The Battle of Neighborhoods

Applied Data Science Capstone Project

Abdulrahman Kayasseh

Introduction

In this project we will help people who are looking for renting an apartment in Vienna. If they are looking to move to Vienna 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 23 districts in Vienna 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 local website with apartment listings (willhaben.at). We clean up the values and calculate the price/m^2 by dividing the price by the size column. The data is pre-processed and we get our first dataframe:

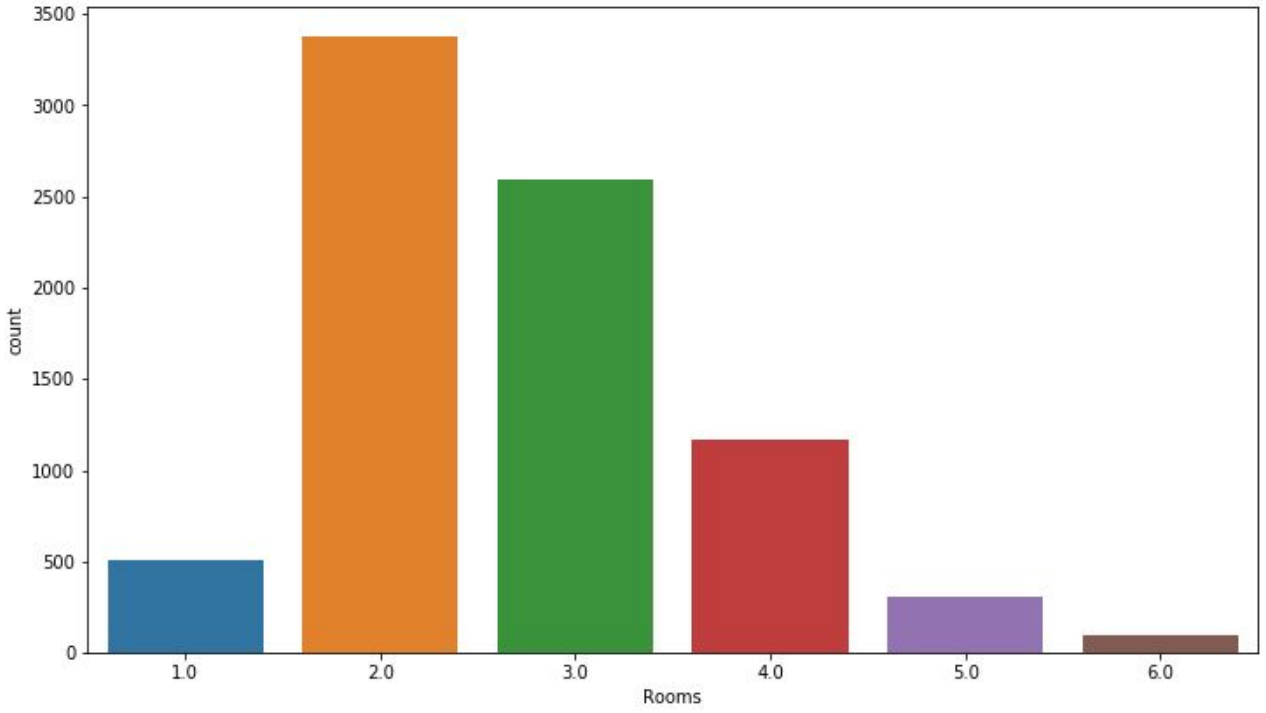
	PostalCode	District	Size	Rooms	Price	Price/m2
0	None	None	130.0	4.0	2152.10	16.55
1	1010	Innere Stadt	104.0	3.0	2260.28	21.73
2	1220	Donaustadt	12.0	1.0	432.00	36.00
3	1120	Meidling	75.0	3.0	840.00	11.20
4	1120	Meidling	43.0	2.0	700.00	16.28

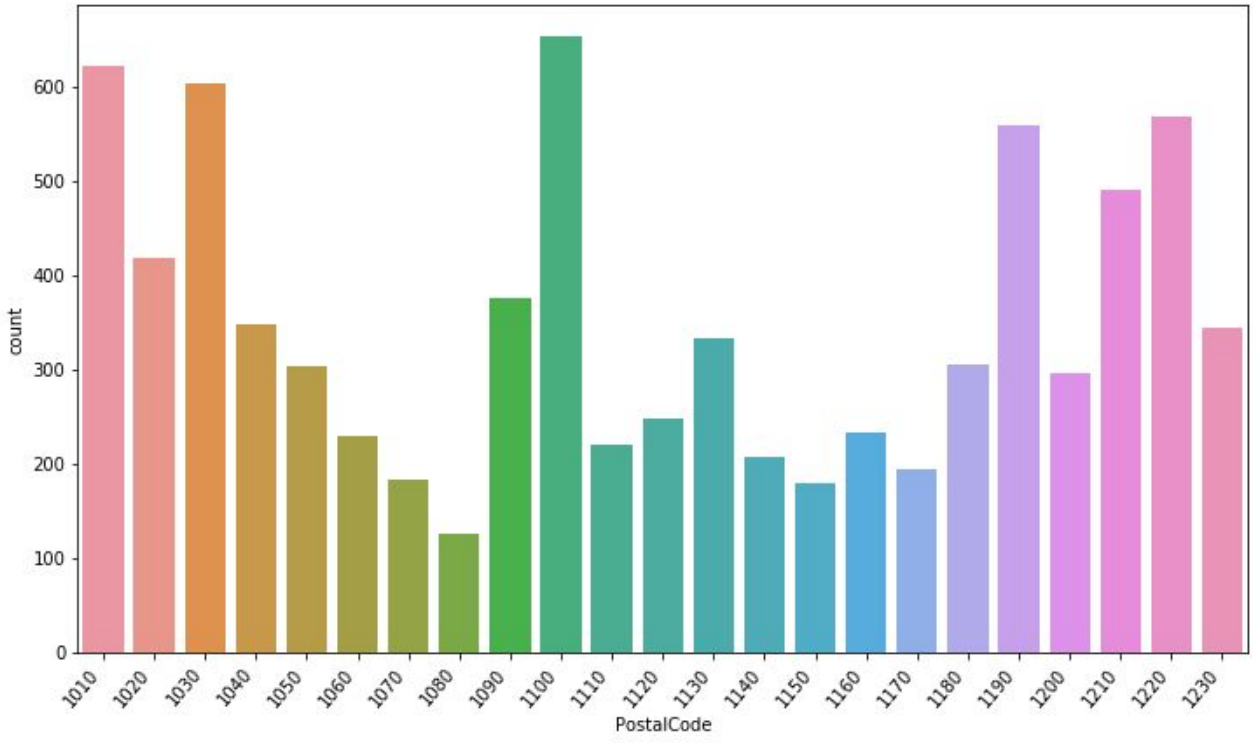
It is further cleaned by removing NA values and outliers, which results in a dataset of 8045 apartments. From this dataset we extract the rows with a unique Postal Code and then using geopy we find the coordinates for each district.

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

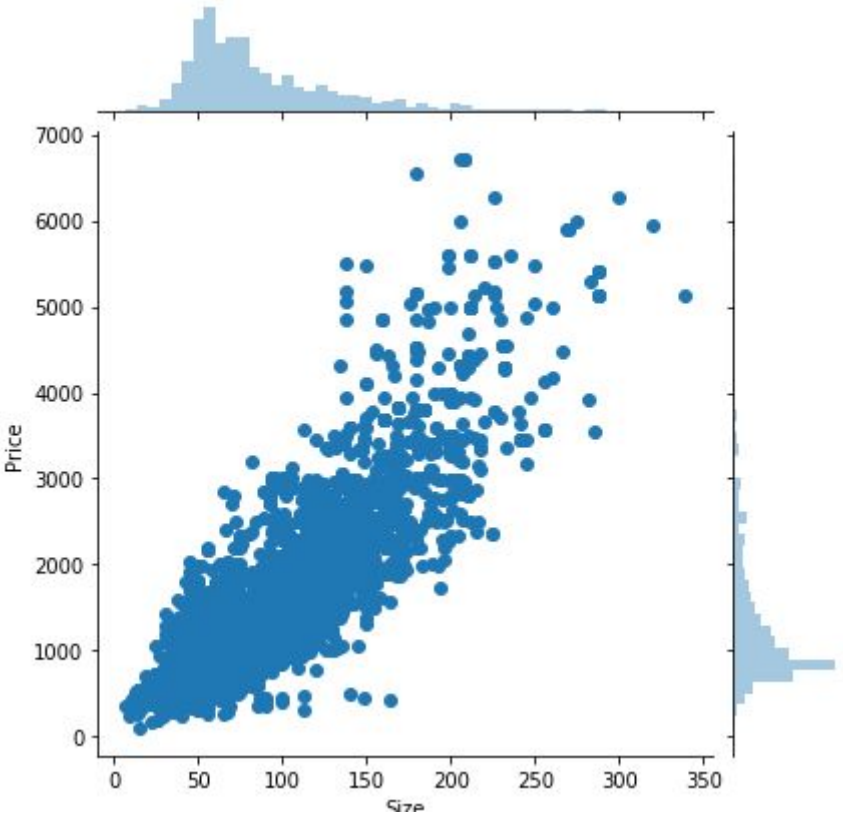
We check the data set for how many apartments per room number we have and how many apartments we have in each district. After removing any outliers we can plot the following charts:



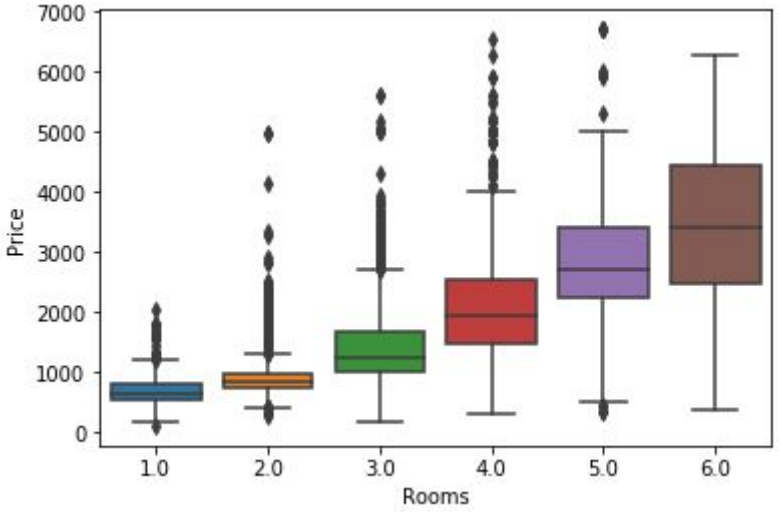


From this we can see that 2 and 3 room apartments are the most common.

We can also see if there is a correlation between the price and apartment size:

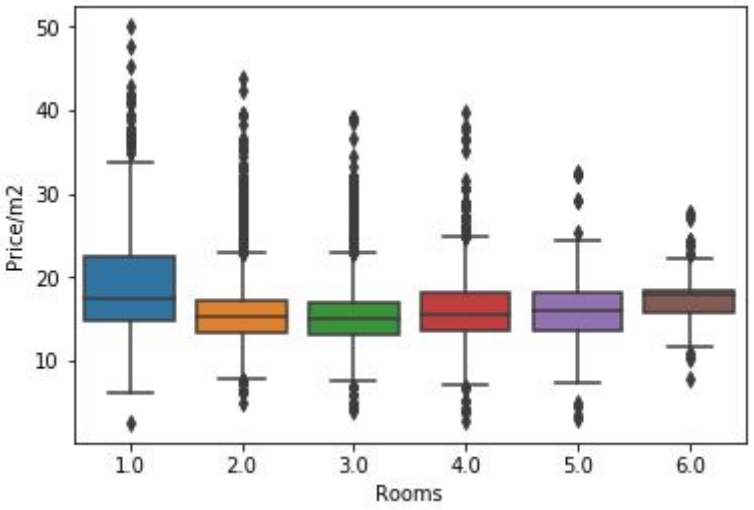


And as expected, the larger the apartment, the higher the monthly rent. Another thing we can look at is the rent per number of rooms:

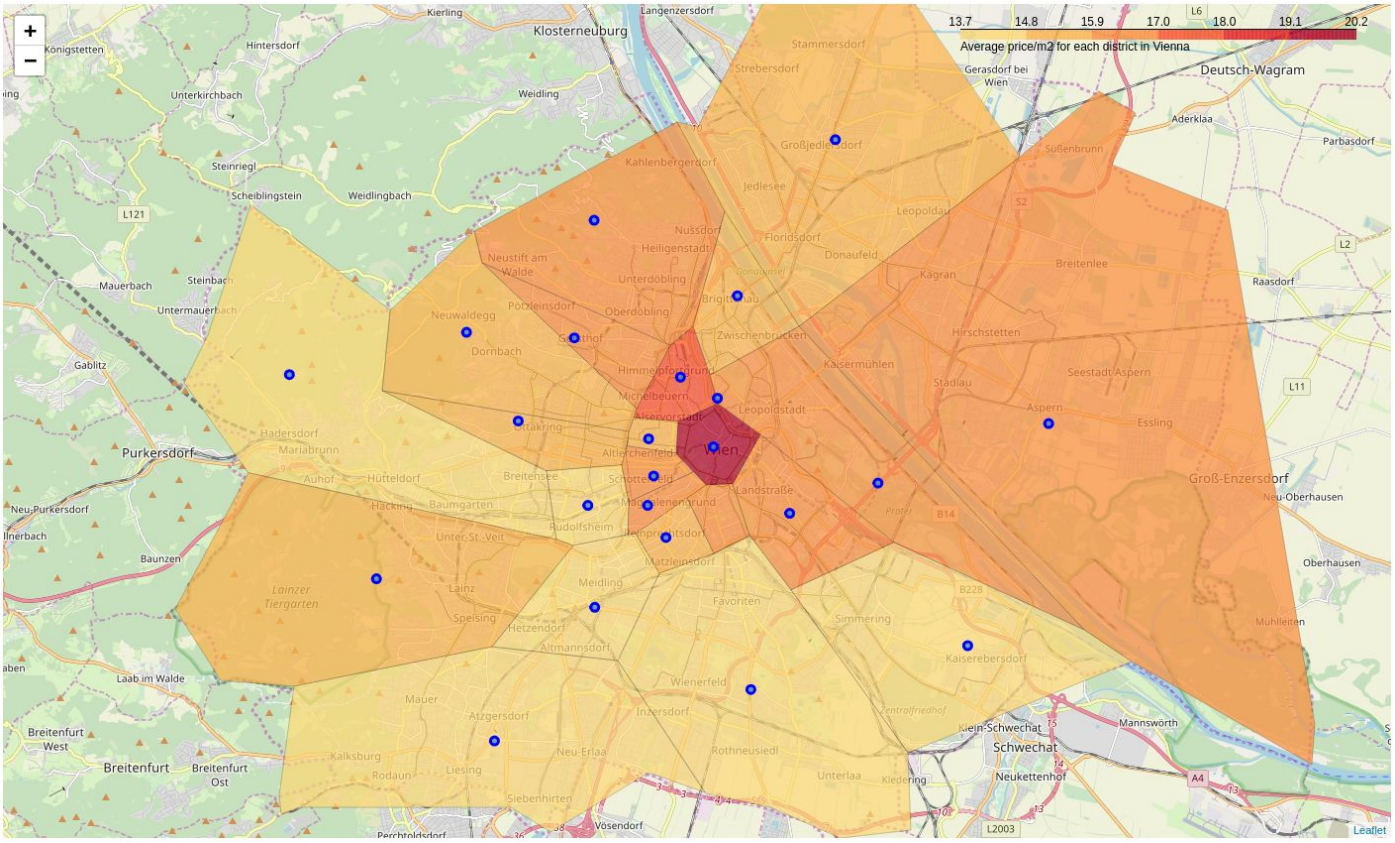


We expect the value of properties to go up as the number of rooms increases. The interesting aspect in this boxplot is that 1 and 2 room apartments are competing in the same price range.

In the following plot we can see that the price/m2 is in the same price range for all size apartments besides for single room apartments where the price/m2 goes even higher.



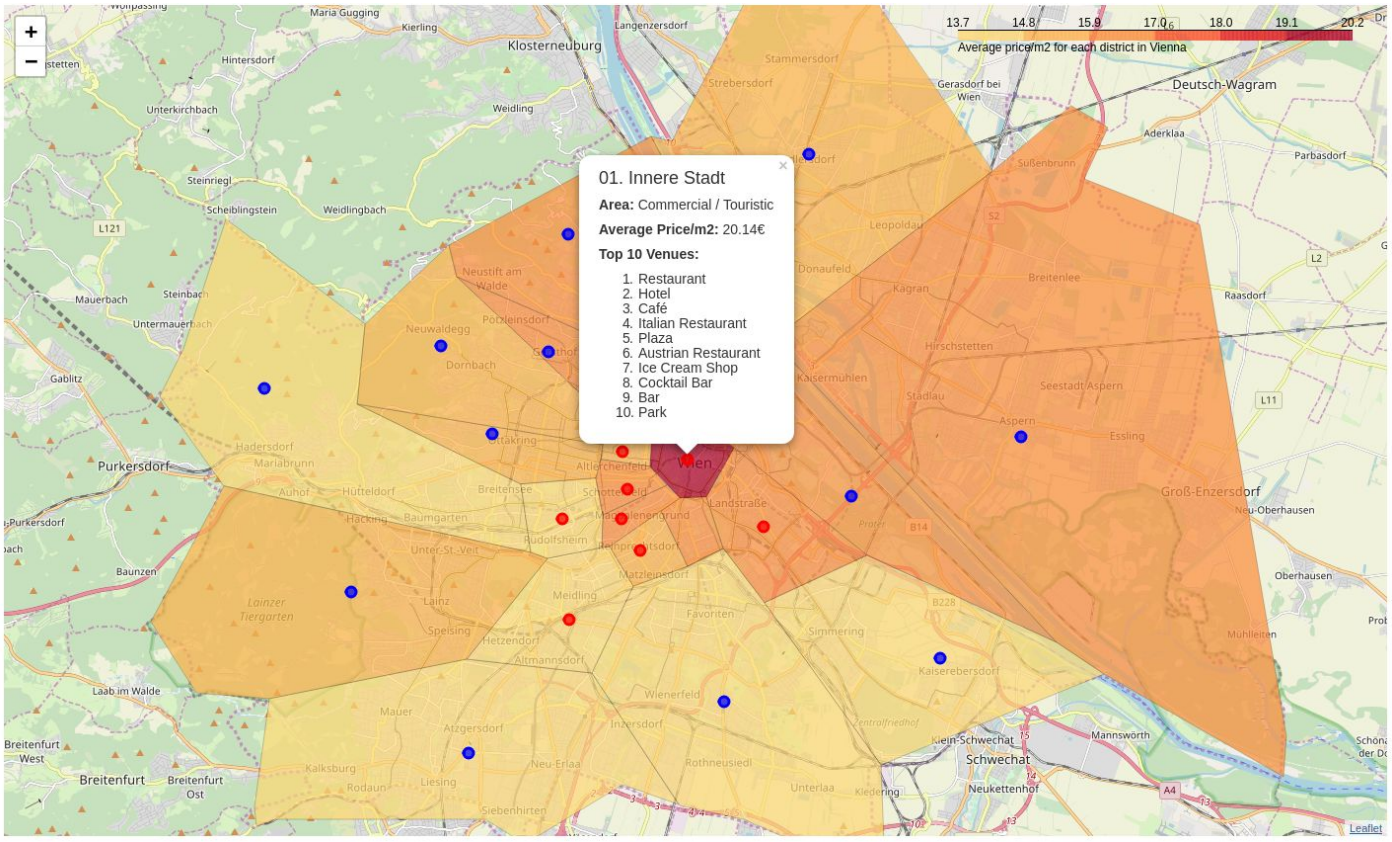
Using the apartment data and the district coordinates we can visualize the average price/m^2 for each district in Vienna. We can produce the following choropleth map:



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 cluster 1 is more residential since it contains lots of parks and supermarkets while cluster 2 is more commercial / touristic and contains many hotels and restaurants. We can visualize all this data on a single choropleth map.

The red markers represent the commercial districts and the blue markers represent the residential districts. The markers give further data on each district such as the average price/m^2 and the top 10 venues for each district.



Discussion & Conclusion

With this map, one could determine for example that the 1st 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 1st district, they could look for apartments in the 12th or 15th district which is similar in venues but has much lower price for renting apartments.