MBAS901: Assessment 2

Exploratory Business Analytics Project

Name: Chinmay Datar

Student No.: 6956361

Word Count: 2700

Content

Introduction

With the advancement in data analytics field every industry is using data to find out useful information and use it to make better and more accurate business decisions to grow. Banks use data too to find potential customers. One way of doing this is by direct marketing. Banks can analyse the data collected from previous campaigns and find out which ideas of the campaign were successful and to redirect their resources in that direction for a better reach.

Banks make profit by borrowing money from its customers and giving them interest. Then they lend this money to other customers and businesses and charge a higher interest rate. Banks make profit from the difference in interest rates. In order to grow on this strategy banks need more customers. Therefore, banks use data to find new customers to increase the bank's sustainable cashflow.

For this, banks need reliable data and analysis. Gaps in data can lead to flawed analytics and loss of resources. For this purpose, it is important to have reliable data to get customer insights and create an effective, successful marketing campaign to get more customers and maximize profit.

Description of the Dataset

The dataset is of a direct marketing campaign of a Portuguese bank. The campaign was run to promote customers to subscribe to their new product (Term Deposit). The direct marketing campaign was conducted over call.

The dataset contains 17 columns consisting of details about customers (Customer ID, Age, Education, Marital Status, Job), customer's current financial status (Balance, Credit default, housing Loan, Personal Loan) and details and outcomes of the previous campaign (last contact month/day, last contact duration, the outcome of the previous marketing campaign, contact type). Our primary variable of interest is 'y' which holds outcome of the campaign.

The details of the columns are as follows:

Table 1: Description of Categories

No.	Variables	Renamed as	Descriptions
1	Contact	Contact	Contact communication type
2	Default		Do customers have any credit that is in default?
3	Loan	Personal loan	Do customers having personal loan?
4	Housing	Housing loan	Do customers have housing loan?
5	Job		Type of job
6	Month		Last contact month of year
7	Marital	Marital status	Marital status
8	Education		Education level
9	Poutcome	Outcome of Previous Campaign	Outcome of the previous marketing campaign
10	y	Subscription	Has the client subscribed a term deposit? (Binary: 'yes', 'no')

Table 2: Description of Measures

No.	Measures	Renamed as	Aggregation	Descriptions
1	Age		Average	Contact communication type
2	Customer_id	Customer ID	Sum	Customer identification number
3	Day		Sum	Last contact day of month
4	Duration		Average	Last contact duration, in seconds

5	pdays		Sum	Number of days that passed by after the client was last contacted from a previous campaign (numeric; -1 means client was not previously contacted)
6	Campaign	Contacts during the campaign	Sum	Number of contacts performed during this campaign and for this client (numeric, includes last contact)
7	Previous	Contacts before the campaign	Sum	Number of contacts performed before this campaign and for this client (numeric)
8	Balance		Average	Available bank balance in existing accounts

Aim of the Analysis

As a business analyst my aim in this report is to find relationships between different columns and our target column 'y' (renamed as subscription). This will help targeting the right audience for the campaign and focus on the consumers that are more likely to subscribe to the product. This will also reduce the overall advertisement costs associated with the campaign. This is achieved by performing exploratory analysis on the above dataset which also consists of results from the previous campaign. By using different graph and visual representations we can identify past trends faster and use it to run an efficient and successful campaign.

Hypothesis

After looking at the data columns we can come up several hypothesis and questions which we can explore and get useful insights which will help the bank run a successful campaign.

Hypothesis 1: How does customer's job and balance affect number of customers who subscribe to term deposit?

Hypothesis 2: In what does customer's age and balance affect number of customers who subscribe to term deposit?

Hypothesis 3: How does loans (personal and housing) and education of the customers affect their decision to subscribe to term deposit?

Hypothesis 4: In what month was the previous campaign most successful in? Can we compare it with the current campaign?

Hypothesis 5: How does number of contacts and duration of call affect the customer's decision to subscribe to term deposit?

We shall go over and analyse the dataset in depth to get answers of all the above questions by visualizing it.

Methods

A. Data Preprocessing:

The dataset provided to us is structured and clean therefore we organise the data for our convenience. We have renamed some of the columns as seen in the tables 1, 2. We further change the type of Customer ID to category. Then we make custom categories by taking age and balance to identify the apparent trends.

We break down age into 6 intervals.

1. 18-30: customers between the age of 18 to 30.
2. 31-40: customers between the age to 31 to 40.
3. 41-50: customers between the age of 41 to 50.
4. 51-60: customers between the age of 51 to 60.
5. 60+: customers above the age of 60.

Next, we create custom category using balance and dividing it into 4 intervals.

1. Negative Balance: Customers with balance less than zero.

2. Low Balance: Customers with balance less than $10,000.
3. Average Balance: Customers with balance between $10,000 and $30,000.
4. High Balance: Customers with balance above $30,000.

We create another calculated item named 'Subscriptions' which calculates the percentage of total subscriptions (This is same as frequency percent. It is created for convenience of the creator).

B. Types of graphs/charts used:

Broadly three kinds of graphs are used for this analysis.

1. Bar graph: Used to compare and show distribution of measures over various categories. Bar charts make it easier to see the highest or lowest measures in a category.
2. Dual Axis Chart: This includes dual axis line charts, dual axis bar-line chart. Dual Axis enables us to more than one measure on the y axes which helps to visualize and compare 2 different measures in a single graph.
3. Line Chart: Used to see the changing trends in the variables over a category or measure. In this report it is used to visualize changes over a category.
4. Pie Chart: This is being used to show distribution of a category, preferably in a percentage value. This is helpful while using in coherence with other graphs as it gives a deeper understanding of the category.

Results and Discussion

Hypothesis 1: How does customer's job and balance affect number of customers who subscribe to term deposit?

We need to first see how job affects subscriptions. To visualize this, we use word cloud.

Fig 1. Word cloud of job

The above word cloud has subscription filter applied to show only number of subscribed customers. From this it is evident that maximum subscriptions are done by customers working in management, technician, and blue-collar jobs.

To further investigate this, we visualise balance and subscriptions on a dual axis bar-line chart and filter by customers who have subscribed.

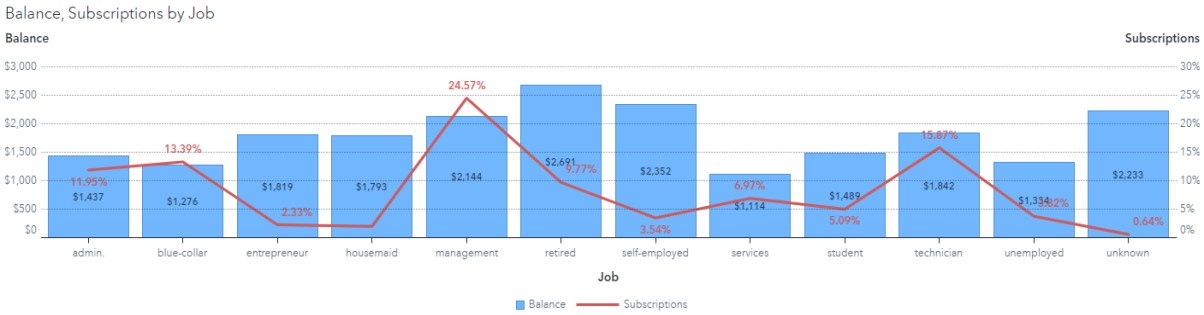

Fig. 2: Balance and Subscription by Job of Customers.

We can see from fig 2 that about 25% of the customers who have subscribed to the term deposit have a management job with fourth highest average balance ($2144) followed by customers who work as technician making just over 15% of total subscribed customers. As can be noticed even though retired customers have the highest average balance ($2691) they only make up for about 10% of the subscribed customers. Customers who are self-employed and customers whose profession is unknown have second and third highest average balance ($2352 & $2233 respectively) but they make-up for less than 5% of the total subscriptions. It

is also note-worthy that customers doing blue-collar jobs make-up for just more than 13% of the subscriptions.

Customers working in management have a working understanding of financial instruments which is why they have higher average balance and highest percent of subscribers. Similarly, technicians are qualified workers and would prefer safe and attractive financial product such a term deposit.

Bank should target customers who are in management, technicians or do blue-collar jobs as they have the highest subscription rates making-up for than 50% of subscribers.

Job categories with highest subscription rate are management, technician, blue-collar, admin and retired.

Insight: Marketing can be focused on customers with the above mentioned jobs as they are more likely to subscribe than the others.

Hypothesis 2: In what does customer's age and balance affect number of customers who subscribe to term deposit?

We now use balance intervals and age intervals to see which age group has the highest subscription rate. By evaluating two variables at the same time enables an analyst to draw more comprehensive insights and corelations from the data which otherwise wouldn't have been possible. To visualise this, we use bar chart with colors to show different intervals of balance. We also use a pie chart to show the total subscriptions from a particular age group.

From fig 3(a), it is clear that maximum subscriptions come from customers with low balance in every age interval. Customers with average or high balance account for less than 1% of total subscriptions while customers with negative balance account for 8% of subscriptions and customers with average balance accounting for rest 90%. More than 50% of the subscribed customers with average balance are under the age of 40. It can be seen that the percentage of subscribed customers decreased when working age increased.

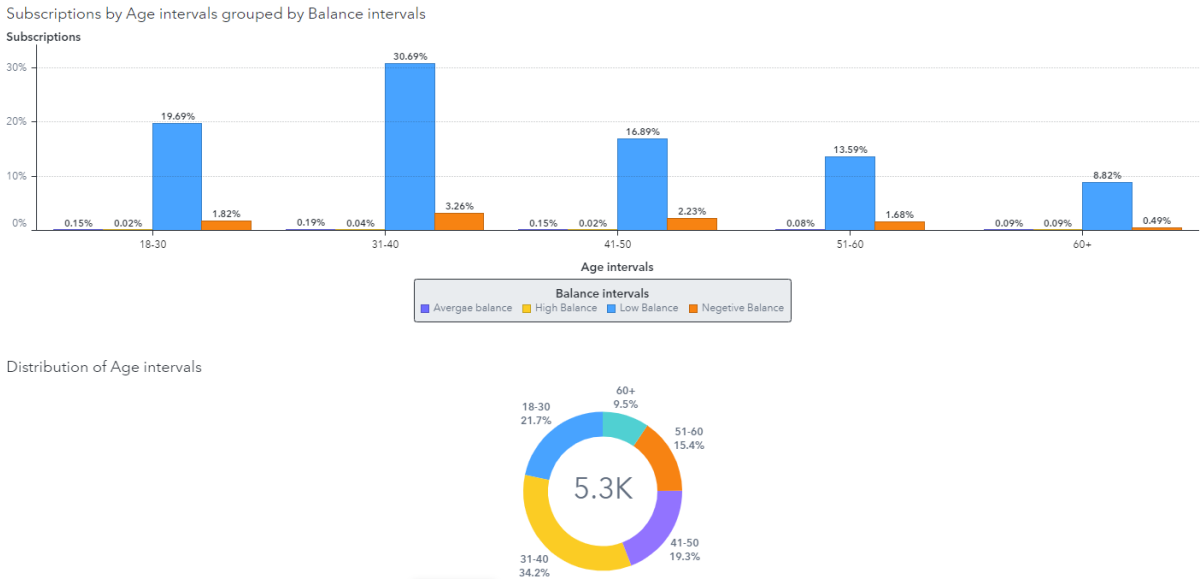

Fig. 3: a. Subscriptions by age intervals and colored by balance intervals. b. Subscriptions by age intervals.

Customers below the age of 30 often have unstable income, insufficient financial knowledge, and capacity so they take high risks instead of investing in low-risk financial instruments such as term deposit. This is because they don't have any dependents.

Customers above the age of 30 tend to have low balance since majority of their money is invested in assets which is evident from the above graph. Since term deposit is one of the safest investments, young and middle-aged customers will want to subscribe to it. The above numbers tell us the same story. 75% of subscribers are below the age of 50 years. Customers above the age 60 years do not find term deposit that appealing enough to subscribe to as they are retired and don't have a constant inflow of cash.

Insight: It is evident that the bank should prioritize customers with negative and low balance to get more subscribers. Additionally, looking at the number of subscribers the bank needs a new strategy to get the customers with average and high balance to subscribe to the product.

Hypothesis 3: How does loans (personal and housing) and education and marital status of the customers affect their decision to subscribe to term deposit?

To visualize this, we use lattice function to group different loan categories. Liabilities such as loans and credit dept can influence a person's ability to subscribe to products such as term deposit.

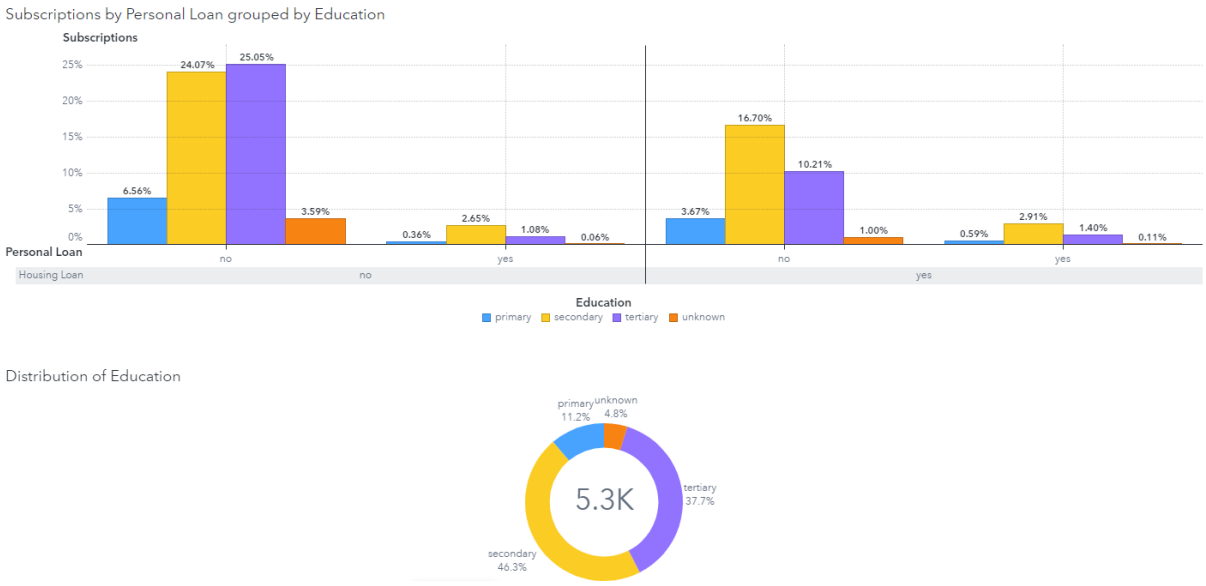

Fig. 4: a. Subscriptions by loans and colored by education. b. Subscriptions by education.

From fig 4 (a) we observe that about 60% of customers who have subscribed have no loans and about 30% of subscribers have only housing loan. Rest 10% of subscribed customers have either only personal loan (4%) or both. From figure 4 (b) we further notice that more than 80% of subscribed customers have secondary or tertiary level of education. Out of these 80% customers 50% don't have any loans and 25% have housing loan.

Customers who are well educated and have no liabilities and would want to invest in assets which have less risk associated to it. This is reflected from the graph above. Similarly, well-educated customers with only housing loan would want to invest in assets with less risk. This is different from customers with personal loan as housing loans are secured as in most cases people don't have big balance to spend and due to tax concessions, that you get on housing loans (Depending on the country). This makes customers with housing loans more willing to invest in other assets.

Customers who are well educated and have no liabilities and would want to invest in assets which have less risk associated to it. This is reflected from the graph above. Similarly, well educated customers with only housing loan would want to invest in assets with less risk. This

is different from customers with personal loan as housing loans are secured as in most cases people don't have big balance to spend and due to tax concessions, that you get on housing loans (Depending on the country). This makes customers with housing loans more willing to invest in other assets.

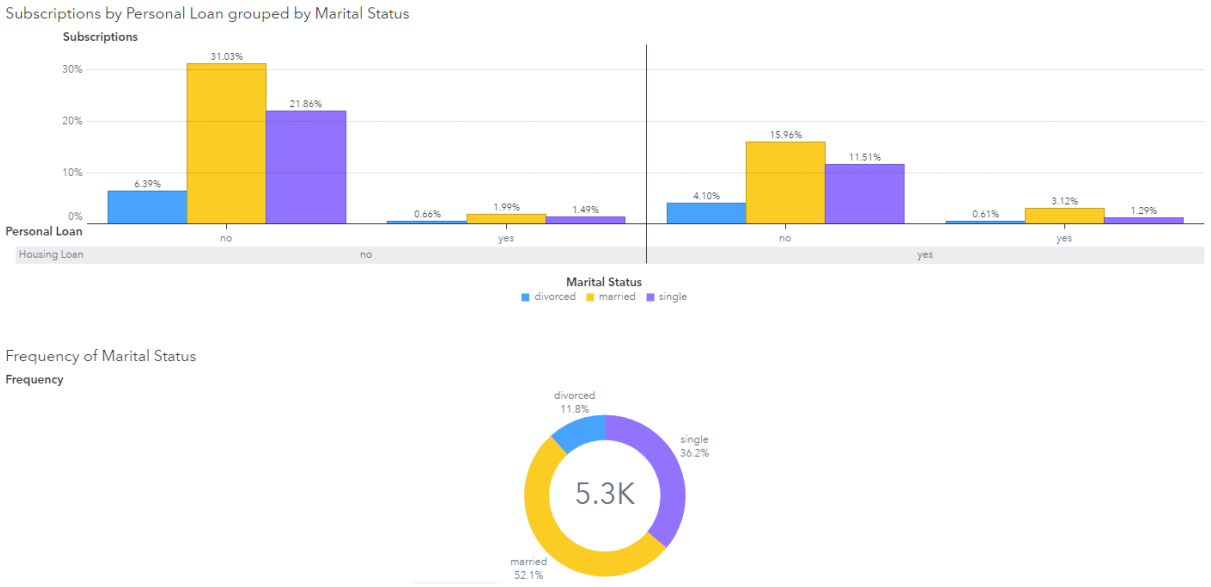

Fig. 5: a. Subscriptions by loans and colored by marital status. b. Subscriptions by marital status.

Similar to fig. 4, fig. 5 shows distribution based on marital status. It can be observed that more than 50% subscriptions from customers who are married and 35% from customers who are single. As single people and married are financially stable and are more willing to invest in financial instruments such as term deposit to get stable return.

Insight: Banks should focus on customers who are well educated and have no loans or housing loan as they account for high percentage of subscribing customers. Among these, banks should check marital status to be single or married as they are more likely to subscribe compared to divorced customers.

Note: Credit default was also investigated for this hypothesis, but the end result shows that only 1% of the customers who have credit default subscribe to the product.

Hypothesis 4: In what month was the previous campaign most successful in? Can we compare it with the current campaign?

Timing of the marketing campaign is crucial as a good but poorly timed marketing campaign We compare number of contacts made during the previous and current campaign and visualize them using line chart. This will show how the strategy changed from the previous marketing campaign.

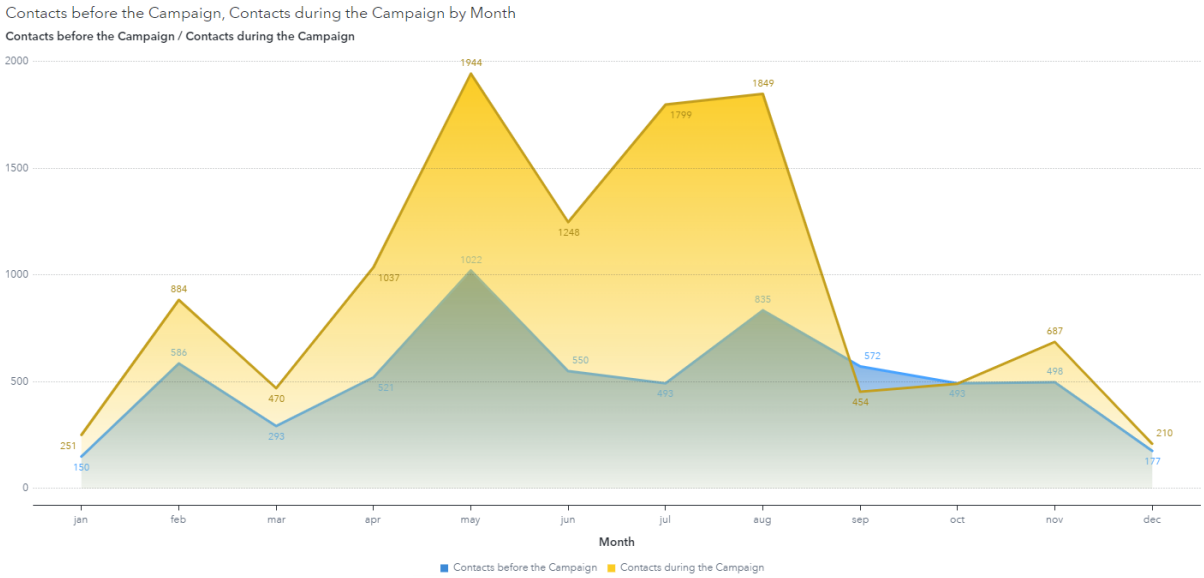

Fig. 6: Contacts made in previous campaign and current campaign.

It is clear from fig. 6 that it is clear that the campaign was influenced by outcome of the previous campaign. From the graph we can see that most subscriptions came in the month of May and August. Number of subscribers in the previous campaign nearly doubled in these months. In the current campaign many customers subscribed in the months of June and July too. The least number of subscribers are from the months of December and January. The number of subscriptions start to rise from March and rapidly falls in September.

During the months of December and January due to holiday season customers are not looking for investment options which is why these months have the least number of subscribers. The bank saw that number of subscribers between the month of May and August were high in the previous campaign and hence the number of subscribers has doubled during these months.

Insight: Banks should spend more on marketing campaign during the months of April to August as customers are willing invest in financial instruments like term deposit. With more focused approach these numbers could reach even higher.

Hypothesis 5: How does number of contacts and duration of call depend on job and affect the customer's decision to subscribe to term deposit?

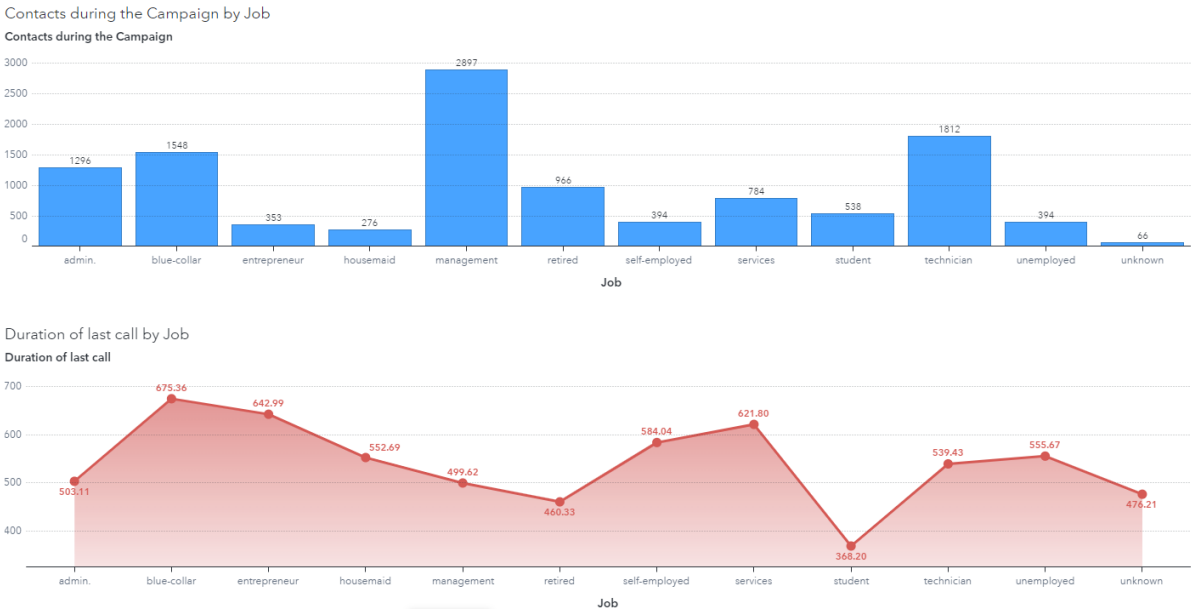

Fig. 7: a. Contacts made during the campaign by Job. b. Duration of last contact by jobs.

We have seen how job and number of contacts affects the number of subscriptions. Now we look at how the average duration of calls combined with the above two factors affect the subscriptions.

As can be seen in fig. 7 customers in management jobs are contacted more than any other job category (2897 times) followed by technicians (1812 times). The average calls for these professions are about 500 seconds and 540 seconds respectively. The highest average call is with customers in blue-collar jobs (675 seconds) and they also have the third highest subscriptions. Less number of calls are made to entrepreneurs and self-employed customers, but the average length of call is more than 600 seconds. Students have the shortest call duration. Call duration with customers who are unemployed is on the higher side while only 3% of the subscribers are unemployed.

Insight: Bank needs to redirect its focus on the target market depending on jobs and other financial assets. As can be seen in fig. 7, long and high number of calls to clients who won't be subscribe wastes bank's resources and time.

Conclusion

Based on the above analysis, the bank should create a profile of customers with common features that can get high number of subscriptions. Customers with highest probability of subscribing will have these features.

- Feature 1: Customers who work in management, as technician and blue-collar jobs.
- Feature 2: Customers between the age of 30 and 40 years.
- Feature 3: Customers have no credit default, personal loan and may or may not have housing loan.
- Feature 4: Customers who have education level of secondary or tertiary.
- Feature 5: Customers who are single or married.

Using several profiles like this can help in more targeted marketing leading to higher percentage of subscribers. This will also lead to reduced marketing costs and more effective marketing strategy. This will also benefit customers by not getting unwanted advertisements.

The timing of the marketing is as crucial as finding the right market for the product. The above analysis shows that the marketing during the months of May to August can be more cost-effective and is more likely to be a successful campaign.

With more detailed customer database bank can provide better services to the customers. Details about the customer such as marital status and job profile reveals the stage of life while balance, credit default, loans shows the financial position the customer is in. This will help bank predict when customer requires what services and products. This will make customers satisfied with the bank and a loyal customer.