

The Battle of Neighborhoods

Capstone Project

Daniela Moniz

1. Introduction

Over the last years, the Portugal's housing market has reached highs with an increase of general house's price of about 9% on an annual basis, mainly in the capital city (Lisbon). Therefore, the main goal of this project is to help those who are looking to buy a house in Lisbon. In this context, an analysis of the variation of houses' costs over the different Lisbon's parishes is considered as well as a differentiation of the places as being more residential or commercial/touristic based on the venues nearby.

2. Data

The information of the Portugal's housing market is based on a Kaggle dataset (Lisbon House Prices) that includes the house price and the price m² for the different parishes in Lisbon. The main features of dataset are presented in Fig.1.

	Id	Condition	PropertyType	PropertySubType	Bedrooms	Bathrooms	AreaAllt	AreaGross	Parking	Latitude	Longitude	Country	District	Municipality	Parish	Price M2	Price
0	101	Used	Homes	Apartment	3	1	76	152	0	38.7792	-9.1186	Portugal	Lisboa	Lisboa	Olivalis	2463	198000
1	102	Used	Homes	Duplex	5	3	190	380	0	38.7056	-9.1784	Portugal	Lisboa	Lisboa	Alcantara	3125	1270000
2	103	Used	Homes	Apartment	1	1	26	52	0	38.7058	-9.1639	Portugal	Lisboa	Lisboa	Estrela	4005	140000
3	104	Used	Homes	Apartment	5	4	185	370	0	38.7466	-9.1640	Portugal	Lisboa	Lisboa	Sao Domingos de Benfica	3412	995000
4	105	Used	Homes	Apartment	7	1	150	300	0	38.7323	-9.1287	Portugal	Lisboa	Lisboa	Arroios	3277	570000

Figure 1 Lisbon House Prices dataset.

Afterwards, this data is pre-processed and visualized in order to find any outliers or missing values. In this case, there was no missing values. In order to detect some outliers and visualize the main features the following Figures were obtained. The Fig. 2 and 3 highlights the variation of number of houses and price for each parish, respectively. Fig. 4 and 5 depicts the variation of houses' prices according to the number of bedrooms, bathrooms and the area net of the houses.

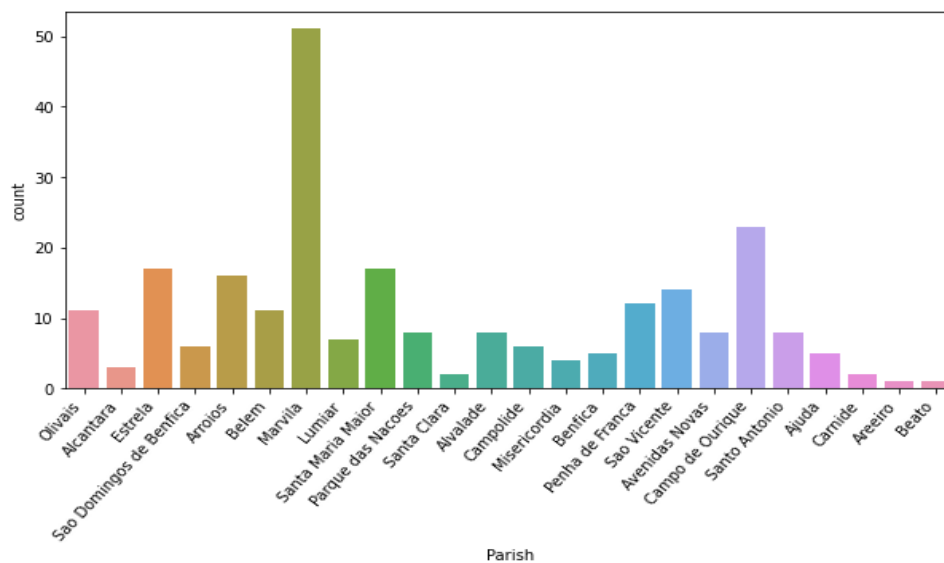


Figure 2 Variation of the number of houses for each parish.

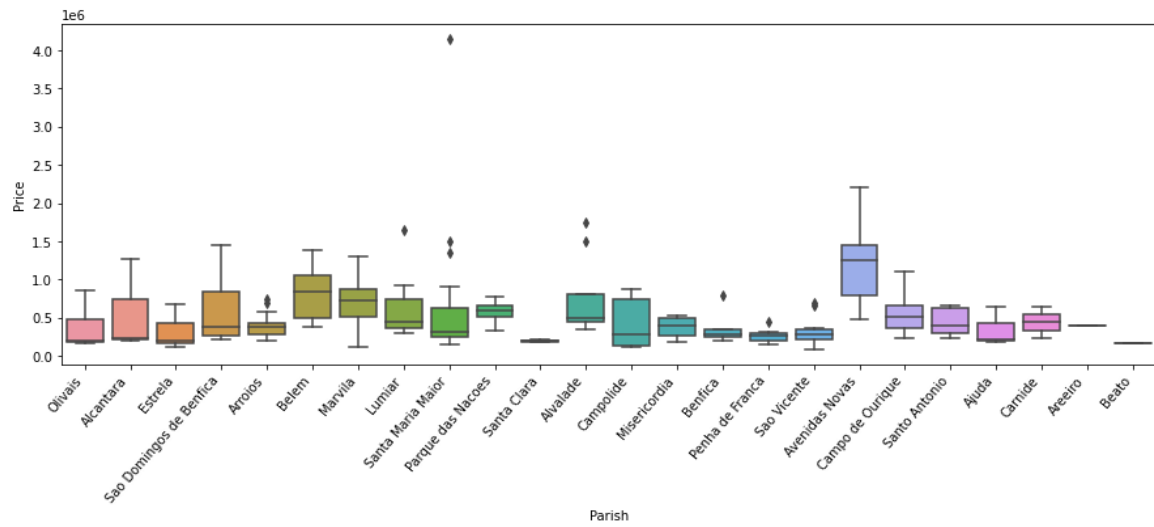


Figure 3 Variation of the houses' price for each parish.

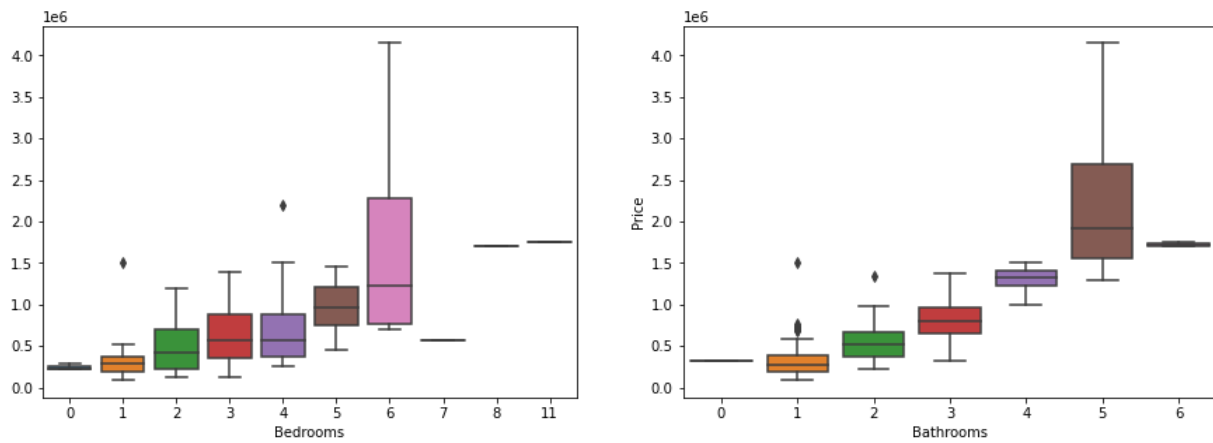


Figure 4 Variation of the house' prices according to the number of bedrooms and bathrooms.

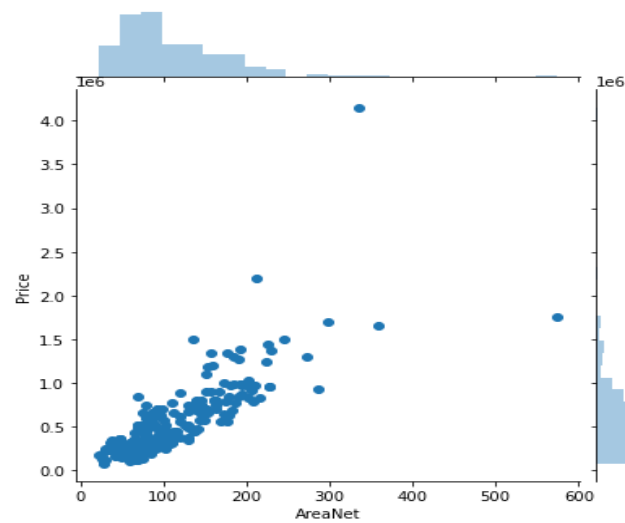


Figure 5 Variation of houses' price according to the area net (m²).

Fig. 2 shows that the number of houses in the dataset for each parish seems to be balanced except for the Marvila parish, probably because there were more houses for selling in that area. On the other side, Fig. 3 indicates that the Avenidas Novas is the parish with the highest houses' prices. From Fig. 4 and 5, it can be seen that on average the houses' prices increase with number of bathrooms, bedrooms and area net of the houses. From the results, there were at least two important outliers with a house price superior to 4 million euros and area net superior to 500 m². Both were removed from the final dataset. The original data also includes the latitude and longitude and this information was kept in order to collect the closes venues provided by the Foursquare application. The final dataset is highlighted in Fig.6, where each row represents the average characteristics of the houses for each parish.

	Parish	Latitude	Longitude	Bedrooms	Bathrooms	AreaNet	Price M2	Price
0	Ajuda	38.704880	-9.200220	1.400000	1.400000	61.800000	3137.0	332500.000000
1	Alcantara	38.705900	-9.180467	2.666667	1.666667	95.000000	3125.0	564966.666667
2	Alvalade	38.748800	-9.143900	3.285714	2.285714	131.857143	3623.0	611428.571429
3	Areeiro	38.743700	-9.129600	2.000000	2.000000	104.000000	3312.0	400000.000000
4	Arroios	38.728031	-9.137150	3.125000	1.375000	97.937500	3277.0	394368.750000

Figure 6 Pre-processed dataset.

3. Methodology

After cleaning and pre-processing the dataset, the first step of the methodology resides in utilizing the Folium package from Python to create a map from Lisbon with the main parishes and adding the average prices and area net of the houses in each area. Based on that, the main venues near each parish are computed via the Foursquare application and a dataset with the main venues is created, as described in Fig. 6.

Parish	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue
0 Ajuda	Supermarket	Pool	Chinese Restaurant	Soccer Field	Gas Station	Fast Food Restaurant	Tennis Court
1 Alcantara	Bakery	Portuguese Restaurant	Café	Mediterranean Restaurant	Dessert Shop	Restaurant	Indian Restaurant
2 Alvalade	Portuguese Restaurant	Bar	Restaurant	Hotel	Supermarket	Bakery	Plaza
3 Areeiro	Portuguese Restaurant	Indian Restaurant	Pizza Place	Stadium	Electronics Store	Bakery	Motorcycle Shop
4 Arroios	Café	Portuguese Restaurant	Hotel	Hostel	Middle Eastern Restaurant	Supermarket	Plaza

Figure 7 Dataset with main venues for each parish.

With this data, the *k*-means algorithm is performed with the aim of dividing the parishes into commercial and residential areas according to the venues nearby. This allows to present a final map that beyond the details of the average prices and area net of the houses in each parish also highlights if it is a residential or commercial area.

4. Results

The first result presents the map from Lisbon highlighting the average prices and area net of houses for each parish in Lisbon, as shown in Fig. 8. Once the *k*-means algorithm is performed, this map plot is enhanced with clusters information, as described in Fig. 9. By analyzing the clusters provided by the *k*-means algorithm, it can be seen that the cluster 2 is more residential with more supermarkets, gardens and bakeries nearby, whereas cluster 1 can be considered as more commercial/touristic since the most important venues near that areas are hotels/hostels and restaurants. In Fig. 9, the red markers represent the commercial parishes while the purple points characterize the residential areas.

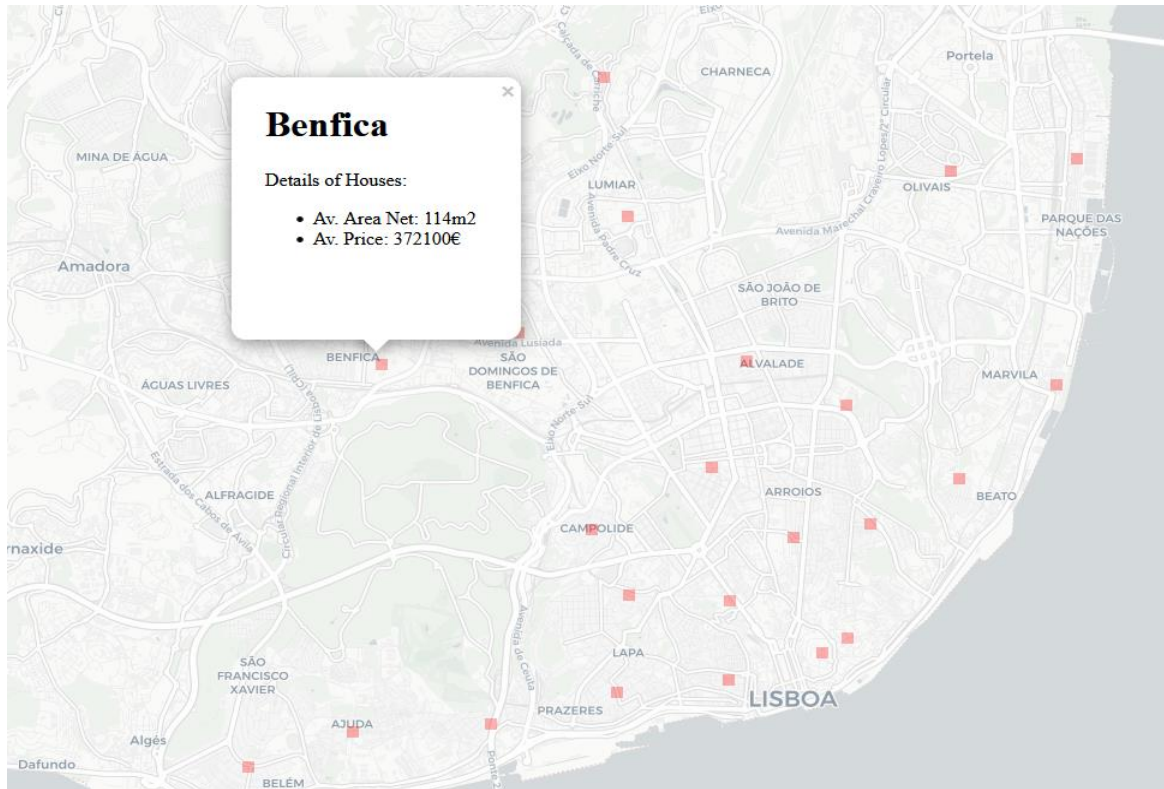


Figure 8. Map of Lisbon with the houses' information for each parish.

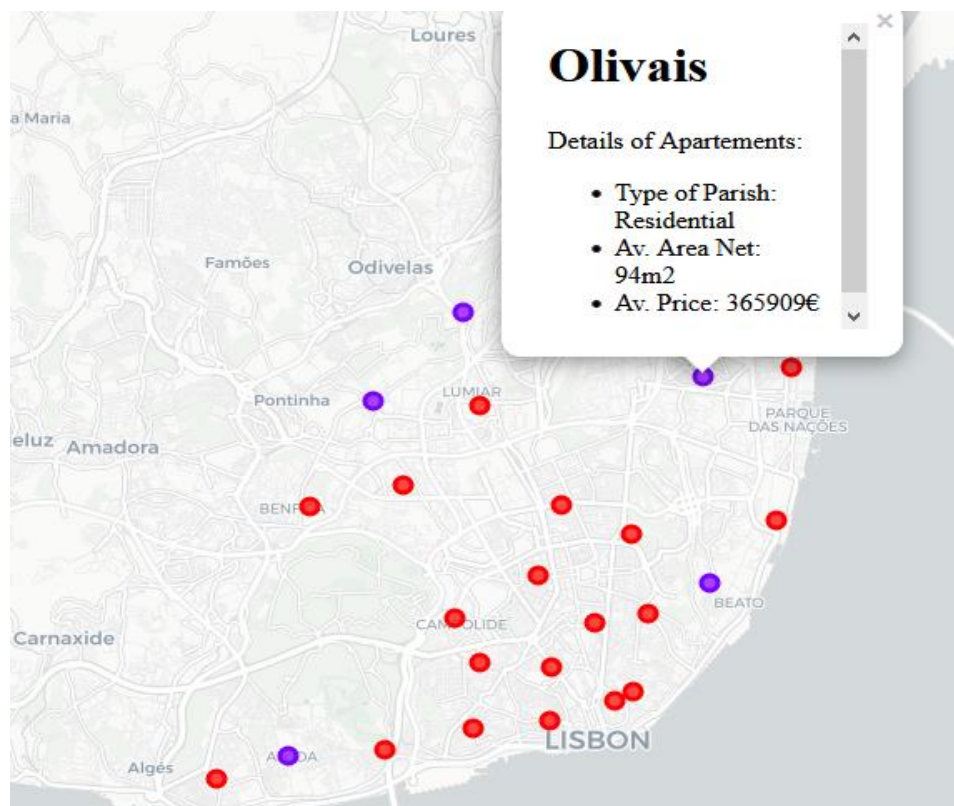


Figure 9 Map of Lisbon with the houses and clustering information for each parish.

5. Discussion/Conclusions

In this study, it is analyzed the variation of houses' prices for each parish in Lisbon that could help someone trying to find a house to buy in Lisbon. Moreover, it also highlights the most appropriate areas for residential or commercial/touristic purposes based on a clustering algorithm that is performed according to the main venues near each area.