Continuous Assessment 1

Enterprise Database Technologies

4th Year Computer Science

Sean Shelley X00121161 | Robert Osborne X00109797

Contents

[Introduction 3](#_Toc508794098)

[Commentary on Overall Findings and Observations 3](#_Toc508794099)

[Histograms for Numeric Variables Overlaid with the Target Variable 4](#_Toc508794100)

[Age by Target 4](#_Toc508794101)

[Campaign by Target 4](#_Toc508794102)

[Duration by Target 5](#_Toc508794103)

[PDays by Target 5](#_Toc508794104)

[Consumer Confidence Index by Target 6](#_Toc508794105)

[Consumer Price Index by Target 6](#_Toc508794106)

[Euribor by Target 7](#_Toc508794107)

[Nr.Employed by Target 7](#_Toc508794108)

[Previous by Target 8](#_Toc508794109)

[Bar Charts for Categorical Variables Overlaid with the Target Variable 9](#_Toc508794110)

[Contact by Target 9](#_Toc508794111)

[Day of Week by Target 9](#_Toc508794112)

[Education by Target 10](#_Toc508794113)

[Housing by Target 10](#_Toc508794114)

[POutcome by Target 11](#_Toc508794115)

[Martial by Target 11](#_Toc508794116)

[Job by Target 12](#_Toc508794117)

[Month by Target 12](#_Toc508794118)

[Loan by Target 13](#_Toc508794119)

[Outlier in the Data 14](#_Toc508794120)

[Appendix: 15](#_Toc508794121)

[Scatter Plots for Each Numeric Pair 15](#_Toc508794122)

# Introduction

The following piece of work, for this assignment, was created by Robert Osborne and Sean Shelley

It included the use of the ‘*Banking*’ dataset from Moodle.

## Commentary on Overall Findings and Observations

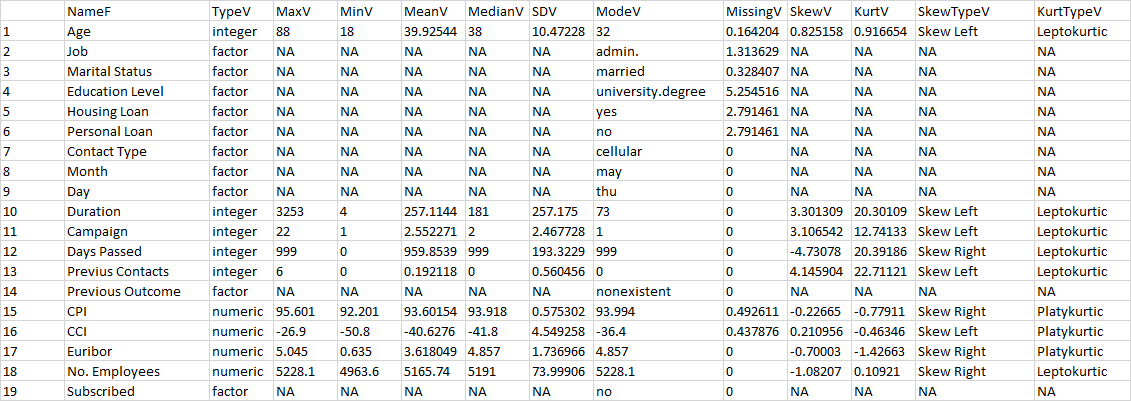


Figure 1: Results Data Frame

From our preliminary observations we can see that attributes of type factor do not show whether they are skewed left or right, unlike the numeric or integer variables. Most numeric variables are skewed to the left.

There is a significant amount of N/A values in relation to factor variables, in our results data frame.

## Correlation Between Predictor Variables

A screenshot of a cell phone

Description generated with high confidence



# Histograms for Numeric Variables Overlaid with the Target Variable

## Age by Target

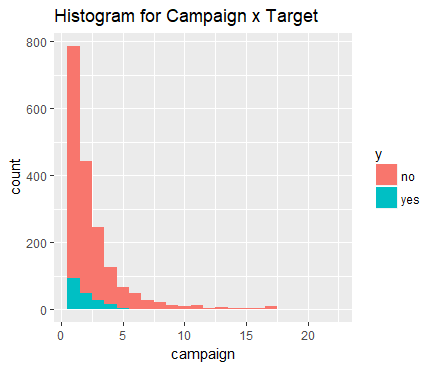
A picture containing screenshot

Description generated with high confidence

The Age for the Banks marketing campaign is heavily focused on those around the age of 30. From the information displayed in the histogram, we see that most people contacted by the marketing campaign, did not subscribe to the term deposit.

The graph also shows a skewness towards the younger age demographic. While also being lightly kurtosis, the graph is also Leptokurtic in natural meaning it is positive, with a tall or high point. As the graph starts tailing off as the age of the clients increase.

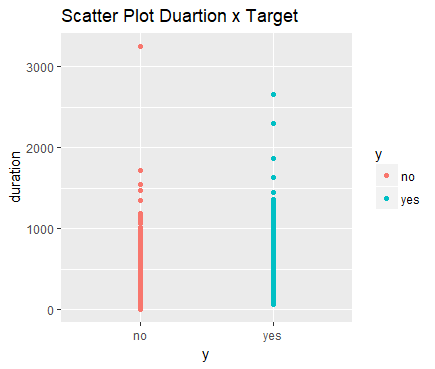
## Campaign by Target



Form this graph we can that the first clients contacted, where contacted the most, with around 800 calls from the marketing campaign. With very few of those who the campaign first contacted subscribing to the term deposit.

The graph shows a skewness towards the left, as show in the graph with the first client receiving the most phone calls. It is also lightly kurtosis as the campaign stars called move from client to client.

## Duration by Target

A picture containing text, map

Description generated with high confidence

Duration of most of the phone calls the marketing campaign made were very short. Most of which were less than 1000 seconds. Like with the other graphs, most did not subscribe to the term deposit.

Skewness can also be seen in the Histogram as it is skewed to the left, towards the shorter phone calls. The kurtosis type is Leptokurtic as the Histogram has a tall peak, resulting it being positive.

From the Histogram and Scatter Plot we can see that there is an outlier in our data. One phone call last over 3000 seconds.

## PDays by Target

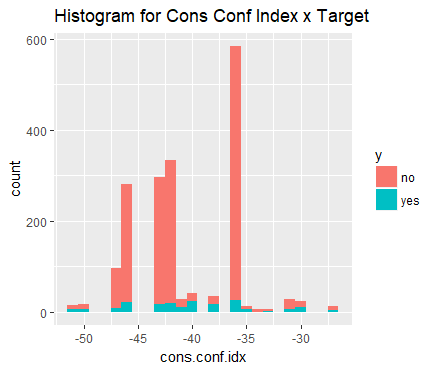
A screenshot of a cell phone

Description generated with high confidence

From this graph we can see that the number of days passed since the clients last received a call from the Bank’s marketing campaign has been less than 10, however the number of clients that received a class for the first time is significantly higher.

We also see that is skewed to the right and because the peak of the number of people who received calls for the first time is so high, it is also has a kurtosis type of Leptokurtic.

## Consumer Confidence Index by Target



From this graph we can see how optimistic Clients are in the state of the economy or their personal finances, through their spending and savings. This graph also skews to the left, it is also a positive graph, as it has a high peak.

Meaning that is kurtosis type is Leptokurtic. This graph also shows that most people the marketing campaign contacted didn’t subscribe to the term deposit, even in times when they were more optimistic about their finances and the economy.

## Consumer Price Index by Target

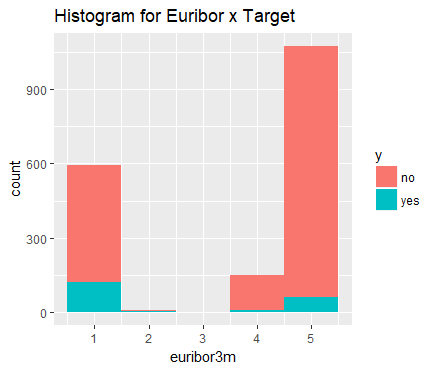
A picture containing text, screenshot

Description generated with high confidence

This graph shows the Consumer Price Index and that Clients subscribed more to the term deposit, that the marketing campaign were pushing. Mainly when the CPI was between 93 and 94. However, overall most did not subscribe to the term deposit.

It is a positive graph, as it has a high peak, which means it is a leptokurtic, kurtosis type. This graph also skews the right.

## Euribor by Target



This graph highlights the number of subscribers the marketing campaign, when compared to the Euribor rate at the same time the campaign ran. We can see that there are two main peaks, one when the Euribor rate was at 1 and again when it was a 5.

Most subscribers to the term deposit came in when the Euribor was at 1, but overall the not moany of the clients contacted by the campaign subscribed to the term deposit.

This graph has a kurtosis type of Leptokurtic, as it has a high peak when the Euribor rate is at 5. It also skews to the right.

## Nr.Employed by Target

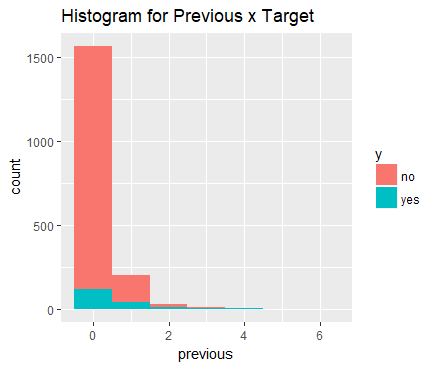
A picture containing text

Description generated with high confidence

From this graph we can see that the Number of Employees increase after each quarter during the campaign. We can also see that the graph skews to the right, which highlights the increase in Number of Employees.

This graph could also be a kurtosis type of Leptokurtic, as its highest peak is when the Number of Employees is over 5200.

## Previous by Target



This graph shows number of time a client was contacted before this new marketing campaign began. As we can see the first few clients were contacted and multiple occasions. This graph also skews to the left. Most of those who were contacted previously, did not subscribe to the term deposit

Again, like the other graphs the have been created, this graph has a kurtosis type of Leptokurtic.

# Bar Charts for Categorical Variables Overlaid with the Target Variable

## Contact by Target

A screenshot of a cell phone

Description generated with high confidence

From this bar chart we can see that the way in which clients were contacted, by the marketing campaign was mainly through cellular devices. With calls from a telephone being the second most popular way the clients were contacted.

This chart shows that no calls were made to clients, by way of a cell. It also shows that more clients subscribe to the term deposit over cellular device than telephone and cells.

## Day of Week by Target

A screenshot of a cell phone

Description generated with high confidence

This chart shows that the most calls to clients were made on Thursday, with most of the clients that subscribed to the banks term deposit on Monday, when compared to the other days of the week.

## Education by Target

A screenshot of a cell phone

Description generated with high confidence

This chart shows the more clients with University Degrees were contacted the campaign. With clients with Basic 6 years of education being contacted the least, not including NA.

People with a University Degrees also showed the highest numbers that subscribed to the term deposit.

## Housing by Target

A screenshot of a cell phone

Description generated with high confidence

From this chart is shows that a higher number of people with bank loans were contacted by the marketing campaign, than compared to those without bank loan for housing.

Both groups with a housing loan and without a housing loan shared roughly an equal amount of subscriptions to the bank’s term deposit.

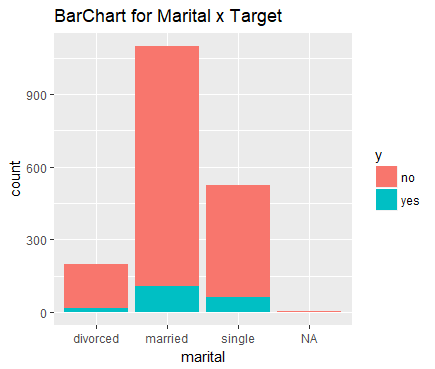
## POutcome by Target

A screenshot of a cell phone

Description generated with high confidence

The out of the previous campaign for clients is mainly non-existent. As shown by this chart.

## Martial by Target



Most clients contacted by the campaign were married, with clients that were single coming as the second most contacted. People who were divorced or widowed, received the lowest number of calls from the campaign.

Married people also had the higher numbers, in terms of subscribing to the term deposit. While overall, clients mainly rejected the term deposit.

## Job by Target

A screenshot of a cell phone

Description generated with high confidence

From the Chart we can see that most of those contacted by the campaign had jobs as Administrators, followed closely by Blue Collar workers.

Admins show the strongest numbers in terms of subscribing to the term deposit., when compared to the other job types.

## Month by Target

A screenshot of a cell phone

Description generated with high confidence

The month of May shows that most of calls, by the campaign, were made in this month and it also has the most number of people who have subscribed to the term deposit.

## Loan by Target

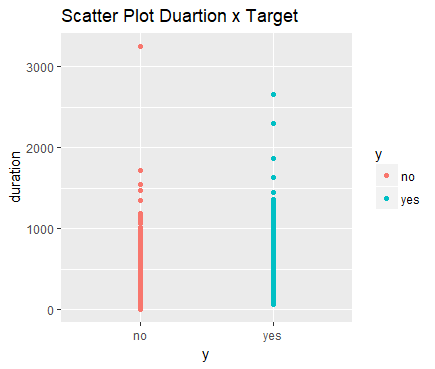
A screenshot of a cell phone

Description generated with high confidence

From this chart we can see most of the people contacted by the marketing campaign for the bank have no personal loans. But showed the higher number in terms of clients subscribing to the term deposit of the bank, when compared to the other bars in the chart.

People with no bank loans also showed the higher number of clients that didn’t subscribe to the term deposit after the communications with the marketing campaign.

# Outlier in the Data

A picture containing text, map

Description generated with high confidence

As Mention in the Histograms for Numeric Variables Overlaid with the Target Variable above, in the Duration by Target graphs, we can see graphically that there is an outlier in the data.



Statistically, we can see that the Max Value for Duration by the Target is 3253 seconds. Which we can see in the graph and scatter plot. With a Min Value of 4, the mean for this graph is 257.1144 seconds.

# Attributes most or least linked to the Target variable

# Appendix:

# Scatter Plots for Each Numeric Pair

|  |  |  |  |
| --- | --- | --- | --- |
|  | A close up of a piece of paper  Description generated with high confidence | A close up of a piece of paper  Description generated with high confidence | A close up of a white background  Description generated with high confidence |
| A screenshot of a cell phone  Description generated with high confidence | A screenshot of a cell phone  Description generated with very high confidence | A close up of a white background  Description generated with high confidence | A screenshot of a cell phone  Description generated with high confidence |
| A close up of a piece of paper  Description generated with high confidence | A screenshot of a cell phone  Description generated with very high confidence | A screenshot of a cell phone  Description generated with high confidence | A close up of a white wall  Description generated with high confidence |
| A screenshot of a cell phone  Description generated with high confidence |  | A picture containing photo  Description generated with high confidence | A screenshot of a cell phone  Description generated with very high confidence |
| A screenshot of a cell phone  Description generated with very high confidence | A picture containing sky  Description generated with high confidence | A close up of a white wall  Description generated with high confidence | A screenshot of a cell phone  Description generated with very high confidence |
| A screenshot of a cell phone  Description generated with high confidence | A picture containing outdoor  Description generated with high confidence | A picture containing photo  Description generated with high confidence | A screenshot of a cell phone  Description generated with very high confidence |
| A screenshot of a cell phone  Description generated with high confidence | A screenshot of a cell phone  Description generated with high confidence | A close up of a piece of paper  Description generated with high confidence | A screenshot of a cell phone  Description generated with very high confidence |
| A screenshot of a cell phone  Description generated with high confidence | A screenshot of a cell phone  Description generated with high confidence |  | A close up of a piece of paper  Description generated with high confidence |
|  | A picture containing photo  Description generated with high confidence | A close up of a device  Description generated with high confidence |  |