Analysis of Online Food Delivery Process by Zomato using Data Science

Name: Kolipaka Preethi

Track: Python- Data Science

CHUBB – INTERN

ABSTRACT:

People all around the world are preferring online platforms to complete all of their vital work and avoiding going to the marketplaces or to the job owing to the spread of airborne disease, which has brought the entire world to a standstill.

The current world is deteriorating due to the spread of the corona virus, and the entire country is going through a difficult time due to the lockdown, which has put a stop to any country's desire of multilateral growth.

Swiggy, Zomato, Uber Eats, and other food-related online giants include Swiggy, Zomato, and Uber Eats, among others. The goal of studying the Zomato dataset is to have a clear understanding of the elements that influence each restaurant's overall rating, as well as to identify distinct types of restaurants in different locations.

This report uses data science to demonstrate data analysis on the Zomato dataset assisting website in order to aid in better ideas for future marketing strategy. This study addresses the demands of consumers looking for the best cuisine in the country and which regions of the country have the most restaurants serving such cuisine.

INTRODUCTION:

Zomato is one of the most important online foods ordering business in today's generation. It is a kind of an online guide book for users of Zomato. This site mainly focuses on the restaurants of Bangalore and helps the user categorized the different places and choose the best among them in means of menus, dishes, localities, and several other mediums. In order to understand the data set, there is a necessity of procedure by means of machine learning concepts which would help in analysing the data set in all different aspects. We could develop marketing strategy by noticing the client comfort through the result of deep analysis performed by data science.

DATA COLLECTION:

Data science is the study of the systematic removal of indistinct and useful patterns and knowledge from data, towards research advancement, organizational decision-making which enables a computer-controlled society. Accordingly, scholars advocate for data-intensive science, discovery science or data-driven science, in which: It is suited to make sense of massive interconnected datasets, overcoming problems of small samples, and scarce data. Interdisciplinary research is promoted. Complete models and theories about complex systems, rather than elements of it, are possible.

The data collected for this research is from the secondary data sources (Kaggle). Since the complexity and quantity of the data leads, need to perform data pre-processing in order to get an accurate result. The first step in data analysis is to prevent missing data, then the co-relation method is performed to foremost the data visualization process.

APPROACH:

- The various python libraries such as Numpy, pandas, Matplotlib are used for the purpose of mathematical calculations, extraction of data and visualization respectively.
- The extracted dataset was found to have null values which are then effectively handled to achieve accurate analysis rather than removing them from the dataset.
- The datasets to be used for analysis was found to be highly skewed. Although skewness will not have much effect on descriptive analysis, it is handled for the purpose of predictive analysis to accomplish better decisions from the model. The log transformation was used to handle skewness in the datasets.

DATA PRE-PROCESSING:

1. Counting Missing Values for different columns

```
url                                0
address                            0
name                               0
online_order                       0
book_table                         0
rate                            7775
votes                              0
phone                           1208
location                          21
rest_type                        227
dish_liked                     28078
cuisines                          45
approx_cost(for two people)      346
reviews_list                       0
menu_item                          0
listed_in(type)                    0
listed_in(city)                    0
dtype: int64
```

- Dropping Unnecessary columns.
- Converting "votes" and "approx_cost_for_2_people" into numeric(int).
- Removing comma and changing approx_cost_for_two_people as int.
- Removing Restaurant data's where rate = 'NEW'.
- Removing '/5' in rate column.

2. Information on Original Zomato Dataset.

```
url                                  51717 non-null object
address                              51717 non-null object
name                                 51717 non-null object
online_order                         51717 non-null object
book_table                           51717 non-null object
rate                                 43942 non-null object
votes                                51717 non-null int64
phone                                50509 non-null object
location                             51696 non-null object
rest_type                            51490 non-null object
dish_liked                           23639 non-null object
cuisines                             51672 non-null object
approx_cost(for two people)          51371 non-null object
reviews_list                         51717 non-null object
menu_item                            51717 non-null object
listed_in(type)                      51717 non-null object
listed_in(city)                      51717 non-null object
dtypes: int64(1), object(16)
```

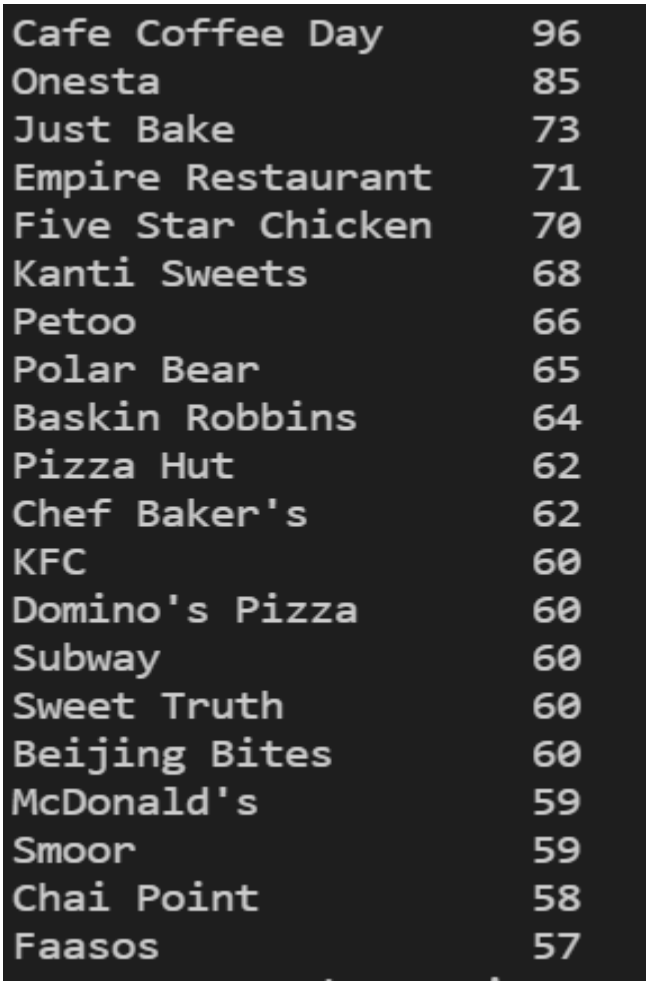
From above, it is found that the column "dish liked" has more than 50% values missing.

ANALYSIS AND VISUALIZATION:

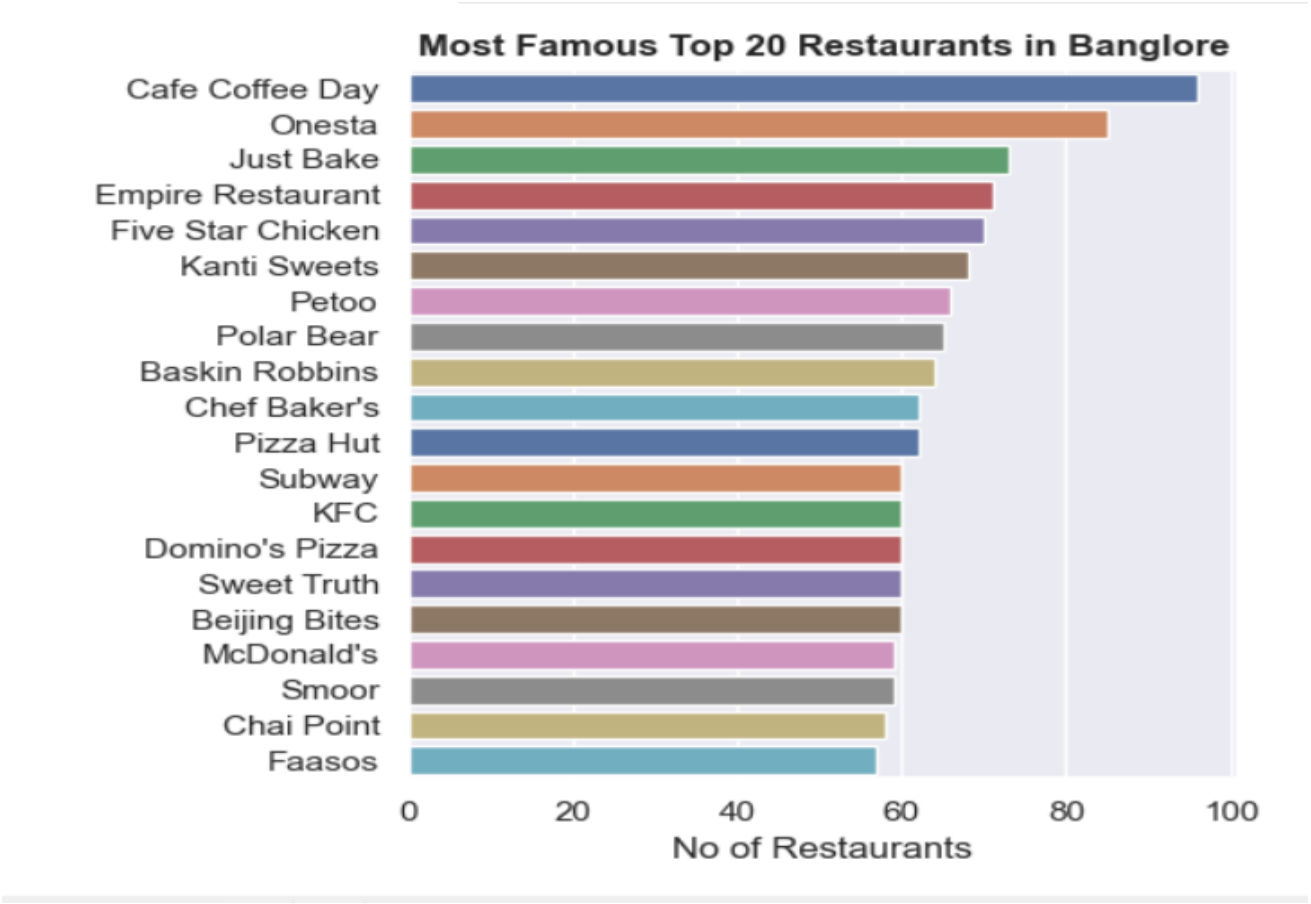
1. Most Popular Top 20 Restaurants in Bangalore

PURPOSE

- In Order to Know the top famous restaurants to choose by the chains they have in in Bangalore.



Cafe Coffee Day	96
Onesta	85
Just Bake	73
Empire Restaurant	71
Five Star Chicken	70
Kanti Sweets	68
Petoo	66
Polar Bear	65
Baskin Robbins	64
Pizza Hut	62
Chef Baker's	62
KFC	60
Domino's Pizza	60
Subway	60
Sweet Truth	60
Beijing Bites	60
McDonald's	59
Smoor	59
Chai Point	58
Faasos	57



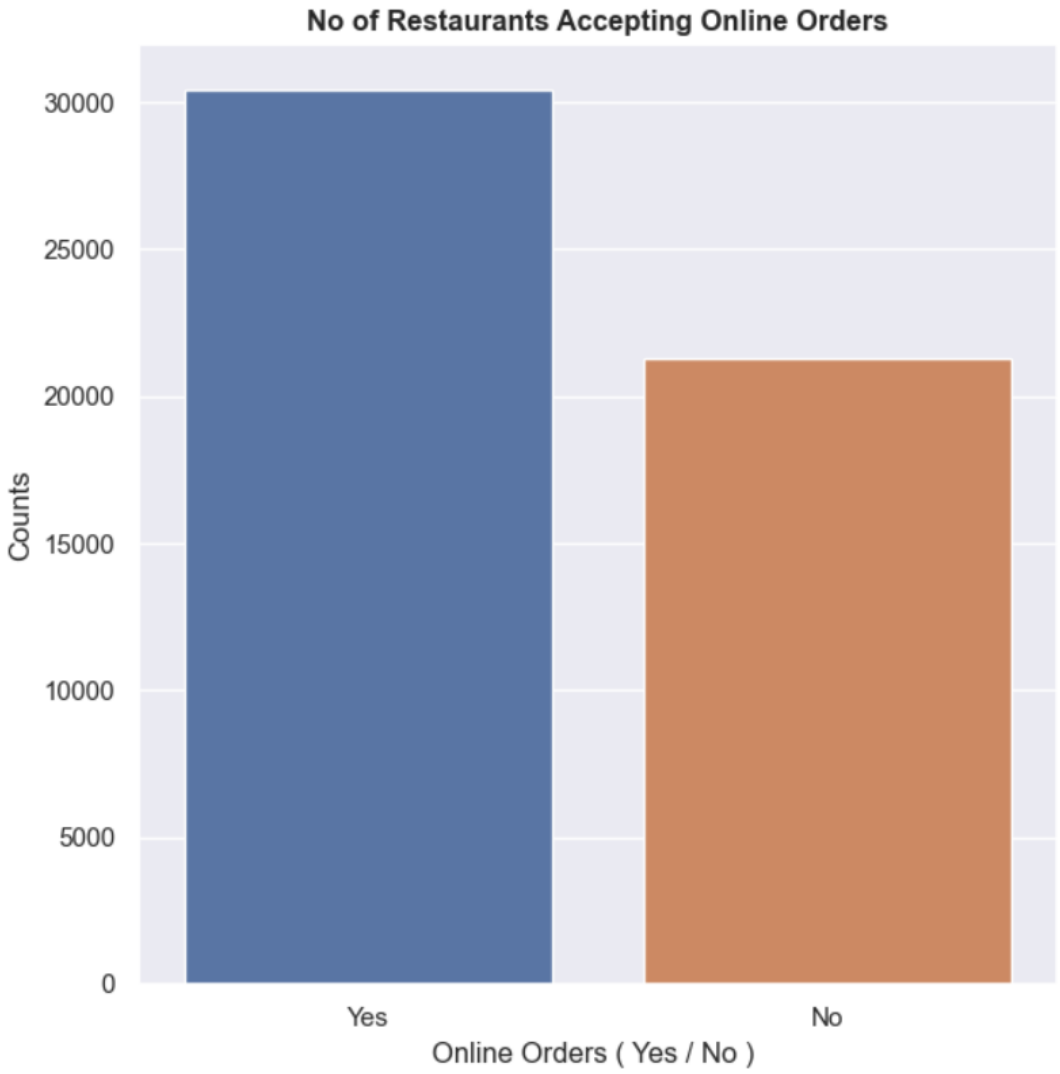
INFERENCE:

- By the above graph here we can see that “Cafe Coffee Day “has the most popular restaurant chains and Onesta has almost 90 popular chains in Bangalore. They have almost 100 restaurants in Bangalore.

2. Restaurants Accepting Online Orders

PURPOSE

- To Know the Restaurants Which are accepting Online Orders and their Count for Online vs Offline orders.



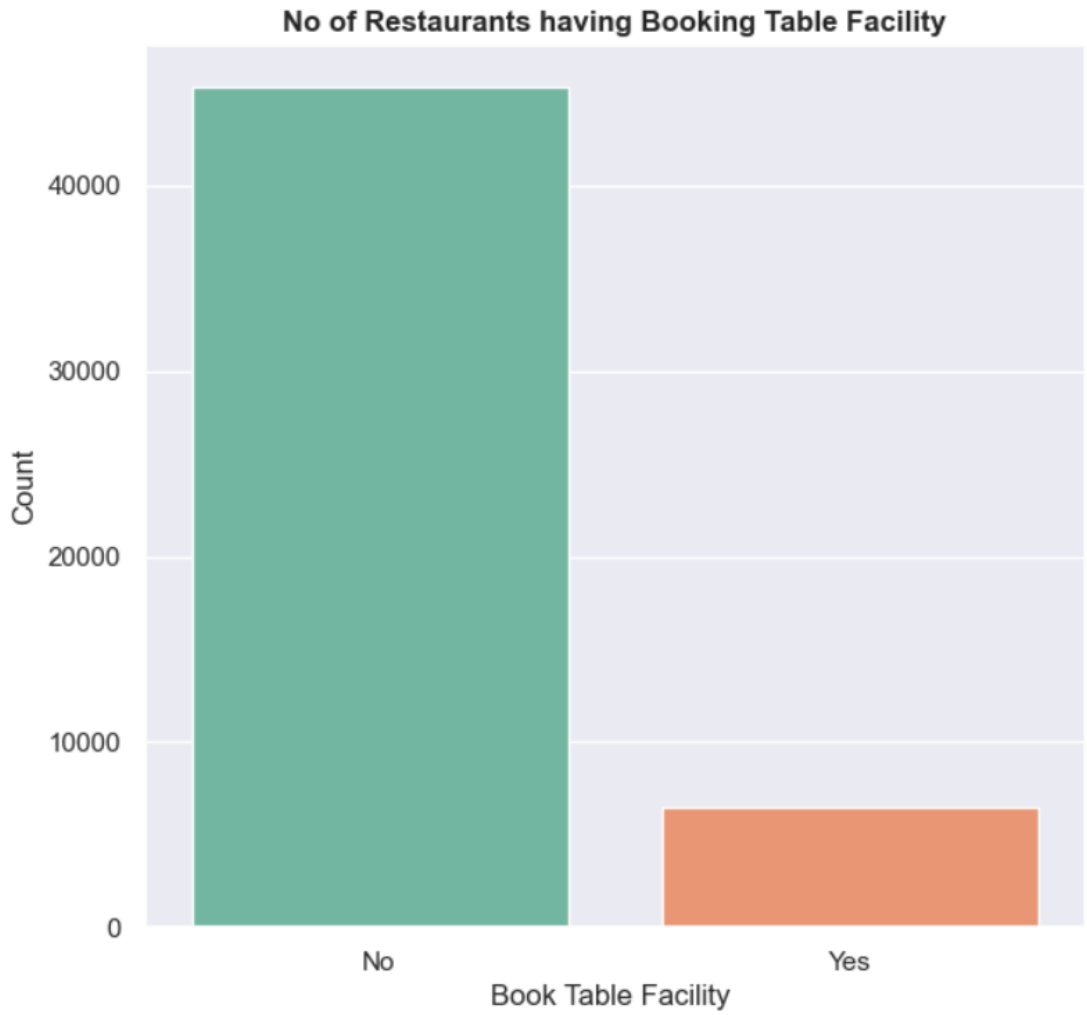
INFERENCE:

- We Can Say that Majority of the Restaurants has online Order facility i.e 30444 restaurants has Online Order Facility.

3. Restaurants Having Book Table Facility

PURPOSE

- To Know the Count of Restaurants Which are accepting Booking Table Facility



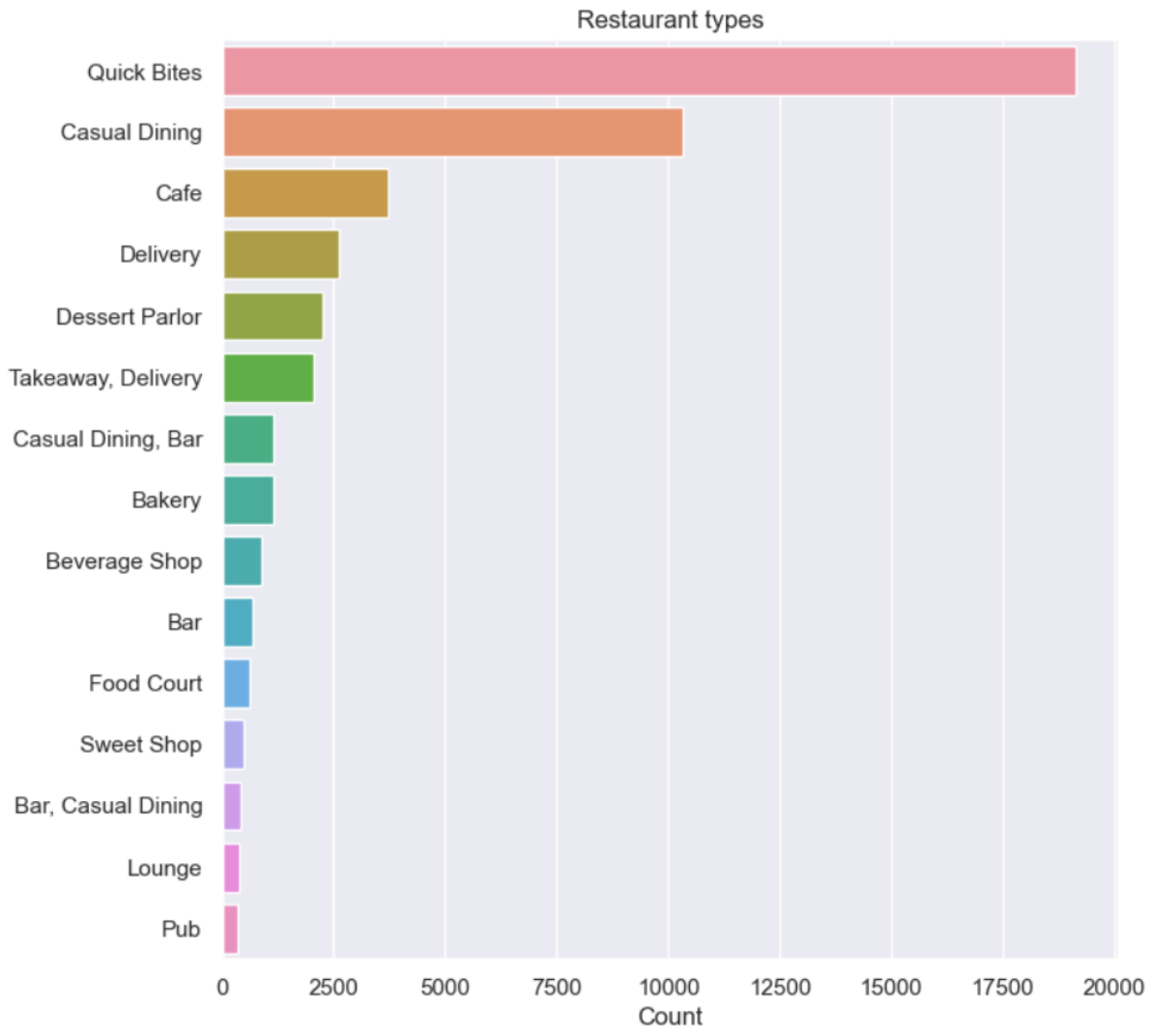
INFERENCE:

- We Can Say that Majority of the Restaurants Doesn't have book table facility i.e only 6449 restaurants has Book Table Facility.

4. Most Popular Restaurant Types in Bangalore

PURPOSE

- To Know the top 15 Most popular types of restaurant types in Bangalore



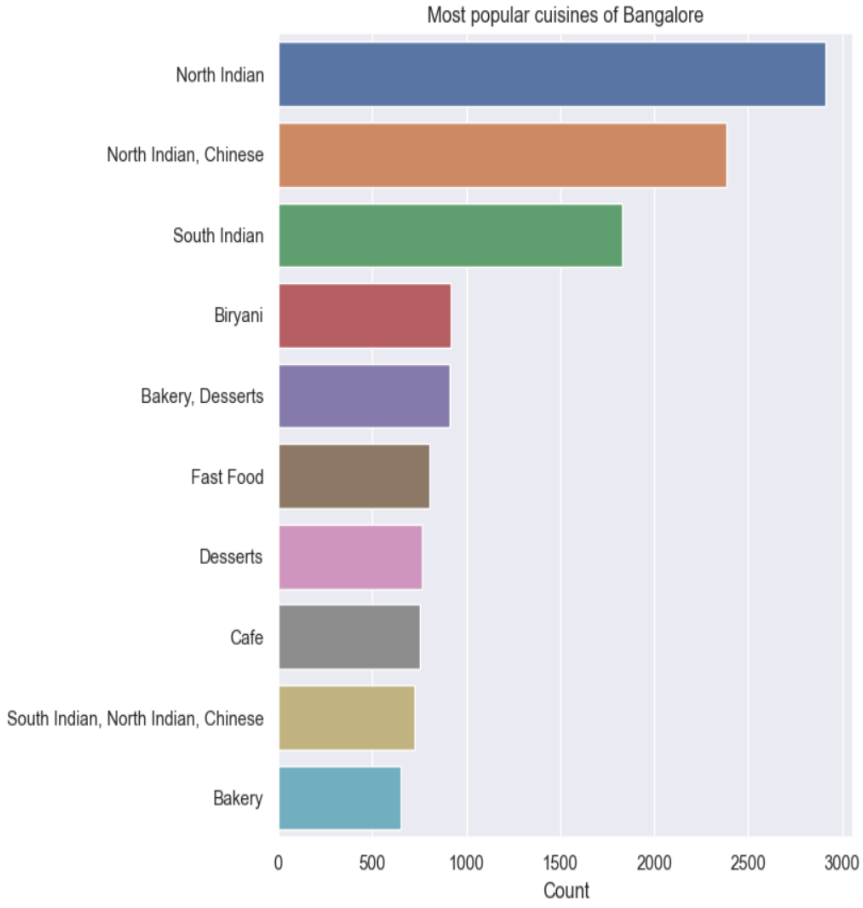
INFERENCE:

- We can Observe that Quick Bytes, Casual dining and Cafe are some top 3 Popular Restaurant Types

5. Most Popular Cuisines in Bangalore

PURPOSE

- To Know the Most Popular cuisines in Bangalore with respect to the count of restaurants.



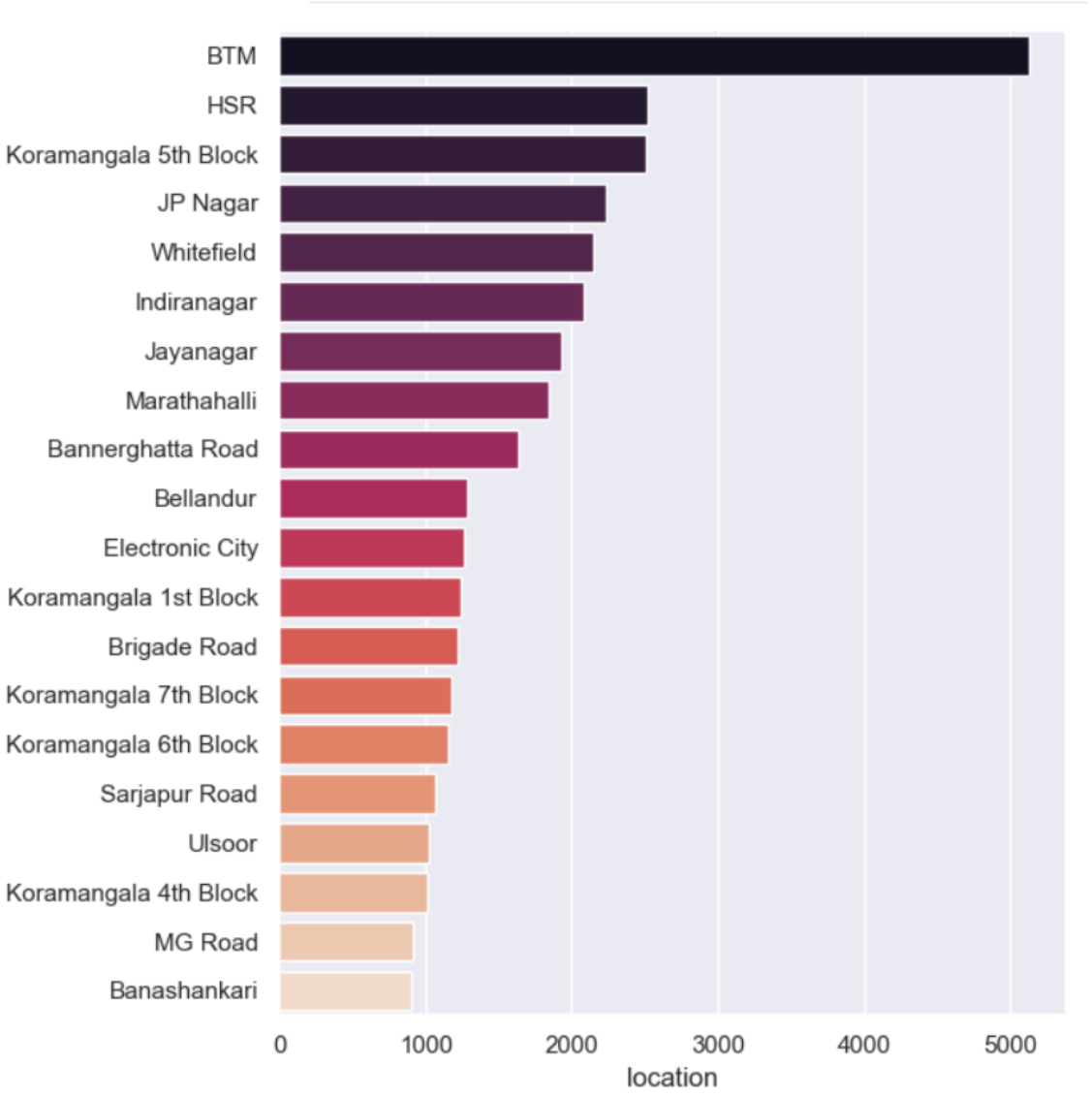
INFERENCE:

- We Can Say that Most popular Restaurants cuisines are of North Indian, Chinese, South Indian , Biryani , Desserts etc.

6. Foodie Areas in Bangalore

PURPOSE

- To Know the Most Popular Foodie Areas according to the restaurant counts in Bangalore.



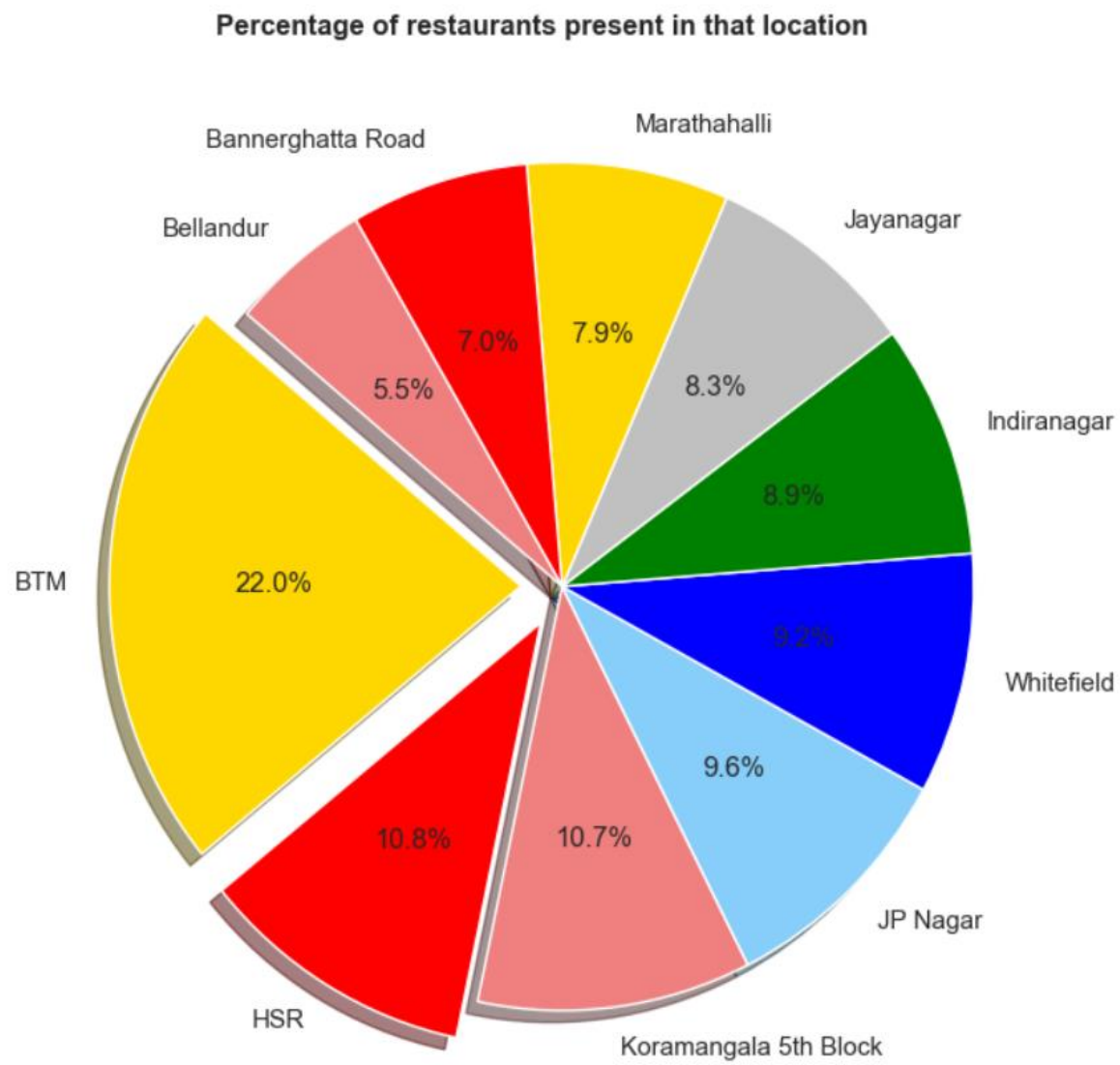
INFERENCE:

- By the above graph we can say that BTM , HSR , Karmangala 5th block are some of the top 3 popular foodie areas with respect to their counts in Bangalore

7. Percentage of Restaurants in Foodie Areas in Bangalore

PURPOSE

- To Know the Most Popular Foodie Areas according to the restaurant counts and their percentage.



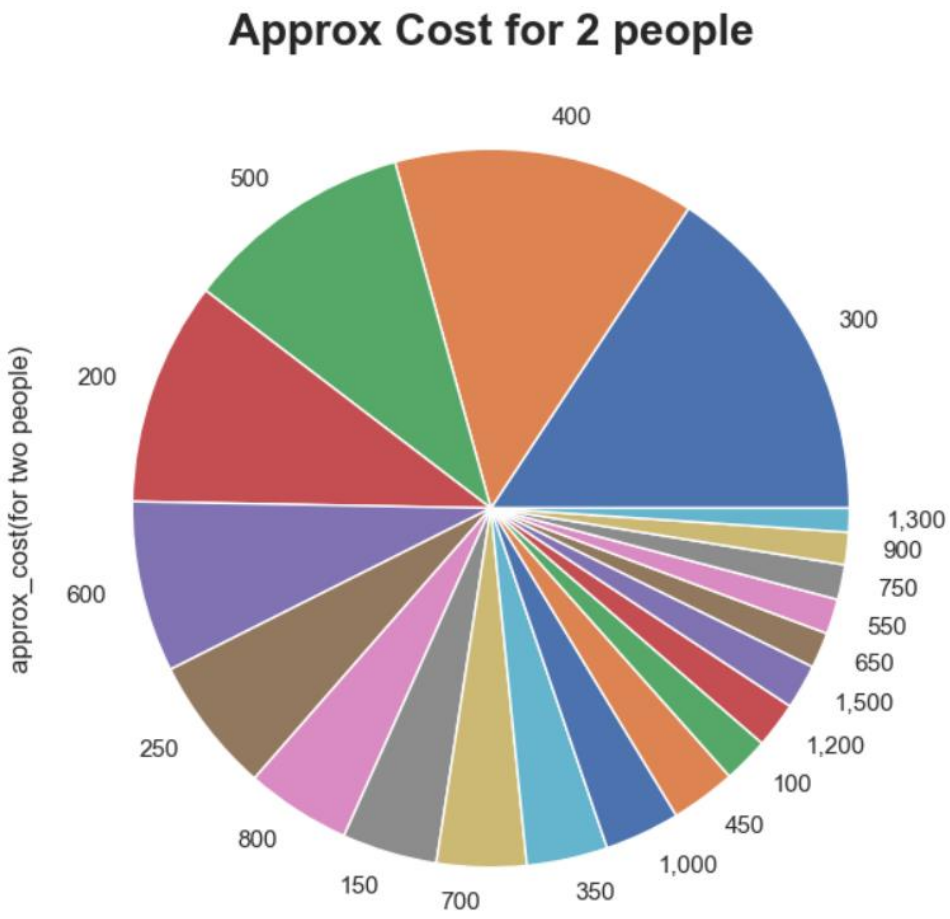
INFERENCE:

- By the above graph we can say that BTM with 22.0%, HSR with 10.8%, Karmangala 5th block with 10.7 % are some of the top 3 popular foodie areas with respect to their percentages in Bangalore

8. Average Cost for Two People

PURPOSE

- To Know the Average Cost for Two People with percentage.



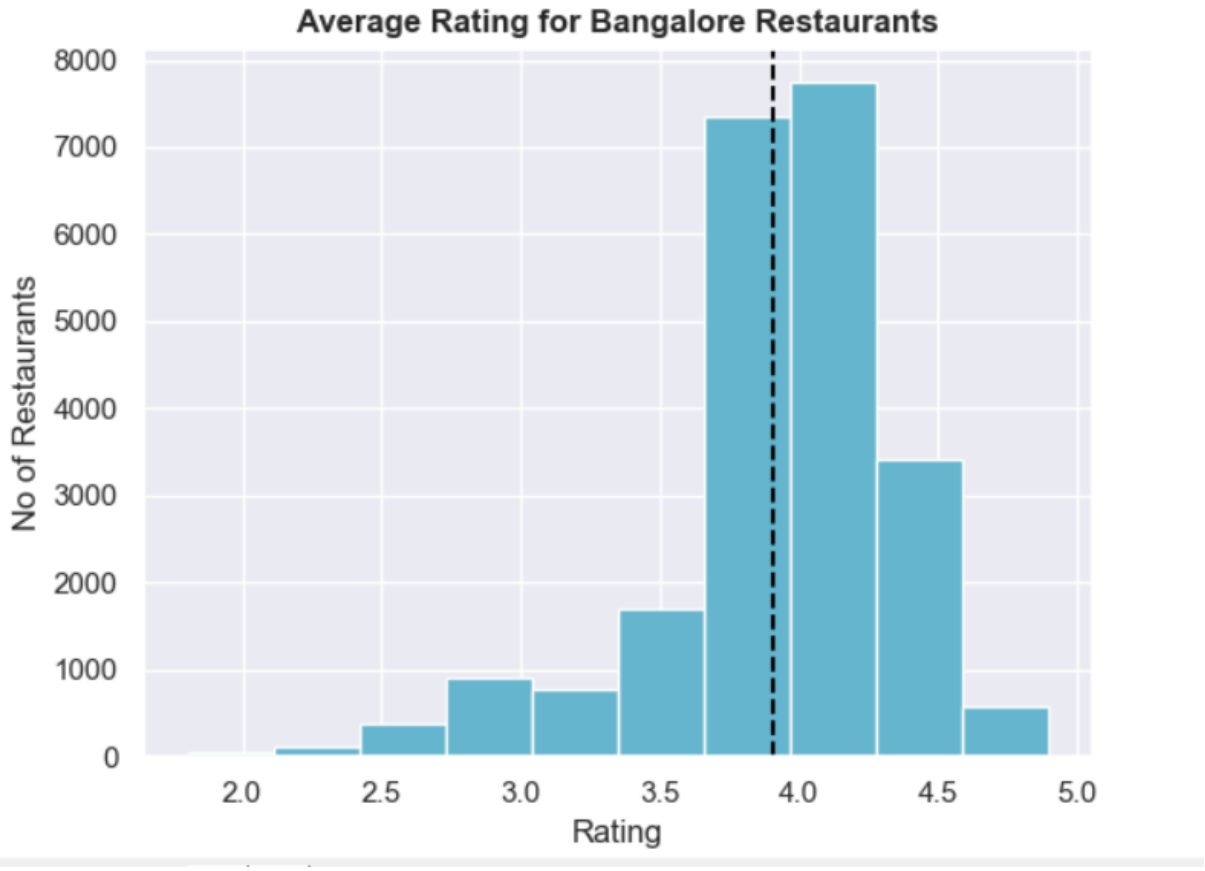
INFERENCE:

- By the above chart there is 17.86 percentage chances that for two persons the approximate cost will be 400 and 17.04 % chance that the cost will be 300 and so on.

9. Average Rating Per Restaurant in Bangalore

PURPOSE

- To Know the Average rating per restaurant and count of the restaurants having the same average rating in Bangalore



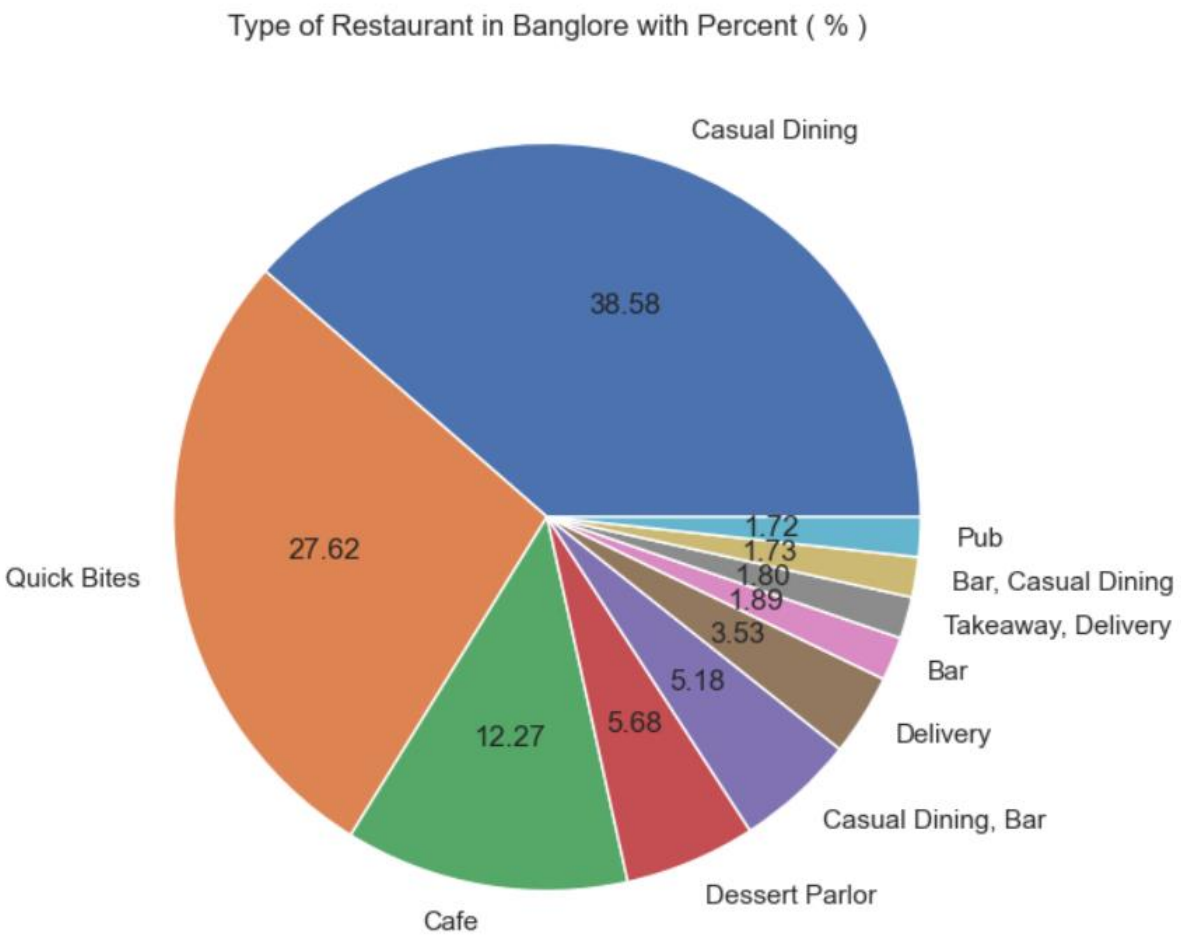
INFERENCE:

- The Average Rating for Bangalore Restaurants is 3.9 approximately.

10. Type Of Restaurants and Percentage Bangalore

PURPOSE

- To Know about the type of restaurants along with the percentage of its type in Bangalore.



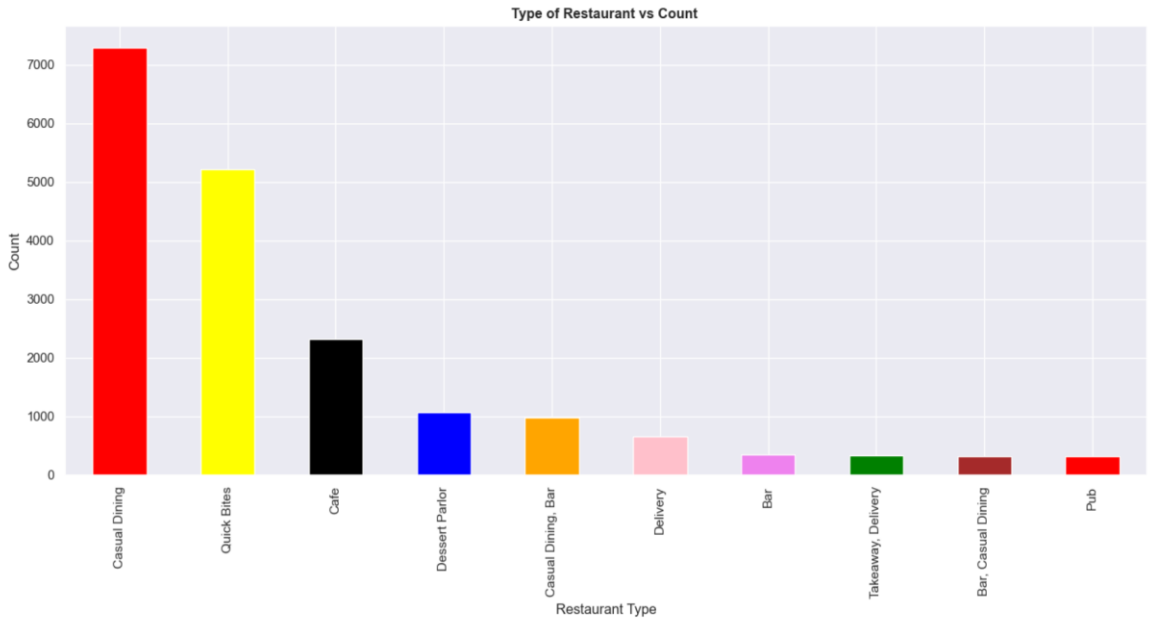
INFERENCE:

- The Restaurant type of casual Dining are of 38.58%, Quick bite are of 27.62 %, Café are of 12.27 % in Bangalore.

11. Type Of Restaurant and Count in Bangalore

PURPOSE

- To Know the count of different types of restaurants available in bangalore.



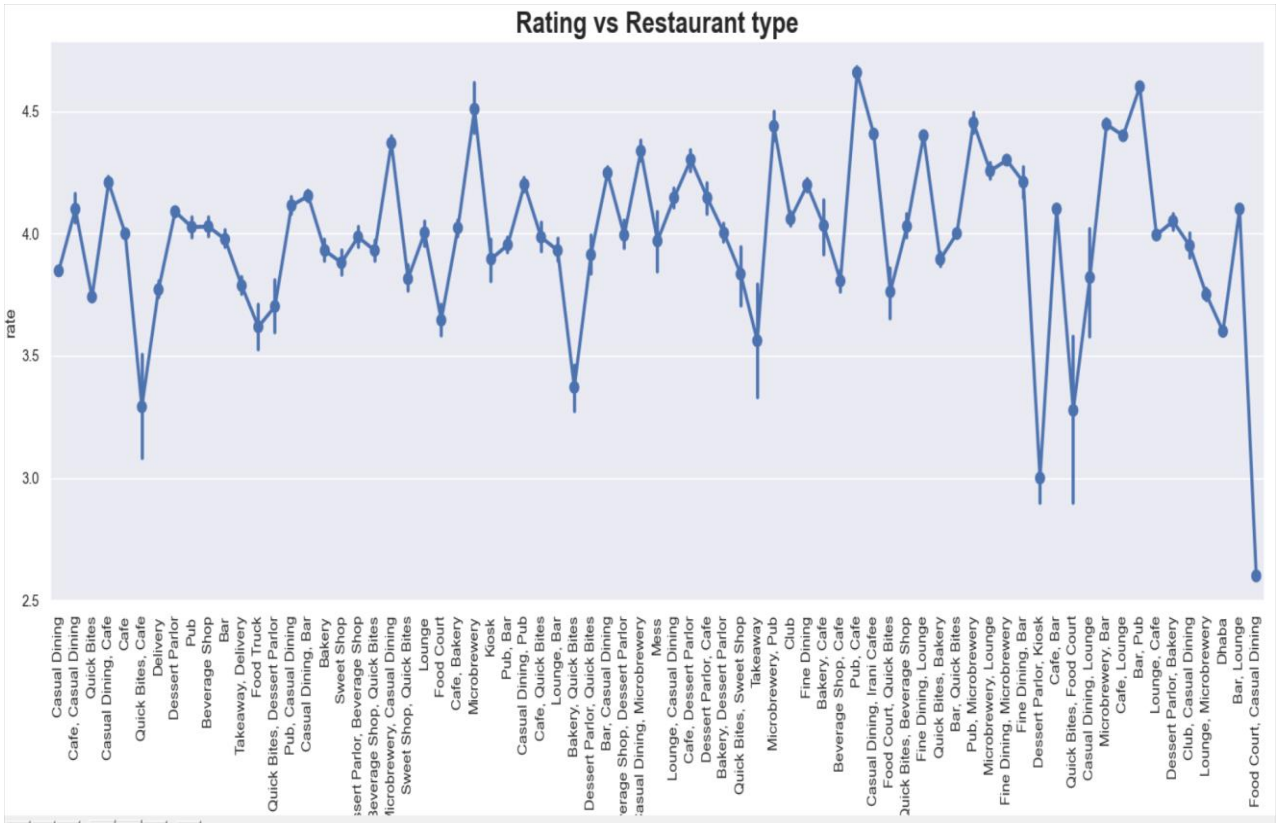
INFERENCE:

- From the above graph we can infer that more than 7000 restaurants are of casual dining type and below 1000 restaurants are of pub type.

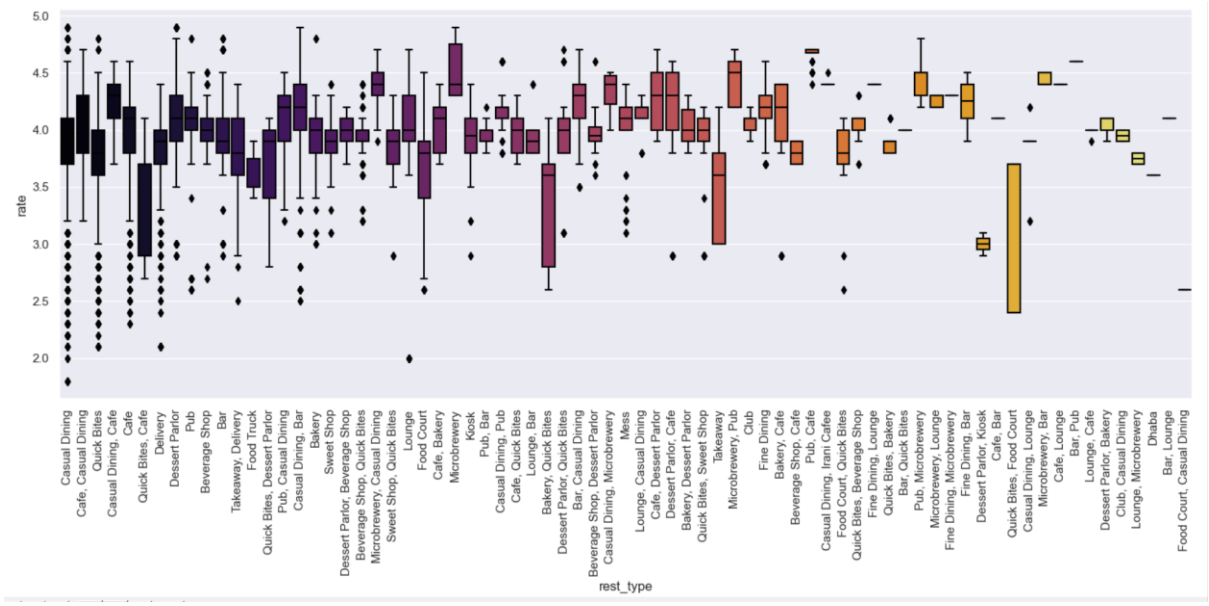
12. Ratio Between Rating and Restaurant Type in Bangalore.

PURPOSE

- To Know the ratings for different type of restaurants.



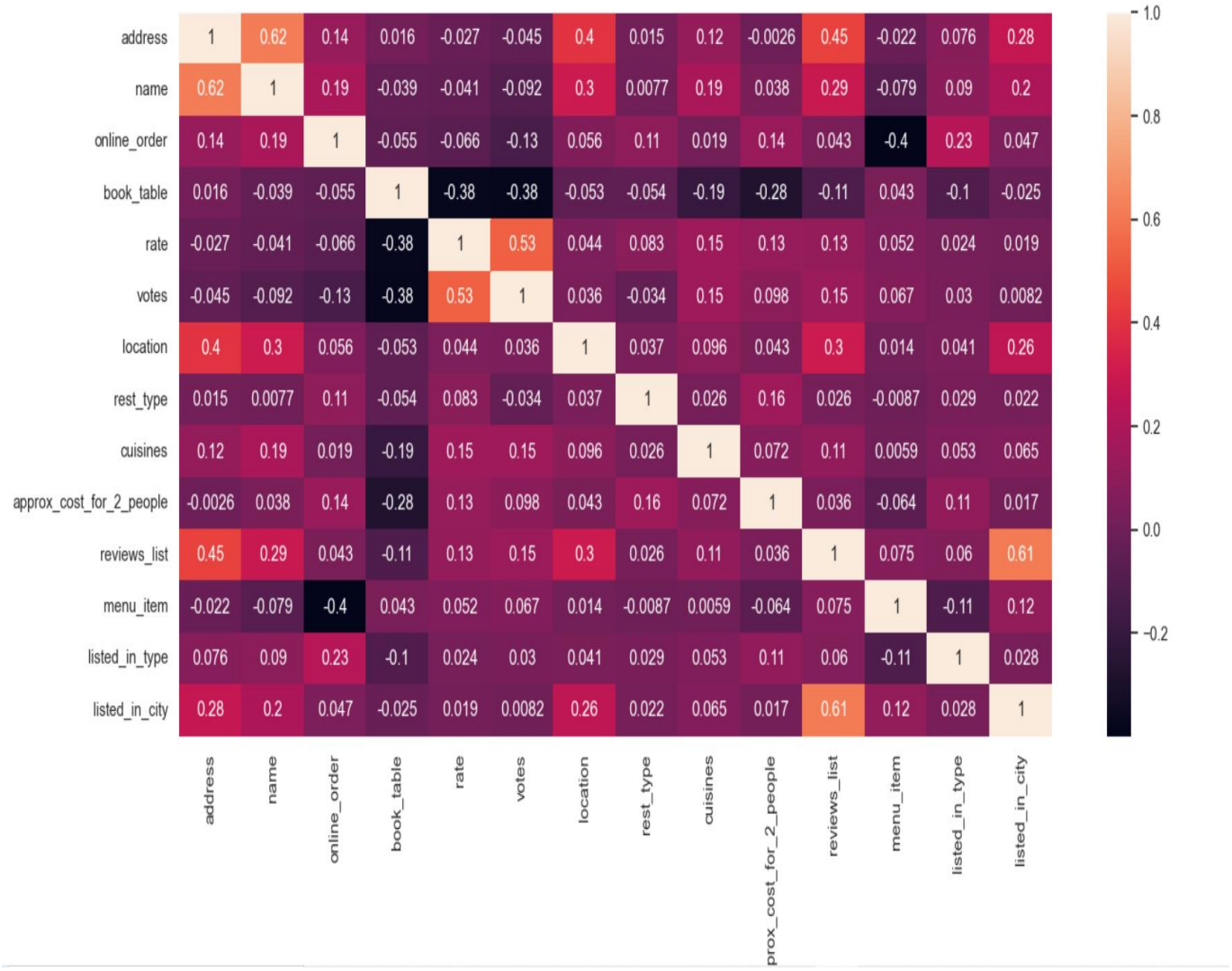
BOX PLOT FOR RATING Vs RESTAURANT TYPE



INFERENCE:

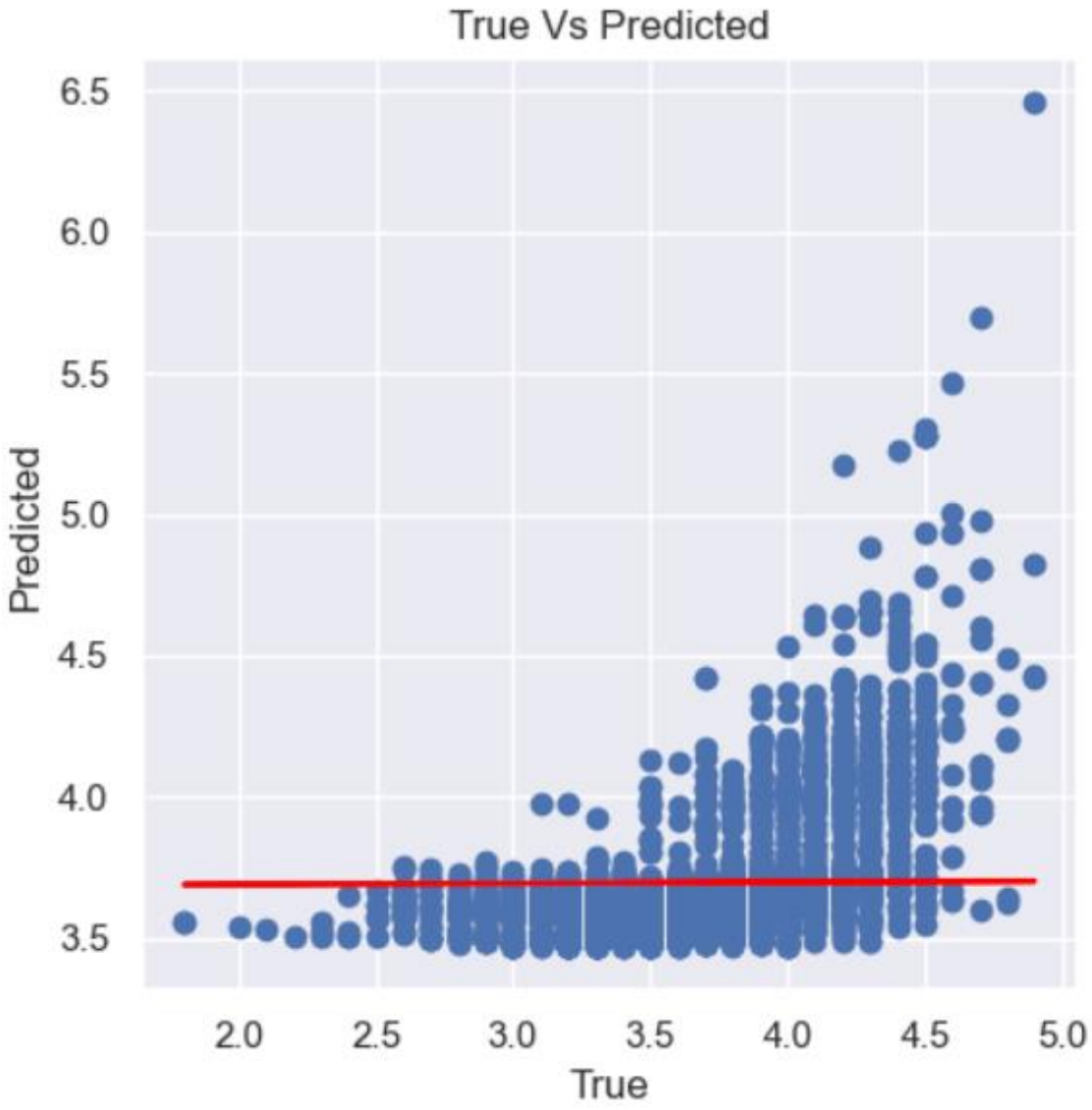
- From the above plot we can infer that, the peak value i.e., the rating of above 4.5 is obtained for the restaurant type of PUB, Café and lowest rating i.e. below 3.0 is obtained for food courts etc.

CORRELATION BETWEEN DIFFERENT VARIABLES



The highest correlation is between name and address which is 0.62 which is not of very much concern Splitting dataset into train & test

LINEAR REGRESSION ANALYSIS:



By applying Linear Regression, we get

m and c of True Vs Predicted are 0.003861819559095663, 3.682772868374157

R- Square Value: 0.2914084406705565

We get 29% accurate by applying linear regression for the training and predicted values.

PREDICTION OF RATING USING DECISION TREE REGRESSION

R-squared is a statistical measure of how close the data are to the fitted regression line. It is also known as the coefficient of determination, or the coefficient of multiple determination for multiple regression.

The definition of R-squared is fairly straight-forward; it is the percentage of the response variable variation that is explained by a linear model. Or:

R-squared = Explained variation / Total variation

R-squared is always between 0 and 100%:

- 0% indicates that the model explains none of the variability of the response data around its mean.
- 100% indicates that the model explains all the variability of the response data around its mean.

After Applying decision tree regression, we get R-Squared value as 0.8458051293469493.

Which Indicates the model predicts 84 % accurate with the help of by applying decision tree regression.

PREDICTION OF RATING USING RANDOM FOREST REGRESSION

R-squared is a statistical measure of how close the data are to the fitted regression line. It is also known as the coefficient of determination, or the coefficient of multiple determination for multiple regression.

The definition of R-squared is fairly straight-forward; it is the percentage of the response variable variation that is explained by a linear model. Or:

R-squared = Explained variation / Total variation

R-squared is always between 0 and 100%:

- 0% indicates that the model explains none of the variability of the response data around its mean.
- 100% indicates that the model explains all the variability of the response data around its mean.

After Applying decision tree regression, we get R-Squared value as 0.8684086685577226

Which Indicates with the help of random forest regression the predicted rating for the restaurants is 86.8 % accurate.

INFERENCE:

- After applying the above three regressions for true and predicted rating by considering the attributes, we can say that among the three regressions prediction of rating using random forest regression produces better accurate result of 86% when compared to decision tree regression and Linear regression.

CONCLUSION:

The Zomato dataset was analysed to gain some insights of the food industry and food culture in Bangalore. We deduce some great results like which city has the greatest number of restaurants and hence highest number of foodies, we also saw that which city has high number of well-rated restaurants, what all cuisines are mostly preferred by people etc. This will help in future business persons to start their business.

FUTURE SCOPE:

For my future work I plan to :

- Implement machine training to predict which location is best for food business.
- Combine this dataset with specific data about different foods and people that visiting the restaurant for evaluating daily report.
- Making my own dataset by crawling data from website.
- Learn more about data sciences in general to improve my understanding.

REFERNCES:

1. https://www.kaggle.com/himanshupoddar/zomato-bangalore-restaurants
2. **https://www.ijmtst.com/volume6/issue08s/24.IJMTSTCIET69.pdf**