Machine Learning:

About the Data:

Zomato is an Indian multinational restaurant aggregator and food delivery company founded by Deepinder Goyal and Pankaj Chaddah in 2008. Zomato provides information, menus, and user reviews of restaurants as well as food delivery options from partner restaurants in select cities.

Attributes:

1. URL - Website of the Zomato for each restaurant. - Object datatype
2. Address - Address of the Restaurant. - Object datatype
3. Name - Name of the restaurant. - Object datatype
4. Online Order - The customer ordered the menu online or not. - Object datatype
5. Book table - The customer has booked the table or not. - Object datatype
6. Rate - Rating of the restaurant that has by the customer. - Numerical datatype
7. Votes - The votes have been given by the customer to the restaurant. - Numerical datatype
8. Phone - Contact number of the Restaurant. - Object datatype
9. Location - The city name where the restaurant is located. - Object datatype
10. Rest Type - The type of restaurant. - Object datatype
11. Dish liked - Dishes liked by the customer from the restaurant. - Object datatype
12. Cuisines - The cuisines that have been prepared by the restaurant. - Object datatype
13. Approx Cost for two people - The approximate cost of the customer for 2 people. - Number datatype
14. Reviews list - The reviews made by the customers on the restaurant. - Object datatype
15. Menu Item - The menu items that are usually available at the restaurant. - Object datatype
16. Listed in (type) - Contains the type of the meal. - Object datatype
17. Listed in (city) - This contains the neighborhood in which the restaurant is listed. - Object datatype

Section A: ***Title: Regression model***

Problem:

Restaurants from all over the world can be found here in Bengaluru. From the United States to Japan, Russia to Antarctica, you get all types of cuisines here. Delivery, Dine-out, Pubs, Bars, Drinks, Buffet, Desserts you name it and Bengaluru has it. Bengaluru is the best place for foodies. The number of restaurants is increasing day by day. Currently, it stands at approximately 12,000 restaurants. With such a high number of restaurants. This industry hasn't been saturated yet. And new restaurants are opening every day. However, it has become difficult for them to compete with already established restaurants. The key issues that continue to pose a challenge to them include high real estate costs, rising food costs, shortage of quality manpower, fragmented supply chain, and over-licensing.

Objective:

The newly started companies are not able to decide the cost that would happen per two people for once. So the Zomato company has a good analyst team who can predict the cost per two customers for one time so that the newly started restaurants and upcoming restaurants will be well prepared how the restaurant should invest in improving the ambiance and all other stuff to attract the customers. Assume you are the analyst team that Zomato has organized to help new and upcoming restaurants by letting them know the various reasons that customers look for and build a model which able to predict the cost for two people.

Steps that are to be followed:

Step 1: Understand the business problem.

Step 2: Read the data, and convert the data types.

Note: The data set has numerical and categorical data but due to noise(anomaly) in the data, the columns are treated as the object type. And You may feel like converting the features into numerical at this step if not appropriate at this stage, In that case, feel free to convert the variable to the appropriate type in the further step as well based on your way of analyzing the data.

Step 3: Perform the described method for the data, Try to find any essential points from the described analysis. And check the missing values and Duplicate records. Impute the missing values in the best way possible.
Note: To impute the missing values with parameters, You must find the best parameter.
Hints:

- Check the distribution using plots. And check the Skewness, Kurtosis and etc.

Step 4: Once the basic preprocessing is done like converting the data types, missing value imputation, and duplicate rows. perform the EDA(Exploratory Data Analysis) on the data to find the various factors that will help to understand the cost per two persons.

Step 5: Make a copy of the data set and Perform the preprocessing that require for the model.
Note: You can see many categorical variables with a high number of unique values. Therefore do not keep dropping the variables as the first option, try to create new variables or perform any other feature engineering methods.

Step 6: Perform statistical hypothesis testing on features to get an idea of whether features are impacting the target variables.

Step 7: Split the dataset into train and test data sets and Perform the scaling on both sets if necessary.

Step 8: Build the base model.

Step 9: Understand how the model is performing, Perform feature engineering again if needed. Do feature selection. Try with various models like a parametric and nonparametric models. Once you choose the final model, rebuild the model with best parameters.
Note: If you are performing with Linear models, check the model is fulfilling the assumptions.

Step 10: Based on your understanding of the model and EDA analysis, Explain the business understanding.

Section B: ***Title: Classification***

Problem Statement:

The model that you built in the above case study has gone to deployment and Zomato has been impressed with your data analysis and Zomoto has been believing that your analysis going to be impactful. Now Zomato has been observing the orders happening online and offline, Due to offline orders, Zomato is not able to attract customers with diverse items and offers, and the user subscription also getting low. so it has decided to give you the project on the same. Now the problem statement is that Zomato wants to know whether the customer would order the orders online or offline so that Zomato can take further strategies to improve the online order.

Objective:

The Aim is to classify the orders that have been ordered online and offline. And identify the patterns that lead to orders online orders as well as offline. Your model should be able to classify the classes effectively.

Steps that are to be followed:

Step 1: Understand the problem statement, Identify the metric for the problem.

Step 2: Use the data sets that have been cleaned data types and missing values in **section A step 3.**

Step 3:

i) Explore the data and find the hidden patterns in the data that affect your objective(target variable).

ii) Find the features that would impact the target variable.

Step 4: Do the preprocessing in a way that impacts the models that you going to build.

Note: If you feel the few preprocessing steps that you did in section 1, feel free to use those steps.

Step 5: Prove statistically about step 2 point ii.

Step 6: Split the data into train and test sets.

Step 7: Build the base model and Identify your metric based on the problem.

Step 8: Try to improve your metric by trying different models, changing the feature engineering methods, feature selection, etc.

Step 8: Choose the final model and tune the model.

Step 9: Write your business interpretation. And explain the reasons behind choosing the final model and how the chosen model performs well compared to other models.