DVT Mini Project- Approach

A Portuguese banking institution wants to understand the factors that influence customers to subscribe to term deposits. They have provided us with their direct marketing campaign data, which includes customer demographics, previous interactions with the bank, and whether or not the customer subscribed to a term deposit. Our task is to analyze this data and provide insights into customer behavior and campaign outcomes, which will help the bank improve its future marketing strategies. The dataset contains 45,211 instances and 16 attributes, including customer demographics, previous interactions with the bank, and whether or not the customer subscribed to a term deposit.

Here are the details of the attributes in the dataset:

- age: The age of the customer (numeric).
- job: The type of job the customer has (categorical: 'admin.', 'blue-collar', 'entrepreneur', 'housemaid', 'management', 'retired', 'self-employed', 'services', 'student', 'technician', 'unemployed', 'unknown').
- marital: The marital status of the customer (categorical: 'divorced', 'married', 'single', 'unknown').
- education: The level of education of the customer (categorical: 'basic.4y', 'basic.6y', 'basic.9y', 'high.school', 'illiterate', 'professional.course', 'university.degree', 'unknown').
- default: Whether or not the customer has credit in default (categorical: 'no', 'yes', 'unknown').
- balance: The current balance of the customer (numeric).
- housing: Whether or not the customer has a housing loan (categorical: 'no', 'yes', 'unknown').
- loan: Whether or not the customer has a personal loan (categorical: 'no', 'yes', 'unknown').
- contact: The type of contact the customer was reached through (categorical: 'cellular', 'telephone').
- day: The day of the month the customer was last contacted (numeric).
- month: The month the customer was last contacted (categorical: 'jan', 'feb', 'mar', ..., 'nov', 'dec').
- duration: The duration of the last contact with the customer, in seconds (numeric).
- campaign: The number of contacts performed during this campaign for this customer (numeric).
- pdays: The number of days that passed by after the customer was last contacted from a previous campaign (numeric).
- previous: The number of contacts performed before this campaign and for this customer (numeric).
- y (target variable): Whether or not the customer subscribed to a term deposit (binary: 'yes', 'no').

Perform the following operations: -

Data Preparation

- Import the bank_data Dataset into Power BI: (Hint: by default indexing column will be added to uniquely identify each row/record as there are no unique identifiers in the table, rename this respective column and make use of it for the python visuals so that no record should be considered as duplicate) **(3 Marks)**

- Check for any null values or inconsistencies in the data and refill missing values if any using python.
  (Hint: The column age contains missing values which can be imputed based on the average age of

different job types eg: customer whose job is `admin` have an average age of 39, so the places where the age is missing for a customer whose job is `admin` we can insert 39 as age.) **(5 Marks)**

Analysis:

- What is the proportion of customers who subscribed to a term deposit, broken down by Job? Create a pie chart to display the same. You can create using the drag-and-drop option of power bi or u can make use of `python visual` and write the python code for the same in power bi. **(3 Marks)**
- What is the most common education level among customers in the dataset? Create a bar chart to show the number of customers by education level. **(3 Marks)**
- Create a box plot for detecting outliers in the age column using python visual (hint: make sure you add a unique identifier as well along with the age column while working with python visual as it will delete the value of age where the age is same, considering its duplicate). **(3 Marks)**
- Create a Line chart showing the total number of campaigns done in different month with respect to contact type(i.e contact column).**(3 Marks)**

Data Preparation
Import the bank_data Dataset into Power BI:

Approach: Use the data import functionality in Power BI to load your dataset.

Analysis

What is the proportion of customers who subscribed to a term deposit, broken down by Job?

Approach: Create a pie chart using Power BI's visualization tools to represent the proportion of customers who have subscribed to a term deposit by job category. Use the drag-and-drop interface to set the job category as the legend and the subscription status as the value, ensuring the pie chart accurately reflects the proportions.

What is the most common education level among customers in the dataset?

Approach: Construct a bar chart to visualize the number of customers by education level. This will help identify the most common education level among the customers. Use the visualization tools to drag the education level to the axis and set the count of records as the value.

Create a box plot for detecting outliers in the age column using python visual:

Approach: Add a Python visual to your report and input the age data along with the manually added unique identifier. Write a Python script to create a box plot, which will be used to detect outliers in the age column.

Create a Line chart showing the total number of campaigns done in different months with respect to contact type:

Approach: Develop a line chart to display the total number of campaigns conducted each month, broken down by contact type. Use the visualization tools to place the month on the axis, the campaign count in the values area, and the contact type in the legend, allowing for a temporal analysis of campaign activities by contact method.