IBM - Applied Data Science Capstone
Capstone Project – Battle of the Neighborhoods

Airbnb properties in the District of Porto, Portugal

1. Introduction

1.1 Background

Airbnb is an online-based company that connects people looking for accommodation (Airbnb guests) to people looking to rent their properties (Airbnb hosts) on a short-term or long-term basis. The rentals properties include apartments (dominant), homes, boats, and whole lot more. Renters are presented with a good selection of listings and can filter by criteria like price, number of bedrooms, room type, and more. Tourists are mostly motivated to book Airbnb accommodations because of their low cost, convenient location, and household amenities. They are generally less motivated by the opportunity to interact with the host or other locals, or by the promise of an authentic, local experience. For hosts, participating in Airbnb is a way to earn some income from their property, but with the risk that the guest might do damage to it. For guests, the advantage can be relatively inexpensive accommodations, but with the risk that the property won't be as appealing as the listing made it seem.

1.2 Business Problem

Porto is a city located in the north of Portugal by the outlet of the Douro River. It has elegant neighborhoods and large villas sitting on narrow cobbled streets. There is a lot of properties available in the Airbnb platform in the District of Porto. Exploratory analysis could help determine the best places (cities around, neighborhood) to stay and minimize the options of properties. This project aims to conduct some exploratory data analysis using the Foursquare API and the Airbnb list of properties to produce a small list of properties to be analyzed based on the client requirements.

1.3 Target Audience

I am planning to spend a month or two in the District of Porto (Portugal) and I am looking for a reasonable accommodation from Airbnb. There are a lot of properties (approximately 12,000) and analyze each one available will spend many time. So, the

main objective of this analysis is to produce a small list of properties to be analyzed in the Airbnb web platform, helping the property decision process and saving time.

This would interest anyone who wants to travel to the cities around Porto, Portugal.

2. Dataset

2.1 Data sources

The dataset used for this project comes from Inside Airbnb: http://insideairbnb.com/get-the-data.html.
The dataset that was employed was named *listings.csv*; it is a detailed data set with 106 attributes, a few of the attributes being: price per day (which will hereafter be simply referred to as price), number of beds, property type, neighborhood, cleaning fee, security deposit, host's ratings score, etc.

The data contains a total of 12,005. Each row in the data set is a listing available for rental in Airbnb's site. The columns describe different characteristics of each listing (features). The geographic dataset named *neighborhood.geoson* (available in Inside Airbnb) was used to create exploratory maps of the location. It contains the coordinates of each neighborhood group present in the *listings.csv* file.

I've used the Foursquare API to explore neighborhoods in the District of Porto. The Foursquare explore function will be used to get the most common venue categories in each neighborhood. The following information were retrieved for the 8 top venues from each neighborhood group:

- Venue ID
- Venue Name
- Coordinates
- Categories Name

2.2 Feature selection

The Figure 1 shows the beginning of the data frame created from the listing.csv file. It was read 12,005 samples containing 106 features. Many of the features is not necessary for our analysis, so a selection of the 26 principal features was realized.

	id	listing_url	scrape_id	last_scraped	name	summary	space	description	experiences_offered	neighborh
0	41339	https://www.airbnb.com/rooms/41339	20200321155050	2020-03-21	Porto city flat near the sea	Here you'll find all you need for your holiday...	Apartment facing Southeast, with a big bedroom...	Here you'll find all you need for your holiday...	none	In the have Serr
1	55111	https://www.airbnb.com/rooms/55111	20200321155050	2020-03-21	Fontielas Houses |Floor1 in House with shared ...	First Floor in House with shared Swimingpool a...	The first floor in house with shared pool. I...	First Floor in House with shared Swimingpool a...	none	
2	70925	https://www.airbnb.com/rooms/70925	20200321155050	2020-03-21	APARTMENT WITH THE BEST CITY VIEW	Two separate bedrooms are an undeniable advan...	Apartment with the best view of the Porto city...	Two separate bedrooms are an undeniable advan...	none	The hous supermar
3	73828	https://www.airbnb.com/rooms/73828	20200321155050	2020-03-21	Fontielas Houses |Floor0 in House with shared ...	Piso no rés-do-chão em moradia com piscina par...	The first floor in house with shared pool. Ou...	Piso no rés-do-chão em moradia com piscina par...	none	
4	76436	https://www.airbnb.com/rooms/76436	20200321155050	2020-03-21	Go2oporto @ River Side	Elegant and modern apartment, facing south in ...	Comfortable space, in one of the most beautifu...	Elegant and modern apartment, facing south in ...	none	In this enjoy b

5 rows × 106 columns

Figure 1: Data frame created from listings.csv file.

Out of 106 features, 26 features were selected (Figure 2). A few of the important numerical features are:

- *accommodates*: the number of guests the rental can accommodate
- *bedrooms*: number of bedrooms included in the rental
- *beds*: number of beds included in the rental
- *price*: nightly price for the rental
- *minimum_nights*: minimum number of nights a guest can stay for the rental
- *maximum_nights*: maximum number of nights a guest can stay for the rental
- *review_scores_rating: score of reviews that previous guests have left*

A few of the important categorical features are:

- *property_type*: house, townhouse, apartment, condo, hostel, cabin, etc.
- *room_type*: entire home/apt, private room or shared room
- *neighbourhood_cleansed*: neighborhood e.g. Midtown, Harlem, Murray Hill, etc.
- *cancellation_policy*: 6 categories: super_strict_60, super_strict_30, strict_14_with_grace_period, strict, moderate, and flexible.

```
Data columns (total 26 columns):
id                                12005 non-null int64
neighbourhood_cleansed            12005 non-null object
neighbourhood_group_cleansed      12005 non-null object
city                              11899 non-null object
zipcode                           11634 non-null object
latitude                          12005 non-null float64
longitude                         12005 non-null float64
property_type                     12005 non-null object
room_type                         12005 non-null object
accommodates                      12005 non-null int64
bathrooms                         12001 non-null float64
bedrooms                          12000 non-null float64
beds                              11948 non-null float64
price                             12005 non-null object
guests_included                   12005 non-null int64
extra_people                      12005 non-null object
minimum_nights                    12005 non-null int64
maximum_nights                    12005 non-null int64
review_scores_rating              10053 non-null float64
review_scores_accuracy            10046 non-null float64
review_scores_cleanliness         10048 non-null float64
review_scores_checkin             10042 non-null float64
review_scores_communication       10045 non-null float64
review_scores_location            10043 non-null float64
review_scores_value               10042 non-null float64
cancellation_policy               12005 non-null object
dtypes: float64(12), int64(5), object(9)
memory usage: 2.4+ MB
```

Figure 2: Features selected for this analysis.

3. Requirements

Considering the problem, a list of requirements was created to guide the analysis. The accommodation must be in a great location, considering the proximity to bakeries, restaurants and cafes.

Requirements:

- Entire home/apt
- Neighborhood with mean price by night bellow $100
- Properties in neighborhoods that presents a good offer of Café, Bakery and Restaurant
- The most common type of properties in the selected neighborhoods
- Only properties with review score rating equal 100
- Properties that accommodate 3 persons with 2 bedrooms

4. Exploratory Data Analysis

The first requirement is not to share a property, selecting only the room type of 'Entire home/apt'. After this selection, the dataset was reduced from 12,005 to 9,285 samples. The analyzed was conducted using the feature *neighbourhood_group*, which also comprises the cities around Porto.

4.1 Prices by neighborhood

Figure 3 shows the boxplot of the rental prices by night grouped by neighborhood. It is possible observe that there are a lot of expensive properties in Porto and Vila Nova de Gaia neighborhood. The both are the most wanted places to stay when travel to Porto, according to the blog and travel websites. However, we need a place to stay for a long period (1 or 2 months) and because of that cheaper places are preferred.

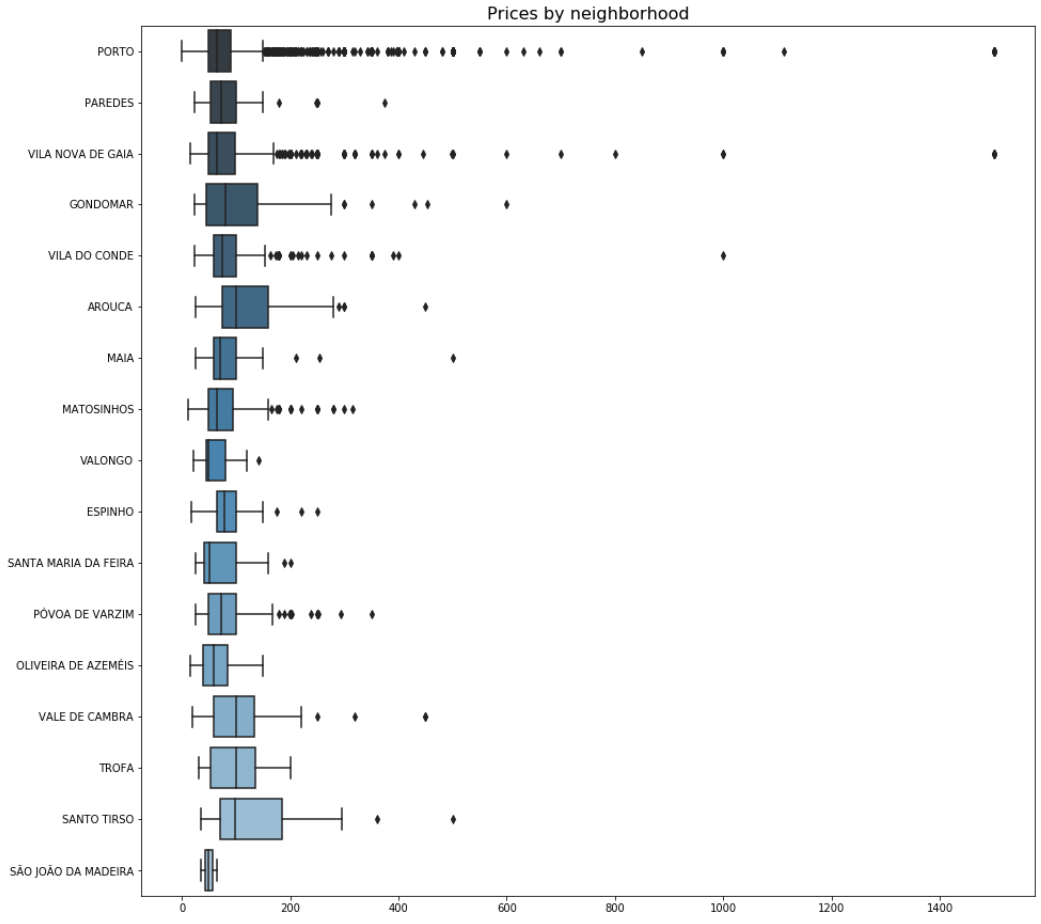

Figure 3: Boxplot of prices in each neighborhood. The mean lines represent the median price, the width is related to range of price and the points represent the supposed outliers.

The mean prices by neighborhood are shown in Figure 4. According to the requirements, only neighborhoods with mean rental prices less than 100 by night must to be analyzed. The Figure 5 shows the selection of neighborhood with mean prices below 100. In Figure 6, it is presented a colored map to show the mean prices by region.

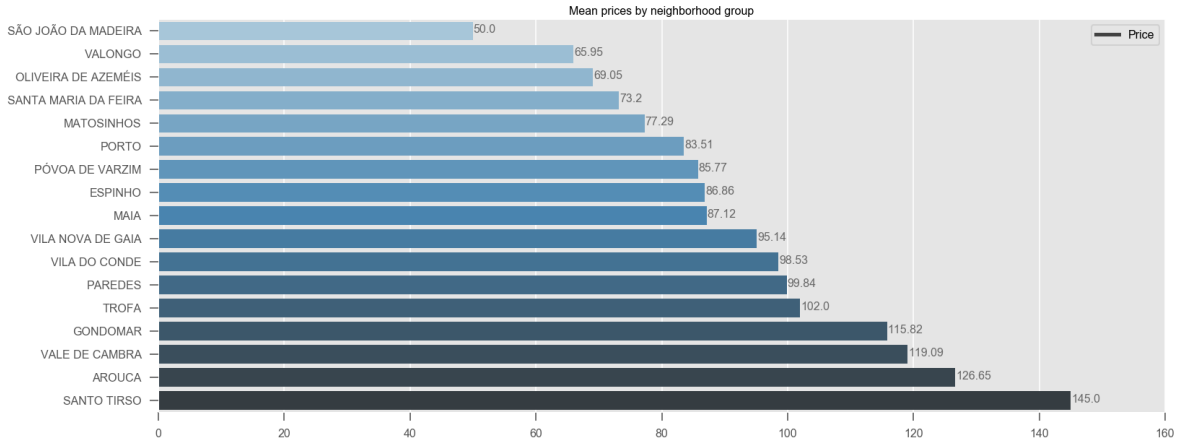

Figure 4: Mean rental prices by neighborhood.

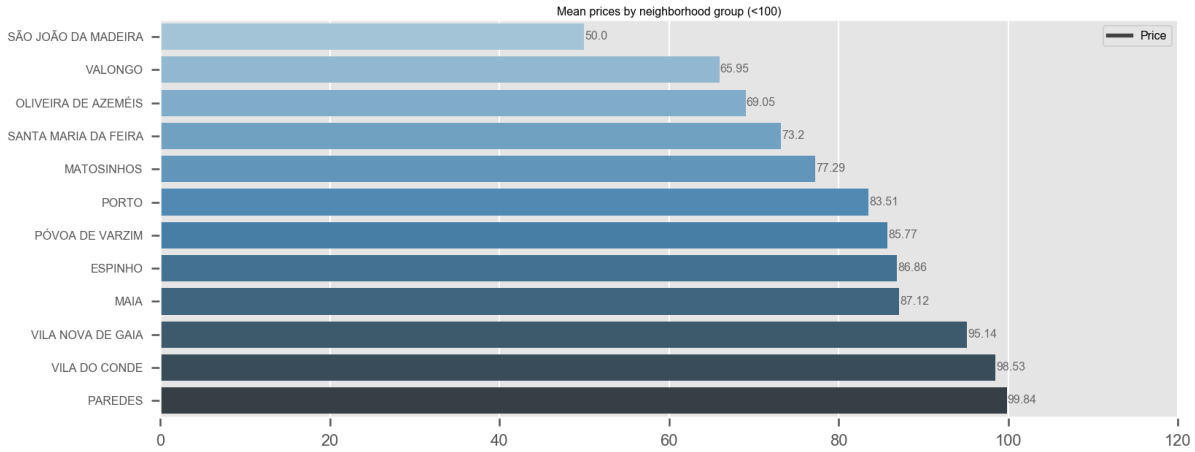

Figure 5: Mean rental prices less than 100 by neighborhood.

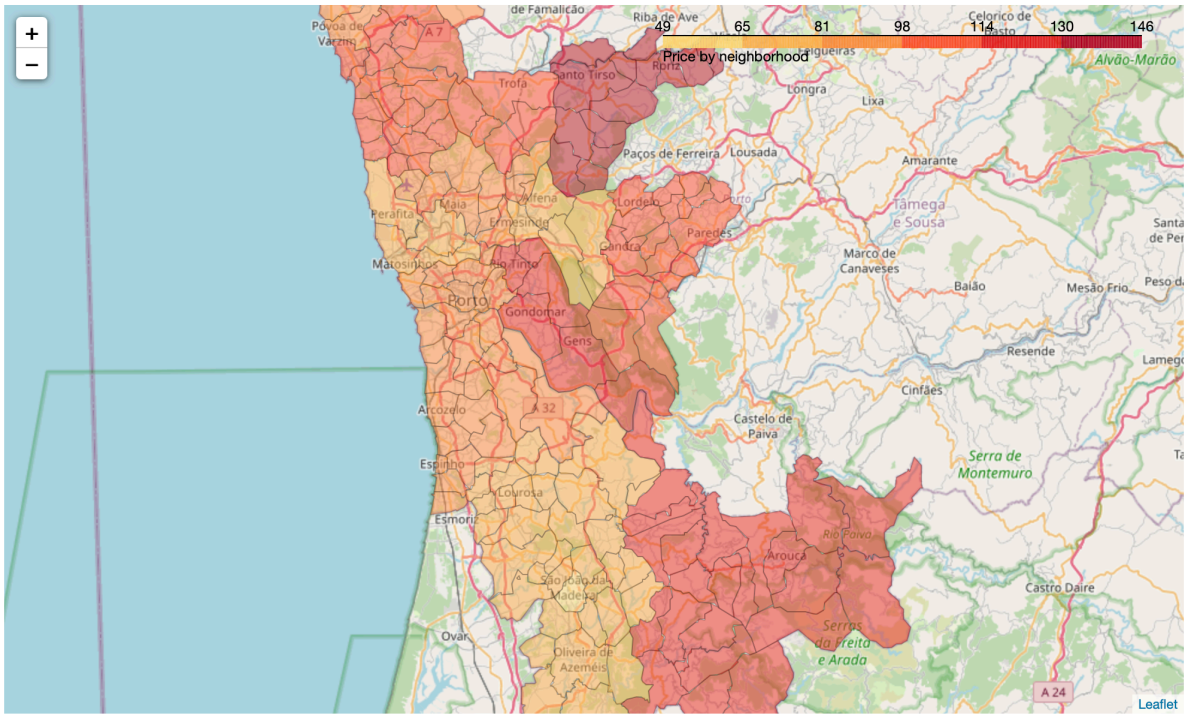

Figure 6: Color map of the mean rental prices of each neighborhood. The highest mean prices are represented by the darkest orange.

4.2 Foursquare exploration

Using the location of each neighborhood, we explore the top venues of them through the Foursquare API. We retrieved information about venues and analyzed the venue category frequency. Table 1 shows the frequency of the category venues by neighborhood. Analyzing the requirements about the amenities, we select the two neighborhoods that better satisfy the conditions: Matosinhos and Vila do Conde. From now, only these two neighborhoods will be considered in the analysis.

Table 1: Venue category frequecy by neighborhood.

	Café	Bakery	Restaurant	Vegan Restaurant	Portuguese Restaurant
Espinho	0.25	0.14	0.11	0	0
Maia	0	0.10	0.15	0	0
Matosinhos	0.25	0.25	0	0	0
Paredes	0	0	0	1	0
Porto	0.04	0.03	0.07	0	0.12

Póvoa de Varzim	0.12	0	0.25	0	0
São João da Madeira	0.14	0.14	0.07	0	0.14
Vila do Conde	0.17	0.17	0.17	0	0
Vila Nova de Gaia	0	0	0	0	0

4.3 Property types

Analyzing the property types of all neighborhoods (Figure 7), Matosinhos (Figure 8) and Vila do Conde (Figure 9), we see that mostly of the properties available for rent by Airbnb in the District of Porto consists of apartments. Considering this, we selected only the apartments in the Matosinhos and Vila do Conde neighborhood.

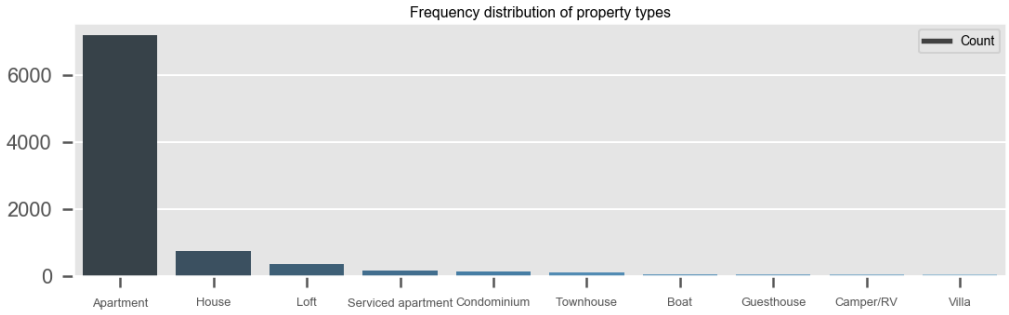

Figure 7: Property types distribution in the District of Porto.

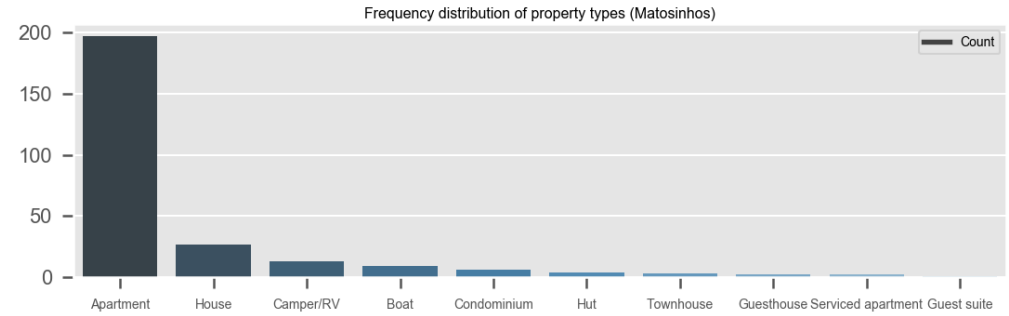

Figure 8: Property types distributions in the Matosinhos neighborhood.

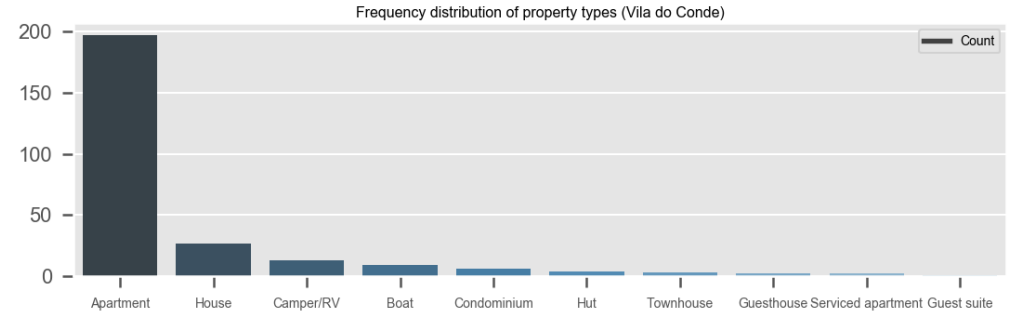

Figure 9: Property types distributions in the Vila do Conde neighborhood.

4.4 Review scores of properties

The distribution of the review scores rating from Matosinhos and Vila do Conde are presented in Figures 10 and 11. First, we select only the apartments that presented review scores rating different from NaN. Mostly of the apartments received scores between 90 and 100. Considering the requirement, only apartments that received score equal 100 were selected.

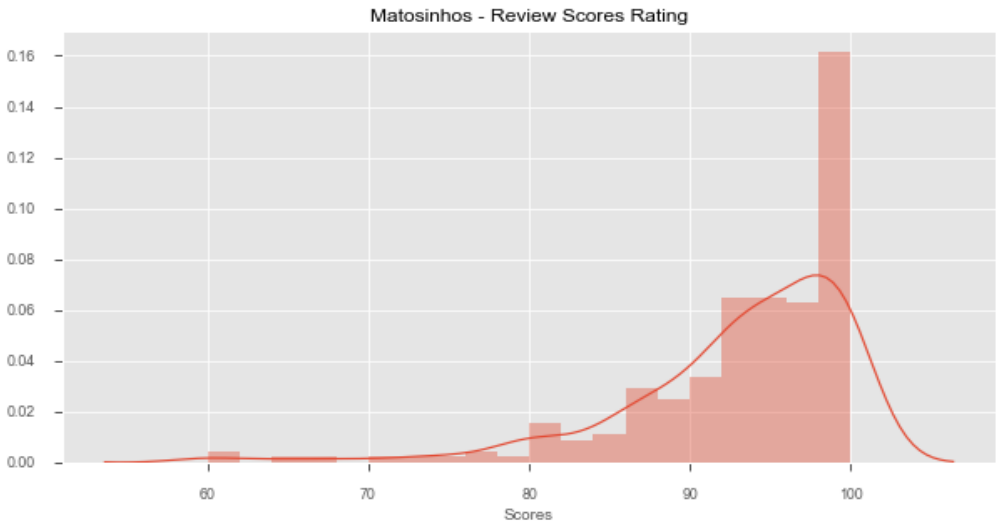

Figure 10: Review scores rating of properties in Matosinhos neighborhood.

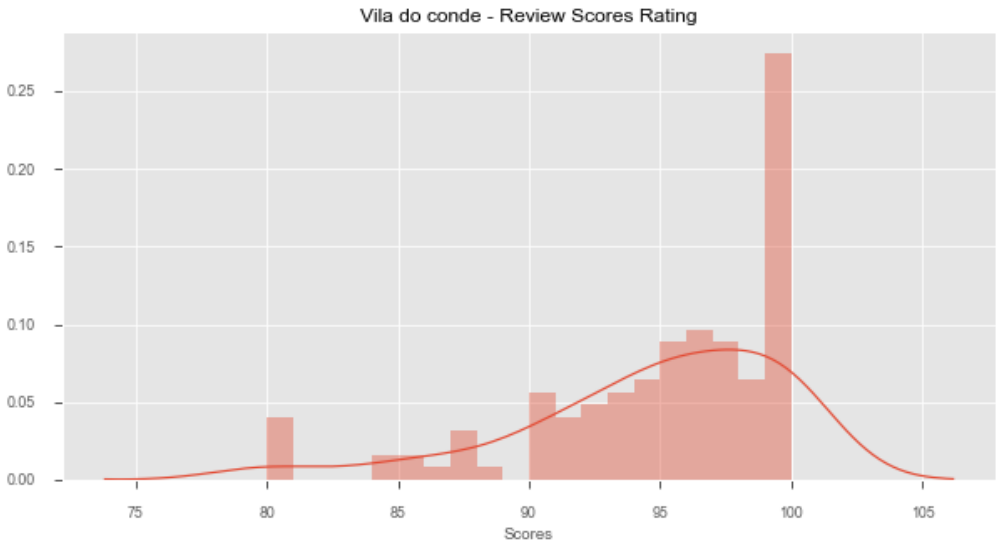

Figure 11: Review scores rating of properties in Vila do Conde neighborhood.

4.5 Selection of properties

The requirements were applied to the final data frame, as showed in Figure 12. As a result, a list of five properties was defined (Figure 13), being 3 in Matosinhos and 2 in Vila do Conde neighborhood.

```
#Cancellation policy = flexible or moderate
#Maximum_nights > 30
#Bedrooms >=2
#Price/guest <= 30

df_matosinhos_final = df_matosinhos[((df_matosinhos['cancellation_policy'] == 'flexible') |
                                     (df_matosinhos['cancellation_policy'] == 'moderate'))  &
                                    (df_matosinhos['maximum_nights'] > 30) &
                                    (df_matosinhos['bedrooms'] >= 2) &
                                    (df_matosinhos['prices/guest'] <= 30)]

df_matosinhos_final
```

Figure 12: Selection of dataset.

	id	neighbourhood_cleansed	neighbourhood_group	city	zipcode	latitude	longitude	property_type	room_type	accommodates	bathrooms	b
0	36166471	São Mamede de Infesta e Senhora da Hora	MATOSINHOS	São Mamede de Infesta	4465	41.18306	-8.62918	Apartment	Entire home/apt	5	2.0	
1	36525996	Matosinhos e Leça da Palmeira	MATOSINHOS	Leça da Palmeira	4450	41.19487	-8.69144	Apartment	Entire home/apt	4	3.0	
2	39226932	Matosinhos e Leça da Palmeira	MATOSINHOS	Matosinhos	4450-010	41.17981	-8.68333	Apartment	Entire home/apt	7	2.0	
3	12560419	Árvore	VILA DO CONDE	Árvore, Vila do Conde	4480-113	41.33326	-8.72321	Apartment	Entire home/apt	5	2.0	
4	20710821	Vila do Conde	VILA DO CONDE	Vila do Conde	4480	41.36160	-8.75744	Apartment	Entire home/apt	5	1.5	

Figure 13: Final list of the selected properties.

5. Conclusions

The purpose of this project was to identify Porto properties and neighborhoods to spend a month or two in the District of Porto.

By calculating amenities density distribution from Foursquare data, we have selected 2 neighborhoods that satisfy the requirements. Combining this result with Airbnb dataset, a list of five properties was produced to be analyzed:

https://www.airbnb.com/rooms/12560419

https://www.airbnb.com/rooms/20710821

https://www.airbnb.com/rooms/36166471

https://www.airbnb.com/rooms/36525996

https://www.airbnb.com/rooms/39226932

Analyzing the five properties resulted from the exploratory analysis of Airbnb and Foursquare data, I selected a property to stay during my travel to the District of Porto, showed in Figure 14.