

Coursera - Applied Data Science Capstone

The Battle of Neighborhoods

Rafał Śliż

March 29, 2021

Table of Contents

1.	Introduction	2
2.	Data acquisition	2
3.	Methodology	3
4.	Results	6
5.	Discussion and conclusion	7

1. Introduction

In this project we will try help people who are looking for renting an apartment in Warsaw, the capital of Poland. If someone is looking to move to Warsaw, they can see in report:

- 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 18 districts in Warsaw 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

2. Data acquisition

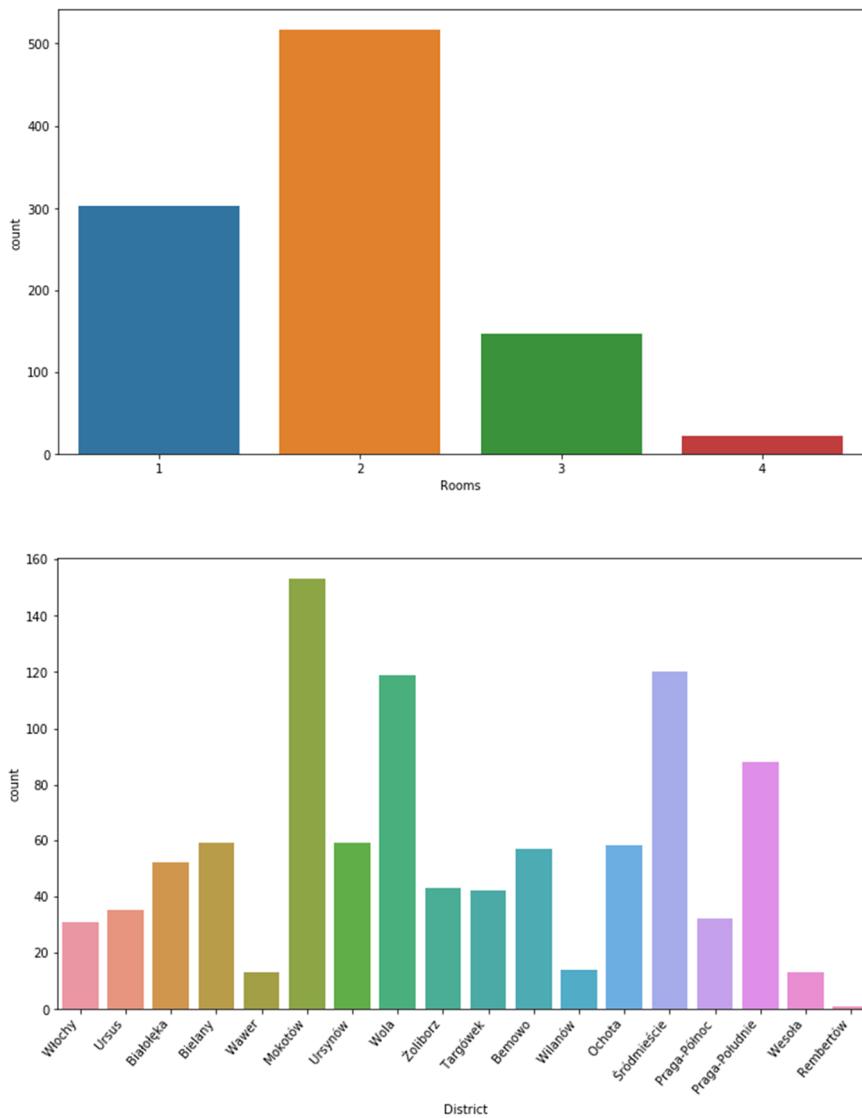
The data on apartments like: district, size, number of rooms, price were collected by scraping a local website with apartments advertisement www.olx.pl. I scraped the values and calculated the price per m² by dividing the price by the apartment size. The data was pre-processed and I got a first dataframe:

	District	Size m2	Rooms	Price	Price/m2
0	Włochy	49.00	2	2500.0	51.02
1	Ursus	37.07	2	1800.0	48.56
2	Białołęka	37.00	2	2300.0	62.16
3	Bielany	88.00	4	3499.0	39.76
4	Wawer	130.00	3	3500.0	26.92

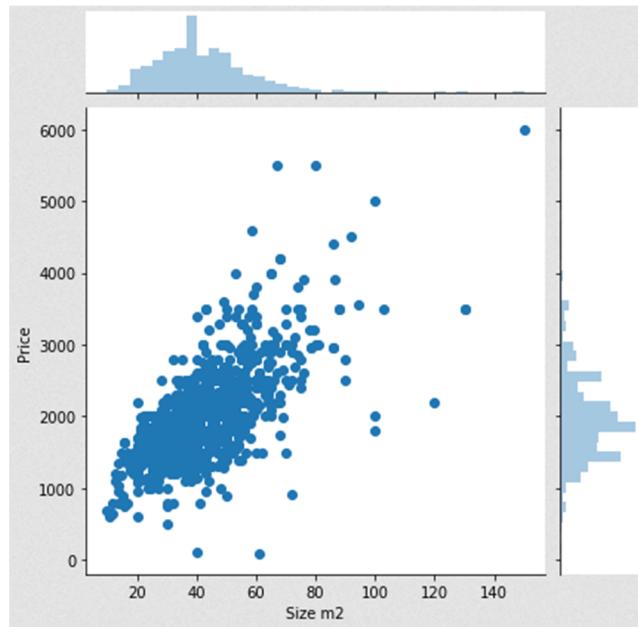
The above dataset was clean by removing N/A values and outliers, which results in a dataset of 989 apartments. In the next step, data about the location of each district was added. Using Foursquare API I collected the closest venues (supermarket, restaurant, park, etc.) and selected the top 10 venues for each district. After the data collection I was able to run k-means clustering to cluster the districts into residential and commercial areas and visualize all the data on a single choropleth map.

3. Methodology

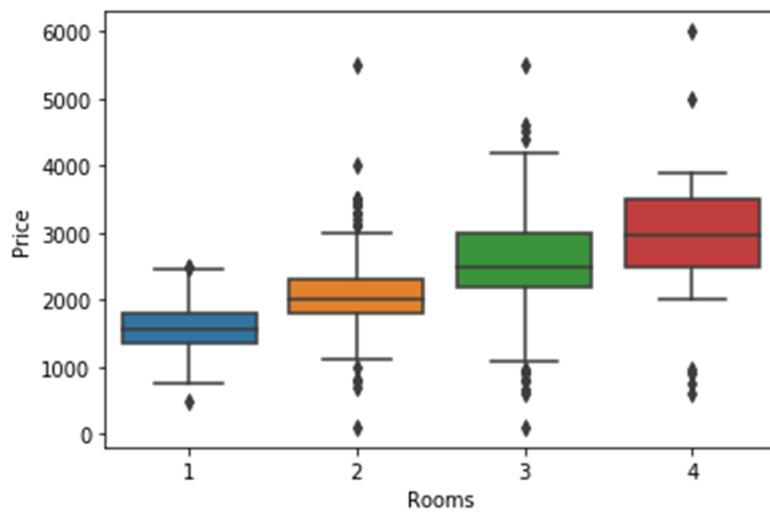
I check the data set for define how many apartments there are with particular number of rooms and how many apartments there are in each district. After removing any outliers I can plot the following charts:

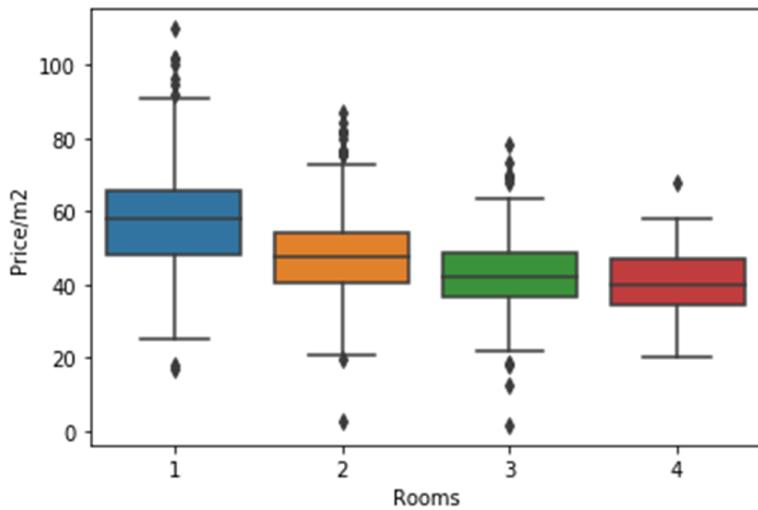


The above charts show that two-room flats are the most common. On the chart below we can see that there is a correlation between the price and apartment size.

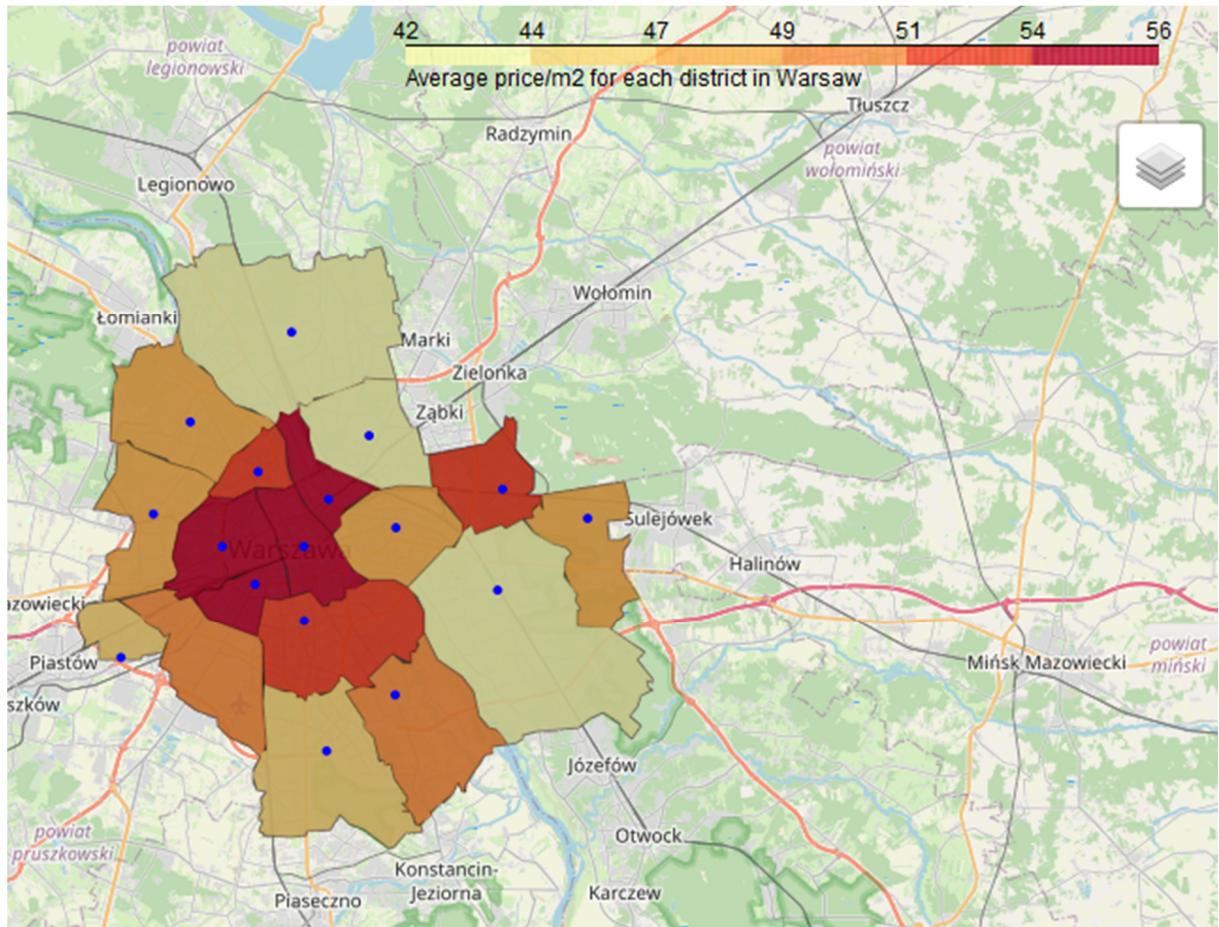


In the charts below, we can see that the price of an apartment increases with the number of rooms, as we would expect. While the price of one square meter decreases with the increase in the number of rooms in the apartment. For apartments with three and four rooms, the price per square meter is similar.





Using the apartment data and the district coordinates I visualized the average price per square meter for each district in Warsaw. It is presented in the choropleth map below :

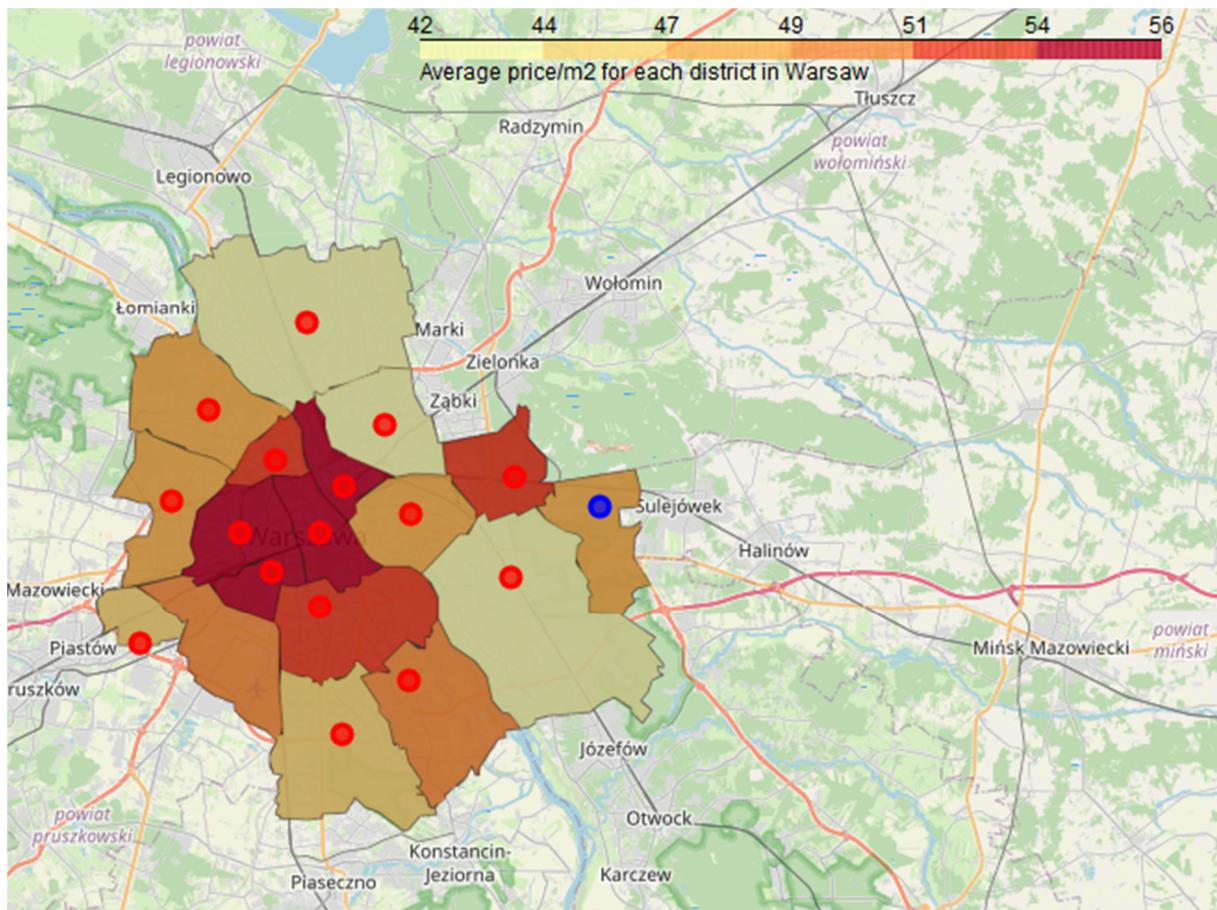


As we can see on the map above, the highest average price per square meter of a flats is in the districts of Praga-Północ, Śródmieście, Ochota, Wola. These are districts located in the city center.

4. Results

This analysis of apartments in Warsaw was made on the basis of data collected from Foursquare, Wikipedia and the local website with advertisements olx.pl.

After collecting the data, I found the most common venues (supermarket, restaurant, park, etc.) and I selected the top 10 venues for each district. In the next step I ran k-means clustering to cluster the districts. By analyzing the clusters we can see that cluster 1 have the most green spaces like forest and park. We can say that cluster 1 is a suburban area. Cluster 2 is more urbanized. All the above data is visualized on one a single choropleth map below. The red markers represent the urbanized districts - cluster 2, and the blue markers represent "the green" district - cluster 1. The markers give further data on each district such as the average price per square meter and the top 10 venues for each district.



5. Discussion and conclusion

This analysis makes it possible to determine:

- what size of flats are there in warsaw
- how many flats for rent are in particular districts
- what is the average price of a one-room, two-room, three-room and four-room flat
- what is the price of one square meter of an apartment depending on the number of rooms
- which districts are more urbanized and which have more green spaces
- 10 of the most popular venues in each district

This analysis can help people who are looking for renting an apartment in Warsaw, the capital of Poland