**Restaurant Recommender System : The Art of Choosing the right Restaurant**

IBM Applied Data Science Professional

Debopriyo Nag

**Project Overview**

**Too Many Choices in life kills productivity**

Psychologists and economists studied that an overload of options may actually paralyze people or push them into decisions that are against their own best interest. A similar dilemma occurs if we need to choose a restaurant for an outing as per our interest. Websites like Yelp and Foursquare provides us with a plethora of restaurants classified into different categories to select. Consumers apply filters based on their criteria such as price, rating, reviews or location to narrow down the options but still ends up with a lot. A similar scenario happens while selecting movies to watch online. We have at least 4-5 streaming services to select. If we are successful to choose a streaming service(say Netflix), we have 1000’s of movies to choose from. In the end, we are tired and exhausted.

Thus, a system which browses through a user’s past activities(Restaurants visited, Movies watched) and recommends the top item in a particular category will immensely save users time and energy.

**Problem Statement**

In this problem we will develop a system to suggest the top Restaurant for a particular category in Manhattan borough to an user based on the Restaurants liked/tipped by the user in the past. The same methodology can be replicated for other cities across the world.

The tasks involved are the following:

1. Create similar clusters of Restaurants in Manhattan borough based on the category of the Restaurant and number of users who have liked that Restaurant(*More parameters like price, rating, tips can be used for more accurate recommendation but was omitted as it would require premium account*)

2. Analyse the Restaurant visited or tipped by a particular user. Predict the cluster of each Restaurant and display the most liked Restaurant based on the category.

**Target Audience**

The profile of the Recommender System user is a young, well-educated male.  
According to findings:  
  
44% Gen Y (18-29), 42% Gen X (30-43), 09% Young Boomers (44-53), 05% Senior Boomers (54-65), 0%  Senior (65+)  
  
The average age on those networks is 32  
The Female % is 22%  
Average household income $105,000  
And people that earned a college degree or higher is 70%

**Metrics**

Elbow Method For Optimal k

Using the elbow method to determine the optimal number of clusters for k-means clustering. Dataset A: Dataset B: Parse datasets. K-means is a simple unsupervised machine learning algorithm that groups a dataset into a user-specified number (k) of clusters.

**Data**

Data Exploration (Data or URLs used in the project are using free account of foursquare)

* New York Location Data obtained online(https://cocl.us/new\_york\_dataset) to explore and store venues for every neighbourhood
  + \*features\* key, list of all the neighbourhoods in New York
  + Populated 'Borough', 'Neighbourhoods', 'Latitude', 'Longitude' in “Neighbourhoods” pandas Data frame
  + Filtered “Manhattan borough” data from “Neighbourhoods” and stored in Pandas Data frame “manhattan\_data”
* Foursquare URL (https://api.foursquare.com/v2/venues/explore) to explore venue details of each Manhattan Neighbourhoods
  + Stored ‘Neighbourhoods', 'Neighbourhoods Latitude', 'Neighbourhoods Longitude','Venue ID','Venue', 'Venue Latitude', 'Venue Longitude','Venue Category' data in a pandas Data frame “manhattan\_venues”
* Foursquare URL (https://api.foursquare.com/v2/venues/VenueID/likes) to fetch number of likes each venue in “manhattan\_venues” received
* Foursquare URL (https://api.foursquare.com/v2/users/userid/tips) to fetch venues tipped by a user in the past(Example user id used 484542633)
  + Stored the features 'venue.name','authorInteractionType','venue.id','venue.location.lat' 'venue.location.lng’ ('authorInteractionType' is used to check whether user liked the venue)
* Foursquare URL (https://api.foursquare.com/v2/venues/VenueID/likes) to fetch venue details of venues tipped and liked by the user in the past
  + Stored the features 'Venue ID', 'Venue', 'Venue Latitude','Venue Longitude','Venue Category', 'Likes'

This data will produce the following deliverables  
  
A list of all Restaurants in Manhattan with features consisting of Venue ID, Venue Name, Venue Latitude, Venue Longitude, Venue Category, No. of Users Liked the Venue

A list of all Restaurants tipped and liked by a specific user with features consisting of Venue ID, Venue Name, Venue Latitude, Venue Longitude, Venue Category, No. of Users Liked the Venue

Most Liked Restaurant suggestion in “Manhattan Borough” for every Restaurant category tipped and liked by the user in the past

**Methodology**

Libraries and Packages

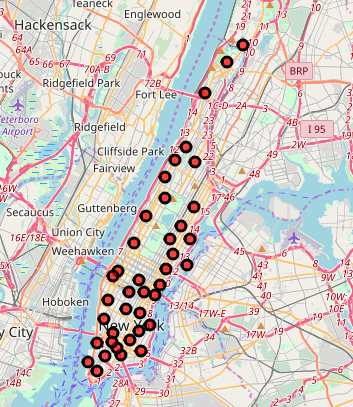
Imported the required libraries and packages like pandas, numpy, sklearn, matplotlib and folium

Data Preparation(Venues of Manhattan)

Downloaded New York location data from <https://cocl.us/new_york_dataset> and extracted features such as Borough, Neighbourhood, Latitude and Longitude.

After extraction, filtered the data to capture only Manhattan Borough in a pandas data frame.

Visualized all the neighbourhood in Manhattan.



Extracted the below features for every neighbourhood in Manhattan and stored in a pandas Data Frame using Foursquare API <https://api.foursquare.com/v2/venues/explore>.

'Neighbourhood'

'Neighbourhood Latitude'

'Neighbourhood Longitude'

'Venue ID'

'Venue'

'Venue Latitude'

'Venue Longitude'

'Venue Category’

Extracted the number of likes a particular venue received using Foursquare API <https://api.foursquare.com/v2/venues/Venue_ID/likes> and stored the same in the previous Data frame.[More Features like venue rating and price could have been added but they were not used because of free account restrictions]

The final Data Frame contained the below features.

'Neighbourhood'

'Neighbourhood Latitude'

'Neighbourhood Longitude'

'Venue ID'

'Venue'

'Venue Latitude'

'Venue Longitude'

'Venue Category’

‘Likes’

Dropper Categorical Features and used one Hot Encoding over “Venue Category” to ultimately fit in K-means Clustering.

Data Preparation(User Data)

Users usually tip a place which they either love or hate. Extracted all the venues tipped by the user using Foursquare API <https://api.foursquare.com/v2/users/user_id/tips>.

Extracted the below features and stored in a data frame.(Extracted only the venues loved by the user)

'Venue ID'

'Venue’

'Venue Latitude'

‘Venue Longitude'

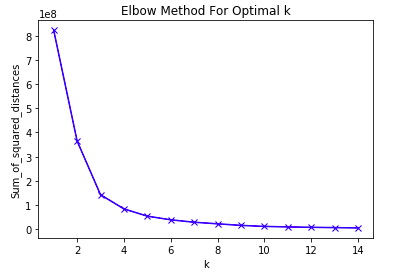
'Venue Category'

'Likes'

Grouped the result based on ‘Venue Category’ to find the most visited venues by the user.

Modelling

Finding the optimal K value using Elbow Method : Used elbow method below to find the optimal value of K for K means Clustering.

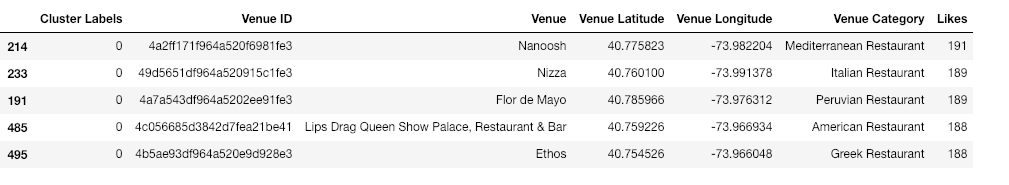


**Results**

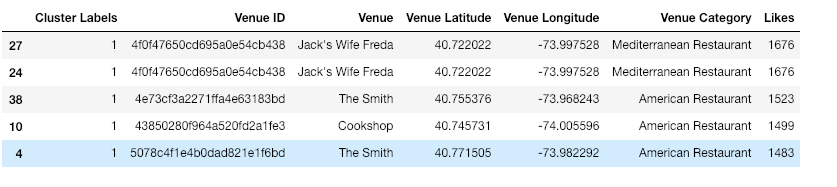
K=5 turned out to be the Optimal value for number of Clusters. Used K-Means Clustering to form similar clusters of Venues in Manhattan.

Only top 5 rows are shown below

Cluster 1(Least Popular : Won’t Come in Recommendations)



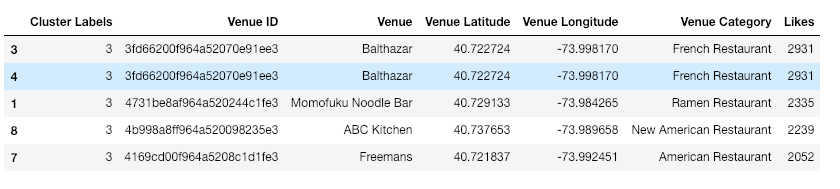
Cluster 2(Moderately Popular : Will Come in Recommendations once Most Popular is Exhausted)



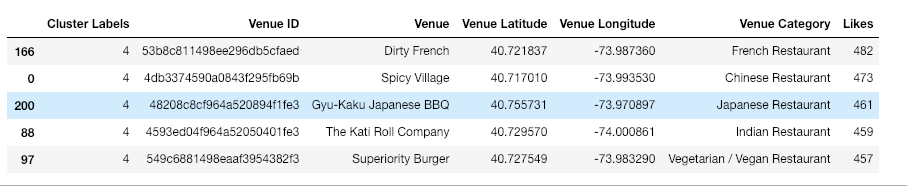
Cluster 3



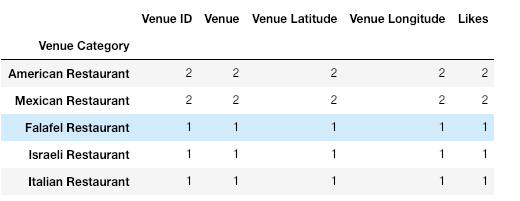
Cluster 4 (Most Popular : Recommendations for User)



Cluster 5



Top 5 Restaurant Types Visited by User(Sample User ID taken '484542633')



**Conclusion**

In this study, I analysed the different clusters of restaurants present in Manhattan based on the Restaurant Category and number of likes(popularity) received by the venue. More features such as ratings of a venue and cost of a Restaurant could have been used to have more accurate recommendations.

After the Clusters are formed, I built an user profile by analysing the most visited Restaurants of the user and matched the same with the clusters of Manhattan.

A recommendation from the cluster is done based on the most popular category of Restaurant visited by the user.

**Thus suggestions/recommendations to user will happen in below order,**

1. **American and Mexican Restaurants from Cluster 4**
2. **Other Restaurants from Cluster 4**
3. **American and Mexican Restaurants from Cluster 2**
4. **Other Restaurants from Cluster 2**
5. **American and Mexican Restaurants from Cluster 3**
6. **Other Restaurants from Cluster 3**
7. **American and Mexican Restaurants from Cluster 5**
8. **Other Restaurants from Cluster 5**
9. **American and Mexican Restaurants from Cluster 1**
10. **Other Restaurants from Cluster 1**