Recommender System for opening new restaurant around Manhattan NY

Coursera Capstone Project

**Name: Souma Suvra Ghosh**

December, 2018

**Table of Contents**

[Introduction 1](#_Toc533372069)

[Business Problem 1](#_Toc533372070)

[DATA Used 1](#_Toc533372071)

[Methodology 2](#_Toc533372072)

[Results 6](#_Toc533372073)

[Discussion 6](#_Toc533372074)

[Conclusion 6](#_Toc533372075)

# Introduction

In this project we will explore if someone wants to open a business at Manhattan Area of New York City what will be the recommended neighborhood.

# Business Problem

The Manhattan area of New York is the economic and administrative center. It is often described as the cultural, financial, media, and entertainment capital of the world. Thus it offers a great opportunity for someone who is trying to open a new shop. However, being a diverse metropolitan it also poses certain challenges to figure out where to start a new business.

It would be nice if there is a recommender system which analyze the different neighborhoods and recommend the best area to open the business. This machine learning recommender system will help the small business owner make an informed decision on best areas to open the business.

# DATA Used

The recommender system will use the Foursquare location data to get details of the presence of similar type of business in each neighborhood. Foursquare rest-based API provide details like Venue Name, Category, customer likes, etc. which are very useful data points to make the decision.

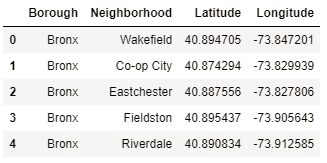
It will then try to use the clustering algorithm to group the neighborhoods and provide addition details of those neighborhoods.

The data from Foursquare will be fetched using the longitude and latitude details of neighborhoods around Manhattan.

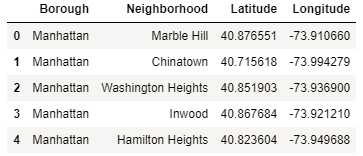
# Methodology

Machine learning Clustering Algorithm is used to group the neighborhoods to find similar pattern. The steps are followed-

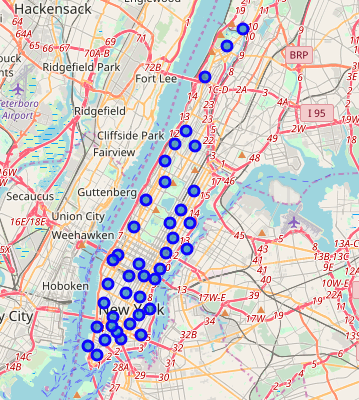
* The data for New York neighborhood along with the longitude and latitude details are obtained from internet. Sample value –



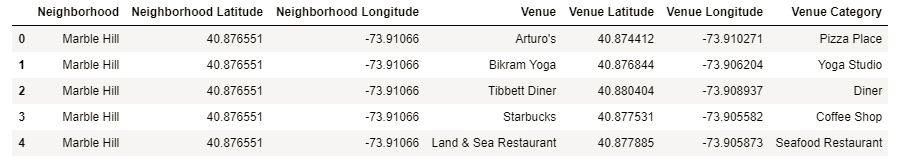
* Next, we selected the neighborhoods for Borough as Manhattan only



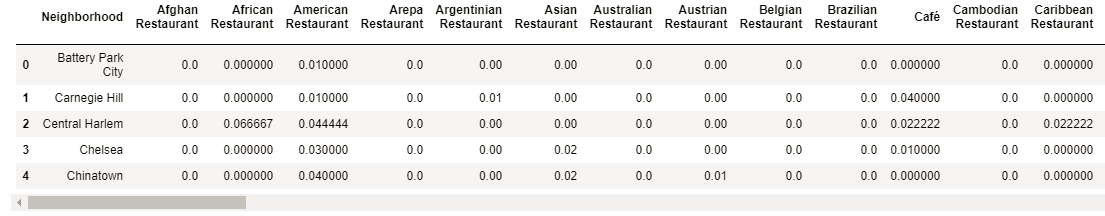
* Using the Python folium package and using the longitude and latitude each neighborhood is plotted on the map.



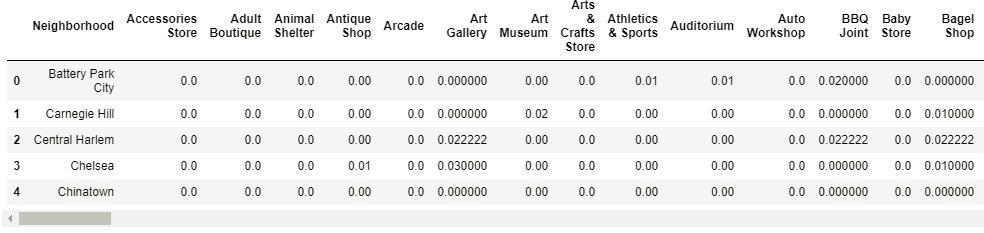
* Next, the venue details from the Manhattan neighborhoods are pulled from Foursquare site. Sample value –



* The venues are segregated into two groups – type RESTAURANTS & OTHERS
* Pandas dataframe is used to obtain the mean of the frequency of occurrence of each category
* Sample frequencies of different restaurants-



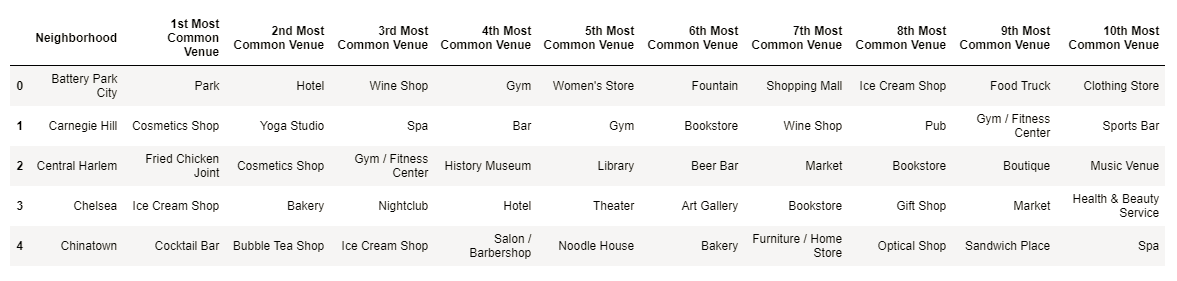
* Sample frequencies of venues other than restaurants –



* Next, we found the top 10 Restaurant types for each neighborhoods. Sample –



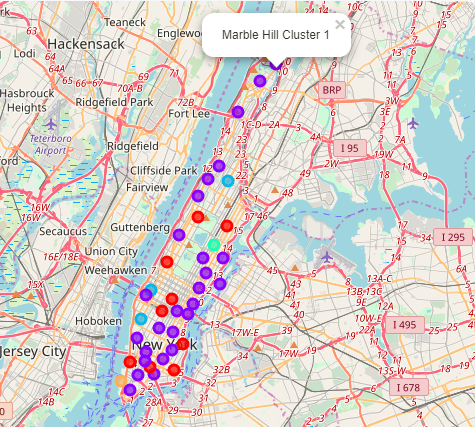
* Similarly, the top 10 venues other than restaurants are also obtained. Sample -



* Using k-mean clustering algorithm grouped the neighborhoods according to similar type of venues (other than restaurants) into 5 clusters. Sample –



* The 5 neighborhood venue clusters as generated by the algorithm are mapped using folium package with color coding –



* Cluster 1:



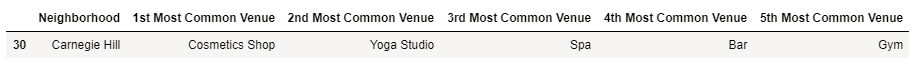
* Cluster 2:



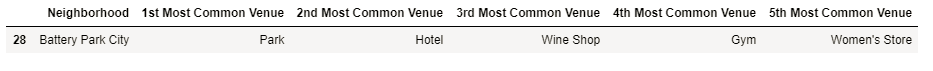
* Cluster 3:



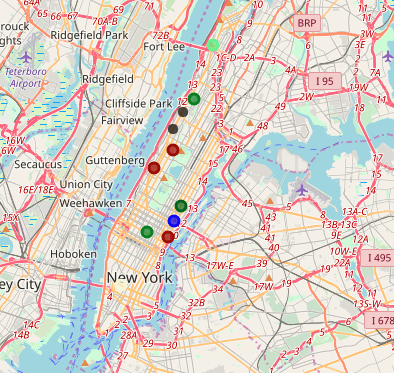
* Cluster 4:



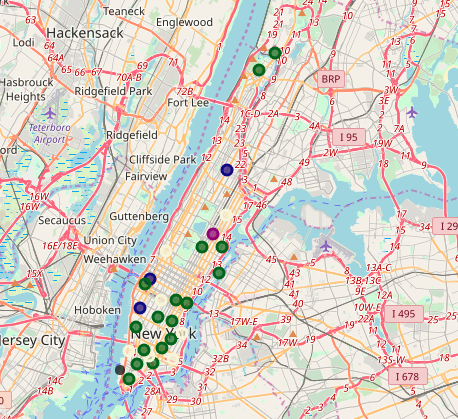
* Cluster 5:



* Next, an input is requested from the user to specify which type of new restaurant is of interest
* Once the input is provided the recommender system checks for the presence of that specific restaurant type in top 5 category for each neighborhoods and plot on the map the neighborhoods-



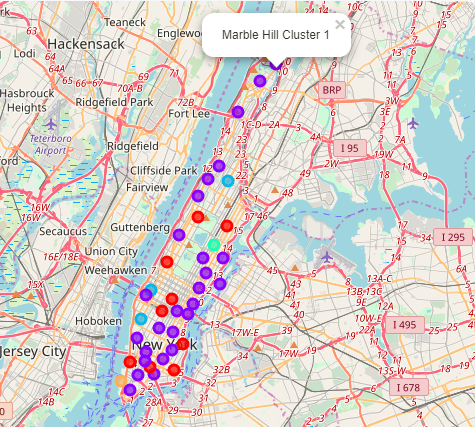
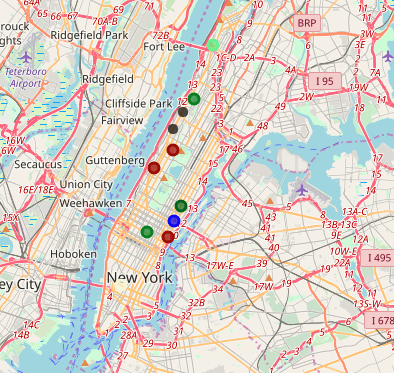
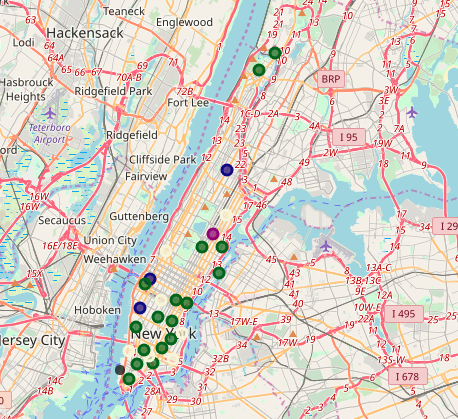
* Next it suggests the neighborhoods where it’s a best fit for opening the specific type of restaurant –



* The user visually looking on the 3 maps can decide which will be the best place to open the new restaurant.

# Results

Using the combination of visual guide from the maps and top venues in each cluster one can make an informed decision on which area is best for opening a new restaurant. The below maps are for new restaurant type – India

Neighborhood Clusters Neighborhood with Indian Restaurants Proposed neighborhoods for new one

The diverse neighborhoods with lots of business venues are in cluster 1 & 2 as seen from the cluster reports.

**Venues at Cluster 1:**



**Venues at Cluster 2:**



# Discussion

* The top 10 venues provide an idea of how diverse the Manhattan area
* The clustering provides an insight of the similarities in the neighborhoods

# Conclusion

Thus using the location data from Foresquare along with Machine Learning algorithm we can use a recommender system that will help to guide small business owners to make informative decision on which is the best neighborhood to a new restaurant.