Group Name: Individual

Name: Maria Contractor

Email, Country, College, Specialization: mariacontractor5@gmail.com , USA , University of Central Florida, Data Science

Batch Code: LISUM22

Submission Date: July 26, 2023

Submitted to: Data Glacier

Group Project: Week 8

Instructions:

1. Problem description

2. Data understanding

3. What type of data you have got for analysis

4. What are the problems in the data (number of NA values, outliers , skewed etc)

5. What approaches you are trying to apply on your data set to overcome problems like NA value, outlier etc and why?

Steps

1. Problem Description

ABC Bank wants to sell its term deposit product to customers and before launching the product they want to develop a model which help them in understanding whether a particular customer will buy their product or not (based on customer's past interaction with bank or other Financial Institution). To solve this problem, we will need to predict whether or not the client will subscribe to a term deposit.

2. Data Understanding & Types of Data for Analysis

The ABC Bank has data that includes bank client data and other attributes. The data includes 21 attributes/columns and has 41188 entries. Some of the attributes include age, job, marital status, housing, and other demographic information. It also includes economic data like price indexes and outcomes of previous campaigns. This data is either categorical or numeric. There is an output variable, which is a binary data value (Y/N).

Tabular data details: bank-additional-full.csv

Total number of observations	41188
Total number of files	1
Total number of features	21
Base format of the file	.csv
Size of the data	5.56 MB

3. Problems in Data & Approaches to Overcome it

This data does not include any NA/missing values. However, there are outliers and skewness in the data. Furthermore, there are duplicates in the data. To solve duplicates, we can remove the duplicates in the data. To solve the problem of outliers, we can remove the outliers in the data. With the use of Jupyter notebook and python, we can solve these issues.