The battle of neighbourhoods

Applied Data Science Capstone Project

Introduction

In this project we will help people who are looking for renting an apartment in Paris.

If they are looking to move to Paris they can see:

● Which district has cheaper rent or,

● They can choose to live in residential or commercial areas and can see for example which residential districts is best

Or, if they already live in one of the 20 districts in Paris they will be able to see:

● If they are paying more than the average price for their apartment

● If there are similar districts to theirs with lower rents Data

The data on apartments: size, number of rooms, address, and price is collected by scraping a French real estate website with apartment listings (seloger.com).

We clean up the values and calculate the price per square meter. It is further cleaned by removing NA values and outliers, which results in a dataset of 7800 apartments.

From this dataset we extract the rows with a unique Postal Code and then using geopy : we find the coordinates for each district.

Using Foursquare we collect the closest venues (supermarket, restaurant, park, etc.) and select the top 10 venues for each district.
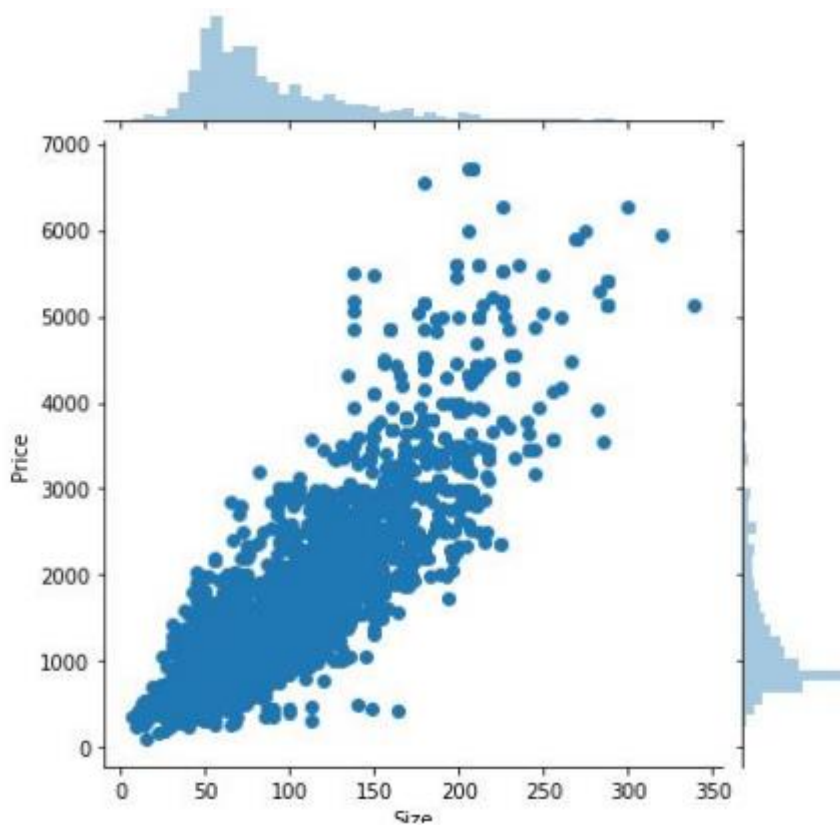
After the data collection we can run k-means clustering to cluster the districts into residential and commercial areas and visualize all the data on a single choropleth map.

Methodology :

We check the data set for how many apartments per room number we have and how many apartments we have in each district. We plot the following charts :
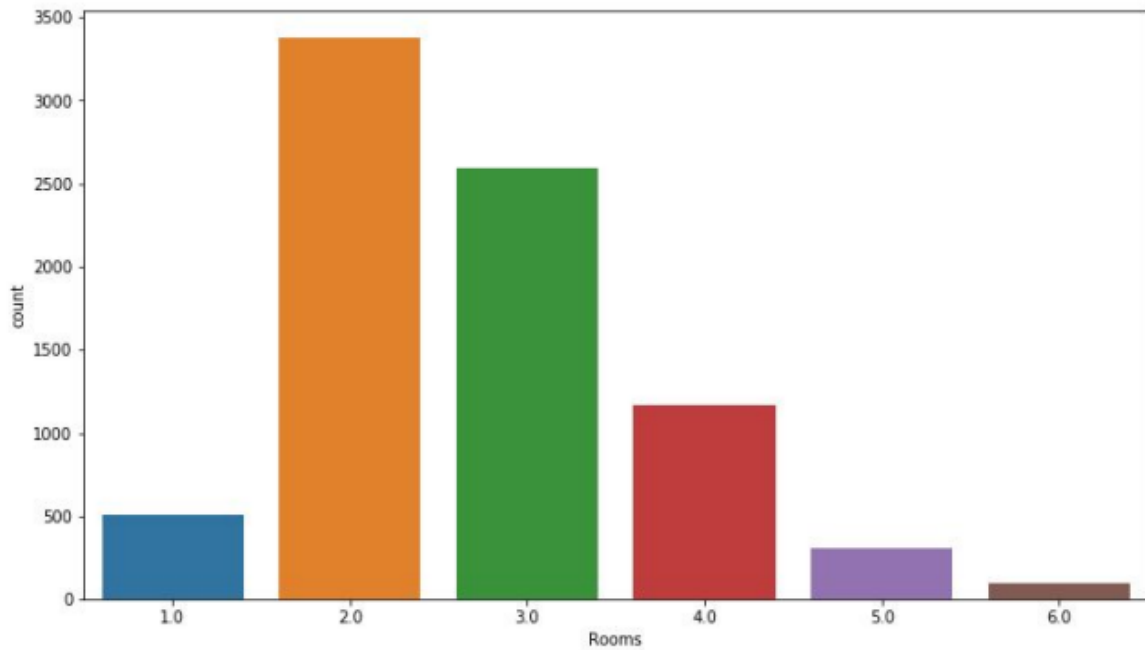
We can also see that there is a price between the rent price and apartment size :

Results

Once Foursquare data is collected. We find the most common venues (supermarket, restaurant, park, etc.) and select the top 10 venues for each district. After the data collection we can run k-means clustering to cluster the districts. By analyzing the clusters we can see

that the 16th district is more residential since it contains lots of parks and supermarkets while the 8th district is more commercial / touristic and contains many hotels and restaurants.

Discussion & Conclusion

One could determine for example that the 16th district is the most expensive district to live in, however by clustering we determined that there are several more similar districts where the price/m2 is significantly lower. Therefore, if someone wants to rent an apartment but cannot afford to live in the 16th district, they could look for apartments in the 12th or 15th district which are similar in venues but has much lower price for renting apartments.