Suggest Location for *new* Bagel Shop in Manhattan, NY

Suraj Kumar

October 2nd, 2019

1   Introduction

1.1   Background

The City of New York is famous for its excellent cuisine. It's food culture includes an array of international cuisines influenced by the city's immigrant history. For a business to be successful, the choice of location is very important. One does not want to open the business in an area where competitors are already present, and it is very dense. Given data set of location, and available restaurants or venues in the location, is it possible to find suitable location for new business opportunity with certain degree of confidence? The project tries to find an approach to evaluate the options of location for any given business say new Bagel Shop with fair degree of confidence.

1.2   Problem

Data that might contribute to determining best location for new business opportunity include existing venues / business in the given area and find how densely they are located. The objective of this capstone project is to analyse and select the best locations in the city of Manhattan, NY to open a new Bagel Shop.

1.3   Interest

Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the business question: if a someone is looking to open a new Bagel Shop, which location will be our recommendation?
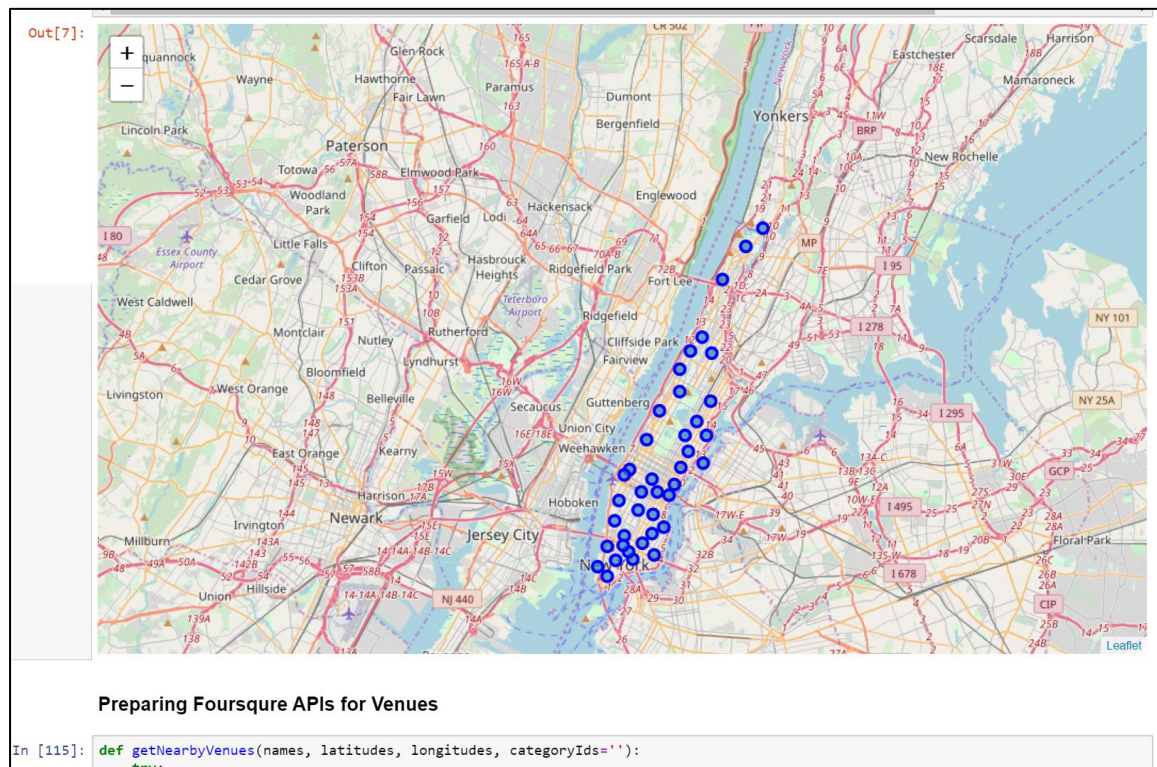
## 2   Data acquisition and cleansing

There were two types of data used:

1. List of Neighbourhood in Manhattan, NY along with Geo Coordinates. This data is available for free in https://geo.nyu.edu/catalog/nyu_2451_34572, and it available geo json formatted file - nyu-2451-34572-geojson.json file
2. To get the list of venues in Manhattan, NY, we next use Foursquare APIs. As developer account was being used, there is rate limit in terms of usage of APIs.
3. The data received from API is in JSON format, and requires converting to Dataframe.

3 Exploratory Data Analyses (Visualisation)

After getting the Neighbourhood data from geo-json file, we plot the points on NY map to see the spread of the neighbourhood and which all areas of Manhattan is covered with these points.

Preparing Foursqure APIs for Venues

```
In [115]: def getNearbyVenues(names, latitudes, longitudes, categoryIds=''):
              try:
```

Post getting the Venues data for Venue Category (`bagel_category = '4bf58dd8d48988d179941735'`) using Foursquare APIs, we plot all venues on NY map to see the spread of venues across Manhattan

Analyzing Neighbourhoods
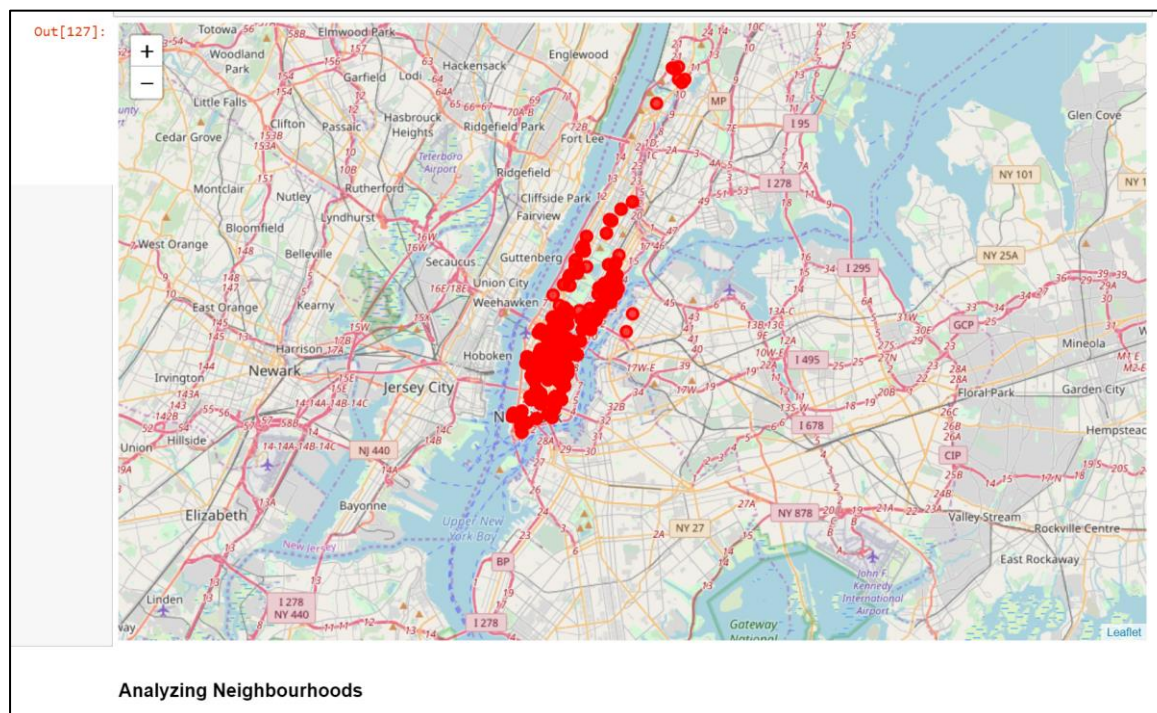
4   Methodology

4.1   Data

Preparing the data which included - acquiring, cleansing, formatting,

a) List of Neighbourhoods in NY, Manhattan along with Geo Coordinates in Dataframe
b) Geo Coordinates for Manhattan, NY
c) Foursquare API results for Venue Listing for Category = Bagel within NY geo-coordinates,
d) The JSON data from API is converted into Dataframe

4.2   Learning Algorithm - Clustering

We perform clustering on the data by using k-means clustering.

K-means clustering algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible. It is one of the simplest and popular unsupervised machine learning algorithms and is particularly suited to solve the problem for this project.

a. We clustered the neighbourhoods into 3 clusters based on their frequency of occurrence for "Bagel Shop".
b. The results will allow us to identify which neighbourhoods have higher concentration of Bagel Shops while which neighbourhoods have fewer number of Bagel Shop.
c. Based on the occurrence of Bagel Shops in different neighbourhoods, we will be able to to answer the question *as to which neighbourhood is most suitable to open new Bagel Shop.*

4.3   Data Analysis

As outcome of our analysis we get 3 clusters of neighbourhoods *based on number of Bagel Shop compared to other shops*

Cluster 0: Neighbourhoods with moderate number of Bagel Shop (🔴)

Cluster 1: Neighbourhoods with low number of Bagel Shop (🟣)

Cluster 2: Neighbourhoods with high concentration of Bagel Shop (🟢)


```
In [170]:  manhattan_bagelshop_merged.loc[manhattan_bagelshop_merged['Cluster Labels'] == 0]
```
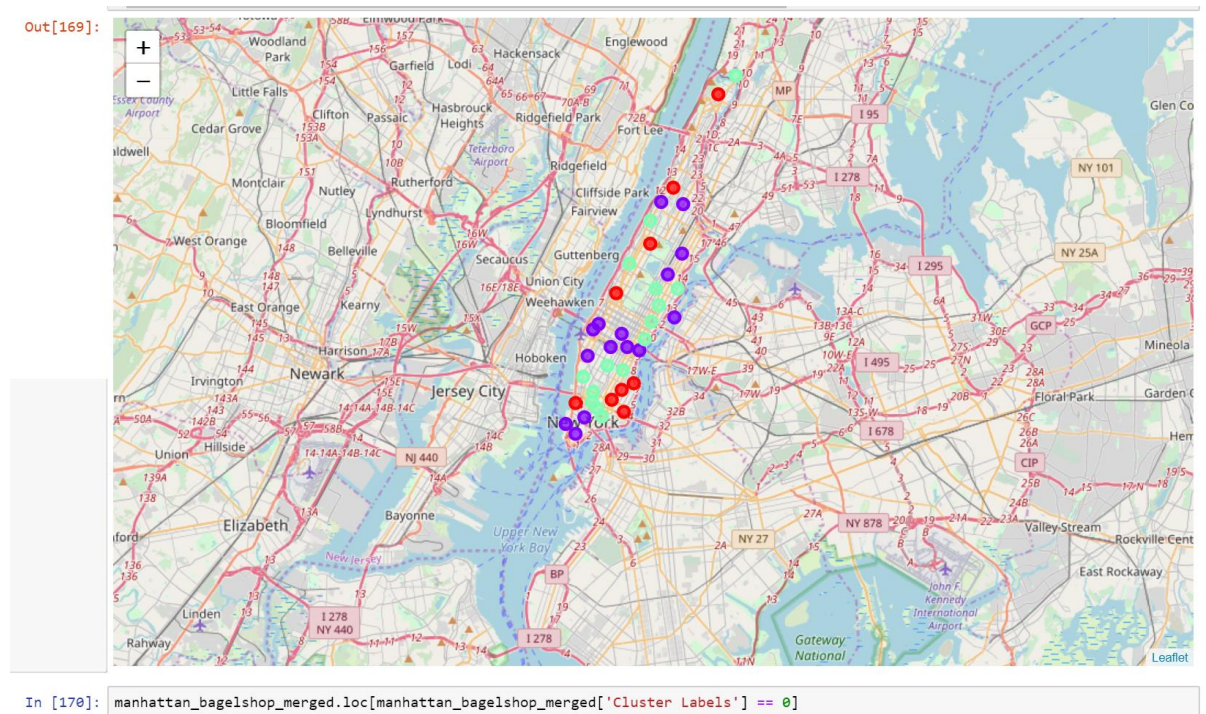
5   Conclusion

Cluster 1 (followed by Cluster 0), neighbourhood represents a great opportunity and high potential areas to open new Bagel Shop as there is less competition from existing Bagel Shops. Meanwhile, Bagel Shops in cluster 2 are likely suffering from intense competition due to oversupply and high concentration of Bagel Shops.

6   Future Direction

For our analysis and project, we have only considered 1 feature for analysis which is frequency of occurrence of Bagel Shops in neighbourhood, but there could be more parameters such as,
- Opening Hours – Breakfast, Lunch or Evening Snacks
- Type of neighbourhood – Residential vs Commercial
- Population Density
- Population Taste and Preferences, etc.

Another point is the data which is used for current analysis is either dated or limited due to developer edition (API). Using production grade data, and large data size, may result in better prediction / suggestion.