Finding the best spot for ice cream shop in Manhattan.

Dmytro Prokopenko

July 4th, 2020

1. **Introduction**

**Business Problem**

As a Gelato lover, I would like to open the ice-cream shop. Moreover, this is a high-margin business if done properly.

With this work I would like to discover what neighbourhood in New York has a lesser number of such shops to open it there.

**Instrumentation**

In this capstone project, I will leverage the FourSquare API and K-means algorithm to find the best place to open a Gelato shop.

To find right K I will use Elbow method.

To visualize the result of my work I will use Folium module.

**Target Audience**

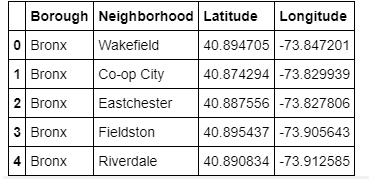
Businessmen who are interested in high margin retail business in New York

1. **Data acquisition and cleaning**

I used New York dataset from this location <https://cocl.us/new_york_dataset>.

It is in json format so I used pandas json module to load the data, selecting only a feature node, cutting he rest.

Then I put it all into padnas dataframe leaving only the following columns:

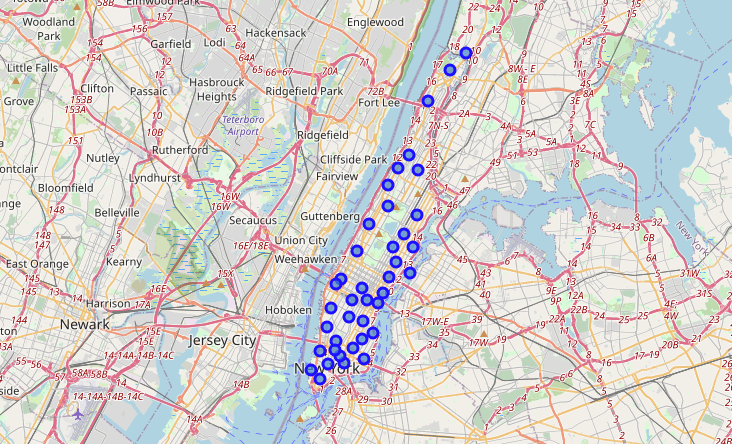


Next I used geopy library to get the latitude and longitude values of New York City.

The geograpical coordinate of New York City are 40.7127281, -74.0060152.

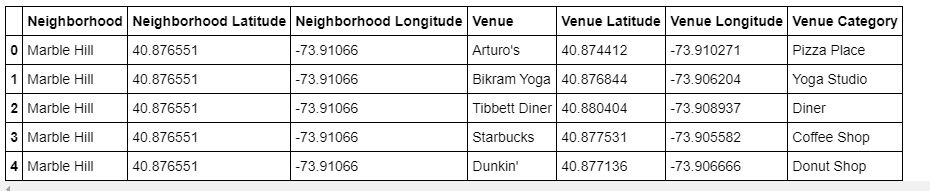
Then I decided to focus on Manhattan area as the most attractive.

Next step is to leverage FourSqaure API to load all venues located in Manhattan.



For one need to register as a developer on their website to get connection information (client ID and secret ID).

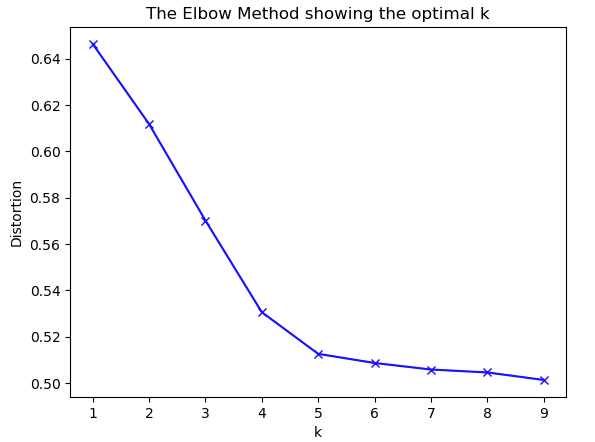
Exploring FourSquare information.



After loading POI from FourSquare for Manhattan, I left only ice cream shop points and grouped them by neighbourhood to see how many of them in each neighbourhood. To remind you – we need to find neighbourhoods with lesser amount of such shops.

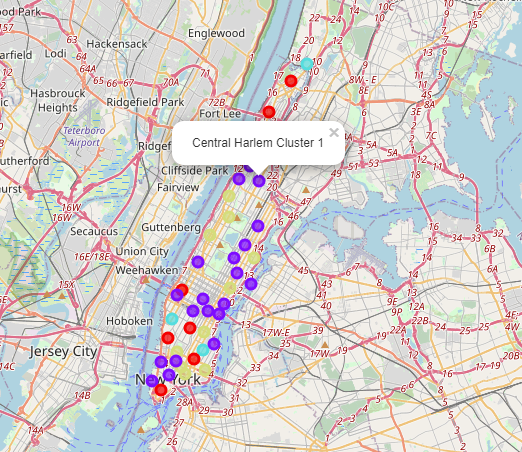
1. **Clustering**

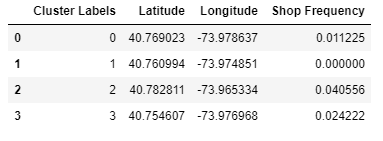
I used K-means algorithm to cluster the neighbourhoods based on number of shops in each of them. To determine right K – I used elbow method. I decided to use 4 clusters.



1. **Visualizing the result**

Finally once we clustered the result we have to show it on the map to determine the best location.





1. **Conclusion**

New York is a very crowded place in terms of venues, including ice cream shops. But still there are vacant locations.

As we can see, 2nd cluster has no ice cream shops at all, and this is the area we should look at.

You can click on any dot on the map to understand to which cluster it belongs to.

In our case we should look at purple dots on the map.

1. **Future directions**

This study is to show how data science instrumentation can be used to solve business problems.

This work can’t be treated as final solution for real life situation, as there more things to improve like take into account business, health and wellness situation of a particular neighbourhood, number of people passing many other aspects for a successful business to work and be profitable.