**EXPERIMENT NO: 2**

**Aim:** To identify Business aspects for an identified domain and perform analysis for the same.

**To do:-**

1. Case study on an identified-Research Paper. (Problem Statement)

2. Dataset Description/sample attribute(s) -

1. Identify target attribute/Dependent and Input/Independent attributes,
2. Identify types of attributes (Nominal, Ordinal, Continuous. Discrete) on which the task is performed.

3. Methodology Used: Different Algorithms are used and Analysis of different techniques for the same

4. Inferences and effects on business/value

**Case study:**

The Swiggy: Bangalore delivery outlet dataset provides us with information on factors influencing the establishment of different types of restaurants at different locations in Bangalore, along with the overall rating of each restaurant. We'd like to find the cheapest restaurant in Bengaluru here. Along with the same, we can discuss different other relationships such as the most expensive restaurant, the best location, and the relationship between cuisine and rating, as well as the cost for two people for a particular cuisine.

Since it is real-time data we would start with our Data Exploratory processes like handling the Nan values, null values, dropping duplicates, and other Transformations. Our target variable is the "Rating" column. We explore the relationship between the other features in the dataset with respect to Rating. We will visualize the relation of all the others depending on features with respect to our target variable, and hence find the most correlated features which affect our target variable. We would then implement the data in various modeling structures such as Random Forest, Linear Regression, and Decision Tree. These modeling will then give us the accuracy of prediction and then we could state which model gives us the most optimized and accurate readings

**Introduction:**

Online food ordering apps are the media by which local hotels and restaurants, chefs, and canteens are delivering takeaway and food parcel directly to consumer's footsteps. Due to increase in working young generation in metro cities and hectic work life culture this type of idea is easily spreading. This system has generated a new dimension to working people’s kitchen. Nowadays consumers are getting more attracted towards online ordering apps rather than home delivery of a specific restaurants. In a process of online food ordering apps there is no human intervention involved which gives it more privacy. Apps are having number of restaurant, chefs kitchens listed with their menu specifically. So the consumers need not to carry pamphlets and menu list for further orders. It gives convenience to order food on click of a button. These apps can be directly downloaded to smart phone which give them more accessibility. By giving your address and profile, payment information account can be created. However the app needs to be downloaded by the customers on their cell phones and register them on the app. Creating profile on apps includes their address and payment information. Apps are having different kind of mode of payments like credit cards, debit cards, cash less accounts and free home delivery. Different apps offer different services, offers, features or restaurants too. Downloaded app used to give some coupons discounts, previous order history, some palette suggestions, recent customers review on restaurants as well as dishes. The market business of Indian food industry is expected to reach $420 billion by 2020, reports BCG. Presently, the Indian food business is around $350 billion in 2019. This sector is coming up with innovative ideas to provide their consumer convenience, satisfaction and retention. There is a huge competition between free home delivery given by particular restaurants and online food ordering apps. Due to introduction of new technology and innovation consumers are having so many options and varieties to choose from. Initially due to adherences there is skeptically, dicey about idea investor were hesitant but after success of concept there are growing number of players in market. Zomato, Swiggy, food panda, Ubereat, Just eat are some main players in this segment. Few big restaurants brand owns their own chain of delivery like Dominos, Pizza hut, fassos, KFC.

Swiggy is an online platform, popularly known for its mobile app, to order food from a wide range of neighbourhood restaurants. Because of offering quick doorstep food services in the minimum time possible, Swiggy has become a prior choice of foodies. In this project, we are analyzing the various aspects with different use cases which covers many aspects of Swiggy Food Delivery Service. It helps in not only understanding the meaningful relationships between attributes, but it also allows us to do our own research and come-up with our findings. The objective of the project is to perform an exploratory data analysis, data pre-processing, & data cleaning and at the end, apply different Data Visualization techniques to get the meaningful insight from the given data. This project aims apply some amazing Python Libraries such as Plotly and WordCloud which will give a boost to our visual understanding of the data.

**Dataset Description**

**Dataset link**: [link](https://www.kaggle.com/datasets/sabinhashmi/swiggy-bangalore-delivery-outlet-data)

In the datasets, we are provided with the following columns (Features) of data.

* Shop\_Name: Name of the Shop/Restaurants.
* Cuisine: Name of the different Cuisines provided by Restaurants.
* Location: Restaurant Area/Location.
* Rating: Rating given by the Customers out of 5.
* Cost\_for\_Two (₹): Approx. Cost of Two people w.r.t. Restaurants.

**Types of attributes:**

**1. Ordinal:**

Cost\_for\_Two (₹)

**2. Discrete:**

Shop\_Name, Cuisine, Location

**3. Continuous:**

Rating

**Data Preprocessing:**

**Exploratory Data Analysis:**

A lot of effort went into the EDA as it gives us detailed knowledge of our data.

Exploratory Data Analysis (EDA) is an approach/philosophy for data analysis that employs a variety of techniques (mostly graphical) to

* maximize insight into a data set;
* uncover underlying structure;
* extract important variables;
* detect outliers and anomalies;
* test underlying assumptions;
* develop parsimonious models;
* determine optimal factor settings

**Classification Algorithms:**

**1. Linear Regression:**

Linear regression attempts to model the relationship between two variables by fitting a linear equation to observed data. One variable is considered to be an explanatory variable, and the other is considered to be a dependent variable. For example, a modeler might want to relate the weights of individuals to their heights using a linear regression model. Before attempting to fit a linear model to observed data, a modeler should first

determine whether or not there is a relationship between the variables of interest. This does not necessarily imply that one variable causes the other (for example, higher SAT scores do not cause higher college grades), but that there is some significant association strength of the relationship between two variables. If there appears to be no

association between the proposed explanatory and dependent variables (i.e The scatter plot does not indicate any increasing or decreasing trends), then fitting a linear regression model to the data probably will not provide a useful model. A valuable numerical measure of association between two variables is the correlation coefficient,

which is a value between -1 and 1 indicating the strength of the association of the observed data for the two variables. A linear regression line has an equation of the form Y = a + bX, where X is the explanatory variable and Y is the dependent variable. The slope of the line is b, and a is the intercept (the value of y when x = 0)

**2. Random Forest:**

Random forests or random decision forests are an ensemble learning technique for classification. Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model. As the name suggests, “Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset.” Instead of relying

on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output. The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting. The below diagram explains the working of the Random Forest algorithm.

**Working Of Random Forest Algorithm**

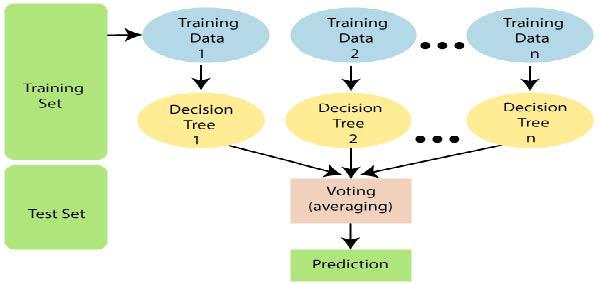
We can understand the working of Random Forest algorithm with the help of following steps –

**Step 1** – First, start with the selection of random samples from a given dataset.

**Step 2** − Next, this algorithm will construct a decision tree for every sample. Then it will get the prediction result from every decision tree.

**Step 3** − In this step, voting will be performed for every predicted result.

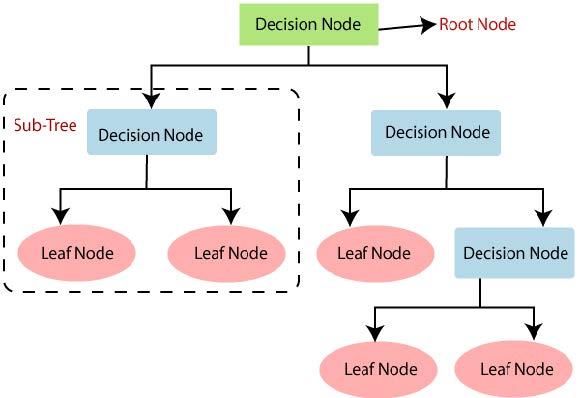
**Step 4** − At last, select the most voted prediction result as the final prediction result.



The pseudo code for a random forest algorithm can split into two stages. First, in which ‘n' random trees are created, this forms the random forest. In the second stage, the outcome for the same test feature from all decision trees is combined. Then the final prediction is derived by assessing the results of each decision tree or just by going with a prediction that appears the most times in the decision trees.

**3. Decision Tree:**

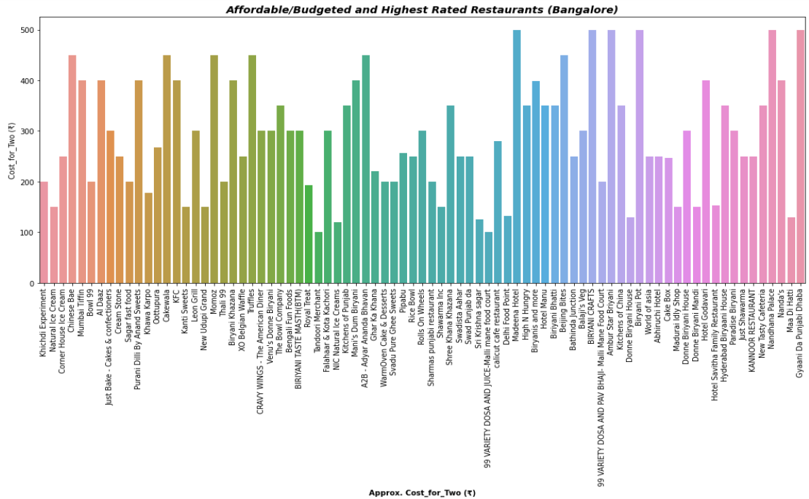
* Decision Tree calculation has a place with the supervised learning algorithms. Decision Tree is a supervised learning technique that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome.
* In a Decision tree, there are two nodes, which are the Decision Node and Leaf Node. Decision nodes are used to make any decision and have multiple branches, whereas Leaf nodes are the output of those decisions and do not contain any further branches.
* The decisions or the test are performed on the basis of features of the given dataset.
* It is a graphical representation for getting all the possible solutions to a problem/decision based on given conditions.
* It is called a decision tree because, similar to a tree, it starts with the root node, which expands on further branches and constructs a tree-like structure.
* In order to build a tree, we use the CART algorithm, which stands for Classification and Regression Tree algorithm.
* A decision tree simply asks a question, and based on the answer (Yes/No), it further splits the tree into subtrees.



**Conclusion**:

In this analysis project, we have analyzed several different use cases for the given dataset to make better business decisions and help analyze customer trends and satisfaction, which can lead to new and better products and services.

We have also analyzed that, we have a total of “82" which are "Budget Restaurants" as well as they are "Affordable". On top of that, we have found out, Most of the Affordable/Budgeted Restaurants are having Excellent Ratings as well. Like, For Approx. Cost of "200", "150", "250", and "450", the Ratings were "4.8", "4.6", and "4.5" respectively. This might be because Most people prefer Affordable/Budget-Restaurants which also provide good quality Cuisine.



And On the other hand, there are a few Expensive Restaurants that don't have that many ratings and they are Expensive too. Those Restaurants Cost around "600" to "800" for Two People are having Ratings between '4.0' 'and 4.1' which is too less as compared to Affordable/Budgeted Restaurants.