



**CATÓLICA
LISBON**

SCHOOL OF BUSINESS & ECONOMICS

UNIVERSIDADE CATÓLICA PORTUGUESA

MSc in Strategy and Entrepreneurship

Big data analytics in healthcare: are end-users ready?

ANDRÉ JORGE MOURA PIRES

ADVISOR:

PROF. PAULO CARDOSO DO AMARAL

Dissertation submitted in partial fulfilment of the requirements for the degree of MSc in
Strategy and Entrepreneurship at Universidade Católica Portuguesa

September 2014

ABSTRACT

TITLE: Big data analytics in healthcare: are end-users ready?

AUTHOR: André Jorge Moura Pires

This dissertation aims to understand if end-users are aware of big data analytics, and given this, if the perceived value of healthcare products that use big data techniques is sufficient to surpass their concerns for sharing personal data. Additionally, it is tested whether they are interested in purchasing such products.

In order to address this topic, the theoretical foundations are based on the Theory of Reasoned Model, which studies human's decision-making process. Based on the data from a questionnaire directed to end-users, a Chi-square test studies if it exists an association between the different variables and Simple Linear Regressions evaluate the strength of the associations.

The results obtained from both type of tests prove that a higher perception of value from health products that require the use of big data technologies (*PV*) is positively correlated with a superior willingness to share personal data (*WTS*), as well as a higher willingness to buy (*WTB*), a positive word of mouth for both sharing data (*WoM_sd*) and purchasing such devices (*WoM_d*).

Finally, three Multiple Regression models are created. The first model explains *WTB* as a positive influence of *PV*, *WTS*, *WoM_d* and *WoM_sd*. The second regression tests the *WoM_d* dimension as a result of *PV*, *WTB* and *WTS*. The third model shows that *WoM_sd* is explained by *PV*, *WTB* and *WTS*. These three models are in line with the previous conclusions obtained from both the Chi-Square test and the Simple Linear Regressions.

RESUMO

TÍTULO: Big data analytics na área da saúde: estarão os consumidores finais prontos?

AUTOR: André Jorge Moura Pires

Esta dissertação tem como objetivo compreender se os consumidores finais estão cientes das técnicas analíticas de big data e, em caso afirmativo, se o valor percebido de produtos na área da saúde que usem tecnologias big data é suficiente para ultrapassar os receios de partilha de dados pessoais. Adicionalmente é testado se estes estão interessados na compra de tais produtos.

Para tal, a abordagem teórica é baseada no modelo Theory of Reasoned Action, o qual estuda o processo de tomada de decisão do ser humano. Com base nos dados obtidos através de um questionário destinado aos consumidores finais, um teste de tabelas de contingência de Qui-quadrado testa se existe associação entre as diferentes variáveis, enquanto regressões lineares simples avaliam a força destas associações.

Os resultados obtidos comprovam que uma maior percepção de valor dos produtos de saúde que exigem o uso de tecnologias big data (*PV*) está positivamente correlacionada com uma maior predisposição para a partilha de dados pessoais (*WTS*), bem como uma maior intenção para a aquisição deste tipo de produtos (*WTB*) e, finalmente, com uma positiva recomendação, tanto para a partilha de dados pessoais (*WoM_sd*) como para a compra de tais dispositivos (*WoM_d*).

Finalmente, são criados três modelos de regressões lineares múltiplas. O primeiro modelo relaciona a dimensão de *WTB* com uma influência positiva de *PV*, *WTS*, *WoM_d* e *WoM_sd*. A segunda regressão testa a dimensão de *WoM_d* associada a *PV*, *WTB* e *WTS*. O terceiro modelo mostra que *WoM_sd* é explicado por *PV*, *WTB* e *WTS*. Estes três modelos estão em linha com as conclusões anteriormente obtidas no teste Qui-quadrado e nas regressões lineares simples.

ACKNOWLEDGMENTS

Ironically these are the hardest words I had to write along this dissertation. For those who know and supported me, know I am extremely thankful for all their help.

I would like to start by expressing my deepest gratitude to my thesis advisor, professor Paulo Amaral. What started as a professor-student relationship in his Management Information Systems course ended up becoming the foundation for my thesis advisory. His guidance and emphasis on rigorous work were at all times crucial.

I am also extremely grateful to professor Francesco Sguera, whom despite not being obliged to provide me any help, was always available. His guidance was essential for the end result, especially in the creation of the questionnaire and definition of the methodological approach of this thesis.

Moreover, I would like to present my deepest thanks to those who took the time for responding to the questionnaire. Without them this thesis could not be completed, as they were a crucial source of data for the final conclusions.

To all my close friends and relatives, I would like to express my sincere gratitude not only for spreading the questionnaire among their networks but especially because during these past six months I experienced for the first time the feeling of working alone and went through some moments of frustration. Their support was vital for keeping me motivated, enthusiastic and focused on the final goal. Among them I want to express a special thank to my friend João Almeida, who worked by my side for many hours, providing not only vital intuitions but also being a constant hearing in times of struggle. Moreover, I also want to express gratitude to my uncle, Rogério Tenreiro, who provided me key insights for the statistic treatment of this dissertation.

Finally, and despite the endless help among all people that supported me during this process there were three people whose support was continuous and most vital.

First, I would like to thank my mother, for long hours of support, for keeping me motivated and focused. Furthermore, her help in correcting my writing was critical to the final text's fluency. I must say that without her this process and end result would not have been the same.

Secondly, my father, whom himself is a thesis advisor and expert in this matter, I would

like to express my deepest gratitude. Having a thesis advisor “at home” and always available was not only important but also an endless source of help for the end outcome.

In that sense, I would like to thank both my parents for the constant support not only during this process but also for providing me the means to be fortunate enough to attend such a prestigious master.

Finally, a special acknowledge to my girlfriend, Sara, who above all, was my first friendly shoulder, source of support, motivation, guidance and help in difficult situations. I must say without her this end result would not be possible.

André Pires

September 2014

TABLE OF CONTENTS

ABSTRACT	i
RESUMO	iii
ACKNOWLEDGMENTS	v
LIST OF FIGURES	xi
LIST OF TABLES	xiii
CHAPTER I. INTRODUCTION	1
1.1. Background	1
1.2. Objectives and the research question	3
1.3. Methodology	4
1.4. Structure	5
CHAPTER II. LITERATURE REVIEW	7
2.0. Literature review structure	7
2.1. Available Data	8
2.2. Big data.....	10
2.2.1. Background and concept.....	10
2.2.2. Big data analytics.....	12
2.2.3. Big data techniques	14
2.2.4. Strategic impact	16
2.2.5. Main challenges	19
2.3. Big data in healthcare	20
2.3.1. Big data for healthcare providers.....	22
2.3.2. Big data for society	25
2.3.3. Big data for end user and patients.....	26
2.4. Chapter summary	29
CHAPTER III. METHODOLOGY	33

3.1. Research question	33
3.2. Research model	34
3.3. Hypotheses development.....	39
3.4. Survey methodology.....	42
3.4.1. Brief summary of survey’s evolution	42
3.4.2. Online surveys	43
3.4.3. Questionnaire Design.....	44
3.5. Data Sampling	47
3.6. Statistical analysis	48
3.7. Chapter summary	50
CHAPTER IV. DATA ANALYSIS	53
4.1. Hypotheses analysis	53
4.1.1 – Hypothesis 1	54
4.1.2 – Hypothesis 2	55
4.1.3 – Hypothesis 3	56
4.1.4 – Hypothesis 4	59
4.1.5 – Hypothesis 5	60
4.1.6 – Hypotheses summary	61
4.2. Linear model.....	62
4.3. Chapter summary	67
CHAPTER V. CONCLUSION AND FUTURE WORK	69
5.1. Conclusion.....	69
5.2. Limitations of the study.....	70
5.3. Future research	71
APPENDIX	73
APPENDIX 1: Questionnaire.....	73

REFERENCES.....	81
-----------------	----

LIST OF FIGURES

Figure 1 - Literature Review Scheme	7
Figure 2 - Theory of Reasoned Action Model	34
Figure 3 - Research Model Scheme	38
Figure 4 - Hypothesis scheme	42
Figure 5 - Questionnaire & Hypotheses scheme.....	45
Figure 6 - Demographic sample distribution	48
Figure 7 - Questionnaire questions and sub-questions.....	50
Figure 8 - Negative or positive inclination in likert scale	53
Figure 9 - Awareness according to Gender.....	54
Figure 10 - Awareness according to age - %	55
Figure 11 - OLS models scheme	68

LIST OF TABLES

Table 1 - Questionnaire characteristics	43
Table 2 - Questionnaire statements and references associated	46
Table 3 - H ₂ 's contingency table - Q1&Q5	56
Table 4 – H ₃ 's contingency table – Q5&Q6	57
Table 5 - <i>Chi-square</i> test results (association between perceived value and individual data sharing categories)	58
Table 6 - H ₃ 's contingency table - Q5& Σ Q8	58
Table 7 - H ₄ 's contingency table - Q5&Q10	59
Table 8 - H ₅ 's contingency table - Q5&Q9	60
Table 9 - H ₅ 's contingency table - Q5&Q10	61
Table 10 - Hypotheses Chi-square test's outputs	61
Table 11 - OLS dimensions explanation	62
Table 12 - OLS simple linear regressions' outputs	63
Table 13 - <i>WTB_w</i> model's output	64
Table 14 - <i>WoM_d</i> model's output	65
Table 15 - <i>WoM_sd</i> model's output	66
Table 16 - Overall hypotheses summary	67

CHAPTER I. INTRODUCTION

1.1. Background

Data *per se* does not change or leverages firm's performance. Information, however, being the outcome of processed data (Cooper, 2014), can become a competitive advantage (Bhatt, Grover, & GROVER, 2005). Moreover, the emergence of technology in the information process, increases internal efficiency and firm's external strategic positioning (Rivard, Raymond, & Verreault, 2006).

With the rise of the technological era, the volume of data generated has grown exponentially (Nunan & Di Domenico, 2013). One example is the fact that all information ever produced by humans until 1999, is the same as the amount of data generated every 9 weeks by the world's largest telescope, the *Square Kilometre Array* (Nunan & Di Domenico, 2013). Such growth is also observed considering that every twenty months the amount of data created doubles (Adler-Milstein & Jha, 2013; Frawley, Piatetsky-Shapiro, & Matheus, 1992). Moreover, previous studies estimate that around 90% of all the data in the world has been produced in the past two years (Adler-Milstein & Jha, 2013; Murphy & Barton, 2014; Nunan & Di Domenico, 2013).

Roger Magoulas firstly defined the concept of big data in 2005, as vast amounts of data, which traditional techniques could not process due to complexity and volume (Groves, Kayyali, Knott, & Van Kuiken, 2013; Ularu, 2012).

Big data is although not just about volume. Several studies identified big data as a combination of three aspects - the three "Vs" - Volume, Velocity and Variety (Chen, Hsinchun, Roger HL Chiang & Storey, 2012; Lenard & Rubin, 2013; Russom, 2011; Zikopoulos & Eaton, 2011). These dimensions will be explored in detail in the literature review chapter, as well as further studies will be addressed in this dissertation describing big data's scope, definition and evolution.

According to (et all Moore, 2013), these changes occurring in today's data create new business possibilities with new approaches for the usage of data.

In order to overcome the challenges and take advantage of big data, new techniques to process, store, and compute it arise (Zikopoulos & Eaton, 2011). Literature defines these techniques as big data analytics (Russom, 2011).

Recent studies describe big data analytics as techniques to process large (from terabytes to exabytes) and multiparty (varies from sensor to social media) data-series that demand constrained storage, process, analysis and visualization methods (Adler-Milstein & Jha, 2013; Chen, Hsinchun, Roger HL Chiang & Storey, 2012).

The capacity of big data techniques to transform data into more accurate sets of information provide companies a potential to improve efficiency, throughout a better and more accurate vision of their business (Adler-Milstein & Jha, 2013; Schroeck, 2012; Ularu, 2012). Moreover, according to (Boyd & Crawford, 2012; Russom, 2011), big data analytics enables companies to discover hidden patterns, resulting in a better and more informed decision-making process.

By enriching the power of information, big data analytics enables the creation of value in several domains and industries (Manyika, 2011). Essentially, all industries that can increase their business functionalities and efficiency through information gain from the use of big data analytics (Russom, 2011), namely technology, government, finance, insurance, manufacturing and retailing (Manyika, 2011). Nonetheless, according to (Manyika, 2011), one of the industries that presents the highest gains from the use of big data analytics is the industry of healthcare.

In fact, the use of big data on healthcare industry is creating new market possibilities (Adler-Milstein & Jha, 2013; Kalapesi, 2013). New digital data sources, combined with new analytics, empower the decision making process, allowing earlier detection of effective treatments, real-time bio monitoring and providing sickness detection forecasting models (Adler-Milstein & Jha, 2013; Ularu, 2012). In the past ten years, medical records have been digitalized, while governments and other public stakeholders have agreed that transparency is needed on medical records' stored data (Groves et al., 2013; Nunan & Di Domenico, 2013). These processes have been creating large amounts of data in digital forms that require new techniques to process the entire data stream.

Enabling big data analytics in healthcare allows the creation of a *learning health system*, which was defined by (Adler-Milstein & Jha, 2013; Nunan & Di Domenico, 2013) as a way to “continuously identify and adopt new approaches to deliver better care at lower costs”. According to a McKinsey study, the use of big data analytics will benefit in value creation to the healthcare industry corresponding to more than \$300 billion annually (Lenard & Rubin, 2013; Manyika, 2011).

However, some key issues can become a threat to the usage of big data technologies. On one hand, data fragmentation can be seen as a barrier to the merger and analysis of all data combined (Adler-Milstein & Jha, 2013; Murphy & Barton, 2014; Nunan & Di Domenico, 2013). On the other hand, privacy issues arise when discussing big data analytics (Murphy & Barton, 2014), in such an extent that today's sharing personal data and big data analytics arise as both a problem and an opportunity.

Nonetheless, the most frequent barrier to big data analytics appointed by past literature refers to privacy concerns and end-users miss-feelings to share their personal data (et all Moore, 2013; Murphy & Barton, 2014; Nunan & Di Domenico, 2013).

Thus, end-users' feelings for sharing their personal data are still an issue on today's society. Despite the existence of several studies emphasizing the gains large organizations achieve when using big data analytics, literature lacks a broader perspective in order to understand if end-users' perception of benefits from using applications driven by big data analytics overcome their miss-feelings to share their personal data.

Having all the arguments presented into consideration, this dissertation aims to study big data opportunities in the healthcare market from the end-users' perspective.

1.2. Objectives and the research question

According to (Ularu, 2012), big data analytics is nowadays a concept heard regularly, which is able to leverage the decision making process, turning firms more efficient and effective in their processes. However, alongside with the increased firm's performance, big data also faces several challenges (Boyd & Crawford, 2012; Kalapesi, 2013). According to (Ularu, 2012), the need for companies to shift to a new data solution's reality and the fact that it is expected to happen a shortage of IT specialists and data-literate managers in the short-term represent a barrier to a wide spread of big data analytics into the market. Moreover, according to (Russom, 2011) current database software lacks on big analytic capabilities and has barriers related to data amount capacities.

However, according to (Boyd & Crawford, 2012; Manyika, 2011; Ularu, 2012) the barrier most associated against the rise of big data analytics is the privacy concerns driven from end-users' miss-feelings when considering to share their personal data.

Despite the challenges faced by big data analytics, (Manyika, 2011) states big data analytics provide opportunities to the healthcare system. Moreover, the author points out three applicabilities that augment gains from the end-user's point of view. The use of such technologies can enable a higher personalized medicine, the creation of better pricing plans and a remote patient control system. Nevertheless, all the three applicabilities require personal data from the patients.

In order to study this topic, firstly it is required to understand if end users perceive the benefits that arise from using big data techniques. In other words, it is needed to understand if applications driven from the use of big data techniques create value to end-users. Secondly, being end-users also data sources, it is necessary to understand if the benefits driven from big data techniques may surpass end user's privacy concerns. The final intent of this dissertation is to understand if there is a possibility to capture value by organizations directly from end-users' purchases of big data applications or services.

Therefore, the research question that this study addresses is:

Are end-users aware of big data analytics and their benefits? Are they interested in using healthcare services driven from big data insights and willing to share their personal data?

1.3. Methodology

As stated in previous sections, this dissertation aims to study end-users' perception about big data applications on healthcare. Moreover, this study aims to understand if end-users' perception of value driven from big data applications on healthcare surpasses their misfeelings to share their personal data. Finally, by testing if end-users are willing to purchase such applications, the possibility for organizations to capture value directly from end-users is analysed.

Aiming to develop an empirical study to understand end-users perceptions, a questionnaire is created with the goal of analysing the humans behaviour' steps until the act of purchase is consummated.

According to (Guo et al., 2007), the Theory of Reasoned Action (TRA) is the most frequent model to test human's behaviour until the actual behaviour is accomplished. Likewise, the concept of perceived value is analysed as a complement to the TRA model,

since it measures customer's perception of a service by establishing a cost-benefit relationship (Sweeney & Soutar, 2001).

Consequently, the questionnaire collects data to understand target end-users behaviour towards big data applications on healthcare, by taking into consideration the measures studied by the TRA model.

To understand if the dimensions analysed are statistically relevant and can be presented as conclusions, a *contingency table Chi-square* test is conducted with the data driven from the questionnaire. Moreover, *OLS Regressions* are computed in order to measure the strength of the associations driven from the *Chi-square* test.

Chapter 3 – *Methodology* – presents a more complete and developed description of the methodological approach adopted in this study.

1.4. Structure

This dissertation presents a five-chapter structure, aiming to answer the research question in a more adequate manner. After chapter I – *Introduction* – that provides a wide overview of the study, the second chapter – *Literature Review* – presents a detailed analysis on the topic in study. Chapter II is divided in 3 sections: 2.1 – *Available data*; 2.2 – *Big data*; 2.3 – *Big data in healthcare*.

After providing the reader with past research about the research topic, chapter III – *Methodology* – explains the research model, testing hypotheses with a correspondent questionnaire. This chapter aims to link the research model to the research question and provides a graphic view of the research model and the interrelation between hypotheses and questionnaire.

The purpose of chapter IV – *Data Analysis* – is to test the hypotheses and present primary conclusions. In order to do so, the hypotheses are validated or not, making possible to design a map of the main findings about the research proposal.

In the final chapter, chapter V – *Conclusion and Future Work* – a concise summary of what has been discussed is presented, along with the main answers to the research question and a brief overview of the study's limitations and future research proposal.

CHAPTER II. LITERATURE REVIEW

2.0. Literature review structure

As stated in section 1.4 – *Structure* – Chapter II has the objective of providing the reader with the academic *status quo* about big data’s topic and its current impact and future opportunities on the healthcare field.

As it can be observed in Figure 1 this chapter has 3 main sections: 2.1 - *Data available*; 2.2 - *Big data*; 2.3 - *Big data in healthcare*.

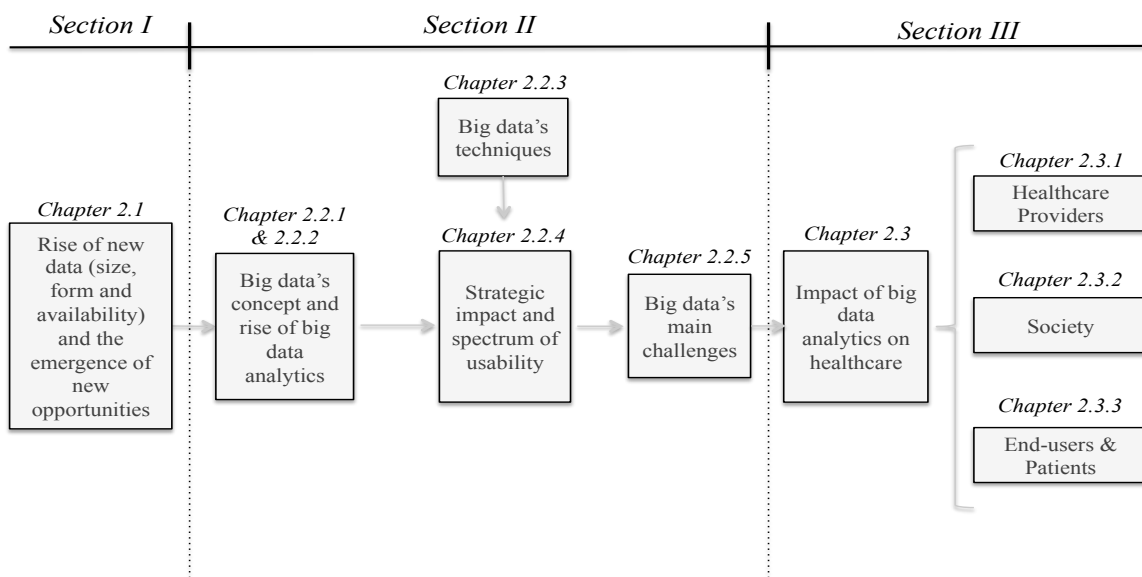


Figure 1 - Literature Review Scheme

According to past literature the world’s amount, form and availability of data is changing in an unprecedented way (Kalapesi, 2013). The section 2.1 – *Available Data* – provides the reader with a deeper knowledge of this topic, the challenges and the opportunities and the changes that today’s digitalized world brings.

The section 2.2 – *Big data* – is divided in 5 subsections. First the concept of big data and the use of big data throughout big data analytics is developed in subsections 2.2.1 – *Background and concept* – and 2.2.2 – *Big data analytics*. In this first subsection the concept of big data is developed and the most frequent big data’s definition on current literature is presented with the understanding of the 3Vs big data’s dimensions (Nunan & Di Domenico, 2013; Russom, 2011; Schroeck, 2012; Zikopoulos & Eaton, 2011). Moreover, the second subsection provides the reader with the *status quo* on current uses

of big data analytics by organizations as well as some challenges associated with big data analytics.

The third subsection 2.2.3 – *Big data techniques* – provides the reader with an overview about the current techniques and usability of big data analytics and the technologies associated to the usability.

The fourth subsection 2.2.4 – *Strategic Impact* – intends to present past research about the current impact that the use of big data techniques can ensure on organizations. Moreover, a spectrum of availabilities and current on-going projects using big data techniques are presented.

Finally, the last subsection presented on this section, 2.2.5 – *Main challenges* – provides the reader with the current barriers faced by big data analytics.

The last section of the literature review chapter 2.3 – *Big data in healthcare* – intends to understand the current and future opportunities of big data analytics on the field of study of this dissertation.

This section starts by presenting problems in the healthcare system and the opportunities driven from the use of big data technologies. Furthermore, a division of the main parties in the healthcare system as well as their benefits and challenges for each party is analysed.

By dividing section 2.3 into 3 subsections: 2.3.1 – *Big data for healthcare providers* – 2.3.2 – *Big data for society* – and 2.3.3 – *Big data for end users and patients* – a clear view is presented of the challenges and benefits, for each party analysed driven from past literature.

To conclude, this chapter presents current research about big data's field of study and the motivations for this dissertation to study big data's applicabilities on healthcare.

2.1. Available Data

The world is changing. Policies, business models and social norms differ from what existed ten years ago; in this regard, data changed in three main dimensions: size; form; and availability (Kalapesi, 2013).

Regarding the first change, size, according to (Ularu, 2012), 2,7 Zetabytes exist today in the digital world, while in 2020 the amount of data generated annually is expected to be

around 35 Zetabytes. Such figures, which embody an exponential growth, are consequence of the rise of the technological Era (Nunan & Di Domenico, 2013). One example that illustrates such growth consists in the fact that all information ever produced by humans until 1999 is the same as the amount of data generated every nine weeks by the world's largest telescope, the *Square Kilometre Array* (Nunan & Di Domenico, 2013). This growth is also observed when considering that every 20 months the amount of data created doubles (Frawley et al., 1992). Moreover, previous studies estimate that around 90% of all the data in the world has been produced in the past two years (Murphy & Barton, 2014; Nunan & Di Domenico, 2013).

The increasing usage of smart-phones and the impact of social networks shaped new data creators. Currently, more than 6 billion of people are interconnected through mobile devices, allowing data to capture individual identity (Kalapesi, 2013). As a consequence, people are no longer just subjects of data, but also data creators. This situation creates a digital footprint of people's path (Kalapesi, 2013). As another consequence of the increasing smart-phone usage, data is today not only collected actively (in surveys, registrations), but also in a passive way (web browsing, location from smart-phones, credit card transactions and sensors) (et al Smas, 2014; Kalapesi, 2013).

Finally, as an outcome of the rise the technological Era, the amount of data is boosted, by the increasing usage of social networks and smartphones (Manyika, 2011).

As a response to the data exponential growth, storage and CPU not only grew in capacity and speed (Manyika, 2011), but they are also decreasing in price (Adler-Milstein & Jha, 2013). Moreover, the ability to store higher amounts of data and for larger periods of time ensures that future data analysis presents higher volumes of data with a more historical content, enriching future insights from big data techniques (Chen, Hsinchun, Roger HL Chiang & Storey, 2012; Rose & Burgin, 2014).

Regarding the second data's change, data form, changed with the emergence and scalability of social networks and smart-phones (Chang, Kauffman, & Kwon, 2013; Chen, Hsinchun, Roger HL Chiang & Storey, 2012; et all Moore, 2013; Kalapesi, 2013; Nunan & Di Domenico, 2013; Russom, 2011). Taking Facebook as an example, it is estimated that people's interactions reach out to 30 billion pieces of content shared on Facebook every month (Manyika, 2011; Russom, 2011).

Nowadays, data fits in diverse forms, and requires contextual information. From text to

numbers, audio, video, sensors, web behaviour, and social networks interactions, data is becoming wider in such a regard, that new conclusions arise when analysing all data combined (Chang et al., 2013; et all Moore, 2013; Gantz, 2007; Russom, 2011).

Finally, the third data change, availability, points out the idea that data has not only grown in figures and form, but also is more available to every people.

According to (Chiarello, Chase, Chickering, President, & Investments, 11AD), today's world faces a democratization of information, throughout the vast spread of internet. Furthermore the author states that most council members of large organizations will face a threat to find human capital specialized in dealing with the challenges that arise with the emergence of new, more diverse and larger quantities of data.

2.2. Big data

2.2.1. Background and concept

The first time the concept of big data was addressed was in 2005. It was described as vast amounts of data, which traditional techniques could not process due to complexity and volume (Russom, 2011; Ularu, 2012).

Recent studies describe big data as techniques to process large (from terabytes to exabytes) and multiparty (varies from sensor to social media) data series that demand storage, process, analysis and visualization methods (Chen, Hsinchun, Roger HL Chiang & Storey, 2012; Roininen, L a hteenm a ki, & Tuorila, 1999).

The most frequent definitions among past literature identify big data as a combination of three aspects - the three Vs - Volume, Velocity and Variety (Chang et al., 2013; et all Moore, 2013; Nunan & Di Domenico, 2013; Russom, 2011; Schroeck, 2012; Zikopoulos & Eaton, 2011).

These three main dimensions of big data are explained below:

- **Volume:** it is the dimension most associated to big data (Russom, 2011) and refers to data's size (Chang et al., 2013; et all Moore, 2013; Schroeck, 2012). However it does not necessary has to do with bytes, terabytes and petabytes, since several companies quantify data in terms of time (e.g. seven years of data available for risk and legal analysis) (Russom, 2011; Ularu, 2012).

In accordance to (Russom, 2011) storage is not today the most complex issue about big data, although the ability to process or move data from one storage facility to other might take large amounts of time. Moreover, nowadays it is not viable to transport large amounts of data from one data set to another, creating the need for the analysis to be processed at distance and therefore requiring higher analytics capacity.

- **Velocity:** this implicates the processing capacity at real or almost real-time speed (et all Moore, 2013; Schroeck, 2012; Ularu, 2012). Data is created at an accelerated pace and firms are currently collecting and storing it at an outstanding pace (Russom, 2011). Nonetheless, several applications are only achievable if the processing capacity is doable at a real-time basis, diminishing the time between data capture and the moment it becomes accessible and it is analysed (et all Moore, 2013; Russom, 2011; Schroeck, 2012). Moreover, according to (Russom, 2011; Ularu, 2012), some activities require real time processing capacity in order to enrich firm's efficiency.

Since traditional methods do not have the capacity to obtain results in real or almost real time speed, big data analytics are needed to provide a proper response to these challenges (Russom, 2011; Schroeck, 2012) .

Additionally the combination of the first V – Volume – with the second V – Velocity – increases the need for high processing ability, as high volumes of data, which require processing ability to be at real or almost real time, increase the need for higher and more complex processing and data mining applications.

- **Variety:** being the third vertex of the big data V's triangle, variety refers to the wide spectrum of data formats (Russom, 2011; Schroeck, 2012). Structured, semi-structured and unstructured data, audio, image, text, social networking data needing contextual information compose nowadays the mixture of data available (et all Moore, 2013; Russom, 2011; Schroeck, 2012; Ularu, 2012). Despite the wide variety of data already being collected by organizations, only recently are they beginning to analyze data from different streams as one large torrent.

In addition, both volume, velocity and variety tend to boost each others (Russom, 2011).

Another vertex was added to big data's definition by (Schroeck, 2012; Ularu, 2012). The fourth "V" – Veracity – represents the level of data's reliability, since some contents are not considered to be a reliable source of information. However as few authors discussed this issue, this dissertation only considers big data as the combination of volume, velocity and variety.

Nonetheless big data is just data, therefore it does not represent any competitive advantage without analysis on top. As (Russom, 2011) said: "Big data is not new, but the effective analytical leveraging of big data is".

Big data analytics consists of techniques able to process big data in all its three dimensions (et all Moore, 2013; Russom, 2011; Schroeck, 2012).

The next subsection - *Big data analytics* - presents an analysis about the rise of big data analytics, its fields of usage and forms of applicabilities.

2.2.2. Big data analytics

As stated in the previous subsection, big data analytics consists of the techniques to operate on big data (Russom, 2011; Ularu, 2012). Despite users being aware of the huge data creation, the concept of big data analytics is still likely to be unknown (Russom, 2011; Ularu, 2012). According to a survey conducted by (Russom, 2011), 65% of participants know the implications of big data analytics but do not recognize the name and 7% of them have not heard about the concept itself.

Organizations perceive big data analytics as a group of techniques to analyze today's large amounts of data, providing new insights to leverage firm's performance (Manyika, 2011; Russom, 2011; Ularu, 2012). Furthermore, according to (Schroeck, 2012; Ularu, 2012), big data analytics exists today in several companies not only as big data analytics' providers but also as customers, purchasing big data analytics insights to improve their businesses.

According to a survey conducted by (Schroeck, 2012), the five most common big data techniques used by organizations are reporting, data mining, data visualization, predictive modelling and optimization services. Such big data techniques can be used by organizations for different purposes: to achieve a higher level of efficiency in their

processes throughout reporting and optimization techniques, or a higher segmentation of customers with the use of data mining technologies and a better and faster response to their customers changes (Manyika, 2011).

However, and despite big data analytics' wide spectrum of capabilities, only 6% of organizations are currently using big data techniques (Manyika, 2011; Schroeck, 2012). Moreover, according to the author, 24% did not use such techniques neither plans of using them in the near future, whereas 47% are beginning to plan and conducting a roadmap to incorporate such capabilities in their strategy. Finally, around 22% of organizations begin to prove the business value of big data techniques, as well as perform an assessment of their technologies and skills.

Despite big data analytics' wide spectrum of capabilities and sectors of activity, several changes are faced (Manyika, 2011; Ularu, 2012). Being a relatively new stream of technology, it is important for both organizations and final users to understand big data opportunities and how it can leverage and influence firm and customer's benefits (Boyd & Crawford, 2012; Manyika, 2011; Schroeck, 2012; Ularu, 2012). Moreover, the learning curve to adopt big data technologies is fast, but requires effort to keep track of the most recent technologies developed (Manyika, 2011; Ularu, 2012). As a matter of fact, previous studies state that only one third of companies have the goal to change its analysis throughout big data analytics technologies, whereas roughly 47% will keep their analytics without big data techniques (Manyika, 2011; Russom, 2011).

Additionally, usually big data technologies require IT specialists to perform and develop analysis that provide useful insights to influence companies' strategy (Manyika, 2011; Russom, 2011). According to (Manyika, 2011), the United States of America alone will face a shortage of 190.000 to 140.000 IT specialists in the short-term. This creates a barrier to the spread of this type of technologies across the entire market.

Another challenge big data analytics faces is its benefits' lack of perception by end users (Manyika, 2011; Russom, 2011), in which value creation by companies might not be followed by its correspondent value capturing (Manyika, 2011).

Finally, the most popular challenge associated to big data is related to end-users' privacy concerns (Boyd & Crawford, 2012; Manyika, 2011; Schroeck, 2012; Ularu, 2012).

This topic is explained in higher level of detail during the *Main challenges* subsection.

2.2.3. Big data techniques

Big data techniques are focused on a wide spectrum of activities and analysis that aggregate, manipulate, analyze and visualize data (Manyika, 2011).

These techniques drive from areas such as statistics, economics, mathematics and computer science (Manyika, 2011). Despite creating value throughout knowledge, they require a flexible and multidisciplinary approach in order to cross knowledge from different areas of expertise (Manyika, 2011). Only so, new findings can emerge from the use of big data analytics (Chen, Hsinchun, Roger HL Chiang & Storey, 2012; Manyika, 2011).

According to (Zikopoulos & Eaton, 2011) organizations that use older data analysis techniques do not retain all the opportunities and insights that data can provide if analysed accordingly. Moreover, this situation is empowered as the volume of data available to organizations is significantly higher than the amount of data they can process and analyze (Zikopoulos & Eaton, 2011). In that sense, several big data techniques emerge as a consequence of scalability problems old analytical technologies have (Manyika, 2011).

Additionally, data-driven organizations have a higher performance on financial and operational results (et all Moore, 2013). Therefore, the use of big data techniques is today essential as organizations require to take the most use of data insights as they can.

In subsection 2.2.1 – *background and concept* – the concept of big data is described in three main domains. To deal with challenges driven from the first V – Volume – several big data techniques emerged. According to (Spiess, T'Joens, Dragnea, Spencer, & Philippart, 2014) NoSQL allows to store and process millions of different accesses in real-time and it is specialized in dealing with unstructured data. Moreover, techniques such as NoSQL use scalable technologies as the example of Cassandra. Another big data technique for process and store data is, according to (Bitterer, 2011) in-memory data bases, which load and process within memory for a faster query and computation performance, being beneficial to deal with large volumes of data.

Such and similar techniques emerged as a response to the need of storing and analysing huge volumes of data with a real-time processing capacity (Manyika, 2011). For instance, Twitter appearance and fast pace spread created a need for the ability to process and store millions of tweets every day.

Another big data technique is data mining (Bitterer, 2011; Manyika, 2011; Tsuchiya, Sakamoto, Tsuchimoto, & Lee, 2012). (Manyika, 2011) describes data mining as “a set of techniques to extract patterns from large datasets by combining methods from statistics and machine learning with database management”. The author affirms that there are four main data mining techniques: association rule learning, clustering, classification, and regression.

Association rule learning discovers relationships between variables within large databases (Manyika, 2011), resulting in common combinations. The author continues by stating that one common application of association rule learning is market basket analysis.

Clustering is, according to (Manyika, 2011), a statistical method to categorize entities dividing a group into smaller groups with similar particularities. However, clustering does not allow real-time processing capacity.

The third main technique used by data mining is classification. In accordance with (Manyika, 2011) this technique allows the ability for real-time processing and to classify data in categories, folding new data objects in previous categories. This ability allows a faster process categorization for example for consumer type.

Finally, the author points out the last technique fitting data mining categorization to be regression. This technique allows understanding the way the value of the dependent variable varies when one or more independent variables are altered, being extremely useful in cases of forecasting models.

However, according to the Variety dimension, big data is frequently on different data sets and in different formats. Thus, (Manyika, 2011) affirms that is required for big data techniques are required to integrate and merge data sets and formats. Likewise, there are three techniques to prepare the different data sets for future analysis.

First, data fusion and integration is a big data technique which integrates and analyze data from multiple sources, allowing more accurate and useful insights than if independent analyses were made (Manyika, 2011). Secondly, natural language processing (NLP) consists on a set of big data techniques that not only understand linguistics but also the context of the content. Moreover, throughout machine learning techniques, NLP allows a continuous learning of new content. Finally the last technique, sentiment analysis, is according to (Manyika, 2011) extremely useful when combining with NLP, as it allows to understand the subjective information (i.e., positive, negative or neutral) and the reason

for it. The author states that, when combined with NLP techniques it allows organizations to analyse social media data, enabling to understand how different customers are perceiving their services/products.

The last vertex of big data – Velocity – requires techniques with high processing capacity. In this regard, (Manyika, 2011) points out network analysis, a technique used to categorize relationships in social networks. Moreover, MapReduce is a software firstly hosted by Google to process large databases in a separate way to accelerate the processing capacity. Finally (Chaudhuri, Dayal, & Narasayya, 2011) points out Complex Event Processing (CEP), a big data technique that allows real-time business intelligence processing. This technique is of extreme use in, for instance, financial transactions, by allowing the processing of huge data sets.

2.2.4. Strategic impact

Section 2.2 – *Big data* – identifies the fields of usage and industries in which big data analytics can be used. Likewise, this sub-section intends to provide the reader with a set of examples of the strategic impact big data analytics can add to organizations.

According to (Spiess et al., 2014), in order to improve customer experience, organizations are firstly required to be able to understand and measure these experiences. In other words, organizations need to understand what drives their customers. The author explores the topic stating that often large organizations have enough data to understand customer's feelings and expectations. Nonetheless, according to (Spiess et al., 2014), there is a need for organizations to use big data technologies in order to take the most of their data, being not only able to increase customers life cycles but also to meet customer's real expectations.

Moreover, in accordance with (Manyika, 2011), customers' segmentation and personalized services have been used for several years as a business practice, however, the appearance of big data technologies with the ability to process substantial higher amounts of data in real or almost real-time leverages these business techniques. For instance, according to the author, big data analytics increases up to 60% retailers' operating margins through a higher segmentation.

(Manyika, 2011) affirms that organizations' current decision-making process depends

mostly on employers' intuition, while the use of big data technologies may replace or support human's decision making process with automated algorithms, minimizing risks, and finding new, hidden insights that previous methods could not reach.

Furthermore, (Chang et al., 2013) affirms that big data analytics enables an unmatched way to do research, as it provides higher amounts of information from a wider variety of sources and in a real-time basis, allowing to achieve new searchable horizons. Likewise, new business opportunities appear.

The use of big data techniques such as data mining allows organizations to determine more accurately different customer segments that most probably feel their needs answered with an offer (Manyika, 2011). Likewise, according to the author, data mining big data techniques allow a market basket, which consists in a technique used by Amazon, among other organizations, which suggests purchases based on prior purchased items. This situation, usually empowered by machine learning techniques, allows the creation of a sustainable model that updates purchasing suggestions at a real-time basis.

The use of NLP techniques in social networks allows the understanding of how information flows and the discovery of how and who are the individuals that influence others (Manyika, 2011). This is extremely useful for marketing campaigns as organizations can target people who influence saving costs by following a more direct and accurate strategy. (Manyika, 2011) points out that the use of several techniques allows organizations to have a positive strategic impact in several dimensions. Reducing costs, improving company's efficiency and effectiveness, by performing the internal and external processes better. Moreover, the use of similar big data techniques allow organizations to comprehend more accurately their customers' needs, which allows a better segmentation, customization and marketing campaigns.

Despite being in the emerging stage, nowadays there are examples of organizations using big data analytics to improve their figures, results and benefits.

In accordance with (Kalapesi, 2013) research, the author provides examples of relevant cases only doable throughout big data technologies. According to the author Refugees United, a NGO refugee family tracing, used to follow methods of "pen and paper", disregarding the possibilities of mobile data and big data techniques in order to trace and regroup refugees. Nonetheless, nowadays the NGO uses a web platform that crosses mobile personal data, creating a breakthrough in the trace of broken families. Moreover,

(Kalapesi, 2013) affirms that, prior to such method, a similar NGO could support 750 refugees' tracing needs per year. However, with the use of mobile personal data, such as nicknames, names, ages, scalable networks based on friends of friends, the amount of data multiplied when crossed with an online platform which allowed Refugee United to help 183,000 families per year.

Moreover, in accordance with (Kalapesi, 2013), the use of transaction data is nowadays a reality to safeguard consumers and vendors from fraud. The author states that this method is expected to identify more than US \$1.5 billion across the globe. In this field, Feedzai, a Portuguese company, is beginning an international path, using a similar business model.

According to (Manyika, 2011), Tesco's loyalty program provides the organizations with high amounts of customers' personal data, that are used to create segmented promotions according to customers. Moreover, the author states that Amazon uses nowadays customers' behaviour based on click and purchase interface to suggest similar items to the final purchase. Another example consists of the company Smart, a wireless player in the Philippines, which crosses data from its market penetration with the average revenue per user to obtain possible smaller markets with higher potential.

The real-time analytics provided by big data technologies addresses today's business a new time frame of response to challenges, allowing early adoption of measures to respond to customers needs (Kalapesi, 2013). According to the author, such feature allows a business opportunity to achieve a more up-to-date vision of customer's needs, perceptions and intentions towards a market or product, and achieve a real-time feedback, enabling fast responses to market changes.

(Kalapesi, 2013) affirms that the methods go beyond the business approaches, as they can also be used by politics to understand citizens' demands and opinions about social topics, enabling an early adaptation of their strategies towards populations' intentions. Moreover, by crossing online search, social media interactions, and mobile data, patterns arise from big data analyses allowing early detection of disease or disasters.

Up to this point this dissertation stated gains from using big data applications resulting in real figures by being able to achieve a better customization, a more precise customers' segmentation, to engage in better pricing models, and even the beginning of predictive models to create a higher impulsive purchase. Likewise, existent cases have been presented to the reader so far. However, research for opportunities to deal directly with

end-users, as data providers, is still required.

2.2.5. Main challenges

During the big data's section, the drivers, benefits and the diverse techniques to use big data have been stated. Nonetheless, according to (Boyd & Crawford, 2012), there are challenges associated with the use of big data analytics that organizations have to face in order to take the most of its benefits.

When discussing the usability and benefits of big data analytics it is important to state that the shift big data causes on previous methods to analyze data requires, according to (Russom, 2011), a need for IT and big data specialists. One of the barriers big data analytics face is, according to the author, a shortage of adequate staff to deal with big data technologies. Moreover, (Russom, 2011) affirms that the costs associated with the replacement of current software for big data software is something to be taken into consideration as it might become a barrier to the entrance of big data technologies in the market.

Additionally, according to (Murphy & Barton, 2014), new legislation will have to be taken into consideration for big data applications. The author states that big data technologies might require similar data from different countries. Nonetheless, (Murphy & Barton, 2014) argues that different legislations within the countries providing data might incur in fraudulent activity.

(Murphy & Barton, 2014) continues by stating that there are also legal problems associated with the storage capacity big data technologies have, as data collected from organizations should be used only for the purpose they firstly defined and then should be destroyed in order to avoid the impact of data breach. Nonetheless, big data technologies enlarge the storage capacity associated in such an extent that it may result in future misuse of data utilizations that can also result in future data breaches and frauds.

As stated in previous subsections, big data analytics' capacity to find unknown and hidden patterns from data analyses that can provide new business insights can also raise legal questions. According to (Murphy & Barton, 2014), if the patterns and correlations driven by big data insights discriminate individuals according to skin colour or religion, the applicabilities found may incur in an conflict of employment law.

Nonetheless, past research maintain a common ground when considering big data's main challenge. Several authors state that privacy concerns and data providers' mistrusts to share personal data are the main barriers to the spread of big data applicabilities (Manyika, 2011; Murphy & Barton, 2014; Nunan & Di Domenico, 2013).

Literature divides data shared into two main categories: data with personal identification and anonymized data (Murphy & Barton, 2014). In this regard, the first allows a higher customization and deep knowledge of the individual itself, while the latter is used for the identification of patterns (Murphy & Barton, 2014). Nonetheless if, at a first sight, the second can be seen as less intrusive to the individual, (Murphy & Barton, 2014) affirms that scientists recently demonstrated that anonymized data can become reidentified, turning it into personal and detail data. Thus, even with anonymized protocols, data providers' mistrusts to share their personal data is understood as the main constrain to big data technologies to flourish on today's market.

Alongside with data providers' mistrusts another challenge big data analytics faces is its benefits' lack of perception by end users (Manyika, 2011), in which value creation by companies might not be followed by its correspondent value capturing (et all Moore, 2013; Lenard & Rubin, 2013; Manyika, 2011; Nunan & Di Domenico, 2013).

Hence, this dissertation aims to understand if end-users perceive applications that require the use of big data analytics as valuable and, if that is the case, if the value created for them surpass end-users' mistrusts to share their personal data. Moreover, end-users willingness to purchase such applications is tested in this study.

2.3. Big data in healthcare

This section intends to provide the reader a deeper knowledge about current big data applications on healthcare as well as future opportunities. Moreover, past research about the reason healthcare market represents an emergent opportunity for big data technologies, is presented.

Nowadays, half of the world's population live in cities (Moore, Gould, & Keary, 2003). According to (Moore et al., 2003), cities are associated with prospects of education, employment, higher number of diversified facilities and services, resulting in better life expectations. Nonetheless, the author states that the increasing migration of people into

cities creates several challenges. In accordance with (Pinger, 2006), pollution driven from industrial work and traffic jams generates health challenges, increasing the pressure on the healthcare system. Moreover, (Moore et al., 2003) affirms that air pollution represents one of the major causes for respiratory infections, being the biggest factor for the death of children up to 5 years in developing countries.

Continuous cities growth pressures metropolises' inhabitants to be pushed into wild and unprepared environments (Moore et al., 2003). As a consequence, the author states that diseases associated with wildlife tend to emerge

According to (Yongliang, 2002) the high level of population density observed in cities leveraged by a higher international mobility between countries triggers a greater propensity for dissemination and spread of pandemics.

Finally, not only cities are often associated with a more sedentary life than rural environments but there is also a higher offer of processed and high caloric food, which results in a widespread of obesity and diabetes (Moore et al., 2003).

Thus, despite offering superior life expectations, metropolises have health problems associated, which results in several challenges the healthcare system has to face.

On one hand, according to (et al Moore, 2013), technology has been a source of trigger solutions, which results in solutions for today's health problems. On the other hand, the healthcare system is becoming more digitalized in the last ten years (Groves et al., 2013), being ready for the impact technology can have on the health market.

In accordance with (Manyika, 2011) the use of big data technologies will benefit healthcare industry with more than \$300 billion annually only in the United States of America.

In accordance with (et al Moore, 2013), the executive vice president and CFO of Baptist Health South Florida states that exists vast benefits of using big data analytics in the healthcare system. Additionally, the author affirms that the executive vice president and the CFO of Geisinger Health System supports that the use of big data technologies allows the creation of a giant data warehouse which enables gathering information from different sources, such physicians, group practices, hospitals and health plans. Geisinger Health System' CFO continues that this would result in endless applicabilities for data mining and future investigation.

(et al Moore, 2013) points that Geisinger is an organization known for its propensity to use data for its Electronic Health Recording (EHR), a real-time, patient-centered record that makes information available instantly and securely to authorized users, going beyond the typical clinical data compendium.

Nonetheless, the use of big data analytics brings benefits and challenges for three different parties - Health providers; Society itself; Patients and end-users (Manyika, 2011).

Despite being intensively interconnected, there are opportunities and challenges driven from the use of big data technologies that can influence only a part of the healthcare system, where opportunities for some are challenges for others.

In accordance with (Manyika, 2011), there are four main independent and low integrated sources of data in the healthcare market: pharmaceutical R&D; clinical data; healthcare provider's data; patient's behaviour data. The author continues by stating that sharing of data provider's data can result in huge opportunities for the three interested parties.

The next three subsections explore in detail the benefits and the challenges for the three different but interconnected parties.

2.3.1. Big data for healthcare providers

(Manyika, 2011) states that, once data sharing between health players stops to be an issue among health data providers, healthcare providers will benefit in cost savings, increased efficiency in their processes, improved treatment effectiveness and a higher productivity.

Furthermore, (Manyika, 2011) states that currently considerable efforts exist to create a comparative effectiveness research with the goal of sharing research and findings across providers. This fact allows the creation of a huge database that combined with the use of big data techniques avoids patient's overtreatment and contributes to a more accurate prescription system (Manyika, 2011). In that way, there is not only a reduction of costs but also an improvement in the number of services provided.

In accordance with (Manyika, 2011) the use of big data techniques as a clinical decision support system allows a higher efficiency and quality of operations, resulting in fewer treatment errors and the consequent lawsuits that might incur in costs. Moreover, the author suggests that there are currently on going support models based on big data techniques that automatically suggest truthful treatments based on previous similar

historic medical records, avoiding diagnostic mistakes.

(Adler-Milstein & Jha, 2013) states that, the use of big data analytics on healthcare industry is creating new health applicabilities and sources of innovation to the healthcare providers. (Adler-Milstein & Jha, 2013) affirms that the combination of big data techniques and its new possibilities with the new data available (both medical and personal data), allows health providers to obtain an earlier detection of effective treatments, a real or almost real time biosurveillance and a higher accurate prediction of diseases. Furthermore, the author states the possibility of appearance of a “learning health system”, with the incorporation of machine learning and data mining techniques, which consists on an uninterrupted process to discover new approaches to deliver better medical service at lower costs.

(Manyika, 2011) talks about the current use of big data techniques for the creation of automated systems in order to achieve a more accurate fraud detection system. In this context the author affirms that from 2 up to 4% of claims tend to be fraudulent or unfounded and defends that using big data techniques on automated fraud detection systems leverages its accuracy avoiding overpayments and saving costs at a real time.

Another benefit for healthcare providers by using big data technologies is, according to (Manyika, 2011), the use of performance-based pricing plans to charge differently not only among patients but also between suppliers.

In accordance with (Manyika, 2011) the use of big data predictive models allows R&D to determine faster the drugs that are not efficient. This situation permits saving three to five years of non-useful research of the typical 13 required for drug development and approval and a more accurate redistribution of resources. Furthermore, clinical trials that tend to be long, can be leveraged by the use of big data techniques in the statistical methods used (Manyika, 2011).

Finally, one of the highest and most time consuming tasks for the healthcare providers is the patient profiling and consequence triage (Manyika, 2011). The use of big data techniques applied to patient profiling allows a higher segmentation and the proper use of a predicative modelling (Manyika, 2011). One example of the benefits for providers of such scenario is the fact that different patients with the same diseases react differently to the same treatment. Therefore (Manyika, 2011) states that a better segmentation allows a faster reaction and proactive behaviour towards different patients, or lifestyle changes,

enhancing also the ability to quantify treatments' effectiveness.

One real example of the benefits from the big data techniques is the case of *Methodist Health System* (et al Moore, 2013). According to the author, the CEO of the organization is beginning to apply the use of big data to health management by capturing hospital and medical data, EHR data and claims files from large health players. By combining both internal and external sources of data, intensive market findings have been achieved, resulting in a better understanding of patients' needs and their experiences with physicians and hospitals. Moreover, (et al Moore, 2013) states that the organization has established partnerships, being one of them one of the major competitors in order to share historical data, resulting in several operational gains for both organizations.

According to (Kalapesi, 2013), the systematic gathering and usage of health and personal data improves clinical efficiency and reduce costs. According to the author a trial, conducted in Sweden, consisting in a platform for clinical improvement, based on the comparison of best practices among providers led to an improvement of the quality of the care. (Kalapesi, 2013) continues by stating that the shared knowledge among providers enabled the identification, dissemination and adoption of best medical practices among all participants. Additionally, the author states that this example improved the quality of the services provided by below average hospitals by 40%. These figures have real-life consequences, as according to (Kalapesi, 2013), if lower performing healthcare providers in the United States of America used the same techniques on just one condition (heart attacks), more than 6.300 lives would be saved.

This dissertation does not aim to provide the reader with an extent of real big data applications on healthcare system, but to pass the idea that current big data applications in healthcare are already benefiting and increasing business performances with strategic and operational impact (Manyika, 2011).

However, in accordance with (Manyika, 2011) healthcare providers' main challenges are not the lack of technology that might arise in some cases or the patients mistrusts, but technology incompatibility and policy difficulties.

Nonetheless, patients' sharing personal data is nowadays a barrier to the implementation of several big data applicabilities, resulting in a field of study that still requires a deeper understanding (Murphy & Barton, 2014).

To conclude, the benefits for current and future applications driven from big data

techniques for health providers can have an unprecedented growth and impact on the way care services are provided (Manyika, 2011). However, there are challenges associated that need to be addressed.

2.3.2. Big data for society

As stated in the section 2.3, the world's increasing population living in cities creates a threat for fast-spread diseases (Moore et al., 2003). People tend to live in increasing crowded spaces, which augments the rapid widespread of diseases spread by respiratory and fecal-oral routes (Moore et al., 2003).

For instance, in April of 2009, Mexico faced the appearance of a new disease in form of a virus (H1N1), also known as bird's flu, which spread extremely quickly and became a pandemic being responsible for several deaths across the world (de Bruin et al., 2014). In accordance with (de Bruin et al., 2014) today's population mobility makes today's systems for monitoring pandemics with the goal of containing the virus in the early stages insufficient.

Despite not being directly linked with health providers or patients, society itself gains from using big data techniques in order to avoid pandemics and diseases-outbreaks faster and more efficiently (Kalapesi, 2013).

One real example that clarifies the benefits for society from the use of big data techniques in the healthcare system is a case, in Kenya, where it was made an experimental trial using mobile data to map and fight the spread of malaria disease. According to the author, by tracking calls and messages, population movements and patterns emerged. When crossed with the information about malaria in the country, it was possible to observe a decrease of the disease's death rate by 25%.

Likewise, (Kalapesi, 2013) affirms that through anonymized mobile phone data' analysis it is possible to comprehend the socioeconomics and well-being of communities. The author points that these applicabilities leveraged by big data techniques can ensure the creation of population movement patterns that enable a fast tracking and the creation of contingency actions in cases of natural disasters or disease-outbreaks. This applicability started in 2009 with Global Pulse working with governments to model the approaches at a smaller level and not global level to ensure real impact (Kalapesi, 2013).

In this real case, (Kalapesi, 2013) states that, by exploring big data techniques with a real-time analysis of online search data, blogs, social media helps understand population opinions about topics in a real-time basis. Such application have a scale of impact of real-time awareness of population's needs and feedbacks to achieve an early and more accurate response to situations of crisis. Moreover, (Kalapesi, 2013) affirms that these applications also spread to business areas with the aim of understanding customers' needs, feedbacks, opinions and changes at a real-time basis.

To conclude, big data techniques not only have currently higher advantages when compared with more traditional methods in the case of health problems that are not directly linked with health providers, but also the continuous growth of cities and consequently pandemic diseases require future spread of usage of big data techniques as a more efficient form of combat to society's health problems.

2.3.3. Big data for end user and patients

As stated in previous sections, most of big data applications in healthcare require patients' personal data combined with medical and health providers' data in order to achieve useful applicabilities (Kalapesi, 2013; Manyika, 2011). However, as explored in higher detail in subsection 2, end-users mistrusts to share their personal data is also the main barrier against the spread of big data applicabilities in the health market (Murphy & Barton, 2014; Nunan & Di Domenico, 2013).

Nonetheless, according to (Manyika, 2011) the use of big data analytics can bring direct benefits for patients. As stated in previous subsection, big data analytics can and are currently prepared to achieve a better patient profiling (Manyika, 2011). The author points out that, besides the benefits for health providers, such applications also allow a customized service towards the patient as a result of better segmentation.

Furthermore, the use of analytics for the creation of a personalized medicine allows health providers early detection and diagnostics before the patient develops symptoms (Manyika, 2011). Likewise, by allowing patients' medical database to be crossed and shared among providers, patients with same symptoms and diagnostics can be better segmented and later cured (Manyika, 2011). This situation happens, as the author points out, from the fact that different molecular signatures and life habits occur in different treatments effectiveness.

In this regard, (Manyika, 2011) mentions the example of drugs dosage that can be adjusted differently for the same disease, which not only allows patients to save drugs expenses but also to minimize side effects driven from drugs usage. Finally, related to personalized medicine, (Manyika, 2011) affirms that such ability can ensure a higher healing ability in long-term diseases such as lung cancer, where early detection can be the difference between life and death.

Another clinical service that can be leveraged by big data techniques is the remote patient monitoring (Manyika, 2011). Such applications consist in a constant data collection, remotely, from patients. According to the author the gains for providers, but especially for patients, are endless. In fact, it allows to determine if patients are truly doing what was agreed, which gives space for future treatment improvements (Manyika, 2011). Taking for example a problem stated in previous sections largely associated with the high migration to cities – increasing obesity – (Moore et al., 2003). Health diet programs have to be adapted from patient to patient, and similar big data application allows to understand if patients are in fact doing what was agreed, and if not, it is possible to change the treatment.

Additionally (Manyika, 2011) estimates that in the year of 2010 in the United States more than 150 million patients suffered from diseases such as diabetes, congestive heart failure and hypertension, being responsible for more than 80% of health costs. (Manyika, 2011) states that, with the usage of remote controlling devices by patients, real-time data can be collected, allowing a larger control of symptoms such as blood-sugar levels, walking habits, heart-beat, and even “chip-on-a-pill” technology, which allows a digital tracker of patients’ consumption of prescribed drugs. In practical terms, by using data collected from remote monitoring systems, it is possible to lower patient in-hospital bed days, to cut emergency department visits, and to reduce long-term health complications and bureaucratic services (Manyika, 2011). In other words, such big data applications allow to save lives and to reduce the time spent in the hospital.

The increasing and fast improvement made by the use of big data techniques in the healthcare market (Moore et al., 2003) provides large benefits not only for health providers, but also for society and patients.. In that regard (Biesdorf & Niedermann, 2014) conducted a survey stating that more than 75% of the respondents “would like to use digital healthcare services, as long as those services meet their needs and provide the level of quality they expect”.

According to Biedorf's survey regarding patients' use of technology in healthcare services, there are two conclusions that contradict what the author was expecting. Firstly, despite younger people are more opened to the use of technology concerning health services, older people also desire the use of such channels, being the main difference the fact that younger people prefer the use of social media and smartphones, while older people prefer websites and e-mail. Secondly, the core features that patients desire to be fulfilled by technology in the healthcare market are a better service efficiency, large and more access to information, the integration with other channels, and a more customized service (Biesdorf & Niedermann, 2014).

Thus, applications and devices that use technology in the healthcare market are predictable to be well received by patients.

To conclude, not only patients are starting to use technology channels and devices regarding their health, but also the use of big data techniques leverage the service' features patients prefer (Biesdorf & Niedermann, 2014; Manyika, 2011). The use of big data techniques in health services allow a higher patients' segmentation and therefore a more customized service, the decrease of time spent in hospitals by patients, better medical treatments at lower costs, and even early detection of diseases enabling an improved health system that increases treatment accuracy.

According to (Atluri, Rao, & Varanasi, 2014), four new trends are emerging creating market for devices that incorporate sensors in fashion, skin, training and normal life habits equipment. This generates an unprecedented rich and complete set of personal data, enabling the creation of new businesses for the healthcare market. In accordance with the author, such devices can be in form of bracelets, smartphone apps, among others. (Atluri et al., 2014) continues by pointing out that such devices allow diagnostics to be made, not only with larger amounts of data, but also at a real time basis, detecting, for instance, cancer and kidney conditions in their early stages.

(Atluri et al., 2014) states that, such devices use wearable sensors that incorporate biosensors permits for example to obtain accurate measures of patients' muscle energy, breathing pace, heart rate and the severity of physical impact when doing sports, for example. The author proceeds and states that, by incorporating these data sources in the health market, medical decisions with the use of big data techniques such as progressive data integration, and health algorithms, can ensure the foundation of an unprecedented

active health monitoring allowing patients not only to save costs in the long term, but also to ensure them with a healthier life.

Nonetheless, such applicabilities require an extent amount of patient's personal data (Biesdorf & Niedermann, 2014; et all Moore, 2013; Kayyali, Knott, & Van Kuiken, 2014; Manyika, 2011).

Therefore, this dissertation aims to study this literature gap. The purpose is to study not only if customers are aware of the use and benefits of big data in healthcare, but also if the value created from such applications surpass patients' mistrusts for sharing personal data. Finally, it is tested if there are possibilities for value capturing by organizations throughout patients' willingness to purchase such devices.

2.4. Chapter summary

Chapter II – *Literature review* – addressed past literature about several themes. Section 2.4 intends to present a brief overview of the main topics to retain.

The spread of the Internet and digital world, changed the amount of data created at an unprecedented pace in such an extent that (Frawley et al., 1992) affirms the amount of data created doubles every 20 months. Additionally, data changes go beyond size as form and availability tend to gain relevance (Kalapesi, 2013). Roger Magoulas firstly defined these continuous and fast pace changes, which modified the concept of data, in 2005, introducing the concept of big data.

Big data's concept consists in a triangle of three main dimensions – the three Vs (Russom, 2011; Zikopoulos & Eaton, 2011). Firstly, Volume refers to the amount of data which does not necessarily means bytes, but also time and storage capacity (Ularu, 2012). Secondly, Velocity refers to the real or almost real-time pace data is created nowadays (Schroeck, 2012). Finally, Variety consists in the wide diversity of data formats, which can go from structured to unstructured, audio to text and image and social networking to sensor-data (Russom, 2011).

Moreover, the combination of these three dimensions leverages and augments the complexity to store, process and analyze today's data (Russom, 2011).

Big data analytics are a group of emerging and already established techniques and technologies to deal with the complexity of the new data challenges that older techniques

cannot address (Russom, 2011).

Despite being in the early stages, there are already several techniques that provide organizations with new applicabilities or improve old ones (Manyika, 2011). By dealing essentially with data, and having the goal to transform data into useful information, big data analytics can be applied by organizations in a wide variety of industries (Manyika, 2011).

Organizations benefit from using big data techniques as they allows a better and more precise customer segmentation, which enhances a higher and more accurate customer customization (Manyika, 2011). Moreover, the creation of automated and machine learning mechanisms allows the decision-making-process to be more accurate, minimizing risks and finding new hidden useful insights that could not be achieved with previous methods (Manyika, 2011). Furthermore, through automated algorithms and efficiency programs finding, big data technologies enable organizations to turn their internal processes more efficient and effective, reducing not only costs but increasing margins (Manyika, 2011; Russom, 2011). Finally, the emergence of big data analytics creates new business models and opportunities (Manyika, 2011).

However, big data faces several changes that need to be addressed, being the most referenced in past literature the end-users privacy concerns and users' mistrusts to share their personal data (Murphy & Barton, 2014; Nunan & Di Domenico, 2013).

According to (Manyika, 2011) one of the industries that is expected to benefit more from the use of big data techniques is the industry of healthcare, with a value creation of more than \$300 billion annually.

(Kalapesi, 2013; Moore et al., 2003) affirm that there are three parties that can distinctly benefit from the use of big data techniques in the healthcare market, namely health providers, society itself, and the focus of this dissertation, the patients.

Today's use of big data techniques in the healthcare market allows several benefits for patients (Biesdorf & Niedermann, 2014). However, most techniques to leverage patients' benefits require patients personal data (et all Moore, 2013).

In accordance with (Atluri et al., 2014) there are wearable devices that use incorporated biosensors allowing precise measures of patients' muscle energy, breathing pace, heart rate and the severity of physical impact when doing sports, for instance. By combining

patient's real-time data with big data techniques such as remote monitoring systems, a distance monitoring is achievable in such an extent that it allows to lower patient time spent in hospitals and to reduce long-term health complications, which results ultimately in better health results with less costs (Manyika, 2011). Moreover, the use of such devices allows doctors the early detection of illnesses (even before symptoms, in some cases) (Manyika, 2011). In other words, such big data applications allow saving lives and reduce the time spent in the hospital.

Nonetheless, the use of patient's data and end-users sharing of their own personal data is (according to literature) the major barrier to the spread of big data techniques in healthcare (Murphy & Barton, 2014; Nunan & Di Domenico, 2013).

Taking all the above into consideration, this dissertation aims to study this literature gap. The purpose of this dissertation is to study not only if end-users are aware of the use and benefits of big data in healthcare, but also if the value created from such applications surpass patients' mistrusts for sharing personal data. Finally, it is tested if there are possibilities for value capturing by organizations throughout patients' willingness to purchase such devices.

CHAPTER III. METHODOLOGY

3.1. Research question

Recent research about big data's foundations, its fields of applicability, the impact on organizations and its main challenges is addressed in Chapter II. Additionally, this dissertation narrows the research topic to the industry of healthcare, as according to (Manyika, 2011) it is one of the main industries that is expected to benefit the most from the use of big data techniques.

In section 2.3 – Big data in healthcare – it is shown that literature lacks an approach that focus on end-users and their perceptions about big data applicabilities in healthcare.

Having all the above into consideration, this dissertation aims to study four dimensions concerning big data analytics, healthcare and end-users.

According to (Russom, 2011; Ularu, 2012) end-users are not aware of big data analytics, and therefore cannot understand the benefits that arise from applications driven from such techniques. In this context, regarding the first dimensions, this dissertation aims provide the reader with a descriptive status quo of the level of awareness of big data analytics. The second dimension, largely related to the first one, is the understanding if end-users perceive value from the use of big data techniques in the healthcare market.

However, end-users tend to have a mistrust when asked to share their personal data with providers, being this the main barrier big data technologies have to surpass in order to spread across businesses (Murphy & Barton, 2014; Nunan & Di Domenico, 2013). Thus, the third dimension this dissertation aims to test is if end-users' mistrusts for sharing personal data are surpassed by the value perceived by big data applications.

Finally, although storage and processing capacity are decreasing their costs per unit, big data techniques are still costly for organizations (Adler-Milstein & Jha, 2013). In that regard, companies are required to be able to capture value by selling big data applications directly do end-users in order to become motivated to sell such devices. Thus, the fourth dimension intends to understand if end-users are willing to purchase such devices to test if there is a possibility for organizations to capture value of big data techniques directly from end-users.

Therefore the research question of this dissertation is:

Are end-users aware of big data analytics and their benefits? Are they interested in using healthcare services driven from big data insights and willing to share their personal data?

3.2. Research model

The Theory of Reasoned Action (TRA), developed by Fishbein and Ajzen, analyses human's behaviour towards any action (Guo et al., 2007; Ramayah, Rouibah, Gopi, & Rangel, 2009).

Several studies applied the TRA method, or an evolution driven from it, to analyse several human behaviours towards an issue. The wide spectrum of scopes and topics the TRA model can address go from individuals' decisions towards diets (Roininen et al., 1999), to human's behaviour about safety of sexual intercourse (Bowen, Williams, McCoy, & McCoy, 2001). Furthermore, (Ramayah et al., 2009) used TRA model to study intention among Malaysian investors to use Internet stock trading, while (Xie, 2011) developed an analysis about the human's acceptance of green information technology.

As it can be observed in Figure 2, the TRA model measures the human's decision-making process, from the point the individual is aware of something, until the actual behaviour is consummated.

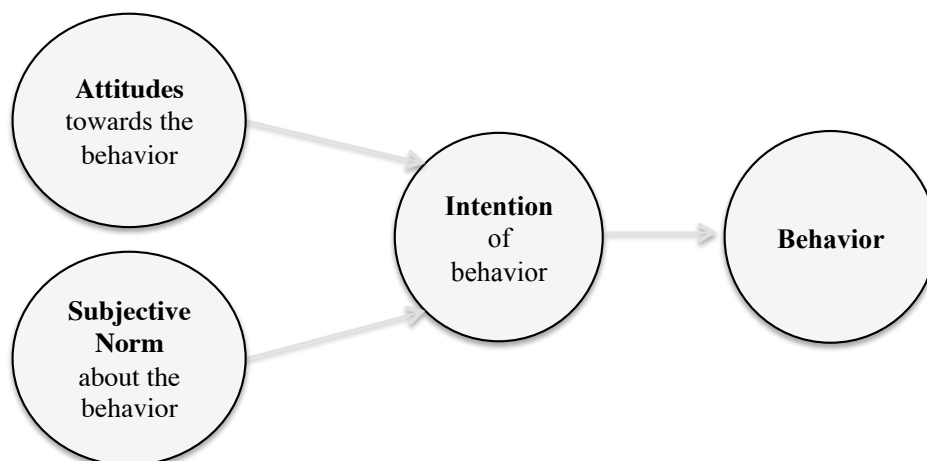


Figure 2 - Theory of Reasoned Action Model

The TRA model is divided in four main dimensions: Attitudes; Subjective Norm; Intention; Behaviour (Ramayah et al., 2009).

The first dimension – Attitudes – is divided into two smaller sub-dimensions. Attitude towards the object and attitude towards the behaviour (Ramayah et al., 2009). The author continues by explaining that in the TRA model the first attitude sub-dimension refers to the individual's beliefs about a particular object, while the later refers to the individual's belief of the consequences of the behaviour itself. One example that can illustrate these sub-dimensions is the individual's beliefs that doing exercise is good for one's health, but the consequences of such behaviour will lead to situations of discomfort.

Other studies simplify the attitude dimension with no subdivision in smaller sub-dimensions and describe attitude as the positive or negative perception of an individual towards an object (Ajzen & Fishbein, 2000).

The second dimension of the TRA model – Subjective Norm – consists in the social pressure individuals experience when doing a certain behaviour (Ramayah et al., 2009). For example, and continuing with the example presented above, one might not feel motivated to practice exercise, but if one's friends, family, work colleagues pressure him/her to do it, the final behaviour is partially a consequence of the subjective norm of the ones surrounding the individual.

(Ramayah et al., 2009) divides the concept of subjective norm into two previous behaviour dimensions, namely injunctive norm and descriptive norm. The first refers to the social surroundings approval of the individual doing certain behaviour, while the second refers to the behaviours normally executed by others the individual is connected with (Ramayah et al., 2009).

According to several authors, both attitude and subjective norm dimensions have a positive and direct correlation towards the third dimension of the TRA model – intention (Guo et al., 2007; D. Mishra, Akman, & Mishra, 2014; Ramayah et al., 2009; Shimp & Kavas, 1984). (Ramayah et al., 2009) described intention as “the perception of individuals that a particular behaviour will be performed”.

The final stage of the TRA model is the actual behaviour, as it can be observed in Figure 2 (Ramayah et al., 2009). The author states that a strong intention, also measured by the willingness, is highly positively correlated with the individual's consummation of the behaviour itself.

For a better understanding, let's consider the previously mentioned example of the practice of exercise. The way an individual would go from the state of awareness to the actual behaviour would consist in firstly understanding that exercise is useful for one's health. Secondly, by being pressured or motivated by others surrounding the individual, he/she would empower the intention to practice exercise. Finally the individual would pass to the intention stage, for instance, by choosing a proper outfit to run, ending in a one-hour running. By providing such an elementary example, this dissertation aims to provide the reader with a clear view of the TRA model.

Another key concept for understanding human's behaviour is the concept of perceived value (Gallarza & Gil Saura, 2006; Sweeney & Soutar, 2001; Xie, 2011). The concept of "perceived value" is defined as "the customer's overall appraisal of the utility of a product (or service) on the basis of perceived benefits and perceived costs". (Gallarza & Gil Saura, 2006; Sweeney & Soutar, 2001) state that the concept of perceived value is divided into three dimensions: emotional value; social value; and functional value. The first dimension, emotional value, is related with the utility driven from the feelings generated by the service/product, (e.g. the branding effect). The social value dimension consists in the way others perceive one's usage of a product or service (e.g. youngsters perceive smoking as something useful due to social pressure) (Gallarza & Gil Saura, 2006). Finally, the authors describe the functional value dimension as the price-quality relationship and customer's willingness to pay.

Despite the importance of the customer's perception of value, further studies state that not only service/product perceived quality is important to instigate purchase intentions, but also the level of satisfaction is essential for purchased decision (Gallarza & Gil Saura, 2006). However, it is important not to confuse perceived value with satisfaction, as perceiving value can happen during the entire purchase process, while satisfaction is a characteristic that only occurs after the acquisition is done and the evaluation of the consumed good is finished (Westbrook & Oliver, 1991). Therefore, perceived value is different from satisfaction, being viable for an individual to perceive value without having purchased the good or service (Sweeney & Soutar, 2001).

The concept of perceived value is intrinsically related with TRA model's dimensions of attitudes and subjective norm. First, attitudes, which is related to the positive or negative perception towards the object, is consistent to the functional value of the concept of perceived value. Secondly, subjective norm, which is related to social influence towards

one's decision, is highly similar to the emotional and social dimensions of perceived value concept.

Having all of the above into consideration, this dissertation aims to study the human behaviour according to the TRA model, by assuming that a high perception of value induces a positive intention to proceed to the actual behaviour, which in this study is ultimately the purchase by end users or patients of devices that use big data techniques for healthcare purposes. This assumption is based on the positive relationship between the two first TRA model' dimensions and the third – intention – (Guo et al., 2007; D. Mishra et al., 2014; Ramayah et al., 2009; Shimp & Kavas, 1984).

The two final concepts used by this dissertation to study end-users' full intrinsic process from the state of awareness until the willingness to purchase, are the concept of value creation and value capturing.

For the sake of readability, from now on, this dissertation mentions only patients when considering both end-users and patients.

The concept of value creation is characterized by (D. Mishra et al., 2014; Priem, 2007) as the innovation from organizations that increases customer's perception of value, encouraging a purchasing act. In that regard, the concept of value created is connected to the concept of perceived value.

In order to clarify these two concepts, one might take the example used above, the practice of exercise. Organizations create new types of shoes that are adequate to run, adding new features such as dampers, which increases the value of the product. Throughout such innovation, customers perceive a higher value from the product, which ultimately results in a higher willingness to pay.

Additionally (Quan, Hu, & Hart, 2003) states that every organization needs to create value so that customers perceive quality from their products and services in order for the purchasing intention to occur. Thus, perceiving value is a critical aspect for purchase decisions (Quan et al., 2003).

However, besides the perception of value driven from the value created from products or services, companies are required to be able to capture value. (Priem, 2007) described the concept of value capturing as “the appropriation and retention by the firm of payment made by customers in expectation of future value from consumption”.

One example that clarifies the difference between both concepts is the Facebook’s case. Facebook creates value as a social network for the users, however, their gross stream of revenues do not come from them. Taking this into consideration, Facebook creates value for users, but does not capture value from them. However, users themselves are a massive database that organizations use to make segmented publicity of their products according to user’s social networks habits. Therefore, in this case, Facebook is not only able to create value, but also to capture it.

In Figure 3 a summary of the research model can be observed. It is divided in three rows and four columns.

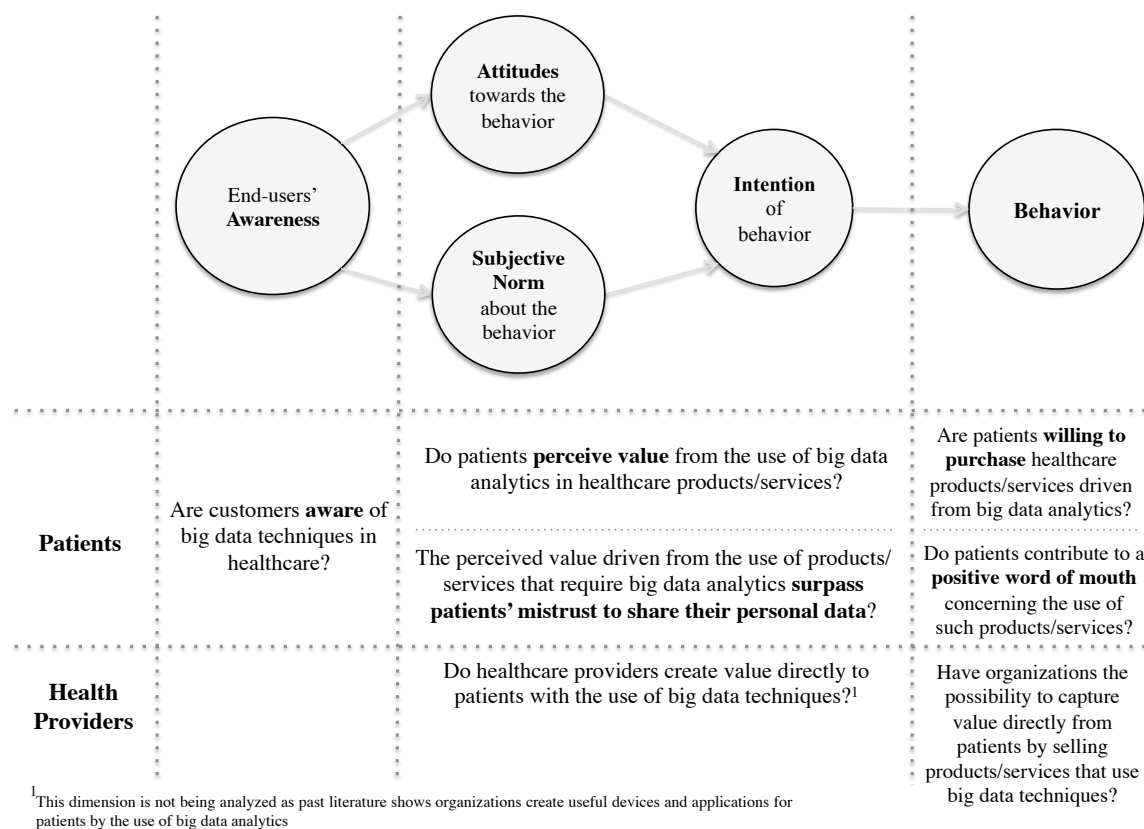


Figure 3 - Research Model Scheme

The first row represents the description of the TRA model with an antecedent dimension – awareness – since it has been stated above that, for a human to do a behaviour, he/she must be aware of the existence of the behaviour itself.

The second row evaluates patients’ decision making process according to former literature and the different TRA model’ dimensions. Firstly, patient’s state of awareness towards the existence of products or services in the healthcare market is tested. Secondly,

two dimensions are analysed according to the first three TRA's dimensions: patients' perception of value towards the use of big data techniques for health purposes and if the perceived value surpasses patients' mistrust. Finally, and to test the last TRA dimension, the final column of the line referring to patients addresses issues related to patients' willingness to purchase and if their perception of value creates positive word of mouth about the use of big data techniques in healthcare products/services.

Finally the last row tests organizations' possibility to capture value, throughout patients' purchase intentions. As it can be observed, in Figure 3, the dimension of value created is not tested as former literature addresses this issue and it is not the aim of this dissertation.

By having such structure this dissertation ensures a relationship between the TRA model, patients and organizations dimensions.

The next section shows that the hypotheses proposed are formed with academic research as its foundations.

3.3. Hypotheses development

In order to explore and reach the final conclusions for the research question, hypotheses are established. According to past literature the different phases of human's behaviour are evaluated.

The first step towards human behaviour is awareness (Guo et al., 2007; Johannessen & Olsen, 2010). The author defends that awareness was positively correlated with self-reported sustainable behaviour. Moreover, according to (O. Kwon, Lee, & Shin, 2014; Lin, Sher, & Shih, 2005), awareness is the first step of the purchase intention process.

Therefore, the first hypothesis of this dissertation is:

H₁: End-users are **aware** of the big data concept and the use of such technologies on the healthcare market

As stated in former literature, big data analytics have a wide spectrum of applications, in different fields, throughout the use of different technologies (Lin et al., 2005; Manyika, 2011). Moreover, based on subsections 3.3.1 and 3.3.2 of this dissertation, customer-

oriented applications based on big data analytics are presented, as well as the benefits that arise from such solutions.

In order to test intention's behaviour, the second phase of the human behaviour's process is related to the concept of perceived value (Gallarza & Gil Saura, 2006; Lin et al., 2005; Sweeney & Soutar, 2001). Moreover, according to the TRA model, the concepts of attitude and subjective norm, which are related to the three dimensions of perceived value, are the direct antecedent of intention (D. Mishra et al., 2014).

Both theories' concepts are related to the concept evaluated within this study. In accordance to (Oh, 2000), the dimension of value creation is linked to the concept of perceived value.

Consequently, in order to test if awareness influences a positive perception of value by end users, the second hypothesis is created:

H₂: End users' **awareness** of the use of big data technologies on healthcare industry **increases the perceived value** of the healthcare products/services by patients

In particular case of big data domain, human's behaviour theory lacks of analyses to external challenges, as behaviour concerns are the main issue related to big data applications (Lenard & Rubin, 2013).

Therefore, in order to understand if customers are willing to share their personal data, the third hypothesis is:

H₃: An **increased perception of value** by patients is expressed as **higher level of end users' willingness to share** their personal data for healthcare purposes

As mentioned in section 3.2, the last phase of the TRA model – behaviour – is not studied due to lack of data. Therefore, value capturing, the domain studied, is related to the concept of intention (Guo et al., 2007; Johannessen & Olsen, 2010).

In order to know if there is a possible source of revenue from big data technologies with B2C approach, the fourth hypothesis is created.

Therefore, the fourth hypothesis is:

H₄: An increased perception of value by patients from the use of big data technologies in healthcare services increases customer's willingness to buy

Finally, according to (Lin et al., 2005), a positive word of mouth creates higher future purchase intentions, therefore word of mouth is extremely important. Related to the social dimensions of the concept of perceived value and the subjective norm from the TRA model, word of mouth reflects the level of perception by customers (Johannessen & Olsen, 2010; Lin et al., 2005) about products/services.

Therefore, the fifth hypothesis is:

H₅: An increase perception of value by patients has a positive influence in a form of positive word of mouth contributing to the organizations' value capturing

Note: this hypothesis is composed by word of mouth for the purchase of such devices and for sharing data

Figure 4 synthetises the way the hypotheses created test the three dimensions of awareness, value creation and value capturing, according to past literature measurements of perception of value and TRA model to study these dimensions.

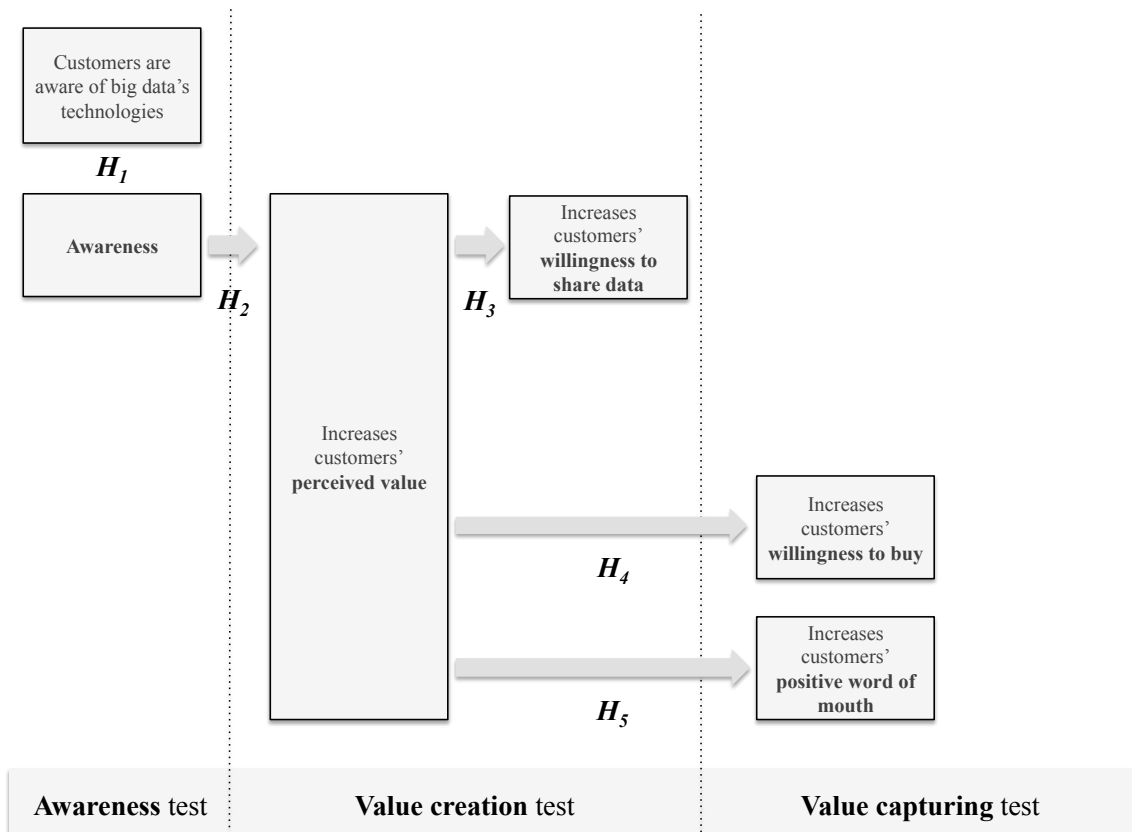


Figure 4 - Hypothesis scheme

3.4. Survey methodology

3.4.1. Brief summary of survey's evolution

The first survey is dated from 1788, when a questionnaire was sent to the ministers of all communities of the Church of Scotland (Leeuw, 2005; Szolnoki & Hoffmann, 2013). In the beginning of the twentieth century, surveys became more popular, having the first face-to-face interview with statistics purposes occurred in 1912 (Leeuw, 2005; Szolnoki & Hoffmann, 2013). Moreover, between the 40s and the 70s, face-to-face interviews and mail surveys were the most common data collection models (Szolnoki & Hoffmann, 2013). Nonetheless, according to the author, with the rise of telephone rates, telephone surveys emerged, not only due to the higher quality of data collected, but also due to lower costs of the collection process.

However, all the methods described above had some limitations concerning the speed of spreading. In the 90s, with the high Internet usage rates observed, online surveys started emerging (Rindfleisch, Malter, Ganesan, & Moorman, 2008; Szolnoki & Hoffmann,

2013)..

According to (Leeuw, 2005; McAuley & Edward, 1994; Szolnoki & Hoffmann, 2013), the most common survey forms can be compared by the characteristics observed in Table 1.

Table 1 - Questionnaire characteristics

	Face-to-face	Mail (traditional)	Telephone	Web Internet
Geographical flexibility	Low	High	High	High
Anonymity	Low	Medium	Low	High
Ability to follow-up	High	Medium	High	Low
Suitability for complex surveys	High	Low	Medium	High
Response rate	High	Low	Medium	Low
Speed	Low	Medium	Medium	High
Cost	Very High	Medium	High	Low

Despite having the lowest response rate and lacking ability to pursuit a follow-up, web surveys constitute the less costly method, with the highest spreading speed, and the ability for preserving anonymity.

3.4.2. Online surveys

Online population has been increasing, enabling online surveys to become more accurate (Leeuw, 2005). Online questionnaires ensure a quick response, being at the same time cost efficient and fast (Szolnoki & Hoffmann, 2013). Furthermore, it allows a randomization of answers choices, being useful to address sensitive issues (Szolnoki & Hoffmann, 2013). However, according to the author, online surveys lack the ability of the interviewer to clarify questions or issues and can achieve high biased patterns. Note that, this dissertation does not use the randomization ability that online surveys may have, since the presented questions are not disorganized groups but liker scales.

Moreover, (Szolnoki & Hoffmann, 2013) affirms that two main online surveys models exist: the online panel survey, using quota sampling; and the online survey with snowball sampling. The first ensure representativeness, whereas the latter allows a rapid growth

with almost no costs associated.

Additionally, to (Rindfleisch et al., 2008), there are two types of survey research: cross-sectional and longitudinal. The first is used to gather information on a population at a single point in time, while the latter gathers data over a period of time.

This dissertation, due to time and cost constraints and in order to ensure a high number of responses and a diversified sample, opts for an online survey with a snowball sampling and for a cross-sectional analysis.

3.4.3. Questionnaire Design

Opting for a questionnaire that intends to collect data throughout online devices, no interaction with the respondent happens. Therefore, most of the questions are numerical, based on a 7-point likert scale. In accordance with (McAuley & Edward, 1994) it is preferable to use a 7-point likert scale than a smaller scale. Thus this dissertation's questionnaire uses a 7-point likert scale values. Moreover, the scale used consists in an odd number in order not to force the respondent to choose a side, as the survey is long and the respondent might get annoyed.

The questionnaire has 16 questions, being represented from Q1 to Q16. Q1 to Q12 are related to the hypotheses tested, while Q13 to Q16 collects social-demographic data. The survey starts with a concise explanation of its aim and it finishes by providing the opportunity for the respondent to be informed about the conclusions achieved by receiving an email. It is divided into four main parts. Q1 and Q2 compose the first segment and inquire the respondent about big data's awareness and his/her willingness to share their own data. The second part of the survey narrows the field of big data's applications to the healthcare market with Q3 and Q4. The third part starts by presenting a concrete application that uses big data technology on the healthcare market and inquires the respondent about their perception of value, willingness to share their data, willingness to buy and willingness to spread about such applications to others. This segment of the survey is the longest (Q5 to Q12), as it tests if there is value created to customers that can be captured by organizations from applications that use big data technologies in the healthcare market. Finally, the last segment of the survey asks the respondents about their social and demographic characteristics.

Figure 5 links the questionnaire's questions to the correspondent hypothesis being tested (already presented in Figure 4).

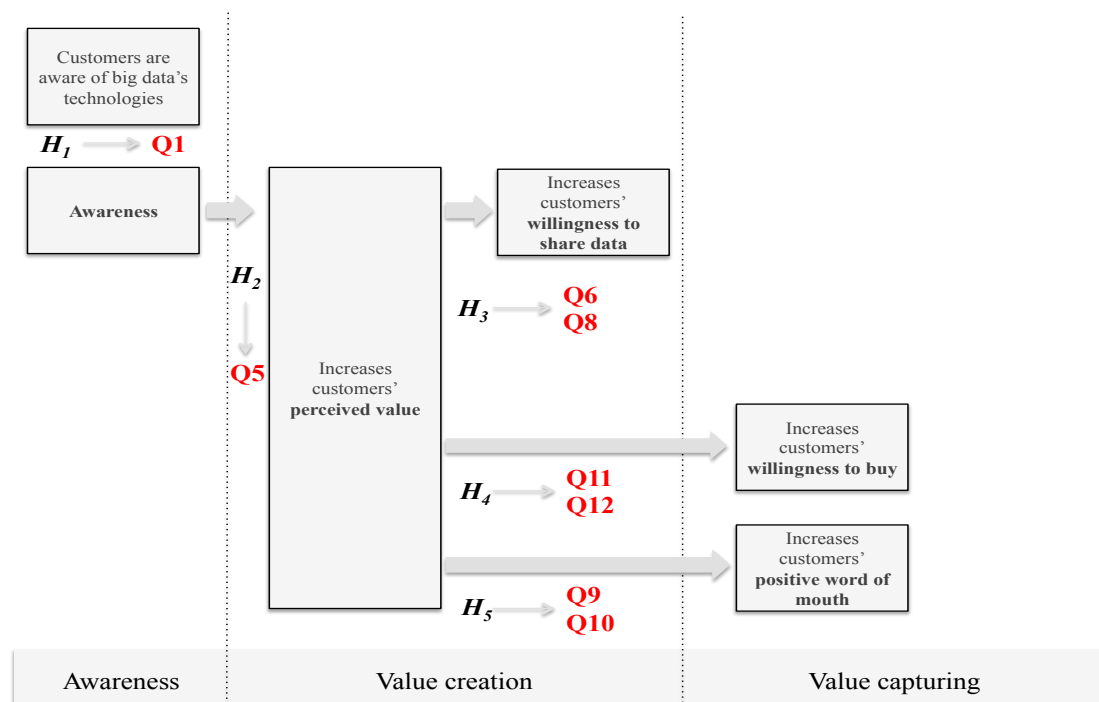


Figure 5 - Questionnaire & Hypotheses scheme

As it can be observed in Figure 5, Q1 is responsible for the test of the first hypothesis. It is important to mention that Q1 is composed by four sub-questions, being the last two dimensions associated with big data analytics and allow to test the veracity of the overall awareness test.

Q4 and Q5 test the second hypothesis and evaluate the perceived value dimension. According to (Holbrook, 1999), inquiring the respondent if he/she would enjoy something allows to test their perception of value towards an object.

Table 2 - Questionnaire statements and references associated

Questionnaire Statment	Reference
From 1 to 7 how much do you agree with the sentence bellow:	
Awareness	
I am aware of big data's concept	(Xiao Tong and Hawley 2009)
Perceived Value	
Using health providers' services that use big data analytics is something I would enjoy use	(Holbrook 1999)
I would find a similar application as the bracelet described above useful (...)	Adapted from (Swerrney and Soutar 2001)
Willingness to buy	
I woul intend to buy a similar application (...)	Adapted from (Chen & Li, 2010)
Positive Word of mouth	
I would say positive things about (...)	(Goyette et al., 2010)
I would recomend the bracelet to someone (...)	(Goyette et al., 2010)

Additionally, the third hypothesis is tested by analysing Q2, Q6, and Q8. Both Q2 and Q8 analyze end-users' mistrust to share data separately according to different data dimensions, while Q6 is a global data-sharing question. The fourth hypothesis is analysed by Q11 and Q12 where the first one measures an association between perceived value and end-users' willingness to buy, while the latter quantifies the amount the respondents are willing to spend on applications that used big data techniques for health purposes.

Finally, the last hypothesis, which refers to the relationship between perception of value and end-users' positive word of mouth is tested through Q9 and Q10. The first question refers to a positive word of mouth referring similar devices for healthcare purposes, while the second refers to word of mouth for sharing personal data.

Additionally, to ensure a higher reliable academic research, the questionnaire evaluates the dimensions studied according to former literature, as it can be observed in Table 2.

3.5. Data Sampling

The use of Internet in the diffusion has the advantage of ensuring a rapid, less costly and a wider variety of respondents. These characteristics are associated with the snowball dissemination technique already mentioned in this dissertation (Szolnoki & Hoffmann, 2013). However, the use of such technique is also linked to a biased response rate as the starting point and the environment of dissemination is usually tied up to the disseminator reach.

Therefore, despite trying to ensure a data normalization between gender and age, the snowball effect creates a biased sample due to reach limitations. Moreover, time and cost constrains create a barrier and the questionnaire is diffused mainly through social media (e.g. Facebook, Twitter, Blogs) and personal email.

Before spreading the survey through Internet-based applications, a pre-test is conducted in order to evaluate respondents' reactions and adjust the survey according to the diffusion methods. This process allows to adjust the questionnaire for the hypotheses to be analysed.

The first sample starts with a total of 543 responses. However, a dropout rate of 48% makes the useful sample to become significant smaller, as a considerable part of the total respondents do not finish the questionnaire. After a premier analysis 70 survey responses were recovered, resulting from surveys that are at least 95% completed. This process enabled a final sample of 331 respondents.

As it can be observed in Figure 6, there are gender and age biases. From the total sampling, 201 are male respondents and 130 are female ones. Additionally, due to the diffusion means, 73% of the respondents correspond to an age comprised between 18 and 30 years old.

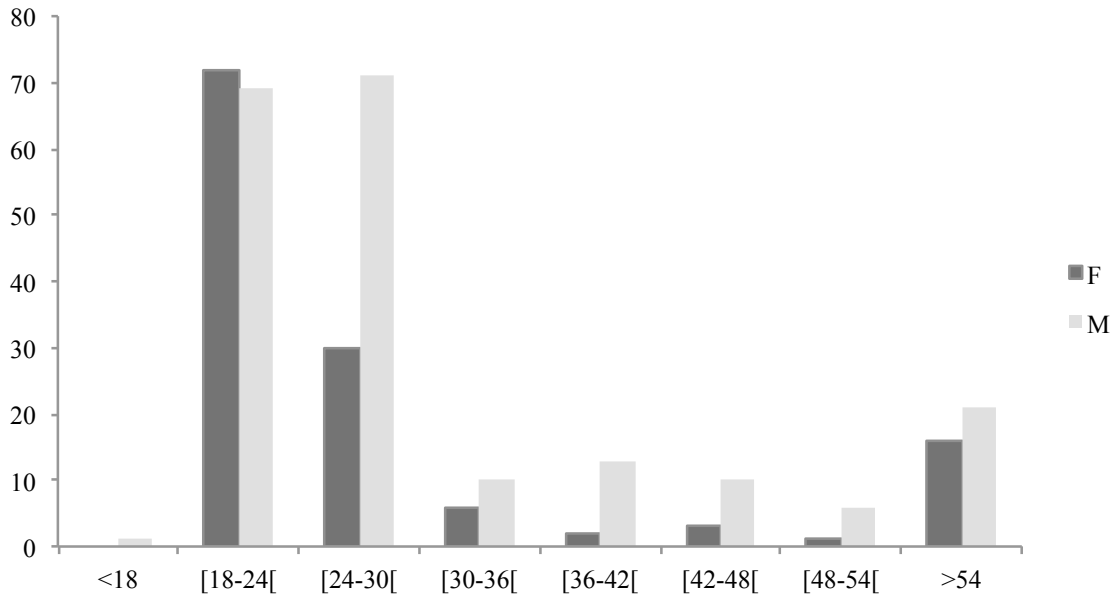


Figure 6 - Demographic sample distribution

Despite the gender and age gaps, the results obtained permitted the analysis to proceed and conclusions to be drafted from the final output.

3.6. Statistical analysis

Four of the five hypotheses (H_2 , H_3 , H_4 , H_5) presented in this dissertation are a relationship between two dimensions (e.g. awareness and perception of value).

The survey conducted is based on an ordinal scale (Jr & Boone, 2012). In this situation and being the hypothesis to be tested relational ones, the use of means, standard deviations and Pearson r 's to test the veracity of the hypothesis is considered to be inappropriate (Jr & Boone, 2012). Therefore, the use of a statistic test that uses ordinal data is necessary.

In that sense, contingency tables *Chi-square* analysis is a test that can be computed with an ordinal scale (Jr & Boone, 2012). As so, the *Chi-square* test is conducted for each of the four relational hypotheses, as it measures if there is an association between different dimensions. Taking for instance (Semedo et al., 2003) uses such analysis to test the independency between blood agar media and incubation conditions in hemolytic assays.

Subsequently, the *Chi-square* analysis is applied to compare the observed distribution

with the null hypothesis distribution, that is, the theoretical distribution that assumes independency between the two dimensions. This comparison tests if there is (or not) an association between the two dimensions. Thus, the null hypothesis states that both variables are independent, while the alternative hypothesis affirms that the dimensions are not independent, as one influences the second in a positive or negative manner (Zar, 2014).

Having all of the above into consideration, according to (Zar, 2014), the *Chi-Square* analysis formula is the following:

$$\chi^2 = \sum \sum \frac{(f_{ij} - \hat{f}_{ij})^2}{\hat{f}_{ij}}$$

Equation 1 - *Chi-Square* test formula

In Equation 1, the f_{ij} denotes the observed frequency and the \hat{f}_{ij} the theoretical distribution where both dimensions are assumed to be independent. For this test, and assuming a level of significance of 5%, the critical value is of 3,841.

Each dimension might have associated different values, which in this dissertation are represented by the likert scale values (from one to seven). In order to the *Chi-Square* test to be valid its requirements need to be fulfilled.

The *Chi-Square* analysis requirements are the following:

- Each observation is independent, which is satisfied if the sample is random (Iman & Conover, 1983);
- Frequencies with the value of zero are not accepted (Zar, 2014);
- Absolute frequencies cannot be under the value of five (Iman & Conover, 1983).

To conclude, the contingency table analysis tests if there is an association between the different dimensions analysed in this dissertation.

However, according to (Jr & Boone, 2012), if the likert scales use a neutral value, and do not create bias suggestions on the values' subtitles, it can be approximated to interval scales. Likewise, if the question presents several sub-questions related – Figure 7 – by the Central Limit Theorem, the sum of the sub-questions creates a interval scale, which allow to the use of parametric statistics (Jr & Boone, 2012).

	Strongly Disagree	Disagree	Somewhat Disagree	Neither Agree nor Disagree	Agree	Strongly Agree	Strongly Agree
	1	2	3	4	5	6	7
Sub question 1 of question 1	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sub question 2 of question 1	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sub question 3 of question 1	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sub question 4 of question 1	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Figure 7 - Questionnaire questions and sub-questions

Thus, by creating interval scale data it is possible to measure the strength of the associations driven from the *Chi-square* test.

The second test this dissertation computes is an OLS Regression in order to test the strength of the association. Finally, an explanatory to answer the research question is developed.

3.7. Chapter summary

Chapter III addresses the methodological approach of this dissertation. This section intends to present a brief overview of the main topics to retain.

This study aims to respond the research question:

Are end-users aware of big data analytics and their benefits? Are they interested in using healthcare services driven from big data insights and willing to share their personal data?

As the focus of this essay is the end-users' decision-making process towards the use of big data applications in the healthcare market, a TRA model – Figure 2 (in section 2) – is taken into consideration in order to analyse the steps end users go through before the actual purchase occurs (Ramayah et al., 2009). Alongside with the TRA model, the concept of perception of value is linked with the first two TRA' dimensions, attitudes and subjective norm (Gallarza & Gil Saura, 2006). In Figure 3 (in section 2) it can be observed an overall connection between the theoretical approach and the dimensions analysed in this dissertation.

Following the methodological approach, five hypotheses are developed to test the outcomes of the research question. The first one is the only hypothesis that is not a relational one, while the last four intend to test the association between two dimensions, as it can be observed in Figure 4 (in section 3).

The first hypothesis tests if end-users are aware of big data's concept and techniques. Additionally, the second hypothesis intends to study if there is any association between end-users' awareness and their perception of value. Finally, the last three hypotheses relate end-users' perception of value with willingness to share their personal data, willingness to buy devices that require the use of big data techniques, and if its associated with a positive word of mouth.

Qualtrics, an online platform, is the form chosen for conducting the survey, which is spread through social media and email. The use of a snowball diffusion method allows a less costly and rapid process to gather enough respondents to proceed to the data analysis and the study of the research question (Szolnoki & Hoffmann, 2013).

After the data collection, the sample size is of 331 respondents, in which 201 are male and 130 are female. Figure 6 (in section 5) shows the sample discriminated by age and gender.

In order to test if there are an association between the relational hypotheses, a *Chi-Square* test analyses if the variables in question are independent from each other (Zar, 2014).

Finally, in order to test the strength of the associations resulting from the *Chi-square* tests *OLS Regressions* are computed.

The next chapter, Chapter IV – *Data Analysis* – presents the results obtained for each hypothesis, and the main findings associated with the study.

CHAPTER IV. DATA ANALYSIS

4.1. Hypotheses analysis

The four relational hypotheses tested are validated by the *Chi-square* test. The ideal scenario is to have all the values from the likert scale (from one to seven) for both dimensions. However, if doing so, one of the last two contingency table analysis' requirements are violated. Thus, in order to satisfy all the test requirements, this dissertation analyses the four last hypotheses by grouping the dimensions in smaller groups of possible values: positive or negative inclination towards a subject.

In order to ensure a conservative approach, a negative inclination is marked for every opinion scoring below or equal to four in the likert scale point, while a positive inclination stands for the remaining values, being them 5 – Somewhat agree; 6 – Agree; 7 – Strongly agree, as it can be observed in Figure 8.

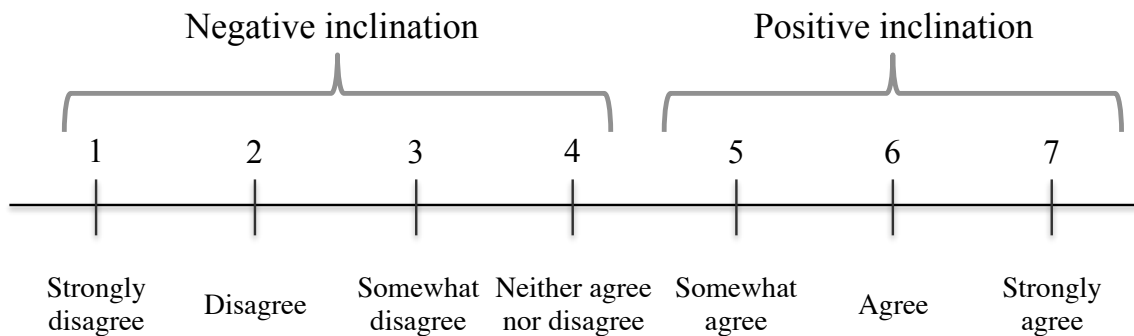


Figure 8 - Negative or positive inclination in likert scale

By including the neutral value with the explanatory subtitle of “Neither agree nor disagree” in the negative inclination side, the conservative approach is ensured. Furthermore, the requirement that all the sub-questions within one question have to be higher than four in order for the dimension to be perceived as positive empowers the conservative approach.

Taking the example of the awareness dimension analysed in Q1: the respondents that are considered to be aware of big data concept and techniques are those who score higher than 4 in all the four sub-questions of Q1.

Moreover, in some cases, three possible outputs of a dimension are analysed by adding a

neutral output of value equal to a score of four in the likert scale (in this case, the negative inclination includes only the scores below or equal to three in the likert scale), as there is a need for understanding if the acceptance of the null hypothesis occur due to the conservative approach taken in previous analysis.

4.1.1 – Hypothesis 1

The first hypothesis of this dissertation is the only one that is non-relational, as former literature does not have a clear definition neither for awareness figures, nor for precedents of the awareness stage. Therefore, a descriptive observation is presented based on the survey conducted.

As it can be observed in Figure 9, Q1 characterizes the sample population in terms of awareness about big data concept and its techniques. In order to do so, this dissertation assumes a conservative approach for every group characterization, meaning that only the respondents who score higher than four in every sub-question related to the subject in analysis are considered to have a positive inclination towards the subject. Regarding Q1, all the respondents who answer higher than four in all four sub-questions are considered to be aware of big data concept and techniques.

Having the above into consideration, Figure 9 shows that 189 of the total respondents, which represent 57,1% of the total sample, are aware of big data concept and its techniques.

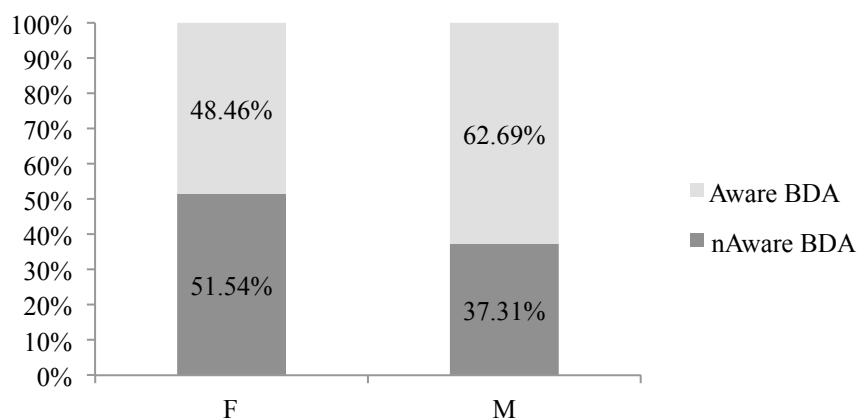


Figure 9 - Awareness according to Gender

If divided by gender, 48,5% of female respondents are aware of big data concept and techniques against 62,7% of male respondents. Moreover, perceptually, regarding the age intervals, Figure 10 shows that in all age intervals, except the first one that only represents one respondent, more than 50% of the sample population is aware of big data concept and techniques.

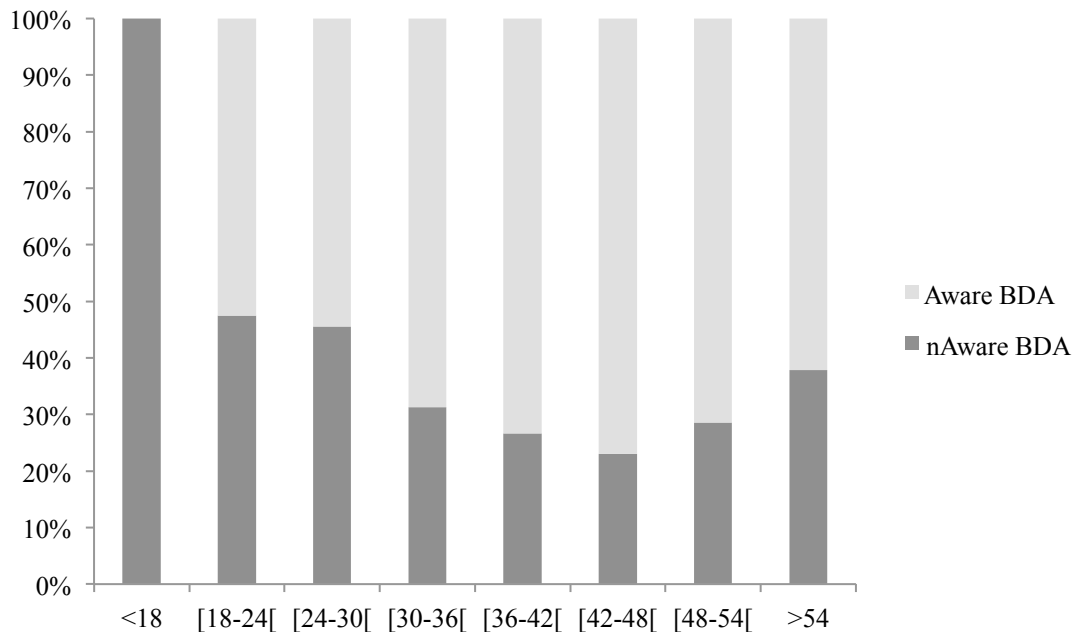


Figure 10 - Awareness according to age - %

Note: the first age interval has only one respondent, and therefore should be ignore for conclusions.

Thus, an overall view of the descriptive status of the first hypothesis shows that more than half of the respondents are aware of big data concept and techniques, situation that occurs in all age categories. However, gender discrimination shows a male awareness tendency, as 62,7% of male respondents are aware of big data while female percentage lowers to 48,5%.

4.1.2 – Hypothesis 2

The second hypothesis is the first relational one of this dissertation. It tests if end users' awareness of big data techniques influences patients' perceived value of healthcare products/services that require big data technologies.

As stated in section 3.7, to test association between two dimensions, a *Chi-square* test is

conducted. After dividing Q1 into the two inclinations the two awareness's polar outcomes are achieved. The first one is "Aware BDA", which represents the respondents that are aware of big data and the second is "nAware BDA" and represents those who are not.

Additionally, Q5 divides the respondents in those who have a positive perceived value towards devices that requires the use of big data technologies – PV – and the ones who have a negative inclination towards the perception of value dimension – nPV.

Table 3 - H₂'s contingency table - Q1&Q5

H2: Aware BDA-PV - Chi Square test

	<i>PV</i>	<i>nPV</i>	Σ
<i>Aware BDA</i>	140	49	189
<i>nAware BDA</i>	82	60	142
Σ	222	109	331
<hr/>			
Chi-square test result	p-value		
9,79	1,758E-03		

Table 3 shows the contingency table of the dimensions awareness and perception of value, where the combined observed frequencies of *Aware BDA*, *nAware BDA*, *PV*, and *nPV* are shown.

The *Chi-square* test applied to these figures results in a value of 9,79, which guarantees a p-value of 0,0018, rejecting the null hypothesis. This rejection allows the conclusion that that there is an association between awareness of big data' concept and techniques and end-users' higher perception of value of devices that use big data techniques in the healthcare market, and, therefore, the second hypothesis is validated.

4.1.3 – Hypothesis 3

Hypothesis three tests a relationship between perception of value and end-users' willingness to share data. Three contingency tables tests are made to test the main barrier of big data technologies spread. Firstly, Q6 asks the respondents a general question about their feelings towards sharing personal data. Secondly, Q8 divides the prior general question into sub categories of data type (Q8.1 – Location data; Q8.2 – Financial; etc.). Finally, a new dimension is created that describes respondents with a positive inclination

towards sharing data as the ones who score higher than four in all Q8's sub-questions.

Table 4 – H₃'s contingency table – Q5&Q6

H3: PV-WTS_g - Chi Square test

	WTS_g	nWTS_g	Σ
PV	121	101	222
nPV	58	51	109
Σ	179	152	331
Chi-square test result			
	p-value		
	0,05		8,244E-01

Table 4 shows the contingency table of the dimensions *PV*, *nPV*, positive inclination towards sharing data (general question) – *WTS_g* – and negative inclination towards sharing data (general question) – *nWTS_g*.

Computed the *Chi-square* test with the data showed in Table 4, a test result of 0,05 is obtained, which ensures a p-value of 0,8244. Such result accepts the null hypothesis and therefore there is no association between perception of value and willingness to share data if asked in a general manner.

As there is no association between end-users' perception of value and willingness to share data, when asked with a general question, the survey' Q8 asks the respondents the same question but divided in individual seven data categories: location; financial; personal (weight, height, age, gender); general (email, address, phone number); medical; walking and sports habits; and daily data.

Table 5 shows that, when testing the association between the perception of value dimension and the sub categories of data for end-users sharing data, all the hypothesis are validated by rejecting the null hypothesis.

Table 5 - Chi-square test results (association between perceived value and individual data sharing categories)

<i>Data category & Q8 associated</i>	<i>Chi-square test result</i>	<i>P-value</i>	<i>Hypothetehsis outcome</i>
<i>Q8.1 - Location Data</i>	<i>12,322</i>	<i>4,477E-04</i>	<i>Reject null hypothesis</i>
<i>Q8.2 - Financial Data</i>	<i>8,117</i>	<i>4,385E-03</i>	<i>Reject null hypothesis</i>
<i>Q8.3 - Personal Data</i>	<i>23,98</i>	<i>9,736E-07</i>	<i>Reject null hypothesis</i>
<i>Q8.4 - General Data</i>	<i>4,108</i>	<i>4,268E-02</i>	<i>Reject null hypothesis</i>
<i>Q8.5 - Medical Data</i>	<i>11,233</i>	<i>8,033E-04</i>	<i>Reject null hypothesis</i>
<i>Q8.6 - Walking and sports Data</i>	<i>14,264</i>	<i>1,589E-04</i>	<i>Reject null hypothesis</i>
<i>Q8.7 - Personal daily Data</i>	<i>4,545</i>	<i>3,302E-02</i>	<i>Reject null hypothesis</i>

As the results from obtained from the previous two analyses are contradictory, a third test is conducted by creating a new dimension. For an easy comparison, the new dimension gathers all the Q8's sub-questions. To ensure a conservative approach, it is considered as positive inclination all the respondents who score higher than four in all seven Q8's sub-question. On the other hand, negative inclination comprises all the remaining responses.

Table 6 shows the contingency table's values. It is important to notice that only 40 respondents (12,1% of the total) scored all Q8's dimensions higher than four. This happens as only 15% of total respondents have a positive inclination towards sharing financial data.

Table 6 - H₃'s contingency table - Q5&ΣQ8

H₃: PV-WTS_{av} - Chi Square test

	<i>WTS_{av}</i>	<i>nWTS_{av}</i>	<i>Σ</i>
<i>PV</i>	<i>34</i>	<i>188</i>	<i>222</i>
<i>nPV</i>	<i>6</i>	<i>103</i>	<i>109</i>
<i>Σ</i>	<i>40</i>	<i>291</i>	<i>331</i>

<i>Chi-square test result</i>	<i>p-value</i>
<i>6,62</i>	<i>1,007E-02</i>

Nonetheless, it is being tested if a positive perception of value is associated with a higher willingness to share data, which can be positively concluded with the contingency table obtained.

Despite the discrepancy between respondents willing to share and not willing to share, the *Chi-Square* test can be computed as all its requirements are fulfilled. The values observed in the contingency table of Table 6 result in a statistic test of 6,62 which represents a p-value of 0,0101. By rejecting the null hypothesis it can be concluded that there is a positive association between end-users' perception of value of big data applications in healthcare and their willingness to share data for care purposes.

4.1.4 – Hypothesis 4

The fourth hypothesis tests if a higher perception of value of the use of big data technologies in the healthcare industry increases end-users' willingness to buy such applications.

As in prior analyses, a *Chi-Square* test is conducted. The first dimension is already explained in section 4.1.1, while the second dimension analysed, willingness to buy, is examined through Q11 and it is divided into two tendencies: a positive willingness to buy – *WTB_w* – and a negative tendency towards the purchase of such applications – *nWTB_w*.

Table 7 - H₄'s contingency table - Q5&Q10

H4: PV-WTB_w - Chi Square test

	<i>WTB_w</i>	<i>nWTB_w</i>	Σ
<i>PV</i>	138	84	222
<i>nPV</i>	44	65	109
Σ	182	149	331

<i>Chi-square test result</i>	<i>p-value</i>
14,03	1,799E-04

In Table 7 it can be observed the contingency table of the dimensions awareness and perception of value.

The *Chi-square* test result of 14,03 guarantees a p-value of 0,0002, rejecting the null hypothesis. This rejection is highly significant and allows to conclude that there is a positive association between perception of value of big data' techniques in the healthcare system and willingness to buy applications that use big data technologies.

This hypothesis also validates the possibility for organizations to capture value directly for end-users by selling applications that use big data techniques for health purposes.

4.1.5 – Hypothesis 5

The fifth hypothesis tests if a higher perception of value of the use of big data technologies in the healthcare industry is associated with a positive word of mouth (*WoM*) from end-users.

This hypothesis is tested according to two different *WoM*' subjects. The first is related to the use of devices that require big data techniques in the healthcare market, while the latter refers to an incentive for sharing data. On one hand, the first test is related with organizations' ability to capture value, as according to (Lin et al., 2005) a positive *WoM* is reflected in an increase of the perceived value from the ones surrounding the individual and consequently their willingness to buy. On the other hand, the second test is related to end-users' privacy concerns, which is, according to current literature, the main barrier big data applications are required to overcome in order to achieve a B2C approach (Manyika, 2011; Murphy & Barton, 2014).

The first *WoM* subject – big data devices in healthcare – is divided into two inclinations, being them positive word of mouth – *WoM_d* – and a negative inclination towards recommending such devices – *nWoM_d*.

Table 8 - H₅'s contingency table - Q5&Q9

H5: PV-WoM_d - Chi Square test

	<i>WoM_d</i>	<i>nWoM_d</i>	Σ
<i>PV</i>	147	75	222
<i>nPV</i>	33	76	109
Σ	180	151	331
Chi-square test result		p-value	
38,07		6,839E-10	

Table 8 shows the combined absolute frequencies of *PV*, *nPV*, *WOM_d*, *nWOM_d*. With a *Chi-square* test result of 38,07 the p-value is zero, meaning that the null hypothesis is rejected and therefore it exists a positive association between perception of value and a positive *WoM* towards using devices that require big data techniques in the healthcare

market.

The second *WoM* subject – sharing data – is also divided into two inclinations, being them positive word of mouth – *WoM_sd* – and a negative inclination towards recommending such devices – *nWoM_sd*.

Table 9 - H₅'s contingency table - Q5&Q10

H5: PV-WoM_sd - Chi Square test

	<i>WoM_sd</i>	<i>nWoM_sd</i>	Σ
<i>PV</i>	119	103	222
<i>nPV</i>	36	73	109
Σ	155	176	331

<i>Chi-square test result</i>	<i>p-value</i>
12,43	4,224E-04

Above, Table 9 shows the combined absolute frequencies of *PV*, *nPV*, *WoM_sd*, *nWoM_sd*. The *Chi-square* test generates a result of 12,43, and consequently a p-value of 0,00042. Such p-value ensures the rejection of the null hypothesis, meaning that it exists a positive association between perception of value and a positive *WoM* for end-users to share their personal data.

4.1.6 – Hypotheses summary

This section enables an overall vision of the hypotheses outputs and results trough Table 10

Table 10 - Hypotheses Chi-square test's outputs

<i>Hypothesis</i>	<i>Dimensions related</i>	<i>Chi-square test result</i>	<i>P-value</i>	<i>Hyphotehsis outcome</i>
<i>H2</i>	<i>Aware BDA - PV</i>	9,787	1,758E-03	<i>Reject null hypothesis</i>
<i>H3</i>	<i>PV - WTS_g</i>	0,049	8,244E-01	<i>Accept null hypothesis</i>
	<i>PV - WTS_av</i>	6,623	1,007E-02	<i>Reject null hypothesis</i>
<i>H4</i>	<i>PV - WTB_w</i>	14,030	1,799E-04	<i>Reject null hypothesis</i>
<i>H5</i>	<i>PV - WoM_d</i>	38,066	6,839E-10	<i>Reject null hypothesis</i>
	<i>PV - WoM_sd</i>	12,430	4,224E-04	<i>Reject null hypothesis</i>

Table 10 shows that all hypotheses are validated as for this dissertation purposes H3 is tested with the dimensions of *PV-WTS_av*, which ensures that the dimensions analysed are not independent from each other, and their association effect occurs in positively manner (e.g. positive awareness of big data concept and techniques is associated with a higher end-users' perception of value of big data applications in the healthcare market).

4.2. Linear model

This section starts by presenting a brief summary of the associations presented in section 4.1 but computed through a linear basic regression. Afterwards three multi variable linear regressions are presented as an explanation of the research question.

In order to a better comprehension of the OLS models presented along this section Table 11 shows a brief summary of the variables used.

Table 11 - OLS dimensions explanation

Variable symbol	Variable explanation	Variable computation	Data type
<i>Aware_BDA</i>	End-users are aware of big data concept and techniques	<i>Q1's average</i>	<i>Interval scale</i>
<i>PV</i>	End-users' perceived value from big data applications in healthcare	<i>Q5's average</i>	<i>Interval scale</i>
<i>WTS_av</i>	End-users' willingness to share their personal data for healthcare purposes	<i>Q8's average</i>	<i>Interval scale</i>
<i>WTB_w</i>	End-users' willingness to buy healthcare applications that require big data techniques	<i>Q11</i>	<i>Aproximated to an interval scale</i>
<i>WoM_d</i>	Positive word of mouth for big data applications in the market of healthcare	<i>Q9's average</i>	<i>Interval scale</i>
<i>WoM_sd</i>	Positive word of mouth for sharing personal data for healthcare purposes	<i>Q10's average</i>	<i>Interval scale</i>

Table 11 presents an overall summary of the outputs driven from the linear regressions according to this dissertation's hypotheses.

Table 12 - OLS simple linear regressions' outputs

<i>Hypothesis</i>	<i>Dimensions related</i>	<i>OLS Regression</i>	<i>β p-value</i>	<i>R Square</i>	<i>F - statistic</i>
<i>H2</i>	<i>Aware BDA - PV</i>	<i>4,274+0,203*Aware BDA</i>	<i>1,714E-05</i>	<i>5,091%</i>	<i>17,648</i>
<i>H3</i>	<i>PV - WTS_g</i>	<i>2,283+0,448*PV</i>	<i>4,388E-09</i>	<i>9,584%</i>	<i>34,872</i>
	<i>PV - WTS_av</i>	<i>1,058+0,609*PV</i>	<i>1,138E-18</i>	<i>20,759%</i>	<i>86,188</i>
<i>H4</i>	<i>PV - WTB_w</i>	<i>0,049+0,784*PV</i>	<i>4,171E-26</i>	<i>28,524%</i>	<i>131,296</i>
<i>H5</i>	<i>PV - WoM_d</i>	<i>1,258+0,698*PV</i>	<i>2,612E-41</i>	<i>42,158%</i>	<i>239,789</i>
	<i>PV - WoM_sd</i>	<i>1,443+0,57*PV</i>	<i>1,857E-17</i>	<i>19,418%</i>	<i>79,280</i>

Despite the R Square's lower values of the two first regressions, it can be seen that the p-values of the t statistic on the β s are always lower than 0,05 and therefore statistically significant.

Table 12's first regression measures the association between *Aware BDA* and *PV*. It can be observed that for every unitary increment from the *Aware BDA* dimension there is an increase of 0,203 in the end-users' perception of value.

Additionally, from the three dimensions driven from *PV*, *WoM_d* is the dimension that is more explained with a linear model, with a R Square of 42,158%. Furthermore the dimension that presents a higher increment – 0,784 – for every unitary increase on the *PV*'s dimension is *WTB_w*.

Nonetheless, simple linear regressions tend to be insufficient to evaluate a dimension. In this regard, and according to (Ramayah et al., 2009) the last dimension of TRA's model is behaviour. This dissertation cannot test an actual behaviour, however an intention is often correlated with the happening of the behaviour (Guo et al., 2007).

Therefore this dissertation's first model - Equation 2 - explains end-users willingness to buy big data applications in the market of healthcare.

The theoretical model is explained by the dimensions of *PV*, *WTS_av*, *WoM_d*, *WoM_sd* and it can be observed below:

$$WTB_w = \alpha + \beta_{PV} * PV + \beta_{WTS_{av}} * WTS_{av} + \beta_{WoM_d} * WoM_d + \beta_{WoM_{sd}} * WoM_{sd} + e$$

Equation 2 - WTB_w model

The dimensions presented in the model have two lines of thinking. Firstly, and as it is presented in section 4.1.4, a higher PV is associated with a higher WTB_w . Additionally, according to (Lin et al., 2005) a positive WoM_d increases not only the end-users' perception of value but is also associated with an increment in customer's willingness to buy. Secondly, (Kalapesi, 2013; Murphy & Barton, 2014) affirm that big data's main barrier are end-users' mistrusts to share their personal data. Therefore, it is reasonable to think that the higher the WTS_{av} , the higher is the WTB_w .

Table 13 - WTB_w model's output

OLS - WTB_w Regression				
	β	<i>Sdt. Error</i>	<i>t</i>	<i>p-value</i>
(Constant)	-1,140	0,303	—	—
PV	0,323	0,068	4,716	3,571E-06
WTS_{av}	0,458	0,084	5,450	9,971E-08
WoM_d	0,140	0,058	2,414	1,634E-02
WoM_{sd}	0,195	0,071	2,756	6,180E-03
<hr/>				
<i>F-test</i>	<i>R Square</i>	<i>Adjusted R Square</i>	<i>d.f</i>	<i>ESS</i>
109,459	57,321%	56,797%	326	458,369

Table 13 - WTB_w model's output shows the results of WTB_w multi linear regression. First it can be observed that all the β s in the regression are statistically significant as the t -student individual significant test outputs present p-values lower than 0,05. In fact, the β s of the dimensions PV and WTS_s are highly significant while the remaining β s are very significant. Moreover, the F-test, which analyses the level of global significance of the model, has an output of 109,459 for 326 degrees of freedom. This value highly surpasses the critical value of the F distribution and thus the model is globally significant. Having an output of 57,321% value for R Square, shows that the model explains more than 57% of the dimension WTB_w .

A curiosity that emerges from the regression achieved is that the dimensions of PV ($\beta = 0,303$) and WoM_d ($\beta = 0,140$) have in fact less impact than questions related with the

share of data ($\beta_{WTS_{av}} = 0,458$ and $\beta_{WoM_{sd}} = 0,195$). Such results only empower the necessity for healthcare organizations to deal with end-users privacy concerns in order to be able to capture value from a wider range of products in the market of healthcare.

The second model - Equation 3 - of this dissertation tests WoM_d as a result of PV , WTB_w and WTS_{av} , as it can be observed in the model formula below:

$$WoM_d = \alpha + \beta_{PV} * PV + \beta_{WTB_w} * WTB_w + \beta_{WTS_{av}} * WTS_{av} + e$$

Equation 3 - WoM_d model

It makes sense to analyze these dimensions as positive word of mouth is linked to a higher perception of value and willingness to buy (Lin et al., 2005). Therefore, this model tries to understand in what extent this situation occurs within the healthcare market.

Furthermore, for a respondent to have a higher score of value in questions related with positive word of mouth, it is expected that he/she has a higher perception of value of big data applications in the market of healthcare. Additionally, being share personal data such a delicate issue for end-users, a higher predisposition to share personal data is expected to be associated with a higher positive word of mouth.

Table 14 shows the output of the model tested.

Table 14 - WoM_d model's output

OLS - WoM_d Regression				
	β	<i>Sdt. Error</i>	<i>t</i>	<i>p-value</i>
<i>(Constant)</i>	1,066	0,203	–	–
<i>PV</i>	0,167	0,034	4,861	1,819E-06
<i>WTB_w</i>	0,291	0,033	8,794	8,411E-17
<i>WTS_av</i>	0,369	0,044	8,376	1,633E-15
<hr/>				
<i>F-test</i>	<i>R Square</i>	<i>Adjusted R Square</i>	<i>d.f</i>	<i>ESS</i>
181,924	62,533%	62,189%	327	268,713

Firstly it can be seen that all the β s in the regression are statistically significant as the *t*-student individual significant test outputs are a *p*-value lower than 0,05. As a matter of fact all the dimensions are considered highly significant due to the lower *p*-values.

Furthermore, the F-test has an output of 181,924 for 327 degrees of freedom, which ensures that the model is globally significant.

The R Square of 62,533% shows that the variations of the *WoM_d* dimension are 62,5% explained by the current model.

Again, as in the previous model, a curiosity that emerges from the regression achieved is that the dimensions of *PV* ($\beta = 0,167$) and *WTB_w* ($\beta = 0,291$) have in fact less impact than the *WTS_w* ($\beta = 0,369$).

With these results the third model - Equation 4 - is conducted, which tries to understand the *WoM_sd* dimensions.

$$WoM_{sd} = \alpha + \beta_{PV} * PV + \beta_{WTB_w} * WTB_w + \beta_{WTS_{av}} * WTS_{av} + e$$

Equation 4 - *WoM_sd* model

It is important to state that a multi linear regression model considering the *WTS_av* dimension as a dependent variable is not conducted. This results from the fact that, according to this dissertation's theoretical model, *WTS_av* only depends on one variable, the perceived value.

Table 15 - *WoM_sd* model's output shows the *WoM_sd* regression outputs.

Table 15 - *WoM_sd* model's output

OLS - *WoM_sd* Regression

	β	<i>Sdt. Error</i>	<i>t</i>	<i>p-value</i>
(Constant)	0,946	0,251	–	–
<i>PV</i>	0,455	0,043	10,627	7,452E-23
<i>WTB_w</i>	0,340	0,041	8,298	2,814E-15
<i>WTS_av</i>	0,027	0,055	0,485	6,277E-01
<i>F-test</i>	<i>R Square</i>	<i>Adjusted R Square</i>	<i>d.f</i>	<i>ESS</i>
164,123	60,091%	59,725%	327	373,460

Having the global results in the same line as previous models, what is surprisingly is the fact that the *WTS_av* dimension is not statistically significant. Despite being the dimension that, apparently, has the most common ground (sharing data) it is the one not

to be significant for the model in question.

This situation only empowers the need for healthcare organizations understand end-users' privacy concerns.

4.3. Chapter summary

Table 16 shows a relational hypotheses' summary, for both the *Chi-square* and the *OLS Regression* computation.

Table 16 - Overall hypotheses summary

<i>Hypothesis</i>	<i>Dimensions related</i>	<i>Chi-square test</i>	<i>p-value</i>	<i>OLS Regression</i>	<i>p-value</i>
<i>H2</i>	<i>Aware BDA - PV</i>	9,787	1,758E-03	4,274+0,203* <i>Aware BDA</i>	1,714E-05
<i>H3</i>	<i>PV - WTS_g</i>	0,049	8,244E-01	2,283+0,448* <i>PV</i>	4,388E-09
	<i>PV - WTS_av</i>	6,623	1,007E-02	1,058+0,609* <i>PV</i>	1,138E-18
<i>H4</i>	<i>PV - WTB_w</i>	14,030	1,799E-04	0,049+0,784* <i>PV</i>	4,171E-26
<i>H5</i>	<i>PV - WoM_d</i>	38,066	6,839E-10	1,258+0,698* <i>PV</i>	2,612E-41
	<i>PV - WoM_sd</i>	12,430	4,224E-04	1,443+0,57* <i>PV</i>	1,857E-17

According to the p-values achieved it can be seen that all the hypotheses tested with the *Chi-square* analysis are validated, except for the third hypothesis *PV – WTS_g*. This is explained due to the fact that respondents do not feel comfortable sharing their personal data when asked in a general question. However, the same test realized with a conservative approach for the dimension *WTS_av* enables the conclusion that end-users' willingness to share personal data for healthcare purposes is in fact associated with a higher perception of value.

When computed with OLS Regression all hypotheses are validated with a high level of significance.

The second part of chapter IV is the creation of three models that explain the dimensions of *WTB_w*, *WoM_d*, and *WoM_sd*. In order to compare the three models, the Adjusted R Square is used. Although the value is similar, the second model - EQ2 - is the one with the highest value obtained (62,189%).

Figure 11 - OLS models scheme shows that the three models created for the purpose of explaining the research question having as its foundation the TRA model.

$$\text{EQ2: } WTB_w = \alpha + \beta_{PV} * PV + \beta_{WTS_av} * WTS_av + \beta_{WoM_d} * WoM_d + \beta_{WoMsd} * WoM_sd + e$$

$$\text{EQ3: } WoM_d = \alpha + \beta_{PV} * PV + \beta_{WTB_w} * WTB_w + \beta_{WTS_av} * WTS_av + e$$

$$\text{EQ4: } WoM_sd = \alpha + \beta_{PV} * PV + \beta_{WTB_w} * WTB_w + \beta_{WTS_av} * WTS_av + e$$

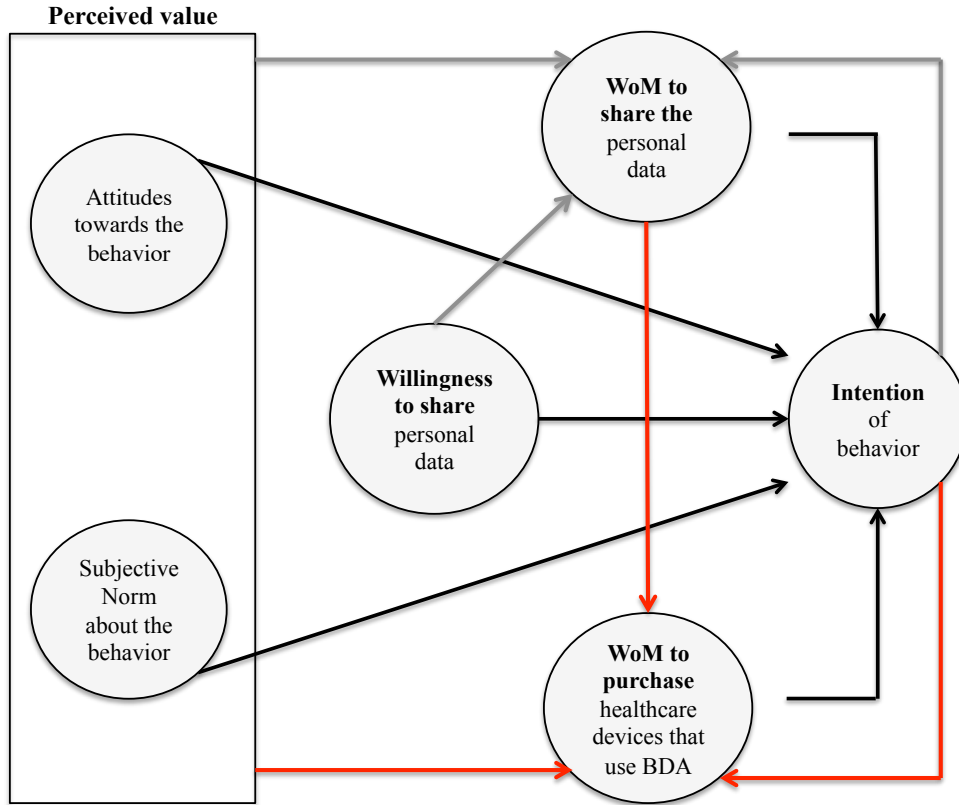


Figure 11 - OLS models scheme

It can be observed that each model (Equation 2, Equation 3, Equation 4) is represented by a different colour.

To conclude, Chapter IV proves that there is association among the dimensions tested, and that a TRA model explains end-users' decision-making process towards healthcare's products/services that require the use of big data analytics.

CHAPTER V. CONCLUSION AND FUTURE WORK

5.1. Conclusion

The use of big data techniques for developing products to be used in the healthcare market has, according to (Manyika, 2011), a potential impact of \$300 billion annually, as it is mentioned in section 2.3. Additionally, the approach to deal directly with end-users composes part of such figures.

Nonetheless the complexity of big data creates several barriers when considering selling products directly to end-users, as it is explained in section 2.2.2. First of all, it is required to understand if end-users are aware and perceive a positive value from healthcare products that require the use of big data techniques (Russom, 2011). Secondly, the perception of value *per se* might not ensure an intention to purchase (section 2.5), nor a positive word of mouth towards such products (section 3.2), as a sensible issue has to be addressed - end-users' privacy concerns for sharing personal data, as it is mentioned in section 2.5 (Murphy & Barton, 2014).

In this regard, this dissertation provides a useful contribution to literature, as it addresses the big data topic through the end-users' point of view.

The survey conducted allows several conclusions, which are developed in chapter 4. Firstly, the test results show that for a higher state of big data awareness, it is associated with a higher perception of value from products that use big data techniques in the healthcare market (section 4.1.2). This conclusion allows to outline that patients might not purchase such products due to the fact that they are not aware of such applications and not because they do not perceive value from them. In that regard, marketing and awareness campaigns are required.

Secondly, this dissertation proves that there is a positive association between a higher perception of value and a higher willingness to share personal data for healthcare purposes (section 4.1.3). Despite the fact that results vary on this question, the final conclusion this dissertation wants to transpire is that end-users tend to be more willing to share data if they perceive value from the products they deal with.

Thirdly, section 4.1.4 shows that a higher perception of value is also associated with a higher willingness to buy. This is of relevant importance, especially when combined with

the point discussed above. Not only a higher perception of value leads to higher propensity to share data, but it also increases end-users willingness to buy. In this regard, solving problems of awareness and ensuring a higher perception of value might allow to directly solve organizations' problems when considering a B2C approach of healthcare products that require big data analytics. Not only end-users are willing to provide their data (which is essential for the product performance) but also are they ensuring organizations' value captured as a source of revenues.

Moreover, from the three models created (section 4.2), it can be concluded that the dimensions are interconnected, as end-users' willingness to purchase is positively correlated with a higher perception of value, willingness to share data, and positive word of mouth for both big data products purchase and sharing data.

In that regard the research question is positively answered, as end-users are somewhat aware of big data concept and techniques, and perceive enough value from products driven from the use of such techniques in order to surpass their privacy concerns for sharing data.

The issues and the conclusions achieved in this dissertation ensure that organizations have the possibility to capture value directly from end-users by using big data techniques in the development of products/services.

5.2. Limitations of the study

Firstly, the results achieved in this study, despite going in the same direction as the theoretical model of support tend to create bias responses. In fact, according to (Szolnoki & Hoffmann, 2013) the snowball online survey method, creates a higher probability of gathering a biased sample. Additionally, this dissertation's context did not have the means and times required by a larger study.

Furthermore, a dropout rate of almost 50% creates strong limitations to the increase of the sample size, which is desirable for the final number of useful respondents. Despite not being the ideal method to spread the survey, Internet empowers the number of respondents.

Also, a statement has to be made about the choice of target of respondents. Ideally, and in an environment without constraints, the sample has to be thought as the representation of

the target population, which in this case is the healthcare market. The fact that the respondents obtained are restricted to the dissertation researcher's circle of life shortens the sample diversification.

Secondly the dimensions analysed by the theoretical model (TRA model) can be stratified into more complex models, as according to (Ramayah et al., 2009) this model can become more complex and with antecedent dimensions to the attitudes and subjective norm dimensions. But again, the time constraints call for the creation of a simpler model.

Thirdly, as stated above, this dissertation does not test the last dimension of the TRA model (behaviour), and supposes that intention is highly correlated with the behaviour consummation (Guo et al., 2007). However, it would be interesting to understand this last relationship between these two dimensions.

Socio-demographic indicators, especially income and social status, should be taken into consideration as it might influence the purchase of products that are seen today as premium. This dissertation opted not to ask such questions, since they are often seen as intrusive. Eventually, more dimensions could be added to the model, such as current life conditions, or previous diseases.

Finally, it has to be pointed out the fact that the survey conducted is only available in one language (English). The lack of a multi-lingual approach creates a barrier to the population targeting especially due to the fact that most respondents are believed to be Portuguese.

5.3. Future research

Big data is today a field of research and also of business opportunities that constantly foster new lines of thinking.

One topic slightly neglected in this dissertation is related to the privacy concerns and therefore future research should address this topic. Firstly, being the more critical dimension of big data (Murphy & Barton, 2014), it would be interesting to understand what motivates end-users mistrusts towards sharing their personal data. Moreover, it would also be important to study in which extent do end-users are aware of the amount of data they share today (e.g. Facebook information or mobile phone constant tracking).

A point should also be made for organizations' point of view and their ability to capture

value. As this dissertation could not test the actual customer's purchase, it would be interesting to understand the figures health providers and big data product sellers for health purposes can achieve by selling directly to end-users.

Related to the point stated above, future studies could address if the use of big data techniques in the healthcare market could become a competitive advantage for organizations.

Finally, it could be interesting to spread the test's scope to other industries. For instance the manufacturing industry, which according to (Manyika, 2011) is the industry that presents more potentialities with the emergence of big data techniques and applications.

APPENDIX

APPENDIX 1: Questionnaire

9/11/2014

Qualtrics Survey Software

Default Question Block

Dear participant,
I am a student from Católica Lisbon School of Business & Economics and I am currently working on my master thesis about Big Data. Your input is essential to my research. It only takes about 5 minutes and I would like to thank you in advance for your contribution. All data collected is anonymous and the conclusions presented in my master thesis, can be shared with you if interested.

Thank you!

From a scale of 1 to 7, how much do you agree with the sentences below?

	Strongly Disagree 1	Disagree 2	Somewhat Disagree 3	Neither Agree nor Disagree 4	Somewhat Agree 5	Agree 6	Strongly Agree 7
I am aware of big data concept.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I believe big data analytics can leverage services' benefits.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I believe high customisation is valuable for customers.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I believe real-time applications can increase services' value.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Big data concept represents the characteristics of today's data with the rise and spread of Internet, smart-phones, social-networks, sensors and all the data stored from different sources. In the recent years, big data analytics have been developed as a response to data's changes. Throughout the use of such techniques, a highly wide variety of applications started to emerge. However, these techniques require data from a wide variety of sources, being one of them the customer's personal data.

According to the different dimensions, from a scale of 1 to 7, how much would you be willing to share your personal data?

	Not At All 1	Probably Not 2	Maybe 3	Neutral 4	Probably 5	Most Probably 6	Very Much So 7
Location data	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Car usage data	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Financial data	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Personal data (weight, height, age, gender)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
General data (Address, email, contact)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Medical history	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Walking and sports habits data	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Healthcare industry - The following questions are related to the healthcare system

According to the different dimensions, from a scale of 1 to 7, how would you rate your level of satisfaction in the healthcare system?

	Completely Dissatisfied 1	Mostly Dissatisfied 2	Somewhat Dissatisfied 3	Neutral 4	Somewhat Satisfied 5	Mostly Satisfied 6	Completely Satisfied 7
Time waiting	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Prevention measures	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Constant monitoring	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Cost	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Cost/benefit feeling	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

From a scale of 1 to 7, how much do you agree with the sentences below?

	Strongly Disagree 1	Disagree 2	Somewhat Disagree 3	Neither Agree nor Disagree 4	Somewhat Agree 5	Agree 6	Strongly Agree 7
Information is important to increase health provider services' quality	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Information related to health is only useful for doctors' decisions	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Information driven from big data analytics increase health services'	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

<https://acplbusiness.co1.qualtrics.com/WRQualtricsControlPanel/Ajax.php?action=GetSurveyPrintPreview&T=25yrIe>

1/4

<p>quality</p> <p>Using health providers that use big data analytics is something I would enjoy use</p> <p>I would say big data analytics' functionalities can leverage health services</p>							
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<p>Big data on healthcare</p> <p>Now, imagine a bracelet (or any other wearable device) was available in the market that would measure your weight, heart beat, walking habits, quality of sleep and other useful metrics. Such bracelet, would alert you from getting heart attacks, better control over diet and consuming habits, among other applications. It would also provide doctors a more complete medical history, with integrated global information and not just from analyses or visible symptoms. This would enable better and more informed decisions from the health providers, maximising the treatments' accuracy.</p> <p>When considering big data applications on healthcare, from the business point of view, such technologies could enable a higher efficiency, reducing costs, which can be channelled towards a more customer-oriented service. By integrating medical histories from different sources, a more precise medical record is achievable, resulting in more accurate medical decisions. Moreover, it increases back-office efficiency, reduces waiting line time creates a better patient profiling which results in gains for both entities (healthcare service providers and patients).</p> <p>From a scale of 1 to 7, how much do you agree with the sentences below?</p>							
	Strongly Disagree 1	Disagree 2	Somewhat Disagree 3	Neither Agree nor Disagree 4	Somewhat Agree 5	Agree 6	Strongly Agree 7
Using a similar application as the bracelet described above would enhance my effectiveness in controlling my health	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I would find a similar application, as the bracelet described above, useful in controlling my health	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Using a similar application as the bracelet described above would improve my control on my health	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<p>From a scale of 1 to 7, how much are you willing to provide your own personal data for a healthcare purpose if a similar application as the bracelet described above existed?</p>							
Not At All 1	Probably Not 2	Maybe 3	Neutral 4	Probably 5	Most Probably 6	Very Much So 7	
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<p>From the list below, according to the different dimensions, which are the causes for not being available to share your personal data with the healthcare system?</p>							
	Strongly Disagree 1	Disagree 2	Somewhat Disagree 3	Neither Agree nor Disagree 4	Somewhat Agree 5	Agree 6	Strongly Agree 7
I have privacy concerns.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I do not think it is useful.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I do not want to be bothered with suggestions and contacts.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I do not think there are real benefits from such applications	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<p>From a scale of 1 to 7, how much would you be willing to share your personal data according to the different dimensions, if such application existed?</p>							
	Not At All 1	Probably Not 2	Maybe 3	Neutral 4	Probably 5	Most Probably 6	Very Much So 7
Location data	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Financial data	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Personal data (weight, height, age, gender)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
General (Address, email, contact)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Medical history	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Walking and sports habits data	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Personal daily medical record	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

From a scale of 1 to 7, how much do you agree with the sentences below?

	Strongly Disagree 1	Disagree 2	Somewhat Disagree 3	Neither Agree nor Disagree 4	Somewhat Agree 5	Agree 6	Strongly Agree 7
I would say positive things about the bracelet described above to other people	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I would recommend the bracelet described above to someone who wishes to improve his/her health and life style	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I would encourage friends and relatives to purchase a similar application as the bracelet described above	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I would encourage friends and relatives to use a similar application as the bracelet described above	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

From a scale of 1 to 7, how much do you agree with the sentences below?

	Strongly Disagree 1	Disagree 2	Somewhat Disagree 3	Neither Agree nor Disagree 4	Somewhat Agree 5	Agree 6	Strongly Agree 7
I would say positive things about sharing my personal data to such application	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I would recommend to those who wish to improve their health and life style to share their personal data to such application	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

From a scale of 1 to 7 how much would intend to buy a similar application?

Not At All 1	Probably Not 2	Maybe 3	Neutral 4	Probably 5	Most Probably 6	Very Much So 7
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Considering such application existed, how much would you be willing to pay for it (in €)?

0	[0-30[[30-60[[60-90[[90-120[[120-150[More than 150
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Gender

Male

Female

Age

<18	[18-24[[24-30[[30-36[[36-42[[42-48[[48-54[>54
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Do you own a smartphone?

Yes

No

Do you wish to receive a summary of the survey conclusions?

Yes

No

Please write your email below

Thank you very much for your participation.

REFERENCES

- Adler-Milstein, J., & Jha, A. K. (2013). Healthcare's 'Big Data' Challenge. *The American journal of Managed Care*, 19(7), 537–539.
- Ajzen, I., & Fishbein, M. (2000). Attitudes and the Attitude-Behavior Relation: Reasoned and Automatic Processes. *European Review of Social Psychology*, 11(1), 1–33. doi:10.1080/14792779943000116
- Atluri, V., Rao, S., & Varanasi, S. (2014). New trends in smart devices. *Mckinsey&Company*.
- Bhatt, G. D., Grover, V., & GROVER, V. (2005). Types of information technology capabilities and their role in competitive advantage: an empirical study. *Journal of Management Information \dots*, 22(2), 253–277.
- Biesdorf, S., & Niedermann, F. (2014). Healthcare's digital future. *Mckinsey&Company*.
- Bitterer, A. (2011). Hype Cycle for Business Intelligence, 2011, (August).
- Bowen, A. M., Williams, M., McCoy, H. V., & McCoy, C. B. (2001). Crack smokers' intention to use condoms with loved partners: intervention development using the theory of reasoned action, condom beliefs, and processes of change. *AIDS care*, 13(5), 579–594. doi:10.1080/09540120120063214
- Boyd, D., & Crawford, K. (2012). CRITICAL QUESTIONS FOR BIG DATA. *Information, Communication \& Society*, 15(5), 662–679. doi:10.1080/1369118X.2012.678878
- Chang, R. M., Kauffman, R. J., & Kwon, Y. (2013). Understanding the paradigm shift to computational social science in the presence of big data. *Decision Support Systems*. doi:10.1016/j.dss.2013.08.008
- Chaudhuri, S., Dayal, U., & Narasayya, V. (2011). An overview of business intelligence technology. *Communications of the ACM*, 54(8), 88. doi:10.1145/1978542.1978562
- Chen, Hsinchun, Roger HL Chiang, & Storey, V. C. (2012). BUSINESS INTELLIGENCE AND ANALYTICS: FROM BIG DATA TO BIG IMPACT. *MIS Quarterly*. Dec2012, 36(4), 1165–1188. 24p. 1 Diagram.
- Chen, S., & Li, J. (2010). Examining Consumers' Willingness to Buy in Chinese Online

- Market. *Journal of Computers*, 5(5). doi:10.4304/jcp.5.5.815-824
- Chiarello, G., Chase, J., Chickering, J., President, V., & Investments, F. (11AD). Big Data : Big Opportunities to Create Business Value. *Group of EMC*.
- Cooper, P. (2014). Data, information, knowledge and wisdom. *Anaesthesia \& Intensive Care Medicine*, 15(1), 44–45. doi:10.1016/j.mpaic.2013.11.009
- de Bruin, E., Loeber, J. G., Meijer, A., Castillo, G. M., Cepeda, M. L. G., Torres-Sepúlveda, M. R., et al. (2014). Evolution of an influenza pandemic in 13 countries from 5 continents monitored by protein microarray from neonatal screening bloodspots. *Journal of Clinical Virology*, 61(1), 74–80. doi:10.1016/j.jcv.2014.06.020
- Desouza, K. C. (2008). Understanding knowledge as a commons: From theory to practice.
- et al Smas, L. (2014). *Urban Planning and Big Data – Taking LUTi Models to the Next Level?*
- et all Moore, K. D. (2013). The big deal about big data. *healthcare financial mangement*.
- Frawley, W. J., Piatetsky-Shapiro, G., & Matheus, C. J. (1992). Knowledge discovery in databases: An overview. *AI magazine*, 13(3), 57–70.
- Gallarza, M. G., & Gil Saura, I. (2006). Value dimensions, perceived value, satisfaction and loyalty: an investigation of university students’ travel behaviour. *Tourism Management*, 27(3), 437–452. doi:10.1016/j.tourman.2004.12.002
- Gantz, J. F. R. D. (2007). The Expanding Digital Universe.
- Goyette, I., Ricard, L., Bergeron, J., & Marticotte, F. (2010). e-WOM Scale: word-of-mouth measurement scale for e-services context. (P. R. Messinger & D. Galletta, Eds.)*Canadian Journal of Administrative Sciences / Revue Canadienne des Sciences de l'Administration*, 27(1), 5–23. doi:10.1002/cjas.129
- Groves, P., Kayyali, B., Knott, D., & Van Kuiken, S. (2013). The “big data” revolution in healthcare. *McKinsey Quarterly*.
- Guo, Q., Johnson, C. A., Unger, J. B., Lee, L., Xie, B., Chou, C.-P., et al. (2007). Utility of the theory of reasoned action and theory of planned behavior for predicting Chinese adolescent smoking. *Addictive behaviors*, 32(5), 1066–1081. doi:10.1016/j.addbeh.2006.07.015
- Holbrook, M. B. (1999). Consumer value: a framework for analysis and research.

- Iman, R. L., & Conover, W. J. (1983). *A modern approach to statistics*. John Wiley & Sons.
- Johannessen, J.-A., & Olsen, B. O. R. (2010). The future of value creation and innovations: Aspects of a theory of value creation and innovation in a global knowledge economy. *International Journal of Information Management*, 30(6), 502–511. doi:10.1016/j.ijinfomgt.2010.03.007
- Jr, H. N. B., & Boone, D. A. (2012). Analyzing Likert Data. *Journal of Extension*, 50.
- Kalapesi, C. (2013). Unlocking the value of personal data: From collection to usage. *World Economic Forum, Tech. Rep.*, (February).
- Kayyali, B., Knott, D., & Van Kuiken, S. (2014). The big data revolution in health care: Accelerating value and innovation. *Mckinsey&Company*, 1–6.
- Kwon, O., Lee, N., & Shin, B. (2014). Data quality management, data usage experience and acquisition intention of big data analytics. *International Journal of Information Management*, 34(3), 387–394. doi:10.1016/j.ijinfomgt.2014.02.002
- Leeuw, E. D. D. (2005). To Mix or Not to Mix Data Collection Modes in Surveys, 21(2), 233–255.
- Lenard, T. M., & Rubin, P. H. (2013). The Big Data Revolution : Privacy Considerations. *Technology Policy Institute*, (December).
- Lin, C.-H., Sher, P. J., & Shih, H.-Y. (2005). Past progress and future directions in conceptualizing customer perceived value. *International Journal of Service Industry Management*, 16(4), 318–336. doi:10.1108/09564230510613988
- Manyika, J. C. M. B. B. J. D. R. R. C. B. A. H. (2011). Big data: The next frontier for innovation, competition, and productivity. *McKinsey Global Institute*.
- McAuley, & Edward. (1994). The Subjective Exercise Experiences Scale (SEES): Development and Preliminary Validation. *Journal of Sport & Exercise Psychology*, 16, 163–177.
- Mishra, D., Akman, I., & Mishra, A. (2014). Theory of Reasoned Action application for Green Information Technology acceptance. *Computers in Human Behavior*, 36, 29–40. doi:10.1016/j.chb.2014.03.030
- Moore, M., Gould, P., & Keary, B. S. (2003). Global urbanization and impact on health.

International journal of hygiene and environmental health, 206(4-5), 269–278.
doi:10.1078/1438-4639-00223

Murphy, B. M., & Barton, J. (2014). From a Sea of Data to Actionable Insights : Big Data and What It Means for Lawyers. *Intellectual Property \& Technology Law Journal*, 26, 8–18.

Nunan, D., & Di Domenico, M. (2013). Market research and the ethics of big data. *International Journal of Market Research*, 55(4), 505. doi:10.2501/IJMR-2013-015

Oh, H. (2000). The Effect of Brand Class, Brand Awareness, and Price on Customer Value and Behavioral Intentions. *Journal of Hospitality \& Tourism Research*, 24(2), 136–162. doi:10.1177/109634800002400202

Pl ger, J. (2006). Urban planning and urban life: Problems and challenges. *Planning Practice and Research*, 21(2), 201–222. doi:10.1080/02697450600944673

Priem, R. L. (2007). a Consumer Perspective on Value Creation. *Academy of Management Review*, 32(1), 219–235. doi:10.5465/AMR.2007.23464055

Quan, J. J., Hu, Q., & Hart, P. J. (2003). Information Technology Investments and Firms' Performance--A Duopoly Perspective. *Journal of Management Information Systems*, 20(3), 121–158.

Ramayah, T., Rouibah, K., Gopi, M., & Rangel, G. J. (2009). A decomposed theory of reasoned action to explain intention to use Internet stock trading among Malaysian investors. *Computers in Human Behavior*, 25(6), 1222–1230. doi:10.1016/j.chb.2009.06.007

Rindfleisch, A., Malter, A. J., Ganesan, S., & Moorman, C. (2008). Cross-Sectional Versus Longitudinal Survey Research: Concepts, Findings, and Guidelines. *Journal of Marketing Research*, 45(3), 261–279. doi:10.1509/jmkr.45.3.261

Rivard, S., Raymond, L., & Verreault, D. (2006). Resource-based view and competitive strategy: An integrated model of the contribution of information technology to firm performance. *The Journal of Strategic Information Systems*, 15(1), 29–50. doi:10.1016/j.jsis.2005.06.003

Roininen, K., Lahteenmäki, L., & Tuorila, H. (1999). Quantification of consumer attitudes to health and hedonic characteristics of foods. *Appetite*, 33(1), 71–88.

doi:10.1006/appe.1999.0232

Rose, J., & Burgin, M. (2014). Disrupting Health Care through Big Data and Predictive Analytics. *Managed Care Outlook*, 11–13.

Russom, P. (2011). BIG DATA ANALYTICS. *TDWI BEST PRACTICES REPORT*.

Schroeck, M. S. R. S. J. R.-M. D. T. P. (2012). Analytics: The real-world use of big data. How innovative enterprises extract value from uncertain data, IBM Global Business Services.

Semedo, T., Almeida Santos, M., Martins, P., Silva Lopes, M. F., Figueiredo Marques, J. J., Tenreiro, R., & Barreto Crespo, M. T. (2003). Comparative Study Using Type Strains and Clinical and Food Isolates To Examine Hemolytic Activity and Occurrence of the *cyl* Operon in Enterococci. *Journal of Clinical Microbiology*, 41(6), 2569–2576. doi:10.1128/JCM.41.6.2569-2576.2003

Shimp, T. A., & Kavas, A. (1984). The theory of reasoned action applied to coupon usage. *Journal of Consumer Research*, 11(3), 795–809.

Soto-Acosta, P., & MeroñO-Cerdan, A. L. (2008). Analyzing e-business value creation from a resource-based perspective. *International Journal of Information*

Spiess, J., T'Joens, Y., Dragnea, R., Spencer, P., & Philippart, L. (2014). Using Big Data to Improve Customer Experience and Business Performance. *Bell Labs Technical Journal*, 18(4), 3–17. doi:10.1002/bltj.21642

Sweeney, J. C., & Soutar, G. N. (2001). Consumer perceived value: The development of a multiple item scale. *Journal of Retailing*, 77(2), 203–220. doi:10.1016/S0022-4359(01)00041-0

Szolnoki, G., & Hoffmann, D. (2013). Online, face-to-face and telephone surveys—Comparing different sampling methods in wine consumer research. *Wine Economics and Policy*, 2(2), 57–66. doi:10.1016/j.wep.2013.10.001

Tsuchiya, S., Sakamoto, Y., Tsuchimoto, Y., & Lee, V. (2012). Big data processing in cloud environments. *Fujitsu Sci. Tech. Journal*, 159–168.

Ularu, E. G. P. F. C. A. A. V. M. (2012). Perspectives on Big Data and Big Data Analytics. *Database Systems Journal*, III(4), 3–14.

Westbrook, R. A., & Oliver, R. L. (1991). The Dimensionality of Consumption Emotion

Patterns and Consumer Satisfaction. *Journal of Consumer Research*, 18(1), 84.
doi:10.1086/209243

Xiao Tong, & Hawley, J. M. (2009). Measuring Customer Based Brand Equity: Empirical Evidence from the Sportswear Market in China. *Journal of Product & Brand Management*, 4, 262–271.

Xie, X. (2011). Service Quality Measurement from Customer Perception Based on Services Science, Management and Engineering. *Systems Engineering Procedia*, 1, 337–343. doi:10.1016/j.sepro.2011.08.051

Yongliang, Z. (2002). 7. Several Reflections on Problems in Zhejiang's Urbanization. *Chinese Economy*, 35(2), 71–76. doi:10.2753/CES1097-1475350271

Zar, J. H. (2014). *Biostatistical Analysis*.

Zikopoulos, P., & Eaton, C. (2011). *Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data*.