Group members: -

1. Divya Pulivarthy 3) Naga Sai Surya Prakash Reddy Kovvuri
2. Sowjanya Tammali 4) Faisal

Important & Interesting Experiences in this Project: -

* Data Science Model Building is all about iterating over efforts by enhancing application techniques and learning from previous mistakes.
* In this Project, the major challenge we faced was “Dealing with a lot of categorical values.”
* Initially, we created dummies out of all categorical variables and performed data visualization, which increased the dimensionality of the dataset.
* We applied visualization techniques like heatmaps, correlation analysis, and density plots on the dataset to understand the underlying relationship between variables.
* We implemented label encoding to convert some object/string categorical variables into numerical categorical variables so that the ML models can produce results efficiently.
* The dataset has been further refined by following the below-mentioned steps:

Replacing all the missing values with the Mode of the respective columns. This improved the efficiency of the ML Models.

* We attempted to predict the restaurant rating based on the previous users ratings, food ratings, and service ratings and also user preferences of cuisine, mode of payment and ambiance choices, etc., which can be recommended to users based on all the above factors.

One Important takeaway from this project is how to deal with huge categorical independent variables.

**Methodology:**

In this Project, we tried different “Regression Models” such as

a) Multiple Linear Regression and

b) KNN Regression Model,

and found that Regression models do not hold well with many categorical variables.

Classification Models help discover a model or function that helps separate the data into multiple categorical classes, i.e., discrete values. Hence, we have applied various Classification Models, such as:

1) 'k-NN' Classifier

2) 'Gaussian Naive Bayes'

3) 'Bernoulli Naive Bayes'

4) 'Decision Tree'

5) 'RF with Hyperparameter Tunning'

6) 'Logistic Regression without Class Weight'

7) 'SVC'

8) 'XGBoost'.

And among the above-mentioned Classification Models, **XGBoost** **Classification Model** turned out to predict the data with better accuracy of “**55.8%”, a** better Precision score of “**59.31%”** and a better AUC of “**65.85%**”.

**Recommendations based on Ratings:**

In this type of recommendation system, we recommend only those items which are most popular among the users, i.e., which are liked by most of the users. As for example, In Restaurants & consumer Dataset, we have a rating\_final.csv file which contains UserID, PlaceID, rating, food\_rating, service\_rating attributes.

Although the relationship between all the variables is not linear, it turned out that rating is linearly dependent on food\_rating and service\_rating. Apart from food-rating and service\_rating the Dependent Variable is also dependent on variables such as Mode of Payment, cuisines offered by the Restaurant, Parking lot, Ambiance etc.

**Conclusion and Observations:**

Here we observe that, the features which affect the rating most are food-ratings, service\_ratings, Mode of Payment, Cusines, Parking lot, Ambiance. We see that a list of features like parking, Other Service, Price, Dress Code, Accessibility, Area have less affect on the ratings.

i. Food Rating

Food rating has major affect as users prefer type of food, taste of food offered by the restaurants and we see users prefer restaurants offering Mexican cuisines.

ii. Service Rating:

Service rating has major affect on the rating of the model as users prefer Restauarnts with Good service ratings.

iii. Parking

As we can see that parking affect the combined rating but not food rating. The parking has many options like: 'parking\_lot\_none', 'parking\_lot\_public', 'parking\_lot\_valet parking', 'parking\_lot\_yes' different values. In this, parking\_lot\_public affect the restaurant and service rating both because it only matters for the car owner.

iv. Price

The price is totally depends on the budget of the people. Mostly medium and high-price restaurants have better ratings. Hence, it does not have major toll affect on the over-all Performance of the model.

v. Dress Code

The dress code has no much significant affect on the rating as we see that most of the restaurant have no dress code.

vi. Other Services

We see that the restaurant having variety of other services having good rating as compare to those having internet only. So, other services increase the rating.

vii. Accessibility

Accessibility doesn’t affect the rating more, because sometimes the area where the people prefer close space as compared to open space, for cuisine, what type of food people prefer is depends on which rating we are predict. As per example, fast food restaurant not have good rating.