

# Capstone Project – The Battle of Neighborhoods

---



## 1. Project description

Moving to Warsaw is the most common direction of migration in Poland, many people search for a good house price and good rating schools for their childrens. This project focuses on analyzing the features of neighborhoods and comparative analysis between them. The features include mean price per squared meter, school rating and occurring venues.

The aim of this project is to help choose the most suitable neighborhood and to get awerness of the place before moving in.

## 2. Location

Warsaw is the capital and the largest city of Poland. It is located in the center of the country, which makes it easily accessible from the other parts of Poland. The city's agglomeration is highly populated due to its economic potential and the best job

opportunities (the average salary in Warsaw is higher than anywhere in the country). Rich and diverse educational offer attracts people from all over the world.

### 3. Data

Informations about Warsaw's districts was obtained from **Wikipedia** [1], they contain the names of the districts, number of inhabitants, area and density of population. Geographical coordinates of the city and neighborhoods were obtained through the **Geopy** library. In the next step a **Foursuwar API** was used which, based on geographical coordinates, returned places in the area along with their exact location and category. Information about Warsaw school grades comes from <http://egzaminy.edu.pl/> [2]. Mean price per square meter comes from <https://nieruchomosci.dziennik.pl/> [3].

### 4. Methodology

#### Data cleaning

- Read data from html, csv and txt files
- Drop cells with missing data
- Rename columns
- Fix data types

#### Map

- Found coordinates of Warsaw and their neighborhoods
- Use Folium to mark our data on map

#### Folium venues

- For each neighborhood found 50 places in radius 1000m
- Group places by their category
- Found top 10 common venues in every neighborhood

## Additional data

- From 50 best school we pick the best one for every neighborhood
- Added mean price per square meter for every neighborhood
- Merge all data

## Cluster data

- Use K-mean with 6 clusters
- Create map with clusters
- Data analysis

## Data visualization

- Use matplotlib to plot charts

## 5. Results

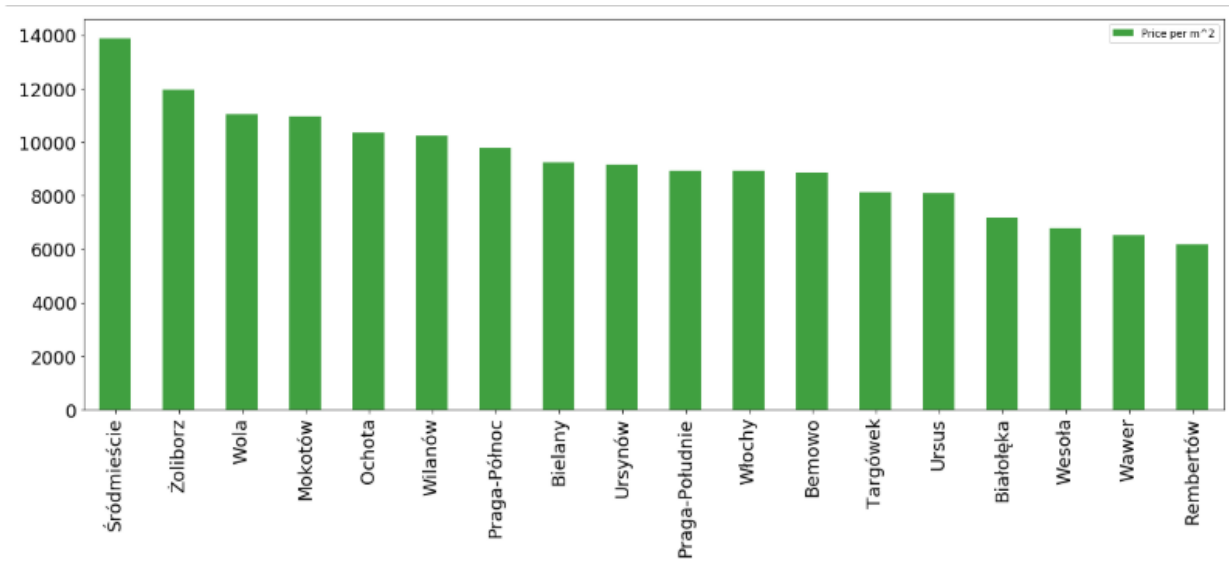


Fig 1. Mean price per m<sup>2</sup>.

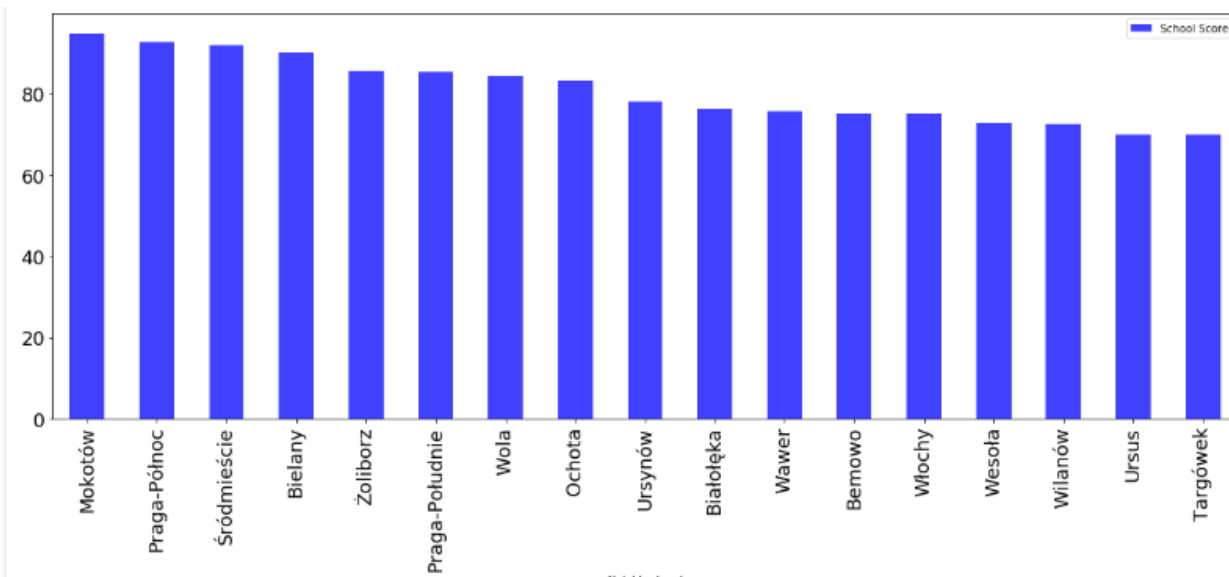


Fig 2. Best school rating.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Bemowo	Grocery Store	Pizza Place	Burger Joint	Middle Eastern Restaurant	Supermarket	Flea Market	Dessert Shop	Playground	Italian Restaurant	Pharmacy
1	Białoleka	Bus Station	Business Service	Train Station	Tennis Court	Zoo	Donut Shop	Eastern European Restaurant	Dumpling Restaurant	Drugstore	Discount Store
2	Bielany	Café	Bus Stop	Park	Bookstore	Paper / Office Supplies Store	Asian Restaurant	Tram Station	Ice Cream Shop	Indian Restaurant	Beer Bar
3	Mokotów	Café	Grocery Store	Italian Restaurant	Modern European Restaurant	Cupcake Shop	Burger Joint	Indian Restaurant	Kebab Restaurant	Farmers Market	Eastern European Restaurant
4	Ochota	Pharmacy	Pizza Place	Italian Restaurant	Hotel	Ice Cream Shop	Burger Joint	Pub	Electronics Store	Bus Station	General Entertainment
5	Praga-Południe	Supermarket	Hotel	Electronics Store	Park	Fast Food Restaurant	Farmers Market	Multiplex	Cosmetics Shop	Coffee Shop	Pharmacy
6	Praga-Północ	Zoo Exhibit	Fast Food Restaurant	Café	Music Venue	Diner	Plaza	Coffee Shop	Convenience Store	Bus Station	Bus Stop
7	Targówek	Bus Station	Market	Liquor Store	Pizza Place	Clothing Store	River	Donut Shop	Dumpling Restaurant	Drugstore	Diner
8	Ursus	Sporting Goods Shop	Park	Clothing Store	Lingerie Store	Supermarket	Chocolate Shop	Pizza Place	Ice Cream Shop	Food & Drink Shop	Hotel
9	Ursynów	Café	Pizza Place	Gym / Fitness Center	Supermarket	Sushi Restaurant	Burger Joint	Discount Store	Bus Line	Coffee Shop	Modern European Restaurant
10	Wawer	Bus Line	Bus Station	Flower Shop	Café	Pool	Donut Shop	Eastern European Restaurant	Dumpling Restaurant	Drugstore	Zoo
11	Wesoła	Pizza Place	Market	Grocery Store	Supermarket	Italian Restaurant	Café	Discount Store	Eastern European Restaurant	Dumpling Restaurant	Drugstore
12	Wilanów	Italian Restaurant	Café	Gym	Historic Site	Garden	Lake	Chinese Restaurant	Pharmacy	Pizza Place	Polish Restaurant
13	Wola	Pizza Place	Grocery Store	Chinese Restaurant	Italian Restaurant	Restaurant	Falafel Restaurant	Camera Store	Pub	Plaza	Bistro
14	Włochy	Supermarket	Hotel	Tram Station	Bed & Breakfast	Coffee Shop	Bus Station	Eastern European Restaurant	Restaurant	Bowling Alley	Shopping Mall
15	Śródmieście	Café	Bookstore	Dessert Shop	Hotel	Greek Restaurant	Hostel	Israeli Restaurant	Polish Restaurant	Plaza	Hotel Pool
16	Żoliborz	Park	Italian Restaurant	Beer Bar	Coffee Shop	Plaza	Café	Diner	Grocery Store	Cocktail Bar	Playground

Table 1. Most common venues

Basing on data presented on Fig 1,2 and Table 1 Kmean algorithm group neighborhoods in clusters presented in Fig 3.

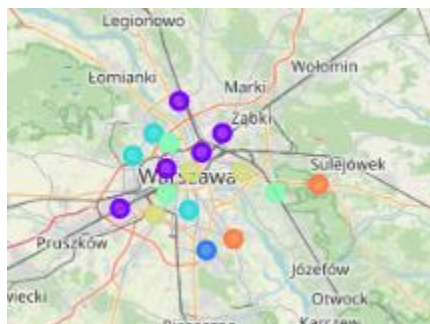


Fig 3. Map of clusters

	Neighborhood	Population	Area in km²	Density per km²	Latitude	Longitude	Cluster Labels	School Score	Price per m²	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Mokotów	217 683	6146	3542	52.19539	21.00850	3	94.90	10966.0	Café	Farmers Market	Modern European Restaurant	Italian Restaurant	Bistro
1	Praga-Południe	179 836	8036	2238	52.23633	21.09640	5	85.37	8936.0	Electronics Store	Hotel	Fast Food Restaurant	Coffee Shop	Park
2	Ursynów	150 668	3441	4379	52.15418	21.03786	2	78.06	9148.0	Café	Pizza Place	Burger Joint	Gym / Fitness Center	Supermarket
3	Wola	140 958	7319	1926	52.23903	20.97123	1	84.33	11024.0	Grocery Store	Pizza Place	Chinese Restaurant	Café	Italian Restaurant
4	Bielany	131 910	4079	3234	52.27697	20.94778	3	89.97	9225.0	Café	Park	Bookstore	Bus Stop	Grocery Store
5	Targówek	124 279	5131	2422	52.27726	21.06594	1	69.88	8114.0	Bus Station	Market	Pizza Place	Liquor Store	Clothing Store
6	Bemowo	123 932	4967	2495	52.25269	20.91244	3	75.20	8842.0	Italian Restaurant	Indian Restaurant	Plaza	Supermarket	Playground
7	Śródmieście	115 395	7411	1557	52.23560	21.01037	5	92.02	13888.0	Café	Hotel	Bookstore	Greek Restaurant	Coffee Shop
8	Białołęka	124 125	1699	7304	52.31097	20.99324	1	76.16	7180.0	Bus Station	Business Service	Tennis Court	Train Station	Zoo
9	Ochota	82 774	8516	972	52.21314	20.97069	4	83.22	10357.0	Electronics Store	Hotel	Italian Restaurant	Business Service	Department Store
10	Wawer	77 205	969	7970	52.21505	21.15758	4	75.71	6500.0	Bus Line	Pool	Bus Station	Gym	Café
11	Praga-Północ	64 113	5614	1142	52.25680	21.02976	1	92.72	9769.0	Zoo Exhibit	Fast Food Restaurant	Bus Station	Café	Plaza
12	Ursus	60 112	6422	936	52.19786	20.89114	1	69.97	8088.0	Park	Ice Cream Shop	Sporting Goods Shop	Fast Food Restaurant	Supermarket
13	Żoliborz	52 293	6174	847	52.26572	20.97483	4	85.57	11969.0	Park	Café	Italian Restaurant	Diner	Grocery Store
14	Włochy	42 862	1497	2863	52.19141	20.95212	5	75.15	8910.0	Hotel	Supermarket	Tram Station	Grocery Store	Bed & Breakfast
15	Wilanów	40 060	1091	3673	52.16566	21.08649	6	72.55	10229.0	Italian Restaurant	Gym	Café	Fast Food Restaurant	Steakhouse
16	Wesoła	25 439	1109	2294	52.22292	21.23074	6	72.78	6785.0	Market	Grocery Store	Pizza Place	Supermarket	Italian Restaurant

Table 2. Merged data

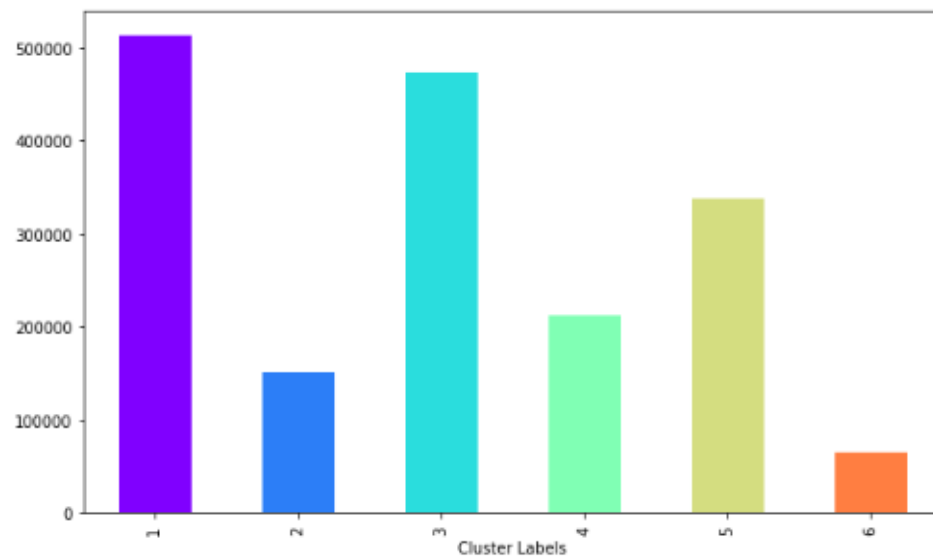


Fig 4. Population in each cluster.

## **6. Discussion**

In my opinion the best neighborhood for living contains Cluster 1. School rating in comparison with price of houses has the favorable ratio and easy access to public transport. The district stands out the most is Praga-Północ with second best school and price in the middle of the other districts

## **7. Conclusion**

In this project, through a K-means cluster algorithm we separate the neighborhood into 6 clusters. Using the charts above decision leading to a particular neighborhood based on average house prices and school rating can be made

## 8.Sources

[1] [https://pl.wikipedia.org/wiki/Podzia%C5%82\\_administracyjny\\_Warszawy](https://pl.wikipedia.org/wiki/Podzia%C5%82_administracyjny_Warszawy)

[2] <http://egzaminy.edu.pl/rekrutacja-do-liceum/rankingi-szkol-warszawskich/ranking-gimnazjow-publicznych-i-niepublicznych-w-warszawie-z-najlepszymi-wynikami-na-egzaminach-2016-r-portalu-egzaminy-edu-pl>

[3] <https://nieruchomosci.dziennik.pl/kupno-i-wynajem/artykuly/594650,ceny-mieszkan-kupie-mieszkanie-w-warszawie-nieruchomosci-dom-pieniadze.html>