Low Level Design (LLD)
Restaurant Rating Prediction for Zomato

Basawanayya Hiremath

Objective:

The main goal of this project is to perform extensive Exploratory Data Analysis(EDA) on the Zomato Dataset and build an appropriate Machine Learning Model that will help various Zomato Restaurants to predict their respective Ratings based on certain features.

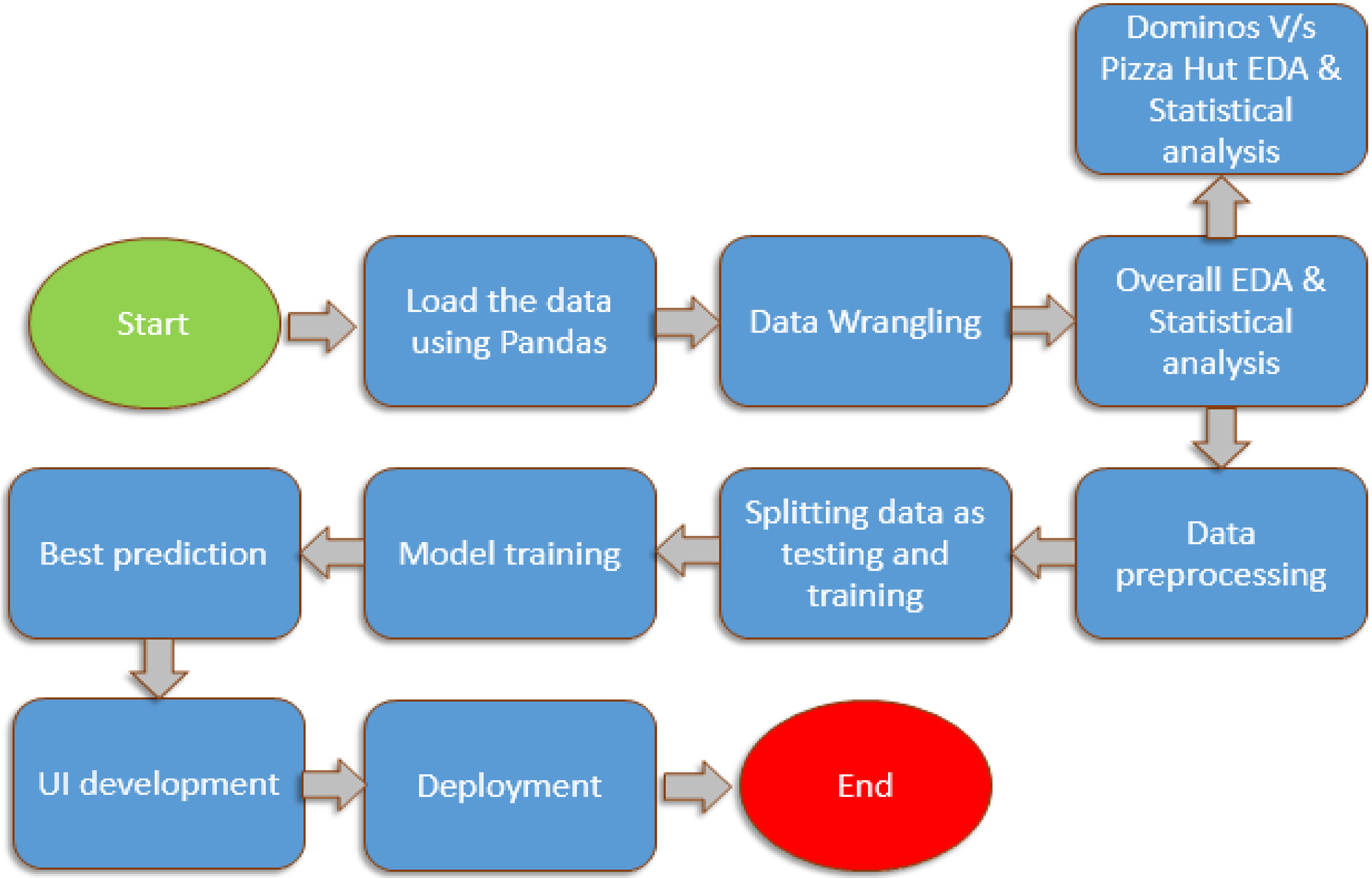
Benefits:

- Comparison of overall graphs
- Predictive model
- Comparison of Dominos and Pizza Hut
- Real world use case

Data Sharing Agreement :

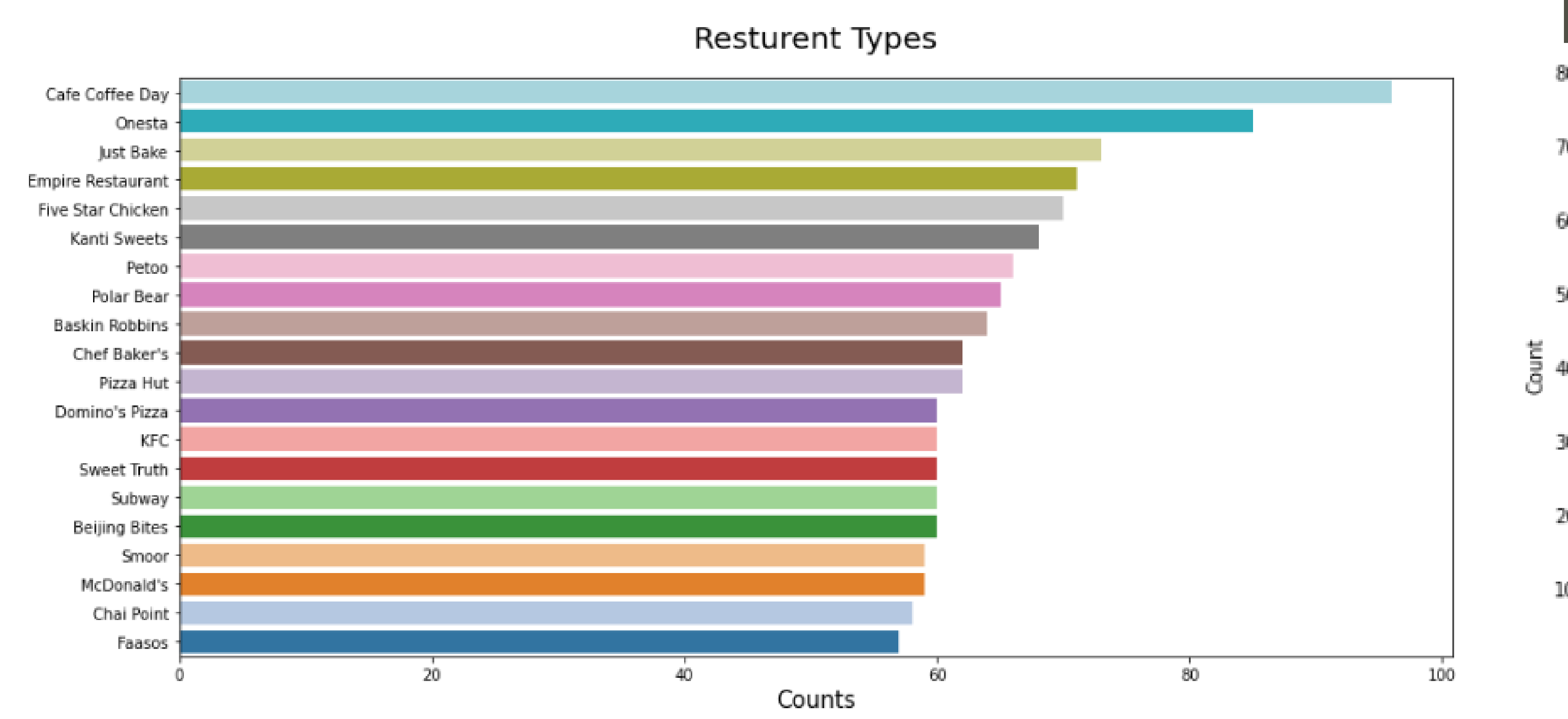
- File name:- **zomato.csv**
  - Link for the dataset is give below
  - https://www.kaggle.com/himanshupoddar/zomato-bangalore-restaurants
- Shape of the data 51717*17
- Number of Columns 17
- Number of Rows 51717
- Column data type

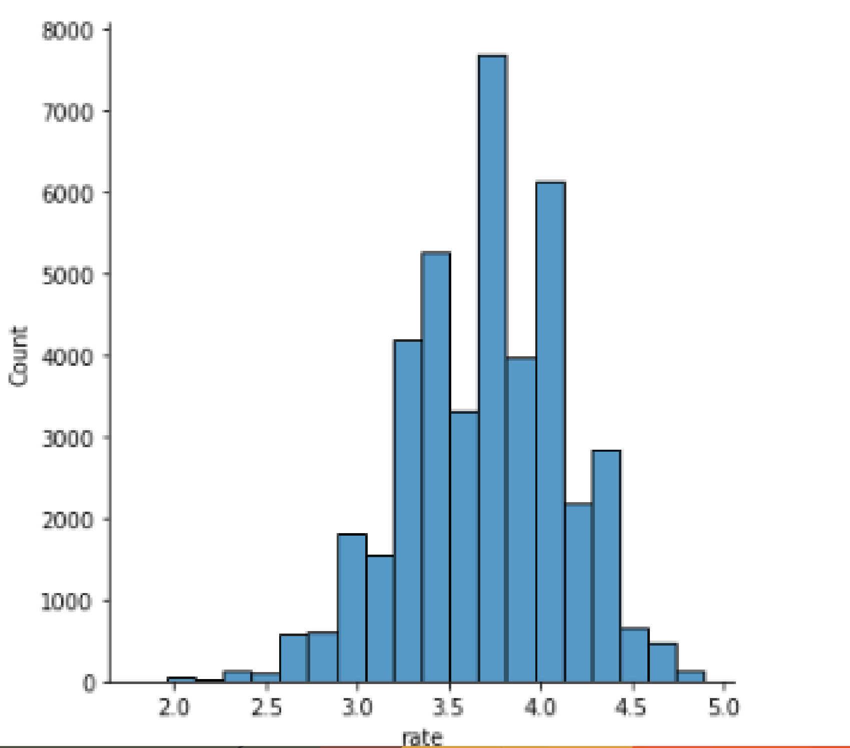
0	url	51717 non-null	object
1	address	51717 non-null	object
2	name	51717 non-null	object
3	online_order	51717 non-null	object
4	book_table	51717 non-null	object
5	rate	43942 non-null	object
6	votes	51717 non-null	int64
7	phone	50509 non-null	object
8	location	51696 non-null	object
9	rest_type	51490 non-null	object
10	dish_liked	23639 non-null	object
11	cuisines	51672 non-null	object
12	approx_cost(for two people)	51371 non-null	object
13	reviews_list	51717 non-null	object
14	menu_item	51717 non-null	object
15	listed_in(type)	51717 non-null	object
16	listed_in(city)	51717 non-null	object

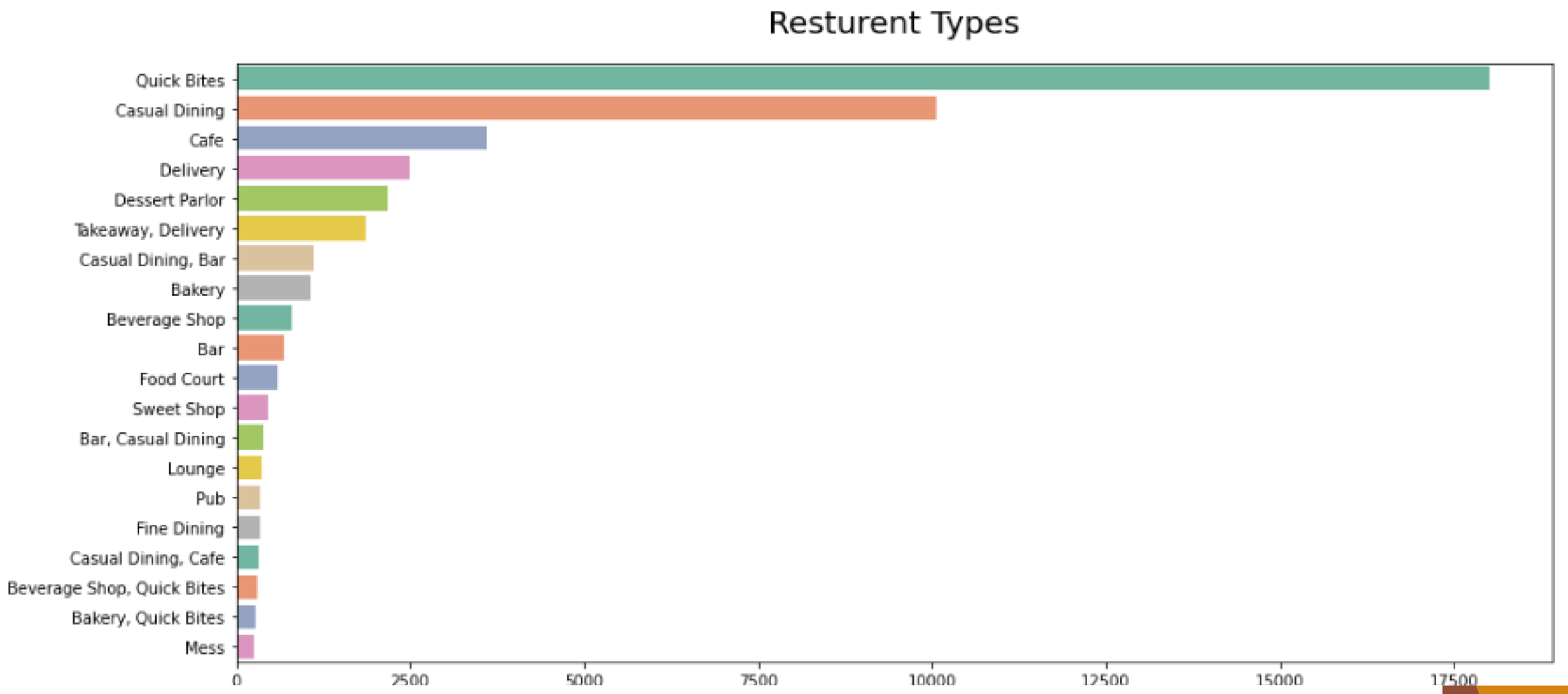

Start
Load the data using Pandas
Data Wrangling
Overall EDA & Statistical analysis
Dominos V/s Pizza Hut EDA & Statistical analysis
Data preprocessing
Splitting data as testing and training
Model training
Best prediction
UI development
Deployment
End

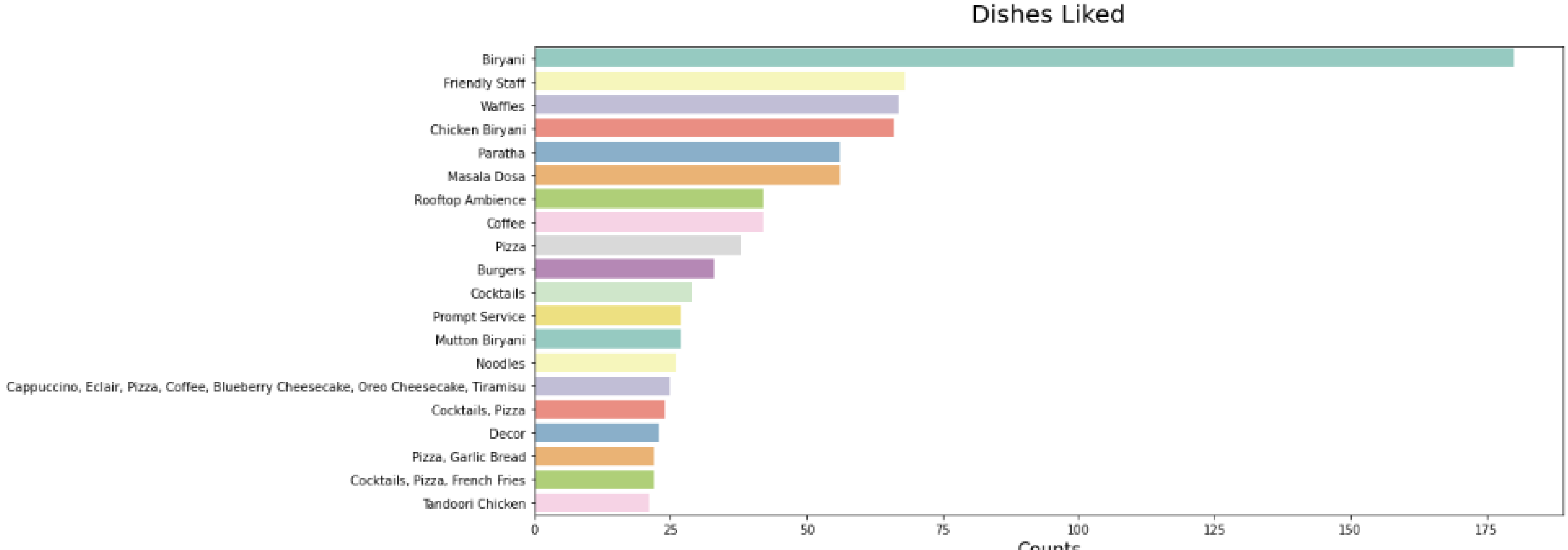
EDA:

Overall expletory data analysis is preformed on the data and subset Dominos and Pizza Hut data extracted form the data and same EDA is preformed. In both cases univarient and bivarient analysis is performed.

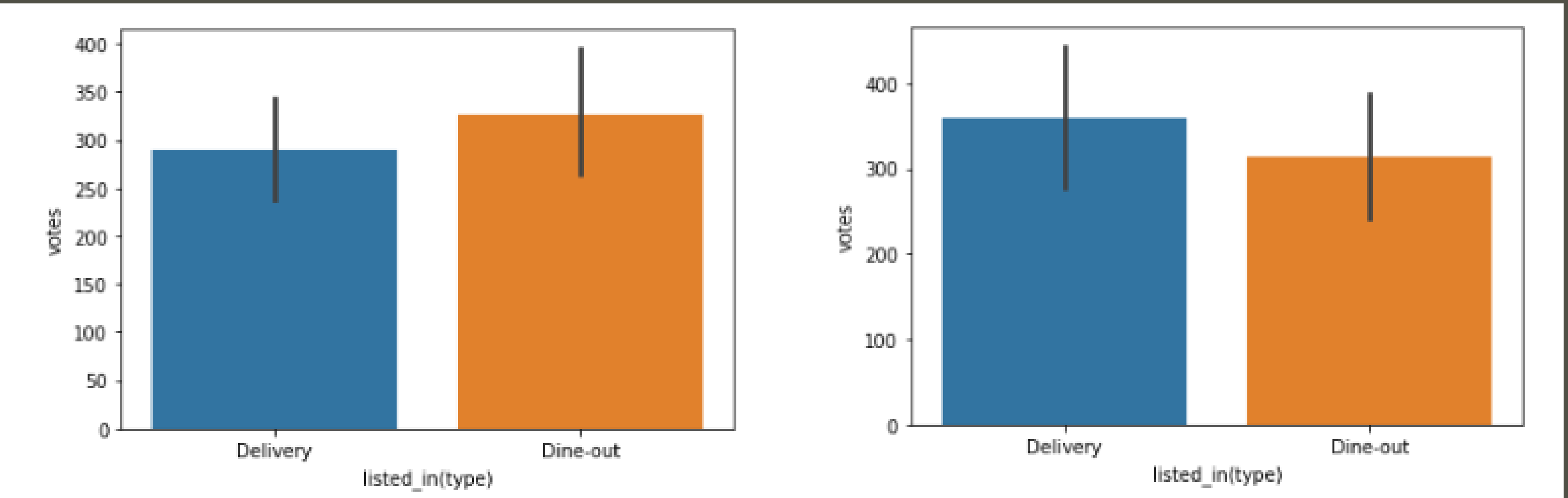


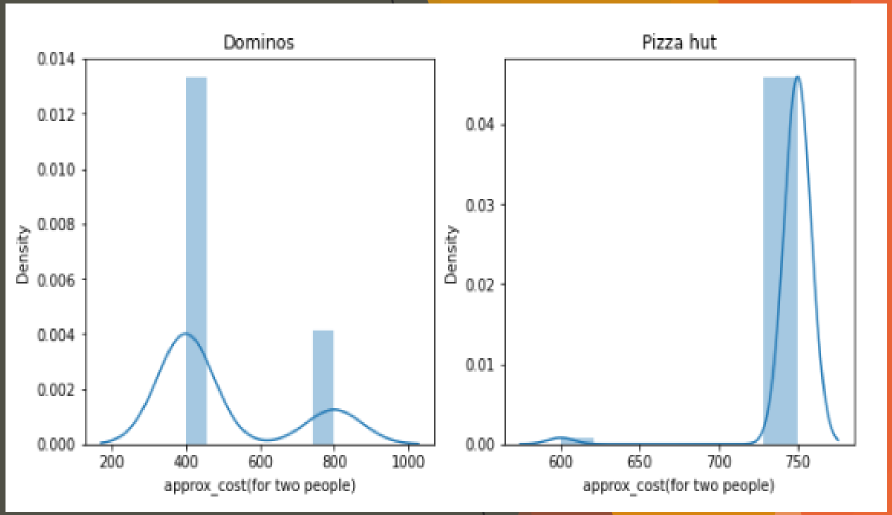


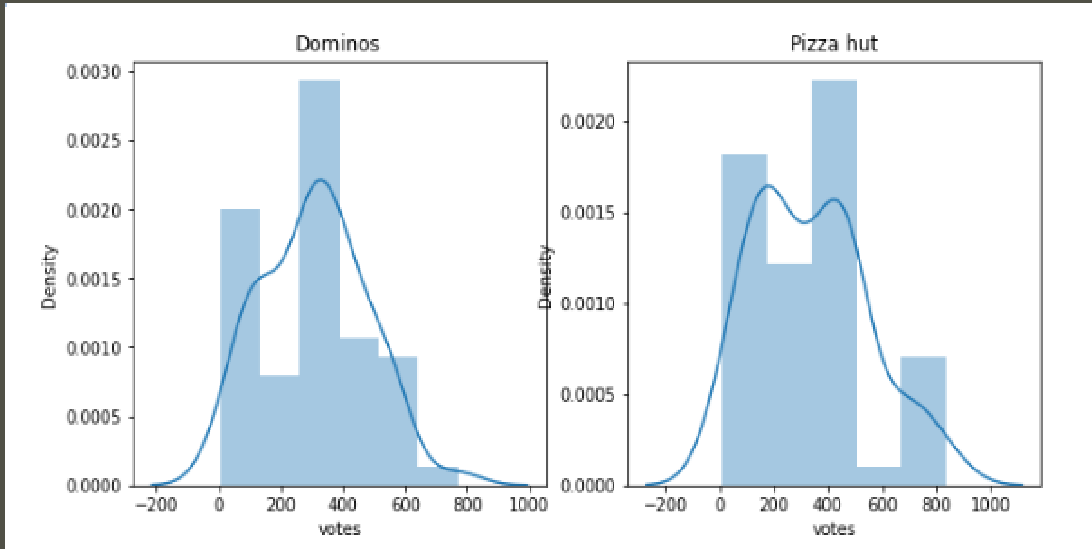

Resturent Types
Quick Bites
Casual Dining
Cafe
Delivery
Dessert Parlor
Takeaway, Delivery
Casual Dining, Bar
Bakery
Beverage Shop
Bar
Food Court
Sweet Shop
Bar, Casual Dining
Lounge
Pub
Fine Dining
Casual Dining, Cafe
Beverage Shop, Quick Bites
Bakery, Quick Bites
Mess
0
2500
5000
7500
10000
12500
15000
17500

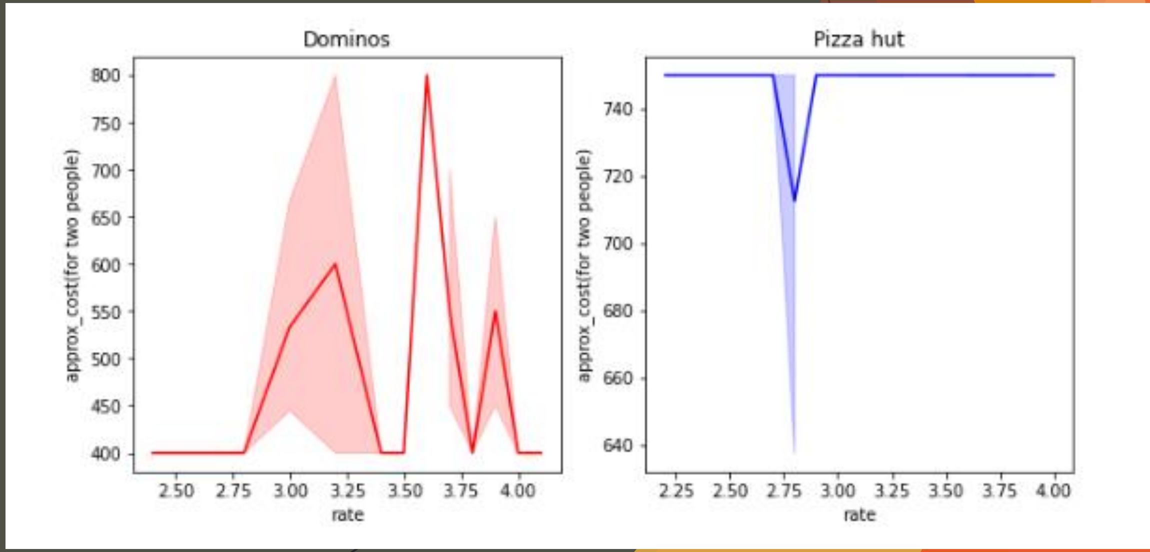

Dishes Liked
Biryani
Friendly Staff
Waffles
Chicken Biryani
Paratha
Masala Dosa
Rooftop Ambience
Coffee
Pizza
Burgers
Cocktails
Prompt Service
Mutton Biryani
Noodles
Cappuccino, Eclair, Pizza, Coffee, Blueberry Cheesecake, Oreo Cheesecake, Tiramisu
Cocktails, Pizza
Decor
Pizza, Garlic Bread
Cocktails, Pizza, French Fries
Tandoori Chicken
0
25
50
75
100
125
150
175
Counts

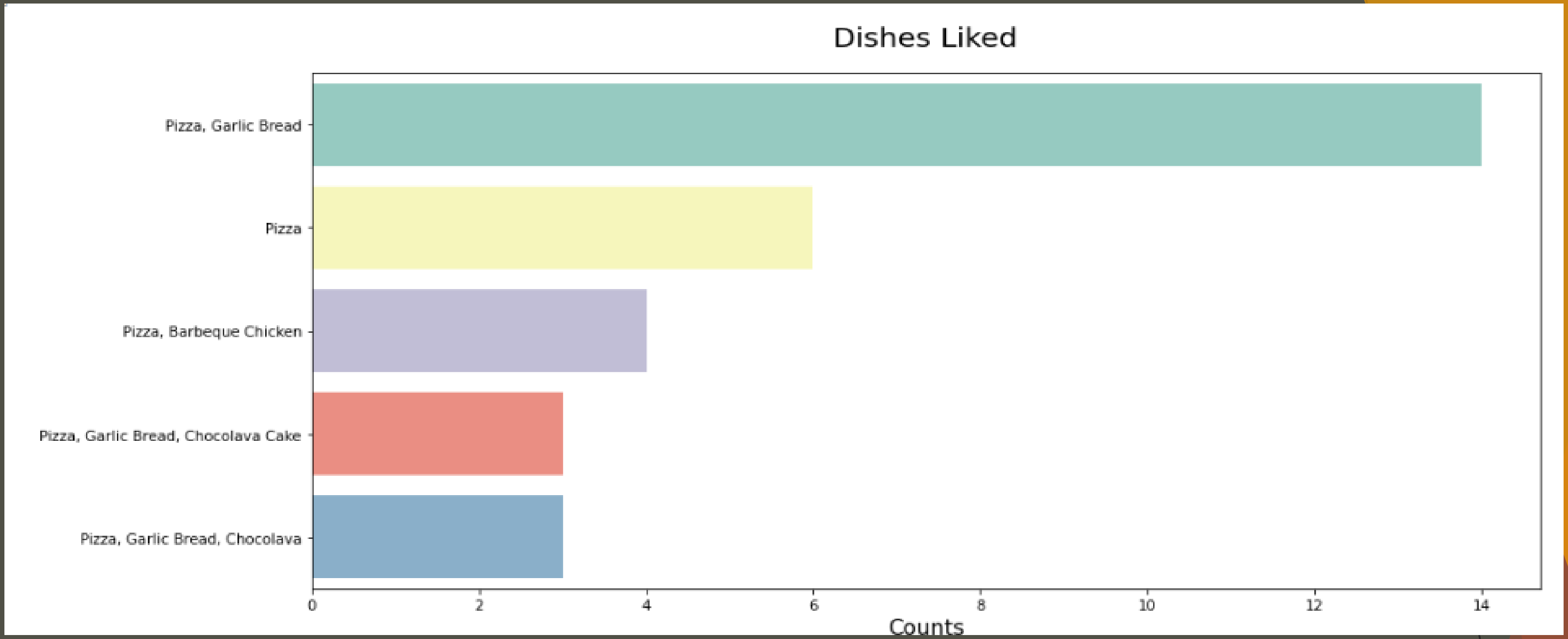
Dominos V/s Pizza Hut



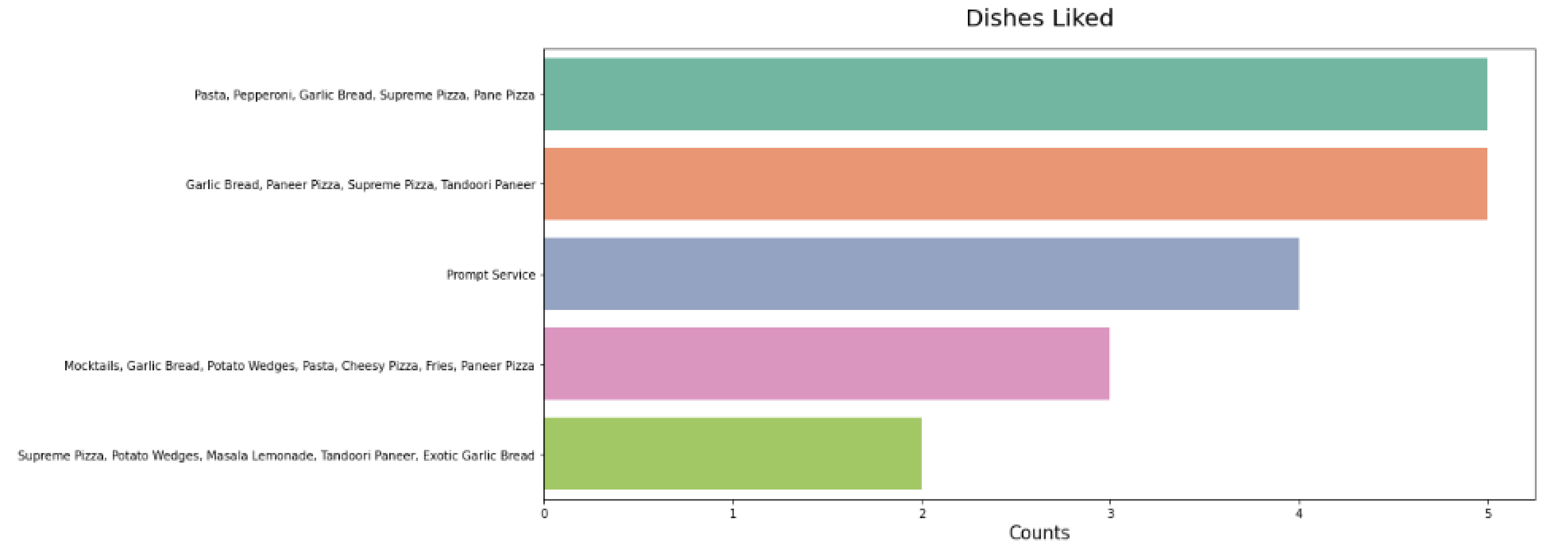








Dominos



Pizza Hut

- For both Dominos and pizza hut table booking is not available.
- online order for Pizza Hut is more compared to Dominos.
- Resturent type, Dominos is considered as Quick Bites but pizza hut is conisdered as Casual dininga and Quick bites.
- locations vs votes for Dominos.
  - MG road, brigade road have very low votes and Banashankari, Koramangala 5th block etc have more votes.
- Locations vs votes for pizza hut.
  - BTM, Marathahalli have very low votes and Rajainagar, old airport etc have more votes.

- Dine out rating for Dominos has more density then Pizza hut and for delivery Dominos have less density compared to pizza hut.
- Cost for two people in Dominos is about 400rs/-(high density) and 800rs/- (low density) but in pizza hut it is about 700rs/-.
- This graph is very interesting, in Pizza hut it is almost people paid about 700rs/- are both happy and unhappy but in Dominos we can see range of customer who rated from 3 to 4.

- **Features:** - url, address, name, online_order, book_table, rate, votes, phone, location, rest_type, dish_liked cuisines, approx_cost(for two people), reviews_list, menu_item, listed_in(type) and listed_in(city).

In the data we can see that only votes feature is int, all other are objects.

- **Data wrangling:-** Some features required cleaning for example 'rate' feature is in format '4/5' which is object, and contains entries like 'NEW', '-' etc. this feature has to clean in order to analysis.
- **EDA:** - overall EDA (uni-variate, bi-variate analysis) is performed on most of the features.
- **EDA of Dominos V/s Pizza Hut:** - comparison is done on both Dominos and Pizza Hut by plotting several graphs and analysis.

- **Data Preprocessing**: - In this step some features are dropped like phone number, location, url etc. After that featuring engineering is performed on object features. Here label encoding is used to transform the data. Only online_order, book_table, rate, votes, location, rest_type, dish_liked, cuisines, approx_cost(for two people), listed_in(type) andlisted_in(city) are used to predict.
- **Splitting data as training and testing**: - Data is splitting as 80% and 20% as training and testing respectively.

- **Model training:** - Various models like linear regression, random forest and XGboosting machine learning algorithms are used to produce best R2 score. Hyperparmeter is performed to increase the score. Best score model is chosen and converted and saved in pickle file. Score of Linear regression is 0.308, Random forest is 0.8737 and XGBoosting is 0.9120 with n_estimators=1000.
- **UI development:** - Flask and Html is used to design to be displayed in a web browser.
- **Deployment:** - Model is deployed on local host.

Q & A:

Q1) What's the source of data?

- The data for training is available in kaggle with link

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

Q 2) What was the type of data?

- The data was the combination of numerical and Categorical values.

Q 3) What's the complete flow you followed in this Project?

- Refer slide 4th for better Understanding.

Q 4) After the File validation what you do with incompatible file or files which didn't pass the validation?

- No such file exist.

Q 5) How logs are managed?

- Database is not integrated.

Q 6) What techniques were you using for data pre-processing?

- Removing unwanted attributes
- Checking and changing Distribution of continuous values
- Cleaning data and imputing if null values are present.
- Converting categorical data into numeric values.

Q 7) How training was done or what models were used?

- Splitting of data as 80% traning and 20% testing is preformed, Various models like linear regression, random forest and XGboosting machine learning algorithms are used to produce best R2 score. Hyperparmeter is performed to increase the score. Best score model is chosen and converted and saved in pickle file. Score of Linear regression is 0.308, Random forest is 0.8737 and XGBoosting is 0.9120 with n_estimators=1000

Q 8) How Prediction was done?

- Prediction is done using UI build using flask and html.

Q 9) What are the different stages of deployment?

- When the model is ready to use it is deployed in Heroku .
- All files are uploaded in Github.