Coursera - Applied Data Science Capstone

The Battle of Neighborhoods

Rafał Śliż

March 29, 2021

Table of Contents

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 raport:

- 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 m^2 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.