Neighborhood analysis for restaurant opening in Wroclaw, Poland

By Rafal Dziendziel, July 2019

1. Introduction/Business Problem

A client of my advisory company, a restaurant business investor, has requested for analysis regarding the best location to open her next restaurant. She wants to open a second restaurant in the Polish city Wroclaw. The first restaurant was open three years ago in a neighborhood called "Biskupin-Sępolno-Dąbie-Bartoszowice" and has been a great success. The restaurant is very popular and has received great online reviews from customers.
The client is looking for another neighborhood of Wroclaw that would be a best fit to open a second restaurant, based on the below criteria:

- The neighborhood has to be similar to "Biskupin-Sępolno-Dąbie-Bartoszowice" to maximize chances of repeating the first restaurant's success.
- The neighborhood should have minimal number of existing restaurants and their average ratings should be minimal as well in order to make competition easier.

2. Data

The data for the analysis will be taken from the flowing sources:

- Wikipedia page with the administrative split of Wroclaw's neighborhoods containing list of all 48 Wroclaw neighborhoods
- Geo-py web-service allowing to retrieve geo-localization data (longitude and latitude) for each of the neighborhoods
- Foursquare location platform to retrieve information about:
    - All types of venues in the neighborhoods to allow finding similar areas to "Biskupin-Sępolno-Dąbie-Bartoszowice"
    - Food services venues in similar neighborhoods (including their customer ratings) to allow identifying areas with least and worse-rated restaurants.

3. Methodology

For my analysis I have used Jupyter notebooks and Python programing language with multiple open source libraries.

In order to get to the conclusion of the analysis and provide recommendations to the client, I have followed the below steps:

1. Getting Wroclaw neighborhoods data
2. Getting information about nearby venues of each neighborhood
3. Clustering neighborhoods using K-means clustering algorithm
4. Identifying neighborhoods similar to "Biskupin-Sępolno-Dąbie-Bartoszowice"
5. Analyzing and comparing food venues in the similar neighborhoods

Below is the detailed description of each of the steps:

1. Getting Wroclaw neighborhoods data

In order to get the neighborhood dataset relevant for the analysis I firstly had to download the list of all Wroclaw neighborhoods from the Wikipedia page: https://pl.wikipedia.org/wiki/Podzia%C5%82_administracyjny_Wroc%C5%82awia

I used Pandas library to extract the table.

Here's the table on the Wikipedia page:

Osiedle administracyjne (od 1991)	Liczba mieszkańców w tys. 2017(2008)	Zmiana liczby ludności 2008-2017	Dawna dzielnica	Uwagi: osiedla, jednostki przestrzenne	Nazwa niemiecka (do 1945)
Gajowice	24,8 (27,6)	▼ 10,4%	Fabryczna		*Gabitz*
Gądów-Popowice Południowe	25,8 (26,4)	▼ 2,3%	Fabryczna	• Gądów Mały • Popowice	• *Klein-Gandau* • *Pöpelwitz*
Grabiszyn-Grabiszynek	13,8 (13,1)	▲ 5,3%	Fabryczna	• Grabiszyn • Grabiszynek	• *Gräbschen* • *Leedeborn*
Jerzmanowo-Jarnołtów-Strachowice-Osiniec	2,1 (1,8)	▲ 16,7%	Fabryczna	• Jerzmanowo • Jarnołtów • Strachowice • Osiniec	• *Hermannsdorf* • *Arnoldsmühle* • *Schöngarten*
Kuźniki	5,9 (6,2)	▼ 4,8%	Fabryczna		*Schmiedefeld*
Leśnica	28,1 (21,4)	▲ 31,3%	Fabryczna	• Leśnica • Stabłowice • Złotniki • Marszowice • Ratyń	• *Deutsch-Lissa* • *Stabelwitz* • *Goldschmieden* • *Marschwitz* • *Rathen*

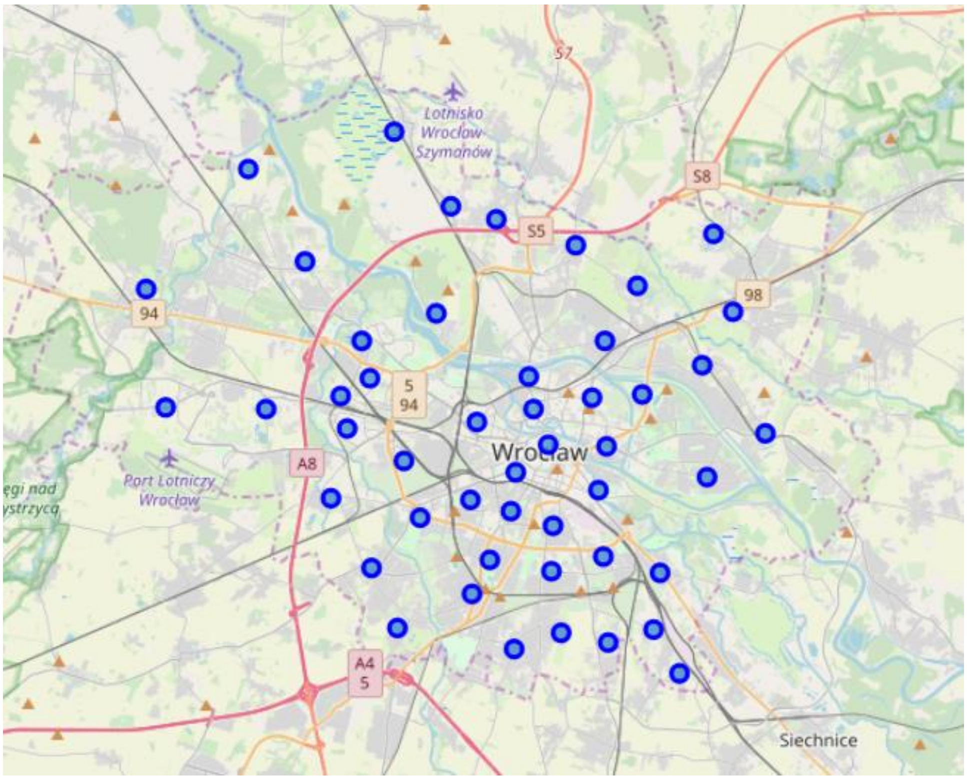
And here it is a sample of the dataset imported to Jupyter notebook:

	Osiedle administracyjne(od 1991)	Liczba mieszkańcóww tys. 2017(2008)	Zmiana liczbyludności 2008-2017	Dawna dzielnica	Uwagi: osiedla,jednostki przestrzenne	Nazwa niemiecka(do 1945)
0	Gajowice	24,8 (27,6)	10,4%	Fabryczna	NaN	Gabitz
1	Gądów-Popowice Południowe	25,8 (26,4)	2,3%	Fabryczna	Gądów Mały Popowice	Klein-Gandau Pöpelwitz
2	Grabiszyn-Grabiszynek	13,8 (13,1)	5,3%	Fabryczna	Grabiszyn Grabiszynek	Gräbschen Leedeborn
3	Jerzmanowo-Jarnołtów-Strachowice-Osiniec	2,1 (1,8)	16,7%	Fabryczna	Jerzmanowo Jarnołtów Strachowice Osiniec	Hermannsdorf Arnoldsmühle Schöngarten
4	Kuźniki	5,9 (6,2)	4,8%	Fabryczna	NaN	Schmiedefeld

The data had to be cleaned (only name of the neighborhood column was relevant) and later enriched with geographical coordinates of each of the neighborhoods. For that purpose, I have used GeoPy Python library which allowed me to get the following table with longitude and latitude columns:

	neigh	lat	long
0	Gajowice	51.0964	17.0033
1	Gądów-Popowice Południowe	51.1284	16.9608
2	Grabiszyn-Grabiszynek	51.0914	16.9823
3	Jerzmanowo-Jarnołtów-Strachowice-Osiniec	51.1206	16.8743
4	Kuźniki	51.1238	16.9482

With the above data I was able to visualize all neighborhoods on the map of Wroclaw. For that purpose, I have used Folium Python library:



2. Getting information about nearby venues of each neighborhood

Next step was to use the Foursquare (a global geographical data provider) API to get information about nearest venues in each of the neighborhoods.

In order to use the API a free developer account needed to be created on Fousquare.com. Once that has been done, I had to reference my Foursquare credentials and use the Request python library to perform relevant web-service calls using my previously collected Wroclaw neighborhood data.

I have used a 1000-meter radius setting for a venue to be classified as a nearby venue of a given neighborhood.

873 venues were extracted for all neighborhoods. The sample of data can be seen below:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	ID	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Gajowice	51.096399	17.003334	4db295398154eb510decbce2	Cukiernia Spychała	51.100183	17.008219	Dessert Shop
1	Gajowice	51.096399	17.003334	515f0616498ea7a8f7758d6d	Piwnica	51.097530	17.013686	Beer Store
2	Gajowice	51.096399	17.003334	551c0d8e498e496f8a5e5ace	Pizzeria Bube	51.089989	16.995464	Pizza Place
3	Gajowice	51.096399	17.003334	51aa2724498e042d1058a7ee	Dalat II	51.101980	17.012398	Vietnamese Restaurant
4	Gajowice	51.096399	17.003334	50c8f748704333c9c6e9b1ed	Centrum Historii Zajezdnia	51.096725	16.991440	History Museum

3. Clustering neighborhoods using K-means clustering algorithm

Then, as per the clients request in order to identify the neighborhoods similar to "Biskupin-Sępolno-Dąbie-Bartoszowice", I decided to use K-means clustering machine learning method. However, before I could do that, I had to prepare my data set so that it can be used by the algorithm.

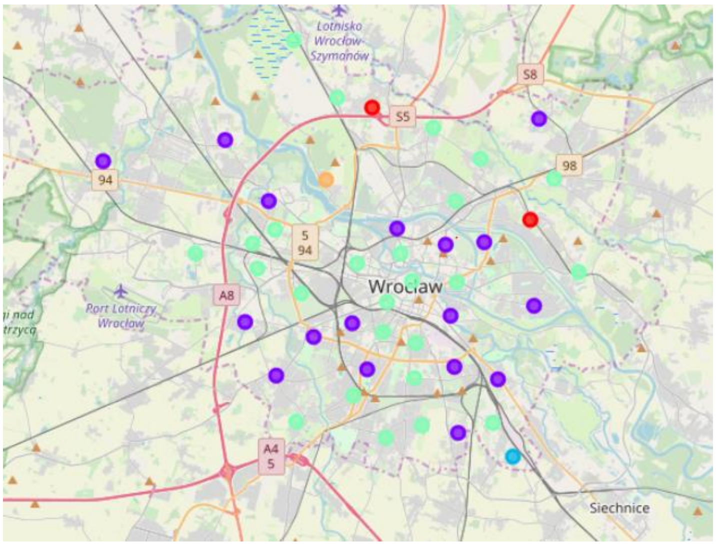
After the data transformation, my dataset consisted of neighborhood names and 10 most common venue categories (i.e. Restaurant, Dance Studio, Concert hall etc.):

[66]:

	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	Bieńkowice	Pet Service	Waterfront	Discount Store	Fast Food Restaurant	Farmers Market	Farm	Electronics Store	Eastern European Restaurant	Dumpling Restaurant	Donut Shop
1	Biskupin-Sępolno-Dąbie-Bartoszowice	Bus Stop	Pizza Place	Park	Tram Station	Grocery Store	Molecular Gastronomy Restaurant	Food	Dessert Shop	Supermarket	Convenience Store
2	Borek	Tram Station	Supermarket	Eastern European Restaurant	Hotel	Italian Restaurant	Park	Polish Restaurant	Cosmetics Shop	Café	Dance Studio
3	Brochów	Hotel	Train Station	Food & Drink Shop	Park	Grocery Store	Bus Station	Cosmetics Shop	Creperie	Convenience Store	Concert Hall
4	Gaj	Gym / Fitness Center	Grocery Store	Italian Restaurant	Fast Food Restaurant	Park	Food & Drink Shop	Bakery	Food Truck	Diner	Restaurant

With that data I was able to execute the k-means clustering algorithm. I used a popular Sklearn Python library to do that and chose to cluster the Wroclaw neighborhoods into 5 clusters.

The map below (I used Folium again) illustrates the clustering results (each color is a different cluster):



4. Identifying neighborhoods similar to "Biskupin-Sępolno-Dąbie-Bartoszowice"

When looking at the map we can notice that “Biskupin-Sępolno-Dąbie-Bartoszowice” neighborhood has been classified as Cluster 1.:

Therefore, the next step was to identify other neighborhoods that have been classified as Cluster 1 as well:

```
Out[72]: ['Gajowice',
          'Grabiszyn-Grabiszynek',
          'Leśnica',
          'Maślice',
          'Muchobór Wielki',
          'Oporów',
          'Pilczyce-Kozanów-Popowice Północne',
          'Borek',
          'Jagodno',
          'Księże',
          'Przedmieście Oławskie',
          'Tarnogaj',
          'Kleczków',
          'Pawłowice',
          'Biskupin-Sępolno-Dąbie-Bartoszowice',
          'Ołbin',
          'Zacisze-Zalesie-Szczytniki']
```

This already provides a partial answer to questions asked by the client- what are the similar neighborhoods to “Biskupin-Sępolno-Dąbie-Bartoszowice”. My recommendation will definitely be to open the new restaurant in one of the above neighborhoods as they are all similar to the area where the first, successful restaurant is situated. But which of the above list should be chosen? For that I had to go deeper in my analysis.

5. Analyzing and comparing food venues in the similar neighborhoods

The client’s second question was about the similar neighborhood with the lowest competition level. The lowest competition neighborhood is defined as one which has the lowest number of food venues nearby and also, those venues are lowest rated by the clients (based on online reviews).

At this moment my dataset contained only information about venue sub-category i.e. ‘Vietnamese Restaurant’ or ‘Dessert Shop’. Therefore, I first had to identify the list of all sub-categories that fall into general "Food" category in order to filter my dataset to show all relevant food venues.

For that I had to use Foursquare API again and download the overall Categories hierarchy used by the service. The results were returned in a JSON format and after a number of parsing steps I identified a list of all Food sub-categories for venues:

```
: ['Afghan Restaurant',
  'African Restaurant',
  'Ethiopian Restaurant',
  'American Restaurant',
  'New American Restaurant',
  'Asian Restaurant',
  'Burmese Restaurant',
  'Cambodian Restaurant',
  'Chinese Restaurant',
  'Anhui Restaurant',
  'Beijing Restaurant',
  'Cantonese Restaurant',
  'Cha Chaan Teng',
  'Chinese Aristocrat Restaurant
  'Chinese Breakfast Place',
  'Dim Sum Restaurant',
  'Dongbei Restaurant',
  'Fujian Restaurant',
  'Guizhou Restaurant',
  'Hainan Restaurant',
```

That allowed me to filter the venues dataset of Cluster 1 neighborhood.

Next, another Foursquare API call had to be made to get the customer ratings for each of the food venues in Cluster 1 neighborhoods.

4. Results

With the data acquired in steps described above, I was able to construct the below table listing all neighborhoods similar to "Biskupin-Sępolno-Dąbie-Bartoszowice", count of the food venues in each of them and average rating of food venues. The table is sorted by the count and then by mean_rating:

	Neighborhood	count	mean_rating
10	Pilczyce-Kozanów-Popowice Północne	1	6.200000
13	Zacisze-Zalesie-Szczytniki	1	7.800000
4	Jagodno	1	NaN
8	Maślice	1	NaN
12	Tarnogaj	2	6.000000
6	Księże	2	6.500000
7	Leśnica	2	6.800000
11	Przedmieście Oławskie	2	7.000000
5	Kleczków	3	6.700000
3	Grabiszyn-Grabiszynek	5	7.033333
0	Biskupin-Sępolno-Dąbie-Bartoszowice	5	7.100000
2	Gajowice	5	7.425000
9	Ołbin	10	7.550000
1	Borek	11	6.825000

We can see that the neighborhood “Pilczyce-Kozanów-Popowice Północne” has only one food venue and its rating is only 6.2 out of 10 meaning that the competition level (as defined before) is the lowest.

Therefore, my recommendation to the client is to open her next restaurant in that area.

5. Discussion

As mentioned above my recommendation, based on the client criteria of choice my recommendation to the client will be to open the new restaurant in the “Pilczyce-Kozanów-Popowice Północne” neighborhood.

However, this neighborhood currently has only 1 food venue, which may mean that it might not be very popular to go to eat-out in that area, so the client base may only be limited to local inhabitants of the area.

Another neighborhood to consider could be Tarnogaj or Kleczków that have a little bit more venues (2 and 3 respectively), but not as many as Ołbin or Borek (10 and 11), and relatively low average ratings (6.0 and 6.7).

Those two areas might not be very much competitive in terms of number of competitors and their quality but could be more popular among ‘Wroclawians’ to go eat-out in.

	Neighborhood	count	mean_rating
10	Pilczyce-Kozanów-Popowice Północne	1	6.200000
13	Zacisze-Zalesie-Szczytniki	1	7.800000
4	Jagodno	1	NaN
8	Maślice	1	NaN
12	Tarnogaj	2	6.000000
6	Księże	2	6.500000
7	Leśnica	2	6.800000
11	Przedmieście Oławskie	2	7.000000
5	Kleczków	3	6.700000
3	Grabiszyn-Grabiszynek	5	7.033333
0	Biskupin-Sępolno-Dąbie-Bartoszowice	5	7.100000
2	Gajowice	5	7.425000
9	Ołbin	10	7.550000
1	Borek	11	6.825000

6. Conclusion

My analysis managed to deliver results desired by the client plus it generated some additional considerations (described in the Discussion section).

For more in depth analysis, the results of this study could be accompanied by an additional study on, for example: venue rental space availability and costs in the similar neighborhoods as well as other factors such as population, profile of inhabitants (average income etc.) assuming such data is available.