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

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 venue for an outing as per our interest. Websites like Yelp and Foursquare provides us with a plethora of venues 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(venues 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 venue for a particular category in Manhattan borough to an user based on the venues 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 venues in Manhattan borough based on the category of the venue and number of users who have liked that venue(*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 venues visited or tipped by a particular user. Predict the cluster of each venue and display the most liked venue 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 Venues 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 venues 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 Venue suggestion in “Manhattan Borough” for every venue category tipped and liked by the user in the past

**Exploratory Visualization**

The below plots will be delivered by the notebook

Fig. 1 A map of Manhattan with latitude and longitude values of each venue marked

Fig. 2 Elbow plot to find the optimal value of K for k-means clustering