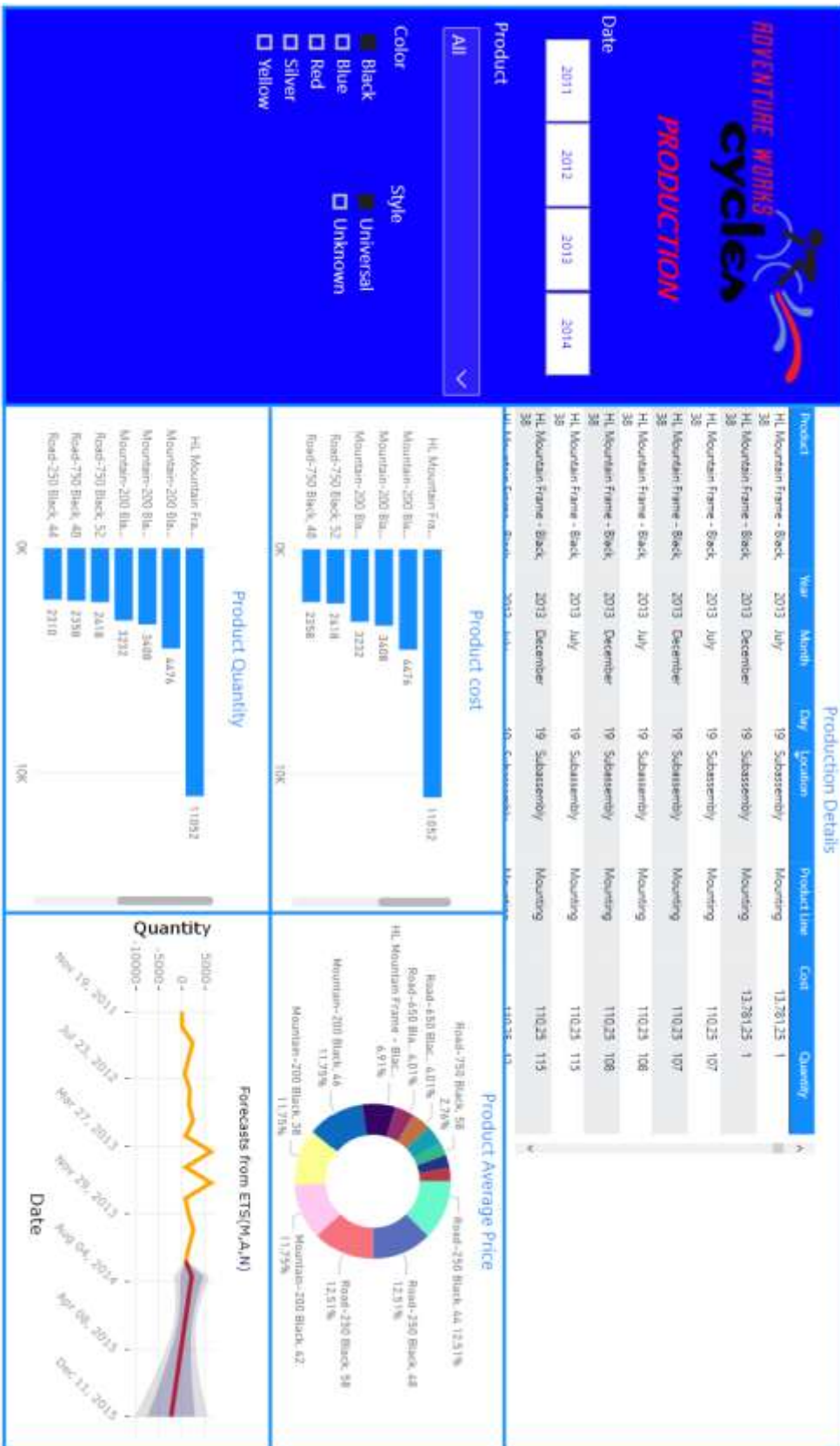


Figura 17 - Página da Produção na Dashboard

Apêndice 4

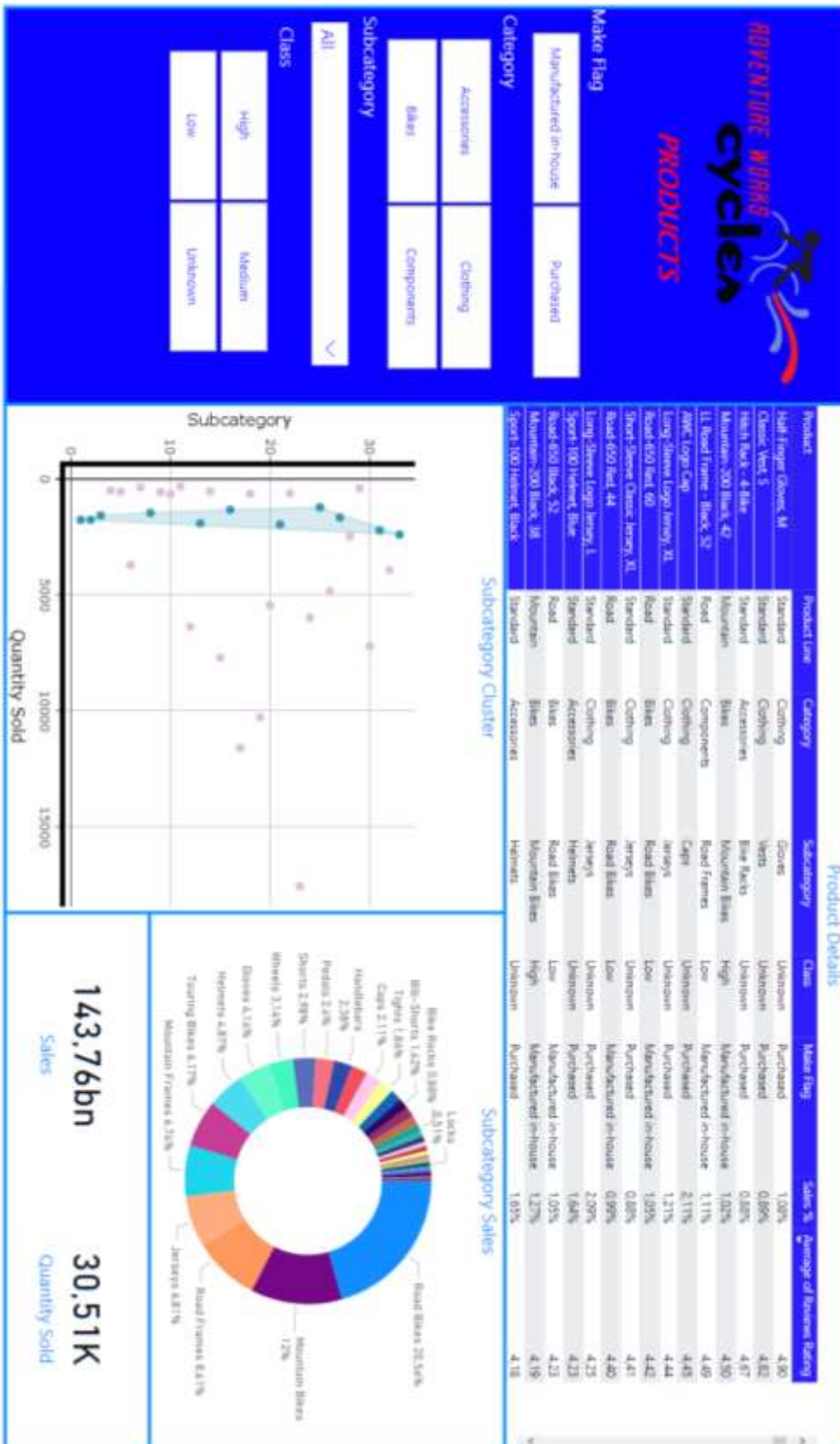


Figura 18 - Página dos Produtos na Dashboard

Apêndice 6

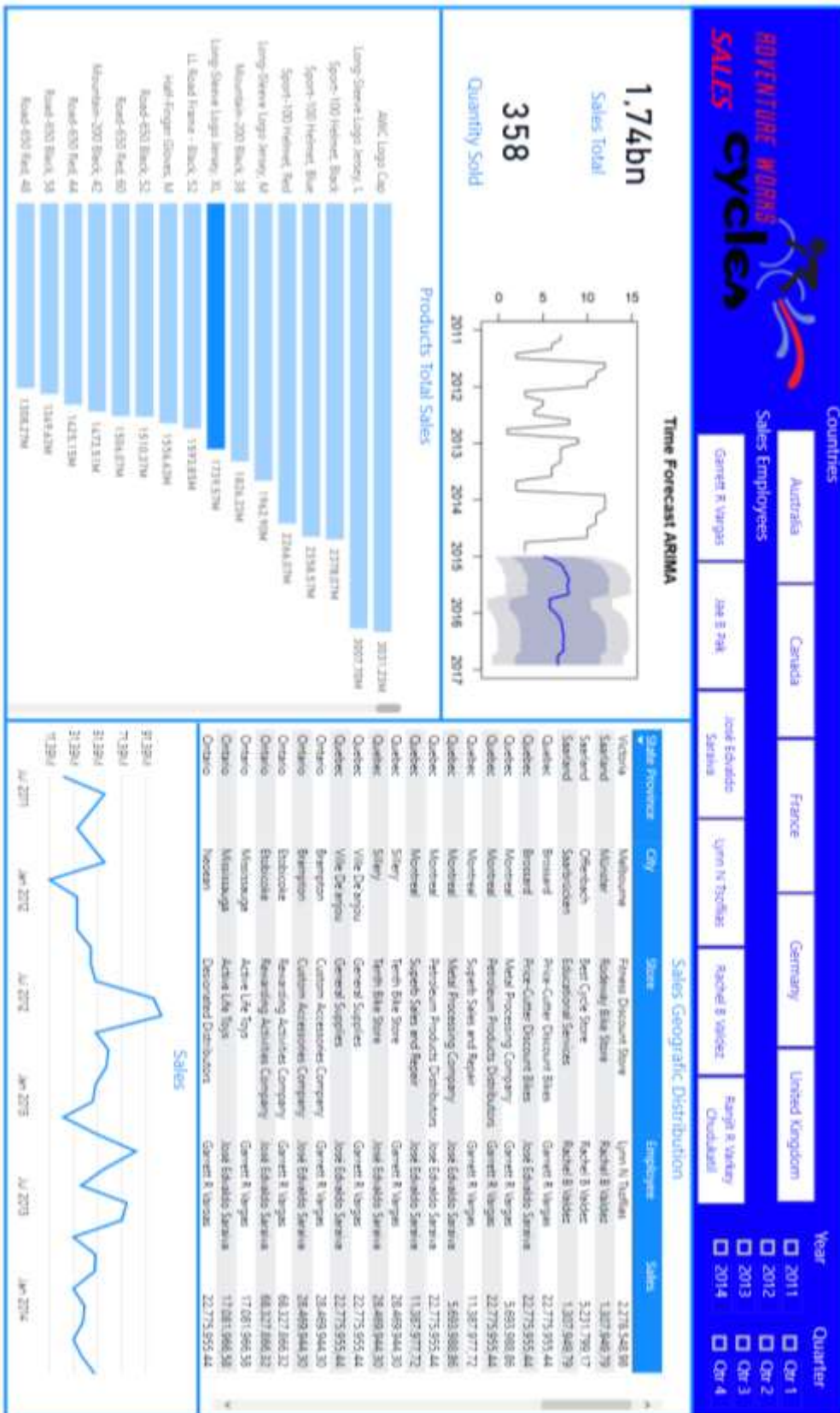


Figura 20 - Página das Vendas na Dashboard