AUTHOR	Manish Narwal
UNDER GUIDANCE OF	Dr. Carlo Lipizzi
COURSE	EM-624
ASSIGNMENT	Final Project
TOPIC	Zomato India Restaurant Data Analysis
DATE	11/28/2020

RESEARCH QUESTION:

Zomato is the India based food delivery app which is now available in around 25 countries including the United States and Australia. The app is being used widely in India since it was founded in 2008. The app also prompts users to give out ratings and write reviews which will play an important part of our analysis.

For any businessman who wants to start his own restaurant in India, the process can get tedious at times. And for the process to be smooth and the right suggestion to be given to the businessman, data analysis plays a vital role.

In my project I am going to be performing data analysis on Zomato All India dataset to analyze the following factors:

- Analyze which are top 10 restaurants in India based on the ratings given by the users, which will give us the clear picture of the competition in the restaurant market of India.
- Analyze what are the price ranges of the top-rated restaurants in India, which will us an idea of the budget which most people have.
- Analyze what are the most common cuisines in India and what are some of the important factors of the top-rated restaurants.
- Analyze which are the top 10 cities in India with the highest rated restaurants which will help us analyze where our market lies. These cities will be crucial as its highly probable that the business will run more smoothly right here given that it's a big market.
- Of these, top 10 cities we will analyze few of these cities so get some insights on what type of restaurant can possibly be profitable.

To summarize it, through data analysis I will conclude what type of restaurant is thriving with high rating (assuming high ratings is excellent customer satisfaction and hence excellent business) and which city in India will be a good fit for that kind of restaurant.

DATA DESCRIPTION:

The dataset was taken from Kaggle. The following is the **URL**:

https://www.kaggle.com/darshangandhi/zomato-india-dataset/download

The **Original Source of Data** is **Zomato India**

The Dataset contains 211,944 rows of data and 26 variables.

The following is the screenshot of output of df.info() which shows the variable names

```
RangeIndex: 211944 entries, 0 to 211943
Data columns (total 26 columns):
 #   Column                Non-Null Count   Dtype
---  ------                --------------   -----
 0   res_id                211944 non-null  int64
 1   name                  211944 non-null  object
 2   establishment         211944 non-null  object
 3   url                   211944 non-null  object
 4   address               211810 non-null  object
 5   city                  211944 non-null  object
 6   city_id               211944 non-null  int64
 7   locality              211944 non-null  object
 8   latitude              211944 non-null  float64
 9   longitude             211944 non-null  float64
 10  zipcode               48757 non-null   object
 11  country_id            211944 non-null  int64
 12  locality_verbose      211944 non-null  object
 13  cuisines              210553 non-null  object
 14  timings               208070 non-null  object
 15  average_cost_for_two  211944 non-null  int64
 16  price_range           211944 non-null  int64
 17  currency              211944 non-null  object
 18  highlights            211944 non-null  object
 19  aggregate_rating      211944 non-null  float64
 20  rating_text           211944 non-null  object
 21  votes                 211944 non-null  int64
 22  photo_count           211944 non-null  int64
 23  opentable_support     211896 non-null  float64
 24  delivery              211944 non-null  int64
 25  takeaway              211944 non-null  int64
dtypes: float64(4), int64(9), object(13)
```

Fig (1): All Variable names

Not all variables were required for our analysis. The following is the description of the important variables of the analysis:

1) **name**: This is the names of the restaurants
2) **establishment**: This is the type of restaurants like Bar, Café, Lounge, Casual Dining etc.
3) **city**: Name of the city in which the given restaurant is situated
4) **cuisines**: the options for cuisines available
5) **price_range** : This is range of price (cost for 2 individuals on average). The following are the price ranges in Indian Rupees (1 Indian Rupees is approximately 0.014 US Dollars)
   0 to 350 Rs is Price Range 1
   351 to 750 Rs is Price Range 2
   751 to 1850 Rs is Price Range 3
   1850 Rs above is Price Range 4
6) **highlights**: These are the key services available at the restaurants like Wi-fi, Indoor delights etc

7) **aggregate_rating**: These are the average ratings given to the restaurant by the user
8) **votes**: These are the number of users who rated/reviewed the restaurant

DATA PREPARATION:

For a good start and better analysis, I took only the data which had **aggregate_ratings** of greater than 3.0 and minimum **votes** of 100. This is because any restaurant above score of 3.0 out of 5 is considered decent restaurant. For the votes, I feel that at least 100 user thresholds should be there for the restaurant to be considered decent one, especially considering the large population of India.

After this I was left with only 104,470 rows of data which is still good. I then cleaned highlights and cuisine column by removing commas from column values. There were few missing values in the establishment column and I filled it out using the forward fill function. Now there was a need to check the null values if there are any.

```
In [48]: df_final.isnull().sum()

Out[48]: res_id                 0
         name                   0
         establishment          0
         url                    0
         address                1
         city                   0
         city_id                0
         locality               0
         latitude               0
         longitude              0
         zipcode            68377
         country_id             0
         locality_verbose       0
         cuisines              13
         timings              239
         average_cost_for_two   0
         price_range            0
         currency               0
         highlights             0
         aggregate_rating       0
         rating_text            0
         votes                  0
         photo_count            0
         opentable_support     19
         delivery               0
         takeaway               0
         dtype: int64
```

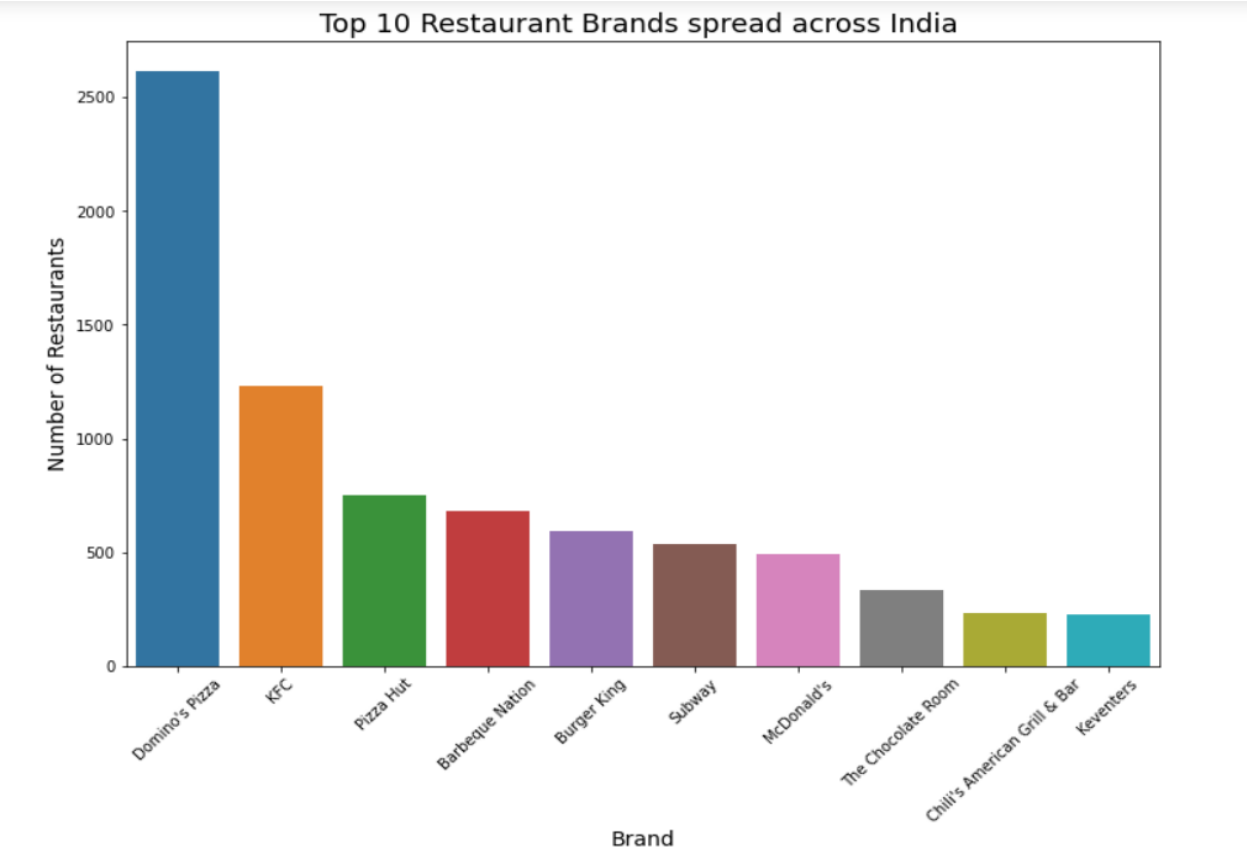
Fig (2): Checking null values in the dataset

It can be observed that there are null values in zipcode, cuisines, timings and opentable_support variable columns.But these are not going to significantly impart my analysis since I am not going to use the variables zipcode, timings and opentable_support.And for cuisines I am going to select only the most common occurring words hence there is no need to clean the null values as it will not impact my analysis.

STEPS FOLLOWED FOR ANALYSIS:

1) Import all the necessary libraries and open the CSV file of the dataset.
2) Select the data of “aggregate_rating” >3.0 and “vote” > 100 to make the analysis more productive and filter anomalies out.
3) Fill missing values in “establishment” column.
4) Plot a chart for top 10 highest rated restaurant brands all over India.
5) Plot a chart to get the idea of price ranges which most users prefer, and which most good restaurants have.
6) Plot the word cloud to get most common cuisine and most common services which the restaurants in India offer.
7) Plot a chart to see which kind of restaurant type is famous among the users and which they have liked so much.
8) Plot a chart to see which cities in India has good restaurants.
9) Analyze some of those cities and come up with the conclusion.

RESULTS OF ANALYSIS:



Fig(3): Top 10 Restaurant chains in India

Top Restaurant Chains:

From the **Fig(3)** it can be seen that all these restaurant chains are multinational brand with Domino's having most chains out of all other major brands like KFC, Pizza Hut, Burger King etc. So, opening a pizza or fried chicken restaurant is probably not a good idea as the major players are already dominating the fast-food market in India.

Price Range Preferences:

Now we'll see what kind of price range the good restaurants have in India which will give us the idea of the preferred budget of the users in India.

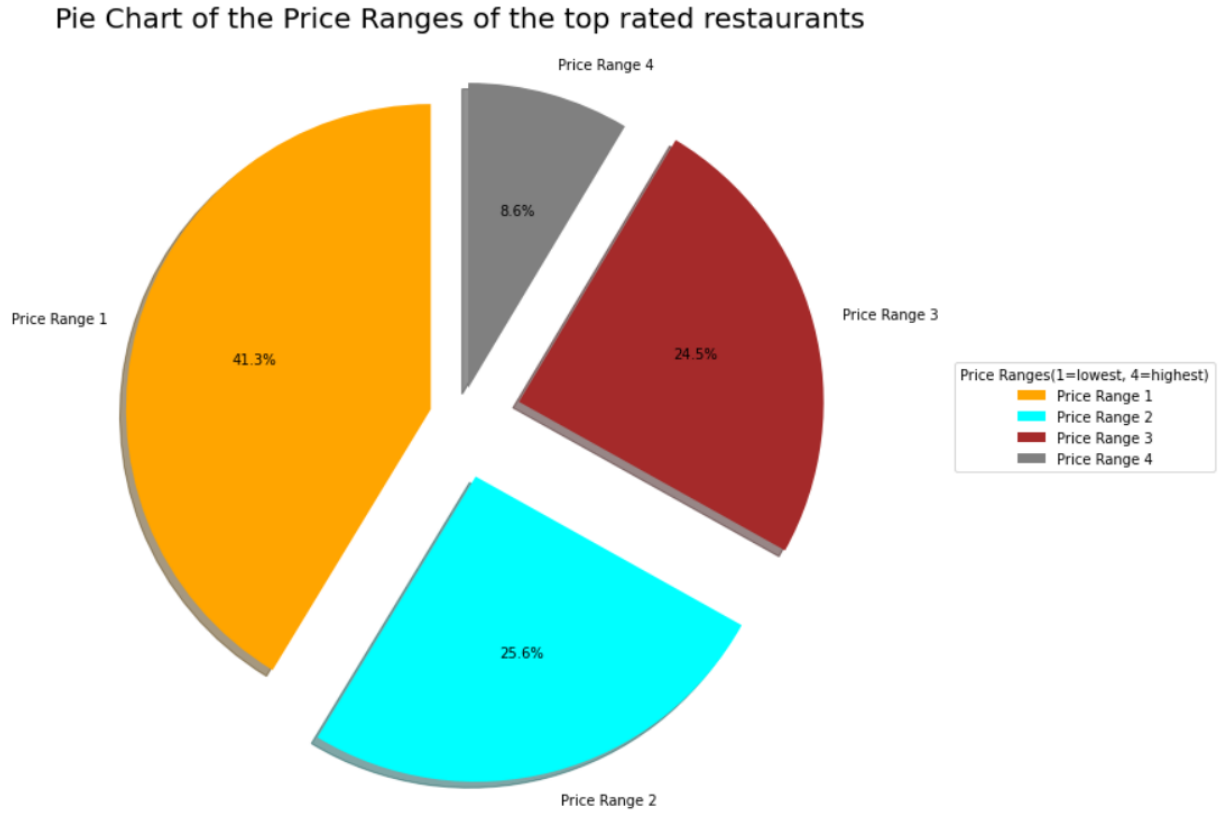


Fig (4): Price Range of the highly rated restaurants

It can be observed that most of the good restaurants have average cost for two in Price Range 1 which is Rs.1 to Rs. 350 and for Price Range 2 and 3 are almost equal. From this, it is safe to say that majority of the people in India will prefer restaurants which offers meal for two at a cost range of Rs.1 to Rs.500 (a little over Range 1). This can be a good point of suggestion which can be made to a businessman planning to start restaurant in India. As he/she will not benefit much profit if the price range is 3.

Common Services at restaurants in India:

Fig (5): Word Cloud of Highlights of Services

From the Word Cloud, it can be observed that apart from major words (Available, Table, Required, Reservation which are important in restaurant business) there are interesting words like Entertainment, Wi-fi, desserts, Bakes, cocktails, Veg etc. show the kind of services of the highly rated restaurants. It is important to convey the stakeholder (the person who wants to set up restaurant in India) this information and recommend them to highly consider having these services.

Common cuisines at restaurants in India:

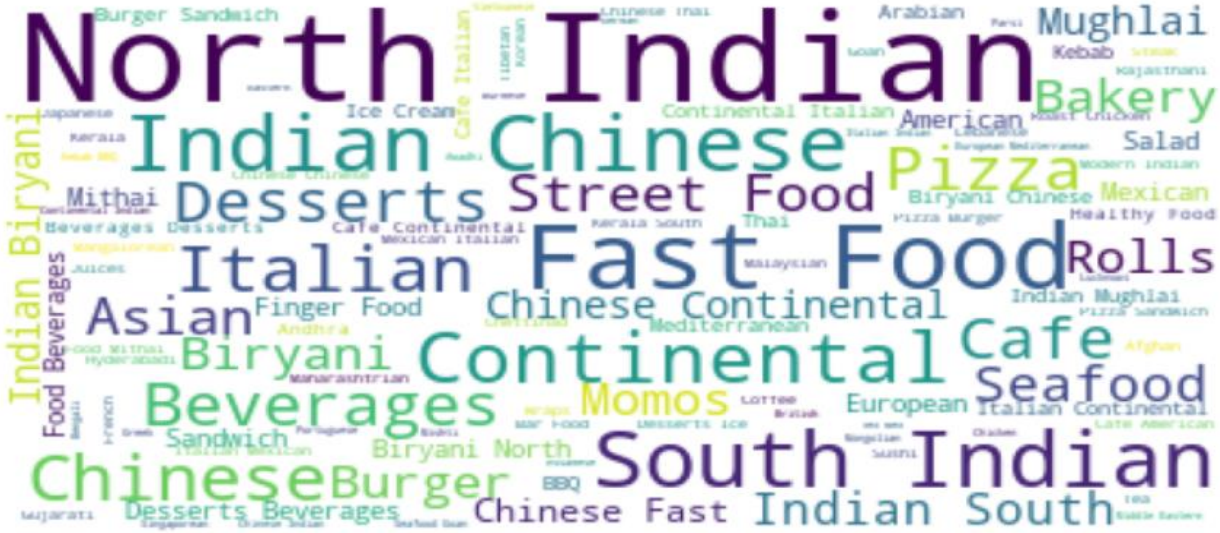


Fig (6): Word Cloud of most common cuisines

From the above word cloud, it is observed that the North Indian cuisine is one of the most common cuisine in any of the top-rated restaurants. So, according to the analysis, our stakeholder should consider having at least some of the North Indian, Chinese, South Indian, Continental, Biryani on the menu. This will highly increase profits and good ratings.

Kind of restaurant popular among the users:

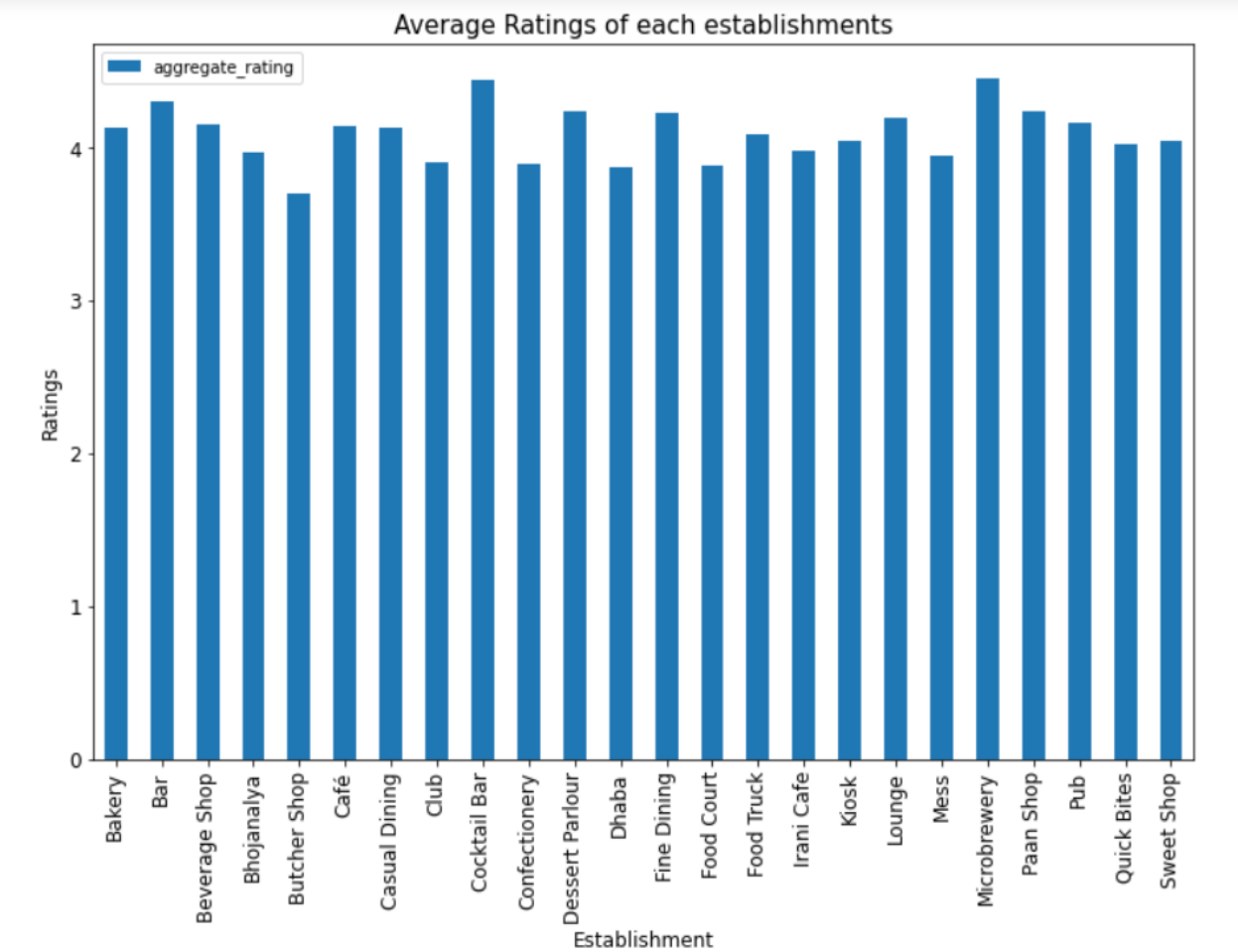


Fig (7): Bar Plot of most popular restaurant kind which are high rated

From the Bar Plot, it can be observed that the Cocktail Bar and Microbrewery type of restaurants are very much highly rated in India. That is followed by other Bars and Fine Dining. From these observation the choices are narrowed down to only few which are cocktail bars primarily and fine dining as secondary option.

One of the reasons why India has highly user rated Bars and Pubs is because majority of the Indian population falls in the age group 18-45 years. This is another good insight which needs to be considered by the stakeholders.

Top 10 cities which has great market for restaurants:

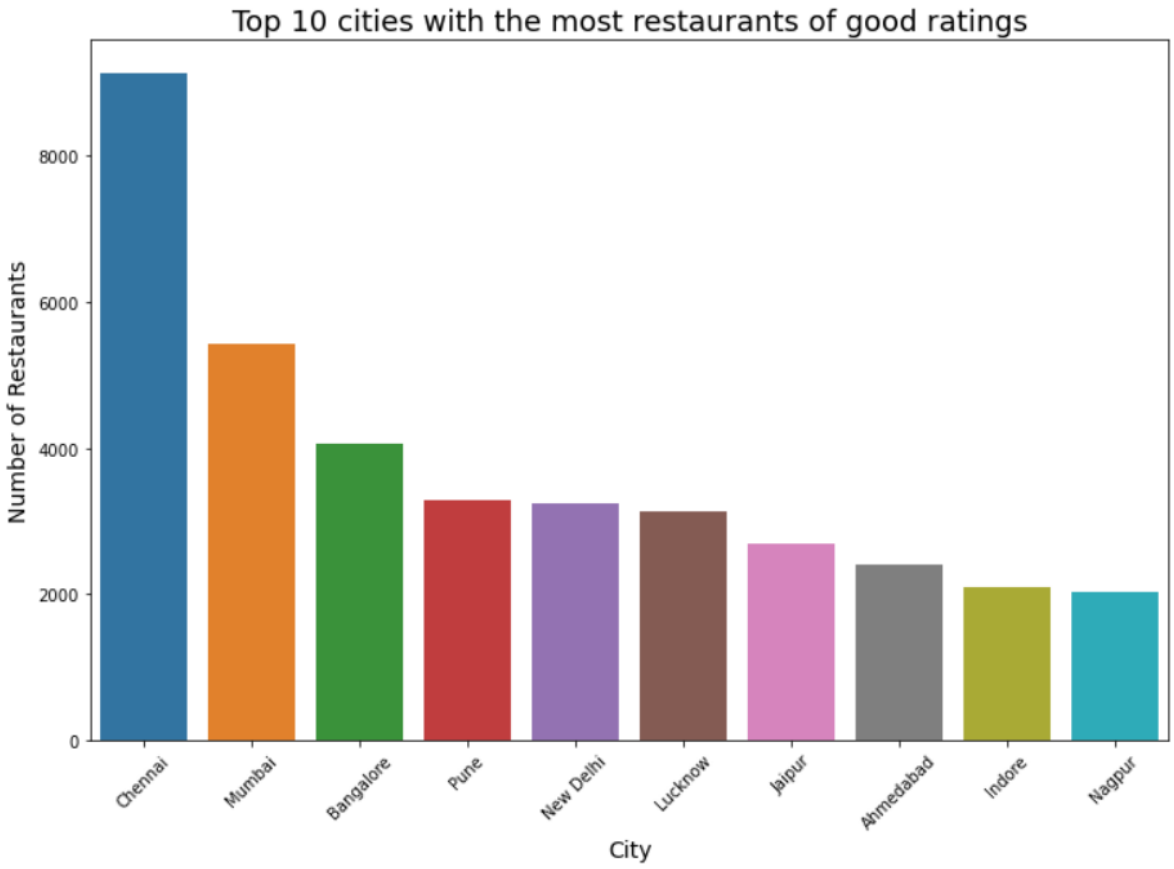


Fig (8): Top 10 cities with great market for restaurants

From the **Fig (8)** , Chennai , Mumbai and Bangalore are the top 3 cities with most quality restaurants. The stakeholders need to target these cities for maximum profit and good returns on investments. In my further analysis, I will analyze the top 3 cities to see what kind of restaurants they have, and which one will be in demand.

Chennai Analysis:

The following is the bar plot of the number of restaurants in Chennai according to the restaurant types:

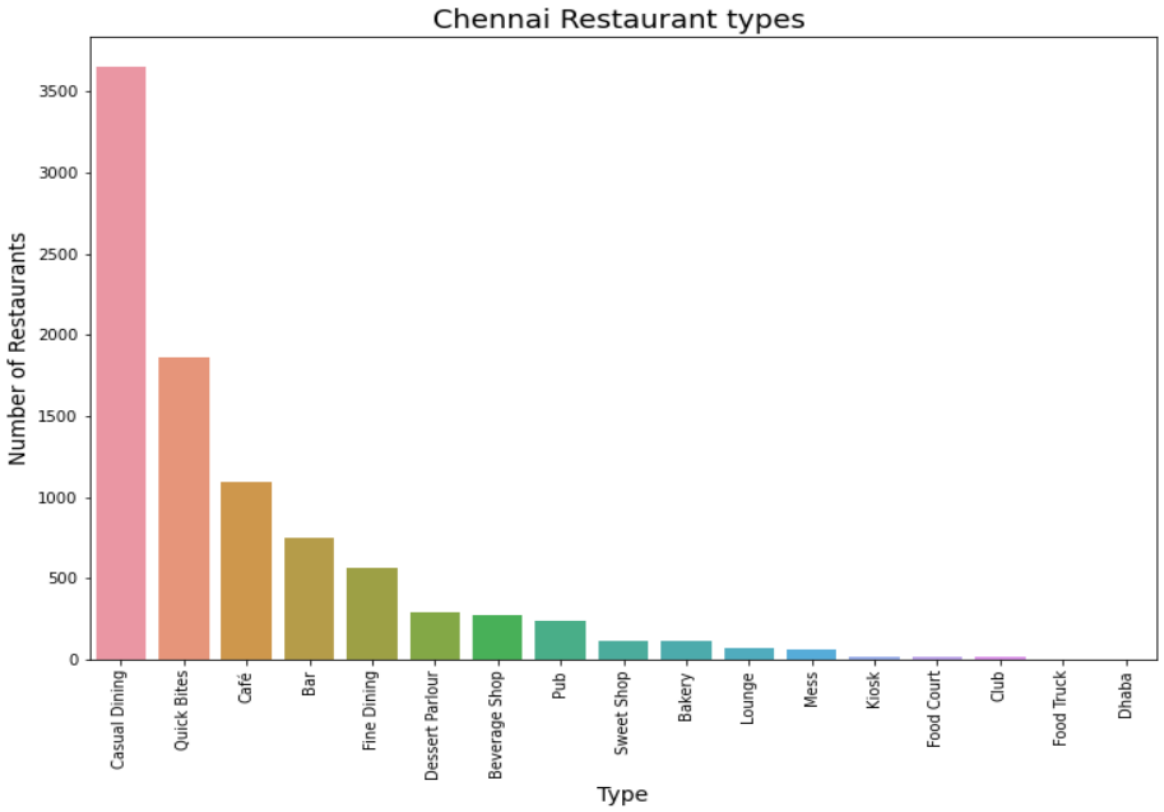


Fig (9): Top restaurant types in Chennai

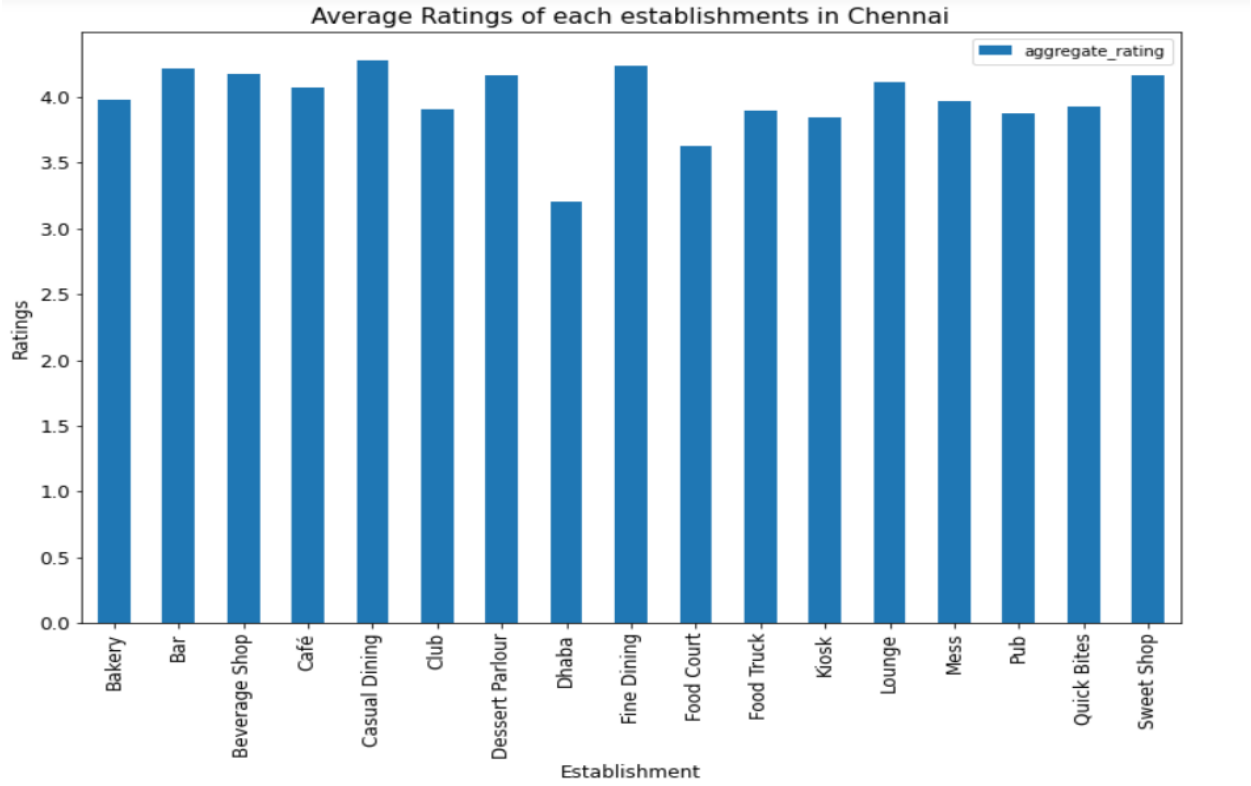


Fig (10): Average ratings of restaurant types in Chennai

From the above graphs of restaurants in Chennai, it is clear that Casual Dining is quite popular in Chennai as it has more restaurants dedicated to it but there comparatively less bars and pub and even less food trucks and dhaba.

Bangalore Analysis:

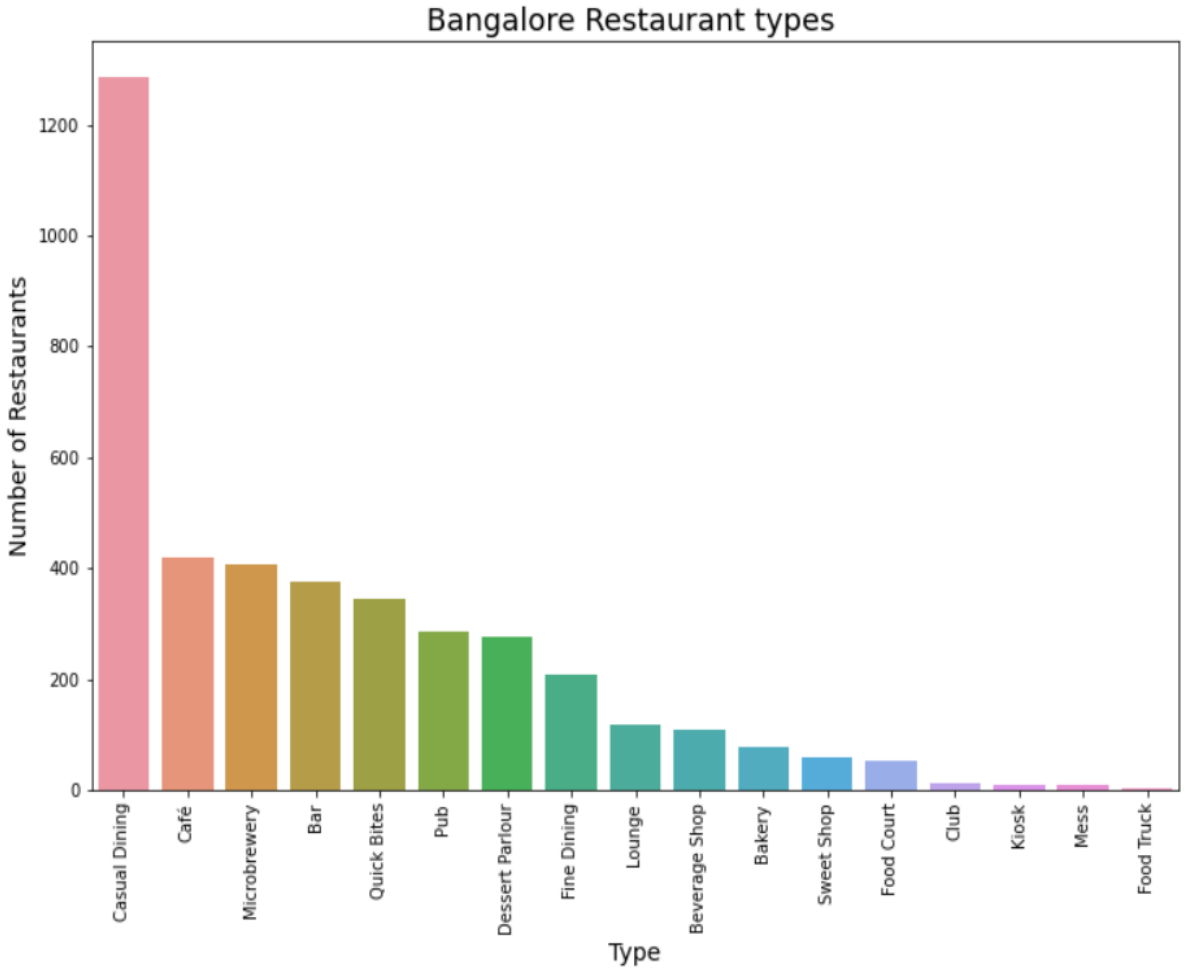


Fig (11): Top restaurant types in Bangalore

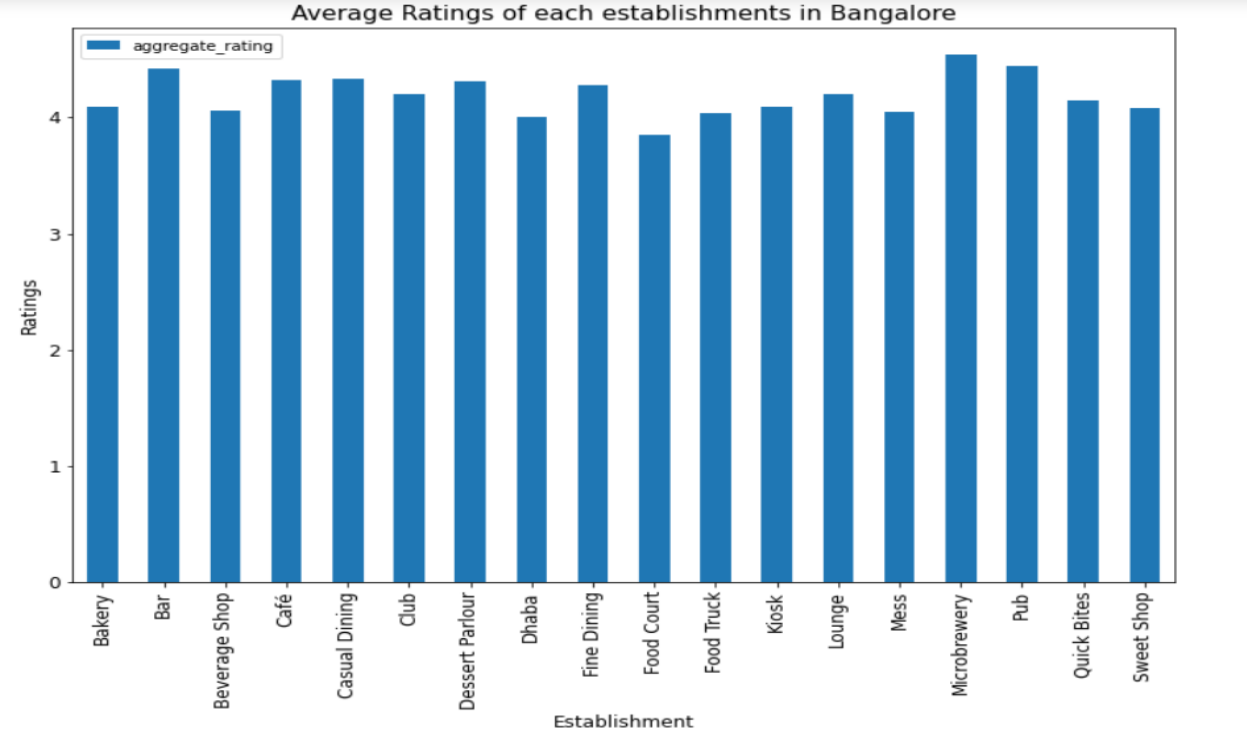


Fig (12): Average ratings of restaurant types in Bangalore

From the two plots of Bangalore, Casual Dining options in the city are extensive compared to those of Bars, pubs and food trucks. Bangalore being a big city, is very surprising that it has less number of bars and pubs.

Mumbai Analysis:

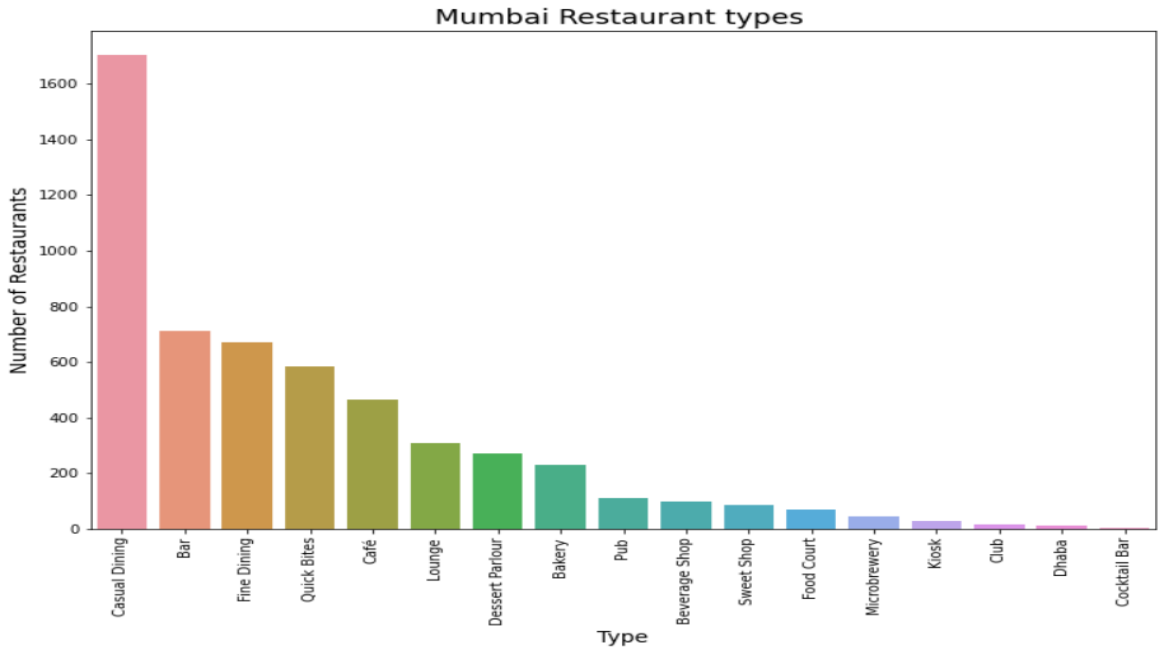


Fig (13): Top restaurant types in Mumbai

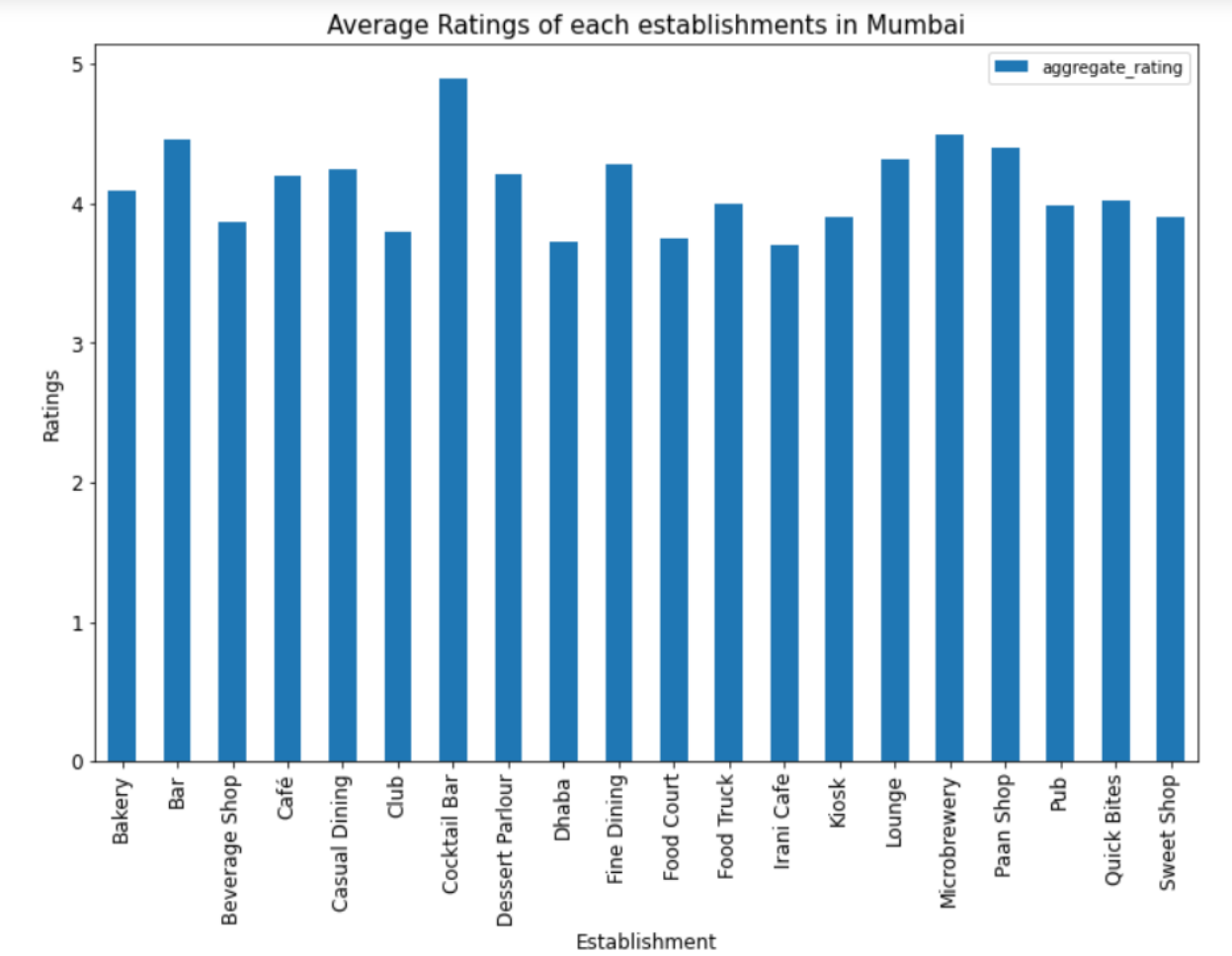


Fig (14): Average ratings of restaurant types in Mumbai

From both the plots of Mumbai city, it is observed that it once again follows the same pattern as other two cities. But it is to be noted that Mumbai has Microbrewery and Cocktail Bars which are also very highly rated. This makes the city of Mumbai even more competitive for stakeholders as there is decent amount of variety and hence more competition here.

CONCLUSION OF THE ANALYSIS:

After the extensive analysis, it is clear that there is vast competition in the Casual Dining type of restaurants. The following seems to be some of the options which can be suggested to the stakeholders:

- Set up a mocktail bar or microbrewery in Chennai. As it is already a popular option in the country it is surprising that there is not one in big city like Chennai. Where the demand is bound to more. A short survey in Chennai could give a very nice idea about what people think of mocktail bar. But according to the analysis, it will become popular in the city and thus yield profit for the stakeholders.
- Set up pub in Mumbai. With Mumbai being the most populated city in India, there is not a lot of option for nightlife or weekend party. With less pubs the ratings are also low, suggesting that there is high demand for a good pub. A pub with good vibes, services and quality will surely thrive in the city like Mumbai.

These two options will certainly bring profits for the stakeholders and provided the stakeholders keep up with the customer satisfaction and increasing competition in the restaurant market they would benefit a lot.