**THE BATTLE OF NEIGHBORHOOD.**

**WARSAW, POLAND.**



Marcos Almeida.

June 20th, 2020

**SUMMARY:**

1. Introduction: Business problem.
2. Data used and source.
3. Methodology.
4. Results.
5. Discussion.
6. Conclusion.
7. ***Introduction: Business problem.***

The purpose of this study is to make a research about where to open a business in a particular city with the most convenient location taking into account city area, population (and therefore, potential customers) or level of competence of similar businesses. I've decided to focus my project in the city of Warsaw, capital of Poland. Our customer wants to establish a business in the city, but he doesn't know which kind of business. Anyway it will be an activity related to food, like a restaurant, a cafeteria or similar, since he has previous experience in such areas. He doesn't want a business which is very present in the city, to prevent from suffering a huge competence since he is new in the city of Warsaw. Instead he wants something with an average number of venues.

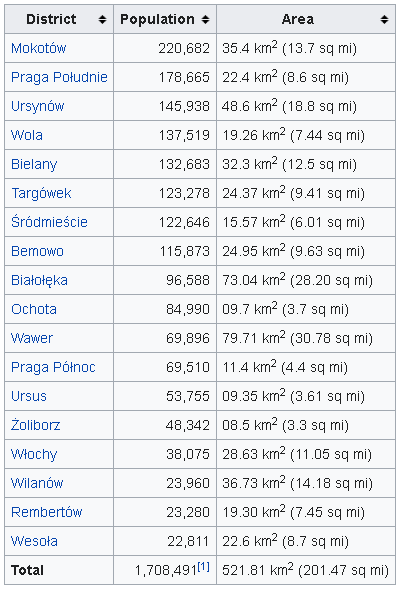
So the main concern we must face prior to seeking a business is, in fact, to determine which food-related business is more appropriate for our customer to open based on his preferences.

1. ***Data used and sources.***

The main source of data comes from the Wikipedia, which provides a table with the districts of Warsaw (https://en.wikipedia.org/wiki/Districts\_of\_Warsaw).

In order to develop this research I imported numpy library to handle data in a vectorized manner, pandas which is a library for data analsysis and matplotlib for graphical usage.

The coordinates of the city were obtained with Nominatim, whereas BeautifulSoup, which is a Python package for parsing HTML and XML documents, was used to extract the table from the Wikipedia containing the data of the Districts of Warsaw, for a later transformation into a Pandas data frame:



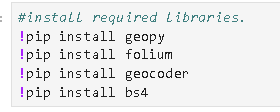
Folium library was used to create leaflet maps of the city of Warsaw.

Foursquare API was used to search for venues in the city as well as their ratings.

Scipy and Agglomerative Clustering were also used to cluster the districts.

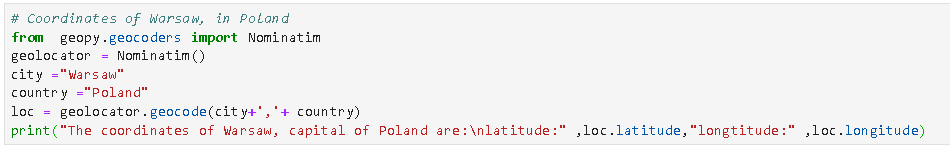
1. ***Methodology***

I started by importing the essential libraries in Python:



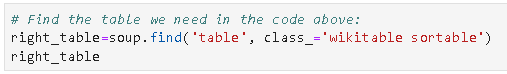
Then I imported several libraries such numpy, pandas or matplotlib, as mentioned above.

One of the main things we must do when working with cities is to get its geographical coordinates. In the case of Warsaw it was done with the next code:



The next step is getting the table from the Wikipedia. However, since there are more tables and data involved because it also shows historical data, I first had to import the source code of the site (which is very extent) and later extract the content of the required table:

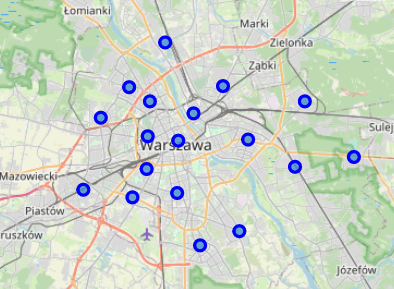




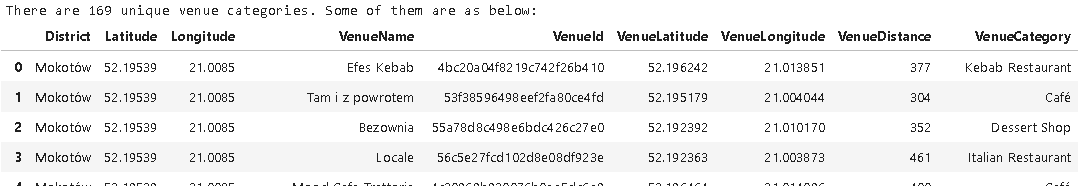
After transforming into a dataframe and adding coordinates of each district, we get the next table:

This table shows all the 18 districts in Warsaw, with their respective population and coordinates. We’ve got an important information to take into account when making a decision about where to place the business, because we can determine the number of potential customers based on the population or how far is a specific district from the city center, that is, the closer the better

Last, but no least, with this table a map was created using folium:

The blue dots on the map represents each district of Warsaw.

So now our next concern is search for venues in Warsaw. By using Foursquare we can get all the venues in the city with the corresponding name and category:

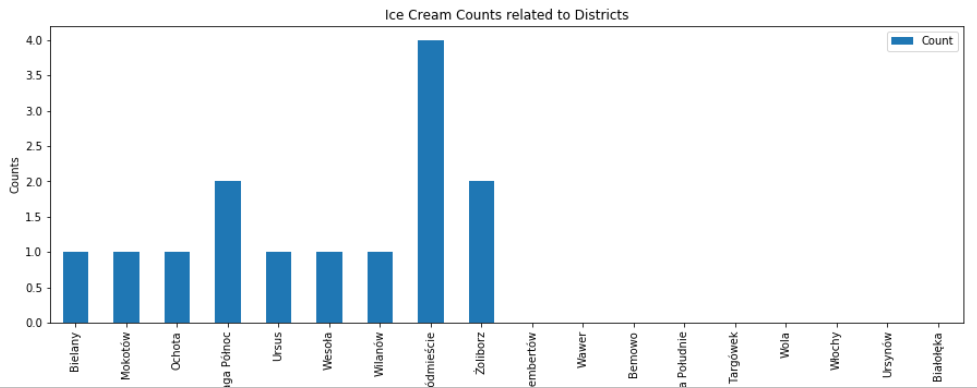


Obviously we must sort all these data out before working with them. We can classify each venue category and count them all to know the total number of each venue and the level of presence in the city. This is important because, as the customer requested, he would like to establish a business which is not very present in the city:

So according to the purpose of the study and taking into account the results from the table, we should discard a business like a Café or an Italian Restaurant, which presence are very high in the city.

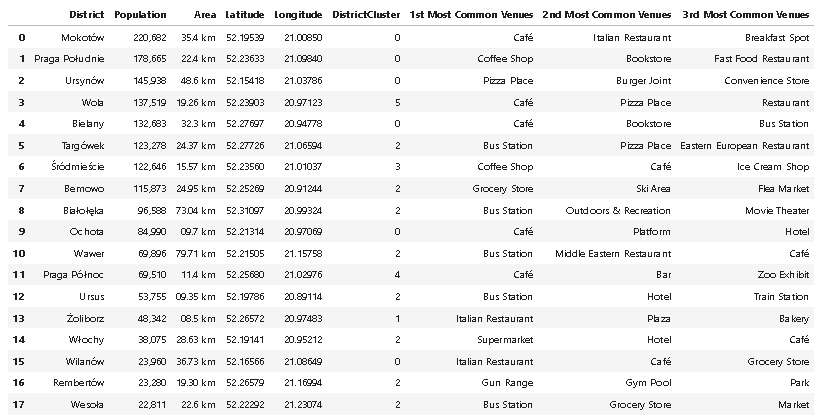
So our customer could open a business with an average presence in the city, like a Pizza Place, Grocery Store or an Ice Cream Shop. Based on his interests the option of the Ice Cream Shop is the more convenient.

After the decision of an Ice Cream Shop was made, we must determine the number of this type of venue in each district, to know the level of competence and the location of each district from the city center:

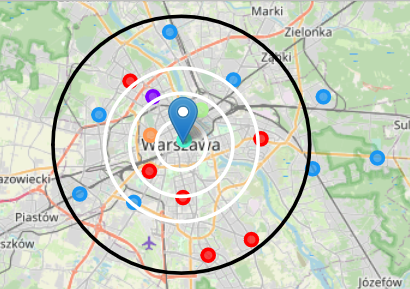
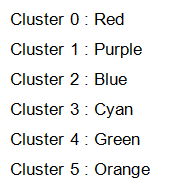


According to the graph above, the 14 ice cream shops are concentrated in the first 9 districts of Warsaw. So the customer could open this business in one district with no ice cream shops and at the same time, one district as close to the city center as possible. That would be the perfect balance.

I decided to cluster these data into 6 clusters taking into account the number of venues in each district.



And then plot the corresponding map:

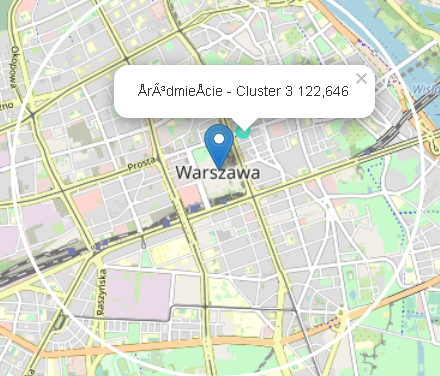
Radius are 2km, 4km, 6km and 10km

1. ***Results***

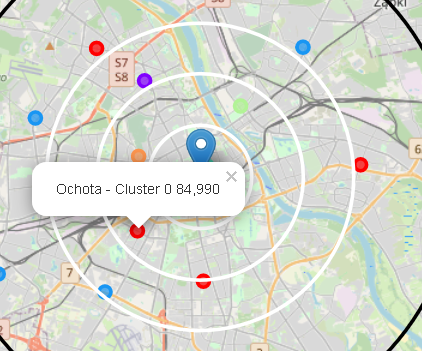
From the previous tables and maps, we found that there are 14 Ice Cream Shops in the city of Warsaw, with 4 of them in the District of Śródmieście, then 3 in Praga Północ and Żoliborz. There are 9 districts with no Ice Cream Shops, which should be taken into account since there are no competence for the business of our customer.

1. ***Discussion***

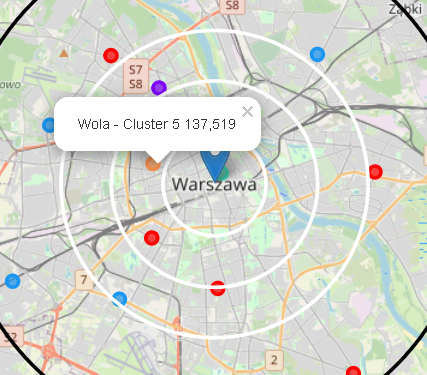
Taking into account the previous tables and maps, we have two main suggestions about opening an Ice Cream Shop in Warsaw:

1. In Śródmieście district cluster 3, there are 4 Ice Cream shops. The city center is a good advantage, but the customer must compete with others.
2. We can also think on the Ice Cream Shop as a complement to venues like Restaurants or Cafés, so clusters 0 and 5 seems to be the more convenient for that since there are more venues like these and almost no Ice Cream Shops.  
   So we can split this option into this:

1: In cluster zero, both Ochota and Mokotow districts have 1 Ice Cream Shop each, although Ochota is closer to city center, but less population (84,990) and Mokotov is a bit far but with much more population (220,682):



2: In cluster 5, the Wola district with 137,519 habitants is even closer from the city center than the ones in cluster zero and also with no Ice Cream Shops.



1. ***Conclusion***

After analysing all these data and considering the customer purpose, we suggested him to open an Ice Cream Shop. About which district to establish this business in, the final conclusion would be to open the venue in Śródmieście district, in the city center, where there are 4 Ice Cream shops, or in Wola district which is not city center but close (in a 4km radius) without any competence in the business purposed.