PHOENIX
INNOVATIONS

Problem Statement

For this coding assignment, you are provided with data in an Excel sheet (Dataset.xlsx). You need to use this data to answer the following questions using the Python programming language. The goals of this assignment are to understand your thought process while solving a data science problem and to evaluate the accuracy of your approach.

Data details

The dataset contains a total of 18 columns and 12,330 rows. It is a mixed dataset containing numerical and categorical variables. The dependent variable is "Target," which is a binary variable. You need to develop a model to predict it.

Questions

1. Perform Exploratory Data Analysis (EDA) to uncover patterns in the data. Some example questions that could be answered are:
    a. What is the distribution of the Target variable?
    b. What different patterns are present for visiting customers?
    c. What is the correlation between exit and bounce rates?

    You can develop more questions and answer them using analysis/visualizations. This will help us understand your data analysis skills.

2. Identify which variables are important for predicting the Target variable.

3. Develop machine learning or deep learning models to predict the Target variable. Compare the models and select the best one for this dataset.

Submission

Please follow these guidelines while submitting your solution:

1. Submit an executable Jupyter notebook. We will execute the received notebook to check the validity of the code.

2. Prepare a short report highlighting your design decisions and thought process behind selecting algorithms/methods.

3. Include a short section (200-300 words) on the impact of the results on the business.

4. Your submission should be a compressed file containing the dataset, Jupyter notebook, and report.