IBM Data Science Professional Certificate

Course 9: Capstone

Final report

Predicting property prices of Lisbon’s boroughs using Foursquare data

Prepared by:

Luis Domingues

25/May/2019

Table of Contents

Abstract

Property prices have been rising in Portugal and particularly in Lisbon. Property is not only an essential asset for life but is also an investment instrument. For these reasons, being able to predict property prices is extremely useful.

In order to predict property prices in Lisbon's boroughs, Foursquare data was combined with other data on area and population of each of Lisbon's boroughs. Due to the high number of different venue categories returned by Foursquare, these were combined into a smaller set of more significant categories.

The features used to predict property prices of Lisbon's boroughs were chosen to maximize the correlation between features and the predicted variable (property price). The features used were *'Food and drink'*, *'Hotels and accomodation'*, *'Arts, entertainment and nightlife'*, *'Shopping'*, *'Population'*, *'Public buildings'*, *'Area'*, *'Outdoors'*, and *'Athletics and sports'*. The property prices of Lisbon's boroughs were then predicted using multiple linear regression. The fit obtained was reasonably good. The R^2 was 0.732, and the residuals were randomly distributed.

1 Introduction

Property prices have been steadily rising in Portugal since 2014, after a significant decrease in the period 2011-14, following the international crisis of '08 [1]. The main reasons for this increase are: increase in tourism activities, lower prices when compared with other European cities with similar touristic potential, legislation changes, which sparked an increase in urban rehabilitation, low interest rates environment, not enough properties to let, which stimulates buying for letting [1], [2]. Property prices rose by 6.09% y-o-y in 2018, to an average price of 1,220 €/sqm, with the highest increase in Lisbon metropolitan area, where prices rose by 7.9% [1]. The highest property prices are also in Lisbon, where the median is currently 2877 €/sqm [3].

Even though the improvement in economic conditions has been pointed out as one of the reasons for the increase in property prices [1], it has also been argued that it is mainly due to foreign investment. The reasons for this argument are that currently, it takes an average of 45.9 years to pay out an apartment in the centre of Lisbon with an average salary, and also because any non-EU citizen is granted Portuguese citizenship for 5 years if they invest >500 000 € in real-estate, with an option to renew permanently after that period. This means that property prices are expected to keep rising, provided foreign investors remain interested. Property in Portugal is still viewed as relatively inexpensive to investors and the number of transactions has been steadily increasing since 2014 [1], so it seems that indeed prices are expected to keep rising for the next couple of years.

Since property is, not only a place of residence but can also be an investment instrument, being able to predict property prices is extremely useful. Ideally, property prices would be predicted by readily available information.

The goal of this work is thus to use amenities' data, *e.g.* the number of restaurants in a borough, in conjunction with demography and geographical data of Lisbon's boroughs to predict the property prices in each of Lisbon's boroughs. Amenities' data will be retrieved by Foursquare with the remaining data on demography and geography being taken from Wikipedia.

2 Data collection

Different types of data were collected for this work. The data collected is described below with their respective sources:

- Lisbon boroughs' info:
  - Names [4]
  - Location (lat., long.) [4]
  - Property prices [5]
  - Population [4]
  - Area [4]
- Lisbon venues' info
  - Name [6]
  - Location [6]
  - Category [6]

The data on Lisbon's boroughs' info is presented in Appendix A. Since there are 182 venue categories returned by Foursquare, with a very high sparsity (many zeros), these categories were grouped in a set of 11 more meaningful reduced categories. The mapping between Foursquare categories and the reduced categories is presented in Appendix B. The reduced categories considered are presented below:

- Arts, entertainment and nightlife
- Food and drink
- Supermarkets and groceries
- Shopping
- Historic sites and museums
- Hotels and accomodation
- Athletics and sports
- Transport
- Public buildings
- Health and education buildings
- Outdoors

In total 1007 venues were retrieved from Foursquare. The data was then encoded and grouped by borough. The encoding was done on the venue categories to convert the data from categorical to binary, with the venues categories as the columns. This facilitated the grouping per borough, which consisted in summing all the venues in the same borough. This means that for each borough there is information on the number of venues for '*Shopping*', '*Food and drink*', '*Transport*', etc.

3 Methodology

The methodology used in this work, as well as the algorithms used, and the assumptions made are presented in this section.

3.1 Methods and algorithms

The main goal of the work is to predict property prices (€/m^2) from a set of features. The full set of features is the following:

- Population
- Area (km^2)
- Population density (inhabitants/km^2)
- Number of venues in category '*Arts, entertainment and nightlife*'
- Number of venues in category '*Food and drink*'
- Number of venues in category '*Supermarkets and groceries*'
- Number of venues in category '*Shopping*'
- Number of venues in category '*Historic sites and museums*'
- Number of venues in category '*Hotels and accomodation*'
- Number of venues in category '*Athletics and sports*'
- Number of venues in category '*Transport*'
- Number of venues in category '*Public buildings*'
- Number of venues in category '*Health and education buildings*'
- Number of venues in category '*Outdoors*'

As the goal is to be able to predict a continuous variable, multiple linear regression was used. The main question is which features to use in the correlation. The methodology used is explained in detail in the following sections.

3.1.1 Understanding the data

After analysing the data using descriptive statistics, an unsupervised clustering algorithm was used in order to try to better understand it.

- **Section methods:** Descriptive statistics and k-means clustering
- **Section goal:** To better understand the data

3.1.2 Determine factors to be used to predict property prices

In order to determine which factors to use in the multiple linear regression, the correlation factor of each feature with the property price was determined. Then the data was split into two sets (75% train; 25% test) and different multiple linear regression models were fitted to the train set each by adding the remaining factor with highest correlation factor. The optimal number of features corresponded to the maximum of R^2.

- **Section methods:** Correlation factors and multiple linear regression
- **Section goal:** Determine which factors to use to predict borough's property prices

3.1.3 Predict property prices of Lisbon's boroughs

Using the number of factors determined in section 3.1.2, the multiple linear regression model was fitted to the train dataset, which contained 75% of the data. The model was then validated, by predicting the remaining 25% of data, and plotting the residuals.

Multiple linear regression was used because the goal was to predict a continuous variable using a set of multiple features.

- **Section methods:** Multiple linear regression
- **Section goal:** Predict boroughs' property prices

3.2 Assumptions

- Foursquare data is correct and complete.
  This is a central assumption to this work as the majority of features are provided by this service.
- The radius of search was adequate.
  When querying the Foursquare database, a radius of search must be provided. In order to avoid overlap between boroughs, the radius used corresponded to the radius of the circumference with the lowest borough area. Even though there is a risk that some venues may not have been included in some boroughs, using the minimum radius was considered the fairest approach. Since all boroughs were queried with the same radius, the number of venues returned can be seen as a number of venues per area, with the area being the same for all.

4 Results and discussion

There is a total of 24 boroughs in Lisbon. The location of the centre of the boroughs can be seen in Figure 1. The complete borough data is presented in Appendix A.

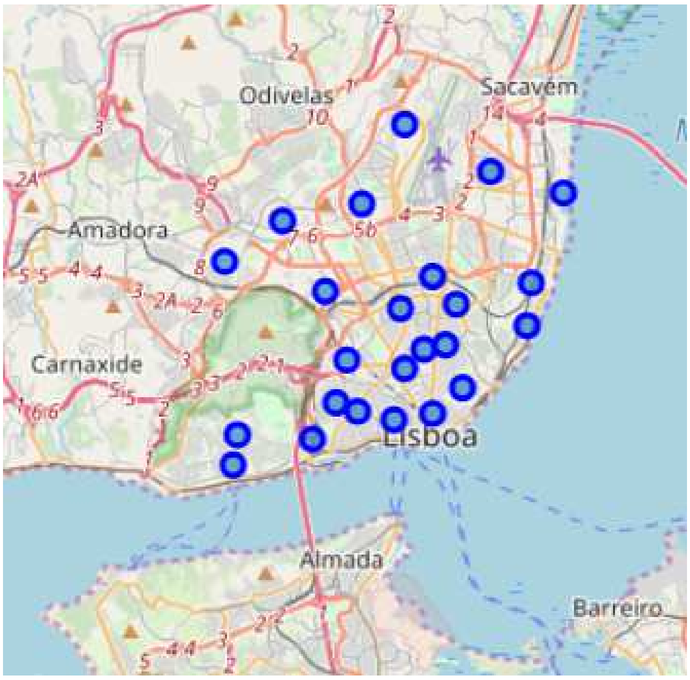

Figure 1 – Map of Lisbon including borough centres.

4.1 Understanding the data

In order to better understand the data prior to any further usage, both descriptive statistics and k-means clustering were used. Table 1 presents some descriptive statistics on price, population and area.

Table 1 – Descriptive statistics on Lisbon boroughs data.

Statistics	Price (€/m^2)	Population	Population dens. (inhab./km^2)	Area (km^2)
mean	2771.75	23029	6608	4.17
std	667.60	9519	3193	2.40
min	1543.00	11836	1584	1.49
25%	2381.00	15430	4548	2.39
50%	2741.00	20578	5769	3.19
75%	3155.75	31693	7704	5.37
max	4105.00	45605	14860	10.43

Histograms for price, population and area are presented in Figure 2. As can be observed in Figure 2, none of the variables seems to follow a normal distribution. Nonetheless, for price the mean (2771.75 €/m^2) corresponds to the bin with more observations, which is not the case for population and area. For both these features, the first three bins (with lowest values) have the highest number of observations. This indicates that price is differently distributed than population and area.

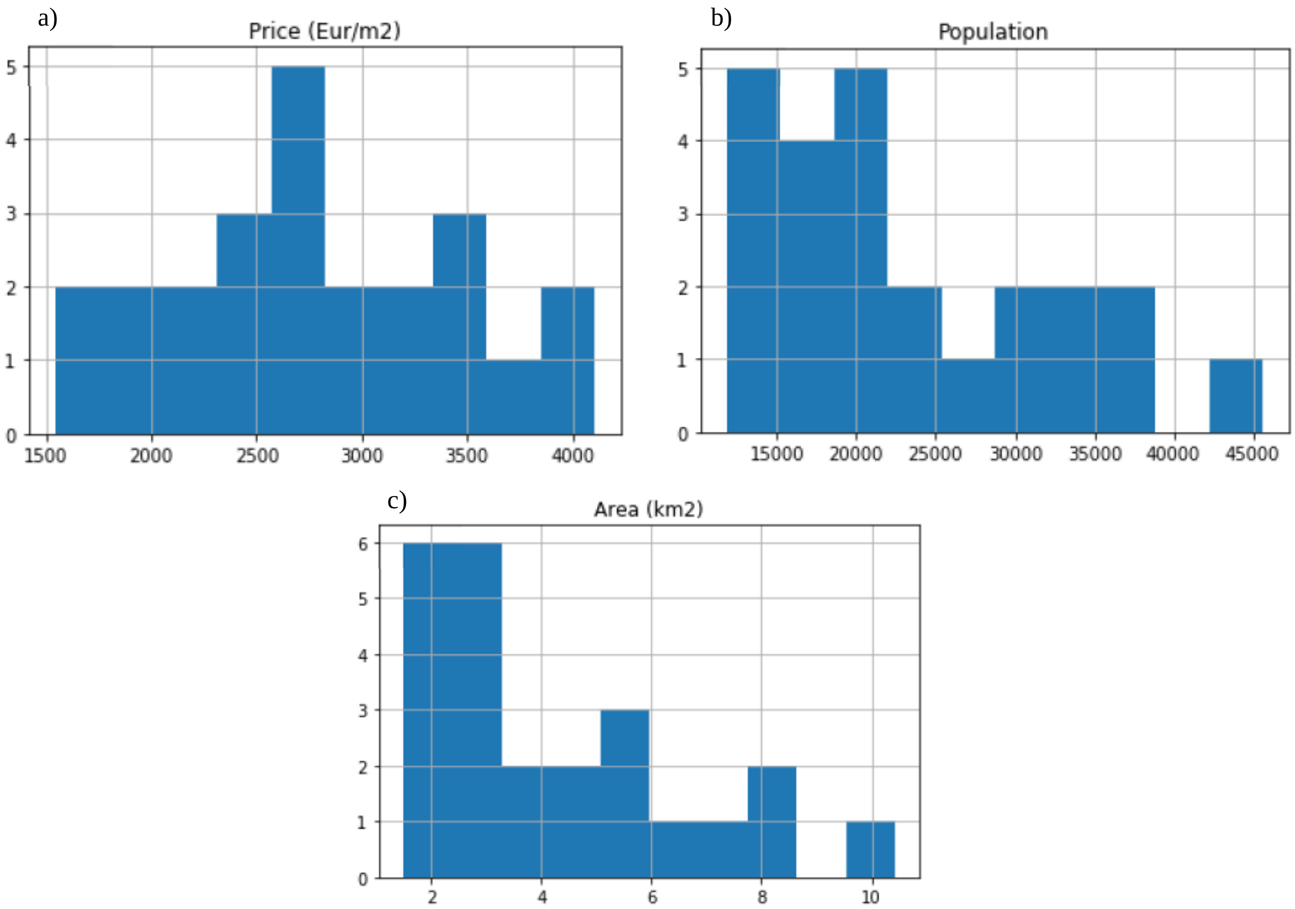

Figure 2 – Histograms for different features (number of bins = 10)

The number of observations for each (reduced) category is presented in Table 2. As can be seen in Table 2, the category with highest number of venues is *'Food and drink'*, with *ca.* 60% of retrieved venues. This is followed by distantly by *'Arts, entertainment and nightlife'* and *'Hotels and accomodation'*. These three categories combined account for *ca.* 80% of total venues retrieved by Foursquare. The number of venues in other categories seems suspiciously low, especially *'Transport'*, which including bus stops has only 13 occurrences, *'Health and education buildings'* and *'Supermarkets and groceries'*. Foursquare seems biased towards venues in the *'Food and drink'*, which includes all sorts of restaurants and bars (see Appendix B).

Table 2 – Number of venues for each category.

Category	**Number of venues**
Shopping	80
Food and drink	604
Arts, entertainment and nightlife	113
Athletics and sports	30
Transport	13
Outdoors	22
Hotels and accomodation	90
Public buildings	19
Historic sites and museums	11
Health and education buildings	4
Supermarkets and groceries	21
Shopping	80
TOTAL	1007

In order to better understand the data an unsupervised clustering algorithm was also used. The algorithm chosen was k-means.

One of the main problems when using k-means is choosing how many clusters to use. In this work, the number of clusters to be used was determined by the "elbow" method. Using this strategy, the optimal number of clusters corresponds to the cluster number after which the sum of intracluster distances squared starts to decrease at a lower rate. The variation of the sum of intracluster distances squared with the number of clusters is presented in Figure 3. As can be observed in Figure 3 the optimal number of clusters is either 3 or 4. In order to simplify the analysis, the optimal number of clusters chosen in this work was 3.

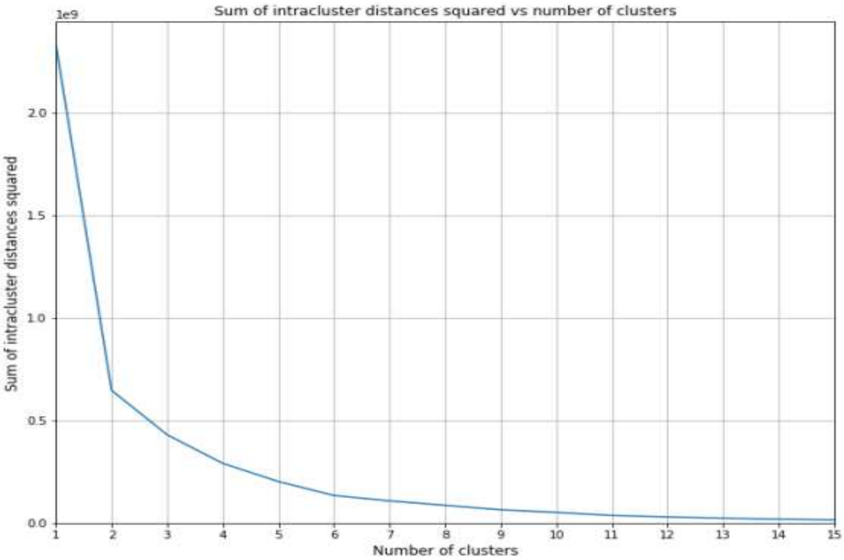

Figure 3 – Variation of the sum of intracluster distances squared with the number of clusters.

The algorithm was fitted again to all data, using the 3 clusters, which corresponds to the optimal number of clusters. The clustered boroughs can be seen in Figure 4 and the full list of boroughs and their respective clusters can be found in Appendix C.

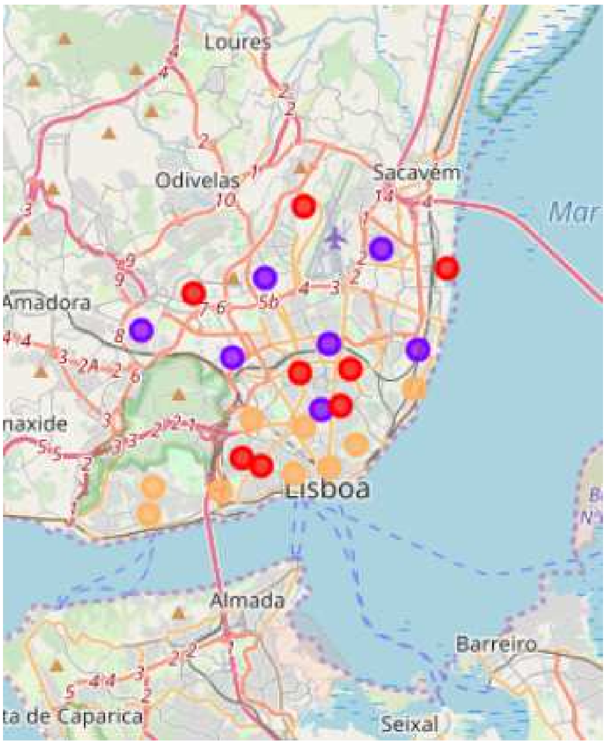

Figure 4 – Clustered boroughs obtained using k-means (k=3).
Cluster 1 (red); cluster 2 (purple); Cluster 3 (orange).

In order to better understand how the boroughs were clustered, the cluster centres can be analysed. These are presented in Table 3. Analysing Table 3 it can be seen that the cluster with highest property price is cluster 3. Geographically, this cluster corresponds to boroughs closer to the river, which is usually a prime location, as can be seen in Figure 4. It is also interesting to notice that cluster 3 has the lowest population and population density, which might indicate, that boroughs in this cluster might have bigger and more luxurious properties.

As can be observed in Table 3, cluster 2 has the lowest property price, while clusters 1 and 3 have similar prices. Table 3 shows that the differences when comparing categories *'Food and drink'*, *'Arts, entertainment and nightlife'* and *'Hotels and accomodation'* follow the same trend as the price. Cluster 3 has the highest price and also the higher value for *'Food and drink'* (29.56), followed closely by cluster 1, which has a value of 28. Cluster 3 comes very far with only 16.29. The same trend is observed for *'Arts, entertainment and nightlife'* and *'Hotels and accomodation'*, with cluster 3 having the highest value, followed closely by cluster 1 with cluster 2 a distant third. These three categories seem to be the most significant, as the combined sum of venues accounts for *ca.* 80% of the total venues retrieved from Foursquare. For this reason, the other categories were not analysed in the cluster analysis.

In summary, clustering reveals a set of boroughs (belonging to cluster 2), which have significantly lower property prices, less venues in *'Food and drink'*, *'Arts, entertainment and nightlife'* and *'Hotels and accomodation'* categories, when compared to boroughs in other clusters. It is harder to distinguish between boroughs in clusters 1 and 3 as comparing these categories they are fairly similar. Cluster 1 has nonetheless both a higher population and population density.

Table 3 – Cluster centres for k-means with 3 clusters.

Feature	Cluster 1	Cluster 2	Cluster 3
Price (€/m^2)	2864.0	2384.8	2990.7
Population (inhab.)	21837	35811	14147
Population dens. (inhab./km^2)	7833	7078	5153
Area (km^2)	3.273	5.939	3.588
Shopping	4.88	2.00	3.00
Food and drink	28.00	16.29	29.56
Arts, entertainment and nightlife	3.63	2.29	7.56
Athletics and sports	1.75	1.57	0.56
Transport	0.63	0.57	0.44
Outdoors	1.38	0.71	0.67
Hotels and accomodation	3.13	1.29	6.22
Public buildings	0.63	0.43	1.22
Historic sites and museums	0.13	0	1.11
Health and education buildings	0	0.43	0.11
Supermarkets and groceries	1.38	1.00	0.33

4.2 Determine factors to be used to predict property prices

With a better understanding of the data gathered, the next issue is which features to use to predict the property prices of Lisbon's boroughs. This issue was addressed firstly by determining the correlation factors of each feature with the property price. These features were then ranked and used to fit different multiple linear regressions, by adding every feature one at a time by the order in which they were ranked. The features which maximised the R^2 of the linear regression were chosen as the best.

Figure 5 presents the correlation factors of each feature with the property price. The features which have the highest correlation are *'Food and drink'*, *'Arts, entertainment and nightlife'* and *'Hotels and accomodation'*. As discussed in section 4.1, these are the categories with the highest number of venues and were shown by k-means to correlate well with price, so this result was expected.

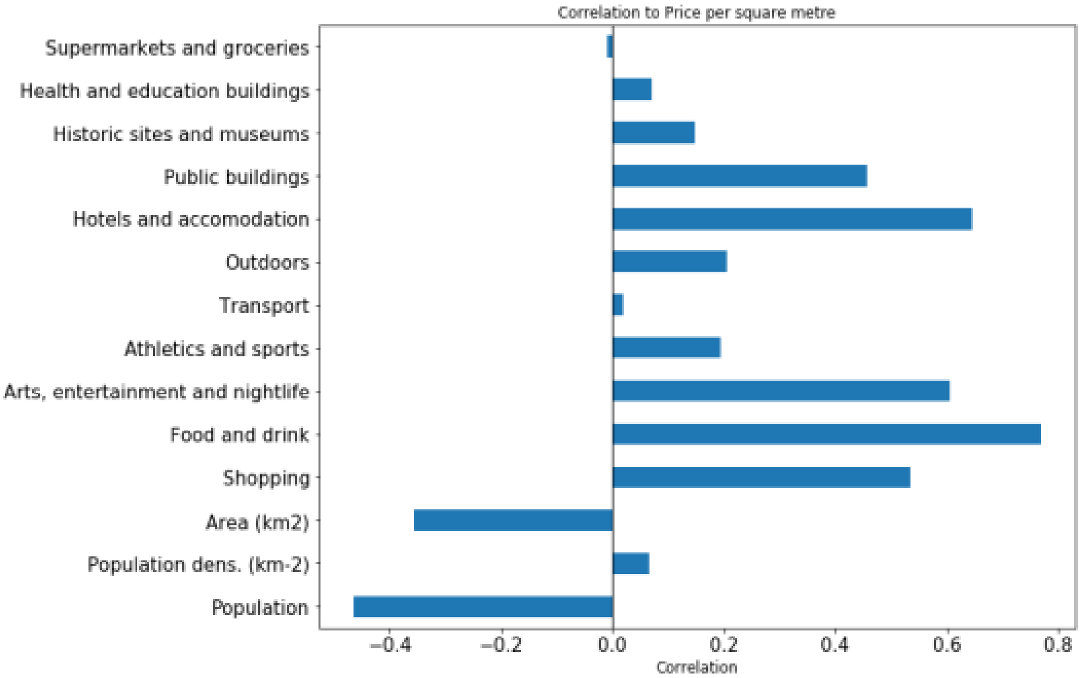

Figure 5 – Correlation factors with property price.

All the features were then ranked by highest absolute value of correlation factor and added one by one to multiple regressions. The variation of the R^2 with the number of features is presented in Figure 6. As can be seen in Figure 6, the optimum number of features to use in 9. The nine features used are presented in Table 4.

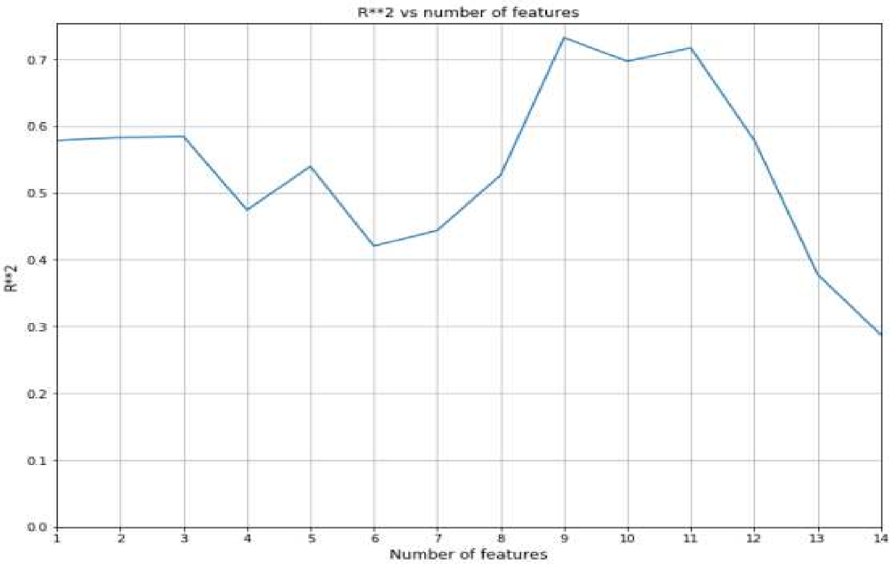

Figure 6 – Variation of R^2 with the number of features.

Table 4 – Features used to predict property price and their respective correlation factors with the predicted variable.

Feature	Absolute correlation factor
Food and drink	0.767761
Hotels and accomodation	0.644956
Arts, entertainment and nightlife	0.605600
Shopping	0.534702
Population	0.465174
Public buildings	0.457006
Area	0.356242
Outdoors	0.204509
Athletics and sports	0.194633

4.3 Predict property prices of Lisbon's boroughs

Using the optimal number of features, presented in Table 4, the desired multiple linear regression used to predict property prices of Lisbon's boroughs was determined. The coefficients obtained are presented in Table 5.

Table 5 – Coefficients of multiple linear regression used to predict property prices using the 9 optimal features.

Variable	Coefficient
Intercept	2559.86
Food and drink	-4.0805
Hotels and accomodation	44.4691
Arts, entertainment and nightlife	34.5041
Shopping	28.4188
Population	-0.0141
Public buildings	110.1465
Area	-23.3758
Outdoors	148.1406
Athletics and sports	84.5804

Multiple linear regression was used because the goal was to predict a continuous variable using a set of features. The data was divided in train and test datasets using a split of 75% and 25%, respectively. The R^2 of the multiple linear regression was found to be 0.732. The parity diagram showing the comparison between predicted and observed values for both train and test datasets is shown in Figure 7. It can be concluded from Figure 7 that the fit is reasonably good and the test dataset is predicted reasonably well. The residuals are presented in Figure 8 and seem to be randomly distributed.

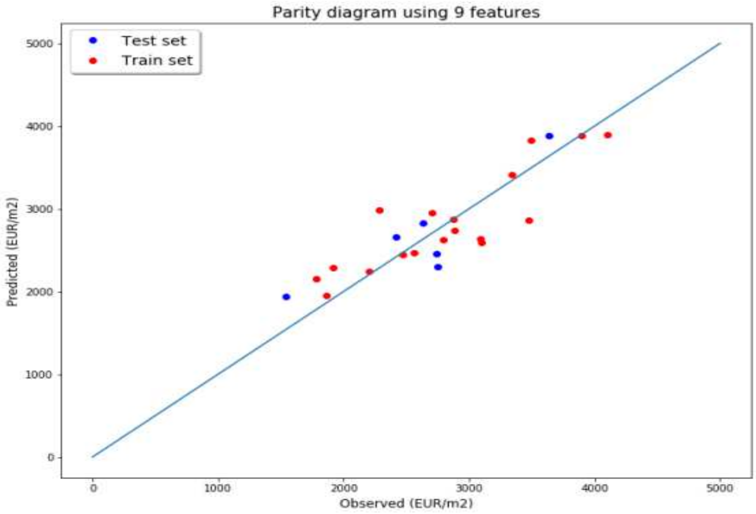

Figure 7 – Parity diagram for predicted property price using multiple linear regression with features in Table 4. Train and test datasets were 75% and 25% of the full dataset, respectively. $R^2 = 0.732$

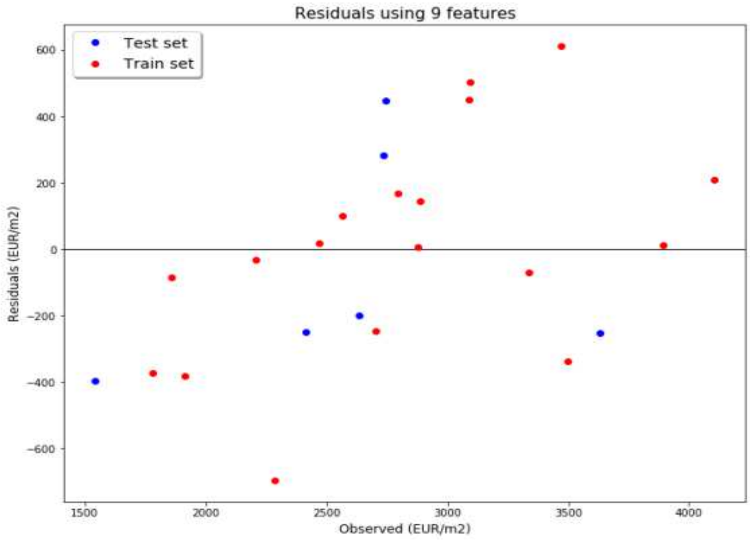

Figure 8 – Residuals of price predicted by multiple linear regression with features in Table 4.

5 Conclusion

In this work, Foursquare data was combined with other data on area and population of each of Lisbon's boroughs to try and predict their property prices.

Due to the high number of different venue categories returned by Foursquare, these were combined into a smaller set of more significant categories. It was found that Foursquare is somewhat biased towards returning venues in categories *'Food and drink'*, *'Arts, entertainment and nightlife'* and *'Hotels and accomodation'*, as they contain *ca.* 80% of the total venues returned.

In order to better understand the data, k-means clustering was used, with 3 clusters. This number was obtained by the "elbow" method, and corresponds to the number of clusters after which only small improvements in sum of intracluster distance square are obtained. It was concluded that cluster 3, which had the highest property prices also had more venues in the categories *'Food and drink'*, *'Arts, entertainment and nightlife'* and *'Hotels and accomodation'*. This showed that these categories correlate well with property price.

The features used to predict property prices of Lisbon's boroughs were chosen to maximize the correlation between features and the predicted variable (property price). The features used were *'Food and drink'*, *'Hotels and accomodation'*, *'Arts, entertainment and nightlife'*, *'Shopping'*, *'Population'*, *'Public buildings'*, *'Area'*, *'Outdoors'*, and *'Athletics and sports'*.

Finally, the property prices of Lisbon's boroughs were predicted using multiple linear regression. The fit obtained was reasonably good. The R^2 was 0.732, and the residuals were randomly distributed.

The obtained correlation can hopefully become a useful tool to predict the evolution of property prices in Lisbon's boroughs, by continuously monitoring the evolution of each of the features.

6 References

[1] C. D. Lalaine, “Investment Analysis of Portuguese Real Estate Market,” *GlobalPropertyGuide*, 06-Feb-2019. [Online]. Available: https://www.globalpropertyguide.com/Europe/Portugal/Price-History. [Accessed: 21-May-2019].

[2] N. Capital, “Existe uma bolha imobiliária em Lisboa? Sim, existe, e o Porto está a caminho.,” *Nomera Capital*, 24-Aug-2018. [Online]. Available: https://www.nomeracapital.pt/pt-pt/2018/08/24/existe-uma-bolha-imobiliaria-em-lisboa-sim-existe-e-o-porto-esta-a-caminho/. [Accessed: 21-May-2019].

[3] R. Godinho, “Casas em Lisboa são 20 vezes mais caras do que na Pampilhosa da Serra,” *Negócios*, 31-Jan-2019. [Online]. Available: https://www.jornaldenegocios.pt/empresas/imobiliario/detalhe/casas-em-lisboa-sao-20-vezes-mais-caras-do-que-na-pampilhosa-da-serra. [Accessed: 21-May-2019].

[4] Wikipédia, “Lista de freguesias de Lisboa,” 2019. [Online]. Available: https://pt.wikipedia.org/wiki/Lista_de_freguesias_de_Lisboa. [Accessed: 10-Apr-2019].

[5] N. Carregueiro, “Mapa: Só numa freguesia de Lisboa e Porto os preços das casas não sobem mais de 10%,” *Negócios*, 30-Oct-2018. .

[6] “Foursquare API,” *Foursquare*, 2019. [Online]. Available: https://developer.foursquare.com/. [Accessed: 22-May-2019].

Appendices

Appendix A. Lisbon data used in this work.

Borough name	Price (Eur/m2)	Latitude	Longitude	Population
Ajuda	2564	38.707500	-9.198333	15617
Alcantara	2413	38.706389	-9.174167	13943
Alvalade	3087	38.746944	-9.136111	31813
Areeiro	2701	38.740278	-9.128056	20131
Arroios	2793	38.728889	-9.138889	31653
Avenidas Novas	3338	38.738889	-9.145833	21625
Beato	1913	38.734722	-9.105833	12737
Belem	2877	38.700000	-9.200000	16528
Benfica	2205	38.751111	-9.202222	36985
Campo de Ourique	3095	38.715278	-9.166944	22120
Campolide	2633	38.726389	-9.163333	15460
Carnide	2745	38.760833	-9.183611	19218
Estrela	3472	38.713333	-9.160000	20128
Lumiar	2469	38.765278	-9.158611	45605
Marvila	1543	38.745278	-9.104167	37793
Misericordia	3894	38.711389	-9.148056	13044
Olivais	1860	38.773611	-9.117500	33788
Parque das Nacoes	3496	38.768056	-9.093889	21025
Penha de Franca	2285	38.730000	-9.131667	27967
Santa Clara	1780	38.785278	-9.145000	22480
Santa Maria Maior	3632	38.712778	-9.135556	12822
Santo Antonio	4105	38.724167	-9.145000	11836
Sao Domingos de Benfica	2737	38.743611	-9.170000	33043
Sao Vicente	2885	38.719444	-9.126389	15339

Appendix B. Mapping of Foursquare categories to reduced categories.

Foursquare category	Reduced category
Accessories Store	Shopping
African Restaurant	Food and drink
Aquarium	Arts, entertainment and nightlife
Argentinian Restaurant	Food and drink
Art Gallery	Arts, entertainment and nightlife
Arts & Crafts Store	Shopping
Asian Restaurant	Food and drink
Athletics & Sports	Athletics and sports
Auto Garage	Transport
BBQ Joint	Food and drink
Bagel Shop	Food and drink
Bakery	Food and drink
Bar	Arts, entertainment and nightlife
Beach	Outdoors
Bed & Breakfast	Hotels and accomodation
Beer Bar	Arts, entertainment and nightlife
Beer Garden	Food and drink
Bistro	Food and drink
Bookstore	Shopping
Botanical Garden	Outdoors
Boutique	Shopping
Brazilian Restaurant	Food and drink
Breakfast Spot	Food and drink
Brewery	Food and drink
Bubble Tea Shop	Shopping
Buffet	Food and drink
Burger Joint	Food and drink
Bus Station	Transport
Cable Car	Transport
Café	Food and drink
Candy Store	Shopping
Cantonese Restaurant	Food and drink
Capitol Building	Public buildings
Casino	Arts, entertainment and nightlife
Castle	Historic sites and museums
Cheese Shop	Food and drink
Chinese Restaurant	Food and drink
Chocolate Shop	Food and drink
Church	Historic sites and museums
Clothing Store	Shopping
Cocktail Bar	Arts, entertainment and nightlife
Coffee Shop	Food and drink
College Academic Building	Health and education buildings
Concert Hall	Arts, entertainment and nightlife
Convenience Store	Food and drink
Cosmetics Shop	Shopping

Creperie	Food and drink
Cultural Center	Arts, entertainment and nightlife
Dance Studio	Athletics and sports
Dessert Shop	Food and drink
Dim Sum Restaurant	Food and drink
Eastern European Restaurant	Food and drink
Electronics Store	Shopping
Empanada Restaurant	Food and drink
Event Space	Arts, entertainment and nightlife
Exhibit	Arts, entertainment and nightlife
Falafel Restaurant	Food and drink
Farmers Market	Supermarkets and groceries
Fast Food Restaurant	Food and drink
Fish & Chips Shop	Food and drink
Flea Market	Shopping
Flower Shop	Shopping
Food	Food and drink
Food Service	Food and drink
Food Stand	Food and drink
Food Truck	Food and drink
Fountain	Public buildings
French Restaurant	Food and drink
Frozen Yogurt Shop	Food and drink
Fruit & Vegetable Store	Supermarkets and groceries
Furniture / Home Store	Shopping
Garden	Outdoors
Gas Station	Transport
Gastropub	Food and drink
Gay Bar	Arts, entertainment and nightlife
General Entertainment	Arts, entertainment and nightlife
Gift Shop	Shopping
Gourmet Shop	Shopping
Grocery Store	Shopping
Gym	Athletics and sports
Gym / Fitness Center	Athletics and sports
Gym Pool	Athletics and sports
Health Food Store	Shopping
Hobby Shop	Shopping
Himalayan Restaurant	Food and drink
Historic Site	Historic sites and museums
History Museum	Historic sites and museums
Hostel	Hotels and accomodation
Hot Dog Joint	Food and drink
Hotel	Hotels and accomodation
Hotel Bar	Arts, entertainment and nightlife
Ice Cream Shop	Food and drink
Indian Restaurant	Food and drink
Indie Movie Theater	Arts, entertainment and nightlife
Italian Restaurant	Food and drink
Japanese Restaurant	Food and drink

Jewelry Store	Shopping
Juice Bar	Arts, entertainment and nightlife
Karaoke Bar	Arts, entertainment and nightlife
Kitchen Supply Store	Shopping
Lounge	Arts, entertainment and nightlife
Market	Supermarkets and groceries
Medical Center	Health and education buildings
Mediterranean Restaurant	Food and drink
Metro Station	Transport
Mexican Restaurant	Food and drink
Middle Eastern Restaurant	Food and drink
Miscellaneous Shop	Shopping
Mobile Phone Shop	Shopping
Motel	Hotels and accomodation
Motorcycle Shop	Shopping
Movie Theater	Arts, entertainment and nightlife
Museum	Historic sites and museums
Neighborhood	Public buildings
Nightclub	Arts, entertainment and nightlife
Noodle House	Food and drink
Office	Public buildings
Organic Grocery	Supermarkets and groceries
Other Great Outdoors	Outdoors
Other Nightlife	Arts, entertainment and nightlife
Paper / Office Supplies Store	Shopping
Park	Outdoors
Pastry Shop	Food and drink
Performing Arts Venue	Arts, entertainment and nightlife
Persian Restaurant	Food and drink
Pet Store	Shopping
Pharmacy	Health and education buildings
Pie Shop	Food and drink
Pizza Place	Food and drink
Playground	Outdoors
Plaza	Public buildings
Pool	Athletics and sports
Pool Hall	Athletics and sports
Portuguese Restaurant	Food and drink
Resort	Hotels and accomodation
Restaurant	Food and drink
Road	Transport
Roof Deck	Arts, entertainment and nightlife
Russian Restaurant	Food and drink
Salad Place	Food and drink
Sandwich Place	Food and drink
Scenic Lookout	Arts, entertainment and nightlife
Sculpture Garden	Arts, entertainment and nightlife
Seafood Restaurant	Food and drink
Shopping Mall	Shopping
Smoke Shop	Shopping

Snack Place	Food and drink
Soccer Field	Athletics and sports
Soccer Stadium	Athletics and sports
Soup Place	Food and drink
South American Restaurant	Food and drink
Spanish Restaurant	Food and drink
Speakeasy	Arts, entertainment and nightlife
Sporting Goods Shop	Shopping
Sports Club	Athletics and sports
Stadium	Athletics and sports
Steakhouse	Food and drink
Supermarket	Supermarkets and groceries
Sushi Restaurant	Food and drink
Swiss Restaurant	Food and drink
Tapas Restaurant	Food and drink
Tea Room	Food and drink
Tennis Court	Athletics and sports
Thai Restaurant	Food and drink
Theater	Arts, entertainment and nightlife
Theme Park	Arts, entertainment and nightlife
Thrift / Vintage Store	Shopping
Toy / Game Store	Shopping
Train Station	Transport
Tram Station	Transport
Vegetarian / Vegan Restaurant	Food and drink
Wine Bar	Arts, entertainment and nightlife
Wine Shop	Shopping
Wings Joint	Food and drink
Women's Store	Shopping
Yoga Studio	Athletics and sports
Zoo	Arts, entertainment and nightlife

Appendix C. Table of Lisbon boroughs and their respective cluster.

Borough name	Cluster
Ajuda	3
Alcantara	3
Alvalade	2
Areeiro	1
Arroios	2
Avenidas Novas	1
Beato	3
Belem	3
Benfica	2
Campo de Ourique	1
Campolide	3
Carnide	1
Estrela	1
Lumiar	2
Marvila	2
Misericordia	3
Olivais	2
Parque das Nacoes	1
Penha de Franca	1
Santa Clara	1
Santa Maria Maior	3
Santo Antonio	3
Sao Domingos de Benfica	2
Sao Vicente	3