



KADİR HAS UNIVERSITY  
GRADUATE SCHOOL OF SOCIAL SCIENCES  
NEW MEDIA DISCIPLINE AREA

**THE ADOPTION AND INFLUENCES OF BIG DATA IN  
TOURISM INDUSTRY IN TURKEY**

MERVE KÜÇÜKBATIR

SUPERVISOR: ASSOC. PROF. DR. ÇİĞDEM BOZDAĞ

MASTER'S THESIS

ISTANBUL, JANUARY, 2018

[MERVE KÜÇÜKBATIR]

[MASTER' S THESIS]

[2018]

# **THE ADOPTION AND INFLUENCES OF BIG DATA IN TOURISM INDUSTRY IN TURKEY**

MERVE KÜÇÜKBATIR

SUPERVISOR: ASSOC. PROF. DR. ÇİĞDEM BOZDAĞ

MASTER'S THESIS

Submitted to the Graduate School of Social Sciences of Kadir Has University in partial fulfillment of the requirements for the degree of Master's in the Discipline Area of New Media under the Program of New Media.

ISTANBUL, JANUARY, 2018

I, MERVE KÜÇÜKBATIR;

Hereby declare that this Master's Thesis is my own original work and that due references have been appropriately provided on all supporting literature and resources.

NAME AND SURNAME OF THE STUDENT

MERVE KÜÇÜKBATIR

DATE AND SIGNATURE



## ACCEPTANCE AND APPROVAL

This work entitled **THE ADOPTION AND INFLUENCES OF BIG DATA IN TOURISM INDUSTRY IN TURKEY** prepared by **MERVE KÜÇÜKBATIR** has been judged to be successful at the defense exam held on **09.01.2018** and accepted by our jury as **MASTER'S THESIS**.

Assoc. Prof. Dr. Çiğdem Bozdağ (Advisor)

Kadir Has University



Assoc. Prof. Dr. Erkan Saka

Istanbul Bilgi University



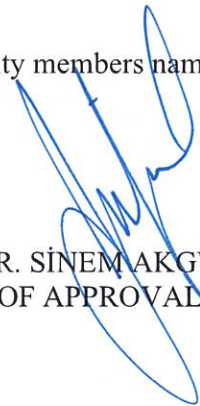
Asst. Prof. Dr. Ayşe Binay Kurultay

Kadir Has University



I certify that the above signatures belong to the faculty members named above.

PROF DR. SİNEM AKGÜL AÇIKMEŞE  
DATE OF APPROVAL: 9/1/18



## TABLE OF CONTENTS

<b>TABLES LIST</b> .....	<b>iv</b>
<b>FIGURES LIST</b> .....	<b>v</b>
<b>ABSTRACT</b> .....	<b>vi</b>
<b>ÖZET</b> .....	<b>vii</b>
<b>INTRODUCTION</b> .....	<b>1</b>
<b>1. LITERATURE REVIEW</b> .....	<b>5</b>
1.1 What is Big Data? .....	5
1.2 The Impact of Big Data .....	7
1.3 How to Get Value from Big Data .....	11
1.4 Opportunities in Tourism Design and Advantages of Using Big Data in Tourism Industry .....	17
1.5 Critique of Big Data .....	20
1.5.1 The lack of skilled workforce .....	21
1.5.2 Data privacy .....	24
1.5.3 Data security .....	25
<b>2. RESEARCH METHODOLOGY</b> .....	<b>28</b>
2.1 Research Design .....	28
2.2 Data Collection .....	29
<b>3. RESEARCH FINDINGS AND ANALYSIS</b> .....	<b>33</b>
<b>CONCLUSION</b> .....	<b>51</b>
<b>SOURCES</b> .....	<b>55</b>
<b>APPENDICES</b> .....	<b>59</b>

## TABLES LIST

Table 3.1.	Number of hotels and average yearly price (World-wide)	36
Table 3.2.	Share in total number of hotels (World-wide)	36
Table 3.3.	Price change over time (World-wide)	37
Table 3.4.	Number of hotels and average yearly price (Turkey)	38
Table 3.5.	Share in total number of hotels (Turkey)	38
Table 3.6.	Price change over time (Turkey)	39
Table 3.7.	Search Report of Metglobal	41
Table 3.8.	Funnel Report of Metglobal	43



## FIGURES LIST

Figure 1.	International tourism	2
Figure 1.1.	Interest over time in the phrase big data	8
Figure 1.2.	What happens in an internet minute	12
Figure 1.3.	Different steps in data mining	15
Figure 1.4.	Different functional areas within an organization	16
Figure 1.5.	Tourism-related data and its application	18
Figure 1.6.	Conceptual classification of big data challenges	21
Figure 1.7.	The impediments to becoming more data driven	22
Figure 1.8.	Sectors with the ability to use and obtain value from big data	23
Figure 3.1.	Behavioral model of forecasting tourist behavior	48



## ABSTRACT

KÜÇÜKBATIR, MERVE. *THE ADOPTION AND INFLUENCES OF BIG DATA IN TOURISM INDUSTRY IN TURKEY*, MASTER'S THESIS, Istanbul, 2018.

Today's common belief on big data says all sorts of industries will become much more efficient through the data-driven evidence. The tourism industry is one of those industries which thrive on information. Together with the adoption of big data in tourism industry, there is a number of opportunities arising which might help organizations in providing predictive modelling analysis and key insights. These insights can provide businesses with more integrated view of the whole industry and accordingly much more enhanced relationship with customers through a focused and responsive approach to customers' needs and preferences. This study concludes that predictive analytics can help businesses in anticipating what is likely to happen in the future, and comprehend the fundamental causes of different results, and thus correlate different insights for future forecasting. However, the way of gaining more opportunities is not only related with putting the technologies related with big data into use, but also related with the expertise of the personnel/data team.

**Keywords:** big data, data analysis, data-driven insights, tourism industry

## ÖZET

KÜÇÜKBATIR, MERVE. *BÜYÜK VERİNİN TÜRKİYE'DE TURİZM SEKTÖRÜNDE KULLANIMI VE TÜRKİYE'DE TURİZM SEKTÖRÜNE ETKİLERİ*, YÜKSEK LİSANS TEZİ, İstanbul, 2018.

Günümüzün büyük veriyle ilgili yaygın inancı, her sanayinin veri odaklı kanıtlarla daha verimli hale geleceğini söylüyor. Turizm endüstrisi, bilgi üzerinde gelişen endüstrilerden biridir. Turizm endüstrisinde büyük verinin kullanımıyla birlikte, öngörülü modelleme analizi ve anahtar bilgiler sunma konusunda kuruluşlara yardımcı olabilecek birtakım fırsatlar ortaya çıkmaktadır. Bu bilgiler, işletmelere tüm sektörün daha bütünsel bir görünümünü sağlayabilir ve bununla birlikte müşterilerin ihtiyaç ve tercihlerine odaklanmış ve duyarlı bir yaklaşımla müşterilerle çok daha gelişmiş ilişkiler kurma fırsatı verebilir. Öngörülü analitik, işletmelere gelecekte neler olabileceğini tahmin etmede, farklı sonuçların temel nedenlerini kavramada ve dolayısıyla farklı içgörülerini gelecek tahmini için ilişkilendirmede yardımcı olabilir. Bununla birlikte, daha fazla fırsat kazanmanın yolu, yalnızca büyük veri ile ilgili teknolojilerin kullanımıyla değil aynı zamanda veri ekibinin uzmanlığıyla da ilgilidir.

**Anahtar Sözcükler:** büyük veri, veri analizi, veri güdümlü içgörü, turizm sektörü

## INTRODUCTION

“Data have become a torrent flowing into every area of the global economy” is a well-said quote to define the age we live in (McKinsey Global Institute, 2011, p.1). Big data is one of the most popular and frequently used terms to define the rapid growth of data in the modern age. The common belief on big data is that the analysis of big data will change all sorts of industries by bringing new opportunities with the data-driven evidence. What attracts people’s attention regarding the big data is more accurate analysis, confident decision-making, reduction in cost and risk, etc. It has been highlighted quite often in the reports on the big data revolution, that businesses can benefit from using big data in their business operations. But, how? “The idea of data creating business value is not new, however, the effective use of data is becoming the basis of competition” (Ernst and Young, 2014, p. 1). People leave many digital traces behind on the Internet through social network sites, smartphones, PCs, and laptops and therefore; large amounts of data become available to use especially for companies to track drivers of sales like shopping patterns, purchasing behavior, recommendation, etc. and then, to analyze those influencers for producing more effective service.

By many academics, researchers, managers and policy-makers, one of the most cited definitions of big data has been produced as “Big data represents the information assets characterized by such a high volume, velocity and variety to require specific technology and analytical methods for its transformation into value” (Wolfert et al., 2017, p.70). Today’s common belief on big data, which says industries will become much more efficient through the use of big data, is a turning point for the world in which lots of the countries started to embrace the power of big data. Currently, many leading companies are already explored how they can create value by using big data and all the others should learn how to do the same if they want to keep their part in the competition of innovation, productivity, and growth. In all sectors and industries, there is an important opportunity

to boost their productivity and efficiency at a significant level through the effective use of big data (McKinsey Global Institute, 2011).

One of those above-mentioned industries that will be changed by big data in terms of the way they operate and compete is travel and tourism industry. Travelling today is radically different than travelling in let's say 1990s, and this is the natural outcome of the widespread use of the Internet and the internet-related technologies such as big data. Tourism industry, as an infrastructural and productive one, has the opportunity of using big data for smart travel experiences with an optimal level of exploitation of maximum data. The 2017 report of the United Nations World Tourism Organization (UNWTO), which presents an overview of international tourism, shows that there has been a growth in tourism globally, and these ever-increasing number of travelers leave their digital fingerprints.



**Figure 1. International tourism**

All of the online research of travelers produces millions of breadcrumbs, namely data, for companies. And, what is left for data scientists and analysts is to follow those breadcrumbs and to market, and of course to make the operators sustain in the business. Data will give the businesses the chance for growing with the right analytics tools as well as the chance for providing the tourists and the travelers with convenient, ideally the best, traveling experiences.

The tourism industry is one of the industries which thrive on information. Because travelers using mobile technologies leave traces behind and that dispersed information can be used by tourism organizations to interconnect and to forecast. The use and the analysis of big data in travel and tourism industry have potential to provide the business insights for innovation. The benefits of big data can be huge with the right approach and it should not be underestimated by tourism organizations. Today, many countries with the appropriate tourism potential seems like part of a cutthroat competition in order to obtain an advantage in the market. And like all sorts of industries, the worlds of travel and hospitality is changing by big data analysis. Together with the adoption of big data in tourism industry, there is a number of opportunities arising which might help organizations in providing predictive modelling analysis and key insights. These insights can provide businesses with more integrated view of the whole industry and accordingly much more relationship with customers through a focused and responsive approach to customers' needs and preferences. To understand customers' needs and preferences better can end up with serving better through enhanced customer focused decisions. Therefore, it is quite vital for travel and tourism industry to efficiently manage the millions of traces regarding the travel behavior of the consumers which is generated by the industry each and every single day.

Together with the right approach, the tourism industry can learn a lot from the benefits of big data. Therefore, the prior objective of this study is to see the outcomes of the use of big data as a competitive advantage tool in this competition. After giving the preliminary background information on big data by going through previously published related work in the literature, this study is aimed to understand the real outcomes of the use of big data by presenting the cutting-edge developments of analytics in travel and tourism industry.

What will be introduced in this dissertation is a contribution to the discussions about how big data is transforming different industries, together with datasets and applications provided by the leading companies operating in this sector; Trivago and Metglobal which use big data very effectively and actively for destination marketing and management. In the light of such information, this study will examine and attempt to answer the following question as the main research question in order to explore the benefits of big data:

- How does big data affect tourism industry in terms of discovering and determining insights and of providing comprehensive results?

And following sub-research questions in order to support the answer to the main research question:

- What are the advantages of using big data in tourism demand forecasting in terms of providing a competitive advantage for businesses?
- What are the problems that the tourism industry faces regarding big data in implementations of projects?

This study by no means represents the final word in the field of big data, but rather, what is anticipated with this research is a humble attempt to contribute to an evolving field by technological developments alongside associated concerns and challenges, and to help readers understand the opportunities in the use of big data in tourism industry in the overall competitive landscape. Thus, this work is classified into five chapters. The first chapter presents an introduction to the study field and points out the potential gains of tourism industry together with the use of big data. In the second chapter, it is elaborated previously published related work in the literature with related concepts. The third chapter explains the methodology employed by this research. The following chapter includes the findings of this study as well as the analysis. And the last chapter concludes this dissertation with the summary of the study and the implication of the research.

# CHAPTER 1

## LITERATURE REVIEW

### 1.1. WHAT IS BIG DATA?

The world has been turning into a digital space. In today's world; together with the widespread use of the Internet, people manage, share and store their lives online. The tremendous data; like social media shares, text and record keeping files, photos, videos, etc. -basically, everything people provide on the Internet- waiting to be processed in a meaningful format, are around now. In order to collect and transmit information on what people do, those large amounts of data are gathered through a variety of sources and devices like internet server records, internet logs, computers, tablets, smartphones, smartwatches and/or other wearable devices, etc. Big data, through these sources, comes in three forms; structured data, unstructured data, and a combination of these two; semi-structured data (Introduction to Big Data - Module 2 – Beyond the Hype, 2016). This phenomenon, which changes people's understanding of the world and their place in it, is called big data. So, what is big data? Due to the broad interest in big data from different groups, there are several attempts to define big data. Although the concept of big data is continually evolving and being reconsidered and there is not a universally accepted definition so far,

From a macro perspective, big data can be regarded as a bond that subtly connects and integrates the physical world, the human society, and cyberspace. Here the physical world has a reflection in cyberspace, embodied as big data, through Internet, the Internet of Things, and other information technologies, while human society generates its big data-based mapping in cyberspace by means of mechanisms like human-computer interfaces, brain-machine interfaces, and mobile Internet. (Jin et al., 2015, p.59)

Basically, in Bernard Marr's words; "big data refers to our ability to collect and analyze the vast amounts of data we are now generating in the world" (Marr, n.d.). Ernst and Young offers an alternative definition as follows: "Big Data refers to the dynamic, large

and disparate volumes of data being created by people, tools and machines. It requires new, innovative, and scalable technology to collect, host and analytically process the vast amount of data gathered in order to derive real-time business insights that relate to consumers, risk, profit, performance, productivity management and enhanced shareholder value” (2014, p. 2).

Big data is not something that you can define with certain amounts of gigabytes or petabytes; but rather it is something that an individual’s or a firm’s storage capacity or ability to analyze data is exceeded by the amount of data they have (Introduction to Big Data - Module 1 - What is Big Data?, 2016). Although the word ‘big’ sounds like referring to volume, what big data refers to is not only the ever-increasing size of data sets that people can analyze, but also the speed at which it is created and used as well as the different types and structures of data that people can analyze now.

Despite the fact that there is not a unified definition of big data at this moment, there are some elements which are common across the different definitions. To describe big data, people have been using Bernard Marr’s widely accepted five Vs characterization: Volume, Velocity, Variety, Veracity and the emerging V; Value. To have a detailed look at 5 Vs as the concepts that describe the dimensions of big data;

- **Volume:** What is meant by volume is the vast amount of data generated every second; namely all the emails, tweets, Facebook messages, photos, videos and sensor data that people generate and share.
- **Velocity:** It basically refers to the speed at which new data is produced and the speed at which data moves around. Through the technology of big data people now able to analyze the data while it is being generated without the need of putting it into databases.
- **Variety:** What is implied with variety is the different types of data people can use now, in other words; the mixture of structured and unstructured data. Today, 80 percent of the world’s data is unstructured, namely photos, videos and social media updates, and it is hard to put this type of data into tables or relational

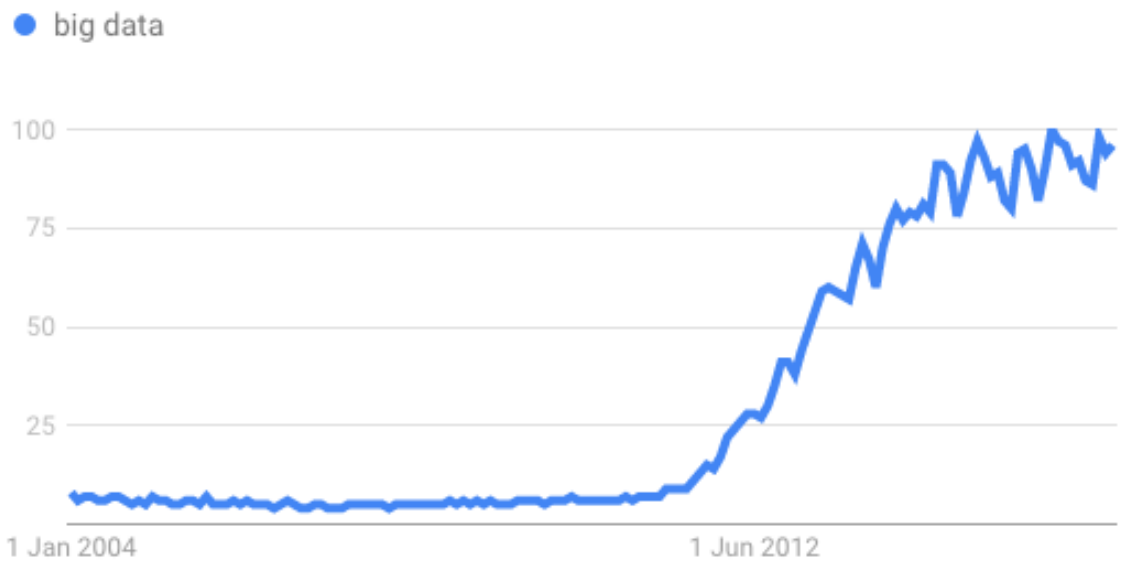


databases. However, together with the technology of big data it is possible to harness different types of data.

- **Veracity:** It refers to the data quality which needs to be preserved with the increasing volume of data since the quality and the accuracy are not easily controllable when it comes to the different forms of big data.
- **Value:** How people will turn all this information into knowledge is the most problematic part of big data. So, the last V of big data, which matters the most, refers to the ability of turning data into value (2014).

## **1.2.THE IMPACT OF BIG DATA**

Big data is today's one of the most frequently discussed topics. On the Google search engine, when the phrase "big data" is searched by people all over the world for the period from 2004 until today, it can be easily observed on Figure 2.1 that there is a serious amount of upward movement after 2010.



Worldwide. 2004 - present.

**Figure 1.1. Interest over time in the phrase big data**

To have an appreciation for why so many people are interested in big data, we can have a look on how big data has begun transforming many areas of business, industry, research and many other parts of people's lives. The importance and the strength of big data comes from its potential of providing a window into the lives of customers that any business has never imagined before. The articulation of this point of view regarding big data in Chris Anderson's article, "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete", is like this; "There is now a better way. Petabytes allow us to say: "Correlation is enough." We can stop looking for models. We can analyze the data without hypotheses about what it might show. We can throw the numbers into the biggest computing clusters the world has ever seen and let statistical algorithms find patterns where science cannot" (2017).

A similar argument regarding the scientific analysis comes from Marc Prensky as follows; "...scientists no longer have to make educated guesses, construct hypotheses and models, and test them with data-based experiments and examples. Instead, they can mine the

complete set of data for patterns that reveal effects, producing scientific conclusions without further experimentation...”(2009, p.6).

However, it should also be noted that there is not a comparison between correlations and causation in terms of the importance. As it is stated in Fulvio Mazzocchi’s article, “Could Big Data be the end of theory in science?”, although data mining techniques increase the capacity to find relevant patterns and to uncover complex structures within the huge amounts of data, these correlations can tell and alert that something is happening but they are not enough to understand why it is happening (2015, p. 1252). The crucial part of understanding why something is happening is related with the necessity of reaching a certain level of knowledge for more confident practical applications and more reliable predictions (Mazzocchi, 2015, p. 1252). In addition to Mazzocchi’s arguments, one of the problems regarding the spurious correlations is pointed out by Boyd and Crawford as follows; “Too often, big data enables the practice of apophenia: seeing patterns where none actually exist, simply because enormous quantities of data can offer connections that radiate in all directions” (2012, p. 668).

The excitement that big data has generated both in the business press and in the larger culture is quite understandable since there are easily observable effects of big data such as scientific breakthroughs, new business models, and great societal transformations. There is no end for the possible applications of big data, therefore, today, not only technology based or technology related companies but almost all enterprises have realized the power of information, and; they have been already using big data to better understand and predict customer behavior and optimize and improve business processes. Using detailed databases regarding the behavior and lifestyles of customers by companies is beneficial not only for better targeting customers, but also for creating such innovative products like playlists, newsfeeds, next-best offers, and recommendation engines for items ranging from airplane tickets to couples. There is no doubt that the large amounts of data are the raw material for all and many more of these innovations. Therefore, the idea of that big data is not relevant to a company is not acceptable for businesses of all sizes, large and small, which want to get one step ahead and therefore cannot risk being left behind.

Big data might have many different use cases. To have a clear understanding on how big data is impacting people and businesses at different levels, we can have a look at several examples. When people search for or buy a product on Amazon, they will notice that Amazon starts making recommendations related to the product they searched for. This is a common application of big data, recommendation engines. Not only Amazon, but also companies like Spotify and Netflix use similar algorithms based on big data to make explicit recommendations based upon customer preferences and customers' historical behavior. Another recommendation engine application is of Google's recommendations based on the big data on a user's device (Introduction to Big Data - Module 1 - Big Data in Business, 2016).

While the impact of big data on consumers is like this, an impressive example on the business side can be shown as follows. The television business is full of surprises. A gold-plated director, a bankable star and a popular concept might not result in success. However, together with the use of big data the success can be guaranteed in advance. The famous TV series, House of Cards proved an immediate hit when it debuted on the video streaming service Netflix on 1<sup>st</sup> of February in 2013. When it became an immediate hit on the release of its first season, the strange thing about the series was that the creative direction was inspired by data since the show was a product of big data analytics. While explaining the reasons which gave the company enough confidence to go ahead and to spend millions, Netflix's Chief Communications Officer Jonathan Friedland said in his interview with The New York Times; "Because we have a direct relationship with consumers, we know what people like to watch and that helps us understand how big the interest is going to be for a given show. It gave us some confidence that we could find an audience for a show like House of Cards" (Carr, 2013).

As a large streaming service provider with 33 million subscribers worldwide, the company knows what people are watching, when, where and with what kind of device they are watching; so, at the end Netflix knows more about the viewing habits of the audience than they can imagine (Carr, 2013). By the right approach in using data insights as in the case of House of Cards, data can be a that much powerful tool for businesses as

well. Any business in any industry has a chance to reinvent itself by turning huge amounts of data into smart data.

### **1.3. HOW TO GET VALUE FROM BIG DATA**

Data science is an evolving field which requires some skills like cleaning, mining, and analyzing huge amounts of data in order to extract meaningful insights of value and thus to provide comprehensive results. Those insights derived through an exploratory analysis are used to make decisions related to a business and/or scientific case. But first, it should be noticed that there is a fundamental distinction regarding big data that needs to be reminded: data and information are not the same thing. But rather, data can be considered as a raw material waiting to be interpreted in a database for the information it holds.

However, all these is of course possible with the availability of those large amounts of data. Today people do not worry about the lack of data as they used to in the past. Today's "overwhelming amount of complex and heterogeneous data pouring from any-where, any-time, and any-device", can be explained even as a data deluge (Sivarajah et al., 2016, p.263). The internet activities of the users all over the world can be seen as main reason of the current explosion of data. Together with the increase in the internet speed and in the access to the Internet, these numbers and correspondingly big data will continue to grow. To explain the indications of this growth, and to drill down to what is happening in an internet minute and to have an idea about how great the scale of the Internet is, Lori Lewis and Chadd Callahan create the graphic below, Figure 2.2, each year:

# 2017 *This Is What Happens In An Internet Minute*



Figure 1.2. What happens in an internet minute

However, although big data is something generated each and every moment by each and every device connected to the Internet, it will be useless until people know how to analyze that data in a correct and efficient way. In spite of the increasing availability of data, the problem of analysis still exists. As it is stated in the article of Hassani and Silva, “big data and predictive analysis goes hand in hand in the modern age with companies focusing on obtaining real time forecasts using the increasingly available data” (2015, p. 6). Therefore, it can be argued that deciding how to use these increasingly available data is the biggest problem in terms of the ways people approach big data to get value from it.

“As a large mass of raw information, Big Data is not self-explanatory. And yet the specific methodologies for interpreting the data are open to all sorts of philosophical debate. Can the data represent an ‘objective truth’ or is any interpretation necessarily biased by some subjective filter or the way that data is cleaned?” (Bollier, cited in Boyd and Crawford, 2012, p. 667).

To highlight the importance of this problem, Shafiee and Ghatari states that “the analytical approach of big data emphasizes the capacity of data collection and analysis with an unprecedented extent, depth and scale for solving the problems of real life and uses it” (2016, p. 1). Since the interpretation is at the heart of data analysis process; to avoid misinterpretation and to make data analysis more effective, all these limitations and biases of which data is subject to need to be understood and outlined (Boyd and Crawford, 2012).

“Five years ago, most companies collected data that were a part of their daily transactions and stored them in a database. This data was used primarily to keep track of operations or forecast needs “(Parise, Iyer and Vesset, 2012). A proactive approach to manage digital transformation is essential to keep the largest brands and businesses relevant and competitive in the future. Together with a winning digital strategy; a company can have a vision for how integrated digital marketing can support their business, the best resources to compete, and a performance improvement. Concordantly; in global economy, almost all big companies have started relying on feedbacks from their customers, business operations and their interior processes to find out new opportunities for sustainable growth. “In fact, according to a recent Gartner study, three-quarters of companies have already invested in big data or plan to within the next two years” (Deighton, 2017).

Currently, no one disagrees with the idea that various kinds of data lakes and data warehouses exist with their own systems to store and organize big data, but the problem is to achieve generating insights. At this stage, what is important for companies decided to rely on data they collected is to focus on the relevant type of data which will be valuable to their industry for not drowning in a bunch of data. Since it is vital to “know and account for the weaknesses in that data”, companies should know “where data is coming from” in

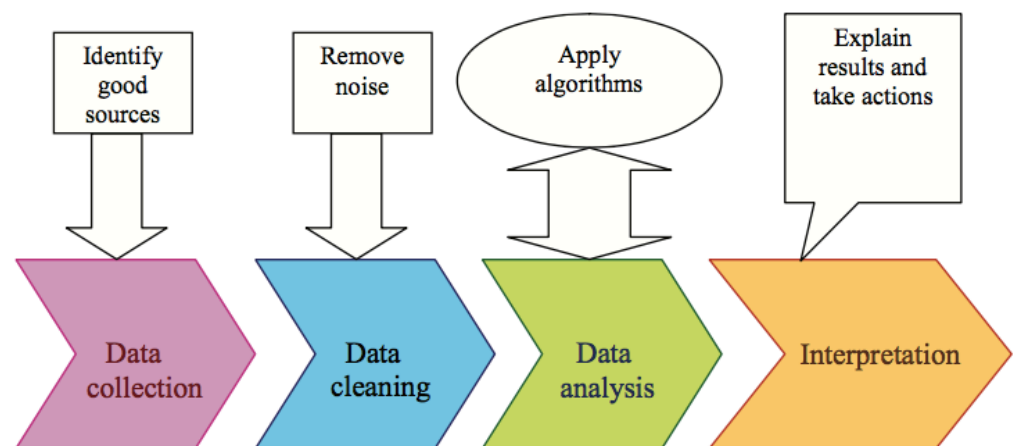
order to “make statistical claims about a data set” (Boyd and Crawford, 2012, p. 668). “The type and content of big data can vary by industry and thus have different value propositions for each industry” (Russom, 2013). Thus, it can be said that a data scientist and/or data analyst should have clear objectives before gathering the data and then start collecting the data in accordance with these objectives. After collecting the data, the next step could be to consider the possible outcomes and then accordingly to analyze that data. Since it should be remembered that “regardless of where big data is generated from and shared to, with the reality of big data comes the challenge of analyzing it in a way that brings big value” (Sivarajah et al., 2016, p. 264). To come up with a process model regarding making sense out of big data, the following steps might be useful. A skillful data scientist can start with determining the problem of the business, then continue with deciding the project objectives and collecting the relevant data in accordance with the objectives. After exploring and analyzing the data by building a model, he/she can end up by making decisions and taking action. (Introduction to Big Data - Module 2 – Beyond the Hype, 2016)

It is clear that simply collecting big data will not be enough to unleash the potential value in it. The real challenge of big data analysis is “abundance, exhaustivity and variety, timeliness and dynamism, messiness and uncertainty, high relationality, and the fact that much of what is generated has no specific question in mind or is a by-product of another activity” and it is a bit difficult to implement without high-powered computation and new analytical techniques including “artificial intelligence and expert systems that have sought to produce machine learning that can computationally and automatically mine and detect patterns and build predictive models and optimize outcomes” (Kitchin, 2014, p.2). To understand how to cope with the challenge of the analysis part of the process, we can also have a look at Shelly Blake-Plock’s words;

A human brain is not well-suited to the digestion of mountain-sized data. ... Human brain input is too slow to handle this kind of big data alone. Machine brains, on the other hand, are excellent at this kind of task. These machine brains are not all-purpose, all-knowing universal AI, but instead, specialized algorithms that can learn to identify patterns, correlations and indicators through training and repetition. Machine brains can be designed to handle highly specialized datasets, spinning up hundreds of concurrent threads to manage the kind of data ingestion speed needed to truly eat mountains. (2017)



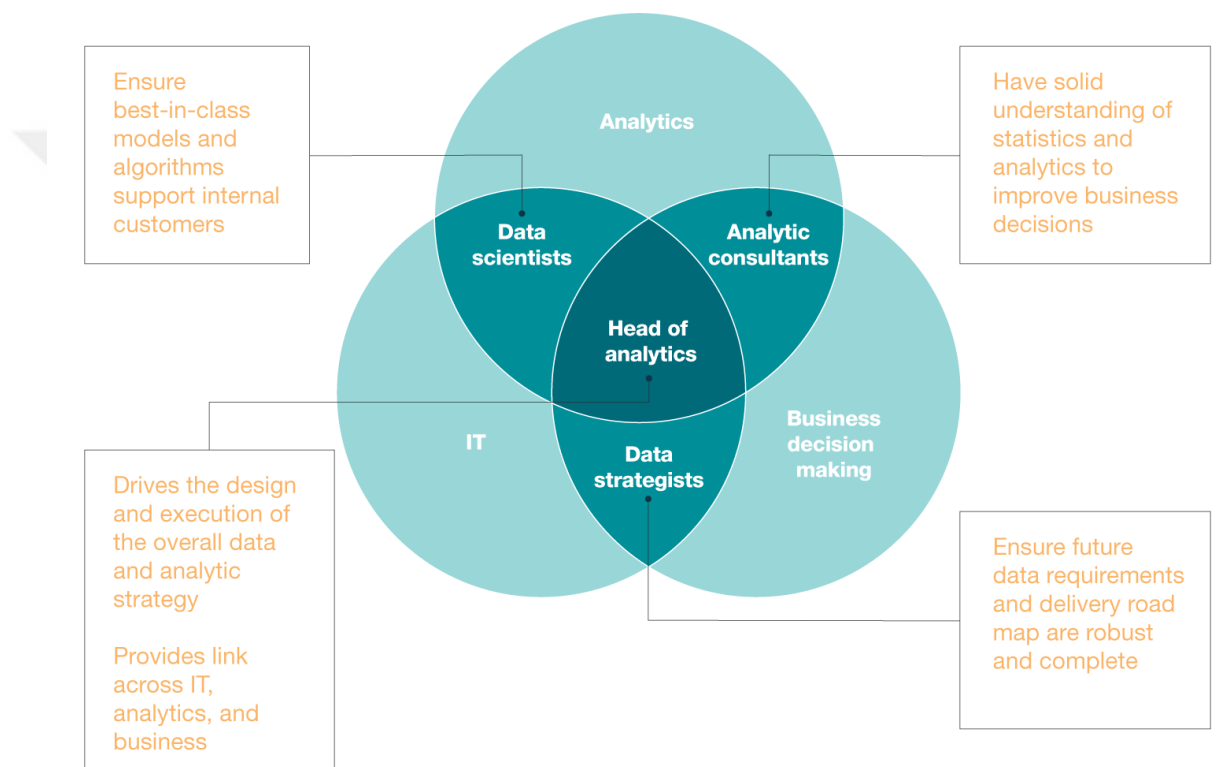
The reason behind this is related with the definition of analytics for business which is expressed as “the extensive use of data, statistical and quantitative analysis, using explanatory and predictive models to drive fact-based business management decisions and actions” (Ernst and Young, 2014, p.6). If so, what are the steps involved in data analysis process? Before a data analyst jumps into writing an algorithm, he/she needs to spend some time with his/her data. The process starts with a question he/she wants to answer or a problem he/she wants to solve. The next step of the process can be data wrangling phase which has two parts as data acquisition and data cleaning. At first; what is required is to acquire the data that he/she needs to answer his/her question or to solve his/her problem, and then it comes to investigate the data and to clean up any problems found. The next phase is data exploration. In this phase, a data analyst starts spending some time in getting familiar with the data and thus starts building his/her intuition regarding it and finding patterns. Once he/she gets familiar with the data, he/she wants to make some predictions or to draw conclusions about it. This face generally requires statistics or machine learning that are beyond the focus of this research. And finally, communication phase, as the last step of this process, involves the necessity of communicating what a data analyst found with other people since the findings will as useful as they are transferred or shared. (Buckey, 2017). This multiplicate process of translating the aggregated data into meaningful information is depicted with the following scheme (Bose, 2009, p.936);



Copyright © 2009, IGI Global, distributing in print or electronic forms without written permission of IGI Global is prohibited.

**Figure 1.3. Different steps in data mining**

However, to fully recognize the value of big data is not limited to data mining process, but also linked to identifying connections between different functional areas since what is needed to implement big data is multiple roles in multiple area of expertise. To bridge different functional areas within a given organization and to put the process explained above into practice, the need for and the necessity of a collaboration among these different functions are described with the visual below (Ariker, Breuer and McGuire, 2014);



**Figure 1.4. Different functional areas within an organization**

Considering the explanations given throughout this section, the high level steps of data analysis process, namely the way of getting the most from big data can be respectively summarized as; to bring together a group of people who have the complementary skills, to let these people define a strategy in accordance with the problems and/or needs of the company, to collect the relevant type of data and to perform analytics in order to render the data into insights and finally to turn these insights into offers and/or messages in the marketplace. In short, the effective use of big data is up to a successful combination of

technology, people, and process. Once this process is achieved successfully, “it enables organizations to meet stakeholder reporting demands, manage massive data volumes, create market advantages, manage risk, improve controls and, ultimately, enhance organizational performance by turning information into intelligence” (Ernst and Young, 2014, p.6).

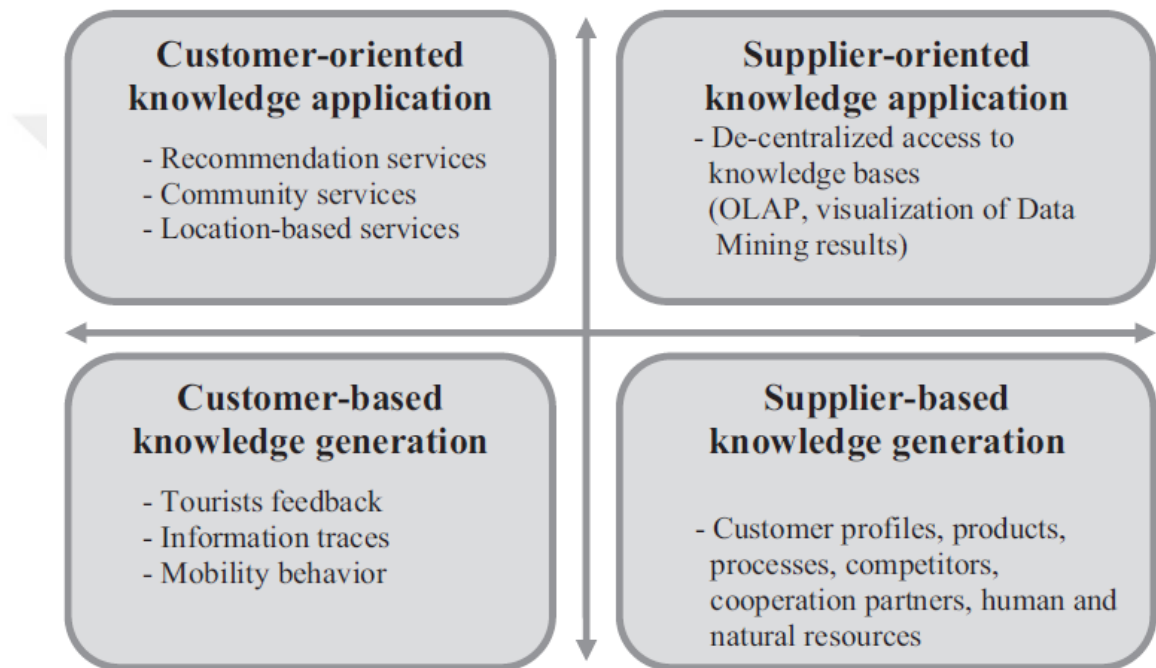
#### **1.4. OPPORTUNITIES IN TOURISM DESIGN AND ADVANTAGES OF USING BIG DATA IN TOURISM INDUSTRY**

In previous sections, it has been pointed out that big data has the possibility of enhanced insights and so better decision-making across all industries and what is needed is just to take the opportunity as it comes. The more industries will put the technologies related with big data into use, the more opportunities will appear. But, how can this be possible specifically for travel and tourism industry?

Millions of people travel around the world every day for several reasons including but not limited to business, vacations, and sightseeing (Bose, 2009). While international travel continues to grow strongly, thanks to internet technologies, there has been some changes occurred in tourism industry (Www2.unwto.org, 2017). These changes occurred through internet technologies are both on the supply side and the demand side, and the outcomes are detailed as follows; “Tourists expect that tourism companies actively implement modern technologies as the part of the value chain. Modern trends and technologies generated a whole set of new tools, such as recommendation systems. In addition, new technologies help tourism companies to establish one-to-one connection with tourists, thus increasing the customer loyalty” (Bach, Schatten and Marušić, 2013, p. 26).

Customers’ interaction with these internet technologies, for example; “all travel-related activities, like searching and trip planning, reservation and booking, service consumption as well as feedback provision in community web-sites (e.g. social media platforms) or through online surveys”, has engendered digital traces (Fuchs, Höpken and Lexhagen,

2014, p. 198). And, as a consequence, a huge amount of data is collected. In terms of the source, it can be either customer based data or supplier based data. Supplier based data is “about goods that tourists consume such as tickets, hotel rooms, entertainments, and so forth”, while customer based data “consists of aggregate information about tourists to service providers” (Bose, 2009, p. 936). Those changes, and the generated data and its application as a result of those changes are pieced together with the following figure (Fuchs, Höpken and Lexhagen, 2014, p. 200);



**Figure 1.5. Tourism-related data and its application**

Well, what is all these for? According to Ulrike Gretzel, “innovation deficiencies in the tourism industry and the need for greater market orientation, strategic innovation as well as collaboration” is all for “to take advantage of big data and stay globally competitive” (n.d., p. 1). If this is the main objective with big data, the main challenge for the tourism industry can be expressed as “to translate these technological opportunities into revenue and competitive advantage” (Gretzel, n.d., p. 2). If so, what is needed to make this aim real? Companies within the tourism sector need to have the required organizational capabilities to embrace big data as well as the analytical mindset it necessitates (Gretzel, n.d.).

In the previous section, the steps involved in data analysis process were summarized. In the case of tourism industry too, data mining process is a very beneficial tool to analyze data and, as might be expected, it goes through the same steps. Tourism data mining also starts with the collection of the most suitable data from the most appropriate sources. The suitable and relevant data to provide more efficient services can be “travelers’ hotel stays, purchase transactions, and customer information” (Song and Liu, 2017, p. 16). However, “it is possible to cross-reference tourism big data with other sources such as social media and open public data” (Song and Liu, 2017, p.16). Organizations need to know about the relationship between tourism activities and preferences of tourists to plan tourism infrastructures (Bose, 2009). “When the aggregated data about the tourists is presented in the right way, analyzed by the correct algorithm, and put into the right hands, it could be translated into meaningful information for making vital decisions by tourism service providers to boost revenue and profits” (Bose, 2009, p. 936).

To express some advantages of big data usage in tourism industry Song and Liu enumerate the following elements. One of the advantages of using such innovative methods in tourism industry is reliability since big data is based on the real actions of the users so it enables to get comprehensive results rather than biased conclusions. Yet another advantage is phrased as real-time data in other words now-casting. It is used to describe the use of contemporaneous activities. Different types of real-time data streams can be exemplified as Google search queries, data on credit card purchases, the trucking and shipping of packages, mobile phone usage, etc. Thanks to these advantages, any demand in tourism can be addressed instantaneously and organization can remain relevant to consumers (2017).

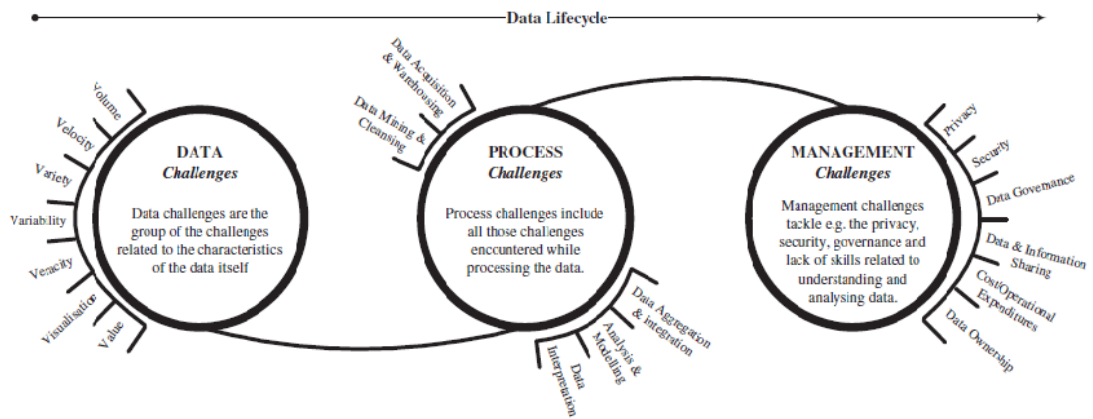
All these changes and resultant advantages have led to a prevalent belief of that big data can aid the creation of an authentic connection between consumers and tourism organizations in order to succeed an improvement customer service and support through predictive analytics (Song and Liu, 2017). In this sense, predictive analytics can be a strong “foundation of well-informed, highly efficient, and deeply satisfying interactions that benefit both customers and businesses” (Song and Liu, 2017, p. 29). According to

Song and Liu, the reason behind this is the more insight offered by predictive analytics in order to understand customer preferences. However, they continue with a big but by saying the success of predictive analytics is up to the quality of the feedback mechanisms and the quality of big data regarding consumer behavior (Song and Liu, 2017). The importance of feedback mechanisms is expressed in the following way; “Feedback in the tourism industry is important in the quest to identify customer preferences and deliver positive experiences. Soliciting customer feedback is one of the most important elements in achieving high company growth and building a strategy around better meeting customer needs” (Song and Liu, 2017, p. 21).

Thus, the increase in the use of feedback mechanisms can result in “commercial messages which are quick, unique, focused, and personal” (Song and Liu, 2017, p.21). In addition to feedback mechanism, the importance of recognizing consumer behavior has been mentioned as a requirement for the success of predictive analytics. Today, companies are experiencing an “unprecedented flux in consumer behavior, customer expectations, and company business models created by technologies that is simultaneously disrupting established businesses and spawning new ones” (Song and Liu, 2017, p. 20). These changes in the customer-business relationship and using tourism big data to derive more insights into customer behavior will end up with a customer-business connection.

## **1.5. CRITIQUE OF BIG DATA**

Despite the rise of big data with its unprecedented insights and opportunities across all industries, “...not all authors agree that big data is a revolutionary phenomenon” (Hassani and Silva, 2015, p. 6). “Like other socio-technical phenomena, big data triggers both utopian and dystopian rhetoric” (Boyd and Crawford, 2012, p. 663). There is no doubt that if companies and organizations are willing to capture the full potential from big data, they need to touch upon significant challenges. Some of these challenges are grouped into three categories consisted of data challenges, process challenges, and management challenges by Sivarajah et al. as it is seen in Figure 2.6 (2016, p. 265):



**Figure 1.6. Conceptual classification of big data challenges**

There are some concerns and questions that need to be addressed such as the existence of the analytical talent, privacy, and security. While the tremendous benefits of big data such as cost and risk reduction, improved efficiency and better decision-making process are more than welcomed, those concerns and questions mentioned above cannot be disregarded but rather needs to be critically interrogated.

### **1.5.1. The Lack of Skilled Workforce**

To analyze large amounts of data and to make decisions based on the findings, in other words, to make the most of big data, there has to be an analytical talent. In previous sections, it is pointed out that the existence of qualified data scientists is one of the key factors in the implementation of big data applications. Figure 2.7, which shows the results of a survey where “respondents were asked to select three obstacles to the widespread adaption of analytics in their organization”, corroborates this idea (LaValle et al., 2011, p. 25).

## THE IMPEDIMENTS TO BECOMING MORE DATA DRIVEN

The adoption barriers organizations face most are managerial and cultural rather than related to data and technology.



Figure 1.7. The impediments to becoming more data driven

What is expected from the analytical talents, to be more precise what is meant by talent is the ability to understand the questions, problems, strategic challenges of a given organization and thereafter translating them into the design of data analysis projects (Introduction to Big Data - Module 2 – Beyond the Hype, 2016). “While companies don’t often think about talent in terms of value chains, the skill and capability links between people are crucial for unlocking the full value of advanced analytics” (Ariker, Breuer and McGuire, 2014). If we compare the situation in the past with today’s availability of data and the data tools together with the ability to reach and to store them in a very cheap and ubiquitous way; it can be clearly argued that there would not be a better time to work as a data scientist or a data analyst. Here, the question is whether or not we have enough skilled workforce. “According to a McKinsey survey, only 18 percent of companies

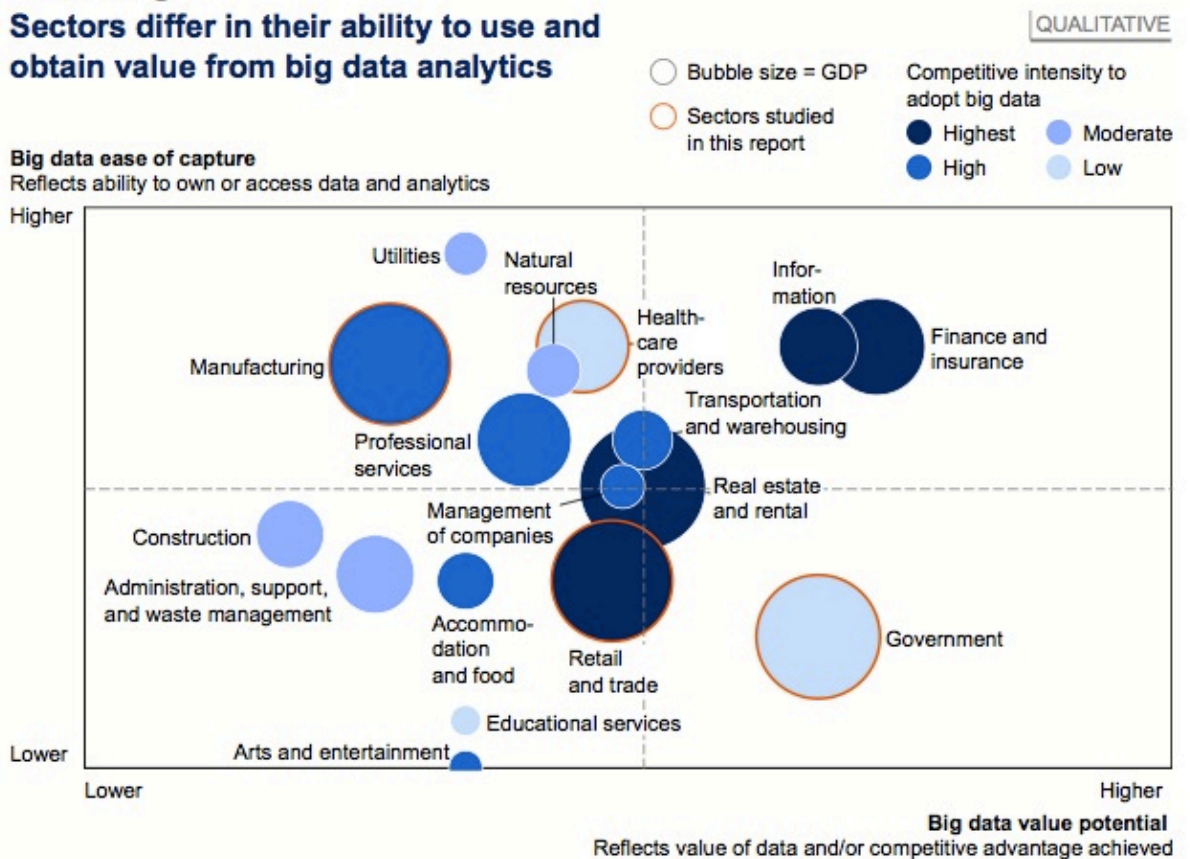


believe they have the skills necessary to gather and use insights effectively” (Ariker, Breuer and McGuire, 2014).

According to another research which highlights the impending data scientist shortage, “by 2018 the United States will experience a shortage of 190,000 skilled data scientists, and 1.5 million managers and analysts capable of reaping actionable insights from the big data deluge” (Davis, 2013). Considering the data explosion mentioned in the previous sections together with this statement above, the demand for skillful practitioners in the wide range of industries can be visualized as (Davis, 2013);

### Exhibit 15

#### Sectors differ in their ability to use and obtain value from big data analytics



SOURCE: US Bureau of Economic Analysis; McKinsey Global Institute analysis

Figure 1.8. Sectors with the ability to use and obtain value from big data

Thus, it is important to note that one of the key challenges of remaining competitive in the market and maximizing efficiencies across all industries is to build out a strong data science team.

### 1.5.2. Data Privacy

It is clearly evident that big data-powered services are quite beneficial for our lives. However, it should be questioned that they are beneficial at what risk to people's privacy. As Boyd and Crawford state, "...big data is seen as a troubling manifestation of ... enabling invasions of privacy..." (2012, p. 664). That's why, one of the most critical parts of the big data era is data privacy issue since how data is collected, retained, used, and disclosed requires a great thought and as it is understood the data is centered around people. "Many citizens around the world regard this collection of information with deep suspicion, seeing the data flood as nothing more than an intrusion of their privacy" (McKinney Global Institute, 2011, p.1). As the value of big data-powered apps grows, it becomes more apparent that there is and will be even more visible trade-off between utility and privacy especially for some categories of personal data which can be called more sensitive to share. In his article named *3 Massive Big Data Problems Everyone Should Know About*, Bernard Marr uses these words to explain the all-pervasive effect of the personal data usage;

We're now at the point where even a total technology boycott may no longer fully protect us. Unless, of course, you choose to walk everywhere you go, wear a different mask every day (to foil face-recognition technology) and use only cash (that you never deposit in a financial institution). Succeeding at navigating the modern world without technology is quite tricky and won't necessarily protect your privacy 100%. (2017a)

Well then, what constitutes personal data? According to the General Data Protection Regulation, "any information related to a natural person or 'Data Subject', that can be used to directly or indirectly identify the person. It can be anything from a name, a photo, an email address, bank details, posts on social networking websites, medical information, or a computer IP address" (EU GDPR Portal, 2017). It is preferred to think in a benign way regarding the use of this data, however, there is always the risk for this data to be used for evil purposes. If so, what could be the best way to regulate online privacy rules? What policies and principles should be established to govern big data privacy? One of the

solution recommendations can be to make the service providers inform their customers on which information will be collected and then how this information will be shared or used. “Currently, each region (European Union, USA, etc.), government and enterprise handles privacy and data protection in a different way” (Isaca.org, 2013). If we look at European Union regulations, we see that the General Data Protection Regulation provides “data transparency and empowerment of data subjects” (EU GDPR Portal, 2017). And, when it comes to penalties, the following statement is kind of a distant signal of a stronger regulation on its way; “under GDPR organizations in breach of GDPR can be fined up to 4% of annual global turnover or €20 Million (whichever is greater). This is the maximum fine that can be imposed for the most serious infringements e.g. not having sufficient customer consent to process data or violating the core of Privacy by Design concepts” (EU GDPR Portal, 2017).

Businesses should establish strict controls and privacy policies in compliance with the legal framework of a given region because a privacy breach is likely to occur if there is an unauthorized collection, use, and/or disclosure of personal data. It is inarguably certain that personal data gathering has to be in compliance with data protection laws and consequently there will not be any legal conflict with the customers. However, at the same time, it should be implemented in a way organizations are still able to create an effective and competitive usage of big data.

### **1.5.3. Data Security**

Despite the revolutionary contributions of it, from a security point of view, another major concern regarding big data is the protection of data privacy. It has been pointed out in the previous section that there is a trade-off between the sensitive data, which can be personal information or confidential business information, and utility; and this situation poses the problem of how to protect this data and to prevent data breaches. Even if people agree to their data being used in return for the service or product of a company, can or should they trust that company to keep their data safe? A definition arises from this question, the definition of data loss prevention. It is “the strategy used to ensure that sensitive data is not lost, misused, or accessed by unauthorized users” (Lord, 2017). And as a

consequence, this strategy generates the need for technological and policy tools. What is expected from these tools are to protect all kind of confidential information and to prevent people from willingly or accidentally sharing this information and consequently putting the company at risk.

What is the source of this security flaw? The main reason behind data security problem is the integration of big data and cloud storage (Buttler, 2017). The effect of a possible data breach and its consequences are well explained by Guillermo Lafuente; “Because of the big amount of data stored, breaches affecting big data can have more devastating consequences than the data breaches we normally see in the press. This is because a big data security breach will potentially affect a much larger number of people, with consequences not only from a reputational point of view, but with enormous legal repercussions” (2014).

Since big data is a relatively new phenomenon, we do not have already existed policies and procedures. Although there is not a huge variety of practices to manage large amounts of data from a security perspective and, it is an undeniable fact that to find an easy answer to this security problem; there are some solutions which vary in methodologies derived from the primary states of data; data at rest, data in motion, and data in use. The key objectives associated with these three states of information are respectively as follows; “to locate and catalog sensitive information stored throughout the enterprise, to monitor and control the movement of sensitive information across enterprise networks, and to monitor and control the movement of sensitive information on end-user systems” (Isaca.org, 2010).

So, it is evident that organizations face with some security problems from analytics applications to infrastructures. When it comes to mitigate those challenges, it can be recommended that

Policy makers should focus on providing guidance for secure use of Big Data systems in the critical sectors... Big Data providers or vendors should invest in compliance with security standards for their products (devices, services, cloud etc)... The competent authorities of the critical sectors should encourage vendors to offer security authentication mechanisms and protocols in their products... The standardisation bodies should adapt existing or create new security standards for Big Data... Industry players and vendors should invest more into

enhancing technical security skills of the staff on Big Data through trainings and certifications. (Naydenov et al., 2015, p.26)

Thus, it can be concluded by saying that big data requires big security to ensure its potential benefits.



## **CHAPTER 2**

### **RESEARCH METHODOLOGY**

In this chapter, what will be provided is the methodology used to answer the research question of the thesis. What is expected through the provided information in this chapter is to help future studies in this field with similar objective. In the light of this motivation, Research Methodology chapter consists of the following sections; research design, and data collection.

#### **2.1. RESEARCH DESIGN**

To be able to answer the research questions, the following steps were performed. After having discussed the theoretical framework through the literature related with big data and its use in tourism industry, it has been preferred to use a mixed method of qualitative and quantitative research methods by making interviews with subject matter experts and analyzing the collected datasets and reports of the contributor companies (Metglobal and Trivago). To clarify the challenges and opportunities of big data and working with it, qualitative research method is used since it gives the opportunity to explore the subject of a given study with the respondents' perspectives. Although qualitative method provides the meaning of practices; statistics and data, when it comes to reflect the state of the use of big data in specific fields, using quantitative research method is a must as it offers the framework and background information which helps to put the findings into context.

Here, the importance of conducting this kind of a research in Turkey should be mentioned. Thanks to the favorable location and existing potential, "Turkey is currently the 6th most popular tourist destination in the world, attracting more than 30 million tourists annually

and continuing to show positive growth year-on-year” (Invest.gov.tr, n.d.). Although the country has been dealing with political tensions and unpredictable shocks from terrorist attacks over years, tourism industry showed resilience, contributing direct GDP growth of 4.1% (Travel&Tourism Economic Impact 2017 Turkey, 2017). Therefore, a sustainable development in tourism industry supported with the current technologies is crucial for Turkey since its economic contribution is at a substantial level. To examine the outcomes of the use of big data in tourism industry in a country like Turkey would be beneficial in terms of understanding the current state of big data projects in tourism industry and implementing the findings on different countries or regions with similar conditions.

## **2.2. DATA COLLECTION**

I gathered data through collecting and analyzing index and reports from the leading companies operating in travel and tourism industry; Trivago and Metglobal which use big data very effectively and actively for destination marketing and management, and conducting online interviews with subject matter experts. The index used in this study is the most expensive and cheapest cities ranked by average hotel rates and trivago hotel price index (tHPI) which is a monthly index on current hotel prices in cities, regions and countries. To present results of the index and to make it easier interpreting the index, based on a small segment of Trivago’s big data which shows price fluctuations and accommodation trends, some tables are produced and presented in the following chapter. Together with the index shared by Trivago, some reports provided by Metglobal are used in this dissertation; the funnel report which shows all the movement and the counts of user traffic including all steps from the search phase to booking phase on an hourly basis and the search report which shows the hotel, agency and supplier-based distribution of all search that Metglobal receive. In addition to all these index and reports, I conducted online interviews with professionals who work closely with data at Metglobal to provide deeper understanding on this matter via the questions about the current state of the use of big data, and the analytical strategy of their organization, and significant challenges of big data projects that tourism industry faces.

### ***Trivago***

Trivago, as it is described on its website, is a global brand with a mission of “the traveler’s first and independent source of information for finding the ideal hotel at the best rate” (trivago US, 2017). Trivago provided me with the following index and analysis; the most expensive and cheapest cities ranked by average hotel rates and trivago hotel price index (tHPI) which is a monthly index on current hotel prices in cities, regions and countries. The other index and analysis produced by Trivago can be listed as follows; trivago hotel price disparity index (tHPDI) which shows the percents of money that people can save in each destination if they compare prices with trivago, trivago Hotel Online Reputation Ranking and trivago Hotel Destination Trends (tHDT) which is an analysis about the destination trends for consumers worldwide based on hotel researches of trivago users. The answer to the question of how Trivago produces all these index and reports is given on the website as “Data is central to the success of our business” (trivago US, 2017).

### ***Metglobal***

In addition to the index by Trivago mentioned above, I obtained some reports and analysis from Metglobal. Metglobal, as the other contributor of this dissertation, is “a world leading technology powerhouse that aims to provide individuals and businesses with simplified experiences that improve productivity through multiple online effectiveness tools” (Metglobal.com, 2017). The company houses a variety of brands. These brands in their portfolio, HotelsPro, otel.com and tatil.com, are “established in online travel and travel related technologies allowing their partners to operate in over 205 countries and territories across the globe with an inventory of over 500.000 hotels and 20.000 destinations” (Metglobal.com, 2017). What makes Metglobal one of the contributors of this study is its data centric approach as it is clearly seen via the following statement; “At Metglobal, we make our decisions based on data, we love data! We have a big data ecosystem which helps all business teams make data driven decisions. With the help of our analytical capabilities and advanced dashboards, our partners can also have access to detailed sales monitoring tools” (Metglobal.com, 2017).



In addition to information provided on the company's website, Özlü used these words while explaining what big data does mean for Metglobal; "Big data means "power to understand clearly customer behavior and create opportunity to serve them better". Best way to provide better products and services is to understand customer needs and behaviors deeply; what they need, how they want to be treated, which type of customers need what" (via email 25 August 2017).

The aforementioned reports and analysis which shows how big data facilitates Metglobal activities in everyday use are the following ones; funnel report and search report. These reports show respectively, on an hourly basis, all the movement and the counts of user traffic including all steps from the search phase to booking phase, and the hotel, agency and supplier-based distribution of all search that Metglobal receive (Hakan Olgun, via email 19 October 2017).

### ***Online Interviews***

Sample index and reports from Trivago and Metglobal are supported with the online interviews conducted with professionals who work closely with data at Metglobal to provide deeper understanding on this matter. Interview questions were sent via e-mail and the respondents replied back through the same way. Subject matter experts with whom I conducted online qualitative interviews during August and September 2017 were:

1. Christina Leni, Global Public Relations Manager, Trivago
2. Nuri Özlü, Interim Managing Director, otel.com
3. Nezhil Yalabik, Data Science – Big Data Project Manager, Metglobal
4. Hakan Olgun, Data Science Team Lead, Metglobal
5. Seda Alacan, Data Scientist, Metglobal

The questions asked were about the current state of the use of big data, and the analytical strategy that their organization has been implementing, and significant challenges that they think big data projects in travel and tourism industry have been facing.

Conducting interviews online via email has its own advantages and disadvantages just like each approach and method in research has pros and cons. The most important advantage of conducting email interviews for this study was the independence of place and time. My sources were so occupied for a phone or in-person interview at a specific time especially during the extreme busyness of the summer season. As it is clearly stated in Opdenakker's work, "...busy interviewees do not have to identify a mutually convenient time to talk to each other" since it is based on voluntariness (2017, p. 9). Nevertheless, the prominent disadvantage of using email was that "...the chance of a spontaneous answer to a question is smaller, because the interviewee has more time to reflect on the question" (Opdenakker, 2017, p. 9). Anyway, this concern has a positive aspect as such it limits the potential problem of distractibility of interviewee. During a synchronous interview, external factors might cause distraction and accordingly might prevent participants from giving their full attention and being fully engaged with the interview. However, external factors are not that much important in online interviews since the participant can reply at any time.

## CHAPTER 3

### RESEARCH FINDINGS AND ANALYSIS

This research study discussed the adoption and the influences of big data in tourism industry. It investigated the impact of big data in terms of facilitating every day activities of organizations and making them stay globally competitive. To remind the research questions asked at the beginning of this paper, the main question and the sub-research questions were respectively the following ones: “How does big data affect tourism industry in terms of discovering and determining insights and of providing comprehensive results?”, “What are the advantages of using big data in tourism demand forecasting in terms of providing a competitive advantage for businesses?”, and “What are the problems that the tourism industry faces regarding big data in implementations of projects?”. Based on the findings through the conducted literature review, the answers given via online interviews and the analysis of the index and reports from Trivago and Metglobal, this study aimed to understand how big data is transforming tourism industry. The findings of this study are drawn from the research results which have been analyzed with a mixed method presented in the previous chapter and linked to the research questions.

#### **RQ1: How does big data affect tourism industry in terms of discovering and determining insights and of providing comprehensive results?**

“While data and analytics are revolutionizing our world, and changing the way we live and work, it is the insights that they provide that make them powerful” (Marr, 2017c). There is no doubt that big data does affect tourism industry as well as the world we live in terms of discovering and determining insights and of providing comprehensive results. And, it affects while “a lot of tourism data is sitting unused in data silos instead of being shared, compared and transformed into business insights” (Gretzel, n.d., p.1). The

following words of Yalabık exhibit the potential effect of big data; “Many sources from planets to microorganisms are creating immeasurable amount of data in our universe every second. If we consider uncollected data as wasted, in this wasted bulk many patterns, many records are hidden. The more we can analyze and process this bulk there is no doubt that we can understand and identify more” (via email 6 October 2017).

But why do people act like data have not been there and part of the impact of information technology up until now? To answer this question, it has been confirmed by Özlü that currently we have mature enough systems to govern and secure large amounts of data while empowering end users to analyze that data (via email 25 August 2017). A much more realistic answer to the same question comes from Olgun; “We can say yes to whether or not we have mature enough systems only for a limited capacity. However, this capacity enlarges day by day. As a result, we can functionally use big data for empowerment of end users but will be much more comprehensive in the future for sure” (via email 6 October 2017).

As it is repeated several times in this paper, “the tourism industry thrives on information” (Benckendorff, Sheldon and Fesenmaier, 2014, p. 2). And the vast amount of data waiting for turning into information can provide data driven evidence to make decisions based upon analysis rather than guess-work or past experience (Frederiksen, 2012, p.345). Although “tourism big data enriches the knowledge of tourism businesses’ target market and is very useful for analyzing the consumers’ demand for different tourism products and services” (Song and Liu, 2017, p. 16), it should also be noted once again that tourism industry cannot be merely driven by consumer behavior (Hakan Olgun, via email 6 October 2017). While the collection of the most suitable and relevant data is important, it should not be ignored that tourism is highly fragile to all external impacts and not solely driven by consumer behavior (Nezih Yalabık, via email 6 October 2017). Today, companies are able to gain insight into customers’ decision-making processes by keeping track of shopping patterns, purchasing behavior, recommendations and many other sales drivers (Song and Liu, 2017). Yalabık confirms the argument which says different types of data have different value propositions for different sectors while explaining the ways to mine maximum value from the rapidly-expanding data as follows; “determining needs

directly related to the desired insight areas, building close relations with business departments to discover the problems of the organization which can be solved via data centric solutions, and upkeep with the new technological advancements and practicing them continuously” (via email 25 August 2017). For example, as in the case of this research, different types of data that operators generally collect and analyze in tourism industry are user searches, web traffic, click heat maps, hotel prices, customer reviews, etc. (Nuri Özlü, via email 25 August 2017). The validity of this statement on the side of Trivago is expressed by Leni while explaining how they extract the data at Trivago with these words:

We aggregate prices from more than 200 booking websites and 1.4 billion visits per year. We save every search and query. That is why we are in a place to analyze in depth the travel behavior of online consumers worldwide. We do have a dedicated team, BI, that digs into the data, analyzes those and then hands it over to the Communications people that they share them with media and consumers. (via email 25 April 2017)

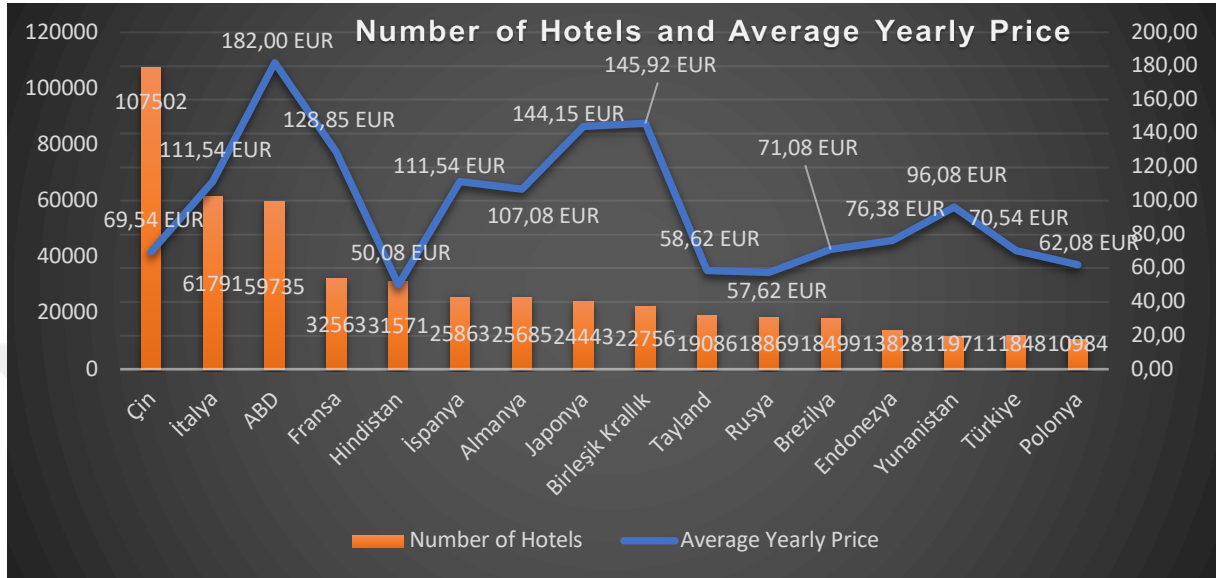
By this way, big data can disclose travel trends that help organizations make decisions about their marketing strategy (Thompson, 2016). In line with this explanation, Trivago provided me with the following index and analysis; the most expensive and cheapest cities ranked by average hotel rates and trivago hotel price index (tHPI) which is a monthly index on current hotel prices in cities, regions and countries.

Since visualization is an effective way to present results, based on a small segment of Trivago’s big data which shows price fluctuations and accommodation trends, I produced the following tables; Table 3.1, Table 3.2, Table 3.3, Table 3.4, Table 3.5 and Table 3.6, to make it easier interpreting the index shared.

The tables below, Table 3.1 and Table 3.2, prepared based on the highest number of hotels all around the world on a country basis and the price changes from April 2016 to April 2017. The highest number of hotels is seen in China whereas the lowest average yearly price change is seen in India. Italy and USA are the other countries that follow China in respect to number of hotels. It might be expected for the case that while number of hotels are increasing, average yearly price change should be in tendency to descent. However, Table 3.1 shows that there is no sign of a negative correlation. This might be the cause of the type of number of hotels enhancing the portfolio for the given countries. For instance,

the number of hotels in Greece is more than Turkey, however, the average yearly price change of Greece is greater than Turkey's.

**Table 3.1. Number of hotels and average yearly price (World-wide)**



The graph below shows the share of each country that subject the research. As supporting the data which is illustrated in Table 3.1, China, with 44.37%, is the country that has the highest share in total number of hotels. The total shares of China, Italy and USA are account to 60.32%, represents majority of the hotels.

**Table 3.2. Share in total number of hotels (World-wide)**

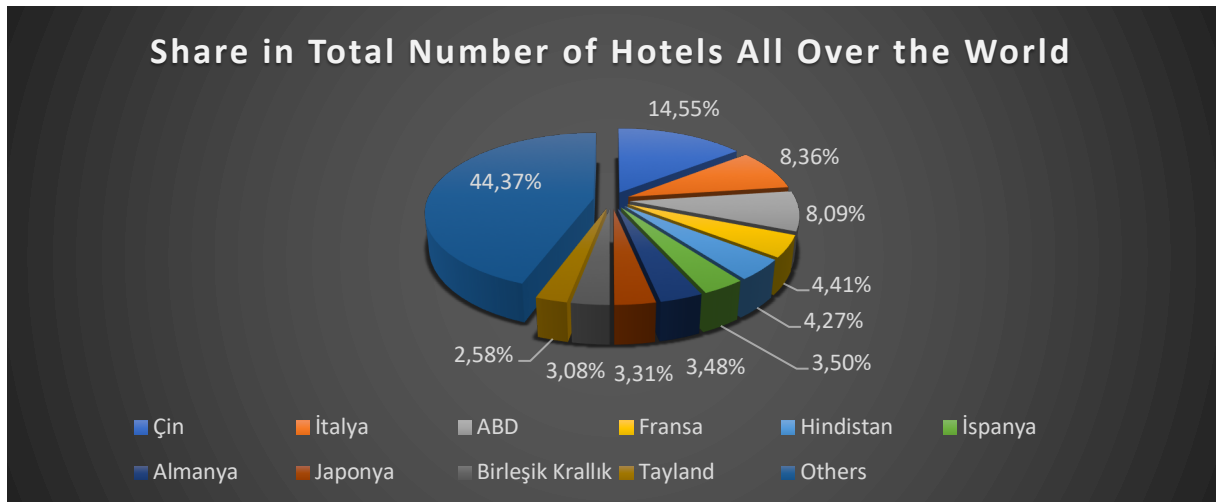
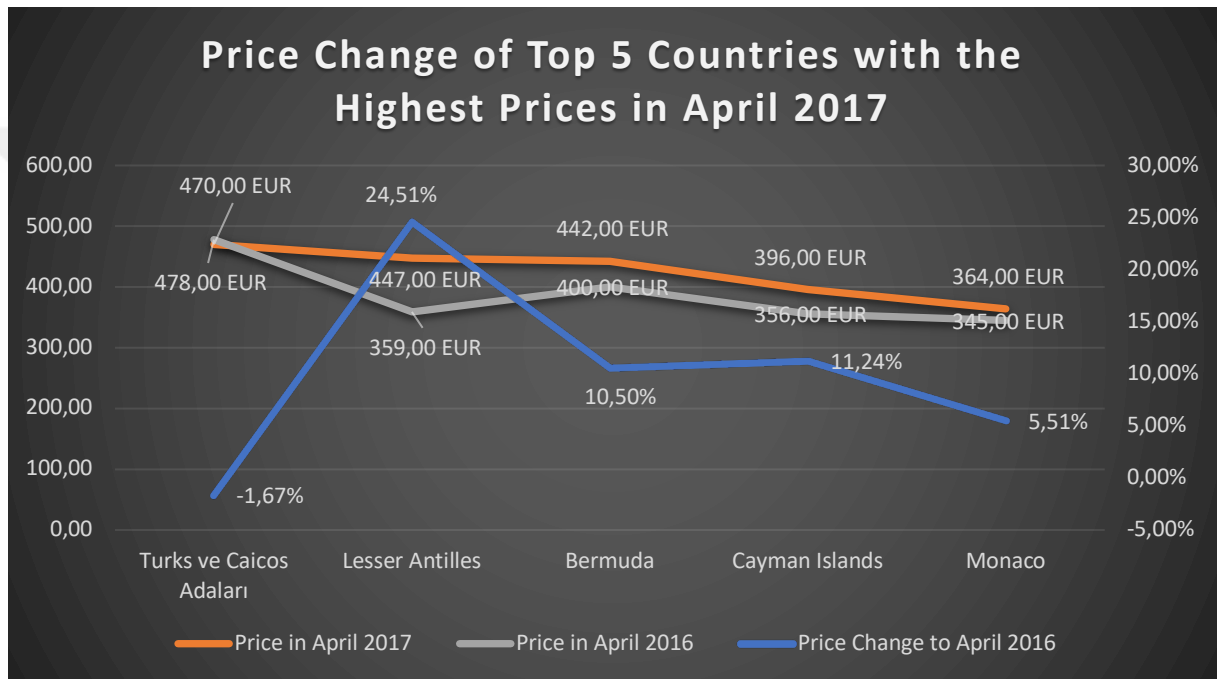


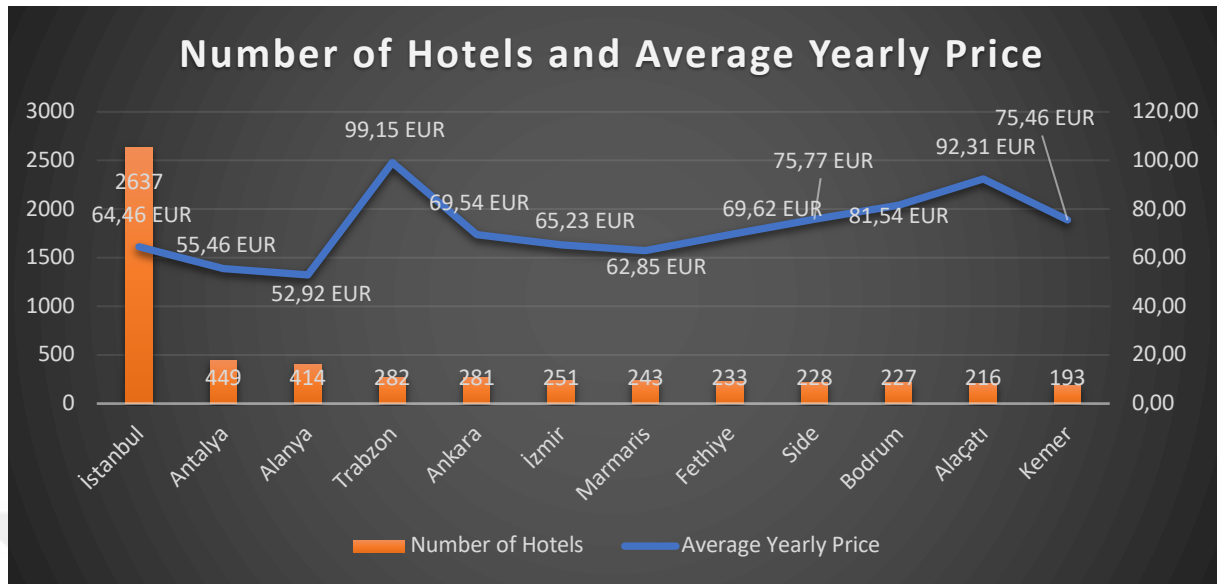
Table 3.3 indicates the price change of the top five countries with the highest prices in April 2017 compared to the prices in April 2016. Trends of price increases or decreases compare to previous years easily could be obtained by looking at analysis given in the table below. While the hotels in Lesser Antilles have more chance to make more profitable (24.51% increase in price) business in 2017, compare to 2016 this is not that much possible for the hotels in Turks and Caicos Islands where prices dropped 1.67%.

**Table 3.3. Price change over time (World-wide)**



And, the following tables are prepared by the same logic, namely they are showing the highest number of hotels in Turkey on a city basis and their price chances from April 2016 to April 2017. Another strong example of negative correlation has shown in Table 3.4 which shows number of hotels and average yearly price changes by city. İstanbul, Antalya and Alanya are cities which contribute most to the number of hotels in Turkey, compared to the other cities subject to the research. On the other hand, when average yearly price changes are carefully evaluated, it can be easily seen that Trabzon has the highest change with 99,15 Euro while this change account to 62,85 Euro for Marmaris. No matter number of hotels in declining tendency, average yearly price change might go up or diminish.

**Table 3.4. Number of hotels and average yearly price (Turkey)**



For centuries, Istanbul has been one of the most popular destinations for visitors which also has not changed during this study. As it is indicated in Table 3.5, 23.71% of Turkey’s hotels is in Istanbul which is followed by Antalya with the share of 4.04%. Alanya, county of Antalya, has more hotels than Ankara and İzmir.

**Table 3.5. Share in total number of hotels (Turkey)**

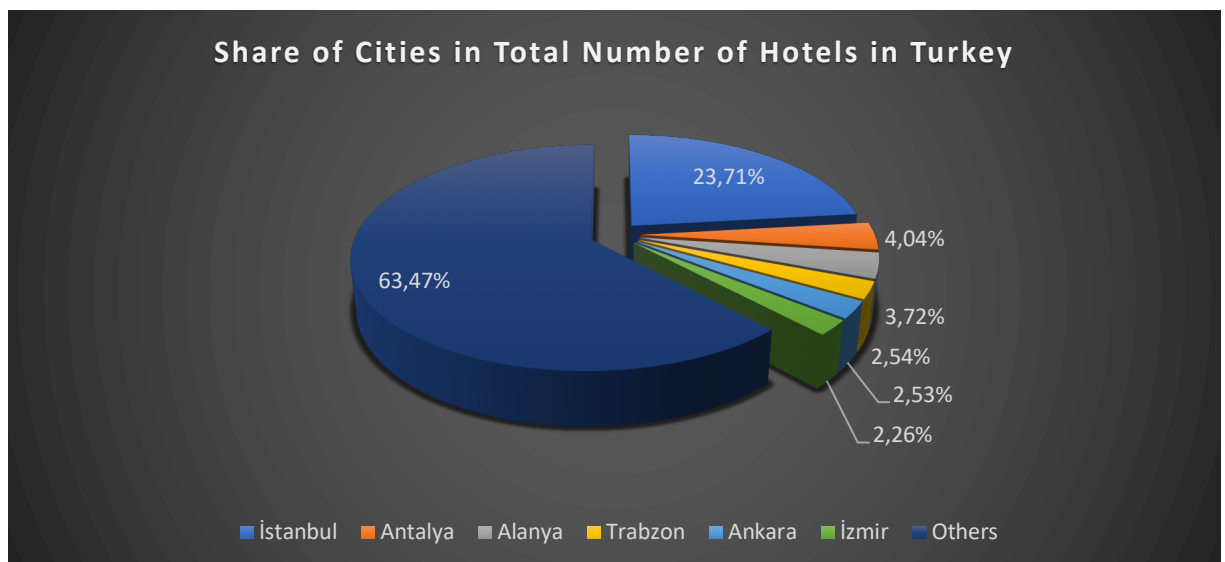
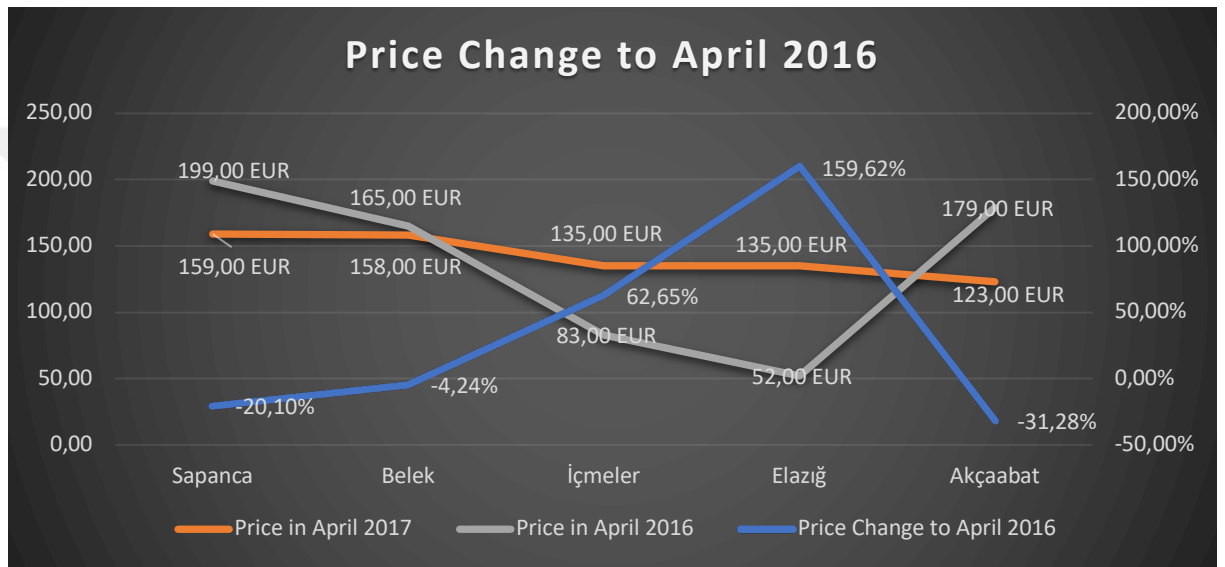


Table 3.6 indicates the price change of the top five cities in Turkey with the highest prices in April 2017 compared to the prices in April 2016. Tourism is one of the most vulnerable



industries in Turkey which easily could affect from any image damaging incident, such as political tension and geopolitical risks. This serves as basis to Belek case, one of the famous destination in Turkey which possessed to decline in the prices. In 2016, the hotels in Belek have made more profit than the year 2017, due to decrease of 4.24% in hotel prices. Such analysis enables executives and decision makers to conduct statistical predictions to take efficient measures.

**Table 3.6. Price change over time (Turkey)**



What makes Trivago capable of collecting the large amounts of data which holds insights needed for producing such index and many more like this is the global trend of using meta-search engines. “54% of Chinese, 36% of American and 35% of British travellers use meta-search engines to compare rates” (McCartan, 2017). It is because of the liberating way of meta-search engines by strategic alliances with key channels and then by “adopting semantic searches and ensuring near accurate customized price comparisons based on user preferences in an easy-on-the-eye display” (McCartan, 2017). It is often claimed throughout this paper that big data have an incredible capacity for providing answers to any question asked regarding people’s behaviors, views and feelings (d’Amore, Baggio and Valdani, 2015). Actually, these kinds of statements are more than just claims since to unravel a complex problem by combining all the available data is a great benefit for those who are able to exploit (Bedeley and Nemat, 2014). It is also confirmed by Olgun that

We definitely think that we get more efficient and positive results/feedback with the use big data compared to the previous technologies. Now we can clearly see how we had been restricted in the past while trying to improve our projects with data compressed only into tablature or relational databases. Thanks to Big Data we can process data including text files, visual files, and think outside the box. For example; thanks to text mining, we can extract customers' positive and negative comments on the hotels, and so we have the chance to suggest a hotel that attracts more attention to that user. (via email 6 October 2017)

A recent study conducted by TripAdvisor indicates that those efforts are not going down the drain by stating that “50% of their respondents agreed that hotel price comparison or meta-search saves time and helps to find the right price for their preferred hotel” (McCartan, 2017). Providing travelers with one-stop solutions is a reason for preference and at the end of the day it turns out big data everywhere. According to d'Amore, Baggio and Valdani, “when thousands of series of data are available, in fact, the probability to find a significant correlation between any two series, even if made of completely random numbers, can be as high as 90%” (2015). Then what? What kind of changes/improvements in service quality should be expected to observe after increasing the use of big data in a company? Alacan puts it this way; “Thanks to our big data applications, we can response consumer needs and requests much more accurate. We have improved our control over our marketing operations and we are dealing way more effective price discrepancy issues compared to previous years” (via email 6 October 2017).

Metglobal achieve these improvements via their data analytics division called collective intelligence consisted of three sub groups; data engineering team that collects data and create pipelines and storage architecture, data science team that creates data driven tools using machine learning methods and statistical analysis techniques, and data warehouse team that creates reports according to business needs and provide insights related to whole operation (Nezih Yalabık, via email 6 October 2017). At this juncture, it would be appropriate to share some reports provided by Metglobal which show how the use of big data facilitates their activities in everyday use.

Table 3.7 is a view from main dashboard of search report. It is a big data based work which shows the hotel, agency and supplier-based distribution of all search that Metglobal

receive, and the conversion rate of these search to booking. In general, dashboards “reflect actual last-quarter sales” but also “show what sales could be next quarter under a variety of different conditions - a new media mix, a price change, a larger sales team, even a major weather or sporting event” (LaValle et al., 2011, p. 28). In the case of Metglobal, the report is used for tasks including; grouping the hotels, regulating marketing expenditures, and investigating the size and the efficiency of the transactions among Metglobal, agency, and provider sides (Hakan Olgun, via email 19 October 2017). The values in Table 4.7 can be considered “backward looking and reveal what has already occurred” (Sivarajah et al., 2016, p. 275). By reading the results of this report; such as comparing conversion for hotel-based search against supplier-based conversion, and by analyzing the factors, Metglobal can comprehend the fundamental causes of different results on different channels and thus correlate different insights for future forecasting.

**Table 3.7. Search Report of Metglobal**

Date				
Search Date	Response Search	booking	Conv Rate	
15/10/2017	1.384.028.112	2.801	0,0002%	
16/10/2017	1.463.469.167	4.575	0,0003%	

Agency				Provider				
Consumercode	Agency Name	Response Search	booking	Search To Book	Providername	Response Search	booking	Conv Rate
AG-006276-NA	Super Coucou	7.434	11	0,1480%	Total	2.847.497.279	718.956	107,4915%
AG-006837-YD	Amoma Call Center & CUG (B2B)	233.632	7.184	3,0749%	Count	55	55	55
AG-007325-PY	Nikparvaz Talaei USD	329.406	543	0,1648%	expedia	1.475.173.615	1.103	0,0001%
AG-009458-MZ	BookMyRooms	383	4	1,0444%	hbeds	297.391.930	101.102	0,0340%
AG-009928-DZ	Viraxom Viagens	69.324	13	0,0188%	gta	167.407.971	74.262	0,0444%
AG-011723-KA	Canada One Travel	7.262	28	0,3856%	lotsofhotels	156.354.487	93.476	0,0598%
AG-013250-IO	VIKING XML	9.252	1	0,0108%	compass	83.451.785	39.684	0,0478%
AG-016042-IX	wuzhoumaryou	664.020	279	0,0420%	tourico	77.529.243	84.384	0,1088%
AG-020319-RO	Hotelskip.ir	103.720	312	0,3008%	miki	74.026.507	15.312	0,0207%
AG-020690-ES	LetsFly	1.129.753	3.969	0,3513%	hotusa	58.144.336	38.234	0,0658%
AG-021681-GS	B.D. Graiver	182.567	169	0,0926%	dotw	56.449.194	15.524	0,0275%
AG-021737-QL	Mid East USD	198.451	149	0,0751%	jacob	49.128.009	21.408	0,0436%
AG-022330-AD	Cancelon.com	411.780.464	53.660	0,0130%	getaroom	45.261.540	34.312	0,0758%
AG-022894-EA	Booked.net B2B Coral	240.026	170	0,0708%	smyrooms	39.670.879	11.812	0,0298%
AG-024373-PQ	TripSymphony_ILT	1.146	0,0000%		hbeds_apitude	36.715.454	31.895	0,0869%
AG-025682-OO	RKL Travel B2B	3.555	0,0000%			30.783.656	1.265	0,0041%

Hotel							
Hotelcode	Hotel Name	Destination	Response Search	Cache Response	NonCache Response	booking	Search To Book
Total			2.847.497.279	1.475.173.615	1.372.323.664	713.846	9088,9994%
Count			224059	224059	224059	224059	
107155	Wyndham New Yorker Hotel	New York	1.363.320	826.070	537.250	416	0,3035%
1071bf	Hotel Pennsylvania	New York	1.072.804	477.794	595.010	749	0,0698%
138542	The Manhattan at Times Square	New York	1.054.032	565.696	488.336	281	0,0267%
106022	Hyatt Regency Dubai	Dubai	1.046.286	501.146	545.140	248	0,0237%
107161	Wellington Hotel	New York	1.034.711	583.684	451.027	192	0,0186%
107078	Millennium Broadway Hotel Times Square	New York	1.016.088	572.792	443.296	144	0,0142%
154829	Conrad Dubai	Dubai	983.206	474.464	508.742	395	0,0402%
107311	Le Parker Meridien New York	New York	965.723	536.407	429.316	114	0,0118%
171627	Carlton Downtown (Formerly Warwick Dubai)	Dubai	953.286	454.359	498.927	150	0,0157%
139693	Ramada Deira	Dubai	939.801	468.798	471.003	58	0,0062%
107357	Park Lane Hotel	New York	922.809	468.314	454.495	65	0,0070%
16c348	Dorsett Shepherds Bush	London	922.049	451.088	470.961	215	0,0233%
137521	Westin New York Grand Central	New York	904.253	529.978	374.275	174	0,0192%
104202	Grand Plaza Serviced Apartments	London	902.785	473.161	429.624	63	0,0070%
189832	Hotel Edison	New York	892.675	488.299	404.376	141	0,0158%
106407	Marriott Dubai	Dubai	890.660	500.743	389.916	144	0,0163%

It can be argued that identifying the search trends of the users is a window of opportunity to decide for marketing investment by appealing to customers’ search behavior. Together with these kinds of reports, it is possible for businesses “to see their customers’ purchases, payments and interactions” and this enables businesses “to listen to customers’ unique

wants and needs about channel and product preferences” (LaValle et al., 2011, p. 28). The data compiled in such kind of a search report can be used to perform detailed analysis, so businesses can better understand their customers and customers’ purchase and/or booking journeys. Then they can test new solutions to make their business function better.

Table 3.8 below is part of the company's funnel report. Before looking at the report, it can be beneficial to explain what a funnel report is. “A funnel is a sequence of goals arranged in such a way as to reflect the actual flow of visitors which results in a major business objective being achieved” (Visual Website Optimizer, 2017). These business objectives can be any action of value to the business, however in this case the business objective is all steps from the search phase to booking phase. It might be better to point out the impact of funnel reports by saying “funnel reports are important for anyone who wants to pinpoint and act upon pitfalls in their processes” (Woopra, n.d.). What can be seen in Metglobal’s funnel report is all the movement and the counts of user traffic including all steps from the search phase to booking phase on an hourly basis. So, they become able to see in which step they lost most of the traffic on their website and to intervene accordingly (Hakan Olgun, via email 19 October 2017). And, if there is an unexpected difference below the average rates, they become able to check the possible reasons that may cause these unexpected results and to take corrective measures for improvement (Hakan Olgun, via email 19 October 2017). These reports are part of descriptive analytics method which defines “the current state of a business situation in a way that developments, patterns and exceptions become evident” (Sivarajah et al., 2016, p. 266). Table 3.8 reveals what is happening in the business, namely, the details of user traffic by indicating a different value in each column like search count, hotel detail page view, book count, etc. Thus, Metglobal can track the conversion rate of each individual business goal and one the mentioned goals can be, let’s say, a visitor’s completing a transaction like booking a room. This kind of reports are not only used for tracking conversions but also for understanding “the financial strength of a business at a given point of time” or for comparing “it with others or its own across different point of time” as it can be seen in the first two columns of Table 3.8 (Sivarajah et al., 2016, p. 275). It is vital to monitor these multiple processes across different times since Metglobal generates this kind of different metrics on a regular basis.

**Table 3.8. Funnel Report of Metglobal**

<b>Date (grand funnel)</b>	<b>Hour (grand funnel)</b>	<b>Search Count</b>	<b>Hotel Detail Page View</b>	<b>Book Form View</b>	<b>Hotel Detail To Book Form %</b>	<b>Book Form Submit</b>	<b>Book Form To Submit %</b>	<b>Book Count</b>
5.10.2017	0	525.424	869	113	13,00%	13	11,50%	8
	1	547.455	979	158	16,14%	8	5,06%	6
	2	452.145	949	176	18,55%	9	5,11%	7
	3	374.804	893	151	16,91%	12	7,95%	5
	4	394.489	870	153	17,59%	19	12,42%	14
	5	426.891	854	150	17,56%	7	4,67%	6
	6	536.109	880	113	12,84%	4	3,54%	2
	7	549.956	1.036	152	14,67%	11	7,24%	5
	8	548.832	1.185	151	12,74%	7	4,64%	5
	9	492.553	1.214	175	14,42%	16	9,14%	13
	10	547.827	1.389	182	13,10%	7	3,85%	4
	11	541.234	1.427	189	13,24%	14	7,41%	9
	12	572.852	0	0	Null	0	Null	0
	13	657.430	1.636	220	13,45%	14	6,36%	11
	14	703.922	1.684	258	15,32%	15	5,81%	11
	15	659.678	1.711	258	15,08%	19	7,36%	16
16	691.418	1.693	274	16,18%	15	5,47%	9	

	17	683.372	1.586	216	13,62%	13	6,02%	10
	18	721.571	1.777	240	13,51%	18	7,50%	14
	19	737.353	2.076	248	11,95%	20	8,06%	14
	20	671.807	1.846	243	13,16%	16	6,58%	12
	21	534.883	1.551	200	12,89%	16	8,00%	13
	22	551.453	0	0	Null	0	Null	0
	23	555.626	910	159	17,47%	9	5,66%	6
6.10.2017	0	499.268	922	144	15,62%	9	6,25%	6
	1	503.452	944	147	15,57%	6	4,08%	3
	2	460.520	1.050	177	16,86%	13	7,34%	10
	3	362.532	1.012	172	17,00%	8	4,65%	7
	4	363.056	910	151	16,59%	3	1,99%	1
	5	440.047	0	0	Null	0	Null	0
	6	355.882	593	94	15,85%	5	5,32%	4
	7	502.209	872	107	12,27%	6	5,61%	4
	8	560.974	903	113	12,51%	4	3,54%	1
	9	480.555	1.085	156	14,38%	10	6,41%	4
	11	535.285	1.264	171	13,53%	11	6,43%	7
	12	538.783	1.388	196	14,12%	8	4,08%	4
	13	603.377	1.488	221	14,85%	8	3,62%	5
	14	676.235	1.438	205	14,26%	8	3,90%	4
	17	642.438	1.421	196	13,79%	11	5,61%	8

	18	635.187	1.576	199	12,63%	10	5,03%	6
	19	605.801	1.661	202	12,16%	15	7,43%	13
	20	609.518	1.675	228	13,61%	12	5,26%	6
	21	497.923	1.293	149	11,52%	4	2,68%	3
	23	507.173	842	130	15,44%	11	8,46%	7
7.10.2017	0	438.842	814	112	13,76%	3	2,68%	1
	1	435.468	861	125	14,52%	7	5,60%	7
	3	359.895	832	107	12,86%	5	4,67%	4
	4	331.713	737	100	13,57%	8	8,00%	4
	5	431.329	797	123	15,43%	7	5,69%	7
	6	479.956	858	115	13,40%	6	5,22%	4
	9	491.881	1.167	151	12,94%	4	2,65%	3

So; to sum up, it is good to say over that “big data can reveal travel trends that help you make decisions about your marketing strategy” (Thompson, 2016). Thus, businesses can “deliver travelers world-class travel experiences with the most desired prices, personalized recommendations and content, globally expanding quality inventory and deep local understanding” (Metglobal.com, 2017).

**RQ2: What are the advantages of using big data in tourism demand forecasting in terms of providing a competitive advantage for businesses?**

In tourism literature, it is argued that tourism is “a complex social phenomenon where knowledge is the essential basis for tourism development and competitiveness” (Fuchs, Höpken and Lexhagen, 2014, p. 199). So, it can be argued that the value of an organization depends on the data within it. While data is that much important, Olgun

answers the question of “What difference do you think the big data will make in the tourism sector?” by saying;

Today, we think that the information is hidden in data itself rather than the written sources. We think that the increase in the amount of data together with the increase in the speed of collecting of data day by day brings considerably important changes along to the tourism industry as well as to all industries. Big data; can help in many business processes, such as building search engines for e-commerce websites, capturing what customers want on a reservation request, reaching the customer's goal and providing budgetary profitability, as well as bringing the transaction speed to significantly higher levels. (via email 6 October 2017)

Regarding the power of available data in terms of providing sustainable development and then of remaining competitive in the industry, it should be noted that small-scale startups can force big companies in the market by using big data, regardless of company size. The reason behind this state is related with the fact that what matters in terms of turning big data to an opportunity is the volume and operability of the data (Seda Alacan, via email 6 October 2017). Therefore, small-scale companies can actually achieve growth as long as they can give enough importance to analyzing the data they obtained. In short, the motto here is that the bigger your data, the better system works.

Companies which create and implement big data strategies effectively are able to stand to gain a competitive advantage. At this point, it can be argued that the opportunities to gain competitive advantage through the use of big data are diverse. As it is stated in the work of Hassani and Silva, “big data and predictive analysis goes hand in hand in the modern age with companies focusing on obtaining real time forecasts using the increasingly available data” (2015, p. 6). To make it clear, it should be mentioned that “predictive analytics is concerned with forecasting and statistical modelling to determine the future possibilities” (Sivarajah et al., 2016, p. 266). Alacan also pointed out that big data can be used for making predictions based on previous consumer behaviors and foresee development of new tourism trends while answering how the use of big data in tourism demand forecasting does provide a competitive advantage for companies (via email 6 October 2017). In other words, “organizations need to know what is happening now, what is likely to happen next and what actions should be taken to get the optimal results” (LaValle et al., 2011, p. 22). However, to carry these possible outcomes into effect and to unlock more insights by blending data, companies and organizations need



some sources such as; skilled workforce, extensive amount of data from both operational and financial sources, and technological infrastructure or technology infrastructure partners (Nezih Yalabık, via email 25 August 2017). So, what is important is the way businesses develop the necessary tools for exploiting the ever-increasing data for their own advantage. It has been pointed out throughout this paper that to get the most from big data is directly related with the infrastructure of an organization, skilled workforce and expertise (Nezih Yalabık, via email 6 October 2017).

It is often argued that the characteristics of the tourism industry, including “the variety of services it involves, the intangible and perishable nature of its many products, and the inseparable relationship between its production and consumption require the generation, storage, co-ordination, and analysis of information,” indicates the potential benefits from the fast evolution of big data related technologies (Pan and Yang, 2015, p. 43). To elaborate the benefits of big data for tourism industry, Song and Liu states that; “The bounty of tourism big data has the potential to deliver new and more highly informed inferences about human activity and behavior that will give the tourism industry a big boost and benefit not only customers but also those who participate in the tourism industry” (2017, p. 14).

However, this can be possible only with a correlational relationship since “[...] when merging a large amount of data from a wide variety of sources, one can gain insight on unexpected patterns that might not be otherwise disclosed by a limited number of conventional sources” (Pan and Yang, 2015, p. 44). These various sources of the tremendous amount of data that tourists generate are listed by Pan and Yang as “tourism website’s analytics data, a hotel mobile app’s log data, call center logs, the amount of foot traffic in the city, the sales records of travel services, search engine query volumes, social media mentions, location data from cell phones, GPS and photos, etc.” (2015, p. 44). The activities of the tourists in different stages of travel are gathered in the following behavioral framework which is connecting a tourist’s behavior and the different types of big data sources (Pan and Yang, 2015, p. 45):

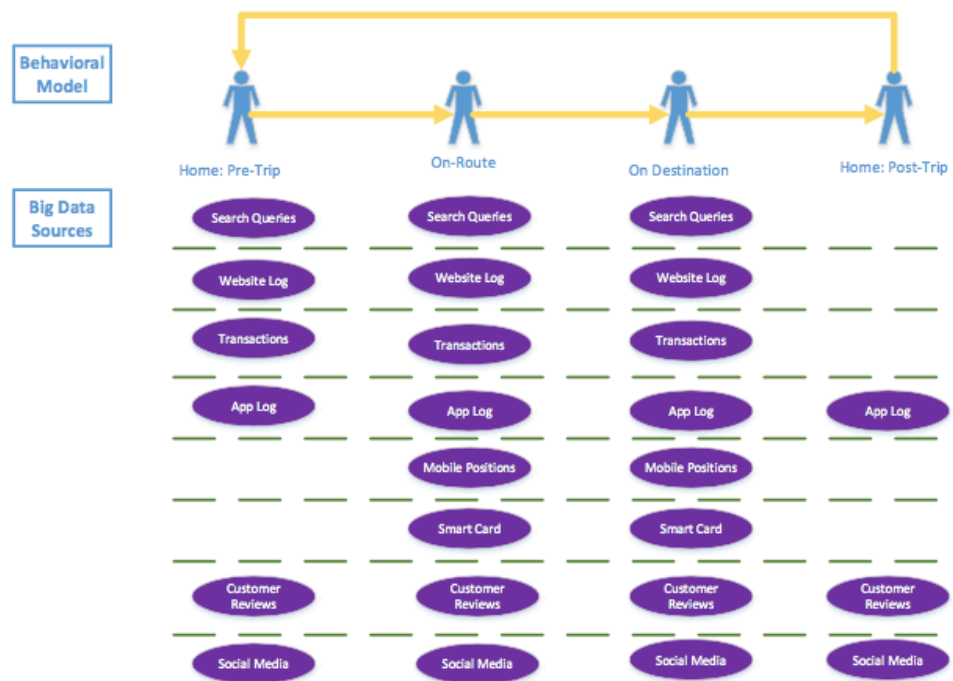


Figure 1. A Behavioral Model of Forecasting Tourist Behavior with Big Data

### Figure 3.1. Behavioral model of forecasting tourist behavior

This kind of an information will be useful in monitoring and forecasting tourists' activities since they indicate the tourists' potential choices by enabling businesses to anticipate what is likely to happen in the future. So, the tourism organizations which move quickly and become adapters of big data related projects can enjoy some opportunities including understanding tourist behavior, forecasting tourist activities, personalizing of service and improving customer experience, etc. (Pan and Yang, 2015). And, all these opportunities can end up with a competitive advantage.

### **RQ3: What are the problems that the tourism industry faces regarding big data in implementations of projects?**

It has been presented throughout this study regarding big data that it poses both opportunities and challenges for businesses. In the sections which focus mainly on the concerns and questions, it has been shown that the availability of big data itself is not

enough to bring problems to an end. It might be better to put it like this way by seeing big data as a “social, cultural, technological, and ethical phenomenon that is not all good, all bad, or constantly neutral” (Song and Liu, 2017, p. 26). Also, it should be noted that “the potentials of using big data are endless but restricted by the availability of technologies, tools and skills available for big data analytics” (Sivarajah et al., 2016, p. 265).

Together with the pervasive developments in data technologies, some concerns and challenges has emerged. In the section 2.5. Critique of Big Data, it has been worked through that despite the tremendous benefits of big data such as cost and risk reduction, improved efficiency and better decision-making process, there are some concerns and questions that need to be addressed such as the existence of the analytical talent, privacy, and security. While these challenges mentioned above are valid and applicable for all big data projects in all industries indiscriminately, by adding a few more Song and Liu clarifies the industry-specific challenges for tourism industry as follows;

The first challenge is the difficulty of identifying the right data and determining how to best use it. The second challenge is to find the right talent capable of both working with the new technologies and interpreting the data to find meaningful business insights, and the third is to overcome the obstacle of data access and connectivity, which requires the right platforms to aggregate and manage big data. The fourth problem is how to find new ways of leveraging big data. The final concern is the security of big data and how to keep the advantage of using such data. (2017, p. 26)

To overcome the challenges regarding the adoption of big data and to ensure the attainment of the possible gains and opportunities in tourism industry, what should be on top of the list of the organizations can be specified as; a deep understanding of big data and a proactive approach to take advantages of big data (Song and Liu, 2017). To be more specific; firstly, organizations need to have an overall understanding regarding “where the data come from, what audience will be consuming the data, and how that audience will interpret the information” (Song and Liu, 2017, p. 26). Secondly, since most part of the information in big data is real-time data, big data should be used timely for forecasting and decision-making process (Song and Liu, 2017). Olgun also proposed a solution as follows;

There is always the probability of failure for each and every project. In line with the needs of the company; data technology, budget constraints, and lack of in-house technical expertise can often lead to project failures for businesses. However, it possible for big data projects not to fail, but rather go beyond piloting and experimentation phases through a good statistical analysis process by using the existing technology. (via email 6 October 2017)

When it is analyzed in a more careful and neutral way, it can be seen that there is a number of issues regarding the big data phenomenon. “In absence of a very clear research objective and a likewise rigorous data collection plan, the risk of discovering meaningless effects or deceptive outcomes is quite high” (d'Amore, Baggio and Valdani, 2015). Yes, there are outstanding advantages of using big data in tourism industry, but to gain these advantages, organizations need “a particularly high care in the handling of any investigation that uses information coming from the vastly populated online world” (d'Amore, Baggio and Valdani, 2015). Therefore, it can be concluded by saying that the success of organizations depends on how they deal with the various challenges of big data such as the shortage of skilled workforce, data privacy, data security, and a few more which is not interrogated in a detailed way in this study. It can be achieved by using appropriate methods and approaches to capitalize the benefits of big data. The increasingly pervasive effect of big data and its value is unquestionable but businesses should come up with optimal ways while deriving advantages from big data.

## CONCLUSION

This research has been done to examine the outcomes of the use of big data as a competitive advantage tool in travel and tourism industry. In this regard, it has been outlined that the digital trace of this era that has been generated by people while using digital technologies is big data. Accordingly, it is noted that the access to ever-increasing volumes of data, and data scientist's ever-increasing technological ability to mine the data for meaningful insights are the driving forces of this new era. By looking at increasing interest in big data, it is argued that in the years to come this phenomenon will continue to grow even bigger. However, it is evident that there are some constraints that should be overcome to ensure the lucrative outcomes of big data. Those constraints are listed as the shortage of skilled data scientist, data privacy, and data security. And correspondingly, it is pointed out that the key for generating and though enhancing effective and fruitful outcomes with big data is to embrace those concerns and questions and then to develop better techniques with sufficient effort.

Today, it is crystal clear that everyone is of the same mind regarding the importance of data. Big data is a source of knowledge generation and thus innovative insights for companies when it is managed, processed and analyzed in an appropriate way. The reason behind this is the fact that data increases reliability compared to intuition because decisions based on intuition carry an important level of risk. Here, it can be mentioned the difference between traditional and big data analytics. While in big data, sources are both structured and unstructured including text, video and audio in addition to numbers and categories, it is structured in traditional data. Since it is relatively hard to perform queries and analysis with big data, this makes it excessively outsized to process by using traditional methods and creates the need of analytic tools that are consistent, vigorous and proficient of being automated. Additionally, to be more clear in terms of the size of data sets, it should be pointed out that when it comes to big data, we are talking about petabytes and exabytes of data whereas it is megabytes and gigabytes for traditional data. That's why, one of the biggest obstacles that any business might face when trying to get the most

from big data is to draw meaningful insights from such large stores of data, especially data which is poorly organized. This obstacle causes a difference in approach. Big data approach starts with harvesting and storing data and later looking for patterns with or without a specific question in mind. What comes next is to look deeper into these patterns and to analyze the data. At this stage, it is likely to find correlations which might prompt unexpected results and/or insights that can be used to optimize and improve the functions and operations of a business. Therefore, it can be concluded that big data analysis can be utilized in a much more in-depth way compared to traditional methods. Related with this note, the importance of the adoption of big data technologies in Turkey can be summed up as the following way. In order to place development of tourism industry on a sustainable course, some efforts need to be pursued and organizations can benefit from big data regarding the objective of being continuous optimization and improvement. The analysis of the large amounts of data available on Turkey's tourism industry can be used in such a manner to support new developments and studies aligned to global standards. The other countries with the same objective in their tourism strategy can follow the findings of such kind of a study for benchmarking as a tool to gain competitive insight evidence-based views of performance and increased organizational learning.

Although the lack of enhanced versions of reports limit the scope of this analysis, this research supported and enriched all the findings by previous studies. Based on the findings from interview with subject matter experts, and the collected datasets and reports of the contributor companies, the conclusion of this research can be summed up in the following way. Although big data is an emerging phenomenon, it includes a huge potential for tourism industry and its importance should not be underestimated by organizations performing in tourism industry. Together with the right approach, tourism organizations and the industry in general can learn a lot about the preferences of the consumers. Currently, tourism industry has already stepped on a path of data-driven approach to obtain competitiveness. With the digital traces of travelers left behind, it is quite possible to get valuable insights on practically any aspect of travel planning. The aforementioned lucrative outcomes of big data, which is intended through above-cited insights in tourism industry, can be listed as; much more accurate response to consumer needs and requests, improved control over marketing operations, and more affective price

discrepancy, etc. Since the level of analytics achieved by big data techniques are a lot more detailed than traditional models, these approaches put businesses on the path of predictive analytics. Predictive analytics can help businesses in anticipating what is likely to happen in the future. It makes this through defining the current state of a given business in the form of reports and dashboards as in the case of Metglobal. As a reminder, search and funnel reports shared by Metglobal were used to see the sources of all search that Metglobal receive and the details of the user traffic. By these kinds of reports businesses can comprehend the fundamental causes of different results on different channels and thus correlate different insights for future forecasting. Big data technologies progressively replaced mass marketing with commercial messages which are personal, unique and focused. By data clustering, it is possible to understand the exact borders of locations of interests and thus it is more likely to better understand the tourists' needs, preferences, behavior patterns, and habits. What big data technologies offer are not limited with predictive tools to forecast future trends, they also offer real time data to anticipate and respond to customers' needs immediately. However, one of the main outcomes of this research indicates that the way of gaining more opportunities is not only related with putting the technologies related with big data into use, but also related with the expertise of the personnel/data team. At this point, it should be better to mention once again that there are some challenges of working with big data including but not limited to the shortage of skilled data scientist, data privacy, and data security. Additionally, the evidences shared in previous sections show that together with some practices like "keeping up with the changing environment, expanding connections for information and experience share, and sparing more time to understand business needs better", businesses can consult big data to determine when and where to invest marketing campaigns (Nezih Yalabık, via email 6 October 2017). Since the power of big data is about "how you use the information you glean from the data you collect", the insights can revolutionize the way businesses function (Marr, 2017b).

Although the research has reached its aims, since certain limitations are inevitable in a research, it should be noted that this study also has limitation and readers should be aware of them and interpret the outcomes presented in this dissertation within the context of these limitations. The limitation of this study was related with one of the main sources of

this research, datasets and reports of the contributor companies. The resources that the researcher needed to have to do a more comprehensive work were enhanced versions of the collected datasets and reports. In this regard, the potential areas the researcher could not be able to explore caused a lack of access to a better instrument and/or relevant information. This limitation could not be fully overcome by using the method chosen to gather the data, the confidential company reports. Related with this explanation, it should be pointed out that the weaknesses and limitations of this research have indicated the following areas as recommendations for further work. Many different adaptations and experiments have been left for the future work due to lack of enhanced versions of the collected datasets and reports. Future study concerns deeper analysis of the mechanisms presented throughout this research, new proposals to try different methods, or simply curiosity. This thesis has been mainly focused on the influences of the use of big data. In addition to that, due to the relevance of the thesis subject and the questionnaire used in this thesis, it would be interesting to see the results from tourism industry in a couple of years to ascertain how much the adoption and influences of the use of big data has changed over time and to make comparisons.



## SOURCES

- Anderson, C. 2017, 'The End of Theory: The Data Deluge Makes the Scientific Method Obsolete', *Wired.com*, viewed 15 May 2017, <https://www.wired.com/2008/06/pb-theory/>.
- Ariker, M., Breuer, P. and McGuire, T. 2014, 'How to get the most from big data', *Digital McKinsey*, December, viewed 28 September 2017, <http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/how-to-get-the-most-from-big-data>.
- Bach, M., Schatten, M. and Marušić, Z. 2013, 'Data Mining Applications in Tourism: A Keyword Analysis', *Proceedings of the Central European Conference on Information and Intelligent Systems*, Varaždin, pp. 26-32.
- Benckendorff, P., Sheldon, P. and Fesenmaier, D. 2014, *Tourism information technology*. 2nd ed. Wallingford: CABI.
- Blake-Plock, S. 2017, 'Where's The Value In Big Data?', *Forbes.com*, viewed 25 November 2017, <https://www.forbes.com/sites/forbestechcouncil/2017/04/14/wheres-the-value-in-big-data/#1adcd60530da>.
- Bose, I. 2009. 'Data Mining in Tourism' in M. Khosrow-Pour, (ed.), *Encyclopedia of Information Science and Technology*, 2nd ed. Hershey: IGI Global, pp.936-940.
- Boyd, D. and Crawford, K. 2012, 'Critical Questions for Big Data', *Information, Communication & Society*, 15(5), pp.662-679.
- Buckey, C. 2017, *Intro to Data Analysis*, online video, viewed 2 October 2017, <https://classroom.udacity.com/courses/ud170/lessons/5430778793/concepts/53961386120923>.
- Buttler, P. 2017, *10 Challenges to Big Data Security and Privacy – Dataconomy*, 10 July, viewed 7 October 2017, <http://dataconomy.com/2017/07/10-challenges-big-data-security-privacy/>.
- Carr, D. 2013, 'Giving Viewers What They Want', *The New York Times*, 25 February, viewed 22 August 2017, <http://www.nytimes.com/2013/02/25/business/media/for-house-of-cards-using-big-data-to-guarantee-its-popularity.html>.
- Cumulus Media 2017, *What happens in an internet minute in 2017?*, image, viewed 22 September 2017, <https://www.weforum.org/agenda/2017/08/what-happens-in-an-internet-minute-in-2017>.
- Davis, P. 2013, *McKinsey Report Highlights the Impending Data Scientist Shortage*, viewed 4 October 2017, <https://content.pivotal.io/blog/mckinsey-report-highlights-the-impending-data-scientist-shortage>.
- d'Amore, M., Baggio, R. and Valdani, E. 2015, 'A practical approach to big data in tourism: a low cost Raspberry Pi cluster', *Proceedings of the 22nd International Conference on Information Technology and Travel&Tourism*, Lugano.
- Deighton, A. 2017, 'How To Unlock The Value In Your Company's Big Data', *Forbes.com*, viewed 25 November 2017, <https://www.forbes.com/sites/forbestechcouncil/2016/12/27/how-to-unlock-the-value-in-your-companys-big-data/#21db80505fba>.

- Ernst and Young 2014, *Big data Changing the way businesses compete and operate*, pp. 1-27, viewed 1 August 2017, [http://www.ey.com/Publication/vwLUAssets/EY\\_Big\\_data\\_changing\\_the\\_way\\_businesses\\_operate/%24FILE/EY-Insights-on-GRC-Big-data.pdf](http://www.ey.com/Publication/vwLUAssets/EY_Big_data_changing_the_way_businesses_operate/%24FILE/EY-Insights-on-GRC-Big-data.pdf).
- EU GDPR Portal 2017, *Frequently Asked Questions about the GDPR*, viewed 6 October 2017, <http://www.eugdpr.org/gdpr-faqs.html>.
- EU GDPR Portal 2017, *Key Changes with the General Data Protection Regulation*, viewed 6 October 2017, <http://www.eugdpr.org/key-changes.html>.
- Frederiksen, L. 2012, 'Big Data', *Public Services Quarterly*, 8(4), pp.345-349.
- Fuchs, M., Höpken, W. and Lexhagen, M. 2014, 'Big data analytics for knowledge generation in tourism destinations – A case from Sweden', *Journal of Destination Marketing & Management*, 3(4), pp.198-209.
- Google Trends 2017, image, viewed 15 May 2017, <https://trends.google.com/trends/explore?date=all&q=big%20data>.
- Gretzel, U. n.d., *Technology and Tourism: Building competitive digital capability*, viewed 10 October 2017, <https://www.aheadguest.com/assets/pdf/TechnologyTourism.pdf>.
- Hassani, H. and Silva, E. 2015, 'Forecasting with Big Data: A Review', *Annals of Data Science*, 2(1), pp.5-19.
- Introduction to Big Data - Module 1 - What is Big Data?* 2016, online video, viewed 29 September 2017, [https://courses.cognitiveclass.ai/courses/course-v1:BigDataUniversity+BD0101EN+2016\\_T2/courseware/bd64ccdf56ad4ea1afe870e26d583038/](https://courses.cognitiveclass.ai/courses/course-v1:BigDataUniversity+BD0101EN+2016_T2/courseware/bd64ccdf56ad4ea1afe870e26d583038/).
- Introduction to Big Data - Module 2 – Beyond the Hype* 2016, online video, viewed 29 September 2017, [https://courses.cognitiveclass.ai/courses/course-v1:BigDataUniversity+BD0101EN+2016\\_T2/courseware/bd64ccdf56ad4ea1afe870e26d583038/](https://courses.cognitiveclass.ai/courses/course-v1:BigDataUniversity+BD0101EN+2016_T2/courseware/bd64ccdf56ad4ea1afe870e26d583038/).
- Invest.gov.tr n.d., *Tourism - Invest in Turkey*, viewed 15 December 2017, <http://www.invest.gov.tr/en-US/sectors/Pages/WellnessAndTourism.aspx>.
- Isaca.org 2010, *Data Leak Prevention*, viewed 7 October 2017, [http://www.isaca.org/Knowledge-Center/Research/Documents/Data-Leak-Prevention\\_whp\\_Eng\\_0910.pdf?regnum=395889](http://www.isaca.org/Knowledge-Center/Research/Documents/Data-Leak-Prevention_whp_Eng_0910.pdf?regnum=395889).
- Isaca.org 2013, *Privacy & Big Data*, viewed 6 October 2017, [http://www.isaca.org/Knowledge-Center/Research/Documents/Privacy-and-Big-Data\\_whp\\_Eng\\_0813.pdf?regnum=395363](http://www.isaca.org/Knowledge-Center/Research/Documents/Privacy-and-Big-Data_whp_Eng_0813.pdf?regnum=395363).
- Jin, X., Wah, B., Cheng, X. and Wang, Y. 2015, 'Significance and Challenges of Big Data Research', *Big Data Research*, 2(2), pp.59-64.
- Kitchin, R. 2014, 'Big Data, new epistemologies and paradigm shifts', *Big Data & Society*, 1(1), p.1-12, viewed 22 November 2017, <http://journals.sagepub.com/doi/full/10.1177/2053951714528481>.
- Lafuente, G. 2014, 'Big Data Security - Challenges & Solutions', *MWR InfoSecurity*, viewed 7 October 2017, <https://www.mwrinfosecurity.com/our-thinking/big-data-security-challenges-and-solutions/>.
- LaValle, S., Lesser, E., Shockley, R., Hopkins, M. and Kruschwitz, N. 2011, 'Big Data, Analytics and the Path From Insights to Value', *MIT SLOAN MANAGEMENT REVIEW*, 52(2), pp.20-32.
- Lord, N. 2017, 'What is Data Loss Prevention (DLP)? A Definition of Data Loss Prevention', *Digital Guardian*, viewed 7 October 2017,

- <https://digitalguardian.com/blog/what-data-loss-prevention-dlp-definition-data-loss-prevention>.
- Marr, B. 2014, *Big Data: The 5 Vs Everyone Must Know*, viewed 24 November 2017, <https://www.linkedin.com/pulse/20140306073407-64875646-big-data-the-5-vs-everyone-must-know/>.
- Marr, B. n.d., *Whats Really Behind The Big Data Hype?*, viewed 24 November 2017, <https://www.bernardmarr.com/default.asp?contentID=967>.
- Marr, B. 2017, '3 Massive Big Data Problems Everyone Should Know About', *Forbes.com*, 15 June, viewed 5 October 2017, <https://www.forbes.com/sites/bernardmarr/2017/06/15/3-massive-big-data-problems-everyone-should-know-about/#48ba7d9a6186>.
- Marr, B. 2017, 'Why No Company Should Start A Big Data Project Right Now', *Forbes.com*, 17 August, viewed 15 October 2017, <https://www.forbes.com/sites/bernardmarr/2017/08/17/why-no-company-should-start-a-big-data-project-right-now/#3ddffae62d6>.
- Mazzocchi, F. 2015, 'Could Big Data be the end of theory in science?: A few remarks on the epistemology of data-driven science', *EMBO reports*, 16(10), pp.1250-1255.
- McCartan, M. 2017, *The Big Battle - Meta-Search vs OTAs - Hotel Insight*, viewed 18 October 2017, <http://pebbledesign.com/insights/the-big-battle-meta-search-vs-otas>.
- MCKinsey Global Institute, 2011, *Big data: The next frontier for innovation, competition, and productivity*, pp. 1-13, viewed 18 September 2017, file:///C:/Users/mkucuk01/Downloads/MGI\_big\_data\_full\_report.pdf.
- Metglobal | *Business Model*, 2017, viewed 23 October 2017, <http://www.metglobal.com/business-model.html>.
- Metglobal | *Overview*, 2017, viewed 29 September 2017, <http://www.metglobal.com/overview.html>.
- Naydenov, R., Dupre, L., Chalvatzi, E. and Skouloudi, C. 2015, *Big data security-Good Practices and Recommendations on the Security of Big Data Systems*, Heraklion, ENISA.
- Opendakker, R. 2006, 'Advantages and disadvantages of four interview techniques in qualitative research', *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 7(4), pp.1-13.
- Pan, B. and Yang, Y. 2015, 'Monitoring and Forecasting Tourist Activities with Big Data', in M. Uysal, Z. Schwartz and E. Sirakaya-Turk, (ed.), *Management Science in Hospitality and Tourism: Theory, Practice, and Applications*, 1st ed. Charleston: Apple Academic Press, pp.43-62.
- Parise, S., Iyer, B. and Vesset, D. 2012, 'Four strategies to capture and create value from big data', *Ivey Business Journal*, July/August, viewed 12 September 2017, <https://iveybusinessjournal.com/publication/four-strategies-to-capture-and-create-value-from-big-data/>.
- Prensky, M. 2009, 'H. Sapiens Digital: From Digital Immigrants and Digital Natives to Digital Wisdom', *Innovate: Journal of Online Education*, 5(3), pp.1-11.
- Russom, P. 2013, *9 Ways to Get Business Value from Big Data | Transforming Data with Intelligence*, 27 August, viewed 25 November 2017, <https://tdwi.org/articles/2013/08/27/get-business-value-from-big-data-9-ways.aspx>.

- Shafiee, S. and Ghatari, A. 2016, 'Big data in tourism industry', *Proceedings of the 2016 10th International Conference on e-Commerce in Developing Countries: with focus on e-Tourism (ECDC)*, IEEE, viewed 25 November 2017, <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7492979&isnumber=7492960>.
- Sivarajah, U., Kamal, M., Irani, Z. and Weerakkody, V. 2016, 'Critical analysis of Big Data challenges and analytical methods', *Journal of Business Research*, 70, pp.263-286.
- Song H., Liu H. 2017, 'Predicting Tourist Demand Using Big Data', in Xiang Z., Fesenmaier D. (ed.) *Analytics in Smart Tourism Design, Tourism on the Verge*, Springer, Cham, pp.13-29.
- Thompson, A. 2016, *How hotels can use big data to inform marketing strategy*, viewed 13 October 2017, <http://hotelmanager-blog.trivago.com/big-data-marketing-strategy/>.
- Travel&Tourism Economic Impact 2017 Turkey*, 2017, London: World Travel&Tourism Council, pp.1-20, viewed 15 December 2017, <https://www.wttc.org/-/media/files/reports/economic-impact-research/countries-2017/turkey2017.pdf>.
- About us*, 2017, viewed 29 September 2017, <http://company.trivago.com/about/>.
- UNWTO Tourism Highlights: 2017 Edition*, 2017, image, viewed 23 September 2017, <http://www.e-unwto.org/doi/book/10.18111/9789284419029>.
- What are Funnels?*, 2017, viewed 23 October 2017, <https://vwo.com/knowledge/what-are-funnels/>.
- Wolfert, S., Ge, L., Verdouw, C. and Bogaardt, M. 2017, 'Big Data in Smart Farming – A review', *Agricultural Systems*, 153, pp.69-80.
- Woopra, n.d. *Funnel Reports — Woopra*, viewed 23 October 2017, <https://www.woopra.com/docs/manual/funnels/>.
- Sustained growth in international tourism despite challenges* | *World Tourism Organization UNWTO*, 2017, 17 January, viewed 11 October 2017, <http://www2.unwto.org/press-release/2017-01-17/sustained-growth-international-tourism-despite-challenges>.

## APPENDICES



## **APPENDIX A**

### **INTERVIEW QUESTIONS**

1. Does big data really affect the world we live in terms of discovering and determining insights and of providing comprehensive results?
2. What kind of opportunities do people have for accurate analysis and decision-making by data-driven evidence?
3. How does the use of big data in tourism demand forecasting provide a competitive advantage for companies?
4. Do we have mature enough systems to govern and secure big data while empowering end users to analyze that data?
5. What does big data mean for Metglobal?
6. What type of data do you generally collect and analyze?
7. What sources do you have to unlock more insights by blending data?
8. How can you mine maximum value from your rapidly-expanding data?
9. What kind of changes/improvements in service quality have been observed after increasing the use of Big data in the company?
10. What kind of a data analytics team you already have?
11. Do you think you get more efficient and positive results/feedback with the use big data compared to the previous technologies? Can you give examples?
12. What difference do you think the big data will make in the tourism sector?
13. Do you think that small-scale startups will force big companies in the market by using big data, regardless of company size? Or before that, don't small scale companies rather have a disadvantage in terms of access to data?
14. Is it possible for big data projects to fail since they might not go beyond piloting and experimentation phases? What can be the obstacles? (lack of skill or expertise, and a mismatch between the technology strategy and overall company needs, etc.)
15. What can be the best practices that BI and analytics leaders can use to get advanced analytics initiatives off the ground?