The Battle of Neighborhoods

Coursera - Applied Data Science Capstone Project

Rafał Śliż

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 raport:

- Which district has cheaper rent,
- 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:

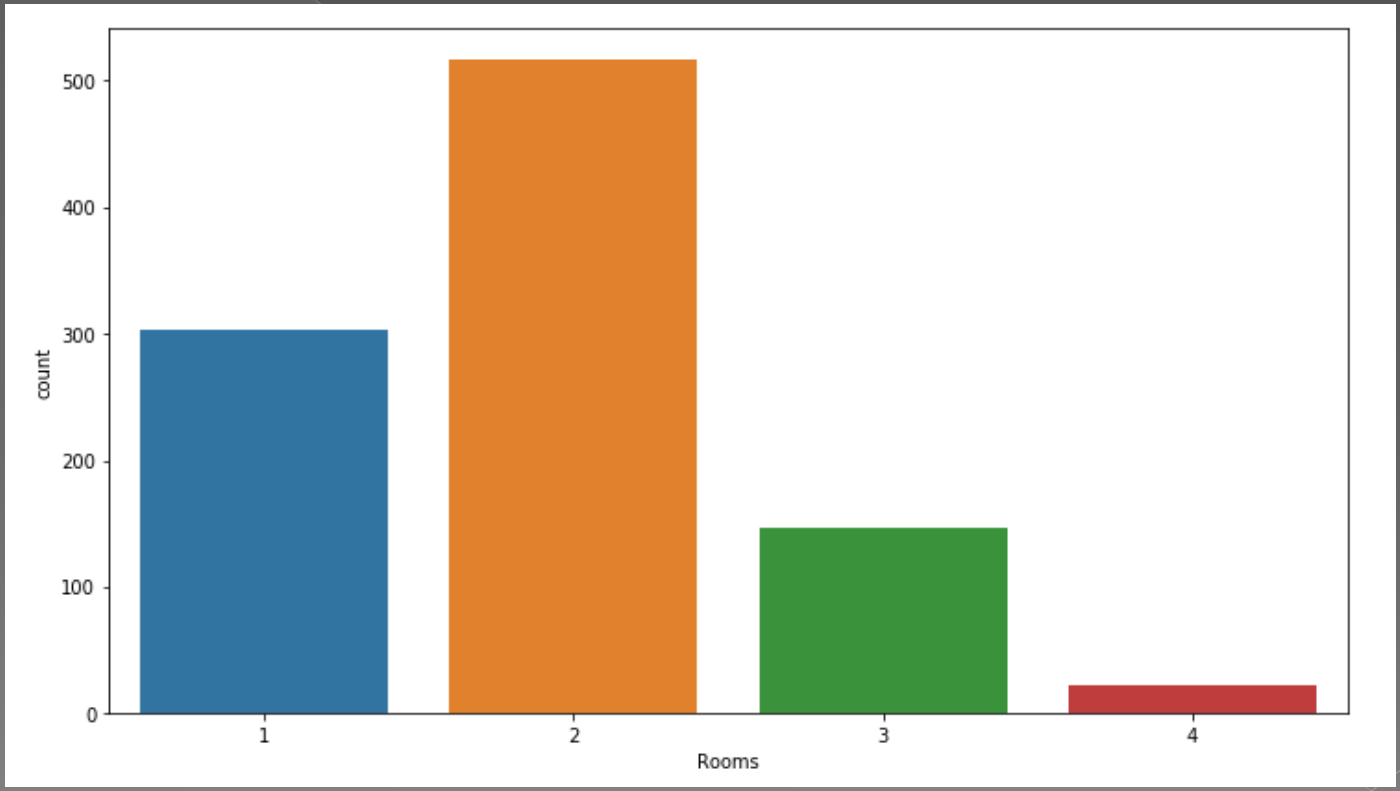
- If they are paying more than the average price for their apartment
- If there are similar districts to theirs with lower rents

Data

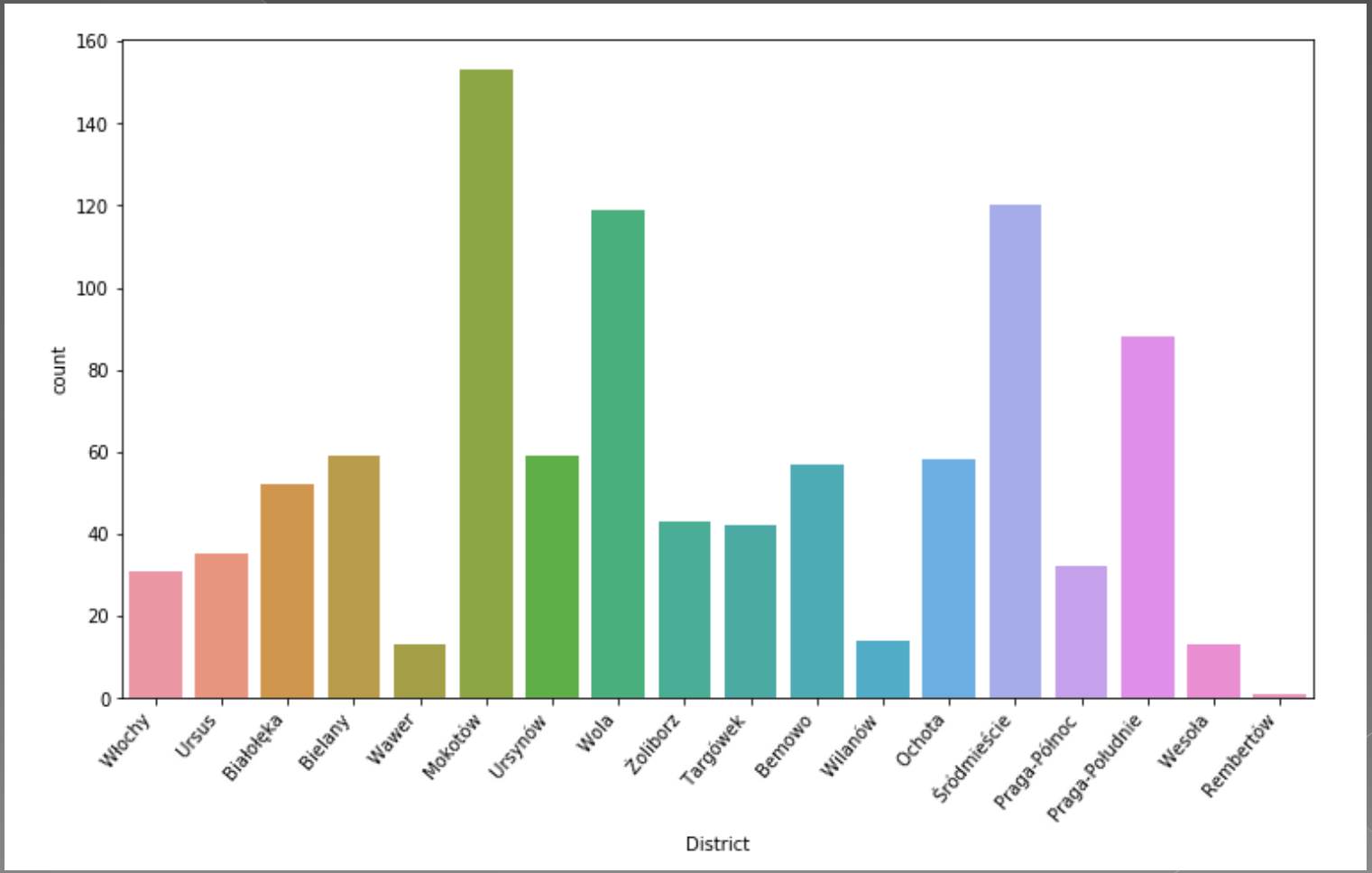
Data on apartments in Warsaw was collected by:

- scraping local website with advertisements www.olx.pl
- scraping Wikipedia
- Foursquare

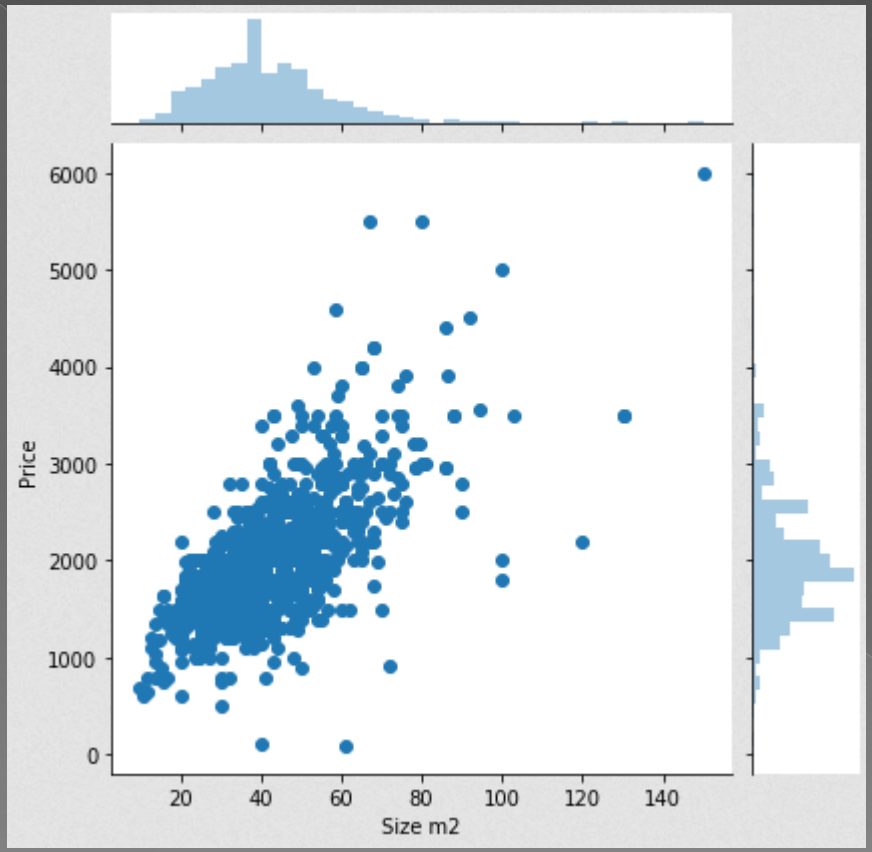
Data



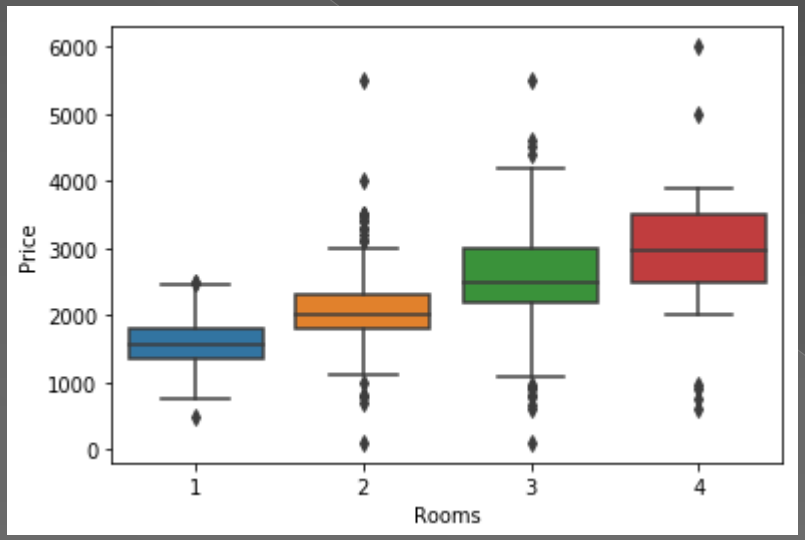
Data

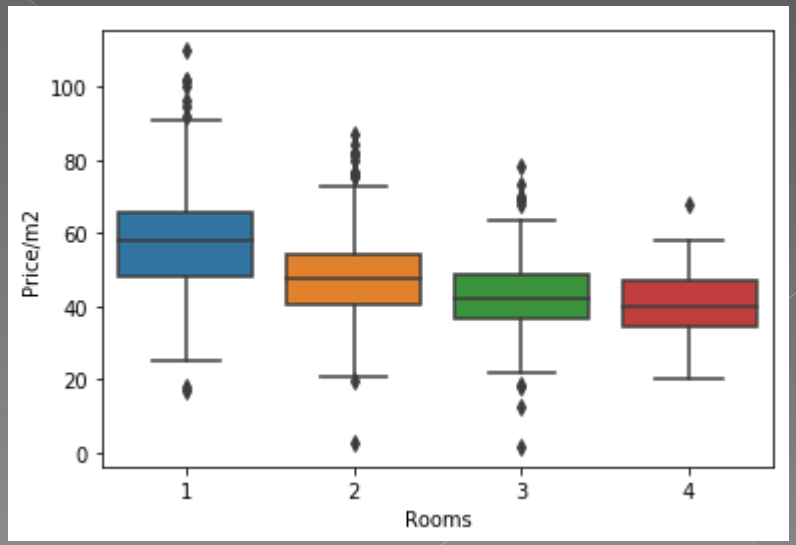

160
140
120
100
80
60
40
20
0
count
Włochy
Ursus
Białołęka
Bielany
Wawer
Mokotów
Ursynów
Wola
Żoliborz
Targówek
Bemowo
Wilanów
Ochota
Śródmieście
Praga-Północ
Praga-Południe
Wesoła
Rembertów
District

Data

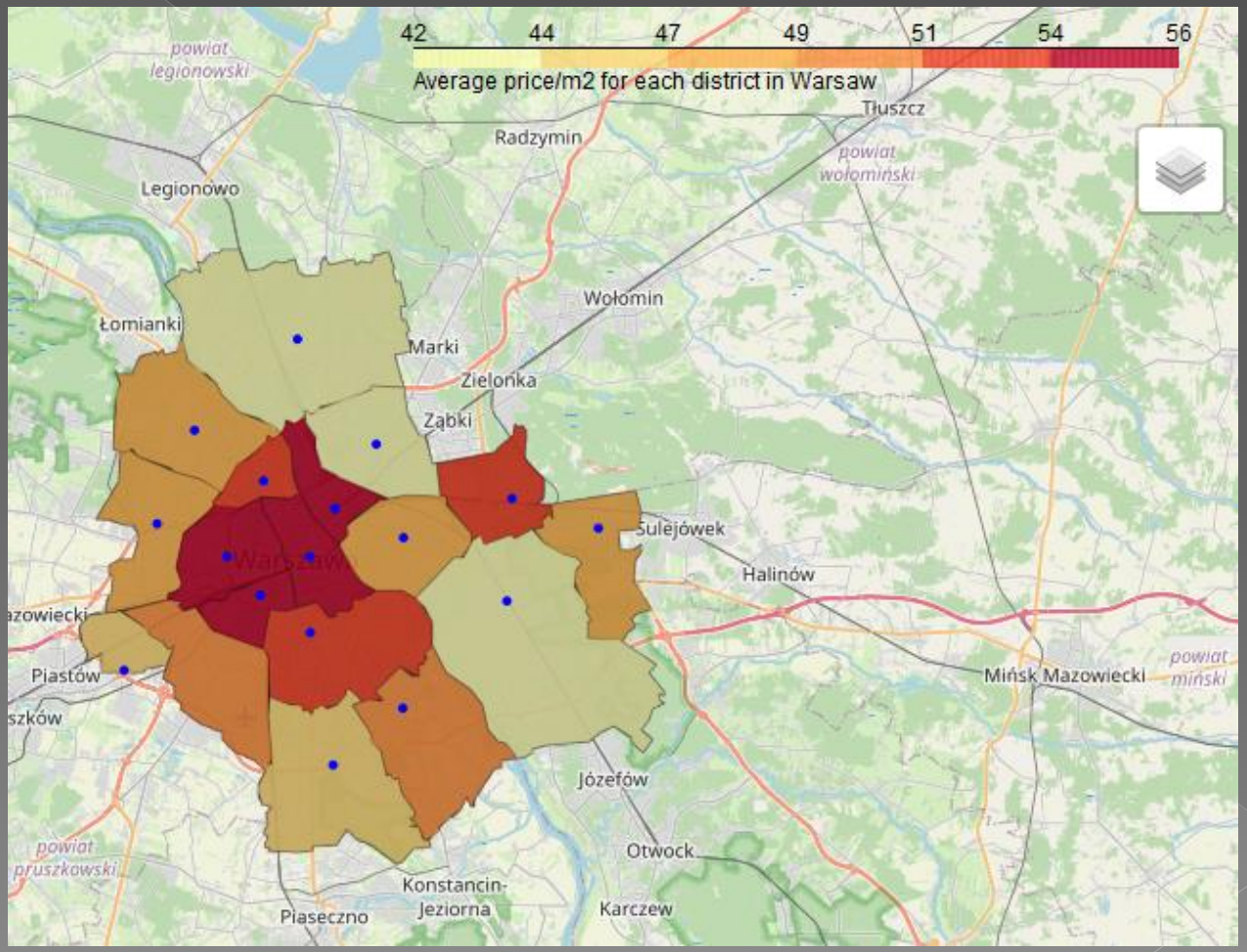

6000
5000
4000
3000
2000
1000
0
Price
20
40
60
80
100
120
140
Size m2

Data

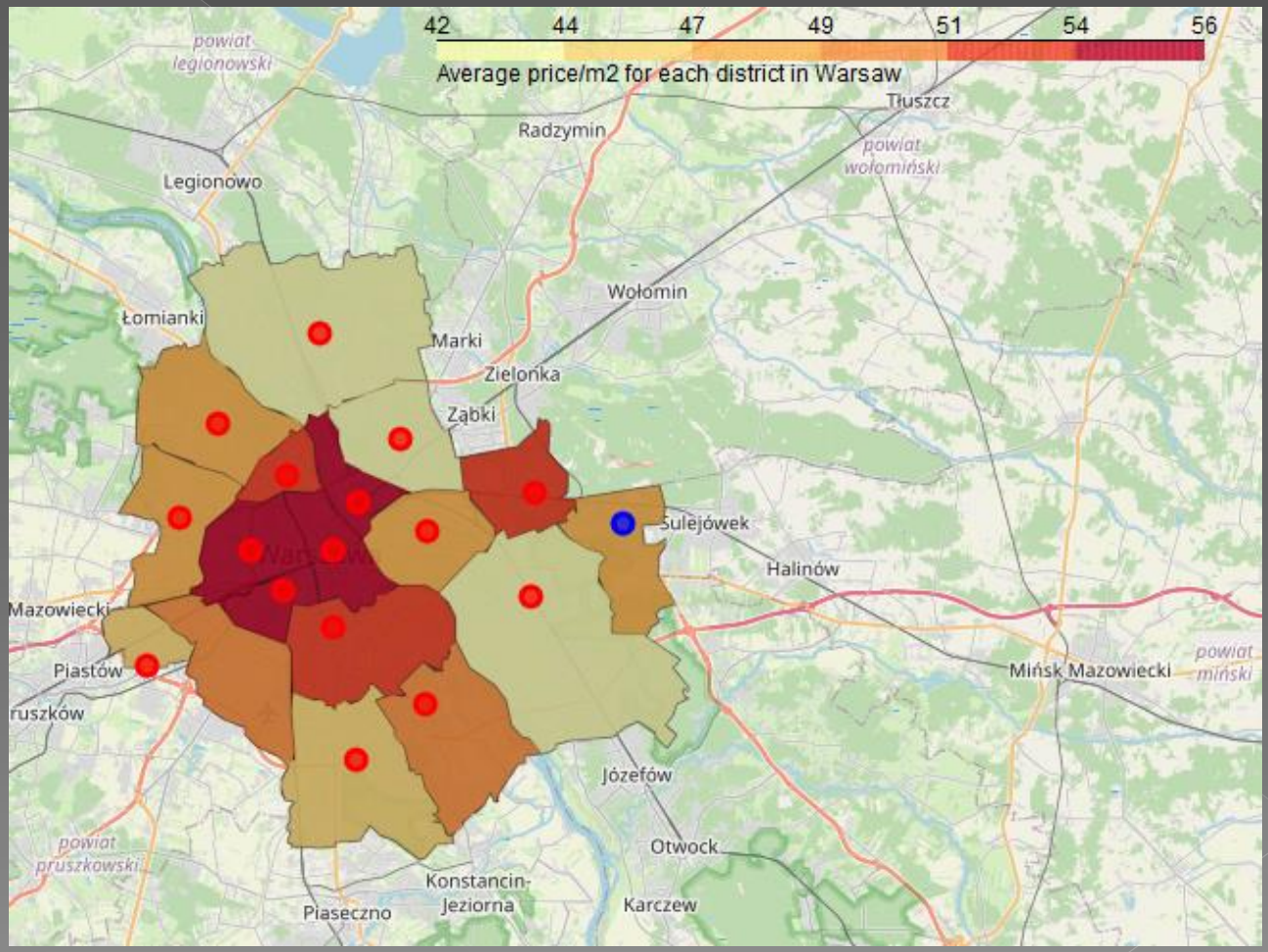

6000
5000
4000
3000
2000
1000
0
Price
1
2
3
4
Rooms


100
80
60
40
20
0
Price/m2
1
2
3
4
Rooms

Data


42
44
47
49
51
54
56
Average price/m2 for each district in Warsaw
powiat legionowski
Tłuszcz
Radzymin
powiat wołomiński
Legionowo
Wołomin
Łomianki
Marki
Zielonka
Ząbki
Sulejówek
Halinów
Piastów
Mińsk Mazowiecki
powiat miński
Józefów
Otwock
powiat pruszkowski
Konstancin-Jeziorna
Piaseczno
Karczew

Results


42
44
47
49
51
54
56
Average price/m2 for each district in Warsaw
powiat legionowski
Tłuszcz
Radzymin
powiat wołomiński
Legionowo
Wołomin
Łomianki
Marki
Zielonka
Ząbki
Sulejówek
Halinów
Mazowiecki
Piastów
Mińsk Mazowiecki
powiat miński
Józefów
Otwock
powiat pruszkowski
Konstancin-Jeziorna
Piaseczno
Karczew

Results

- This analysis of apartments in Warsaw was made on the basis of data collected from Foursquare, Wikipedia and the local website with advertisements www.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.
- 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 aquare meter and the top 10 venues for each district.