**Analysis of Indian and Asian Restaurants in Mumbai**

**BHAGIRADH VAJJALA**

**JANUARY 06, 2021**

**1. Introduction**

**1.1 Background**

According to the [United Nations](https://en.wikipedia.org/wiki/United_Nations), as of 2018, Mumbai is the [second-most populous](https://en.wikipedia.org/wiki/List_of_cities_in_India_by_population) city in the country after [Delhi](https://en.wikipedia.org/wiki/Delhi) and the [seventh-most populous](https://en.wikipedia.org/wiki/List_of_largest_cities) city in the world with a population of roughly 20 million. Mumbai is more famous for tourism, food and gateway of India. As per Indian government population census of 2011, Mumbai was the [most populous city](https://en.wikipedia.org/wiki/List_of_cities_in_India_by_population) in India with an estimated [city proper](https://en.wikipedia.org/wiki/City_proper) population of 12.5 million living under [Municipal Corporation of Greater Mumbai](https://en.wikipedia.org/wiki/Municipal_Corporation_of_Greater_Mumbai). Mumbai is the centre of the [Mumbai Metropolitan Region](https://en.wikipedia.org/wiki/Mumbai_Metropolitan_Region), the sixth most populous metropolitan area in the world with a population of over 23 million. Mumbai lies on the [Konkan](https://en.wikipedia.org/wiki/Konkan) coast on the west coast of India and has a deep [natural harbour](https://en.wikipedia.org/wiki/Natural_harbour). It has the most number of millionaires and billionaires in the world. Mumbai is home to three [UNESCO World Heritage Sites](https://en.wikipedia.org/wiki/List_of_World_Heritage_Sites_in_India): the [Elephanta Caves](https://en.wikipedia.org/wiki/Elephanta_Caves), [Chhatrapati Shivaji Maharaja Terminus](https://en.wikipedia.org/wiki/Chhatrapati_Shivaji_Terminus), and [the city's distinctive ensemble of Victorian and Art Deco buildings](https://en.wikipedia.org/wiki/The_Victorian_and_Art_Deco_Ensemble_of_Mumbai).

**1.2 Problem**

The exploration of the neighbourhoods of Mumbai will help people to make their best decisions. This capstone mainly concentrates on exploring the neighbourhoods in Mumbai. Mumbai is one of the developed cities of India. The result of this project will be the customer can view all the venues of the neighbourhoods of Mumbai.

**1.3 Interest**

Every person who likes Indian food will be interested in this project. People who would like start an Indian restaurant will show their interest in this analysis.

**2. Data Acquisition and Data Cleaning**

**2.1 Data Sources**

The data used in this project is as follows:

* Neighbourhood Data of Mumbai
* Venues Data from Foursquare

Neighbourhood Data of Mumbai has been scrapped from <https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Mumbai>.

The JSON file consisting of all the venues in the neighbourhood has been retrieved using Foursquare API using Foursquare Credentials.

**2.2 Data Cleaning**

The neighbourhoods can be scrapped using web scrapping. The library used for web Scrapping is Beautiful Soup. Beautiful Soup is a Python library for getting data out of HTML, XML, and other mark-up languages. Say you’ve found some webpages that display data relevant to your research, such as date or address information, but that do not provide any way of downloading the data directly. Beautiful Soup helps you pull particular content from a webpage, remove the HTML markup, and save the information. It is a tool for web scraping that helps you clean up and parse the documents you have pulled down from the web. The requests module allows you to send HTTP requests using Python. The HTTP request returns a Response Object with all the response data (content, encoding, status, etc.).The find () and find all () methods are among the most powerful weapons in your arsenal.

The source of the venues is Foursquare API. Every API we need Client Secret, Client ID, version, limit. URL is the most important part of the program. Communicating with the foursquare database is really very easy, all thanks to their RESTful API. You simply create a uniform resource identifier, or URI, and you append it with extra parameters depending on the data that you are seeking from the database. Any call request you make is composed of, we can call this base URI, which is api.foursquare.com/v2, and you can request data about venues, users, or tips. But, every time you make a call request, you have to pass your URL. It is designed to give developers the freedom to adapt to Foursquare API changes on their own schedule. The output is the Json file which can be converted to Dataframe using pandas library.

**2.3 Feature Selection**

As we are interested in Indian and Asian restaurants, it is easy to filter them by applying a condition on ‘Venue Category’ column. In order to apply K-clustering the categorical variables must be turned into multiple binary variables. In order to get an overall idea I have grouped them based on their location. In order to get the probability I have found the mean of all the values. We can easily identify the number of restaurants in the particular neighbourhood.

**3. K-Cluster Algorithm**

K-means clustering is a type of unsupervised learning, which is used when you have unlabelled data (i.e., data without defined categories or groups). The goal of this algorithm is to find groups in the data, with the number of groups represented by the variable K. The algorithm works iteratively to assign each data point to one of K groups based on the features that are provided. Data points are clustered based on feature similarity. The results of the K-means clustering algorithm are:

1. The centroids of the K clusters, which can be used to label new data
2. Labels for the training data (each data point is assigned to a single cluster)

Rather than defining groups before looking at the data, clustering allows you to find and analyse the groups that have formed organically. The "Choosing K" section below describes how the number of groups can be determined.

Each centroid of a cluster is a collection of feature values which define the resulting groups. Examining the centroid feature weights can be used to qualitatively interpret what kind of group each cluster represents.

**3.1 Steps of K-Means Clustering**

* The first step in k-means clustering is the allocation of K centroids randomly (as K=6). Two points are assigned as centroids. Note that the points can be anywhere, as they are random points. They are called centroids, but initially, they are not the central point of a given data set.
* The next step is to determine the distance between each of the data points from the randomly assigned centroids. For every point, the distance is measured from both the centroids, and whichever distance is less, that point is assigned to that centroid. You can see the data points attached to the centroids and represented here in blue and yellow.
* The next step is to determine the actual centroid for these two clusters. The original randomly allocated centroid is to be repositioned to the actual centroid of the clusters.
* This process of calculating the distance and repositioning the centroid continues until we obtain our final cluster. Then the centroid repositioning stops.

**4. Results**

**1) Neighbourhoods of Mumbai**



Figure 4.1 Neighbourhoods of Mumbai

**2) Map of Mumbai**

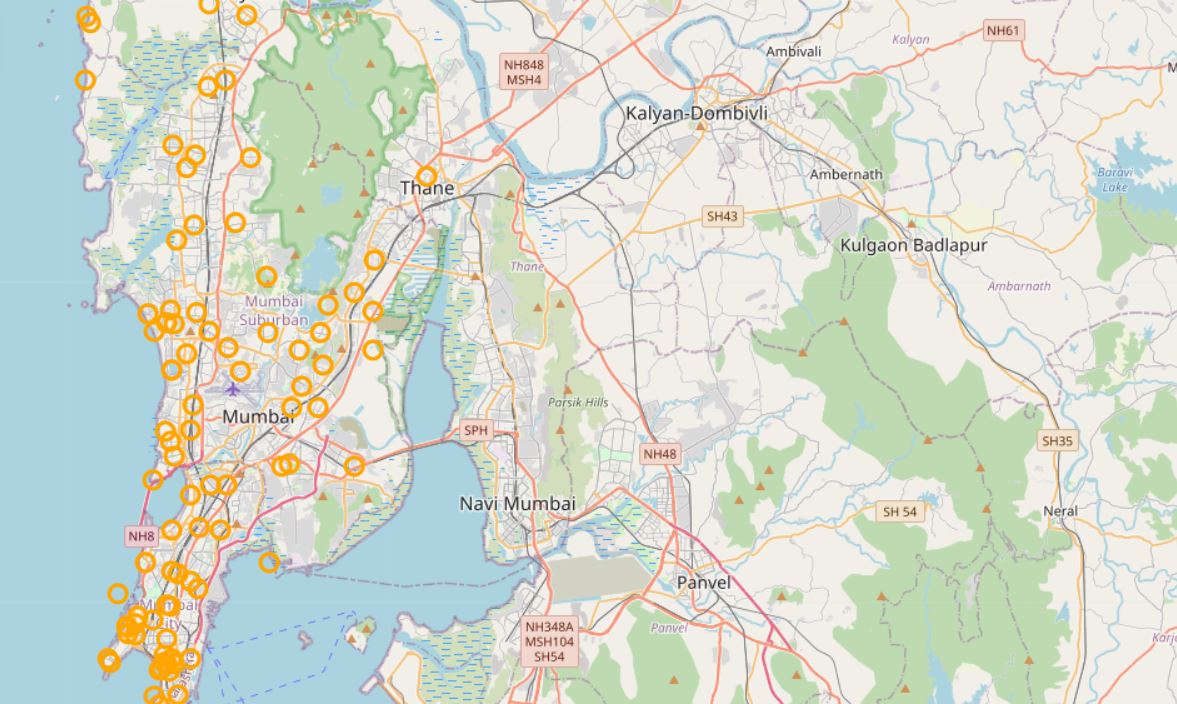


Figure 4.2 Locations of Restaurants on Map

**3) Indian and Asian Restaurants in Mumbai**



Figure 4.3 Lists of Indian and Asian Restaurants

**4) Map of Indian and Asian Restaurants in Mumbai**

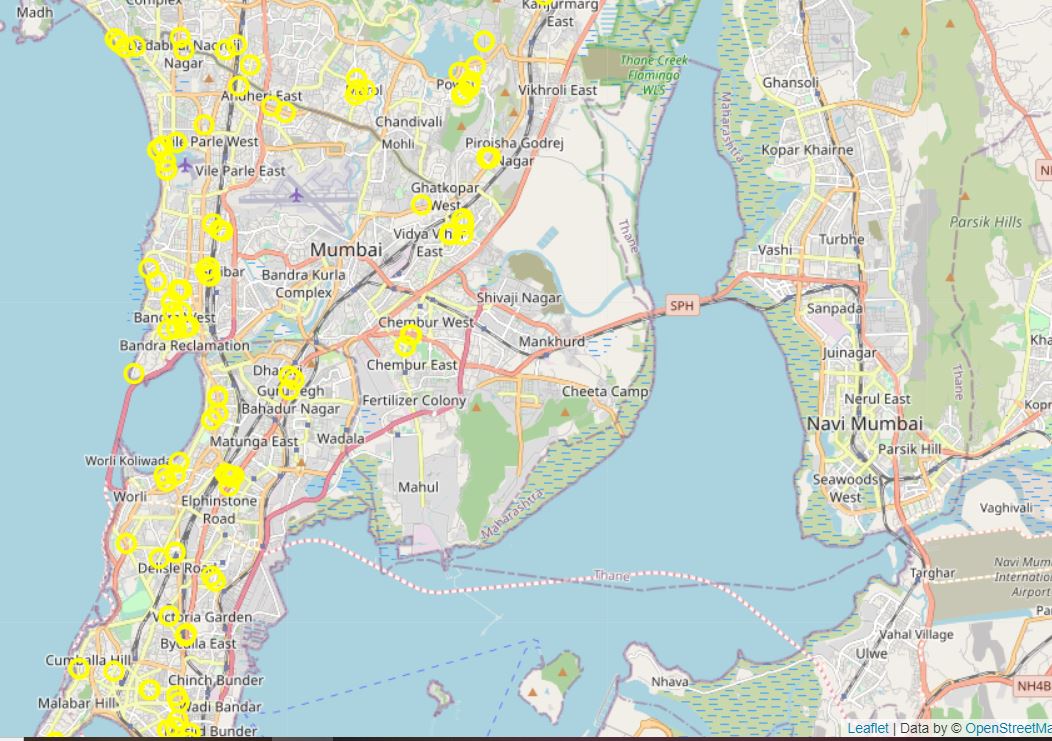
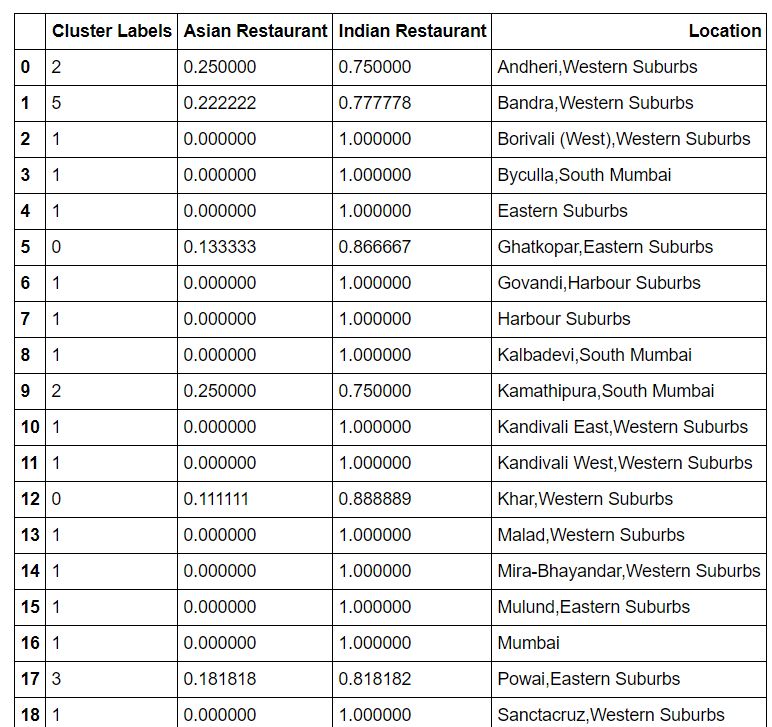


Figure 4.4 Locations of Indian and Asian Restaurants in Mumbai.

**5) Cluster Labels**



‘’’

Figure 5.1 Cluster labels assigned to the Data

**6) Results of Cluster Categories.**

**6.1) Results of K=0**

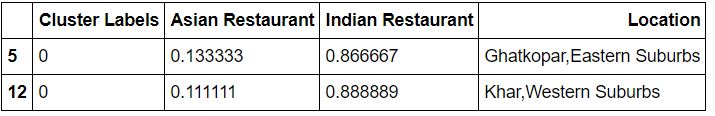


Figure 6.1 Asian and Indian Restaurants with Category 0

**6.2) Results of K=1**



Figure 6.2 Asian and Indian Restaurants with Category 1

**6.3)** R**esults of K=2**

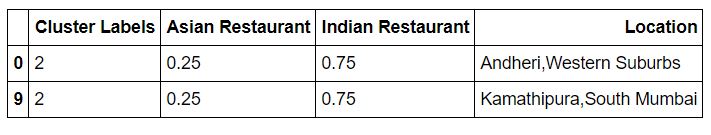


Figure 6.3 Asian and Indian Restaurants with Category 2

**6.4) Results of K=3**

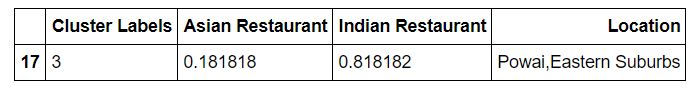


Figure 6.4 Asian and Indian Restaurants with Category 3

**6.5) Results of K=4**

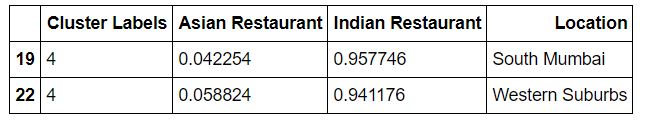


Figure 6.5 Asian and Indian Restaurants with Category 4

**6.6) Results of K=5**

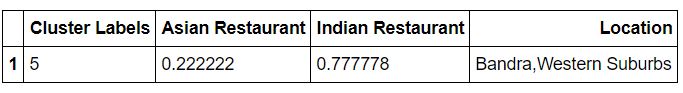


Figure 6.6 Asian and Indian Restaurants with Category 5

**5) Conclusion**

We have analysed all the Indian and Asian Restaurants present in the neighbourhoods in Mumbai. Every person who likes Indian food can view the survey and take necessary decisions.