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	19.249436	72.859628	312006.0
2	Magathane	19.220206	72.866687	306369.0
3	Mulund	19.172554	72.942537	305878.0
4	Vikhroli	19.092161	72.918940	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	325.758621	468.465517	706.810345
1	Dahisar	19.249436	72.859628	312006.0	6.727273	258.945455	591.236364	793.600000
2	Magathane	19.220206	72.866687	306369.0	5.000000	300.085714	783.828571	1037.314286
3	Mulund	19.172554	72.942537	305878.0	6.903226	298.741935	580.758065	864.709677
4	Vikhroli	19.092161	72.918940	234880.0	14.679245	542.509434	963.301887	1260.132075

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	59	4
1	Dahisar	19.249436	72.859628	312006.0	43	5
2	Magathane	19.220206	72.866687	306369.0	30	0
3	Mulund	19.172554	72.942537	305878.0	45	4
4	Vikhroli	19.092161	72.918940	234880.0	59	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:

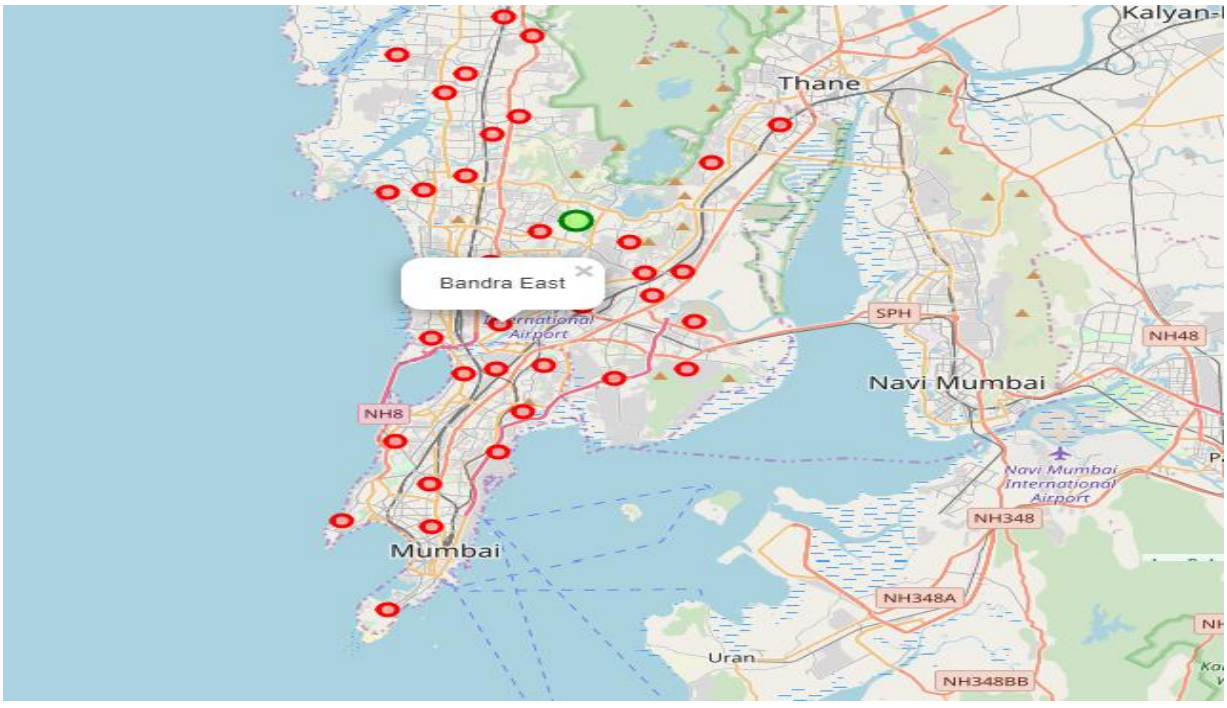

Figure 4: Mumbai Neighbourhood

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.

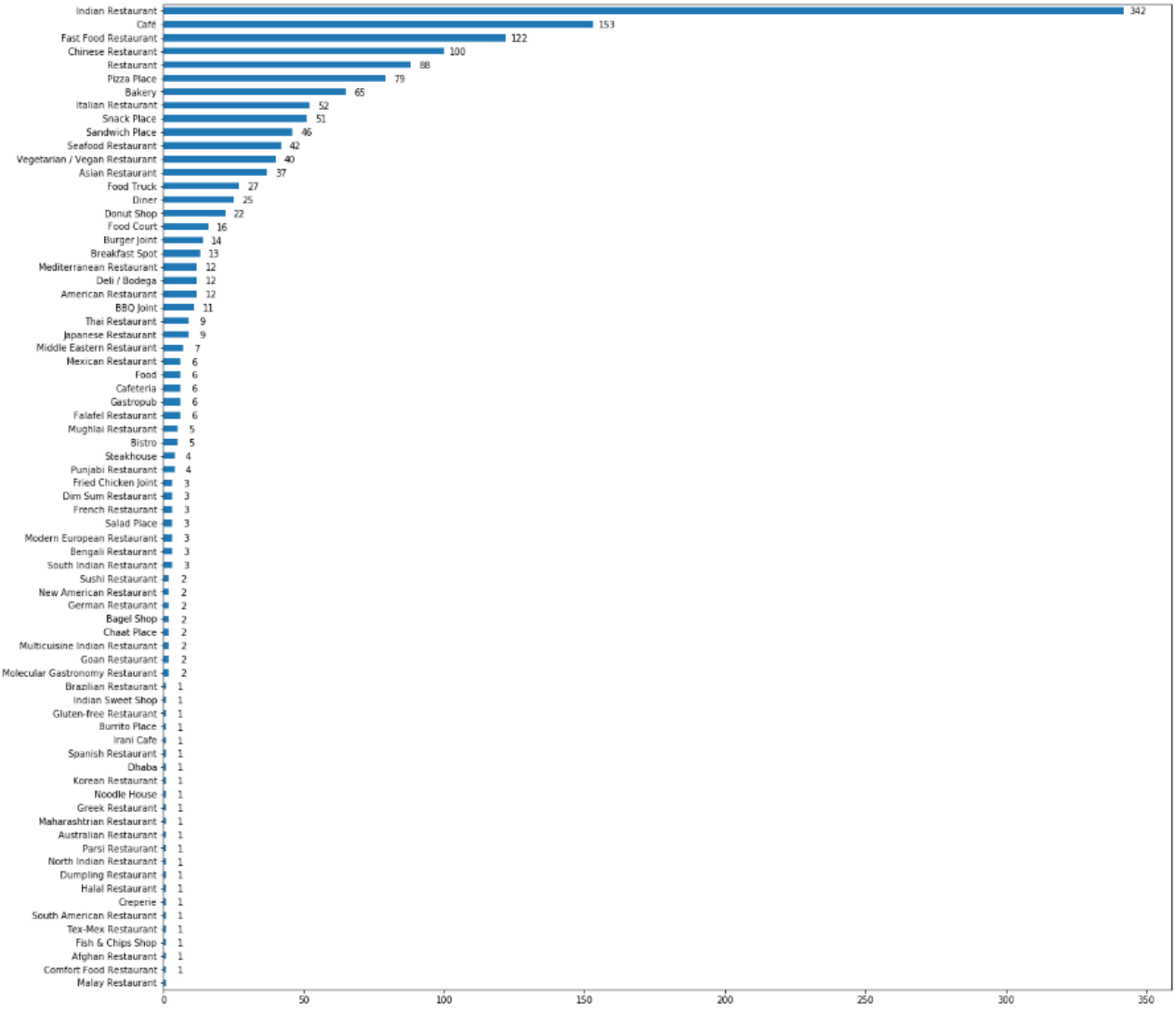

Figure 5: Restaurant Counts

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.

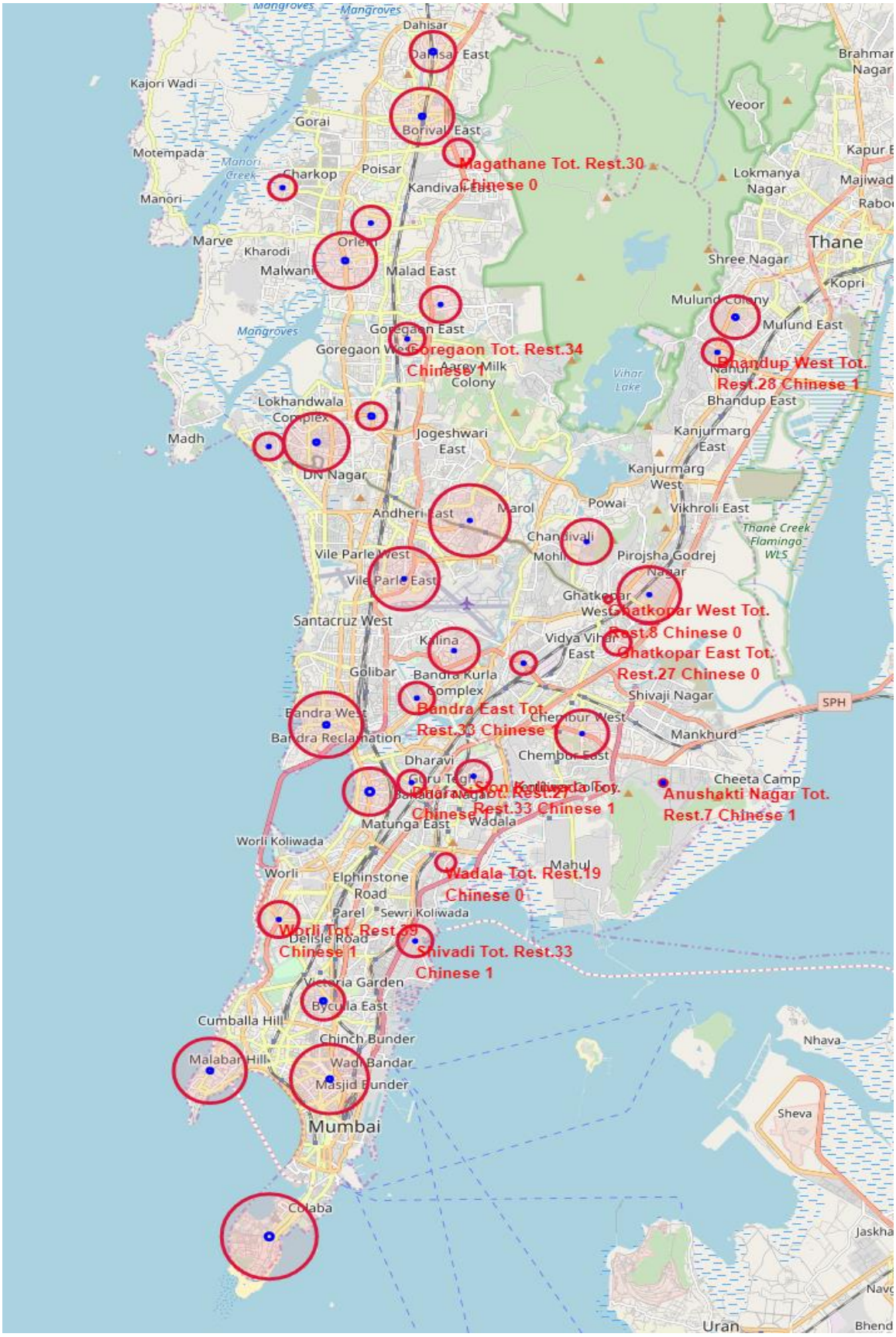

Figure 6: All Restaurant v/s Chinese Restaurant

From the map in figure 5, one can see that Worli (Tot. Rest.39 Chinese 1), Ghatkopar East (Tot. Rest.27 Chinese 0) and Bandra East (Tot. Rest.33 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

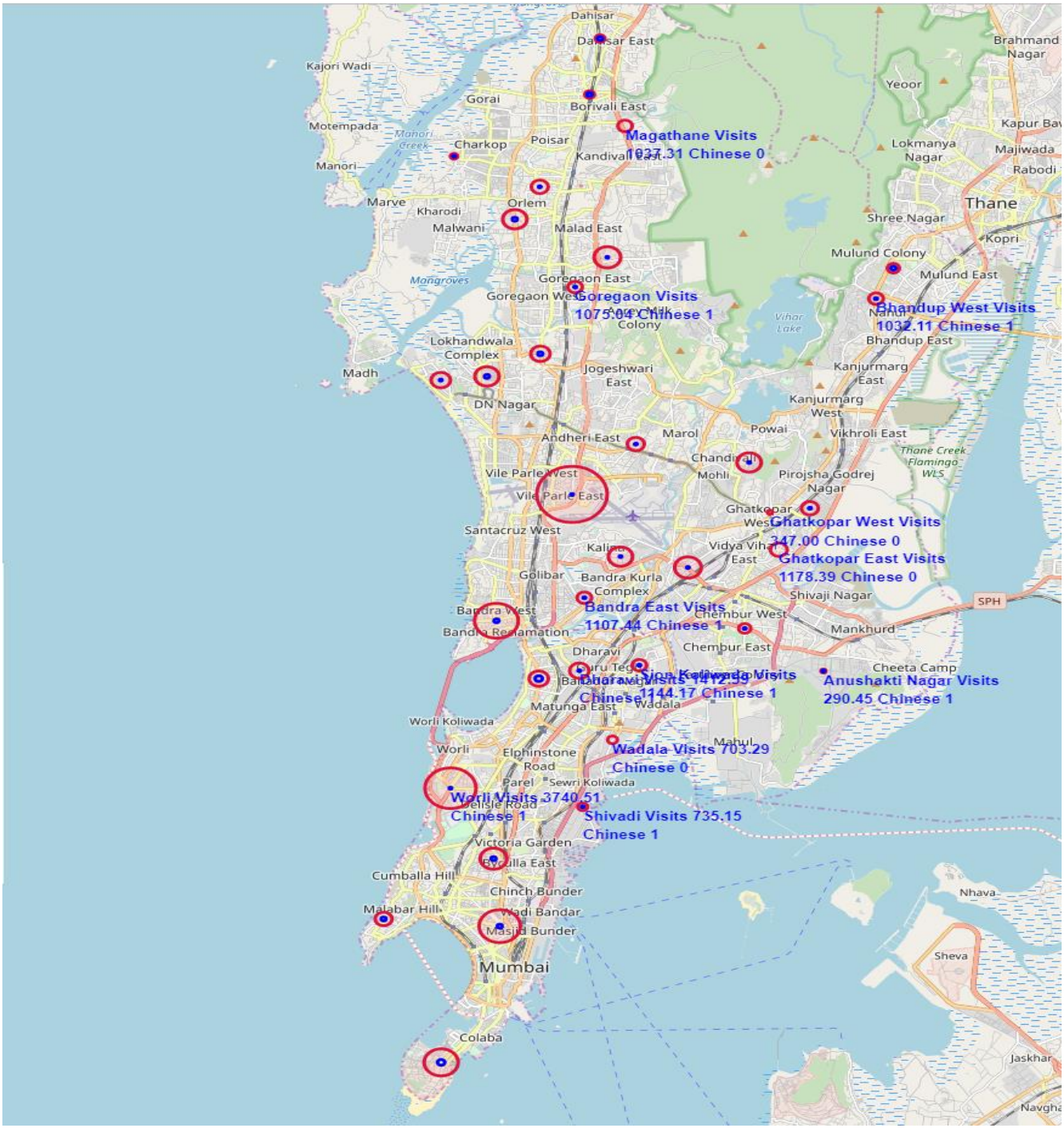

Figure 7: Chinese Restaurant Count v/s Average Visit Count

From the map in figure 7, one can see that again Worli (Avg. Visit Coun: 3740.50, Chinese 1), Ghatkopar East (Avg. Visit Coun: 1178.39 Chinese 0) and Bandra East (Avg. Visit Count: 1107.44 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	59	4	8.672414	325.758621	468.465517	706.810345
1	Dahisar	19.249436	72.859628	312006.0	43	5	6.727273	258.945455	591.236364	793.600000
3	Mulund	19.172554	72.942537	305878.0	45	4	6.903226	298.741935	580.758065	864.709677
4	Vikhroli	19.092161	72.918940	234880.0	59	3	14.679245	542.509434	963.301887	1260.132075
5	Bhandup West	19.162370	72.937632	296610.0	28	1	8.083333	380.222222	799.222222	1032.111111
6	Jogeshwari East	19.143897	72.842771	286826.0	30	6	21.043478	575.608696	1015.956522	1410.217391
7	Dindoshi	19.176218	72.861698	294966.0	38	2	16.600000	757.911111	1584.066667	1977.711111
8	Kandivali East	19.199821	72.842594	270721.0	37	2	7.863636	384.250000	808.250000	1198.568182
9	Charkop	19.210045	72.818349	306447.0	26	3	6.148148	206.851852	333.444444	463.888889
10	Malad West	19.188954	72.835543	297683.0	60	5	15.250000	694.816667	1402.466667	1778.066667
11	Goregaon	19.166257	72.852570	337688.0	34	1	9.695652	380.913043	748.173913	1075.043478
12	Versova	19.135085	72.814610	209407.0	28	3	17.916667	684.750000	1020.458333	1518.791667
13	Andheri West	19.136325	72.827660	256600.0	62	6	18.558824	744.941176	1144.588235	1745.720588
14	Andheri East	19.113645	72.869734	273219.0	76	3	10.679487	640.397436	886.769231	1273.179487
15	Vile Parle	19.096793	72.851695	270735.0	67	3	30.486111	1926.125000	3437.472222	5264.041667
16	Chandivali	19.107491	72.901760	417734.0	49	3	23.078431	827.156863	1255.333333	1813.960784
19	Anushakti Nagar	19.037921	72.922757	241587.0	7	1	3.090909	120.000000	230.000000	290.454545
20	Chembur	19.052211	72.900522	276701.0	50	3	7.752941	293.647059	583.023529	852.400000
21	Kurla	19.072630	72.884472	239207.0	23	3	12.791667	785.208333	1475.791667	2042.375000
22	Kalina	19.076277	72.865400	226508.0	48	3	13.947368	918.368421	1252.947368	1872.500000
23	Bandra East	19.062493	72.855196	226508.0	33	1	10.583333	550.500000	702.416667	1107.444444
24	Bandra West	19.054772	72.830338	226508.0	70	4	34.149425	1465.275862	2153.091954	3225.632184
25	Dharavi	19.038033	72.853759	211000.0	27	1	25.448276	618.068966	950.689655	1412.586207
26	Sion Koliwada	19.039877	72.870748	251322.0	33	1	15.458333	508.208333	781.916667	1144.166667
28	Mahim	19.035385	72.842304	230050.0	50	7	17.729167	551.770833	939.020833	1364.291667
29	Worli	18.998641	72.817360	262507.0	39	1	27.163636	1570.000000	2803.836364	3740.509091
30	Shivadi	18.992413	72.854716	270308.0	33	1	6.169811	331.962264	503.358491	735.150943
31	Byculla	18.975024	72.829518	224686.0	41	4	22.806452	970.612903	1381.096774	2054.290323
32	Malabar Hill	18.954798	72.798452	260510.0	69	4	10.961538	570.948718	836.346154	1329.051282
33	Mumbadevi	18.952389	72.831268	233527.0	74	5	28.837500	1340.975000	2078.412500	3023.150000
34	Colaba	18.906703	72.814712	248586.0	92	7	28.844444	1378.622222	1788.000000	2472.033333

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.

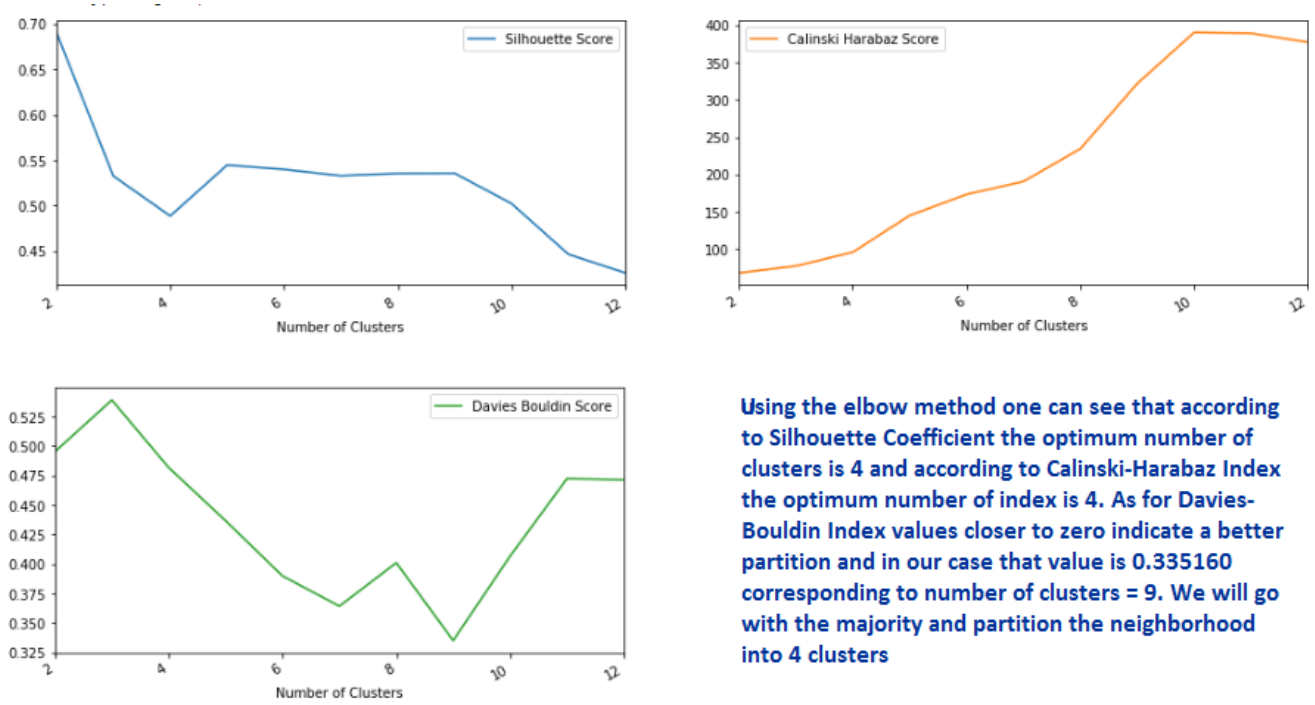

Figure 9: K-Means Cluster Evaluation To Determine Optimal Number of Clusters Value

From figure 9, using the elbow method one can see that according to Silhouette Coefficient the optimum number of clusters is 4 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 0.335160 corresponding to number of clusters = 9. 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: This cluster contains all neighborhoods with high average count values signifying greater accessibility potential. It contains a neighborhood with high accessibility (average count) and only 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
24	Bandra West	19.054772	72.830338	226508.0	70	4	34.149425	1465.275862	2153.091954	3225.632184	0
29	Worli	18.998641	72.817360	262507.0	39	1	27.163636	1570.000000	2803.836364	3740.509091	0
33	Mumbadevi	18.952389	72.831268	233527.0	74	5	28.837500	1340.975000	2078.412500	3023.150000	0
34	Colaba	18.906703	72.814712	248586.0	92	7	28.844444	1378.622222	1788.000000	2472.033333	0

Cluster 2: This cluster contains all neighborhoods with one or more than 1 Chinese restaurants and moderate count values. Not a candidate cluster

	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
4	Vikhroli	19.092161	72.918940	234880.0	59	3	14.679245	542.509434	963.301887	1260.132075	1
6	Jogeshwari East	19.143897	72.842771	286826.0	30	6	21.043478	575.608696	1015.956522	1410.217391	1
7	Dindoshi	19.176218	72.861698	294966.0	38	2	16.600000	757.911111	1584.066667	1977.711111	1
10	Malad West	19.188954	72.835543	297683.0	60	5	15.250000	694.816667	1402.466667	1778.066667	1
12	Versova	19.135085	72.814610	209407.0	28	3	17.916667	684.750000	1020.458333	1518.791667	1
13	Andheri West	19.136325	72.827660	256600.0	62	6	18.558824	744.941176	1144.588235	1745.720588	1
14	Andheri East	19.113645	72.869734	273219.0	76	3	10.679487	640.397436	886.769231	1273.179487	1
16	Chandivali	19.107491	72.901760	417734.0	49	3	23.078431	827.156863	1255.333333	1813.960784	1
21	Kurla	19.072630	72.884472	239207.0	23	3	12.791667	785.208333	1475.791667	2042.375000	1
22	Kalina	19.076277	72.865400	226508.0	48	3	13.947368	918.368421	1252.947368	1872.500000	1
25	Dharavi	19.038033	72.853759	211000.0	27	1	25.448276	618.068966	950.689655	1412.586207	1
28	Mahim	19.035385	72.842304	230050.0	50	7	17.729167	551.770833	939.020833	1364.291667	1
31	Byculla	18.975024	72.829518	224686.0	41	4	22.806452	970.612903	1381.096774	2054.290323	1
32	Malabar Hill	18.954798	72.798452	260510.0	69	4	10.961538	570.948718	836.346154	1329.051282	1

Cluster 3: This cluster contains all neighborhoods with no Chinese restaurants and moderate average count values. This cluster is a possible candidate for picking the neighborhood for opening a 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
1	Dahisar	19.249436	72.859628	312006.0	43	5	6.727273	258.945455	591.236364	793.600000	2
2	Magathane	19.220206	72.866687	306369.0	30	0	5.000000	300.085714	783.828571	1037.314286	2
3	Mulund	19.172554	72.942537	305878.0	45	4	6.903226	298.741935	580.758065	864.709677	2
5	Bhandup West	19.162370	72.937632	296610.0	28	1	8.083333	380.222222	799.222222	1032.111111	2
8	Kandivali East	19.199821	72.842594	270721.0	37	2	7.863636	384.250000	808.250000	1198.568182	2
9	Charkop	19.210045	72.818349	306447.0	26	3	6.148148	206.851852	333.444444	463.888889	2
11	Goregaon	19.166257	72.852570	337688.0	34	1	9.695652	380.913043	748.173913	1075.043478	2
17	Ghatkopar West	19.090806	72.907667	285276.0	8	0	5.600000	171.600000	282.800000	347.000000	2
18	Ghatkopar East	19.078599	72.910146	249807.0	27	0	12.535714	462.214286	766.928571	1178.392857	2
19	Anushakti Nagar	19.037921	72.922757	241587.0	7	1	3.090909	120.000000	230.000000	290.454545	2
20	Chembur	19.052211	72.900522	276701.0	50	3	7.752941	293.647059	583.023529	852.400000	2
23	Bandra East	19.062493	72.855196	226508.0	33	1	10.583333	550.500000	702.416667	1107.444444	2
26	Sion Koliwada	19.039877	72.870748	251322.0	33	1	15.458333	508.208333	781.916667	1144.166667	2
27	Wadala	19.014869	72.863143	194776.0	19	0	6.208333	258.416667	429.625000	703.291667	2
30	Shivadi	18.992413	72.854716	270308.0	33	1	6.169811	331.962264	503.358491	735.150943	2

Cluster 4: This cluster contains all neighborhood with 3 Chinese restaurant and highest average counts. Not a candidate cluster

	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	30.486111	1926.125	3437.472222	5264.041667	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 (12.53 tip counts, 462.21 user counts, 766.89 check-in counts, 1178.35 visit count)
2. Bandra East with just 1 Chinese restaurant and good average counts (10.58 tip counts, 550.30 user counts, 702.05 check-in counts, 1107.08 visit count)

3. Worli, though not part of the same cluster as above two, but with only one Chinese restaurant and high average counts (27.14 tip counts, 1569.89 user counts, 2803.56 check-in counts, 3740.23 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. Three Potential neighborhoods namely **Ghatkopar East, Bandra East** and **Worli** were identified for opening a new Chinese restaurant.

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.