Coursera

Applied Data Science Capstone by IBM

The Battle of Neighborhoods

Warsaw according to the attractiveness of each district

Iwona Czech

June 17, 2020

1. Introduction

1.1. Background

Warsaw is the capital and largest city of Poland. The metropolis stands on the Vistula River in east-central Poland and its population is officially estimated at 1.8 million residents within a greater metropolitan area of 3.1 million residents, which makes Warsaw the 7th most-populous capital city in the European Union. More and more people decides to move to Warsaw weather it's because of work or studies.

1.2. Interest

This project is mainly directed toward people who are planning to rent an apartment in Warsaw. Let's suppose you are that person. When moving to Warsaw you probably need to take into account few factors such as price, area, location of the apartment as well as venues in a particular area. Those kind of research would be also helpful for a flat owner who wants to rent a flat and need to set a rental price.

1.3. Problem

The main purpose of conducted analysis is to get an answers on the following questions:

1. Which district of Warsaw is the most expensive?
2. Which district has the most venues to offer?
3. Is there any correlation between these factors (price, number of venues)?
4. Which district would be an optimal choice?

2. Data

This project rely on public data from Wikipedia, Foursquare API, and a local website with apartment listings. In our analysis we also use a geoJSON file.

Data sources :

1. https://en.wikipedia.org/wiki/Districts_of_Warsaw

2. https://warszawa.nieruchomosci-online.pl

3. A geoJSON file of Warsaw districts.

4. Foursquare API.

For the Warsaw districts data, a Wikipedia page exists that has all the information we need to explore and cluster the neighborhoods in Warsaw. We will scrape the Wikipedia page and wrangle the data, clean it, and then read it into a pandas dataframe so that it is in a structured format. The data on apartments (size, district, and price) is collected by scraping a local website with apartment listings (https://warszawa.nieruchomosci-online.pl).

Using geopy we are able to find the coordinates for each district and then using Foursquare API we collect the closest venues (supermarket, restaurant, park, etc.). Finally, a geoJSON file of Warsaw districts, will be used to build the interactive maps in the right places.

3. Methodology

For convenience, all the libraries were installed and imported at the beginning. The entire analysis was carried out in the Jupyter Notebook using the methods learned during the course. Firstly, we imported the geoJSON file for the each district boundaries visualization. Secondly we scraped the Wikipedia page and as a result we received the following data.

	District	Population	Area
0	Mokotów	220682	35.4 km2 (13.7 sq mi)
1	Praga Południe	178665	22.4 km2 (8.6 sq mi)
2	Ursynów	145938	48.6 km2 (18.8 sq mi)
3	Wola	137519	19.26 km2 (7.44 sq mi)
4	Bielany	132683	32.3 km2 (12.5 sq mi)
5	Targówek	123278	24.37 km2 (9.41 sq mi)
6	Śródmieście	122646	15.57 km2 (6.01 sq mi)
7	Bemowo	115873	24.95 km2 (9.63 sq mi)
8	Białołęka	96588	73.04 km2 (28.20 sq mi)
9	Ochota	84990	09.7 km2 (3.7 sq mi)
10	Wawer	69896	79.71 km2 (30.78 sq mi)
11	Praga Północ	69510	11.4 km2 (4.4 sq mi)
12	Ursus	53755	09.35 km2 (3.61 sq mi)
13	Żoliborz	48342	08.5 km2 (3.3 sq mi)
14	Włochy	38075	28.63 km2 (11.05 sq mi)
15	Wilanów	23960	36.73 km2 (14.18 sq mi)
16	Rembertów	23280	19.30 km2 (7.45 sq mi)
17	Wesoła	22811	22.6 km2 (8.7 sq mi)

There are 18 rows in a table, each of them corresponds to specific district. The first column holds the name of each district, the second column contains number of population, the third one tells us how big is the district area. In the next step we have found geological coordinates (latitude, longitude) for each district. We created 2 new columns which correspond to this values.

	District	Population	Area	Latitude	Longitude
0	Mokotów	220682	35.4 km2 (13.7 sq mi)	52.193987	21.045781
1	Praga Południe	178665	22.4 km2 (8.6 sq mi)	52.237396	21.071258
2	Ursynów	145938	48.6 km2 (18.8 sq mi)	52.141039	21.032321
3	Wola	137519	19.26 km2 (7.44 sq mi)	52.236238	20.954781
4	Bielany	132683	32.3 km2 (12.5 sq mi)	52.285043	20.943949
5	Targówek	123278	24.37 km2 (9.41 sq mi)	52.275192	21.058085
6	Śródmieście	122646	15.57 km2 (6.01 sq mi)	52.232810	21.019067
7	Bemowo	115873	24.95 km2 (9.63 sq mi)	52.238974	20.913288
8	Białołęka	96588	73.04 km2 (28.20 sq mi)	52.319665	21.021177
9	Ochota	84990	09.7 km2 (3.7 sq mi)	52.212225	20.972630
10	Wawer	69896	79.71 km2 (30.78 sq mi)	52.220358	21.137083
11	Praga Północ	69510	11.4 km2 (4.4 sq mi)	52.264884	21.027344
12	Ursus	53755	09.35 km2 (3.61 sq mi)	52.196098	20.882899
13	Żoliborz	48342	08.5 km2 (3.3 sq mi)	52.267594	20.979698
14	Włochy	38075	28.63 km2 (11.05 sq mi)	52.186109	20.948438
15	Wilanów	23960	36.73 km2 (14.18 sq mi)	52.153083	21.110441
16	Rembertów	23280	19.30 km2 (7.45 sq mi)	52.261415	21.162819
17	Wesoła	22811	22.6 km2 (8.7 sq mi)	52.251794	21.229276

Then we Scraped https://warszawa.nieruchomosci-online.pl for apartments data in Warsaw. We extracted relevant data (district, area, and price) by looping through all pages until no more listings are found.

	Area	District	Price
0	102	MokotówWarszawa	5000
1	55	MokotówWarszawa	2000
2	45,28	BiałołękaWarszawa	1700
3	110	UrsynówWarszawa	4990
4	71,60	Miasteczko	2500
...	...	...	...
1491	70	MokotówWarszawa	3800
1492	50	KsawerówWarszawa	2800
1493	58	WolaWarszawa	3700
1494	46	MokotówWarszawa	3300
1495	53	MokotówWarszawa	2400

1496 rows × 3 columns

There were 1496 apartments in the dataset. In the next step we cleaned the data. After cleaning there were 1269 positions left in a dataset.

	Area_x	District	Price	Population	Area_y	Latitude	Longitude
0	102	Mokotów	5000	220682	35.40	52.193987	21.045781
1	55	Mokotów	2000	220682	35.40	52.193987	21.045781
2	66	Mokotów	4500	220682	35.40	52.193987	21.045781
3	72,43	Mokotów	4400	220682	35.40	52.193987	21.045781
4	50	Mokotów	2900	220682	35.40	52.193987	21.045781
...	...	...	...	...	...	...	...
1264	59	Włochy	2800	38075	28.63	52.186109	20.948438
1265	80	Ursus	3000	53755	9.35	52.196098	20.882899
1266	47	Ursus	2200	53755	9.35	52.196098	20.882899
1267	71	Ursus	2400	53755	9.35	52.196098	20.882899
1268	58	Ursus	2500	53755	9.35	52.196098	20.882899

1269 rows × 7 columns

In the next step we calculated the average price and property area for each of the district.

	District	Price	Latitude	Longitude	property_area
0	Bemowo	2785.238095	52.238974	20.913288	64.714286
1	Białołęka	2477.777778	52.319665	21.021177	55.793333
2	Bielany	2830.344828	52.285043	20.943949	66.586207
3	Mokotów	3771.651376	52.193987	21.045781	65.265260
4	Ochota	3246.527273	52.212225	20.972630	64.560000
5	Praga Południe	3063.958333	52.237396	21.071258	53.937500
6	Praga Północ	2655.652174	52.264884	21.027344	49.217391
7	Targówek	2630.952381	52.275192	21.058085	57.809524
8	Ursus	2525.000000	52.196098	20.882899	64.000000
9	Ursynów	2973.085106	52.141039	21.032321	65.957979
10	Wawer	2957.142857	52.220358	21.137082	90.500000
11	Wesoła	1600.000000	52.251794	21.229276	41.500000
12	Wilanów	3915.890411	52.153083	21.110441	72.050685
13	Wola	3465.747126	52.236238	20.954782	56.866667
14	Włochy	3856.125000	52.186109	20.948438	100.250000
15	Śródmieście	4617.880952	52.232810	21.019067	69.262075
16	Żoliborz	3035.306452	52.267594	20.979698	54.161290

Then we used FourSquare API, to get details of venues in a particular area. We got the results as follows.

	District	District Latitude	District Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Mokotów	52.193987	21.045781	Green Caffè Nero	52.198116	21.047283	Café
1	Mokotów	52.193987	21.045781	4fun.tv	52.196089	21.046074	Arcade
2	Mokotów	52.193987	21.045781	Sikorskiego	52.192557	21.047922	Skate Park
3	Mokotów	52.193987	21.045781	Tor stegny 02	52.191261	21.046389	Bus Station
4	Mokotów	52.193987	21.045781	Restauracja Giovanni Sport	52.190349	21.045084	Diner
...	...	...	...	...	...	...	...
211	Wilanów	52.153083	21.110441	La millou	52.151070	21.113367	Baby Store
212	Wesoła	52.251794	21.229276	Korty Tenisowe Wesoła	52.250449	21.224401	Tennis Court
213	Wesoła	52.251794	21.229276	Klimat	52.253798	21.223714	Pizza Place
214	Wesoła	52.251794	21.229276	Warszawa Wesoła	52.253798	21.223625	Train Station
215	Wesoła	52.251794	21.229276	Rondo w Wesołej	52.250317	21.222985	Plaza

216 rows × 7 columns

	District	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	Food & Drink Shop	Tram Station	Bus Station	Supermarket	Italian Restaurant	Sandwich Place	Café	Sporting Goods Shop	Donut Shop	Coffee Shop
1	Białołęka	Business Service	Dim Sum Restaurant	Clothing Store	Cocktail Bar	Coffee Shop	Comedy Club	Convenience Store	Cupcake Shop	Deli / Bodega	Department Store
2	Bielany	Park	Bus Station	Gym	Drugstore	Ice Cream Shop	Metro Station	Coffee Shop	Pharmacy	Flea Market	Basketball Court
3	Mokotów	Café	Skate Park	Arcade	Bus Station	Diner	Lake	Cocktail Bar	Coffee Shop	Comedy Club	Convenience Store
4	Ochota	Hotel	Pharmacy	Basketball Court	Business Service	Park	Burger Joint	Polish Restaurant	Italian Restaurant	Department Store	Dessert Shop
5	Praga Południe	Café	Bus Line	Road	Ice Cream Shop	Dessert Shop	Clothing Store	Cocktail Bar	Coffee Shop	Comedy Club	Convenience Store
6	Praga Północ	Bike Rental / Bike Share	Tram Station	Bus Station	Pharmacy	Comedy Club	Drugstore	Convenience Store	Basketball Court	Plaza	Light Rail Station
7	Targówek	Bus Station	Pet Store	Flea Market	Gym / Fitness Center	Cupcake Shop	Dessert Shop	Clothing Store	Cocktail Bar	Coffee Shop	Comedy Club
8	Ursus	Park	Train Station	Supermarket	Hotel	Dessert Shop	Chocolate Shop	Clothing Store	Cocktail Bar	Coffee Shop	Comedy Club
9	Ursynów	Grocery Store	Supermarket	Italian Restaurant	Dessert Shop	Food Court	Plaza	Bus Stop	Food & Drink Shop	Bar	Coffee Shop
10	Wawer	Gun Range	Food & Drink Shop	Athletics & Sports	Falafel Restaurant	Dim Sum Restaurant	Cocktail Bar	Coffee Shop	Comedy Club	Convenience Store	Cupcake Shop
11	Wesoła	Train Station	Pizza Place	Tennis Court	Plaza	Yoga Studio	Department Store	Chocolate Shop	Clothing Store	Cocktail Bar	Coffee Shop
12	Wilanów	Baby Store	Yoga Studio	Diner	Cocktail Bar	Coffee Shop	Comedy Club	Convenience Store	Cupcake Shop	Deli / Bodega	Department Store
13	Wola	Grocery Store	Café	Soccer Field	Motorcycle Shop	Music Store	Bus Station	Falafel Restaurant	Park	Italian Restaurant	Playground
14	Włochy	Accessories Store	Bed & Breakfast	Hot Dog Joint	Hotel	Paper / Office Supplies Store	Bistro	Café	Bakery	Bar	Comedy Club
15	Śródmieście	Café	Beer Bar	Coffee Shop	Cocktail Bar	Sushi Restaurant	Ice Cream Shop	Pizza Place	Bakery	Nightclub	Mediterranean Restaurant

Next we clustered the districts, later on we examined them.

Cluster 1

```
cluster1 = warsaw_merged.loc[warsaw_merged['Cluster Labels'] == 0, warsaw_merged.columns[[1] + list(range(5, warsaw_
cluster1
```

	District	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	Price	Latitude	Longitude	property_area
12	Wilanów	Cocktail Bar	Coffee Shop	Comedy Club	Convenience Store	Cupcake Shop	Deli / Bodega	Department Store	3915.890411	52.153083	21.110441	72.050685

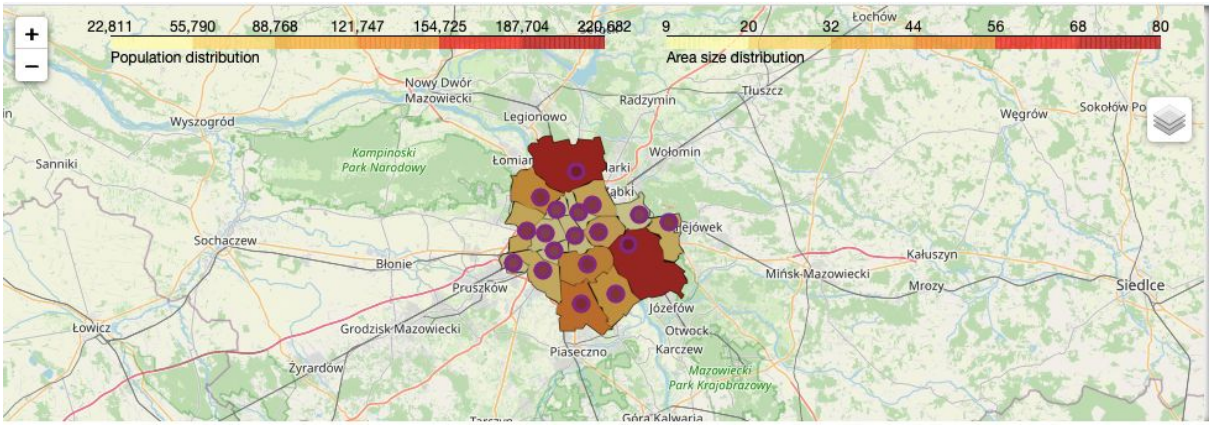
Cluster 2

```
cluster2 = warsaw_merged.loc[warsaw_merged['Cluster Labels'] == 1, warsaw_merged.columns[[1] + list(range(5, warsaw_
cluster2
```

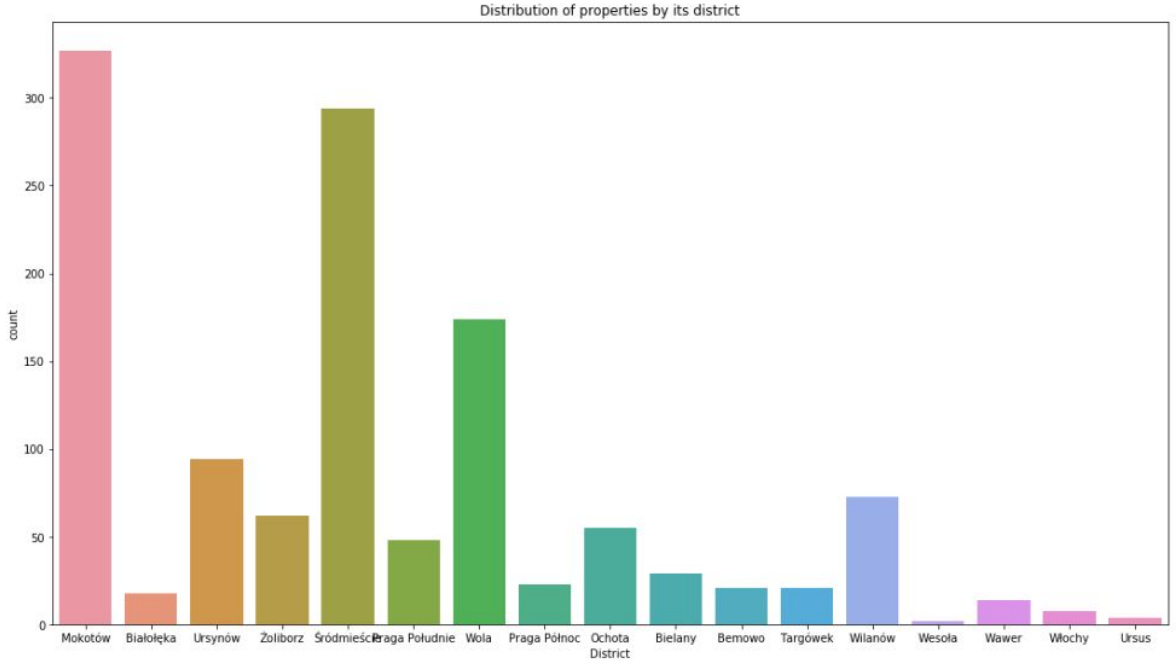
	District	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	Price	Latitude	Longitude	property_area
0	Bemowo	Supermarket	Italian Restaurant	Sandwich Place	Café	Sporting Goods Shop	Donut Shop	Coffee Shop	2785.238095	52.238974	20.913288	64.714286
1	Białołęka	Cocktail Bar	Coffee Shop	Comedy Club	Convenience Store	Cupcake Shop	Deli / Bodega	Department Store	2477.777778	52.319665	21.021177	55.793333
2	Bielany	Drugstore	Ice Cream Shop	Metro Station	Coffee Shop	Pharmacy	Flea Market	Basketball Court	2830.344828	52.285043	20.943949	66.586207
3	Mokotów	Bus Station	Diner	Lake	Cocktail Bar	Coffee Shop	Comedy Club	Convenience Store	3771.651376	52.193987	21.045781	65.265260
4	Ochota	Business Service	Park	Burger Joint	Polish Restaurant	Italian Restaurant	Department Store	Dessert Shop	3246.527273	52.212225	20.972630	64.560000
5	Praga Południe	Ice Cream Shop	Dessert Shop	Clothing Store	Cocktail Bar	Coffee Shop	Comedy Club	Convenience Store	3063.958333	52.237396	21.071258	53.937500
6	Praga Północ	Pharmacy	Comedy Club	Drugstore	Convenience Store	Basketball Court	Plaza	Light Rail Station	2655.652174	52.264884	21.027344	49.217391
7	Targówek	Gym / Fitness Center	Cupcake Shop	Dessert Shop	Clothing Store	Cocktail Bar	Coffee Shop	Comedy Club	2630.952381	52.275192	21.058085	57.809524
8	Ursus	Hotel	Dessert Shop	Chocolate Shop	Clothing Store	Cocktail Bar	Coffee Shop	Comedy Club	2525.000000	52.196098	20.882899	64.000000
9	Ursynów	Dessert Shop	Food Court	Plaza	Bus Stop	Food & Drink Shop	Bar	Coffee Shop	2973.085106	52.141039	21.032321	65.957979
10	Wawer	Falafel Restaurant	Dim Sum Restaurant	Cocktail Bar	Coffee Shop	Comedy Club	Convenience Store	Cupcake Shop	2957.142857	52.220358	21.137082	90.500000
11	Wesoła	Plaza	Yoga Studio	Department Store	Chocolate Shop	Clothing Store	Cocktail Bar	Coffee Shop	1600.000000	52.251794	21.229276	41.500000
13	Wola	Motorcycle Shop	Music Store	Bus Station	Falafel Restaurant	Park	Italian Restaurant	Playground	3465.747126	52.236238	20.954782	56.866667
14	Włochy	Hotel	Paper / Office Supplies Store	Bistro	Café	Bakery	Bar	Comedy Club	3856.125000	52.186109	20.948438	100.250000
15	Śródmieście	Cocktail Bar	Sushi Restaurant	Ice Cream Shop	Pizza Place	Bakery	Nightclub	Mediterranean Restaurant	4617.880952	52.232810	21.019067	69.262075
16	Żoliborz	Plaza	Coffee Shop	Burger Joint	Bakery	Italian Restaurant	Metro Station	Deli / Bodega	3035.306452	52.267594	20.979698	54.161290

4. Results

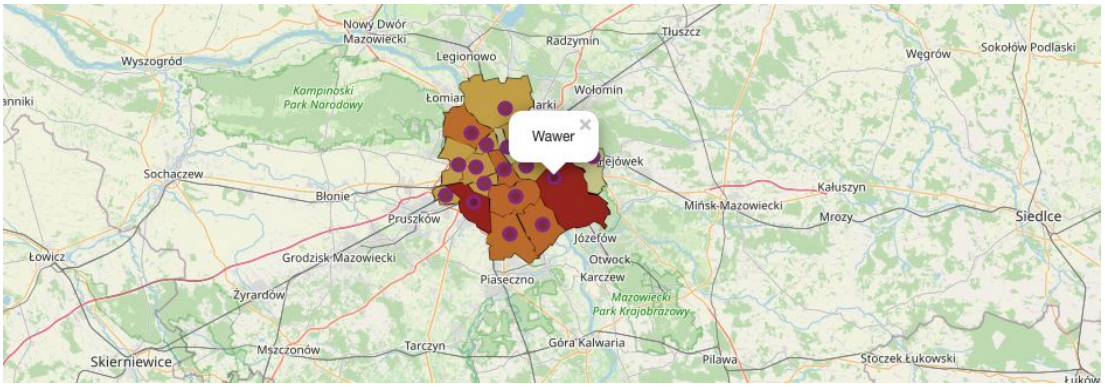
Vizualization all of the districts of Warsaw by its location, population and area size



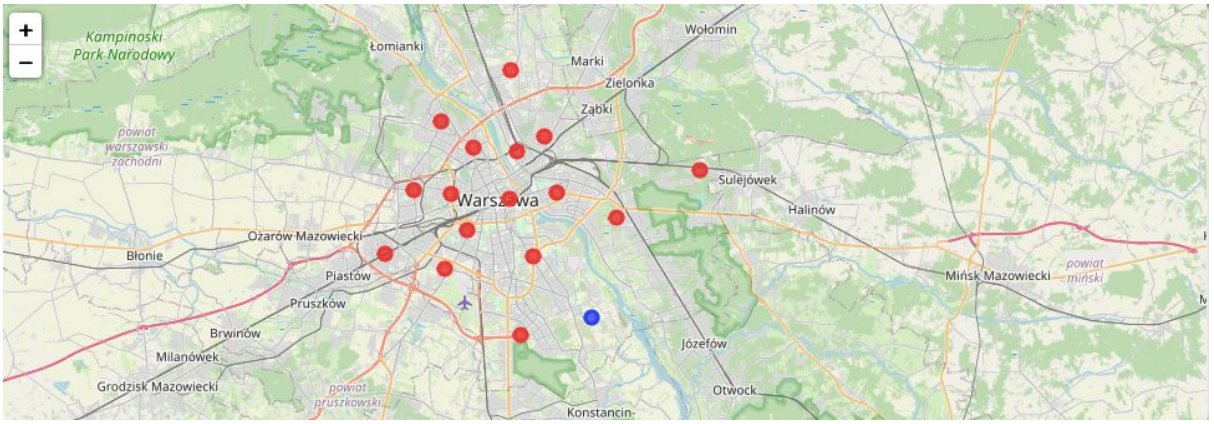
Distribution of properties by its district



Visualize the average price & average apartment size per district on a choropleth map



The above screen shows districts clustered into 2 groups (Cluster 1, Cluster 2).



5. Conclusion

In this study, we analyzed the districts of Warsaw. This paperwork helped us answer the following questions:

1. Which district of Warsaw is the most expensive?

Śródmieście

2. Which district has the most venues to offer?

Śródmieście

3. Is there any correlation between these factors (price, number of venues)?

Yes, the more venues in the district area, the more we need to pay for an apartment.

4. Which district would be an optimal choice?

Śródmieście is the most attractive district to live because of its number of venues. Unfortunately it's also the most expensive one, that's why a person who wants to rent a flat should consider one of the district from **CLUSTER 2,** it will be cheaper but still attractive.