Data Analysis Hackathon

Banking Sector.

Problem Statement:

The bank provides financial services/products such as savings accounts, current accounts, debit cards, etc. to its customers. In order to increase its overall revenue, the bank conducts various marketing campaigns for its financial products such as credit cards, term deposits, loans, etc. These campaigns are intended for the bank's existing customers. However, the marketing campaigns need to be cost-efficient so that the bank not only increases their overall revenues but also the total profit. You need to apply your knowledge of EDA on the given dataset to analyse the patterns and provide inferences/solutions for the future marketing campaign.

The bank conducted a telemarketing campaign for one of its financial products 'Term Deposits' to help foster long-term relationships with existing customers. The dataset contains information about all the customers who were contacted during a particular year to open term deposit accounts.

What is the term Deposit?

Term deposits, also called fixed deposits, are the cash investments made for a specific time period ranging from 1 month to 5 years for predetermined fixed interest rates. The fixed interest rates offered for term deposits are higher than the regular interest rates for savings accounts. The customers receive the total amount (investment plus the interest) at the end of the maturity period. Also, the money can only be withdrawn at the end of the maturity period. Withdrawing money before that will result in an added penalty associated, and the customer will not receive any interest returns.

Your target is to do end to end EDA on this bank telemarketing campaign data set to infer knowledge that where the bank has to put more effort to improve it's positive response rate.

The data set for the project was downloaded from the website **"UCI Machine Learning Repository"** into an excel spreadsheet so that we could convert it into a CSV file. The data set is related to a Portuguese banking institution's marketing campaign. The marketing campaigns were based on telemarketing. The contact information includes date, time and number of contacts made to a customer in order to get the response of "yes" or "no" to their term deposit. The whole data set is the bank's client database consisting of 17 different variables/attributes which is elaborated below.

This is a real dataset collected from a Portuguese bank that used its own contact-center to do direct marketing campaigns to motivate and attract the clients for their term deposit scheme to enhance the business.

Variable	Description
customerid	ID of Each Customer
age	(numeric)
salary	(numeric)
balance	(numeric)
marital	marital status (categorical: 'divorced','married','single')
jobedu	Type of Job and Education
targeted	Target Person (Binary: Yes, NO)
default	has credit in default? (categorical: 'no','yes')
housing	has a housing loan? (categorical: 'no','yes')
loan	has a personal loan? (categorical: 'no','yes')
contact	contact communication type (categorical: 'cellular','telephone')
day	contact day of the months (numeric)

month	last contact month of year (categorical: 'jan', 'feb', 'mar', ., 'nov', 'dec')
duration	last contact duration, in seconds (numeric)
campaign	number of contacts performed during this campaign and for this client (numeric, includes last contact)
p_days	number of days that passed by after the client was last contacted from a previous campaign
previous	number of contacts performed before this campaign and for this client (numeric)
p_outcome	outcome of the previous marketing campaign (categorical: 'failure','other','success')
response	has the client subscribed to a term deposit? (binary: 'yes','no')

Instructions:

1. Data Importing and basic operation for data understanding steps

2. Data cleaning challenge of the Missing Value Data and Outliers treatment

3. Data Cleaning Challenges handling of regex

4. Working on derived metric or new columns

5. Univariate Analysis (Statistical summarization)

6. BiVariate Analysis

7. Visualization plots

  1. Univariate Plots
      - Continuous variable plots
      - Categorical Variable Plots

2. BiVariate Plots
    - Continuous vs Continuous Plots
    - Continuous vs Categorical Plots
    - Categorical vs Categorical Plots

8. Other Challenges like
    - Time and Date based analysis and plots

9. Finally you should be able to interpret the results and comment on them throughout the analysis. (Mandatory if not we won't Consider your solutions).

10. Based on the understanding of the data you should be able to identify important relation between the columns through analysis and possible visualization

NOTE: Don't plagiarize the code during hackathon. If found plagiarizing you won't be allowed for further hackathons and internship opportunities.

ALL THE BEST !!!