The Data Detectives

https://github.com/tassalor1/Bank-Term-Deposit-Prediction

Team member's details:

Name: Vincent Weng
Email: vweng@umich.edu
Country: United States
College/Company: University of Michigan
Specialization: Data Science, NLP

Name: Kefan Yu
Email: ky285@georgetown.edu
Country: United States
College/Company: Georgetown University
Specialization: Data Science, Data Analyst

Name: Connor Walker
Email: walkerconnor14@gmail.com
Country: England
College/Company: N/A
Specialization: Data Science

Name: Virginia Mullins
Email: vlmullin@gmail.com
Country: Lebanon
College/Company: N/A
Specialization: Data Science

Problem Description:
ABC Bank wants to predict whether a customer will subscribe to their term deposit product based on past interactions. They aim to develop a machine learning model to identify customers who are more likely to purchase the product. In other words, they want to shortlist customers whose chances of buying the product is more.

Business Understanding:
After the development of a ML predictive model, we can then make assumptions about the class of customers more likely to purchase the product. This will help the bank focus its marketing efforts on those customers during future marketing campaigns. By narrowing the campaign scope, the bank can reduce costs, save resources, and improve profit margins.

Project Lifecycle:
● Week 7 (ddl: Apr 19): Create Github repo and finish data intake report
● Week 8 (ddl: Apr 26): Understand the data and write approaches of how to process it

● Week 9 (ddl: May 2): Perform data cleaning and transformation
● Week 10 (ddl: May 9): Perform EDA
● Week 11 (ddl: May 16): Create EDA presentation to audiences in business area
● Week 12 (ddl: May 23): Construct models and choose appropriate ones to explain predictions
● Week 13 (ddl: May 30): Select the best solution and create presentation slides

Data Intake Report

Name: Bank Term-Deposit Purchase Prediction
Report date: April 15th 2023
Internship Batch: LISUM19
Version:1.0
Data intake by: The Data Detectives
Data intake reviewer: Virginia Mullins
Data storage location: https://github.com/tassalor1/Bank-Term-Deposit-Prediction

Tabular data details:

Bank-Full.csv

Total number of observations	45211
Total number of files	1
Total number of features	17
Base format of the file	.csv
Size of the data	4547 KB

Bank-Full-Additional.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	4814 KB

Proposed Approach:

Since there are no unique incidence qualifiers (such as name or customer ID), the data is assumed to contain no duplicates or a nearly negligible number of duplicates. There is no missing or null data in the set. However, since much of the data was generated via survey response from customers, some customers declined to give information due to privacy concerns. For the initial data intake, we are treating these non-responses as their own separate category (such as "Unknown" or "None"). Whether these need to be replaced with something more suitable (such as mean, mode, or dropped all together), will be decided during the data cleaning phase.