Battle of Neighborhoods

Introduction

Gdansk is one of the most popular tourist destination in Poland. As stated on the official web page (https://www.gdansk.pl/en/for-tourists):

> *Gdańsk is a pearl of bourgeois architecture, boasting beautiful houses and a unique market. It is a world of cozy streets and historical churches. These are also perfectly preserved fortifications, ranked among the biggest in Europe, and interesting harbour architecture.*

The industrial sections of the city are dominated by shipbuilding, petrochemical & chemical industries, and food processing. The share of high-tech sectors such as electronics, telecommunications, IT engineering, cosmetics and pharmaceuticals is on the rise.

Due to large number of tourist, but also well paid workforce, Gdansk seems to be a good place for people interested in investing in the restaurant industry.

This documents discusses the approach that can be useful for making business decision by potential investors in Gdansk. As it is a high competitive market, the choice of proper destination must be carefully considered, otherwise it can generated losses.

Business problem

Let's suppose that the investor want to open a new dumpling restaurant. The question is where is the best place for establishing the restaurant?

Data

1. List of postal codes in Gdansk can be retrieved from the following **webpage**: https://trojmiasto.onet.pl/kod-pocztowy-gdansk-lista-kodow-pocztowych-artykul/7ftd6ws

   Data are present within the following tag:

   ```
   <p class="hyphenate " data-text-len="27" data-
   scroll="paragraph_1520">80-176;Gdańsk;Sympatyczna; </p>
   ```

   Inside the tag they are presented as:

   ```
   Postal Code;City name;Street name;
   ```

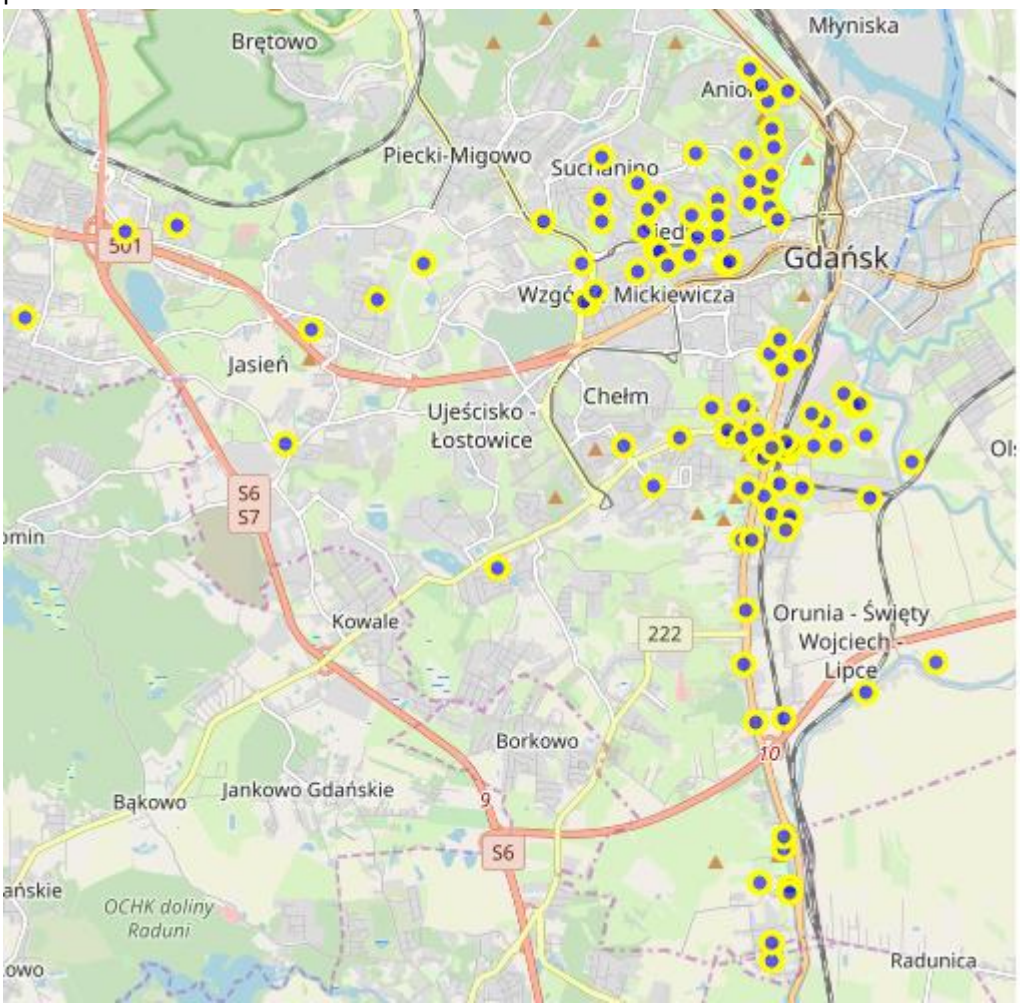
   In the above example: postal code = 80-176, city name= Gdańsk, street = Sympatyczna

   Note, that for one postal code there may be several street names. In this case these rows will be combined into one row with the streets separated with a semicolon
2. Postal codes will be converted into location using data retrieved from the following **webpage**: https://geocode.xyz/. Only location that are known will be further considered.
3. For each location (postal code with coordinates) retrieve 10 most popular venues using **Foursquare API**

Methodology

As the dumpling restaurants are already present in the city the goal is to examine where people are happy to go to this type of restaurant. Then to find similar places where they are still not going there.

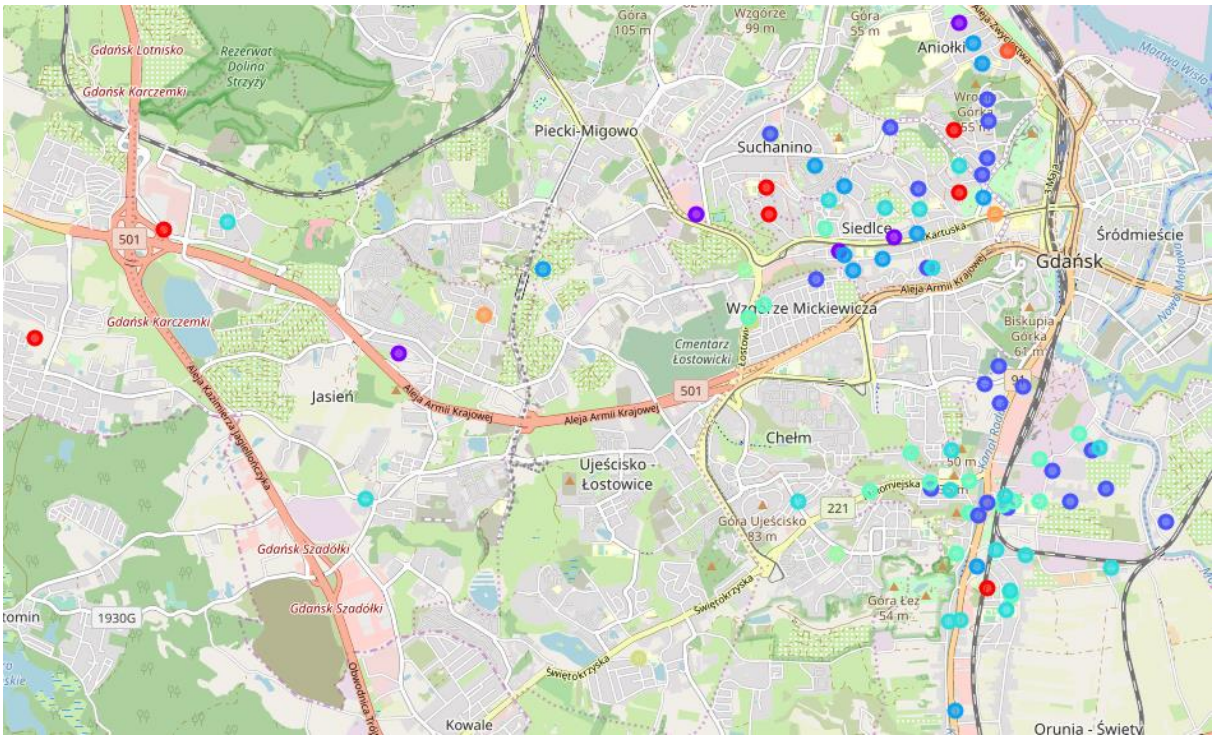
1. For each location (postal code) in Gdansk its location was retrieved. On the map they are presented as follows:



2. For each location retrieve 10 most common venues. Sample results:

	Postal Code	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	80-001	Forest	Train Station	Bed & Breakfast	Vegetarian / Vegan Restaurant	Dessert Shop	Diner	Dumpling Restaurant	Eastern European Restaurant	Electronics Store	Event Space
1	80-003	Sushi Restaurant	Vegetarian / Vegan Restaurant	Forest	Dessert Shop	Diner	Dumpling Restaurant	Eastern European Restaurant	Electronics Store	Event Space	Fast Food Restaurant
2	80-007	Bus Station	Vegetarian / Vegan Restaurant	Forest	Diner	Dumpling Restaurant	Eastern European Restaurant	Electronics Store	Event Space	Fast Food Restaurant	Food
3	80-008	Bus Station	Department Store	Vegetarian / Vegan Restaurant	Forest	Diner	Dumpling Restaurant	Eastern European Restaurant	Electronics Store	Event Space	Fast Food Restaurant
4	80-010	Bus Station	Vegetarian / Vegan Restaurant	Forest	Diner	Dumpling Restaurant	Eastern European Restaurant	Electronics Store	Event Space	Fast Food Restaurant	Food

3. Using K-Means cluster algorithm cluster all locations into 10 cluster according to most popular venues (ignore its physical location). Results of clustering:



4. Select cluster in which “Dumpling restaurant” is already one the **most popular venue**.

Cluster Labels	Dumpling Restaurant
0	0.000000
1	0.000000
2	0.009091
3	0.005747
4	0.000000
5	0.000000
6	0.000000
7	0.000000
8	0.015625
9	0.000000

The results shows that cluster = 8 is the most interesting. It contains only 2 rows:

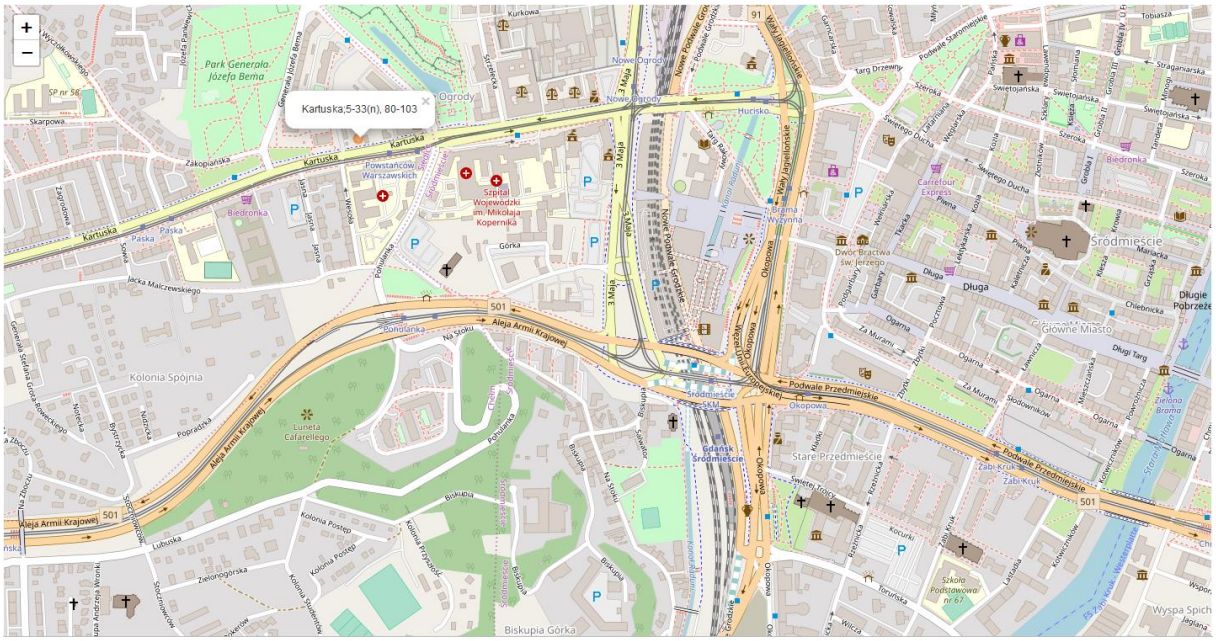
292	80-125	Bus Station	Gym	Grocery Store	Fast Food Restaurant	Gas Station	Vegetarian / Vegan Restaurant	Diner	Dumpling Restaurant	Eastern European Restaurant	Electronics Store
295	80-103	Café	Hotel	Historic Site	Pizza Place	Polish Restaurant	Coffee Shop	Fast Food Restaurant	Chinese Restaurant	Brewery	Theater

5. Select all addresses within the cluster from the previous point where “Dumpling restaurant” is **not among** the most popular venues:

	Postal Code	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
295	80-103	Café	Hotel	Historic Site	Pizza Place	Polish Restaurant	Coffee Shop	Fast Food Restaurant	Chinese Restaurant	Brewery	Theater

Result

The following address was found: 80-103, Kartuska 5-33 Street in Gdansk. The point shown on the map:



Discussion

Best place was found based on already existing data. However there were also 2 other cluster found where dumpling restaurants are quite popular. It might be also interesting to analyze historical records to examine in which places (clusters) a dumpling restaurant once established became popular.

The algorithm is not relevant for brand new ventures (like restaurant of a type that does not exists in the city).

Conclusion

The paper presents the algorithm for selecting the best place in Gdansk for establishing a new ventures of a type, that is already present in the city.