The Battle of Neighborhoods

Capstone Project

Daniela Moniz

Introduction

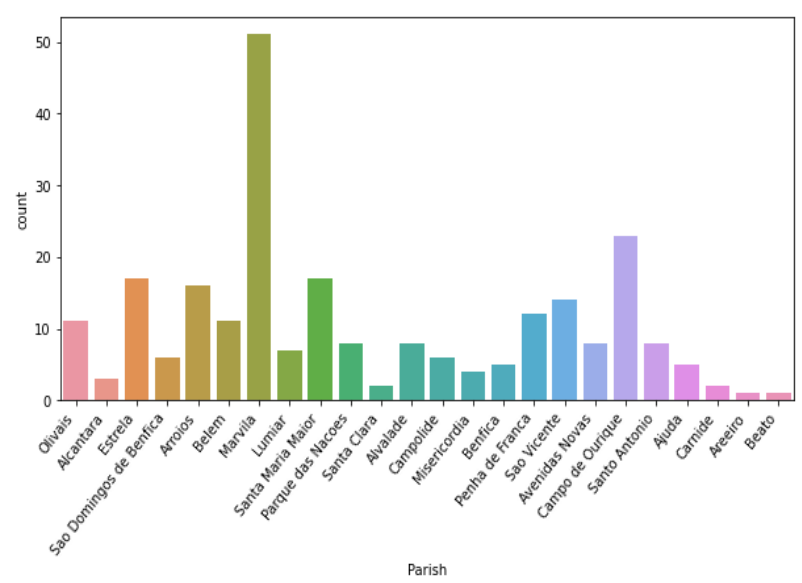
- Portugal's **housing market** has reached **highs** over the last years
- The main goal of this project is to help those who are looking to **buy a house in Lisbon**, Portugal
- Providing an **analysis of the variation of houses' prices** for each parish in Lisbon
- Clustering of the parishes into **commercial and residential** areas.

Data

The information of the Portugal's housing market is based on **Kaggle dataset** (Lisbon House Prices)

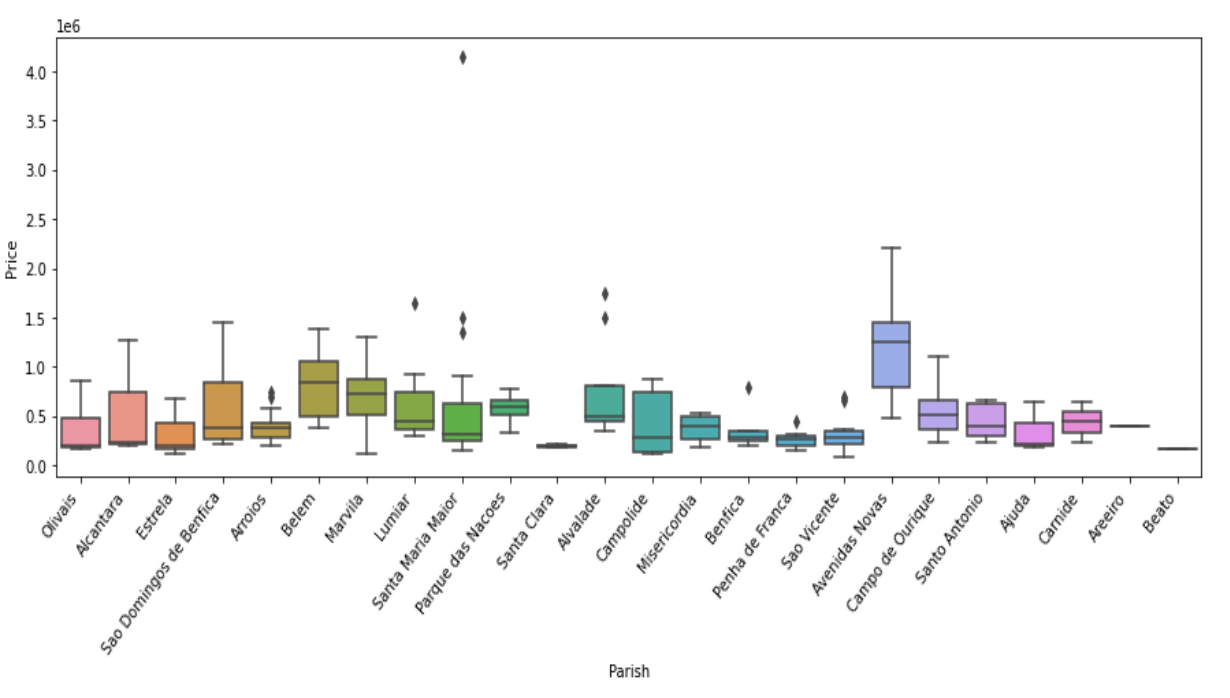
	Id	Condition	PropertyType	PropertySubType	Bedrooms	Bathrooms	AreaNet	AreaGross	Parking	Latitude	Longitude	Country	District	Municipality	Parish	Price M2	Price
0	101	Used	Homes	Apartment	3	1	76	152	0	38.7792	-9.1186	Portugal	Lisboa	Lisboa	Olivais	2463	198000
1	102	Used	Homes	Duplex	5	3	190	380	0	38.7056	-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

This dataset is then **pre-processed and visualized** in order to find missing data and outliers:

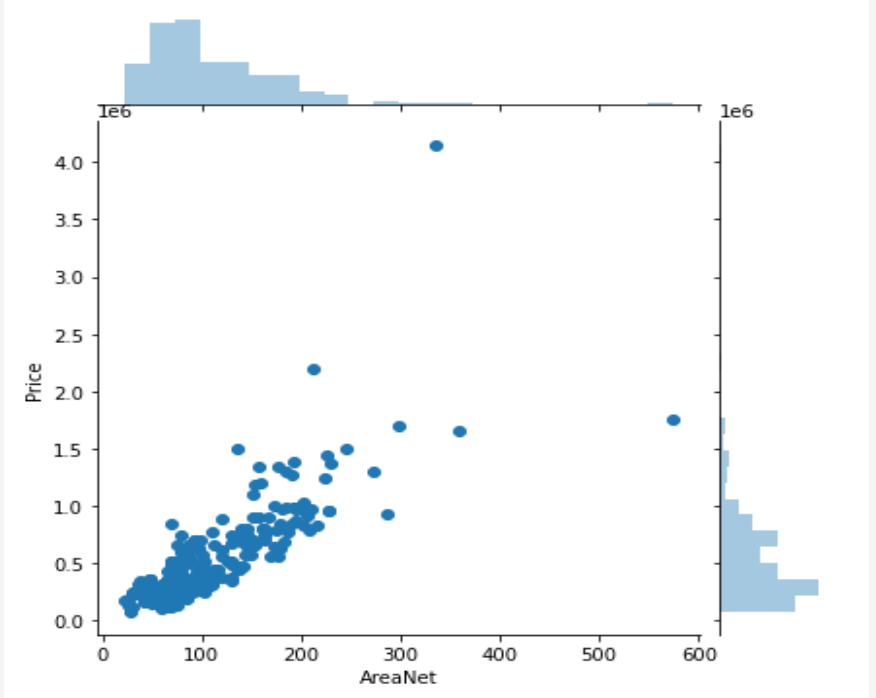


The number of houses in the data seems to be balanced except for **Marvila parish** (There are **more houses** for selling in that area)

Data

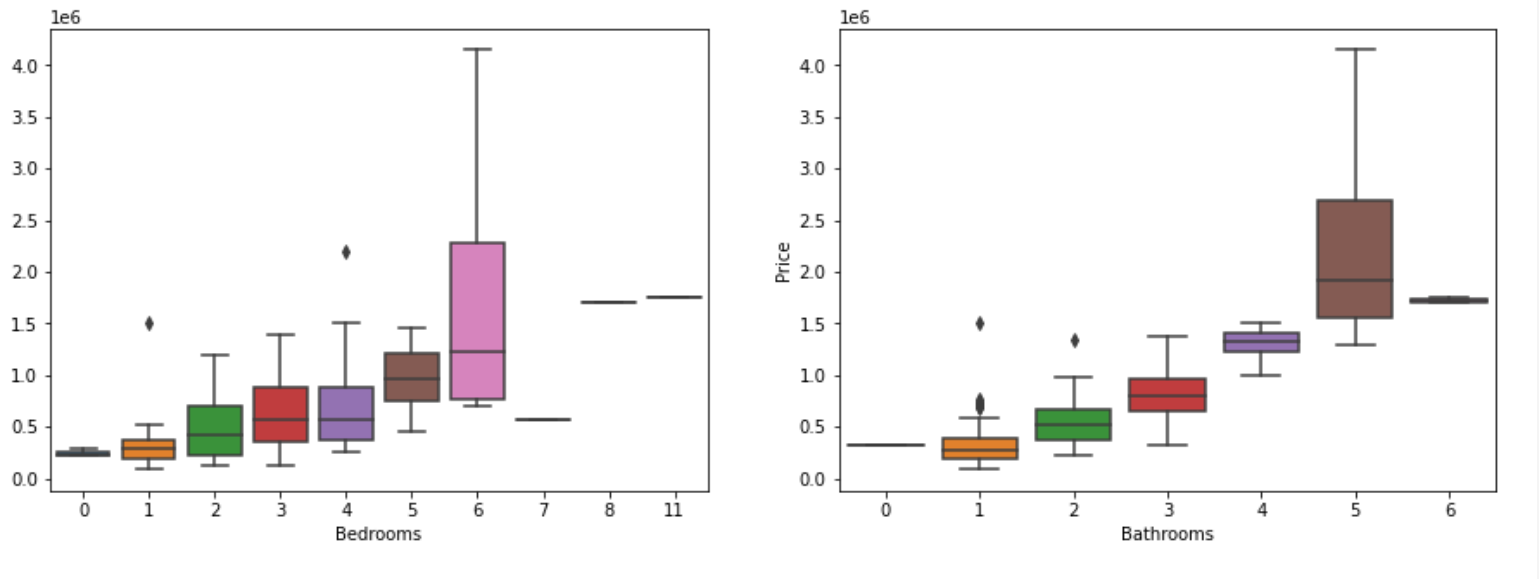


The **Avenidas Novas** is the parish with the **highest houses' price**



Positive correlation between the **Area Net** in m2 and houses' **price**

Data



On average the houses' prices increase with the number of bathrooms and bedrooms

- From the results, it can be seen that are at least **two important outliers** that were **removed** from the dataset

Data

	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

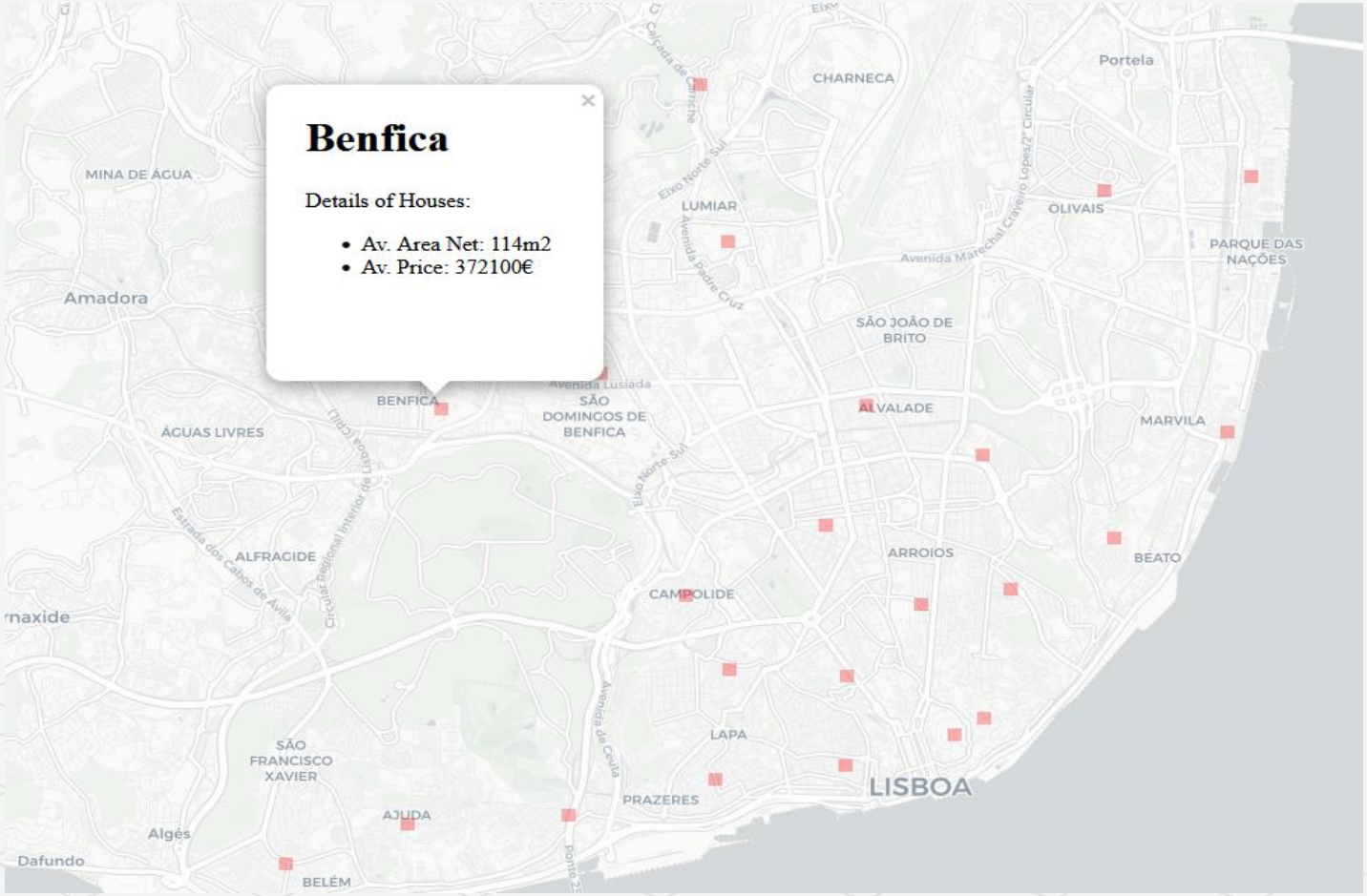
- **Final dataset** utilized where each row represents the **average characteristics** of the **houses** for each parish in Lisbon

Methodology

- The first step resides in utilizing the Folium to create a **map of Lisbon** with the **main parishes** and adding the **average prices and area net** of the houses in each area
- Afterwards, the **main venues near each parish** are computed via the **Foursquare**
- The ***k*-means algorithm** is performed with the aim of dividing the parishes into **commercial and residential areas** according to the venues nearby
- Present a final **Lisbon map** that beyond the houses' characteristics also highlights if it is a residential or commercial area

Results

- The first result presents the **map from Lisbon** highlighting the **average prices and area net** of the houses for each parish in Lisbon

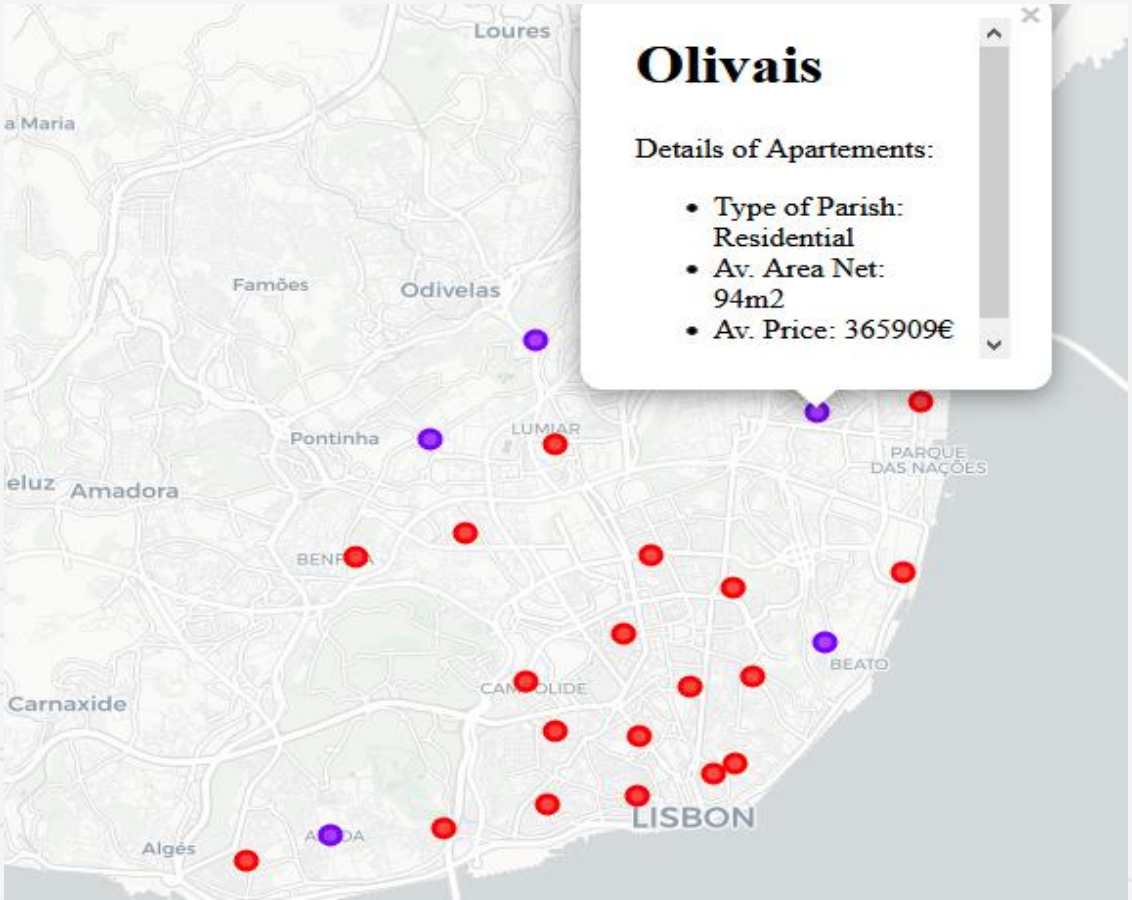


Results

- Once the *k*-means is performed, the map is enhanced with clusters information

Cluster 2 (Purple) is more residential with more supermarkets, gardens and bakeries

Cluster 1 (red) can be considered as more **commercial/touristic** since the most important venues are hotels and restaurants



Discussion

- 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**.

- 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.