Sumit Srivastava

Coursera Capstone Report

April 1, 19

Restaurant Recommendation System

## Introduction

* 1. **Problem Background**

Food is not just a necessity of life. The food we eat represents our culture, tradition, and values. The norms and values of a place can be significantly related to varieties of food available there. While visiting a place one of the most important factors we consider is the varieties of foods available there. But, one of the main problems is that people are unaware of famous restaurants and places available in a specific place. This not only applies to tourists but also to the local people.

Mumbai is the capital and largest city of the India and state of Maharashtra with a population of over 22 million.

The diversity of the cuisine available is reflective of the social and economic diversity of Mumbai. Roadside vendors, tea stalls, Maharashtrian, North Indian, Muslim food, Chinese and Western fast food are all very popular in the city. The Chinese food and the Thai food served in most of the restaurants are can be customized to cater to the tastes of the Indian population. Mumbai can also be called a foodie's paradise because of its vast variety of foods and edibles with a touch of Mumbai’s uniqueness and tradition.

With the increase in a number of restaurants people often get confused about the best-suited restaurant according to their preferences. In addition to that, people face a hard time to find out the best place and food to eat, especially when they are new to that place.

* 1. **Problem Description**

In the past, people obtained suggestions for restaurants from friends or other conventional sources or sites. Although this method is straightforward and user-friendly, it has some severe limitations. First, the recommendations from friends or other common people are limited to those places they have visited before. Thus, the user is not able to gain information about places less visited by their friends. Besides that, there is a chance of users not liking the place recommended by their friends. Second, the information provided by the site can often be biased; thus the information provided cannot always consider being accurate.

## Data Description

To consider the problem we can list the data as below.

* 1. I found the data about Mumbai’s borough and corresponding neighborhood using Wikipedia. The URL is mentioned below.

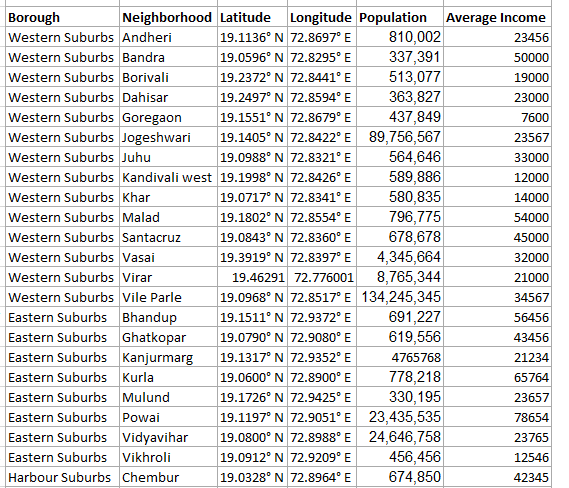
<https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Mumbai>

* 1. Google Map/Google Search was used to identify the latitude and longitude for each of the neighborhood.
  2. Population density for some of the neighborhoods have been found using the information available at the below mentioned URL. The neighborhoods for which the population density was not available over the internet have been assumed.

<http://www.indiaonlinepages.com/population/mumbai-population.html>

* 1. Neighborhood Income is assumed and may be inaccurate but since this is a demonstrating project, the main idea is to get the working model.

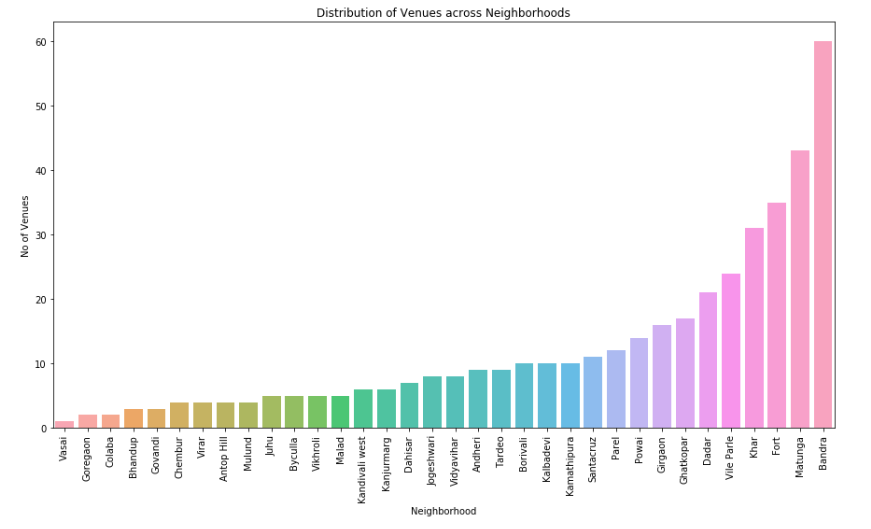
**Snapshot of the data**



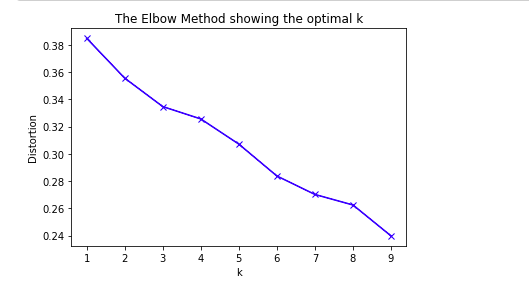
## Methodology:

**Exploratory analysis:**

Scrapping the data from different sources and then combining it to form a single-ton dataset is a difficult task. To do so, we need to explore the current state of dataset and then list up all the features needed to be fetched.

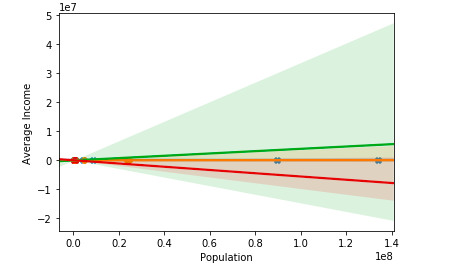
Exploring the dataset is important because it gives you initial insights and may help you to get partial idea of the answers that you are looking to find out from the data.

While exploring the dataset, I found out that Bandra has most number of venues while Vasai has the least.

Also while producing graph for number of cluster, I produced a graph to explore all the values for n\_clusters and then finding the best by exploring the elbow graph.

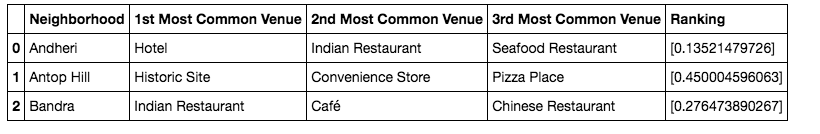
**Inferential analysis:**

Most important factors while building the recommender system were population and income. They are the most import factor because they have a nonlinear relationship according to our dataset.

It needed to make some inferential analysis to understand this nonlinear relationship. As the amount of population increases, it does not necessarily mean that average income of a neighborhood will also increase. It is true to most of the case but also many cases differ to follow this trend. Similarly, a neighborhood with less number of people may not necessarily have less average income. It is possible to have less number of people and more income and vice versa. This can be inferred from the above graph

## Result :

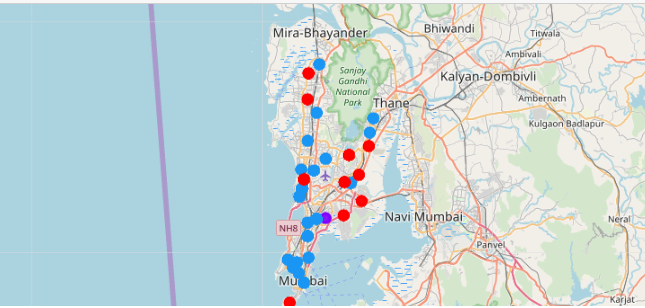
The result of the recommender system is that it produces a list of top restaurants and the most common venue item that the user can enjoy. During the runtime of the model, a simulation was done by taking ‘Andheri’ as the neighborhood and then processed through our model so that it could recommend neighborhoods with similar characters as that of ‘Andheri’.

The following image shows the result:

## Discussion

Since there was a nonlinear relationship between income and population, it can be concluded that we must always perform inferential approach to find relationship among different set of features. Also during clustering, similar neighborhoods must be dumped into the right cluster.

The following graph shows the clusters:



Another observation that we can make is that choosing number of clustering could produce very diverse results. Some may be over fitted or some may be under fitted. Hence analysis of number of clusters must be done. Ref elbow graph in the Methodology section.

## Conclusion

The recommender system is a system that considers factors such as population, income and makes use of Foursquare API to determine nearby venues. It is a powerful data driven model whose efficiency may decrease with more data but accuracy will increase. It will help users to finish their hunger by providing the best recommendation to fulfil all their needs.