**Direct Marketing Campaigns Research Paper**

Glenda Montecillo Hart

Colorado State University Global

MIS581: Capstone: Business Intelligence and Data Analytics

Dr. Jamia Mills

September 4, 2022

**ABSTRACT**

The data set used for our analysis is from the UCI Machine Learning Repository called Bank Marketing data set of a Portuguese bank institution. The data set contains records of direct marketing campaign based on phone calls. We can see that more than one contact to the same customer was required to have the customer subscribed (denoted by “yes) or not (denoted by “no”) subscribed to long term deposit. The goal of this research is to build a predictive model capable of increasing the efficiency of directed campaign for long term deposit subscriptions and to predict which customer and market segment would respond to promotions and marketing campaigns (Apampa, 2016).

Within the banking industry, there is a growing pressure to increase profits and reduce cost (Moro et. al, 2014). We chose this study to provide a useful tool to the banking industry to improve customer experience and to reduce cost (Roy, 2019).

This research paper contains the following sections:

* Introduction
* Objectives
* Overview of Study
* Research Questions and Hypothesis
* Literature Review
* Research Design
* Findings
* Conclusion
* Recommendations

**INTRODUCTION**

One of the industries that uses predictive analytics is the finance industry. By using predictive analytics in finance can help identify trends for planning, forecasting and decision making. It can predict outcomes, predict stock prices, or in our case predict if a customer will subscribe to a term deposit – target variable “yes” otherwise “no”.

Banks are faced with various challenges offering products and service to their customers, such as increasing competition, continually rising marketing costs, decreased response rates, at the same time not having a direct relationship with their customers. In order to address these problems, banks aim to select those customers who are most likely to be potential buyers of the new product or service and make a direct relationship with them. Banks want to select the customers who should be contacted in the next marketing campaigns (Nachev, 2015).

According to Elsalamony (2014), the term direct marketing was coined by Lester Wunderman, considered to be the father of direct marketing, in 1967.

**Problem Statement**

The dataset used in this study is about direct marketing campaigns of a Portuguese banking institution and the campaigns were mostly conducted on phone calls. The goal is to build a model that classifies whether an existing client will subscribe to a term deposit or not based on the analysis of the marketing campaigns the bank performed. Throughout this case study, we will be using the dataset from UCI Machine Learning Repository for determining the success of the bank telemarketing campaign. The dataset contains real world data from a Portuguese retail bank collected from May 2008 to June 2013 (Moro et al., 2014).

**The Dataset**

The dataset used in this study is related with direct marketing campaigns of a Portuguese banking institution and the campaigns were mostly conducted on phone calls. The goal is to build a model that classifies whether an existing client will subscribe to a term deposit or not based on the analysis of the marketing campaigns the bank performed.

The Portuguese bank dataset used in this study was provided by Moro et al. (2014). There were four variants of the datasets out of which we chose “bank.csv” which consists of 4,521 observations and 17 variables with seven numeric variables, 10 categorical variables including the binary response denoted as “y” (see Table 1). There are no missing attribute values. The data summary demonstrates that the binary response class of “y” had 521 instances of “yes” as response, and 4,000 instances responded a “no”. The analysis indicates that 11.7% of the total number of customer contacted during the marketing campaign responded “yes” to the promotion.

**Table 1**: **Feature description for the bank dataset**

|  |  |  |  |
| --- | --- | --- | --- |
| **No.** | **Attribute** | **Description and Values** | **Type** |
| 1 | Age | Age of client in years | Numeric |
| 2 | Job | Type of job: "admin.", "unknown", "unemployed", "management", "housemaid", "entrepreneur", "student", "blue-collar", "self-employed”, "retired", "technician", "services") | Categorical |
| 3 | Marital | Marital status: "married", "divorced", "single"; note: "divorced" means divorced or widowed) | Categorical |
| 4 | Education | Education: "unknown", "secondary", "primary", "tertiary" | Categorical |
| 5 | Default | Has credit in default? | Binary |
| 6 | Balance | Client’s average annual balance | Numeric |
| 7 | Housing | Does the client have housing loan? | Binary |
| 8 | Loan | Does the client have personal loan? | Binary |
| 9 | Contact | Contact communication type: "unknown", "telephone", "cellular" | Categorical |
| 10 | Day | Client’s last contact day | Numeric |
| 11 | Month | Client’s last contact month | Categorical |
| 12 | Duration | Last contact duration, in seconds | Numeric |
| 13 | Campaign | Number of contacts performed for client during the campaign | Numeric |
| 14 | PDays | Number of days elapsed since last contact from previous campaign | Numeric |
| 15 | Previous | Number of contacts performed for the client before campaign | Numeric |
| 16 | POutcome | Outcome of the previous campaign: "unknown", "other", "failure", "success" | Categorical |
| 17 | y | Has the client subscribed a term deposit? | Binary |

*Source:* [*https://archive.ics.uci.edu/ml/datasets/bank+marketing*](https://archive.ics.uci.edu/ml/datasets/bank+marketing)

**OBJECTIVES**

The purpose of this research project is to build a predictive model capable of increasing the efficiency of directed campaign for long term deposit subscriptions and to predict which customer and market segment would respond to promotions and marketing campaigns (Apampa, 2016).

**Benefits from Research**

The following are few of the benefits an organization could use from this research:

* Help focus their efforts to customers and market segment that would respond positively to promotions and marketing campaigns
* Increase capital for selling long-term deposit subscriptions
* Improve relationships and connection with their customers
* Increase the campaign effectiveness by identifying the main characteristics that affect a success (Elsalamony, 2014)

**OVERVIEW OF STUDY**

This research paper is a study on direct marketing campaigns of a Portuguese banking institution. The campaigns were mostly conducted on phone calls. The goal is to build a model that classifies whether an existing client will subscribe to a term deposit or not based on the analysis of the marketing campaigns the bank performed.

**RESEARCH QUESTIONS AND HYPOTHESIS**

**Research Questions**

The goal of this research is to build a predictive model capable of increasing the efficiency of directed campaign for long term deposit subscriptions and to predict which customer and market segment would respond to promotions and marketing campaigns (Apampa, 2016). As part of this section, we will provide two research questions and analyze each question.

1. Can willingness to subscribe to long term deposit be predicted by education, housing, and job variables?

This research question helps to identify customer(s) who is likely to subscribe to long term deposit and develop more targeted marketing campaigns. During the initial investigation using the dataset, we found that customers who were contacted has a university level education, owns a house and has management category job.

1. What is the average age of a customer who is willing to subscribe to a term deposit?

This research question help to identify the age of customer(s) who is likely to respond to the marketing campaign and will acquire the product. During the initial investigation, we found that a number of customers whose age is 60 and above has the same number of “yes” and “no” responses (Wickramanayake et al., 2020).

**Hypothesis Tests**

This section provides an overview of hypothesis including the null and alternate hypothesis tests.

1. Determine whether or not customer(s) who has university level education is more likely to subscribe to long term deposit with the bank?

This question is a way of investigating whether customers with university level education are more inclined towards accepting the long term deposit subscription.

Null (H0): The number of responses (yes or no) is the same whether a customer has university level education or not.

Alternative (Ha): The number of responses (yes or no) is not the same whether a customer has university level education or not.

1. Determine whether or not customer(s) who is more than 60 years is likely to subscribe to long term deposit?

This question is a way of investigating whether customers whose age is 60 years old and above are more inclined towards accepting the long term deposit subscription.

Null (H0): The number of responses (yes or no) is the same whether a customer age is 60 years old or above.

Alternative (Ha): The number of responses (yes or no) is not the same whether a customer age is 60 years old or above.

**LITERATURE REVIEW**

Banks are faced with various challenges offering products and service to their customers, such as increasing competition, continually rising marketing costs, decreased response rates, at the same time not having a direct relationship with their customers. In order to address these problems, banks aim to select those customers who are most likely to be potential buyers of the new product or service and make a direct relationship with them. Banks want to select the customers who should be contacted in the next marketing campaigns (Nachev, 2015).

In this literature review,we found that customers who were contacted has a university level education, owns a house and has management category job.

**RESEARCH DESIGN**

**Methodology**

Research will begin by having a good understanding of the bank dataset which contains of 4,521 rows and 17 columns. There are seven numeric variables and ten categorical variables in the dataset. We will read the dataset into a data frame and pre-process the data for missing values. Efforts will then be made to identify customer(s) who is likely to respond to the marketing campaign and will subscribe to long term deposit subscriptions.

**Methods**

We will be using the following models on SAS Enterprise Miner and SAS Studio to test our hypothesis:

* **Logistic Regression** is an algorithm used for predicting variables with finite set of values. The output is in the form of probability distribution with a value less than one and is based on maximum probability estimation rather than the least squares estimation. Moreover, the idea behind the logistic regression is straightforward: instead of using Y as the dependent variable, we use a function of it which is called the logit (Nachev, 2015).
* **Naïve Bayes** is a classification algorithm based on Bayes theorem. It is assumed that the values of each features in the train datasets are independent of one another. Bayes method learns the conditional probability of each attribute given the class label from the training data and then computes the probability of a class value given the particular instance, and predicting the class value with the highest probability.

**Limitations**

The findings of this research paper pertain only to the direct marketing campaigns of a Portuguese bank and does not represent all other bank institutions. The dataset used is public available for research (Moro et al., 2014).

**Ethical Considerations**

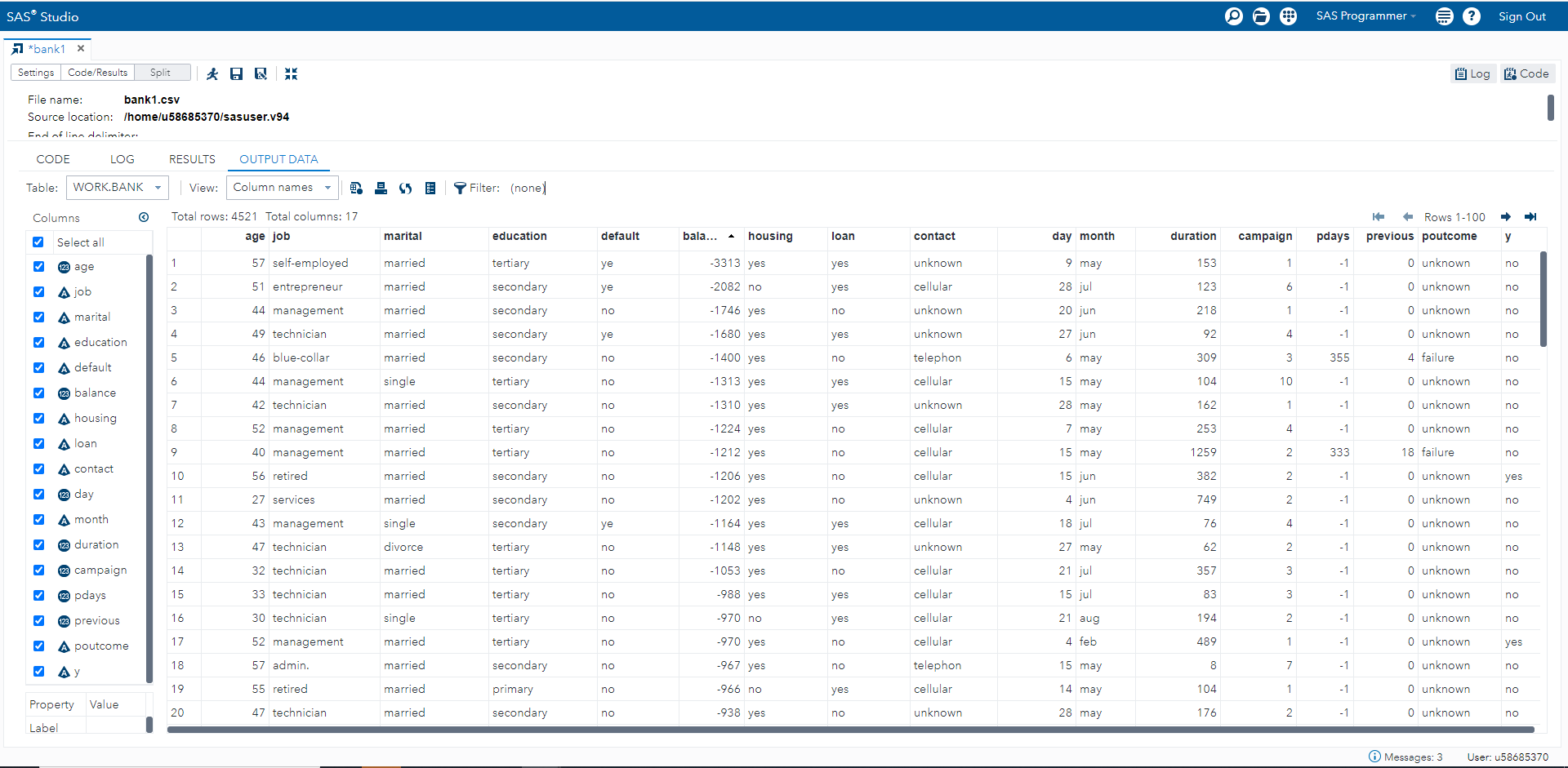
Ethical considerations in research are a set of principles that researchers must adhere to when collecting data from people. The dataset we are using in our research paper has all personally identifiable data is not collected. Therefore we do not know the identities of the participants and we cannot identify any individual participant to their data. This de-identification technique is called anonymization where all PII is simply removed from the dataset (Klose et al., 2020).

**FINDINGS**

The data set contains 4,521 observations and 17 variables where seven are numeric variables and ten categorical variables without missing values. 4,000 records with label “no” and 521 records with label “yes” responded to bank marketing campaign. It is observed that more than one contact to the same customer was necessary in order for a customer to subscribe or not subscribe to a term deposit. In order to test our null hypothesis, we will analyzed the education and age attributes. The age attribute shows that ages for customers range between 19 and 87 years old. Figure 1 present the summary statistics of age attribute including a lower and upper 95% confidence level for mean. The histogram in Figure 2 and the box plot in Figure 3 of age attribute indicate that target of telemarketers were customers from early thirties to late thirties. Customers whose age is above sixty has the same number of responses of “no” and “yes”. The Education attribute is categorized into unknown, secondary, primary and tertiary. As presented in the bar charts in Figure 4, most customers who were contacted has a secondary level education. A sample of the analyzed data set is shown in Table 1.

**Table 1**

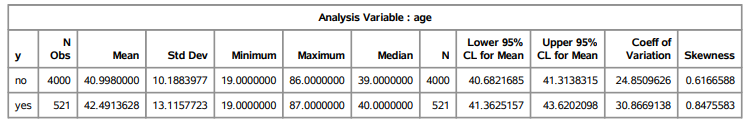
*Bank Data Set Sample*



*Note.* This table shows a sample of the analyzed data set.

**Figure 1**

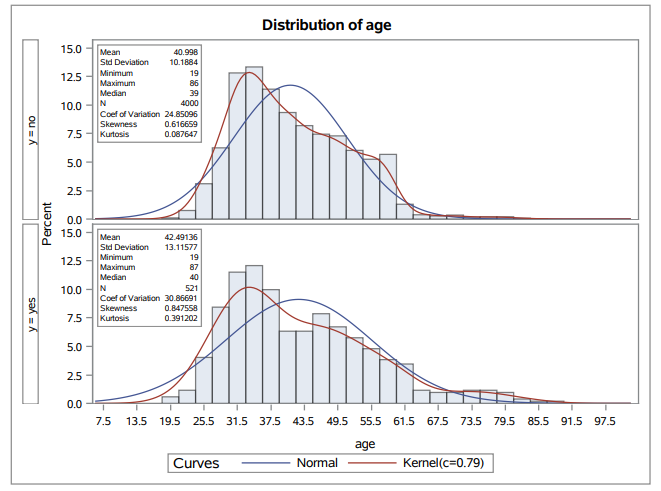
*Age Summary Statistics*



*Note.* This figure illustrates the Summary Statistics output of age attribute.

**Figure 2**

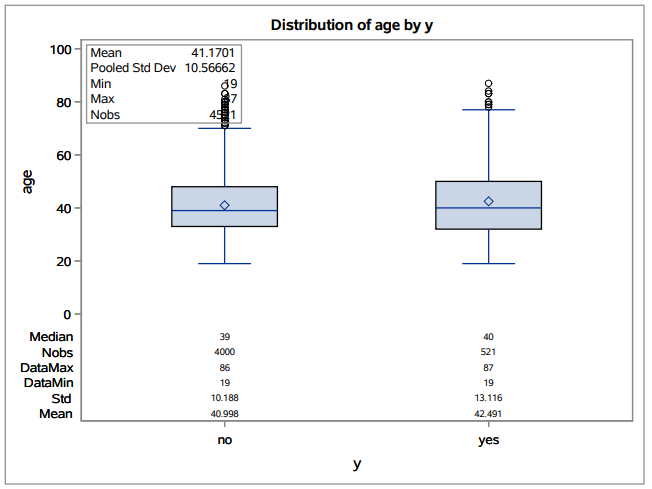
*Age Histogram*



*Note.* This figure illustrates the histogram of age attribute.

**Figure 3**

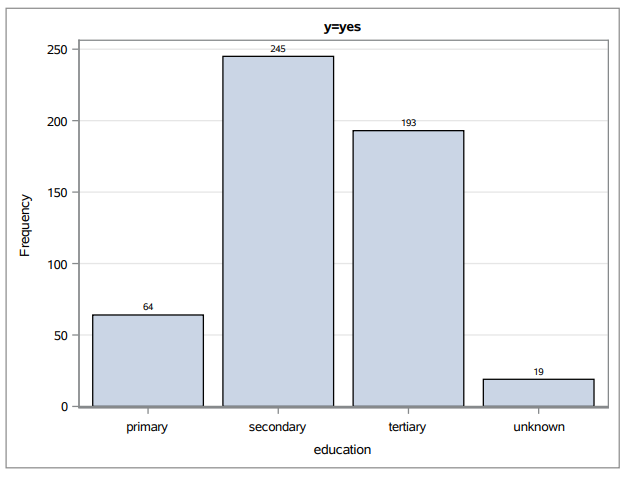
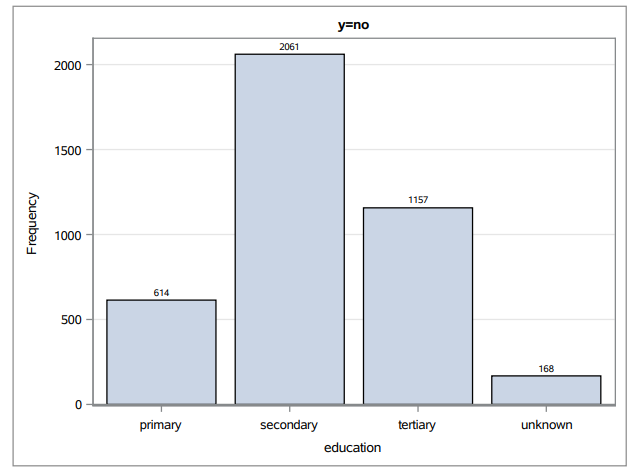
*Age Box Plot*



*Note.* This figure illustrates the comparative box plot of age attribute.

**Figure 4**

*Education Bar Chart*



*Note.* Bar charts of Education attribute group by response denoted by “y”.

**CONCLUSION**

We have analyzed the data set using descriptive analysis on education and age attributes. The findings indicate that were a number of responses from customers whose age is from early thirties to late thirties and has a secondary education level.

Therefore, we can reject the null hypothesis because there are more responses received by telemarketers from customers whose age is between early thirties and late thirties and with secondary education level. There were fewer responses from customers whose age is above sixty years old and has a university education level to whether subscribed (“yes) or not (“no”) subscribed for long term deposit.

The data summary demonstrates that the binary response class of “y” had 521 instances of “yes” as response, and 4,000 instances responded a “no”. The analysis indicates that 11.7% of the total number of customers contacted during the marketing campaign responded “yes” to the promotion, and 88.3% of customers declined to the marketing campaign.

**RECOMMENDATIONS**

We propose that telemarketers target those customers whose age is above sixty years old and with university education level. Telemarketers should convey peace of mind, safe investment as few of their value proposition. They should also prioritize those customers who were part of the previous marketing campaigns.

**REFERENCES**

Apampa, O. (2016). Evaluation of Classification and Ensemble Algorithms for Bank Customer

Marketing Response Prediction. *CSU Global Library*. <https://eds-p-ebscohost-com.csuglobal.idm.oclc.org/eds/pdfviewer/pdfviewer?vid=22&sid=b9dcaf84-51ee-4c7c-9dd0-e1ce6723c2a1%40redis>

Elsalamony, H. (2014). Bank Direct Marketing Analysis of Data Mining Techniques.

*International Journal of Computer Applications*, 85(7).

<https://dx.doi.org/10.5120/14852-3218>

Klose, M., Desai, V., Song, Y., & Gehringer, E. (2020). EDM and privacy: Ethics and legalities

of data collection, usage, and storage. *International Educational Data Mining Society*. <https://files.eric.ed.gov/fulltext/ED607820.pdf>

Li, G. (2021). Financial Early Warning System Model Combining Hybrid Semantic Hierarchy

with Group Method of Data Handling Neural Network for Detection of Banks’ Risks. *Discrete Dynamics in Nature and Society*. <https://doi.org/10.1155/2021/8607667>

Moro, S., Cortez, P., & Rita, P. (June 2014). A data-driven approach to predict the success of

bank telemarketing. *Science Direct*. <https://doi.org/10.1016/j.dss.2014.03.001>

Nachev, A. (2015). Application of data mining techniques for direct marketing. *Computational*

*Models for Business and Engineering Domains*.

<http://www.foibg.com/ibs_isc/ibs-30/ibs-30-p09.pdf>

Roy, S. (2019). Machine Learning Case Study: A data-driven approach to predict the success of

bank telemarketing. *Towards Data Science*. <https://towardsdatascience.com/machine-learning-case-study-a-data-driven-approach-to-predict-the-success-of-bank-telemarketing-20e37d46c31c>

Song, Y., Kong, X., Huang, S., Zhang, C. (2021). Fast Training Logistic Regression via

Adaptive Sampling. *Scientific Programming*. <https://doi.org/10.1155/2021/9991859>

UCI Machine Learning Repository. (n.d.). *Bank Marketing Data Set*. <https://archive.ics.uci.edu/ml/datasets/bank+marketing>

Wickramanayake, D. & Meddegoda, P. (2020). Analyzing previous marketing data of a bank to

maximize its term deposits subscribers in future marketing campaigns. *Research Gate*. <http://dx.doi.org/10.13140/RG.2.2.13386.52161>