UNIVERSITY OF LJUBLJANA
SCHOOL OF ECONOMICS AND BUSINESS

MASTER THESIS

POTENTIAL OF DATA ANALYTICS IN THE TOURISM INDUSTRY: THE CASE OF THE AOSTA VALLEY REGION

Ljubljana, March 2021 RAPHAEL CASTELNUOVO

AUTHORSHIP STATEMENT

The undersigned *Raphael Castelnuovo*, a student at the University of Ljubljana, School of Economics and Business (SEB LU), author of this written final work of studies with the title "*Potential of data analytics in the tourism industry: the case of the Aosta Valley Region"*, prepared under supervision of *Professor Jurij Jaklič* and co-supervision of *Roberto Henriques*

D E C L A R E

1. this written final work of studies to be based on the results of my own research;

2. the printed form of this written final work of studies to be identical to its electronic form;

3. the text of this written final work of studies to be language-edited and technically in adherence with the FELU's Technical Guidelines for Written Works, which means that I cited and / or quoted works and opinions of other authors in this written final work of studies in accordance with the FELU's Technical Guidelines for Written Works;

4. to be aware of the fact that plagiarism (in written or graphical form) is a criminal offence and can be prosecuted in accordance with the Criminal Code of the Republic of Slovenia;

5. to be aware of the consequences a proven plagiarism charge based on the this written final work could have for my status at the FELU in accordance with the relevant FELU Rules;

6. to have obtained all the necessary permits to use the data and works of other authors which are (in written or graphical form) referred to in this written final work of studies and to have clearly marked them;

7. to have acted in accordance with ethical principles during the preparation of this written final work of studies and to have, where necessary, obtained permission of the Ethics Committee;

8. my consent to use the electronic form of this written final work of studies for the detection of content similarity with other written works, using similarity detection software that is connected with the FELU Study Information System;

9. to transfer to the University of Ljubljana free of charge, non-exclusively, geographically and time-wise unlimited the right of saving this written final work of studies in the electronic form, the right of its reproduction, as well as the right of making this written final work of studies available to the public on the World Wide Web via the Repository of the University of Ljubljana;

10. my consent to publication of my personal data that are included in this written final work of studies and in this declaration, when this written final work of studies is published.

Ljubljana, April 24th, 2021 Author's signature: 

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

IoT – Internet of Things

BI – Business Intelligence

BA – Business Analytics

OSN – Online Social Networks

OTA – Online Travel Agency

CRM – Customer Relationship Management

SNA – Social Media Analytics

DMO – Destination Management Organisation

RFID – Radio Frequency Identification

DW – Data warehouse

ETL – Extract, Transform, Loas

TDLAB – Digital Tourism Laboratory

INTRODUCTION

In the era of the Internet of Things (IoT), data usage has become fundamental in any research field: we are witnessing and at the same time being the protagonists of a historical phase in which the accumulation of digital data is growing in such quantities that the only limit is given by the ability to represent them through physical phenomena, usually through a database. Digital data is an artifact that is at the same time technology, service, resource, representation of the world: those of mobile phones, the Internet of things and social networks, and analysis techniques are an inexhaustible source of data; they are born and spread with the promise of potentially representing every aspect of the world. This aspect concerns the change in the way in which, in the world of digital data, we perceive and know reality and its evolution. Digital data is manifested through various representations, of a linguistic type, such as tables or documents organized by means of a structure, or rather of a perceptual type, referring to our senses, such as photographs, images, videos, sounds. In this process, reality is progressively replaced by bits, numbers, encodings, systems of symbols, which often modify the perception of their meaning. The need to provide descriptive, interpretative and analytical tools and models for digital data, is a fundamental theme in Data Science, therefore translates into a need for literacy towards the entire population of a modern country, and for the training of data scientists, in high school and university cycles. The culture that is being created has computer science and statistics in its basic paradigms, but it concerns many other areas, from cognitive sciences to social sciences, economics, legal sciences, and finds application in a vast set of application domains, all those who use digital data. The so-called big data can be an important resource that creates value in society, in the economy, in scientific research. However, they require paradigm changes and the diffusion of a new culture oriented towards datacy, a cultural area of growing importance in all sectors of education and society.

Nowadays most of the tourism areas do not exploit or exploit the economic and social potential of data available on the internet only marginally: a person who is willing to go on a vacation, starts looking for destination on the web, hotels, restaurants, bars, attractions, and so on. One of the most famous study concerning this field is a case study from Barcelona, where they decided to start a project (Estela & Salvador Anton, October 2015) where they used an ad hoc software that collected, analysed, and generated BI almost in real time with Big Data collected from social media feeds, GPS signals, and data from government systems. In order to better understanding the potential of exploiting the data, there is a famous example of how data is being analysed by companies and how are they gaining benefits: Facebook. Millions of accesses every day, billions of chats message and likes, millions of photos uploaded make this social network knowing virtually everything about us. The meaning of all this process is to gather, memorise, analyse and sell to the market which will benefit in the short/medium term (Dimitrios & Aditya, 2015). This potential brings to light relationships between data and invisible phenomena in small volumes: the data comes from a multitude of sources, from social media (behaviours shown on Facebook, Instagram,

Twitter), analytics (through tracking online user behaviours), and transactions (purchases, bookings). Analysing this amount of information has become profitable in the retail and manufacturing world. Tourism has clearly an important role on pointing out where the market is heading, adapting the business strategies and offering a more and more user-friendly product. By reviewing and commenting on social networks, the likes of the services could be outlined, and the products offered by helping travel professionals could calibrate bids to suit their behaviours.

In this type of industry, analytical analysis can be used in particular to define already acquired customers, customise their travel experiences and their stays, create personalised and customized bids on their prospect tastes in order to increase their reach and also to generate user's key interest information which will be transmitted through the media. Consumers and tourism product providers would see all the benefits of using big data: personalised marketing and targeted product designs are extremely powerful opportunities for both groups. It is readily apparent that data can bring ad hoc special products and services targeted to customers (Pries, February 2015), for instance, big data analysts can capture information of consumer's main interests from the amount and kind of photos posted on Facebook or other social networks (e.g., a tourism provider could push information about local ski destinations or ski resorts when they obtain a picture contains a mountain landscape).

Because of this, one of the purposes of this thesis is to explain what the advantages of Data Analytics in Tourism are: by using the right approach, the tourism industry can learn about consumer preferences and can use this information and insight to build connections with individual travellers. It is clear that being able to offer travelers the right service or product at the right time is essential because advertising will only translate into conversions, while if a targeting strategy is in place, the results can be much better. We find ourselves in a period where the quantity and the flow of data regarding the behaviour of the consumers, their expectations, and the business models of companies created thanks to technologies are simultaneously interrupting old and consolidated models to give space to new and more sophisticated one (Liu & Haiyan, 2017). However, tourism analysis shows significant changes in the relationship between businesses and their customers so we could use them to provide superior buying and support experiences to enhance customer choice and expectations. The catalyst of recognising customers' behaviour is their massive use of phones, apps, and other social media, which are currently playing a key role in gathering raw data and distributing easy access to useful and relevant information. Data contains a lot of details about customers, some of which are quite obvious, while others may still need to be processed and requirements created by large data are essential in retailing because the use of business processes is guaranteed by new communication channels, unprecedented service delivery options and sources: the data can be called the new gold as in fact, data represents the new economic paradigm to start any type of company. The role of digital in tourism is constantly evolving and the ability to integrate, manage and analyze large amounts of data

from different sources, both internal and external to the perimeter of the organization, is essential both for the number of actors involved and for the heterogeneity of data released by tourists and visitors through specialized sites, social networks, blogs and communities. From these multiple touch points, it is possible to draw valuable information in terms of preferences, choices, opinions. Having these data and analyzing it by drawing business information means acquiring competitive advantages to improve the relationship with customers, develop new services, optimize internal processes. At all levels of governance, it is necessary to acquire and improve digital skills to attract and manage tourist flows and ensure an excellent and lasting experience. This series of data coming from multiple sources and grouped together, are called Big Data, that means a set of datasets that are extremely large and complex that traditional analysis processes and commonly used software are inadequate to capture, analyse, manage and process in a reasonable amount of time (Matzat, 2012). There are many studies and problems concerning the analysis, research, archiving, transfer, privacy of Big Data, for instance "Big data case study - Netflix". Netflix is the American most loved entertainment company specialised in streaming video on demand that can predict what exactly its customers would enjoy watching. By definition, Big Data analytics is the fuel that ignites the "recommendation engine" designed to serve this purpose and more recently, it has begun to position itself as a content creator and not just a distribution method, thanks to the multitude of data it owns. Netflix's recommendation engines and new content decisions are powered by data points, in particular the titles customers watch, how often playback is stopped and ratings. This way it' s possible and easier for companies to know without making assumptions but founding their decisions on Big Data.

The problem on which I will focus my attention on this paper is how data analytics could be used in the tourism market to provide benefits to potential business owners and public body working in a specific area, in this case, the Region of Aosta Valley (Italy) and how the same approach could be used in any touristic context. In the following paper, a representation of the current state is made trying to highlight focus on what are the problems related to the lack of information and which solutions could be used; by using business intelligence tools, it has been possible to access and analyse data sets and to present analytical findings with details about information that was not previously known. The lack of data is described and the steps to obtain them are shown. For this reason, a research question has been made and some goals had to be reached, as following:

***The Research question is*:** what is the potential of using data analytics in the tourism industry?

***And the goals are*:**

1. Collection and cleaning the available public data from the Aosta Valley Region offices
2. Creation of a database with statistical arrivals and destination of tourists in the years 2014 - 2019

3. Targeting the classified tourists depending on the year using BI tools
4. Definition of the missing data for a 360-degree study and empirical hypothesis to show a complete analysis
5. Make a Social Network Analysis and the usefulness of its use

In recent years, with the advent of new technologies and the proliferation of information that is left on the web by individuals, companies, public bodies, it has been possible to think that these could be exploited: the goal is to obtain economic and social benefits both for the users who uses the services, and for those who can analyze their data. Once this point is made clear, it will be presented, from a technical point of view, how the public data is recovered in the offices in charge of the Region: these has been then stored, saved, and analyzed.

The first chapter has been written to outline how the tourist market has evolved over the years and specifically in the last decades in Italy, what are the big data, how they are born, their characteristics, how these will be always used more frequently by companies and public bodies. Following this, the advent of the internet and web 2.0 has been considered into account and the impact that these new technologies had on tourism markets was addressed. A sub-chapter has been created concerning the collection, targeting, real time use and predictive use of the collected data in this work; the use of business intelligence tools, what they are and what is their connection to big data are essential to carry out a study as complete as possible. The last part of the theoretical chapter addresses the theory of sentiment analysis within social media: what are they, how are they used, and what are their objectives and characteristics.

The second chapter analyses the Research Methodology: considering as a case study the Region of Aosta Valley (Italy), it explains how the data was collected yearly and broken down according to the type of accommodation in 7 touristic areas and how the database was created: the starting point were various statistical tables collecting arrivals and presences, the number of days of the stays, in the different areas of the Region; the usage of a business intelligence software, has allowed to exploit more information than it would have been possible seeing in the initial tables only. During the gathering phase, only raw data was available from the touristic offices, so the paper analyses a statistical aggregation with the aim of finding additional information compared with the initial data. After this, the tools used for the visualization of the data and how these can give more information with a real prospective for predictive analysis is proposed. Finally, a Social Media Analytics methodology was carried out which consists in the collection, integration and analysis of data from the social networks of a real business profile that works in the Region. The paper presents the use of tools applied to the analysis of Facebook and Instagram profiles of a small boutique hotel (called Maison La Saxe) located in the Mont Blanc area: the objectives are showing the potential of social networks from an analytical point of view, how targeting can be used in this context to attract more customers and how the social network analysis model can be also replicated in other turistic areas. Through the data generated by social networks, it was possible to carry out in-depth analyzes within respecting the public's behaviour and

his performance and online positioning of individual profiles, characters, brands, companies, as well as studying the debate existing on the net on specific issues and topics. This allows very precise targeting and truly data-driven content creation and marketing campaigns.

The third chapter illustrates the case of analysis taken into consideration: Aosta Valley. It's the smallest Italian Region located north-west of the peninsula, on the border with France and Switzerland, in the Alps. The economy of Aosta Valley is mainly based on the tertiary sector, in particular on tourism which, thanks to to the autonomy conferred by a special statute, plays a role of primary importance in economy of the Aosta Valley. As a touristic vocation the sectors involves thousands of commercial establishments such as hotels, restaurants, bars, shops. For this particular geographic peculiarity, the case study has the aim to understand the behavior of the tourists who choose this alpine Region for their vacation by predicting possible and further changes in market demands.

The fourth chapter describes the case analysis containing the design and implementation of the data extraction and analysis with business intelligence tools. The specifications of the project are described by the technologies used for the combination, storage and visualisation of the listed data and the final dataset created is described. After this phase of segmentation and implementation from a mere data point of view, it is shown to what extent the available statistical data could be used for research. The last part of the chapter instead deals with the study of social network analysis exclusively for the real and specific case of a small boutique hotel located in one of the 7 areas: the Mont Blanc. In this phase, they have been analyzed the characteristics and the information extracted from the social networks Facebook and Instagram and their usefulness.

The fifth and last chapter describes the analysis carried out in the fourth chapter and it graphically represents the specific results. Furthermore, the weaknesses of the work, along with the human and personal limitations of the study, are pointed out with a particular focus on the data that any public body should gather to have a 360 degrees analysis, especially credit card transactions, gps tracking, transport information. The best-case scenario is also considered: all the missing data in the previous model would be hypothetically retrieved and the potentials of how this information could change the way of tourism in the future will be.

1 THEORETHICAL BACKGROUND

Data analysis is the science of analyzing raw data in order to draw conclusions about that information (Yangyong & Yun). Many of the analysis techniques and processes have been automated into mechanical processes and algorithms that work on raw data for human consumption whose data analysis techniques can reveal trends and metrics that would otherwise be lost in the mass of information. This information can then be used to optimize processes to increase the overall efficiency of a company or system.

Data analytics is a broad term that encompasses many different types of data analysis - any type of information can be subjected to analytical techniques to obtain information that can be used to improve things such as, for example, manufacturing companies that register the runtime, downtime and work queue for various machines and analyze data to better plan workloads so that the machines operate closer to maximum capacity. However, data analysis can do much more than highlight production bottlenecks (Frankenfield, 2020). The gaming companies, thanks to the use of data analysis, apply reward mechanisms to the players in such a way as to always maintain concentration during the game; companies that study human interaction base and organize content on it, in such a way as to induce viewing a certain area or clicking.

1.1 Evolution of the Touristic Market in Italy

In Italy, since the beginning of the twentieth century, tourism has been set up: in 1919 ENIT, the National Tourist Board, was founded. The following year it will help to found CIT - Italian Tourism Company, created to promote Italy abroad (Laney, 2001).

With the end of the war and the increase in people's mobility there is a strong development of tourism all over the world: Italy is starting to be considered an important destination for foreigners, who prefer seaside and mountain resorts. In this period UNWTO, World Tourism Organisation, begins the annual publication of data on tourism in the world (ISTAT, Annuario Statistico Regionale, s.d.).

In the '70s tourism in Italy is identified with the holiday: stays are designed as luxury and have an average duration of 20/30 days while the flows are highly seasonal and concentrated in the summer. The favorite destinations are sea (58.6%), art cities (16.4%) and mountains / hills (11.4%). According to UNWTO data, in 1970 Italy was the first country in the world for tourists: between 1960-1975 the number of visitors in our country rose from 132.336 mln to 291.780 mln, thanks above all to foreigners. In the history of modern tourism, the '80s mark the transition from holiday to organised holiday. The total duration of the holiday does not decrease, but is spread throughout the year, thus varying the seasonal nature of the flows. In Italy there is a fluctuating trend in the arrivals of foreigners, but there is also a remarkable growth of the internal market: between 1975-85, presences go from 291.780 million to 337.402.732 million thanks to the Italians. It is in the 90s that tourism becomes an indispensable asset, becoming truly a mass phenomenon on a global level. The motivations of travel begin to diversify, and the concept of "tourism" is affirmed. According to UNWTO data, Europe continues to be the first destination: in 1995 arrivals reached 310.8 million but the other destinations also started to increase significantly. In these years Italy loses the podium of the most desired destinations, finishing in 4th place (today occupies the 5th). Since the 2000s, tourism has continued to diversify, while the trend of self-organised travel has established itself, thanks to the arrival of the Internet and the phenomenon of holiday

homes. The decline in spending capacity worldwide has changed the type of holiday, with a decrease in time spent and the emergence of low cost offers, both in transport and in stays.

Over the past few years, technological advances have led to an immeasurable increase in the material present in the datasets, so much so that, more and more frequently, the term Big Data is used to describe this enormous amount of data: the concept refers to a large data set, even unstructured and fast moving, such that it cannot be managed with traditional approaches. The amount of data generated today, actually, abnormal: from mobile phones to credit cards used for purchases, from television to storage necessary for computer applications, from intelligent infrastructures in cities to sensors mounted on buildings, on public and private means of transport and so on. The first organizations that began to process similar types of data were those online: companies such as Google, eBay, LinkedIn and Facebook were built, from the beginning, around Big Data, which for them represented not an end as much as, rather, an important tool. In all likelihood, this happened because, dealing with a large amount of data in less structured formats, they had to adapt to the use of new technologies and attempt innovative management approaches.

1.2 Data Analytics, Business Intelligence & Big Data

The term "Data Science" was first introduced in 1974 by the Danish computer scientist Peter Naur in his book Concise Survey of Computer Methods as an evolution of the concept of datalogy used by Naur himself a few years earlier to contrast it with the more limiting concept of computer technology. In this first definition, Naur understands data science as a discipline relating to the management and manipulation of data as it is presented, placing little emphasis on the possibility of extracting valuable information from the data itself. With the advent of big data and the idea of "data value" typical of this paradigm, the very concept of Data Science has evolved, thus becoming a holistic science, whose founding principle is not mere data management, but a wider exploitation of the large heterogeneous amount of data coming from different sources. which should therefore be understood as a transversal discipline, which includes both the spheres of computer science, statistics and mathematics, as in the original meaning, and a set of more managerial skills, linked to the more recent need to know how to read, interpret and capitalize on data for business purposes.

The terms "Business intelligence" (BI) and "Data Analysis" (DA) are often used by companies and organizations in an exchangeable way, but in reality, there is a big distinction between the two: on the one hand we have BI which through of the processes of extraction, transformation and loading, collect data from an unlimited number of sources and subsequently organize and centralize them in a single repository whose goal is to look at historical data to describe past actions. It refers to technologies, processes, software, and techniques for the collection, integration and presentation of information obtained in such a way as to support business decisioners. The Data Analysis uses has the purpose of predicting the actions that will happen or that may happen in the future. To have a clear idea between

the two types of analysis, it is more correct to discuss what you want to achieve as there are three categories of analysis processes:

- *Descriptive analysis* involves transforming raw data into a form that makes it easy to understand and interpret, rearrange, sort, and manipulate to generate useful information. It is a preliminary phase of data processing that creates a summary of historical data to provide knowledge and almost always prepares the data for further analysis. Some examples could be the sale of products, customer relationships, operations carried out.
- *Predictive analytics* is the practice of extracting information from existing data sets in order to determine patterns and predict future outcomes and trends. It is closely linked to data mining and machine learning, as it uses data models to make predictions, where machines acquire historical and current information and apply them to a predictive model, stating that a certain event has a certain probability of happening. Examples include product sales forecasts or suggestions from retailers about what you might want to read, view or buy.
- *Prescriptive analysis* facilitates users to "prescribe" various possible actions to implement and guide the activity towards a solution and attempts to quantify the effect of future decisions to recommend possible results before they are actually adopted. It not only predicts what will happen, but also explains why it will happen and provides recommendations on actions that exploit these predictions.

The distinction between BI and DA at this point is clearer, in which all descriptive analyzes are part of Business Intelligence, in predictive analysis both BI and DA are included in which company management thanks to data visualization software, as for example Tableau, Microsoft Powe BI, Looker, can design their own reports and generally only the DA is included in the prescriptive analysis. Data science requires skills in both mathematics and statistics where scholars acquire large data sets to which they apply algorithms to organize and model them in such a way that they can be subsequently used for predictive studies. Data Analytics relies on algorithms, simulations and quantitative analysis to determine relationships between data that are not evident on the surface which does not happen with BI: rather than answering questions about what happened, data analytics tries to understand why things happened. Business intelligence also studies and analyzes processes in real time in order to help business managers achieve their business goals, while business analytics supports the change of actions and processes that companies implement. The two analysis models require prior data preparation, in which a phase of collection, cleaning, classification and targeting is required. Once the cleaning phase has been carried out, the data is stored in one or more archives that have characteristics such as to allow the execution of reports: this container is called the data warehouse. Nowadays, more and more efforts have been made to decentralize servers in such a way as to have a cloud infrastructure that allows scalability, i.e. the ability of a system to increase its performance if new resources, both software and hardware, are provided to that system. The data stored in this way therefore represent the unique starting point for business reports whether they are studied through Business

Intelligence or with Data Analysis. Both BI and data analytics require a data warehouse-based analytics stack, with data piped through an ETL tool, acronym of Extract, Transform and Load, the steps of reprocessing and storing the data are called transformation and loading respectively: together with the extraction what is called ETL by the acronym Extract, Transform and Load. The re-processing of data must follow precise business rules, rules that the company has adopted over the years.

Having said that, it can be said that there is not such a clear difference between Business Intelligence Business Analytics: the purpose of both is to provide knowledge to solve problems, both short and long term, where however BI allows, thanks to very precise rules to define unequivocally what happened in the past and a nice short period allowing us to hypothesize the trend in the future. The BA, on the other hand, looks for the reasons why certain choices and strategies have been positive or negative and also tries to hypothesize future scenarios.

Nowadays, data is everywhere and is part of our daily life in more ways than most of us realize and its existing digital quantity, which we create with every action we take, is growing exponentially. The study of data, such as BI and DA is increasingly connected to the amount of data that can be recovered: these three terms are often interchanged frequently in all sectors but, although they have some similarities to each other, the meanings are different. As mentioned above, data science that concerns the processing of unstructured and structured data is a field that includes everything that contemplates the cleaning, preparation and analysis of data: it is the sum of statistics, mathematics, programming, logical resolution of problems, acquisition of information different from the standard and which requires particular characteristics, cleaning, preparation and aggregation of data. This generic term therefore includes various techniques used to extract insights and information from raw data. On the other hand, there are big data, significant volumes of data that cannot be processed effectively with traditional applications: their study begins with raw data, generally not aggregated and for which it is often impossible to archive all internal of a single terminal. Gartner provides the following definition of big data: "Big data is high-volume, high-speed, or wide-variety information assets that require innovative forms of cost-effective information processing that enable better understanding, decision-making and process automation" (Gartner).

1.3 Advantages of Data Analysis in Touristic Market

The revolution that the use of data represents for each sector is perhaps even more significant for the tourism industry. In fact, among the companies that first used this resource, there are mainly airlines. For example, British Airways has decided to invest in the deep knowledge of its customers by collecting online and offline information from loyalty programs to combat competition. In this way it is possible to understand the most frequent needs and problems of travellers and develop more effective proposals and solutions. Other companies,

such as Swiss Air, Air France-KLM and Lufthansa, use data to improve revenue management strategies, and even several hotel chains have started implementing operations based on the use of Big Data. Hilton, for example, introduced the use of a Balanced Scorecard to understand what factors drive organisational performance. Thanks to this activity, he was able to identify the correlations between the degree of client satisfaction and their behaviour. Some hotels, on the other hand, use stem platforms that can favour data analytics algorithms that continuously analyse the building and the use of electricity, efficient energy management, coming to reduce costs by at least 10-15%. Even large OTAs (online travel agencies) do not overlook this aspect: Expedia, for example, is making significant investments in this area, considered the keystone for the future of the trip.

The potential for tourism, therefore, is particularly great: to collect, homogenise, extrapolate and correctly interpret the data set representing the 'trace # of behaviour', choices and also the "sentiment" of tourists that will be based not only on the spontaneous comments of travellers on review platforms (an inevitably restricted pool that no longer reflects the "true average tourist"), but oriented on the analysis of further unconditional data such as habits and lifestyles, preferences, the real flows of tourism information. Furthermore, data for tourism offers important information not only on collective behaviour, but also on the relationship between places, things and people. According to what recently confirmed by the TDLAB (Digital Tourism Laboratory), people's daily behaviours are always characterised by some form of digital intermediation that, in fact, feeds huge data streams: Big Data. When analysed with more complex algorithms, these articulated and diversified data make it possible to substantially implement the decision-making processes of tourism companies, but also to improve the offer by adequately responding to the complexity of the demand.

The Online Social Networks (OSN), for example, are not only a powerful tool for the promotion and marketing of tourist offers, but given their incredible diffusion, they are also (and perhaps above all) an extraordinary source of 'information on tourists' preferences, on their activities, or on how they give value to what is offered to them. The resource represented by the analysis of all the online platforms that host users, comments and discussions on their travel impressions, or the implicit investigation of the 'traces' that they leave during their holidays. Travellers are, in fact, more and more social and digital: 91% of the users who have access to the Internet, have booked online at least one product or service in the last 12 months and use the engines research as the main source through which to search or plan a holiday, 42% use a mobile device (smartphone, tablet, etc.) to plan, book, inquire (33% in 2012), 68% search online before deciding location and mode of his journey. Not only: the use of the internet is essential for the tourist in the phase of inspiration (61% is informed through the Internet), but especially in the planning phase (80% use the Internet) and in the fruition, once at destination (58% use online sources to evaluate activities and services, while 40% directly create new content and share it).

Moreover, thanks also to the information available online, it is possible to provide a more accurate evaluation of the actual consistency of the tourist flows, analysing their activity in

the space of the social media. It is not difficult to understand that this is an opportunity of crucial importance for example, that today, to measure the tourist flow of a country, we still rely on a "traditional" count, or the number of visitors hosted by "classic" accommodation facilities, while the phenomenon of alternative accommodation (private houses, couchsurfing, farms, religious establishments, etc.) is developing rapidly, and this causes a substantial part of tourists who choose these alternative accommodation facilities is not properly accounted for in Italy. Since the statistics on the tourism trend are made available with months of delay, having them available in almost real time would allow to act on time, to find corrective measures, to study the historical series and to know what it could happen in the immediate future. Tourists, like everyone, are aware and unaware producers of Big Data and digital traces: a structured analysis of these data could therefore represent a very useful predictive tool.

Obviously, the analytical practice is not free of criticism; from more than one point the need for a rigorous approach to the analysis of data coming from social media has to be highlighted in order to avoid errors. However, the enormous added value represented, in terms of knowledge, from the correct transformation of such a large amount of data into useful indications, is evident: this means that, through the analysis of Big Data, complex phenomena could be explained by combining all the information which come from all the available sources, and it is evident that this translates into an extraordinary advantage for the companies and the reference markets. It is even more evident in the travel industry: every booking in a hotel, every flight purchased, every car rental, every transaction performed, or every train booked, basically every activity that includes a smartphone, a GPS, a credit card, etc. leaves behind a trail of data of considerable importance. More and more often, the organisation of a trip is discussed in areas dedicated to blog sites, where tourists tell their experiences, highlighting the positive or negative aspects of web containers that are visible to everyone. Big Data is therefore considered by many an incredible opportunity to predict or influence behaviour, opinions and feelings; moreover, understanding a customer's travel experience is essential to understand what, in a tourist offer, must be added, improved or eliminated. The biggest advantage will be the possibility of being able to make decisions in real time, a resource that can prove decisive in a sector like the tourism industry, where the time factor is often decisive.

In short, therefore, the advantages offered by the analysis of this type of data are, on one hand, of a strategic nature, because the Data Analysis make it possible to know the reputation of a given structure, of a territory, of a service or itinerary; on the other hand, of an operative nature, because all the information collected and analysed can lead to the maximisation of the satisfaction of the tourist, through a personalisation of his travel experience and offer. This apparently simple information brings with it an incalculable value, represented by the possibility of optimising its policy by finalising it to an improvement in reputation.

1.4 Sentiment Analysis on Social Network

Often the web is compared to a square, a huge digital space where everyone can discuss, get information, communicate. This like all squares is not silent, on the contrary: imagining for a moment to increase the volume to all the conversations that take place online, you will get an incessant noise, a scream made of languages, tones, volumes, different registers; A digital babel where it is hardly possible to distinguish individual speeches. In this context, a company must understand the sentiment of its target, that is, it must know what people say about its products, services, the events it organizes every year to increase fans or the updates it offers periodically. There is a way to start from the conversations made on social networks, the places that more than any other are dedicated to the exchange of information, dialogue and discussion around one or more topics. The growth of social networks and the birth of generalist platforms such as Facebook, Twitter or specific that is aimed at a particular category of users, such as LinkedIn has allowed a huge amount of people to discuss the most disparate topics, sharing opinions, comments on the Net, discussions of all kinds, also and above all around brands. Social networks today are real black boxes for companies: they contain fundamental information, which at the same time is difficult to discover. For this reason, to understand and capture this information, a branch of data analysis called Social Network Analytics (SNA) was developed: a young methodology that is widely used in the social sciences and which has inevitably developed in the economic field and marketing sector. Nowadays, carrying out this activity is of fundamental importance for those involved in digital marketing, since, once the metrics have been analyzed, it is possible to understand how to make the most of the enormous potential of social media. The data circulating within them can be of two types:

- *Horizontal*: opinions are linked to a specific geographical space; it is possible to establish these data thanks to the geolocation of mobile media.
- *Verticals*: people's opinions are considered based on certain characteristics such as culture, professional role, musical tastes and so on. In this case, people who are distant from each other are faced on the same theme.

It is not easy to extrapolate these data and there are, however, specific sentiment analysis software capable of capturing opinions relating to a single brand or other elements and transforming them into readable and understandable data. This huge amount of can then be organized and classified so that companies can exploit them to improve their marketing strategy. Thanks to media analysis, it is possible to obtain numerous data and information on visitors, customers, potential customers or simple followers (Appel, Grewal, Hadi, & Stephen, 2019). Some of the features are obtain information on the users' whereabouts, obtain information on the gender and age of visitors, identify the most searched and used keywords, identify the most used keywords in relation to a specific item and service for sale, or on certain posts, links, etc., check the times and days in which the social network is most visited by users, monitor general user behaviour. Social Media Analytics is a digital

monitoring activity that allows to verify the effectiveness of one or more social networks through a detailed analysis of the data flow. To do this, various tools are available, some free and others for a fee, or the programs for analyzing data already installed in the social media can be used. Furthermore, the Social Media Analysis collects and analyzes the metrics not only of the most famous social networks, but also of all the social media on the web, so for example online newspapers, forums and blogs. Various data can be obtained, in particular the five most important can be identified:

- *Textual data:* comments, posts
- *Data relating to the analyzed network*: Facebook friends, Twitter or Instagram following system
- *Actions*: shares, reactions, "likes"
- *Hyperlinks:* Hyperlinks
- *Useful data for SEO* (Search Engine Optimization) purposes

Knowing how to analyze social media allows, therefore, to save on paid ads, to get to know the customers better and, consequently, to be able to build a marketing strategy based on the various buyer personas. Social Media Analytics activities can be divided into three phases: data identification, data analysis and interpretation of the information obtained. In each of these phases it is essential to try to answer to the so-called 5: who/what, where, when, why, how. Before starting the analysis, however, an objective must be set, which can be the resolution of a problem, the monitoring of a certain aspect of the social network or simply an overview of the site's progress. Identifying the data is the first essential step, and until it is interpreted, the data has no meaning; for this reason, it is good to immediately identify the main attributes to consider: type of content such as text, audio, video, image, the source from which social network or site that data comes, the time that is the period being analyzed of the collected data, the data owner meaning of identifying if the data belongs to a public or private user. These are just some of the main attributes to consider initially. After identifying them, the data can be analyzed: in the data analysis phase, the tools or programs pre-installed in social media (Facebook Insight, Twitter Analytics, for example) actively come into play: these software are able to facilitate the task of the social media manager or anyone who is in charge of monitoring site data, by developing a data model that organizes the contents and gives an overview of how they interact with each other. There are several types of models, one that is widely used is undoubtedly the word cloud in which the most used and sought-after keywords appear larger and more centered within the cloud. The third and final phase is that of the interpretation of the information obtained from the analysis, that is the step in which the work passes into the hands of the human mind and no longer in that of social media analytics software. All the analyzed data must now be considered useful to implement the right marketing strategy, to adjust the shot on a certain activity, or for any other need. In order to better interpret the information, they are organized in a graph, taking into consideration the goal that has been set for ourselves since the beginning of the analysis. There are three criteria when creating this chart:

- *Target*: understand the target audience since there are many types of public and, consequently, different types of interactions with them
- *Framework*: make sure that the graph is semantically and syntactically correct
- *Storytelling*: it is difficult to understand social media analysis well, so to better assimilate the information obtained, it can help writing a sort of report of the whole analysis, so that it is clearer to the analyst

Once this is done, the data can finally be collected and represented in the chart: pie chart, bar chart, line chart or scatter chart are some of the examples that social media managers use thanks to the use of software that allow to measure analytically the exchange of data, metrics, interactions and everything you need to know about your profile or page.

1.5 The impact of Web 2.0 and Social Media

The tourism sector has undergone an important evolution with the development and diffusion of web 2.0 technologies. For this reason, it is useful to understand how social networks, which allow the sharing, planning and participation of customers and interested parties in the creation of value, are changing the marketing processes of small, medium and large enterprises. The demand of information from social media is getting stronger day by day but not only that: it affects quite decisively what happens in the whole Net.

This was already happening in the Web 1.0 world as regards the possibilities offered by travel suppliers, packages and tourist services. In fact, in this case, the Bocconi-Mobissimo observatory (Antonioli Corigliano & Rodolfo, 2009), showed how online transactions made by Italian users, only about 30% was carried out on Italian sites, the rest went to foreign suppliers. The analysis of the use of Web 2.0 functions was carried out on a sample (chosen at random) of 1428 Italian tourist sites divided into the following categories: hotels, bed & breakfast and farmhouses, travel agencies and tour operators, public bodies, APT etc., associations, consortia, wine roads, itineraries etc., restaurants and other services.

Despite the enormous success among users, the Web 2.0 world until a few years ago still did not seem to be "digested" by most Italian and foreign tour operators where the results were all quite similar among the main players in the sector. This "evolution" of the web has meant that many operators have been reluctant for some time to invest and to approach this new trend massively. After a few years, however, all operators in the sector began to understand the potential and began to implement actions and policies that also involved all the information deriving from the web 2.0 within their studies. The demand, as said, is there and tour operators in many parts of the world work hard to intercept this demand. Competition on the web has few boundaries and few restrictions and a traveler can be greatly influenced in his choices by factors completely different from what it has usually used and consider.

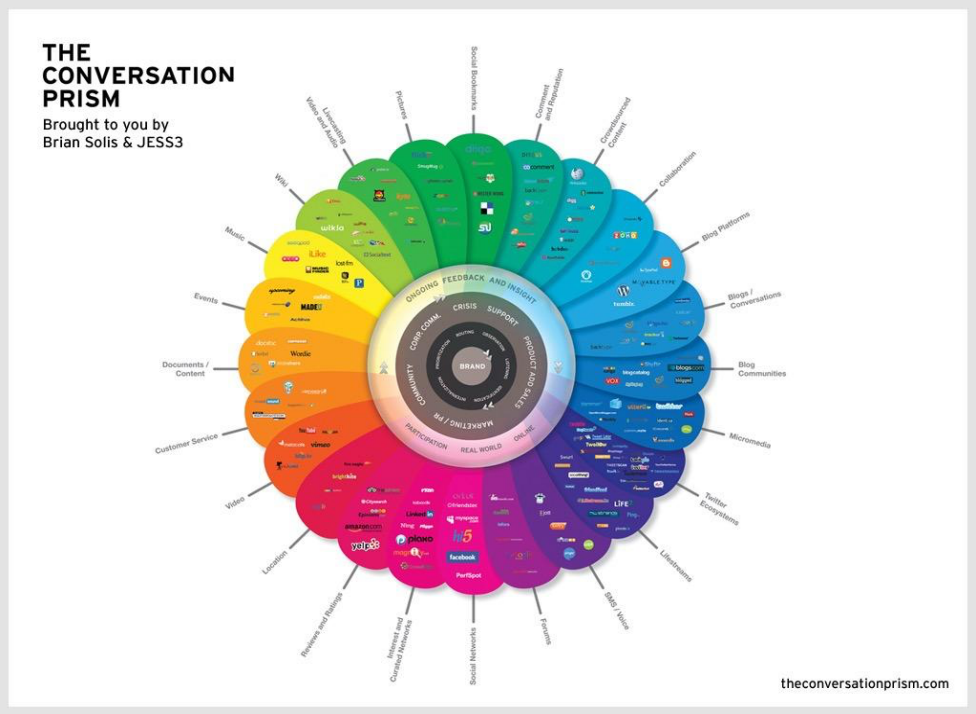
The flattening of the online offer had the immediate result of transforming what is offered to the market into goods and therefore being forced to fight practically only on price. It seems

to date back to a few years ago, until 1996, when Michael Bloch wrote his open letter to a travel agent who subtitled him on Business Travel News: survival tips in the electronic age (Travel Appeal, s.d.). There the author tried to draw travel agencies' attention to some initiatives, such as the combined one of Microsoft and American Express for the creation of a digital travel agency, strongly recommending not to refuse new online tools in advance, but rather to make them their own by using them to highlight their skills and diversify their offer. As has been discovered today, those who govern the vast majority of online markets do not seem to see significant slowdowns, but on the contrary, there are still increases.

Nowadays, the fashion of the moment is called social media: it may be possible, though very unlikely, that this frenzy will deflate in a short time, or that these tools are not suitable for conducting certain business, or that they are more suitable for the exchange of gossip or that they have more problems than others on sensitive issues like privacy or security, but the fact that hundreds of millions of people use them every day and spend most of their time there has some kind of meaning. Ignoring them, underestimating them and not trying to understand how and when to use them, risks being a serious mistake and leading to another economic disaster for a sector that has already been tried enough by recent financial events and a market that does not care about certain ideas. For this reason, the challenges and opportunities offered by web 2.0 have been accepted by many companies in the tourism sector who have not been afraid but have exploited the innovations created by web 2.0, so as to obtain advantages and give life to tourism 2.0. With the birth of Web 2.0, there has been a change in the way people interact with the WEB, if before (WEB 1.0) The Internet age is a tool used to obtain information, today it is used and considered as a platform for interact, have new ideas and creativity, confront other people. Therefore, a new audience was born, no longer "consumers", but by users who are becoming increasingly aware, active and content producers; It is no longer the company that tells the consumer what, where and why to buy, but the consumer determines the new rules of the "game". Users registered new digital technologies, podcasts, blogs, wikis, videos, digital media for sharing photos to exchange opinions, experiences, comments on products or services, to reinvent or modify messages recorded by companies.

- The most used tourist Social Media by users are:
- Of reviews: Trivago, Zoover, Tripadvisor;
- Of online tourist guides: Nextstop, Dopplr, Where I have been;
- Forum: Frommer's, Fodor's, Thorn Tree;

Figure 1: The conversation PRISM



Source: Solis (2008)

As it is possible to see from the image, the prism of the conversation arises from the need to classify the communicative flow of the web, distinguishing within its different types of use of social tools and allowing users to orient themselves by choosing, from time to time, platforms and applications that are suitable for your needs and purposes. The structure of the Prism consists of 5 concentric circles. The center is the single person: YOU. The first link represents the 4 main activities through which it is possible to re-elaborate and add value to experiences and information from social networks and communities by Listening and Learning from others and then co-creating and involving. The second link: it concerns the understanding of the effects of online activities in terms of "resonance", "relevance" and "reach" and also of "social capital", "popularity", "influence". The third link: it represents one's contribution to the dialogue on the web that forms one's own Brand (personal brand), one's person (digital identity) and one's community. The last two links constitute the real Prism, made up of about 230 applications and online platforms divided into 28 categories.

Many travelers use Facebook, Twitter and Flickr to review their holidays which, thanks to these social networks, allows status updates and / or upload photos on company pages. This has great prominence in the immediate responses of the tourist market, and for this reason OTAs (online travel agencies) such as Expedia and Booking.com, low-cost airlines and everything related to the tourist offer have been born over the years. It is in fact an increasingly effective aggregator of online services that have the peculiarity of understanding how consumers operate and are influenced by Social Media when making tourist choices.

Social networks are now the essential tools for travelers and tourism businesses: to understand the link between travel and the use of social media, the social travel startup Tripl has collected numerous information from the main operators in the sector and from sites such as Quantcast and TripAdvisor which according to the survey, showed that (Laney, 2001):

- 72% of all social network users access their social profiles daily when they travel
- 200 million passengers in the year 2017 have booked flights that allow them to connect
- 69% of all travel companies reported increasing traffic from Facebook
- 46% of all travel companies have detected growing traffic from Twitter, and the 5 largest airlines have 2,566,000 Facebook fans
- 50 million reviews have been published in 2017 on TripAdvisor
- Both Foursqaure and Gowalla register check-in on 6 continents

This information is useful to understand the greatness of this world and how their use has become intrinsically within each of us.

1.6 Segmentation, targeting and predictive analytics

Marketing as it was understood until recently has been completely revolutionized: the recognition that consumers have of the needs, needs, resources, preferences and purchasing behaviors are different and has therefore led mass marketing to evolve and embrace targeted marketing. Kotler and Armstrong (Kotler, 2000) describe target marketing as a customer-oriented marketing strategy because its goal is to create value for target customers and therefore it provides the necessary knowledge and tools to develop "the right relationships with the right. customers". Targeted marketing involves three main stages: market segmentation, targeting, and positioning.

The action of market segmentation thanks to the data collected is crucial for all those aspects that fall within the marketing and strategic planning processes thanks to which managers can study targeted and specific products and services for certain consumers, so as to obtain an advantage both competitive than economic. Currently there are two types of segmentation: a priori, when the variables that allow segmentation are known in advance of the market trend, and post doc, when you are not aware of the market and customers and are therefore used basic parameters for segmentation. Marketers have used various criteria as the bases of segmentation. There are various criteria that are used to segment the markets when carrying out a marketing analysis: these include demographic parameters (such as age, sex, family status, income), geographical parameters, behavioral within a certain business (for example the frequency of use of a card, loyalty understood as quantity and quality of purchases) and psychographic (such as lifestyle, personality characteristics) (Brynjolfsson, 2012). Normally, those who study behavior and marketing carry out a segmentation using several criteria in order to identify more homogeneous groups of consumers with transversal

characteristics. Multinational companies, in addition to studying and segmenting the internal market with the aforementioned parameters, work for the segmentation of the international market, that is, they identify distinct international targets based on segmentation criteria such as geographical position (for example, grouping countries by region), economic, political, legal and cultural factors. Travel and tourism agencies carry out studies on the segmentation of national and international markets since they are aimed only at internal, but also mainly at external ones. For this reason, a tourist destination, having analyzed its catchment areas, may need to address consumers from specific countries (such as, for example, Spanish and/or Russian) or specific intercontinental regions (for example Asia or Central America.) Some multinational companies are gradually changing the type of segmentation, adopting an intermarket one (Laney, 2001), that is, segmentations are created based on the needs and behaviors of customers, even if they come from different places and countries. This means that the places and characteristics of the countries of origin are not taken into account, but only focus on people and how they look at the product or service. In the case of tourism for example, this can be seen when the destination is in the mountains and the goal is to attract customers from anywhere but who are all interested in the services offered by nature and its services offered.

Segmentation is one of the most investigated areas when it comes to marketing research aimed at the tourism market. It allows to identify distinct groups of tourists, who, like in any other market research, do not respond in a homogeneous way and in a single class with respect to marketing activities. This means that the different services offered to customers make segmentation a useful and absolutely necessary tool to respond to the normal temporal evolution and to the services offered by the competition. A study on segmentation in the tourism market revealed that psychographic parameters are the most used segmentation criteria (75%) followed by behavioral ones (21%) and a mix of both (4%). The most used actions concern the segmentation of tourists using variables such as demographics, socio-economic situations and lifestyles. In the case of the tourism market, the recommended parameters to be used and taken into consideration for the segmentation certainly concern the demographic aspects, the activities that can be carried out in the destination, the travel expenses and how much the customers can spend, the benefits these can draw from it and the motivation that drives them to search and stay in the destination. In the case of Norwegian tourists (Liu & Haiyan, 2017), which include gender, age and marital status, great differences were found regarding adventure travel when segmentation was carried out, where some parameters such as employment and age were very different between the various segments and it was noted that this greatly affected the possibility of spending. On the other hand, however, no differences were noted with regard to the macro segments marital status, sex, occupation, age, education, income with the search for segments involving risk assumptions. These demographic parameters, albeit significant and of great relevance, have not, however, been able to verify whether in the present case they have actually been useful or not, and for which further investigations are necessary to demonstrate with certainty what demographic data are useful in the study of tourists for certain contexts. It is also not to be excluded that

the various segments, albeit deriving from different categories, may seek similar experiences and this means that the demographic data alone are not sufficient to give exhaustive answers to the questions. Because of this problem, there is no literature that well defines, and segments travel and tourism activities and this means that the standard parameters used are always the demographic ones, bringing with them the problems mentioned. However, there are some cases in which it has been tried to carry out segmentation of tourists based on sporting, cultural, or other activities: Sung (Sung, Alastair, & O'leary, 2000) has highlighted and created some standard parameters that can be useful for tourism research purposes, that is: non-extreme natural travel, extreme challenges, winter natural travel, high-risk travel, moderately risky travel. These parameters, as mentioned, can be applied to all demographic categories of tourists without distinction. They further argued that activities should be taken into account when studying adventure traveller segments because they are associated with consumer preferences. Furthermore, it was highlighted that the attractions need to be considered as they have a great influence on travelers who practice adventure sports because these are closely related to tourist choices and these have been shown to have different peculiarities in the demographic and socio-economic segments (Giammaria, Adolfo, & Andrea, 2017). In the cases in which high-risk attractions have been analyzed, they have shown that they have a clear and essential correlation with tourists from 11 different countries (Toedt, 2013), unlike the attractions that have a low risk.

As mentioned, therefore, segmentation in activities is very relevant and common in many situations: this is often carried out during post hoc segmentation, where one of the parameters under consideration is always referred to physical activities. This parameter, in the present case, called "active tourist" was a major factor associated with age and the type of social class during the segmentation of tourists in Scotland (Chen J. S., 2003), in Austria (Kotler, 2000) and "wellness tourists" in the Czech Republic (Dolnicar, 2006). Analyzing senior tourists, it was also noted that physical activity is also important and has different peculiarities for travel motivations. The category of "enthusiastic tourists" found in Sellick's study (Sellick M. , 2008) found that the feedback was significantly higher when they were influenced by physical activity than the other parameters of senior tourists. Likewise, older people who stayed to participate in some sporting activities were one of the targetings identified by Connie Mok and Thomas Iverson (Mok & Iverson, 2000). Another feature and parameter to consider is the length of the stay where Isabelle Frochot and Alanaistar Morrison (Frochot & Morrison, 2000) defined two parameters for the classification of tourists: "short-term visitors", that is, those who have stayed from 1 to 6 nights and "long-term visitors", that is, those who have stayed 7 or more nights. Considering these two parameters, marked differences in satisfaction were noted between "short-term" and "long-term" visitors as well as first-time tourists who have already stayed there (Sellick M. C., 2002). Tourists who have stayed up to six days are generally less satisfied with the perceived quality of services and similarly with the perception of the cost of their trip than long-term visitors (Frochot & Morrison, 2000); repeat visitors show better satisfaction feedback than first-time visitors (Sellick M. C., 2002) and changes in general feedback based on length of

stay, gender, and time of booking (Sung, Alastair, & O'leary, 2000). Satisfaction and pleasure, according to Abraham Pizam (Pizam, 2004), are highlighted thanks to significant differences between segments in the tourism markets, while demographic constants show irrelevant differences.

Tsiotsou Rodoula and Vasioti Eleytheria (Tsiotsou & Vasioti, 2006) carried out a study regarding ski resorts and the results showed that images of the locality, both primary and secondary, should be studied in order to identify the correct positioning and effective strategies to be implemented. The primary images have the same characteristics for all destinations, such as a ski station, while the secondary images concern the peculiarities of each place so as to differentiate it: the morphology of the mountain, a particular wellness service, the type of services offered. The latter are intended to show an advantage over the competition. That said, the positioning of a tourist destination, whether it is mountainous, maritime or city, should be based on the data obtained thanks to the segmentation and analysis of the market target of tourists who have already stayed in the place together with the analysis of the positioning. The positioning of the location should be corresponding and consistent with the chosen market segments and with the services offered as opposed to competing ones (Frochot, 2005).

As has been said, technologies have accelerated changes in tourists and destinations in a very incisive way, and the tourism industry is also starting to exploit these tools so as to take advantage even more incisively than the current one thanks to greater diffusion. to a much larger audience than in the past, and thanks to the ease of use. These characteristics have meant that the markets are constantly evolving, increasing their catchment areas more and more. Unlike in the past, traveling has become a habit of many millions of people, albeit with different incomes and without the need to have large amounts of money. The use of the enormous amount of information available can certainly help hospitality public decision makers, business owners, including hotels, bars, restaurants, shops, so as to identify the right market in terms of guest feedback. to improve the processes. Through social networks, blogs, and other websites it is possible to obtain feedback and other important information from users so that socio-economic choices can reflect and continually improve. In the case of accommodation facilities, data collection is facilitated thanks to various portals such as Online Travel Agencies that allow the customer to book and at check out to give feedback to the structure so as to allow managers to be able to improve every detail. Thanks to the use of data, it is finally possible to have a complete picture of the competitors, to see which tourist choices the other locations have put in place and what other competing businesses are implementing so as to act accordingly. This feedback and reviews system allows you to absorb a lot of information, not only by looking at your own but also and especially those of your competitors, whether they are positive or negative. Thanks to this, it is possible to adapt, improve, learn what is positive about competitors.

These three fundamental aspects concern the use of a multitude of data which, following specific technical criteria, are able to give the end user an overview of how information is

collected, divided and subsequently used to carry out a predictive analysis of what is it will likely happen in that particular market.

1.6.1 Business Analysis for Tourism

The combination of business analytics, data mining, data visualization, data tools and infrastructures, as well as best practices to enable organizations to make more data-driven decisions is called Business Analysis. In practice, you know you have acquired modern business intelligence when you have a comprehensive view of your organization's data and use it to stimulate change, eliminate inefficiencies, and rapidly adapt to market and supply changes. These tools enable companies to make better decisions by showing current and historical data within the business context. Analysts can leverage Business Analysis to provide performance and competitor benchmarks to help the organization run smoother and more efficiently. In addition, analysts can easily identify market trends to increase sales or revenue. When used effectively, the right data can be useful for everything from compliance to hiring.

While management systems, also called "operational" or "transactional", are mainly concerned with entering and updating data, the purpose of Business Intelligence systems is to extract from this "raw material" made up of data all more or less information hidden that these contain. Clearly the two categories have different purposes, requirements and require completely different design techniques. The areas to which it is possible to apply data analysis represent almost all the business sectors, even if historically the analytical CRM area has seen the greatest number of applications in the field. The more companies produce information, the more they need to present it in a harmonious way to support decision-making processes; the business analysis it mainly addresses the management (strategic, tactical, operational) or the knowledge worker, providing them with the necessary tools to make decisions and solve problems. A BI system meets functionality and design requirements that go far beyond those of a normal reporting environment that is part of a management application. In particular, a BI system must have the following characteristics:

- *Ease of use:* present data in a format that is easy to read and interpret, where it is possible to navigate on the data following analysis paths and which makes extensive use of graphs. Field names must be easily understood by the end user
- *Speed*: ability to process large volumes of data with near instantaneous response times thanks to the use of data modelling, storage and indexing techniques aimed at analyzing rather than updating data.
- *Integration*: integrate data from different sources, both internal and external to the company. The integration process must be reliable and tested, so that users can rely on the data present in the DW. If the data from the operational systems are not clean and reliable, they must go through a data cleansing and certification process before being inserted into the DW.

- *Historicization*: keeping the history of changes undergone by certain selected attributes, to allow contextualized historical analysis
- *Identification of trends and anomalies*: the tools must facilitate the identification of trends in the data, for example by comparing different periods and products. These operations are possible only with the use of interactive tools that allow you to perform drill down / drill up (display of details on a certain data) and slice & dice (change of the analysis dimensions on the two axes).
- *Subject orientation:* presenting the data in order to provide the vision of a business process (supply chain, sales, quality), crossing the boundaries of the individual areas of the management systems.
- *Scenario simulation*: in some cases (budgeting, forecasting and planning applications) it must be possible to set scenarios and then compare them with the real values ("actual")
- *Independence from the IT department*: the analysis and reporting tools must allow end users to create the reports they need themselves
- *Adaptability over time*, understood as the ability to resist the inevitable evolutions of the company reality, operational systems and analysis needs
- *Security*: it must be possible to control access to data, which in many cases include highly confidential information, in a tight and flexible way.

These tools bring advantages for visual analysis and data visualization where people are attracted to visual stimuli and immediately notice patterns or colour differences. Visualizations show data in a more accessible and understandable way. Collected in dashboards, they allow you to tell a story and highlight trends or patterns that would not be easily recognizable with manual analysis of raw data. This accessibility enables further reflection on the data, leading to a greater impact on the business.

Many applications today, and for what is foreseeable in the near future, provide useful and effective results for the analysis of tourist behaviours, movements and desires, for general business intelligence and, for some algorithms, there are examples that make more effective different forecasting activities (Mariani, Baggio, Fuchs, & Höepken, 2018). Furthermore, areas such as recommendation systems can exploit the large amount of data that otherwise would have been difficult to use with traditional methods (Mehrbakhsh, Karamollah, Mohsen, & Vahid, 2017). The idea behind BI techniques is that an algorithm can discover general rules or characteristics in large data sets using a generalized approach without being explicitly designed for some specific work (Witten, Frank, & Hall, 2016). A BI algorithm does not make preliminary assumptions about the possible relationships between variables, but sometimes guided by a few examples, it processes the data and discovers configurations that can be used to identify features, classify elements or make predictions. The broad set of algorithms can be roughly divided into two main families: supervised and unsupervised algorithms. The first are those that require some sort of 'training'. A set of pre-classified (or pre-labelled) data is given as input to the algorithm that deduces the characteristics of the various elements and creates a model. The model is then verified and validated using part of

the original data and comparing the initial classification with that derived from the model created. Several algorithms can be tried until the desired or necessary levels of accuracy are achieved. Examples of supervised machine learning algorithms are regression and classification models (naïve Bayes, support vector machine, k-neighbours, etc.).

In the case of unsupervised learning, the model is prepared independently from the input data. These algorithms are typically recursive, with successive approximations and refinements being performed until verification shows that the specified or expected level of accuracy is achieved. The clustering algorithms (hierarchical or k-means) and those used for dimensionality reductions (factor analysis) belong to this class. The distinction, however, is not always so clear and there are mixed forms. Furthermore, some techniques involve a 'reinforcement' which consists in starting with a supervised or unsupervised learning procedure and reusing the output, once verified and validated, adding it to the input and updating, if necessary, the model created and the previous conclusions and thereby greatly improving the accuracy and precision of subsequent results. This dynamic process can be repeated many times allowing, for example, a continuous refinement of the obtained or obtainable classifications. Another class of algorithms are the so-called 'deep learning'. It is a specialized area of machine learning that refers to algorithms inspired by the structure and function of the brain called artificial neural networks (ANNs). Some believe this to be the closest technique to the original goals of the wider field of artificial intelligence, which aims to develop systems and machines that mimic the cognitive functions of a human brain (Arel, Rose, & Karnowski, 2010); (LeCun, Bengio, & Hinton, 2015). DL architecture has been applied in computer vision, automatic speech recognition, natural language processing, audio recognition and transformation or creation of synthetic digital objects (images, audio, video) (Chen, et al., 2019).

In the tourism sector, applications are available that combine, for example, image recognition capabilities with augmented reality for the design of travel guides tourist attractions (Xiaozhou, Zhe, Chengqi, Yun, & Jing, 2019) and software and mobile apps that provide reasonable translation services are available (e.g. www.deepl.com/translator, or Google translator) DL systems, and in particular those for speech recognition, are now commonly used in robotic devices such as those humanoid robots that begin to appear in hotels around the world and that generally help concierges and front desks to answer questions and customer requests (Bowen & Morosan, 2018). Furthermore, DL techniques are proving effective in improving tourism demand forecasting capabilities (Law, Li, Davis, & Xin, 2019).

1.6.2 Business Intelligence Visualisation Tools

Business Intelligence software (or BI software) is a class of IT applications that process and analyse business data to produce quality insights and help integrate the greeting status the companies (Turecková, 2002). BI software uses a variety of formulas and metrics to

measure, it appears and relates company records and selects to distinguish the strengths and weaknesses of each company. The main functions of these systems are the collection, management and reporting of data, and some of them also evaluate the functionality and performance of employees. The intervention of business intelligence programs is to allow the simple and deeper interpretation of large volumes of business data, discovering growth opportunities and improving the effective strategies provided on these in-depth analytical analyses: it can provide companies with a huge competitive advantage in the market compared to competitors, in addition to long-term stability. These software platforms can provide online, current and predictive historical companies with the various operations of their business and therefore, it is not surprising that many companies, whether they are small start-ups, medium-sized companies or businesses, use these tools to support the offer of business decisions, whether operational or strategic in nature. Apps and systems are used to analyse and transform big data into usable business intelligence, which is particularly beneficial for large companies with complex structure and organisation as they need solid and well-integrated solutions that reveal the picture of their activities, in particular those signalled by important trends and opportunities and reveal the risk at an early stage to help them avoid serious financial damage.

Four major benefits and requirements of this type of software should be in mind when choosing a program's core features (Sellick, 2004):

- *Ability to integrate information from different business systems*: the data concerning the businesses that allow strategic choices to be made are generally on various terminals, where, for example, the economic data need to be seen from a financial perspective. For this reason, integrated business intelligence systems need to rapidly and dynamically aggregate information from different sources in order to respond to business questions.
- *Reporting and analysis*: the solutions offered by business intelligence software must have the ability to be dynamic and allow for an exploration of data so as to allow a better understanding of trends. With the evolution of technology, the mere company report has been overcome allowing various actions and strategies to predict future trends and scenarios in order to obtain a better final result.
- *Historical analysis and reporting:* one of the objectives of business intelligence software is to provide an overall view of the data so as to allow better business efficiency over time by showing the reasons. Within the company, the decision-making flow includes a series of data aimed at understanding and growing it over time and this thanks to the peculiarity of being able to map and analyze historical data over the years, albeit very complicated. Thanks to these features, the program has the ability to extract, manipulate and integrate a lot of information, which is absolutely normal in corporate databases.
- *Future projection and forecasting*: thanks to the historical data that the software is able to manipulate and integrate, it is possible to set future economic choices and projects: this will therefore allow to analyze the processes in progress, and to set new ones. The

goal is to be able to predict future scenarios with a certain degree of certainty, allowing the best possible success.

In our case we used the Tableau software: this is because it is one of the most used data visualization software in the Business Intelligence sector (Neal, 2008).It allows you to aggregate raw data in an understandable and analyzable format. The data analysis software allows the creation of intuitive dashboards and worksheets whose data that is extrapolated can be understood by professionals and at all levels of the business so as to allow any person without certain skills to create a customized dashboard

The feature of Tableau is:

- *Data Blending*: is a method of merging data from multiple sources and merges information from a secondary data source and displays it with the data from the primary source directly in the dashboard. There are different methods of combining data, each with advantages and disadvantages: relationships are the default method in Tableau that can also be used between tables with different levels of detail as these are flexible and adaptable. The join function combines different tables by adding multiple columns of data tables with similar rows: this can however lead to some problems such as data duplication if the tables have different levels of detail. Another basic function involves combinations, which unlike relationships or joins, never actually combine data but query each data source independently: the results are collected at the appropriate level and displayed visually together in the view. Combinations can therefore manage levels of detail and analyze different data sources.
- *Real time analysis*: the software makes it possible to understand and share insights from rows of data that arrive in real time in a visual and interactive way, and with a click of the mouse, it displays the data in a way that makes sense whether it is a pie chart, a detailed map or anything in between.
- *Collaboration of data*: one feature of the software is to be able to quickly and securely share data, reports and dashboards within the company or to customers and this regardless of whether the data is saved in a local terminal, on the cloud or on a hybrid infrastructure.

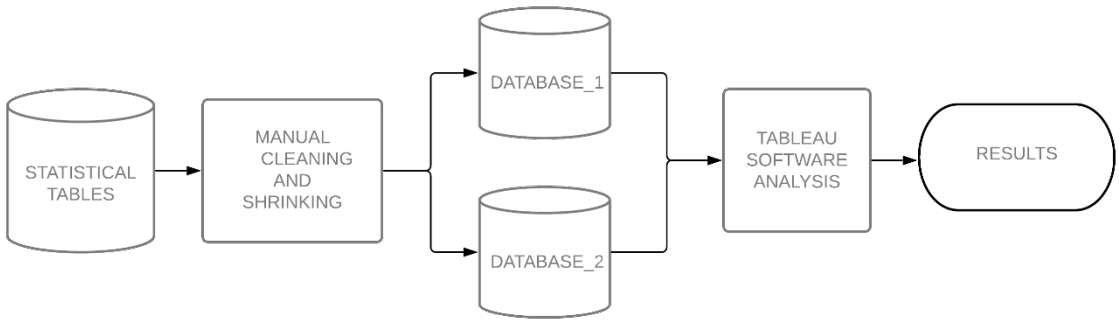
A fundamental aspect that is taken into account is the fact that the software doesn't require technical or any specific skills to operate which allowed its use without having any notion of certains programming languages. At the moment Vodafone Italy is using Tableau for studying tourism in other several Regions of Italy using GSM signal and tracking: it extracts and unsices archived data from different sources that can be extracted from any platform such as database, excel, pdf and from complex databases such as Oracle or cloud databases such as Amazon web services, Microsoft Azure SQL databases, Google Cloud SQL and various other data sources that can be linked in real time or extracted using Tableau's data engine, Tableau Desktop. In our case, as our data source comes from excel files saved locally and we had to use Tableau Desktop where the data was pulled up and develop in

visualisations. Another aspect that has been considered is that the created dashboards are shared with the users as a static file, so the users who receive the dashboards views the file using Tableau Reader and this can be useful when working remotely. The data from the

2 RESEARCH METHODOLOGY

This chapter includes the research methodology of the dissertation: in more details, in this part it outlines the methods of data collection, the research process, the type of data analysis and the ethical considerations of the project. The research held with respect to this dissertation is an applied one, but not new. Rather, several pieces of previous academic research exist regarding the use of Data Analytics for promoting and managing tourist destinations, not only for a specific area, but for all the tourist destinations placed all over the world. In this case the purpose is to explain the reasons why data are increasingly fundamental for any type of business in the tourism sector. To do this, a precise sequence of actions was followed to explore the information collected and, using business intelligence software, to obtain information that was previously hidden.

Figure 2: Reasearch Methodology Flow for statistical analysis



Source: Own Work.

In order to focus on a practical and real case, the Valle D'Aosta Region was taken as an example and the information available to the Regional offices was analyzed. Thanks to these offices, it was possible to retrieve a series of statistical information concerning the accommodation on the entire territory creating two ad hoc databases. Together with this, a social analysis of a real Facebook and Instagram page of a company operating in the tourism sector was also carried out.

2.1 Data Collection from Statistical Aggregation

As regards the first step of the analysis, or the acquisition of data, various tables available have been collected through the Regional Office of Tourism and Statistics. These were first of all normalised and inserted in a single archive that allowed to have under two source the

useful information and that could be used for the purpose of the study. The data was collected manually as well as the standardisation, which due to the tables of different types and formats has made it possible to create the most homogeneous tables possible. The type of data collected is business generated, i.e., those that are produced internally by the touristic offices that comes from the hotels are collected daily through an online form where each lodge must complete: it contains the number of arrivals with nationality and the number of nights of accommodation. Once these are collected, the offices divide this information into different tables on the basis of predetermined parameters, such as the area of interest, the type of accommodation facility. Due to their formatting, they are quantitative and semi-structured and therefore need to be transformed into structured data so that the analysis software can interpret them.

The type of analysis that has been applied is descriptive: this is due to the fact that the amount of information available is related to a small dataset that does not allow to carry out a predictive study. Thanks to the descriptive analysis, the raw data are synthesised and described and make it something that can be interpreted by human beings: specifically, the past events of our dataset are analysed, where for past events we refer to the annual information of tourist arrivals. These studies are useful in that they allow organisations to learn from past behaviours and to help them understand how they could affect future results and to analyse, at the aggregate level, what is happening at the level of flows.

As mentioned, in depth statistical data was studied: the statistics are unstructured, whose aim is to identify the type of guests, country they come from, how long they stay, which month. Starting from the regional statistical tables, it was first necessary to try to have them in a format that the analysis software could interpret them. For this all the data, which previously were on pdf format have been transformed into excel. Once these data were included in the various tables, an attempt was made to discard futile and unusable data and to group them in other tables. To do this, the need arose to find the correlation between the various tables so that once joined, they can be analysed having in fact created an aggregation.

The following tables refer to the tables as they are presented on the Regional official website in the statistics section and which, some of them, have been taken as a reference.

Table 1: List of statistical data tables considered

Arrivals by type of hotel and category - Aosta Valley - Years 2014-2019
Attendance by type of hotel and category - Aosta Valley - Years 2014-2019
Arrivals in hotel and non-hotel establishments by tourist area - Aosta Valley - Years 2014-2019
Arrivals of Italian and foreign customers in the region's accommodation facilities by tourist area and month - Aosta Valley - Year 2019
Attendance in hotels and extra-hotel establishments for tourist area - Valle d'Aosta - Years 2014-2019

Presence of Italian and foreign customers in the region's accommodation facilities by tourist area and month - Valle d'Aosta - Year 2019

Table continues

Table 2: List of statistical data tables considered (Cont.)

Arrivals and presences of Italians by tourist area and some regions of origin - Absolute and percentage values - Valle d'Aosta - Years 2014-2019
Arrivals and presences of foreigners by tourist area and for some countries of origin - Absolute and percentage values - Valle d'Aosta - Years 2014-2019
Size and capacity of hotel and non-hotel hospitality establishments - Valle d'Aosta - Years 2014-2019
Size and capacity of hotel and non-hotel accommodation establishments by category and month - Valle d'Aosta - Year 2019
Average stay of customers in accommodation facilities in the region by type of structure - Valle d'Aosta - Year 2019

Source: ISTAT *(2014 - 2019)*

The only public and free source on which it was possible to find information it's the regional statistical office. On a monthly basis, the arrival and stay information for each individual municipality of all hotels and non-hotel accommodation is grouped together. These too are divided by type. In the case of hotels, the data concerning, for each municipality, the arrivals and stays of each is divided by the number of stars. In the case of non-hotel accommodation, we refer to: tourist apartments, farmhouses, guesthouses, bed & breakfasts, campsites, holiday villages, holiday homes, youth hostels, alpine huts, rest areas. This information is also catalogued by destination (the area). As for the typology of tourists, the information at our disposal is divided into 2 categories: Italians and foreigners. As far as Italians are concerned, it is possible to know the province they belong to, while as regards foreigners, it is possible to know the country of origin. Having a database with complete information from 2014 to 2019, the work was focused on these six years, which is considered an adequate time on which to base oneself. It should be noted that for the year 2020, there is still no complete information available. In the continuation of the text some considerations will be made regarding the topicality of the information that invalidate the validity of any study concerning predictive analyses.

In detail, the tables considered useful for the purpose of the study are:

– Tourist movement in Aosta Valley by tourist district. This table indicates the arrivals and presences of all tourists for each year by dividing them by geographical tourist area of destination. It also shows the annual change (positive or negative) compared to the previous year.

- Tourist movement in Aosta Valley for main foreign origins. This table indicates arrivals and presences by foreign tourists, divided by the Country.
- Tourist movement in the Aosta Valley for the hotel and non-hotel sectors. This table summarises the total of monthly tourists in the 2 categories of accommodation facilities, whether they stayed in hotels or non-hotels

Another interesting source for the study concerns cableway installations. This data, which is detected on a seasonal basis (but considered annually as the cable cars works only during winter) whereby season we mean indicatively the extremes that are located between the end of November and the beginning of May depending on the state of snow and the area. Two tables have been identified as useful and are:

- Number of passes per ski area. This table refers annually to the number of passes on each ski area.
- Number of presences (first admissions) per ski area. Unlike the previous table, in this there are only the first entrances, that is the single tickets sold.

Note that the ski areas do not coincide with the tourist area, as there can be multiple ski resorts in a single area (for example, in the Mont Blanc area there are both Courmayeur and La Thuile).

The data from the statistical offices on the website of the Department of Tourism of the Valle D'Aosta Region had to be skimmed with the redundant and not useful ones: this type of study was done manually, going to discard the tables and leaving only those that could be useful and that could be crossed between them. This fundamental step has to be done because otherwise it is not possible to carry out a data crossing study, but it remains a mere statistical information. This data is collected monthly and annually for what concerns the accommodation facilities while they are collected annually for what concerns the ski lifts of the ski resorts divided into the 7 tourist areas. For privacy reasons (DGPR law, https://gdpr-info.eu/) it is not possible to have free data available through social networks which therefore makes the study as a starting point to have a clear idea of the tourists in the place; However, this problem is solved by businesses and public bodies that can purchase data packages from social network companies, telecommunications companies that follow the European privacy rules.

Once the tables were chosen, they were merged together so that the analysis software was able to interpret the imported data. These tables were firstly studied empirically: once the potentially useful and non-redundant ones were selected for the purpose of the study, having available tables with equal columns but with different values divided between Italians and foreigners, these have been merged into the same tables. These new tables were saved normalised so that the analysis software could, once imported, carry out the case study. The new tables that have been processed are as follows:

- Based on the year, from 2014 to 2019, the number of arrivals divided between Italians and the first 15 foreign countries of origin in the various types of hotel (number of stars) and type of non-hotel accommodation (apartments and "others").
- Based on the year, from 2014 to 2019, the number of arrivals divided between Italians and the first 15 foreign country of origin with the chosen destination (Aosta, Gran Paradiso, Gran San Bernardo, Monte Bianco, Monte Cervino, Monte Rosa, Valle Centrale)

Two new databases have been created using a mathematical formula that has allowed the union of the two previous ones in order to have a database containing all the statistical information of the arrivals for each destination and each type of accommodation facility: one for hotels and one for what concerns the other types of accommodation. With these new tables ready to be used, they have been imported into the Tableau Desktop software: the header of each column is automatically interpreted and made such that it is not interpreted as a value within the table (the fields of the years interpreted as a date and not as a number).

Table 3: Aggregation of raw tables

Tourist movement in Aosta Valley by tourist district	Tourist movement in Aosta Valley for main foreign origins	Tourist movement in the Aosta Valley for the hotel and non-hotel sectors
Based on the year, from 2014 to 2019, the number of arrivals divided between Italians and the first 15 foreign countries of origin in the various types of hotel (number of stars) and type of non-hotel accommodation	Based on the year, from 2014 to 2019, the number of arrivals divided between Italians and the first 15 foreign country of origin with the chosen destination	

Source: Own Work.

With this data available it is possible to know how many tourists from a country stayed in a certain type of accommodation in a particular area: various examples of data crossing are proposed to demonstrate how even only a few data can be useful to be able to start targeting customers in a given area.

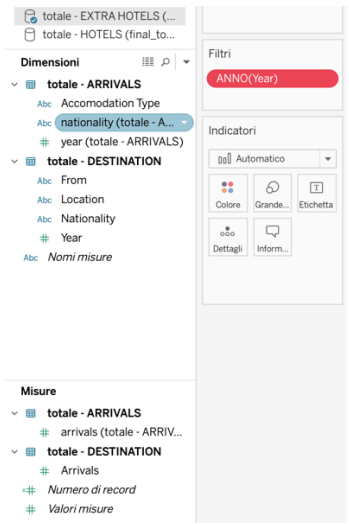
As one of the goals is to create a standard model that can be implemented in the future with additional information such as data from transactions and types of credit/debit cards, data on GSM tracking through the analysis software it was possible to have some visual graphics and some useful information that went beyond the statistics. As said, the model is useful to have a starting point for carrying out a detailed study of the tourist in the seven areas, where it is necessary to add the mentioned above information. The segmentation is necessary to identify homogeneous groups of consumers in order to satisfy their needs, desires and

preferences more specifically than a mass marketing strategy could do while at the same time increasing marketing efficiency and effectiveness.

2.2 Business Intelligence Tool

Once the connection inside tableau software was set up, the data source fields appear on the left side of the workbook in the Data window. Data blending is done sheet by sheet and is established when a field from a data source is used in the view secondary. To create a combination in a workbook that is already connected to at least two data sources, dragging a field from a data source onto the sheet, making it the primary data source. By switching to the other data source and using a field on the same sheet, making it a secondary data source. An orange link icon will appear in the Data pane indicating which fields are used to combine data sources that requires a primary data source and at least one secondary data source. The first data source used in the view becomes the primary data source and defines the view; This can limit values from the secondary data source - only values with matches in the primary data source are visible in the view. This is comparable to a left join. For example, if the primary data source has a Month field that contains only April, May, and June, any views generated for the months will only show April, May, and June, even if the secondary data source includes values for twelve months. For the software, to know how to combine data from multiple sources, there must be one or more dimensions in common between the data sources. This common dimension is called the link field. Active link fields are identified in the Data pane for the secondary data source with an active link icon, while potential link fields are identified with a broken link icon. For example, in a combination of transactional data and quotas, a geographic field might be the desired link field, so you can analyze a region's quota and performance towards that quota.

Figure 3: DB with fields in Tableau Software



Source: Own Work.

Tableau assigns each field in the data source as a dimension or measure in the Data pane, based on the type of data contained in the field: these are used to create views of the data. With the use of this software different columns of the database have crossed and this has allowed to obtain more information. These, as a matter of simplicity and understanding, were then analyzed thanks to plot graphs that cross the two tables.

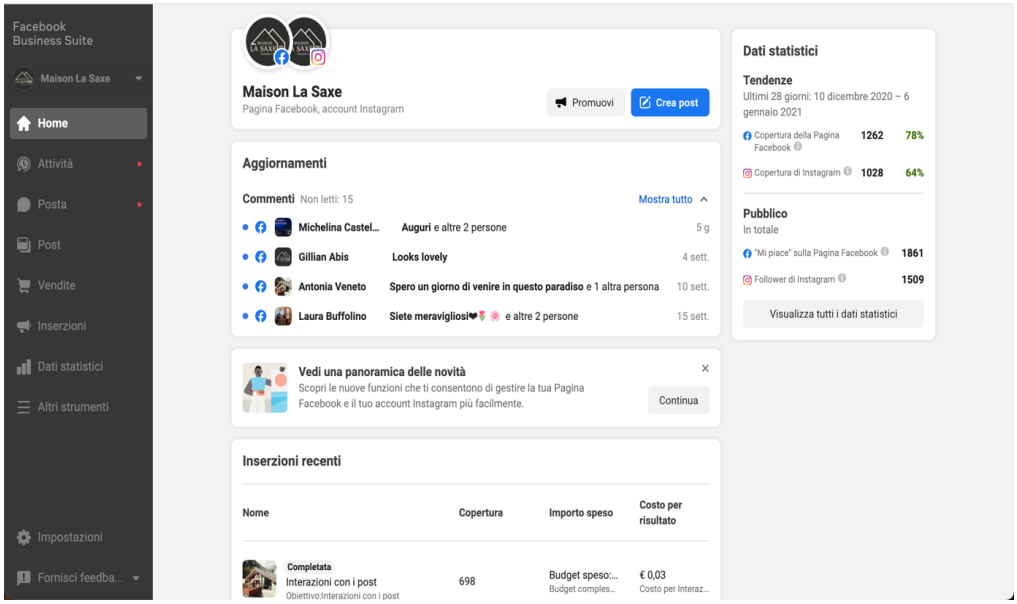
2.3 Social Network Methodology

Thanks to the recent developments in the research on Social Network Analysis, a part of the studies has tried to find concrete applications in terms of tools for diagnosing social groups, information and monitoring of the network in real time. approaches that used Social Network Analysis for the study of feedback methods for monitoring social relations and for intervening on them aimed at improving their quality of the connections between the components, used as a basis for the development of the work of this thesis. Many other examples can be found in the literature.

To carry out the analysis of the main social networks Facebook and Instagram and to find out the reference targets and which are the competitors, once established, it will be important to understand which tags they refer to and how these can be used to increase the audience. Once this information has been obtained, it can be investigated in order to demonstrate the usefulness of the thousands of information found within the social networks themselves and returning to the research question: what the potential of big data in tourism is, and in this case, using social media. This type of analysis can also be replicated by any public body that has a social profile.

In order to carry out this analysis, the management panel introduced by Facebook will be used to control the company pages of both Facebook and Instagram called Facebook business suite.

Figure 4: Facebook Business Suite



Source: Facebook Inc.

It is a unique destination on mobile and on computers that allows companies to access the tools needed to grow on Facebook and Instagram: a single tool to switch between managing one social network and the other. One of the interesting aspects for those who work daily with social networks, integrates numerous functions and connects the content strategy shared between the two social networks more effectively. It includes tools such as post creation, messages, statistics and advertising features, all from a single work environment. Thanks to these tools it is possible to view updates at a glance: view all notices, comments, messages and other important activities that require attention within the Business Suite screen for priority dates and easily manage daily business activities. Thanks to it, it is possible to share content with the Facebook and Instagram communities, or create draft posts for the Facebook news section or for the Instagram feed and schedule their publication at a fixed time and day considered the best to reach the public key. A very important feature is to be able to analyze what works on your profile and expand your audience: thanks to the "Insights" tabs it can be analyzed statistical data on the coverage, interaction and performance of posts, discovering which ones are most appreciated so as to organize and optimize initiatives on each platform. The last feature, but probably the most important, is to be able to increase your audience by promoting a post or creating an advertisement to increase the number of people who see and interact with the content.

The advantages that derive from this application are the time savings: once the content has been uploaded for the social network, it will be the possible to cross-posting, or the automatic publication on both platforms. This therefore allows quick interaction through messages in a simplified way, creating ad hoc shortcuts, to simplify the answers that refer to templates already saved and customized. In order to verify the effectiveness of your own editorial

calendar and posts, the Insights options will be accessible so as to always check coverage, interactions, involvement. This will be accompanied by another tool, Facebook Insights which, thanks to more in-depth statistical data, is able to:

- Know who is the target audience
- Know in detail all the audience's demographics
- Have a general vision of how long the Facebook page works
- Analyze in detail each publication and understand how it has worked
- Detect in detail which publication the audience likes best and which ones works better
- Know if it is reaching the objectives within the facebook strategy
- Compare paid publications with organic ones
- Check what time and during what day the followers are connected
- See what publications don't work in order to not repeat them
- Know if the Fb strategy is working or not

This tool allows to interact with data in a very short period of time, in which, thanks to the engagement, it can be seen the iterations within the previous month. In the analysis phase, thanks to this software a general overview of the Facebook and Instagram pages highlights its peculiarities and market position in the past 28 days. First of all, we want to understand if the page is of real success, that is, it creates an active participation and to do this it is necessary to use some Facebook insights metrics and make a simple calculation, bearing in mind that the average user engagement index is around 2% and this means that if you are below this percentage, the page needs to review the editorial and communication strategy. This information is a metric calculated over a period of a well-defined period of time and is not a "universal" engagement indicator, but certainly a data that can help plan and manage the page.

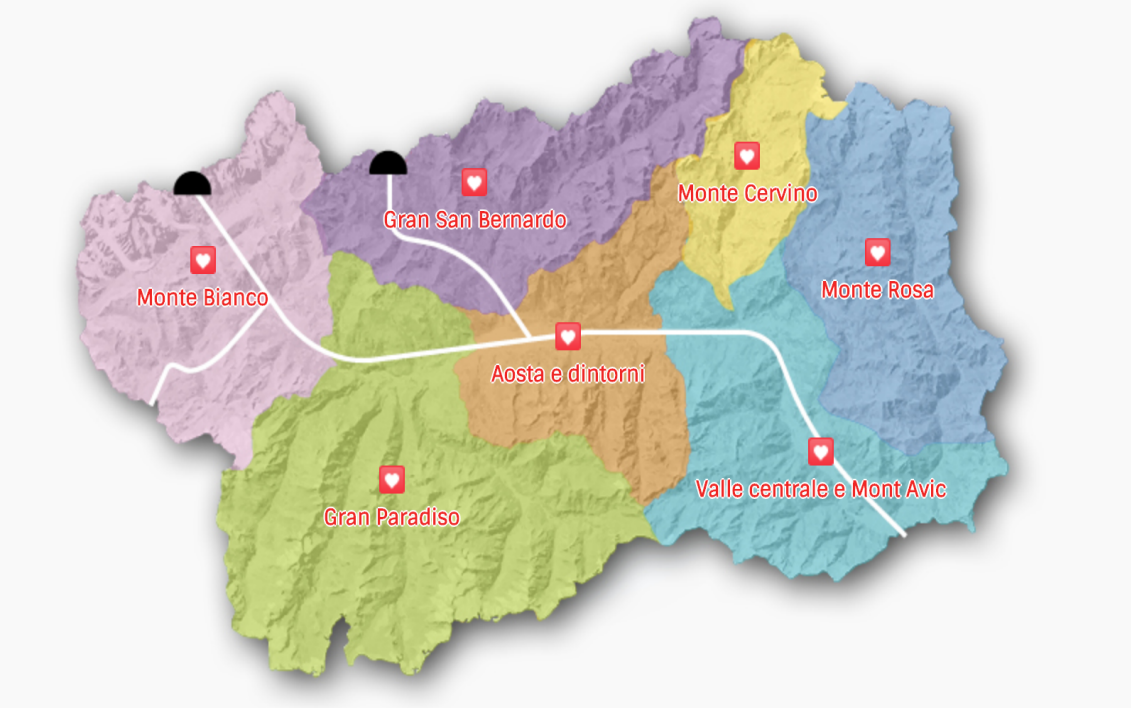
Thanks to these tools, a technical analysis of the Facebook and Instagram pages is carried out, based on statistical insights data: the characteristics are highlighted and how they interact with the page in question. Furthermore, the two social networks are equated: Facebook and Instagram, thanks to which the differences and the same peculiarities are highlighted.

3 CASE PRESENTATION

For the development of this thesis, the area examined and taken into consideration is the Aosta Valley Region: one of the 5 Italian Regions with special status. This choice fell primarily from the fact that it is the writer's place of origin and it's, secondly, a way to brig added value to the community in which I live and work; sure, of the fact that it can be useful for those working in the tourism field that with the advent of all the new technologies has disrupted and changed the market.

Aosta Valley is an autonomous Region with special status located in North-Western Italy: it is the smallest and least populated in Italy but, in its territory, welcomes a great variety of aspects that make it a destination always very appreciated. On its territory they peek out the "Giant of the Alps", such as the Mont Blanc (4.810 m), a mountain carved in granite and second in Europe only to Elbrus, of the Caucasus, peep out on its territory; Mount Cervino / Matterhorn (4.478 m), with its characteristic pyramidal shape; Monte Rosa (4.634 m), second highest mountain in the Alps; and the Gran Paradiso (4.061 m), the only "4,000" entirely included in Italian territory, within the park of the same name that winds between Valle d'Aosta and Piedmont, established to protect trees and some species of animals in of extinction such as ibex, chamois, marmots and stoats. Moreover, churches, castles, Roman monuments and itineraries of faith along the Via Francigena enrich the offer of a region where there is ample space for the expression of a traditional culture that makes use of several derivations, as also bilingualism (Italian French) which is part of regional history and the use of other idioms such as the Franco-Provençal and the Walser with dialects of German stock such as titsch and toitschu.

Figure 5: The 7 touristic areas of Aosta Valley



Source: Ufficio Turismo Regione *(s.d.)*

The Region is divided into 7 Touristic Areas: starting from the border with Piedmont, there is the Central Valley and Mont Avic which offers Alpine environments of great beauty, such as the Champorcher Valley and the Mont-Avic Natural Park, but also numerous cultural destinations, castles and finds from the Roman period. To all this is added the leisure offer of the Casino of Saint-Vincent. Following the Mont Rose Valley from the wide massif which the Ayas / Champoluc valley and the Gressoney valley (Lys Valley) leave, they lead to the

discovery of the culture of the Walser populations, villages with characteristic architecture and vast ski areas. Continuing northwards the is the Valley of the Matterhorn where pleasant places overlook, from flat balconies, on Valtournenche; At the bottom of the mountain stands the Matterhorn, at the foot of which Breuil-Cervinia offers one of the largest ski areas in the Alps. The Aosta basin is located in the heart of the region and embraces an area rich in historical remains, including famous castles and important testimonies of the Roman era. From the bottom of the valley, you can quickly reach Pila, an important ski resort, and other places at high altitude, from which you can enjoy splendid views of the surrounding peaks. Looking south to Aosta there is the Gran Paradis, a mountain that exceeds 4.000 meters of the Gran Paradiso National Park which gives its name to this tourist area which includes, in addition to the valleys of Cogne, Valsavarenche, Rhêmes and Valgrisenche, also the slopes that rise from the bottom of the valley towards Mont Fallère and the Vertosan valley. North of Aosta the valley of the Grand Saint Bernard, the Valpelline and the valley of Ollomont that offer harsh and suggestive panoramas, and authentic examples of local traditions; in the summer the Gran San Bernard pass constitutes an historic passage towards Switzerland, an alternative to the tunnel. The last area, but not least, the Mont Blanc Valley, with its 4.810 meters, is the highest peak in Europe. It is part of an imposing mountain range, which dominates the municipalities of the valley floor - La Salle, Morgex, Pré-Saint-Didier, Courmayeur - which, together with La Thuile, form the so-called “Valdigne".

Considering the mentioned area above the topic in this thesis aims to highlight the potential of touristic data that could be used everyday with the use of internet related directly or indirectly to Aosta Valley Region. Using public information, the research has integrated statistical data that is collected by the Regional Office from the Statistic Italian Agency (ENIT). This office performs various tasks, and the mission is to serve the community through the production and communication of high-quality statistical information, analyses and forecasts and these must be carried out independently and on the basis of rigorous ethical-professional principles and the most advanced scientific standards. The aim is to develop an in-depth knowledge of Italy's environmental, economic and social reality at the various territorial levels and facilitate the decision-making processes of all the subjects of society (citizens, administrators, etc.). One of the functions is to deal with the collection of related data concerning a multitude of economic and social areas such as environment and energy, demographic indicators, average household expenditure and respective income, in information regarding health, education, justice and security, agriculture, services, work and wages, tourism. As far as tourism is concerned, a theme that is addressed in this work, the tourist flows of all the types of accommodation are studied: by law, hotels, b&b, apartments, and other type of lodging are obliged to send this information electronically monthly, so that the competent offices can carry out analyses. In particular in Aosta Valley Region, the information once received is stored and divided into the 7 tourist areas where monthly statistics are processed. In order to provide general information on the flow trend, winter and summer as well as annual seasons are also cataloged. At present, tourist information is saved and divided into different tables following a predefined standard that is repeated for months,

seasons and years. Once cataloged, these are used to see the trends of the flows which concern the country of origin of tourists and their destination and to equate them with previous years. This allows to see the differences both from month to month and the difference of tourists in a specific area over time. Mainly the office has 2 purposes: the first one is to give the community and politicians an idea of the markets, their change over time and the second to give people involved in tourism, that they are private citizens and public administrations, basic information that allow them to be able to focus on marketing, infrastructure and service choices. Furthermore, they are used to understand and study strategies to be pursued to increase the attractiveness of the holiday resorts and to guide the choices of travelers: it is important to evaluate the customer loyalty capacity of the tourist products offered.

The data that are currently stored, analysed and used, however, show that there is a gap between the use made of it and the use that could be made of it if only these were studied more efficiently. This is because the analysis focuses on a mere statistical study without carrying out any subsequent step with today's technologies: over the years the data collected has always been the same, albeit in a different way with manual and paper compilations, but his study. In recent years with the advent of Internet 2.0, big Data and the Internet of things, a great deal of information could be collected and entered into study processes with the aim of enriching the amount of knowledge regarding tourists.

The main tables that are used in this are papers concerning the type of accommodation and the number of tourists who stay: in Table 1 there is the list of the number of hotels divided by stars and respective localization.

Table 4: Number of hotels by tourist area

AREA	HOTEL STARS	QUANTITY
Aosta	1 star hotel	5
	2-star hotel	13
	3-star hotel	28
	4-star hotel	4
Gran Paradiso	1 star hotel	2
	2-star hotel	21
	3-star hotel	29
	4-star hotel	4
Gran San Bernardo	1 star hotel	1
	2-star hotel	4
	3-star hotel	6

Monte Bianco	1 star hotel	4
	2-star hotel	17
	3-star hotel	42
	4-star hotel	18
	5-star hotel	5
Monte Cervino	1 star hotel	2
	2-star hotel	16
	3-star hotel	36
	4-star hotel	13
	5-star hotel	1

Table continues

Table 5: Number of hotels by tourist area (cont)

Monte Rosa	1 star hotel	1
	2-star hotel	12
	3-star hotel	36
	4-star hotel	13
	5-star hotel	1
Valle Centrale	1 star hotel	9
	2-star hotel	6
	3-star hotel	21
	4-star hotel	7
	5-star hotel	1

Source: Own Work.

The same thing happens to the other types of accommodation that are not hotels, such as apartments, b&b, guest houses, refuges, campsites. It is noted that all other types of lodging that are not apartments are always grouped in "others" and are not studied in a timely but general manner.

Table 6: Other type of accommodation by tourist area

AREA	TYPE OF ACCOMMODATION	QUANTITY
Aosta	apartments	36
	others	129

Gran Paradiso	apartments	34
	others	101
Gran San Bernardo	apartments	14
	others	64
Monte Bianco	apartments	26
	others	84
Monte Cervino	apartments	38
	others	61
Monte Rosa	apartments	52
	others	78
Valle Centrale	apartments	15
	others	130

Source: Own Work.

As far as general arrivals in the various tourist areas are concerned, these are divided by year as can be seen from Table 3, where there are both arrivals by Italian citizens and foreigners. Also, in this case there are generic "foreigners" and there is no real subdivision by country of origin.

Table 7: Touristic income per area

AREA	ARRIVALS											
	2019		**2018**		**2017**		**2016**		**2015**		**2014**	
	ITA	FOR	ITA	FOR	ITA	FOR	ITA	FOR	ITA	FOR	ITA	FOR
Aosta	113.340	94.103	118.201	101.423	123.609	99.917	118.695	93.830	110.257	85.904	101.511	80.342
Gran Paradiso	119.352	54.567	116.010	53.758	127.459	54.195	120.507	55.464	110.056	51.593	99.863	48.983
Gran San Bernardo	24.384	23.132	24.800	21.663	25.498	22.407	24.436	20.814	23.771	19.196	20.436	16.668
Monte Bianco	202.760	158.347	196.476	153.018	193.161	139.356	193.603	131.067	178.559	115.217	153.349	105.592
Monte Cervino	87.722	81.352	87.819	80.917	85.929	74.113	83.363	70.031	81.143	63.928	70.458	61.259
Monte Rosa	115.488	51.618	111.623	47.986	120.739	48.749	109.698	47.090	99.126	43.144	86.632	38.368
Valle Centrale	101.450	42.678	100.589	39.924	100.055	37.383	96.214	35.996	87.546	30.670	76.796	26.025

Source: Own Work.

The number of tourists staying divided by type of accommodation are described in Table 4, which shows all types of lodging, in addition to hotels and apartments.

Table 8: Touristic flux per type of accommodation

ACCOMMODATION	Stars	ARRIVALS											
		2019		2018		2017		2016		2015		2014	
		ITA	FOR	ITA	FOR	ITA	FOR	ITA	FOR	ITA	FOR	ITA	FOR
Hotel	1	6.184	3.426	5.979	3.184	6.213	3.131	7.519	3.867	7.348	3.882	7.090	3.751
Hotel	2	59.293	32.712	61.510	35.113	64.598	33.906	64.916	31.907	62.751	30.409	58.670	30.004
Hotel	3	230.682	161.953	238.950	170.494	250.813	166.999	242.416	159.400	233.025	152.957	203.363	144.521
Hotel	4	139.968	104.940	130.773	99.381	125.785	93.907	122.947	85.529	106.143	72.615	97.168	62.622

Table Continues

Table 9: Touristic flux per type of accommodation (cont)

Hotel	5	21.362	17.720	15.928	13.942	14.950	12.153	15.148	14.239	14.151	9.744	9.063	5.456
Residence	2	9.496	3.249	8.748	3.601	9.251	3.330	9.104	3.614	7.973	3.145	6.467	3.364
Residence	3	32.390	11.983	33.591	12.396	33.605	10.629	30.517	10.257	29.043	10.024	25.884	9.268
Residence	4	18.933	7.064	20.642	6.654	18.974	6.483	18.207	9.591	16.792	10.309	14.966	9.825
Farmhouse		201	125	8.987	3.994	10.685	3.780	10.589	3.711	10.184	3.084	9.033	2.232
Room Rentals / Chambres d'hôtes		49.172	26.753	46.404	25.526	44.688	21.904	41.758	18.990	36.173	16.621	30.854	14.652
Bed & Breakfast		16.574	6.435	16.483	6.534	16.110	6.442	14.793	5.633	12.223	3.963	9.475	2.524
Camping	1	8.009	2.398	7.602	2.553	8.215	2.330	8.155	2.176	8.068	2.013	7.166	2.083
Camping	2	22.825	19.797	22.188	19.144	26.801	18.193	25.185	17.563	23.477	16.086	19.299	13.974
Camping	3	8.378	9.787	9.875	7.210	10.902	6.528	10.793	6.263	9.462	5.747	8.018	4.777
Resort	2	6.274	759	0	0	0	0	4.567	303	4.025	389	3.602	398
Resort	3	4.787	608	6.322	706	5.771	448	12.975	6.562	9.059	5.450	7.632	5.018
Apartment		19.343	12.417	18.190	10.404	17.191	8.535	29.054	3.420	29.427	3.245	28.104	3.448
Holiday house		32.305	4.044	34.167	4.385	33.325	3.594	7.591	2.355	7.129	1.843	5.672	1.685

Hostel		6.647	1.551	6.935	1.591	8.186	2.217	1.982	2.301	1.799	1.695	1.115	1.051
Posti tappa / Dortoir		1.885	3.390	1.912	3.240	2.033	3.049	36.155	59.761	30.893	51.729	27.593	50.574
Hut		33.630	66.097	34.222	64.211	34.947	64.440	30.515	4.578	29.361	4.194	27.018	5.040
Rest Area		26.437	3.998	24.639	3.516	31.582	3.559	1.630	2.272	1.952	508	1.793	970
Social Camping		1.274	708	1.471	910	1.825	563	746.516	454.292	690.458	409.652	609.045	377.237

Source: Own Work.

This information is compared and represented in graphs with a focus on trends, without however carrying out mathematical studies and data crossing.

Through the collected data it is possible to have:

- for each tourist area, the number of Italian and foreign tourists
- for each type of accommodation, the number of arrivals
- for each country of origin, the number of tourists arriving by tourist area

There is therefore a gap of information, as there is no precise indication of foreign tourists for each type of lodging for each destination area. For this reason, the information has been studied and more precise data have been extrapolated through mathematical formulas which allow for a more detailed analysis.

Thanks to the increase in the number of information, to be extracted and of knowledge from Big Data and the use of it for the improvement of decision-making activities are subject to the definition of processes that allow to efficiently manage and transform datasets that rapidly grow in volume and variety where each phase of each process changes the state and content of the same, helping to convert moles of data still raw in value and therefore to enrich the analytical model of the data. Carrying out this thesis, an attempt was made to first check whether the data collected can be exploited more, and subsequently, it has been studied what the missing data are, how to find them, and what the desired results are.

4 CASE ANALYSIS

The tourism market is becoming much more dynamic compared with a few years ago: on the one hand the prospects, expectations and needs change more and more over the years and on the other hand technology continues to change the way people search on the web. The evolution in the way people plan their holidays, study the places to stay, get informed

on the web of how the location will be, the services offered, or even just the weather are information that are stored. For this reason, some tools are used to assemble and process these data in order to optimise and facilitate companies that operate and offer services such as shops, bars, restaurants, hotels in order to best meet expectations. In this way, the products offered will always be in step with the evolving market.

4.1 Design

The model developed by means of the software has the purpose of having a sort of "starting point", so that, once additional information has been obtained, these can be incorporated to obtain a basis on which to begin to deepen their knowledge: the model is designed by combining statistical data whose objective is to be able to have a clear idea of the type of tourists who have stayed in the Region and to obtain predictive tourist information that can subsequently used by public bodies and private companies. The potential users of the tool could also be the DMO (destination management organisation, or the coordinated management of all the elements that make up a destination such as attractions, access, marketing, human resources, image and prices) who wish to understand the best strategies to adopt for an optimal publication of the contents and evaluate their presence on the market.

In order to be able to analyse the tables from the statistical database, two tables have been created: the first one concerns hotel accommodation with a breakdown by number of hotel stars and nationality of origin per year. The second one, divided into apartments and "others", which indicates all non-qualified accommodation such as B&Bs, camping, guesthouses, refuges, huts. Each of the two tables in turn brings together 2 large categories of information: the annual arrivals divided between Italians and foreigners with an attached type of accommodation and the chosen destination indicating the nationality of origin and the number of people, as shown in the following tables of which under:

Table 10: Arrivals of tourists by nationality and type of hotel by year

ARRIVALS

year	nationality	arrivals	accomodation_type
2014	italian	7.022	1 star hotel
2014	Italian	57.312	2-star hotel
2014	Italian	204.614	3-star hotel
2014	Italian	97.343	4-star hotel
2014	Italian	9.063	5-star hotel
…	…	…	…

Source: Own Work.

Table 11: Destination of tourists by nationality and tourist area chosen by year

DESTINATION

year	nationality	from	arrivals	location
2014	Italian	Italy	81.415	Aosta
2014	italian	Italy	55.406	Gran Paradiso
2014	Italian	Italy	6.072	Gran San Bernardo
2014	Italian	Italy	121.935	Monte Bianco
2014	Italian	Italy	46.765	Monte Cervino
2014	italian	Italy	52.565	Monte Rosa
2014	Italian	Italy	58.613	Valle Centrale
…	…	…	…	…

Source: Own Work.

In order to create a unique table with arrival and destination, a mathematical formula was used as follows:

$$\frac{Arrivals.arrivals * arrivals.destination}{\Sigma\ arrivals.destination} \quad (1)$$

The formula (1) therefore makes it possible to combine in a single table both arrivals divided by nationality with the type of accommodation and the respective destination. We then proceeded to create 2 tables: one concerning hotel and the other concerning other types of accommodation (apartments and "others").

Table 12: Hotels arrivals by nationality, location and type of accommodation

year	nationality	location	accomodation_type	arrivals
2014	italian	Aosta	1 star hotel	1.352
2014	italian	Aosta	2 star hotel	11.037
2014	italian	Aosta	3-star hotel	39.403
2014	Italian	Aosta	4-star hotel	18.746
2014	Italian	Aosta	5-star hotel	1.745
…	…	…	…	…

Source: Own Work.

Table 13: Extra - Hotels arrivals by nationality, location and type of accommodation

year	nationality	location	accomodation_type	arrivals
2014	italian	Acosta	apartment	9112

2014	italian	Gran Paradiso	apartment	6201
2014	italian	Gran San Bernardo	apartment	680
2014	italian	Monte Bianco	apartment	13647
2014	italian	Monte Cervino	apartment	5234
2014	italian	Monte Rosa	apartment	5883
2014	italian	Valle Centrale	apartment	6560
…	…	…	…	…

Source: Own Work.

The data preparation phase is concluded and the analysis with the use of the tableau software can now begin.

4.2 Specifications

This project has several objectives, all focused on the potential of the use of big data within the tourism sector and in particular we want to study the temporal trend of arrivals with the passing of the years, to determine if indeed which are the countries of the most significant and almost all are the accommodation facilities chosen and which have a growing trend, see if the current data is complete for the purpose of the thesis, see if there is some potential coming from the data and in case understand what is missing for a 360° analysis.

Thanks to these two tables, that bring together all the statistical information, it is possible to start having a sort of targeting of the type of tourists staying in the 7 touristic areas of the Valle D'Aosta Region. Specifically, we have the number of tourists available, divided by country of belonging (the 15 major ones of origin) who have stayed in a certain year in a hotel or other types of accommodation. In the case of hotels, the number of stars is also available, while for the other types of accommodation, it is possible to know who stayed in the apartments and who in the category of B&B, guest houses, hostels, refuges. This information is an excellent starting point: it is possible to know which are the major foreign tourist markets, or rather those who have stayed and in what type of accommodation and it's particularly useful in order to be able to understand, albeit with the limits of the case, what the average daily expenditure is as regards hospitality.

Having this information available, we can say that for a complete analysis of tourists' stays, additional data would be needed compared to those currently in possession and that currently they are not collected by tourist offices or by accommodation businesses. These in detail are:

- Type of customer such as singles, couples, families, group of friends: for each type of accommodation, it would be interesting to find out which customers choose it. This hospitality information would allow businesses to understand how the market evolves and therefore what are the types of accommodation facilities that have a higher rate of stays classified by type of customer. If it is possible to know that over the years, a certain

customer has increasingly preferred a type of accommodation to the detriment of another one, the managers of these last ones could think of adapting the rooms thus maximising their potential.

- Number of nights of stay: categorised by country of origin and type of customer. This information can be used by the entire community in a business mean as the type of guest can be correlated to the number of days of stay depending on the season and intrinsically defines its target.
- Expenditure in each type of accommodation: this information, very important in order to be able to target tourists, allows to know the quality standard of the accommodation facilities and the services connected to it. This obviously depends on the cost of the life of the country but, with a range of prices, both for what concerns the hotels and for what concerns the other types of accommodation, it can be possible to obtain a qualitative and objective customer standard
- Average price of accommodation for each area: a range of prices for each type of accommodation, both hotels and extra hotels, defines a standard of the area and therefore allows the workers to understand which are the most profitable categories of customers
- Expenses for extra services such as bars, restaurants, aperitifs: this information, albeit marginal compared to the above, is useful for having some more details of its guests that stayed in the area.

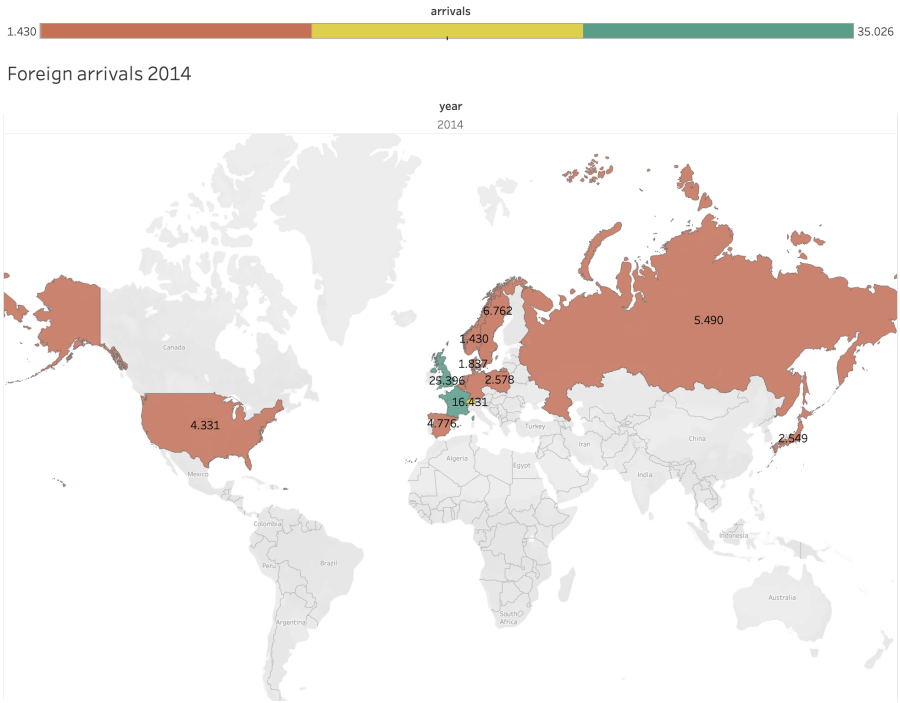
These information's are all closely correlated to each other and it would be desirable for it to be collected with each customer registration. The potential of these data, connected with other information, is certainly an important step in order to have a clear and pragmatic idea of who have been tourists in the area over the months and the years, in case how it has evolved and what are the chosen accommodation facilities: it allows owners or managers of hotels to have certain information on how is evolving and how, most likely, the hospitality market will continue to evolve as it is essential for the area to be able to plan investments both structurally and in marketing, so as to focus mainly on well-defined ones and with supporting data that justify them.

4.3 Implementation

In the first phase of data processing, the goal is to obtain clean and accessible data to be used for the analysis or business processes. The raw data are initially extracted and standardised: the data that will be taken into consideration, the way in which they will be formatted and stored, as well as other key factors that have been defined in the subsequent stages of the process have been defined. Together with this, the deduplication, that is the exclusion and / or elimination of the redundant data, took place manually. Subsequently, before being able to upload the data, these were verified empirically: by carrying out checks to compare similar information. The verification activities allow to further prune the unusable data and to mark any anomalies in systems, applications and data.

The next step is to load the data into Tableau software: this allows to make requests to the database and, through plots, to get answers. First of all, one wonders the trend of arrivals from 2014 to 2019 within the Region both for what concerns the hotels and for what concerns the other types of accommodation. At a purely geographical identification level, it has been decided to work first on the data relating to arrivals in non-hotel accommodation (apartments, bed & breakfast, room rentals, camping's, ext.), that are analysed for the years 2014 - 2016 - 2019. This shows the general variations which will then be analysed in more detail.

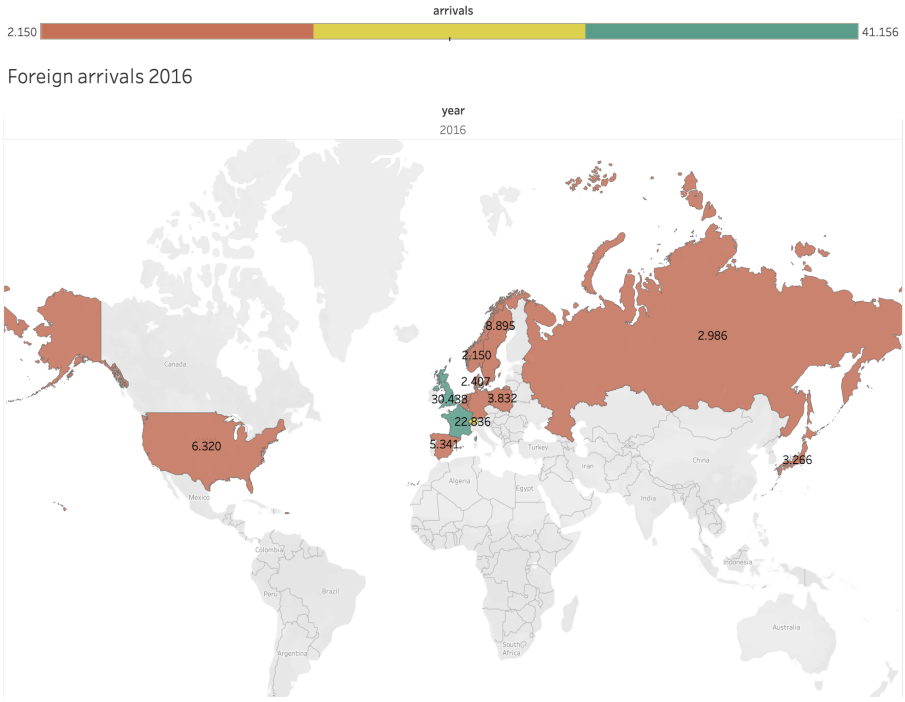
Figure 6: Extra hotel Arrivals in Aosta Valley in 2014



Source: Own Work.

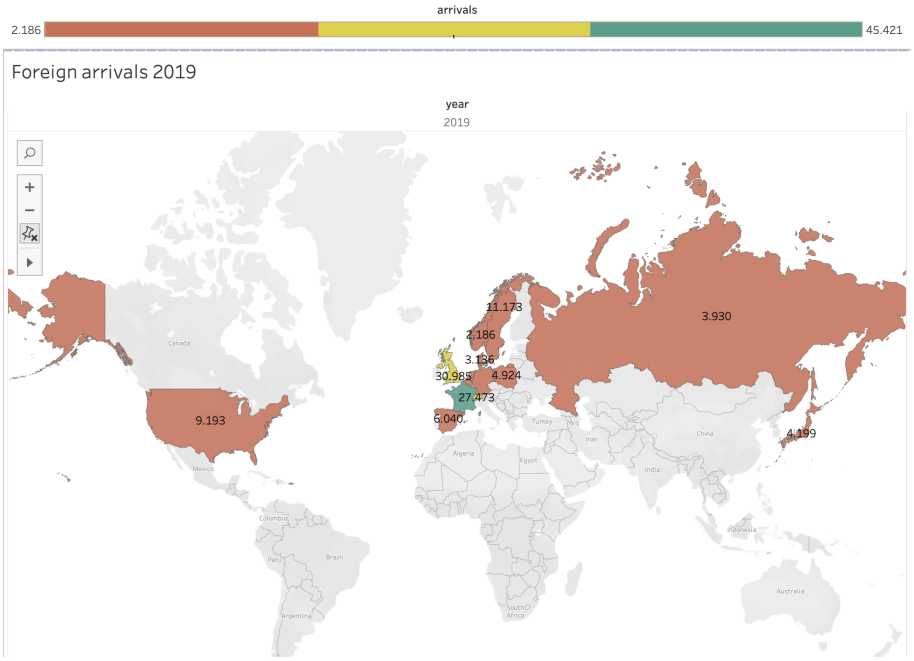
The 15 main countries of origin are highlighted and through a colour scale (brown, yellow and green) that shows which ones have the greatest arrivals in 2014. The same thing happens for the years 2016 and 2019.

Figure 7: Extra hotel Arrivals in Aosta Valley in 2016



Source: Own Work.

Figure 8: Extra hotel Arrivals in Aosta Valley in 2019



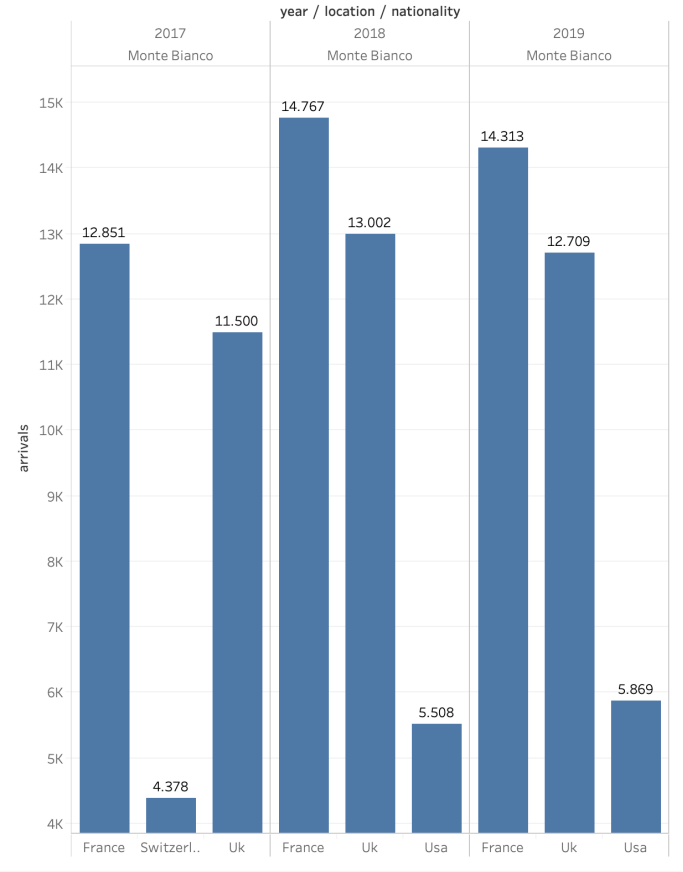
Source: Own Work.

As mentioned, this type of data analysis must try to be as punctual as possible and go to find out for each single tourist area which are the main destinations by country. For this reason, we proceeded to choose a destination, the Mont Blanc area, and to study it in greater detail.

The information requested concerns:

- which are the countries with the largest arrivals in the year 2019, excluding Italy
- considering the main 3 countries of foreign origin which are those that have had a higher arrival rate in the last 3 years
- what were the accommodations that with the greatest arrivals in the year 2019 and compared to the previous 3 years

Figure 9: First 3 country in extra hotel arrivals in Mont Blanc Area 2017 - 2019

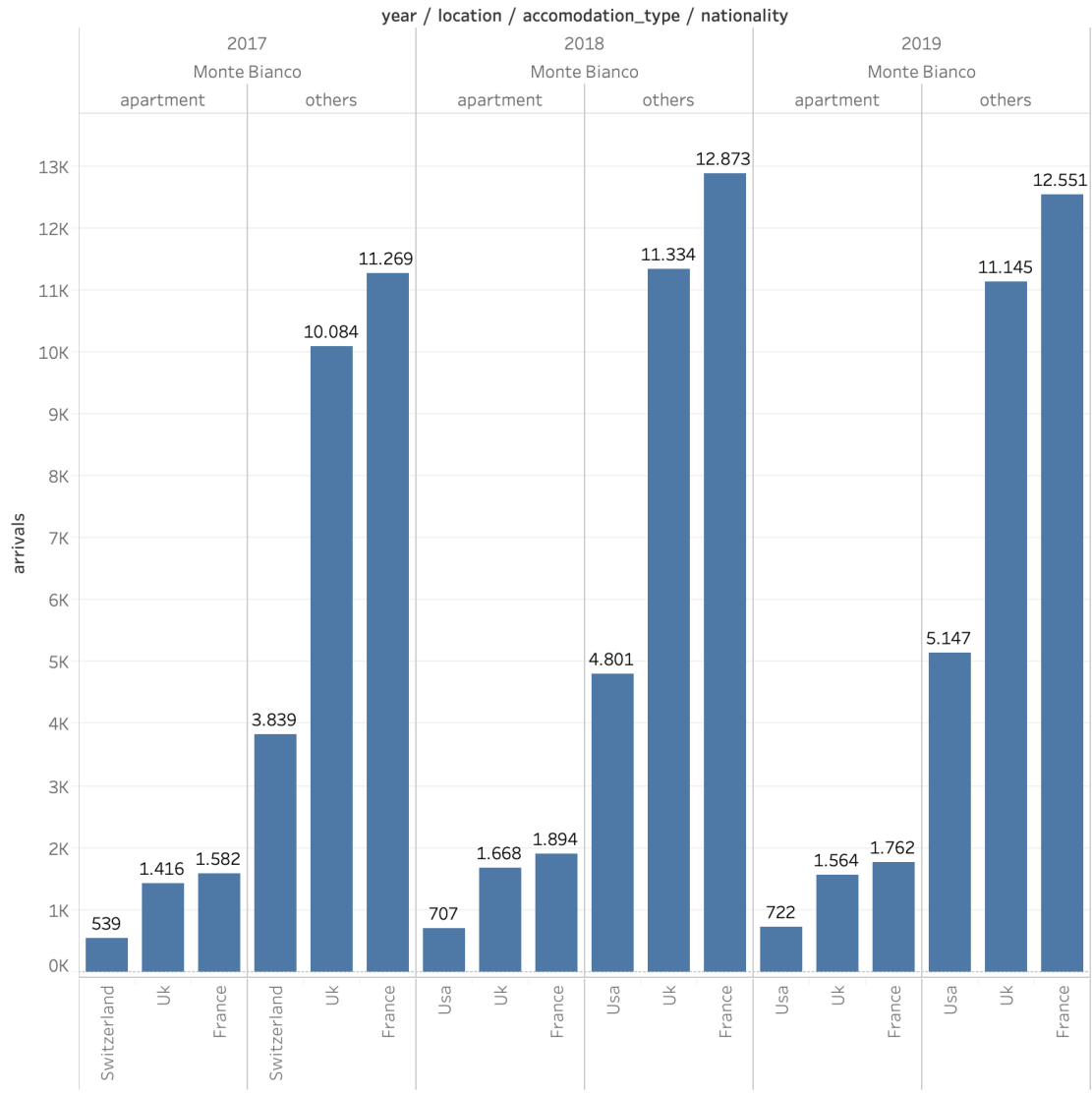


Source Own Work.

From the plot we can see that the largest country of origin in the 3 years is France, where there has been an increase of 13% from 2017 to 2018 and a small decrease of 3% from 2018 and 2019. Increases or decreases below 5% can be considered negligible and that do not affect the analysis: values above this percentage should instead be analysed and to verify what reasons are behind this variation. In this graphic you must always keep in mind that it refers to just one area of the Region and not to the whole tourist areas. As regards the second country of origin, England is confirmed for both years, with values slightly lower than France. Also, in this case, there is an increase between 2017 and 2018 and a slight decrease in 2019. As the third country of origin, we see Switzerland in 2017, which is surpassed by the United States in the following two years. It should be noted that unlike the other countries

considered, in the year 2019 the United States experienced growth, albeit modest. More in detail, having taken into consideration the data concerning the non-hotel accommodation facilities, let's analyse which are the chosen accommodations:

Figure 10: Type of accommodation chosen for first 3 country



Source: Own Work.

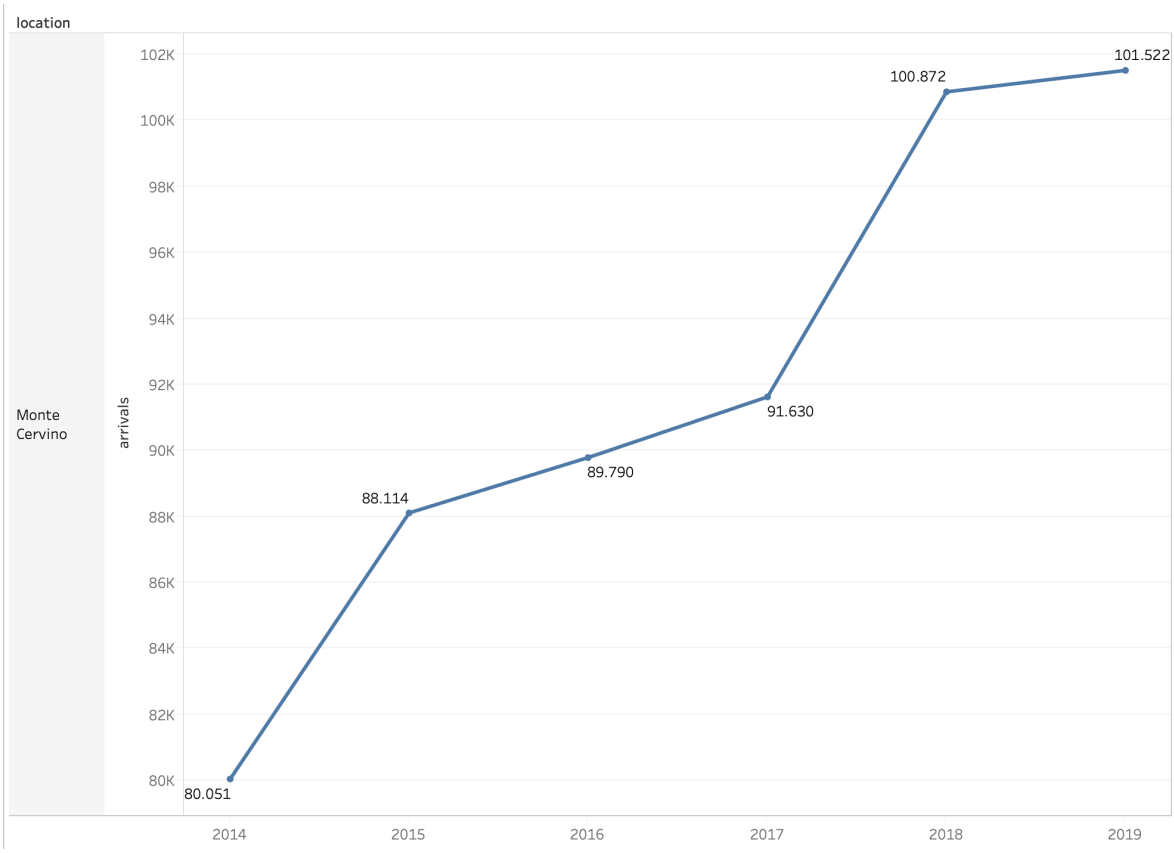
As it is possible to see, in all cases the accommodated "others", that is B&Bs, guest houses, campsites, refuges, hostels are definitely superior to the apartments. As regards apartments, France and the United Kingdom have a slight growth in 2018 and a decrease in 2019, which as mentioned is within a negligible percentage range and it is therefore believed that these values are interpreted as constant in the 3 years. As far as the United States is concerned, we also see in this case a very slight growth, absolutely negligible between the years 2018 and 2019.

If we consider the "others" instead, we can see that for France there was an increase of 13% between 2016 and 2017 and a decrease of 2.5% between 2017 and 2019. If we analyse the 2 other countries have the United Kingdom which increases by 11% in 2017 and remains almost constant in 2019, for the United States there has been an increase of 7% between 2018 and 2019.

This type of analysis can be conducted for any country of origin and respective destination, allowing those who will read the results of these research to have clear ideas on how tourism markets evolve in the hospitality market.

Likewise, the market of the main 4 countries of origin in the hotels in the Monte Cervino area is analysed below.

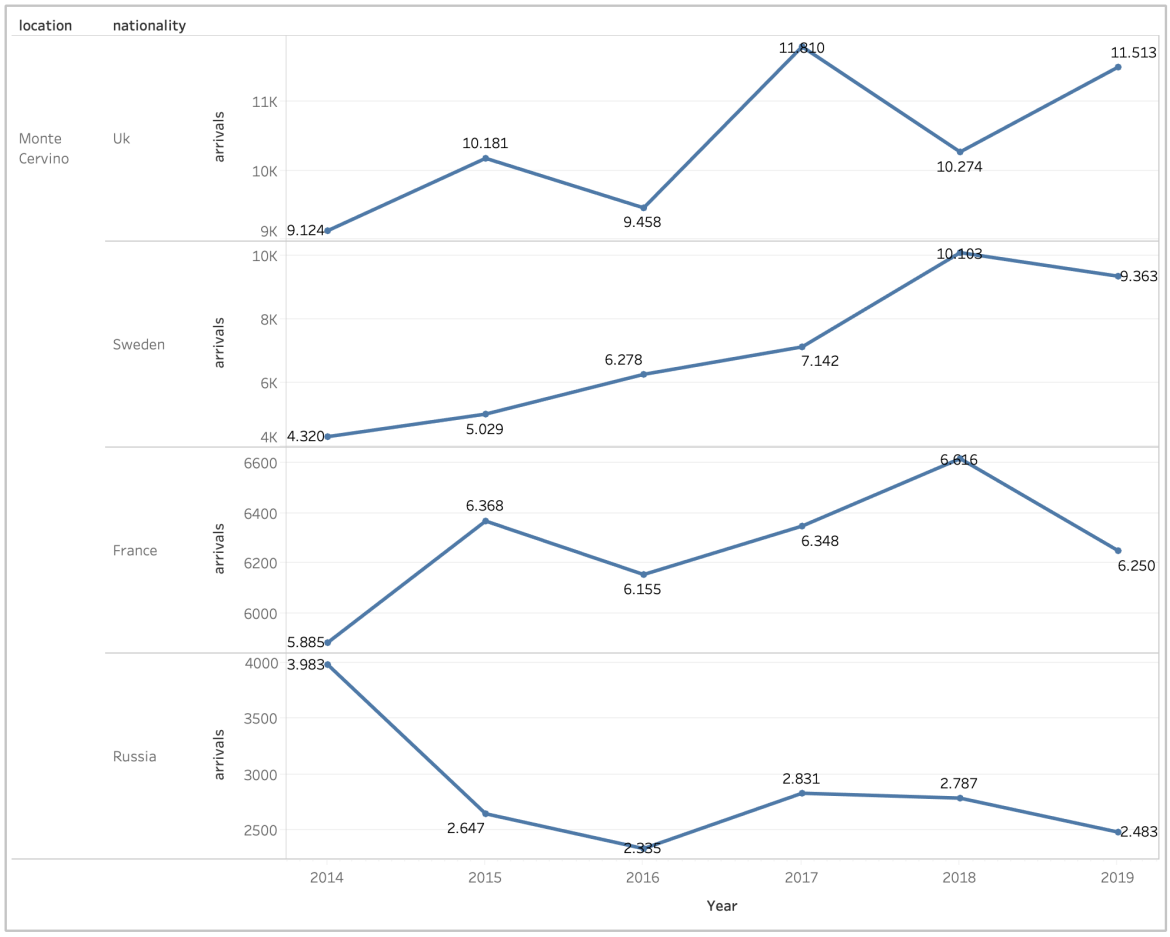
Figure 11: Arrivals in hotels 2014 - 2019



Source: Own Work.

In this first infographic we can see that over the years there has been a continuous increase in tourists in the Monte Cervino area. Between 2014 and 2015 the difference is around 9%, while in the following 3 years the growth remained fairly stable until reaching 2018 where growth was again 9% and then reaching 2019 with negligible new growth. As in the previous analysis, we request which countries have the largest arrivals and respective types of hotels.

Figure 12: Best 4 country arrivals in Monte Cervino Area

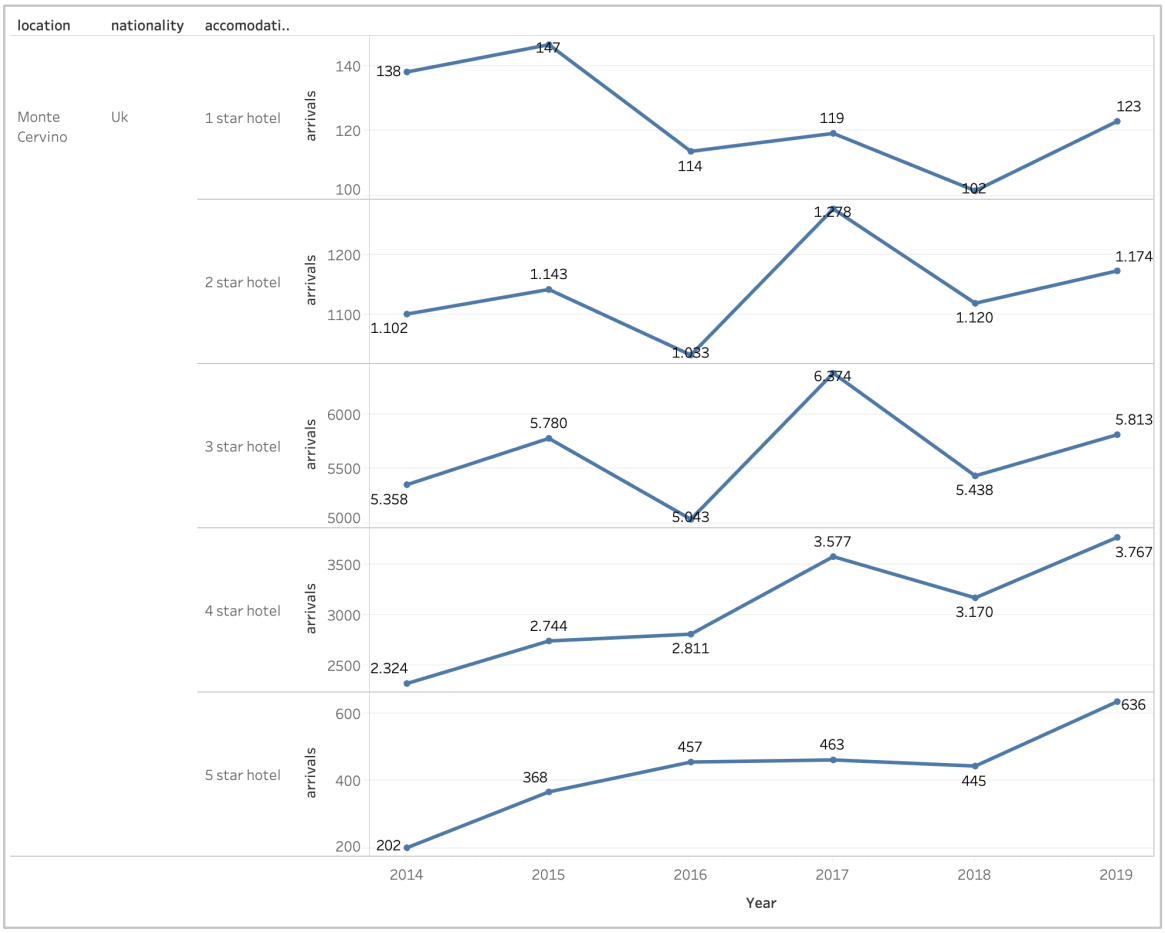


Source: Own Work.

In this infographic, unlike the previous one concerning hotel accommodation, we see that there are substantial differences: first, the country with the largest arrivals, the United Kingdom, has constant fluctuations from year to year, Sweden, second country arrivals have continued growth with a small decrease in the year 2019. France, as well as the United Kingdom undergoes fluctuations and Russia unlike all the other countries decreases. A particularly interesting data by comparing this table with the previous one, shows that the general variations, i.e., of all the countries of origin, with those of the main 4 states, do not match and it can therefore be said that the other 11 countries of origin have a considerable incidence on the total arrivals.

At this point if we analyse the market with the largest arrivals, the United Kingdom, we can see which types of hotels are chosen by the majority.

Figure 13: Type of hotel chosen by United Kingdom tourists in Monte Cervino Area



Source: Own Work.

As we can see, the type of hotel chosen by tourists from the UK in any year are 3- and 4-star hotels: for both of us we have a similar trend with a prevalence of choice over 3-star hotels. This factor is also attributable to the fact that in the Monte Cervino area the number of 3-star hotels is 36 while that of 4-star hotels is only 13: the data certainly affect the number of beds and consequently the capacity between the 2 types.

If we analyse the trend of arrivals in 3-star hotels we note that there are considerable variations from year to year: between 2014 and 2015 there is an increase of 7.5%, followed by a decrease of 13%. The following year there was a 21% growth and a 15% drop in 2018 to end with a 7% increase. It would be interesting to understand what the reason behind these considerable variations with the use of additional data is that implement those already analysed.

As for 4-star hotels, we see that between 2014 and 2015 an increase of 15% followed by a stable + 2%. Following further increase of 21.5% which in 2018 decreases by about 12% and then increases by 16% in the year 2019. Also, in this case we have considerable variations from one year to the next, the causes of which would be to be understood and

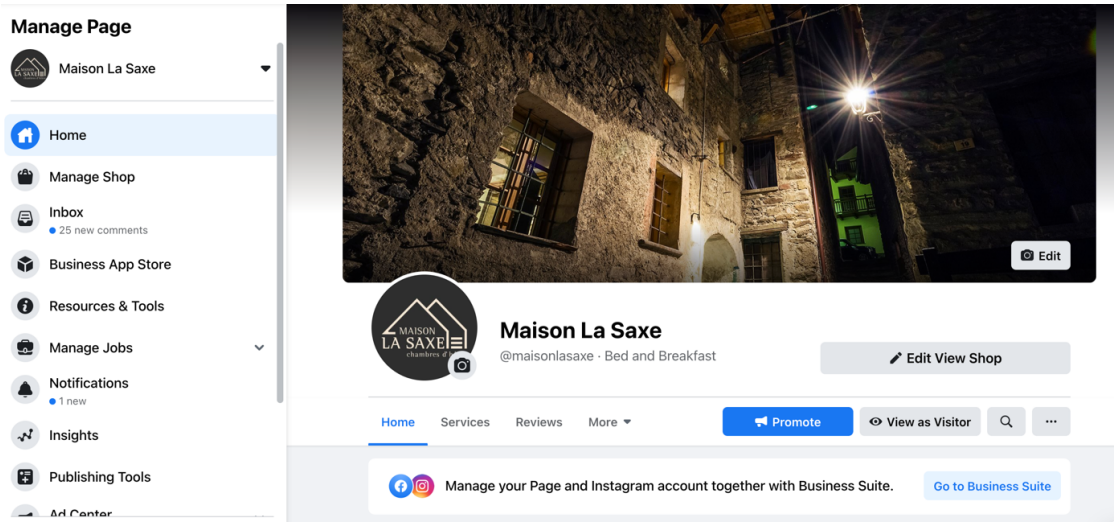
analysed. As we can see from the infographic, this happens on almost all types of hotels, but only for the UK tourist target, while for other tourists this does not happen. Looking at the second reference market, Sweden, we notice a continuous growth from 2014 to 2019, with a slight decrease in the last year. If we take at this point the general infographic of the Monte Cervino area, we will see that, compared with the main reference market, the United Kingdom, there are substantial differences: the number of tourists increases year by year, while as mentioned English tourists, increase and decrease do not follow a stable growth curve.

4.4 Social Network Analysis

Two tools have been used during the social network analysis: Facebook business suite and Facebook Insights.

First you open the Facebook page "Maison La Saxe" to get a general idea of what will be analyzed.

Figure 14: Maison La Saxe Hotel Facebook page



Source: Facebook inc.

And the general "insights" statistics are analysed over a period of 28 days, in this case December 2020. The data that are useful for the purpose of the study are:

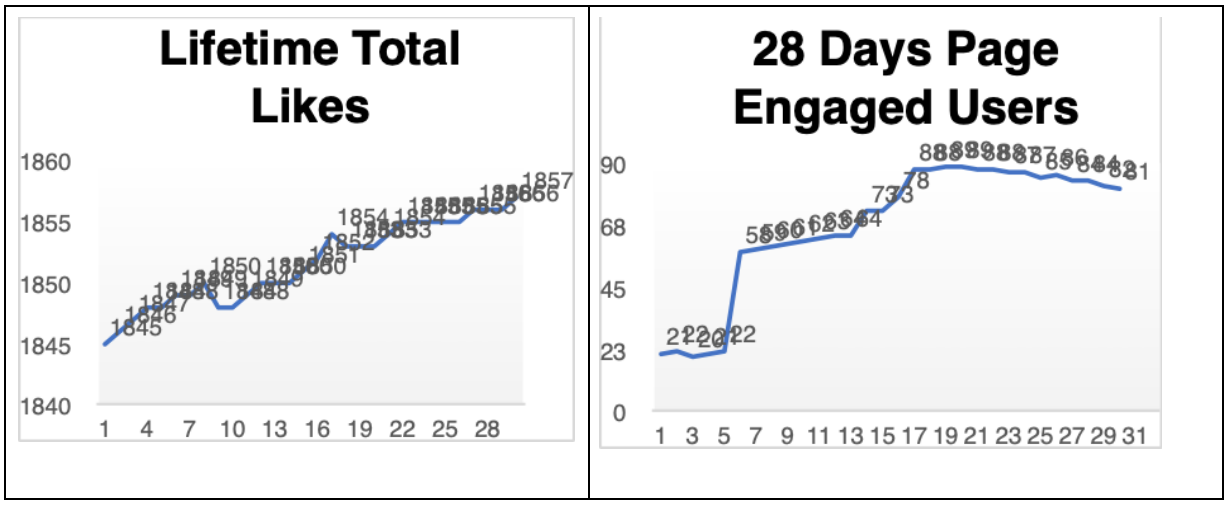
- Lifetime: the total number of people who have liked the Page as unique users
- The number of people who engaged with the page. Engagement includes any click or story created as unique users
- The number of people who had any content from the page or about the page enter their screen through unpaid distribution. This includes posts, stories, check-ins, social information from people who interact with the page and more, as unique users

- The number of people who had any content from the page or about the page enter their screen through with social information attached. As a form of organic distribution, social information displays when a person's friend interacted with the page or post. This includes when someone's friend likes or follows the page, engages with a post, shares a photo of the page and checks into the page, as unique users
- The number of times any content from the page or about the page entered a person's screen through unpaid distribution. This includes posts, stories, check-ins, social information from people who interact with the page and more as total count
- The number of times any content from the page or about the page entered a person's screen with social information attached. Social information displays when a person's friend interacted with the page, post or story. This includes when someone's friend likes or follows the page, engages with a post, shares a photo of the page and checks into the page as total count
- The number of people who had any of the page's posts enter their screen through unpaid distribution as unique users
- The number of people who had any of the page's posts enter their screen with social information attached. As a form of organic distribution, social information displays when a person's friend interacted with the page or post. This includes when someone's friend likes or follows the page, engages with a post, shares a photo of the page and checks into the page as unique users
- The number of times the page's posts entered a person's screen. Posts include statuses, photos, links, videos and more as total count
- The number of times the page's posts entered a person's screen through unpaid distribution as total count
- The number of times the page's posts entered a person's screen with social information attached. Social information displays when a person's friend interacted with the page or post. This includes when someone's friend likes or follows the page, engages with a post, shares a photo of the page and checks into the page as total count
- The number of people who clicked on any of the content. Stories that are created without clicking on the page content (ex, liking the page from timeline) are not included as unique users
- The number of clicks on any of the content. Stories generated without clicks on page content (e.g., liking the page in Timeline) are not included as total count
- Number of times a video has been viewed due to organic reach as total count

These metrics are intended to be able to understand the value of the page and if this in addition to having a good number of likes is also positioned in the market correctly. There is no precise formula for obtaining this data, just as there are no online tools that calculate the value of the page: there is no precise method that can say with certainty how much a page is worth. The value of the Facebook page is given by the so-called conversion to the goal, that is, if you have a company Facebook page and have many likes, comments, interactions, clicks and shares but no one asks for a quote, its value is almost nil. The value

is expressed by the number of contacts that turn into something really useful for real or professional life, such as a deal, a quote, a contract, a job. Having a nice site and not having interactions is the same thing, its value in this case is significantly reduced so the Facebook page and site must interact in order to produce advantages but above all conversions to the goal. If you want to establish the value of the Facebook page, the question that needs to be asked is: how many people have contacted in a given period to give you the opportunity to sell something or how many have even requested a quote? These are the questions that must be answered to understand if the Facebook page is of value or not, having a high number of fans without them creating engagement and conversions towards your business goals is equivalent to having a page with very little value. To calculate the value of the Facebook page it will be necessary to evaluate the number of fans but above all how many of these create interactions and convert into useful actions for the business.

Table 14: Facebook likes and engagement



Source: Own Work.

Interesting is the calculation that verifies over a certain period of time, if the page has a good engagement compared to the total number of likes.

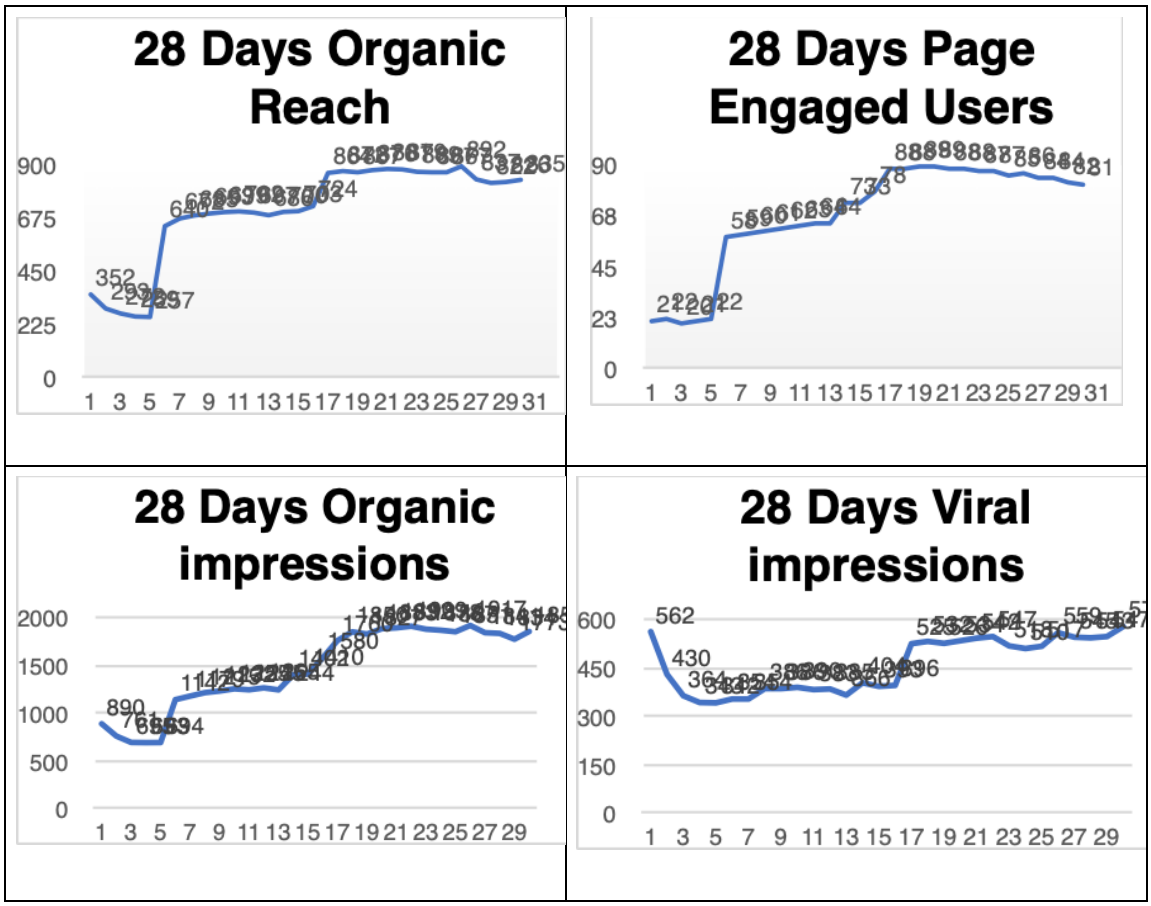
$$\frac{PTPA}{N.ofFans} * 100 = \frac{81}{1857} * 100 = 4{,}36 \qquad (2)$$

Where PTPA is the people talking about the page: this parameter shows that the page has a good engagement compared with the number of likes as it has to be more than 2. Many companies have many likes, but these are not followed by a high engagement and in fact makes the page at least useless and needs to review its marketing strategy.

$$4{,}36 > 2{,}00 \qquad (3)$$

Some of the most important statistics are shown in the tables below, which allows to have a clear idea of how it has evolved, in the previous 28 days:

Table 15: Facebook statistics

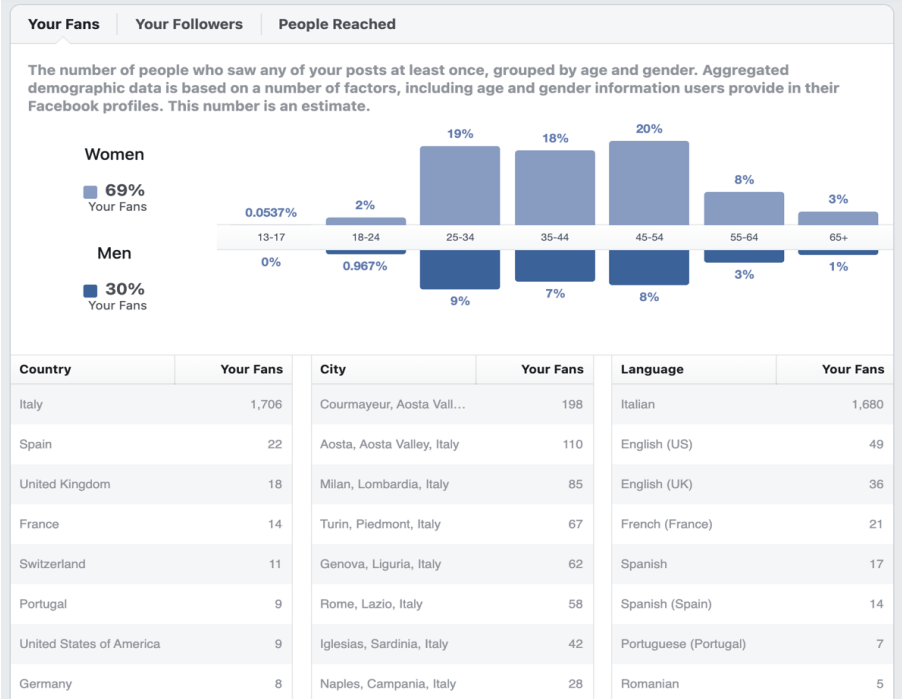


Source: Own Work.

As it can been seen from the tables above, the indicators in the 28 days are almost all growing, that is, it shows that the page is in good health and continually attracts new users who subsequently talk about it, share and interact with it.

Unfortunately, it sometimes happens that the fan base is not exactly what expect: for example, for a hotel it is not useful to have a majority of fans who live in the immediate surroundings or who belong to nationalities that are not at all in line with the target audience of the destination and it is on the basis of these considerations that the revision work begins. The aim is to attract an audience in line with the business objectives. To do this, an editorial strategy capable of involving the right people interested in the products or services offered by the company need to be created.

Figure 15: Demographic data Fans



Country	Your Fans
Italy	1,706
Spain	22
United Kingdom	18
France	14
Switzerland	11
Portugal	9
United States of America	9
Germany	8

City	Your Fans
Courmayeur, Aosta Vall…	198
Aosta, Aosta Valley, Italy	110
Milan, Lombardia, Italy	85
Turin, Piedmont, Italy	67
Genova, Liguria, Italy	62
Rome, Lazio, Italy	58
Iglesias, Sardinia, Italy	42
Naples, Campania, Italy	28

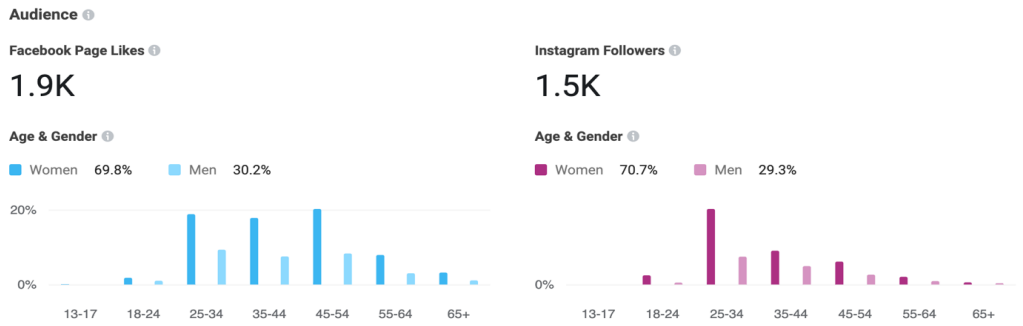
Language	Your Fans
Italian	1,680
English (US)	49
English (UK)	36
French (France)	21
Spanish	17
Spanish (Spain)	14
Portuguese (Portugal)	7
Romanian	5

Source: Own Work.

As expected, most of the fans are resident in Italy, whose prevalence is on the place where the hotel resides, with Courmayeur, the city where the hotel is, with a high number of fans. If we look at the gender aspect, the high prevalence is by women with almost 70%: this factor is very important because it points out that women mainly look at this hotel and most of them are in the 25-54 age. In most Italian families, the managers of the house are still women, who manage the various expenses for furniture and lifestyle, cooking and entertainment, caring for the interests of each member. This is a responsibility that requires purchases thought out in every aspect and the evaluations of women's purchases are called 'spiral': more articulated than the linear thought that generally guides men's, more impulsive and hastier. It seems that women, considered as a pool of customers, tend to evaluate several factors before making a decision that includes price, functionality, safety, durability, but also ethics and sentiment. This commitment of time and energy leads to greater brand loyalty: the female world is generally more predisposed to this type of chatter and at the same time gives more credit to third parties to complete the idea about a particular brand.

If we compare the metrics obtained by Facebook with those of Instagram, we will see that:

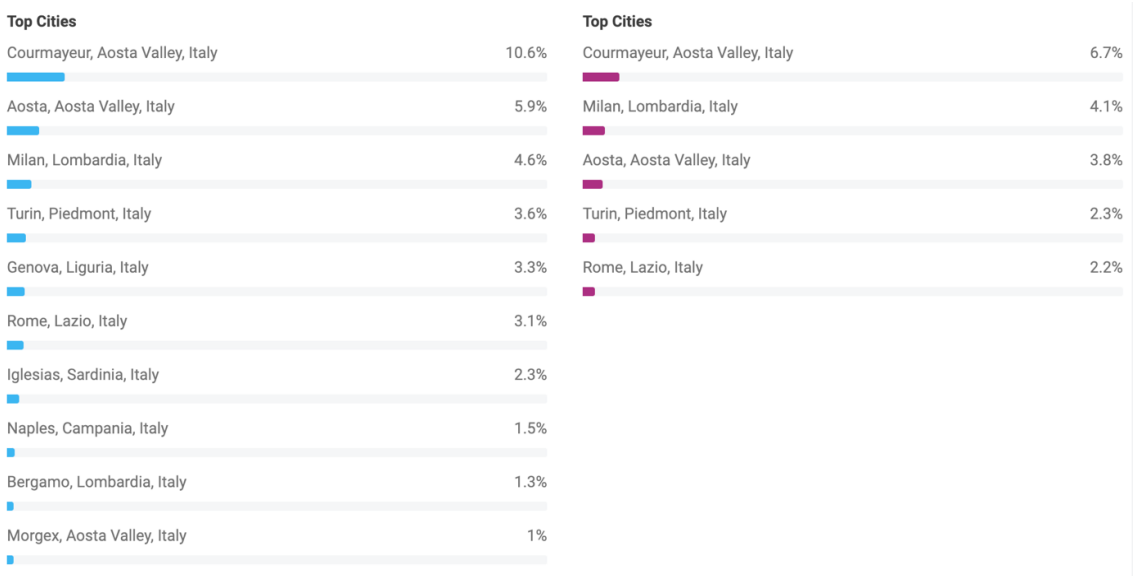
Figure 16: Facebook and Instagram Audience



Source: Own Work.

There is likely to be a similar feedback, with both curves growing and a marked distinction between men and women with similar percentages of 30% -70%. A particularly interesting fact that differentiates the two social networks is the age: in Facebook the age range of users is constant between 25 and 54 years, while in Instagram the age range is mostly in the 25-34 range, while the others are clearly inferior. This data can also be referred to the general target of users on the platform, consisting mainly of younger people and more inclined to social media of this kind.

Figure 17: Facebook and Instagram Top Cities



Source: Own Work.

If we compare the origin, we will also have a high number of people on the Instagram page who are in the municipality of Courmayeur with about 10% for what concerns the Facebook page and 7% for what concerns the Instagram page. These percentages, if contextualized in the entire market, indicate that only a small percentage of the audience lives in the locality, while the rest is diversified in a fragmented way.

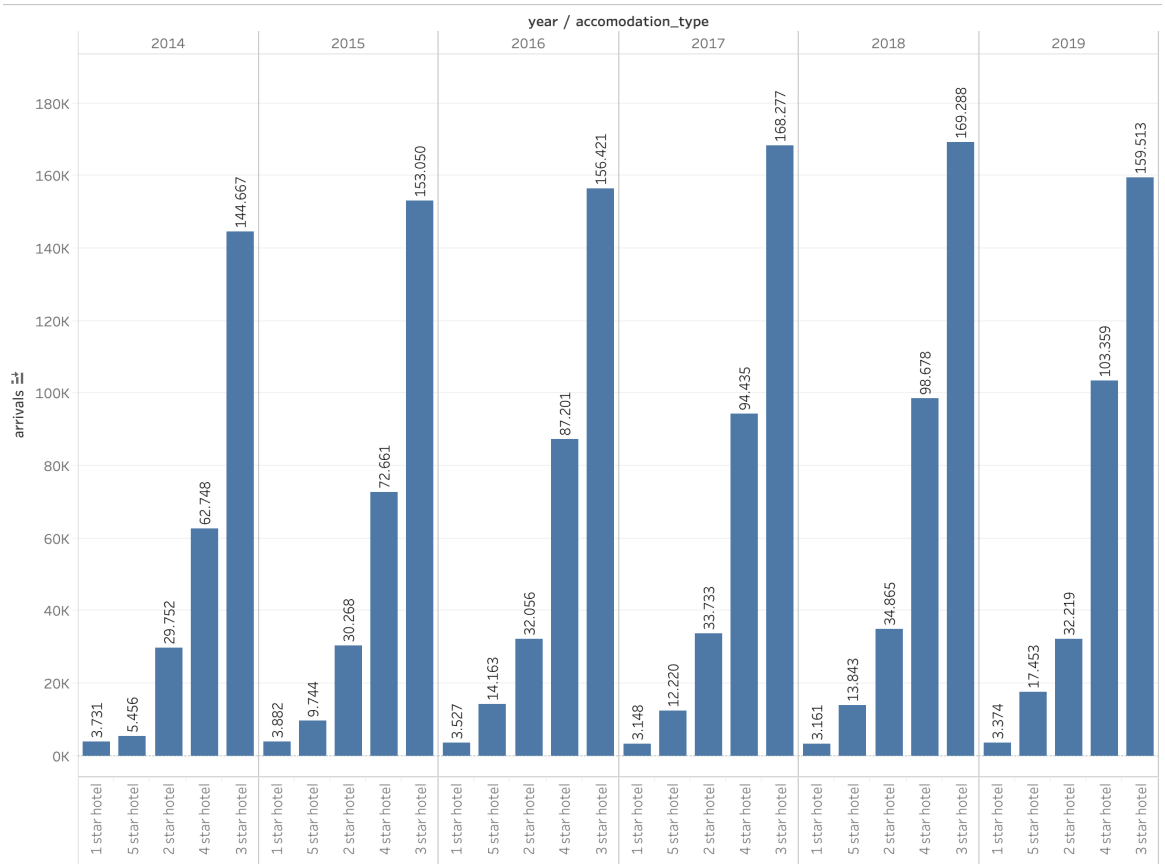
5 DISCUSSION

After collecting the data, manipulated and processed in a single dataset, it is possible to answer the questions that have been set as goals.

5.1 The Data Analysis

First, it has been classified the tourists who stayed in the hotels between 2014 and 2019: from the elaborated information it shows that the type of tourist who stays generally in the Region prefers 3- and 4-star hotels. At a general level, from one year to the next one, there are increases with regard to 5-star lodges and decreases on 1 and 2-star hotels. This leads to think that the type of customers over the years are more affluent and have a growing spending capacity. The destinations where most tourists go are Mont Blanc, followed by Aosta and Gran Paradiso. After them it's the turn of the Monte Cervino, Monte Rosa and the Central Valley, and in the last instance the Grand San Bernardo. There is an increase in tourists that somehow follows the growth trend of the hotels: in the following image, it is already evident from the year 2014 that 3- and 4-star hotels are largely preferred by most tourists.

Figure 18: Type of hotel selected in 2014 by nationality



Source: Own Work.

With this statistical information it is possible to make a first sort of targeting, being able to affirm the type of accommodation chosen by the majority of tourists over the years. This phenomenon has grown ever more and then decreased by a few units in the year 2019. As said, it is also noted that, there has been a large increase in the number of people in 5-star hotels: in 6 years, it has gone from 5456 at 17453 units, positioning itself as the best growth rate compared to all other types of hotels.

As regards the type of tourists, it is possible to make a classification based on the accommodation facilities, both hotels and other types of lodging, with the prevalence of those chosen for each year. Concerning other types of lodging, those chosen in any year analysed are B&Bs, guest houses, hostels and with a much smaller number there are the apartments. However, it should be noted that this last data could be distorted, as the apartments given for tourist rental, commonly called “Airbnb", are not counted and that they are not included in the official ISTAT data because they are considered private lodgement and not as an entrepreneurial activity and is therefore not included in the lists. This factor considerably influences the veracity of the study as regards the apartments and for this reason, it is believed that the information regarding the apartments is not reliable since the exact and complete number of the entire dataset is not available. For this reason, it is believed that the information regarding the apartments is not reliable since the exact and complete number of the entire dataset is not available, so it’s not considered.

During the first year analysed, in 2014, the number of Italian tourists was clearly higher than any other foreign country of origin. This is followed by the foreign countries of France, the UK and Switzerland. It can be assumed that the French-speaking countries are in considerable volumes probably due to their geographical proximity, being neighbouring and being able to arrive with its own means of transport. If considering the hotels, it can be seen that the typologies chosen are the 3 and 4 stars, with a great prevalence for the first ones; This general information can be studied in greater depth by going to analyse the 7 tourist areas from which it is possible to notice those with more flows and which attract more people in a monthly or seasonally base. In this case, it should be noted that the reference target is mainly Italian and as regards foreign countries, most of the income refers to neighbouring countries.

Targeting is the selection of one or more segments of consumers who consider themselves appropriate for their business and the identification of the products suitable to meet their needs: to be able to apply it to the data processed, it is necessary to see the three main approaches. Mass Targeting refers to the implementation of a single strategy with which we intend to satisfy the whole market or a large part of it and is the lowest cost strategy since a single product is offered on the whole market globally. Its weakness lies in the fact that a single product can hardly satisfy the heterogeneous needs of people with different ages, cultures, spending capacities and lifestyles. Mass marketing is suitable for products that require few adaptations from one market to another and that can be communicated with simple and universal messages. The target marketing provides specific strategies and

products designed for each market segment identified and is the most expensive and demanding method to implement but allows you to accurately meet the specific needs of each reality. To implement such a strategy, it is often preferable to have an on-site presence to get to know the characteristics of the target segment in depth. Mass customisation marketing can be considered a middle ground between the two previous ones, with a series of products that with the necessary adaptations manage to satisfy all segments of the market. With the available data we can therefore carry out a mass targeting: this allows first of all to know which type of tourist are referring to and what the needs are. With the available data we have carried out a mass targeting: this allows first of all to know which type of tourist are referring to and what the needs are. Subsequently, with the average cost of each type of lodging available, it is possible to know the shopping range of the various tourists in each individual area.

To carry out a market targeting study, it is needed to find further information which is not currently considered: it is necessary to define the target, which is the goal that must be achieved with communication and the marketing plan. In other words, they are the potential recipients of an advertisement, as well as the potential buyers of the products or services. To identify it, we proceed with a segmentation of the market and with the subsequent choice of one or more segments to which to turn, depending on the objectives. Market segmentation does not lead to the definition of a single target, but of different "target groups": a set of consumers united by similar elements such as age, habits, income, etc. For this reason, for example, the target of the marketing sector may differ from that of communication: in the case of marketing, these are the consumers that you intend to keep or acquire; in the second case, it is the users of your communication. Furthermore, in terms of optimising efforts and resources, once the reference target has been reached, this will probably also influence further market segments, through word of mouth or one's own behaviour. Identifying the target means finding a niche of people really interested in what you want to offer them: people who will benefit from what you do because, by identifying with them, they will respond to their real needs. Knowing the interlocutors also allows them to be coherent, that is, to do and offer what your customers expect, without dispersing resources in actions that risk proving unsuccessful. Defining the target means, in fact, understanding the characteristics and tastes of a specific market segment: who the customers are, their characteristics, what drives them to purchase, what opinion they have of the brand, their level of satisfaction and so on. This operation allows to adopt the best marketing and communication strategies, in order to achieve the set objectives. In other words, knowing the ideal customer allows to reach him and turn him into a buyer. Of course, the target will have to be studied on an ongoing basis, as markets and consumers are constantly changing, as well as everyone's needs and habits. By keeping up, you can adapt strategies to market trends.

The tourist information necessary to carry out, a complete study which concerns the hospitality department are the following:

- number of nights of stay by type of stay
- average price of each accommodation facility in each area
- type of lodging booking made, whether directly or through OTA
- average expense for extra services inside the lodging
- type of customers: singles, couples, families, group of friends, for each accommodation facility and geographical area

This basic information could be collected as of now without great use of resources and would allow for meticulous targeting of customers. Thanks to the use of the mentioned missing data, it is possible to carry out a target marketing study referring to the accommodation sector, showing which are the most profitable segments so as to be able to define which markets to orientate on and how to adapt the services offered. The analysis allows, at the hospitality level, to study which are the most profitable markets by type of accommodation: for each nationality, it is possible to have information regarding the selected accommodation facilities, those that have had major increases or decreases and hypothesise how it will evolve the following year thanks to previous data. If, at this point, a more detailed analysis is desired, there is some information which at present cannot be found. The industry of tourism thrives on information: big data can deliver up-to-date and immensely informed inferences regarding behaviour and human activity that enhances the tourism industry. Tourists leave various digital traces behind when using mobile technologies on the web and through every tourist, enormous amounts of data are present about everything that is relevant to different stages of travel - before, between and after a voyage. Most of the information is external in nature such as social network feeds or in the form of Twitter. In the following table we can see the data that should be gathered in this type of research for a more punctual and categorises analysis:

Table 16: Data for a 360° analysis

Sentiment analysis and profiling from social networks
How the booking of the accommodation was made, through OTA or direct booking: calls, email, website
How the area was found: was known before, through friends, via online advertising, advertising in specialised magazines etc.
How many days of holidays: for each season and month based on the nationality of origin
Why this location was chosen: type of services offered, panoramas, places of interest
Type of vehicle used to arrive: private shuttle, shared shuttle, car rental, owned car
Once in the area, what customers do during the winter season: does everyone go skiing? Do they do cross-country skiing?
How many are the access of foreigners and Italians, divided by nationality who access the ski slopes

Table continues

Table 17: Data for a 360° analysis (cont.)

How many ski tickets are purchased online and on site
What are the most used services in the area: local bus, mobile app
How many days they ski: whether they ski every day of their stay or not
When they ski what do they stop to eat: sandwich or restaurant
When they go skiing do they rent the equipment, where they come from
Dinner in hotel / restaurant? After they go out?
During the summer season, what do tourists do? Do they walk, relax, bike, e-bike?
How is the typical day going? Tracking points of interest
Are they customers who return or always change?
What are the services that are used and what are the services that are missing that should be implemented with customer feedback
How much customers spend in hotels, restaurants, bars, shops, in percentage and net values
Reviews of places, both public and private services

Source: Own Work.

All this series of data could be collected by different figures, both within the accommodation facilities and through commercial agreements with companies operating in the financial transaction market. In support of this, an entity that collects information would be needed so that it could be analysed.

Assuming that it can be disposed of the above data, an analysis in its entirety goes to cover all the areas in which tourists are involved: from the search for information on the location prior to arrival, in search of accommodation. Subsequently, once on site the data collection allows timely information on the type of customer (singles, couples, families, groups of friends), the chosen accommodation, how much is the average expense in the accommodation, bars, restaurants, stores, and if they make other extra purchases. To this information, during the winter season and for better targeting, it should add data concerning winter sports, and therefore access data to the ski facilities, the number of days of skiing, whether the equipment is rented or not, the type, how much is the average expenditure on the district. And again, if the user buys a sandwich or prefers to stop at the restaurant and how much is the average shopping.

It is estimated that most of the data in the world is not structured and not organised by default and it comes from text data, such as email, support tickets, chats, social media, surveys, articles and documents. These texts are generally difficult, lengthy and expensive to analyse, understand and order. Sentiment analysis systems allow companies to make sense of this sea of unstructured text by automating business processes, obtaining useful information and

saving hours of manual data processing, in other words, making teams more efficient. Thanks to this it is possible to identify people to improve services: by capturing customers who feel strongly negative towards the product or service, customer service can manage problems in a specific way. It is also possible to monitor customer sentiment over time, or to monitor customer sentiment associated with specific aspects of the business is more effective than simply monitoring the NPS. If you matched demographics and other quantitative data, you could segment your customer base and consider their feelings separately. Another aspect to consider is keeping track of how a change in the product or service affects the way customers feel and how the company changes. Sentiment analysis has become known as one of the most reliable tools to effectively listen to social media chatter about a brand and helps to know what people are saying, to collect collective opinions and to say if action must be taken to maintain perception of a positive brand.

With this given data, empiric hypothesis of this data and final result, assuming to be able to use the aforementioned data, we would get a very detailed study that would include a wealth of information that would make the results as specific as possible. In detail, it could be known exactly in a specific tourist area, divided by period according to the month, who stays, how and how much it spends, what are the most sought-after tourist attractions, the most requested. It would also be possible to know how customers spend their money, in which types of shops, in which restaurants, in which bars. Thanks to the use of GPS tracking, always respecting privacy, it is also possible to see how tourists move within the territories, in what way, if on foot, with means of transport, with bicycles. All this makes the data more and more numerous and consistent and thanks to a precise analysis strategy it is possible to have a real database available which contains all the tourist nameplates, the destination locations, how much they spend in the hospitality department, how much they spend in the commerce department, how much in the restaurant and bar department. The main goal for those who work in the sector is to create profit from customers who stay and use the services and it is therefore necessary that expectations are always met and that the products offered are always in step with the requests. The potential of the use of big data within the tourism market, the ideal case of complete data collection would appear as follows: for each month to know in each of the 7 tourist areas the evolution of arrivals by Italian and foreign tourists. By collecting social network data, it could be discovered that a specific type of customer searches for a certain area, and specifically requires knowing where it is geographically located and how to get there from the main neighbouring airports: Turin, Geneva or Milan. The logistics to get to the chosen destination is very important so that guests can buy fly tickets and any transfers to and from the airport. The second interesting fact is the search for the type of accommodation: online searches are normally carried out through the Online Travel Agencies that allow to filter the accommodation facilities according to the needs. Specifically, it is possible to choose the type, the price range and the services offered. By gathering this information regarding the bookings made, the customer is started to be tracked in the area prior to arrival in the resort. Once the data prior to the arrival of the client have been obtained, it will be necessary to collect the information regarding the stay, and

therefore, the data regarding the accommodation chosen based on the type of client, the number of days of stay, the range of expenditure and what the tourist does during the stay. Crossing the data of the arrivals of a specific month for example in the winter season with the number of passes on the ski resorts in the same month, it could be known how many tourists are going skiing in the area. This is partially correct because the numbers of people who work in the ski resort (and who enters in the resort), the people who live in the locality and the people who have a seasonal subscription are not taken into account, that's why it would be effective if the date is gathered directly from the ski area company when selling the tickets (both online and on-site). To by-pass this problem, it could be asked to the ski area offices to highlight the different types of tickets that are sold during the days, and the months. All this information has a specific purpose, that is to predict the future target of customers that will arrive in the area. Through business intelligence and data mining learning software the aim is precisely to group as much data as possible, divided into precise categories so as to be able to see the trend and the various variations over time and thus understand mathematically, how this evolution will continue or not over time.

In carrying out this work, 2 study models were presented: the first concerns the geographical area of the Aosta Valley where the data were highlighted and which were found, the peculiarities, which are their characteristics and missing equals. The second model, on the other hand, highlights what additional information is missing in order to carry out a 360-degree study and how to reach the goals. In mountain tourist contexts the first proposed models, the one that uses only hotelier data, can be applied with some care, adding and deleting further specific information of the destination considered: this allows those who must perform a data analysis to have a solid list of information necessary to obtain a database as complete as possible with which precise studies can be carried out on the target customers who have researched the area, which is present in the area and who stayed there. For each place that is analysed, the context in which it is located must be studied beforehand, so as to adapt the information to be found necessary, such as points of interest for tourists, museums access, or other particular attractions. While staying in a resort, tourists access the Internet to confirm the details of their visit, check their emails, participate in social networks and so on. This activity leaves a sort of fingerprint of the many places they visit and often access the web through the structures in hotels, hostels, restaurants and some open spaces to connect to the Internet via free WIFI networks and therefore their activity on social networks can reflect especially this type of space. To replicate the second model to any destination, be it mountain, sea, city, some data collection operations should be carried out that need to be adapted to the area. The essential information that must be collected during any analysis whose objective is to target the type of clients concerned: type of accommodation where the clients stay, type of guests (singles, families, couples, groups of friends), number of days of stay. Further data that must be collected are internet searches, GPS logs, bank card transactions, records of mobile phone activity, social networks, meteorological data, images recorded with video cameras. As we have had the opportunity to explain the construction of a database with as much information as possible it is fundamental to be able to carry out any

analysis, but we must not forget that the only unstructured data are of no use if not contextualised and correctly stored: the main objective of this phase of data collection is to pass from information to knowledge. The importance of having different sources from which to draw from to recover data is therefore crucial and fundamental to have an overview as comprehensive as possible. Thanks to this, it is possible to have the different target groups of tourists who come to spend their holidays in the area, and thanks to which it is possible to set up marketing and adaptation actions with respect to requests.

If we consider social analysis, we have seen that thanks to business profiles it is possible to extract information in a few simple steps. However, these need to be elaborated and contextualized. Social networks represent for companies and public bodies a communication channel that cannot be ignored in order to implement an effective strategy and day after day they increase their importance. Users who have a profile on at least one of them are always increasing, the number of contents shared with their network of contacts grows exponentially and the length of time spent on social networks becomes more and more consistent. In addition to all this, we must add the influence that the introduction of devices on the market, such as smartphones and tablets, has had on the market, which make the use of social networks even easier and more immediate. Considering all these aspects, we begin to understand the role that social networks can play in a communication strategy aimed at a specific target of customers in which the main function is to communicate. The greatest difficulty for any type of company is knowing what to communicate, but above all understanding what users expect us to communicate. In reality, before deciding what to communicate, we should understand what objectives we want to achieve through our messages. Surely the contents will aim to raise awareness of the brand as a whole, therefore the hotel, its services, the area in which it is located and much more. Their power also lies in the fact of reaching targets of any age: research shows that the age group most present on social networks goes from 20 to 55 years old, or a group of people who have an income. Finally, it can be shown that these data, extracted for free, are useful and demonstrate the potential of using social media in the real world.

5.2 Ethical Considerations

The famous scandal triggered by the documents made public by Edward Snowden who revealed to the world the enormous extent of the collection, analysis and use of data by the NSA, the American National Security Agency, has attracted public attention regarding the delicate trade-off between privacy risks and opportunities arising from Big Data. The NSA case raises concerns about "somnambulism in a surveillance society" (Gael & Muller, 1999): the user data, in fact, are aggregated, analyses, used, and in some cases, sold to third parties: people are only marginally involved in this chain and, at most, receive free services, in exchange for the final granting of their personal data. Some recent analyses show that this model will no longer be sustainable in the medium term, this on the one hand, because governments and authorities are aiming to present increasingly stringent regulations on the

collection of data, on the archive, and on their use. On the other hand, this approach does not offer a holistic and complete perspective as the collected data are fragmented: in fact, every company or entity only has data referring to its users and only on specific aspects and for this reason, the extraordinary benefits of Big Data are mitigated by privacy concerns.

Finding the right balance between privacy risks and the benefit of using Big Data is one of the main challenge's lawmakers are studying. In other words, it will be necessary to achieve a connection between the diverging needs of scientific research, public health, national security, law enforcement, efficient use of resources, on the one hand, and privacy rights, equality and freedom of expression of individuals, on the other. We are faced with a scenario in which we need to understand if the most important asset is for general safety and therefore undergo for example an omniscient surveillance and a decision-making process carried out with automated algorithms that allows us to cure a fatal disease or prevent terrorist attacks. Similarly, to the sensors that populate the physical environment of our planet, social services have emerged that populate the web and that through sensors can potentially be used in sociometry to evaluate different indicators: political opinions, consumer preferences, user mood, public opinion about fundamental issues and so on (Neal, 2008). In practice, these dynamic and real time services such as Facebook, Twitter, have published in the years Exabyte of data regarding the mood of people, through so-called status updates. In addition to this, the location sharing services like Foursquare, Facebook Places, have emerged that allow detailed information on the movements of users with unprecedented capillarity, which leads to a deep knowledge and a very geo-spatial understanding detail of the people who use it. The information on the shared location can be text based (for example, "Andrew has been located at 5000, Forbes Avenue, Pittsburgh, PA"), or it can be "map-based", that is, a service in which the location of the user is represented as a point on a map. The widespread use of geolocation technologies therefore poses new challenges regarding the protection of personal data and privacy as it allows third parties to locate and track people and objects anywhere and at any time, for example in an emergency and rescue. Although geolocation technologies are now part of our daily life, such as the way of collecting tolls, when using magnetic cards for public transport, in case of entry and exit badges for access to buildings, through the use radio frequency identification (RFID), in shops or libraries. These technologies are often incorporated into our mobile devices, invisibly and remotely connected to networks at any time of the day. Furthermore, not only devices such as smartphones, laptops, I pads and tablets reveal where we are, when and what we are doing, they also allow telecommunications companies or Internet service providers to record our activities and trips. By building a complete mapping of the position and duration of our activities, they have the possibility of sending extremely private data to third parties for analysis and possibly for subsequent customer profiling. The electronic information and privacy centre in the United States explains that Facebook Places has made available the data on the position of its users to Facebook's commercial partners and third parties: the company can in fact sell the profiles of its users to widespread commercial companies, which

thanks to this information they exercise highly targeted advertising policies towards users (Baloglu, Pekcan, Chen, & Santos, 2008).

CONCLUSION

Tourists leave a digital "footprint" in most of their activities and these new data sources now make it possible to analyse the behaviour of tourists in the places they visit. They take a large number of photographs and upload them to photo-sharing services, make payments with bank cards, speak and send messages via their mobile phones, are active on social networks and so on. All this activity produces a huge amount of digital data (Big Data) that can be analysed to study behavioural models (Asakura & Iryo, 2017) whose large part are geolocated and therefore the activity of tourists can be analysed spatially.

The contribution that this thesis wants to give is to discover what is the potential of the use of big data in tourist markets; in particular, what relationship is there between the information available from public and private bodies and the evolution of the market, that is, tourists and localities, through an analysis of the data that are necessary taking the Valle D'Aosta Region as a reference base. This information is useful for those who will have to make decisions in the tourism and marketing field in order to have a precise address for tourists in the area. From a theoretical point of view, the advantages of using big data with concrete examples and its potential are explained; it is mentioned how the market has evolved with the advent of social networks and web 2.0 and how people have adapted to this phenomenon. Next, we explain how to collect data, primary and secondary, and how the collection of all this information must then be processed and transformed into knowledge. By knowledge, in this case, we mean the targeting of the tourist so that he can know him, know his lifestyles, his desires, what he hopes for. Finally, the theme of data lifecycle is addressed, from acquisition, extraction, saving, modelling and interpretation. From a practical and managerial point of view, there are within the work, the steps to be carried out in order to carry out a study as complete and targeting as tourists who have stayed, who are in the area and who will most likely be staying in one of the 7 tourist areas of the Aosta Valley Region. All this would allow the tourism market to be studied in a new, different and in-depth manner and would help both public bodies and private businesses to adapt to evolution. It starts from the istat data already present in the archives and their union and modification so as to create a database, and spi explain what information should be considered such as GPS data, points of interest and in fact user traceability, while maintaining anonymity and privacy, a point mentioned from an ethical point of view. In order to highlight the usefulness of the data already in possession of the tourist offices and as a starting point, some practical examples are made which demonstrate the use that can be made of it and the resulting benefits.

The analysis is initially based on the total sample of hotels and other types of lodging, extracted from the regional tourism website: the first objective of the research was to

determine whether the data available to the tourist offices are useful and whether, through their analysis, you can get more information. We can say that, as a starting point for any data study of a location, this information is essential and allows you to make an initial targeting of the type of customer who has stayed. From the same analysis it also emerged that it would be desirable to have more easily available information about the lodgement available so as to target tourists in greater depth and punctually. Subsequently, the missing data concerning the accommodation sector were analysed, highlighting those that could be used and that would give added value to the location, concluding that an in-depth study that goes beyond mere statistical data allows to create value in terms of possible socio-economic choices. Also, in the study phase of tourists, as the first type of lodging there are hotels, divided into stars and extra-hotel facilities such as B&Bs, guest houses, campsites, hostels and apartments. A database was first created with all the information and then tables were extracted which, through a mathematical formula, would cross the arrival data in a given location with the type of lodging. This leads to having, for each country of origin, the respective destination, and allows to define which are the most popular countries of origin. Also, in the study phase of tourists, as the first type of users who stayed between 2014 and 2019, the hotel stars were divided and it was concluded that most tourists, in any area considered, prefer to stay in a 3-star hotel subsequently 4-star. It was then noted that over the years there has been a continuous increase in the number of people who have stayed in 5-star hotels. This leads to the assertion that the customer who stays in the region is on average wealthy and has a fairly high propensity to spend. As a further analysis of the data, two queries were made to the visualisation software: the first concerns the first 3 foreign countries of origin in non-hotel accommodation facilities. In first position, there is France with a big difference between the apartments and the other types of lodging and this is very likely due to the fact that the Valle D'Aosta region is on the border with France and Switzerland, and this means that the travel between neighbouring countries is rapid. Switzerland ranks third for the first few years, while in the past few years it has been overtaken by the United States. Secondly, the four main countries of foreign origin in hotels were analysed in the area of the Matterhorn: these countries are the United Kingdom, Sweden, France and Russia. In almost all cases, with the exception of Russia, there are annual growths. Furthermore, we focused on the UK market and it was highlighted that the preferred hotels are the 3-star ones, followed by the 4-star ones and this is also related to the number of hotels in the area. During the study and analysis phase of the project, various problems emerged related to the limitations that these result from the analysis of the data in possession: the problems encountered are mainly related to the data. The first difficulty was that of having a series of statistical data, in this case the initial tables are relatively small and precise, but which give little information. Linked to this problem there is the lack of data: some important information, not being in the public domain or missing, could not be analysed. In particular, the missing information concerns what tourists during their stay do, to answer three questions: where, how, when. First, where and how the tourist moves during his stay, then to answer how, for example, there is no information about payments on the spot, how much a tourist spends in the territory, in what kind of store, bar, restaurant, hotel.

As for the question of when, there is no information regarding the temporal aspect of the stay, such as when tourists move during the day, in what time they spend their money. Next to this essential information, others can be placed side by side to have an optimal targeting: since the Aosta Valley is a mountain region, where the winter season is the master, one might wonder, how many tourists are there once arrived in the chosen place, they decide not to ski or prefer to carry out other activities not related to downhill skiing. As for skiers, there is no information about the type and level. This information is stored by private companies that are expected to place material orders for subsequent years based on the feedback received. This is precisely a limitation because not having a complete tourist database it is not possible to target it so as to plan general investments and not of the individual company. The combination of the results regarding income statistics compared to those of the single economic realities could have a major socio-economic implication and could allow asset managers to invest more precisely by decreasing the risk of offering services and / or products that they are not considered interesting by potential customers. A social analysis concerning a real Facebook and Instagram business page of a Hotel was carried out, the main purpose of which was to analyse the type of market and to verify if the pages were correctly positioned in the market both from a visual and a visual point of view. marketing. This analysis, albeit of a small economic reality, was aimed at demonstrating the usefulness of the information obtainable through social networks and thanks to which it is possible to tag and understand the market.

Lastly, an attempt was made to identify the data necessary to carry out a 360-degree analysis, which allow to concretely define the potential of the use of the data in the tourism market. Therefore, confirming the fact that the more information you have available, the more you can have a clear idea of what tourists are in the locality the more you can make administrative and economic choices that will prove successful always bearing in mind that the data must be analysed correctly. One of the limits of the project are the accuracy of the data: the correctness of the values in which the starting point of the accuracy lies in their integrity. These sometimes turned out to be errors. Furthermore, the absence of duplication: the fields, records or tables must not be duplicated, this, in addition to requiring double maintenance, negatively affects the quality, since it is possible a lack of synchronisation between the copies of the data. The quality of the data must be checked and verified, and the process of data quality must indicate output which are the data with levels of integrity, completeness, consistency and accuracy considered acceptable and which are to be improved instead. The concept behind data quality is simple, if base the analyses and, consequently, the decision-making processes, on poor quality data, the results obtained will also be of poor quality: we would risk making decisions inefficient, or even completely wrong. The Privacy and data ownership: the dissemination and availability of large amounts of data does not indicate that it is ethical to use them. Facebook, and social media in general, make it possible to extract sensitive data, which could also be used inappropriately. Not only that, so often the so-called electronic traces of their own movements are left, so there is the possibility that the movements are monitored. With further and more complete information (for example by

starting a data analysis project with credit card companies or telephone companies with GPS traceability) it is possible to have a more detailed and sophisticated analysis.

From the overall picture thus far outlined, it clearly emerges that a correct use of Big Data requires knowing how to interpret this enormous amount of data available on the web and brings with it considerable advantages in terms of competitiveness and management effectiveness for companies in the tourism sector; it could be imagined that in the coming years, it will be normal to arrange, in real time, as well as weather forecasts, even forecasts for arrivals of tourists and their needs and preferences. However, having these datasets available is not automatically equivalent to having reliable statistical data in hand and usable in the company field and for this reason it is necessary that their treatment takes place according to rigorous methods and from the established value, both from a technical point of view. Only through rigorous and complex statistical analyses, in fact, it is possible to discover hidden models, highlight unknown correlations and thus obtain other potentially strategic information not immediately obtainable from the whole of the data. Too often, therefore, it is forgotten that even it possible to extract all the useful information referable to a certain "object" from a social network like Facebook, it is not said that these, no matter how large the amount of data, can be considered actually referable to everyone, precisely because the set of Facebook users is far from being considered a truly realistic reference population, and, therefore, is ill-suited to represent it correctly.

The real problem, perhaps, lies in the fact that the current methodological knowledge does not yet seem able to allow an effective and efficient use of data analytics. In short, we are still far from abandoning the traditional methods of collecting and creating statistical data, which, at present, allow us to satisfy those requirements of relevance, reliability and repeatability that are indispensable for obtaining good quality information that can be used in decision-making. This, of course, does not mean that the analysis should not be taken into account; on one hand, precisely because, with the possible methodological advances, the potential for use will increase exponentially; on the other, in that, in any case, it would be a great mistake to ignore such a mass of information, especially in relation to some aspects that are ill-suited to being quantified with traditional surveys (think, of all, of the sentiment analysis). It becomes essential, however, to know not only its potential, but also its limits, since it is only on these conditions that it is possible to make that indispensable qualitative leap from the mere collection of information to the knowledge of the phenomena we are investigating.

In conclusion, it can be said that the above study tried to demonstrate how the use of data, albeit few available, is useful and that the use of data analysis in tourism research has enormous potential for both tourists who want to stay in a certain area, both for workers in the workplace, creating a meeting point between supply and demand.

REFERENCE LIST

1. Antonioli Corigliano, M., & Rodolfo, B. (2009). *e-Tourism italiano 2008.* Milano.

2. Appel, G., Grewal, L., Hadi, R., & Stephen, A. (2019, October 12). The future of social media in marketing. *Journal of the Academy of Marketing Science*, p. 79–95.

3. Arel, I., Rose, D., & Karnowski, T. (2010). Deep machine learning - a new frontier in artificial intelligence research. *Computational intelligence magazine*, 13-18.

4. Asakura, Y., & Iryo, T. (2017). Analysis of tourist behavior based on the tracking data collected using a mobile communication instrument. *Transportation Research Part A: Policy and Practice*, 684-690.

5. Baloglu, S., Pekcan, A., Chen, S.-L., & Santos, J. (2008). The Relationship Between Destination Performance, Overall Satisfaction, and Behavioral Intention for Distinct Segments. *Journal of Quality Assurance in Hospitality & Tourism*, 149-165.

6. Bowen, J., & Morosan, C. (2018). Beware hospitality industry: the robots are coming. *Worldwide Hospitality and Tourism Themes*, 726-733.

7. Brynjolfsson, D. (2012). *Big data: The management revolution.*

8. Chen, J. S. (2003, January). Market Segmentation by Tourists' Sentiments. *Annals of Tourism Research*, p. 178-193 .

9. Chen, S.-C., Sitharama, I., Maria, P. R., Mei-Ling, S., Yudong, T., Haiman, T., Samira, P. (2019). A survey on deep learning: Algorithms, techniques, and applications. *Computing Surveys*, 92.

10. Connie, M., & Thomas, I. (2000). Expenditure-based segmentation: Taiwanese tourists to Guam. *Tourism Management*, 299-305.

11. Dimitrios, B., & Aditya, A. (2015). Smart Tourism Destinations Enhancing Tourism Experience through Personalisation of Services. *Information and Communication Technologies in Tourism 2015*, 377-389.

12. Dolnicar, S. (2006). *Data-driven Market Segmentation in Tourism – Approaches, Changes Over Two Decades and Development Potential.*

13. Estela, M. R., & Salvador Anton, C. (October 2015). Barcelona Case Study. *Journal of Destination Marketing & Management Volume 4, Issue 3*, 162-172.

14. Frankenfield, J. (2020, July 1). *Data Analytics.* Retrieved March 17, 2020 from https://www.investopedia.com/terms/d/data-analytics.asp

15. Frochot, I. (2005). *A Benefit Segmentation of Tourists in Rural Areas: A Scottish Perspective* .

16. Gael, A., & Muller, T. (1999). *Delighting the Senior Tourist.*

17. Gartner. (s.d.). *Big Data*. Retrieved February 05, 2020 from: https://www.gartner.com/en/information-technology/glossary/big-data

18. Giammaria, D. P., Adolfo, B., & Andrea, C. (2017). *In viaggio con un click. Nuovi strumenti di marketing digitale per il settore turistico.* Duende.

19. Isabelle, F., & Alastair, M. (2000). Benefit Segmentation: A Review of Its Applications to Travel and Tourism Research. *Journal of Travel & Tourism Marketing* , 21-45.

20. ISTAT. (2014 - 2019). *Annuario Statistico Regionale*. Retrieved January, 2020 from: http://www.regione.vda.it/statistica/pubblicazioni/annuari/annuario2018/SITE/INDEX.HTM

21. ISTAT. (s.d.). *Annuario Statistico Regionale*. Tratto da Annuario Statistico Regionale: https://www.istat.it/it/archivio/presenze+turistiche

22. Kotler, P. (2000). *Marketing Management: The Millennium Edition.* Prentice Hall.

23. Laney, D. (2001). *3D Data Management: Controlling Data Volume, Velocity, and Variety", Gartner, file No. 949*

24. Law, R., Li, G., Davis, K. C., & Xin, H. (2019). Tourism demand forecasting: A deep learning approach. *Annals of Tourism Research*, 410-423.

25. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 436-444.

26. Liu, H., & Haiyan, S. (2017). Predicting Tourist Demand Using Big Data. In X. Zheng, & F. Daniel, *Analytics in Smart Tourism Design* (p. 13-29). Springer.

27. Mariani, M., Baggio, R., Fuchs, M., & Höepken, W. (2018). Business Intelligence and Big Data in Hospitality and Tourism: A Systematic Literature Review. *International Journal of Contemporary Hospitality Management*, 3514-3554.

28. Matzat, R. (2012, January). "Big Data": Big Gaps of Knowledge in the Field of Internet Science. *International Journal of Internet Science*, p. 1-5.

29. Mehrbakhsh, N., Karamollah, B., Mohsen, R., & Vahid, R. (2017). A recommender system for tourism industry using cluster ensemble and prediction machine learning techniques. *Computers & industrial engineering*, 357-368.

30. Mitchell. (1997). *Machine Learning.* New York: McGraw Hill.

31. Neal, J. (2008). The effect of length of stay on travelers' perceived satisfaction with service quality. *Journal of Quality Assurance in Hospitality and Tourism*, 167-176.

32. Pizam, A. (2004). *The Relationship between Risk-Taking, Sensation-Seeking, and the Tourist Behavior of Young Adults: A Cross-Cultural Study.*

33. Pries, D. (February 2015). *Big Data Analytics: A Practical Guide for Managers.*

34. Sellick. (2004). *Travel Motives Scale.*

35. Sellick, M. C. (2002). Discovery, Connection, Nostalgia: Key Travel Motives Within the Senior Market. *Journal of Travel & Tourism Marketing*, 55-71.

36. Solis, B. (2008). *Conversation Prism*. Retrieved April, 2020 from: https://conversationprism.com/

37. Sung, H., Alastair, M., & O'leary, J. (2000). Segmenting the Adventure Travel Market by Activities: From the North American Industry Providers' Perspective. *Journal of Travel & Tourism Marketing*, 1-20.

38. Sung, H., Morrison, A., & O'Leary, J. (2000). *Strategic Marketing in Tourism Services.*

39. Toedt, M. (2013). *Big Data – Challenges for the Hospitality Industry: 2nd Edition.*

40. Travel Appeal. (s.d.). *Officina Turistica*. Retrieved June 10, 2020 from: http://www.officinaturistica.com/2010/12/back-to-future-a-letter-del.html

41. Tsiotsou, R., & Vasioti, E. (2006). Using Demographics and Leisure Activities to Predict Satisfaction with Tourism Services in Greece. *Journal of Hospitality & Leisure Marketing*, 69-82.

42. Turecková, J. (2002). Positioning the destination product 'Southern Moravia'. *Journal of Vacation Marketing*, 247-262.

43. Ufficio Turismo Regione Valle D'Aosta. (s.d.). *Love VDA*. Retrieved March, 2019 from http:// www: https://www.lovevda.it/it/scoprire/aree-turistiche

44. Witten, I., Frank, E., & Hall, M. (2016). *Data Mining: Practical machine learning tools and techniques.* Cambridge: Morgan Kauffman.

45. Xiaozhou, Z., Zhe, S., Chengqi, X., Yun, L., & Jing, Z. (2019). Mobile AR Tourist Attraction Guide System Design Based on Image Recognition and User Behavior. In

W. Karwowski, & T. Ahram, *Intelligent Human Systems Integration 2019* (p. 858-863). San Diego, CA.

46. Yangyong, Z., & Yun, X. (s.d.). *Defining Data Science.* Tratto da Arxiv.org: https://arxiv.org/pdf/1501.05039.pdf

APPENDICES

Appendix 1: Povzetek (summary in Slovene language)

Danes, v dobi interneta stvari, je uporaba podatkov postala temelj kateregakoli raziskovalnega področja. Pri raziskovanju v turizmu so se na nekaterih področjih vprašanja poglobila prav zaradi zbiranja podatkov iz različnih virov z namenom izboljšanja družbeno-ekonomskih vidikov z izkoriščanjem potenciala podatkovne analitike. S pravilnim pristopom se lahko turistična panoga seznani s preferencami potrošnikov ter uporabi te informacije oz. vpogled za vzpostavitev povezav s posameznimi popotniki in oblikovanje prave storitve ali izdelka ob pravem času.

Da bi prikazali potencial podatkovne analitike je v magistrskem delu obravnavan praktičen in dejanski primer regije Valle D'Aosta v Italiji, kjer je bilo mogoče zbrati vrsto statističnih podatkov o nastanitvenih zmogljivostih na celotnem ozemlju. Ti podatki so bili pozneje preučeni in z ročnim postopkom čiščenja in odstranjevanja je bilo mogoče ustvariti dve zbirki podatkov, ki sta bili izhodišče za analizo oz. oblikovanje informacij s programsko opremo za vizualizacijo in analizo podatkov, česar ne bi bilo mogoče doseči z neposredno uporabo podatkov, kot so bili pridobljeni na začetku procesa. Ta model bi bilo mogoče uporabili v kateremkoli drugem turističnem kontekstu in zaradi svoje vsestranskosti omogoča enostavno prilagajanje razvoju trga.

Vzporedno s tem je bila izvedena tudi analiza podatkov z družbenih omrežij, konkretno s Facebook in Instagram strani podjetja, ki deluje v turistični panogi v isti regiji. To je bilo koristno za prikaz raznovrstnih podatkov, ki jih je mogoče zbrati, in možnosti za ustvarjanje dodane vrednost podjetjem z njihovo analizo, npr. z analizo ciljnih trgov in identifikacijo ustreznih načinov za komunikacijo.

Med študijo je bilo mogoče izpostaviti koristnost zbiranja čim večjega števila podatkov iz različnih virov, kot so statistični podatki, podatki GPS, socialna omrežja, da bi zagotovili čim bolj celovit vpogled v stanje oz. čim boljše razumevanje ciljnih skupin in vrst turizma na določenem področju.

Med analizo se je jasno izkazalo, da je za pravilno uporabo podatkov potrebna pravilna interpretacija oz. razumevanje ogromne količine razpoložljivih podatkov. Potem ima podatkovna analitika potencial za velike koristi v smislu povečanja konkurenčnosti in učinkovitosti podjetij v turističnem sektorju. V prihodnosti lahko pričakujemo, da bo povsem običajno, da možno poleg vremenske napovedi v realnem času pripraviti tudi napovedi prihodov turistov, njihove potrebe in želje. Vendar, kot je razvidno iz opravljenega dela, razpoložljivost te množice podatkov ne pomeni samodejno zagotavljanja zanesljivih in uporabnih informacij, zato je nujno, da se njihova obdelava izvaja z uporabo rigoroznih metod in uporabo ustrezne tehnologije. Tako je mogoče odkriti skrite vzorce, izpostaviti neznane korelacije in pridobiti druge potencialno strateške informacije, ki jih ni mogoče takoj dobiti iz začetnega nabora podatkov.