Name:
Student ID:
Course:
University:

Abstract

Data is king in today's world. If you use it wisely, it can have a big impact on your business; if you don't, you'll be left behind in no time in this fast-paced world. And one of the ways that an organization can increase its business success is by efficiently capturing and analysing consumer data in order to improve the customer experience.

The finance industry is one of the sectors that has been changed the most by recent Machine Learning advances. Predicting asset prices is one thing, but predicting whether a consumer would subscribe to a term deposit is another. Machine learning has the potential to be a very useful method for increasing profitability.

For our Research Data has been collected from a Portuguese banking institution's which conducted a survey via direct marketing campaigns (phone calls). The classification aim is to determine whether or not the client can subscribe for a term deposit (target variable y). Parameters such as Job type, Marital Status, age, Education Levels etc. have been represented using Data Visualization in Python, EDA, etc. A comparative study is often performed to determine which variables have the greatest impact. We've even done certain data analysis as part of this. Finally, we have focused on the limitations and common obstacles that have been encountered through the surveyed works in order to create a vision on this.

1.0 Introduction:

We are given data from a Portuguese banking institution's direct marketing campaigns (phone calls). The classification aim is to determine whether or not the client can subscribe for a term deposit (target variable y).

Our research's problem statement is to demonstrate the significance of different aspects of this marketing campaign, as well as how people are reacting to it in terms of subscribing to the product.

In order to conclude in response to our problem statement, the proposed method would include processing a given dataset, performing data cleansing and per-processing on the data, and then using data visualization techniques.

Research Workflow
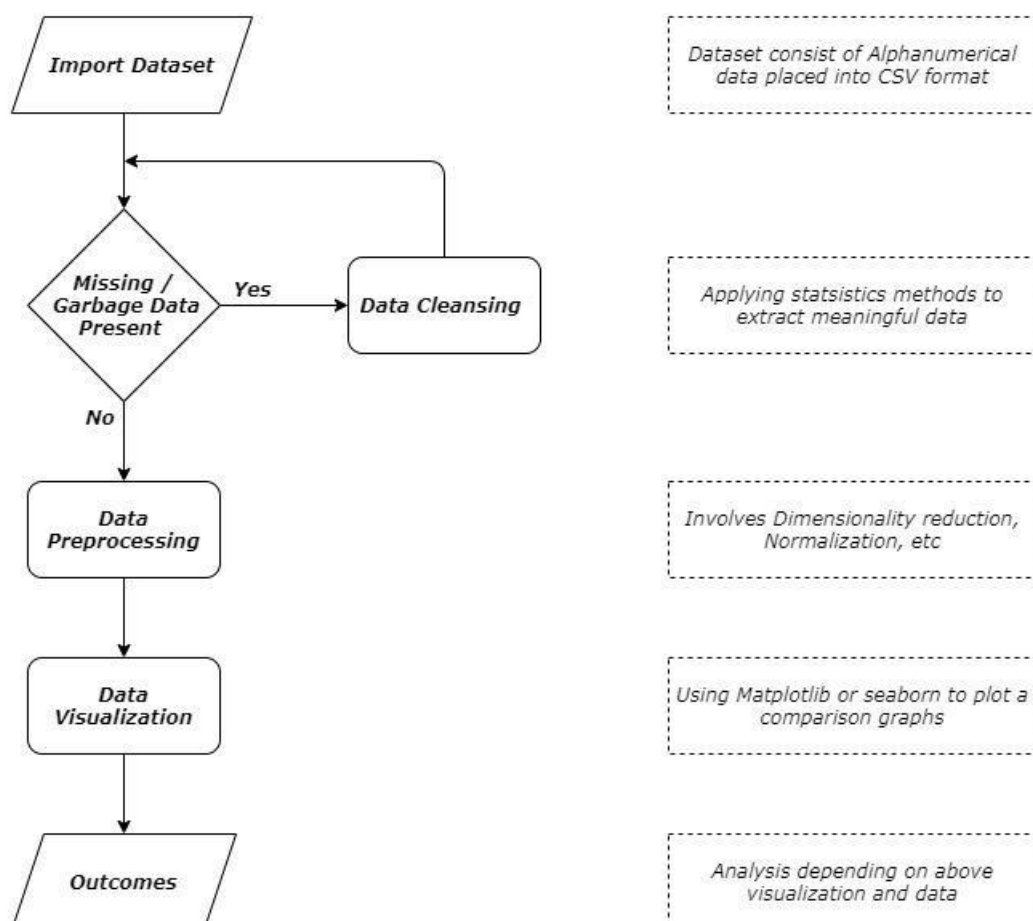
1.0.1 Research Questions:

1. How will age Group will be one of the factors used to find a pattern?
2. How much % of people are already undergone this marketing campaign?
3. How Education level performs major role to have enough knowledge when comes in under such survey.
4. The ratio of customers having various types of loan already such as Housing Loan or Personal Loan.
5. What is the outcome of the previous campaigns held in order to increase the product Subscription rate?
6. The Product subscription varying according to Marital status which influences Customers to buy the Bank Term Deposit.
7. How Various Job types at their levels affects customers to subscribe the Product.
8. How may times the same customer has been approached and what is the frequency of the same?
9. What can be the best Contact Medium which helps to attract customers to subscribe the Product?
10. Overall Customer Product Subscription variation depending on all the factors provided.

1.0.2 Objective:

We identify the major challenges involved while conducting the marketing campaign survey for the Banking Institution. This survey completely based on the Phone calls in order to find promising directions for further analysis. This review also elaborates the benefits of data science which help us to analyse people's engrossment towards such marketing schemes based on different parameters we have been given in the dataset.

1.0.3 Research Methodology:

We will follow various data science activities in order to achieve our objective which we will be elaborating in further points.

1.0.4 Expected Outcome:

We want to conclude the final Product Subscription Success and Failure ratio that is being taken on the basis of phone call survey and the way people are responding to it throughout all the campaigns held.

2.0 Tools, Datasets and Preliminary Analysis:

Data analysis is nothing more than the process of examining collected data in order to derive valuable information using various analysis methods in order to uncover trends that assist us in making decisions.

2.0.1 Programming Language and Tools

Python Intro

Data manipulation, processing, cleaning, crunching, and modelling are all part of the Python for Data Analysis system.

It's also a useful, up-to-date introduction to scientific computing in Python that's optimized for data-intensive applications and includes automated memory management.

Libraries:

NumPy:

NumPy (Numerical Python) is an open-source Python add-on module that allows you to use random number generators, linear algebra routines, Fourier transforms, and other functions.

NumPy vectorization, indexing, and NumPy arrays are much faster and more flexible than the other built-in Python data structures for storing and manipulating data.

Pandas

Where datasets were combined and duplicated records were removed, Pandas provides rich data structures and functions designed to make working with structured data quick, simple, and expressive.

Pandas can manipulate data in a variety of ways. Panda has a number of features, including the ability to reshape, slice, and dice data, as well as perform aggregations.

Matplotlib / Seaborn Paragraph

Matplotlib is the most prominent Python library for plotting and discovering data. It can draw various plots and other visual analytics features such as zoom, pan, and update.

Seaborn is a data visualisation library written in Python.

It's also very interactive, allowing us to create statistical graphical representations that aid in understanding data trends.

2.2   Dataset:

1. The source dataset we have given is related to direct marketing campaigns of a Portuguese banking institution. The marketing campaigns were based on phone calls. Often, more than one contact to the same client was required.
2. There are a total 21 columns in the given dataset which is structured data consisting 20 input feature and one output feature.
3. The type of data we have collected is having categorical, continuous features combined.
4. There is only one column which is dependent as dependent while rest of the columns we can treat as independent columns.
5. Similarly, there are other columns which have continuous data like age, Month, duration, campaign which will come under independent variables.
6. Also, columns such as job, marital status, education, loan status will be considered categorical data.
7. One of the major advantages is given dataset do not consist any garbage or missing value.

2.3   Initial Data Analysis:

Applying statistical and logical methods to extract data, normalize the data structure, represent data using images, charts, graphs, and plots, and analyze statistical forecasts, likelihood of trends or patterns to draw concrete conclusions are all part of data analysis.

Two main methods for data analysis:
1. Qualitative Analysis
2. Quantitative Analysis

In our research paper we are focusing on the Qualitative analysis method as we have an input dataset where target feature having values as 'Yes' or 'No. Hence, this will be treated as Binary classification.

2.3.1 Data Cleansing:

1. There are no Null or Missing values in the given dataset.   We   can   check   null   value existence by following command:

df.isnull()

2. As we can see there are some duplicate rows in our datasets, we will remove the duplica tes before moving forward.

df = df.drop_duplicates()

2.3.2 Data Pre-processing:

Data pre-processing is done in order to model the input data in a format that can be easily analyzed.
Techniques are as follows:

Aggregation: It implies merging two or more attributes into a single attribute, resulting in data reduction and faster data processing.
For example, in our dataset, we can combine columns like housing loan and personal loan into one attribute as they both are showing kind of loan taken by the customers.

Sampling: Sampling is nothing but the subset of original data. I can be represented by means of data which will be close to the original value only. Thus, we can apply sampling on age column to make age groups having buckets which will make easier to process further.

3.0   EDA (Exploratory Data Analysis):

EDA is the approach we can take in terms of using various statistical and logical techniques to define a certain trend in current data and then building a potential data forecast using graphs, maps, and plots, among other things.

Different techniques to be followed by EDA are as follows:
1. Finding Central Tendency of the data (Mean, Mode, Median)
2. Finding skewness of the data (Asymmetry n the data)
3. Finding Standard deviation and variance.
4. Covariance and correlation of data.

All of the above strategies will be illustrated using a variety of plots, maps, and graphs, which we will go through in greater detail in the Data visualization section.

3.1   Descriptive Statistics:

A brief overview of the input dataset is given by descriptive statistics. It will assist us in articulating attribute characteristics. As a result, findings will be generalized into particular patterns or trends.
Descriptive statistics is a sub part of EDA only.

It Consists following sub areas:
• Mean: It is the average of total numerical values.
• Median: It is the midpoint of the dataset or the centre of the sample set.
• Mode: Most often used value of the dataset.
• Percentage: It is the position of values in the dataset.
• Frequency: the number of times a value is found.
• Range: Difference between highest and lowest values.

In our dataset, we can represent age group distribution using histograms or box plot which comes under categorical data only. Similarly, we can present Outcome variation of campaigns held based using pie charts.

3.2   Data Visualization:

The graphical representation of observed trends or patterns is known as data visualization. We also conducted pre-operations on the mapping of graphics marks and input data values. It's highly interactive in order to develop a clear viewpoint on our research.

The various visualization approaches we used to reflect our research question analysis are listed below:

1. Comparative ratio varying according to the job type of customers towards subscribing the product is shown by **horizontal column chart.**

Interest rate to buy Bank Term Deposit subscription depending on the Job Types
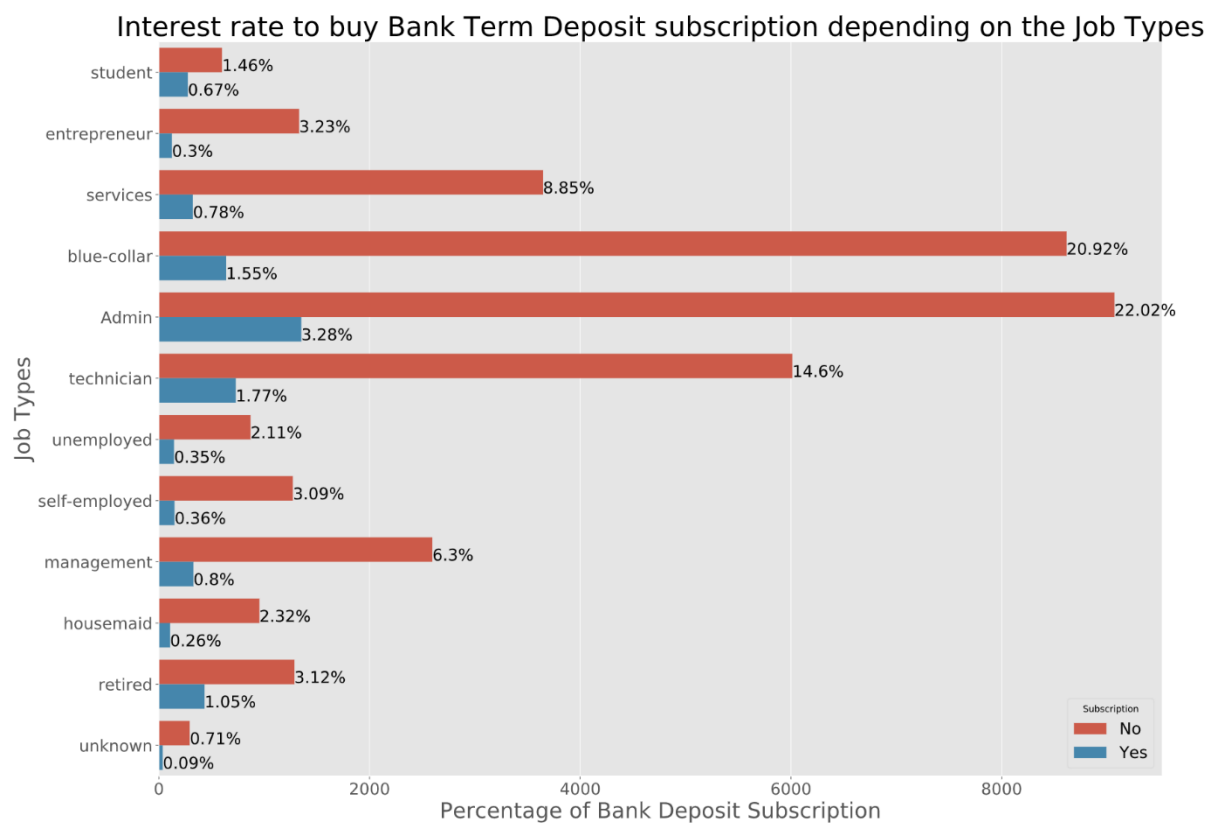
2. **Funnel Chart** showing product Subscription trend with respect to Education levels among Customers.


Thus, it helps us to understand awareness of such product subscription schemes related to education knowledge.

3. Variation of Loan taken by Customers which affects to subscribe the Product presented by plotly **Gauge Chart**.

4. **Bar chart along with the Table** is showing the Contact medium used while taking the survey to attract people towards subscribing the product.

Contact Type affecting to draw clients attention to subscribe Bank Term deposit
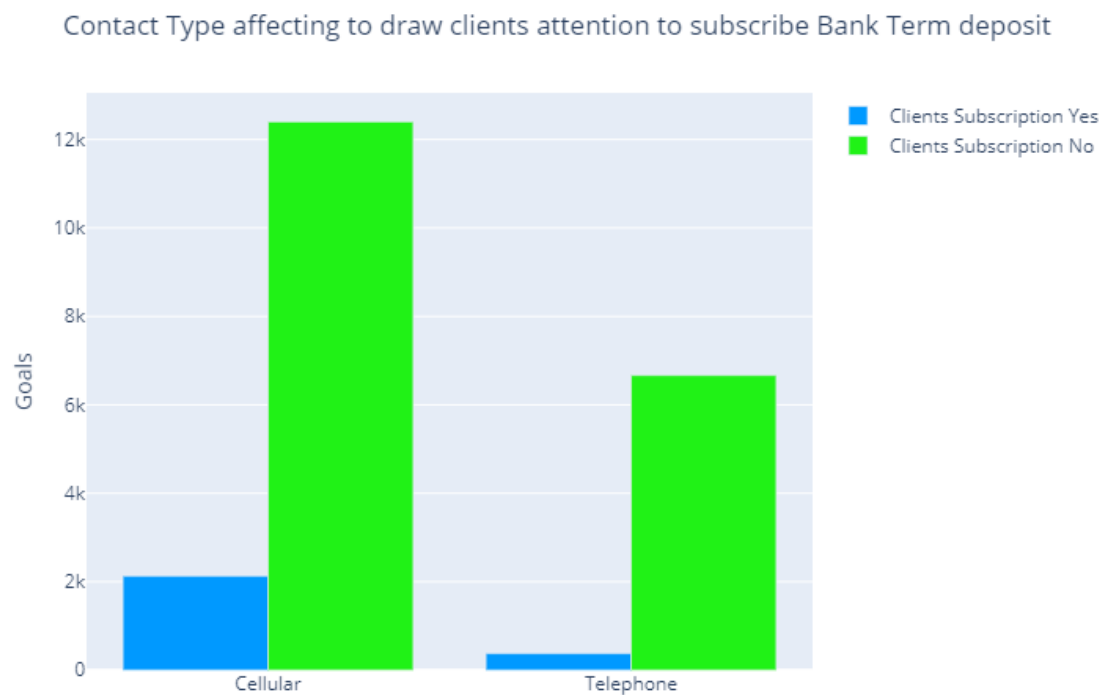
Contact_Type	Cellular	Telephone
Yes	2130	377
No	12402	6667

This shows us Cellular Contact medium is more reliable than the Telephonic one. Also, graph is showing both bar chart along with table of the information.

5. The campaigns which held before having different outcomes resulting to buy the product displayed using **Pie Charts**.
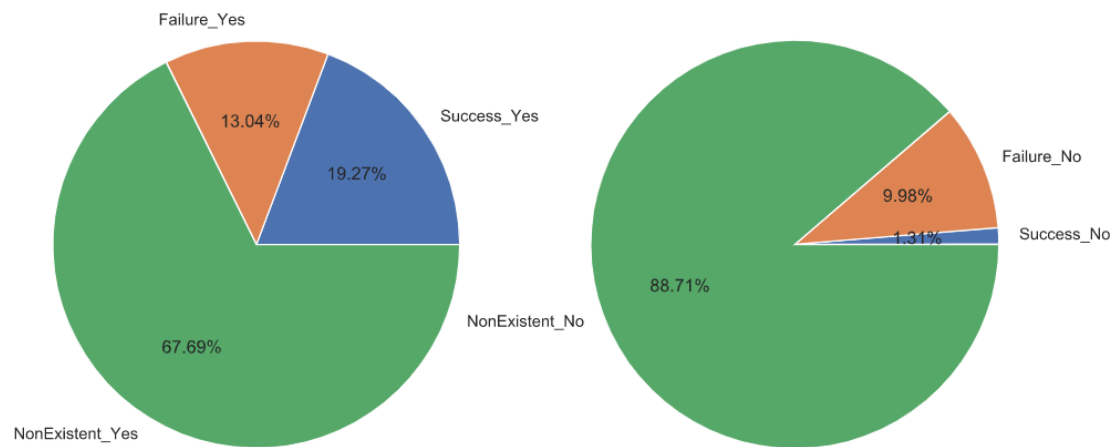
6. Similarly, we can show graph articulating the product subscription rate depending on marital status of the customers represented using Line Chart.

7. Variation of the Campaign held for same customers depending on the months of the year by using boxplot.

8. Overall Product Subscription rate taking all the dependent features in consideration. Hence, we can see the success and Failure rate using Bar Chart.

4.0   Summary and Conclusion:

In this paper, we operated with a dataset that included a variety of factors that were taken into account when performing a marketing survey for a Portuguese banking institution.
We analyzed data using data science approaches, illustrating the significance of all of these strategies. We have published the findings of our report in a variety of ways.
Patterns of observation in relation to various aspects of this Marketing Campaign, which aids us in developing a clear perspective to interpret the trend.

5.0   References:

[Moro et al., 2014] S. Moro, P. Cortez and P. Rita. A Data-Driven Approach to Predict the Success of Bank Telemarketing. Decision Support Systems, Elsevier, 62:22-31, June 2014