Coursera Capstone

IBM Applied Data Science Capstone

Opening a Bar in Oslo, Norway

BY: IULIAN CATALIN COSTACHE

AUGUST 2020

Introduction

- A friend of mine wants to open a bar in Oslo.
- Oslo is a vibrant city and people like to hang around in bars with friends. Since he worked as a bartender for many years, he would like to have his own bar.
- The question is where the new bar should be open?

Business Problem

- The objective of this capstone project is to analyse and select the best locations in the city of Oslo, Norway to open a successful new bar.
- Using data science methodology and machine learning techniques like clustering, this project aims to provide a solution to the most important question: in what location should be the bar open?

Data

To solve the problem, we will need the following data:

- List of neighbourhoods in Oslo.
- Latitude and longitude coordinates of those neighbourhoods. This is required to plot the map and to get the venue data.
- Venue data, particularly data related to bars. We will use this data to perform clustering on the neighbourhoods.

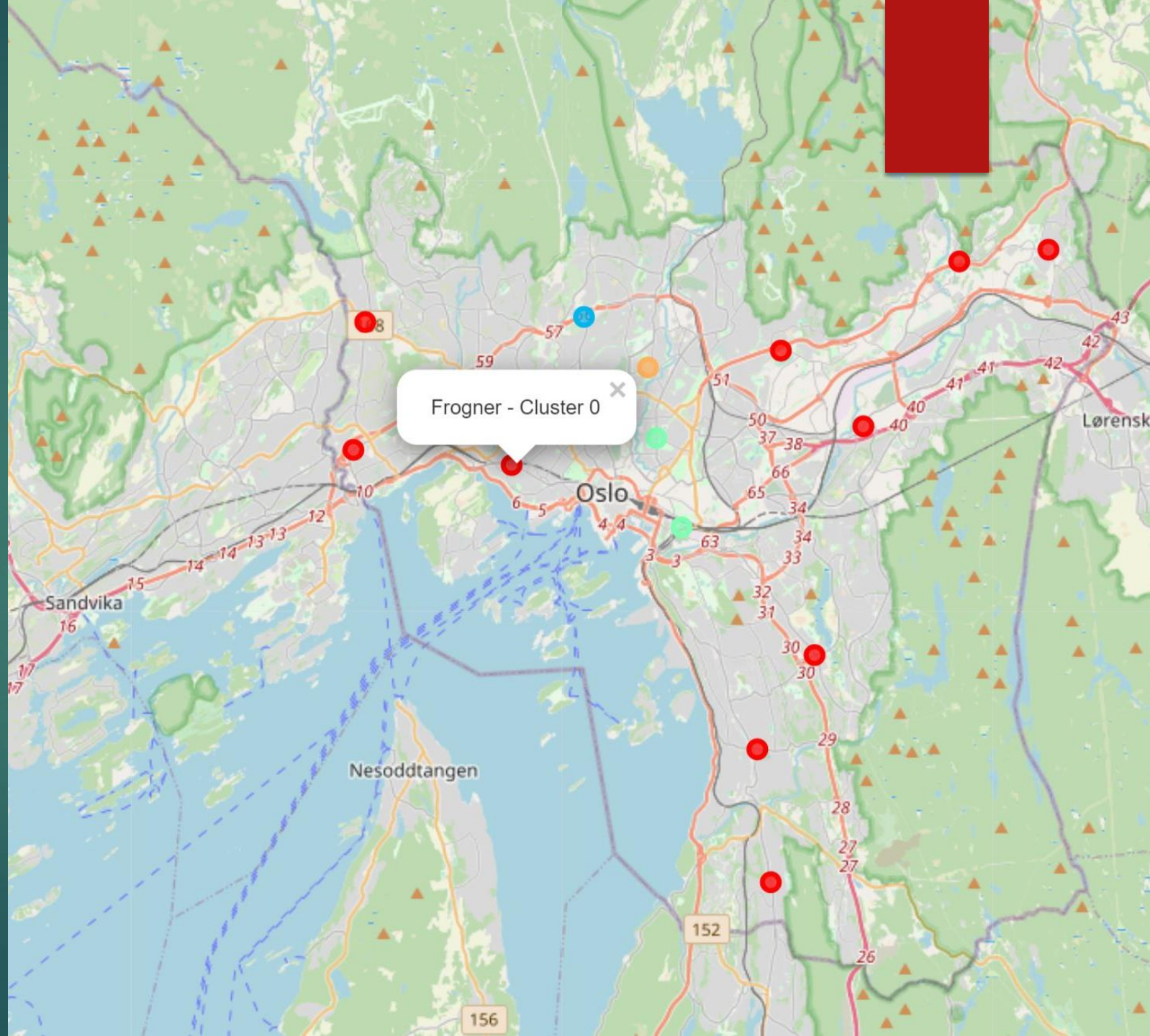
- Sources: Wikipedia and Forthsquare API

Methodology

- We need to get the list of neighbourhoods in the city of Oslo.
- We will do web scraping using Python requests and beautifulsoup packages to extract the list of neighbourhoods data.
- We need to get the geographical coordinates in the form of latitude and longitude in order to be able to use Foursquare API
- We will populate the data into a pandas DataFrame
- Visualize the neighbourhoods in a map using Folium package.
- Analysis

Results

- Cluster 0: Neighbourhoods with few or none bars, the biggest cluster
- Cluster 1: Neighbourhoods with high number of bars
- Cluster 2: Neighbourhoods with few bars
- Cluster 3: Neighbourhoods with the highest number of bars
- Cluster 4: Neighbourhoods with many bars



Discussion

- As observations noted from the map in the Results section, most of the bars are concentrated in the central area of Oslo city, with the highest number in cluster 3 and almost none in cluster 0.
- However, there is a exception with Frogner neighbourhoods. Frogner is locate in the centre of Oslo, but it has very few bars, compared with the other central neighbourhoods.
- This represents a great opportunity and high potential area to open new bar as there is very little to no competition from ones. Meanwhile, bars in cluster 3 are likely suffering from intense competition due to oversupply and high concentration of bars. From another perspective, the results also show that the oversupply of bars mostly happened in the central area of the city.
- Therefore, this project recommends that a new bar to be open in Frogner.

Conclusion

- In this project, we have gone through the process of identifying the business problem, specifying the data required, extracting and preparing the data, performing machine learning by clustering the data into 5 clusters based on their similarities, and lastly providing recommendations for where will be the best to open a new bar in Oslo.