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).

Business understanding:
Bank wants to use ML model to shortlist customer whose chance of buying the product is more so that their marketing channel (tele marketing, SMS/email marketing etc.) can focus only to those customers whose chance of buying the product is more.

This will save resource and their time (which is directly involved in the cost (resource billing)).

Data Intake Report

Name: Bank Marketing (campaign)
Report date: 09/02/2022
Internship Batch: LISUM11: 30
Version: 1.0
Data intake by: Priyadarshani Kamble
Data intake reviewer: NA
Data storage location: https://archive.ics.uci.edu/ml/datasets/Bank+Marketing/bank-additional-full.csv

Tabular data details:

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	5699kb

Note: Replicate same table with file name if you have more than one file.

Data Wrangling:

Step 1 - Import data and Explore the data

- The data is available in csv file.
- The Key attributes include age, job, marital, education, default, balance, housing loan, contact, day, month, duration, campaign, pdays, previous, poutcome, y.
- There are 41188 records and 21 Features.
- There are 10 numeric columns and 11 Categorical Columns
- Checked the no of unique values in each columns. If the feature has constant value or 1 value then such columns can be dropped as they do add any value in model building. The dataset does not any such columns.

Step 2-Data Cleaning:

- Checked for missing values. No missing data found.
- 12 duplicate rows found. Deleted the duplicate records.

I dropped the unwanted features, handled missing data, removed outliers.

Below columns were dropped as per the findings in EDA :

- emp_var_rate
- nr_employed
- default
- pdays

Outlier removing:
Handle the outliers by taking 99 percentile data into consideration for the columns

- Age
- Duration
- Cons_conf_idx
- Campaign