Dr. Vishwanath Karad
MIT WORLD PEACE UNIVERSITY | PUNE
TECHNOLOGY, RESEARCH, SOCIAL INNOVATION & PARTNERSHIPS

Mini Project Report

On

Zomato Restaurant Analysis

By

Aniket Kumar – PF45

Manish Yadav – PD15

Rahul Arya – PE20

Under the guidance of

Prof. Shweta Joshi

MIT-World Peace University (MIT-WPU)

Faculty of Engineering

School of Computer Engineering & Technology

*** 2020-2021 ***

ABSTRACT

Online food delivery has become an inseparable part of our lives now. Every other we look to order something or the other to eat from Zomato. Being cautious and improve safety is not a simple instruction anymore. Every day there is a new offer on some or the other restaurant and we have so much to choose from now that we can't decide anything at all.

In our project we will be analyzing various things such as:-

a. Which franchise has the highest number of restaurants on Zomato?

b. How many Restaurants are accepting online orders?

c. How many Restaurants have a book table facility?

d. Which location has the highest number of Restaurants?

e. How many types of Restaurant types are there?

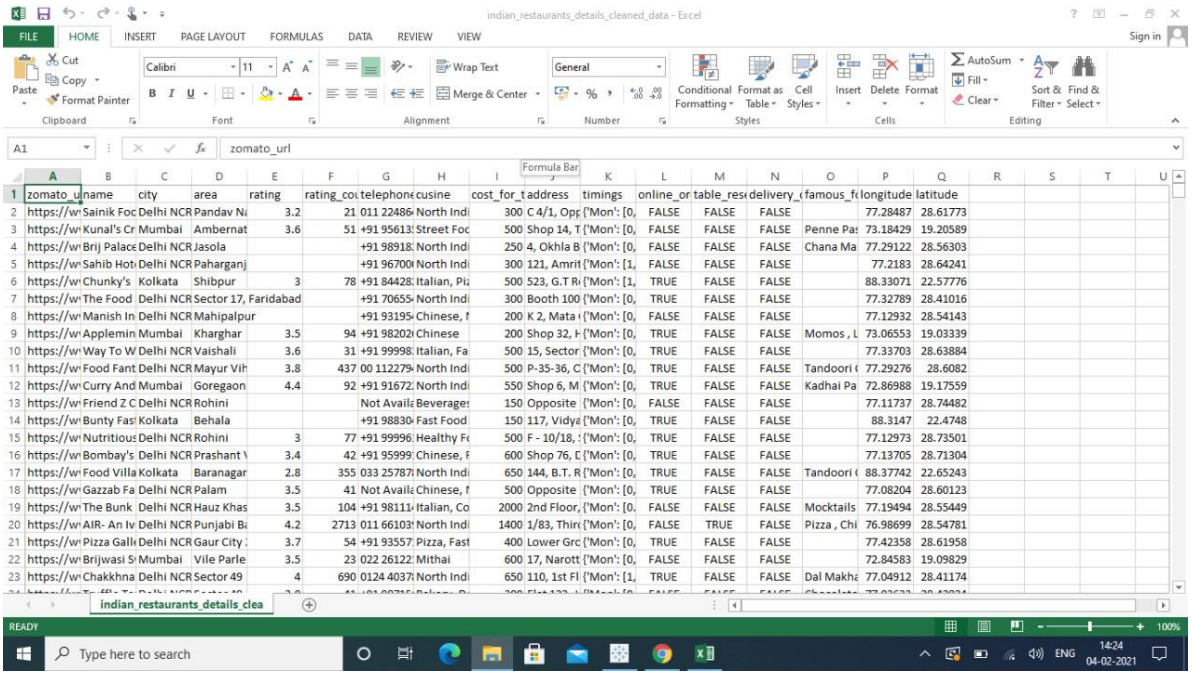
f. What is the most liked Restaurant type?

g. What is the Average cost for 2 persons at a particular restaurant?

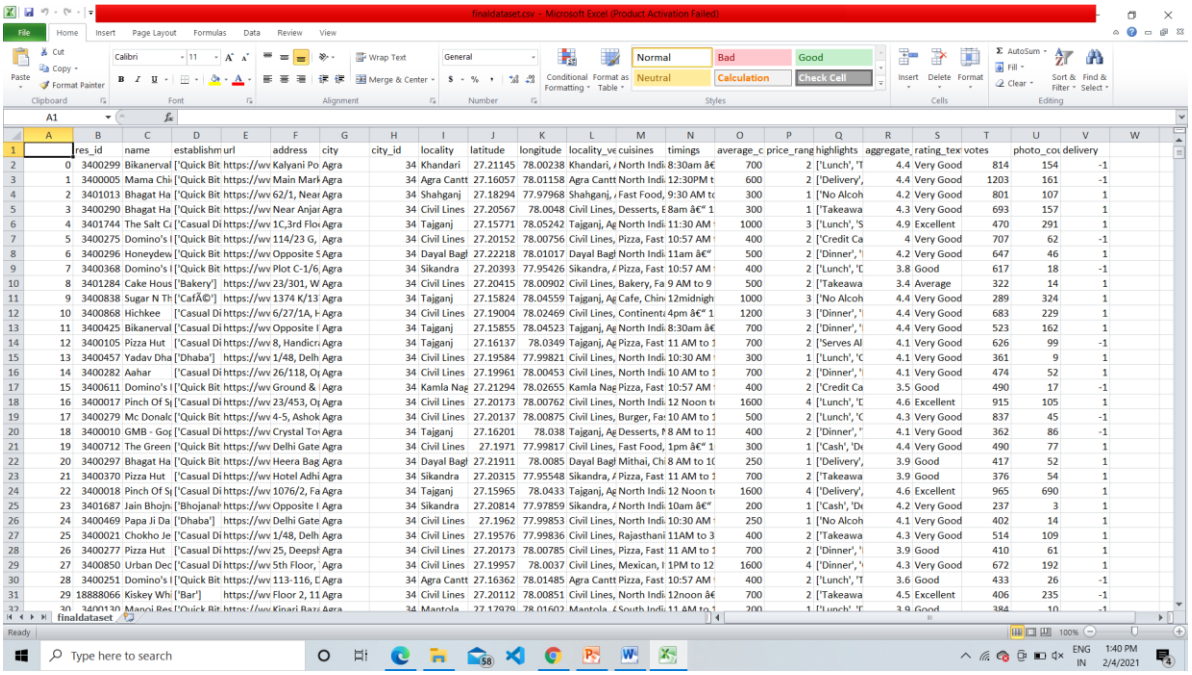
h. What is the most liked cuisine type?

- **Dataset details**

1. Name of Dataset: indian_restaurants_details
2. Dataset Size: 224520 Rows and 17 Columns (180 MB).
3. Dataset format: Comma separated value file (CSV).
4. Description: Data in this dataset belongs to "Zomato Media Private Limited". Basically the data has all the Zomato restaurants with their ratings, votes and other crucial data attributes to do research work.



Original Dataset



Dataset after Preprocessing

Column names of the dataset are:-

- zomato_url
- name
- city
- area
- rating
- rating_count
- telephone
- cuisine
- cost_for_two
- address
- timings
- online_order
- table_reservation
- delivey_only
- famous_food
- longitude
- latitude

Data Visualization using Python

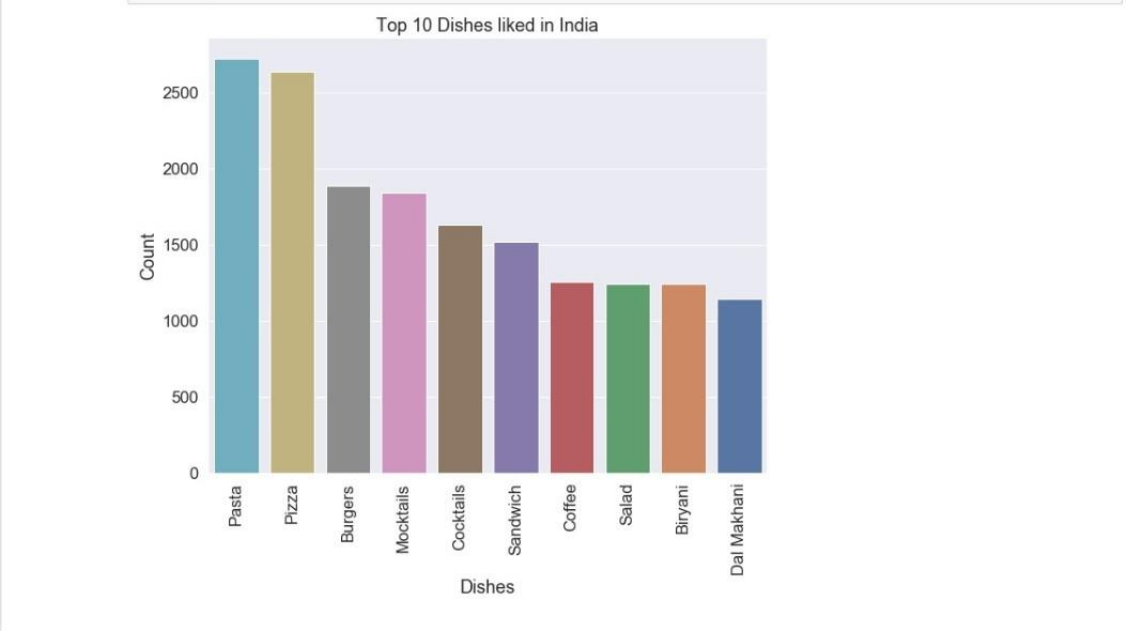
- Data visualization is the discipline of trying to understand data by placing it in a visual context so that patterns, trends and correlations that might not otherwise be detected can be exposed.

- Python offers multiple great graphing libraries that come packed with lots of different features. No matter if you want to create interactive, live or highly customized plots python has an excellent library for you.

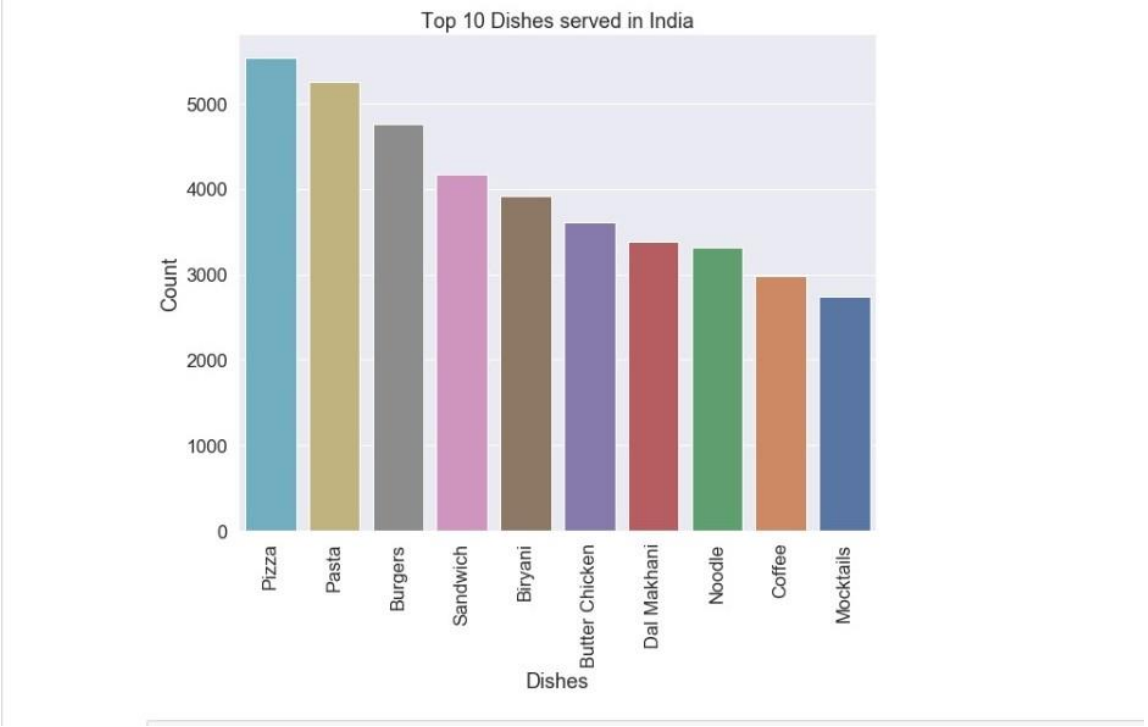
Bar Chart

A bar chart can be created using the bar method. The bar-chart isn't automatically calculating the frequency of a category so we are going to use pandas value_counts function to do this.

The graphs created by us are as follows:-



Top 10 Most liked dishes in India

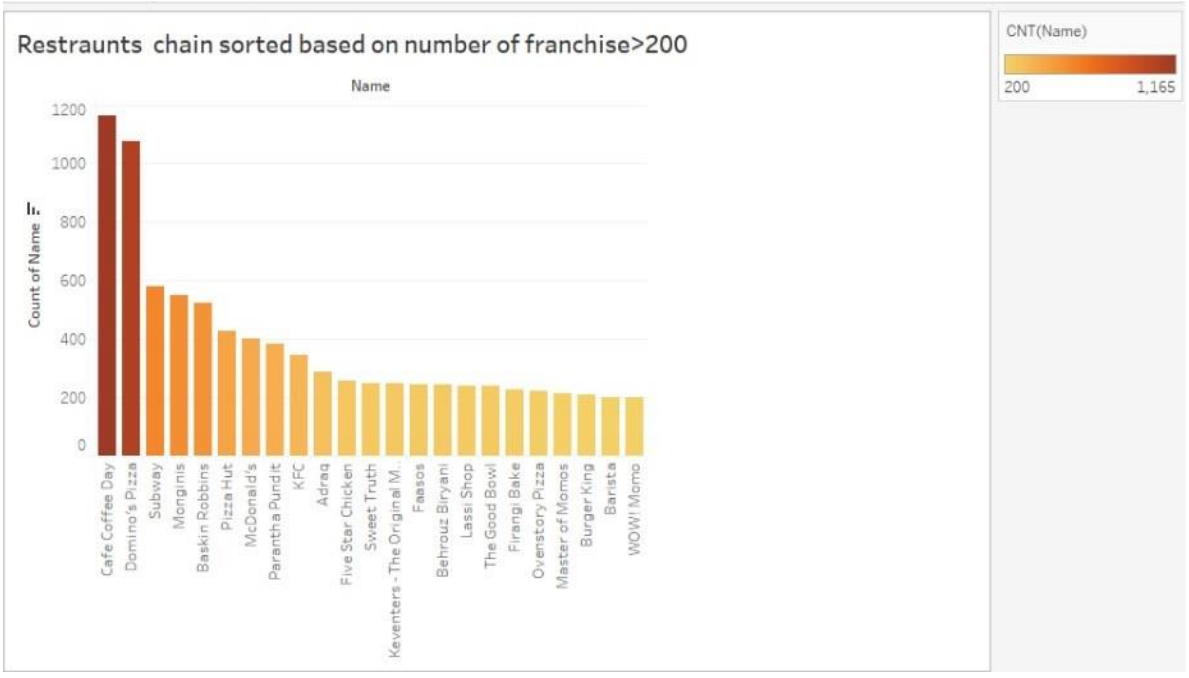


Top 10 Most served dishes in India

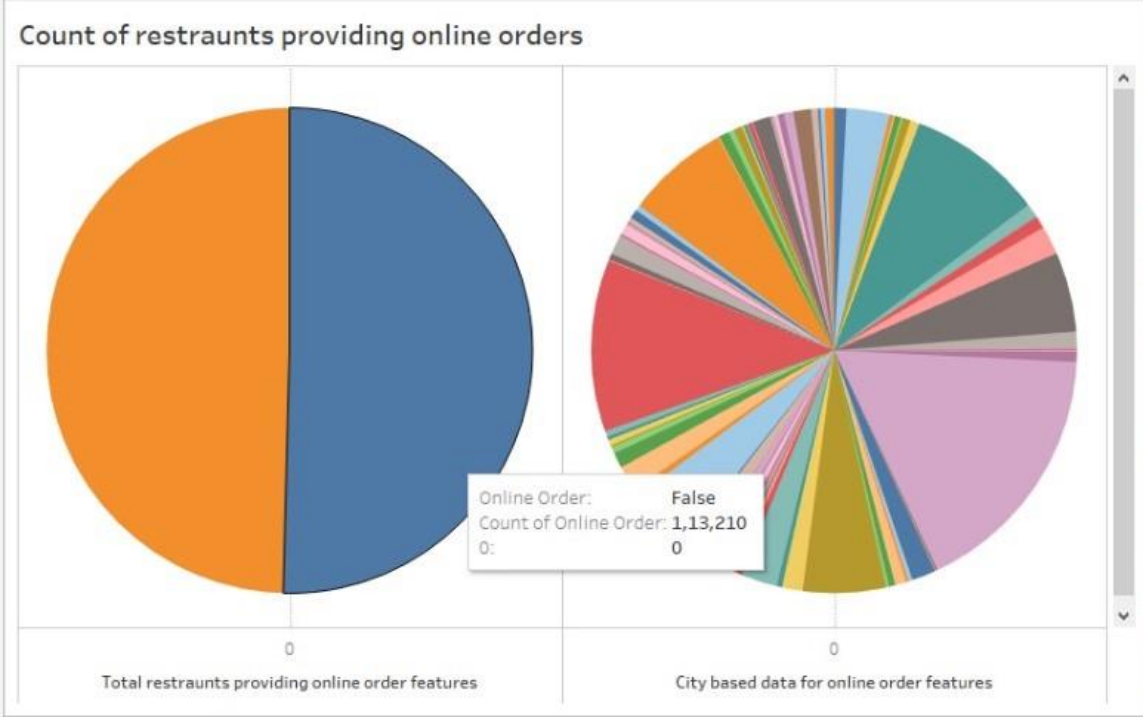
TABLEAU TOOL

- Tableau is a powerful and fastest growing data visualization tool used in the Business Intelligence Industry.
- It helps in simplifying raw data in a very easily understandable format.
- Tableau helps create the data that can be understood by professionals at any level in an organization.
- It also allows non-technical users to create customized dashboards.
- Data analysis is very fast with Tableau tool and the visualizations created are in the form of dashboards and worksheets.
- The best features of Tableau software are
    - Data Blending
    - Real time analysis
    - Collaboration of data

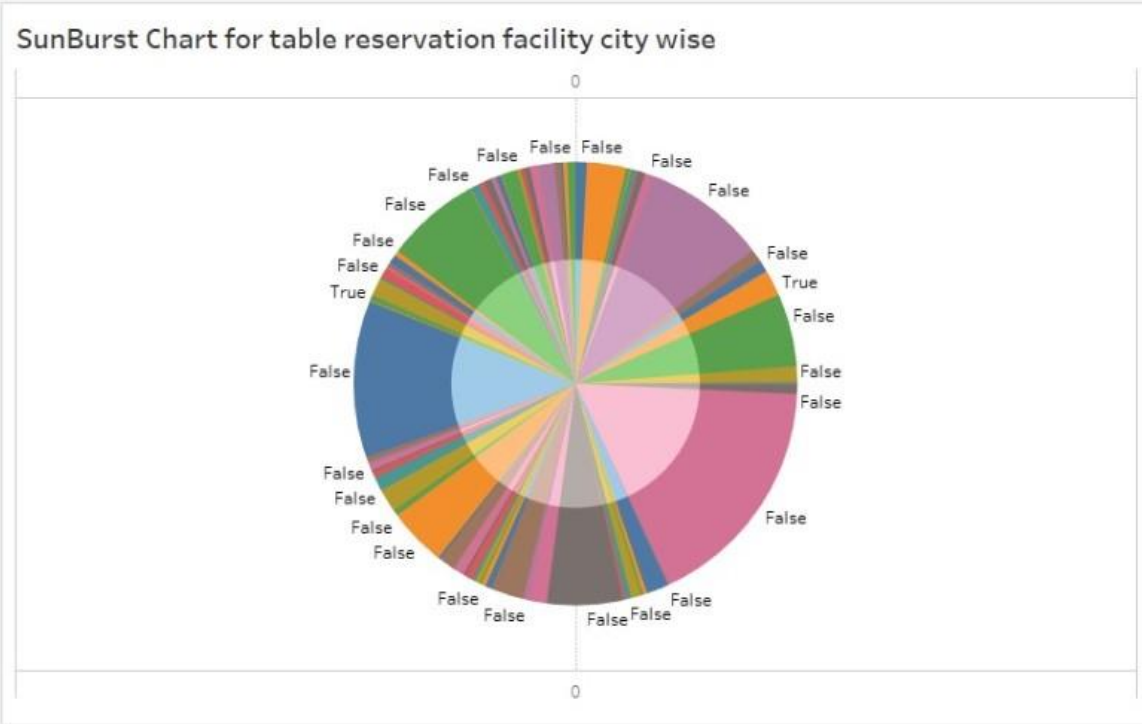
Our Analysis results :-



Restaurant chains with over 200 restaurants

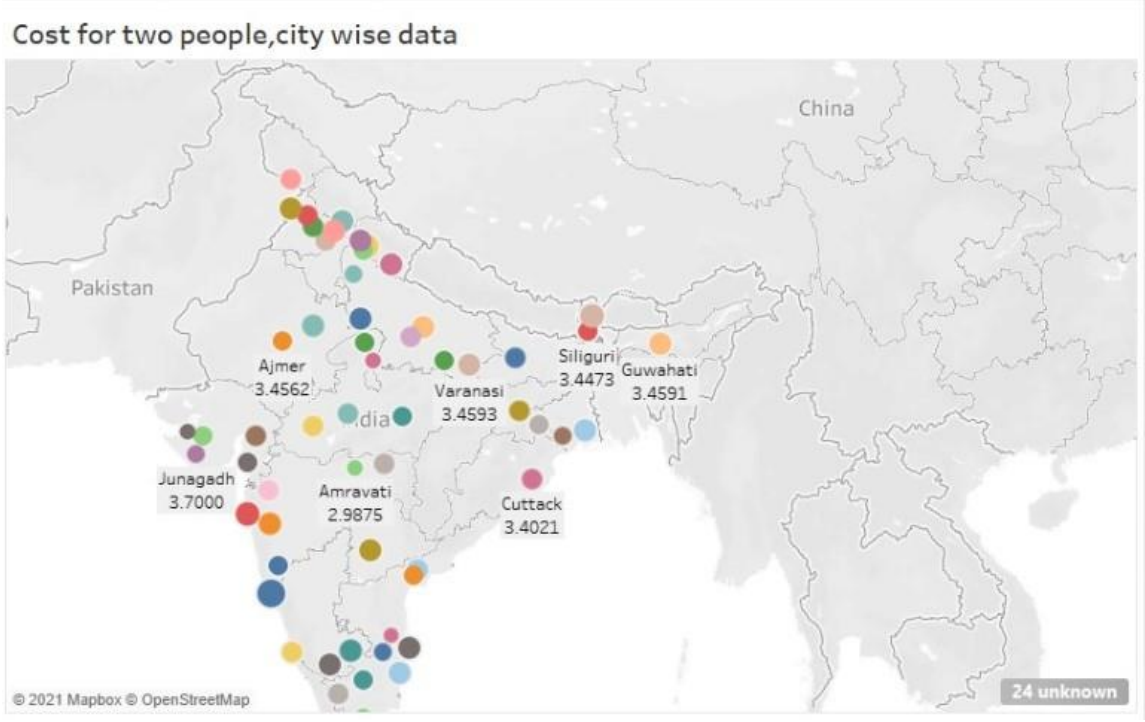


Restaurants Providing Online Orders



City Wise Table Reservation

Total Number of Restaurants



Cost for Two People

Machine Learning Algorithms Used:

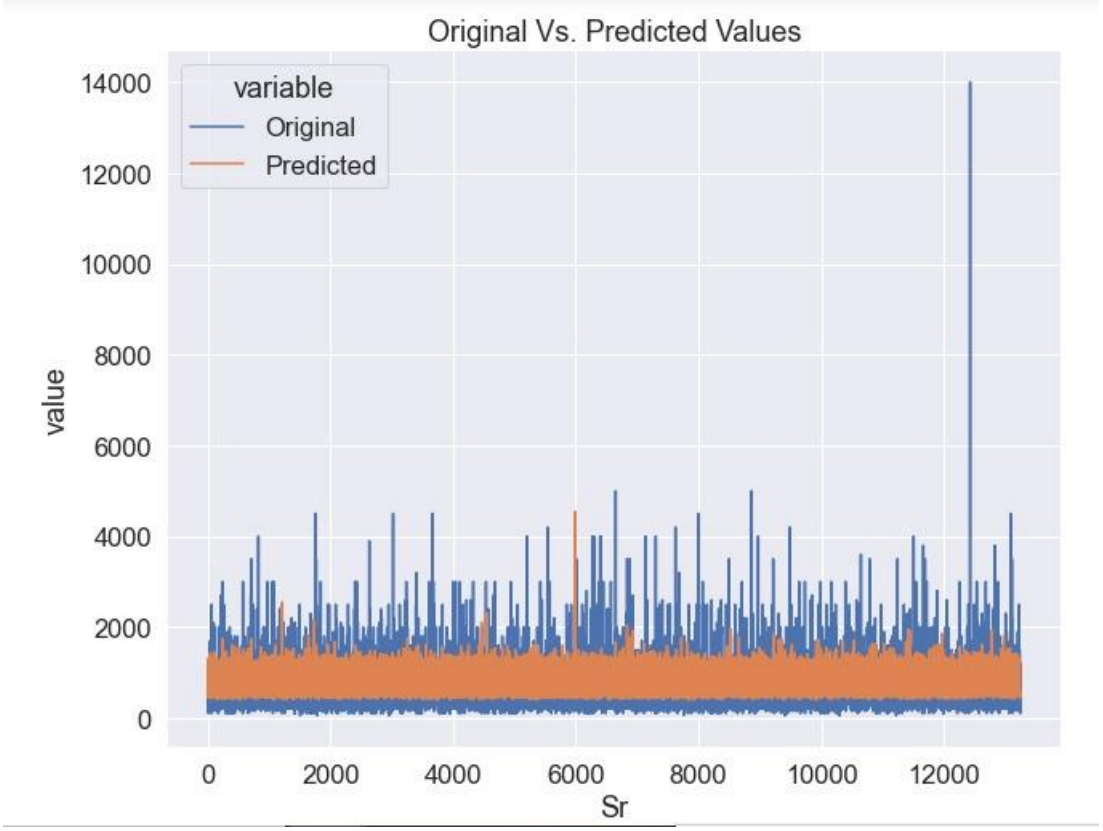
Linear Regression :-

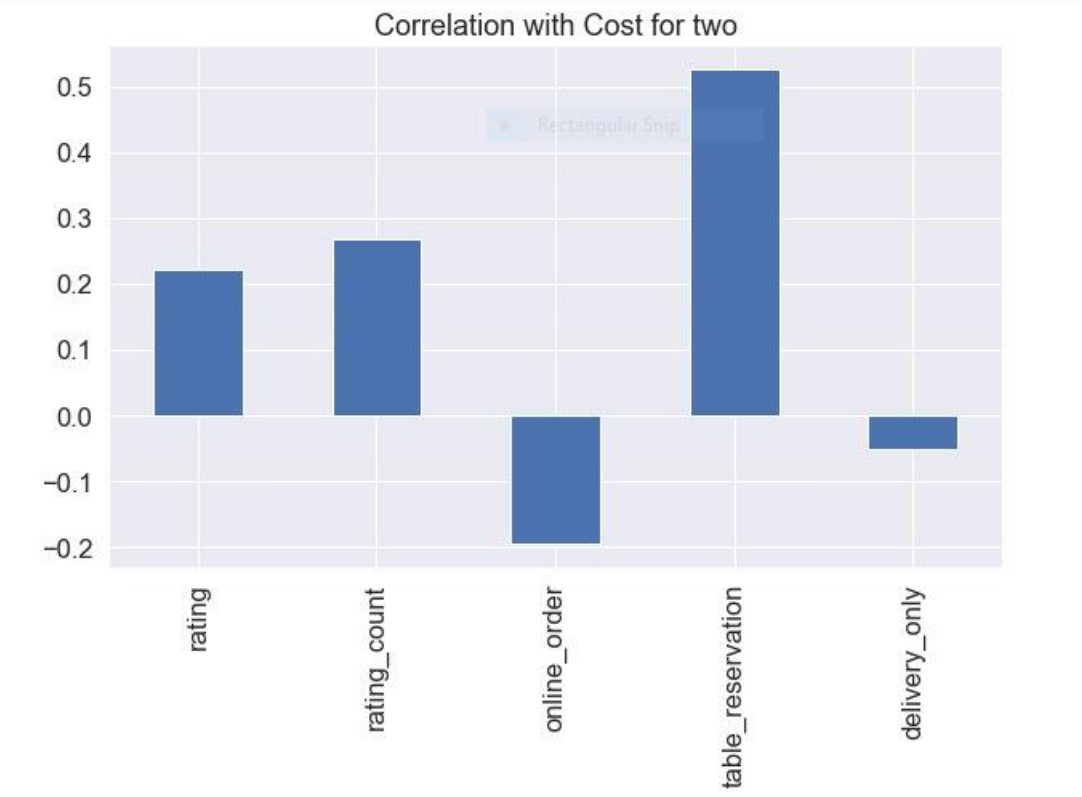
- Linear regression is an attractive model because the representation is so simple.
- The representation is a linear equation that combines a specific set of input values (x) the solution to which is the predicted output for that set of input values (y). As such, both the input values (x) and the output value are numeric.

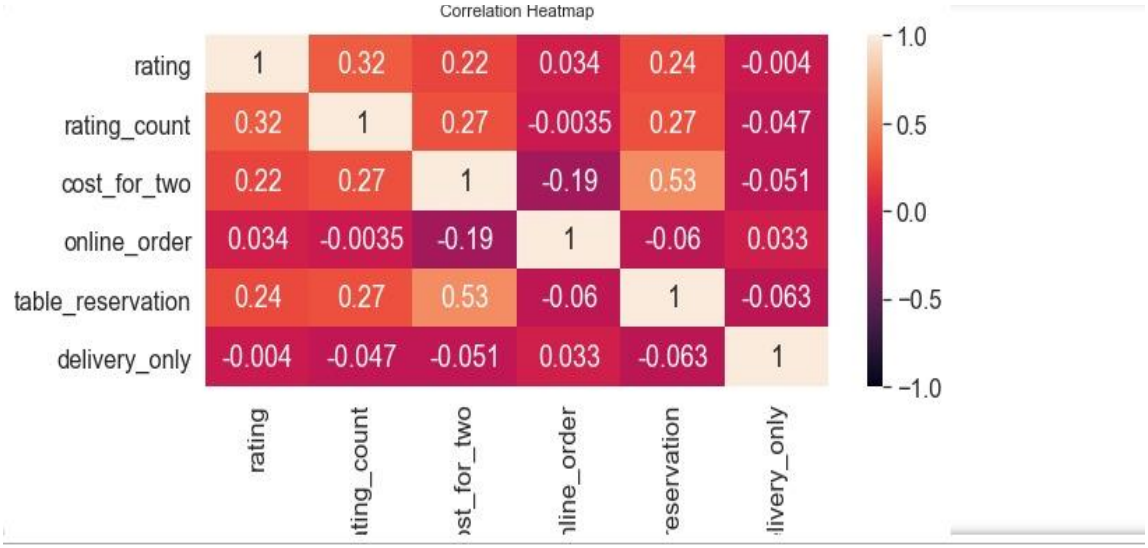
Logistic Regression :-

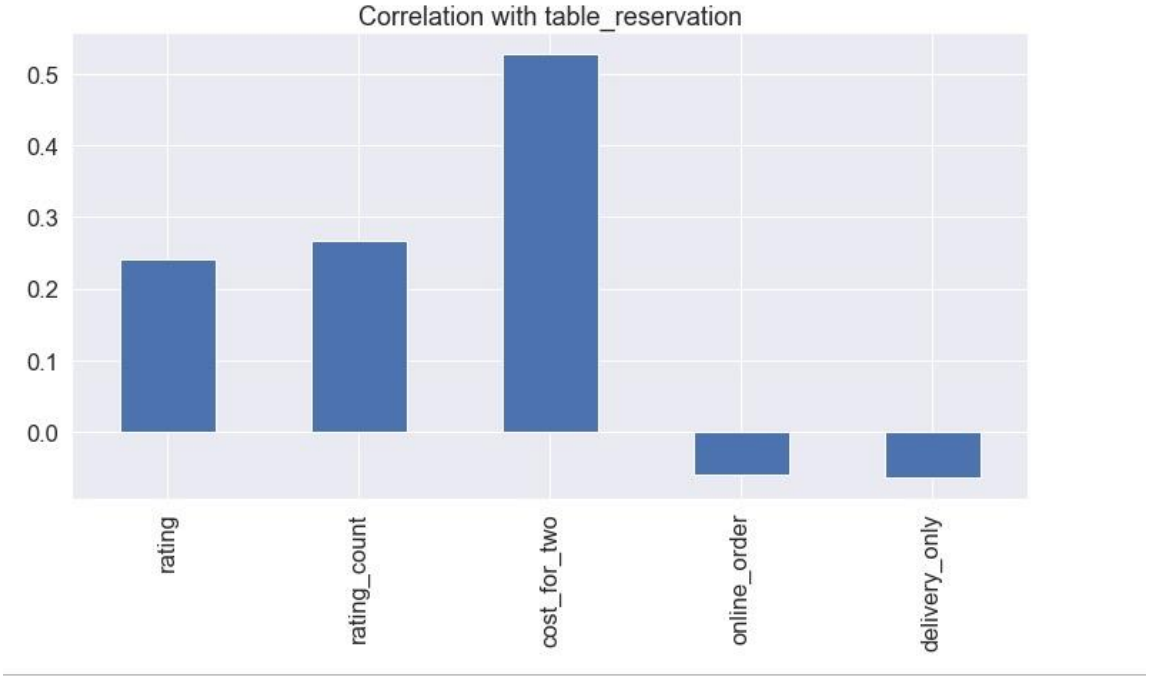
- Logistic regression is a statistical model that in its basic form uses a logistic function to model a binary dependent variable, although many more complex extensions exist.
- In regression analysis, logistic regression (or logit regression) is estimating the parameters of a logistic model (a form of binary regression).

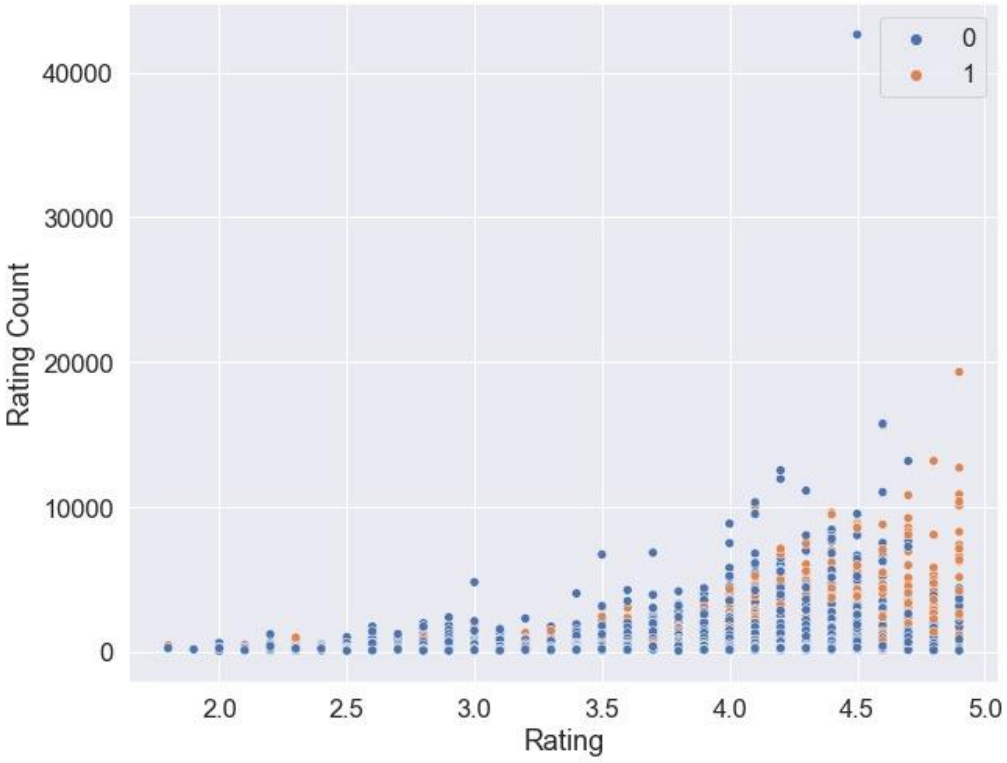
The analysis done by us are as follows:-






Correlation Heatmap
rating
rating_count
cost_for_two
online_order
table_reservation
delivery_only
1
0.32
0.22
0.034
0.24
-0.004
0.32
1
0.27
-0.0035
0.27
-0.047
0.22
0.27
1
-0.19
0.53
-0.051
0.034
-0.0035
-0.19
1
-0.06
0.033
0.24
0.27
0.53
-0.06
1
-0.063
-0.004
-0.047
-0.051
0.033
-0.063
1
1.0
0.5
0.0
−0.5
−1.0


Correlation with table_reservation
0.5
0.4
0.3
0.2
0.1
0.0
rating
rating_count
cost_for_two
online_order
delivery_only



Conclusion :-

We have displayed all the data in a graphical format and used various methods to do it. This gives us a better understanding about the type of dishes and restaurants people in India like. We now understand much better what ratings are dependent on and how different methods can be used to represent a dataset and analyze various important results.