Opening a new restaurant in Warsaw, Poland

Tomasz Klonecki

March 2020

1 Introduction

Nowadays, for many people, restaurants, coffee places and bars are a great way to relax and enjoy themselves during working days and weekends. From Monday to Friday it's usually a place to have a business conversation or just lunch during office hours and during weekends these places can be occupied from morning to evening with people that take pleasure in experiencing new cuisine or having a nice time with their families. On the other hand for business owners there are multiple problems, such as their new venue location. In this work I will focus on solving that problem with Machine Learning methods.

1.1 Business problem

The objective of this project is to analyse and select the best locations in the city of Warsaw, Poland to open a new restaurant. Using Machine Leaning methodology such as clustering. The final analysis will answer the question: Where is the best spot to open new restaurant/bar/... in Warsaw.

1.2 Target Audience

This project is mostly useful for small gastronomic business owners, franchise restaurants, restaurant chains and other types of companies that want to open a new place in Warsaw, Poland.

2 Data

To answer to the business question: Where is the best spot to open new restaurant/bar/... in Warsaw, Poland?, I will use following data:

- List of neighbourhoods in Warsaw, Poland with their geolocation (geojson file).
- Venue data, data related to any gastroniomic businesses in Warsaw, Poland.
- Data about number of residents in each neighborhood in Warsaw, Poland.

2.1 Sources of data and methods to extract them

At the begining, I will use various methods for web scrapping to collect list of neighborhoods with multiple informations. Very good dataset containing all polygons of Warsaw neighborhood could be found here:

`https://github.com/andilabs/warszawa-dzielnice-geojson/blob/master/warszawa-dzielnice.geojson`

I download the geojson file and transform it into wanted format. Then I will use the Wikipedia website, to get all necessary information about residents:

`https://pl.wikipedia.org/wiki/Podzia%C5%82_administracyjny_Warszawy`

After that, I will use Foursquare API to get venues data for whole city. Foursquare has one of the largest database of 105+ million places and its use worldwide by multiple companies. Foursquare API will provide information about latitude and longitude of venues, but also category and name.

3 Methodology

Firstly, we need collect all necessary data. The problem with *geojson* file is, that it cannot be transformed into pandas DataFrame very easily. I used few loops and I transformed *geojson* file into DataFrame for future use. Other data comes in more or less proper format, so no more unexpected transformations is required.

Next I will calculate the centre of each neighbourhood to use limited Foursquare API to get as many results as possible. I do it by aggregating DataFrame of geojson file with mean latitude and longitude.

Then I use Foursquare API to find TOP 100 venues within a range of 5 km from each neighbourhood centre. It results with almost 1500 venues. Then I provide small analysis about venues categories and remove some, that are not connected to gastronomical business.

Then I load the data about amount of people living in each neighbourhood and I visualise all of the data loaded before on a map:

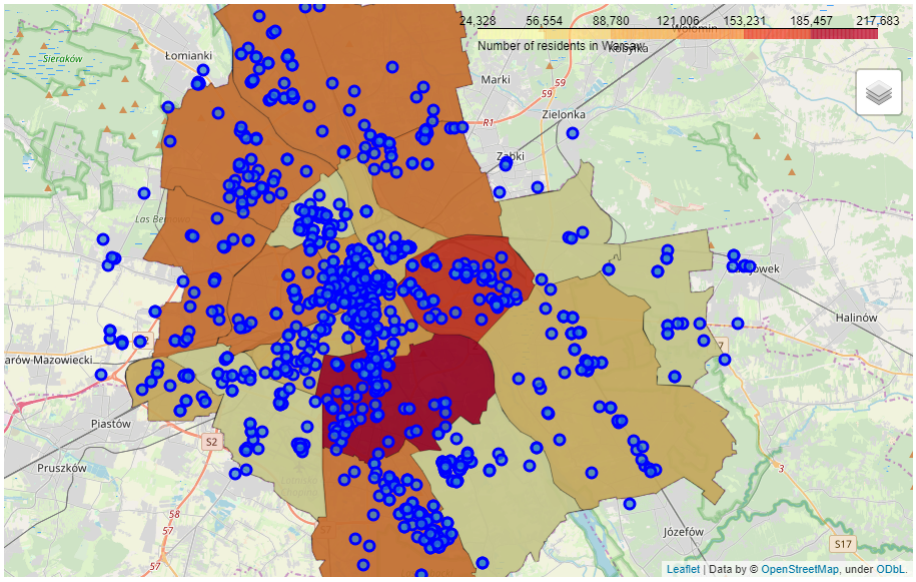

Figure 1: Number of inhabitants in Warsaw and location of "circles" of venues

Then I use *K-means* clustering to find "circles" of venues such as shopping malls, neighbourhoods of restaurants and all other gastronomical places. I don't use elbow method, because it says we need 5 clusters and it is just way to less

for this analysis. I choose number of clusters based on my experience in this city, $n\_clusters = 50$.

Then I visualise results on a map again:

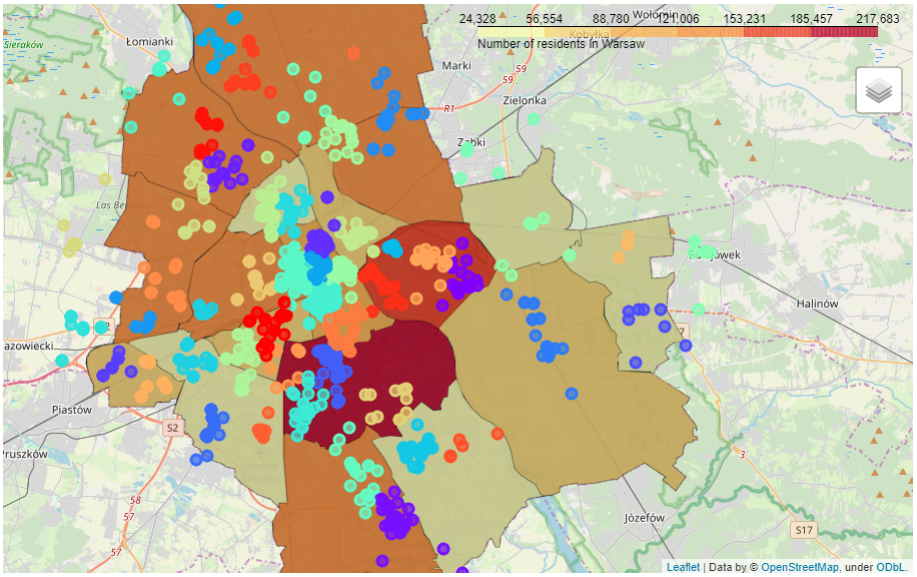

Figure 2: Number of inhabitants in Warsaw and location of venues

At the end I classify which neighbourhoods contain which "circles" of venues and I produce following Data Frame, where ratio is:

$$ratio = \frac{number\_of\_circles}{number\_of\_residents} * 10000 \tag{1}$$

Ratio tells us how many venues exists in each neighbourhood for 10 000 residents. So less the ratio, better the neighbourhood for new business.

	Neighborhood	Residents	DensityOfResidents	Area	Count	ratio
16	Wesoła	25439	1109	2294	2	0.786194
14	Włochy	42862	1497	2863	3	0.699921
7	Śródmieście	115395	7411	1557	6	0.519953
15	Wilanów	40060	1091	3673	2	0.499251
17	Rembertów	24328	1261	1930	1	0.411049
8	Białołęka	124125	1699	7304	5	0.402820
13	Żoliborz	52293	6174	847	2	0.382460
12	Ursus	60112	6422	936	2	0.332712
6	Bemowo	123932	4967	2495	3	0.242068
9	Ochota	82774	8516	972	2	0.241622
4	Bielany	131910	4079	3234	3	0.227428
1	Praga Południe	179836	8036	2238	4	0.222425
0	Mokotów	217683	6146	3542	4	0.183753
5	Targówek	124279	5131	2422	2	0.160928
11	Praga Północ	64113	5614	1142	1	0.155975
2	Ursynów	150668	3441	4379	2	0.132742
10	Wawer	77205	969	7970	1	0.129525
3	Wola	140958	7319	1926	1	0.070943

Figure 3: Final ratio of neighbourhood

4 Results

The results from the K-means clustering show that the most valuable neighbourhood to open a new gastronomical place would be Mokotów, where there is just 0.13 "circle" of venue per 10 000 residents. Next neighbourhood is Wola, where ratio reaches 0.14, which is not much different. So in the end I would divide these neighbourhoods in 2 groups:

- **Insatiable (good for new business):** Where ratio is lesser then 0.2
- **Saturated (bad for new business):** Where ratio is greater then 0.2

Final decision:

- **Insatiable:** Mokotów, Wola, Praga Północ, Bemowo, Ursus, Praga Południe,
- **Saturated:** Bielany, Ochota, Ursynów, Białołeka, Żoliborz, Rembertów, Wilanów, Wawer, Śródmieście, Włochy, Wesoła

5 Discussion

As observations noted from the results section, the best locations are located in most crowded neighbourhoods, but some of them are excluded such as Ursynów. Which shows us that this method could be reliable and the recommendations are targeted correctly.

6 Conclusion

In this project, I have gone through the process of identifying the business problem, specifying all requirements and performing Machine Learning solution to it. The answer to business question was produced and any relevant stakeholders can take an opportunity to use these results to open their new venue in Warsaw, Poland.