JANUARY 2020

AIRBNB DASHBOARD

DATA VISUALIZATION
NOVA IMS

Ana St. Aubyn | r2016713
André Queda | r2016658
Manuel Viegas | r2016674
Miguel Cordeiro | r2016702

1. INTRODUCTION

This project used data about Airbnb in three cities: **Lisbon, Amsterdam and Paris**. The goal is to study the way the platform is being used in cities of Europe. We wanted to see the way the offer for housing is dispersed through the cities, the different kinds of options and how the pricing and rating differs from city to city.

In order to get the data, we used the **public repository for Airbnb data – Inside Airbnb**. Inside Airbnb is an independent, non-commercial set of tools and data that allows you to explore the way the housing platform works in cities around the world. The data was retrieved in November 2019.

To see the dashboard just follow this *link*. The correspondent GitHub repository can be accessed through this *link*.

2. DATA

We downloaded three datasets for each city – **calendar, neighborhoods and listings**. The first one has data that varies through time, so we used it to estimate the average price of housing in each month for each city. Neighborhoods is a dataset that only contains the name of all the neighborhoods that exist in each city. Lastly, the listings dataset contains most of the information. A brief description of the variables in this dataset provided in the following table (table 1):

VARIABLE	TYPE	DESCRIPTION
id	int64	Listing ID
name	object	Name of the listing as it is in the website
host_id	int64	Host ID
host_since	object	Date in which the host started renting
host_response_time	object	Categories describing how fast the host is answering to messages
host_response_rate	int32	Percentage describing how fast the host is answering to messages
host_is_superhost	bool	Binary variable that tells if the host is a superhost (1) or not (0)
host_total_listings_count	int32	Number of listings owned by the host
neighbourhood	object	Name of the neighborhood where the listing is inserted
longitude	float64	Longitude of the place where the listing is
latitude	float64	Latitude of the place where the listing is
is_location_exact	bool	Binary variable that tells if the location of the listing is exact (1) or not (0)
room_type	object	Categories for the different types of housing available
rating	float64	Average rating based on the reviews of the listing
price	float64	Average nightly price for the whole year
score_communication	float64	Average rating for communication only
score_location	float64	Average rating for location only
score_clean	float64	Average rating for cleanliness only

Table 1 - Variable Description For Listings

When it comes to the calendar data, it only contains three variables: *listing_id*, *month* and *price*. The first one, matches the *id* existent in the listings dataset and then, it has the price for each listing for each month of the year.

3. INSPIRATION

In November of 2019, some members of this project group attended a **workshop** at NOVA IMS where Airbnb data was provided as an example. This workshop was hosted by SDG Group and was called From Data To Insights, using Qlik Sense to build a dashboard.

Working with data about tourism is very interesting, as this sector is growing throughout the whole world. In 2019, between January and September, international tourist arrivals rose by four million[1]. Keeping this in mind, subsectors related to tourism like housing are probably also rising.

After seeing the amount of freely available data about Airbnb and the theme's timeliness, we started searching for dashboards using this data. The first ones we found were made with Tableau:

- *https://public.tableau.com/en-us/s/blog/2015/07/analyzing-airbnb-data*
- *https://public.tableau.com/profile/suwei.yang#!/vizhome/LowFiDec3_V161/Dashboard4*

Pretty much everything that can be done with Tableau, can also be done with Dash, so we took some ideas from the two dashboards mentioned above.

Then, we searched directly on the *Dash Gallery* to take some inspiration for the layout and callbacks.

4. LAYOUT

We decided to do a dashboard layout from scratch, not using any of the layouts available in the Plotly Dash Gallery. We used html components to divide the areas in our dashboard.

In order to provide a general view of the data before presenting the theme itself, we decided to present a map at the top of the dashboard. This way, the user can see the huge amount of airbnbs that exist and be aware of the dimension of the data.

Only then we present the title and a brief description of the project, alongside with the Airbnb logo. All the charts follow this description.

On the very top of the dashboard we have the name of the cities available. The user can click on these names and see data referring to each city.

[1] (2019, December 18). *Tourism's Rising Intensity*. Retrieved January 13 from *https://www.industryglobalnews24.com/tourisms-rising-intensity*

5. VISUALIZATIONS

To summarize the prices for the neighborhoods, we showed the information in the simplest way: we created a function that returns the name of the most expensive neighborhood (according to average prices) and another one that returns the name of the cheapest neighborhood. These names are shown in order for the user to know where are the most expensive and the cheapest listings.

As mentioned before, we wanted to see the evolution of the prices throughout the months. With this in mind, we created a **line chart** where the x axis is the months and the y axis is the average price of all the listings in that month. Months are represented by the three first letters of that month (e.g. JAN, FEB, etc.). The image below shows the layout of this chart.

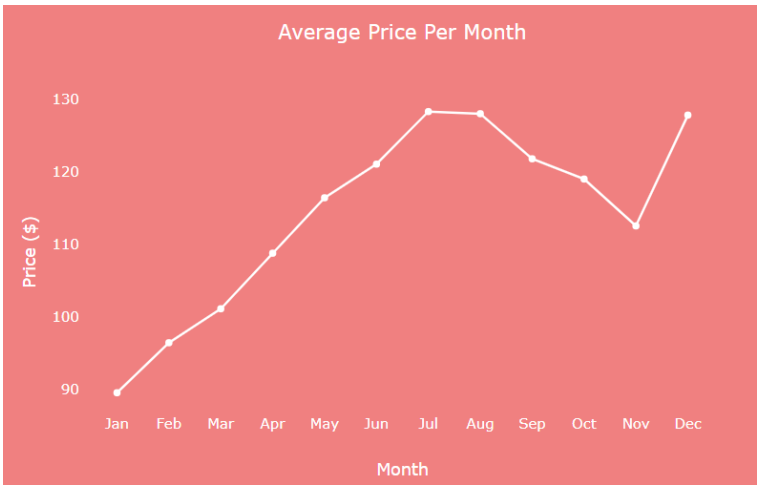

Figure 1 - Line Chart For Price and Months

As we can see from the figure above, and as expected, the highest prices occur during the high season (July and August for the summer and December for the New Year's Eve). Figure 1 represents data from Lisbon.

We also wanted to have an idea of how many different types of housing and how many listings for each kind exist. With this in mind, we created a **stacked bar chart** representing the percentage of each type of housing in the total number of listings. For Lisbon, we concluded that the type with the most listings is the Private Room.

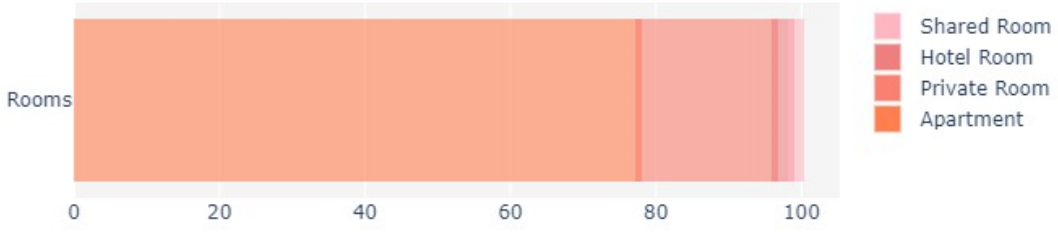

Figure 2 - Stacked Bar For Room Type

In order to investigate the average rating, we created a **radar chart** with all the three dimensions of the evaluation: communication, cleanliness and location. From this chart we can clearly see that, in Lisbon, the average ratings are very high.

Since we wanted to see the most popular neighborhoods, we created a **bar chart** where the x axis is the neighborhood and the y axis is the total number of listings for each neighborhood. We plotted the top 15 neighborhoods. As expected, historical ones are usually the most popular ones.

Finally, because we wanted to have a general panorama of the locations of the listings, we did a **map** where each dot represents a listing. The user can zoom in on the wanted area. Since there are a lot of listings, we made the dots a little transparent, which means that darker areas have more listings.

5. INTERACTIVITY

The map on the top of the dashboard can be zoomed and adjusted to the desired area. Also, when putting the cursor on top of a dot, information about the respective listing is shown: name, rating and price.

Also, the cursor shows information about the values in all the charts when placed on top of the bars, lines or any dot in any graph.

We also have on top of the dashboard, options for the viewer to select the city that he wants to analyze.

6. CONCLUSION

We wanted to make sure that this dashboard is directed for decision support. Keeping this in mind, we wanted to make it easy and intuitive to read so that the users can rapidly come to conclusions. We made it minimalist and clean so that nothing gets in the way of the visualization. Just by clicking on little buttons, the user can see a general panorama for one of the three cities available.

S I M P L I C I T Y I S K E Y

The first obvious conclusion we took out of this project is that there is a huge number of listings in each city, confirming our idea that tourism is growing.

The monthly price variation for each city differs a lot. However, they all have in common the higher values for the month of December. As mentioned before, the most expensive months for renting an Airbnb in Lisbon are the summery ones – July and August. In Paris, December is the most expensive month, followed by June. However, there is a big gap between the average value for December and June. In Amsterdam, the most expensive month is April, because of the King's Day. Amsterdam is also the most expensive city of the three analyzed.

Both for Paris and Amsterdam, the most expensive neighborhood is the most central one. However, in Lisbon, the most expensive neighborhood is actually on the surroundings, closer to other attractions like the Buddha Eden Garden.

In all of the three cities, the average ratings are pretty high (always above 9). In Paris, the lowest dimension of the rating is the cleanliness, with an average value of 9.2. The total average rating is 9.3. In Lisbon, the lowest dimension is also the cleanliness, with a value of 9.4. The same happens for Amsterdam.

The most common type of housing in all the three cities is the private apartment. In Lisbon, this kind of housing represents 77.6% of all the listings. It represents 86.8% in Paris and 69.9% in Amsterdam.

The neighborhood with most listings registered for Paris is Buttes-Montmarte. In Lisbon it is Santa Maria Maior. Both of these neighborhoods are historical ones in the center of the cities, so it makes sense that they are the most popular ones.

7. REFERENCES

(2019, December 18). *Tourism's Rising Intensity*. Retrieved January 13 from *https://www.industryglobalnews24.com/tourisms-rising-intensity*

Munzner, T. (2014) Visualization Analysis and Design. CRC Press

Dash User Guide. (2020). Retrieved 14 January 2020, from *https://dash.plot.ly/*