**Project -7**

**Exploratory Data Analysis and Logistic Regression on Banking Data**

**Objective:**

In this project is to perform exploratory data analysis (EDA) on a banking dataset and create visualizations to better understand the patterns and relationships within the data. Additionally, we implement a Logistic Regression model to predict whether a customer use the term of deposit or not.

**Tools and Libraries Used:**

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

from sklearn.model\_selection import train\_test\_split

from sklearn.linear\_model import LogisticRegression

from sklearn.metrics import classification\_report, confusion\_matrix

**Data Source:**

we used the data in this project in the 'bank-full.csv' file. It contains all information about customers and whether they used to a term deposit.

**Exploratory Data Analysis:**

**Loading Data**:

We used the Pandas library to load the data.

**EDA**(Exploratory Data Analysis):

Checked the first 5 rows of the dataset to get an initial understanding.

Explored the columns and their names using data.columns.

Examined general information about the dataset using data.info.

Checked the data types of each column with data.dtypes.

**Plotting or Visualization:**

Created a count plot to visualize the distribution of customers who choose the term deposit ('yes' or 'no').

Set title: 'Distribution Deposit'.

**Correlation Matrix:**

Computed the correlation matrix to understand the relationships between numerical features.

Created a heatmap for better visualization.

Set title: 'Correlation Matrix'.

**Saving Plots:**

Saved the created plots, such as the count plot and correlation heatmap, as images ('graph.png').

Logistic Regression Model:

**Data Preparation:**

Split the data into features and target variable (X and y).

Divided the data into training and testing sets using train\_test\_split.

Logistic Regression:

Trained the model on the training set.

**Model Accuracy:**

Classification report and confusion matrix to evaluate the model's performance.

Checked the accuracy of the model.

Saving Outputs:

**Zip Files:**

Created zip files 'Bankfull.zip' for graphs

Zipped the 'graph.png' image into 'Bankfull.zip'.

Zipped the Jupyter Notebook files ('Bank\_prj\_7.ipynb' and 'Advance\_Proj\_7.ipynb') into 'Bankdata.zip'.

**Conclusion:**

This project helped into the banking data through exploratory data analysis and builds a predictive model using logistic regression. The saved plots and zip files make it convenient to share and reproduce the analysis.