**Bank Marketing**

Predict customers who will open a term deposit account

**Problem Statement:**

To find the best strategies to improve the marketing campaign of a banking company and to study the effectiveness of older campaign and measures to improve it. We will analyze the results of the last marketing campaign, identify the patterns to develop future strategies.

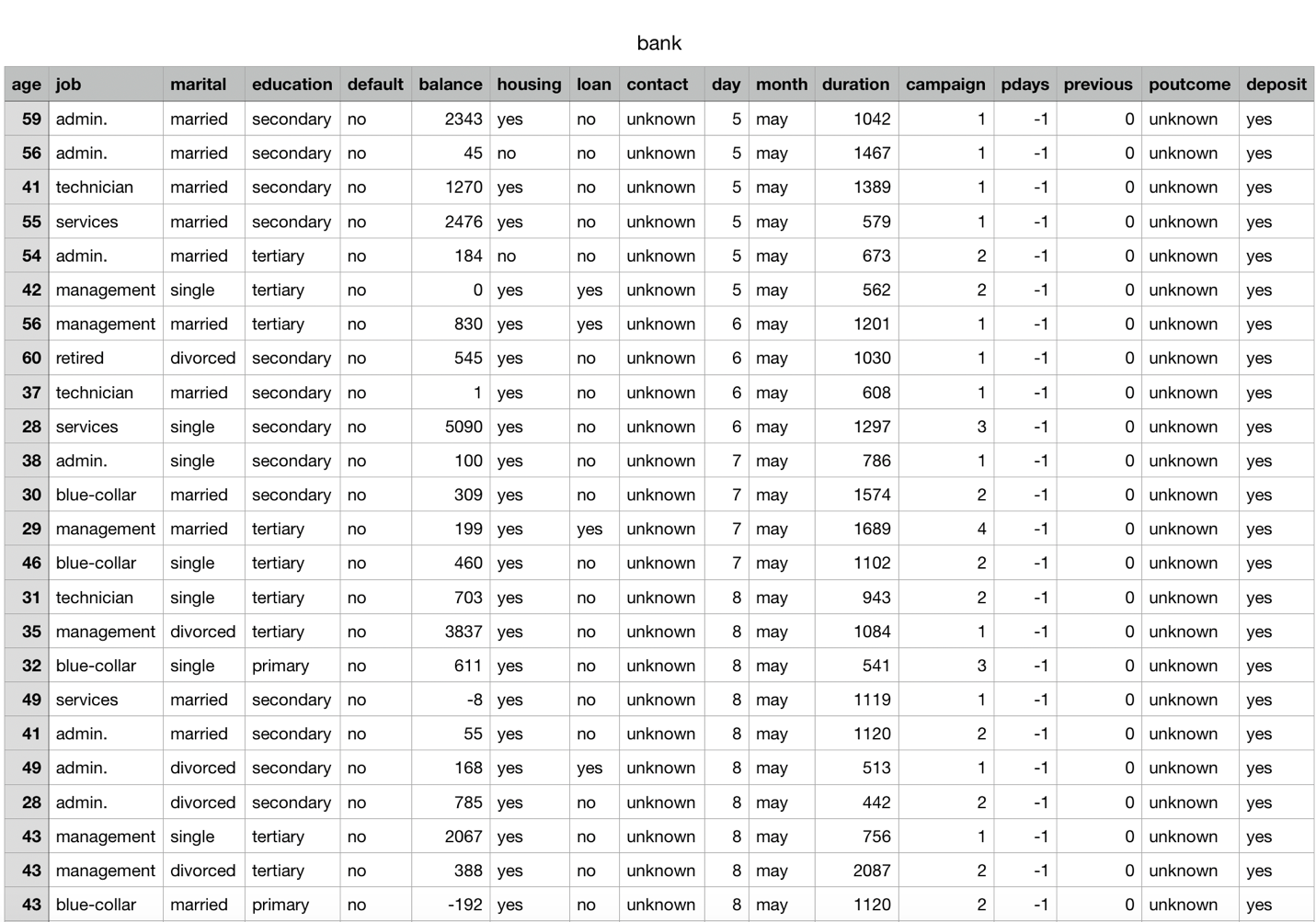
**Implication**:

Just like any other industry, marketing is an utmost important aspect of success.  
Successful Marketing strategies will provide a bank with potential clients and that is what anyone would want while stepping into the banking sector. Marketing will let customers know about a bank, and promotional marketing strategies will attract customers to get associated with a bank.

**Date Source:**

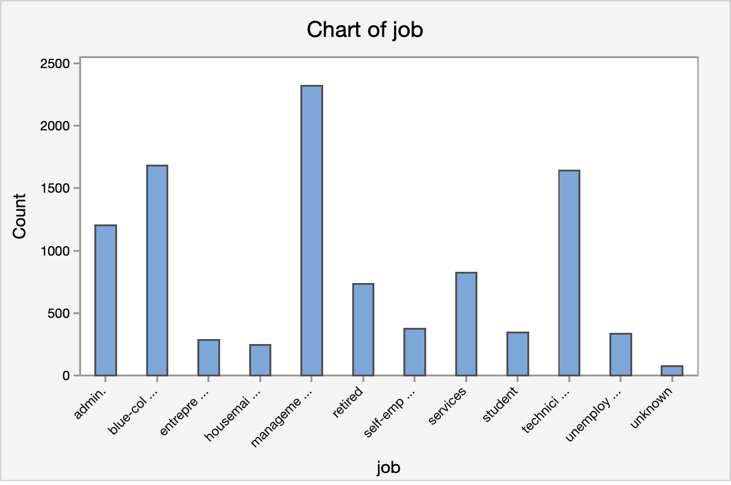
<https://www.kaggle.com/rouseguy/bankbalanced>

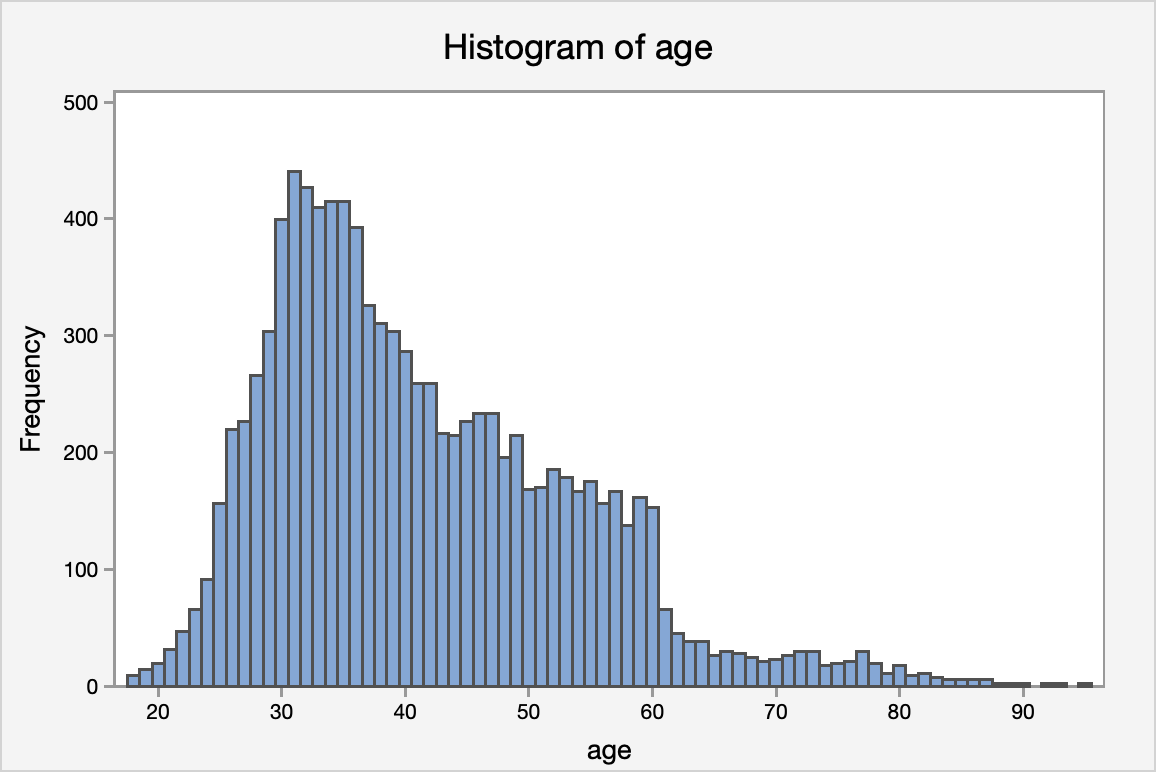
**Dataset Snapshot:**

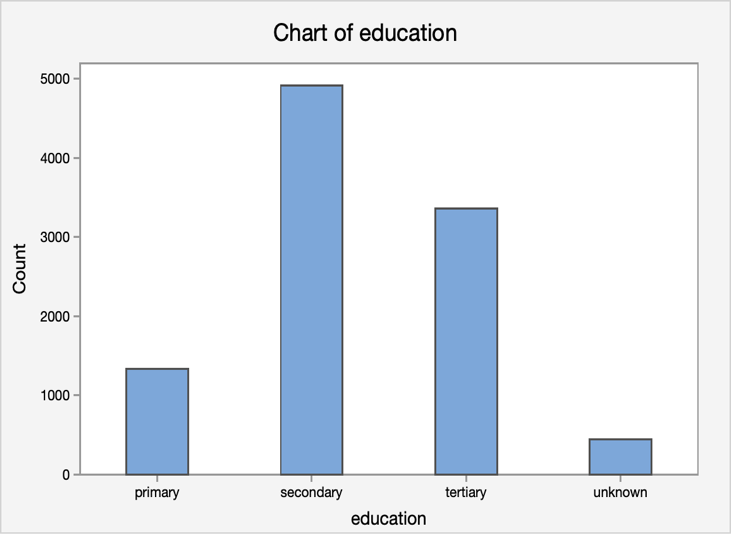
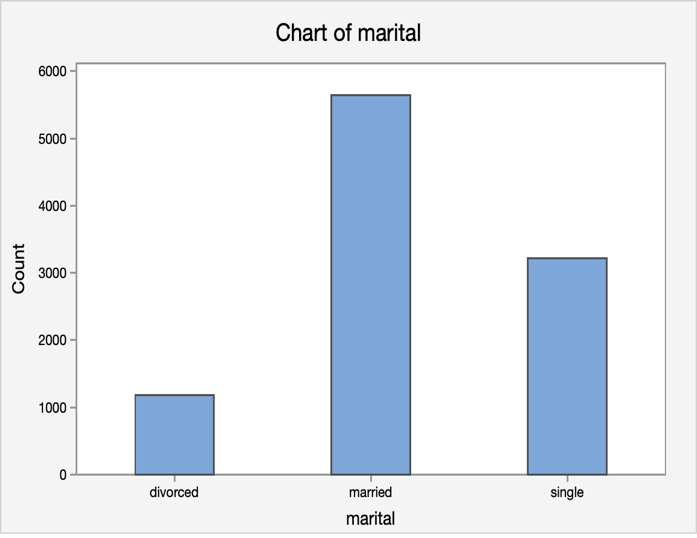


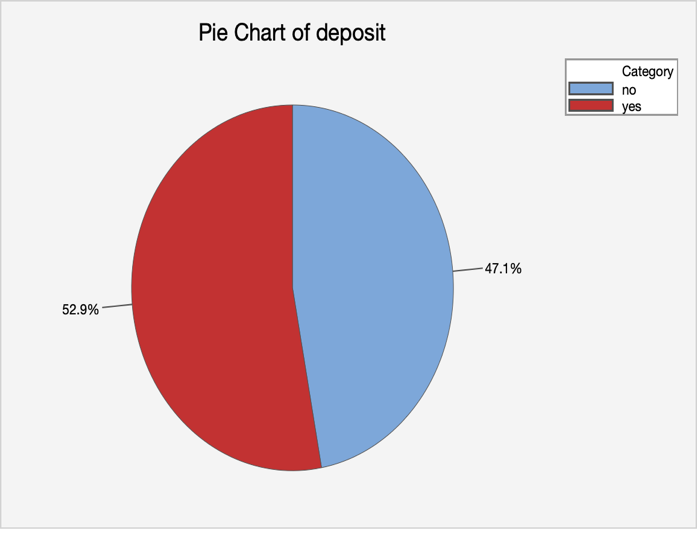
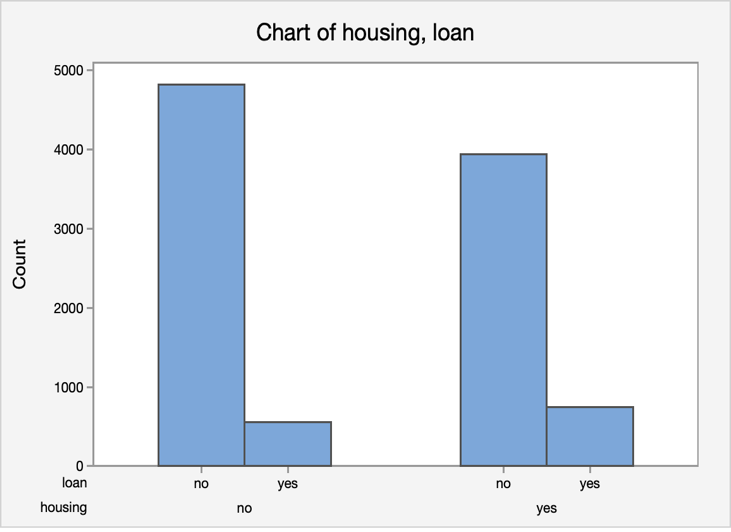
**Data Description:**

There are 17 features including the output and we have around 11.2k unique set of records denoting the number of customers in the dataset.









**Bank client data:**

1 - **age:** (numeric)  
2 - **job:** type of job (categorical: 'admin.','blue-collar','entrepreneur','housemaid','management','retired','self-employed','services','student','technician','unemployed','unknown')  
3 - **marital:** marital status (categorical: 'divorced','married','single','unknown'; note: 'divorced' means divorced or widowed)  
4 - **education:** (categorical: primary, secondary, tertiary and unknown)  
5 - **default:** has credit in default? (categorical: 'no','yes','unknown')  
6 - **housing:** has housing loan? (categorical: 'no','yes','unknown')  
7 - **loan:** has personal loan? (categorical: 'no','yes','unknown')  
8 - **balance:** Balance of the individual.

Related with the last contact of the current campaign:

9 - **contact:** contact communication type (categorical: 'cellular','telephone')   
10 - **month:** last contact month of year (categorical: 'jan', 'feb', 'mar', ..., 'nov', 'dec')  
11 - **day:** last contact day of the week (categorical: 'mon','tue','wed','thu','fri')  
12 - **duration:** last contact duration, in seconds (numeric). Important note: this attribute highly affects the output target (e.g., if duration=0 then y='no'). Yet, the duration is not known before a call is performed. Also, after the end of the call y is obviously known. Thus, this input should only be included for benchmark purposes and should be discarded if the intention is to have a realistic predictive model.

Other attributes:

13 - **campaign:** number of contacts performed during this campaign and for this client (numeric, includes last contact)  
14 - **pdays:** number of days that passed by after the client was last contacted from a previous campaign (numeric; 999 means client was not previously contacted)  
15 - **previous:** number of contacts performed before this campaign and for this client (numeric)  
16 - **poutcome:** outcome of the previous marketing campaign (categorical: 'failure','nonexistent','success')

Output variable (desired target):  
17 - **y** - has the client subscribed a term deposit? (binary: 'yes','no')

**Data mining techniques to be used:**

Clustering and classification are the main type of Data mining techniques that we will be using to segment our data and study the findings and model our marketing strategy to gain more customers enrolling into Term deposit.

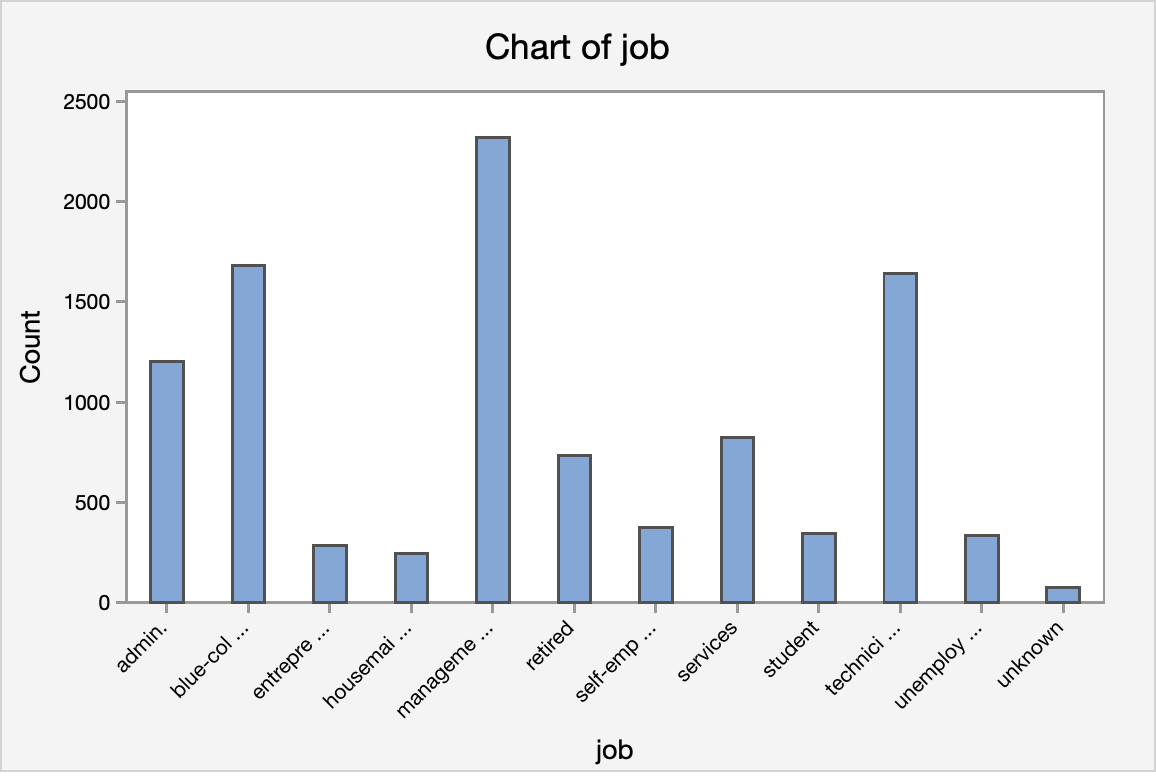
**Data Preprocessing:**

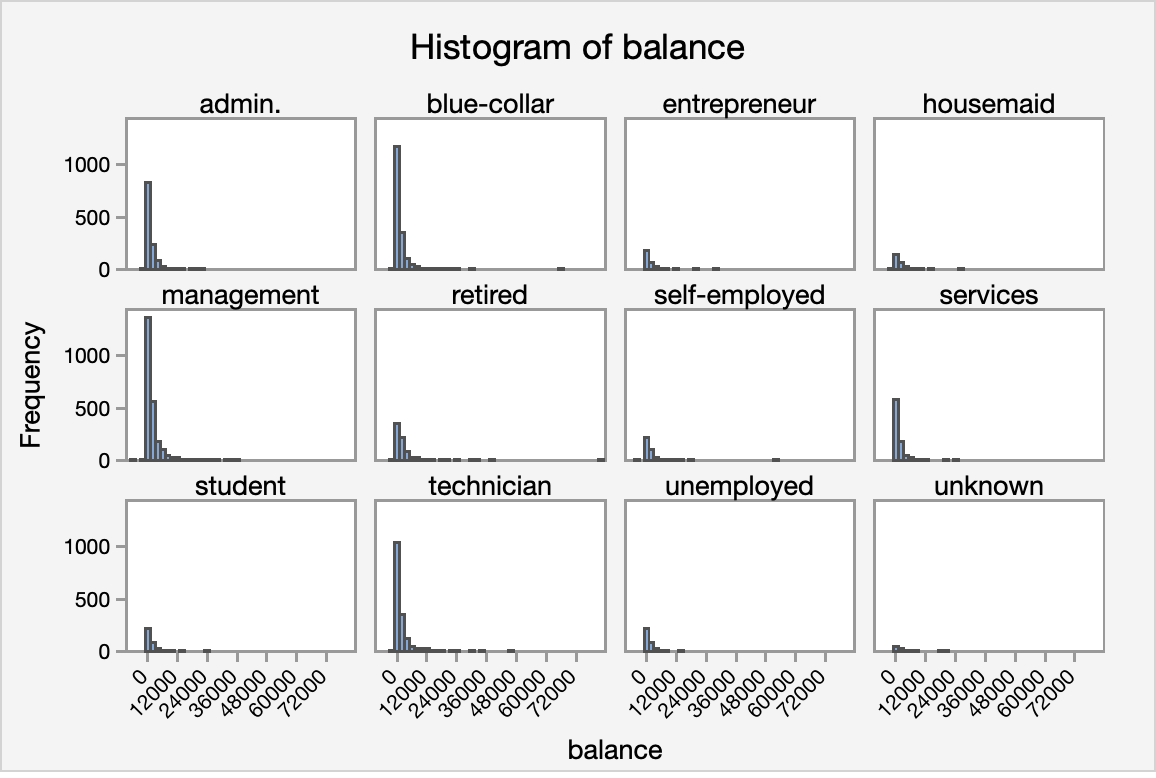
* We removed the columns and ‘poutcome’ and ‘contact’ from the dataset as more than 30% of the data in these columns were empty or unknown.
* We also converted yes and no in our predicate column ‘Deposit’ into binary form as 1s and 0s.

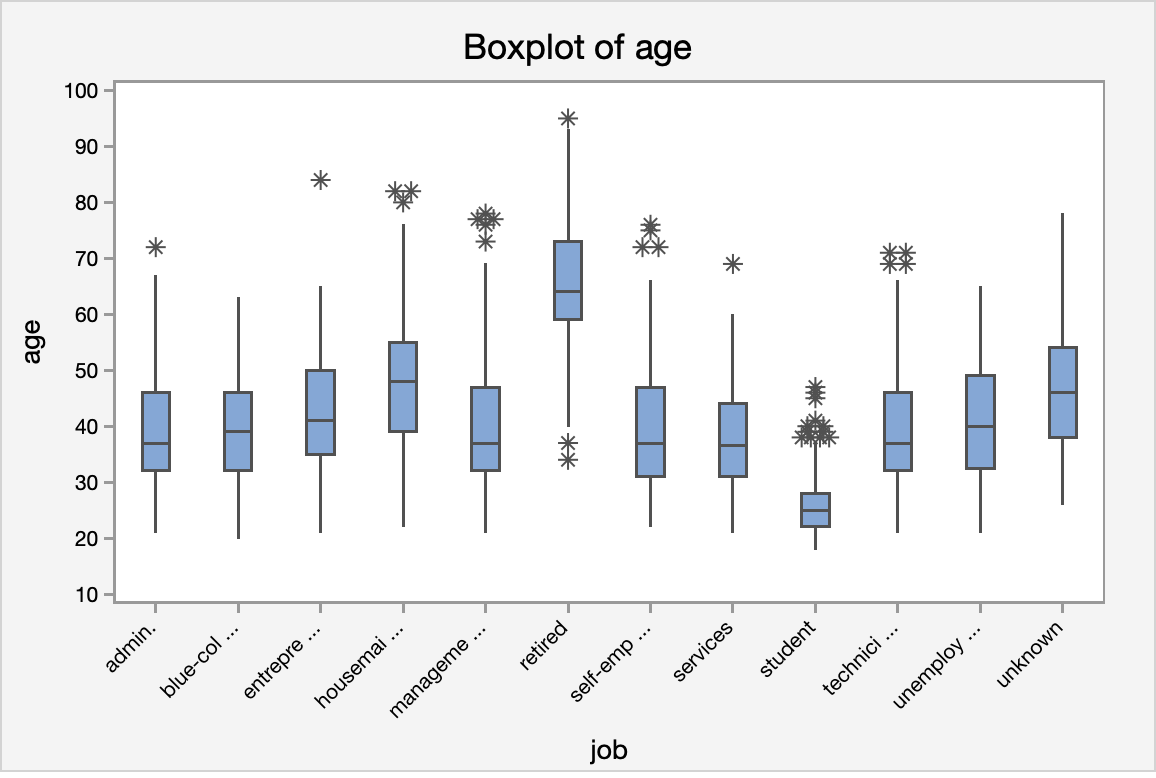
**Data Analysis:**

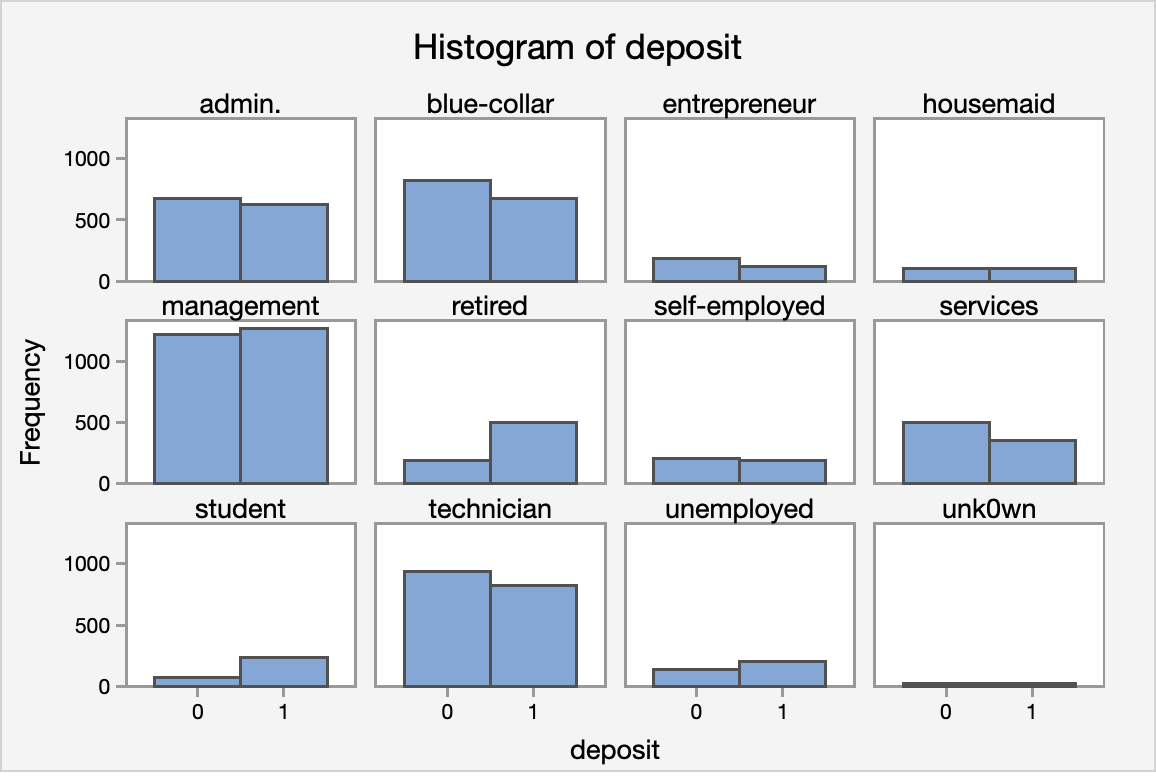
### **Analysis by Occupation:**

* **Number of Occupations:**Management is the occupation that is more prevalent in this dataset.
* **Age by Occupation:**As expected, the retired are the ones who have the highest median age while student are the lowest.
* **Balance by Occupation:**Management and Retirees are the ones who have the highest balance in their accounts.
* **Deposit by Occupation**: We could see that Retirees and Students tend to take the term deposit in more compared to others.

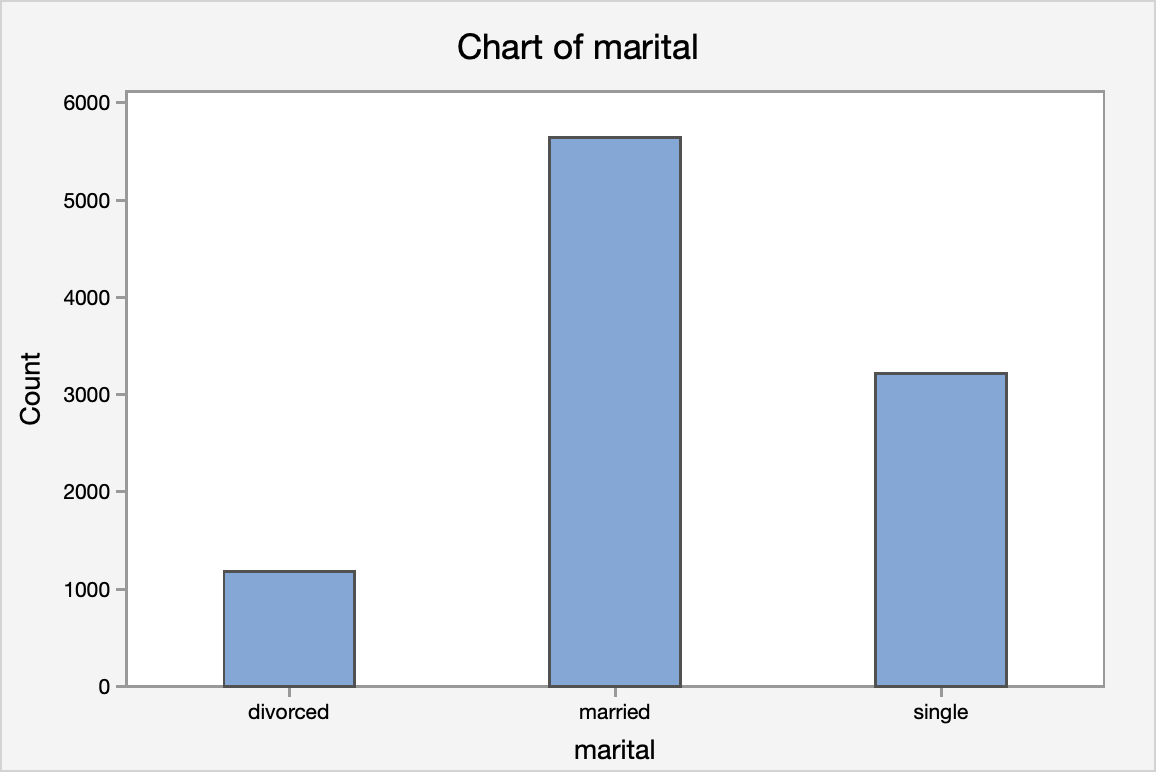


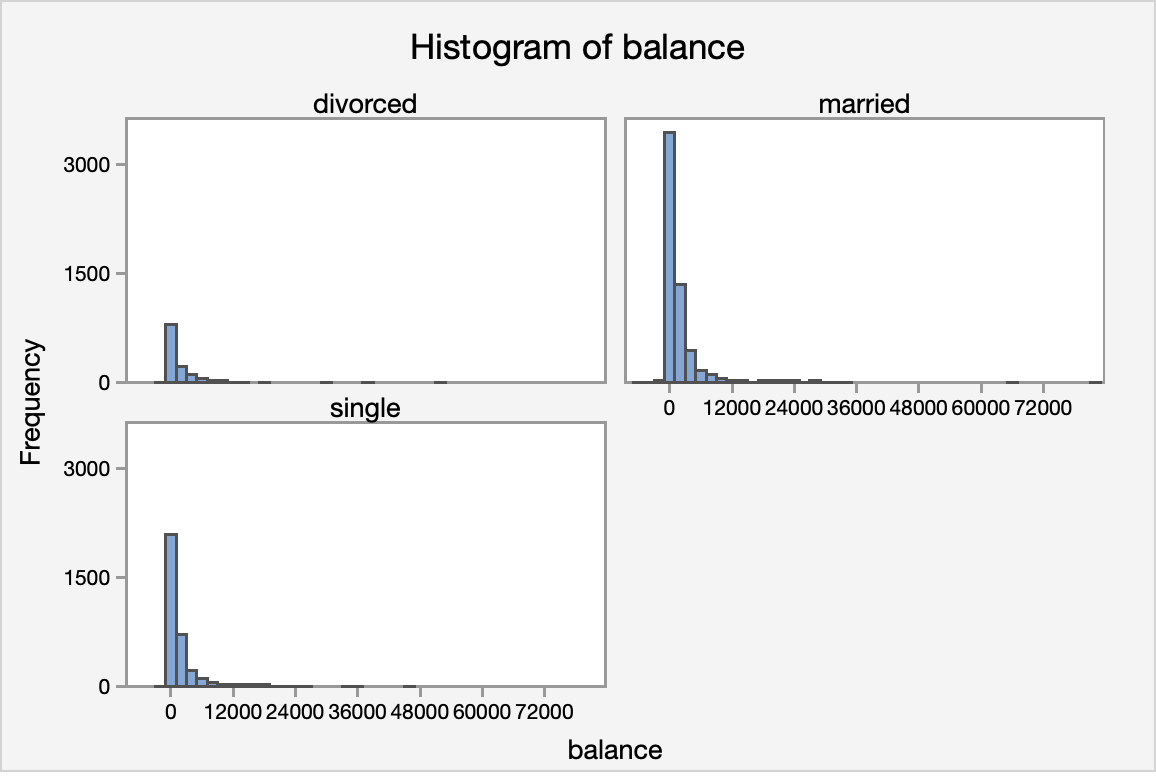






**Analysis by Marital Status:**

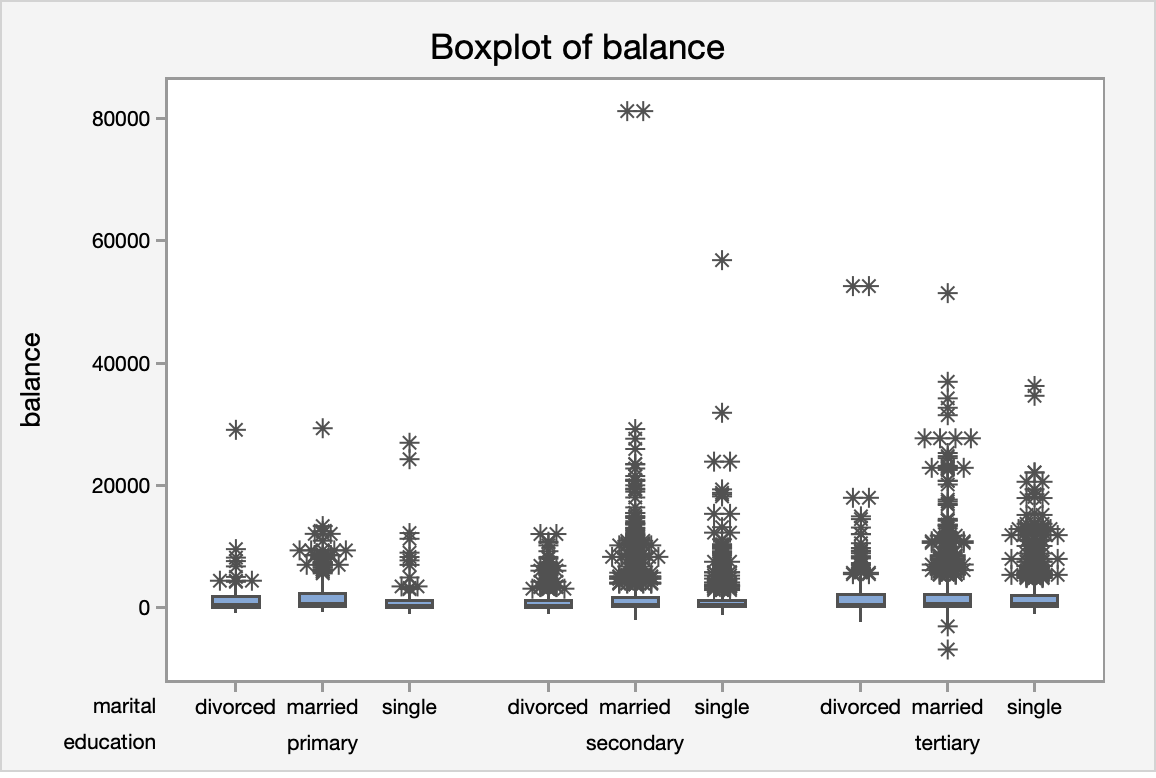


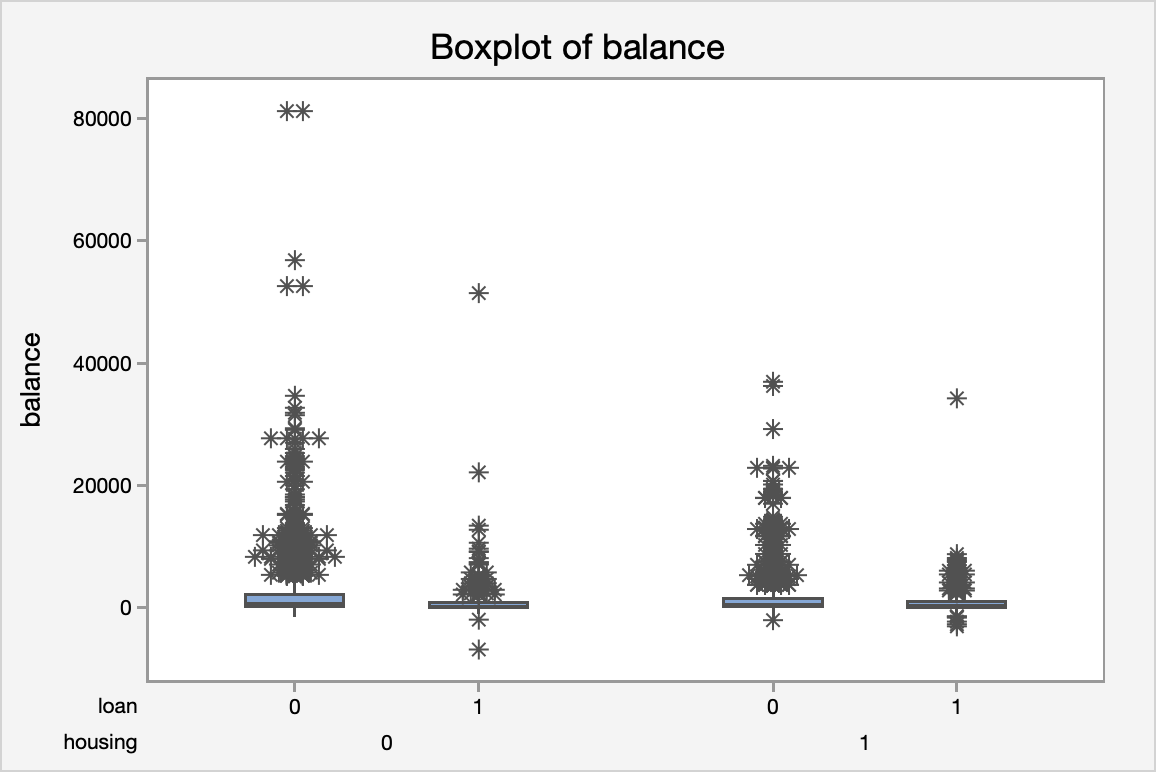


We didn't find any significant insights other than most **divorced individuals**are broke. No wonder since they have to split financial assets! Nevertheless, since no further insights have been found we will proceed to clustering marital status with education status. Let's see if we can find other groups of people in the sample population.

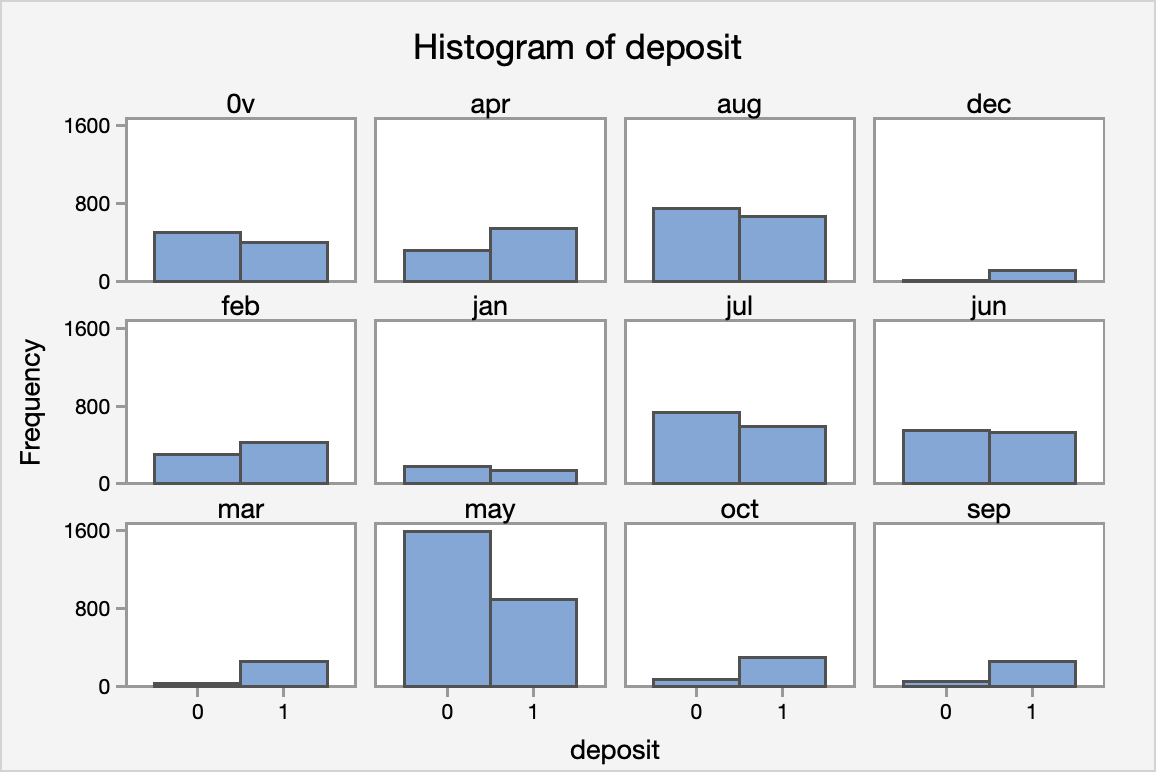
### **Clustering Marital Status and Education:**

* **Marital Status:**As discussed previously, the impact of a divorce has a significant impact on the balance of the individual.
* **Education:**The level of education also has a significant impact on the amount of balance a prospect has.
* **Loans:**Whether the prospect has a previous loan has a significant impact on the amount of balance he or she has.



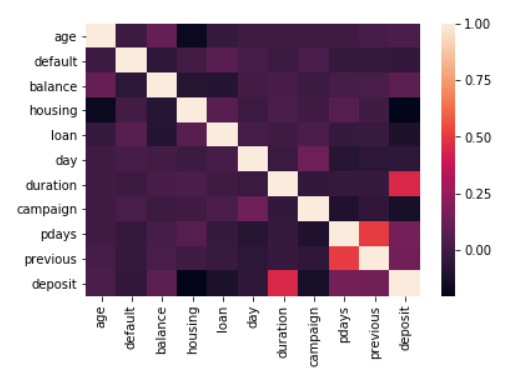


**Analysis based on Month:**



We saw that the the month of highest level of marketing activity was the month of **May**. However, this was the month that potential clients tended to reject term deposits offers (Lowest effective rate: -34.49%). For the next marketing campaign, it will be wise for the bank to focus the marketing campaign during the months of **March, September, October and December.** (December should be under consideration because it was the month with the lowest marketing activity, there might be a reason why december is the lowest.)

**Correlation Heatmap:**

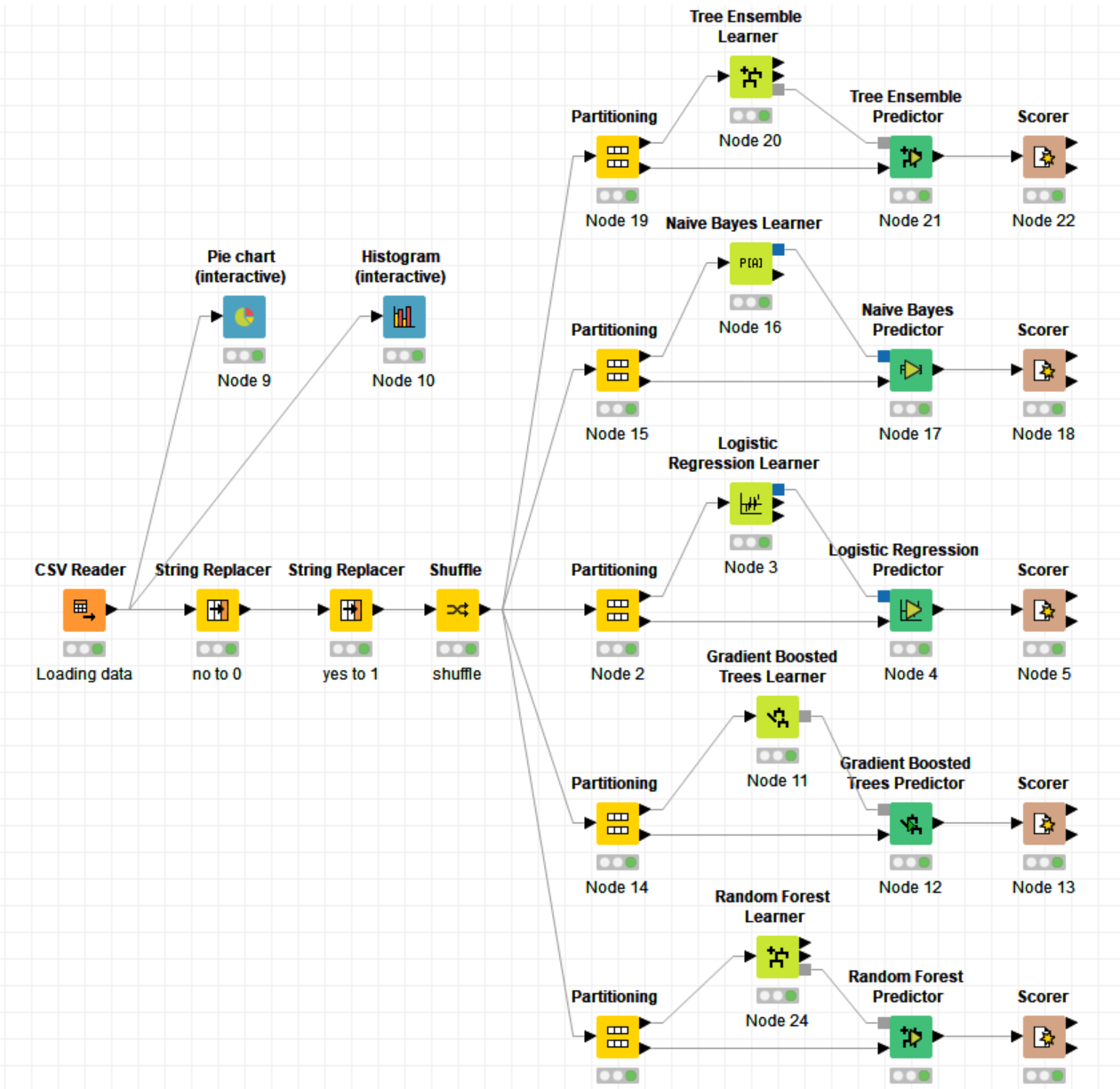


We try to check If there is any attribute correlated and we find that only duration and deposit has some correlation. This is obvious as the there is high chance that a person will accept the plan if he has the interest to listen to the caller.

**Predictions using Different Algorithms:**

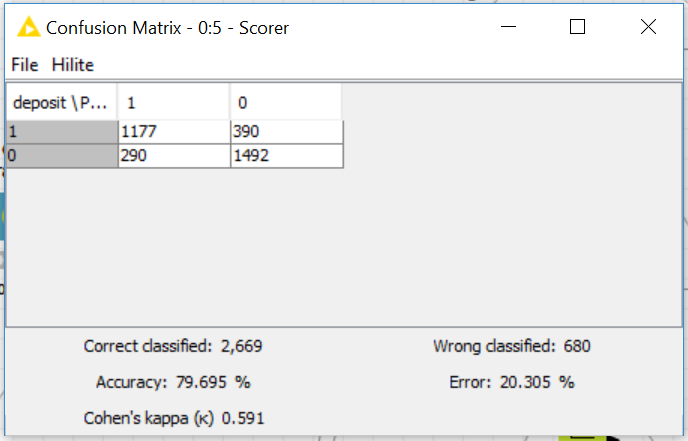
We tried to develop models with different prediction algorithms using KNIME.

**KNIME MODEL:**

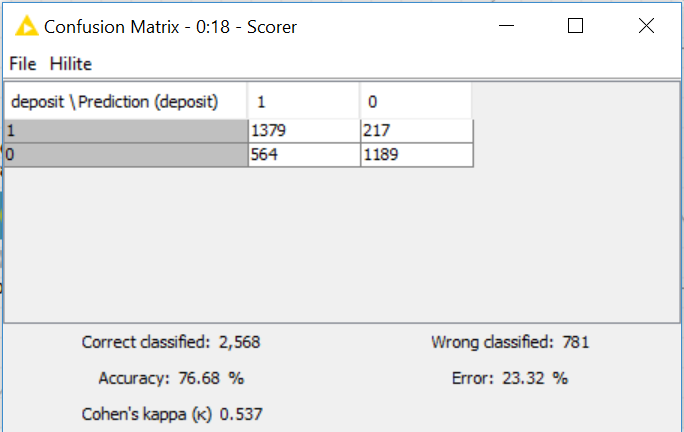
****

We used Tree Ensemble, Naïve Bayes, Logistic Regression, Gradient Boosted, Random Forest and got different accuracy for each.

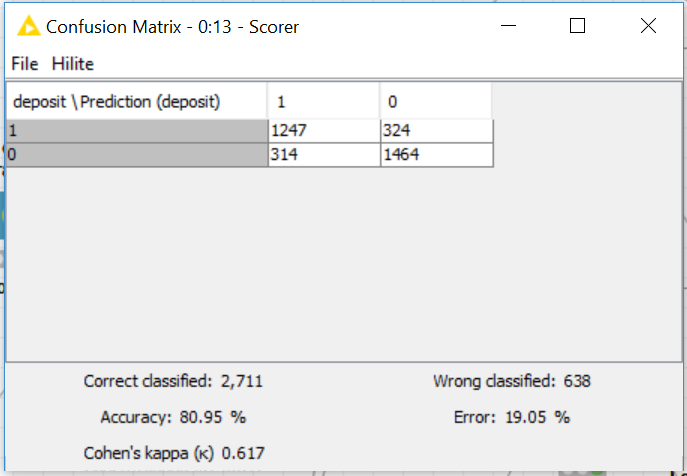
Logistic Regression:



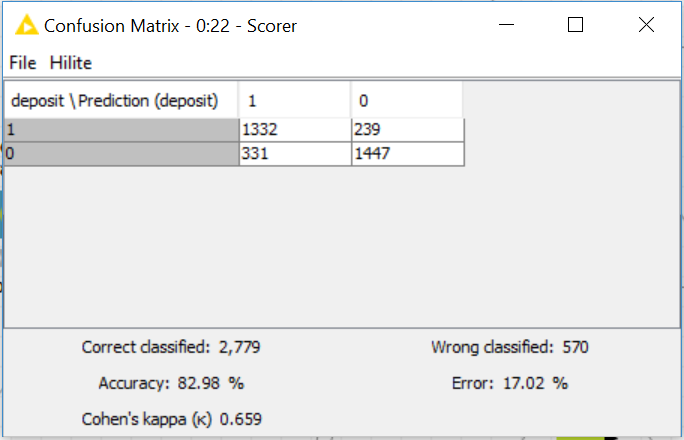
Naïve Bayes:



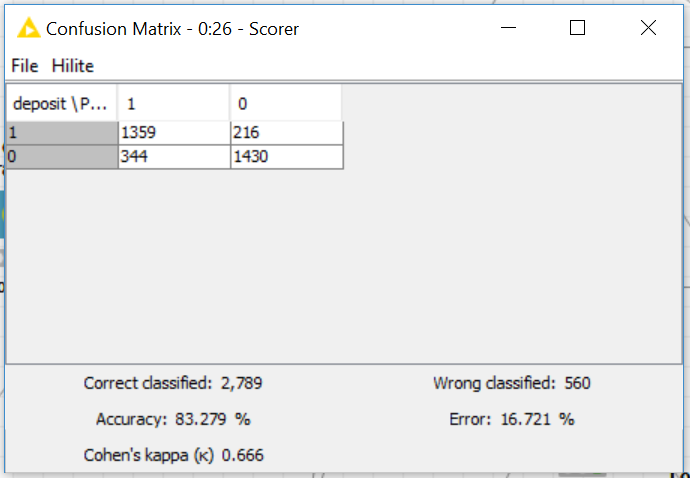
Gradient Boost:



Tree Ensemble:



Random Forest:



Out of all we find Random Forest giving better results and accuracy.

## **Solutions for the Next Marketing Campaign (Conclusion):**

1) **Months:** We saw that the the month of highest level of marketing activity was the month of **May**. However, this was the month that potential clients tended to reject term deposits offers (Lowest effective rate: -34.49%). For the next marketing campaign, it will be wise for the bank to focus the marketing campaign during the months of **March, September, October and December.**   
  
2) **Season:** Potential clients opted to subscribe term deposits during the seasons of **fall** and **winter**. The next marketing campaign should focus its activity throughout these seasons.   
  
3) **Age Category:** The next marketing campaign of the bank should target potential clients in their 20s or younger and 60s or older. The youngest category had a 60% chance of subscribing to a term deposit while the eldest category had a 76% chance of subscribing to a term deposit. It will be great if for the next campaign the bank addressed these two categories.   
  
4) **Occupation:** potential clients that were students or retired were the most likely to subscribe to a term deposit. Retired individuals, tend to have more term deposits in order to gain some cash through interest payments. Students were the other group that used to subscribe term deposits.  
  
5) **House Loans and Balances:** Potential clients in the low balance and no balance category were more likely to have a house loan than people in the average and high balance category. The next marketing campaign should focus on individuals of average and high balances in order to increase the likelihood of subscribing to a term deposit.

By combining all these strategies and simplifying the market audience the next campaign should address, it is likely that the next marketing campaign of the bank will be more effective than the current one. We can use the model developed in KNIME to catalyze decision making for the bank marketing team.

**Reference**:

<https://thefinancialbrand.com/57325/ten-financial-marketing-priorities-trends/>

<https://www.ngdata.com/10-bank-marketing-strategies/>

<https://medium.com/@sifium/machine-learning-types-of-classification-9497bd4f2e14>

<https://www.knime.com/documentation>