

Course:

IBM Applied Data Science Capstone

Survey for opening Food Joint in Mumbai Neighbourhoods,
India

By: Vishal Ochani

Aug 2020



Table of Content

| | |
|------------------------------------|----|
| 1. Business Problem----- | 3 |
| 2. Background----- | 3 |
| 3. Target Audience----- | 4 |
| 4. Data----- | 4 |
| 4.1 Data Requirement----- | 4 |
| 4.2 Data Extraction and Usage----- | 5 |
| 5. Methodology----- | 7 |
| 5.1 Exploratory Analysis----- | 8 |
| 5.2 Statistical Inference----- | 11 |
| 6. Result----- | 12 |
| 7. Discussion----- | 14 |
| 8. Conclusion----- | 15 |

1. Business Problem:

This Capstone is intended to analyse and fetch the appropriate location in Mumbai's Neighbourhood which will be suitable to open a food joint for business man who don't want to invest huge amount as required to start Restaurants but still want to earn well.

Basically need is for small food joint, food truck, take away counters etc.

2. Background:

People in India love to have different variety of food especially when they are roaming /travelling outside may it be alone, with family or friends.

Usually people travel using public transport (Trains and Buses) for work, as it is known fact that Mumbai is financial capital of India and it has many immigrants from various states of India. Most of the people uses Local trains to commute from one place to another and likes to have some quick food near the station which is cheap and best.

Mumbai has lot of tourist places such as Beaches, Film City, Zoo, Temples, Marine drive etc. which attracts huge crowds on weekends along with enjoying the places they are also found of some yummy food to satisfy their taste buds.

This business of small fast food joint, Food Truck, Chaat items has a lot of potential provided the selection of Area is appropriate.

Proper Area with lot of crowd will make sure the food joints will make a quite good business.

3. Target Audience:

1. Business minded people those who are interested to do side hustle on weekends.
2. Young college students who wants to try their luck in food business.
3. People who can cook very well but are low on budget.
4. People who make pickles and other preserved food stuff at home and want to sell it.
5. People who want to expand their food chains in different areas can start off with small investment, it will make it clear if their food is accepted in different areas.

4. Data:

4.1 Data Requirement:

1. List of Mumbai's Neighbourhood:

This will help to identify the neighbouring areas with which further analysis can be done

2. Latitude and Longitude of Neighbouring areas:

This is required to plot areas on map and also for getting the venues in these areas.

3. Venues in each Neighbourhood:

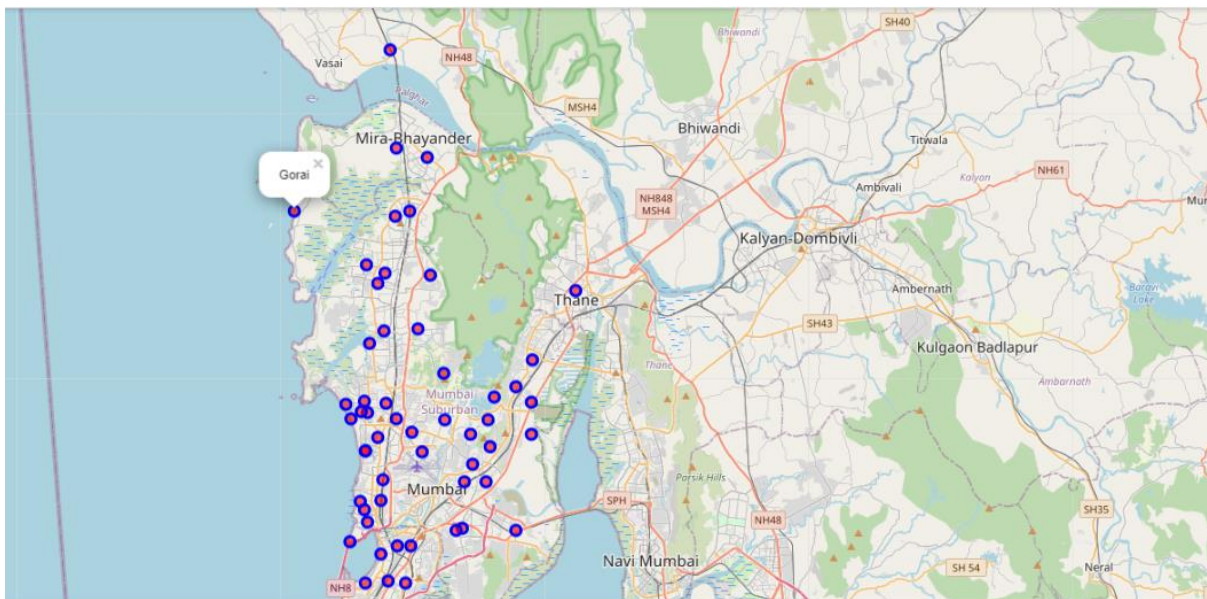
This is the most important data, as this will get the insights of the places where crowd is attracted.

4.2 Data extraction and Usage of data to solve the problem:

Data will be extracted by using following websites

- Wikipedia: To get the list of neighbouring areas of Mumbai along with its Latitude and Longitude, it will be extracted by using python web scrapping (using pandas)
- Foursquare.com: This site will help us to extract venues for each neighbouring area, data will be extracted by using foursquare API

The venues data has been plotted on Map as shown below



Usage of Data:

- Once we extract the data, we will merge both the data mentioned above and make the DataFrame using pandas
- Now we will have Neighbouring Area and Venue category along with their co-ordinates, sample data shown below

| | Area | Area Latitude | Area Longitude | Venue Name | Venue Category | Venue Latitude | Venue Location |
|---|--------|---------------|----------------|------------------|--------------------|----------------|----------------|
| 0 | Amboli | 19.1293 | 72.8434 | Cafe Arfa | Indian Restaurant | 19.128930 | 72.847140 |
| 1 | Amboli | 19.1293 | 72.8434 | 5 Spice , Bandra | Chinese Restaurant | 19.130421 | 72.847206 |
| 2 | Amboli | 19.1293 | 72.8434 | Subway | Sandwich Place | 19.127860 | 72.844461 |
| 3 | Amboli | 19.1293 | 72.8434 | Cafe Coffee Day | Coffee Shop | 19.127748 | 72.844663 |
| 4 | Amboli | 19.1293 | 72.8434 | Delhi Zaika | Halal Restaurant | 19.132159 | 72.844406 |

- Will filter out the Venue categories based on the fact if they are tourist place or Station, as these are the places with huge crowd.

| | Area | Art Gallery | Beach | Beer Garden | Boat or Ferry | Bus Station | Chaot Place | Comedy Club | Food Truck | Garden | History Museum | Light Rail Station | Monument / Landmark | Park | Performing Arts Venue | Playground | Snack Place | Train Station |
|---|------------------|-------------|-------|-------------|---------------|-------------|-------------|-------------|------------|--------|----------------|--------------------|---------------------|------|-----------------------|------------|-------------|---------------|
| 0 | Chakala, Andheri | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 1 | D.N. Nagar | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 2 | D.N. Nagar | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 | Lokhandwala | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4 | Marol | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |

- Will wrangle the data if required
- In the new filtered dataset will apply machine learning algorithm to find the clusters using K-means cluster.
- Each cluster will gives us insight about the areas suitable for opening food joints
- Based on cluster wise insights, decision is to be taken to select appropriate location

5. Methodology

We are intended to explore the venues in nearby Areas of Mumbai and decide the appropriate location to open the Food Joint which can be profitable if the location can attract the good number of Crowd

Usually in any place the huge number of crowd will be on below locations:

1. Public Transport (Bus/Train Station)
2. Tourist place
3. Parks/Gardens

Will Analyse the Public transport and already available food joints in the particular area, so as to arrive at accurate conclusion

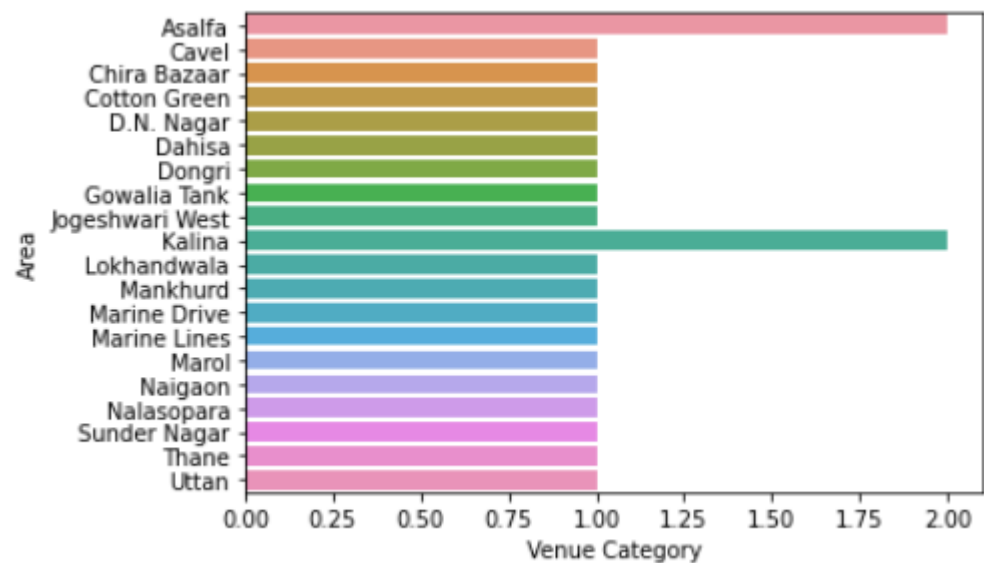
Our focus will be to extract the crowd attracting places and make the cluster out of those places using **KMeans clustering algorithm** and to conclude which clusters/Area are best suitable for food joint

Here we will use Clustering, since it will have different areas clubbed together based on the similarities, then it will be easy to check which all areas have more number of tourist places and public transport and help to identify appropriate location for food joint.

5.1 Exploratory Analysis:

1. Area wise count of Public Transport:

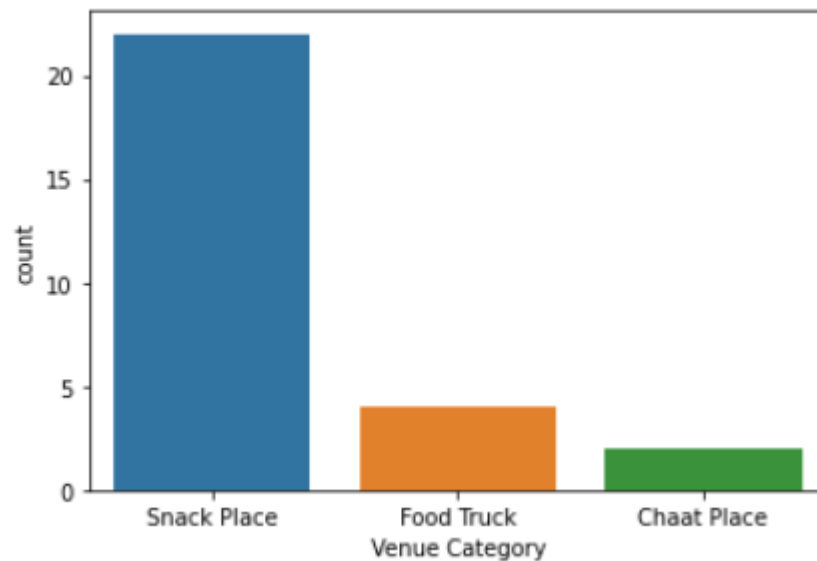
- We have filtered the data frame having all the Station bus/train
- Then we have grouped the filtered data by Area on count of station in each Area



- The above Graph shows the Area vs Venue category count and Venue category here is public transport.
- Areas namely **Kalina** and **Asalfa** have more than 1 Station, this will be useful for conclusion once we have clusters

2. Already Available Food Joints:

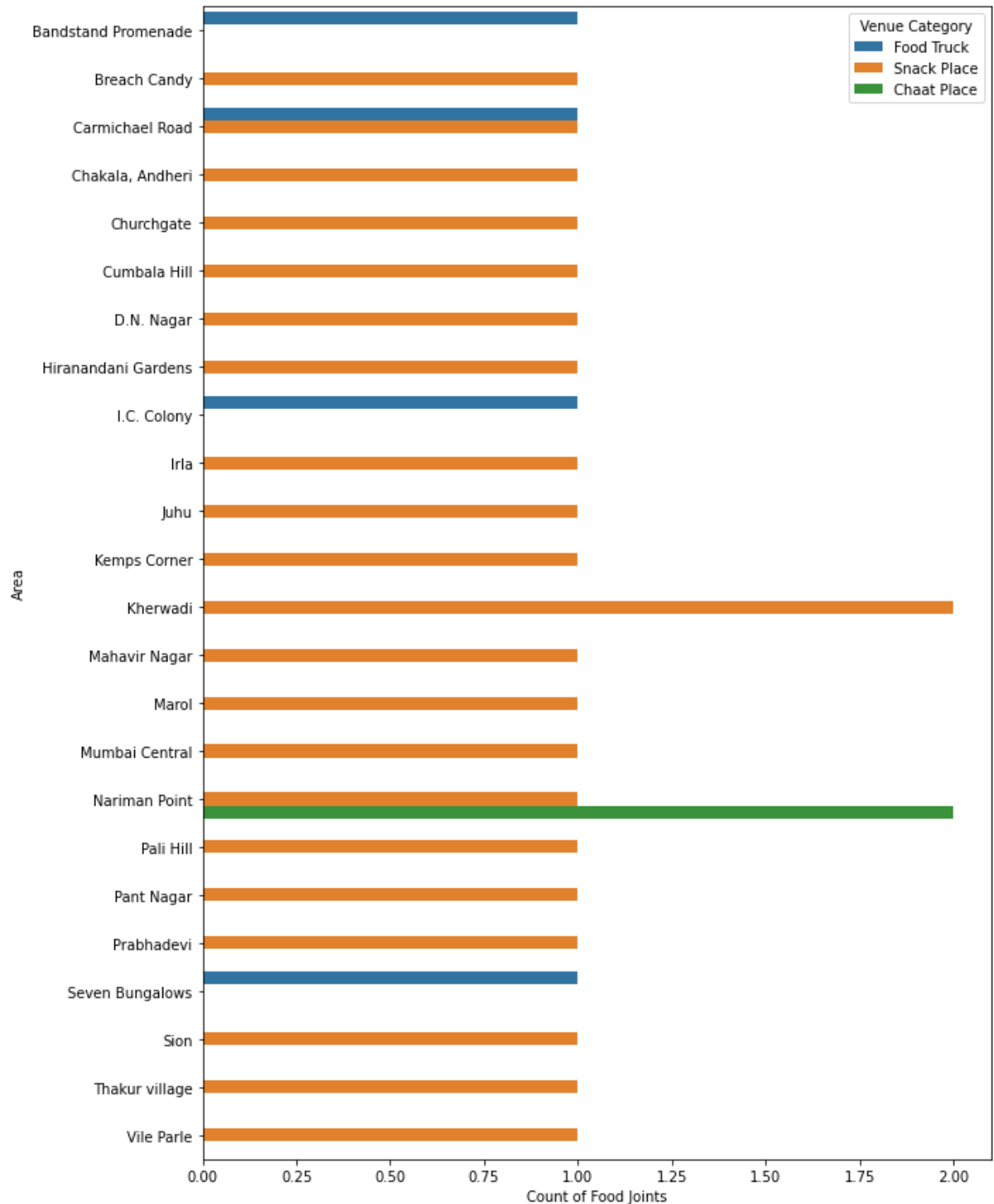
- In the data frame there are already places that has the food joints which we are exploring for
- Filtered data frame having food joints only
- Lest count the different type of food Joints, to check which type of food Joint has great potential



- There is great potential for Food Truck and Chaat place

3. Area wise food joint count:

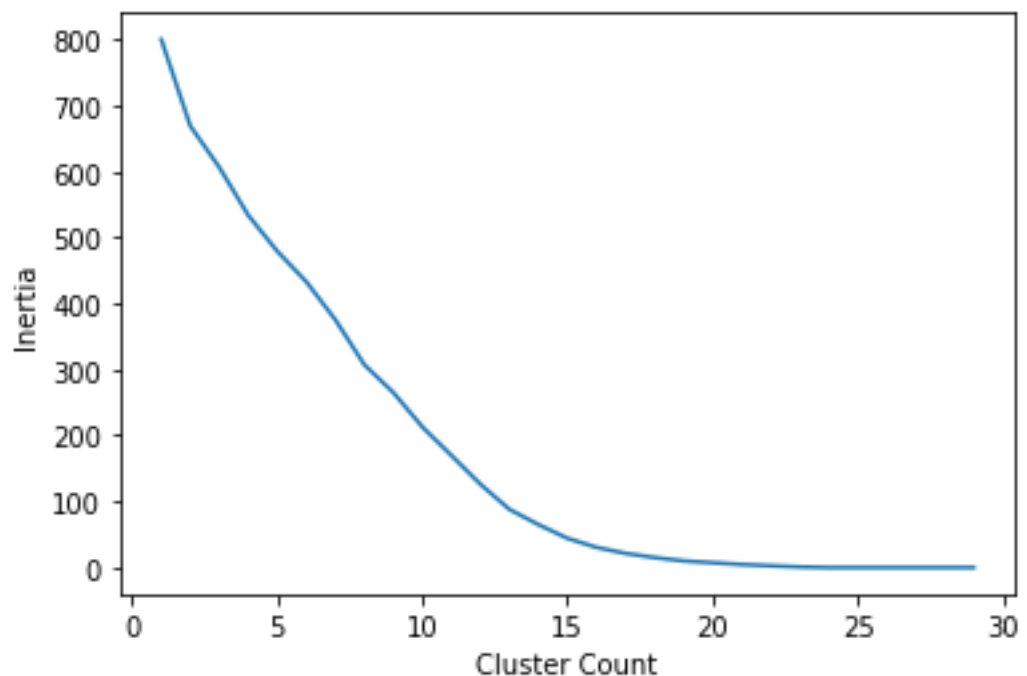
- We can find the area wise count of food joint by filtering the data frame to have food joint and groupby with Area
- This will help to understand which area has competition and areas of potential business



- **Nariman Point** and **Charmichael Road** are the areas with 2 Food joints, will keep these area under observation in clusters. Rest areas have good opportunity.

5.2 Statistical Inference:

- For applying KMeans clustering Algorithm, we need to identify the optimum number clusters required to the appropriate output.
- We will use Inertia and Elbow method to identify the number of clusters
- Inertia is the sum of squared error for each cluster. Therefore the smaller the inertia the denser the cluster
- As per the Elbow method the number of cluster at the elbow where it starts just changing the curve is the optimum value.



- It seems 14 is the optimum value of cluster, as at this value elbow is formed

6. Result:

Now that we have all the data in hand, we got the clusters derived the insights as well.

Let's discuss the details from our analysis

1. Area wise Public Transport:

From the barplot we got two areas Kalina and Asalfa having more than 1 Station, both these areas are in cluster 4

These areas have added advantage as more the Station more the crowd, will keep these places in mind while taking decision

2. Area vs Food Joint Bar plot:

Nariman Point and Charmichael Road are the two areas which have more than 1 food joint already available

Nariman Point is in cluster 4

Charmichael Road is in cluster 0

Competition is already high in these areas, will need to check if cluster 4 and cluster 0 has good crowd or else we can drop these areas.

Cluster Insights:

Clusters 2,3,5,7,8,11,12,13 are not worth for opening food joints as these areas lack in crowd attracting areas.

Now we are left with

Cluster 0:

Has only Two areas has tourist place (Beach) those are Juhu and Vile Parle

Cluster 1: Has more areas compared to cluster 0 which can attract crowd

There are Beaches, Bus station and playground, both the areas in cluster 1 are good for Food Joint as both of them don't have any food joint

Cluster 4: It has Station only, meaning people looking for part time food business

Suitable Areas: Kalina and Lokhandwala

Cluster 6: One of the Area Gowalia Tank looks attractive, as it has Bus Station, Park, Museum

Cluster 9: It has Station only, meaning people looking for part time food business

Suitable Areas: All areas have Station, area to be selected as per crowd by exploring it for few days

Cluster 10: Seems to have good tourist and crowd attracting places, it has

Beach, Boating and Arts venue and it has good potential for Food joint

7. Discussion:

From the Above Result

- **Cluster 0** is ruled out as it has only beaches, meaning little less crowded as tourist have beaches in other areas as well with few more attraction and also in rainy seasons there will be no business in beach areas.
- **Cluster 1 and 10** has beaches, boating, bus station and playground, both of these cluster looks very lucrative as they have multiple crowding attracting places, one or the other places will keep generating business in all seasons
- **Cluster 4 and 9** are good bet for part time food business, this is really great for those who don't want to put their all-time in food business and are interested in side hustle or it may be useful for women to do both business as well as be home maker.
- **Cluster 6** as standalone looked good but compared to other cluster this isn't much interesting as it has Museum and Station compared to others will have low crowd.

8. Conclusion:

As our intention was to find out the best location to have food joint, we wanted to have areas with good crowd so that business can grow and people roaming or visiting such places get benefits of having delicious food at very lower rates

It has two Categories:

1. Full time business:

Cluster 1 and Cluster 10 are best suited for Full time business.

Considering Mumbai has beaches and Tourist usually enjoy going for Beaches and Boat Rides, so cluster 10 has all the required places to attract good crowd

Food Joint recommendation for Full time business goes to Cluster 10 and Area is **Bandstand Promenade**

2. Part time Business:

Cluster 4 and Cluster 9 are having good number of Public Transports and both the clusters will have great crowd in peak hours (while commuting to/from office or colleges)

Food Joint Recommendation goes to Cluster 4's Area **Kalina** and **Asalfa** as both has two station each.

While one can explore Cluster 9 as well but it should strictly based on the crowd behaviour

Bonus recommendation is to focus on “**Food Truck**” in both as they are very less in number and can be located from place to place. Also every type for food can be prepared in food truck and is movable.