IBM Data Science Professional Certificate
Capstone Project Report

Aykut Atakan

05 JUN 2019

Introduction and Business Problem

- Ekrem wants to open a burger restaurant in Warsaw, Poland.
- Warsaw has 18 districts and we are deciding which districts is more suitable for Ekrem to open the restaurant
- Ekrem wants to consider other burger restaurants, number of high schools, number of universities and number of Office in determining the location.

Data Sources

- Districts are taken from Wikipedia.org. To be precise, the link is below.

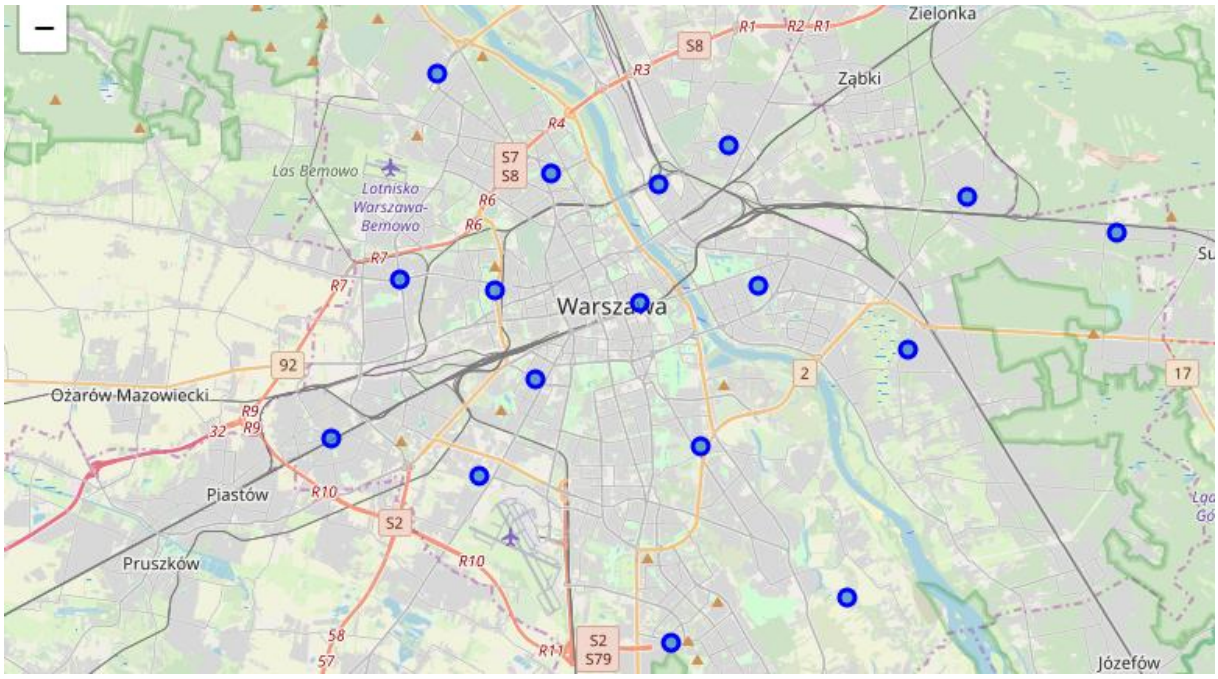
https://en.wikipedia.org/wiki/Districts_of_Warsaw

- The coordinates (latitude, longitude) of these districts of Warsaw are from Open Street Map APIs
- Following venue data is pulled from foursquare API.
    - Burger restaurants in Warsaw
    - High schools in Warsaw
    - Universities in Warsaw
    - Offices in Warsaw

Methodology

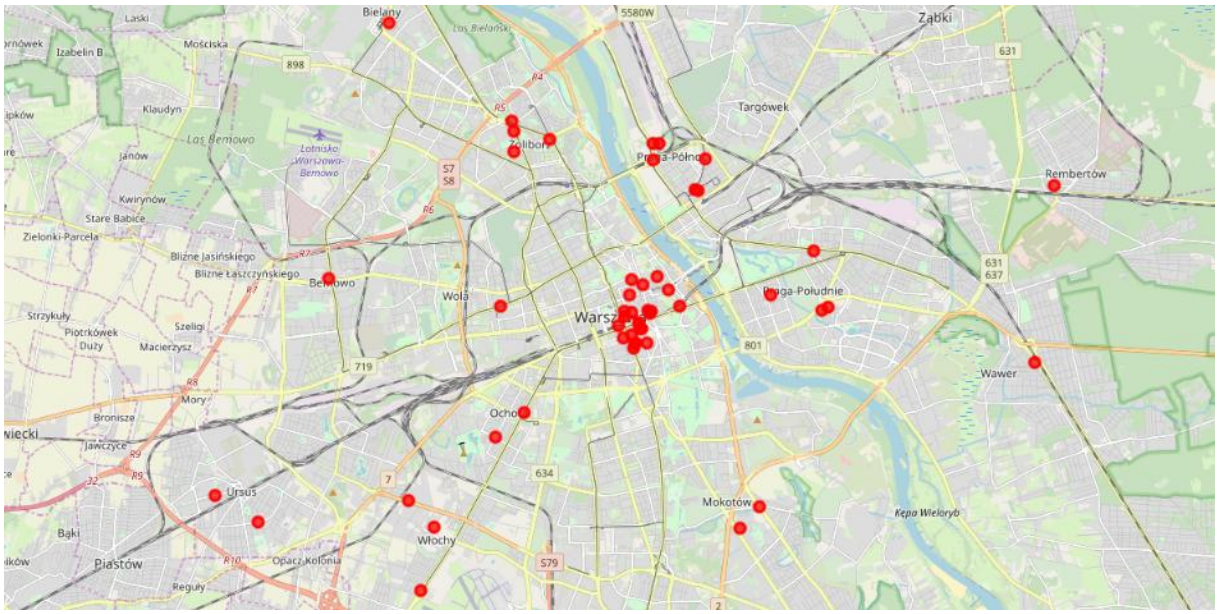
- For each locality, all office, school, university and burger restaurant venues data have been collected from Foursquare.
- Then for each locality, the sums of the office, school, university and burger joints were computed.
- For each of this 4 categories, a weight (or penalty) has been defined according to what Ekrem considers the most important.
    - Burger Joints have been weighted with -1, since Ekrem wants to avoid concurrence.
    - Schools have been weighted with 1, since student are good customers.
    - Universities have been weighted with 1.5, since students are good customers.
    - Offices have been weighted with 2, since employees are even better customers.
- Note that the weights can be modified according to the importance of each category.
- Lastly, a score was computed for each locality as the weighted sum of the number of venues in each of the 4 categories (school, university, office, burger joints).

Districts of Warsaw, Poland

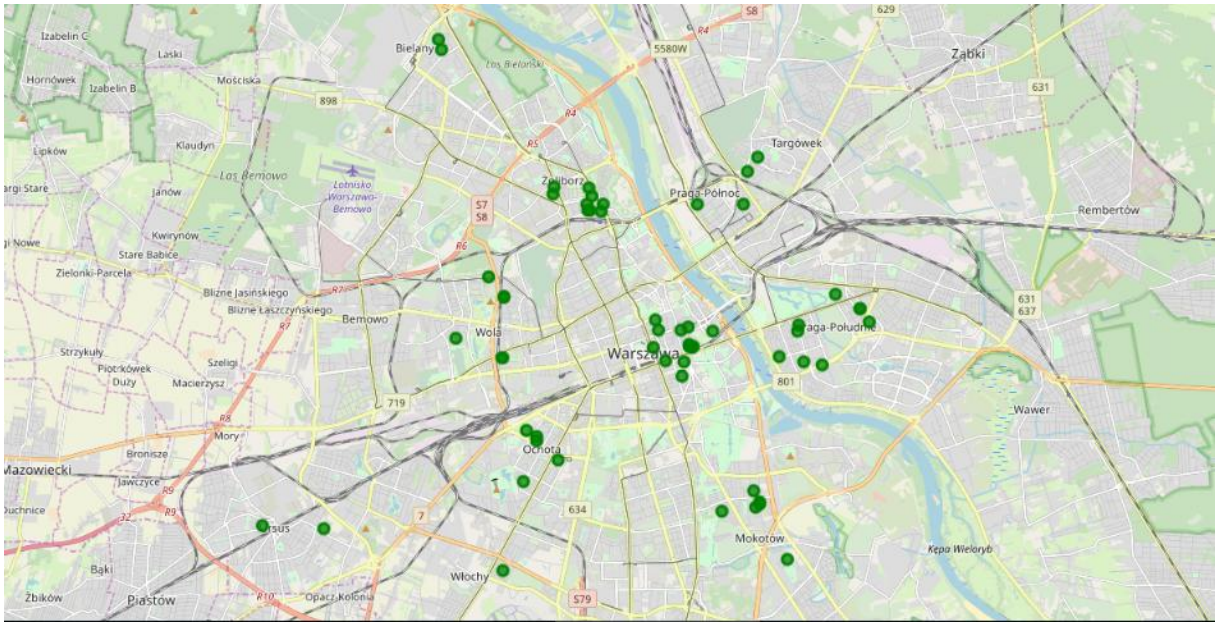


	Districts	Latitude	Longitude
0	Mokotów	52.193987	21.045781
1	Praga Południe	52.237396	21.071258
2	Ursynów	52.141039	21.032321
3	Wola	52.236238	20.954781
4	Bielany	52.294652	20.929980
5	Targówek	52.275192	21.058085
6	Śródmieście, Warsaw	52.232810	21.019067
7	Bemowo	52.238974	20.913288
8	Białołęka	52.319665	21.021177
9	Ochota	52.212225	20.972630
10	Wawer	52.220358	21.137083
11	Praga Północ	52.264884	21.027344
12	Ursus, Warsaw	52.196098	20.882899
13	Żoliborz	52.267594	20.979698
14	Włochy	52.186109	20.948438
15	Wilanów	52.153083	21.110441
16	Rembertów	52.261415	21.162819
17	Wesoła	52.251794	21.229276

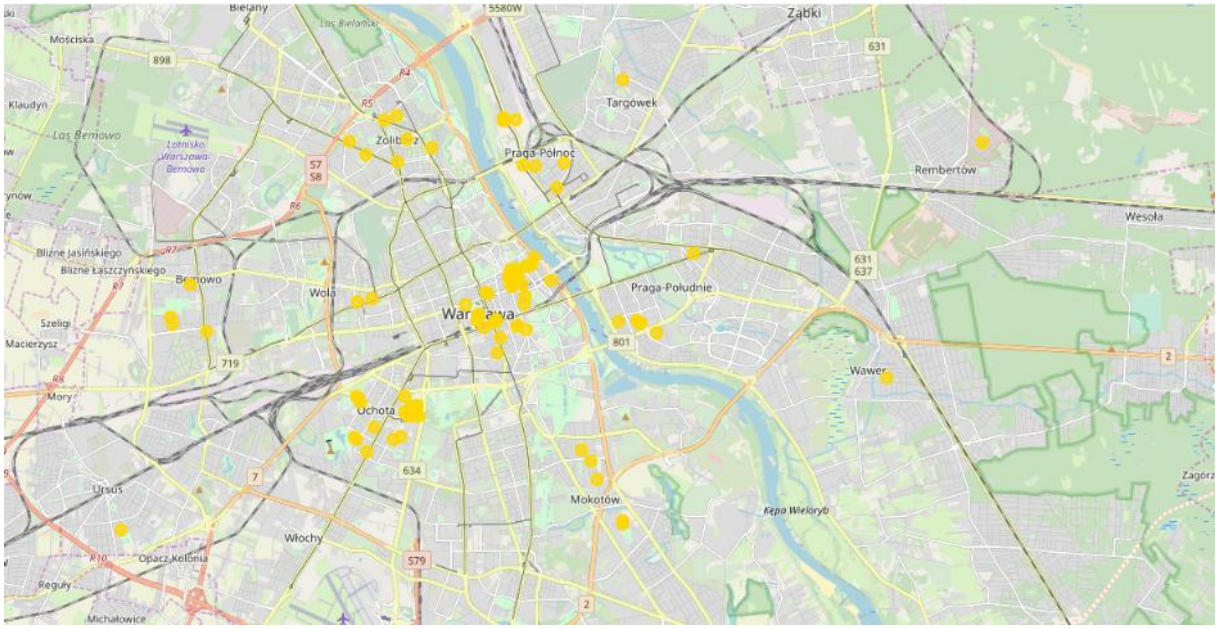
Burger Restaurants



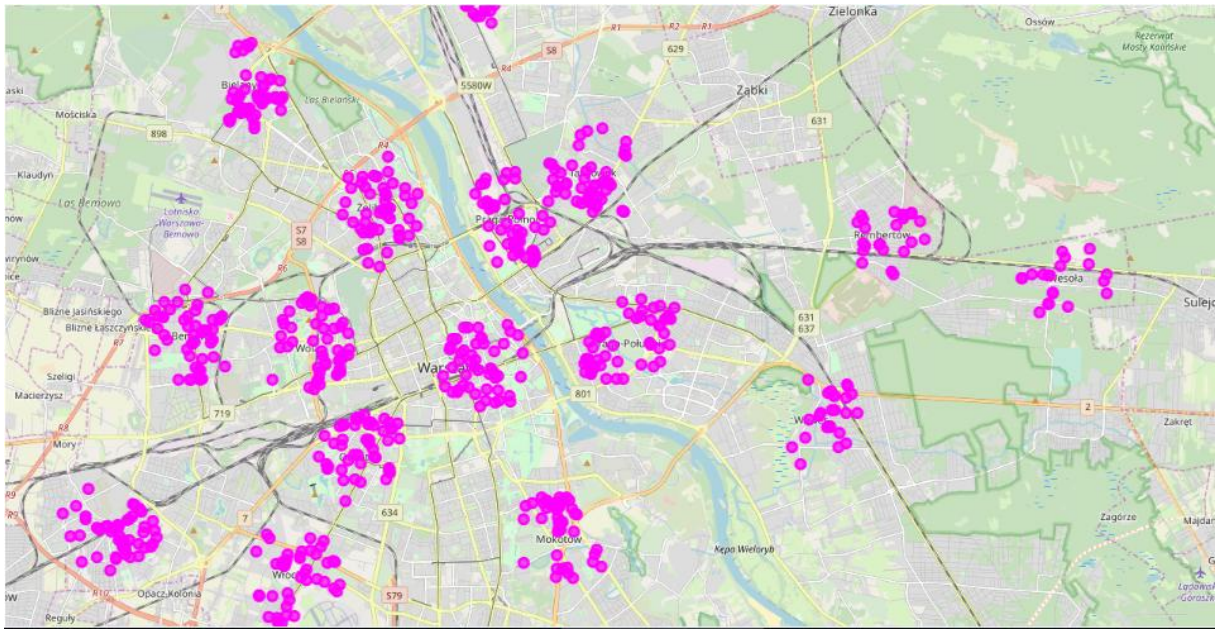
Highschools



Universities



Offices



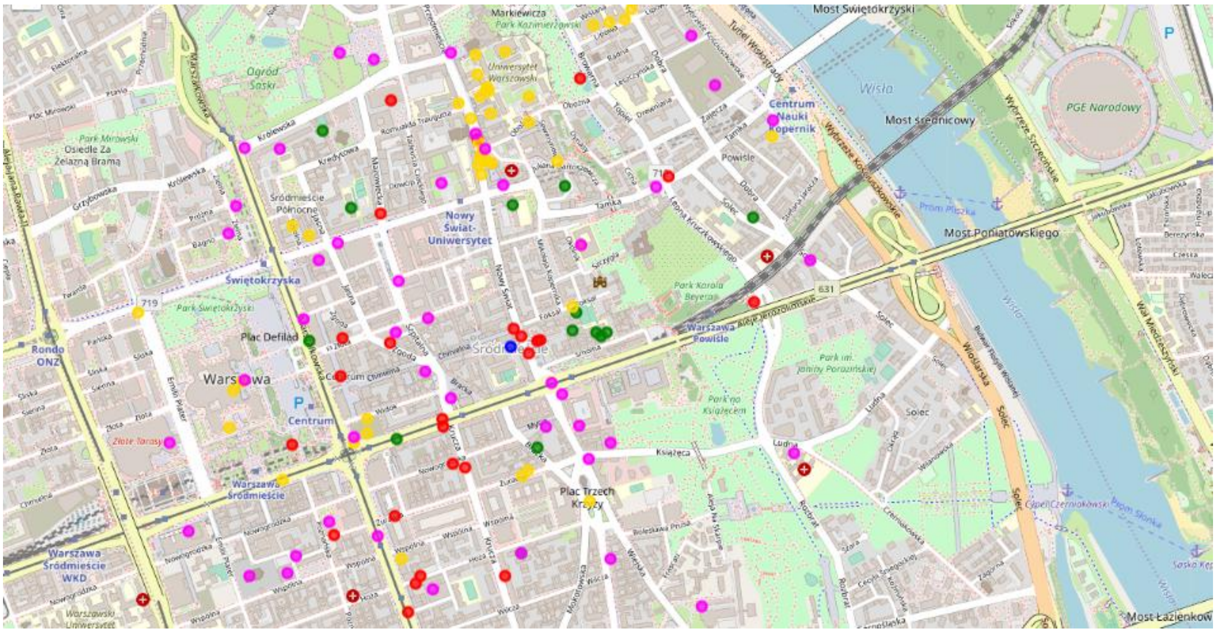
Breakdowsn of the high schools, universities, burger restaurants and offices by districts

	Districts	Latitude	Longitude	Burger	High Schools	Universities	Offices
0	Mokotów	52.193987	21.045781	2.0	6.0	5.0	47.0
1	Praga Południe	52.237396	21.071258	4.0	9.0	6.0	49.0
2	Ursynów	52.141039	21.032321	1.0	2.0	4.0	48.0
3	Wola	52.236238	20.954781	1.0	6.0	2.0	49.0
4	Bielany	52.294652	20.929980	1.0	2.0	0.0	49.0
5	Targówek	52.275192	21.058085	0.0	1.0	1.0	48.0
6	Śródmieście, Warsaw	52.232810	21.019067	25.0	14.0	44.0	50.0
7	Bemowo	52.238974	20.913288	1.0	0.0	4.0	49.0
8	Białołęka	52.319665	21.021177	0.0	0.0	0.0	23.0
9	Ochota	52.212225	20.972630	2.0	5.0	19.0	50.0
10	Wawer	52.220358	21.137083	1.0	0.0	1.0	27.0
11	Praga Północ	52.264884	21.027344	6.0	3.0	9.0	49.0
12	Ursus, Warsaw	52.196098	20.882899	2.0	2.0	1.0	47.0
13	Żoliborz	52.267594	20.979698	4.0	9.0	7.0	50.0
14	Włochy	52.186109	20.948438	3.0	1.0	0.0	47.0
15	Wilanów	52.153083	21.110441	0.0	0.0	0.0	3.0
16	Rembertów	52.261415	21.162819	1.0	0.0	1.0	24.0
17	Wesoła	52.251794	21.229276	0.0	0.0	0.0	20.0

Scores after weighted computation

	Districts	Score
6	Śródmieście, Warsaw	155.0
9	Ochota	131.5
13	Żoliborz	115.5
1	Praga Południe	112.0
11	Praga Północ	108.5
3	Wola	106.0
0	Mokotów	105.5
7	Bemowo	103.0
2	Ursynów	103.0
4	Bielany	99.0
5	Targówek	98.5
12	Ursus, Warsaw	95.5
14	Włochy	92.0
10	Wawer	54.5
16	Rembertów	48.5
8	Białołęka	46.0
17	Wesoła	40.0
15	Wilanów	6.0

- The Locality with the best score is **“Śródmieście, Warsaw”** with 155.0, being the best option.


Most Świętokrzyski
Markiewicza
Park Kazimierzowski
Uniwersytet Warszawski
Ogród Saski
Park Mirowski
Osiedle Za Żelazną Bramą
Królewska
Centrum Nauki Kopernik
Most średnicowy
PGE Narodowy
Wisła
Powiśle
Śródmieście Północne
Nowy Świat-Uniwersytet
Tamka
Prom Pliszka
Most Poniatowskiego
Świętokrzyska
719
Park Świętokrzyski
Park Karola Beyera
631
Rondo ONZ
Plac Defilad
Warszawa Powiśle
Park im. Janiny Porazińskiej
Warszawa
Centrum
Śródmieście
Solec
Park na Książęcem
Złote Tarasy
Książęca
Warszawa Śródmieście
Plac Trzech Krzyży
Cypel Czerniakowski
Prom Słonka
Warszawa Śródmieście WKD
Warszawski Uniwersytet
Most Łazienkow

Recommendations:

The following analysis can be improved with following extensions:

- Consider more categories. For example like "Night life" which is also a good source for customers. But also like "Restaurants", which even if not burger joints may be some concurrence if too many.
- In the Locality itself, it can also be computed the distance between all the venues in order to find a place with the most number of potential customers.
- Using smaller geographical areas like Neighborhoods could improve the accuracy for the scores.