

MGI

Master Degree Program in
Information Management

ANALYSIS OF AIRBNB IN PORTUGAL IN THE LISBON AND PORTO REGIONS

Data Integration and Analysis through Fabric

Ana Margarida Correia Tomaz

Project Work

presented as partial requirement for obtaining the Master's Degree in Information Management

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

ANALYSIS OF AIRBNB IN PORTUGAL IN THE LISBON AND PORTO REGIONS

Data Integration and Analysis through Fabric

by

Ana Margarida Correia Tomaz

Master Project Work presented as partial requirement for obtaining the Master's degree in Information Management, with a specialization in Business Intelligence.

Supervised by

João Bruno Morais de Sousa Jardim, PhD, NOVA Information Management School

July, 2024

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration. I further declare that I have fully acknowledged the Rules of Conduct and Code of Honor from the NOVA Information Management School.

Lisbon, 15-07-2024

ACKNOWLEDGEMENTS

I would like to thank Professor Bruno Jardim, my Masters colleagues and teachers for their invaluable patience and feedback.

This endeavour would not have been possible without my loving parents, Rute, Cátia, Maria and family.

Special thanks to my dearest friends and to my futsal team in Juventude Ouriense.

Finally, I'd like to mention my team colleagues at Celfocus and my department colleagues at Banco Finantia.

ABSTRACT

Tourism is one of Portugal's main economic drivers. The country attracts tourists because of its climate, gastronomy and safety. There are more and more requirements and peculiarities that accommodation must meet to provide a pleasant experience.

Airbnb is a platform that connects travellers with different types of accommodations. Through this application, the user can access different accommodations around the world and make a reservation.

This project aims to process, work and analyse the data provided by the website Inside Airbnb - Get the Data for Portugal, focusing on the Lisbon and Porto areas. Through this information, the expectation is to build an analysis of the influence of this platform in Portugal, discovering patterns, trends and relationships between prices, reviews, number of customers, type of accommodation, region, among other relevant data.

In the first phase, an analysis of the collected data will be carried out in order to define the appropriate structure for processing and analysing the data. Then, using Microsoft Fabric, a pipeline will be created to load the data into a data lake, transform the collected data, establish links between the information, and create new measures to aid in this analysis. This will set the stage for the creation of a report in Power BI from which the desired insights, such as prices, location, accommodation characteristics and the comparison of the Lisbon and Porto regions, could be made.

KEYWORDS

Fabric; Data Lakehouse; Airbnb; Tourism; Portugal; Accommodation platforms

Sustainable Development Goals (SDG):



TABLE OF CONTENTS

1. Introduction.....	1
1.1. Motivation	1
1.1.1. Tourism in Portugal	1
1.1.2. Airbnb	4
1.1.3. Tourism in Lisbon and Porto.....	5
1.2. Objectives and Methodology	6
1.2.1. Research Questions	6
2. LITERATURE REVIEW.....	8
2.1. Related Work.....	8
2.1.1. Touristification and Gentrification of Cities	8
2.1.2. Airbnb impact on housing market.....	9
2.1.3. Sharing economy	10
2.1.4. COVID-19 impact on the housing market	11
2.1.5. Inside Airbnb dataset	12
2.1.6. Literature Review Resume	13
3. Methodology	15
3.1. Project Planning.....	15
3.2. Business Requirements	15
3.3. Data Source Description	16
3.3.1. General Statistics About the Data	18
3.3.2. Data review.....	18
3.4. Dimensional Modelling.....	20
3.4.1. Business Process and Granularity	20
3.4.2. Dimension Tables	20
3.4.3. Fact Tables	21
3.5. ETL Design and Development.....	21
3.5.1. Microsoft Fabric	21
3.5.2. ETL Process Development	21
3.5.2.1. Pipeline.....	22
3.6. Data Lakehouse	24
3.7. Dashboard in Power BI	26
4. Results and discussion	28
4.1. Host's Information's	30

4.2. Airbnb's Information	32
4.3. Localization information.....	33
4.4. Prices information	34
4.5. Answer business questions	35
5. Conclusions, Limitations and future works	38
Appendix A	46

LIST OF FIGURES

Figure 1 - Companies: total and by industry (PORDATA - Empresas, 2023)	2
Figure 2 - Tourist accommodation establishments: total and by type of establishment(PORDATA - Alojamentos turísticos, 2023)	2
Figure 3 - Travel and tourism balance in percent of GDP (PORDATA - Balança de viagens e turismo em % do PIB, 2023)	3
Figure 4 - Tourist accommodation establishments: total and by type of establishment (PORDATA - Alojamentos turísticos, 2023)	4
Figure 5 - Passenger traffic at airports: Lisbon, Porto and Faro (PORDATA - Tráfego de passageiros nos aeroportos, 2023)	5
Figure 6 - Tourist accommodation: total and by type of accommodation (PORDATA, Alojamentos turísticos, 2023).....	6
Figure 7 - Economic and Demographic changes and the Airbnb locations (Aritenang & Iskandar, 2023).....	9
Figure 8 - Density and overnight stays in short-term Rental Accommodations (Batalha et al., 2022).....	12
Figure 9 - Example of reviews from different listings (Alsudais, 2021)	12
Figure 10 - Example of licenses	19
Figure 11 – Example of Host’s with Pedro as host name	20
Figure 12 – ETL Process Development	22
Figure 13 - Airbnb Pipeline	22
Figure 14 - dimDetail Data Flow	23
Figure 15 - Data Lake Airbnb Analysis	25
Figure 16 - Portugal - Airbnb Data.....	29
Figure 17 - Portugal – Host’s	30
Figure 18 - Portugal - Airbnb's.....	32
Figure 19 - Portugal - Localization	33
Figure 20 - Lisbon – Historic Centre by Google Maps vs Number of Airbnb’s.....	33
Figure 21 - Porto – Historic Centre by Google Maps vs Number of Airbnb’s	34
Figure 22 - Portugal – Prices.....	35

LIST OF TABLES

Table 1 - Research Questions and Objectives	7
Table 2 - Literature Review resume	13
Table 3 - Business Problem into Business Needs	15
Table 4 - Inside Airbnb Lisbon, Portugal (Get the Data, 2023)	16
Table 5 - Inside Airbnb Porto, Portugal (Get the Data, 2023)	17
Table 6 - Statistics about the data dimensions	18
Table 7 - Statistics about the data fact table	18
Table 8 – Data Lakehouse tables.....	25
Table 9 – Pipeline execution process	28
Table 10 - Answer business questions	35

LIST OF ABBREVIATIONS AND ACRONYMS

GDP	Gross Domestic Product
PIB	Produto Interno Bruto
ETL	Export Transform Load
Fabric	Microsoft Fabric

1. INTRODUCTION

1.1. MOTIVATION

Tourism plays a very important role in the Portuguese economy and contributes to the development of the regions. This is due to its attractiveness as a safe country with a mild climate, many beaches, and rich landscapes (Secretaria Geral da Economia, 2023).

In recent years, we have seen an increasingly unstable world, with a pandemic and two wars, in addition to constant climate change. The COVID 19 pandemic forced changes in social life, work, and many adaptations in terms of tourism. Tourists had to be reassured, with a strong emphasis on hygiene. In 2022 and 2023, with the wars in Ukraine and in Israel, the economy has suffered great repercussions, causing a significant increase in the cost of living for families, difficulties in housing with increased interest rates and inflation (Silva, P. R. da., 2021).

These economic and social changes can encourage the creation of new housing, as families with more than one home may now be forced to find another source of income to maintain the same lifestyle or avoid greater hardship.

In these situations, Airbnb makes it much easier to list a home on a platform used nationally and internationally, allowing hosts to reach more people. As a customer, people can have accommodations with a wide range of characteristics and sometimes at more attractive prices.

Through this project, it is expected to identify patterns in accommodation, such as price relationships with location, number of rooms or duration of accommodation on this platform, and to understand whether significant changes have occurred since the COVID 19 pandemic and the outbreak of war in Ukraine and Israel.

The choice of Lisbon and Porto in Portugal is related to the fact that they are the two largest cities in the country, both have airports with many daily flights, and it will be interesting to see if we can identify very large discrepancies between these two tourist cities.

1.1.1. Tourism in Portugal

Portugal is known for its strong tourist attraction, highlighted in studies for the strong contribution of the tourism sector to GDP, Gross Domestic Product (World Travel & Tourism Council (WTTC), 2018).

The country has to offer beach regions and landscapes throughout its coast and interior, rural tourism regions, cities with Museums, Castles, and other points of interest, it has an attractive climate, security and is also known for its traditional food. In terms of security, according to information shared by the Turismo de Portugal website, the country is known for being the sixth safest country in the world; (Pandit, 2022).

Companies related to the tourism sector, such as accommodation, restaurants and similar, have shown growth in recent years, more pronounced between 2013 and 2019, as we can see in the Figure 1 taken from PORDATA (PORDATA - Empresas, 2023).

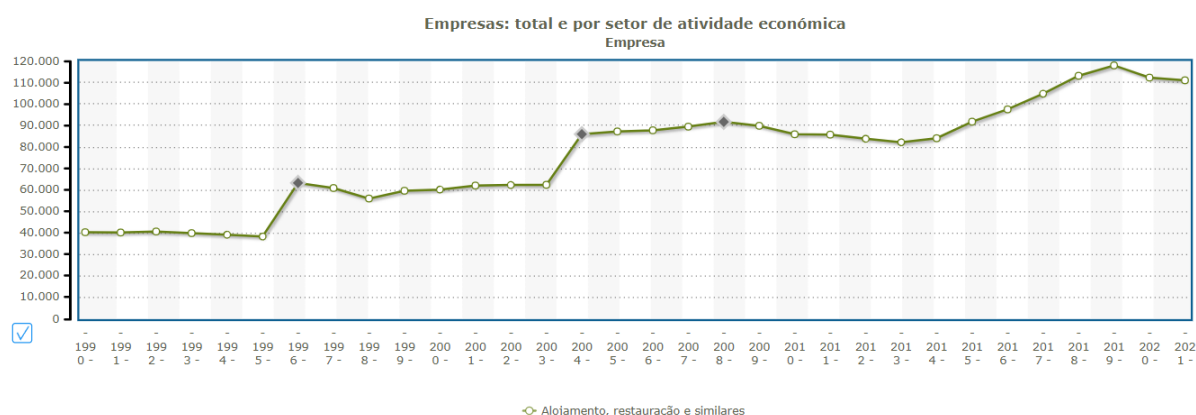


Figure 1 - Companies: total and by industry (PORDATA - Empresas, 2023)

The concept of tourist accommodation has become very common and increasingly attractive for both customers and owners of accommodation, as the associated costs are reduced, for example, in terms of employees and licenses. It becomes quite practical for an individual who has a house available to monetize it and obtain extra income.

In the Figure 2, obtained from the website PORDATA (PORDATA - Alojamentos turísticos, 2023), we can see an increase in the number of tourist accommodations since 2012. Only with a drop in 2020, which can be justified by the COVID 19 pandemic, but which recovered in the following years.

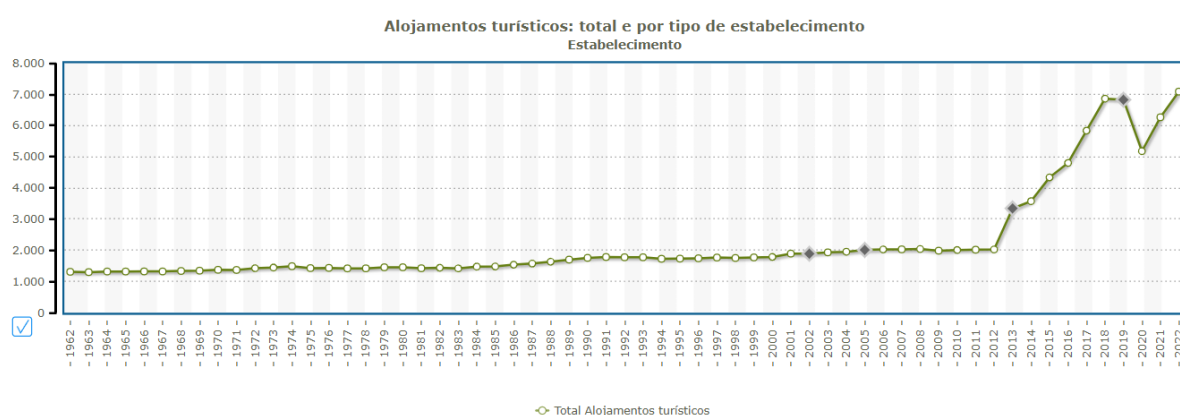


Figure 2 - Tourist accommodation establishments: total and by type of establishment (PORDATA - Alojamentos turísticos, 2023)

The Figure 3, with information obtained from the PORDATA website, allows a 10-year view of the evolution of the GDP ratio in tourism. It is possible see the decline that occurred from

2019 to 2020, and that 2022 was the year in which the GDP ratio was highest, both in terms of exports and imports.

Rácio - %			
Anos	Saldo	Exportações	Importações
2012	3,4	5,1	1,8
2013	3,6	5,4	1,7
2014	4,1	5,9	1,8
2015	4,6	6,5	1,9
2016	5,0	6,9	1,9
2017	5,9	7,9	2,1
2018	6,1	8,3	2,2
2019	6,1	8,5	2,4
2020	2,5	3,8	1,4
2021	(R) 3,0	(R) 4,7	1,7
2022	Pro 6,4	Pro 8,7	Pro 2,3

Figure 3 - Travel and tourism balance in percent of GDP (*PORDATA - Balança de viagens e turismo em % do PIB, 2023*)

Legend: (R) - Data rectified by the responsible entity; Pro - Provisional value

In this project, a certain type of accommodation is considered more relevant, such as local accommodation and guesthouse tourism since they do not always have an online website and are more likely to use the platform. It is therefore important to understand how these two types of accommodation are evolving.

In the Figure 4, taken from the PORDATA website, we can see that local accommodation represents more than fifty percent of all tourist accommodation, and that if we add rural accommodation tourism, we have around seventy percent of tourist accommodation.

Anos	Estabelecimento		
	Alojamentos turísticos		
	Total	Alojamento Local	Turismo de habitação e no espaço rural
2017	5.840	2.663	1.419
2018	6.868	3.534	1.469
2019	↓ 6.833	↓ 3.223	↓ 1.687
2020	5.183	2.240	1.374
2021	6.271	2.811	1.631
2022	7.095	3.277	1.793

Figure 4 - Tourist accommodation establishments: total and by type of establishment
(PORDATA - Alojamentos turísticos, 2023)

1.1.2. Airbnb

“Our mission is to create a world where you can belong anywhere. From cabins to castles to cooking classes, browse through millions of Listings to find the ones that fit the way you like to travel. Learn more about a Listing by reviewing the description and photos, the Host profile, and past Guest reviews. If you have questions, just message the Host.” (Airbnb Terms of Service, 2023)

Airbnb is an online platform that allows its users easy access to share or search for accommodations or experiences.

This offer is made available as an advertisement, which excludes any responsibility of Airbnb in the relationship and contract between the supplier and the customer.

In this way, it is up to each accommodation to guarantee the quality of the information published, the service and the experience provided, and to ensure that the accommodation has a good rating, generated by reviews and comments, and thus attract greater interest from other customers.

Customers have the responsibility to analyse each offer and understand if it meets their needs and if the experience will be as expected, for example by looking for a higher rating and checking comments from other customers.

The customer has a greatly improved search service that allows them to apply various filters in terms of host, accommodation or experience, dates, location, prices and display results by relevance. This feature is built on the experience of other customers and the experience of the customer doing the research.

The available data will be focused on accommodation since, the experiences have great potential and are a service that could complement accommodation, but at this point we do not have specific data on experiences available to analyse.

1.1.3. Tourism in Lisbon and Porto

The choice of the Lisbon and Porto regions for this project is related to the fact that Lisbon is the capital of Portugal and Porto is considered a second capital in the north of the country. These two regions are responsible for a huge flow of people, since they concentrate many large companies, have the largest and most prestigious professional and educational offer, receive a large number of passengers at their airports and, in terms of tourism, also present a very wide and varied range of places to visit, experiences, accommodation, restaurants, bars and others.

As can be seen in the Figure 5, taken from the PORDATA website, in 2022 Lisbon welcomed around fifty percent of the passengers who travelled to Portugal at its airport and Porto welcomed around twenty-two percent of the passengers at its airport. In other words, only about twenty-eight percent of the passengers who travelled to the country did not pass through one of these two cities.

Indivíduo

Anos	Passageiros				
	Total	Aeroporto de Lisboa	Aeroporto do Porto	Aeroporto de Faro	Outros aeroportos
2017	52.713.187	26.676.552	10.790.271	8.728.876	6.517.488
2018	56.286.907	29.045.733	11.942.333	8.687.064	6.611.777
2019	60.114.157	31.184.594	13.112.453	9.010.860	6.806.250
2020	18.392.550	9.267.968	4.436.370	2.208.276	2.479.936
2021	25.598.844	12.148.972	5.841.819	3.265.182	4.342.871
2022	Pro 56.762.273	Pro 28.261.883	Pro 12.637.645	Pro 8.170.715	Pro 7.692.030

Figure 5 - Passenger traffic at airports: Lisbon, Porto and Faro (*PORDATA - Tráfego de passageiros nos aeroportos, 2023*)

The Figure 6, obtained from the PORDATA website, shows the comparison regarding the number of accommodations in Portugal between the years 2009 and 2022. It is immediately apparent that the number of lodgings will increase significantly in the regions shown, with Lisbon and Porto seeing a tripling of the number of lodgings.

Tourism is a strong driver of economic activity, allowing regions that focus more on this sector to develop their structures, transports and attractions, which in turn leads to an exponential increase in the number of accommodations and competition.

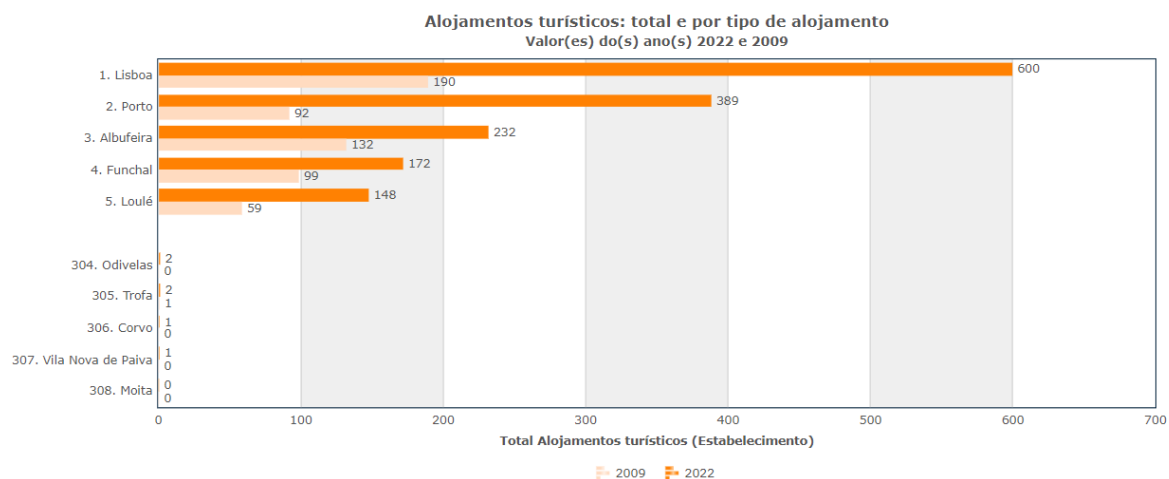


Figure 6 - Tourist accommodation: total and by type of accommodation (PORDATA, *Alojamentos turísticos*, 2023)

1.2. OBJECTIVES AND METHODOLOGY

This project aims to provide a detailed analysis of the trends and patterns of Airbnb accommodation in Lisbon and Porto. This information will help to understand the evolution of this concept in Portuguese tourism, as well as aspects that can be considered for the best relationship between customer and accommodation provider.

The analysis will be based on the information provided by the Inside Airbnb website (Get the Data, 2023) and will therefore essentially focus on the processing and analysis of information about accommodations registered by hosts on the Airbnb platform.

The project will be developed using a quantitative research methodology. The XLSX files will be extracted from the Inside Airbnb website, uploaded, and transformed to create a structured database of fact and dimension tables using a star schema modelling approach. This ETL process will be based on Microsoft Fabric, which provides an integrated environment with tools for storing, transforming, analysing data, and more. Data analysis will therefore be performed using Microsoft Power BI. The available data can be found on the Inside Airbnb website. This data consists of listings, calendars, reviews, and supplemental information.

The creation of ETL allows the data processing and analysis process to be automated, making it possible to load new data as it becomes available. In addition, the use of Fabric brings a new approach, as it is a recent tool, and it is difficult to find projects that have been developed on this basis.

1.2.1. Research Questions

As mentioned above, this project aims to discover patterns and analyse the accommodations available on the Airbnb platform in Portugal, so it was necessary to define specific and measurable questions to build an analysis that meets expectations. The Table 1, present the questions that this study aims to answer.

Table 1 - Research Questions and Objectives

Research Questions	Objectives
Analysing the context of Airbnb in Portugal, in the regions of Lisbon and Porto	Understand the impacts verified on prices according to the characteristics, reviews, and hosts.
	Discover patterns between Lisbon and Porto Airbnb's.
	Identify the specifications of Lisbon and Porto Airbnb's.
	Understand the evolution of years in terms of accommodation.

After defining the main research points, an investigation and analysis of works related to Airbnb's and tourism will be carried out. After this analysis, the collected data will be studied, the work structure to be followed will be defined, data collection and processing processes will be created, through Fabric, so that it is possible to create a report that facilitates data analysis.

After creating the necessary conditions for the analysis, the results will be collected, discussed and related to the work analysed in the literature review, and in a final phase, the conclusions, limitations and recommendations for future work will be presented.

2. LITERATURE REVIEW

2.1. RELATED WORK

The literature review presented sought to conduct an analysis of works that studied common points with those of this project, namely in terms of the Airbnb platform, location, prices, as well as in terms of the data set used.

2.1.1. Touristification and Gentrification of Cities

Lisbon has seen an increase in tourism in recent years, with the number of hotels, short-term accommodation and airport arrivals doubling. This effect is allowing the city to attract more investment and is encouraging the restoration of the historic centre.

With this increase, we are witnessing a touristification of Lisbon, described as *“In cities, it involves the penetration of tourism into residential areas, causing a tacit land use change in which residents are displaced by transient populations* (Estevens et al., 2023). According to the information presented in the article "The Role of the State in the Touristification of Lisbon", the number of households in the historic centre of Lisbon has decreased significantly in the last 40 years, and there has been an increase in the number of tourist establishments and accommodation.

The expansion of tourism has been strongly promoted by the state, through the provision of regulations and a fiscal framework conducive to the construction of any property in the historic centres.

Touristification is driven by the state and the distribution of work, which develops in the areas and leads families to move to the periphery due to high housing prices. City centres are developed and have the highest quality of services or are out of reach for most households.

According to the article "Tourism gentrification and P2P accommodation: The case of Airbnb in Bandung City", *"Gentrification is a spatial process referred to as the concentration of social, economic, and environmental changes resulting in the development of an area that looks different from others"* (Aritenang & Iskandar, 2023).

There is a relationship between gentrification and tourist attraction. Airbnb, as a sharing economy, plays a notable role in this regard, as more accommodation is created and advertised in tourist destinations, driving change in the regions. With the development of the region, the number of available accommodations also increases, and there is a tendency for those responsible for accommodation to become more professional.

“The sharing economy arises as an emerging urban economy in business environments that utilise the role of digital platforms in managing knowledge and connections” (Aritenang & Iskandar, 2023).

As shown in Figure 7, in the regions where this sharing economy is developing, such as the Bandung region in Indonesia (a major tourist destination), it appears that the increase in the number of Airbnb listings is accompanied by simultaneous economic and demographic changes. Economic changes occur in the region where there is a greater concentration of Airbnb, and demographic changes are visible in the peripheral regions of this greater concentration. There is a tendency for secondary cities to change only in terms of quality of life and tourist attractions, with no social or cultural changes for the local population.

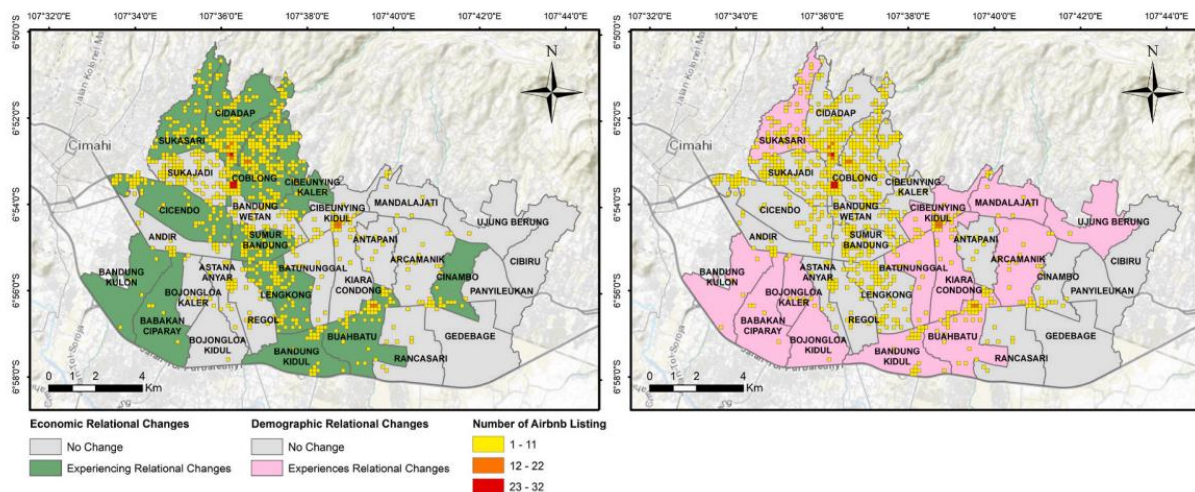


Figure 7 - Economic and Demographic changes and the Airbnb locations (Aritenang & Iskandar, 2023)

2.1.2. Airbnb impact on housing market

Airbnb has an economic impact on society, with relevance to local housing. The advertisements published in each region can cause a significant increase in the value of rents and house prices in the region and surrounding areas, forcing low-income families to move out of these areas.

The number of lodgings is a pattern that has a major impact on the region, as hosts tend to buy more and more lodgings, creating a monopoly.

Hosts tend to expand their empires and professionalize their work in providing this type of short-term accommodation. These trends have an impact on rents and house prices, the latter being more significant. There is a strong impact on the number of households in a situation of relative poverty, which are no longer able to sustain themselves in the regions with the highest concentration of tourist accommodation.

According to the article “The impact of Airbnb on residential property values and rents: Evidence from Portugal”, a one percentage point increase in Airbnb's share corresponds to a three-point seven percent increase in housing prices (Franco & Santos, 2021). There were also larger effects in historic centres and more tourist-attractive regions.

For the owners, tourist accommodation is seen as a profitable investment with fewer risks compared to long-term accommodation. Although this long-term lodging activity displaces families with lower incomes from the above-mentioned regions, this effect also allows for the restructuring of historic centres, as they have become an attractive investment. Older houses are being restored or rebuilt. Attracting people with higher incomes provides more goods and services of higher quality and sophistication.

The State can play a decisive role in mitigating the impact of Airbnb's in the regions, but the strong investment capacity and the contribution of investors to the rehabilitation of centres means that the measures adopted do not have a major impact on large owners, who have the financial capacity to better comply with any standard.

2.1.3. Sharing economy

Most of the revenue generated by Airbnb is distributed to only ten percent of hosts, and that the measures implemented by the government have caused a slowdown in their area but cannot prevent them from becoming large hosts.

In terms of gender, some differences are identified between men and women, with women earning about twelve percent less than men and black hosts earning about twenty-two percent less. Thus, we cannot assume that the discrepancy is caused specifically by the gender, because exits the possibility of those listings to be part of families and only the man or the woman is registered in the platform as host.

The professionalization of hosts is supported by the platform, which provides support tools to assist large-scale hosts. The Airbnb accommodation markets show great inequalities and there is a domain of hosts that tend to become more and more professional, with the strategies adopted by the government being insufficient to stop this growth and increasingly imposing its intervention at the political level, given its growing influence.

The activity of the hosts has a strong influence on the ratings given to them, encouraging them to have more visits, their owners to have more income and, consequently, there is a tendency for them to increase the number of properties they own and invest in better properties and provide better conditions.

The article "The greedy unicorn: Airbnb and capital concentration in 12 European cities" note that the distribution of the highest socioeconomic class is associated with a greater number of advertisements, causing a greater concentration in certain points (Anselmi et al., 2021). In an economy strongly dominated by the economic power of large investors, it seems that they are using Airbnb to penetrate markets and strengthen their presence in large cities. Some of these cities are trying to fight this trend through regulation, which again may not be effective given their investment capacity to adopt any standard. The concentration of Airbnb is a growing trend and is also driven by the increase in the number of comments and users (Anselmi et al., 2021).

The article "Optimal pricing strategies for Airbnb when competing with incumbents on two sides" (Zheng et al., 2023) aims to show how Airbnb's compete with nearby hotels and long-term accommodations, trying to discover what prices and strategies should be practiced making a profit. This study examines several aspects related to guest preferences, the limitations of each accommodation, the relationship between supply and demand, and the relationship between hotel prices and long-term accommodation rents (Zheng et al., 2023).

It was found that it is unthinkable to adopt low price and low revenue strategies if the goal is to make a profit. Among the different strategies studied, it is emphasized that Airbnb's adopt pricing strategies based on guest preferences. High-price strategies should be used when the goal is to drive long-term rentals out of the market.

Price alignment and competition between accommodation types can create unfair competition and have major economic and social impacts, which may require government intervention to limit prices, long-term rentals, or fees charged across platforms.

2.1.4. COVID-19 impact on the housing market

In recent years, we have witnessed several disasters with irreparable damage to society at all levels. In 2019, with the emergence of the COVID-19 pandemic, a highly transmissible virus that caused many deaths around the world, and in 2023, at the beginning of the normalization after the pandemic, we saw the arrival of a war between Ukraine and Russia.

COVID-19, due to its transmissibility and the restrictions adopted worldwide to stop its transmission, is a very relevant factor in terms of tourism. During the pandemic, movement between countries was quite restricted, consumption patterns changed due to greater health concerns, and it was necessary to adapt conditions in all tourist accommodation.

As suggested in Figure 8, in terms of rentals, after the pandemic, rental prices decreased, and the number of available accommodations increased. In terms of accommodation, there was a decrease in the number of overnight stays, an increase in the number of long-term accommodations, the existence of a direct relationship between accommodation prices and demand, and that smaller houses are subject to more negative impacts (Batalha et al., 2022).

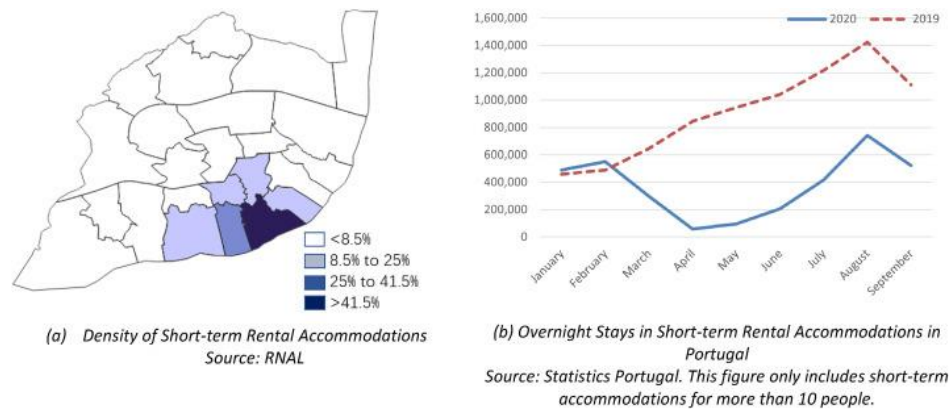


Figure 8 - Density and overnight stays in short-term Rental Accommodations (Batalha et al., 2022)

2.1.5. Inside Airbnb dataset

The basis for the development of this project is the Inside Airbnb website, so it is important to verify the information available about the source. The Inside Airbnb platform collects data on accommodations, experiences, and reviews provided by users.

The article "Incorrect data in the widely used Inside Airbnb dataset" (Alsudais, 2021) aims to analyse the dataset available on Inside Airbnb, as it has been used in several studies and has not been properly verified. The inaccurate data is related to the reviews added to the platform by its users, as it may not be considered accurate when verifying the review and the page from which the listing was supposedly obtained.

For example, in Figure 9 there is an error where the information includes reviews with the same ID but belonging to different listings, such as Places and Experiences. The existence of different versions of data can make its correct validation difficult, and it is very important that any investigation refers to the version of the information collected.

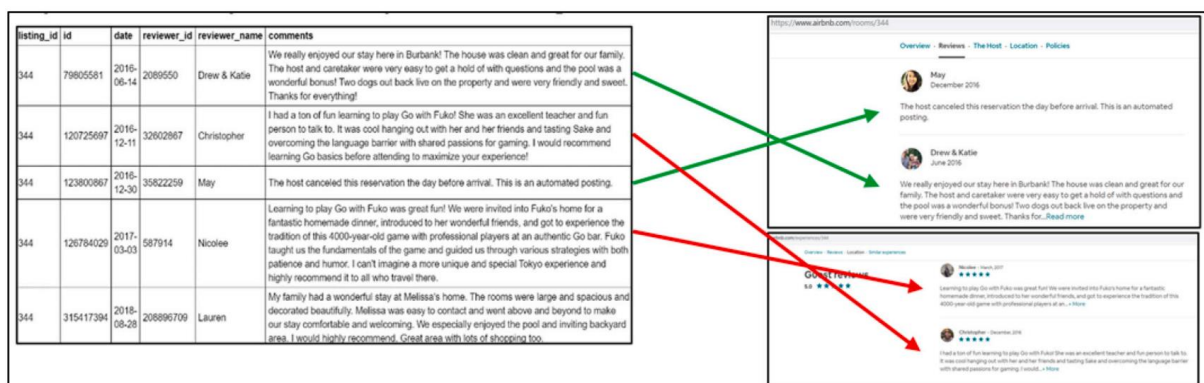


Figure 9 - Example of reviews from different listings (Alsudais, 2021)

Source: (Alsudais, 2021)

Various risks can be associated with any data processing project, which is why we have tried to understand the main aspects to be considered through the article (Varela, 2021). We can find several risks associated with data science, among which we highlight communication failures between the parties involved, errors in the definition of research questions, lack of documentation, poor quality of data or its translation, and loss of knowledge of the business.

In this sense, it is essential to ensure adequate planning and the definition of strategies that act to prevent possible risks. In a reality where the volume of data is increasing and the sources are innumerable, it is important to guarantee the updating and identification of data, to have properly trained work teams to process data at the level of data science and their understanding of the business.

Article (Varela, 2021) highlights the importance of specifying the source, the extraction date, and creating a structured and well-documented project that meets the needs encountered in any project. The aspects verified are very important when analysing the Inside Airbnb dataset, as they lead to a more careful analysis and attention to the correct treatment of the data.

2.1.6. Literature Review Resume

Following the literature review presented in section 2.1, Table 2 was prepared to facilitate quick understanding and knowledge of the topics presented.

Table 2 - Literature Review resume

Topic	Year	Author(s)	Topic(s)	Study Objective(s)	Reference
Touristification and Gentrification of Cities	2023	Ana Esteve, Agustín Cocola-Gant, Antonio Lopez-Gay, Fabiana Pavel	The role of the state in the touristification of Lisbon	Analyse the development of historic centres and the role of the state in tourism.	(Esteve et al., 2023)
	2023	Adiwan Aritenang, Fahlan Zahara Sitta Iskandar	Tourism gentrification and P2P accommodation: The case of Airbnb in Bandung City	Analyse the relationship between gentrification and the increase in the number of advertisements in the city of Bandung.	(Aritenang & Iskandar, 2023)
Airbnb impact on housing market	2021	Sofia F. Franco, Carlos Daniel Santos	The impact of Airbnb on residential property values and rents: Evidence from Portugal	Assessing the impact of increased Airbnb participation on house prices.	(Franco & Santos, 2021)
	2023	Seonjin Lee, Hany Kim	Four shades of Airbnb and its	This study Analyses the economic impact of	(Lee & Kim, 2023)

Topic	Year	Author(s)	Topic(s)	Study Objective(s)	Reference
			impact on locals: A spatiotemporal analysis of Airbnb, rent, housing prices, and gentrification	Airbnb on society, with relevance to local housing.	
Sharing economy	2021	Guido Anselmi, Letizia Chiappini, Federico Prestileo	The greedy unicorn: Airbnb and capital concentration in 12 European cities	Examine the political economy associated with the concentration of Airbnb, using Inside Airbnb as a source.	(Anselmi et al., 2021)
	2022	Petter Törnberg	How sharing is the “sharing economy”? Evidence from 97 Airbnb markets	Analyse the level and distribution of income in the Airbnb market, considering aspects such as the gender and race of accommodation hosts.	(Törnberg, 2022)
	2023	Qing Zheng, Da Ke, Xuan Li	Optimal pricing strategies for Airbnb when competing with incumbents on two sides	Show how Airbnb's compete with nearby hotels and long-term accommodation and try to work out what prices and strategies should be used to make a profit.	(Zheng et al., 2023)
The impact of covid-19 on the housing market	2022	Mafalda Batalha, Duarte Gonçalves, Susana Peralta, João Pereira dos Santos	The virus that devastated tourism: The impact of covid-19 on the housing market	Verification of the impact of the COVID-19 pandemic on short-term rentals, using the Lisbon city centre as a reference and verifying the robustness at the level of Porto.	(Batalha et al., 2022)
Inside Airbnb dataset	2021	Abdulkareem Alsudais	Incorrect data in the widely used Inside Airbnb dataset	Analyse the dataset available on Inside Airbnb, as it has been used in multiple studies and has not been properly verified.	(Alsudais, 2021)
	2021	Varela, Ana Cristina Afonso	Risk management in data science projects in Portugal	Investigation to create a base list of the risks of Data Science projects and their surrounding factors.	(Varela, 2021)

3. METHODOLOGY

The project will be developed using a quantitative research methodology and to develop the Business Intelligence solution will be used the Kimball Lifecycle methodology (Kimball Group, 2024). Kimball allows the creation of a model oriented for the business with first came to support decision making, and to have a structured and efficient data model. The phases of Kimball are Project Planning, Business Requirements Definition, Technical Architecture Design, Dimensional Modelling, ETL Design and Development, Data Lakehouse Deployment and, Maintenance and Growth.

3.1. PROJECT PLANNING

The main objective of this project is to analyse data containing information about the Airbnb platform in Portugal. The csv files will be extracted from the Inside Airbnb website (Get the Data, 2023), uploaded and transformed to create a structured database of fact and dimension tables using a star schema modelling approach. This ETL process is based on Microsoft Fabric, which provides an integrated environment with tools for storing, transforming, and analysing data. Data analysis is therefore performed using Microsoft Power BI (Microsoft, What is Microsoft Fabric?, 2023); (Microsoft, Understanding star schema and the importance for Power BI, 2023). This project takes a different approach to other projects that have been done, as Fabric is a new tool and there are no projects that have been developed in this way.

The available data can be found on the Inside Airbnb website (get-the-data, 2023). This data consists of listings (Airbnb information), calendar, reviews, and complementary information.

3.2. BUSINESS REQUIREMENTS

Four business problems were initially defined for the development of this project. In Table 3, several requirements were defined for each problem to guide the development of the project and provide measurable objectives.

Table 3 - Business Problem into Business Needs

Business Question	Business Needs
Prices relations with Airbnb characteristics, reviews, and hosts.	Does region affect the value of accommodations? Is there a relationship between the average price of accommodation and the score? What are the 10 accommodations with the highest average price in Lisbon and Porto? What is the maximum and average price per quarter? In which quarter of the year is the average price the highest?

Business Question		Business Needs
Portuguese Airbnb characteristics.		What is the most common type of property in Portugal and by Lisbon/Porto region?
		What is the most common type of accommodation in Portugal and by region Lisbon/Porto?
Airbnb differences between Lisbon and Porto.		Which city has the most properties in Portugal and by region Lisbon/Porto?
		What is the average price of accommodation in Portugal and by region?
The evolution of Airbnb and the differences between old and new hosts.		How many accommodations and hosts are there in Portugal and by region of Lisbon/Porto? And what is their average price?
		What is the variation in host registrations detected over the past 5 years?
		In terms of Airbnb's characteristics, are there significant differences when analysing older host registration years?

3.3. DATA SOURCE DESCRIPTION

The data was extracted from the Inside Airbnb website, the Get the data page, and Excel file with date information since 2009 until 2124. The downloaded files are in csv and zip format. On Inside Airbnb page can find listings for different cities. In this case, studied the cities of Lisbon and Porto, where are obtained 7 listings for Lisbon and 6 listings for Porto.

The Table 4 and Table 5 present the region, file name and description about the information obtained through the platform for the cities analysed in this project.

Table 4 - Inside Airbnb Lisbon, Portugal (Get the Data, 2023)

Country/City	File Name	Description
Lisbon	listings.csv.gz	Detailed Listings data
Lisbon	calendar.csv.gz	Detailed Calendar Data
Lisbon	reviews.csv.gz	Detailed Review Data
Lisbon	listings.csv	Summary information and metrics for listings in Lisbon (good for visualisations).

Country/City	File Name	Description
Lisbon	reviews.csv	Summary Review data and Listing ID (to facilitate time based analytics and visualisations linked to a listing).
Lisbon	neighbourhoods.csv	Neighbourhood list for geo filter. Sourced from city or open source GIS files.
Lisbon	neighbourhoods.geojson	GeoJSON file of neighbourhoods of the city.

Table 5 - Inside Airbnb Porto, Portugal (Get the Data, 2023)

Country/City	File Name	Description
Porto	listings.csv.gz	Detailed Listings data
Porto	calendar.csv.gz	Detailed Calendar Data
Porto	reviews.csv.gz	Detailed Review Data
Porto	listings.csv	Summary information and metrics for listings in Porto (good for visualisations).
Porto	reviews.csv	Summary Review data and Listing ID (to facilitate time based analytics and visualisations linked to a listing).
Porto	neighbourhoods.csv	Neighbourhood list for geo filter. Sourced from city or open source GIS files.
Porto	neighbourhoods.geojson	GeoJSON file of neighbourhoods of the city.

The listing.csv.gz contains detailed information about the Airbnb, such as location, type of accommodation, number of beds, number of bathrooms, and rating information. The calendar.csv.gz contains information about pricing and availability for the next twelve months from the data reference date. The reviews.csv.gz contains information about the reviews and comments that users have added to the listing. The listing.csv has information about the Airbnb, the file has the same information as the gz file, but with fewer columns. The reviews.csv contains information about the listing ID and the review date. The neighborhoods.csv contains information about the neighborhoods, making the association between neighborhood_group and neighborhood_group.

3.3.1. General Statistics About the Data

The data available on the Inside Airbnb platform (Get the Data, 2023) used for this project has a reference date of December 17, 2023. The files were extracted on February 1, 2024, but the available data is not updated daily, so the reference date is the same as before.

The Table 6 provides a summary of the data used and its distribution by category and region. In reviewing these datasets, the Lisbon region has more listings and hosts, with a total of 36,352 Airbnb listings in Portugal. The number of hosts per region differs from the total number because the same host can have Airbnb's in different regions.

Table 6 - Statistics about the data dimensions

Region	Airbnb	Hosts	Location	Reviews
Lisbon	22,751	9,130	134	1,373,517
Porto	13,601	5,190	173	780,921
Total	36,352	14,237	307	2,154,438

Table 7 shows a brief analysis of the data collected and used to construct the fact table. Therefore, for the project it is considered important to check the number of rows loaded per region and in total, the minimum date displayed, the maximum date and the average price. The minimum date refers to the extraction date, since the data is presented for the following year. We can see that, similarly to the number of accommodations, the number of price records is super in Lisbon.

Table 7 - Statistics about the data fact table

Region	Lines	Minimum Date	Maximum Date	Average Price
Lisbon	8,314,817	2023-12-17	2024-12-24	155.615440
Porto	4,972,435	2023-12-17	2024-12-23	134.033386
Total	13,287,252	2023-12-17	2024-12-24	147.538874

3.3.2. Data review

Given the information that there are few validations of the dataset used as a source for this project, (Alsudais, 2021), is important to perform some checks. The analysis of the project focuses in part on price, so is verified the existence of two listings with a price of zero dollars and one with a price of fifteen thousand dollars.

In this sense, using the Airbnb website, has been verified that for one of the Airbnb's with a price of zero dollars there is no availability to reserve the accommodation, which justifies the value, and for the other, at the current date, it has a price of one hundred fifty dollars per night. The extraction refers to December 17, 2023, so the accommodation might not be available on the date of the data extraction, but this information is not available to say it for sure.

Regarding the accommodation with the highest value per night of fifteen thousand dollars, this is confirmed on the website, being a luxury loft in Lisbon.

In the Airbnb license information, identify licenses that appear to be false, as in the Figure 10. However, since is not possible determine the veracity of these licenses, the analysis only determines whether there is an indication of a license or not.

`select distinct license from dimDetail order by license asc`

essages **Results** Download Excel file Explore this data (preview)

ABC license ↓
null
-U-001999-1*AND
0
0/AL
00/AL
0000/AL
00000
00000/AL
000000
000000/AL
0000000
0000000/AL
00000000
0000002907948
00001
000012345/AL

Figure 10 - Example of licenses

In this research, it was essential to aggregate information not only by name, but also by host_id, since there are several users, whose name is a Portuguese first name, as shown in the Figure 11.

```
10 select * from dimHost where host_name='Pedro'
```

Messages Results Download Excel file Explore this data (preview) ✓

	12L host_id	ABC host_url	ABC host_name	host_since	ABC host_location	ABC host_about
1	33287282	https://www.airbnb.com/users/show/33287282	Pedro	2015-05-13	Lisbon, Portugal	Hello! I come from Lisbon, where I live and work. I enjoy a lot
2	131888156	https://www.airbnb.com/users/show/131888156	Pedro	2017-05-25	Lisbon, Portugal	We love to receive and meet new people, and want you to fee
3	131193518	https://www.airbnb.com/users/show/131193518	Pedro	2017-05-20	Lisbon, Portugal	Somos o Pedro e Ana, gostamos muito de viajar e ser bem rec
4	20179403	https://www.airbnb.com/users/show/20179403	Pedro	2014-08-17	Lisbon, Portugal	Hi. I'm Pedro and I'm 38 years old. I was born in Lisbon and ha
5	19595047	https://www.airbnb.com/users/show/19595047	Pedro	2014-08-06	Lisbon, Portugal	Hello everyone, my name is Pedro, Im 34 and Im Portuguese. I
6	522369919	https://www.airbnb.com/users/show/522369919	Pedro	2023-06-27	Lisbon, Portugal	Soy un hablador empedernido, contador de historias. Me gust
7	67648423	https://www.airbnb.com/users/show/67648423	Pedro	2016-04-17	Lisbon, Portugal	Step into the world of our extraordinary family. We're Pedro ar
8	93464728	https://www.airbnb.com/users/show/93464728	Pedro	2016-09-04	Lisbon, Portugal	Hello. Did you know hospitality and friendliness are some of t
9	50396857	https://www.airbnb.com/users/show/50396857	Pedro	2015-12-02	Lisbon, Portugal	Keep Calm and Enjoy! Two young architects who started to bu
10	55164620	https://www.airbnb.com/users/show/55164620	Pedro	2016-01-18	Lisbon, Portugal	My name is Pedro, I was born and always lived in Lisbon. I hav

Figure 11 – Example of Host's with Pedro as host name

3.4. DIMENSIONAL MODELLING

3.4.1. Business Process and Granularity

The model presented will have the ability to display information at different levels of granularity, with a high level of granularity that allows information to be drilled down. In this model, present two fact tables with information about Airbnb prices for the year following the extraction date, the corresponding date and number of available nights, and the foreign keys that allow connection to the dimensional tables, and about Reviews with the comments that have been added to each Airbnb.

3.4.2. Dimension Tables

Dimensional tables are composed of their primary keys, respective IDs, and dimensional characteristics. The dimDetail table displays details about the accommodation. The dimDate table shows the calendar from 2009, the first year shown in the host_since column, to 2124, 100 years from now. The dimHost table allows you to view information about the host/user responsible for hosting. The dimLocation table contains information about the location of the accommodation.

For a better analysis and interpretation of the data, calculated columns were created whose formulas are presented in Appendix A.1. and hierarchies were created in the dimDate, dimLocation, dimDetail and dimHost tables to allow a presentation in which the data are related and follow a hierarchical arrangement, these are presented in Appendix A.2.

In order to create a solution that is easier to interpret, measures were also added, Appendix A.3., where variables are created with the average, maximum, minimum price, number of hosts and Airbnb's and average ratings. A KPI based on the number of hosts was created to understand the growth trend of their annual registration, Appendix A.4.

3.4.3. Fact Tables

The fact tables for this project are factAirbnb and factReviews. According to the identified business needs, the fact table used in this analysis will be factAirbnb, which contains information about the prices charged by each Airbnb in the year following the extraction. The factReviews table shows the comments that have been added to each shelter. This table is not used in the analysis because it only contains reviews, and only want to extract the number of reviews for each Airbnb.

3.5. ETL DESIGN AND DEVELOPMENT

3.5.1. Microsoft Fabric

The choice of Microsoft Fabric as a tool for data processing and analysis in this project is due to the fact that this tool allows working in the cloud, reading data of various structured and unstructured types. Fabric is a tool that allows the integration of all data science in a single environment and, due to its simplicity, allows easy implementation of integration processes. In this case, Data Engineering is used to create the data lake and data Lakehouse where the data is loaded, Data Factory is used to create the data import, transformation, export flows in pipelines and use Power BI for reporting.

In this project, two data lakes will be created, one where the initial files will be loaded into tables and another where the final tables will be loaded after transformation. The second data lake will allow the storage of a large amount of structured and unstructured data without the need for transformation.

Parquet files are data storage formats that are represented in the data lake and allow for greater efficiency in terms of space and better database performance. The pipeline refers to a sequence of processes in which multiple steps are performed, such as reading, processing, and ingesting data.

3.5.2. ETL Process Development

Figure 12 shows the steps taken to develop this project, from collecting the data to creating the dashboard.

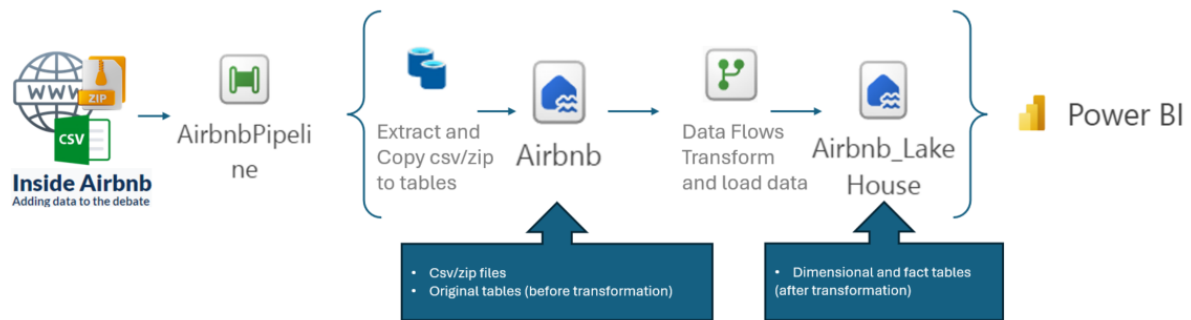


Figure 12 – ETL Process Development

3.5.2.1. Pipeline

The Figure 13 shows the pipeline used to work the data on this project, that consists of data copy tasks and data flow tasks. The tasks have a defined execution order that aims to correctly implement an ETL process with data extraction, transformation, and loading.

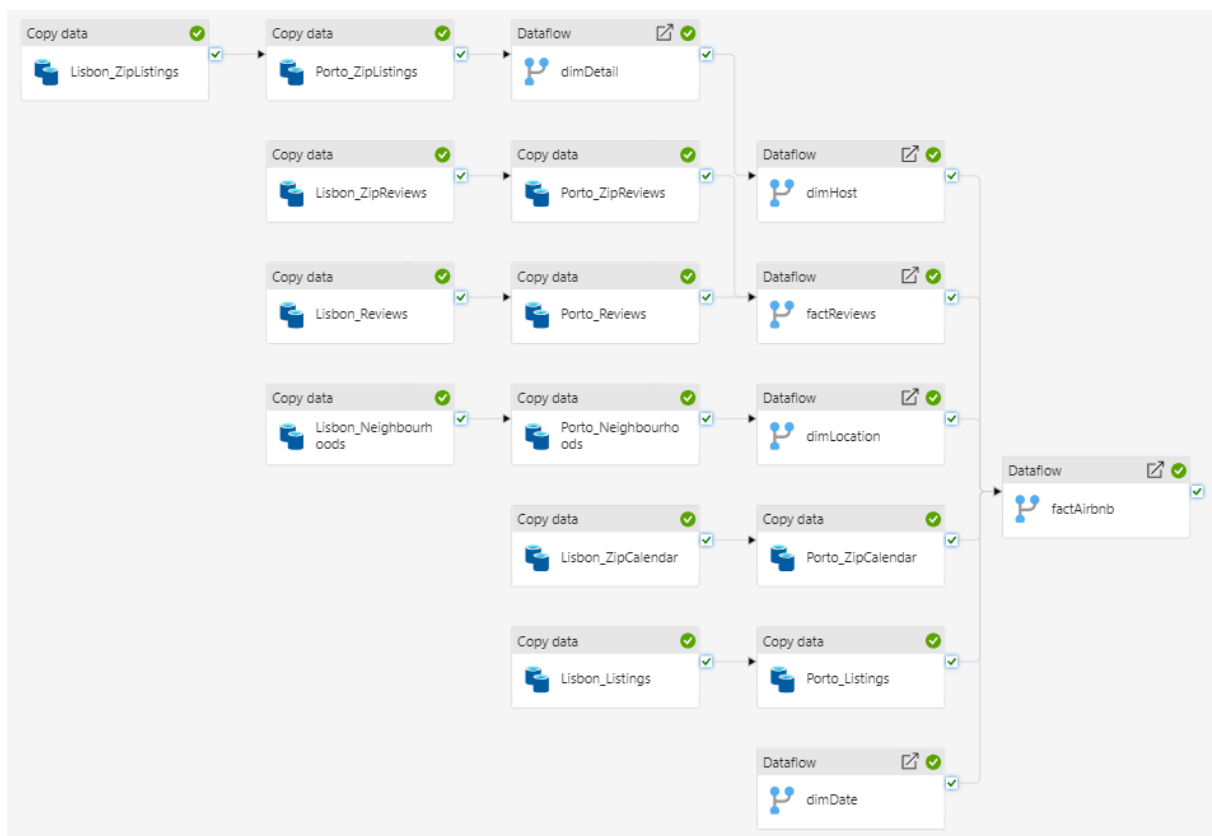


Figure 13 - Airbnb Pipeline

The data copy tasks retrieve the data from CSV and zip files and copy it into tables to create the parquet files. This process creates a Lisbon_Porto column that returns a value of 1 if the file is from Lisbon and a value of 2 if the file is from Porto.

Thus, the process starts by copying the data related to Lisbon into a table, then adding the data related to Porto into the same table, this order is important so that the first process cleanses the data from the table and the second only adds the data for that region.

Once these steps have been successfully completed, it is possible to move on to the dataflow. In the first line, after loading the information from Lisbon and Porto into the base table, we have the dataflow for creating and maintaining the dimDetail and dimHost tables. This precedence is defined by the fact that both tables use the same source table for data collection.

The same chronological order is presented for loading the dimReviews and dimLocation tables. As for the dimDate table, it starts at the same time as the other dataflows because it does not need to be preceded by any process. This is due to the fact table, which is only loaded after all tasks have been successfully completed, as it is built based on the remaining tables.

In order to ensure that the data lake used for data analysis is as intended, guaranteeing the integrity of the data and its proper handling, a data flow was created to load the information into each Airbnb_LakeHouse table individually.

The dimDetail table is sourced from the Airbnb ListingCSV data lake table. From this, only columns with Airbnb specific information were selected, and columns with host information and other data not intended for use in this analysis were removed. In this process, it was important to change the data types to ensure that the values were correctly classified in the final table, and then load the data into the final table.

The Figure 14 shows an example of the dataflow created for the dimDetail table.

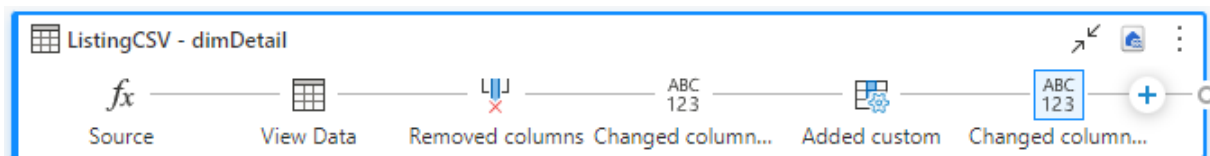


Figure 14 - dimDetail Data Flow

The dimHost table was loaded using the dimHost_DataFlow dataflow, which started by reading data from the ListingCSV source table of the Airbnb Lake house. Then the unnecessary columns are removed from this table, such as specific information about Airbnb. After this removal, it was necessary to modify some types of columns to ensure that they would be correctly parameterized, two columns calculated with the year of creation of the host (YearHost) and the corresponding month (MonthHost) were added, formulas in APPENDIX A - 1.1., with the aim of facilitating the final reading of the data.

Considering that the source contains information on all Airbnb's and that each host can have more than one Airbnb, in this treatment the removal of duplicates was added and then the information was loaded in the final table.

The data flows of the dimReviews, dimLocation and dimDate tables follow the same pattern as the previous ones, using the ReviewsCSV, Neighborhoods tables and an Excel file as sources. Next, the data types of the columns are converted where necessary because they do not match the type of records that make up the columns.

In the dataflow of the dimLocation table, it was necessary to add more information relevant to the processing, such as the country, which in this case was filled with "Portugal", since this project only work with data from Lisbon and Porto, two Portuguese regions. The "District" column has been added, based on the column loaded when importing the files, which indicates the region to which the "Lisbon" or "Porto" information refers.

After these operations, the information is loaded in the corresponding final table, replacing the information. In all these dataflows, the existing information in the table is replaced by a replace when loading the data into the final table.

The fact table has a more complex load because it uses as its source the dimDate, dimLocation tables available in the Airbnb_LakeHouse data lake and the ListingCSV, CalendarCSV tables available in the Airbnb data lake. The fact table contains information about the price of Airbnb's in the year after the extraction and the corresponding date. As a fact table, it must allow the connection with the dimension tables through foreign keys, so the dimension table IDs are collected from the source tables through merges performed between the tables and loaded into it according to the record, thus ensuring the integrity of the data. Once the records are collected, the correct column types are checked and defined, and the layout of the columns is changed to facilitate data interpretation.

3.6. DATA LAKEHOUSE

The schema used in this database is the Star schema, with two centralized fact tables and normalized dimensions, as shown in the Figure 15.

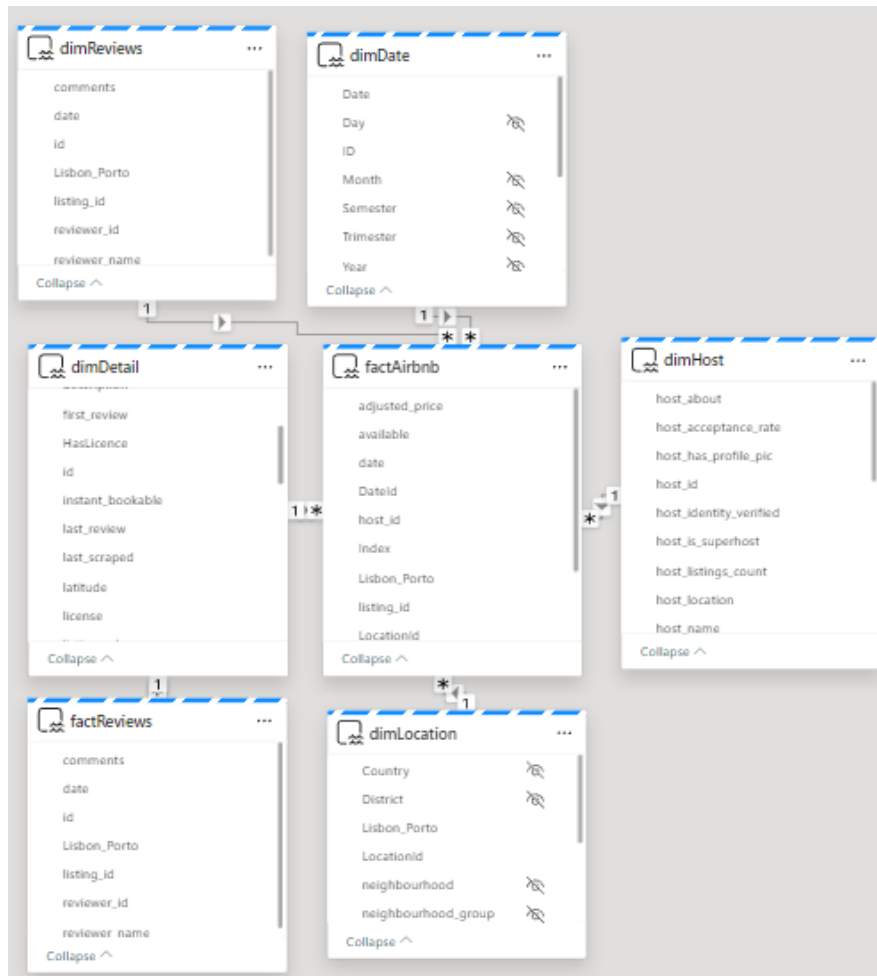


Figure 15 - Data Lake Airbnb Analysis

The tables used in the data lake house aim to collect the necessary insights for a better analysis of Airbnb's in Portugal and allow the information from the different tables to be related to each other. In Table 8, we can better understand the information expected from each table.

Table 8 – Data Lakehouse tables

Table Name	Description and purpose
dimHost	This table presents data related to each host registered on the platform, so its purpose is to know the hosts, their seniority and relate them to the different Aribnb's they have.
dimDetail	This table contains the information about each Airbnb, as their location and type or number of rooms, with the purpose of characterizing the Airbnb's published in Portugal.

Table Name	Description and purpose
dimDate	This table contains information about data that characterizes each day to identify its month, semester, quarter, or year. The table complements the analysis to refer to a temporal space.
dimLocation	The location table provides more details about the neighborhood, since the location is in the Lisbon or Porto region and its group. It is important to analyse the data according to their location.
dimReviews	This table contains information about the reviews, such as the user and the date. In this project, this table is not used because possible insights, such as the number of reviews or date, are not one of the main goals of the project analysis.
factAirbnb	This is the table that lists the different dimensional tables and allows you to make a study of the prices for the following year according to the characteristics obtained from the other tables.
factReviews	This table contains information about the reviews, such as the user or the comment. In this project, this table is not used because possible insights, such as the number of reviews or the text contained in the comments, are not one of the main goals of the project analysis.

3.7. DASHBOARD IN POWER BI

The creation of the dashboard in Power BI aims to facilitate the reading and interpretation of the data, so the creation meets the business needs identified in Table 3 - Business Problem into Business Needs, using an appealing layout and an intuitive interface.

On the first sheet, the objective is to show a summary according to the main characteristics that an accommodation can present, making it possible, through the selection of filters by

region and by year of creation of the host, to check the main differences, with a main header showing a count at the same time of the number of Airbnb's, host's and the average price.

The second sheet shows an analysis of the hosts, indicating the top five with the most registered Airbnb's, the total response time of Portuguese hosts and the evolution of the number of hosts registered in the last five years.

The third sheet presents the analysis of Airbnb's, with information on instant bookings, license ownership, verification of the top five property types and room types, and a list showing the number of listings and average price per property type.

The fourth sheet concerns the location, presenting a filter by region, where through the map it is possible to see the places where there is a greater concentration in the number of Airbnb's, and a graph is also presented with the top five concentration of Airbnb's and the average price associated with the location. The last sheet presents the price analysis, with an analysis of the maximum and average price per quarter, the average price by region, by ranking, and information on the Airbnb's with the highest average prices.

Regarding this report, it is important to note that the reference date of the data has been placed in the bottom right corner so as not to mislead the reader and to make it clear that the data is not in real time.

4. RESULTS AND DISCUSSION

During the execution of the project, a pipeline was created in Microsoft Fabric to load and work with the data. In Table 9, it was possible to see the summary of a process execution, where we can see that loading the Airbnb fact table takes the longest with about 10 minutes and 4 seconds. The fact table is a more complex load because it joins multiple tables to create foreign keys and because it has a large number of records.

Table 9 – Pipeline execution process

Activity name	Activity status	Error	Run start	Duration
factAirbnb	Succeeded		5/8/2024, 7:40:35 PM	10m 4s
dimHost	Succeeded		5/8/2024, 7:38:38 PM	1m 33s
factReviews	Succeeded		5/8/2024, 7:37:30 PM	3m 3s
dimDetail	Succeeded		5/8/2024, 7:37:02 PM	1m 35s
dimLocation	Succeeded		5/8/2024, 7:36:59 PM	1m 4s
Porto_ZipCalendar	Succeeded		5/8/2024, 7:36:59 PM	36s
Porto_ZipReviews	Succeeded		5/8/2024, 7:36:56 PM	33s
Porto_Reviews	Succeeded		5/8/2024, 7:36:39 PM	23s
Porto_ZipListings	Succeeded		5/8/2024, 7:36:38 PM	23s
Porto_Neighbourhoods	Succeeded		5/8/2024, 7:36:37 PM	21s
Porto_Listings	Succeeded		5/8/2024, 7:36:37 PM	21s
Lisbon_ZipCalendar	Succeeded		5/8/2024, 7:36:03 PM	55s
Lisbon_ZipListings	Succeeded		5/8/2024, 7:36:03 PM	34s
Lisbon_Listings	Succeeded		5/8/2024, 7:36:03 PM	33s
Lisbon_ZipReviews	Succeeded		5/8/2024, 7:36:03 PM	52s
dimDate	Succeeded		5/8/2024, 7:36:03 PM	1m 34s
Lisbon_Reviews	Succeeded		5/8/2024, 7:36:03 PM	35s
Lisbon_Neighbourhoods	Succeeded		5/8/2024, 7:36:03 PM	34s

In this project, a report was created to facilitate the reading and analysis of the data. In Figure 16 is showed a summary of the information obtained regarding the regions of Lisbon and Porto, in Portugal.



Figure 16 - Portugal - Airbnb Data

The data used in this project refers to Portugal, with the districts of Lisbon and Porto as a reference. When analysing the data, the number of Airbnb's is more than double the number of Host's, which means that users of the platform, represented here as Host's, have more than one listing (represented here as Airbnb). The KPI indicates that the number of hosts in the last year is greater than the number in the previous year.

The average price charged by Airbnb's in Portugal over the next year is one hundred forty-seven dollars and fifty-three cents, although there are accommodations with prices between zero dollars and fifteen thousand dollars. The score given to the value is very positive, as it has an average score of four and fifty-eight, on a scale of zero to five.

In terms of the type of property most advertised in Portugal, has "Entire Rental Unit" and the type of room "Entire home/apt". In location, Cedofeita - Porto is the place with the most listings in Portugal, and that the average location score is four and seventy-one, on a scale of zero to five.

One of the main objectives of this project is to understand the differences between Airbnb's in Lisbon and Porto. In this sense, the report has a button to change the information between Lisbon, Porto and total, representing the scenarios for the two regions and the combination of them.

It appears that Lisbon has sixty-four percent of hosts and sixty-three percent of Airbnb's. Divide the number of Airbnb's by the number of hosts, it allows us to know that, on average, each host has approximately between two and three Airbnb's in Lisbon and Porto.

In both regions, the most advertised accommodation features are Entire Rental Unit property and Room type Entire home/apt. However, the average price is higher in Lisbon with a value of one hundred fifty-five dollars and sixty-six cents and a maximum price of fifteen thousand dollars, and in Porto the average price is one hundred thirty-three dollars and ninety-five cents, and the maximum price is ten thousand dollars. In terms of score ratings, the average rating is the same in both regions. In terms of hosts, it was possible to verify an increase of eight hundred seventy-nine in Lisbon and six hundred twenty-eight in Porto compared to last year.

The Lisbon region with the most listings is Santa Maria Maior and Porto Cedofeita, both of which are part of the historic centres of the cities. This point confirms the information analysed on the Literature Review.

4.1. HOST'S INFORMATION'S



Figure 17 - Portugal – Host's

This study pretends to understand if there is a tendency to create a monopoly for certain hosts, as also analysed in the literature review. In this way, the Figure 17 present the top five hosts with the most Airbnb's, which include Feels Like Home, Host, BnBird, Blueground and GuestReady.

As mentioned in the article "How sharing is the "sharing economy"? Evidence from 97 Airbnb markets" (Törnberg, 2022), there is a tendency for some hosts to become professional, creating a monopoly. Looking at the five cases with the most Airbnb's, has Feels Like Home with 381 listings, Host with 262, BnBird with 192 in Lisbon and Porto, Blueground with 190 listings only in Lisbon and GuestReady with 175 listings only in Porto.

The response time of the host is something to be taken into consideration by the customer, as it improves the interaction between the provider and the customer, so with the information available in this project it was possible to check the response time and found that most of them take around an hour and only a small part take a few days or more. A large part of the hosts did not specify the time taken for this extraction, which could be an important factor in this characterization of the hosts.

Regarding the evolution of the number of hosts, it is important to understand whether significant changes occurred in Portugal in terms of registration of new accommodations during periods of major economic and social changes, such as the COVID-19 pandemic or the start of the war in Ukraine.

As it was possible to see in the graph that analyses the number of hosts in the last five years, between 2019 and 2022 the number decreased significantly compared to the previous period, and that after this phase it gradually increased, reaching a peak in July 2023. Thus, confirm the theory previously presented in the article "The virus that devastated tourism: The impact of covid-19 on the housing market" (Batalha et al., 2022), in which it is possible to notice the impact felt on the economy.

Regarding the war in Ukraine, the impact felt in this specific analysis of the number of hosts is not relevant, since the war began in February 2022 and the number of registrations did not suffer a sharp drop.

In addition to these aspects, with this analysis it was possible to identify something that appears to be a pattern, since the number of registered hosts is relatively higher in the months of May to August and tends to decrease significantly in the month of December. A situation that occurs regardless of the pandemic or war period, although on a different scale.

4.2. AIRBNB'S INFORMATION

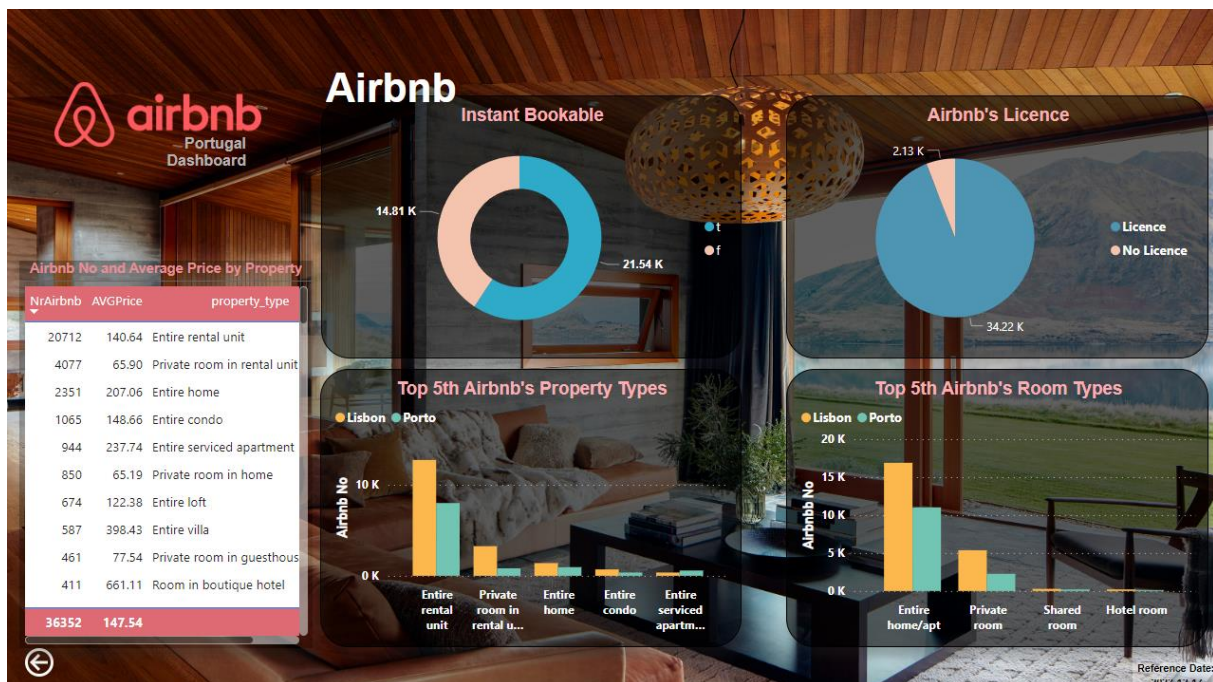


Figure 18 - Portugal - Airbnb's

When analysing Airbnb's, the objective is to have some detail of their characteristics and note pitfalls. For example, it was found that 21.54K of the listings are immediately bookable, about sixty percent of the total. This condition guarantees that the customer automatically keeps the reservation from the moment they make it, without having to wait for a confirmation.

Regarding the Airbnb license, we found that only five percent of the listings do not have a license, which in my opinion is a good indicator of quality, since registered and licensed accommodations may be subject to inspections and must present the conditions established to obtain the license.

4.3. LOCALIZATION INFORMATION



Figure 19 - Portugal - Localization

In the literature review, the works analysed indicate that the tendency is to have more Airbnb's in historical, technological and tourist centres, and families tend to move to the periphery. Based on the analysis carried out, in Portugal, the largest number of advertisements are located in historic centres and coastal locations.

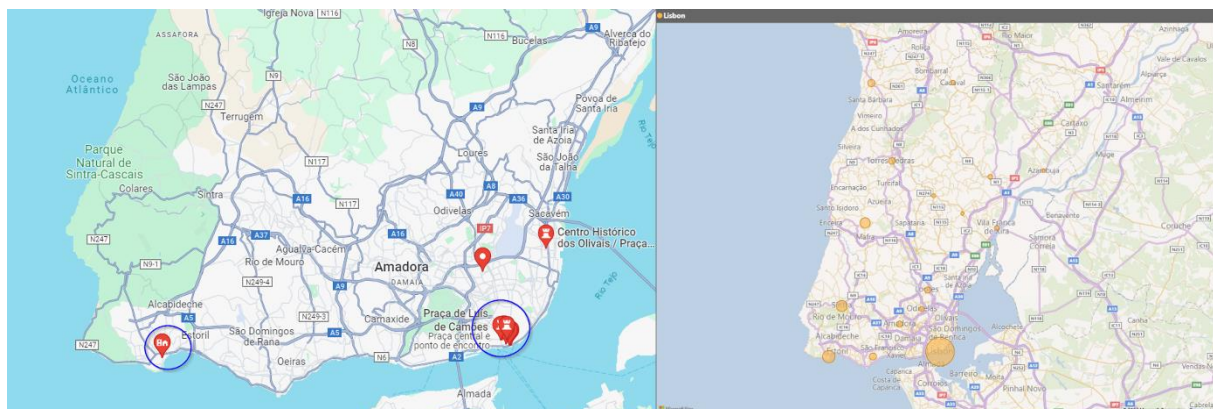


Figure 20 - Lisbon – Historic Centre by Google Maps vs Number of Airbnb's

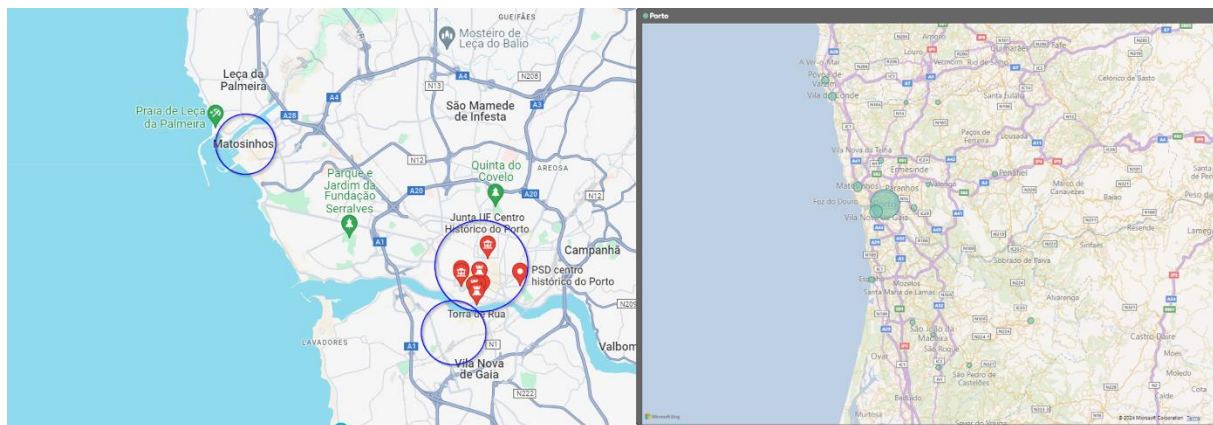


Figure 21 - Porto – Historic Centre by Google Maps vs Number of Airbnb's

According to the "Top 5th Localization" table, we can see that Santa Maria Maior - Lisbon is the region with the most Airbnb's in Lisbon, with 3271 listings. Cedofeita - Porto is the region with the most Airbnb's in Porto and between Lisbon and Porto, with 6987 listings. The next three regions with the most listings are in the district of Lisbon, in Misericórdia, Arroios and Cascais e Estoril.

4.4. PRICES INFORMATION

An analysis of prices allows us to verify that Lisbon has a higher average price than Porto, with an average of one hundred fifty-five dollars and one hundred thirty-three dollars respectively. The price does have significant differences according to the score, but above score two the price shows more variations. Seasonality indicates that the average price tends to increase, and that the average price in the month in which the listing is extracted is lower than in the following months.

If we look at the two regions of Lisbon and Porto separately, the price evolution trend is slightly different. In Lisbon, there is an increase between the first and third quarters and a decrease in the fourth quarter, while in Porto the average remains the same between the first and third quarters and increases in the fourth quarter. Lisbon has an increase between the first and third quarters and a decrease in the fourth quarter, while in Porto the average remains the same between the first and third quarters and increases in the fourth quarter.

When analysing the ten accommodations with the highest prices in the year, Lisbon presents a range between five thousand dollars and ten thousand three hundred dollars, while Porto presents a range between three thousand nine hundred dollars and ten thousand dollars, which also reflects the difference in average prices. The Lisbon region with the highest average price is Sintra and the lowest average is Amadora. In Porto, Gondomar has the highest average and São João da Madeira the lowest.



Figure 22 - Portugal – Prices

4.5. ANSWER BUSINESS QUESTIONS

At the beginning of the project, several questions were defined to which the answers should be determined. The answers to these questions are presented in Table 10.

Table 10 - Answer business questions

Business Question	Business Needs	Answers
Prices relations with Airbnb characteristics, reviews, and hosts.	Q1. Does region affect the value of accommodations?	A1. Yes, there is a difference between regions, as Lisbon has a higher average price than Porto, Figure 22.
	Q2. Is there a relationship between the average price of accommodation and the score?	A2. No, there is a price peak at a score of 3.29, but there is no increase in the average price at higher scores, Figure 22.
	Q3. What are the 10 accommodations with the highest average price in Lisbon and Porto?	A3. 10 Accommodation with higher average price: Villa in Sintra, Hotel in Porto, Apartment in Porto, Boutique Hotel in Lisbon, Serviced Apartment in Vila Nova de Gaia - 9 beds 4 baths, Serviced Apartment in Vila Nova de Gaia - 4 beds 2 baths, Serviced Apartment in Vila Nova de Gaia - 5 beds 2 baths, Boutique Hotel in Lisbon, Home in Valadares,
	Q4. What is the maximum and average price per quarter? In which quarter of the year is the average price the highest?	

Business Question	Business Needs	Answers
		Boutique Hotel in Lisbon, Figure 21 and Figure 22. A4. In the last quarter of 2023, the maximum price is 15k and the average price is 147.08, in the first quarter of 2024 we have 15k maximum and 147.46 average, in the second of 2024 15k and 147.56, in the third of 2024 15k and 147.6 and in the last of 2024 15k and 147.61. We verified that the average price is increasing and that the period in which the price was lowest was in December 2023, the extraction reference date, Figure 22.
Portuguese Airbnb characteristics.	Q5. What is the most common type of property in Portugal and by Lisbon/Porto region? Q6. What is the most common type of accommodation in Portugal and by region Lisbon/Porto?	A5. The most common property type in Portugal is the Entire Rental Unit, which is the same for Lisbon and Porto, Figure 17. A6. The most common accommodation type in Portugal is Entire Home/Apt, which is the same for Lisbon and Porto, Figure 17.
Airbnb differences between Lisbon and Porto.	Q7. Which city has the most properties in Portugal and by region Lisbon/Porto? Q8. What is the average price of accommodation in Portugal and by region? Q9. What is the average price of accommodation in Portugal and by region? Q10. How many accommodations and hosts are there in Portugal and by region of Lisbon/Porto?	A7. The region in Portugal with most properties is Cedofeita. In Lisbon is Santa Maria Maior and in Porto Cedofeita, Figure 17. A8. The average price in Portugal is 147.54, in Lisbon is 155.66 and Porto 133.95, Figure 17. A9. Portugal has about 36k Airbnb's and 14 hosts. Lisbon has about 22k to about 9k hosts and Porto has 13k to 5k, Figure 17.
The evolution of Airbnb and the differences between old and new hosts.	Q10. What is the variation in host registrations detected over the past 5 years? Q11. In terms of Airbnb's characteristics, are there significant differences when analysing older host registration years?	A10. In recent years, we have seen several peaks, one of which, with a lower number of registrations, coincided with the year with the greatest impact of the COVID 19 pandemic. Since May 2022, there has been an increase in the number

Business Question	Business Needs	Answers
		of registrations, with some fluctuations, Figure 19. A11. Not answered.

5. CONCLUSIONS, LIMITATIONS AND FUTURE WORKS

The main objective of this project was to analyse Airbnb's in Portugal using Microsoft Fabric, to identify patterns and make comparisons between the regions of Lisbon and Porto. In order to understand the context and ensure an adequate and structured follow-up of the project, an analysis of tourism in Portugal and Airbnb was carried out and then the research questions were defined.

With the initial objective defined and the questions that set the conductor for the investigation, a process of investigation of related literature was followed, which allowed us to bring good insights and better understand the panorama surrounding the topic addressed, as well as obtaining information about the dataset used in this project.

Then, following the Kimball lifecycle methodology, a structured process of knowledge of the dataset, collection, transformation and reading of data through Microsoft Fabric was created, which culminated in the creation of dashboards in Power BI, aimed at facilitating the reading of data and gaining insights from it.

Based on the data obtained, an analysis was carried out, relating it to the literature initially collected and providing answers to the questions raised, thus achieving the main objective, despite some perceived limitations. An important limitation is that the information is not extracted in real time, which can lead to discrepancies between the data and what is displayed on the website at that time. Also, the extraction at this point is manual, as it was necessary to click on the download links to extract the files.

In the development of the project, it is important to consider the time consumption, so the number of records extracted from the calendar file obtained from the Inside Airbnb source causes some performance problems when making changes to the factAirbnb_DataFlow data flow.

The information provided by the platform only presents data on hosts, existing Airbnb's, reviews and prices for the following year, and it would be important to also consider historical information on bookings made and values.

With respect to Airbnb's licenses, the information may be skewed, as there are cases where the license has a value of "0" or "00/AL," for example. However, we cannot confirm the accuracy of this information.

In future work, will be important to combine this analysis with data from Inside Airbnb, which contains information about past stays and prices. This work was done by manually extracting files from the platform and manually uploading them to the data lake, so it would be important for this work to automate this process so that manual intervention in the process is not necessary. It will be important to get more information about the data, find more ways to relate it, link the data to tourist information about Portugal, such as flights.

In the future, it will be interesting to develop a predictive data mining project that will show the expected evolution according to the data collected, allowing to understand what is expected and help in decision-making at both the host and tourism levels.

BIBLIOGRAPHICAL REFERENCES

- Airbnb. (2019, June 25). Airbnb reinventa as Viagens de Luxo com a Airbnb Luxe. Airbnb Newsroom. <https://news.airbnb.com/pt/airbnb-reinventa-los-viajes-de-lujo-con-airbnb-luxe/>
- Airbnb | Vacation rentals, cabins, beach houses, & more. (n.d.). Airbnb. Retrieved 16 March 2024, from https://www.airbnb.com/rooms/sw_skeleton
- Airbnb Logo png | PNGEgg. (n.d.). Retrieved 14 February 2024, from <https://www.pngegg.com/pt/search?q=airbnb+Logo>
- Alsudais, A. (2021b). Incorrect data in the widely used Inside Airbnb dataset. *Decision Support Systems*, 141, 113453. <https://www.sciencedirect.com/science/article/abs/pii/S0167923620302086?via%3Dihub>
- Alugueres de Condomínios e Casas de Veraneio. (2024, January 26). Airbnb. <https://www.airbnb.pt>
- Anselmi, G., Chiappini, L., & Prestileo, F. (2021). The greedy unicorn: Airbnb and capital concentration in 12 European cities. *City, Culture and Society*, 27, 100412. <https://www.sciencedirect.com/science/article/abs/pii/S1877916621000424?via%3Dihub>
- Aritenang, A. F., & Iskandar, Z. S. (2023). Tourism gentrification and P2P accommodation: The case of Airbnb in Bandung City. *City, Culture and Society*, 35, 100548. <https://www.sciencedirect.com/science/article/abs/pii/S1877916623000474?via%3Dihub>
- Batalha, M., Gonçalves, D., Peralta, S., & Pereira dos Santos, J. (2022). The virus that devastated tourism: The impact of covid-19 on the housing market. *Regional Science and Urban Economics*, 95, 103774. <https://www.sciencedirect.com/science/article/pii/S0166046222000059?via%3Dihub>
- bradleyschacht. (2023, May 23). *Data warehouse tutorial—Introduction—Microsoft Fabric*. <https://learn.microsoft.com/en-us/fabric/data-warehouse/tutorial-introduction>
- Conversor de moeda | Banco de Portugal. (2024, January 31). <https://www.bportugal.pt/page/conversor-de-moeda>
- Da Graça Pardal Domingues Miranda, L. C. (2021, December 10). Risk management in data science projects in Portugal. <https://repositorio.iscte-iul.pt/handle/10071/23892>
- davidiseminger. (2023, February 27). Understand star schema and the importance for Power BI - Power BI. <https://learn.microsoft.com/en-us/power-bi/guidance/star-schema>
- Ding, K., Niu, Y., & Choo, W. C. (2023). The evolution of Airbnb research: A systematic literature review using structural topic modeling. *Heliyon*, 9(6), e17090.

[https://www.cell.com/heliyon/fulltext/S2405-8440\(23\)04298-6?_returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS2405844023042986%3Fshowall%3Dtrue](https://www.cell.com/heliyon/fulltext/S2405-8440(23)04298-6?_returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS2405844023042986%3Fshowall%3Dtrue)

Estevens, A., Cocola-Gant, A., López-Gay, A., & Pavel, F. (2023). The role of the state in the touristification of Lisbon. *Cities*, 137, 104275. <https://www.sciencedirect.com/science/article/pii/S0264275123000872?via%3Dihub>

Filieri, R., Milone, F. L., Paolucci, E., & Raguseo, E. (2023). A big data analysis of COVID-19 impacts on Airbnbs' bookings behavior applying construal level and signaling theories. *International Journal of Hospitality Management*, 111, 103461. <https://www.sciencedirect.com/science/article/pii/S027843192300035X?via%3Dihub>

Franco, S. F., & Santos, C. D. (2021). The impact of Airbnb on residential property values and rents: Evidence from Portugal. *Regional Science and Urban Economics*, 88, 103667. <https://www.sciencedirect.com/science/article/abs/pii/S0166046221000272?via%3Dihub>

French, S. (2024, January 11). Are Airbnbs More Cost-Effective than Hotels? NerdWallet. <https://www.nerdwallet.com/article/travel/are-airbnbs-more-cost-effective-than-hotels>

Get the Data. (2023, September 11). <http://insideairbnb.com/get-the-data/>

Gomes, A. F. C. (2022). Is Portugal an accessible tourist destination? Factors that influence the perception of disabled people regarding Portugal as an accessible tourist destination.

Google Maps. (n.d.). Google Maps. Retrieved 20 February 2024, from <https://www.google.com/maps/search/centro+historico+lisboa/@38.7380973,-9.1881572,13z/data=!3m1!4b1?authuser=0&entry=ttu>

Gouveia, Feliz. (2021). *Base de Dados Fundamentos e Aplicações* (2nd ed.). FCA - Editora de Informática, Lda.

gsaurer. (2023a, June 30). What is Microsoft Fabric—Microsoft Fabric. <https://learn.microsoft.com/en-us/fabric/get-started/microsoft-fabric-overview>

gsaurer. (2023b, November 15). O que é o Microsoft Fabric—Microsoft Fabric. <https://learn.microsoft.com/pt-pt/fabric/get-started/microsoft-fabric-overview>

Gyódi, K., & Nawaro, Ł. (2021). Determinants of Airbnb prices in European cities: A spatial econometrics approach. *Tourism Management*, 86, 104319. <https://www.sciencedirect.com/science/article/pii/S0261517721000388?via%3Dihub>

He, J., Li, B., & Wang, X. (Shane). (2023). Image features and demand in the sharing economy: A study of Airbnb. *International Journal of Research in Marketing*. <https://www.sciencedirect.com/science/article/pii/S0167811623000174?via%3Dihub>

HERTE E. & BADULESCU A. (2023). RESEARCH ON THE DISTRIBUTION OF AIRBNB ACCOMMODATION IN MAJOR EUROPEAN DESTINATIONS. *Annals of the University of Oradea, Economic Science Series*, 32(1), 154–164. DOI: 10.47535/1991AUOES32(1)011

Inside Airbnb Data Dictionary. (2022, August 21). Google Docs. https://docs.google.com/spreadsheets/d/1iWCNJcSutYqpULSQHINyGlnUvHg2BoUGoNRIGa6Szc4/edit?usp=sharing&usp=embed_facebook

Janssens, B., Bogaert, M., & Van den Poel, D. (2021). Evaluating the influence of Airbnb listings' descriptions on demand. *International Journal of Hospitality Management*, 99, 103071. <https://www.sciencedirect.com/science/article/abs/pii/S0278431921002140?via%3Dihub>

jianleishen. (2023a, July 17). Excel format in Azure Data Factory—Azure Data Factory & Azure Synapse. <https://learn.microsoft.com/en-us/azure/data-factory/format-excel>

jianleishen. (2023b, September 20). How to copy data using copy activity—Microsoft Fabric. <https://learn.microsoft.com/en-us/fabric/data-factory/copy-data-activity>

Kimball DW/BI Lifecycle Methodology. (2024, July 15). Kimball Group. <https://www.kimballgroup.com/data-warehouse-business-intelligence-resources/kimball-techniques/dw-bi-lifecycle-method/>

LANDY, B. (2021). PLANET AIRBNB. *Fast Company*, 244, 68–107.

Kadi, J., Plank, L., & Seidl, R. (2019). Airbnb as a tool for inclusive tourism? *Tourism Geographies*, 24(4–5), 669–691. <https://www.tandfonline.com/doi/full/10.1080/14616688.2019.1654541>

Kovacs, L. (2023, April 11). Como funciona Airbnb para proprietários? *Tecnoblog*. <https://tecnoblog.net/responde/como-funciona-airbnb-para-proprietarios/>

Lee, S., & Kim, H. (2023). Four shades of Airbnb and its impact on locals: A spatiotemporal analysis of Airbnb, rent, housing prices, and gentrification. *Tourism Management Perspectives*, 49, 101192. <https://www.sciencedirect.com/science/article/abs/pii/S2211973623001204?via%3Dihub>

Luxury stay in Lisboa, Portugal. (n.d.). Airbnb. Retrieved 16 March 2024, from <https://www.airbnb.com/rooms/36517153>

Mahyoub, M., Ataby, A. A., Upadhyay, Y., & Mustafina, J. (2023). AIRBNB Price Prediction Using Machine Learning. 2023 15th International Conference on Developments in eSystems Engineering (DeSE), 166–171. <https://doi.org/10.1109/DeSE58274.2023.10099909>

- Mateus, C., & Mateus, I. B. (2021). Does a VAT rise harm the tourism industry? Portuguese evidence. *Tourism Management*, 83, 104234. <https://www.sciencedirect.com/science/article/abs/pii/S0261517720301606?via%3Dihub>
- Mody, M. A., Jung, S., Dogru, T., & Suess, C. (2022). How do consumers select between hotels and Airbnb? A hierarchy of importance in accommodation choice. *International Journal of Contemporary Hospitality Management*, 35(4), 1191–1218. <https://doi.org/10.1108/IJCHM-12-2021-1551>
- Moro, T. R. D. (2022). Data science na prática: Transformando problemas de negócio em soluções de dados. <https://lume.ufrgs.br/handle/10183/245536>
- NelGson. (2023, November 15). Ciência de dados no Microsoft Fabric—Microsoft Fabric. <https://learn.microsoft.com/pt-pt/fabric/data-science/data-science-overview>
- Pandit, P. (2022, June 20). 10 most peaceful countries in 2022 Global Peace Index. Vision of Humanity. <https://www.visionofhumanity.org/10-most-peaceful-countries-in-the-world-in-2022/>
- PennyZhou-MSFT. (2023, November 15). Crie seu primeiro pipeline de dados para copiar dados—Microsoft Fabric. <https://learn.microsoft.com/pt-pt/fabric/data-factory/create-first-pipeline-with-sample-data>
- periclesrocha. (2023, September 18). Ingesting data into the warehouse—Microsoft Fabric. <https://learn.microsoft.com/en-us/fabric/data-warehouse/ingest-data>
- PORDATA - Alojamentos turísticos: Total e por tipo de alojamento. (2023a, September 21). <https://www.pordata.pt/municipios/alojamentos+turisticos+total+e+por+tipo+de+alojamento-746>
- PORDATA - Alojamentos turísticos: Total e por tipo de estabelecimento. (2023, September 22). <https://www.pordata.pt/db/portugal/ambiente+de+consulta/tabela>
- PORDATA - Alojamentos turísticos: Total e por tipo de estabelecimento. (2023b, September 21). <https://www.pordata.pt/portugal/alojamentos+turisticos+total+e+por+tipo+de+estabelecimento-2562>
- PORDATA - Balança de viagens e turismo em % do PIB. (2023, September 22). <https://www.pordata.pt/db/portugal/ambiente+de+consulta/tabela>
- PORDATA - Conheça o seu Município. (2023, September 21). <https://www.pordata.pt/municipios>

PORDATA - Dormidas nos alojamentos turísticos por 100 habitantes. (2023, September 21). <https://www.pordata.pt/municipios/dormidas+nos+alojamentos+turisticos+por+100+habitantes-761>

PORDATA - Dormidas nos alojamentos turísticos: Total e por país de residência do hóspede. (2023, September 21). <https://www.pordata.pt/portugal/dormidas+nos+alojamentos+turisticos+total+e+por+pais+de+residencia+do+hospede-2611>

PORDATA - Empresas: Total e por setor de atividade económica. (2023, December 18). <https://www.pordata.pt/portugal/empresas+total+e+por+setor+de+atividade+economica-2856-246175>

PORDATA - Tráfego de passageiros nos aeroportos: Lisboa, Porto e Faro. (2023, August 28). <https://www.pordata.pt/db/portugal/ambiente+de+consulta/tabela>

Secretaria Geral da Economia - Portugal é o 5o país com maior contributo do turismo para o PIB. (2018). <https://www.sgeconomia.gov.pt/noticias/portugal-e-o-5-pais-com-mais-forte-contributo-do-turismo-para-o-pib.aspx>

Portugal entre os seis países mais seguros do mundo. (2022, June 22). <https://www.turismodeportugal.pt/pt/Noticias/Paginas/portugal-entre-os-seis-paises-mais-seguros-do-mundo.aspx>

prlangad. (2023, May 23). Create a Warehouse—Microsoft Fabric. <https://learn.microsoft.com/en-us/fabric/data-warehouse/create-warehouse>

Quirós, C., Riveros, K., & Erazo, R.-F. (2023). Airbnb vs booking: Determinants of the use of collaborative and traditional online accommodations. *Applied Economics Letters*, 30(10), 1369–1373. <https://www.tandfonline.com/doi/full/10.1080/13504851.2022.2055737>

Rental unit in Porto · ★4.86 · 1 bedroom · 1 bed · 1 bath. (n.d.). Airbnb. Retrieved 16 March 2024, from <https://www.airbnb.com/rooms/20428997>

Santos, Maribel yasmina, R., Isabel. (2017). *Business Intelligece da informação ao conhecimento* (3rd ed.). FCA - Editora de Informática, Lda.

Shabrina, Z., Arcaute, E., & Batty, M. (2021). Airbnb and its potential impact on the London housing market. *Urban Studies*, 59(1), 197–221. <https://doi.org/10.1177/0042098020970865>

Silva, P. R. da. (2021). *O futuro do turismo: O turismo de Portugal pós COVID* [masterThesis]. <http://repositorio.ulusiada.pt/handle/11067/6201>

Singh, N., & Azevedo, A. B. (2021). Geography of AirBnb in Barcelona and Lisbon: A comparative study. *Finisterra*, 56(118), Article 118. <https://revistas.rcaap.pt/finisterra/article/view/23725>

Tang, L., Xu, Z., & Lyu, X. (2023). More popular, more listings? Effects of popularity on Airbnb host expansion. *International Journal of Contemporary Hospitality Management*, 35(5), 1650–1669. <https://www.emerald.com/insight/content/doi/10.1108/IJCHM-03-2022-0276/full/html>

Teixeira, M. R. M. (n.d.). A Economia da Partilha: O caso Airbnb.

Termos de Serviço—Centro de Ajuda da Airbnb. (2023, January 25). Airbnb. <https://www.airbnb.pt/help/article/2908>

Terms of Service—Airbnb Help Centre. (2023, January 25). Airbnb. <https://www.airbnb.com/help/article/2908>

Törnberg, P. (2022, April 21). How sharing is the “sharing economy”? Evidence from 97 Airbnb markets | PLOS ONE. <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0266998>

Tran, T. H., & Filimonau, V. (2020). The (de)motivation factors in choosing Airbnb amongst Vietnamese consumers. *Journal of Hospitality and Tourism Management*, 42, 130–140. <https://doi.org/10.1016/j.jhtm.2019.10.011>

TRAVEL & TOURISM ECONOMIC IMPACT 2018 PORTUGAL - Relatório do World Travel & Tourism Council (WTTC). (2018). https://www.sgeconomia.gov.pt/ficheiros-externos-sg/wttc_portugal2018-pdf.aspx

USD to EUR Exchange Rate. (2024, February 5). Bloomberg.Com. <https://www.bloomberg.com/quote/USDEUR:CUR>

Villa in Barcarena · ★4.79 · 2 bedrooms · 2 beds · 1 bath. (n.d.). Airbnb. Retrieved 16 March 2024, from <https://www.airbnb.com/rooms/41304319>

Xue, L., Leung, X. Y., & Ma, S. (David). (2022). What makes a good “guest”: Evidence from Airbnb hosts’ reviews. *Annals of Tourism Research*, 95, 103426. <https://www.sciencedirect.com/science/article/abs/pii/S0160738322000779?via%3Dihub>

Zheng, Q., Ke, D., & Li, X. (2023). Optimal pricing strategies for Airbnb when competing with incumbents on two sides. *International Journal of Hospitality Management*, 108, 103392. <https://www.sciencedirect.com/science/article/abs/pii/S0278431922002584?via%3Dihub>

APPENDIX A

Appendix A.1. Calculated columns

Calculated columns in table dimHost:

- YearSinceHost = Text.From(Date.Year([host_since]))
- MonthSinceHost = Text.From(Date.Month([host_since]))

Calculated columns in table dimLocation:

- Country = "Portugal"
- District = if [Lisbon_Porto]="1" then "Lisbon" else "Porto"
- LocationID = Index Column

Appendix A.2. Hierarchies

- dimDate: Day → Month → Trimester → Semester → Year
- dimLocation: Neighbourhood → Neighbourhood_group → District → Country
- dimDetail: Beds → Accomodates → Room_type → Property_type
- dimHost: host_since → MonthHost → YearHost

Appendix A.3. Measures in factAirbnb table

- AVGPrice = AVERAGE(factAirbnb[price])
- AVGREviewLocation = AVERAGE(dimDetail[review_scores_location])
- AVGREviewRating = AVERAGE(dimDetail[review_scores_rating])
- AVGREviewValue = AVERAGE(dimDetail[review_scores_value])
- MinPrice = MIN(factAirbnb[price])
- MaxPrice = MAX(factAirbnb[price])
- NrAirbnb = DISTINCTCOUNT(factAirbnb[listing_id])
- NrHosts = DISTINCTCOUNT(factAirbnb[host_id])
- NrHostsYearLess1 = CALCULATE(COUNT(dimHost[host_id]),
CONVERT(dimHost[YearHost], INTEGER)=CONVERT(Year(TODAY())-2, INTEGER))

Appendix A.4. KPI Hosts number

Value: NrHosts

Tendency: YearHost

Destination: NrHostsYearLess1

