Bank Marketing

1. Introduction:
   * Background:

In recent years, the banking industry has become increasingly competitive, with banks vying for customers in a crowded marketplace. As a result, it has become more important than ever for banks to develop effective marketing strategies that target the right customers with the right messages. One key challenge that banks face is determining which customers are most likely to subscribe to specific products, such as term deposits, which are a type of savings account with a fixed term and interest rate.

* + Purpose:

The purpose of this case study is to analyze the "Bank Marketing" dataset and develop insights that can help a bank improve its marketing efforts related to term deposits. Specifically, we will explore the factors that influence whether a customer will subscribe to a term deposit and identify the characteristics of customers who are most likely to convert. By doing so, we hope to help the bank develop more effective marketing strategies and ultimately increase its sales of term deposits.

* + Research Questions:

To achieve this purpose, we will explore the following research questions:

* + - What are the most important factors that influence whether a customer will subscribe to a term deposit? We will use statistical techniques such as logistic regression to identify the key drivers of term deposit subscriptions, and we will explore how these factors interact with each other.
    - What are the characteristics of customers who are most likely to convert, and how can these customers be targeted more effectively? We will use clustering and decision tree analysis to identify groups of customers who are most likely to subscribe to a term deposit, and we will explore how these groups differ in terms of demographic, behavioral, and attitudinal characteristics. We will also explore which marketing channels and messages are most effective for each group of customers.
  + Significance:

The insights developed in this case study have the potential to be highly valuable for the bank, as they can inform more effective marketing strategies and ultimately lead to increased sales of term deposits. Moreover, the techniques used in this case study are applicable to other marketing contexts beyond banking, making the findings relevant to a wider range of industries. Overall, this case study demonstrates the power of data analysis and machine learning techniques to drive business value and improve marketing outcomes.

1. Business Problem:
   * Target Audience:

The bank in this case study is looking to target customers who are most likely to subscribe to a term deposit, which is a type of savings account with a fixed term and interest rate. Term deposits are an important source of revenue for the bank, as they provide a stable and predictable source of funding that can be used to support the bank's lending activities.

* + Marketing Strategies:

The bank currently uses a variety of marketing strategies to reach potential customers, including direct mail, telemarketing, and email campaigns. However, the bank has limited resources and wants to focus its efforts on the most effective strategies. To do so, the bank needs to better understand which marketing channels and messages are most effective for different types of customers, and how to tailor its marketing efforts to maximize conversions.

* + Sales Goals:

The bank's sales goal is to increase the number of term deposits it sells, with a particular emphasis on targeting new customers who have not previously subscribed to a term deposit. The bank wants to achieve this goal in a cost-effective manner, by identifying the most effective marketing channels and messages and focusing its efforts on customers who are most likely to convert. Moreover, the bank wants to retain its existing customers and encourage them to renew their term deposits when they mature, by developing targeted retention strategies that address the specific needs and preferences of each customer segment.

* + Challenges:

There are several challenges that the bank faces in achieving its sales goals. For example, there is a high degree of competition in the banking industry, and customers are bombarded with marketing messages from multiple banks and financial institutions. Moreover, customers have different preferences and needs, and may be more or less receptive to different types of marketing messages. To overcome these challenges, the bank needs to develop a deep understanding of its customers and tailor its marketing efforts accordingly. This requires a data-driven approach that uses advanced analytics techniques to identify the key drivers of term deposit subscriptions, as well as the characteristics of customers who are most likely to convert. By doing so, the bank can develop more effective marketing strategies that increase conversions and generate higher revenue.

1. Data Description:
   * Data Source:

‘bank-marketing.csv’ - The "Bank Marketing" dataset is sourced from a direct marketing campaign conducted by a Portuguese banking institution between May 2008 and November 2010.

* + Data Characteristics:

The dataset contains 41,188 records of bank customers who were targeted in a marketing campaign for a term deposit. The data includes a range of demographic, behavioral, and attitudinal variables, as well as information on the customers' interactions with the bank and the outcomes of the marketing campaign. The dataset has a mix of categorical and numerical variables, and there are some missing values that will need to be imputed or dropped.

* + Data Dictionary:

The dataset includes the following variables:

* + - Bank customer information: age, job, marital, education, default, housing, loan, contact
    - Associated with the most recent interaction of the ongoing marketing campaign: contact, month, day\_of\_week, duration
    - Additional features: campaign, pdays, previous, poutcome
    - Socioeconomic contextual features: emp.var.rate, cons.price.idx, cons.conf.idx, euribor3m, nr.employed
    - Target variable: y

Details of Variables:

* “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 the customer has credit in default (categorical: 'no', 'yes', 'unknown')
* “housing”: whether the customer has a housing loan (categorical: 'no', 'yes', 'unknown')
* “loan”: whether the customer has a personal loan (categorical: 'no', 'yes', 'unknown')
* “contact”: the contact communication type (categorical: 'cellular', 'telephone')
* “month”: the month of the year when the customer was last contacted (categorical: 'jan', 'feb', 'mar', ..., 'nov', 'dec')
* “day\_of\_week”: last contact day of the week (categorical:  
  'mon', 'tue', 'wed', 'thu', 'fri')
* “duration”: the duration of the last contact in seconds (numeric)
* “campaign”: the number of contacts performed during this campaign and for this customer (numeric)
* “pdays”: the number of days that passed by after the customer was last contacted from a previous campaign (numeric, -1 means the customer was not previously contacted)
* “previous”: the number of contacts performed before this campaign and for this customer (numeric)
* “poutcome”: the outcome of the previous marketing campaign (categorical: 'failure', 'nonexistent', 'success')
* “emp.var.rate”: employment variation rate - quarterly indicator  
  (numeric)
* “cons.price.idx”: consumer price index - monthly indicator (numeric)
* “cons.conf.idx”: consumer confidence index - monthly indicator  
  (numeric)
* “euribor3m”: euribor 3 months rate - daily indicator (numeric)
* “nr.employed”: number of employees - quarterly indicator (numeric)
* “y”: whether the customer subscribed to a term deposit (binary: 'yes', 'no')

1. Data Science Process:
   * Data Preparation:
     + In this stage, we will clean and preprocess the data to prepare it for analysis and modeling. This includes handling missing values, encoding categorical variables, scaling numerical variables, and handling outliers if necessary.
     + We will also split the data into training and testing sets, with the majority of the data being used for training the model, and a smaller portion reserved for testing the model's performance.
   * Exploratory Data Analysis (EDA):
     + In this stage, we will perform exploratory data analysis to gain insights into the data and identify patterns or relationships between variables. We will use various statistical and visualization techniques to examine the distribution of the variables, the correlation between the variables, and any potential outliers or anomalies.
   * Feature Engineering:
     + In this stage, we will create new features or transform existing features to improve the performance of the model. This includes creating interaction terms, transforming variables, and scaling or standardizing the features.
     + We will also conduct feature selection to identify the most important variables that contribute to the prediction of term deposit subscription. We will use techniques and feature importance from the selected model to select the most relevant features.
   * Model Selection:
     + In this stage, we will experiment with several classification models, including Logistic Regression, Decision Tree, Random Forest, and Gradient Boosting, and evaluate their performance using cross-validation and various performance metrics such as AUC, precision, recall, and F1 score.
     + We will select the best performing model for further analysis based on its performance on the test dataset.
   * Model Evaluation:
     + In this stage, we will evaluate the performance of the selected model on the test dataset, which was not used during the training or feature selection phase, to assess the generalizability of the model. We will also visualize the results using various metrics such as Confusion Matrix, ROC curve, and precision-recall curve to better understand the performance of the model and identify any potential areas for improvement.
   * Model Deployment:
     + In this stage, we will deploy the selected model into production and integrate it into the bank's marketing system. We will also monitor the performance of the model and update it as needed to ensure that it continues to provide accurate predictions.

By following this data science process, we can build a robust and accurate predictive model that can provide valuable insights into the factors that influence customers' decisions to subscribe to a term deposit and help the bank improve its marketing strategies.