**Introduction to Project :**

Amidst COVID-19 pandemic, one of the most grown business companies in India has been Zomato. The Zomato founder said that the revenue of the company grew 105 per cent in FY20 as compared to FY19, while costs only grew 47 per cent in the corresponding period.People being not able to eat their favourite savoury foods due to strict lockdown instructions made delivery boys to go and fetch their food for the same from their loved restaurants.The reason for selection of the project is the same one for visualizing their data as to what aspects of this company services does attract people and thus helping indian food chain of restaurants to scale their business in the same manner.

**Problem Definition :**

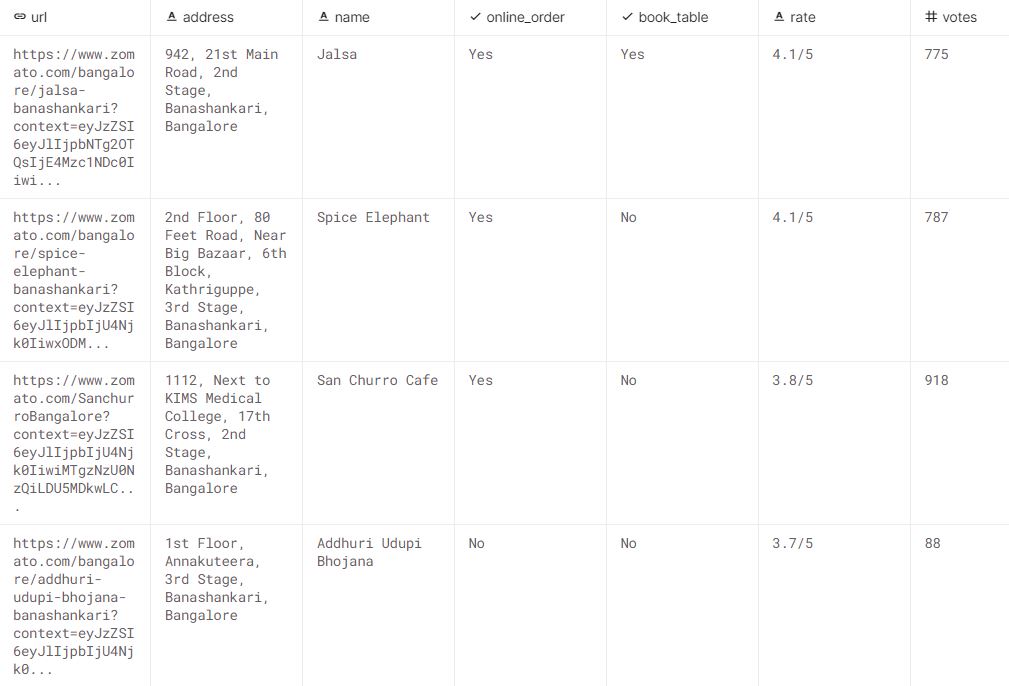
The basic idea behind Zomato Restaurants Analysis and Prediction project was to get a fair idea about the factors affecting the aggregate rating of each restaurant, establishment of different types of restaurant at different places, Bengaluru being one such city has more than 12,000 restaurants with restaurants serving dishes from all over the world. With each day new restaurants opening the industry hasn't been saturated yet and the demand is increasing day by day. In Spite of increasing demand it has become difficult for new restaurants to compete with established restaurants. Most of them serve the same food. Bengaluru being an IT capital of India. Most of the people here are dependent mainly on the restaurant food as they don't have time to cook for themselves. 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? Does the entire locality love vegetarian food? If yes then is that locality populated by a particular sect of people for eg. Jain, Marwaris, Gujaratis who are mostly vegetarian. This kind of analysis can be done using the data, by studying different factors.

**Data Warehousing and Mining Tools Used :**

1. Jupyter Notebook
2. Colab Notebook
3. Orange Tool
4. mySQL database server

**Dataset Used :**

<https://www.kaggle.com/himanshupoddar/zomato-bangalore-restaurants/>



(51717 rows X 17 columns)

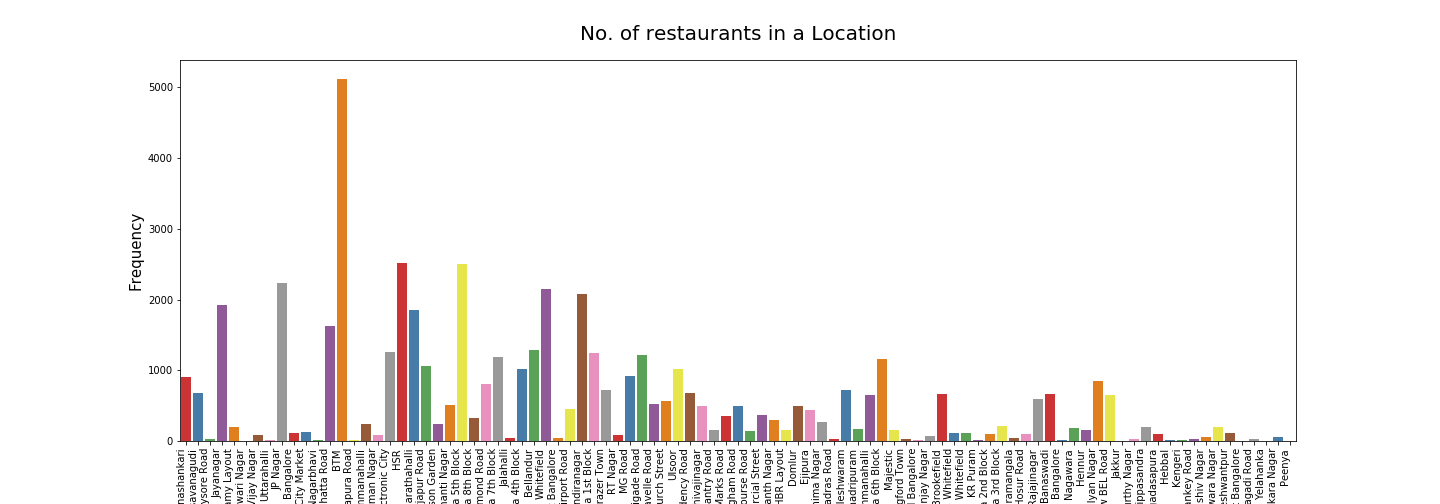
**Methodology Employed / Algorithms Implemented :**

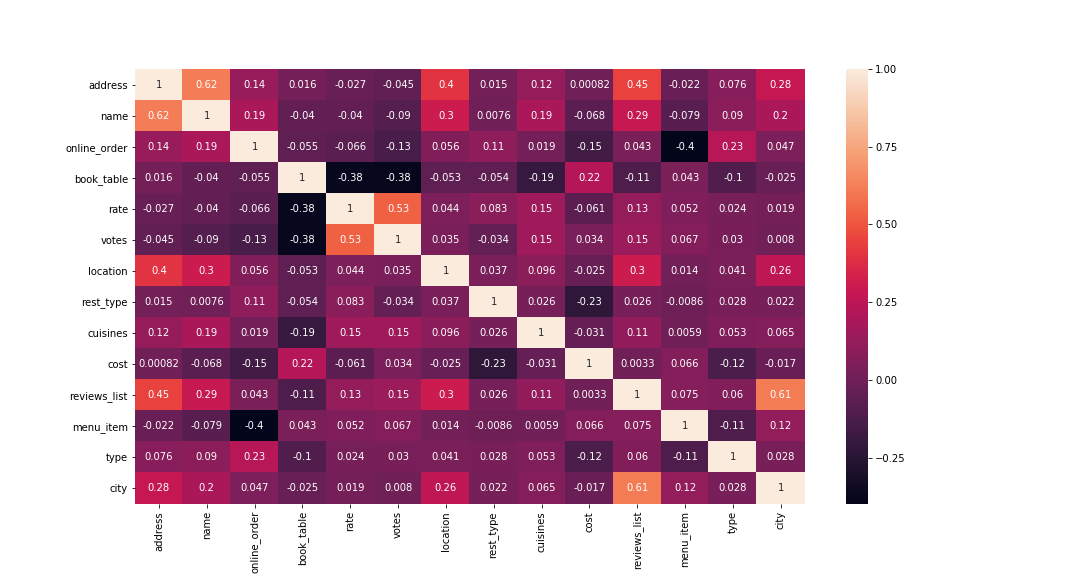
1. Linear Regression
2. Decision Tree Regression
3. Random Forest Regression
4. Extra Tree Regression

**Flow Diagram :**

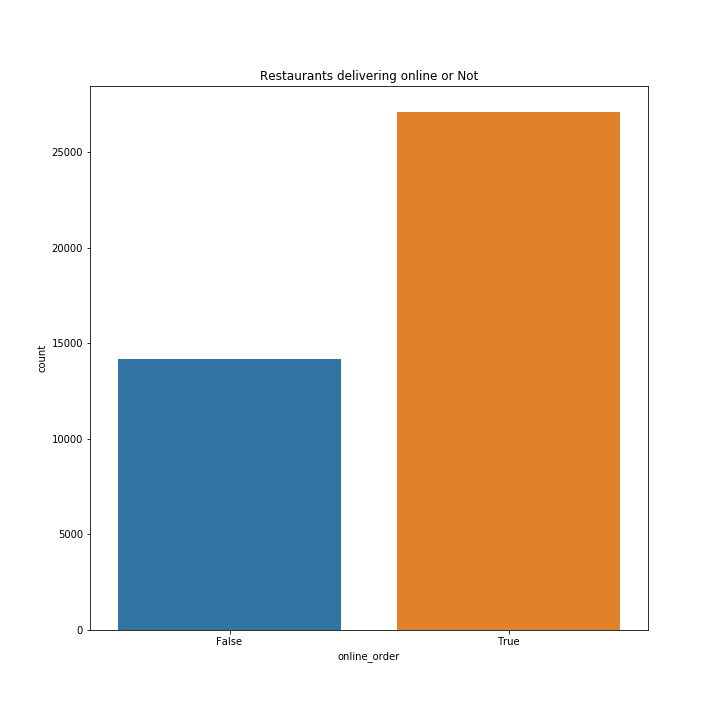
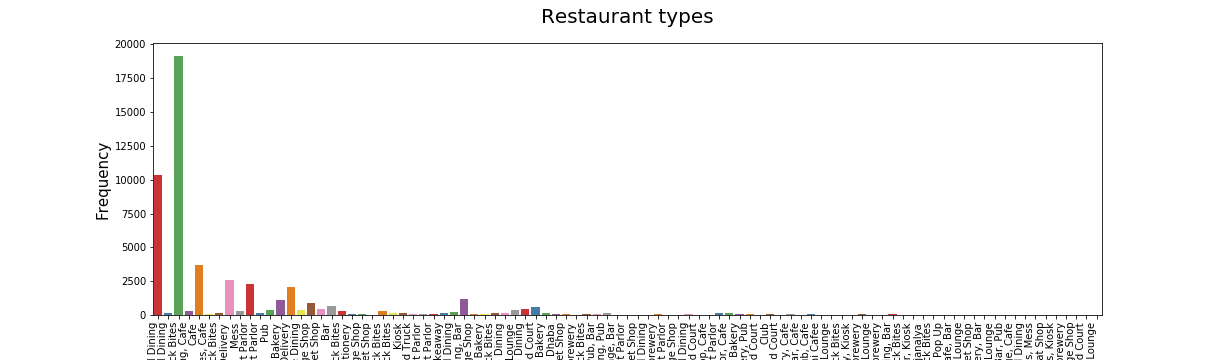
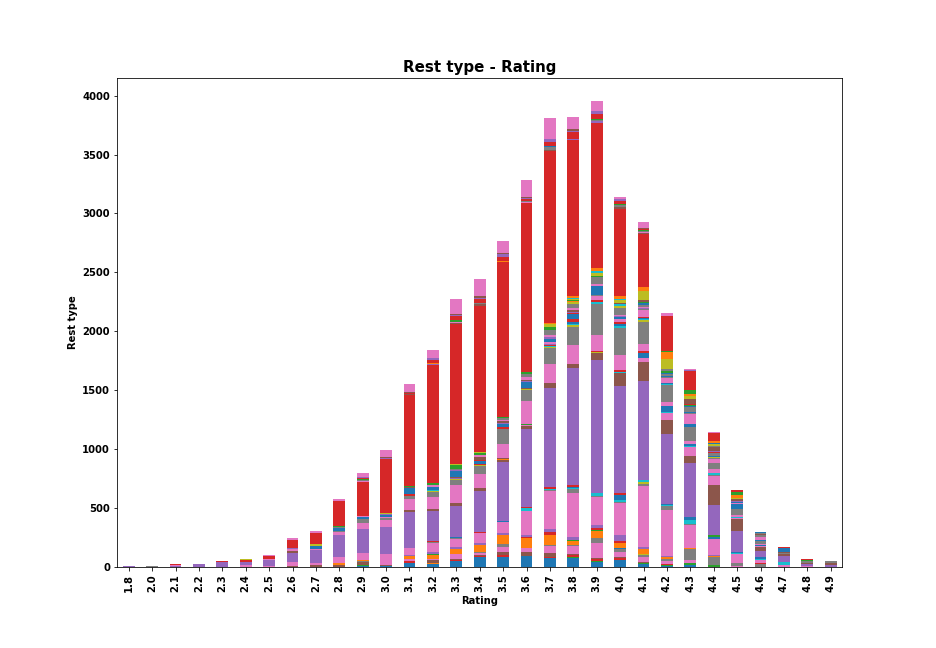
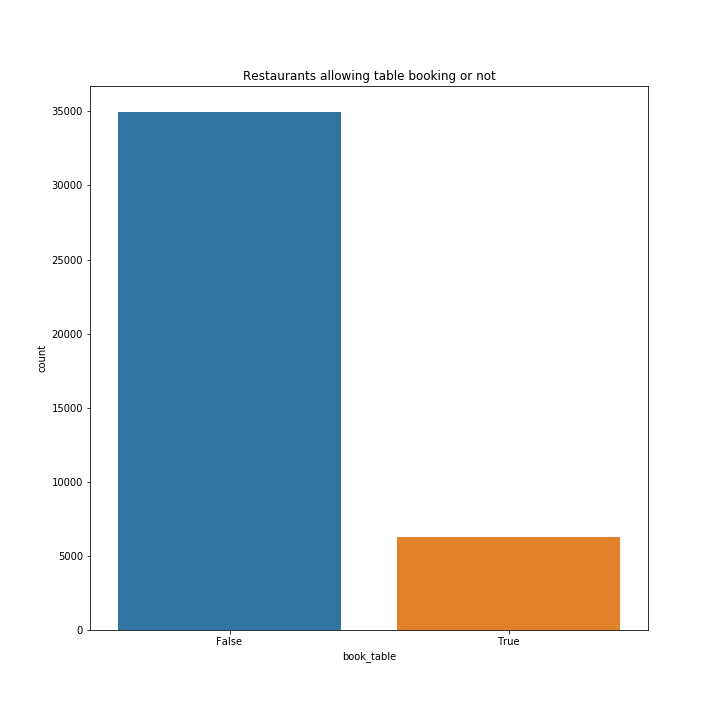
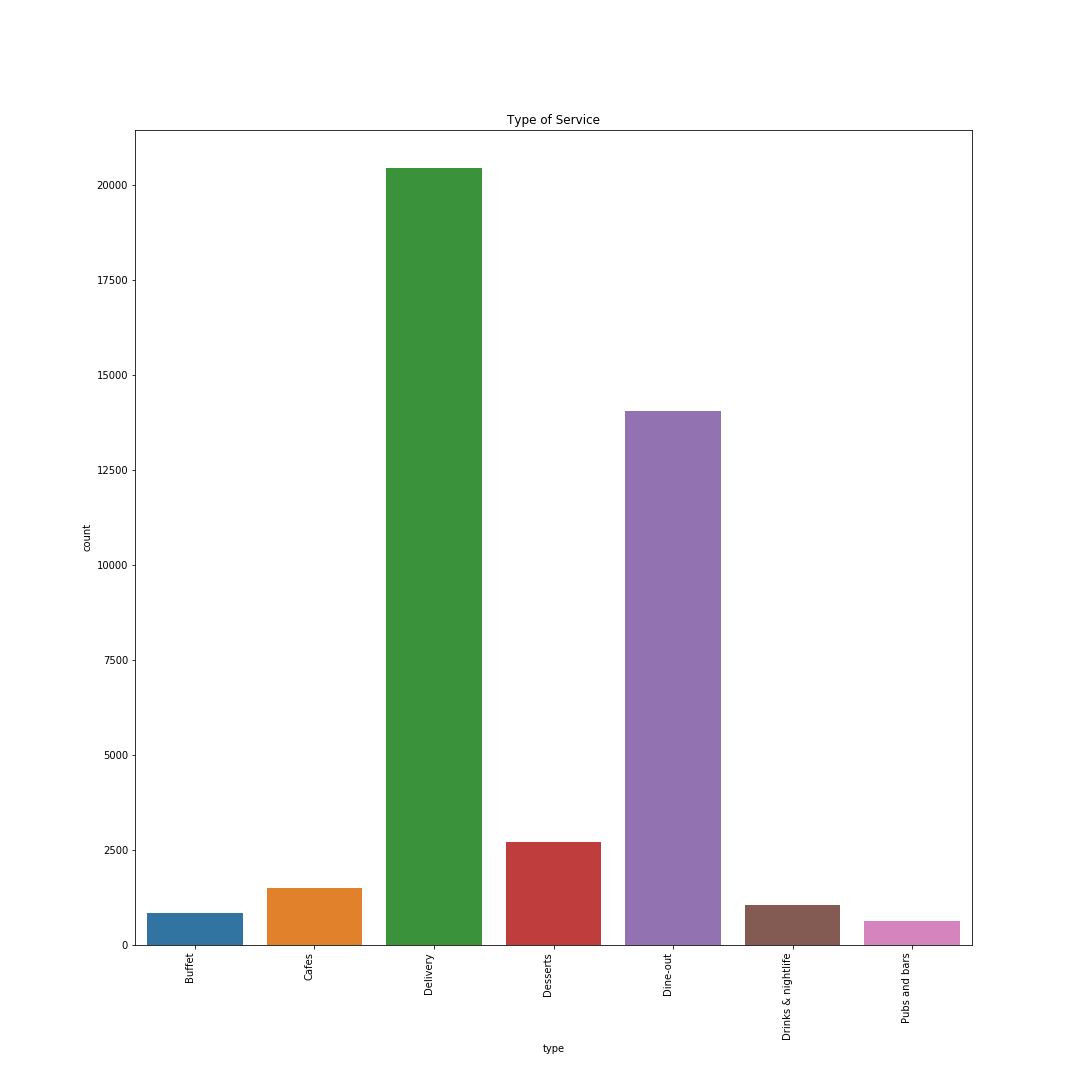
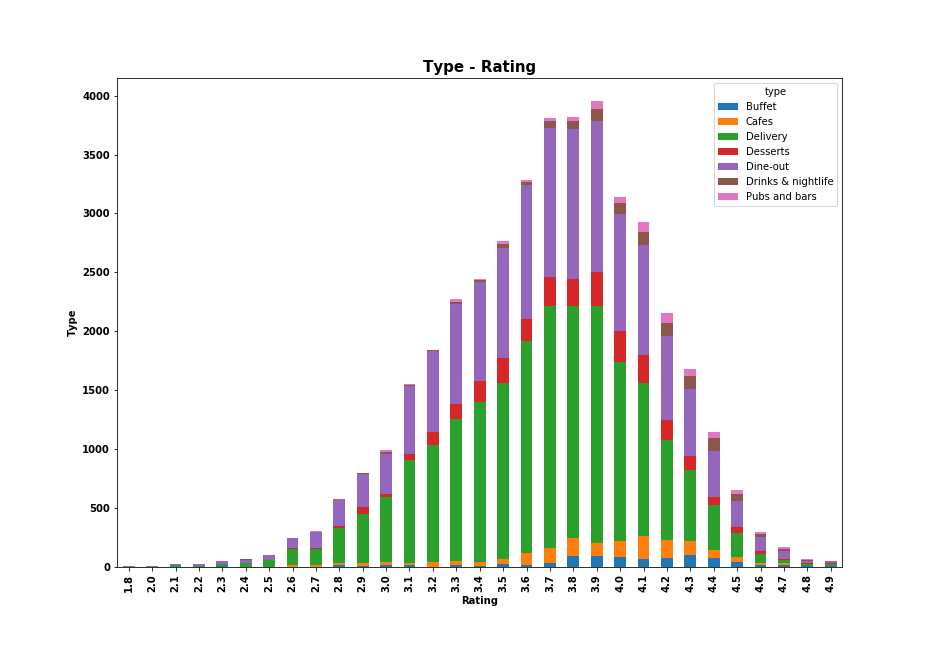
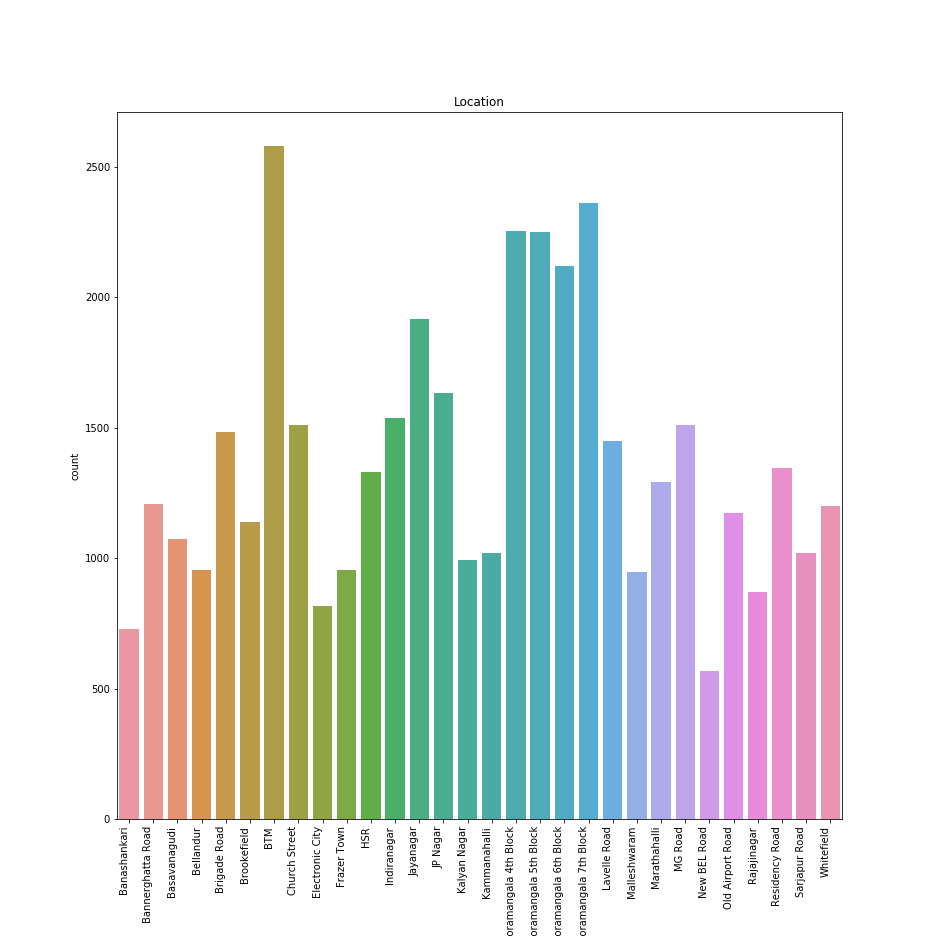
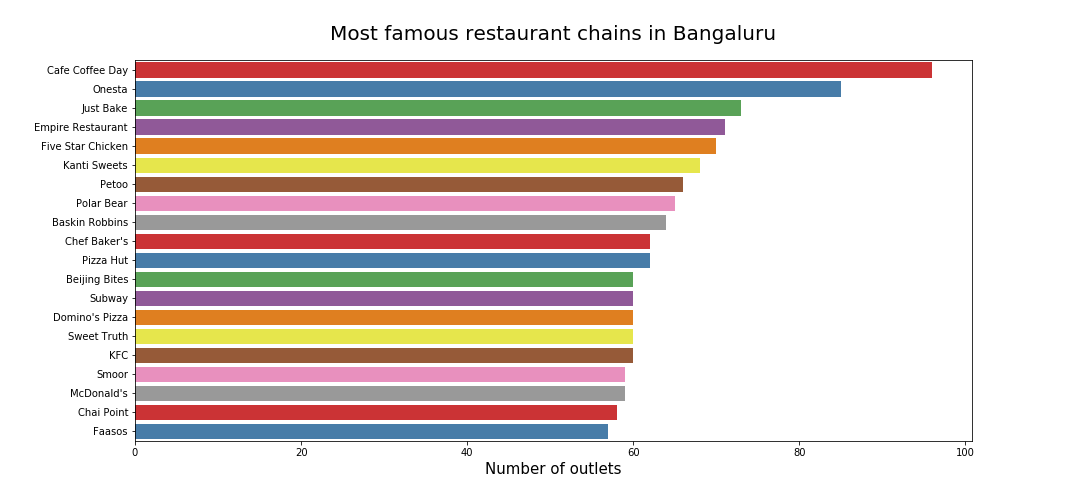
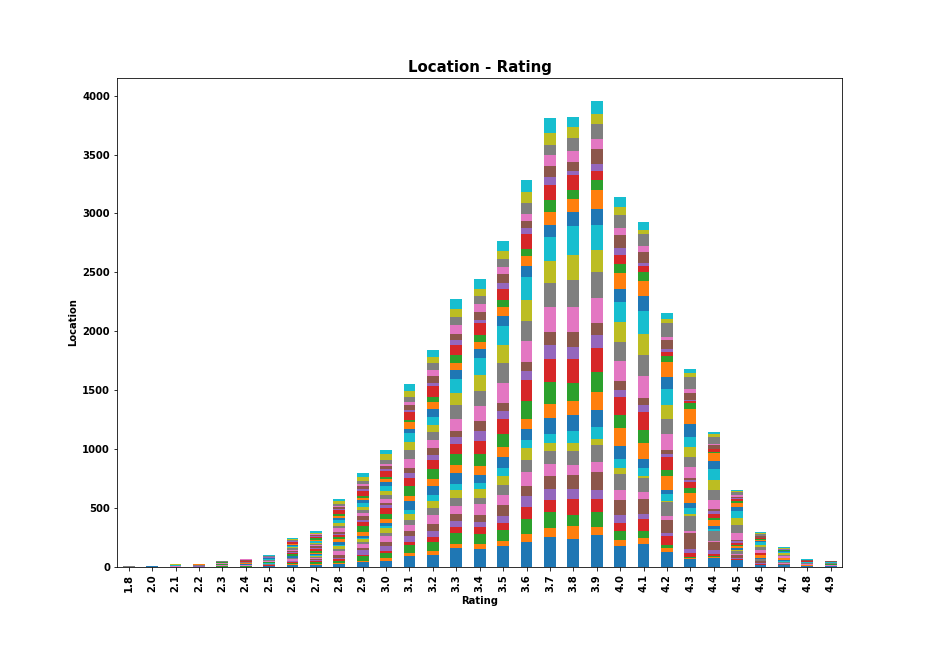
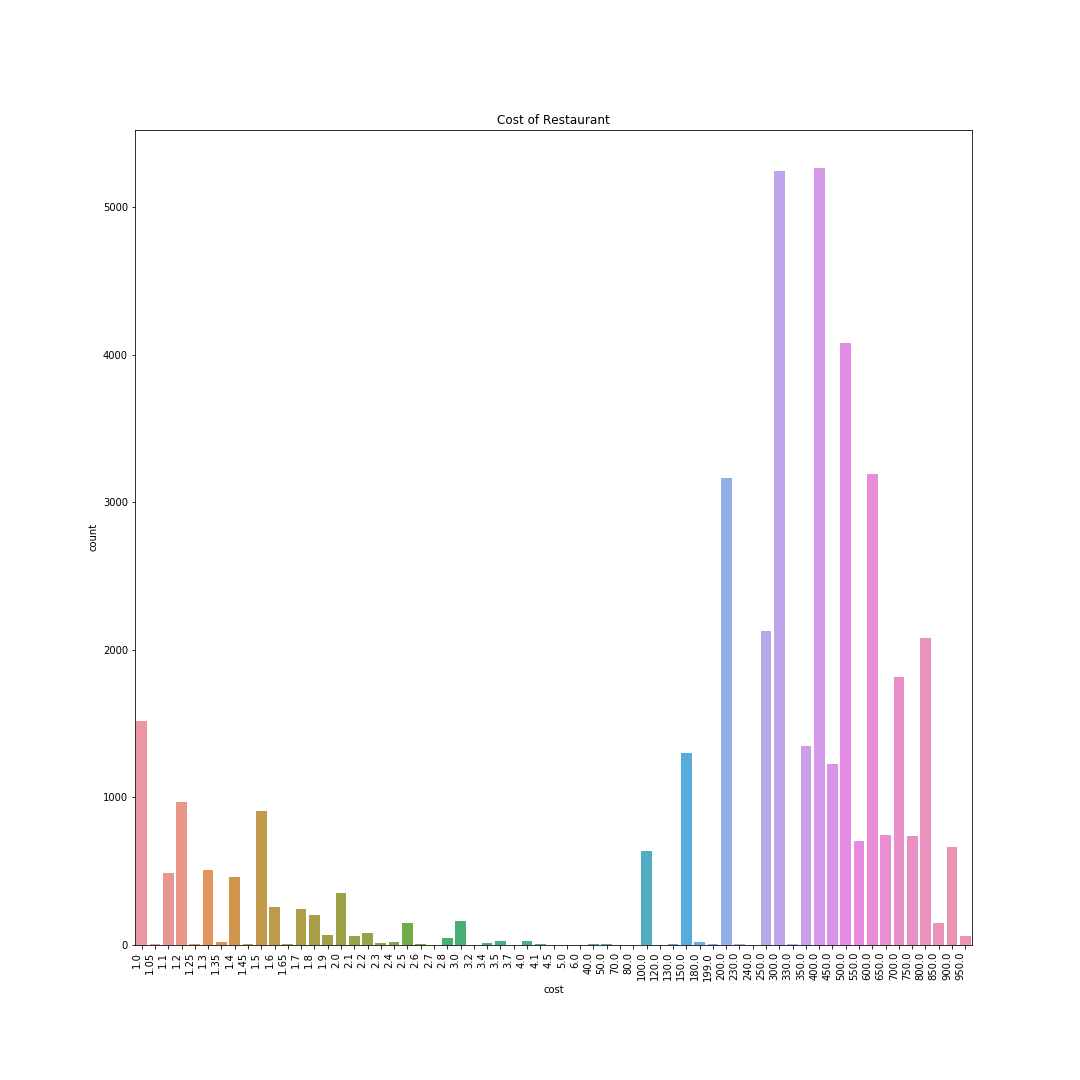
**Implementation Details :**

1. Loading the dataset - Load the data and import the libraries.
2. Data Cleaning -
   * Deleting redundant columns.
   * Renaming the columns.
   * Dropping duplicates.
   * Remove the NaN values from the dataset
   * Some Transformations(Changing data types of some columns)
   * Removing '/5' from Rates
   * Encode the input Variables
   * Get Correlation between different variables
   * Defining the independent variables and dependent variables
3. Regression Analysis -
   * Linear Regression
   * Decision Tree Regression
   * Random Forest Regression
   * Extra Tree Regression
4. Data Visualization - Using plots to find relations between the features.
   * Restaurants delivering Online or not
   * Restaurants allowing table booking or not
   * Table booking Rate vs Rate
   * Best Location
   * Relation between Location and Rating
   * Restaurant Type
   * Gaussian Rest type and Rating
   * Types of Services
   * Relation between Type and Rating
   * Cost of Restaurant
   * No. of restaurants in a Location
   * Restaurant type
   * Most famous restaurant chains in Bengaluru

**Results(Screenshots) :**

****

**Correlation between different variables**

****

**Comparison of evaluation measures using various algorithms :**

|  |  |
| --- | --- |
| MODEL | R2\_SCORE |
| Linear Regression | 0.27 |
| Decision Tree Regression | 0.85 |
| Random Forest Regression | 0.88 |
| Extra Tree Regression | 0.94 |

**Conclusion :**

Thus, we have effectively analysed the dataset and visualized according to different parameters and predicted the rating of restaurants based on those parameters.As a result, we have successfully studied various Data Warehousing and Mining concepts and applied algorithms to our problem statement Zomato Restaurants Analysis and Prediction on their dataset.

**References :**

1. <https://towardsdatascience.com/>
2. <https://www.kaggle.com/>
3. <https://scikit-learn.org/stable/>
4. <https://matplotlib.org/>
5. <https://seaborn.pydata.org/>
6. <https://pandas.pydata.org/>