Data analysis to calculate the area of Oslo, where it is best to open a kebab restaurant.

Anton Ammosov

November 30, 2019

1. Introduction

1.1 Background

Our client is an immigrant from Russia, his name is Anton, he wants to open his point of sale of kebabs. He approached me to explore potential locations for him to open his new kebab restaurant in Oslo. He believes that in Norway there are still free places to open his restaurant. He wants us to analyze the market and find the best place to place a restaurant in Oslo.

1.2 Business problem

Obviously, there are many restaurants in Oslo. So we will need to find places that are not already crowded. However, we also have to find areas where there are very few or no kebab restaurants in the vicinity of a potential location. Location does not matter to our client. He would like to have a restaurant with high pedestrian traffic every day of the week. He States that he wants to make sure it's close to downtown if possible.
Using our expertise in science and data analysis, we will provide some potentially promising locations. We will describe the advantages and disadvantages of each location so that our client can make the right decision.

1.3 Interest

Obviously, Anton would be very interested in accurately predicting the areas of Oslo where it is best for him to open his kebab restaurant, this will allow him to better open a restaurant and gain a competitive advantage to run his business.

2. Data acquisition and cleaning

2.1 Data sources

Based on the business challenge presented above, here are the factors that affect a potential location:

Number of existing restaurants in the area (i.e. density) Number and distance to kebab restaurants in the area (our direct competition) The distance of the neighborhood / district from the city center Instead of using pre-defined areas in Oslo, we use regularly located locations that are centered around the city center to identify these areas. We chose all the places near high schools, universities, offices. We will rely on the following data sources to help us make a recommendation:

1. Potential areas will be generated using the algorithm, and approximate addresses of these generated areas will be obtained using Wikipedia parser and coordinate query using OpenStreetMap
2. The number of restaurants and schools, universities, offices (e.g. kebabs, high schools, universities, offices) as well as their location in each candidate area will be obtained using the Foursquare API
3. The coordinates of Oslo city centre will be obtained using the OpenStreetMap API

2.2 Data cleaning

Since the data we parsed with Wikipedia, it was necessary to clear the excess information from the data that we received. For this we used BeautifulSoup

```
]: ### Wikipedia request
from bs4 import BeautifulSoup
website_url = requests.get('https://en.wikipedia.org/wiki/List_of_boroughs_of_Oslo').text
soup = BeautifulSoup(website_url,'lxml')
print(soup.prettify())
```

```
My_table = soup.find('table',{'class':'wikitable sortable'})
My_table
```

2.3 Foursquare Data

Foursquare classifies places (e.g. arts and entertainment, food). For our task we are interested in a category "kebab", "high schools", "universities", "offices" which is identified by a unique alphanumeric code. We use these unique alphanumeric codes to help us identify potential direct

competitors. As well as potential customers. The categories and their alphanumeric codes are located at: https://developer.foursquare.com/docs/resources/categories.

2.4 Merging data sources

As mentioned, Wikipedia will allow us to define the districts of Oslo. Then, using the OpenStreetMap API, we will be able to determine the exact coordinates of the areas. This defines our boundaries for analysis. Foursquare data allows us to display restaurants within certain boundaries. We get the total number of kebab restaurants within our particular neighborhood (and each neighborhood). Thus, in each area we can calculate the density of kebab restaurants and the percentage of those restaurants that are our direct competitors. Also, using the foursquare API, we will be able to determine where our potential customers are high school students, University students and office building workers.

3. Exploratory Data Analysis

3.1 Data parsing

```
Localidades = []
for link in links:
    Localidades.append(link.get('title'))

print(Localidades)

del Localidades[0:2]

Localidades
```

```
['Alna', 'Bjerke', 'Frogner', 'Gamle Oslo', 'Grorud', 'Grünerløkka', 'Nordre Aker', 'Nordstrand, Norway', 'Sagen
haugen', 'Stovner', 'Søndre Nordstrand', 'Ullern', 'Vestre Aker', 'Østensjø']
```

```
['Frogner',
 'Gamle Oslo',
 'Grorud',
 'Grünerløkka',
 'Nordre Aker',
 'Nordstrand, Norway',
 'Sagene',
 'St. Hanshaugen',
 'Stovner',
 'Søndre Nordstrand',
 'Ullern',
 'Vestre Aker',
 'Østensjø']
```

Once we have been able to retrieve, then clear and bring the data to a convenient view we proceed to the next step.

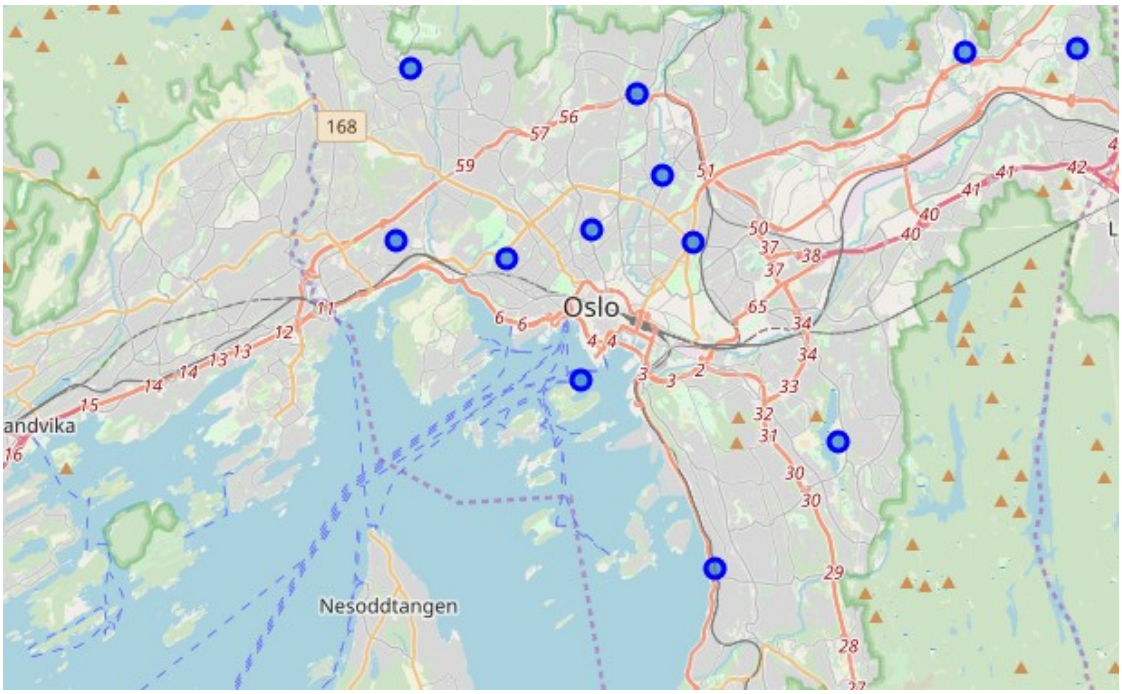
3.2 Obtaining and assigning the coordinates to the areas of Oslo

The api provided by openstreetmaps is the easiest way to obtain coordinates by mapping area names. After receiving the data, we can assign coordinates to the areas that we have parsed from Wikipedia.

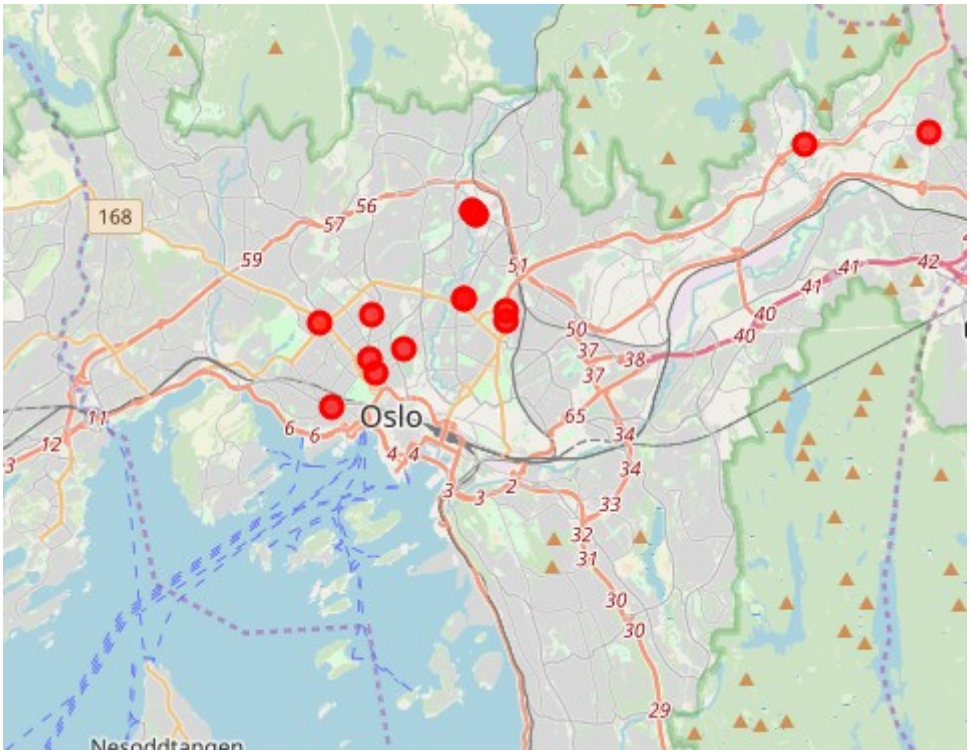
[20]:

	Localidades	Latitude	Longitude
0	Frogner	59.922224	10.706649
1	Gamle Oslo	59.899237	10.734767
2	Grorud	59.961424	10.880549
3	Grünerløkka	59.925471	10.777421
4	Nordre Aker	59.953638	10.756412
5	Nordstrand, Norway	59.863525	10.785830
6	Sagene	59.938273	10.765849
7	St. Hanshaugen	59.927950	10.738958
8	Stovner	59.962140	10.922823
9	Søndre Nordstrand	59.835944	10.798496
10	Ullern	59.925818	10.665132
11	Vestre Aker	59.958300	10.670319
12	Østensjø	59.887563	10.832748

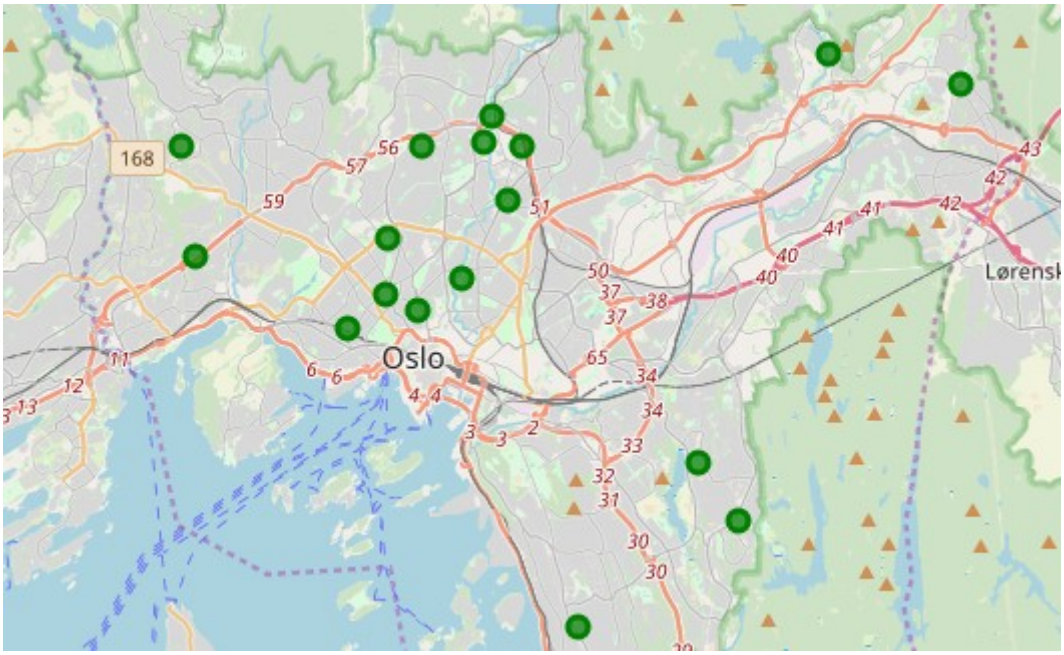
3.3 Building a map of Oslo districts



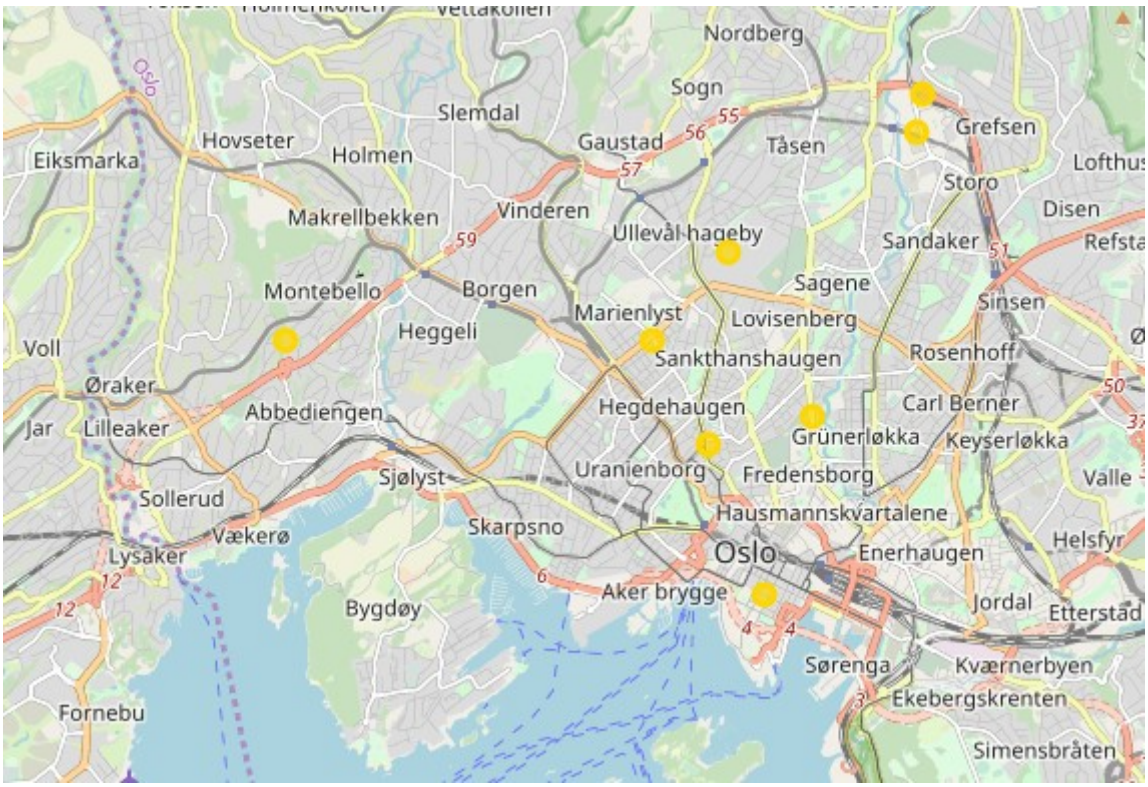
3.4 Parse data with foursquare API and build map



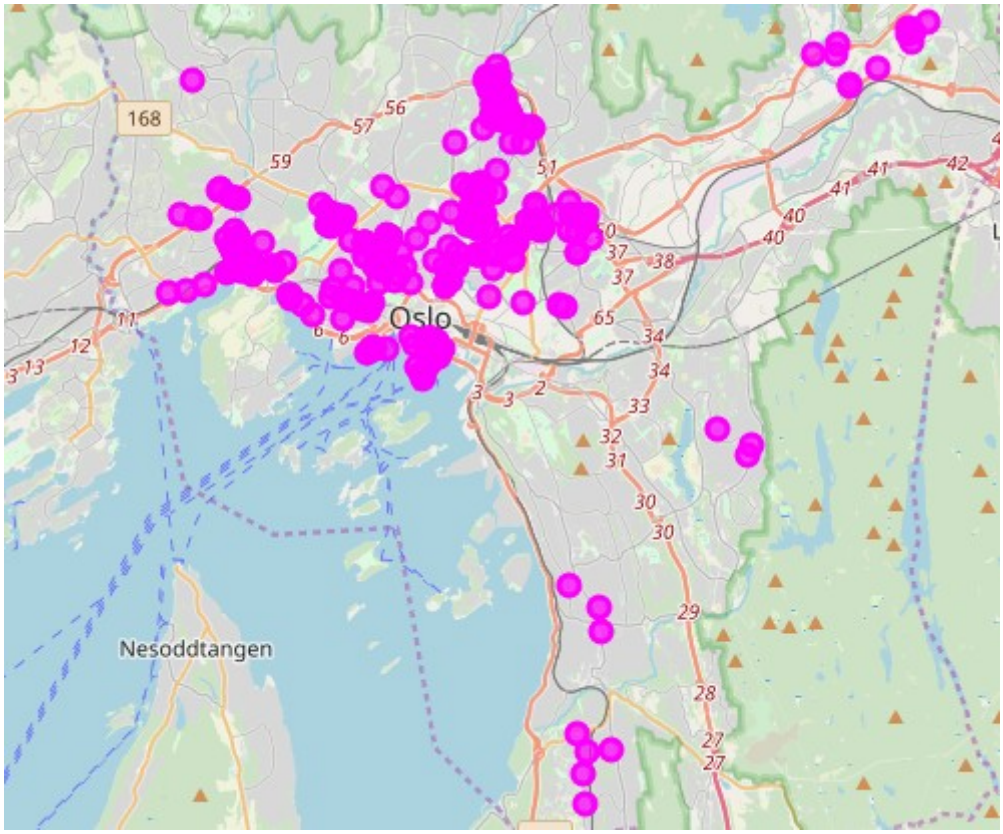
Map of kebab restaurants



Map of high schools



Map of Oslo universities



Map of Oslo offices

4. Predictive Modeling

Since we are trying to find which area would be best for opening a kebab restaurant we need to enter different points into the model for accurate forecasting.

4.1 Comparison of districts by "weight"

[33]:

	Localidad	Latitude	Longitude	Kebab	High Schools	Universities	Offices
0	Frogner	59.922224	10.706649	2.0	1.0	0.0	50.0
1	Gamle Oslo	59.899237	10.734767	0.0	0.0	1.0	49.0
2	Grorud	59.961424	10.880549	1.0	1.0	0.0	6.0
3	Grünerløkka	59.925471	10.777421	3.0	0.0	0.0	46.0
4	Nordre Aker	59.953638	10.756412	2.0	4.0	2.0	50.0
5	Nordstrand, Norway	59.863525	10.785830	0.0	1.0	0.0	3.0
6	Sagene	59.938273	10.765849	3.0	1.0	0.0	50.0
7	St. Hanshaugen	59.927950	10.738958	4.0	4.0	4.0	49.0
8	Stovner	59.962140	10.922823	1.0	1.0	0.0	6.0
9	Søndre Nordstrand	59.835944	10.798496	0.0	0.0	0.0	5.0
10	Ullern	59.925818	10.665132	0.0	1.0	1.0	50.0
11	Vestre Aker	59.958300	10.670319	0.0	1.0	0.0	1.0
12	Østensjø	59.887563	10.832748	0.0	2.0	0.0	3.0

Negative weight, because Anton wants to open a kebab joint and thus wants to avoid concurrence as much as possible if kebab resturand already have on this area is -1.

Positive weight, because high school students are good customers, weight is 1.

Positive weight, because uni students are good customers, weight is 1.5.

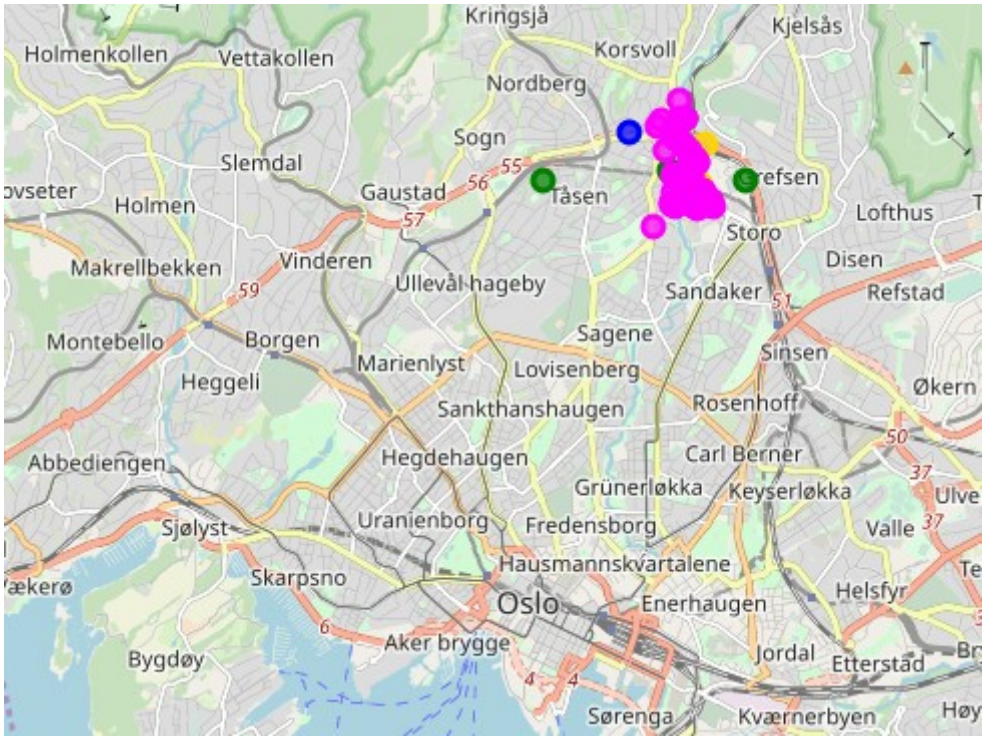
Positive weight because employees are even better customers, weight is 2.

4.2 Result of comparison of districts by «weight»

[36]:

	Localidad	Score
4	Nordre Aker	105.0
7	St. Hanshaugen	104.0
10	Ullern	102.5
1	Gamle Oslo	99.5
0	Frogner	99.0
6	Sagene	98.0
3	Grünerløkka	89.0
2	Grorud	12.0
8	Stovner	12.0
9	Søndre Nordstrand	10.0
12	Østensjø	8.0
5	Nordstrand, Norway	7.0
11	Vestre Aker	3.0

According to the results, the Nordre Aker district is the best for opening a restaurant selling kebabs.



Map with the best area for kebab restaurant