RESTAURANTS

IN MUMBAI

Rhitvik Pasricha

Applied Data Science Capstone

Introduction/Business Problem

According to the website https://worldpopulationreview.com/world-cities/mumbai-population/, Mumbai has a population of about 22 million people which creates a high demand for places like restaurants and cafes where people can socialize in a preordained ambiance.

As stated by a blog on www.torqus.com, Mumbai is known to be a melting pot of all regional tastes of India. It is not a surprise that a city like Mumbai that offers impeccable and varied flavours for all palates and all pocket sizes stands for a unique gastronomical experience. From roadside food stalls to hoity-toity restaurants, Mumbai serves it all.

The overall Indian foodservice industry is estimated to be Rs 5.99 trillion by 2022-23, growing at a compounded annual growth rate of 9 percent, a National Restaurant Association of India (NRAI) report said. Pegging Mumbai's organized foodservice market at Rs 40,480 crore, the report noted it was the highest amongst metros in the country.

Target Audience

Entrepreneurs interested in the opening Fast Food Restaurant in Mumbai, who may require external advice concerning what type of restaurants are in trend and where exactly it should be originated which could create a high chance of being successful. Fast Food Restaurant category is based on personal preference but this analysis can be applied to any business.

Data Description

I used the following data to analyze the city of Mumbai

1. Neighbourhood Data

Using pandas converted the table on https://www.mapsofindia.com/pincode/india/maharashtra/mumbai/ website consisting name of the neighbourhood with respect to its Pin code.

2. Geopy.geocoders

Using geopy.geocoders library to get the latitude and longitude of each neighbourhood in Mumbai.

3. Foursquare API

Using Foursquare API to explore venue information for each neighbourhood in the Mumbai city. Some features extracted include 'Venue', 'Venue Category', 'Venue Latitude', 'Venue Longitude', etc.

Methodology

After extracting the data of Mumbai neighbourhoods and grouping them by their pin codes I got the following data frame containing **182** neighbourhoods.

	Location	Pincode
1	A I staff colony	400029
2	Aareymilk Colony	400065
3	Agripada	400011
4	Airport	400099
5	Ambewadi	400004
6	Andheri	400053
7	Andheri East	400069

I used the **goepy** library to pass the name to get the latitudes and longitudes and merged the data into the above dataset.

	Location	Pincode	Latitude	Longitude
1	A I staff colony	400029	19.151622	72.854981
2	Aareymilk Colony	400065	19.180136	72.908812
3	Agripada	400011	18.975302	72.824898
4	Airport	400099	19.090201	72.863808
5	Ambewadi	400004	19.186776	72.859313
6	Andheri	400053	19.119698	72.846420

The **folium** library was used to visualize the geographic details of Mumbai and superimposed the markers on the map indicating the neighbourhoods with a popup text.
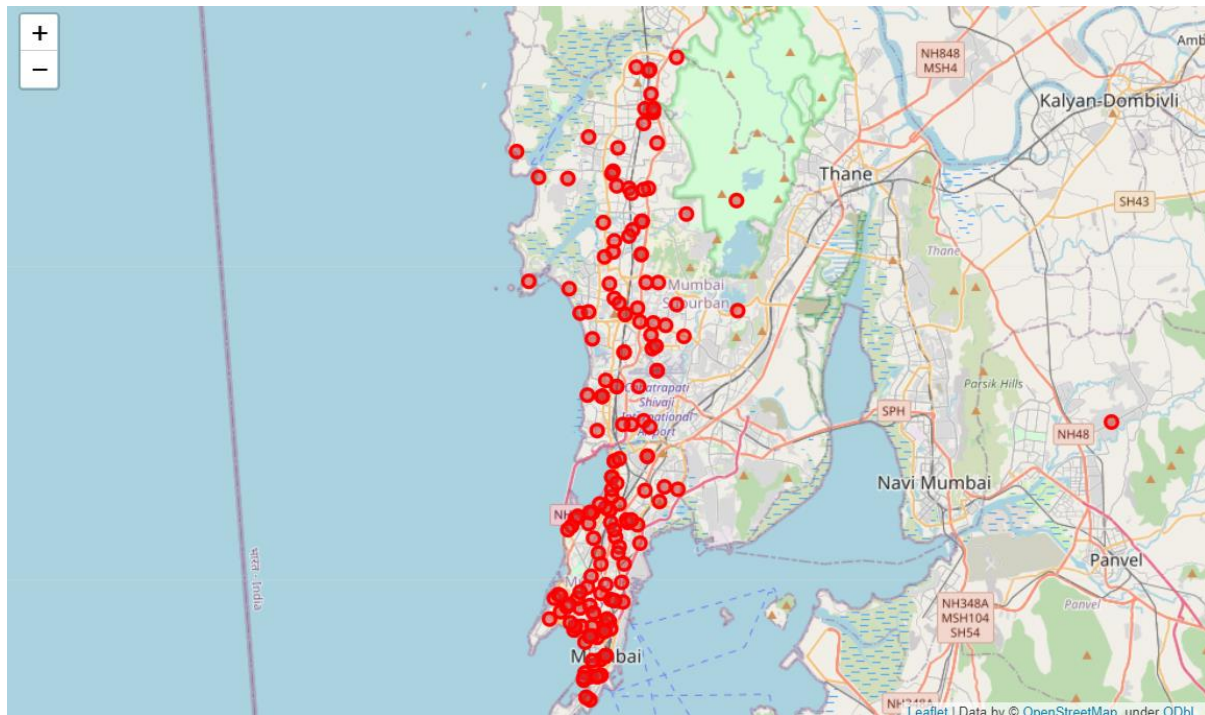

I utilized the Foursquare API to explore the boroughs and segment them. It fetches the venues located nearest to location passed through the URL. I assigned the limit as **200** and the radius as **500** meters for each neighbourhood from their given latitude and longitude information. From the data gathered I created the following data frame consisting of **2738** venues categorized by the venue type.

Exploring the dataset is important because it gives you initial insights and may help you to get a partial idea of the answer that you are looking to find out from the data. By exploring the above dataset I found out that "Indian Restaurants" was the most reoccurring venue category and the "Fast Food Restaurant" category has 80 unique venue locations in the dataset.
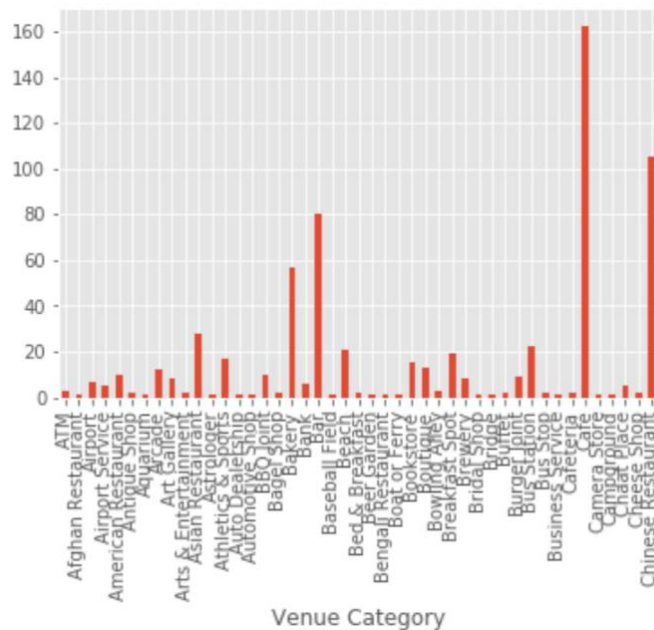
For this study, I used one-hot encoding and standardization to process our categorical variable (venue category) into a form that machine learning algorithms can further classify the data. Since we have some common venue category that is why I decided to choose K -means algorithm for clustering the dataset as it is one of the simplest and popular unsupervised machine learning algorithms. To decide the optimal number of clusters I used the Elbow method.
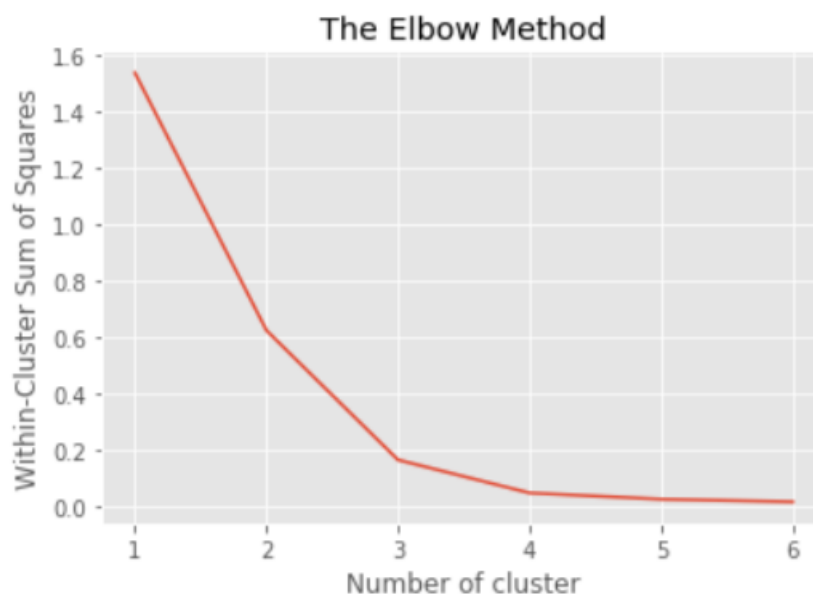
Here is my merged table with cluster labels for each Location.

	Pincode	Location	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue
0	400029	A l staff colony/Santacruz P&t colony	19.151622	72.854981	1	Indian Restaurant	Lounge	Chinese Restaurant	Tea Room	Pizza Place	Plaza	Fast Food Restaurant
3	400099	Airport/International Airport/Sahar P & t colo...	19.103066	72.863560	2	Hotel	Gym	Fast Food Restaurant	Café	Donut Shop	Farmers Market	Farm
4	400004	Ambewadi/Charni Road/Chaupati/Girgaon/Madhavba...	18.954329	72.821730	1	Ice Cream Shop	Fast Food Restaurant	Juice Bar	Train Station	Bus Station	Italian Restaurant	Snack Place
5	400053	Andheri/Azad Nagar	19.128016	72.832513	1	Lounge	Pub	Pizza Place	Indian Restaurant	Athletics & Sports	Liquor Store	Restaurant
6	400069	Andheri East/Nagardas Road	19.130182	72.853080	1	Gym	Coffee Shop	Bike Rental / Bike Share	Bus Station	Athletics & Sports	Bakery	Dessert Shop

Exploring the dataset to see the distribution of venues among the clusters.
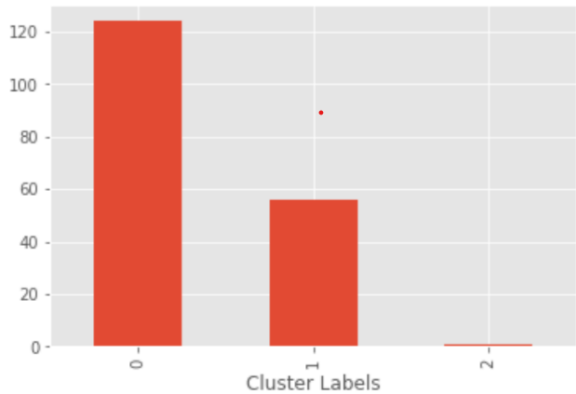
Result

## Cluster 1: Neighborhoods where Fast Food Restaurants aren't a common venue

	Pincode	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
3	400011	Coffee Shop	Athletics & Sports	Platform	Bank	Bakery	Soccer Field	Goan Restaurant	Design Studio	Farm	Falafel Restaurant
4	400099	Airport	Spa	Airport Service	Yoga Studio	Dim Sum Restaurant	Fast Food Restaurant	Farmers Market	Farm	Falafel Restaurant	Event Space
5	400004	Women's Store	Indian Restaurant	Department Store	Business Service	Dim Sum Restaurant	Fast Food Restaurant	Farmers Market	Farm	Falafel Restaurant	Event Space
9	400037	Indian Restaurant	Gym / Fitness Center	Trail	Bus Station	Dessert Shop	Farmers Market	Farm	Falafel Restaurant	Event Space	English Restaurant
10	400005	Bus Station	Yoga Studio	Dim Sum Restaurant	Fast Food Restaurant	Farmers Market	Farm	Falafel Restaurant	Event Space	English Restaurant	Electronics Store

Cluster 2: Neighbourhoods where Fast Food Restaurant are a common venue

	Pincode	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
1	400029	Indian Restaurant	Fast Food Restaurant	Lounge	Pizza Place	Chinese Restaurant	Plaza	Sporting Goods Shop	Electronics Store	Bar	Dim Sum Restaurant
6	400053	Indian Restaurant	Fast Food Restaurant	Café	Restaurant	Burger Joint	Sandwich Place	Food Court	Bakery	Paper / Office Supplies Store	Gastropub
7	400069	Hotel	Indian Restaurant	Bar	Shopping Mall	Camera Store	Fast Food Restaurant	Chinese Restaurant	Pizza Place	Event Space	Dim Sum Restaurant
8	400058	Indian Restaurant	Fast Food Restaurant	Café	Restaurant	Burger Joint	Sandwich Place	Food Court	Bakery	Paper / Office Supplies Store	Gastropub
20	400007	Pizza Place	Fast Food Restaurant	Restaurant	Yoga Studio	Design Studio	Farm	Falafel Restaurant	Event Space	English Restaurant	Electronics Store

Cluster 3: Neighborhoods where Fast Food Restaurant are the most common venue

	Pincode	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
142	400060	Fast Food Restaurant	Yoga Studio	Dessert Shop	Farmers Market	Farm	Falafel Restaurant	Event Space	English Restaurant	Electronics Store	Donut Shop
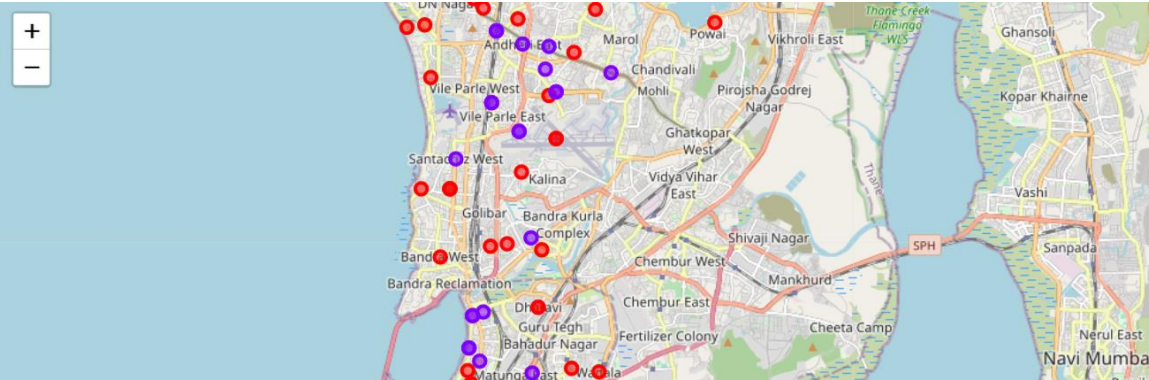
In the final section, I created a map to visualize all clusters on the map of Mumbai. The results are visualized in the above map with Cluster 0 in red color, Cluster 1 in purple color, and Cluster 2 in light green color.

Discussion

The fast-food restaurants' being the most common venue is observed in cluster 2, followed by moderate density in cluster 1 and none in cluster 0.

Therefore, when choosing a neighbourhood to open a fast-food restaurant, it is advisable to consider locations in Cluster 0 and 1 because of the lack of competition and the ability to conquer the market. It is also important to be centrally located in the city to cover the maximum radius of potential customers.

Limitation

This project can be further improved by adding constraints like **population density**, **property costs**, and the proximity of **farmers market**. This analysis allows the developer to provide information on a more profitable place to set up shop for their services. Reducing competition and differentiating your business from those around you allows businesses to develop strategies to grow and thrive. However, to put all these data into this project is not possible to do within a short time frame for this capstone project. Also, the **Foursquare API** used had a limit of 200 calls in a radius of 500m, therefore, all the predictions are based on the data gathered by the Foursquare data platform.

Conclusion

This project was concerned with the data science methodology where one identified a business problem, collected the data required, processed the data, used relevant methods to prepare data for analysis by the appropriate machine learning techniques, and lastly used the results to recommend solutions to relevant entrepreneurs.

Based on the Clusters formed it would be a good idea to open a restaurant in Clusters 0 & 1 since the other clusters already have fast-food restaurants in their vicinities. Also, clusters 0 & 1 have many restaurants in the vicinity (Pizza restaurants, cafe, coffee shop, Indian restaurants) so one will be able to attract a good crowd.