Analysing best location for opening a restaurant in Mumbai, India

Capstone Project
IBM Applied Data Science Certificate Program

Project Report

Kumar Lokesh Singh

April 2020

Introduction

Mumbai (formerly called Bombay) is a densely populated capital city of state Maharashtra on India's west coast. It is the country's financial and commercial centre and its principal port on the Arabian Sea. Mumbai is India's most-populous city, and it is one of the largest and most densely populated urban areas in the world with an estimated city proper population of 12.5 million living under Municipal Corporation of Greater Mumbai. The metropolis has a roadside fast food consisting of Maharashtrian Pav Bhaji, Vada pavs, Dabeli, Panipuri, Bhelpuri, etc. South Indian and Chinese food is also popular in the city. Lebanese, Korean, Thai, Italian, Mexican, Mughlai, Punjabi, Mālvani, and Continental cuisine are also popular in Mumbai.

Mumbai has some of the oldest restaurants in India. Delhi Darbar, Sindhudurg, Highway Gomantak, Samrat, Vitthal Bhelwala, Mahesh Lunch Home, Kailas Parbat, and Adarsh are some of the oldest restaurants in the city.

Mumbai, being the financial capital, has a large migrant population. Members of the migrating families are working with no other family support, hence ordering food from close by restaurants or calling for pre-packaged meals (dabba as called locally) is an accepted thing. These 'Dabbas' are the usually delivered by a special local delivery organization known as the Dabbawallas. Owing to these recent trends, several online food-ordering services have cropped up in the city.

Mumbai has the largest of the organised food service markets pegged at nearly Rs.41, 000 crore. Mumbai has the highest share in the organised market -which mainly accounts for standalone restaurants. Interestingly, in Mumbai, among international cuisines, 33% prefer Italian compared to 29% who like Chinese. People in the island city also prefers south Indian food to north Indian cuisines.

Business Problem

As Mumbai and its sub-urban area is growing, the number of restaurants is increasing day by day. Currently which stands at approximately 13,790 restaurants. With such a high number of restaurants, this industry is yet to saturate. In addition, new restaurants are opening every day. However, it has

become difficult for them to compete with already established restaurants. The key issues that continue to pose a challenge to them include high real estate costs, rising food costs, shortage of quality work force, fragmented supply chain and over-licensing.

Purpose of study

As Mumbai and its sub-urban area is growing, the number of restaurants is increasing day by day. Currently which stands at approximately 13,790 restaurants. The key issues that continue to pose a challenge to them include high real estate costs, rising food costs, shortage of quality work force, fragmented supply chain and over-licensing.

With each day new restaurants, opening the industry is yet to saturate and the demand is increasing regularly. Despite increasing demand, it however has become difficult for new restaurants to compete with established restaurants. Most of them serving similar food and most of the people are dependent on the restaurant food. With such an overwhelming demand of restaurants, it has therefore become important to study the demography of a location. What kind of a food is more popular in a locality? Do the entire locality loves vegetarian food etc.? This kind of analysis can be done using the data, by studying the factors such as

- Approx. price of food
- Location of the restaurant
- Theme based restaurant or not?
- Which locality of that city serves those cuisines with maximum number of restaurants?
- Customers who are striving to get the best cuisine of the neighbourhood
- Is a neighbourhood famous for its own kind of food?

Data acquisition and cleaning

The main data used for this project were from two sources:

- The restaurants data from the Zomato website was obtained from Kaggle website.
- Explore trending venues in a neighbourhood particularly restaurant (Foursquare API).

Other supporting data:

- Coordinates (Geocoder Python)
- https://foursquare.com/explore?mode=url&near=Mumbai%2C%20Mah%C4%81r%C4%81shtra%2C%20India&nearGeoId=72057594039203275&q=Restaurant

Data cleaning - The dataset contains 16 variables all of which were obtained from Kaggle website. The dataset contains details of more than 6000 restaurants in Mumbai in each of its neighbourhood.
Dataset was containing some unwanted columns, which were dropped. All the rows containing null & duplicate data was removed. One of the major challenge was to obtain the geographical coordinates of the neighbourhoods that was necessary in order to generate maps. Hence, a new data frame was created and two more columns for restaurants location latitude and location longitude was obtained.
All columns of the dataset was checked for null and blank values. Columns were further analysed to check data types, arrays and unique values. Requisite conversion of datatypes were carried out as per requirement for analysis. As you can refer to below table for 16 variables with description which was used in this project for data analysis.

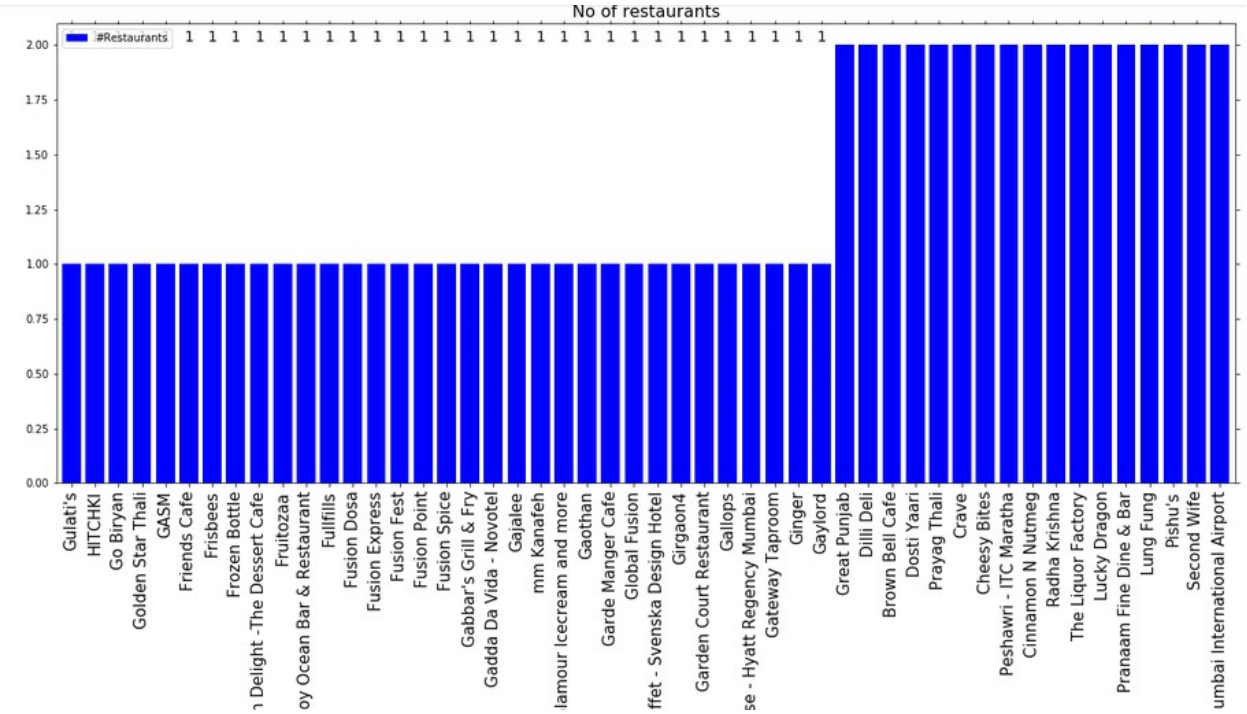
Variable	Type	Description
Cost_for_two	int	contains the approximate cost for meal for two people
Features	object	What are the exclusive features of restaurant and features for which it is known?
Home_Delivery	bool	whether home delivery is available in the restaurant or not

Operational_hours	object	operational timing
Restaurant_Location	object	contains the neighbourhood name in which the restaurant is located in Mumbai
Restaurant_Name	object	contains the name of the restaurant
Restaurant_Type	object	restaurant type
View_Menu	bool	contains the data whether restaurant menu can be viewed
Rating	float	contains total number of rating for the restaurant
Votes	int	User votes given to restaurant
Rating_Category	object	What is the category of rating ranging from excellent, good or fair?
Operational_after_Midnight	bool	Whether restaurant is operational after mid-night or not?
Cuisines	object	Food types separated by commas
Cuisine_count	int	Number of cuisines of a restaurant
Competitors_in_Location	float	Contains number of competitors in that location
Score	float	Overall score of the restaurant

	Cost_for_two	Cuisines	Features	Home_Delivery	Operational_hours	Restaurant_Location	Restaurant_Name	Restaurant_Type	View_Menu
0	1500	Finger Food, Continental, European, Italian	Food Hygiene Rated Restaurants In Mumbai, Best...	False	12noon – 1am (Mon-Sun)	Kamala Mills Compound	Lord of the Drinks	Lounge,Casual Dining	True
1	800	Pizza	Value For Money, Best of Mumbai	False	11am – 12:30AM (Mon-Sun)	Malad West	Joey's Pizza	Quick Bites	True
2	2500	Seafood	Super Seafood, Best of Mumbai	False	Closed (Mon),12noon – 3pm, 7pm – 12midnight...	Bandra West	Bastian	Casual Dining,Bar	True
3	1800	Finger Food, Continental	Where's The Party?, Best of Mumbai, Food Hygie...	False	12noon – 1am (Mon-Sun)	Lower Parel	Tamasha	Lounge,Bar	True
5	1600	Modern Indian, European, Pizza, Grill	Best of Mumbai, Food Hygiene Rated	False	1pm – 1am (Mon-Sun)	Bandra West	JLWA	Casual Dining,Bar	True

Exploratory Data Analysis

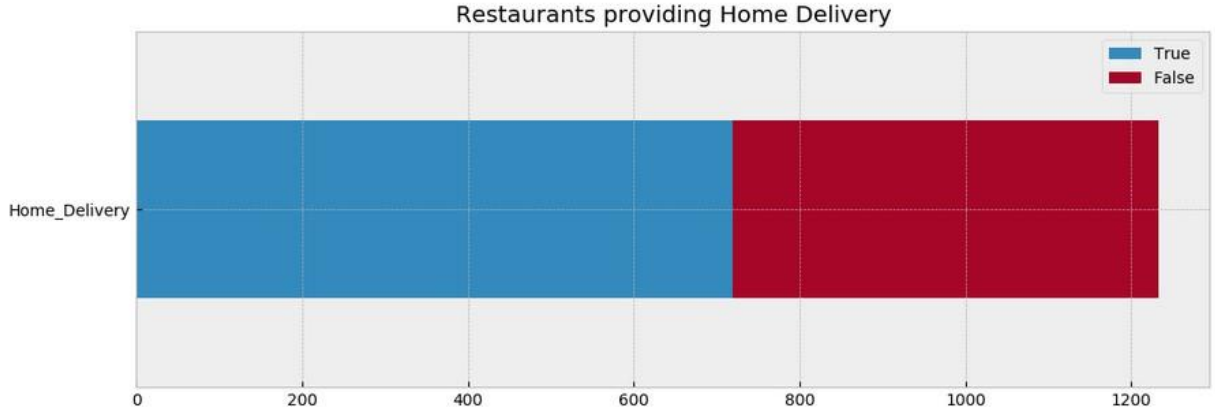
Firstly number of restaurants present in Mumbai and its neighbouring area was found using data and bar plot (below image for your reference).



Consecutively word cloud function from Python was used to obtain most numbers of restaurants in Mumbai (below image for your reference).

As we can see that above word cloud plot shows that Crave has the most number of restaurants in Mumbai, followed by Radha Krishna and The Liquor Factory.

Next we checked for number of restaurants which is providing home-delivery service, as you can see from the below plot that more than 600 restaurants are providing this service in the city.

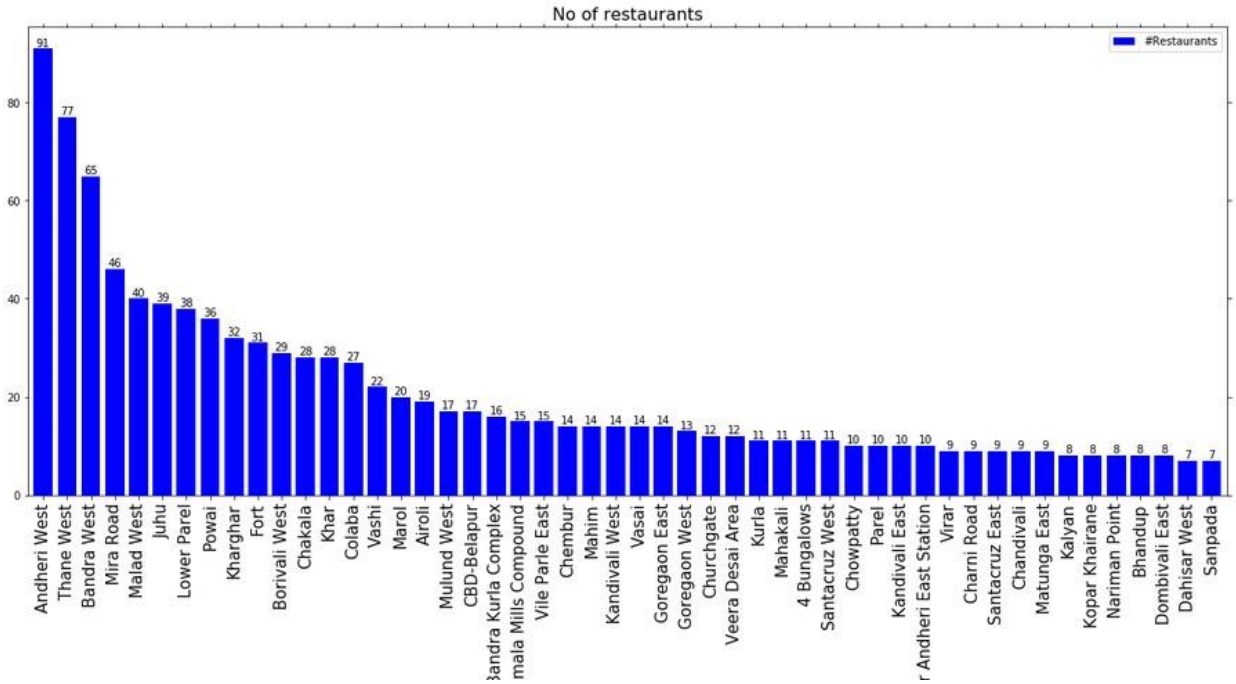


By considering rating of restaurants and we found the average rating for each of the neighbourhood as shown below.

Restaurant_Location	
Vile Parle West	3.916667
Ghatkopar East	3.916667
Thane West	3.931169
Worli	3.940000
Borivali West	3.941379
Goregaon West	3.946154
Marol	3.950000
Grant Road	3.950000
Marine Lines	3.975000
Chakala	3.982143
Nariman Point	3.987500
Bhayandar	4.000000
Wadala	4.000000
Byculla	4.000000
Andheri West	4.002198
Lower Parel	4.002632
Chowpatty	4.010000
Charni Road	4.011111
Sakinaka	4.020000
Dadar West	4.028571
Andheri	4.033333
Khar	4.042857
4 Bungalows	4.045455
Mahim	4.050000
Ghatkopar West	4.066667
Dahisar West	4.071429
Mahalaxmi	4.075000
Ghodbunder Road	4.075000

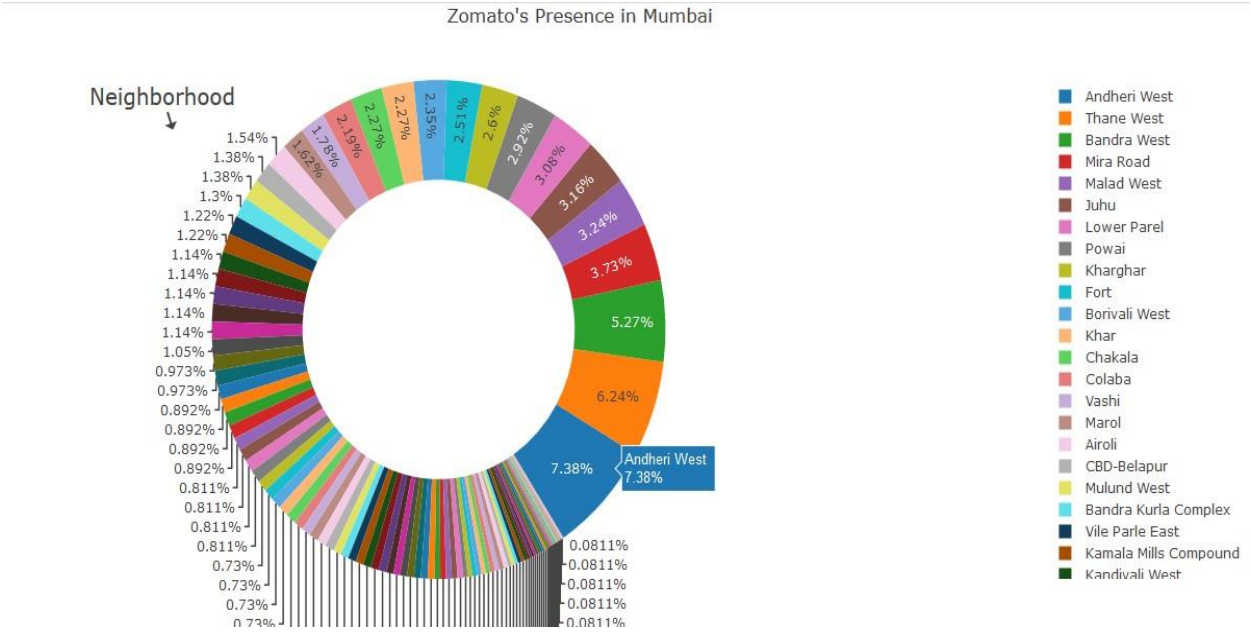
It clearly shows that Kalwa area restaurants are highly rated followed by Bandra Kurla Complex and Kamala Mills Compound

Next, we were interested to know number of restaurants in each neighbourhood of Mumbai. We used *matplotlib* function to generate a bar plot as shown below.



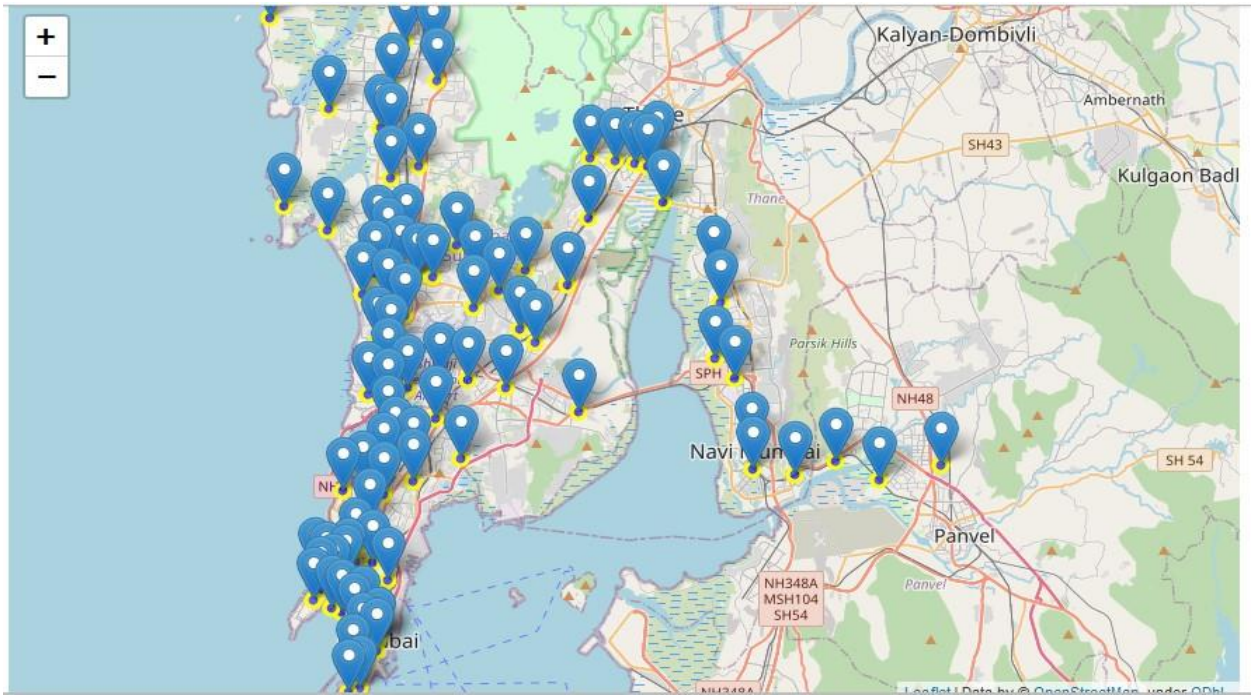
Above plot shows, Andheri West has the highest number of restaurants followed by Thane West of in Mumbai's sub-urban area.

As our data source is Zomato website for Mumbai region, we then explored Zomato' s Presence across Mumbai and its sub-urban area using below shown plot.



Above plot clearly shows that Zomato has its highest presence in Andheri West area of Mumbai.
As the dataset obtained from Kaggle website did not had geographical coordinates so we had use *Nominatim* & *Geocode* functions of Python to obtain the latitude and longitude of the restaurants based on its location and append it to the existing dataset in order to get the coordinates for each of the neighbourhood. Below image shows the new dataset with coordinates. In addition, all the NaN values in the dataset was dropped so that data can visualized on map as shown below.

nu	Rating	Votes	Rating_Category	Operational_after_Midnight	Cuisine_count	Competitors_in_Location	Score	location_latitude	location_longitude
rue	4.6	5974	Excellent	True	1	209.0	71.950295	19.184013	72.841216
rue	4.5	1438	Excellent	False	1	241.0	43.160370	19.058336	72.830267
rue	4.9	3275	Excellent	True	2	125.0	59.778427	18.996332	72.830860
rue	4.8	1989	Excellent	True	4	241.0	50.627941	19.058336	72.830267
rue	4.7	2640	Excellent	True	1	38.0	53.182562	19.067115	72.865724

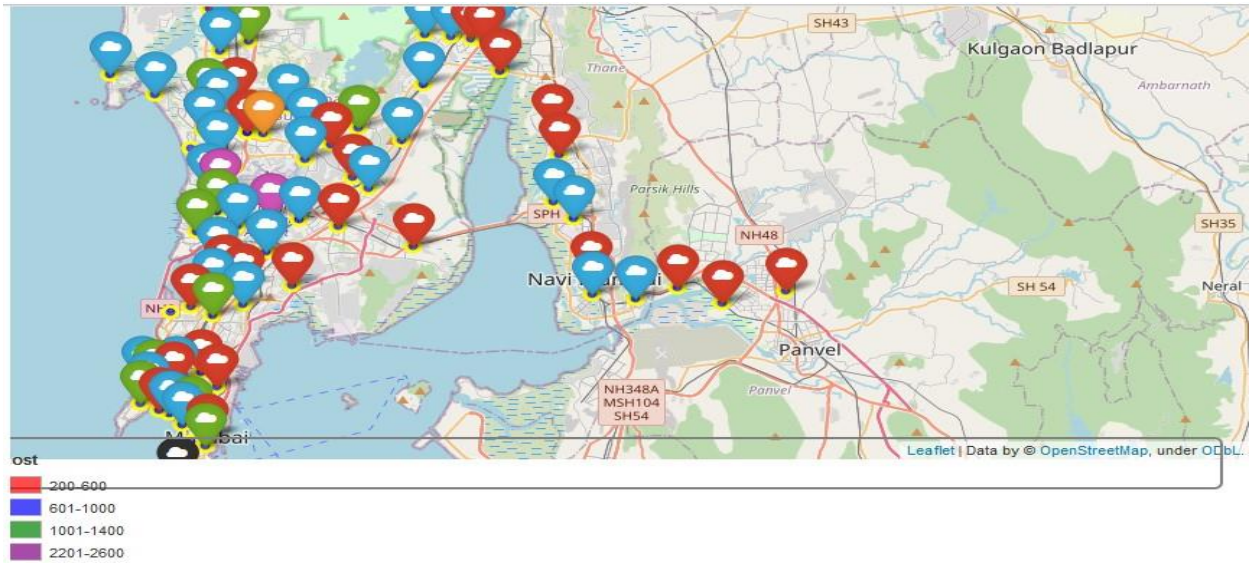


Neighbourhood by cost of food

Next, we checked the cost of food in each neighbourhood of Mumbai and cost of living in Vile Parle East area was highest based on analysis of cost of two persons to eat in this area --as you can see below.

Restaurant_Location	
Vile Parle East	3713.333333
Chakala	2887.500000
Bandra Kurla Complex	2281.250000
Juhu	1852.564103
Colaba	1829.629630
Churchgate	1783.333333
Lower Parel	1532.368421
Parel	1295.000000
Powai	1259.722222
Veera Desai Area	1225.000000
Goregaon East	1142.857143
Khar	1112.500000
Bandra West	1098.461538
Fort	1020.967742
Goregaon West	984.615385
Marol	970.000000
Kurla	940.909091
CBD-Belapur	879.411765
Santacruz West	836.363636
Vashi	831.818182
Mahim	803.571429
Chowpatty	780.000000
Mahakali	740.909091
Andheri West	730.769231
Malad West	680.000000
Thane West	674.675325
Borivali West	665.517241

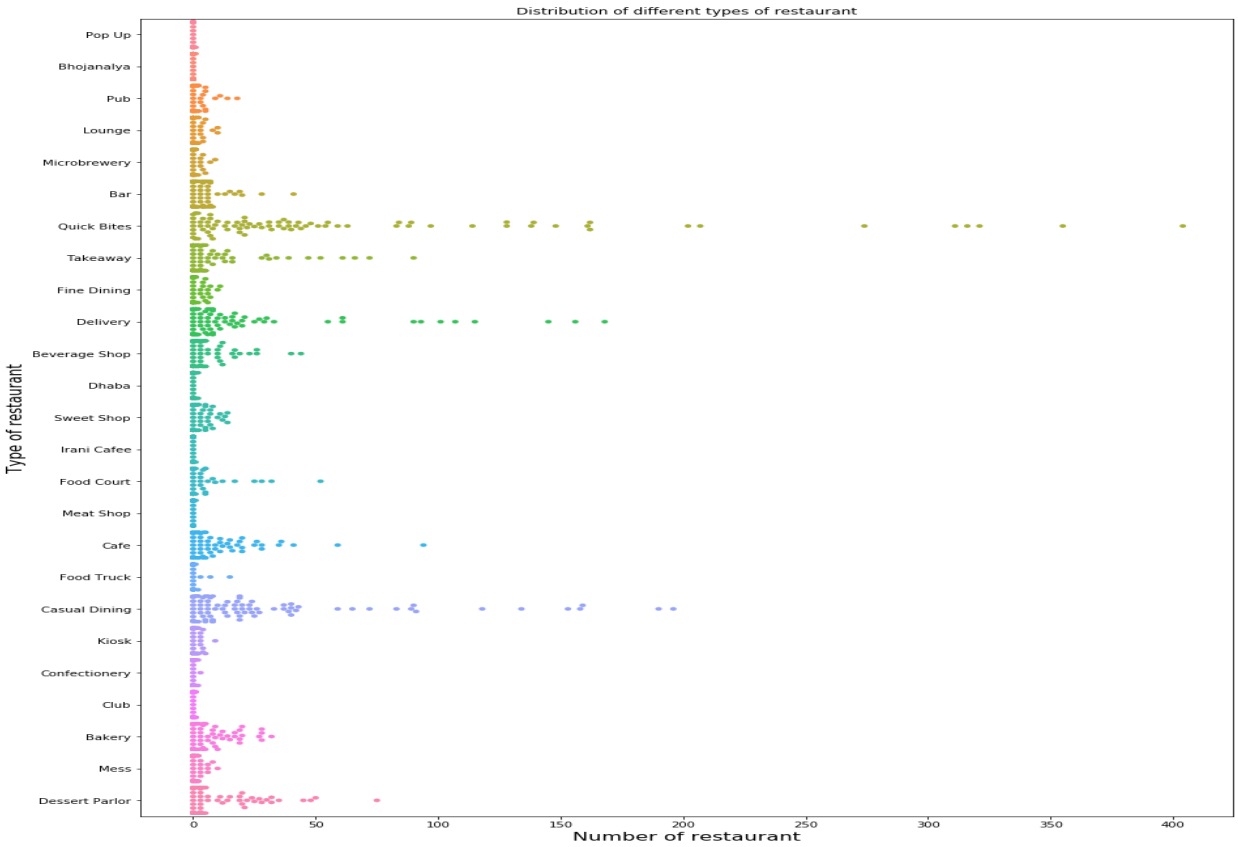
Above obtained data was further used to visualize it on a map as shown below.



It can be seen from above generated map that most of the costly restaurants are located in western, central and south Mumbai areas like - Vile Parle East, Chakala, Bandra Kurla Complex, Juhu, Colaba and Churchgate.

Grouping the neighbourhood by the types of restaurant

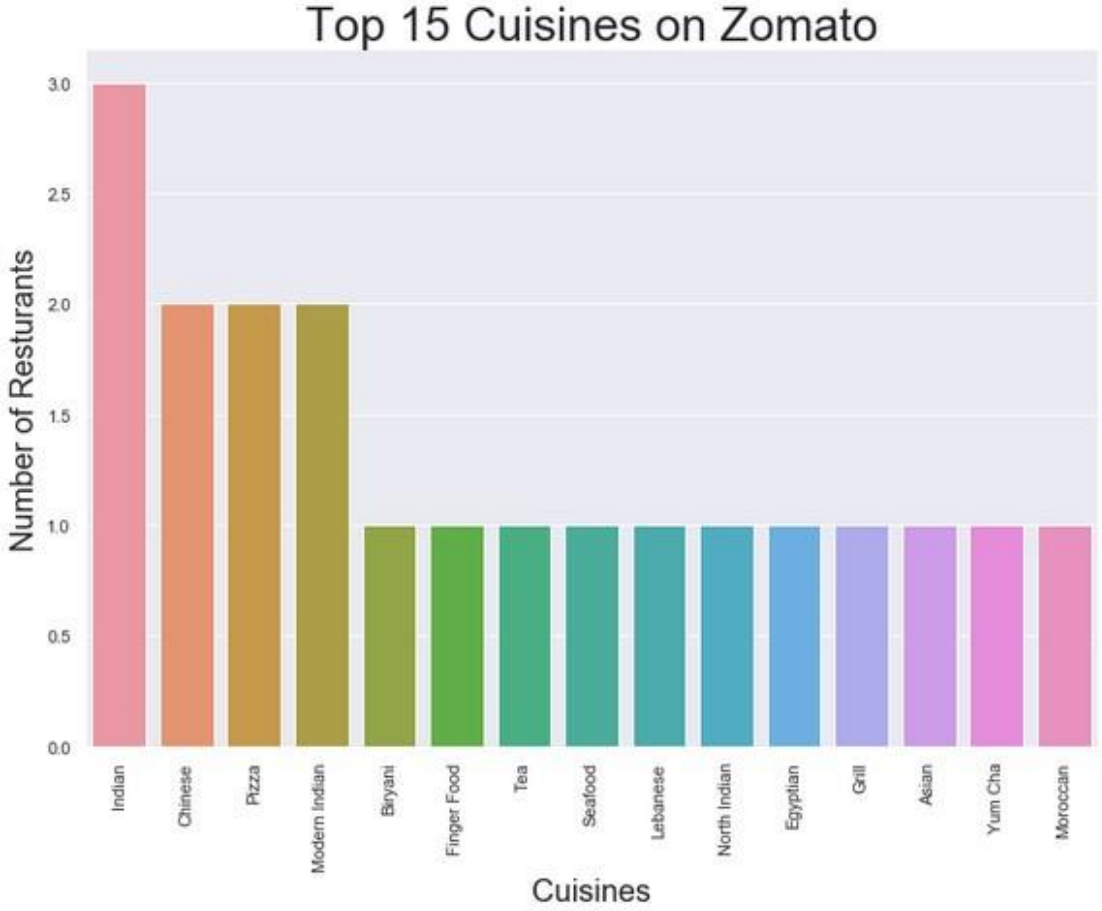
We generated a swarm plot to check the distribution of different types of restaurants showing Quick Bites, Delivery and Casual Dining as the prominent types.

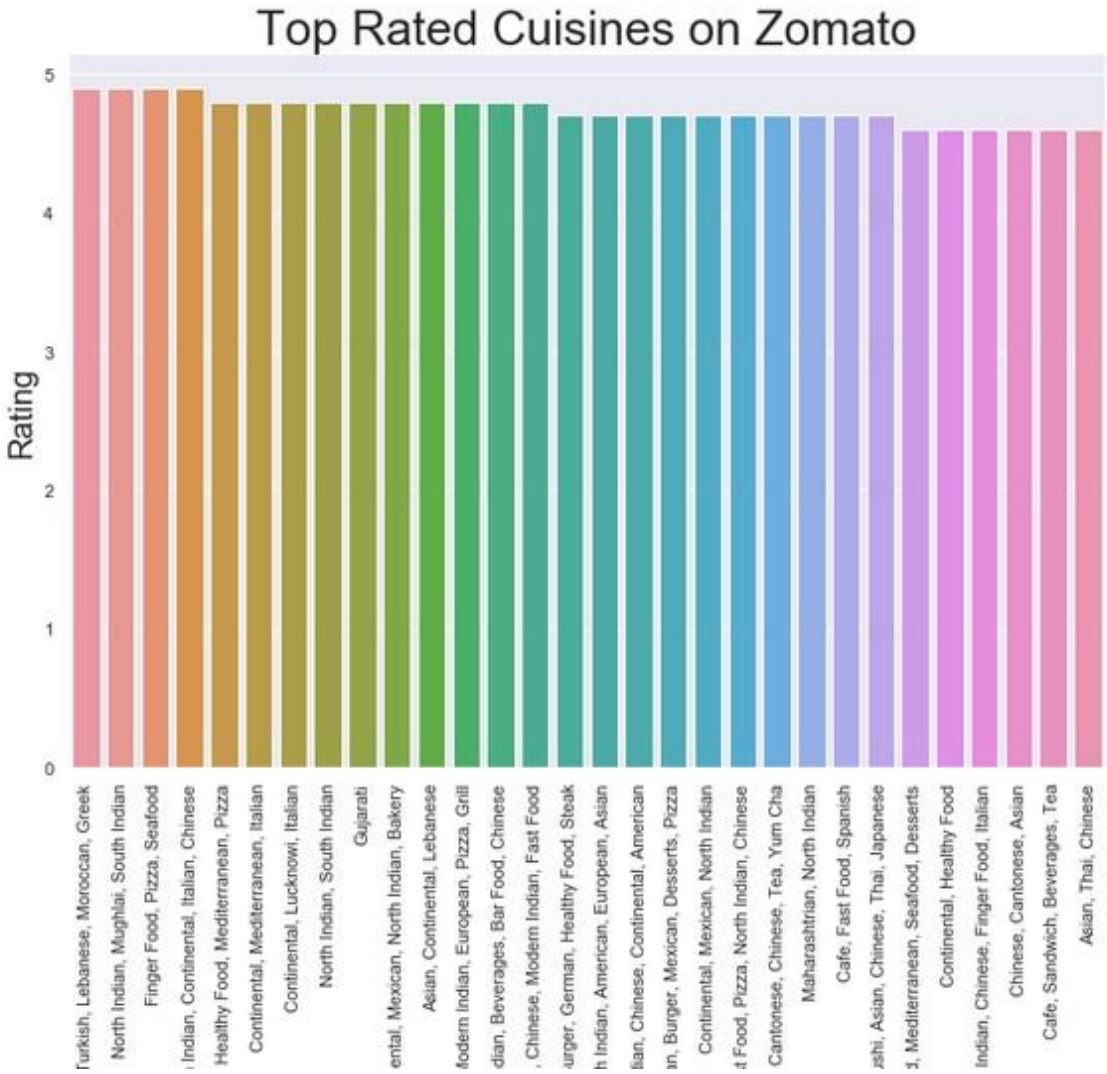


Grouping the neighbourhood by the cuisines

By grouping the neighbourhoods, we checked whether there is any neighbourhood that is more famous for its cuisines using below shown plot.

Simultaneously we explored top fifteen cuisines as well as top rated cuisines on Zomato (shown below).

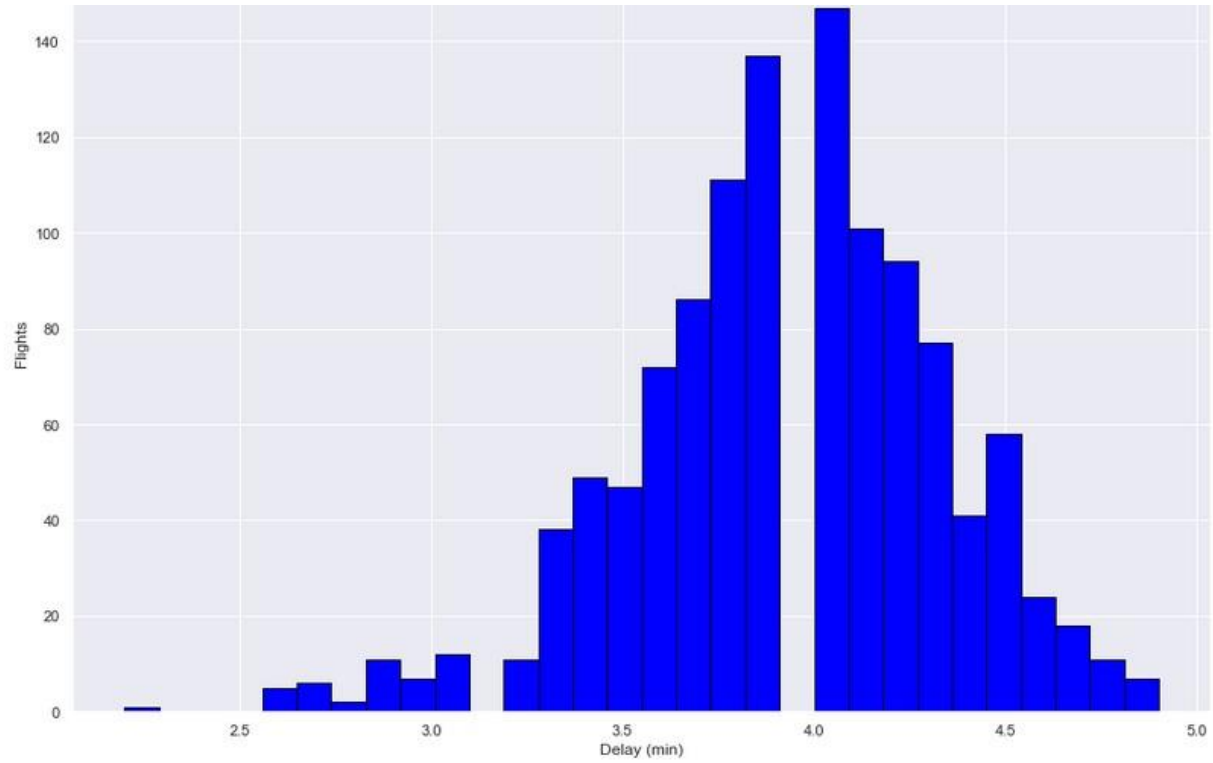




We can observe that in Mumbai Indian dishes are more famous followed by Chinese, pizza and Modern Indian. Rating wise International cuisines are highly rated followed by North Indian, Mughlai and South Indian.

Restaurant Rating

We analysed restaurant rating normal distribution using histogram plot as shown below.



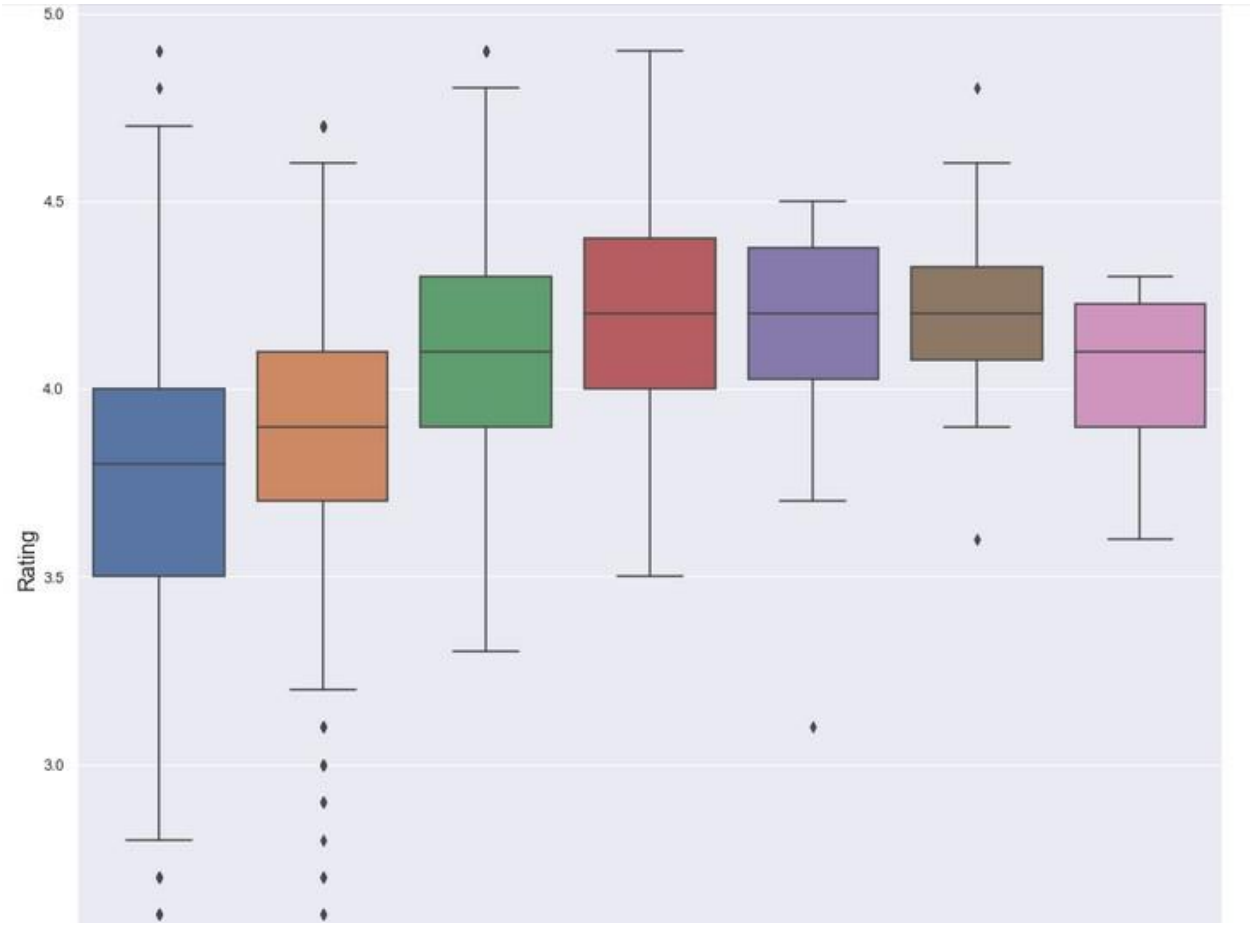
In general, the rating distribution is not normal but in the highest score 4.0 has peak.

Further passing the rating condition greater than 4.8 in our data frame using *new_df[new_df['Rating'] >= 4.8]* . We can see from the above results that most of the high rated restaurants has Continental & North Indian cuisines in their menu; moreover, we can see that the entire higher rated restaurant has higher cost of food too. However, location does not play much important role here.

	Cost_for_two	Cuisines	Features	Home_Delivery	Operational_hours	Restaurant_Location	Restaurant_Name	Restaurant_Type	View_Menu	Rating	Votes
3	1800.0	Finger Food, Continental	Where's The Party?, Best of Mumbai, Food Hygie...	False	12noon – 1am (Mon-Sun)	Lower Parel	Tamasha	Lounge,Bar	True	4.9	3275
4	1600.0	Modern Indian, European, Pizza, Grill	Best of Mumbai, Food Hygiene Rated Restaurants...	False	1pm – 1am (Mon-Sun)	Bandra West	JLWA	Casual Dining,Bar	True	4.8	1989
7	3000.0	Egyptian, Turkish, Lebanese, Moroccan, Greek	Best of Mumbai	False	12noon – 1:30am (Mon-Sat),12noon – 4pm (Sun)	Cuffe Parade	Bayroute	Fine Dining	True	4.9	1037
13	1500.0	North Indian, Mughlai, South Indian	Best of Mumbai, Food Hygiene Rated Restaurants...	False	12:30pm – 3:30pm, 6:30pm – 11pm (Mon-Sun)	Powai	Mirchi And Mime	Casual Dining	True	4.9	5508
15	2200.0	Finger Food, Pizza, Seafood	Best of Mumbai, Where's The Party?, Beer in a ...	False	7pm – 3am (Mon, Tue, Fri, Sat, Sun), Closed (W...	Juhu	Yeda Republic	Bar,Lounge	True	4.9	1922
		Continental	Where's The Party?		12midnight –						

Rating dependencies on price range & cuisines

In this section, we analysed whether restaurant rating is dependent on price range and cuisines using box-plot shown below.



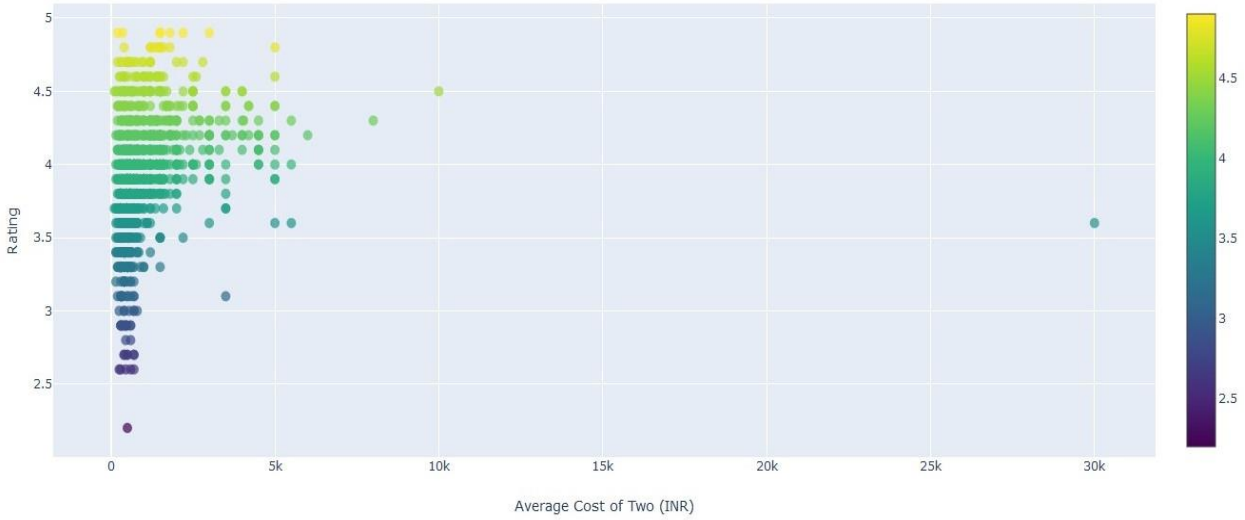
It can be seen that as the price increases the average rating of restaurants also increase. So can we say price does affect rating of a restaurant? Not now, there may be some other factors too. Let us find out -does having more number of cuisines means higher rating of the restaurant? Yes it can be seen below that restaurants having more number of cuisines are rated higher.

```
no_of_cuisine
1    3.849451
2    3.894158
3    3.911194
4    3.946193
5    4.020202
6    3.966000
7    3.944828
8    4.100000
Name: Rating, dtype: float64
```

Statistical Analysis

As we can see from the below plot that, there is no observable linear relationship. At almost every price point, there appears to be both Good and Bad restaurants.

Null Hypothesis: There is no relationship between the rating and price of the restaurant
Alternate Hypothesis: There is some relationship between the two

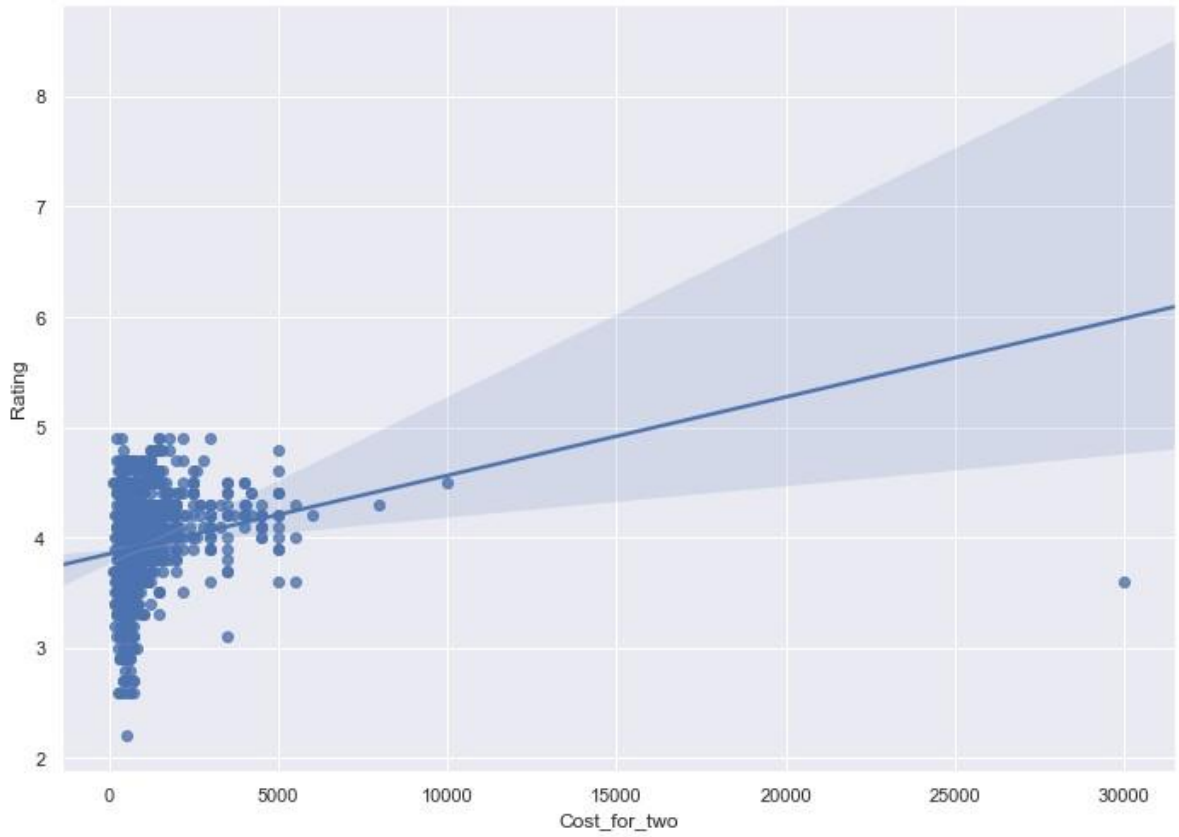



```
Pearson Correlation Coefficient:  0.2268062259131467 and a P-value of:
3.771334182453177e-15
```

The p-values comes out to be much lower than our significance level. Hence, we reject our NULL hypothesis and accept the alternate hypothesis. In addition, our finding can be said to be statistically significant. Therefore, we can say that there is some relationship between the two.

Does more votes means Higher Ratings?

Let us explore this aspect too using statistical analysis and simple linear regression plot as shown below.



As it seems, from a quick glance, there is no observable linear relationship. At almost every price point, there appears to be both low and high cost restaurants.

Let us see the statistical analysis.

Null Hypothesis: There is no relationship between the rating and votes of the restaurant.

Alternate Hypothesis: There is some relationship between the two.

```
pear = new_df.dropna(subset=['Votes', 'Rating'])
```

```
pearson_coef, p_value = stats.pearsonr(pear['Votes'], pear['Rating'])
print("Pearson Correlation Coefficient: ", pearson_coef, "and a P-value of:", p_value)
```

```
Pearson Correlation Coefficient:  0.4481341033447588 and a P-value of:
5.149782109217486e-59
```

The p-value shows result is significant. Hence we reject our NULL hypothesis and accept the Alternate Hypothesis i.e. both the variables are related and our finding was statistically significant.

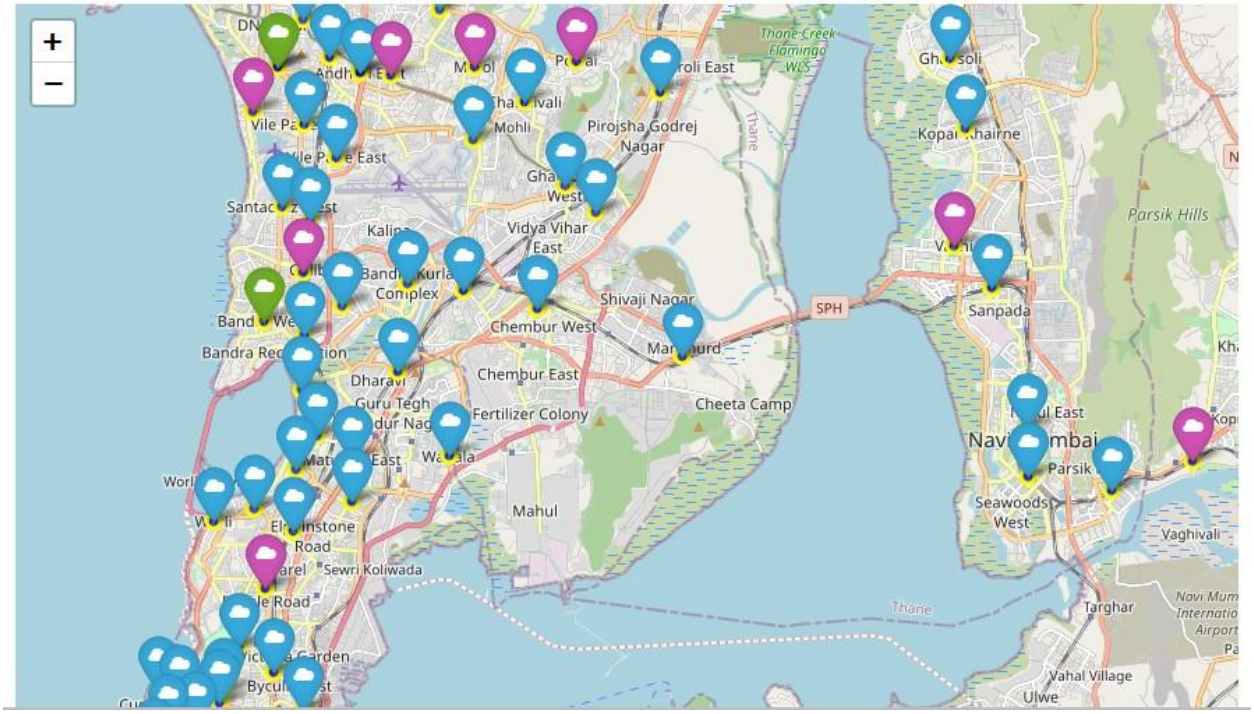
Now let us see the top venues of each areas of Mumbai

Here we will check it using Foursquare API call to check the restaurants, which are more famous by dividing the dataset into five categories. Output result is shown below.

	cat1	cat2	cat3	cat4	cat5
4 Bungalows	Cricket Ground	Café	Coffee Shop	Chinese Restaurant	Parsi Restaurant
Airoli	Ice Cream Shop	Indian Restaurant	Dessert Shop	Indian Restaurant	Playground
Andheri	Bakery	Theater	Pizza Place	Seafood Restaurant	American Restaurant
Andheri East	Bakery	Hotel	Indian Restaurant	Theater	Restaurant
Andheri West	Bakery	Theater	Pizza Place	Seafood Restaurant	Café
...	...	...	...	...	...
Vile Parle East	Club House	Theater	Bakery	Hotel	Seafood Restaurant
Vile Parle West	Theater	Seafood Restaurant	Bakery	Club House	Hotel
Virar	Mountain	Restaurant	Dhaba	Gym	Indian Restaurant
Wadala	Café	Vegetarian / Vegan Restaurant	Indian Restaurant	Gaming Cafe	Bar
Worli	Scenic Lookout	Seafood Restaurant	Athletics & Sports	Lounge	Hotel

94 rows × 5 columns

K-means clustering based on cuisines



On analysing the data it was found that North Indian, Indian, South Indian are available and famous in most (approx. 90%) of the neighborhood of Mumbai. However, among them the key observation are mentioned below.

Cluster 1 or those marked in Blue are the places that are more famous for its North Indian food, Chinese and Fast food and we can see that more number of restaurants are present.

Cluster 2 or those marked in Green are those places that are more famous for its International, Indian, Chinese as well as fast food.

Cluster 3 or those marked in Purple were those areas where Street Food, Chinese, Fast food are least popular.

Conclusion & Future Interpretation

- Quick service restaurants/home-delivery based business model are now mainstream of the Indian food service market, and are growing fast. Fine dining is gaining prominence too. Both multi-cuisine and single-cuisine establishments have seen tremendous growth.
- The model and the idea can be used in real life if we get the access to the ordering data of Zomato.
- Certainly, about entire neighbourhood, which is large, finding out exact data is not possible. However, their ordering pattern and the types of restaurant can find out the taste of a locality.
- Crave has its outlet in most of the neighbourhood of Mumbai followed by other key players. The Indian food service market has seen tremendous changes and disruption when compared to early Nineties when unorganised players and few brands dominated it.

These patterns can be analysed from the data that is seen in the real life too. Thanks to the diversity of Indian food market. The data can also be used to find out the popular food, type of people, residency area etc. to help new restaurants take better data driven decisions.