The Battle of Mumbai Neighborhoods For Next Chinese Restaurant

By

Ziauddin Syed

Contents

The Battle of Mumbai Neighborhoods

For Next Chinese Restaurant

Introduction & Business Problem

Problem Background

Mumbai, formerly **Bombay**, city, capital of Maharashtra state, southwestern India. It is the country's financial and commercial center and its principal port on the Arabian Sea. Located on Maharashtra's coast, Mumbai is India's most-populous city, and it is one of the largest and most densely populated urban areas in the world.

Mumbai, long the center of India's cotton textile industry, subsequently developed a highly diversified manufacturing sector that included an increasingly important information technology (IT) component. In addition, the city's commercial and financial institutions are strong and vigorous, and Mumbai serves as the country's financial hub

Mumbai's business opportunities, as well as its potential to offer a higher standard of living, attract migrants from all over India, making the city a melting pot of many communities and cultures.

The market in Mumbai is highly competitive. It is highly developed city and hence cost of doing business is also one of the highest. Thus, any new business venture or expansion needs to be analyzed carefully. The insights derived from analysis will give good understanding of the business environment which will help in strategically targeting the market. This will help in reduction of risk and the Return on Investment will be reasonable.

Problem Description

Running a restaurant is more than just offering food in exchange for money. One is offering an experience to customers based on items such as the decor, food, and service. As the owner, one must determine the type of experience one wants to give customers. Even after one has decided on the type of restaurant, one needs to make sure there is a market for it, the concept is within budget, and find the right location. **In this report we will be using the data science way to find the best possible location in Mumbai for opening a Chinese restaurant.**

Keeping in line with Mumbai's cosmopolitan atmosphere the scene for dining and eating out in Mumbai is fabulously all-inclusive. The city has an extensive range of fine restaurants serving Indian, French, Italian, Chinese, Thai, Japanese, Lebanese, Arab and Mexican food. The best restaurants of Mumbai run the gamut

from top of the line elegant establishments with the finest selection of wines to family style restaurants to shacks and roadside vendors and stalls. Whatever may be your preference, whoever you are, whatever your wallet size, Mumbai is sure to adapt to your requirements. Thus, one can see that the restaurant market in Mumbai is very competitive and to open a new restaurant and survive in Mumbai it is important to strategically plan the whole process especially selection of location for the new restaurant. Various factors play an important role in the selection of a location for a new restaurant:

1. Visibility
   You may have found the cutest, quaintest location for a restaurant in a historic downtown somewhere, but if you're tucked in a side street with little foot traffic, unless you have amazing marketing tactics, you may have the most Instagram- and Pinterest-worthy space with zero customers:
   - Foot fall
   - Car traffic
2. Parking
   Whether you're opening fine dining or a popular chain, make sure the land you rent or buy is big enough to accommodate parking for your hungry clientele.
3. Space Size
4. Crime Rates
   Crime rates are unglamorous considerations, but if you place your restaurant in a crime-laden area, are your target customers going to visit? High crime rates can make potential customers uncomfortable, and if they feel they'll be mugged walking to their cars, it will only drive away business, no matter how legendary your coq au vin.
5. Surrounding Businesses and Competitor Analysis
   You'll want to do your research surrounding businesses. Are they doing well? Is the area affluent? Is there enough room for your business? Also, you'll want to know what types of restaurants do well in the area; however, you don't want to open a pizzeria if there are four in the area. Areas can only support so many of the same type of restaurant. What will distinguish any new restaurant is excellent service and consistently wonderful food.
6. Accessibility
   There's a reason that major restaurant chains are often located near highway exits: It makes them accessible for customers. Certain restaurants can get away with food or service that isn't the best simply because their locations are so accessible, like restaurants near the Eiffel Tower or Coliseum. There is plenty of foot traffic in urbanized areas, and restaurants only need to attract customers from the street into their business. Most successful restaurants—other than the truly elite—are easy to find, and you will find them in city centers or unique locations throughout the world.
7. Affordability
   Cost is always a bottom-line consideration for any business. If the rent or purchase of the space is more than you'll bring in each month in profits, that location is not feasible at that time. However, if you know that you'll generate business from that location, then you might consider it, but you'll need to be able to afford the upfront costs before you turn profits. Although some risks do pay off, you don't want to be at the point where you're struggling to cover basic costs. A killer location won't make up for driving your restaurant out of business.
8. Safety
9. Population Base

Are there enough people in the area to support your business? There need to be enough people who live in or pass through the area regularly to keep you busy

10. Style of operation
    Is your operation going to be formal and elegant? Or kicked-back and casual? Your location should be consistent with your particular style and image. If your business is retailing, do you want a traditional store, or would you like to try operating from a kiosk or booth in a mall or a cart that you can move to various locations?
11. Proximity to other businesses and services
    Take a look at what other businesses and services are in the vicinity from two key perspectives. First, see if you can benefit from nearby businesses--by the customer traffic they generate--because those companies and their employees could become your customers, or because it may be convenient and efficient for you to be their customer.
    Second, look at how they'll enrich the quality of your company as a workplace. Does the vicinity have an adequate selection of restaurants, so your employees have places to go for lunch? Is there a nearby day-care center for employees with children? Are other shops and services you and your employees might want conveniently located?
12. Utilities and other costs
    Rent composes the major portion of your ongoing facilities expense but consider extras such as utilities--they're included in some leases but not in others. If they're not included, ask the utility company for a summary of the previous year's usage and billing for the site. Also find out what kind of security deposits the various utility providers require so you can develop an accurate move-in budget; however, you may not need a deposit if you have an established payment record with the company.
    If you must provide your own janitorial service, what will it cost? What are insurance rates for the area? Do you have to pay extra for parking? Consider all your location-related expenses and factor them into your decision.

And the list can just go on….

Target Audience

The report would be beneficial to anybody who would like to open a Chinese restaurant in the city of Mumbai. The objective is to find and recommend a neighborhood in Mumbai where a new Chinese restaurant can be started.

Success Criteria

Based on Mumbai city data that is freely available the recommendation will be based on following factors that have been listed above:

1. Accessibility of the neighborhood.
2. Population Base of the neighborhood.

3. Number of restaurants (all inclusive) in the neighborhood.
4. Number of Chinese restaurants in the neighborhood.

Data

Mumbai Neighborhood Data

Mumbai city is divided into regions in various ways based on different administrative requirements like municipal wards, assembly constituencies, etc. However, our requirement for population base is met by the assembly constituency division population data for any other division is not available freely. The data for constituency, hereby referred as neighborhood, will be collected as follows:

1. Neighborhood List will be scrapped from https://en.wikipedia.org/wiki/List_of_constituencies_of_the_Maharashtra_Legislative_Assembly
2. Neighborhood latitude & longitude data will be obtained from the google API https://maps.googleapis.com/maps/api/geocode/json?address

   The header of data obtained is as in figure 1.

	Neighborhood	Latitude	Longitude	Above 18 Population
0	Borivali	19.230733	72.856673	327975.0
1	Dahisar	19.249734	72.859378	312006.0
2	Magathane	19.220206	72.866687	306369.0
3	Mulund	19.172021	72.956351	305878.0
4	Vikhroli	19.091182	72.920864	234880.0

Figure 1

Population Data Base

Total population for each neighborhood is not available freely. However, for each constituency total registered voters (age 18 years and above) is available and this can be used as a proxy for the total population. This data can be obtained from Wikipedia using the link https://en.wikipedia.org/wiki/<Constituency Name>_(Vidhan_Sabha_constituency) and using the 2014 election data. For eg. data for Mahim constituency can be scrapped from the link: https://en.wikipedia.org/wiki/Mahim_(Vidhan_Sabha_constituency). See Figure 1 above for the data.

Accessibility of the neighborhood

Accessibility of a neighborhood can be determined by the average of visit counts of places of interests in the neighborhood. This data can be obtained from **foursquare.com** using the API https://api.foursquare.com/v2/venues/ and using the **stats** data from the result. The header of data obtained is as shown in figure 2

	Neighborhood	Latitude	Longitude	Above 18 Population	Venue Tips Count	Venue User Counts	Venue Checkin Counts	Venue Visits Counts
0	Borivali	19.230733	72.856673	327975.0	8.672414	325.500000	468.241379	706.344828
1	Dahisar	19.249734	72.859378	312006.0	7.254545	347.763636	1044.527273	1464.818182
2	Magathane	19.220206	72.866687	306369.0	5.057143	305.600000	799.971429	1049.685714
3	Mulund	19.172021	72.956351	305878.0	7.758621	281.942529	511.735632	774.701149
4	Vikhroli	19.091182	72.920864	234880.0	12.047619	467.253968	940.809524	1306.793651

Figure 2

Total Number of Restaurants

The total number of restaurants can be obtained from **foursquare.com** using the API https://api.foursquare.com/v2/venues/explore? categoryId= 4d4b7105d754a06374d81259 . Total Chinese restaurants can be filtered from this data. The header of data obtained is as shown in figure 3.

	Neighborhood	Latitude	Longitude	Above 18 Population	All Restaurant Count	Chinese Restaurant Count
0	Borivali	19.230733	72.856673	327975.0	46	4
1	Dahisar	19.249734	72.859378	312006.0	41	3
2	Magathane	19.220206	72.866687	306369.0	22	0
3	Mulund	19.172021	72.956351	305878.0	70	7
4	Vikhroli	19.091182	72.920864	234880.0	57	3

Figure 3

Methodology

The methodology for the process is as follows:

1. Collect data from the sources mentioned in the data section
2. Load data into pandas dataframes for exploratory analysis
3. Do exploratory data analysis to get insights into data as to:
   a. Find category wise restaurant total to ascertain how are Chinese restaurants placed against other categories number wise
   b. Plot all restaurant total and Chinese restaurant total neighborhood wise, Chinese restaurant total and visit counts neighborhood wise. This will give an idea of probable location for opening a new Chinese restaurant
4. Aggregate and merge all the dataframes into one dataframe.
5. Use K-means clustering to get the best neighborhood group for opening Chinese restaurant.

Exploratory Data Analysis

Explore Mumbai's neighborhood on a map

Mumbai's neighborhood data is scrapped from the links mentioned in the data section and stored in dataframe as shown in figure 1. Let's plot this data on a map as follows:

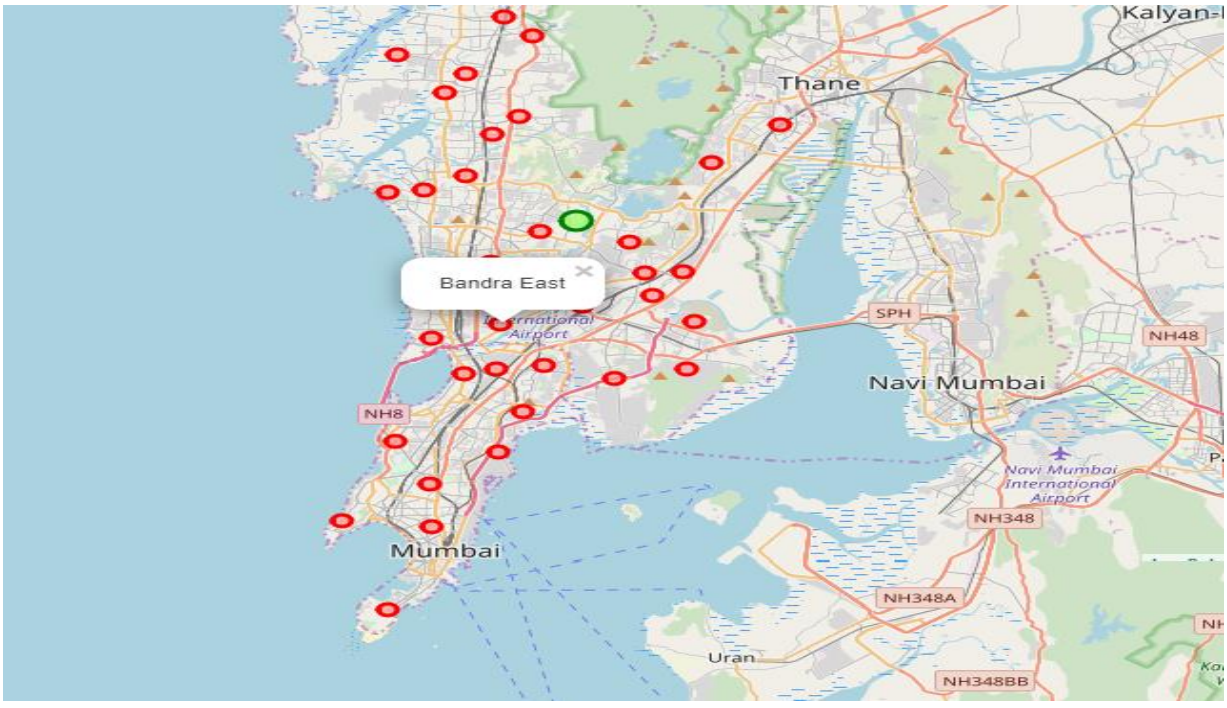

Figure 4: Mumbai Neighbourhood

Explore Restaurant Counts

In following figure one can see that in the case of authenticate cuisine restaurants, Chinese restaurants are most popular after Indian restaurants.

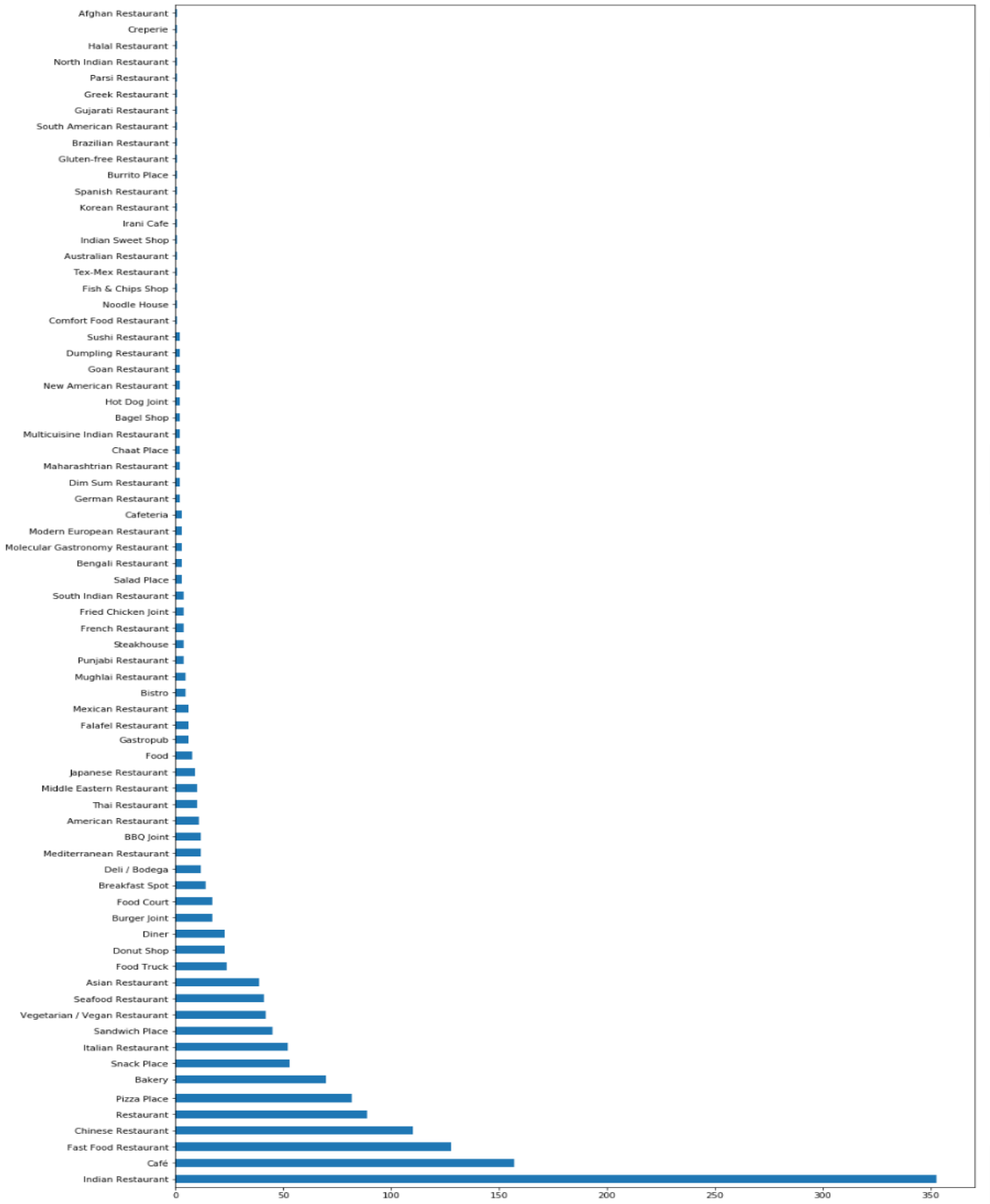

Figure 5: Restaurant Counts

Explore Total Restaurant Count v/s Total Chinese Restaurant Count

Let's explore Neighborhood wise Total Restaurant Count and Chinese Restaurant Count. Neighborhoods with less than 2 Chinese restaurant counts have been explicitly labeled. These neighborhoods could be candidates for opening next Chinese restaurants.

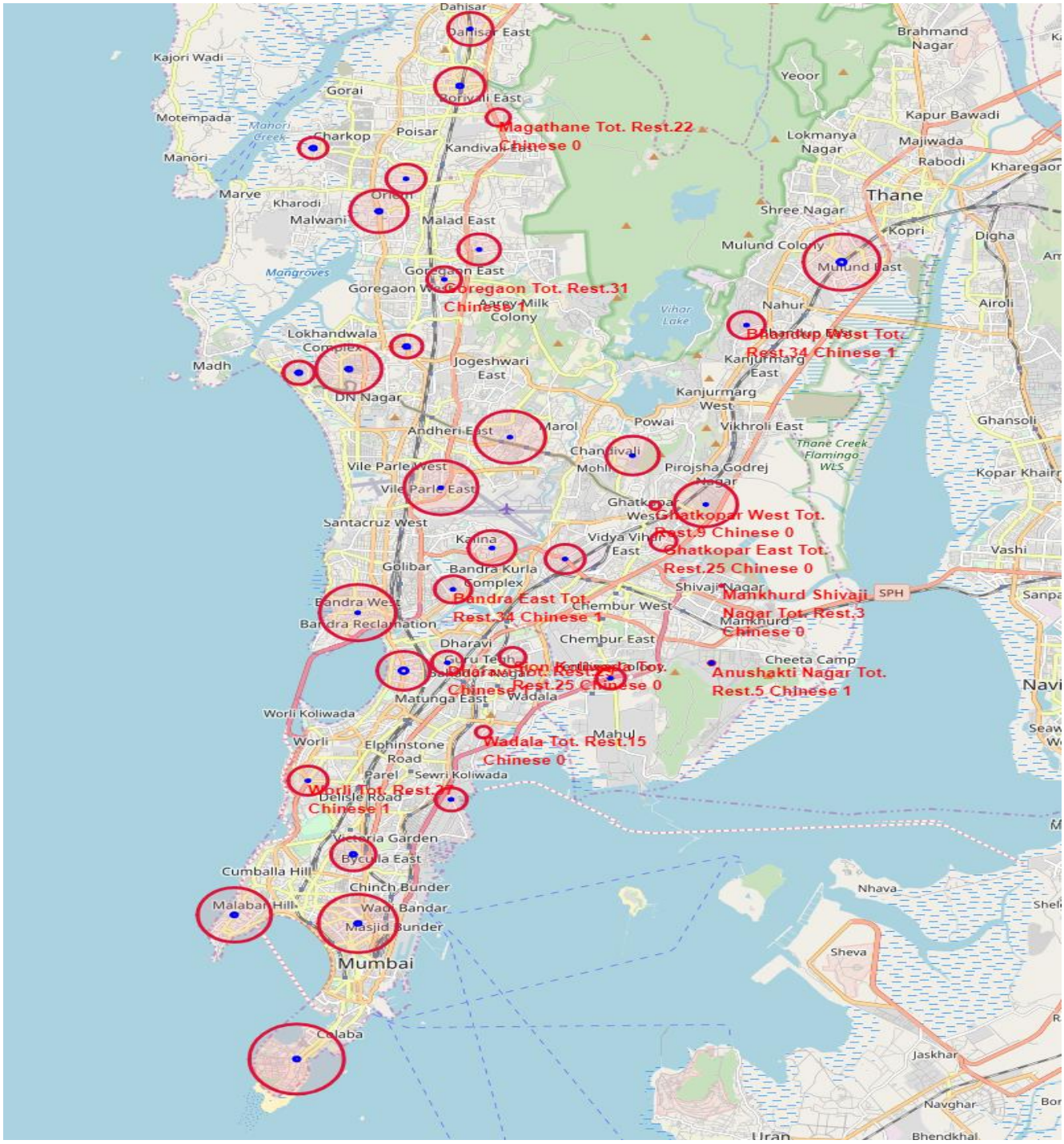

Figure 6: All Restaurant v/s Chinese Restaurant

From the map in figure 5, one can see that Worli (Tot. Rest.37 Chinese 1), Ghatkopar East (Tot. Rest.25 Chinese 0) and Goregaon (Tot. Rest.31 Chinese 1) seem to be good candidates for opening next Chinese restaurant as they have good amount of other kinds of restaurants indicating a good amount of eating-out crowd and have 0 or 1 Chinese restaurants.

Explore Chinese Restaurant Count v/s Average Visit Count

Figure 7: Chinese Restaurant Count v/s Average Visit Count

From the map in figure 7, one can see that again Worli (Avg. Visit Coun:3531.7, Chinese 1), Ghatkopar East (Avg. Visit Coun:1124.94 Chinese 0) and Goregaon (Avg. Visit Coun:1157.7 Chinese 1) seem to be good candidates for opening next Chinese restaurant as they have good visit counts indicating a good amount of crowd visiting places in the neighborhood and have 0 or 1 Chinese restaurants.

Clustering of Neighborhoods

Let's cluster the neighborhood data created to get the set of neighborhoods that would be ideal for opening the next Chinese restaurant. K-mean clustering algorithm will be used to cluster the neighborhoods. The final data after merging data sets described in the data section is as shown in the figure 8

	Neighborhood	Latitude	Longitude	Above 18 Population	All Restaurant Count	Chinese Restaurant Count	Venue Tips Count	Venue User Counts	Venue Checkin Counts	Venue Visits Counts
0	Borivali	19.230733	72.856673	327975.0	46	4	8.672414	325.500000	468.241379	706.344828
1	Dahisar	19.249734	72.859378	312006.0	41	3	7.254545	347.763636	1044.527273	1464.818182
2	Magathane	19.220206	72.866687	306369.0	22	0	5.057143	305.600000	799.971429	1049.685714
3	Mulund	19.172021	72.956351	305878.0	70	7	7.758621	281.942529	511.735632	774.701149
4	Vikhroli	19.091182	72.920864	234880.0	57	3	12.047619	467.253968	940.809524	1306.793651
5	Bhandup West	19.150988	72.931525	296610.0	34	1	5.617021	300.063830	634.276596	853.468085
6	Jogeshwari East	19.143897	72.842771	286826.0	30	5	20.458333	565.500000	982.250000	1382.916667
7	Dindoshi	19.176218	72.861698	294966.0	38	2	16.170213	732.659574	1526.489362	1906.085106
8	Kandivali East	19.199821	72.842594	270721.0	37	2	8.000000	380.209302	810.255814	1199.813953
9	Charkop	19.210045	72.818349	306447.0	26	4	5.224138	190.327586	298.896552	426.362069
10	Malad West	19.188954	72.835543	297683.0	54	6	16.000000	726.385965	1469.561404	1866.491228
11	Goregaon	19.166257	72.852570	337688.0	31	1	10.095238	406.190476	799.809524	1157.761905
12	Versova	19.135085	72.814610	209407.0	28	4	18.458333	686.416667	1022.291667	1520.875000
13	Andheri West	19.136325	72.827660	256600.0	61	6	18.927536	751.710145	1158.579710	1757.565217
14	Andheri East	19.113645	72.869734	273219.0	64	2	10.525000	645.225000	891.887500	1283.900000
15	Vile Parle	19.096793	72.851695	270735.0	67	3	30.875000	1975.277778	3504.541667	5360.555556
16	Chandivali	19.107491	72.901760	417734.0	49	3	23.920000	869.180000	1315.220000	1880.780000
17	Ghatkopar West	19.090806	72.907667	285276.0	9	0	6.000000	201.166667	276.666667	414.000000
18	Ghatkopar East	19.078599	72.910146	249807.0	25	0	11.750000	429.875000	723.281250	1124.937500
19	Mankhurd Shivaji Nagar	19.063745	72.925386	228358.0	3	0	3.000000	370.500000	549.500000	677.000000
20	Anushakti Nagar	19.037921	72.922757	241587.0	5	1	4.416667	211.416667	336.250000	521.500000
21	Chembur	19.032800	72.896357	276701.0	27	2	8.418605	303.302326	684.069767	964.488372
22	Kurla	19.072630	72.884472	239207.0	36	3	11.837838	629.027027	1181.864865	1661.756757
23	Kalina	19.076277	72.865400	226508.0	44	3	14.500000	935.777778	1287.111111	1926.888889
24	Bandra East	19.062493	72.855196	226508.0	34	1	10.794118	561.794118	722.147059	1135.882353
25	Bandra West	19.054772	72.830338	226508.0	70	3	33.022222	1434.911111	2115.122222	3147.200000
26	Dharavi	19.038033	72.853759	211000.0	29	1	26.518519	589.888889	866.888889	1334.555556
27	Sion Koliwada	19.039877	72.870748	251322.0	25	0	15.040000	491.720000	752.720000	1104.440000
28	Wadala	19.014869	72.863143	194776.0	15	0	6.086957	254.130435	407.173913	694.086957
29	Mahim	19.035385	72.842304	230050.0	48	7	19.039216	571.666667	943.784314	1369.098039
30	Worli	18.998641	72.817360	262507.0	37	1	26.672414	1490.000000	2656.724138	3531.775862
31	Shivadi	18.992413	72.854716	270308.0	28	2	7.396552	417.189655	639.465517	976.603448
32	Byculla	18.975024	72.829518	224686.0	42	4	23.300000	1054.533333	1493.233333	2274.233333
33	Malabar Hill	18.954798	72.798452	260510.0	68	4	10.291139	531.607595	787.291139	1236.594937
34	Mumbadevi	18.951971	72.830734	233527.0	73	5	30.300000	1458.175000	2277.312500	3309.112500
35	Colaba	18.906703	72.814712	248586.0	87	6	26.455556	1294.500000	1678.955556	2335.044444

Figure 8: Final Data

Determine Number of Clusters

The number of clusters is determined by running the k-means clustering algorithm with k ranging from 1 to 11 and choosing the optimum value using the metrics: Silhouette Coefficient, Calinski-Harabaz Index and Davies-Bouldin Index. These metrics are used to evaluate a model where ground truth labels are not available as in our case since, we don't have neighborhoods that are pre labeled with rank of choice for opening next Chinese restaurant. For a detail discussion of the aforementioned metrics please refer to https://scikit-learn.org/stable/modules/clustering.html#clustering-performance-evaluation.

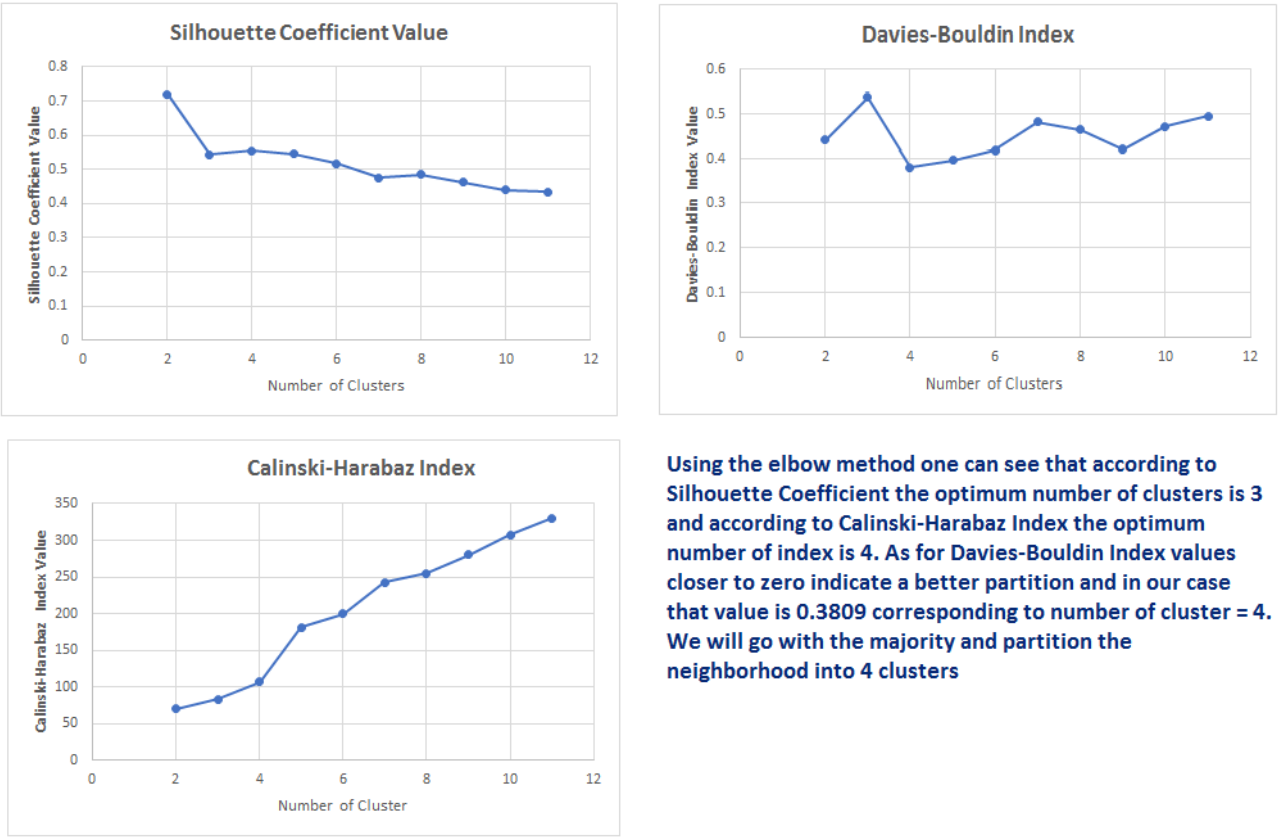

Figure 9: K-Means Cluster Evaluation To Determine Optimal Number of Clusters Value

From figure 9, using the elbow method one can see that according to Silhouette Coefficient the optimum number of clusters is 3 and according to Calinski-Harabaz Index the optimum number of index is 4. As for Davies-Bouldin Index values closer to zero indicate a better partition and in our case that value is 0.3809 corresponding to number of clusters = 4. We will go with the majority and partition the neighborhood into 4 clusters

Results

The Clusters of Neighborhood

Using Number of clusters = 4, the neighborhoods were partitioned into 4 clusters using the K-Means clustering algorithm. The 4 partitions are as follows:

Cluster 1: Neighborhoods with more than 2 Chinese restaurants and moderate count values.

	Neighborhood	Latitude	Longitude	Above 18 Population	All Restaurant Count	Chinese Restaurant Count	Venue Tips Count	Venue User Counts	Venue Checkin Counts	Venue Visits Counts	Cluster Labels
7	Dindoshi	19.176218	72.861698	294966.0	38	2	16.170213	732.659574	1526.489362	1906.085106	0
10	Malad West	19.188954	72.835543	297683.0	54	6	16.000000	726.385965	1469.561404	1866.491228	0
12	Versova	19.135085	72.814610	209407.0	28	4	18.458333	686.416667	1022.291667	1520.875000	0
13	Andheri West	19.136325	72.827660	256600.0	61	6	18.927536	751.710145	1158.579710	1757.565217	0
16	Chandivali	19.107491	72.901760	417734.0	49	3	23.920000	869.180000	1315.220000	1880.780000	0
22	Kurla	19.072630	72.884472	239207.0	36	3	11.837838	629.027027	1181.864865	1661.756757	0
23	Kalina	19.076277	72.865400	226508.0	44	3	14.500000	935.777778	1287.111111	1926.888889	0
32	Byculla	18.975024	72.829518	224686.0	42	4	23.300000	1054.533333	1493.233333	2274.233333	0
35	Colaba	18.906703	72.814712	248586.0	87	6	26.455556	1294.500000	1678.955556	2335.044444	0

Cluster 2: All neighborhoods with no Chinese restaurants and moderate count values (on an average)

	Neighborhood	Latitude	Longitude	Above 18 Population	All Restaurant Count	Chinese Restaurant Count	Venue Tips Count	Venue User Counts	Venue Checkin Counts	Venue Visits Counts	Cluster Labels
0	Borivali	19.230733	72.856673	327975.0	46	4	8.672414	325.500000	468.241379	706.344828	1
1	Dahisar	19.249734	72.859378	312006.0	41	3	7.254545	347.763636	1044.527273	1464.818182	1
2	Magathane	19.220206	72.866687	306369.0	22	0	5.057143	305.600000	799.971429	1049.685714	1
3	Mulund	19.172021	72.956351	305878.0	70	7	7.758621	281.942529	511.735632	774.701149	1
4	Vikhroli	19.091182	72.920864	234880.0	57	3	12.047619	467.253968	940.809524	1306.793651	1
5	Bhandup West	19.150988	72.931525	296610.0	34	1	5.617021	300.063830	634.276596	853.468085	1
6	Jogeshwari East	19.143897	72.842771	286826.0	30	5	20.458333	565.500000	982.250000	1382.916667	1
8	Kandivali East	19.199821	72.842594	270721.0	37	2	8.000000	380.209302	810.255814	1199.813953	1
9	Charkop	19.210045	72.818349	306447.0	26	4	5.224138	190.327586	298.896552	426.362069	1
11	Goregaon	19.166257	72.852570	337688.0	31	1	10.095238	406.190476	799.809524	1157.761905	1
14	Andheri East	19.113645	72.869734	273219.0	64	2	10.525000	645.225000	891.887500	1283.900000	1
17	Ghatkopar West	19.090806	72.907667	285276.0	9	0	6.000000	201.166667	276.666667	414.000000	1
18	Ghatkopar East	19.078599	72.910146	249807.0	25	0	11.750000	429.875000	723.281250	1124.937500	1
19	Mankhurd Shivaji Nagar	19.063745	72.925386	228358.0	3	0	3.000000	370.500000	549.500000	677.000000	1
20	Anushakti Nagar	19.037921	72.922757	241587.0	5	1	4.416667	211.416667	336.250000	521.500000	1
21	Chembur	19.032800	72.896357	276701.0	27	2	8.418605	303.302326	684.069767	964.488372	1
24	Bandra East	19.062493	72.855196	226508.0	34	1	10.794118	561.794118	722.147059	1135.882353	1
26	Dharavi	19.038033	72.853759	211000.0	29	1	26.518519	589.888889	866.888889	1334.555556	1
27	Sion Koliwada	19.039877	72.870748	251322.0	25	0	15.040000	491.720000	752.720000	1104.440000	1
28	Wadala	19.014869	72.863143	194776.0	15	0	6.086957	254.130435	407.173913	694.086957	1
29	Mahim	19.035385	72.842304	230050.0	48	7	19.039216	571.666667	943.784314	1369.098039	1
31	Shivadi	18.992413	72.854716	270308.0	28	2	7.396552	417.189655	639.465517	976.603448	1
33	Malabar Hill	18.954798	72.798452	260510.0	68	4	10.291139	531.607595	787.291139	1236.594937	1

Cluster 3: Neighborhood with more than 1 Chinese restaurant and highest average counts.

	Neighborhood	Latitude	Longitude	Above 18 Population	All Restaurant Count	Chinese Restaurant Count	Venue Tips Count	Venue User Counts	Venue Checkin Counts	Venue Visits Counts	Cluster Labels
15	Vile Parle	19.096793	72.851695	270735.0	67	3	30.875	1975.277778	3504.541667	5360.555556	2

Cluster 4: Neighborhoods with high average counts, also containing a neighborhood with 1 Chinese restaurant

	Neighborhood	Latitude	Longitude	Above 18 Population	All Restaurant Count	Chinese Restaurant Count	Venue Tips Count	Venue User Counts	Venue Checkin Counts	Venue Visits Counts	Cluster Labels
25	Bandra West	19.054772	72.830338	226508.0	70	3	33.022222	1434.911111	2115.122222	3147.200000	3
30	Worli	18.998641	72.817360	262507.0	37	1	26.672414	1490.000000	2656.724138	3531.775862	3
34	Mumbadevi	18.951971	72.830734	233527.0	73	5	30.300000	1458.175000	2277.312500	3309.112500	3

Discussion

Data Exploration and Clustering both point to similar results for the next neighborhood for opening a new Chinese restaurant. These being

1. Ghatkopar East with no Chinese restaurant and good average counts (11.75 tip counts, 429.87 user counts, 723.28 check-in counts, 1124.93 visit count)
2. Goregaon with just 1 Chinese restaurant and good average counts (10.09 tip counts, 406.19 user counts, 799.80 check-in counts, 1157.76 visit count)
3. Worli, though not part of the same cluster as above two, but with only one Chinese restaurant and high average counts (26.67 tip counts, 1490.00 user counts, 2656.72 check-in counts, 3531.77 visit count) is also a good candidate.

As can be seen from figure 5 the Mumbai city has scope not only for Chinese but other authentic cuisine restaurants. An approach like this report can be taken for finding location for other authentic cuisine restaurants.

Conclusion

Mumbai neighborhood data has been collected from Wikipedia and foursquare.com, analyzed and location for Chinese restaurant has been recommended using K-means clustering. Though the data is limited (eg. population data is about 18+ years only) the results are still good and usable.

Better data for parameters used for analysis in this report and data for other parameters mentioned in problem section would yield better results.

Mumbai city has good eating out crowd and hence has scope for other authentic cuisine restaurants. Approach like this report can be used for predicting other cuisine restaurants.