POLITECNICO DI TORINO

Master Degree
in Computer Engineering

Master's Degree Thesis

Design and development of an informative dashboard to monitor urban tourism

Supervisors:
Prof. Silvia Chiusano
Dott. Elena Daraio

Candidate:
Valeria Sottile

A.A. 2019-2020

Acknowledgments

I would like to thank all the people who supported me not only in this last months, but also during all the university years.

I would first thank my thesis advisor Prof. Silvia Chiusano, which gave me the opportunity to work on an interesting project and allowed me to express my strong interest in data analysis. She mentored me with her helpful comments and advices, pushing me to do my best. I would like also to express my thanks to the PhD student Elena Daraio that supported me during my final work.

I thank all my friends, in particular Lucia and Federica, with which I have shared a lot of beautiful moments.

Finally, I must express my very profound gratitude to all my family, in particular to my father and my brother Carmelo who have always believed in me and encouraged me to never give up and to my boyfriend, Francesco, for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of developing and writing this thesis. This accomplishment would not have been possible without them. Thank you.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Urban tourism is a type of tourism activity that represents a central component in the economy and social life of many cities in the world. Tourists visit cities for many purposes: museums, art galleries and historical buildings are the main places of interest; massive events such as festival, concerts, parades, conferences, demonstrations and protests attract huge amount of people. There is also a strong correlation between city tourism and shopping tourism, in particular for people from smaller cities. The tourism industry can make an important contribution to the development of the city and allow better living conditions for its residents and visitors (from [38]). Therefore, tourism managers should continuosly monitor the impact of tourism in the social and economic development of cities in order to better address the main issues of tourist activities. To evaluate the effects of tourism on urban socio-economic aspects, it could be useful to explore the huge amounts of data available nowadays that tourism stakeholders can exploit to make decisions. Different type of data relating to tourist flow and tourist activities in a city can be collected and analyzed to extract knowledge.
To give value to the information obtained from the analysis of data and allow an easy interpretation, it could be useful to produce reports, in particular graphic representations of the indicators of tourist trends.

The research area of this thesis work is *Data Analysis* that is the process of evaluating data by means of analytical or statistical tools to discover hidden information. In particular we refer to *Business Analytics* as a combination of skills, tools and data analysis techniques that allow organizations of every industry to measure and improve the effectiveness of the business strategy.
In the context of urban tourism, the analysis of data concerning tourist flows and turnout of visitors to exhibitions, events and fairs could support decision-making processes on the economic and social development of the city.

The aim of this thesis project is the analysis of some tourist data to measure *Key Performance Indicators* (KPIs) that could provide useful information on urban tourism in Turin. The results obtained from the analysis of the KPIs are represented through graphs and visualized within an informative dashboard which stakeholders and tourist managers could use to monitor the tourist trends of the City of Turin.
The design and development of the dashboard is composed by three main phases: data exploration, data interpretation and data visualization. The first phase was

focused on the research and analysis of heterogeneous open data on tourism in the City of Turin retrieved from several data sources. Each dataset has been explored in order to exploit its characteristics and produce meaningful KPIs that allow to measure the tourism development of Turin.
So, in the data interpretation phase several KPIs have been measured by observing different time granularites. All the KPIs produced from the data interpretation have been represented through several charts programatically built with Python programming language. In this way it is possible to observe the trends of tourism in Turin in a given period of time and consequently make decisions considering the different aspects of urban tourism provided by data.
The final phase was devoted to the design and development of an informative dashboard with the aim of organizing all the produced charts in a logic and meaningful way. So the data visualization phase includes the creation of the *TOurism Analytics Dashboard* (TOAD) application to support users that want to monitor urban tourism trends of Turin City.
TOAD is an interactive informative dashboard with the purpose of providing a set of indicators to measure tourist flows in the City of Turin. The objective of this informative dashboard is to provide a tool that is easy to understand, with a nice visual appeal and that is reactive to user external interactions. It was designed and developed to support stakeholders and tourist managers in the analysis of tourism industry of the Turin City. TOAD represents a powerful tool to improve the effectiveness of the decision-making process on urban tourism in Turin thanks to the graphical representation of significant KPIs extracted from the analysis of tourist data.

This thesis is organized as follows. Chapter 2 is devoted to the description of the reasearch area of this thesis work and the analysis of previous studies on the development of a dashboard. Chapter 3 describes the main characteristics of the TOAD application and the technologies used in the development phase. Chapter 4 provides an overview of the data sources and datasets explored to implement the dashboard. Chapter 5 is focused on the data manipulation and interpretation that includes the creation of the charts to represent the KPIs. In Chapter 6 the logical and graphical organization of the dashboard is provided. Chapter 7 is centered in the evaluation of the results. So all the KPIs measured are discussed in order to extract interesting information. Chapter 8 concludes this elaborate, highlighting the results obtained and the dashboard's functionalities.

Chapter 2

Introduction to Data Analysis and Business Analytics

This chapter is intended to give an overview on the research area of this thesis work project. Firstly, an introduction to *Data Analysis* is provided, in order to give a generalization of what is the topic of this work. In particular, the definition of *Business Analytics* (BA) is given, with an explaination of which are the different business analytics types and which are the main steps applied to perform business data analysis process. Furthermore, a brief description about *Key Performance Indicators* (KPIs) is provided to better understand the importance of defining measures in the context of a business strategy. Since the design and development of an interactive *informative dashboard* is the central focus of this thesis project, specific attention on this technology as a Data Visualization and Data Analysis tool is necessary, reviewing some previous case studies and introducing the functionality of a new innovative dashboard.

2.1 Data Analysis

Data analysis is a universal language with many applications in a world that is becoming more and more data-driven. The endless amounts of data available nowadays represent a powerful knowledge that companies of all the industries can exploit to make decisions. Data analysis is the process of evaluating data by means of analytical or statistical tools to discover hidden information. Data are collected and sorted using these tools and then results are interpreted to make decisions. The final step is the visualization of these results by writing some reports or graphically drawing resasonable charts that make the interpretation easier. Most of the organizations get meaningful advantages from the Data analysis. They collect every day massive amounts of data, exploiting different kind of sources, with the purpose of getting useful insights and building a solid business strategy. So the main targets inside businesses is collecting raw data and forging them in order to make decisions based on specific predefined goals. In particular, we talk about the concept of *Business Analytics* as a combination of skills, tools, and applications that allow companies to measure and improve the effectiveness of core business functions such as marketing, customer service, sales, or IT.

2.2 Business Analytics

In the socio-economic scenario, the easy access to information and knowledge has influenced the continuous changes occcurred in methods of production and in economic relations. Over the years, large amounts of data have been collected by the various organizations thanks to the availability of Internet connections and the adoption of massive data storage technologies. These huge transfers of information offer several advantages to various actors:

- *Individuals*, who can get real-time news in a rapid and easy way, access services and perform on-line commercial and banking transactions;
- *Enterprises*, which can produce services according to the needs of the users and get competitive advantages from the trasformation of raw information into useful knowledge;
- *Public Administration*, which can offer a better service to the citizens, for instance by providing e-government applications, such as on-line payments of tax contributions.

In order to extract knowledge and making decisions about collected data, it is not enough storing data in a systematic and structured way. They should be processed by means of appropriate extraction tools and analytical methodologies that *decision makers* inside an organization need to know. In this context, it is appropriate defining the concept of *Business Analytics*: it is the process of collecting, processing and analyzing business data by means of statistical models and iterative techniques in order to turn raw data into business insights. Business analytics is very used inside companies in order to make data-driven decisions. It is possible to distinguish four types of business analytics, which differ for the complexity by which data are manipulated and the level of the extracted knowledge. All these analytics types are interrelated and implemented in stages (from [37]).

Descriptive Analytics
Descriptive analytics is the interpretation of historical data to better understand changes that have occurred in a business. It tracks specifically designed indexes (Key Performance Indicators) in order to understand what is the current state of the business.
Descriptive analytics gets insight from historic data to draw comparisons. It helps identify points of strength and of weakness and in this way it is possible to inform management strategies. Most of the reported financial metrics exploit descriptive analytics, for example the yearly pricing changes or the montly sales growth, the number of users or the total revenue per subscribers.

Diagnostic Analytics
Diagnostic Analytics takes care about past and current events in terms of *how* and *why*, looking for the factors that influence trends. Usually this kind of business analytics exploits techniques like drill-down, data discovery, data mining and correlations to find the causes of events. So it is mainly concentrated in discovering why certain events may occure and employs techniques such as attribute importance, sensitivity analysis and training algorithms for classification and regression.

Predictive Analytics
Predictive Analytics analyzes trend data to forecast the possibility of future events using statistical models and machine learning techniques. This type of business analytics takes its insight from descriptive analytic results in order to design models that can extract the likelihood of selected outcomes.
An example of applications of predictive analytics is *Sentiment analysis*. This is done by collecting text data coming from social media in order to draw a picture of opinions held by a user. Data could be analyzed to predict the sentiment towards a new subject (positive, negative, neutral).

Prescriptive Analytics
Prescriptive analytics exploits past events to generate recommendations about how to handle similar cases in the future. It relies on a strong feedback system and constant iterative analysis in order to get increasingly stronger correlations between actions and outcomes. The most common application of prescriptive analytics is the implementation of recommendation systems that are in charge of matching options to real-time needs of a customer.

By exploiting these four types of analytics, big data can be dissected, absorbed, and used by organizations in order to face the biggest challenges and achieve competitive advantages.

2.2.1 Business Data Analysis Process

In order to improve the effectiveness of the Business Strategy, a business data analysis process should be composed by some significant steps that lead decision makers to the target achievements. (from [52])

Step 1: Define the Questions
When a business data analysis is required, the right questions must be clear, in order to focus the problem and to achieve the business goals. For example, in public administration, if the aim is to increase the average number of tourist arrivals in a city, a good question to ask could be which are the events, exhibitions, attractions on which it is more convenient to invest?

Step 2: Set Clear Measurement Priorities
It is important to decide what should be measured and how it should be measured. It means that defining KPIs is strongly suggested in order to have a clear overview about data that are needed. For instance, considering the tourism use case, it can be useful to count the number of museum entrance tickets issued. This measure could be more useful if time information are included. Temporal granularities of data aggregations should be chosen in order to better center the final goals.

Step 3: Collect Data
Once questions are fixed and measurement priorities set, data need to be collected. It is important to organize and structure data coming from different data sources with different kinds of information.

Step 4: Data Analysis and Interpretation of results
Data have to be analyzed and processed in order to identify and select the relevant information according to predefined KPIs. To allow a clear interpretation of the results, it might be useful to write tables or to generate charts and graphs. In this way, each KPI is graphically implemented and it is possible to draw conclusions on it. But to give a overall insight about business strategy, all the KPIs need to be displayed together in a unique graphical reporting to facilitate the accessability of all the information needed. This can be done by using for example an informative Dashboard used to make complex data more understandable.

Since the role of KPIs as key measures of business trends is essential in the business analysis process, a more in depth explaination is provided. Therefore, a procedure that could be used to obtain significant KPIs is described.

2.3 KPIs - Key Performance Indicators

In business, the concept of *performance management* is widely used to indicate the way in which an organization keeps on track, makes decisions and drives its strategy. It's very essential for companies to get competitive advantage over other market players in the same industry, so they should well define their objectives and goals and monitor their performance. In this context, organizations need to know about *Key Performance Indicators*.
Key Performance Indicators (or KPIs) are defined as a set of measures, quantifiable values by which it is possible monitoring and managing business strategies and achieving pre-defined goals. In other words, a KPI is a quantitative value that is possible to compare and it helps to evaluate the goodness by which a goal is being reached over time. For instance, in Marketing and Sales industry it might be interesting evaluate the following KPIs:

- **Average Revenue Per Unit (ARPU):** the average customer's revenue from all the sales (from [51]).
- **Customer Acquisition Cost (CAC):** how much does a new customer cost (from [51]).
- **Customer Lifetime Value (CLV):** how much do you expect to earn per customer (from [51]).

Thus, it is important for an organization to make the right considerations about the targets to be achieved and select the key metrics that give meaningful value to business data. For this reason, the design of KPIs is a process that requires a methodological approach and therefore a more in-depth explanation is provided.

2.3.1 Methodology

To build an appropriate business plan and design the KPIs to be measured, companies and organizations need to well define their objectives. Therefore, a methodological procedure composed by seven steps is provided (from [42]).

Step 1 - Define the Goals
As mentioned above, goal setting is the starting point for planning a solid business strategy. KPIs are based on predefined goals. For a company, an example of a goal might be *"increase company profits"*. Since this is a higher level goal, it may be useful to achieve lower level goals before reaching the final result. For example, to increase profits, a good idea may be to improve the level of competence of senior consultants or to implement a sales plan.
Each single objective is designed to contribute to a global strategy: it must be important to make a difference and it must be something that can be measured.

Step 2 - Describe Results
KPIs define the goals to be achieved and so they are concerned with results. In the design of KPIs, it is essential to adopt a results-oriented approach, which leads to a more precise pianification of the goals. Thinking about the objective mentioned above, 'Implement a sales plan', it seems to be a quantifiable value that can be measured but it is still not. It is an activity that can only be measured through the time it takes to execute the plan. The important question is 'why' a sales plan is needed. The answer may be *'to reduce the time taken to convert a qualified lead into a sale'*. Therefore, the goal to be achieved is the time parameter that describes a result, and it is clear to interpret.

Step 3 - Identify Measures
The measurement identification is an essential process for the development of meaningful KPIs. This can be solved by performing these three actions:

1. A clear description of the measure is needed and it should be based on an objective.

2. Measures should be classified according to their importance.

3. The measurement must be calculated.

1. A clear description of the measure is needed and it should be based on an objective - At a preliminary stage, a long and detailed description is needed to achieve the goal of the KPI to be measured. Then short sentences and key words are used to define the measure. Understanding which are the concrete things to measure makes easier the KPI computation process. For this purpose, it can be useful introducing the difference between *lead* and *lag* KPIs. Most of the KPIs are lag measures. It means that KPIs are measured after the event has occurred. Financial measures such as profit, revenue, costs are in this category. Lag measures are most used simply because it is easy to count them and they can prove whether it is success or failure. Lead measures are more difficult to identify but they make the difference in a business strategy because they are the only measures that can be influenced. An example of lead measures are, in case "losing weight" is the intended goal: how many times I go to the gym; how many calories I eat per day. At first, these measures are harder to be focalized but in the long run they can provide significant results.

2. Measures should be classified according to their importance - Making a classification of the important KPIs is needed, in order to ensure the relevant information are discovered. This can be done by observing these three features:

- How effectively this measure may be applicable in that business strategy
- How is important computing this measure in that specific business plan
- How much is easy finding the data to make measurement

These questions should be asked for each KPI, classifying them as High, Medium or Low importance.

3. The measurement must be calculated - In order to compute a KPI, some parameters are considered:

- **Description:** A sentence that describes in details the purpose of the KPI.
- **Label:** A short description, may be using keywords.
- **Owner:** Who is the author of the KPI.
- **Updater:** Who is responsible of updating the KPI when is needed.
- **Calculation:** A mathematical formula that is able to describe how data are combined in order to provide a value such as a percentage.
- **Frequency:** How many times the KPI is calculated.
- **Scope:** What should be included or not, such as a data range.
- **Metrics:** Information about data and data sources employed in the KPI calculation, to provide a complete description of the metrics and to avoid ambiguity.

Step 4 - Define Thresholds
A KPI can help stabilize performance of a process. Once there are enough data to analyze, you can observe some representative values that allow to draw conclusions about the process performance. For example, with a new process, the values initially vary from month to month. The value that the variance should have is unclear but when enough data are available, it is possible to extract an upper and lower limit. These limits become *thresholds*. Usually thresholds are based on targets. Targets are set by using some kind of methods or by inspecting available data. However, the targets are very essential in the definition of thresholds. A target should be reasonable and achievable. In the financial context, the *variance* (threshold) is observed with respect to an expected result (target). Thresholds are the measures by which it is possible to justify the results obtained, deciding whether or not they are acceptable. There are several techniques that allow to manage thresholds. The most common threshold model is the *Red, Amber, Green* (RAG). The RAG model is composed by two threshold points: the KPI should become Green; the KPI should become Red. In the graph in which the KPI is displayed, it is possible to identify three coloured areas:

– Green means the result is acceptable, so the target is achieved

– Amber means there is a problem and it should be examined

– Red means unacceptable result and a rectification is needed

Figure 2.1 shows an application of the RAG method. Considering as example of KPI the day to fulfill an order, fixing the green threshold to 60 days or less and red threshold to 80 days or more, it is possible to instantly observe the current situation and also the historical performance.

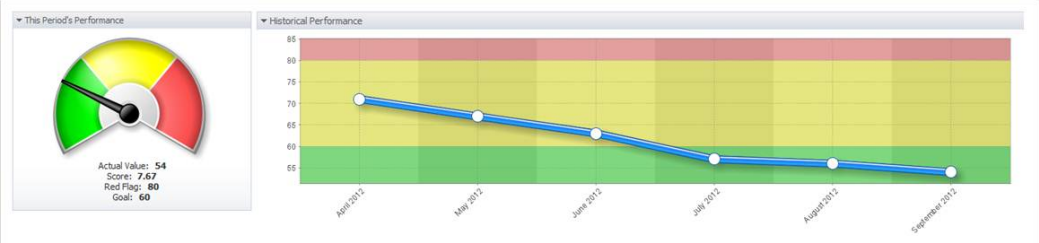


Figure 2.1: Example of RAG (from [3])

Another threshold model is an extension of the RAG method. It is the *Red, Amber, Green, Blue* (RAGB) variant, in which five thresholds are used: the lowest acceptable result; the KPI should become Green; the KPI should become Red; the KPI should become Blue; the highest acceptable result. In this way it is possible to monitor and manage a state of overcoming, most of the system will also put a limit on the most acceptable result. Once again in sales this may be desirable to avoid escape bonus schemes (Figure 2.2) .

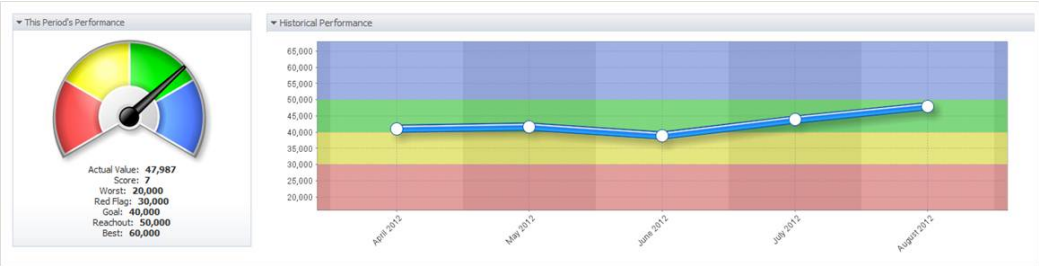


Figure 2.2: Example of RAGB (from [4])

The last threshold example is the *stabilise KPI*. Sometimes KPIs values are not acceptable since they are too high or too low. For instance, in the case of managing the training budget, thresholds are needed in order to not exceed the budget or go below the budget. For this purpose, the best result is defined in order to determine if the achieved results are acceptable or non-acceptable above or under the threshold. Figure 2.3 shows a possible result in the case of training budget example.

Step 5 - Implement KPIs structure and Data preparation
During this phase the KPIs structure must be properly built so that it is possible to identify which are the important data that help to achieve the goals. It can be useful building a structure in which *perspectives*,*objectives* and *KPIs* are well - designed and distinct.

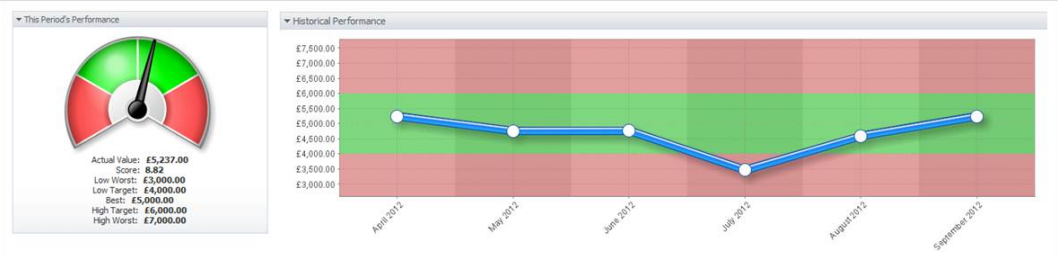


Figure 2.3: Example of stabilise KPI (from [5])

Perspectives represent different ways of looking at organizational performance. Several perspectives needs to be considered to design a business strategy since an organization is a system composed by different elements working together. In a business context, the perspectives could be Financial, Customer, Internal Processes, Learning and Growth. Each perspective presents several objectives to be achieved.
An objective is a top-level result that you want to get to contribute to the overall strategy. For example in the case of the Customer perspective, the objective may be " improve the customer satisfaction". An objective is composed by KPIs to be measured.
Each KPI is a quantifiable value that is important to address the final goal. Examples of KPIs are the following: the percentage satisfaction surveys completed; the average satisfaction survey score. To create the KPIs structure, spreadsheets are commonly used but sometimes they might be difficult to centralise and maintain. There are some software by which it is possible to create scorecards and quickly setup the KPIs structure to be clearly visualized. Once KPIs are clear and well - structured, the data to be employed for getting the intended results are easier to select and extract. So all the data sources are inspected (spreadsheets, databases) to get the values needed and compute the KPIs.
In order to make a meaningful KPIs analysis, it is better to have at least 5-9 months history. In addition, providing automation for data integration could accelerate the accessibility of available information.

Step 6 - Interpret Results
Data must be represented in such a way that it is easy to interpret it and it is possible to make decisions on corporate strategy. For this reason, there are several tools to be exploited such as written reports in which data are displayed through tables or dashboards that visually describe the measured KPIs.
When designing dashboards or reports, it is important to think about who the audience is. People with technical skills are able to monitor more specific details about a process than people with managerial roles. For example, in the case of a CEO, the dashboard should be as simple as possible, highlighting only the most important indicators that best respond to business trends. However, the view of the business is certainly different as compared to an operations manager or a financial director.
An important feature is the readability of the dashboard. The user interface must be intuitive and must have an easy navigation.
As mentioned above, defining thresholds could be a way of keep under control the values of the KPIs. This kind of method is useful to perform a point analysis of a

KPI, but it does not provide an overview of the entire business process. If a KPI moves from an acceptable result towards an unacceptable result, it does not mean that there is a problem.
A better way to interpret results is looking for patterns in the data. This because in a business strategy there are multiple factors that contribute to the overall plan. Thus, patterns are helpful and they can be extracted from the data. If a single KPI provides poor performance, it is necessary looking for related KPIs. Moreover, it is important to check if a KPI is stable and predictable. For example, with regards to sales, if you want to measure the "number of sales per month", it could be useful to identify what the unstable periods are, observing the long-term trend and take an action.

Step 7 - Take Action
After evaluating and interpreting the results achieved, actions need to be taken to address the problems that have occurred. Each action should be linked to a KPI. So only by observing a positive variation of the KPI value it is possible to know if the action has provided a positive effect. The action must be clearly described in order to understand when it is required and why it is necessary performing it. In addition, an action should be performed by looking at multiple factors that caused the problem and not considering only a single short-term anomaly.
Depending on the complexity of the changes to be applied, the actions could be short-term or long-term activities. In the latter case, a strategic initiative is performed that may have an impact on the overall corporate strategy.

In conclusion, this methodological approach to build meaningful KPIs is useful for effectively understanding the importance of performance indicators that help monitor business trends and improve decision making.

2.4 Informative Dashboard

The definition of questions that should be answered in order to achieve specific business goals is the very central part of a business data analysis process. Once data are collected, analyzed and interpreted, there are more chances of getting reasonable results and making right decisions. Furthermore, since it is not enough showing some charts according to the processed data, designing something that works as data visualization tool might be useful to speed up the decision-making process. Each graph, histogram, table needs a contextualization to provide an easy interpretation about the achieved results. In this context, we talk about informative dashboard and its applications. An *informative dashboard* is a collection of visual reports that display KPIs. It is a tool that allows to quickly visualize data and get useful knowledge about some predefined business goals. A dashboard is usually displayed on a web page that is linked to a database which allows the report to be constantly updated. Most of the organizations need to discover valuable insights in order to make decisions. Thus, an informative dashboard helps to give a snapshot of the most important metrics in a easily usable visual format. Well-Designed informative dashboards give serveral advantages for organizations because they can make better use of their data.
Some of the benefits of using data dashboard are the following:

- Dashboards help answer business questions: when complex metrics coming from several sources are required, dashboard is helpful in providing a collection of reports that better summarize all the results.
- Dashboards make data easy to digest: they typically display information using data visualizations like bar charts, pie charts, scatter plots. These visual representations allow the human brain to better process information.
- Dashboards are interactive: with static charts or graphs is not possible to interact with data. Instead dashboard can help the user to drill down on a specific element or KPI in order to find the level of detail needed.
- Dashboards are flexible: they are a collection of visual reports. Flexibility in the way data are accessed is needed, because which reports are displayed is up to the final user.

Nowadays dashboards are very widespread, mainly because there are many predefined Dashboard Software that most of the companies exploit in order to monitor KPIs of some business process. Most common business processes are: Sales, Marketing, Customer Support, Human Resources, Production.
Figure 2.4 shows an example of dashboard in which some KPIs about sales are displayed. It is possible to identify four main aggregations (General, By Region, By Sales Rep, By Segment) by which data are grouped and visualized and this can help the final user to easily decide the level of detail to obtain from results. Furthermore, each single chart is in charge of interpreting a particular KPI and it is possible to select a date range in order to get a snapshot of what happened in that temporal range.

Thus, dashboard represents a powerful way to display data, because several KPIs and visual reports provide to the final user a more meaningful overview than showing only single static charts.

2.4.1 Previous Studies

In order to explore the usefulness of an informative dashboard as support tool in business analysis, five previous developed applications are reviewed according to some criteria: intendend goals and target users, data extraction, data processing and data visualization techniques. These previous dashboard applications are focused on different business areas, but the purpose is to look at their functionalities and to think about new development approaches of a dashboard.

Intended Goals and Target Users
Any application development process requires the definition of objectives to be achieved and the identification of end users. In this context, Table 2.1 provides an overview of the dashboards under analysis, highlightning the objectives for which these support tools have been implemented. It is possible to observe that both target users and goals are clear and well-defined for each of these applications. Each dashboard wants to provide a specific result. For instance, the Diabetes Dashboard (from [68]) aims to improve the efficiency and the accuracy in accessing data of diabetes patients and so reducing the time and effort that Physicians need to get

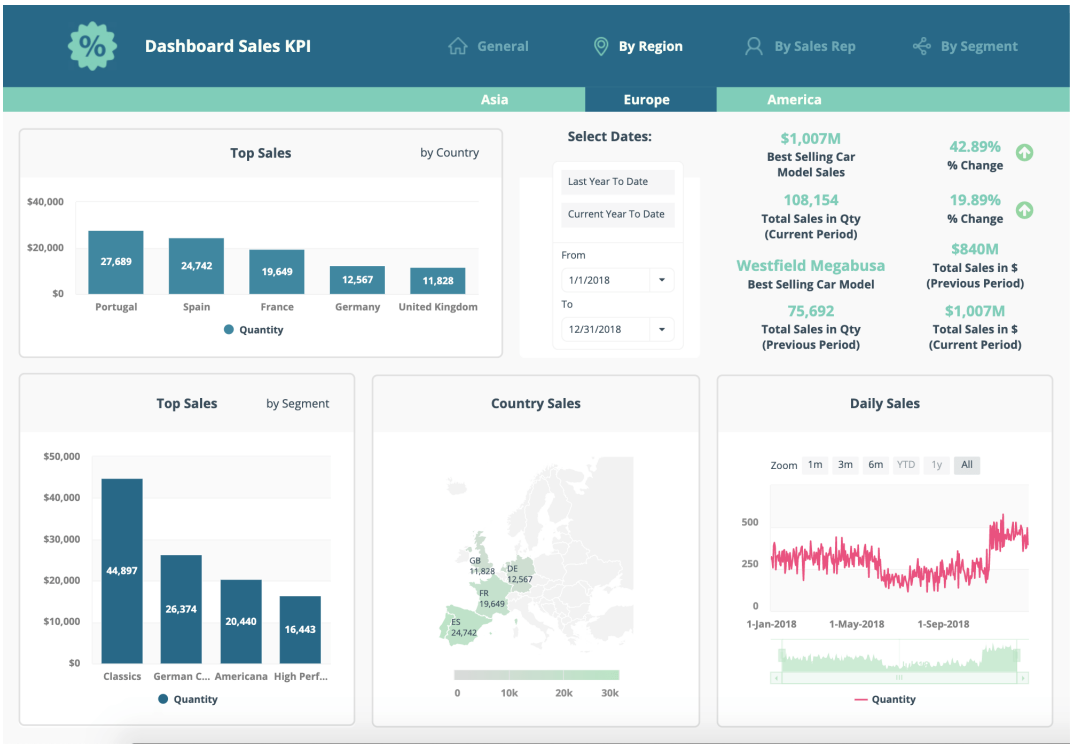


Figure 2.4: Example of a Dashboard (from [6])

information. The 'RE:fine drugs' Dashboard (from [61]) has the purpose of speeding up the process of discovering some existing drugs that can effectively treat new conditions. These two examples make clear the importance of having easy and fast accessibility of data and that it is achievable by means a data visualization tool like a dashboard.
Another important goal that emerges from the analysis is providing insights about trends of specific activities in a way that it can stimulate improvements in the decision-making process. Examples of this are the Learning Analytics Dashboard (LAD) (from [69]) , the Global Sustainable Tourism Dashboard (GSTD) (from [65]) and the Hawaii Tourism Dashboard (HTD)(from [64]) . The LAD is designed to help students monitor their learning strategies and motivate themeselves to achieve better results. The GSTD aims to give a general overview of the Global Tourism trends by means of top-level indicators (e.g. poverty alleviation, carbon emissions, protected areas and tourism planning) which tourism managers can explore to get more understanding of the contribution of tourism industry. The HTD wants to keep track of tourism-related factors (e.g. water quality, visitor numbers and resident sentiments) that can affect the vitality of a tourist destination, in order to extract useful information and improve the management of tourism activities. As for the latter two case studies, the tourism topic is treated with two different perspectives: GSTD wants to monitor economic, environmental and social indicators to address the impact of tourism industry at global level; HTD is focused on the main issues that tourist destinations like Hawaii's islands need to manage to guarantee a sustainable and nice place to visit and live.

Data Extraction and Processing

Since the dashboards under analysis are all examples of visual tools that aim to

Name of the Dashboard	**Target Users**	**Intended goals**
Diabetes Dashboard	Physicians	improving the efficiency and the accuracy in accessing patient-level data for diabetes care
'RE:fine drugs' Dashboard	Researchers	supporting researchers to find some existing drugs that might be used in the care of new deseases
Learning Analytics Dashboard (LAD)	Students	supporting students in their learning strategies and motivating them to improve their performance outcomes
Global Sustainable Tourism Dashboard (GSTD)	Tourism Managers	providing a set of top-level indicators for a more complete understanding of the contribution of the tourism industry
Hawaii Tourism Dashboard (HTD)	Stakeholders of Hawaii Tourism	diagnosing the health Hawaii's tourism an making this information shared among stakeholders of Hawaii tourism

Table 2.1: Target users and intended goals of previous dashboard applications

provide a descriptive representation of information, no complex data processing is performed. Raw data are extracted from appropriate data sources and manipulated according to predefined objectives. Therefore, relevant measures are generated by exploiting several data features. Table 2.2 shows a brief overview about the way data have been processed for the previous dashboard applications. There are cases like the Diabetes Dashboard and the LAD in which data come from only a single resource and this might speed up the data visualization process. Furthermore data are periodically updated since they are retrieved by log-data in the case of LAD and electronic records in the case of Diabetes Dashboard. In the remaining three

dashboards (RE:fne drugs, GSTD, HTD), data are more heterogeneous and so more time and effort are needed to perform data integration and generate a unique dataset.

Name of the Dashboard	**Data Extraction and Processing**
Diabetes Dashboard	manipulation of raw data of diabetes patients extracted from Electronic Health Records (EHRs)
Learning Analytics Dashboard (LAD)	manipulation of log-data retrieved by Learning Management Systems to measure meaningful indicators of student's learning activities
'RE:fine drugs' Dashboard	manipulation of Open datasets (GWAS, PHEWAS, DrugBank) to generate a new dataset of drug repurposing
Global Sustainable Tourism Dashboard	Exploration of different data sources (e.g. World Bank, ICAO, Amadeus booking database) to measure top-level indicators of global tourism industry
Hawaii Tourism Dashboard (HTD)	manipulation of raw data available by the public sector (e.g. tourism government offices, universities, associations) to monitor significant variables of the health of tourism in Hawaii

Table 2.2: Data Extraction and Processing of previous dashboard applications

Data Visualization

The way in which data are displayed is the central part of the analysis of these previous dashboard applications. As we can see in the Table 2.3, most of these visual tools have been developed as websites in which user can interact with the graphs to drill down relevant indicators and make decisions. The key metrics that need to be analyzed and displayed inside the dashboards are grouped by category and accessible via links to multiple pages. In this way users can clearly identify the various perspectives of the analysis and perform more accurate evaluations of the KPIs. In general, website is the most used web-based architecture to create an interactive dashboard.

Name of the Dashboard	Data Visualization
Diabetes Dashboard	single visual plane with use of sparklines or word-sized graphics
Learning Analytics Dashboard (LAD)	single panel in which KPIs are represented by means of scatterplots
'RE:fine drugs' Dashboard	freely interactive Website with a search bar to browse and display the dashboard according to selected parameters
Global Sustainable Tourism Dashboard	a website with a thematic aggregation of indicators that are graphically displayed on multiple pages
Hawaii Tourism Dashboard (HTD)	Website with a graphical representation of variables to monitor, divided by categories

Table 2.3: Data Visualization of previous dashboard applications

2.5 Final considerations

In conclusion of this chapter, some considerations and observations are provided in order to highlight the advantages that a well-designed informative dashboard could offer to target users. Looking at the dashboard application models previously analysed, some key points are emerged:

1. Intended Goals
Defining the objectives to be achieved taking into account which are the target users is the starting point of any kind of application. Thus, in this phase thinking to the questions that need to be asked is very important, according to the needs of the final users. For this purpose, performing preliminary interviews to the users might be a good practice to discover which are the main requests to address. Furthermore, identifying the target users might help to outline the further process steps, making easier the definition of the requirements.

2. Requirements
The design of the requirements leads to the development phase. Hence, the definition of KPIs is needed, in order to find useful information to be extracted and enable the user to decision-making process.

3. Development Process
During this phase, all the data are collected and manipulated in order to extract the information needed. To guarantee easy readability of the contents, appropriate graphical models are chosen according to the key metrics to be displayed. Once

all the graphs are produced, a graphical interface that allow to group them in a meaningful way is needed.

4. Data Visualization
The purpose of building an informative dashboard is to provide users with easy access to all the necessary information, reducing the time taken to recover data and make it more understandable. Therefore, an interactive dashboard is needed, which offers an easy-to-use graphical interface and which provides all the required information. All the displayed graphical reportings need to be placed inside the dashboard in a meaningful way, to guide the target user to an easy and correct interpretation of the information.

All these considerations should be taken into account in the context of any application development, setting the desired objectives and defining the target users in order to build a successful interactive informative dashboard.

Chapter 3

TOurism Analytics Dashboard

This chapter provides a brief introduction of the informative dashboard designed and developed in this thesis project named *TOurism Analytics Dashnoard* (TOAD). Then the main characteristics of the Python programming language and the most relevant libraries are described as technologies used to implement the dashboard. Finally a description of the TOAD architecture is given.

3.1 Introduction

The tourism industry can make an important contribution to the development of the city and allow better living conditions for its residents and visitors. Therefore, tourism managers should continuosly monitor the impact of the tourism in the social and economic development of cities in order to better address the main issues of tourist activities. This can be done starting from the collection and extraction of useful data that might provide significant indicators of the tourism effects on the economic growth of the city. Since data are heterogeneous and in many cases are very crude, it is necessary to find a way to easily explore them and create a visual representation of meaningful information.
In this context, the *TOurism Analytics Dashboard* (TOAD) application is born to support users that want to monitor urban tourism trends of Turin city. TOAD is an interactive informative dashboard with the purpose of providing a set of indicators to measure tourist flows in the City of Turin. Heterogeneous data coming from various open data sources are explored in order to find the KPIs that might indicate the effects of tourism in terms of social life and economic growth of Turin. TOAD supports the user in discovering urban tourism trends according to specific aggregations of data.

As previously said, the definition of the requirements is the most important step in the design of an application. To this end, the creation of TOAD is based on the following criteria:

Intended Goals and Target Users
In the context of tourism industry, an informative dashboard that shows results of tourist data analysis could be a support tool for stakeholders and tourism managers that need to monitor the impact of tourist activities in the economic growth of the City. Therefore the objective of TOAD is to visualize all the information extracted

from the data analysis according to data correlations and support the decision-making process of tourism stakeholders.

Data Extraction and Processing
The tourist data in analysis are retrieved from several open data sources. All these data sources are web data platforms in which it is possible to freeely access and export tabular data. All the data are processed and interpreted in order to extract the key metrics on the urban tourism. Then each KPI is represented through a suitably selected chart.

Data Visualization
To display the KPIs produced an interactive web application is developed. This application corresponds to an informative dashboard organized according to the main topics identified in the data analysis process.

Table 3.1 gives an overview about the target users and intended goals of the application, looking also at the main phases of the development process.

In conclusion, to create a nice and innovative graphical interface and guarantee and easy-to-use dashboard, a web application model is used. A web application is typically designed to be an information tool to support users in the organization of hetereogenous data that can be analyzed to extract knowledge. The objective of this informative dashboard is to provide a tool that is easy to understand, with a nice visual appeal and that is reactive to user external interactions. TOAD is a new way to develop an informative dashboard that is innovative and user-friendly.
The next section describes the technologies and the framework used to develop the web application.

Name of the Dashboard	**Target Users**	**Intended goals**	**Data Extraction and Processing**	**Data Visualization**
TOurism Analytics Dashboard (TOAD)	Stakeholders and tourism managers of Turin tourism	allow stakeholders to monitor tourists flow,turnout in museums and tourism urban movements	manipulation of tabular data coming from open data sources	Interactive web application that shows KPIs according to three top-level indicators

Table 3.1: Features of TOAD

3.2 Technology used: Python

Python is an open source programming language used to implement both standalone programs and software applications in several fields. The top Python's technical strengths are the following (from [66] and [62]):

- It's Object-Oriented: Python supports object-oriented programming mode. Its class model provides features such as polymorphism, operator overloading, and multiple inheritance. In the context of OOP languages, Python is the easiest to use.
- It's Free: Python is completely free to use and distribute, so there are no restrictions on copying it and embedding it in your systems. There are many open-source Python libraries to exploit for Data manipulation, Data Visualization, Statistics, Mathematics, Machine Learning, Natural Language Processing and more.
- It's Portable: Python is designed to run on Windows and Linux environments and it can easily be ported to multiple platforms.
- It's Easy to Learn: as a programming language, it is very simple and readable, since only few lines of code are needed to perform tasks compared to other traditional languages. For this reason it is highly suggested to beginning programmers.
- It's Well-Supported: Python is strongly supported from a large community of users. There are plenty of useful analytics libraries, documentations and support materials available for free.

Today Python programming language is very widely spread in different industries like scientific computing, oil, gas, finance, physics, signal processing, and more. IT companies begin to use Python in application creation, particularly in the areas of data science and data engineering. Python allows the usage of a wide range of third-party packages and utilities that support programmers in data processing tasks, including the analysis and visualization of large datasets. In this context, the main tools used for the dashboard development are briefly described, pointing out their functionalities in the data analysis process.

Data Manipulation: Pandas
Pandas is a great Python open source library that allows to manipulate data contained in spreadsheets and CSV files with just few lines of code. This Python library provides essential data structures like Series and Dataframes that allow the manipulation of datasets. In particular, a Dataframe is a two- dimensional Array visually represented in the form of a table.
Some interesting Python Pandas features are the following (from [34]):

1. Data management: Pandas gives us an efficient and fast way to handle and explore data by exploiting Series and Dataframes structures.
2. Data alignment and indexing: an important factor when we work with data is their organization. Pandas allows to label and organize data by using efficient methods of alignment and indexing.

3. Missing data management: since we usually deal with very raw data, one of the main issues is the occurrence of missing data or values. Pandas provides features to address this type of inconvenience.

4. Data cleaning: Pandas gives the possibility to make data clean. As we said, data are very crude, so it is extremely important cleaning data in a way that not only the code result clear but data is also well-organized.

5. Multiple file formats supported: Pandas supports different types of file formats for reading and writing operations. The most important ones are CSV, Excel and JSON.

6. Join operations on datasets: datasets are efficiently merged, in order to avoid any problems during data analysis.

7. Data aggregations : Pandas offers the GroupBy feature to split data and group it into categories according to some criteria.

8. Perform mathematical operations on the data: there is a Pandas 's function that allows to perform mathematical operation on the data. This is helpful in case the dataset is not properly in the correct order.

Therefore, Pandas is a powerful tool because it provides useful features to manipulate and re-organize data.

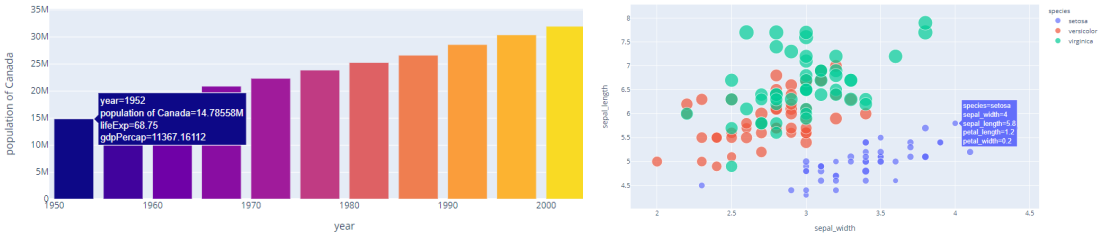
Data Visualization: Plotly Python, Dash and Folium

The framework used to implement interactive graphs is the *Plotly Python* library. It is an interactive, open source graphing library that supports over 40 chart types (e.g. bar charts, line charts, pie charts, scatterplots) applicable to statistical , financial, geographic, scientific, 3-dimensional use cases. Plotly offers some useful features in order to customize graphs as you need: it is possible to set graph size, fonts, colors, axis titles, legend entries, format ticks, make subplots, and many other. Plotly Python is built on top of the Plotly JavaScript library and allows Python developers to create beautiful interactive graphs that can be displayed in Jupiter notebooks, saved as HTML files or put inside a pure Python-built web applications using *Dash* library. Figures 3.1 and 3.2 show examples of some basic charts usage.
Dash is a user interface Python Library useful to create analytical web applications. If you want to perform data analysis, exploration, visualization, modelling and reporting, Dash represents a very powerful tool (from [46]). This library is composed by a rich set of interactive web-based components that make the Dash app code declarative and reactive. The three libraries that provide these web components are the following:

1. *Dash Core components* that includes the functions to display the core components of a Dash application such as dropdown, radio item, checkbox.

2. *Dash HTML compontents* that allow to use HTML features to create and customize the layout of the Dash application using functions written in Python.

3. *Dash Bootstrap components* that is a library that includes Bootstrap components for Dash applications by which it is possible to build styled apps with complex, responsive layouts (from [33]).



(a) Bar chart (from [7]) (b) Scatter plot (from [8])

Figure 3.1: Plotly examples of bar chart and scatter plot

(a) Line chart (from [[9]]) (b) Pie chart (from [10])

Figure 3.2: Plotly examples of line chart and pie chart

So many interactive elements can be implemented (e.g dropdown, radio item, checkbox, etc.) to allow the user an easy access of data. It is possible to customize every aesthetic element of the app: the sizing, the positioning, the colors, the fonts. In order to make the app reactive and bind data analysis code to the user interface, inputs and outputs of the application interface are described declaratively through a decorator. This decorator manages the new inputs coming from the user interaction with a component (e.g. the selection of an item in the dropdown) and provides back an output displayed within the Dash app.
Dash applications are built on top of Flask and React frameworks. A Dash application is a web server running Flask to communicate JSON packets over HTTP requests. To render components, Dash uses React.js, a Javascript user interface library that allow to display changes in the app when some input and output events occur (from [46]). Figure 3.3 is an example of the interaction with a dropdown menu within a Dash application.

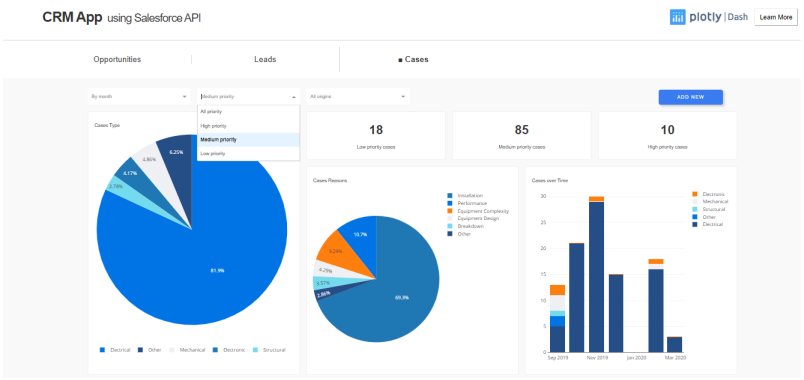
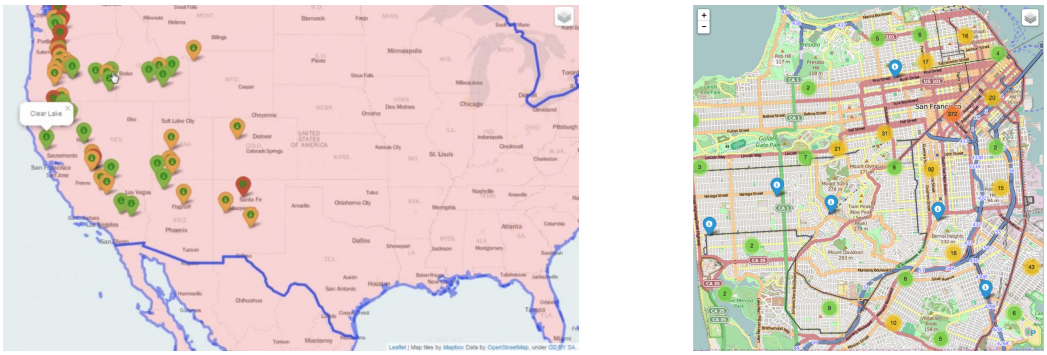


Figure 3.3: Example of Dash app with user interaction (from [11])



(a) Marker and popup (From [12]) (b) Cluster of markers (From [13])

Figure 3.4: Folium map examples of markers

To perform a map visualization of spatial data, *Folium* library is used. Folium is a powerful Python library based on Leaflet Javascript library that allows to create maps and render them as html files. Due to the interactivity feature, this library is very useful for dashboard building. By default, a base map created by Folium has the style of OpenStreetMap but you can customize it by choosing one of the different Leaflet maps. You can pass spatial points to Folium map function to display them by customized markers and popups that describe some information. Marker's color and also marker's icon can be changed according to user needs (Figure 3.4a). There is the possibility to group markers at varaying zoom levels (Figure 3.4b). Folium provides many other features (e.g. chloropeth maps) very useful to perform analysis that imply the use of map visualization.

In conclusion, a brief description of all the technologies applied in the development of TOAD is given to understand the general structure of the web application. As we can see in Figure 3.5, the architecture of the Dash application TOAD is composed by a backend in which app code retrieves data from the file system and processes them. The web server Flask communicates with the web client by exchanging http requests and responses in order to display all the graphical components and the interactive charts within a web page. So the user can interact with several components inside the web application and receive back results from the backend of the app.

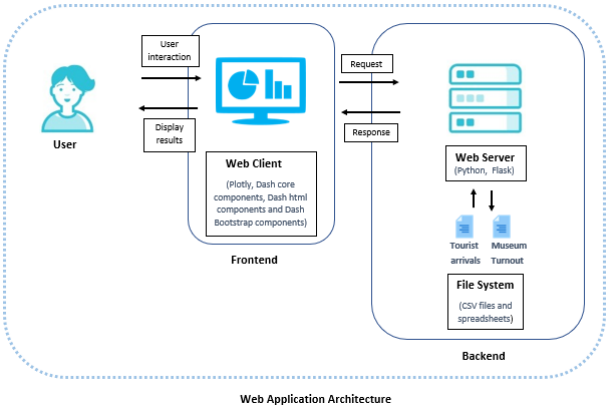


Figure 3.5: Application Architecture (from [32])

Therefore, once the data is manipulated and structured using Pandas library features, they can be interpreted and viewed by means of appropriate Python libraries: Plotly's functions allow to draw interactive data charts; Folium provides tools to build interactive maps to display spatial data; Dash library components are useful to design and build a web-based application. By combining these elements, it is possible to build an interactive informative dashboard.

Chapter 4

Data Exploration

Usually a data analysis process starts with the exploration of the datasets found in order to extract useful knowledge. In fact in this phase all the tabular data coming from different data sources have been analyzed to discover interesting data features and measure meaningful KPIs. All these data are open, it means they have been retrieved by public data plaforms that allow the user to obtain data autonomously. Therefore, a brief description of data characteristics is given.

4.1 Datasets

A dataset is a collection of data. In the case of tabular data, a dataset corresponds to one or more tables composed by columns and rows. A column represents a specific variable or attribute while a row corresponds to a record of the dataset in analysis. The dataset lists values for each variable of each record.
Datasets analysed to generate the dashboard have been retrieved by Open web repositories from which it is possible to freely download tabular data in csv file format in order to perform data analysis. In this context we introduce the concept of *Open data*. Open data are information freely available to everyone with the possibility of using them and republishing without restrictions from copyright or other mechanisms of control. This allows to anyone to get data that could be analysed to build reportings or data dashboards.
A description of the different datasets used in the data analysis process is provided based on the data source from which they come. The five explored data sources are the following:

1. AperTO (from [14])
2. Dati Piemonte (from [15])
3. Fondazione Torino Musei Open data (from [16])
4. Rock Project Open data (from [17])
5. Yacme Open data (from [18])

1. APerTO
AperTO is an open data platform provided by the City of Turin in which data collected by the municipal administration and third parties can be searched, visualized

and downloaded. Data can be selected by topics to facilitate the research. In this case study, four datasets about tourism have been explored. The datasets are the following:

- *Numero di passaggi negli Uffici del Turismo*
- *Tessere Torino+Piemonte Card emesse*
- *Top 10 musei per numero di accessi con Torino+Piemonte Card 2018*
- *Arrivi turisti*

Numero di passaggi negli Uffici del Turismo
The dataset *Numero di passaggi negli Uffici del Turismo* is a table that contains aggregate data regarding the monthly number of people who went to a Turism Office in the years 2014-2018. Figure 4.1 shows a preview of the dataset. This table is composed by three columns: month of the year (Mese), year (Anno) and number of transits in the Turism Offices (n passaggi). The temporal granularity of data is the month of the year. Therefore, for each month of each year in the range 2014-2018, the counting of Turism Offices visitors is given.

MESE	ANNO	N_PASSAGGI
GENNAIO	2014	17166
FEBBRAIO	2014	14752
MARZO	2014	16831
APRILE	2014	23623
MAGGIO	2014	22360
GIUGNO	2014	17891
LUGLIO	2014	23156
AGOSTO	2014	31901
SETTEMBRE	2014	26593
OTTOBRE	2014	22663
NOVEMBRE	2014	19671
DICEMBRE	2014	19511
GENNAIO	2015	15449
FEBBRAIO	2015	13277
MARZO	2015	15148
APRILE	2015	38429
MAGGIO	2015	62940

Figure 4.1: Preview of Numero di passaggi negli Uffici del Turismo Table (from [19])

Tessere Torino+Piemonte Card emesse
Table *Tessere Torino+Piemonte Card emesse* contains tabular data about the monthly number of Torino+Piemonte Cards issued in the years 2014-2018. The Torino+Piemonte Card is a special card that provides several advantages to tourists: free ticket to the most important Turin museums, Royal Residences and other historical buildings; reduced ticket to different cultural sites of Piedmont; special 10% discount on the

standard fare for tourists who take the City SightSeeing Turin bus; reduced ticket for the main tourist services of Turin; discounts on important events such as guided tours in the Piedmont region.
Figure 4.2 shows a preview of the dataset. It is a table in which rows are described by three variables: month (mese), year (anno), number of cards issued (n_tessere). Data are aggregated by a monthly temporal granularity. So, for each month of each year in the range 2014-2018 the number of Torino+Piemonte cards issued is provided.

MESE	ANNO	N_TESSERE
GENNAIO	2014	3038
FEBBRAIO	2014	1533
MARZO	2014	3379
APRILE	2014	8307
MAGGIO	2014	6403
GIUGNO	2014	2683
LUGLIO	2014	2427
AGOSTO	2014	4180
SETTEMBRE	2014	3447
OTTOBRE	2014	3251
NOVEMBRE	2014	2103
DICEMBRE	2014	4055
GENNAIO	2015	2499
FEBBRAIO	2015	989
MARZO	2015	1683
APRILE	2015	5421
MAGGIO	2015	5537

Figure 4.2: Preview of Tessere Torino+Piemonte Card emesse Table (from [20])

Top 10 musei per numero di accessi con Torino+Piemonte Card 2018
Table *Top 10 musei per numero di accessi con Torino+Piemonte Card 2018* presents interesting information to display inside the dashboard. It is a table that provides the ranking in 2018 of the top ten museums in Turin based on the number of museum entrances with Torino + Piemonte card. Figure 4.3 shows a preview of the dataset, in which there are ten rows and four columns. Table columns correspond to: ranking position (posizione), museum name (denominazione), year (anno), total amount of entrances (tot_ingressi). The evaluation parameter that determines the ranking is the number of museum entrances.

Arrivi turisti
Dataset *Arrivi turisti* is composed by four tables, each one recording data of a specific year in the range 2016-2019. Each row of each table provides the monthly number of tourists arrivals and the percentage change compared to the previous year. Figure 4.4 shows a preview of the dataset.

POSIZIONE	DENOMINAZIONE	ANNO	TOT_INGRESSI
1	MUSEO EGIZIO	2018	38891
2	MUSEI REALI	2018	36188
3	MUSEO NAZIONALE DEL CINEMA	2018	26250
4	REGGIA DI VENARIA REALE	2018	24539
5	MUSEO CIV. D'ARTE ANTICA PALAZZO MADAMA	2018	15743
6	BASILICA DI SUPERGA	2018	14666
7	MUSEO NAZIONALE DELL'AUTOMOBILE	2018	11331
8	MUSEO NAZ. DEL RISORGIMENTO ITALIANO	2018	9335
9	MUSEO DELLA JUVENTUS	2018	5648
10	PALAZZINA DI CACCIA DI STUPINIGI	2018	4998

Figure 4.3: Preview of Top 10 musei Torino+Piemonte Card 2018 Table (from [21])

ANNO	MESE	N_ARRIVI	VAR_ANNO_PRECEDENTE_%
2019	1	89937	3,12
2019	2	96933	-3,91
2019	3	136077	14,27
2019	4	138946	7,21
2019	5	133537	4,39
2019	6	110193	2,15
2019	7	118667	8,65
2019	8	105178	5,19
2019	9	134821	9,63
2019	10	152663	8,75
2019	11	137018	9,63
2019	12	127637	6,69

Figure 4.4: Preview of Arrivi turisti Dataset (from [22])

2. Dati Piemonte

Dati Piemonte is a web portal in which it is possible to retrieve open data about Piedmont region. For the aim of the analysis, the dataset retrieved by this data source is: *Flussi turistici per settore complessivo e Comune*. It is a table that includes the yearly number of presences and arrivals of tourists in the accomodations of some municipalities of the Piedmont region. Presences correspond to the number of nights spent by guests in the accomodations while arrivals are the total number of guests of the accomodations in the period considered. The municipalites under analysis must have at least six accommodations. Presences and arrivals values are collected on the basis of two different levels of aggregation: total count of presences and arrivals and count of presences and arrivals distinguished by Italian or foreign nationality. Figure 4.5 shows a preview of the dataset. In particular, data information are described by twelve columns. They are: year (anno), municipality (comune), province (provincia), accomododations (esercizi), beds (letti), italian arrivals (arrivi italiani), italian presences (presenza italiani), foreign arrivals (arrivi stranieri), foreign presences (presenze straniere), total arrivals (arrivi totali),

total presences (presenze totali) and average stay (tmp). The average stay value represents the ratio between the total presences and the total arrivals in the municipality's accomodations. The temporal data aggregation is yearly within the interval 2005-2018. Therefore, for each municipality the yearly number of accomodations, beds , total count of presences and arrivals, count of presences and arrivals based on Italian or foreign nationality and the average stay of the total presences and arrivals are recorded in the table. For the purposes of the analysis, only data relating to Turin municipality have been considered. Furthermore, since the time interval of the collected data is very long, only data about the last five years of the the temporal range have been analysed. Hence the time interval considered for the analysed data is 2014-2018.

anno	comune	provincia	esercizi	letti	arrivi_italiani	presenza_italiani	arrivi_stranieri	presenza_stranieri	arrivi_totali	presenza_totali	tmp
2007	BOBBIO PELLICE	TO	8	345	1180	5279	269	651	1449	5930	4,09
2009	IVREA	TO	10	487	7913	13118	1969	3359	9882	16477	1,67
2012	SUSA	TO	10	307	2580	5506	5067	6457	7647	11963	1,56
2012	SAMPEYRE	CN	14	1181	7342	27816	467	961	7809	28777	3,69
2015	PONZONE	AL	10	174	594	1582	108	235	702	1817	2,59
2014	TORRE PELLICE	TO	8	399	3712	18037	5069	14021	8781	32058	3,65
2018	CANELLI	AT	17	242	4065	6634	2716	7207	6781	13841	2,04
2007	SERRALUNGA D'ALBA	CN	12	140	901	2304	1224	3011	2125	5315	2,5
2013	BAGNOLO PIEMONTE	CN	7	100	463	1703	16	160	479	1863	3,89
2017	FORMAZZA	VB	20	1014	7917	17418	2607	3512	10524	20930	1,99
2008	GIGNESE	VB	7	465	4070	12391	1751	3521	5821	15912	2,73
2010	ASTI	AT	63	1452	26392	50135	16631	37461	43023	87596	2,04
2010	BIELLA	BI	22	1209	35831	89916	11694	31933	47525	121849	2,56
2011	MERGOZZO	VB	10	332	2804	4952	5715	18018	8519	22970	2,7
2012	CANTALUPA	TO	6	228	4177	15935	1582	4490	5759	20425	3,55
2014	REFRANCORE	AT	6	35	288	484	248	545	536	1029	1,92
2016	MURAZZANO	CN	6	46	95	123	257	1068	352	1191	3,38
2017	SINIO	CN	13	166	1141	1791	1738	6169	2879	7960	2,76
2007	CANDELO	BI	8	48	556	1563	110	276	666	1839	2,76

Figure 4.5: Preview of Flussi turistici per settore complessivo e Comune Table (from [23])

3. Fondazione Torino Musei Open data

Fondazione Torino Musei is an organization that takes care about the art collections of Turin. The civic museums that are part of it are *Palazzo Madama-Museo Civico d'Arte antica*, *MAO Museo d'Arte Orientale*, *GAM - Galleria Civica d'Arte Moderna e Contemporanea* and *Borgo Medievale*. In the Fondazione Torino Musei website it is possible to retrieve open data about museums turnout. In particular, data of each museum are collected in tabular form and can be exported by csv file format. So the dataset named *Affluenza del Pubblico* is made up of four tables one per each museum where rows record the daily number of tickets issued. Tickets are counted according to the type of ticket issued. As we can see in Figure 4.6, the ticket types are: full (intero), reduced (ridotto), free (gratuito), subscribers (abbonati), schools (scuole). The full ticket is the ticket with the standard fare; reduced ticket is usually referred to students who are less than 26 or people over 65; free ticket is for people under 18 and people with disabilities; subscribers are people who have museum's subscription; schools tickets are issued for school groups. The first column of each table corresponds to the calendar day on which data was recorded. The time interval of the available data is between January 2013 and January 2017.

4. Rock Project Open data

ROCK Project is an European research project with the aim of supporting the transformation of historic city centres that are afflicted by physical decay, social

Data [gg/mm/aaaa]	Intero	Ridotto	Gratuito	Abbonati	Scuole
01/01/2013	167	49	161	249	0
02/01/2013	162	84	112	323	0
03/01/2013	179	100	143	410	0
04/01/2013	156	57	146	346	0
05/01/2013	153	75	142	299	0
06/01/2013	140	69	259	477	0
07/01/2013	0	0	0	0	0
08/01/2013	42	8	23	58	0
09/01/2013	22	14	7	61	38
10/01/2013	16	21	17	55	64
11/01/2013	20	16	28	47	0
12/01/2013	54	37	31	111	0
13/01/2013	55	56	64	331	0
14/01/2013	0	0	0	0	0
15/01/2013	12	13	25	41	106
16/01/2013	13	66	15	49	69
17/01/2013	28	15	18	55	27
18/01/2013	34	9	15	70	17

Figure 4.6: Preview of Museums Turnout Table (from [24])

conflicts and poor life quality into creative and sustainable districts by exploiting their Cultural Heritage. ROCK stands for Regenaration and Optimisation of Cultural Heritage in creative and Knowledge cities, so it reflects the goal to promote the sustainable development and the economic growth of cities with an important cultural heritage. The ROCK initiative is interested in supporting the economic and cultural growth of 7 cities: Athens, Cluj-Napoca, Eindhoven, Liverpool, Lyon, Vilnius and Turin (from [2]). In particular, *people flow sensors* have been installed by the ROCK project in order to perform crowd analysis in different regions of the selected cities. These sensors measure the number of people passing near the sensor detection area per unit time. So ROCK project sensors provide measurements of the flows of people in the city.
Data collected by these sensors are stored in tabular form and published on a open web-based data platform. The platform allows to freely visualize and download data. As for the city of Turin, data are monthly collected in various csv files that record the number of visitors in the areas of the city where sensors have been installed. Data collected by sensors have been available since March 2019 and are provided month by month. Furthermore, metadata about sensors are provided which record the sensors' geographic coordinates (latitude and longitude). Figure 4.7 shows the regions of Turin in which people flow sensors have been located. Most of the regions have been selected to perform the analysis. They are some of the museums and historical buildings of the city of Turin: Borgo Medievale Parco Del Valentino, Gam Galleria Arte Moderna, Giardini Reali, Mao Museo Arte Orientale, Museo Del Risorgimento, Palazzo Madama, Palazzo Reale, Palazzo Carignano-Appartamenti Principi, Palazzo Carignano-Cortile and the four regions about *Salone Internazionale Del Libro* event (Salone Del Libro-Pad1, Salone Del Libro-Pad2, Salone Del Libro-Pad3, Salone Del Libro-PadOval).
The overall dataset provided by sensors is composed by different Tables generated monthly based on the temporal granularity of the aggregated data and the kind of information to be obtained on the visitors' movements.
The four types of datasets analysed in this thesis project are the following:

– *Daily Footfall* and *Hourly Footfall*: daily and hourly counting of visitors per

each of the selected regions of the city.

– *Daily Duration Distribution*: daily counting of visitors per each of the selected regions of the city and for each duration of stay.

– *Daily Mobility*: daily counting of visitors' movements in the city considering a start and an end region.

– *Daily Returning Visitors*: daily counting of returning and new visitors per each of the selected regions of the city.

_id	Nr	OldName	Region	lat	lon
1	0	Entire Site	All Regions' Sum	45.071192	7.685241
2	1		Borgo Medievale Parco Del Valentino	45.04892	7.684626
3	2		Da Lisa Via Baretti 2	45.058434	7.678281
4	3		Gam Galleria Arte Moderna	45.064285	7.670184
5	4		Giardini Reali	45.072612	7.687534
6	5		Mao Museo Arte Orientale	45.074763	7.680225
7	6		Museo Del Risorgimento	45.06887	7.686102
8	7		Palazzo Madama	45.071117	7.685806
9	8		Palazzo Reale	45.072841	7.686263
10	9		Piazza San Giovanni	45.072979	7.685361
11	10		Salone Del Libro - Pad 1	45.02924	7.665077
12	11		Salone Del Libro - Pad 2	45.029454	7.664369
13	12		Salone Del Libro - Pad 3	45.029726	7.663282
14	13		Salone Del Libro - Pad Oval	45.027107	7.660748
15	14		Via Saluzzo 24	45.059435	7.679075
16	15		Palazzo Carignano - Appartamenti Principi	45.068966	7.685592
17	16		Palazzo Carignano - Cortile	45.069268	7.684849
18	17		Paratissima 2019	45.071383	7.688512
19	18		Office Test 1	45.071192	7.685241
20	19		Office Test 2	45.071192	7.685241

Figure 4.7: Sensors's locations (from [25])

Daily Footfall and Hourly Footfall

Daily Footfall and *Hourly Footfall* are two collections of data that record the counting of visitors per each region in daily and hourly resolutions. Figures 4.8 and 4.9 show previews of the two tables. Each table is composed by four columns: the row identifier, the date in a timestamp format, the region and the number of visitors. In the case of Hourly Footfall table, the counting of visitors is collected per each time slot in the range 00:00-23:00.

Daily Duration Distribution

Daily Duration Distribution tables contain information about the daily number of visitors counted by region and *visitors's duration of stay*. The duration of stay can be defined as a time window in which people are detected by sensors. It is the evaluation parameter to measure the amount of people in the areas of the city in which sensors are located. The duration of stay corresponds to one of these twelve

_id	Date	Region	Visitors
9	2019-10-01T00:00:00	Mao Museo Arte Orientale	24
10	2019-10-01T00:00:00	Gam Galleria Arte Moderna	296
11	2019-10-01T00:00:00	Borgo Medievale Parco Del Valentino	416
12	2019-10-01T00:00:00	Palazzo Madama	407
13	2019-10-01T00:00:00	Museo Del Risorgimento	1127
14	2019-10-01T00:00:00	Palazzo Reale	3929
15	2019-10-01T00:00:00	Giardini Reali	2760
16	2019-10-01T00:00:00	Palazzo Carignano - Cortile	2340
17	2019-10-01T00:00:00	Palazzo Carignano - Appartamenti Principi	0

Figure 4.8: Preview of Daily Footfall Table (from [26])

_id	Date	Region	Visitors
315	2019-10-01T18:00:00	Mao Museo Arte Orientale	0
316	2019-10-01T18:00:00	Gam Galleria Arte Moderna	27
317	2019-10-01T18:00:00	Borgo Medievale Parco Del Valentino	68
318	2019-10-01T18:00:00	Palazzo Madama	41
319	2019-10-01T18:00:00	Museo Del Risorgimento	86
320	2019-10-01T18:00:00	Palazzo Reale	492
321	2019-10-01T18:00:00	Giardini Reali	196
322	2019-10-01T18:00:00	Palazzo Carignano - Cortile	230
323	2019-10-01T18:00:00	Palazzo Carignano - Appartamenti Principi	0

Figure 4.9: Preview of Hourly Footfall Table (from [27])

values: less 1 minute, from 1 to 5 minutes, from 5 to 10 minutes, from 10 to 20 minutes, from 20 to 40 minutes, from 40 to 60 minutes, from 60 to 90 minutes, from 90 to 120 minutes, from 2 to 3 hours, from 3 to 4 hours, from 4 to 6 hours, over 6 hours. In Figure 4.10 it is possible to see how data are aggregated for each region. So given a date and a sensor's position, for each duration of stay the number of visitors is recorded.

Daily Mobility

Daily Mobility tables provide information about the visitors's movements in the city of Turin. In particular, for each day of the month the number of people moving from a start position to an end position is counted. These measurements are performed for each combination of regions. Figure 4.11 shows a preview of a Daily Mobility table in which the Start Region column determines the initial position of visitors (in this case Palazzo Reale) whereas the End Region column indicates the end position of the detected people. Therefore, for each combination the daily amount of visitors is recorded.

Daily Returning Visitors

Daily Returning Visitors tables are generated for storing information about new and returning visitors for each region where sensors have been installed. Returning visitors are people who have already been to a specific region within a 90-day interval. New visitors are people who visit the region for the first time. These two information are very interesting to monitor the flows of people when important events are organized in the city. Figure 4.12 shows a preview of the Daily Returning Visitors

_id	Date	Region	Duration	Visitors
101	2019-10-01T00:00:00	Mao Museo Arte Orientale	from 20 to 40 minutes	2
102	2019-10-01T00:00:00	Mao Museo Arte Orientale	from 40 to 60 minutes	0
103	2019-10-01T00:00:00	Mao Museo Arte Orientale	from 60 to 90 minutes	3
104	2019-10-01T00:00:00	Mao Museo Arte Orientale	from 90 to 120 minutes	2
105	2019-10-01T00:00:00	Mao Museo Arte Orientale	from 2 to 3 hours	0
106	2019-10-01T00:00:00	Mao Museo Arte Orientale	from 3 to 4 hours	0
107	2019-10-01T00:00:00	Mao Museo Arte Orientale	from 4 to 6 hours	0
108	2019-10-01T00:00:00	Mao Museo Arte Orientale	over 6 hours	0
109	2019-10-01T00:00:00	Gam Galleria Arte Moderna	less 1 minute	71
110	2019-10-01T00:00:00	Gam Galleria Arte Moderna	from 1 to 5 minutes	57
111	2019-10-01T00:00:00	Gam Galleria Arte Moderna	from 5 to 10 minutes	40
112	2019-10-01T00:00:00	Gam Galleria Arte Moderna	from 10 to 20 minutes	62
113	2019-10-01T00:00:00	Gam Galleria Arte Moderna	from 20 to 40 minutes	25
114	2019-10-01T00:00:00	Gam Galleria Arte Moderna	from 40 to 60 minutes	15
115	2019-10-01T00:00:00	Gam Galleria Arte Moderna	from 60 to 90 minutes	16
116	2019-10-01T00:00:00	Gam Galleria Arte Moderna	from 90 to 120 minutes	2
117	2019-10-01T00:00:00	Gam Galleria Arte Moderna	from 2 to 3 hours	4
118	2019-10-01T00:00:00	Gam Galleria Arte Moderna	from 3 to 4 hours	2
119	2019-10-01T00:00:00	Gam Galleria Arte Moderna	from 4 to 6 hours	2
120	2019-10-01T00:00:00	Gam Galleria Arte Moderna	over 6 hours	0

Figure 4.10: Preview of Daily Duration Distribution Table (from [28])

_id	Date	Start Region	End Region	Amount Of Visitors
428	2019-10-02T00:00:00	Palazzo Reale	Mao Museo Arte Orientale	0
429	2019-10-02T00:00:00	Palazzo Reale	Gam Galleria Arte Moderna	2
430	2019-10-02T00:00:00	Palazzo Reale	Borgo Medievale Parco Del Valentino	5
431	2019-10-02T00:00:00	Palazzo Reale	Palazzo Madama	235
432	2019-10-02T00:00:00	Palazzo Reale	Museo Del Risorgimento	26
433	2019-10-02T00:00:00	Palazzo Reale	Giardini Reali	762
434	2019-10-02T00:00:00	Palazzo Reale	Palazzo Carignano - Cortile	157
435	2019-10-02T00:00:00	Palazzo Reale	Palazzo Carignano - Appartamenti Principi	0

Figure 4.11: Preview of Daily Mobility Table (from [29])

table in which data of May 2019 have been collected. In this particular month of year 2019 there was an important book Fair in Turin: the *Salone Internazionale Del Libro*. In this context, given the calendar of the event, it is possible to monitor the number of new and returning visitors. As mentioned above, four specific areas within the Salone del Libro location have been considered to install people flow sensors: Salone Del Libro-Pad1, Salone Del Libro-Pad2, Salone Del Libro-Pad3, Salone Del Libro-PadOval. These regions correspond to four pavilions of the book Fair. So, for each pavilion the daily counts of new and returning visitors are recorded.

_id	Date	Region	New Visitors	Returning Visitors
65	2019-05-09T00:00:00	Salone Del Libro - Pad 2	18116	5738
66	2019-05-10T00:00:00	Salone Del Libro - Pad 2	22934	9580
67	2019-05-11T00:00:00	Salone Del Libro - Pad 2	31555	13279
68	2019-05-12T00:00:00	Salone Del Libro - Pad 2	23036	11686
69	2019-05-13T00:00:00	Salone Del Libro - Pad 2	16417	10289

Figure 4.12: Preview of Daily Returning Visitors Table (from [30])

4. Yacme Open data

Yacme Open data is another web-based data source in which informative data about Turin museums can be retrieved. These data can be exported as csv file or by exploiting the *GeoJSON* file format. GeoJSON is an open standard format useful to encode geographic data structures called features. A *Feature* is a geometric object that can have one of the following geometry types: *Point* by which addresses and locations are represented; *LineString* to indicate streets, highways and boundaries; *Polygon* to identify countries, provinces, tracts of land; *MultiPoint*, *MultiLineString* and *MultiPolygon* that are multi-part collections of these types. Feature objects have additional properties used to provide detailed information about data. Sets of Feature objects are contained by *FeatureCollection* objects (from [44] and [45]). A FeatureCollection object can be thought as an array of Features objects.
In Figure 4.13 it is possible to see a preview of the GeoJSON file used to retrieve information about Turin museums. It represents a FeatureCollection object in which there are many Feature objects, each one describing properties of museums. Museum's properties are: name of the museum, address, postal code, city, phone number, fax number, email address and the URL of museum's website. The geometry type of Feature objects is Point since museums are locations identified by geographic coordinates (longitude and latitude) and therefore are represented as geometric points on a geographic map.


```
{
  "type": "FeatureCollection",
  "features": [
  {
    "type": "Feature",
    "geometry": {
       "type": "Point",
       "coordinates":  [ 7.652032,45.081341 ]
    },
    "properties": {
    "DENOMINAZIONE":"MAU - Museo d'Arte Urbana
    "INDIRIZZO":"VIA MUSINE' 19                          ",
    "CAP":10143,
    "CITTA'":"TORINO",
    "TELEFONO":"011 745580 - 335 6398351                        ",
    "FAX":"011 745580                ",
    "EMAIL":"info@museoarteurbana.it                                              ",
    "URL":"http://www.museoarteurbana.it                                                    "
    }
  },
  {
    "type": "Feature",
    "geometry": {
       "type": "Point",
       "coordinates":  [ 7.692383, 45.076727 ]
    },
    "properties": {
    "DENOMINAZIONE":"Archivio Storico Museo Italgas
    "INDIRIZZO":"CORSO PALERMO 3                         ",
    "CAP":10152,
    "CITTA'":"TORINO",
    "TELEFONO":"011 8395312                                 ",
    "FAX":"011 8395315                ",
    "EMAIL":"archiviostorico@italgas.it                                           ",
    "URL":"http://www.italgas.it/home_italgas/la+societa/chi+siamo/archivio+storico/default1.htm"
    }
  },
```


Figure 4.13: Preview of Turin Museums data in GeoJSON format (from [31])

This GeoJSON representation of Turin museum data has been exploited and modified in order to be consistent with the other Turin museums information retrieved by other data sources. As mentioned above about Rock Project data, information on geographic positions of sensors are given. Furthermore, for the aim of the analysis only information coming from sensors installed near museums of the city have been considered. Turin museums explored in the crowd analysis of Rock Project are a small subset of all the museums of the city. For these reasons, a new GeoJSON file has been generated which includes information on Turin museums where people flow sensors have been positioned.
Figure 4.14 shows a preview of the new GeoJSON data format of Turin museums used for the data analysis process.

These data have been saved in a GeoJSON file used to create a map visualization of data by means of the Folium Python library.

```
{
  "type": "FeatureCollection",
  "features": [

  {
   "type": "Feature",
   "geometry": {
      "type": "Point",
      "coordinates":  [7.680225,45.074763 ]
   },
   "properties": {
   "DENOMINAZIONE":"Mao Museo Arte Orientale
   "INDIRIZZO":"Via S. Domenico 9
   "CAP":10122,
   "CITTA'":"TORINO",
   "TELEFONO":"011 4436927
   "FAX":"011 4436918                    ",
   "EMAIL":"mao@fondazionetorinomusei.it
   "URL":"http://www.maotorino.it
   }
  },
  {
   "type": "Feature",
   "geometry": {
      "type": "Point",
      "coordinates":  [7.670184, 45.064285]
   },
   "properties": {
   "DENOMINAZIONE":"Gam Galleria Arte Moderna
   "INDIRIZZO":"Via Magenta, 31
   "CAP":10128,
   "CITTA'":"TORINO",
   "TELEFONO":"011 4429518
   "FAX":"011 4429550                    ",
   "EMAIL":"gam@fondazionetorinomusei.it
   "URL":"http://www.gamtorino.it
   }
```


Figure 4.14: Preview of new GeoJSON data format of Turin museums

Chapter 5

Data Manipulation and Interpretation

Once collected and explored, data must be manipulated in order give them an appropriate structure. A *Data Manipulation* phase is required in which tabular data are extracted and programmatically processed by means of data structures. As previously said, Python programming language provides useful open source libraries like Pandas to make data manipulation. So data coming from CSV files and spreadsheets are read an stored into Dataframe objects and then processed through lists and dictionaries. When data are well structured and the expected results have been obtained, the next step is the *Data Interpretation*. Data Interpretation is the process by which significant KPIs extracted from the analysed data are graphically represented through appropriate graphs. For this purpose, various types of graphs are exploited and applied based on how data are to be interpreted. As mentioned above, the Plotly Python library is used to generate interactive graphs. Furthermore, to visualize some data on a map, the Folium library is exploited.
Data manipulation and interpretation processes are described for each dataset previously introduced.

1. Numero di passaggi negli Uffici del Turismo-Graphic Representation
Data coming from this table do not need particular manipulations since they are well structured. Data are read by using a Pandas function that reads a csv file and stores data inside a Dataframe object. This data structure is parsed to obtain the values to be displayed in an interactive Plotly graph.
As seen in the description of the dataset, data are aggregated with a monthly resolution. Therefore the KPI extracted from the analysis measures the monthly number of people who went to tourism offices in the 2014-2018 interval. The KPI helps to understand the trends over the years of the amount of visitors to tourism offices in Turin. This information can be visualized with a line chart representation in which five lines (one for each year) with different colors are displayed within the graph with months on the x axis and the number of visitors on y axis. Figure 5.1 shows the interactive graph with the user interaction in which we can see for the month of April the amount of visitors per each year.

2. Tessere Torino+Piemonte Card emesse - Graphic Representation
Table *Tessere Torino+Piemonte Card emesse* has a simple structure that is sim-

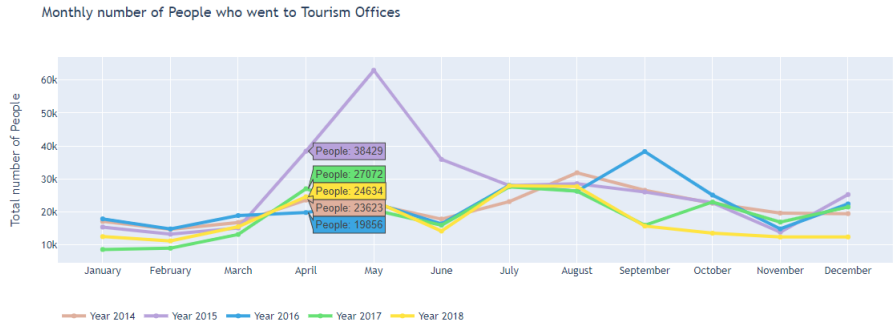


Figure 5.1: Line chart of monthly number of people who went to tourism offices

ilar to the table about visitors in tourism offices since data are aggregated with a monthly resolution. In this case, the KPI measured is the monthly number of Torino+Piemonte cards issued in the range 2014-2018. To display the results obtained by this KPI, a line chart representation is used that helps to understand the trends over the years of the amount of cards issued. For each year a line is drawn in the graph where months are displayed on x axis while the number of cards issued are visualized on y axis.
Figure 5.2 shows the line chart representation of the KPI with a user interaction with the chart. In particular, for the month of April the amount of Torino+ Piemonte cards issued is displayed for each year.

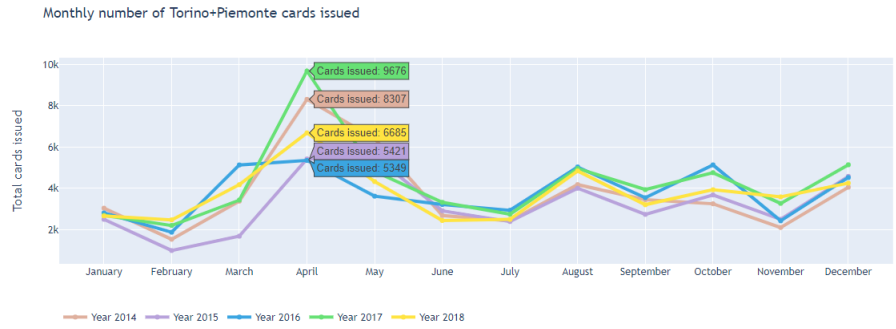


Figure 5.2: Line chart of monthly number of Torino+Piemonte cards issued

3. Top 10 musei per numero di accessi con Torino+Piemonte Card 2018 - Graphic Representation
The table provides the ranking for the year 2018 of the top 10 Turin museums based on the number of entrances with Torino+Piemonte Card. The ranking is a key indicator of the most famous museums in Turin in 2018 and it may be interesting to observe it based on the previous graph relating to the number of cards issued for the year 2018. In order to provide the idea of the ranking, a bar chart representation is used in which the number of entrances to museums is displayed from the highest value to the lowest value (Figure 5.3).

4. Arrivi turisti - Graphic Representation
As regards the dataset *Arrivi Turisti*, data provided by the four tables, one per each year in the range 2016-2019, are parsed by the Pandas function that reads csv files.

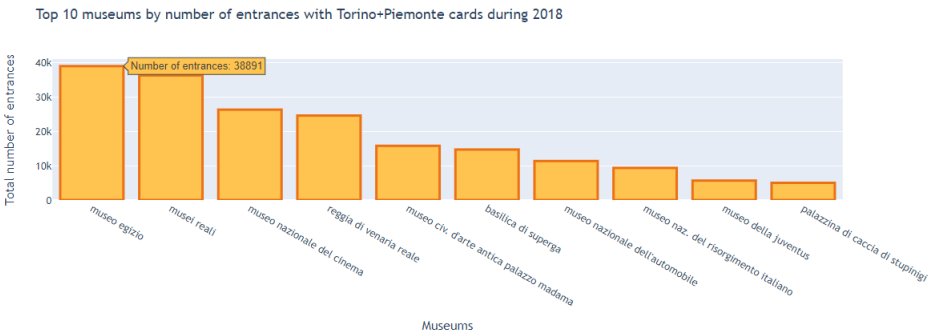


Figure 5.3: Top 10 museums by number of entrances with Torino+Piemonte cards during 2018

In particular for each table, a Pandas Dataframe object is generated from which values about the number of tourists arrivals are extracted to be displayed inside a Plotly graph. In addition, each table provides the *percentage variation* in arrivals compared to the previous year. The percentage variation is a mathematical operation to compare new and old values of a given variable. The variation is computed by subtracting the old value from the new value; this difference is divided by the old value; the result is then converted to percentage by multiplying by 100 and adding a % sign (from [49]). In this case, for a given a month, the new value and old value correspond to the values of tourists arrivals of two consecutive years. The percentage variation is calculated on these two values.
Therefore, the two KPIs extracted from the analysis of this dataset are the monthly number of tourist arrivals in the City of Turin and the percentage variation of tourist arrivals against the previous year.
Since data are aggregated by a monthly temporal granularity, it was thought to use two grouped bar charts in which the x axis corresponds to the months and the y axis corresponds to the tourists arrivals values. For each month, a group of four bars are traced, one per each year. Figure 5.4 shows the bar chart relating to the first KPI. As regards the percentage variation bar chart, it was decided to exploit a bar representation in which all the negative values are visualized as bars under the zero of the axis (Figure 5.5). In this way it is possible to best emphasize the negative values when are present.

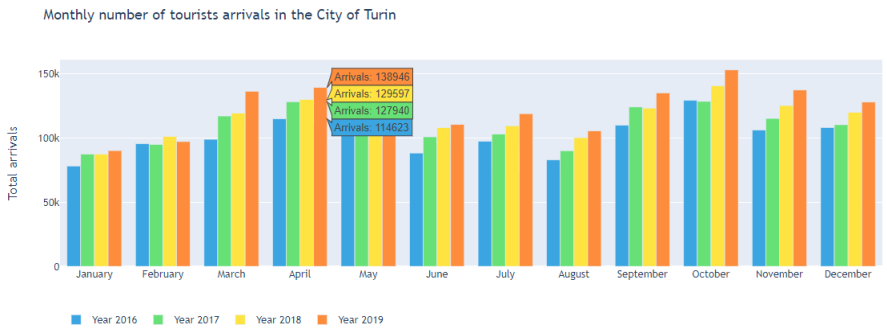


Figure 5.4: Bar Chart of monthly number of tourist arrivals in the City of Turin

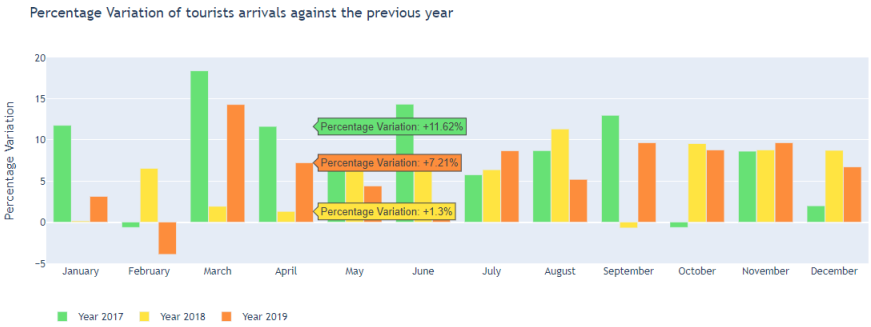


Figure 5.5: Bar Chart of percentage variation of tourist arrivals against the previous year

5. Flussi turistici per settore complessivo e Comune - Graphic Representation

As mentioned above, this dataset provides information about tourist flows in different municipalities of Piedmont region. Since this data analysis process is focused on the urban tourism of Turin, the data of interest are the ones regarding Turin's municipality. Thefore, a data manipulation phase is required to extract the relevant information. A Dataframe object is created to collect the useful data that are filtered based on the municipality's value equal to Turin and the interval of years selected for the analysis that is 2014-2018. In addition, a sorting operation by year is required in order to facilitate access to the data collected. In order to provide useful insight about flow of people in Turin accomodations, the following five KPIs are extracted:

- The number of tourist arrivals
- The number of tourist presences
- The percentage variation of tourist arrivals with respect to the previous year
- The percentage variation of tourist presences with respect to the previous year
- The average stay of tourists

As regards the first two KPIs, data about the counting of presences and arrivals are retrieved from the table. Both presences and arrivals are measured according to three aggregation levels: the total number of tourists, the number of Italian tourists and the number of foreign tourists. These three data features are interpreted by means of two grouped bar charts, one for arrivals and one for presences. Figures 5.6 and 5.7 show the graphic representations of the two KPIs and the user interactions with them. Since the temporal granularity of data is annual, the two graphs represent respectively the interval of years on the x axes and the yearly count of arrivals and presences on the y axes. For each year three bars are drawn corresponding to the three levels of data aggregation.

The third and fourth KPIs about the percentage variation of arrivals and presences are also represented with two grouped bar charts in which each bar corresponds to values of one of the three data aggregation levels.

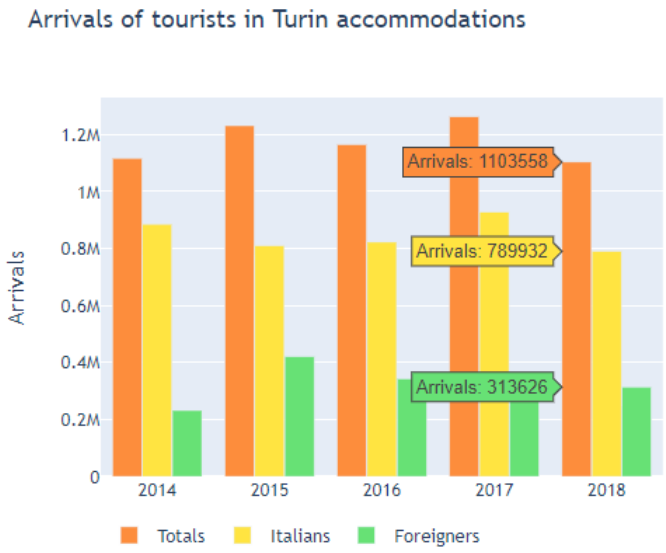


Figure 5.6: Bar chart of number of tourist arrivals

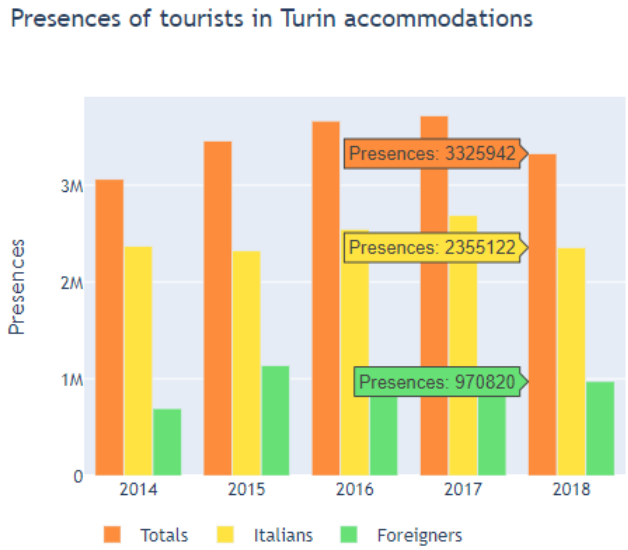


Figure 5.7: Bar chart of number of tourist presences

The last KPI describes the average stay of tourists in Turin accomodations. The *average stay* or Length-of-stay (LOS) is a key parameter of tourism that determines the number of nights spent by guests of the accomodation relative to arrival numbers. It corresponds to the ratio between the number of presences and the number of arrivals in tourist accomodations. The table provides only the average stay computed considering the total number of presences and the total number of arrivals. Therefore, it was thought to also measure the average stay considering the Italian and foreign presences and arrivals. Figure 5.10 shows the graphic representation of the KPI with the user interaction in which three values of average stay are displayed for the three aggregation levels.

6. Affluenza del Pubblico - Graphic Representation

The four tables of the dataset Affluenza del Pubblico provide information on the

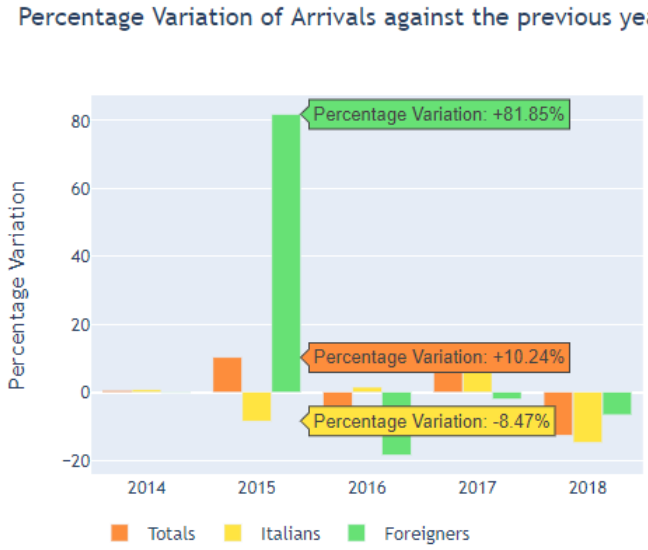


Figure 5.8: Bar chart of percentage variation of tourist arrivals with respect to the previous year

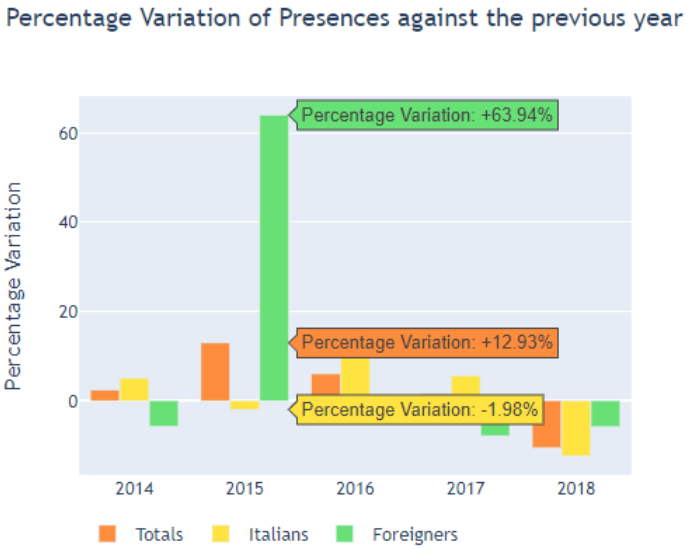


Figure 5.9: Bar chart percentage variation of tourist presences with respect to the previous year

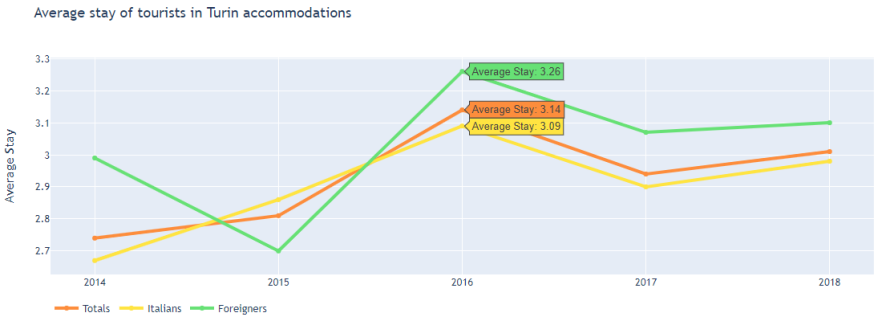


Figure 5.10: Line chart of average stay of tourists

number of tickets issued in the museums of Palazzo Madama, MAO, GAM and Borgo Medievale. As previously seen, data are aggregated based on the ticket type with a daily granularity from January 2013 to January 2017. For each museum, data have been manipulated in order to extract the following KPIs:

- Daily number of entries
- Monthly number of entries
- Yearly number of entries
- Yearly number of entries for a given month of the year

Since the type of data aggregation which provides the count of ticket issued per each ticket category (full, reduced, free, subscribers, schools), it was decided to use a stacked bar chart representation to clearly visualize the trends of each ticket type. The user interaction is essential to access information about museums turnout. For this purpose, the TOAD application provides some graphical components to allow user to set the temporal granularity of data to be displayed.
As for the first KPI, no particular manipulation of data is necessary since they are provided in daily resolution. Data are always filtered according to the date range selected by the user and a Dataframe object is created to structure data and provide a graphical representation of them. Figure 5.11 show the daily number of entries to MAO museum with the user interaction in which for each ticket type the number of visitors for a given date is given . As we can see, the x axis corresponds to the interval of dates of the needed data (in this case the time range goes from 06-16-2016 to 07-16-2016) instead the y axis corresponds to the number of entries.

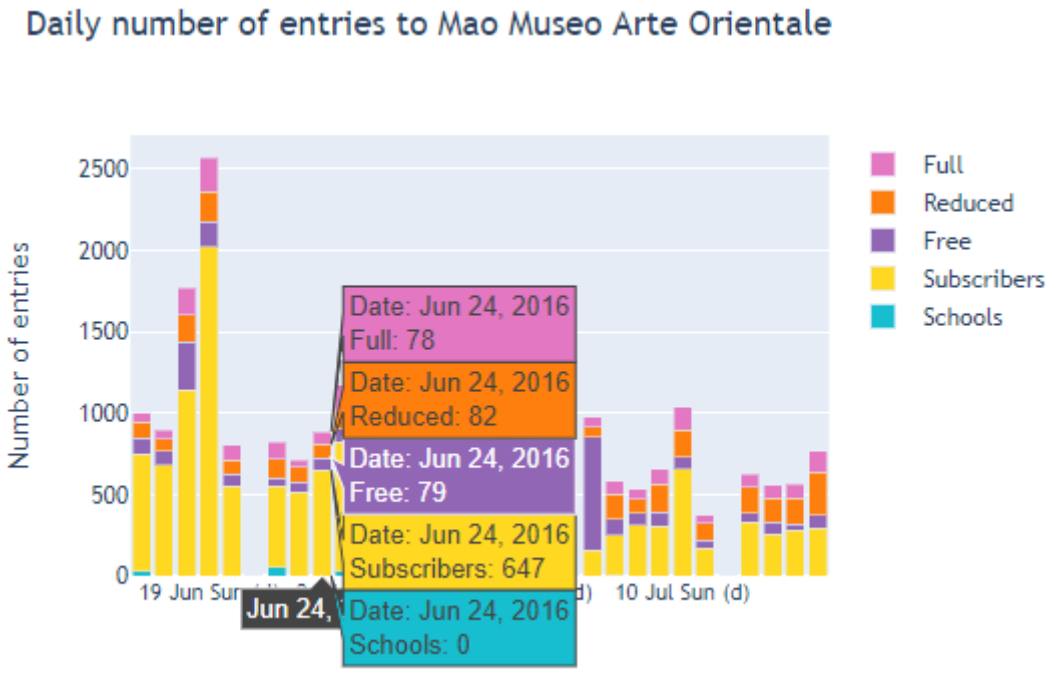


Figure 5.11: Bar chart of daily number of entries to MAO museum

To compute the monthly number of entries to museums of a given year the total

sum per each month is applied. Therefore, a Dataframe object of the tabular data is parsed to build a 2-dimensional array of 5 rows , one per each ticket type and 12 columns, one per each month. This data structure is then used to display a stacked bar chart for each museum in which the x axis represents the months of a given year and the y axis shows the number of entries. Figure 5.12 show for the MAO museum the monthly number of entries for each ticket type issued.

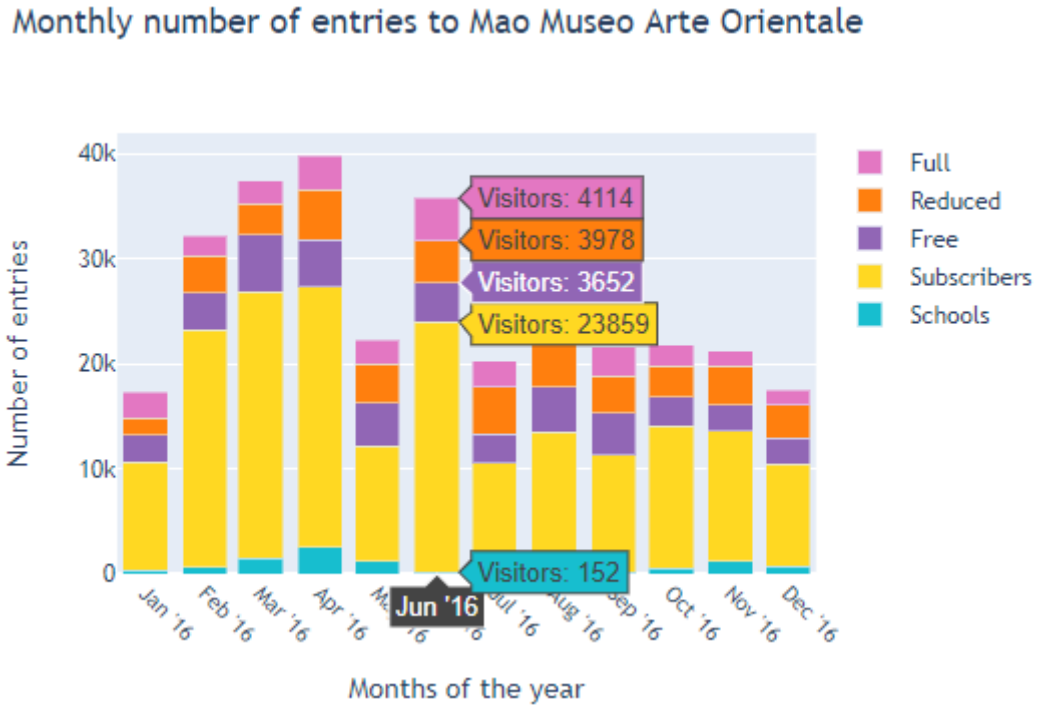


Figure 5.12: Bar Chart of monthly number of entries to MAO museum

Also for the third and fourth KPIs the total sum is used to compute the yearly number of entries for each of the four museums. Therefore, two bar charts are generated: for the yearly number of entries each bar corresponds to the total count of tickets issued in a given year; for the yearly number of entries in a given month of the year the user can observe the trends of that month in different years. The month is selected by the user by means of a dropdown menu element inside the dashboard. In Figures 5.13 and 5.14 it is possible to visualize the two KPIs in the case of the MAO museum.

7. Daily Footfall and Hourly Footfall - Graphic Representation
Daily Footfall and Hourly Footfall are two datasets each one composed by different tables, one for each month in which data are collected by people flow sensors.
For the Daily Footfall dataset two KPIs are measured:

- daily number of visitors per each region
- percentage of visitors in working days and non-working days

The purpose of these two KPIs is to provide to the final user a graphical representation of the number of visitors and the percentage of visitors in working days and non-working days in some Turin regions for a given time interval. To represent the

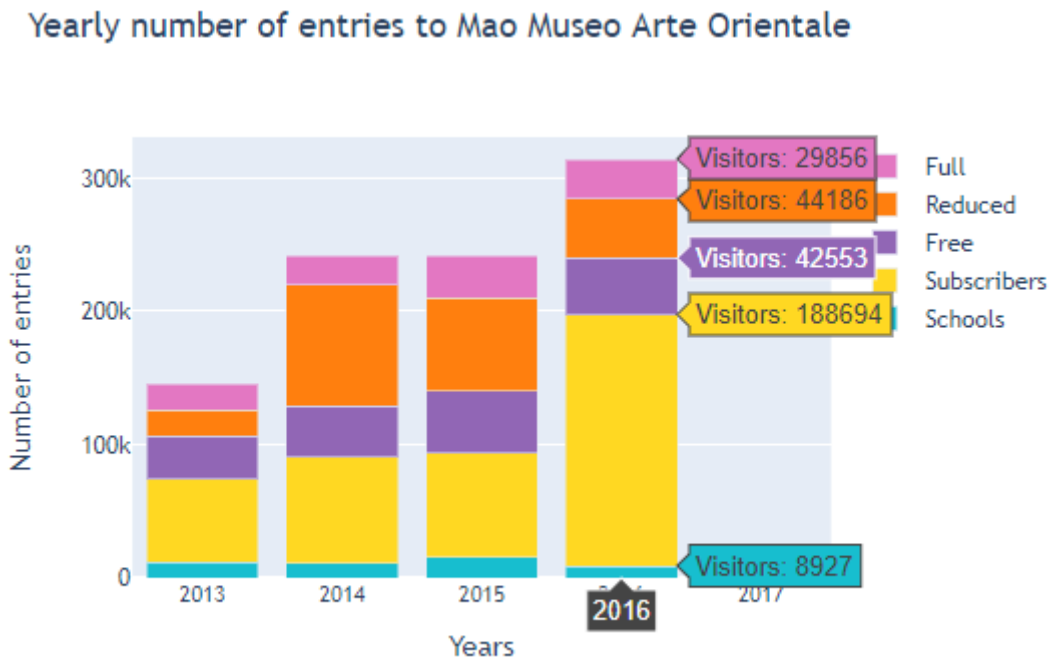


Figure 5.13: Bar Chart of yearly number of entries to MAO museum

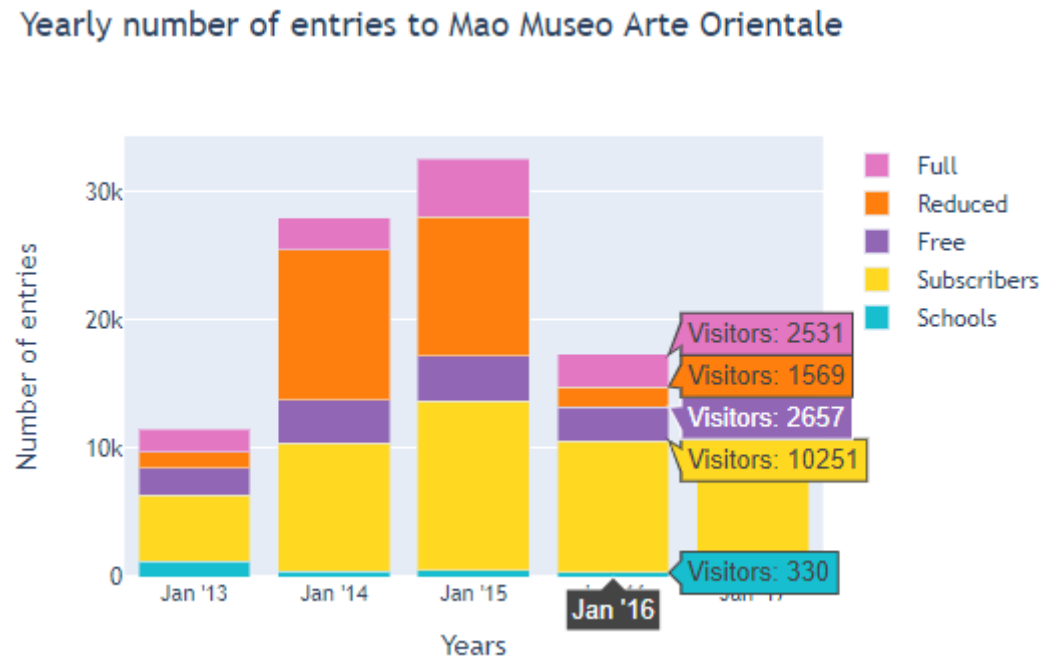


Figure 5.14: Bar Chart of yearly number of entries for a given month of the year to MAO museum

daily number of visitors a stacked bar chart representation is adopted (Figure 5.15) in which each bar represents the number of visitors distinct by region in a given day of a given date range.
To measure the percentage of visitors for each region, the *workalendar* Python library is exploited to check the working days and non-working days on the date range

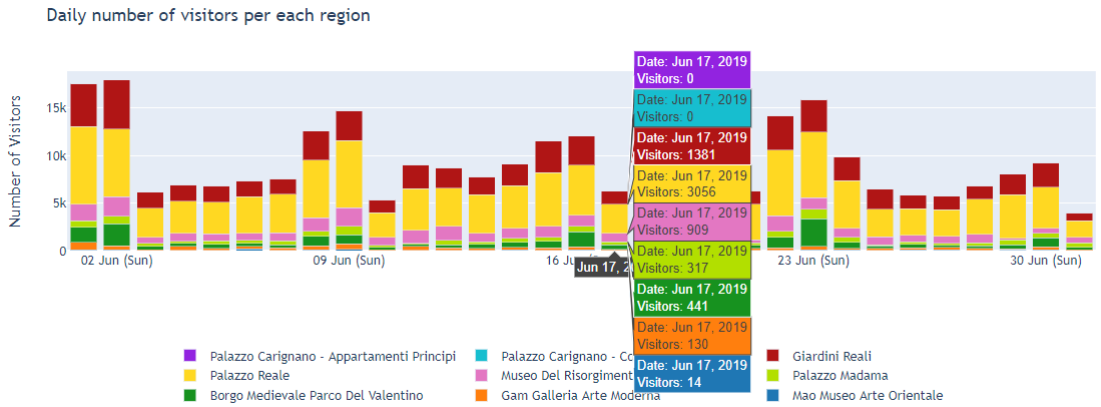


Figure 5.15: Bar Chart of daily number of visitors per each region

selected by the user. To display results two pie charts are used (Figure 5.16), one for values about working days and the other for values about non-working days. Therefore for a given date range selected by the user, all the dataset is processed and filtered in order to display the required results.

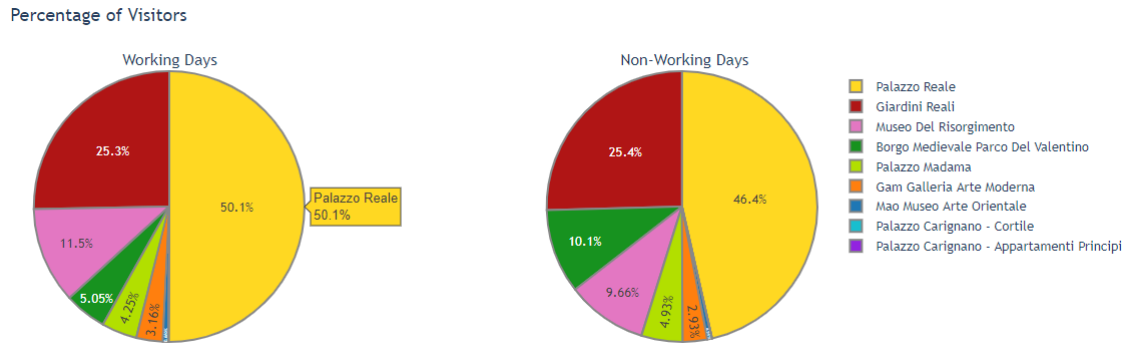


Figure 5.16: Pie chart of percentage of visitors per each region in working days and non-working days

A particular case in which the Daily Footfall dataset is used is to retrieve information about the Salone del Libro event. In this context, given the calendar date of the event, the daily number of visitors (Figure 5.17) and the percentage of visitors (Figure 5.18) to Salone del Libro are extracted.
The Hourly footfall dataset is useful to retrieve the hourly number of visitors for a given date. So data are processed and filtered according to the selected regions and the date choosen by the user. Since time is represented in the form of timestamp, some manipulation is needed in order to split the date value from the hour value and this is useful for a better representation of data. The graphical representation used for these hourly data is the grouped bar chart (Figure 5.19) in which the x axis corresponds to the hours of the day instead the y axis includes the number of visitors. Therefore, for each hour of day, a group of bars is displayed, each representing one of the regions.

8. Daily Duration Distribution - Graphic Representation
The tabular data coming from the Daily Duration Distribution dataset store for

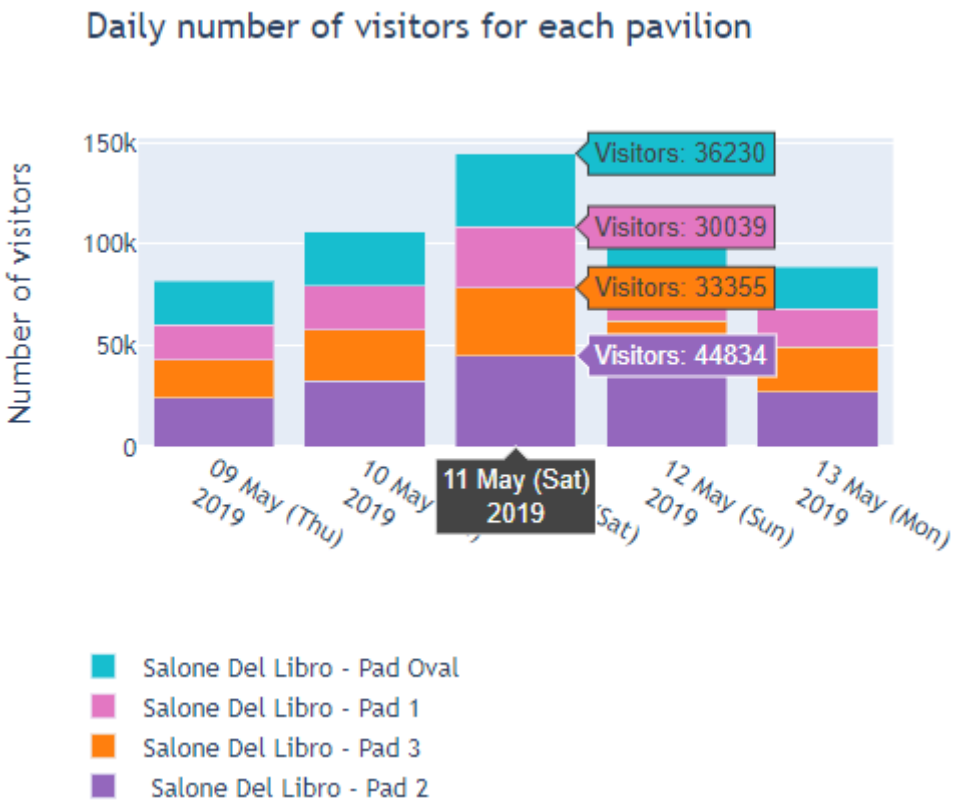


Figure 5.17: Bar chart of daily number of visitors per each pavilion of Salone Del Libro 2019

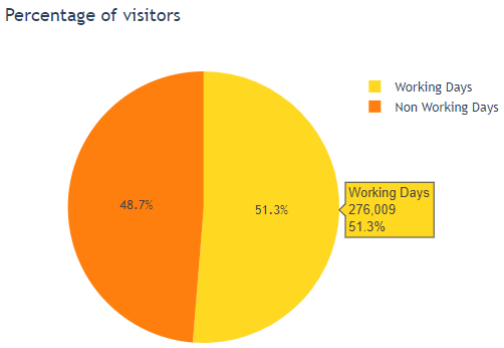


Figure 5.18: Pie chart of percentage of visitors of Salone Del Libro 2019 by working days and non-working days

each region the daily count of visitors distinct by the visitor's duration of stay. In order to better display these information, it was decided to generate a line chart according to some inputs coming from the user interaction. In particular, by setting a date range and a duration of stay, the daily number of visitors is displayed. Figure 5.20 shows the line chart in which each line corresponds to values of a region.

9. Daily Mobility - Graphic Representation

As previously seen, the Daily Mobility dataset provides for each combination of regions the daily number of visitors that move from a start position to an end position. In order to take full advantage of the characteristics of this dataset, it

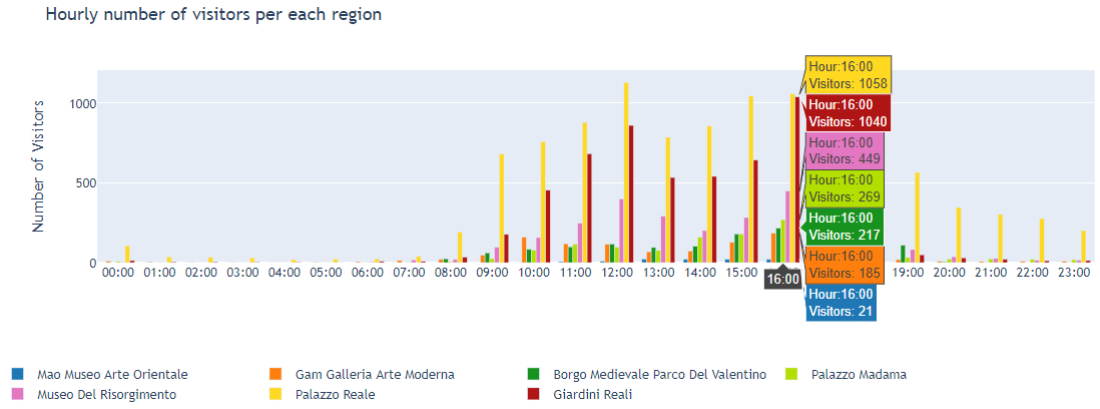


Figure 5.19: Bar chart of hourly number of visitors per each region

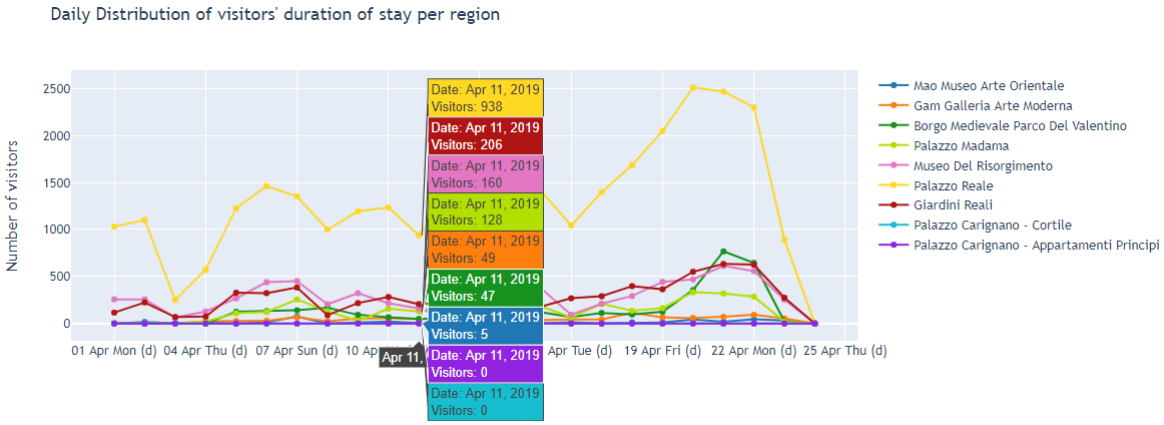


Figure 5.20: Line chart of daily distribution of visitors' duration of stay per each region

was thought to provide the number of visitors by setting a date range and a start position. So the user by selecting a date interval and a start region can visualize inside the dashboard two kinds of representations:

- a stacked bar chart
- a map of the regions

The bar chart representation (Figure 5.21) provides on the x axis the date range selected by the user and on the y axis the amount of visitors that were in the selected start region and were detected in the other regions.
On the map the selected regions are displayed with a colored marker: the start position is identified by an orange marker while the end positions are colored according to a color scale that associates a color with the amount of visitors of the region. In this case the amount of visitors is the sum of the values related to the date range considered. Figure 5.22 shows the map visualization with the user interaction in which a marker is selected and a popup is visualized that provides some information on the region and the number of visitors.

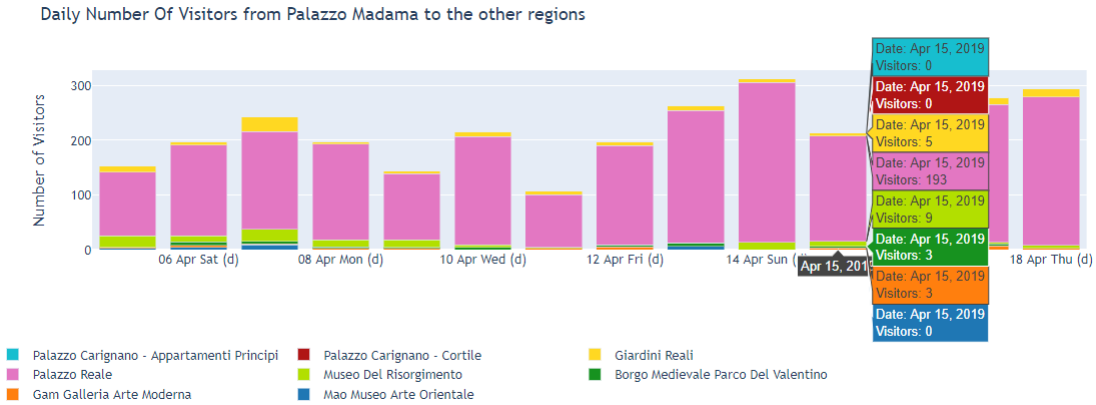


Figure 5.21: Bar chart of daily movements of visitors from a start region to other regions

Figure 5.22: Map of daily movements of visitors from a start region to other regions

10. Daily Returning Visitors - Graphic Representation

The Daily Returning Visitors dataset provides the daily number of new visitors and returning visitors distinct by regions. This kind of information could be useful to monitor the visitors numbers of an event. In this context the dataset is exploited in order to retrieve the new and returning visitors of the Salone del Libro event. Therefore, data are processed and filtered based on the calendar of the event and displayed with a line chart representation. In particular, data are split into two subplots (Figure 5.23), one for representing new visitors and the other for representing returning visitors. So a line is drawn to indicate the amount of visitors for each pavilion of the event.

Figure 5.23: Daily Returning Visitors Salone Del Libro Line Chart with user interaction

In conclusion, the aim of this section is to describe the graphical representations applied according to the data interpretation. Therefore, the measures extracted from the analysed data have been visualized inside interactive graphs that allow the user to easily obtain results.
In the next section an overview about the visual appearance of the TOAD application is given in order to understand the way the various information are displayed inside the dashboard.

Chapter 6

Data Visualization on the Dashboard

The visualization of the graphs inside the dashboard is the last phase of the development process in which all the analysis made on data are provided to the user. The goal is to provide an informative and interactive dashboard, easy to use and understand that the user can exploit to make decisions on urban tourism trends in the City of Turin. To this end, the TOAD dashboard is designed to be a web application in which all the graphical representations are organized into three main sections according to three top level indicators. The three sections are implemented as dropdown menus in which there are several sub sections for browsing the various graphs based on the type of information to be recovered. Each sub section is implemented as a navigation link to access to the page where graphs are rendered. Each page also includes a button showing a popup that indicates references to data sources. In some pages a chart info button is provided in order to help user to understand the parameters represented within the charts.
For each of the three sections a brief descripition of the graphs's organization is provided. This allows to visualize the graphical appearence of the dashboard application. The three sections are the following:

1. Tourists Flow
2. Museum Entry Tickets
3. People Flow Counting

6.1 Tourists Flow

The Tourists Flow section is composed by all the graphs that show information about the tourists flow in the City of Turin. For this purpose, the graphs relating to the arrivals of tourists to the city, the number of tourists who went to tourist offices and the arrivals and presences of tourists in the Turin accomodations are included in the section, divided into three pages that the user can access by using the navigation items within the dropdown menu (Figure 6.1). The Three Pages are:

1. Tourist Arrivals
2. Tourist Offices

3. Tourism Accomodations

The Tourists Arrivals page includes two charts that show respectively the monthly number of tourist arrivals in the City of Turin and the percentage variation of tourists arrivals against the previous year. Figure 6.2 is a preview of the page in which the first graph about number of arrivals is displayed. 6.4

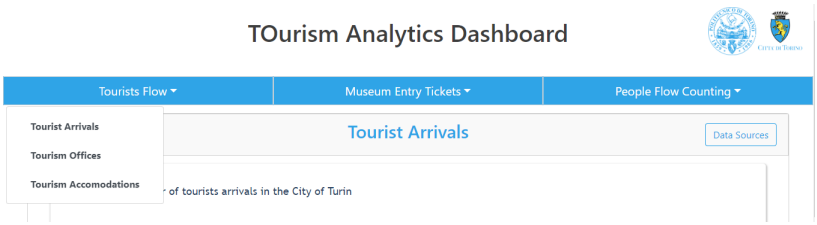


Figure 6.1: TOAD Tourists Flow - navigation items

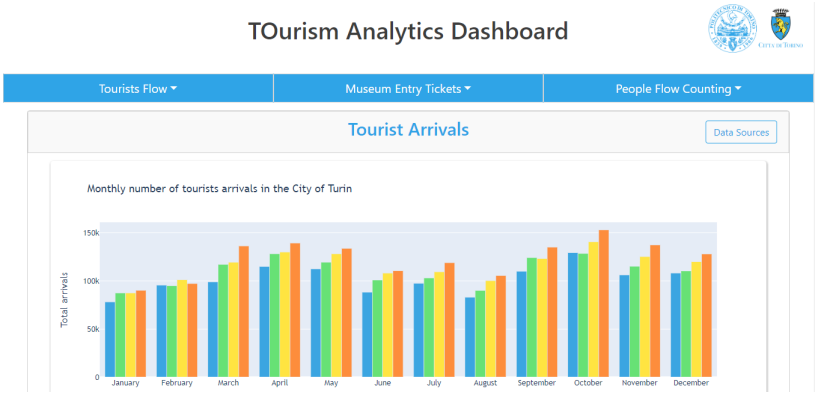


Figure 6.2: TOAD Tourists Flow - Tourist Arrivals

The Tourist Offices page includes the graphical representation of the KPI on the number of people went to tourism offices during the years 2014-2018. Figure 6.3 shows the chart's view within the TOAD application.

Figure 6.3: TOAD Tourists Flow - Tourism Offices

The Tourism Accomodations is composed by all the charts that provide information on presences and arrivals of tourist in tourism accomodations of Turin. Therefore

the five charts on number of arrivals, number of presences, percentage variation of arrivals, percentage variation of presences and average of stay in the tourism accomodations are displayed. Figure 6.4 provides a preview of the page in which the two charts about number of arrivals and number of presences are displayed in a grid layout with two columns.

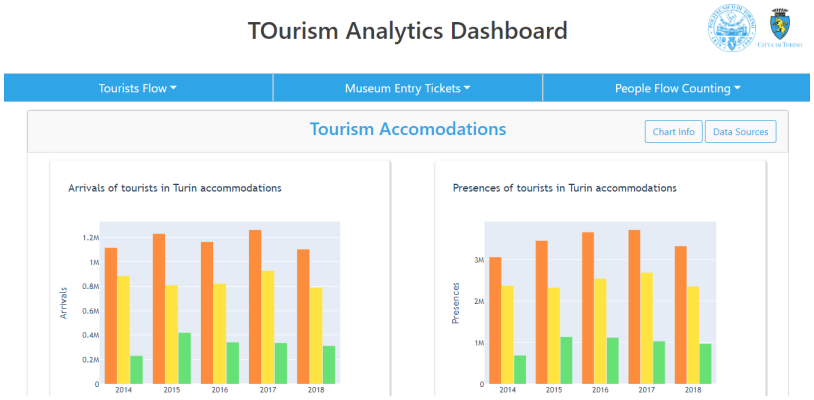


Figure 6.4: TOAD Tourists Flow - Tourism Accomodation

6.2 Museum Entry Tickets

The Museum Entry Tickets section includes the graphs that regard the flow of visitors to museums. The graphs are organized into two pages (Figure 6.5) according to two types of information:

1. Torino + Piemonte Card
2. Museum Turnout

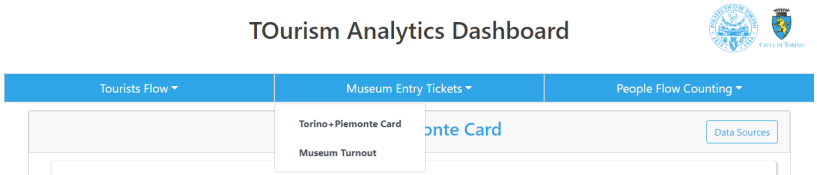


Figure 6.5: TOAD Museum Entry Tickets - navigation items

As regards Torino+Piemonte Card page, two graphs are represented, the one relating to the monthly cards issued in the years 2014-2018 and the one on the ranking of the top 10 museums based on the number of entrances with Torino + Piemonte card. Figure 6.6 is a preview of the Torino+Piemonte Card page in which the first chart displayed is the one on the number of card issued.
The Museum Turnout Page shows all the charts generated by the data analysi of Affluenza del Pubblico dataset relating to four Turin museums (MAO, GAM, Borgo Medievale and Palazzo Madama). As mentioned above, three different temporal granularities (daily, monthly and yearly) are exploited to retrieve the count of ticket issued. In order to help the user to explore data by analysing the different temporal granularities, some graphical components are provided. In particular, to access a

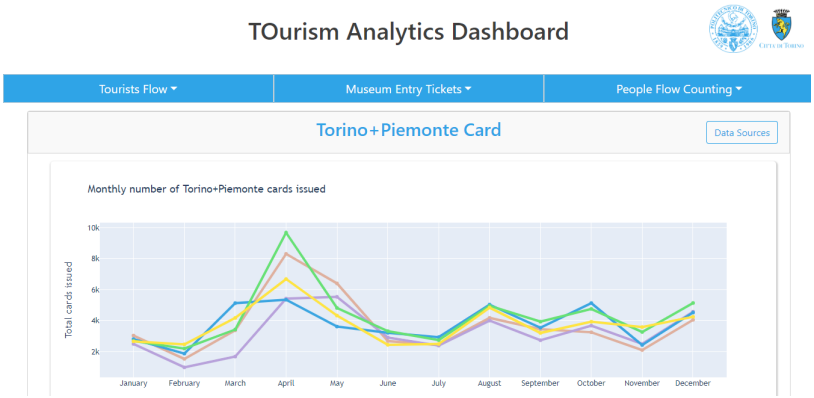


Figure 6.6: TOAD Museum Entry Tickets - Torino+Piemonte Card

daily granularity of data, a date picker range element is implemented in which user can select an interval of dates and obtain the information on the four museums with a daily resolution (Figure 6.7).

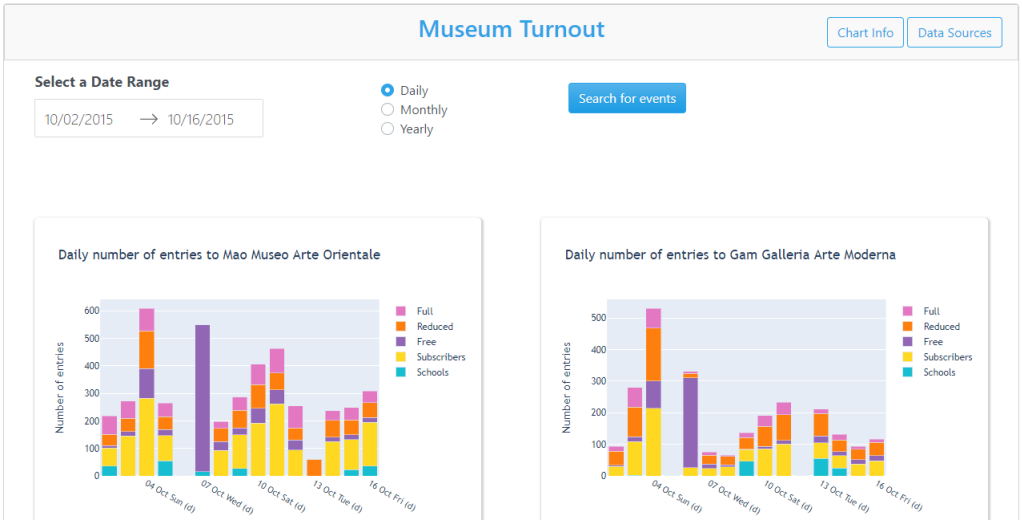


Figure 6.7: TOAD Museum Entry Tickets - Affluenza Pubblico data with daily granularity

To visualize data with a monthly temporal granularity, a dropdown menu is provided by which the user can select a specific year in the range 2013-2017. In this way, the monthly number of tickets issued in that year are displayed for each museum (Figure 6.8).
As regards yearly amount of tickets issued, a dropdown menu is provided to select the month for which data are required. Therefore, the user can select all the months, so that each month of the year is considered to compute the number of tickets issued, or he can choose a specific month to retrieve data of that month over the years (Figures 6.9 and 6.10).

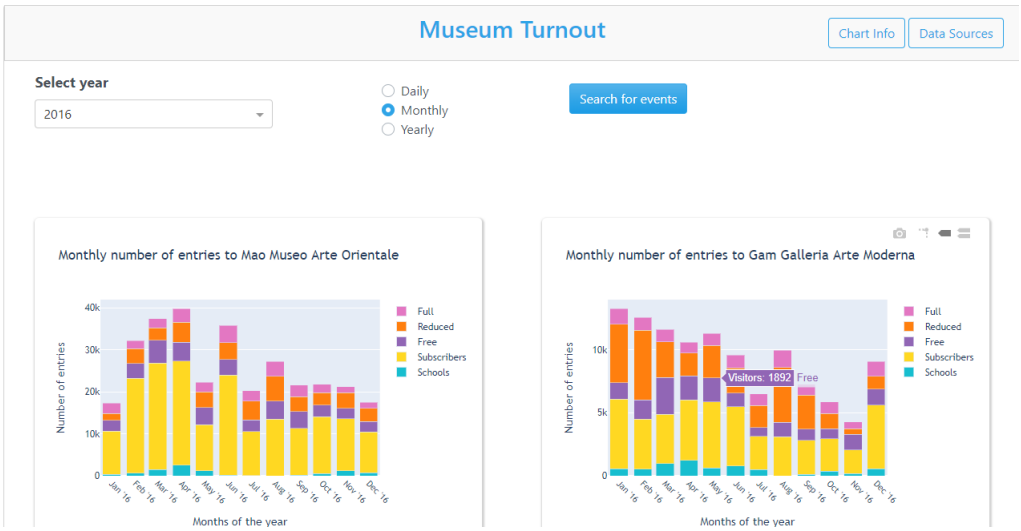


Figure 6.8: TOAD Museum Entry Tickets - Affluenza Pubblico data with monthly granularity

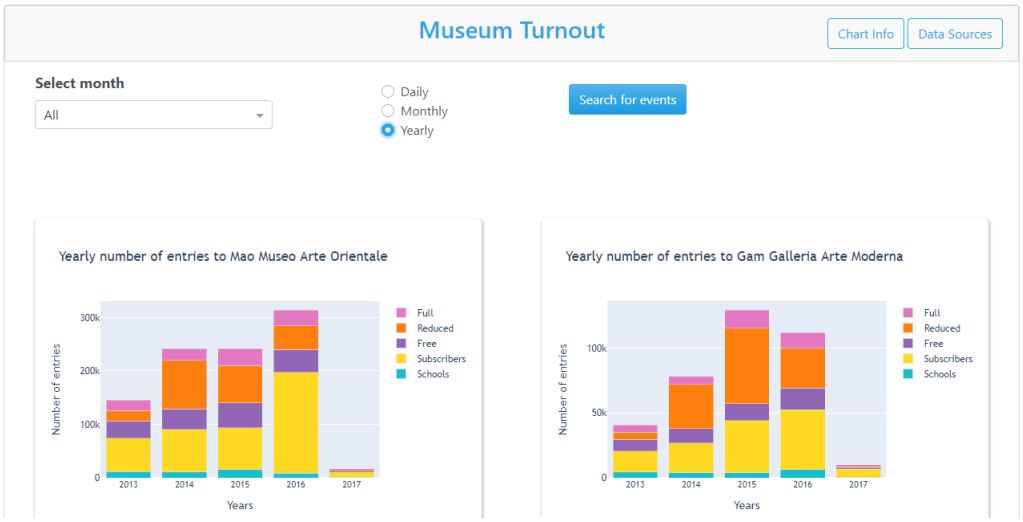


Figure 6.9: TOAD Museum Entry Tickets - Affluenza Pubblico data with yearly granularity for all the months

6.3 People Flow Counting

People Flow Counting is the section where the graphs that represent open Rock Project data are displayed. The section is divided into three subsections (Figure 6.11) which correspond to three different levels of information extracted from data of people flow sensors.
The three subsections are the following:

1. Counting of Visitors
2. Daily Visitor Movements
3. Events

Counting of Visitors subsection includes the graphs about Hourly Footfall, Daily Footfall and Daily Duration Distribution datasets. The graphical representations

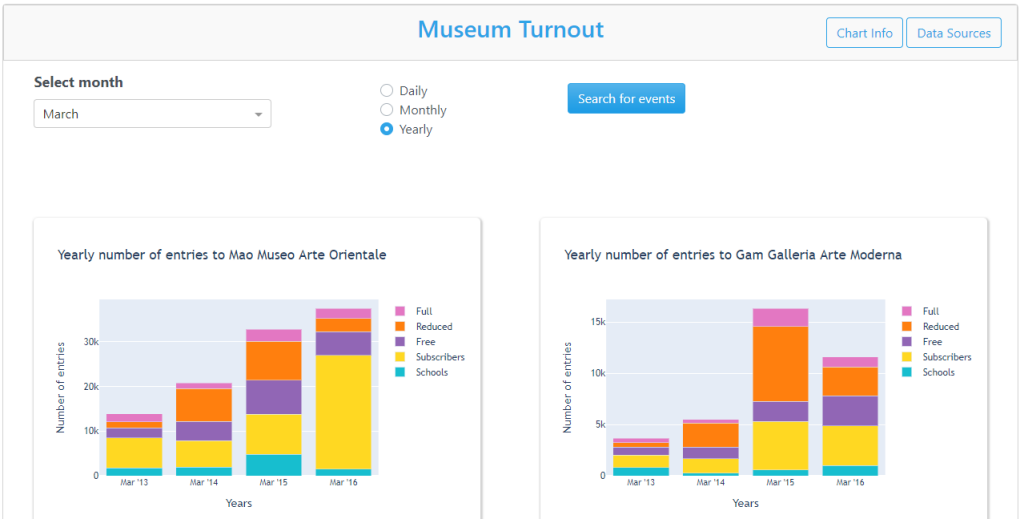


Figure 6.10: TOAD Museums Entry Tickets - Affluenza Pubblico data with yearly granularity for a specific month

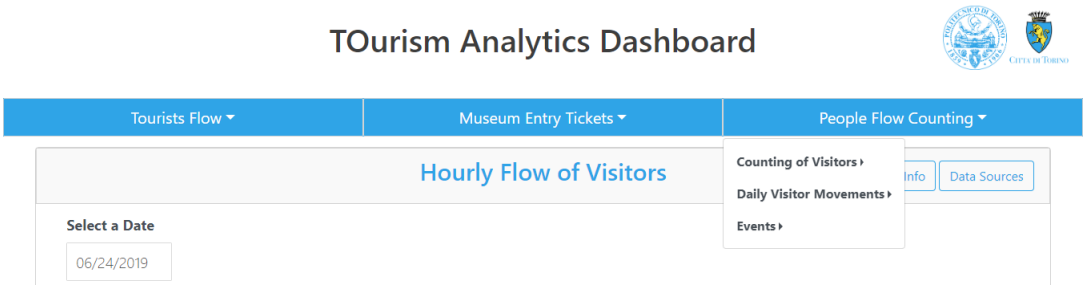


Figure 6.11: TOAD People Flow Counting - navigation items

of the three datasets are shown on three different pages which the user can access through navigation items inside the menu (Figure 6.12).

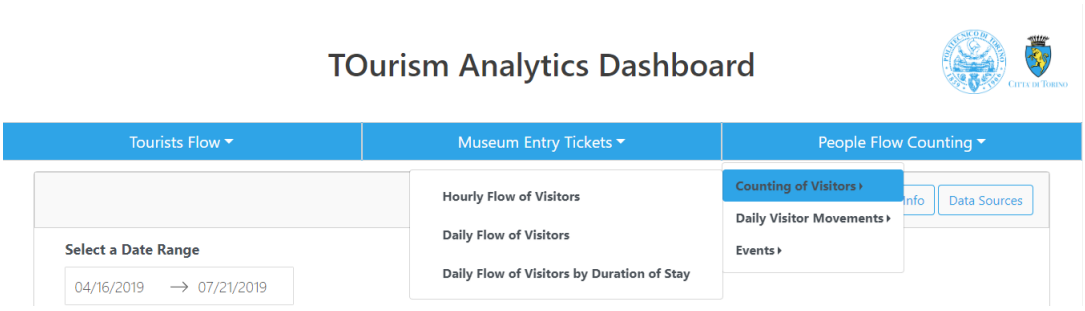


Figure 6.12: TOAD People Flow Counting - Counting of Visitors - navigation items

In order to allow the user to filter data and perform data analysis, some graphical components are needed.
The hourly data coming from the Hourly Footfall dataset are retrieved by selecting a calendar date inside a date picker. Given a date, the hourly amount of visitors per each of the regions are displayed in the grouped bar chart (Figure 6.13).

The graphs relating to the daily number of visitors and the percentage of visitors in

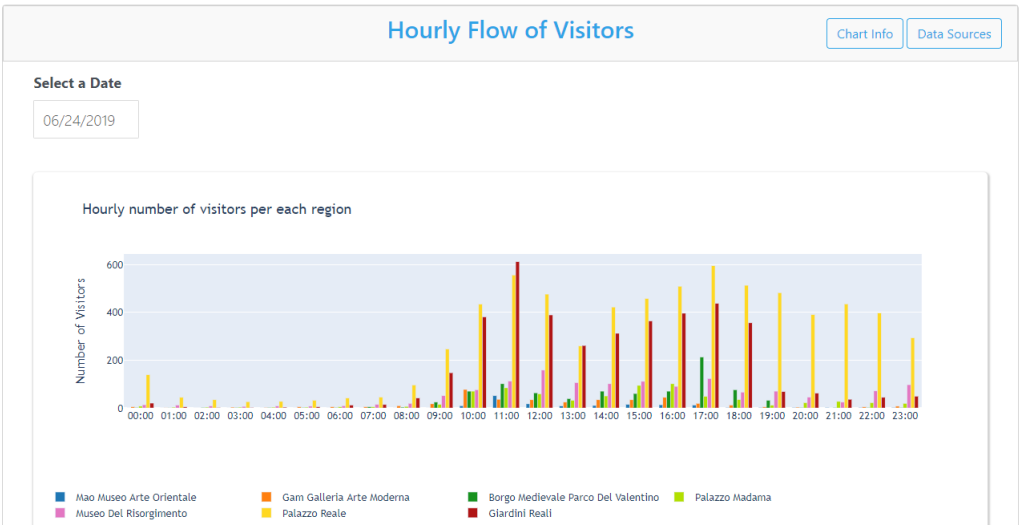


Figure 6.13: TOAD People Flow Counting - Hourly Footfall graph

working days and non-working days include data coming from Daily Footfall dataset which the user can analyse by setting a time interval in a date picker range. The Figure shows a preview of the page in which the two graphs are visualized (Figure 6.14).

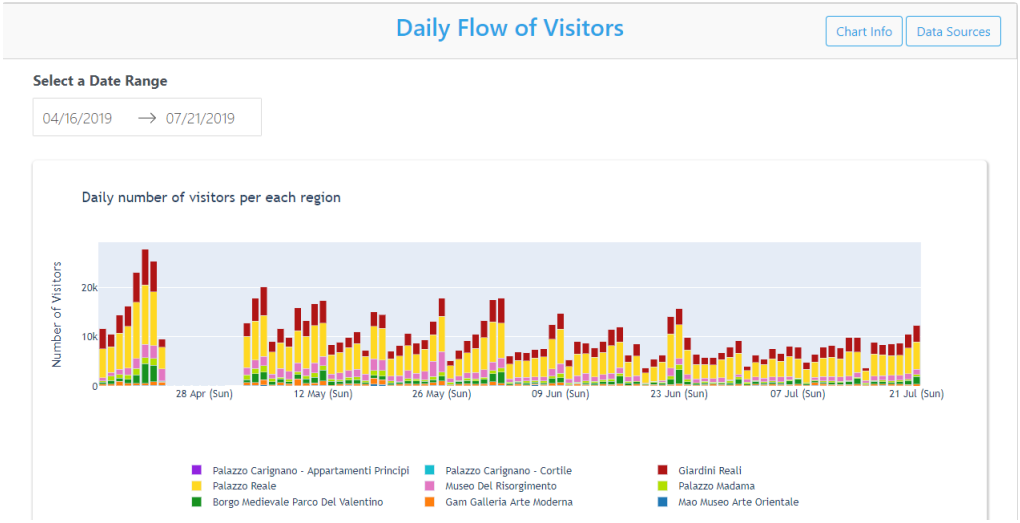


Figure 6.14: TOAD People Flow Counting - Daily Footfall graphs

The graph about the daily number of visitors per duration of stay explores the Daily Duration Distribution dataset and is displayed in a page under the Counting of Visitors subsection. As mentioned above, the graph filters data based on a given a date range and a duration of stay selected by the user. These two input values can be set by means of two components inside the dashboard: a date picker range for selecting a start date and end date and a dropdown menu to choose the duration of stay of visitors (Figures 6.15).

In Visitor Movements subsection there are the two graphical representations of Daily Mobility dataset: a bar chart and a map view which show for each combination of regions the daily number of visitors who move from a start position to an end po-

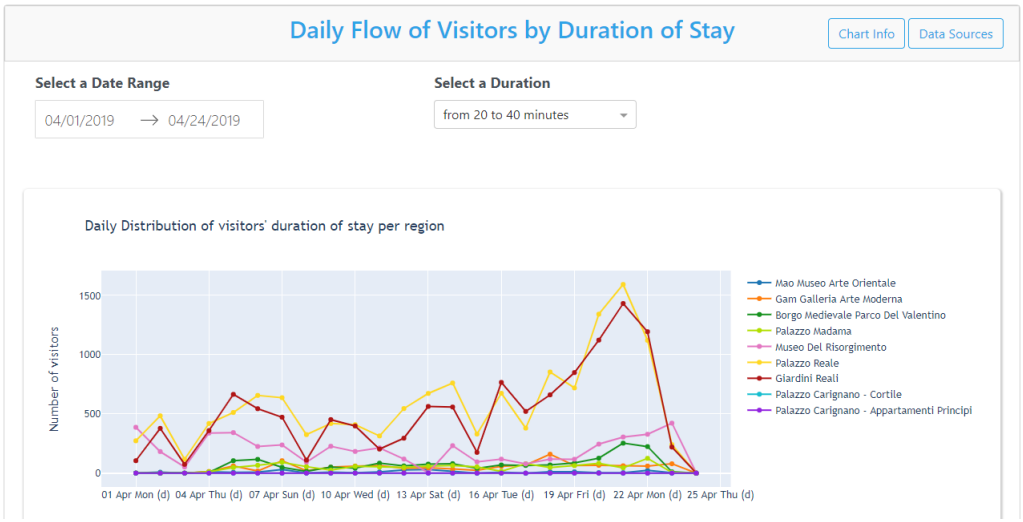


Figure 6.15: TOAD People Flow Counting - Daily Duration Distribution graph

sition. The two charts are displayed into two different pages accessible through the Daily Visitor Movements dropdown menu (Figure 6.16).

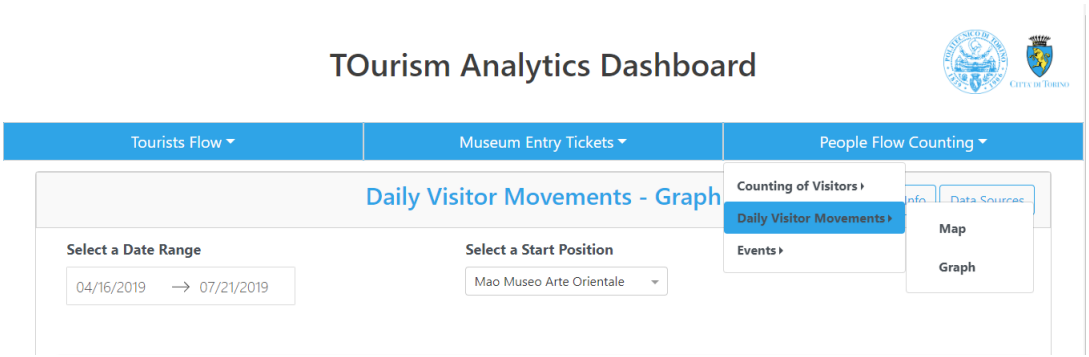


Figure 6.16: TOAD People Flow Counting - Daily Visitor Movements - navigation items

For each graph the user can choose an interval of dates from a date picker range and a start region from a dropdown menu. In this way the user can interact with the charts in order to view visitor movements in the City of Turin. Figures 6.17 and 6.18 show how the two graphical representations are displayed and the graphical components provided inside each page.

The third subsection of People Flow Counting is Events which is devoted to graphs about important events organized in the City of Turin. In particular this section includes the three graphs on the Salone Del Libro 2019 which data are provided from Daily Footfall and Daily returning Visitors datasets. Figure 6.19 shows a preview of the page in which the charts are visualized.

In conclusion, this section gives an overview of the TOAD application layout in which all the graphs previously built are then visualized together on the dashboard. Users can interact with all the graphs, filtering data by means of some graphical tools such as date picker range or dropdown menus that make the app reactive to user inputs. All the sections are well structured to give an idea of the information

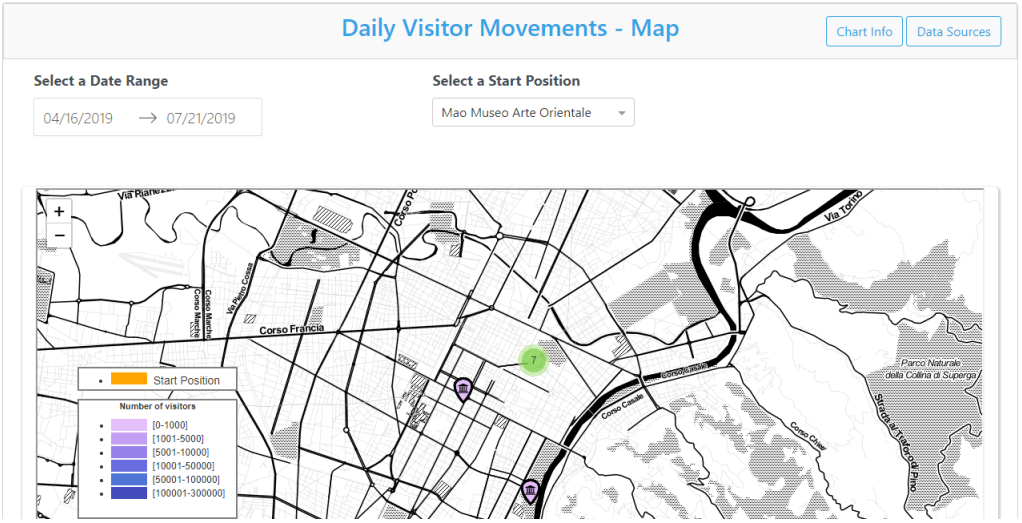


Figure 6.17: TOAD People Flow Counting - Daily Mobility Map Visualization

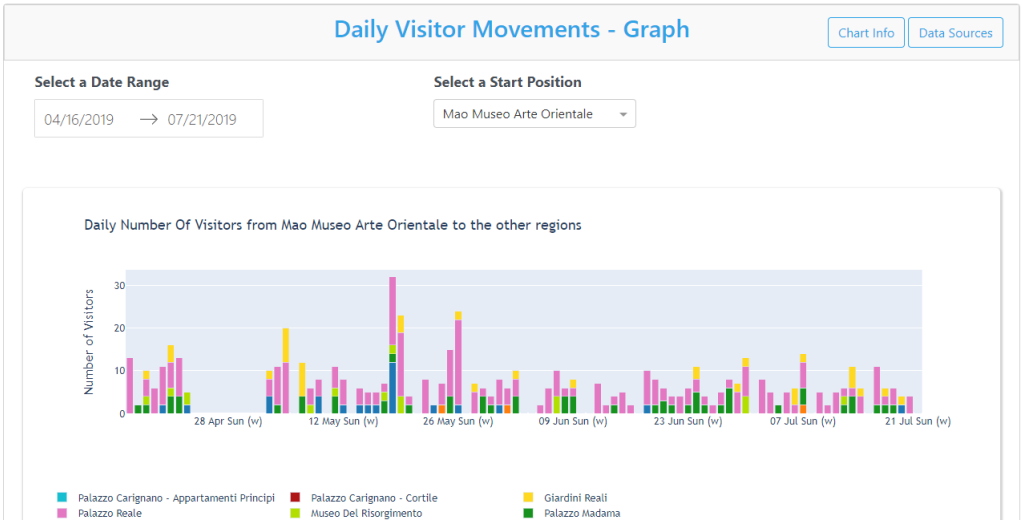


Figure 6.18: TOAD People Flow Counting - Daily Mobility Bar Chart Visualization

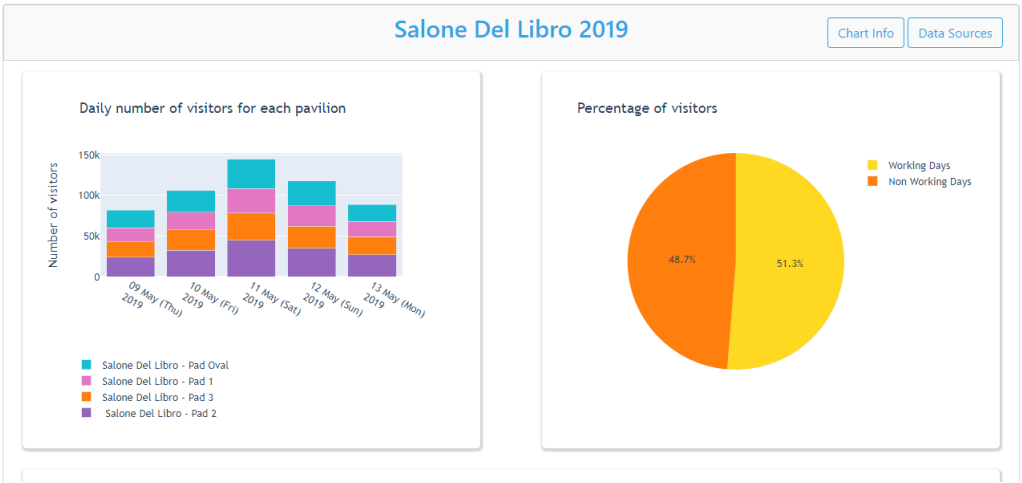


Figure 6.19: TOAD People Flow Counting - Salone Del Libro 2019 Charts

the user can find within the various pages. Therefore, the user can easily navigate the dashboard and get an overview of urban tourism in Turin.

Chapter 7

Evaluation of results

This chapter provides some considerations about the KPIs represented in the graphs. As previously seen, each measure extracted from data analysis process is interpreted by means of specific graphical representations in order to explain the trends over the years about the urban tourism in Turin. The three main topics by which the measured KPIs are grouped inside the TOAD application are explored to perform trend analysis. Therefore, for each section of the dashboard, the graphs are analyzed and compared to extract some useful knowledge about them. Trends are evaluated considering some important events organized over the years by the city of Turin that allow to understand the results obtained which are displayed within the graphs.

7.1 Tourists Flow

The graphs within the Tourists Flow section display the key metrics on the flow of tourists in the City of Turin. In particular, the KPIs extracted from the data analysis are the following:

1. Monthly number of tourists arrivals in the City of Turin
2. Percentage Variation of tourists arrivals against the previous year
3. Monthly number of People who went to Tourism Offices
4. Arrivals of tourists in Turin accomodations
5. Presences of tourists in Turin accomodations
6. Percentage Variation of Arrivals against the previous year
7. Percentage Variation of Presences against the previous year
8. Average stay of tourists in Turin accomodations

The first two KPIs can be analysed together in order to see the monthly trends of the tourists arrivals in the years 2016-2019. Figure 7.1 shows a grouped bar chart in which the monthly number of tourists arrivals is represented for each year in the range 2016-2019. From the graph it is possible to observe a seasonality of the trends: the months with the largest number of arrivals are March, April and May that corresponds to spring months and September, October, November and December which

are the autumn months. In general there is an increase of tourist arrivals over the years.
Looking at the percentage variations graph in Figure 7.2, it is possible to notice significant positive variations for some months of the year 2017 compared to 2016 and for some months of 2019 compared to 2018. In the year 2017, the highest positive variations of the number of arrivals are observed in January, March, June, September and November. Also in 2019 the months of March, September and November present the highest values of percentage variation. There are also negative variations, especially in February 2019 where the number of tourists arrivals is decreased with respect to February 2018. In addition, the values obtained from the first KPI show that the February month presents in general low numbers of tourists arrivals. So we might conclude that this particular month is not so interesting for tourists to visit the City.
In order to give value to these results, some events can be considered important in terms of tourism growth for the City. For instance, in the spring months there are several events organized by the City of Turin. Some of these events are: Festival dell'Oriente and Festival Irlandese in March and Salone Internazionale del Libro in May. Since the amount of tourists arrived in the City increases over the years especially in these months, these events might be an important attraction for tourists.

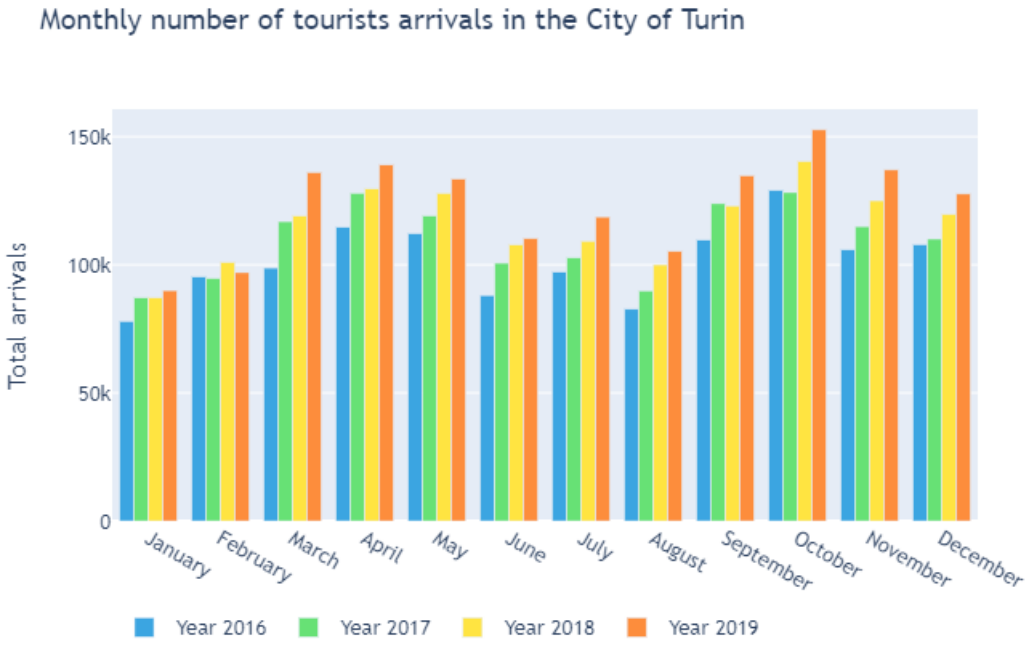


Figure 7.1: Monthly number of tourists arrivals

The third KPI about the monthly number of People who went to Tourism Offices is shown in Figure 7.3. This KPI could allow to implicitly measure the number of tourists that arrive in Turin. For each year in the interval 2014-2018 a line is drawn to represent the number of People. We can see that in general the trends present a seasonality: the months in which there are highest peaks on number of People went

Figure 7.2: Percentage Variation of tourists arrivals against the previous year

to Tourism Offices are March, April and May but also July, August, September, October and December.
It is particularly interesting to observe the peaks on May 2015 and September 2016. In the year 2015 there were two important events: the international reopening of the Egyptian museum in April and the Milan Expo that started in May and lasted until October. These two events were very attractive for tourists from all over the world by which the City of Turin reached an high concentration of visitors. Furthermore, in the month of May the *Salone Del Libro* is the most important event organized by the City, and this could be another meaningful factor for the high value achieved of the number of People that want to visit the City.
As regard September 2016, there is a peak in the number of visitors went to the Tourism Offices. This could be indicator of the success of an important event organized by Turin in that year: the *Salone Del Gusto*. It is a gastronomic event that occurs every two years in September and usually lasts five days. Therefore, if we compare the value of September 2016 with the value of September 2014 in which the event also occurred, we can see that in 2016 the number of People went to Tourism Offices is higher than 2014. Also in Septermber 2018 the Salone Del Gusto event was organized. In this case, if we see the graph, the values in the autumn months of 2018 tends to decrease. Giving a general overview of the results, it is possible to observe especially for values relating to the winter and autumn months that there is a slightly decrease over the years of the People who go to Turism Offices. These values could apparently indicate that the number of tourists arrivals in the City of Turin decreases over the years. But if we look at the values relating to the number of arrivals in 2016 and 2018, we see there is a growth over the years. Thus, a possible interpretation that could be made on the results obtained by the third KPI is that thanks to the diffusion of websites that offer the user all the information

regarding travel bookings, tourist guides, calendar of exhibitions and events, people can organize their trips more independently and consequently Tourist Offices are less frequented by visitors than before.

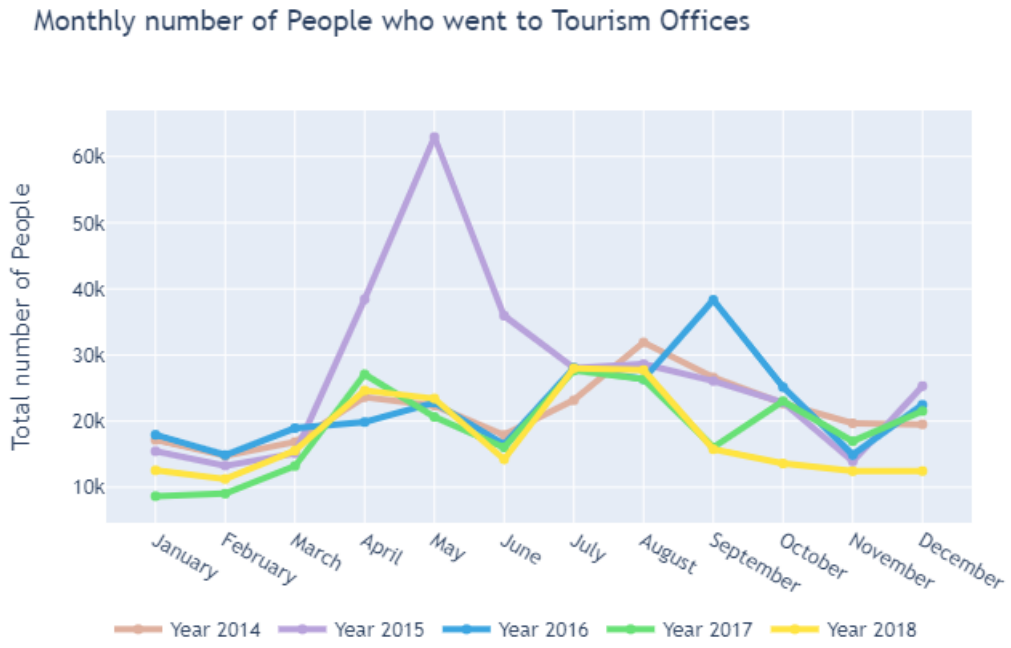


Figure 7.3: Monthly number of People who went to Tourism Offices

The last five key indicators that are part of the Tourists Flow section are the ones which meausure the arrivals and presences in Turin accomodations in the years between 2014 and 2018. As we can see in Figures 7.4 and 7.5, the number of arrivals and presences can be evaluated based on three kind of aggregations: Totals, Italians and Foreigners. Looking at the total amount of tourist arrivals in Turin accomodations, the highest values correspond to the ones in the years 2015 and 2017. All the considerations previously made for the year 2015 can also be observed for the number of tourists in the accomodations since there is a higher value than in other years. This shows that in 2015 there was an important growth for the Turin tourism. Instead, as regards the number of presences, the years 2016 and 2017 record the largest values. It means that in these years the number of nights spent by guests in the accomodations is higher than in the other years.
In general, for both arrivals and presences, the Italian tourists are more than foreign tourists.
Looking at the charts that display the percentage variations for arrivals and presences (Figures 7.6 and 7.7), we can see interesting trends in the year 2015. There are significantly positive variations in the number of foreign tourists with a percentage increase of more than 80% for arrivals and a percentage increase of more than 60% for presences with respect to 2014. These are important results that provide an in-depth view on how 2015 was a crucial year for tourism in the City of Turin. It could therefore be thought that both the reopening of the Egyptian museum and the Milan Expo have attracted the attention of foreign tourists towards the City of Turin.

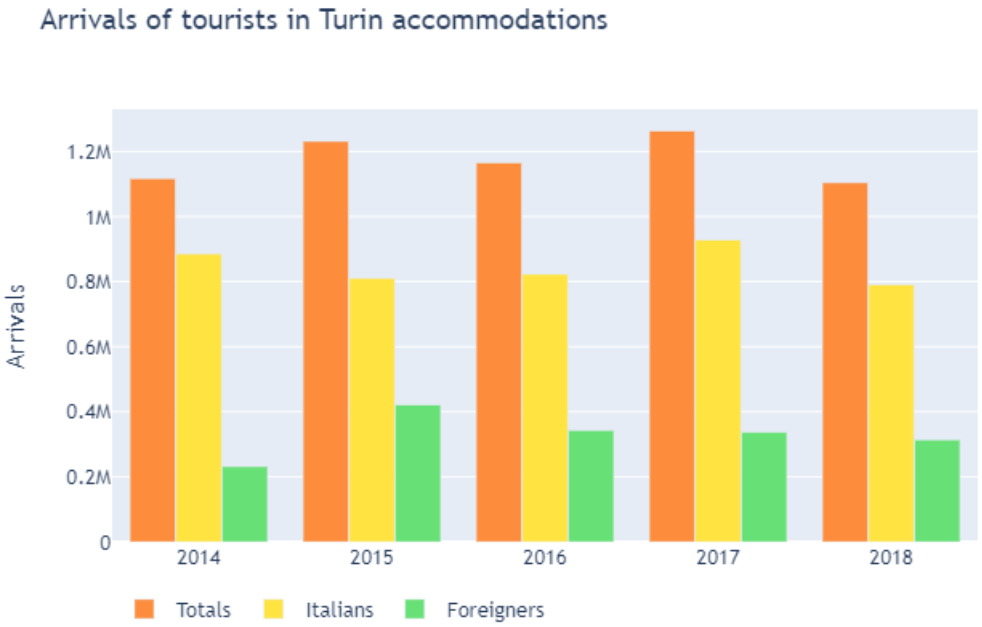


Figure 7.4: Arrivals of tourists in Turin accomodations

Figure 7.5: Presences of tourists in Turin accomodations

In the other years of the interval, it is possible to see negative variations for both foreign arrivals and foreign presences. In particular, the percentage decrease of foreign arrivals in 2016 against the previous year was -18.5%.
The year 2018 presents in general a decrease in the total number of tourist arrivals and presences. But if we look at the first graph that shows the monthly number of

tourists arrivals, we can see that over the months of 2018, the number of arrivals is greater than in 2017. Therefore we can not say that the number of visitors is low in 2018 but it could be possible that most people visited the City only for one day, without staying in a tourist accomodation.

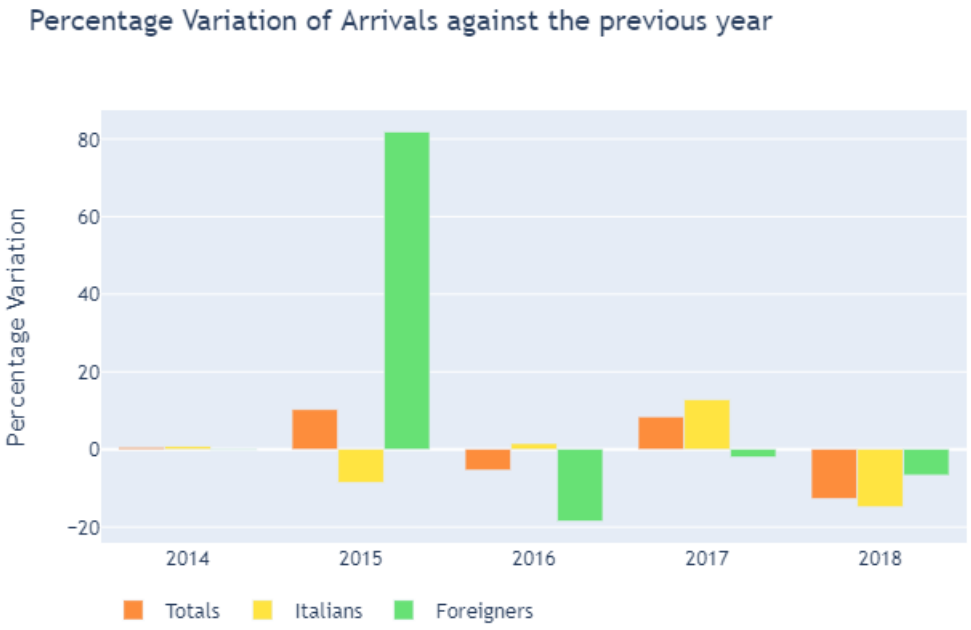


Figure 7.6: Percentage Variation of Arrivals against the previous year

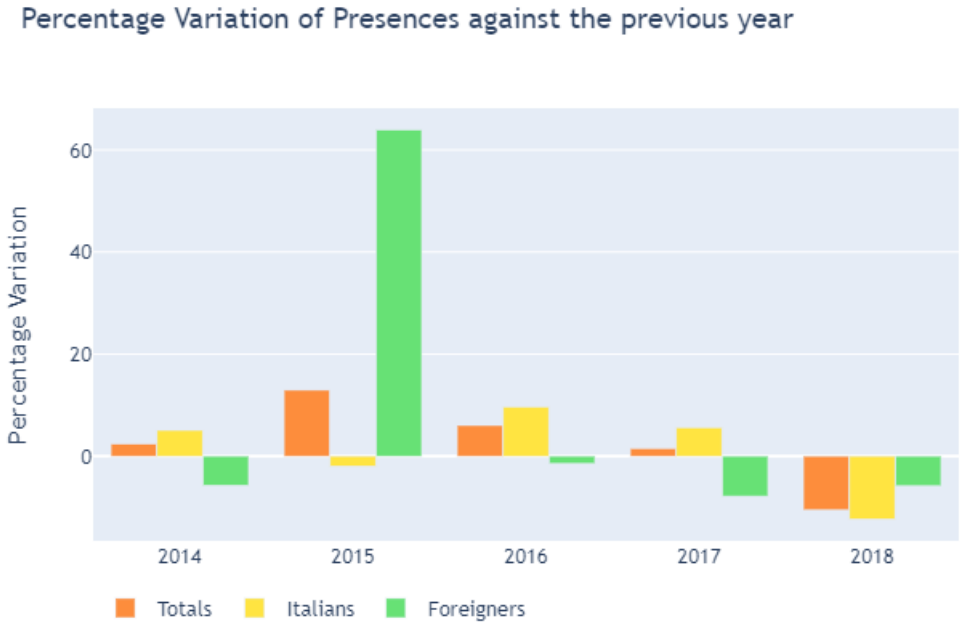


Figure 7.7: Percentage Variation of Presences against the previous year

The graph in Figure 7.8 shows the average stay of tourists in tourist accomodations of Turin. It is possible to observe that, except for 2015, in the other years the average stay of foreign guests of the accomodations is higher than the average stay of Italian guests. This means that during the years 2014-2018 foreign tourists spent more nights in the accomodations and consequently in the city than Italian tourists. In particular the year 2016 registers the greatest value of average stay for foreign visitors.

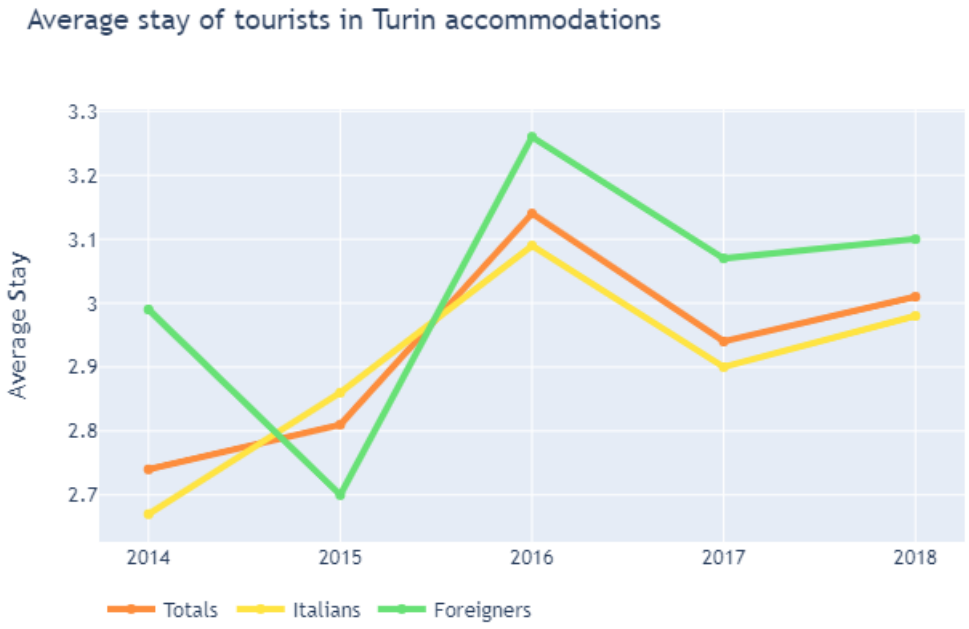


Figure 7.8: Average Stay of tourists in Turin accomodations

7.2 Museum Entry Tickets

The Museum Entry Tickets section includes the charts relating to the entrances to some museums in the City of Turin. In particular the KPIs measured can be classified into the following two groups:

- Counting of Torino+Piemonte cards issued
- Counting of tickets issued in the museums per each ticket type

Counting of Torino+Piemonte cards issued
The KPIs regarding the Torino+Piemonte card are the following:

1. Monthly number of Torino+Piemonte cards issued
2. Top 10 Museums by number of entrances with Torino+Piemonte cards during 2018

The first KPI describes the monthly number of cards issued in the years between 2014 and 2018. As Figure 7.9 shows, trends present a seasonality. The most significant values of the number of cards issued are in the months of March, April, May, August, October and December. For each year in April, the counting of Torino+Piemonte cards is the highest compared to the other months. In particular in April 2017 the number of cards issued was very high (9676). It could be possible that during the Easter's holidays some important museums such as *Egyptian museum* and *Royal museums* were open and tourists took advantage of it to visit them. In general, the spring months represent the period in which there is an high concentration of tourists.
There are three small peaks in the months of August, October and December. For example in December the City of Turin is very suggestive as the *Artist's Lights* are installed and there are various Christmas markets. This could be an attraction for tourists who come to the City to visit museums.
The lowest counts of cards issued are in February. As seen before for the Tourists Flow section, February is not a favorable period for tourism.

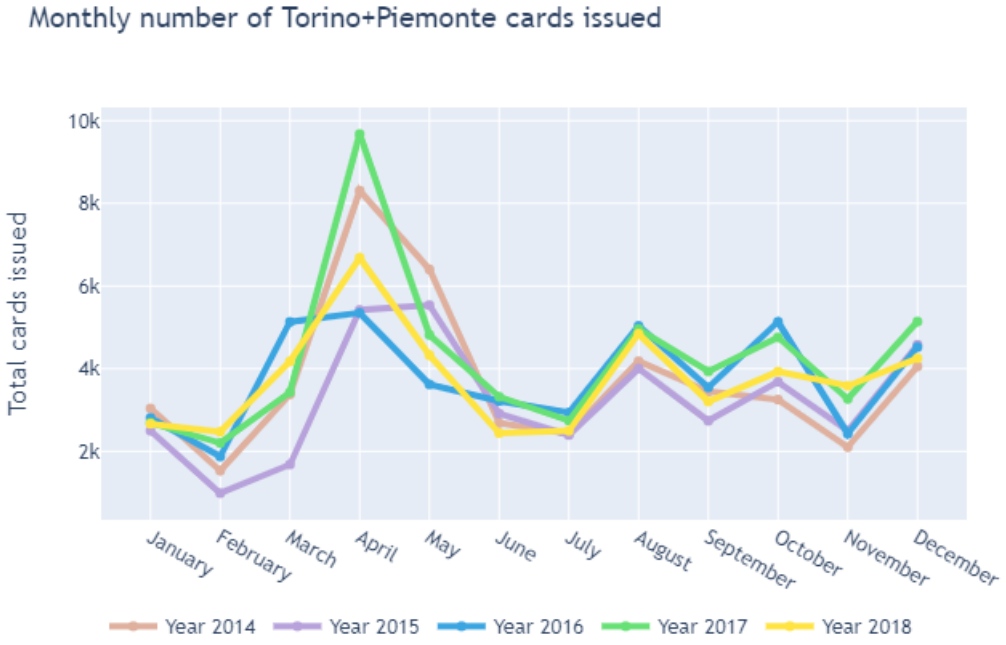


Figure 7.9: Monthly number of Torino+Piemonte cards issued

The second KPI provides the ranking for the year 2018 of the 10 most visited Turin museums based on the number of entrances with Torino+Piemonte card. As we can see in Figure 7.10 in the first four positions we have the *Egyptian museum*, *Royal museums*, *National Cinema museum* and *Royal Venaria Palace* while *Juventus museum* and the *Hunting Lodge of Stupinigi* are the less visited museums in 2018. A possible explaination is that the first four museums are in general the most visited museums of the City of Turin and consequently the most famous in terms of cultural, historical and artistic content. In particular the Royal museums and the Royal Venaria Palace represent the symbol of the Savoy Kingdom of which Turin was the capital. Therefore these historical and cultural buildings are very interesting

to know the history of the City.
Since the international reopening in 2015, the Egyptian museum has been one of the greatest attraction of visitors of Turin because it is the most important in the world after the Egyptian museum in Cairo. So this could be a valid interpretation of its first position in the ranking.
The National Cinema museum is also very popular with tourists in terms of importance in the world and also for the suggestive position within the *Mole Antonelliana* which represent the symbolic monument of Turin.

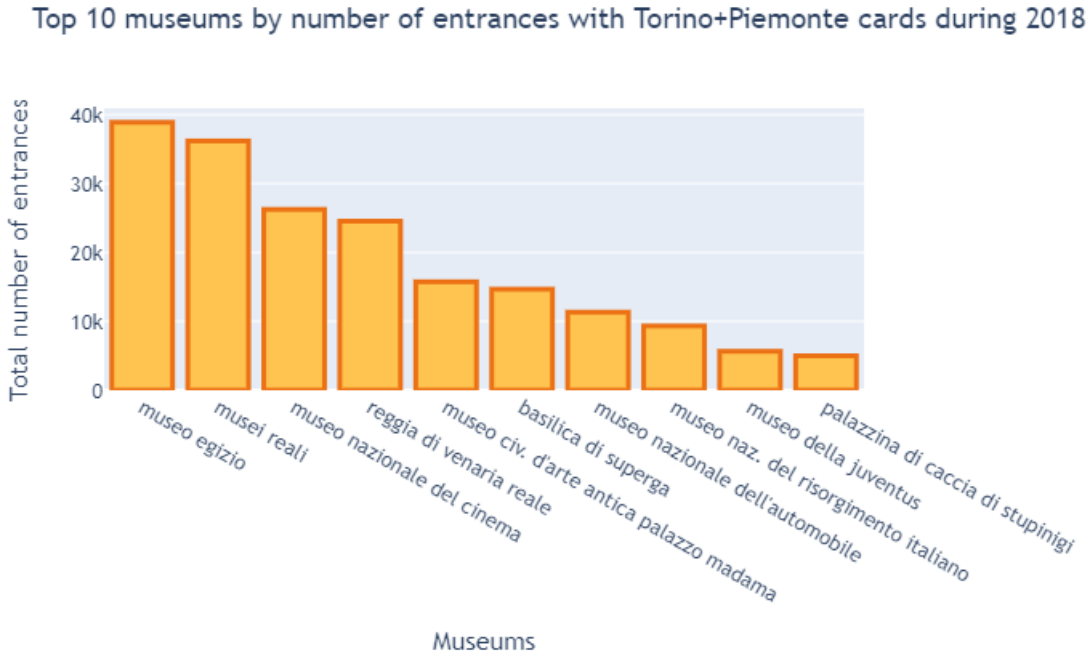


Figure 7.10: Top 10 Museums by number of entrances with Torino+Piemonte cards during 2018

Counting of tickets issued in the museums per each ticket type
The KPIs that measure the counting of tickets issued in the museums per each ticket type are relating to data of Affluenza del Pubblico dataset. For each of the four museums (MAO, GAM, Borgo Medievale and Palazzo Madama) three KPIs are computed which explore three temporal granularity of data.
The KPIs are the following:

1. Daily number of entries
2. Monthly number of entries
3. Yearly number of entries

In order to show how the user should use the charts to obtain useful information, the trends about a specific museum are explored, in this case GAM museum. In particular, for the first KPI, the trends about an important artistic exhibition in the GAM museum are analysed. The exhibition was organized by the GAM museum

from October 2nd 2015 to January 31st 2016 and concerned the art collection of Monet, a famous artist of the Impressionist movement.
The first KPI on the daily number of entries to GAM during the period of the exhibition is displayed in Figure 7.11. As we can see, for each ticket type (full, reduced, free, subscribers, schools) the daily number of entries is counted and is represented with stacked bars. In general, the reduced and season tickets are the most tickets issued. In October and November 2015 there was a low concentration of visitors. A particular day in which only free tickets were issued is November 3rd 2015 (Figure 7.12). Probably in that day the museum decided to offer free entrances in order to promote the exhibition about Monet art work. In December 2015 the highest number of visitors was recorded in the 6th, 7th and the 8th days. In particular in Dicember 6th 2015 the total number of entries was near 1000 in which the most tickets issued were the reduced tickets (Figure 7.13). This high number of visitors could be due to the fact that December 6th 2015 was the first sunday of the month. In Italy each first sunday of the month some museums of each italian region are freely open to the public. This event may have implicitly attracted some visitors to see the exhibition of Monet at GAM.
The last two sundays of January 2016 registered the largest amount of tickets issued. In general, the non working days present the highest distribution of visitors in the museum.
We can notice that in the last week of January 2016 some school groups were organized to visit the GAM museum. In particular the January 28th 2015 was the day with highest number of schools tickets issued during the period of the exhibition (Figure 7.14).

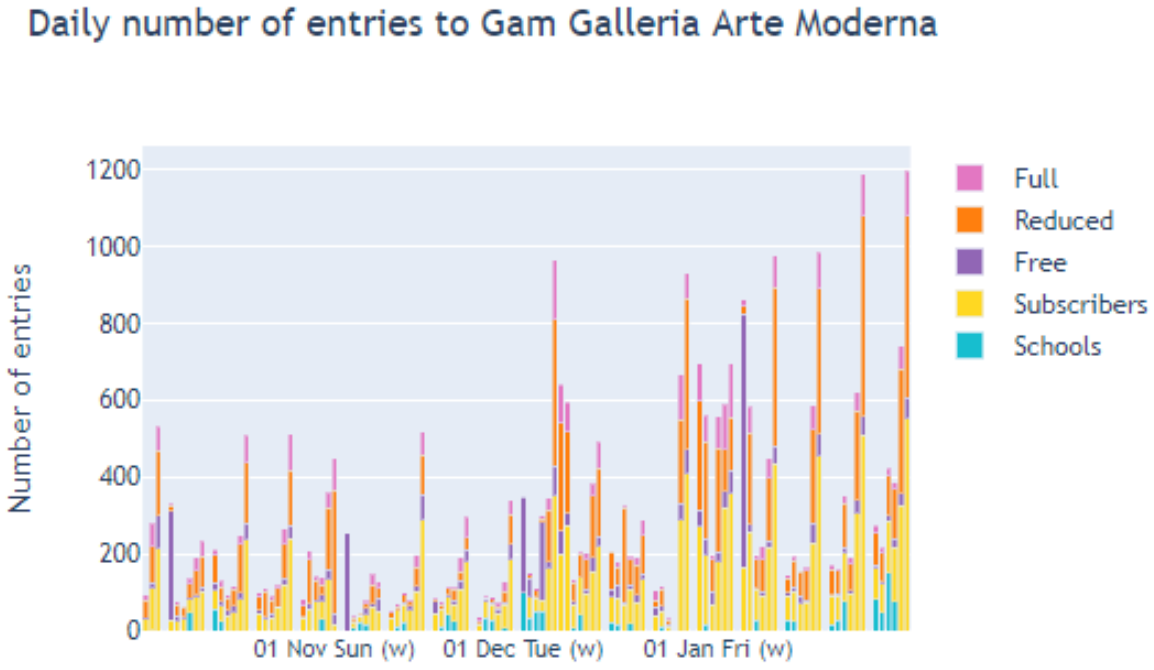


Figure 7.11: Daily number of entries to GAM during the Monet art exhibition 1

The second KPI about the monthly number of entries per each ticket type is described for the GAM museum looking at monthly trends of the years 2015 and 2016. In particular, the two graphs shown in Figures 7.15 and 7.16 are analysed. If we

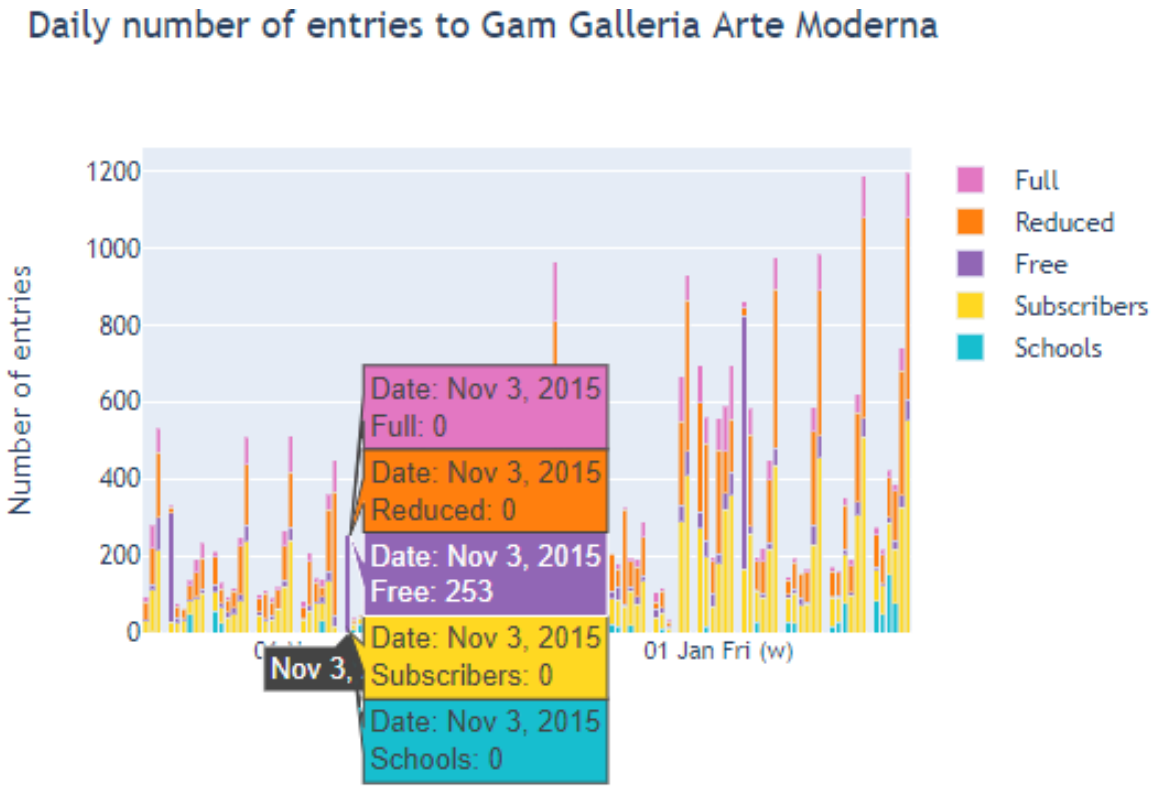


Figure 7.12: Daily number of entries to GAM during the Monet art exhibition 2

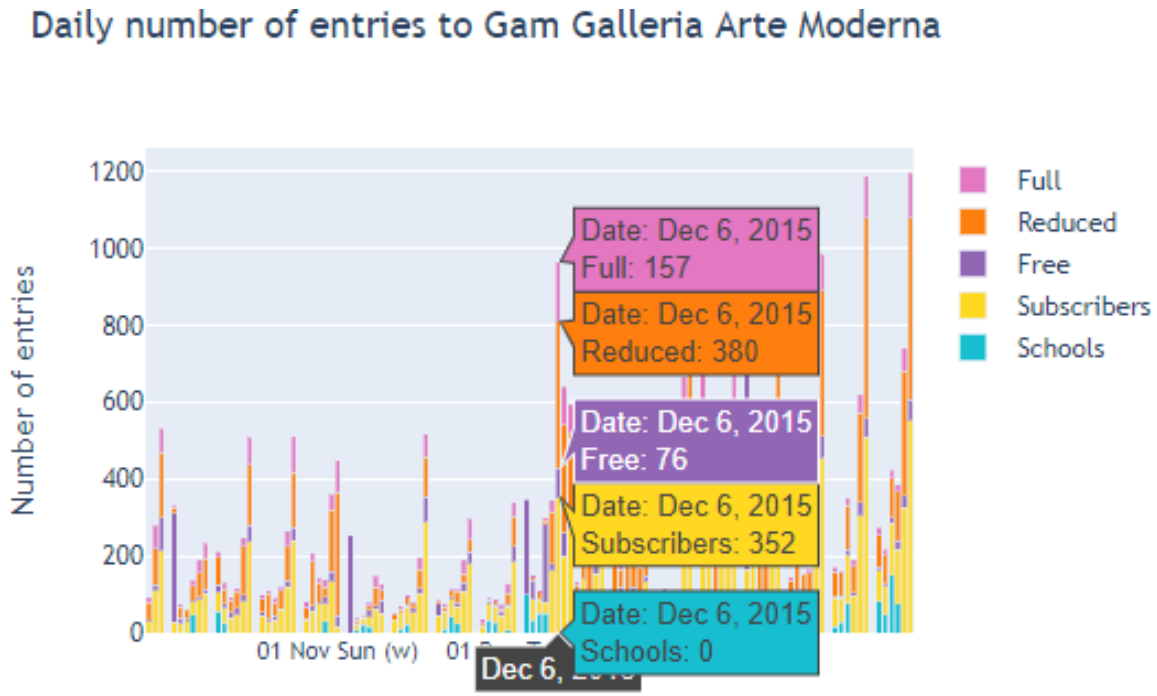


Figure 7.13: Daily number of entries to GAM during the Monet art exhibition 3

look at the two bar charts we can see similar trends over the months. In fact for both the years January, February, March and April are the months in which the GAM museum is most visited. Also in August 2015 and August 2016 the counting of entries is high. The months of July, September, October and November present

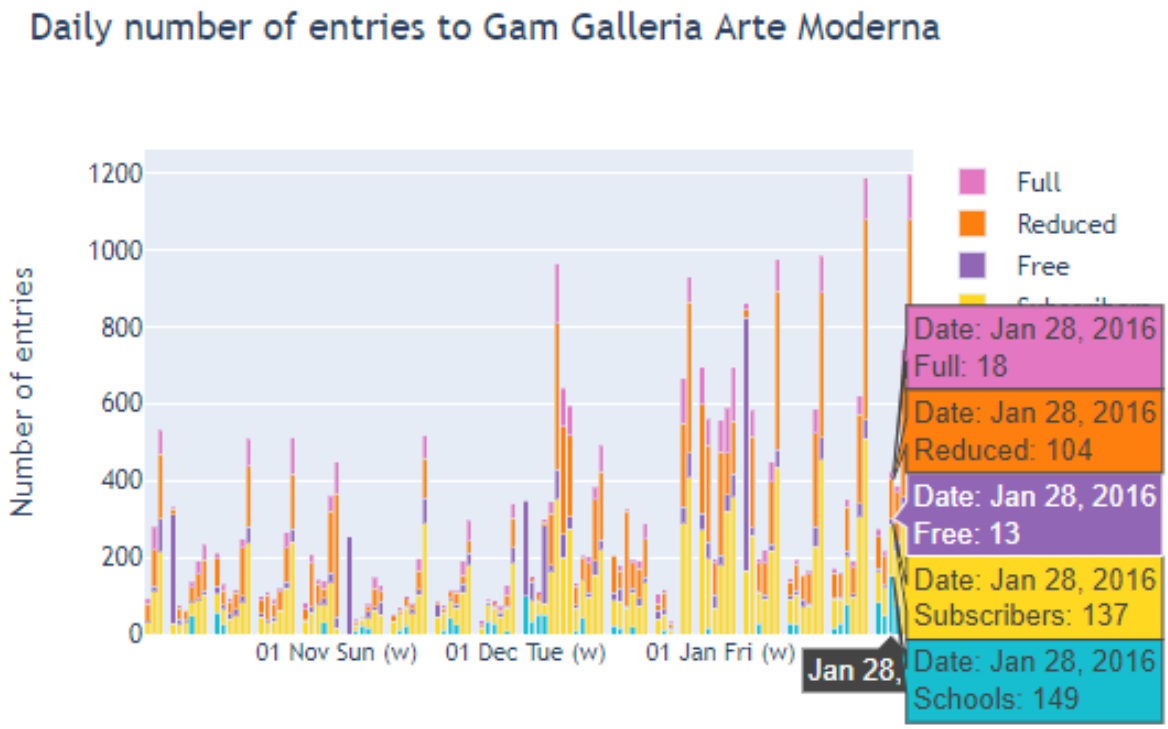


Figure 7.14: Daily number of entries to GAM during the Monet art exhibition 4

low values for both the years. It is interesting to notice that from November to December there are relevant differences in the amount of entries to the GAM museum for both 2015 and 2016.
As regard the month of May, in the year 2015 the total number of entries is very low with respect to May 2016. Looking at the ticket types, in the year 2015 the most ticket issued are of type reduced, instead in year 2016 most of the entrances are with subscriber tickets.
The number of school tickets issued is lower than for other types of tickets. Despite this, we can note that in both 2015 and 2016, the months of March and April are those with a greater number of school tickets issued than in the other months. This is because school trips are often organized in these two months.

The third KPI measures the yearly number of entries per each ticket type. Also in this case we take in example the trends about GAM museum that are displayed in Figure 7.17. Year 2017 presents a very low number of entries because the only available data are relating to the month of January. We can see that the year 2015 presents the highest number of ticket issued, in particular the reduced tickets and the subscriber tickets were the most used. Then in 2016 the trend is decreased but we can see that the number of school tickets is greater than 2015. This can be visualized also in the previous charts with monthly resolution,in which some months of 2016 present a larger amount of entries with school ticket than 2015.

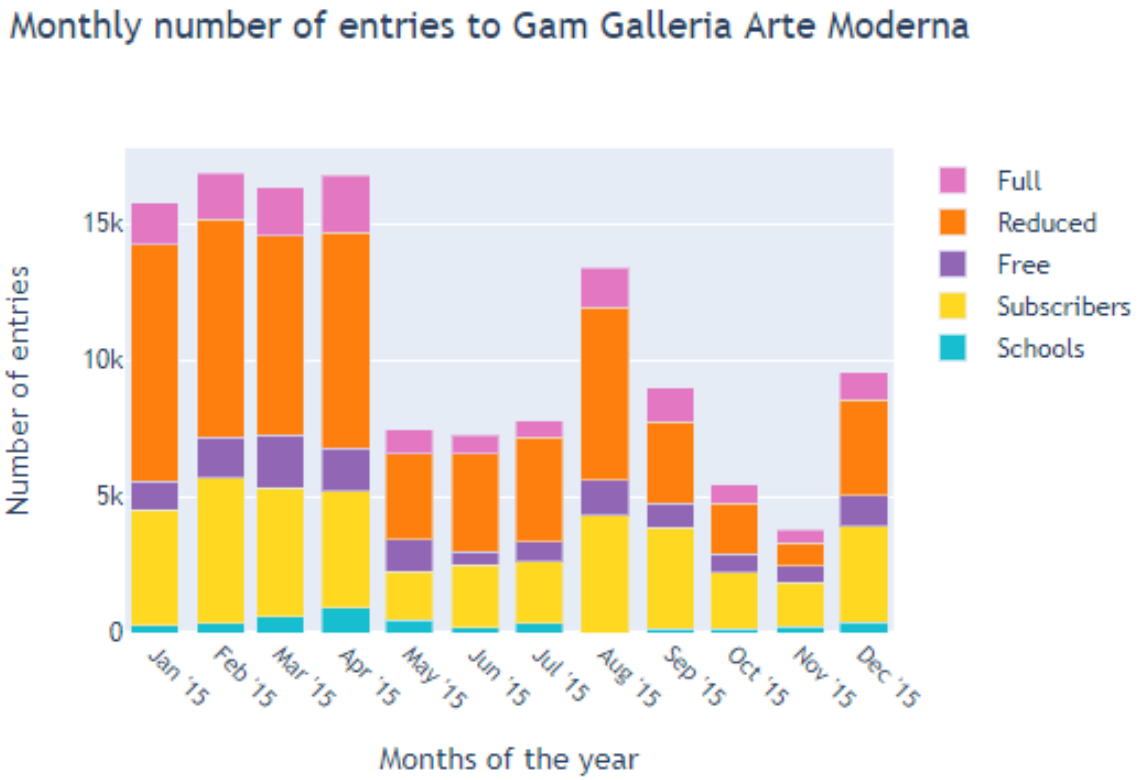


Figure 7.15: Monthly number of entries to GAM during 2015

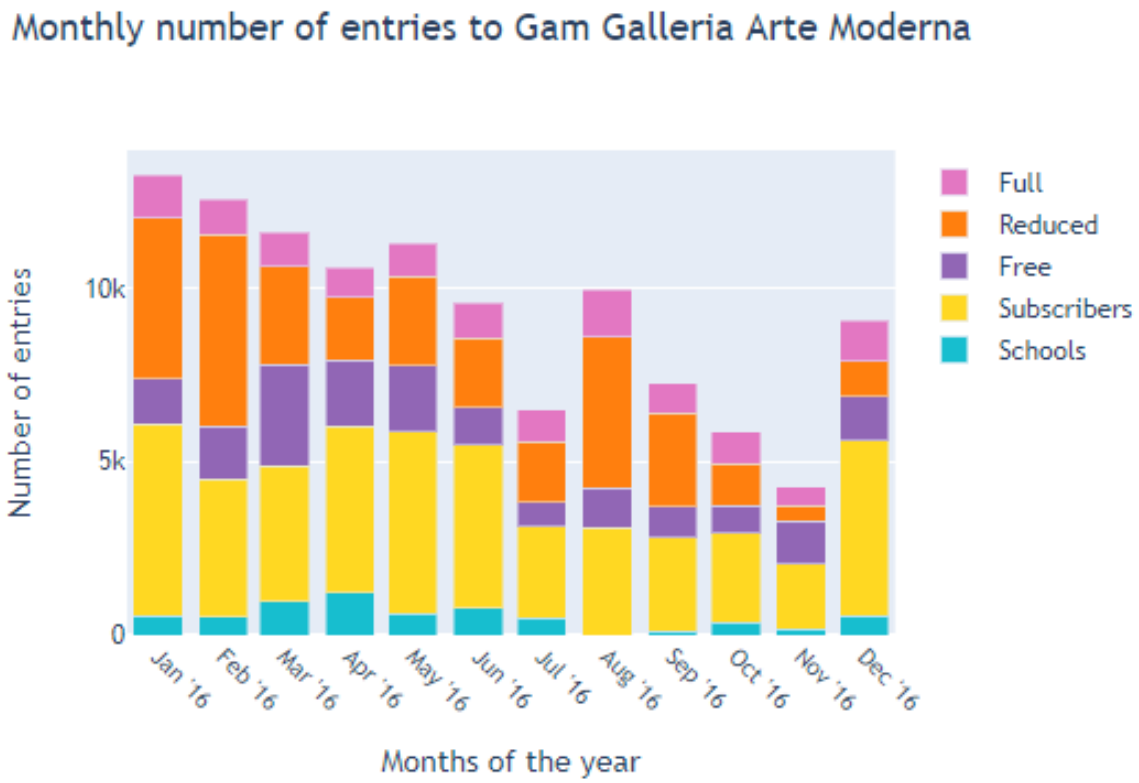


Figure 7.16: Monthly number of entries to GAM during 2016

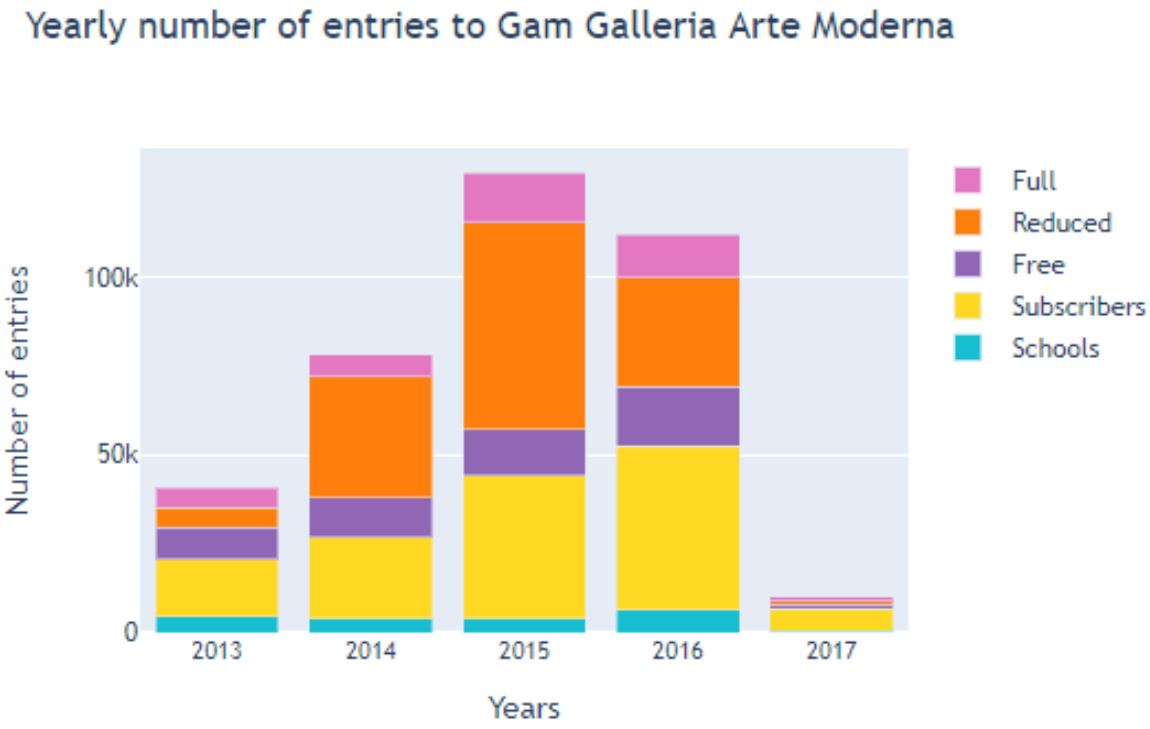


Figure 7.17: Yearly number of entries to GAM

7.3 People Flow Counting

The People Flow Counting section is composed by all the charts generated by the open data coming from people flow sensors of the Rock Project. The KPIs extracted from the data analysis are the following:

1. Hourly number of visitors per each region
2. Daily number of visitors per each region
3. Percentage of visitors per each region in working days and non-working days
4. Daily distribution of visitors' duration of stay per each region
5. Daily movements of visitors from a start region to the other regions
6. Daily number of visitors per each pavilion of Salone Del Libro 2019
7. Percentage of visitors of Salone Del Libro 2019 by working days and non-working days
8. Daily number of new visitors and returning visitors per each pavilion of Salone Del Libro 2019

The hourly number of visitors per each region of the City in which people flow sensors are positioned can be evaluated for a given date that the user can select from the application. An important day for the City of Turin is the *feast of Saint John* that is celebrated each year on June 24th. It is one of the most attractive event of the City for visitors, particularly appreciated for the fireworks and in the last two years

for the flying drone show that closes the celebrations. Figure 7.18 shows the hourly trends about the feast of Saint John for the year 2019. We can see that the regions of Royal Palace and Royal Gardens are the most visited during the day in the time slot from 11.00 to 18.00. This is because these regions are the two main attraction for visitors that go around the City Center. In addition we can consider the fact that in the Royal Museums an important exhibition of works of art by the master Leonardo Da Vinci was in progress in that period. So this could be one of the factors that determine the large amount of visitors to Royal Palace and Royal Gardens.

The artistic exhibition of Leonardo Da Vinci lasted for four months, from April 16th to July 21st 2019. It could be used to evaluate the second, third, fourth and fifth KPIs which measure the number of visitors per each of the selected regions of the City of Turin with a daily resolution. So given the date range of the exhibition, we consider the four KPIs that represent four different ways to count the daily number of visitors.
Figure 7.19 shows that during the period of the exhibition, Royal Palace and Royal Gardens present the largest amount of visitors, in particular in the non-working days. In general in the month of July a low concentration of visitors was registered.

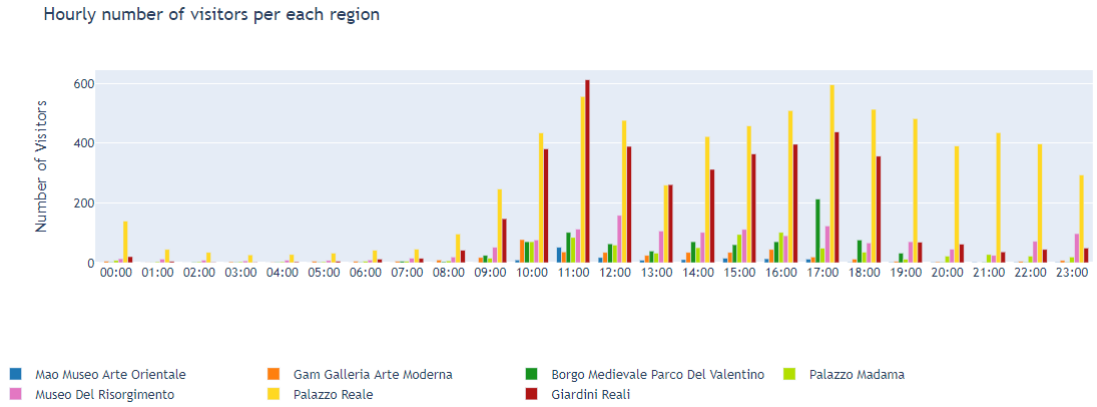


Figure 7.18: Hourly number of visitors per each region

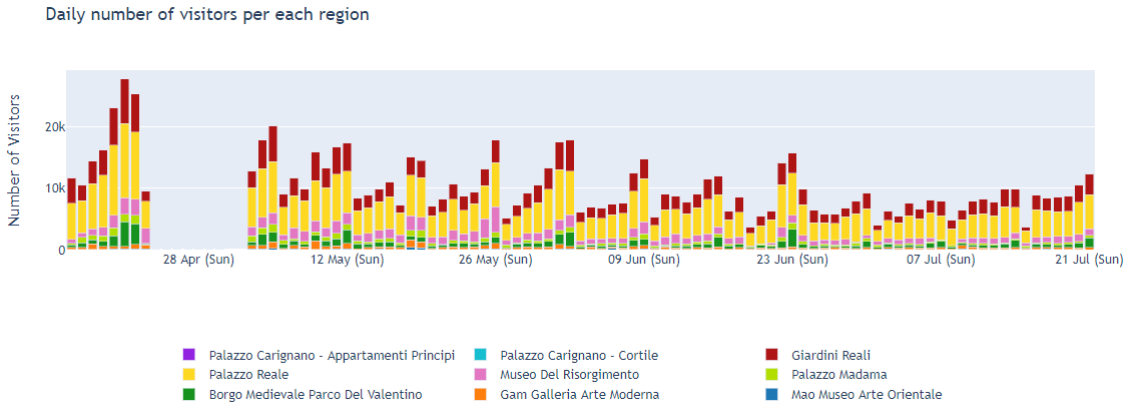


Figure 7.19: Daily number of visitors per each region

The two pie charts shown in Figure 7.20 describe the KPI about the percentage of

visitors distinct per each region in the case of working days and non-working days computed over the date range considered. In both cases we can see that Royal Palace and Royal Gardens represent the highest percentage. So these two places are the most visited during the period of Leonardo Da Vinci exhibition.

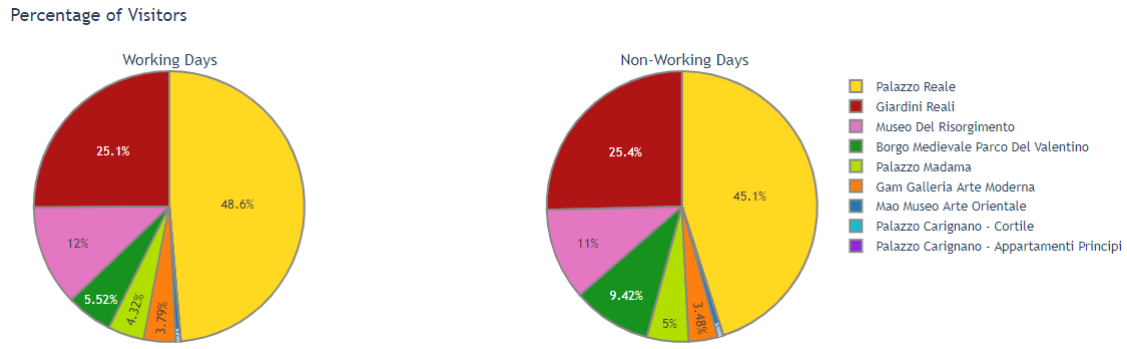


Figure 7.20: Percentage of visitors per each region in working days and non-working days

The fourth KPI on daily distribution of visitors' duration of stay per each region provides the information on the number of visitors in each region given a date range and a duration of stay of people. In the context of the art exhibition of Leonardo Da Vinci, it could be useful to observe the counting of visitors for the regions of Royal Palace and Royal Gardens that implicitly give an estimation of the possible visitors to the exhibition. In particular, it was decided to evaluate the trends over exhibition's date range for two different duration of stay: a duration of 60 to 90 minutes and a duration of 90 to 120 minutes. Figures 7.21 and 7.22 shows the number of visitors for the two duration of stay. If we compare the two plots, we can see that in the first case the number of people is greater than people with a duration of stay between 90 to 120 minutes. This probably because in the second case the duration considered is a bit too long and the number of visitors detected is lower than number of visitors of the first case. The amount of visitors to Royal Palace is not so high compared to the Royal Gardens while interesting high values can be viewed for the Risorgimento museum. This museum is one of the main tourist and cultural attractions of the City of Turin as it represents the history of the Italian Risorgimento and the Unification of Italy. Given that in 2019 the museum was renovated on the occasion of the celebration of the 150th anniversary of the Unification of Italy, this could have attracted more interest from visitors to the city.

The fifth KPI measures the daily movements of visitors from a start region to the other regions. Two types of representations are used: a visualization on the map in which the amount of visitors is aggregated based on the specific date range and a start position of people; a bar chart representation in which the daily amount of people who move from a start region towards other regions is shown. Also in this case we evaluate the KPI using the date interval of the artistic exhibition of Leonardo Da Vinci and we choose as start position of visitors the Royal Gardens in order to observe which are the trends for that given position. Figures 7.23 and 7.24 show the two different representation of the KPI.
In the bar chart we can observe that most of the visitors detected in the Royal Gardens have moved to the Royal Palace. In particular, the highest number of visitors

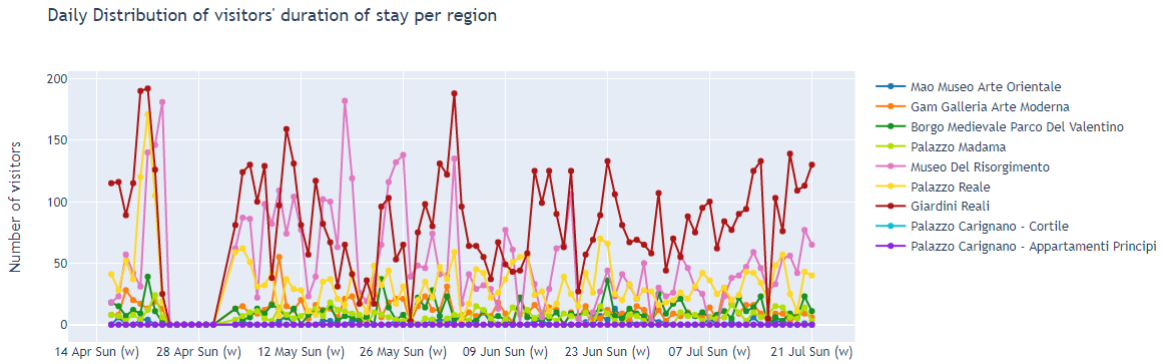


Figure 7.21: Daily distribution of visitors per each region with a duration of stay between 60 and 90 minutes

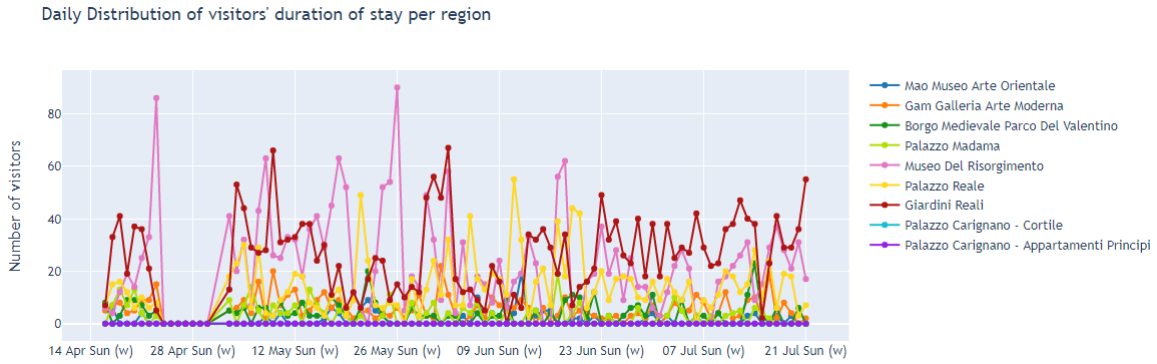


Figure 7.22: Daily distribution of visitors per each region with a duration of stay between 90 and 120 minutes

is registered in the months of April, May and June, especially on non-working days. Even if the Palazzo Madama region is near Royal Gardens, the amount of visitors is very low. In addition we can observe that some visitors of Royal Gardens have moved to Risorgimento museum, especially in the second-half of April and during the month of May.
The map shows the total amount of people that during the period of the Leonard Da Vinci exhibition have moved from Royal Gardens to Royal Palace. With this combination of start region and end region, the Royal Palace corresponds to the end region with the highest value of visitors (84990 visitors). So we can assume that a large percentage of people who went to the Royal Palace also visited Leonardo Da Vinci's art collection.

The last three KPIs concern on the counting of visitors per each pavilion of the Salone Del Libro 2019. Figure 7.25 shows that the highest number of visitors was recorded on May 11th that was a non-working day (saturday) in which the most visited pavilions of the book Fair were pavilion 2 and pavilion Oval. The pavilion 2 was particularly devoted to a reading area for children called BookStock Village in order to promote the reading practice in the young people. The pavilion Oval was composed to stands of some popular publishing houses such as *Mondadori*, *Feltrinelli* and *Einaudi Editori*. Therefore we can assume that all these stands were very interesting for the visitors of the book Fair.

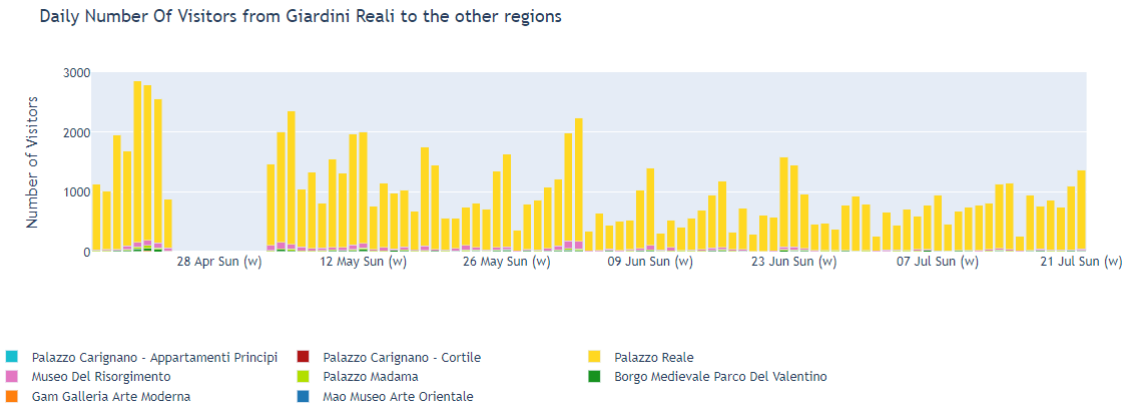


Figure 7.23: Bar chart of daily movements of visitors from a start region to the other regions

Figure 7.24: Map of daily movements of visitors from a start region to the other regions

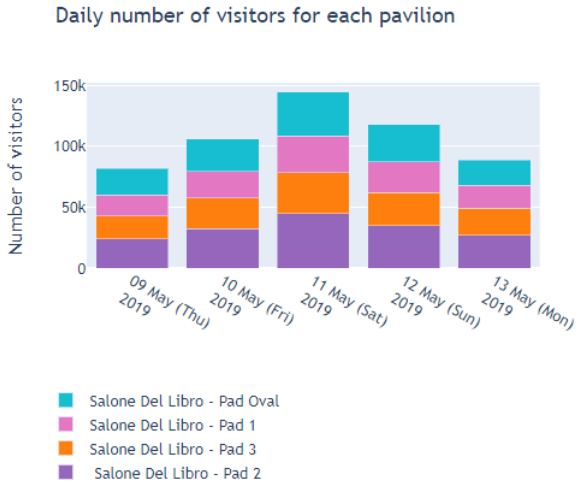


Figure 7.25: Daily number of visitors per each pavilion of Salone Del Libro 2019

Figure 7.26 shows the percentage of visitors in working days and non-working days is not so different but we need to note that working days are more than non-working days. Although the percentage of visitors on working days is slightly higher than the number of visitors on Saturday and Sunday, proportionally to the five days of

the fair we can say that people prefer to visit the Book Fair on non-working days.

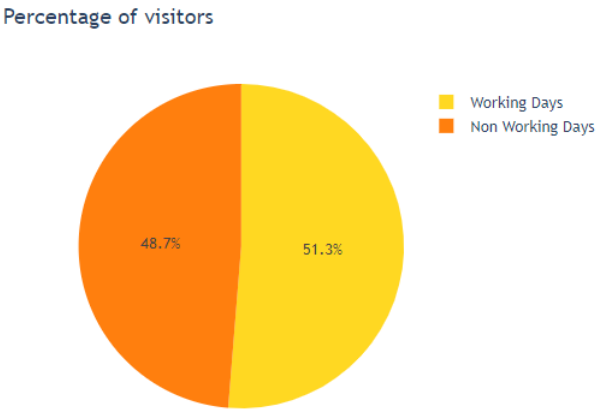


Figure 7.26: Percentage of visitors of Salone Del Libro 2019 by working days and non-working days

Figure 7.27 shows the results of the daily number of new visitors and returning visitors to Salone Del Libro 2019. In the first plot we can see that on Saturday (May 11th 2019) there are high peaks in which we can observe also in this case pavilion 2 and pavilion Oval are the most visited ones. As regards the returning visitors, we can notice that lines which correspond to pavilion 2 and pavilion 3 are the ones with the largest amount of returning visitors.

Figure 7.27: Daily number of new visitors and returning visitors per each pavilion of Salone Del Libro 2019

Chapter 8

Conclusion

In conclusion, this thesis project concerns the reaseach, analysis and interpretation of open data in order to extract meaningful KPIs to display within an informative dashboard. Different datasets about tourism in Turin have been analyzed. Since data are heterogeneous, a data classification is given and three main topics are identified: tourists flow, turnout in museums and tourism urban movements. For each category of data, several KPIs are measured and graphically represented by means of interactive charts. The KPI analysis is used in order to explore the characteristics of data and extract hidden information. These information help to build an overview of the tourism effects on the socio-economic process of the City.
The design and development of a dashboard is the final result which represents a support tool to monitor urban tourism in the City of Turin. Some previous studies about the development of dashboards have been analyzed in order to design a new model. These works have been evaluated in terms of the target users, intended goals and data extraction, processing and visualization. From these previous works, a new informative dashboard was built, the *TOurism Analytics Dashboard* to support stakeholders and tourist managers in the analysis of tourism trends of the City of Turin and improve the effectiveness of the decision-making process. TOAD is a web application that the final users can use and with which they can interact in order to obtain interesting insights about urban tourism which includes the tourist arrivals, flow of visitors to museums and flow of people around the City.

Bibliography

[1] URL: https://www.predictiveanalyticstoday.com/%20open-source-dashboard-software.

[2] URL: https://rockproject.eu/about.

[3] URL: https://static.intrafocus.com/uploads/2014/06/Why-red-amber-green-RAG-1.jpg.

[4] URL: https://static.intrafocus.com/uploads/2014/06/Why-red-amber-green-RAG-2.jpg.

[5] URL: https://static.intrafocus.com/uploads/2014/06/Why-red-amber-green-RAG-3.jpg.

[6] URL: https://www.clicdata.com/wp-content/uploads/2019/06/example-dashboard-028.png.

[7] URL: https://plot.ly/python/bar-charts/.

[8] URL: https://plot.ly/python/line-and-scatter/.

[9] URL: https://plot.ly/python/line-charts/.

[10] URL: https://plot.ly/python/pie-charts/.

[11] URL: https://dash-gallery.plotly.host/dash-salesforce-crm/cases.

[12] URL: https://pythonhow.com/wp-content/uploads/2020/01/Screen-Shot-2020-01-08-at-2.53.09-PM-e1578492236670-1024x525.png.

[13] URL: https://blog.dominodatalab.com/wp-content/uploads/2016/05/mapdrugweekend.png.

[14] URL: http://aperto.comune.torino.it/.

[15] URL: http://www.dati.piemonte.it/?.

[16] URL: https://www.fondazionetorinomusei.it/it/opendata.

[17] URL: https://opendata.rockproject.eu/dataset/torino-dfrc.

[18] URL: http://opendata.yacme.com/.

[19] URL: http://aperto.comune.torino.it/dataset/numero-di-passaggi-negli-uffici-del-turismo/resource/9b095e00-35c8-4ae7-a878-076746e22bcc.

[20] URL: http://aperto.comune.torino.it/dataset/tessere-torino-piemonte-card-emesse/resource/78b5f7cf-1521-4326-bece-e790b03db9eb.

[21] URL: http://aperto.comune.torino.it/dataset/top-10-musei-per-numero-di-accessi-con-torino-piemonte-card-2018/resource/88bf2fa3-3788-4e19-86b4-a30ac53b160b.

[22] URL: http://aperto.comune.torino.it/dataset/arrivi-turisti-2019.

[23] URL: http://www.dati.piemonte.it/catalogodati/dato/1014563-.html.

[24] URL: https://www.fondazionetorinomusei.it/it/opendata.

[25] URL: https://opendata.rockproject.eu/it/dataset/sensor-locations/resource/f64a7f72-2657-43d2-a4e4-5bc2b1bb4488.

[26] URL: https://opendata.rockproject.eu/dataset/torino-dfrc/resource/d05b22bb-8d49-4416-baea-574e10333145.

[27] URL: https://opendata.rockproject.eu/dataset/torino-dfrc/resource/d05b22bb-8d49-4416-baea-574e10333145.

[28] URL: https://opendata.rockproject.eu/dataset/torino-dfrc/resource/d05b22bb-8d49-4416-baea-574e10333145.

[29] URL: https://opendata.rockproject.eu/dataset/torino-dfrc/resource/d05b22bb-8d49-4416-baea-574e10333145.

[30] URL: https://opendata.rockproject.eu/dataset/torino-dfrc/resource/d05b22bb-8d49-4416-baea-574e10333145.

[31] URL: http://opendata.yacme.com/dataset/0b4b7637-5fc7-48cc-9769-c1047c75f794/resource/9cdba2a4-44c5-4d1c-93ed-189ef5e1e0ae/download/torinomusei2014.geojson.

[32] URL: https://reinvently.com/blog/fundamentals-web-application-architecture/.

[33] URL: https://dash-bootstrap-components.opensource.faculty.ai/.

[34] *"15 Latest Pandas Features – What Makes Python Pandas Unique?"* URL: https://data-flair.training/blogs/python-pandas-features/.

[35] *"business analytics (BA)"*. URL: https://searchbusinessanalytics.techtarget.com/definition/business-analytics-BA.

[36] *"Business Analytics vs. Data Analytics: Which is Better for Your Business?"* URL: https://www.talend.com/resources/business-analytics-vs-data-analytics/.

[37] *"Business Analytics: Everything You Need to Know"*. URL: https://www.microstrategy.com/us/resources/introductory-guides/business-analytics-everything-you-need-to-know.

[38] *"City tourism"*. 2019. URL: https://en.wikipedia.org/wiki/City_tourism.

[39] *"Creating interactive crime maps with Folium"*. 2015. URL: https://blog.dominodatalab.com/creating-interactive-crime-maps-with-folium/.

[40] *"Data set"*. URL: https://en.wikipedia.org/wiki/Data_set.

[41] *"Descriptive Analytics"*. URL: https://www.investopedia.com/terms/d/descriptive-analytics.asp.

[42] *"Developing Meaningful KPIs"*. URL: https://www.intrafocus.com/wp-content/uploads/2014/09/Developing-Meaningful-Key-Performance-Indicators-V7.pdf.

[43] *"Folium MarkerClusters and FastMarkerClusters"*. URL: https://medium.com/@bobhaffner/folium-markerclusters-and-fastmarkerclusters-1e03b01cb7b1.

[44] *"GEOJSON"*. URL: https://geojson.org/.

[45] *"GeoJSON"*. URL: https://en.wikipedia.org/wiki/GeoJSON#Geometries.

[46] *"Introducing Dash"*. 2017. URL: https://medium.com/plotly/introducing-dash-5ecf7191b503.

[47] *"Open data"*. URL: https://en.wikipedia.org/wiki/Open_data.

[48] *"Pandas DataFrame Tutorial – A Complete Guide (Don't Miss the Opportunity)"*. URL: https://data-flair.training/blogs/pandas-dataframe/.

[49] *"Percentage Change"*. URL: https://www.mathsisfun.com/numbers/percentage-change.html.

[50] *"Python Pandas Tutorial – Learn Pandas For Data Science in 7 Mins"*. URL: https://data-flair.training/blogs/python-pandas-tutorial/.

[51] *"Sales Key Performance Indicators and Metrics"*. URL: https://www.datapine.com/kpi-examples-and-templates/sales#customer-acquisition-cost.

[52] *"The Data Analysis Process: 5 Steps To Better Decision Making"*. URL: https://www.bigskyassociates.com/blog/bid/372186/The-Data-Analysis-Process-5-Steps-To-Better-Decision-Making.

[53] *"The Four Perspectives of the Balanced Scorecard"*. URL: https://balancedscorecard.org/bsc-basics/articles-videos/the-four-perspectives-of-the-balanced-scorecard/.

[54] *"Torino+Piemonte Card"*. URL: https://www.turismotorino.org/it/il-tuo-viaggio/le-nostre-card/torinopiemonte-card.

[55] *"URBAN TOURISM"*. URL: https://www.unwto.org/urban-tourism.

[56] *"What Is Data Analysis and Why Is It Important?"* URL: https://www.makeuseof.com/tag/what-is-data-analysis/.

[57] *"What is Data Analysis? Types, Process, Methods, Techniques"*. URL: https://www.guru99.com/what-is-data-analysis.html.

[58] *"Why Red Amber Green (RAG)?"* 2013. URL: https://www.intrafocus.com/2013/07/red-amber-green/.

[59] Sonya Hooja. *"Why Python is the Right Programming Language for Data Science"*. 2016. URL: https://datafloq.com/read/why-python-programming-language-data-science/2426.

[60] Christine Lellis. *"What is a Data Dashboard?"* 2019. URL: http://www.perillon.com/blog/what-is-a-data-dashboard.

[61] Soheil Moosavinasab Jeremy Patterson Robert Strouse Majid Rastegar-Mojarad Kelly Regan Philip R. O. Payne Yungui Huang Simon M. Lin. *"'RE:fine drugs': an interactive dashboard to access drug repurposing opportunities"*. 2016. URL: https://academic.oup.com/database/article/doi/10.1093/database/baw083/2630453.

[62] Mark Lutz. *"Learning Python"*. O'Reilly Media, 2009.

[63] Z. Razaq M. Ishaq Bhatti H. M. Awan. *"The key performance indicators (KPIs) and their impact on overall organizational performance"*. 2014. URL: https://www.researchgate.net/publication/266398617_The_key_performance_indicators%20_KPIs_and_their_impact_on_overall_organizational_performance.

[64] Walter Jamieson Sun-Young Park. *"Developing Hawaii Tourism Dashboard"*. 2007. URL: https://repository.usfca.edu/cgi/viewcontent.cgi?article=1016&context=hosp.

[65] Graham Miller Susanne Becken. *"Global Sustainable Tourism Dashboard"*. 2016. URL: http://tourism4sdgs.org/wp-content/uploads/2018/08/DashboardBackgroundReportfinal2.pdf.

[66] John Terra. *"Python for Data Science and Data Analysis"*. 2019. URL: https://www.simplilearn.com/why-python-is-essential-for-data-analysis-article.

[67] Carlo Vercellis. *"Business Intelligence: Data Mining and Optimization for Decision Making"*. John Wiley and Sons Ltd, 2009.

[68] Richelle J. Koopman Karl M. Kochendorfer Joi L. Moore David R. Mehr Douglas S. Wakefield Borchuluun Yadamsuren Jared S. Coberly Robin L. Kruse Bonnie J. Wakefield and Jeffery L. Belden. *"A Diabetes Dashboard and Physician Efficiency and Accuracy in Accessing Data Needed for High-Quality Diabetes Care"*. 2011. URL: http://www.annfammed.org/content/9/5/398.full.pdf+html.

[69] Il-Hyun Jo Yeonjeong Park. *"Development of the Learning Analytics Dashboard to Support Students' Learning Performance"*. 2015. URL: http://www.jucs.org/jucs_21_1/development_of_the_learning/jucs_21_01_0110_0133_park.pdf.