

**Tousif Mohammad Alam**

**Topic: Bank Marketing Analysis**

## Contents

<b>Abstract.....</b>	<b>3</b>
<b>Introduction: .....</b>	<b>3</b>
<b>Methodology .....</b>	<b>3</b>
<b>Data Exploration:.....</b>	<b>6</b>
<b>Discussion: .....</b>	<b>15</b>
<b>Conclusion .....</b>	<b>15</b>
<b>Reference .....</b>	<b>15</b>

## **Abstract**

The aim of the report is discussing the selected dataset and explain in detail and elaborately discuss the finding. The chosen data is “bank-additional-full.csv” by S. Moro, P. Cortez and P. Rita which is found in UCI Machine Learning Repository [1]. The data is related to banking marketing campaigns of the Portuguese Banking Institution from May 2008 to November 2010. The campaign’s aim was finding target customer for fixed-term deposit subscription. From the data, the main target goal is targeting more specific clients who can buy bank project e.g. Fixed Term Deposited.

## **Introduction:**

From the last few decades, the media and technology are increasing every step of our life. Humans are more engaging with technology ever than before. Banking is getting popular past few decades among general people. Banks are providing an enormous option on a day including loan, interest. Banks are more focus on consumer banking and targeting client that they can save and spend from the bank. The clients' data is a medium of analysis to find out targeting customer who more like to deposit money with exclusive interest for a fixed term. The dataset is containing clients personal information which can be helpful to the bank to find specific clients who can be interested in subscribes banks product e.g. fixed-term deposit.

## **Methodology**

The dataset I choose for analysis is the “Bank Marketing Data Set” [1] which is modified by S. Moro, P. Cortez and P. Rita from UCI Machine Learning Repository for the tasks. The datasets contain 21 attributes and 45211 instances. The attributes are age, job, marital status, education, default, housing status, loan, contact, month, day of the week, duration, campaign, days and so on. The target is the dataset is to find out “has the client subscribed a term deposit?”. Based on the dataset the target value has divided the client into two parts who subscribed during the marketing campaign and who are not. The two methodologies are used to find out the result from the dataset are Data Cleaning & Data Exploration.

To begin with data quality assurance process, I loaded the dataset to discover the dataset missing value, duplication, write space, predicting an unknown value and remove unknown/null value. The dataset is a single dataset which contains a large number of values. First of all, I used head () and tall () function in Jupiter Notebook to understand the dataset at a glance.

During the validation, I found out the data set some value which is inappropriate for further analysis. For data quality assurance, I remove inappropriate values using pandas library functions. The cleaning process and function is used to give a good shape of the dataset will example below:

- **Whitespaces:** Whitespaces is a key issue for some values. `str.strip()` function of the pandas' library is used to remove all whitespaces.
- **Name Sensitivity:** The attribute 'blue-collar' is not significant so it is replaced with 'labor'
- **Unknown/Missing Value:** The dataset has no missing value. However, there are lot of unknown value categorical attribute like 'job', 'marital', 'education', 'default', 'loan' & 'contact'. Some unknown values are predicted. So, I use `df.loc ()` to change is value in an appropriate order. For example. In the dataset, if client age is greater than 60 then the client job role predicted as retired. Also, in Education, I predicted some possible value e.g. if customer education is school level than it is replaced by the labor.
- **Drop Unknown Value:** After modifying the previous point, the attribute contains unknown value is dropped by using `drop ()` function
- **Group By:** Some value is contained details information which is not suitable for exploration. So, the value is selected in the same group e.g.: 'basic.4y', 'basic.6y', 'basic.9y' is replaced with 'Primary'. Also, the marital status is the group by 'divorced' with 'single' for data exploration.

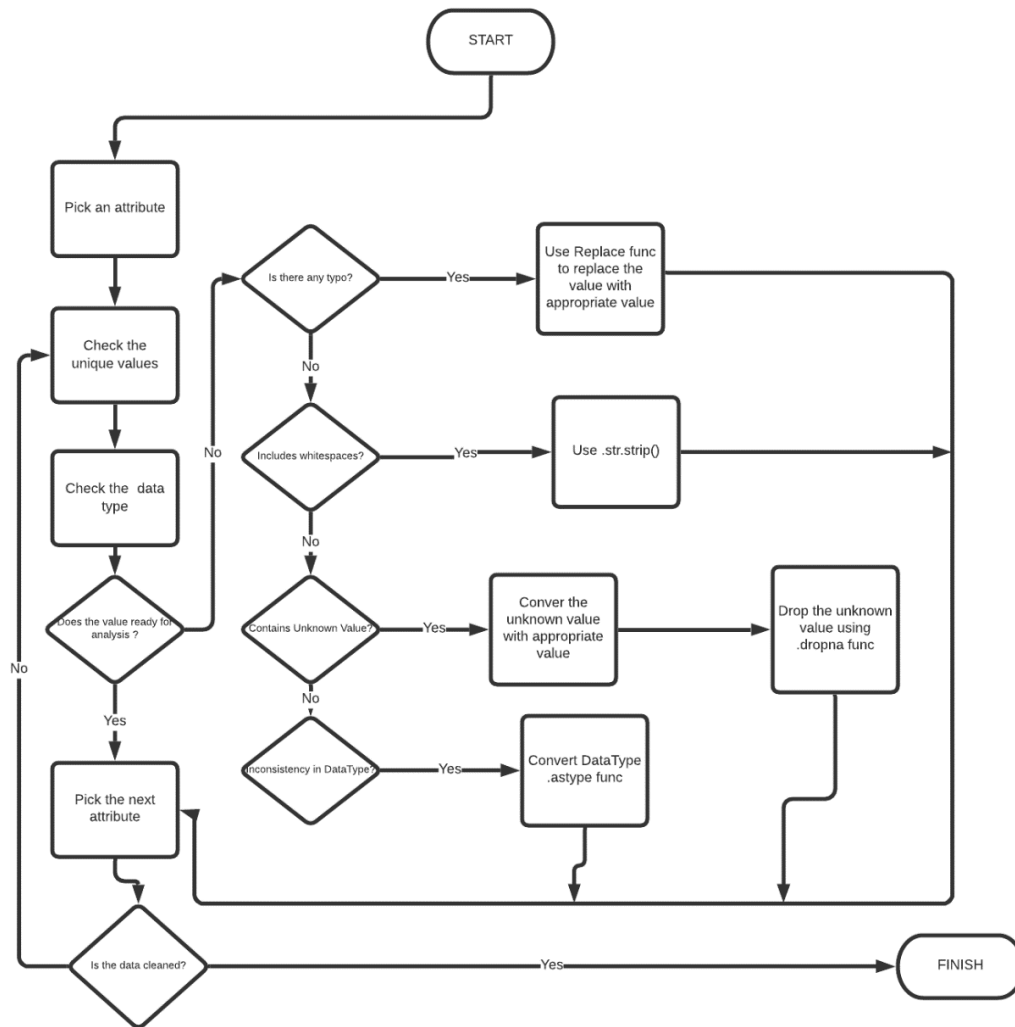


Figure 1: Flowchart of Data Processing

## Data Exploration:

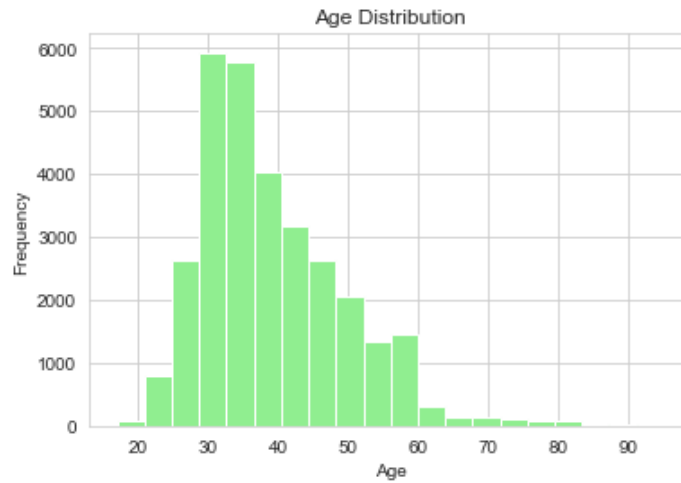


Figure 2: Age distribution

According to Figure 2, the histogram shows the age of the clients in the dataset. The highest number of clients are between 30 years to 40 years old which is more than 10,000. Also, the number of clients is low between 60 to 80 years old. The client numbers are getting increased by age and getting gradually decreased after 45.

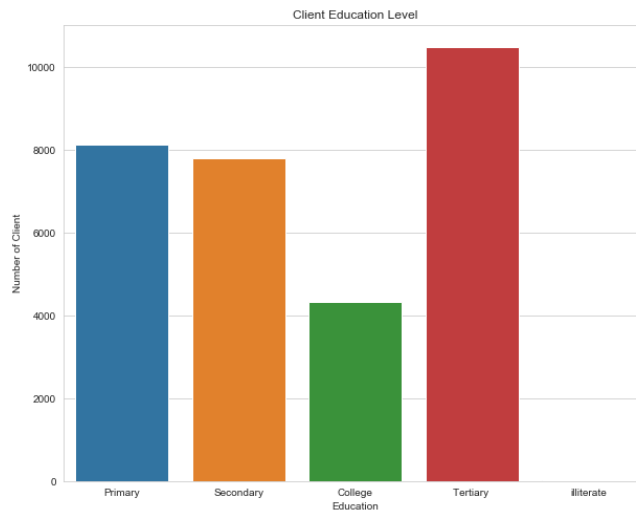


Figure 3: Client Education Qualification

Figure 3 bar chart shows the clients' education qualification level from the dataset. The highest number of clients education level 1 is a primary school which is more than 10,000. The number of

clients who completed primary and secondary education is similar and below 8,000. Some clients did not gain any education at all. The clients who finished their college degree which numbers are more than 4,000.

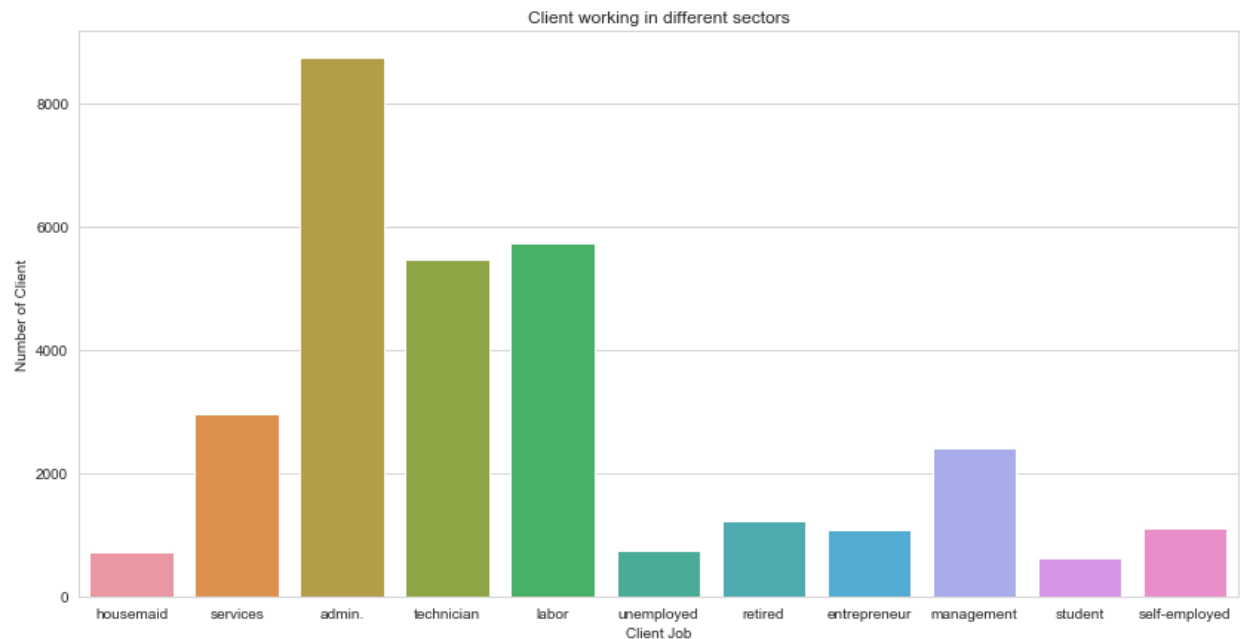


Figure 4: Clients working in different sectors

Figure 4 bar chart demonstrates the clients who work in different occupations. Clients who work in administration is the highest number which is more than 9,000. The number of clients who working as a labor and a technician is quite similar which is more than 5,000. Some clients are currently unemployed or retired which number is below 1,000. Service holder clients number are quite high which is around 3,000. Therefore, we can say that the number of clients working sector is diverse and there are no targeting sectors.

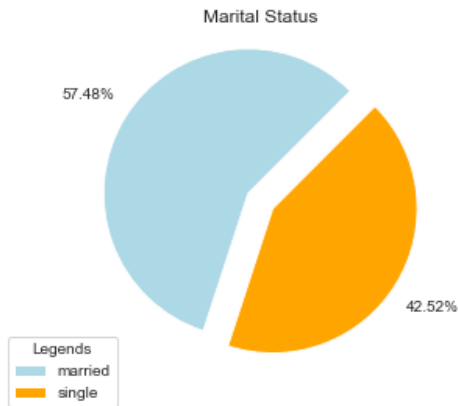


Figure 5: Pie chart client marital status

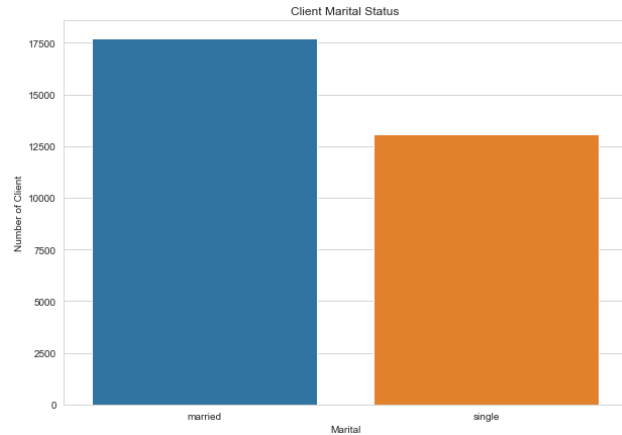


Figure 6: Bar chart client marital status

According to Figure 5 pie chart and Figure 6 bar chart, it indicates the number of clients in the bank who is single and married. The number of clients is married is around 17,500 which is 57.48% of overall clients. The single clients the number is quite low than who are married. The number is 48.52% of the overall clients which is more than 12,500.

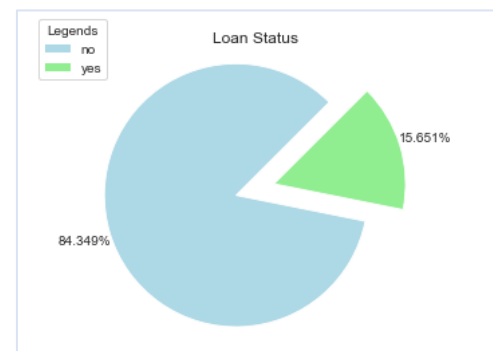
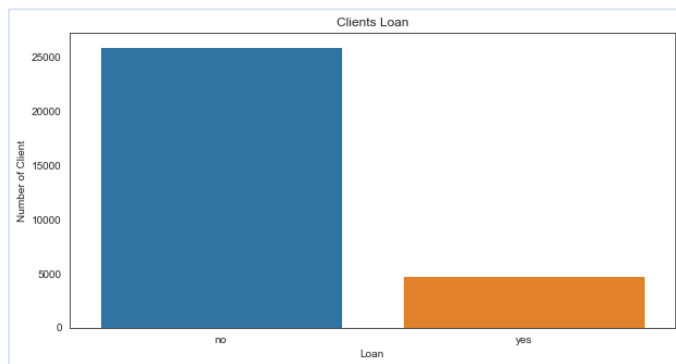


Figure 7 & 8: Clients Loan Status

From Figure 7 and 8, It shows the number of clients in the bank is taking any other loans from the bank. The number of clients is who takes loans from the more than 20,000 which is 84.34% overall clients. Clients who do not take any loan the number is quite small which is around 5,000 and 15.65% of overall clients.





According to Figure 9, the clients who are currently in housing loan in the bank. The number is quite close. More than 16,000 clients are currently in housing loan to the bank.

On the other hands, 14,000 clients did not take any housing loan from the bank.

Figure 9: Bar Chart for Housing Loan

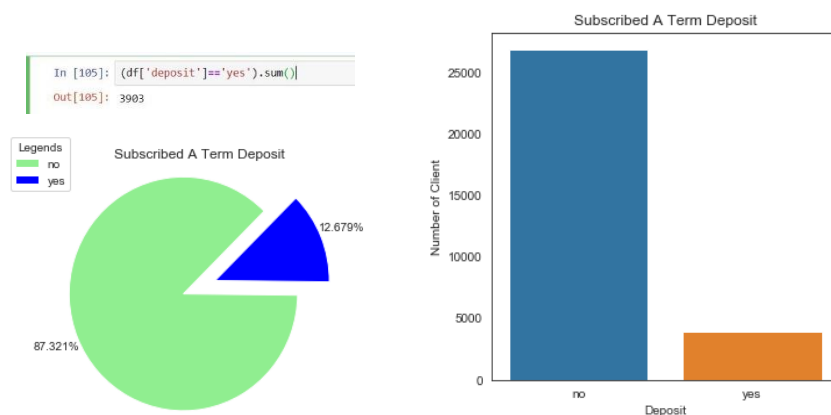
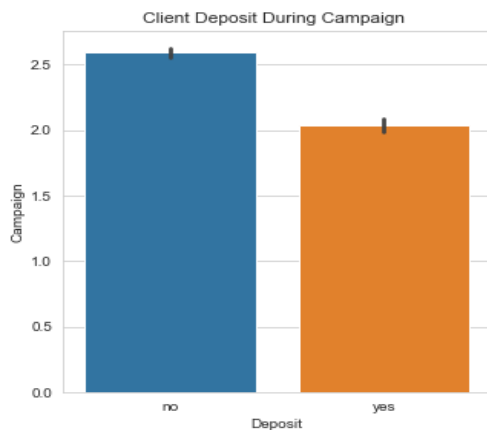


Figure 10 & 11: Client who subscribed a term deposit

According to Figure 10 and 11, the pie chart and bar graph illustrates the clients who has subscribed a term deposit within the period. Firstly, from the bar chart, we can say that more than 25,000 clients do not subscribe to any deposit at all which is about 87.32% overall in the dataset. On the other hands, only a few numbers of clients have a term deposited to the bank which is around 4,000 e.g. 12.67% of overall clients. We can say from the above analysis that, most of the clients use a bank account for the regular transaction purpose.



According to Figure 12, It describes the number of contacts performed during this campaign who are subscribed for a term of the deposit. The number of clients contacted during the period to term deposited is around 20 times who has already had a deposit in the bank.

Figure 12: Number of contacts performed during

According to figure 13 bar graph, the number of clients subscribes for the term of deposit compared to age. It shows that client age between 0 to 40 years old subscribed most during the campaign more than the client who did not subscribe any deposited at all.

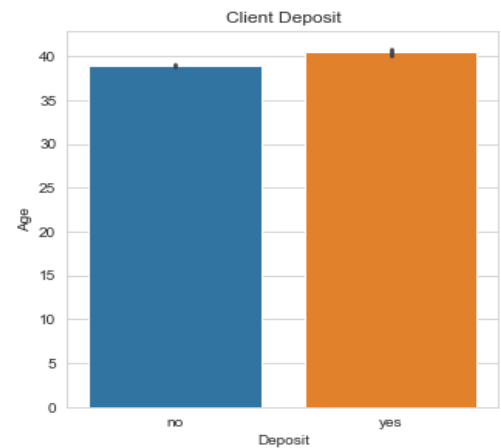


Figure 13: Comparing age deposite

### Research:

The dataset contains several attributes which mainly focus on the client who subscribes for deposit term during the campaign period. There is some research question is built to find out the specific result from the data exploration. Also, it builds a hypothesis about datasets.

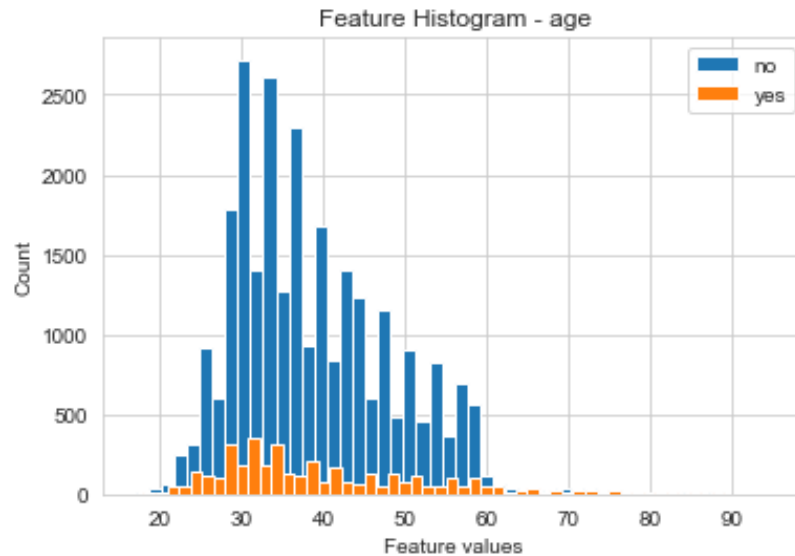


Figure 14: Age histogram comparing deposited subscribe client

The first research question was ‘which age group is most likely depositing their money for a term’. From the histogram in Figure 14, it illustrates clients who subscribed for the term of deposit during the campaign period. Clients whose age is 20 to 35 subscribes most comparing to the client whose age is more than 40. Overall, a small proportion of clients subscribe to the deposited. We can assume that young clients are more into deposit and savings for a better future.

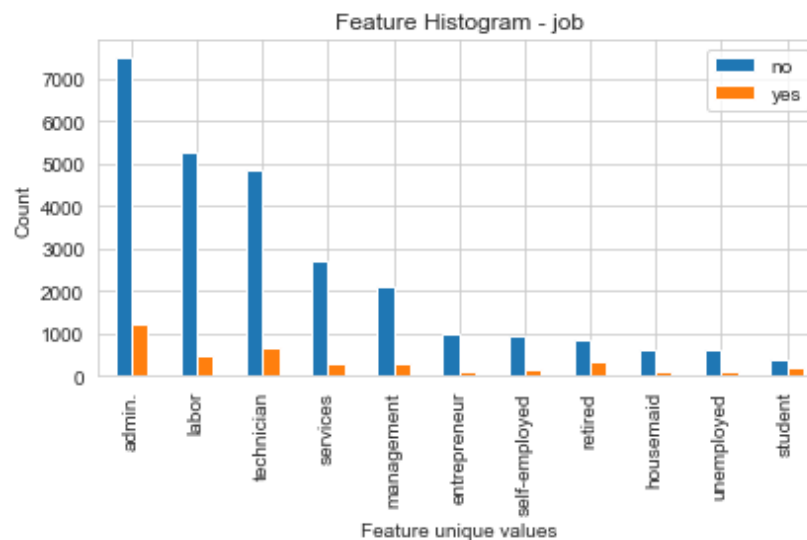


Figure 15: Comparing job and deposit

The research question was “does the clients' occupation matters in case of depositing money”. According to Figure 15 histogram, it demonstrates that clients who are currently doing administration job more likely to subscribe for the deposit. It is quite surprising that the student clients subscribe for deposited more than entrepreneur, housemaid and self-employed. Also, Client who is an administration role is earning more money so they can save money after expenses. We pretend that the new generation is more into saving money than the old generation.

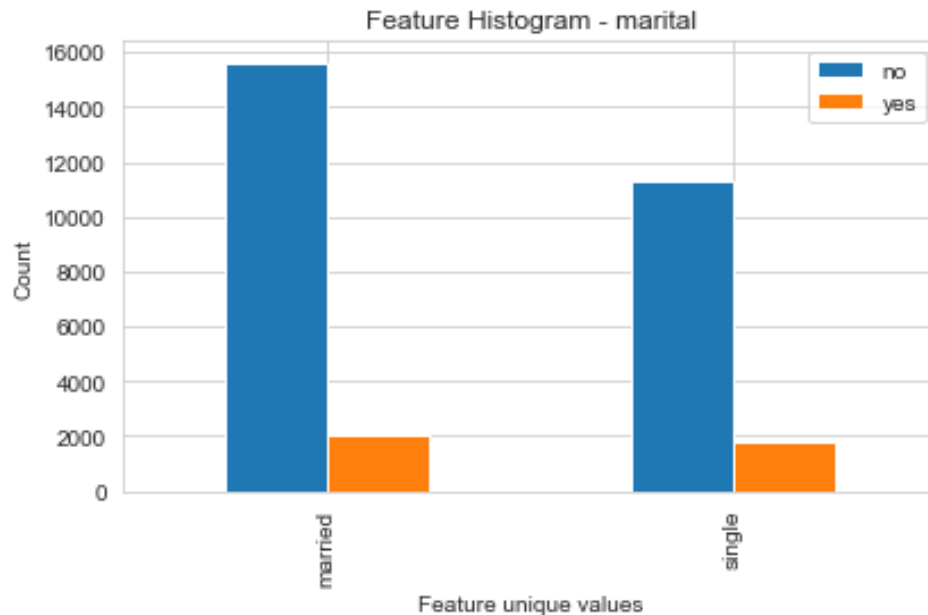


Figure 16: Comparing marital status and deposit subscribe client

Another research question was “how does the marital status impact for despise for a term”. According to Figure 16, the histogram shows that clients who are married are subscribed for the deposit term which is around 2,000. The clients who are currently single and divorced are signed up for the campaign is less than 2000. Overall, we can say that married people are more into saving money than single clients.

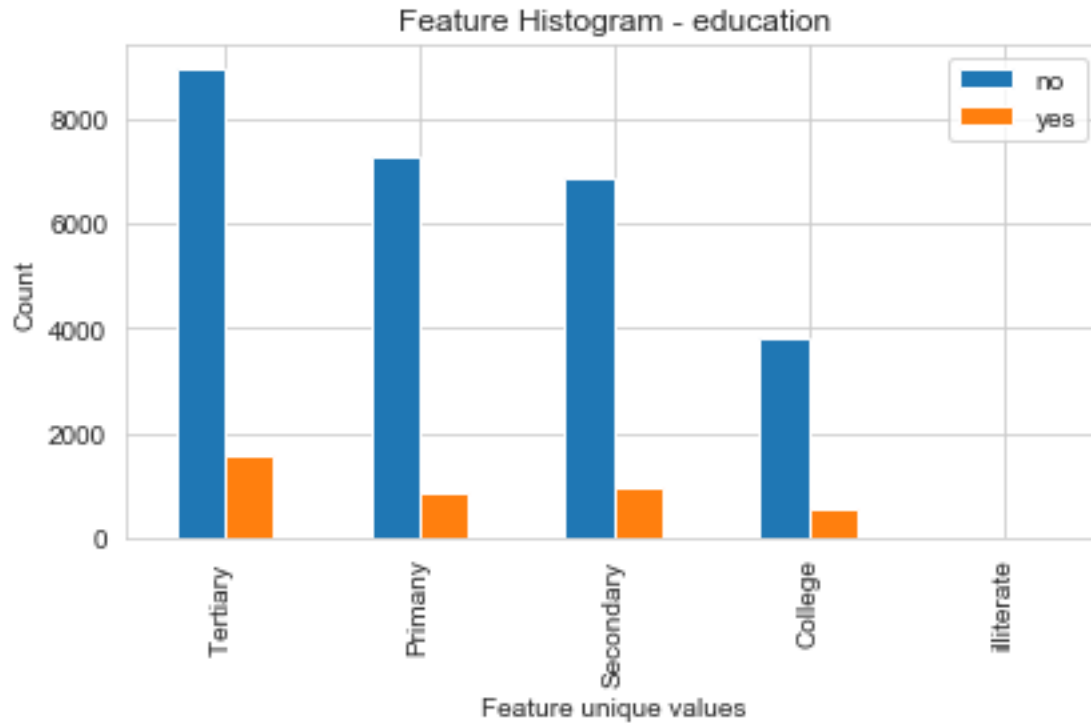


Figure 17: Education and deposit

The next research question was “does educated people more into saving money”. According to Figure 17 histogram, the clients who are completed tertiary education level are more subscribe for the deposit term which is more than 1500. Also, the clients are education level is primary, secondary or college are subscribed for the deposit the number is quite similar and it is less than 1000. So, we can justify that educational client is more into saving money than whose education level is lower. Though, educated people are doing more high paying job and earning more money so they have extra money to save for the future.

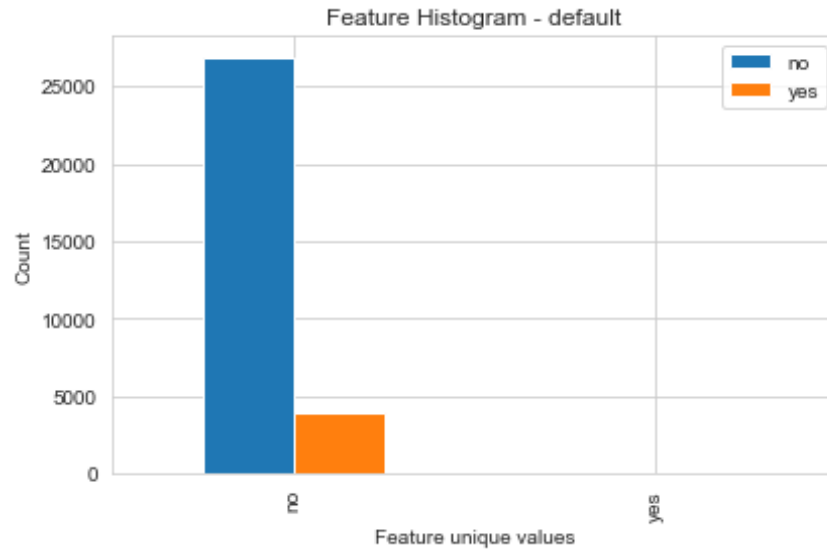


Figure 18: Comparing default and deposit

The research question was “does the credit default loss the subscriber for the deposit or not”. According to Figure 18 histogram, the comparison between clients credit card default and client who subscribe for the deposit term in the campaign. Firstly, clients who have credit card default are not eligible for the program because they already have debt. On the other hands, the number of clients subscribes for deposit is around 4000. Overall, we can say client majority client who does not have any debt still not subscribe for the despite term.

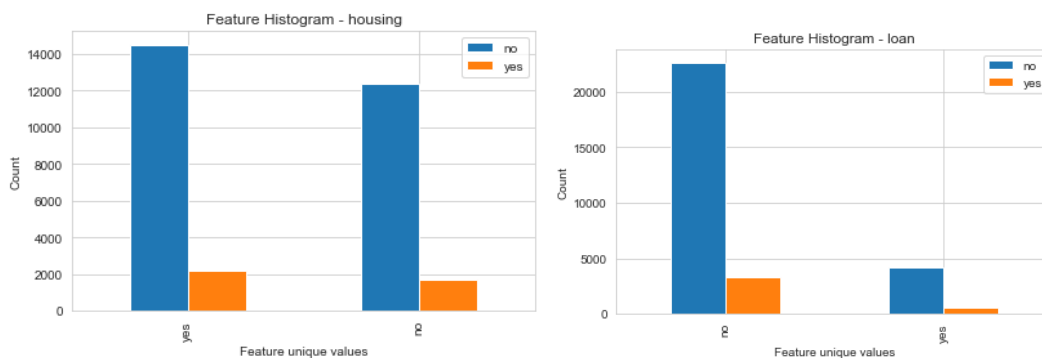


Figure 19 & 20: Loan and Housing loan compering with Deposited

The last research question was “does loan affect people to save money”. According to Figure 19 and 20 histograms, clients who have housing loan still manage to subscribe for the term deposit. The number is more than in 2000. Some clients who do not have any housing loan, but they subscribe to a term deposit which is around below 2000. The number is quite low. The client who

has other loans still handle to manage the term deposit for long term. However, the number is very low which is around 1000. On the other hands, Client who do not have any other loan manage to sign up for the deposit term comparing another client the number is high which is around 3500. We can assume that people unpaid loan mostly affects on subscribe deposit for regular basis.

### **Discussion:**

After comparing, I found a various result that explains in visualization that which types of clients are more suitable for the term deposit. According to that, the Bank marketing team focus on the client and run their marketing campaign. Overall, Client who are currently doing high paying job e.g. admiration, service holders are more into subscribing for fixed-term deposit. Also, the young client whose age is 20-40 is more into saving money than aged clients. But most of the client is using a bank account for a daily basis transaction rather than saving money. Addition to that, married clients are doing more savings than single. So, targeting married clients for marketing would be a great idea to increase the sales of fixed-term deposit. Bank can increase interest rates comparing to other banks so that they can attract more clients to subscribe fixed-term deposit.

### **Conclusion**

In conclusion, Banks are offering various fixed-term deposit for their customer. Although, it is depending on the clients that clients are buying fixed-term deposit or not. But targeting client are the potential buyers of the deposit. However, offering more interest from other bank is increased the interest in buying fixed-term deposit.

### **Reference**

[1] S. Moro, P. Cortez and P. Rita. 2014, *Bank Marketing Data Set*, Data-Driven Approach to Predict the Success of Bank Telemarketing. Decision Support Systems, UCI Machine Learning Repository, viewed 20 October 2020, <<https://archive.ics.uci.edu/ml/datasets/Bank+Marketing>