ISYS 5843 - Project

## Data Dictionary

|  |  |  |  |
| --- | --- | --- | --- |
| # | Attribute | Data type | Illustration |
| 1 | Age | Numeric | Age of customer |
| 2 | Job | Categorical | Job of customer  'admin','unknown','unemployed','housemaid','entrepreneur’ etc. |
| 3 | Marital Status | Categorical | Marital status of customer  ‘married','divorced','single'. |
| 4 | Education | Categorical | 'unknown','secondary','primary','tertiary' |
| 5 | Default Credit | Binary (Categorical) | Yes/No |
| 6 | Balance | Numeric | Annual balance in account |
| 7 | Housing loan | Binary (Categorical) | Yes/No |
| 8 | Personal loan | Binary (Categorical) | Yes/No |
| 9 | Contact type | Categorical | Contact communication type  ‘unknown','telephone','cellular' |
| 10 | Day | Numeric | Last contact day of the month |
| 11 | Month | Categorical | Last contact month of year |
| 12 | Duration | Numeric | Last contact duration in seconds |
| 13 | Contacts Made | Numeric | number of contacts performed during this campaign |
| 14 | Days Passed | Numeric | number of days that passed by after the client was last contacted from a previous campaign (-1 means client was not previously contacted) |
| 15 | Previous | Numeric | number of contacts performed before this campaign |
| 16 | Poutcome | Categorical | outcome of the previous marketing campaign |
| 17 | Subscription | Binary (Categorical) | Yes/No. Has the client subscribed a term deposit? |
| 18 | Employee variation rate | Numeric | Employee variation rate |
| 19 | Consumer price index | Numeric | Consumer price index |
| 20 | Consumer confidence index | Numeric | Consumer confidence index |
| 21 | Credit score | Numeric | Credit score of customers |

## Data cleaning:

First, we analyzed what’s in the data and how it can be used effectively for our analysis. We used MS Excel as a part of data cleaning process.

Important tasks include:

* Filling in missing values
* Conforming data to a standardized pattern

Improving data quality: We used SAS enterprise guide as a part of our exploratory data analysis to find out if there are any inconsistency issues so that the appropriate transformations can be applied.

Transforming and enriching the data:

Here we updated the format or value entries in order to reach a well-defined outcome, or to make the data more easily understood by a wider audience.  Enriching data refers to adding and connecting data with other related information to provide deeper insights.

We have not transformed the data for using it in the software.

Modifications made to dataset:

* Education was modified as below:

Illiterates as “Primary”

Basic4y, Basic6y, Basic 9y, High School as “Secondary”

Professional, University degree as “Tertiary”

* Missing values in Balance column are substituted with mean.
* Missing values in contact type column are substituted with mode.
* The Month column is modified as below:

1🡪Jan

2🡪 Feb

3🡪Mar

And so on.

* Renamed the column names for below:

Pdays 🡪 Days Passed

Emp.var.rate 🡪 Employee variation rate

cons.price.indx 🡪 Consumer price index

cons.conf.idx 🡪 Consumer confidence index

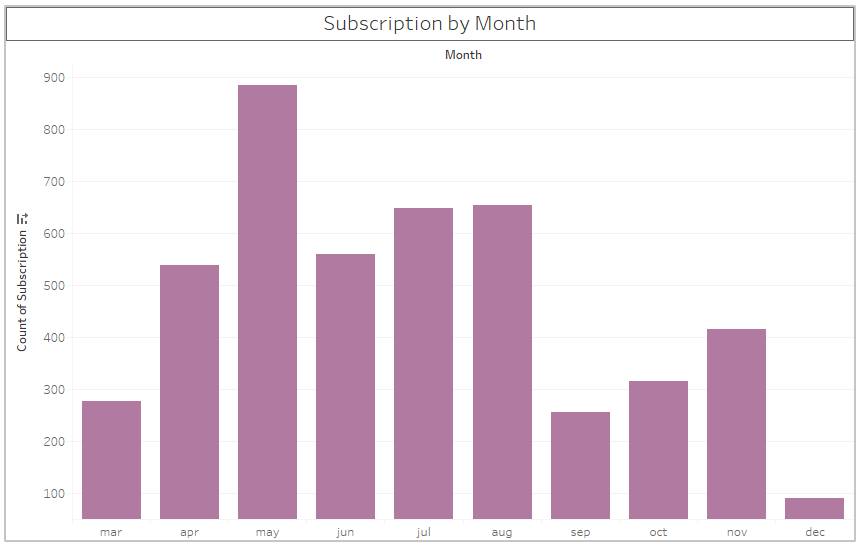
campaign 🡪Contacts made

* Default, Housing loan, personal loan, subscription columns are modified as:

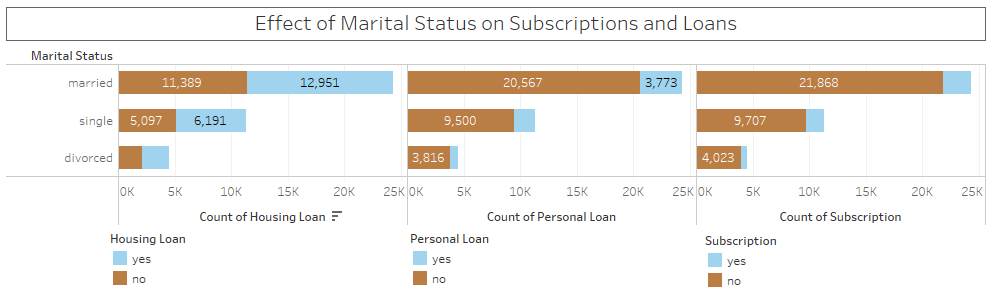
Y🡪 “Yes”

N🡪 “No”

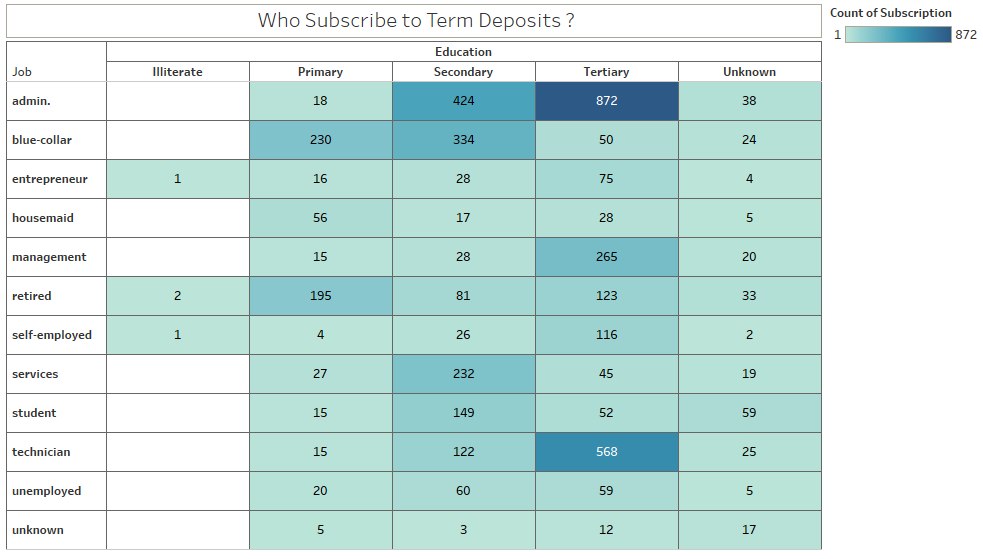
## Exploratory data analysis:



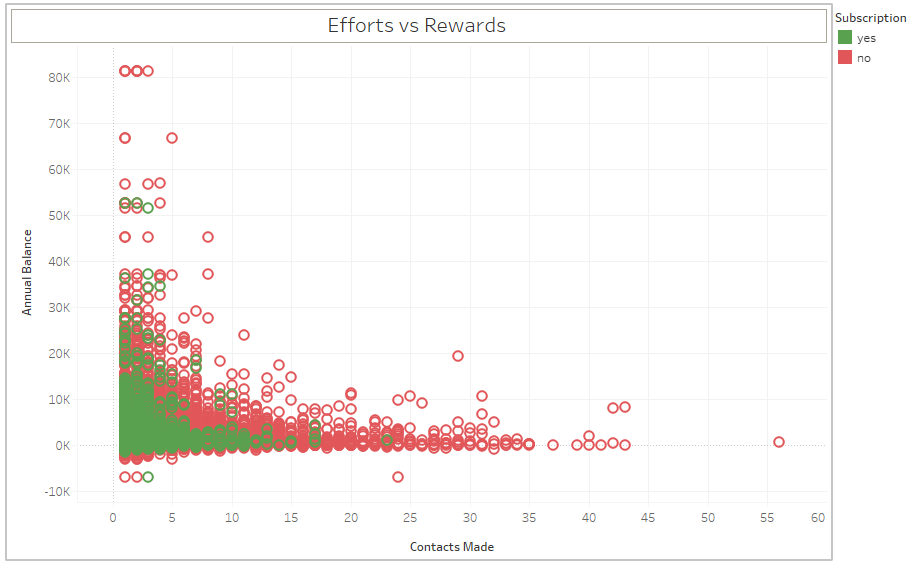
We are tracking subscriptions to term deposits by month and notice that the mid months, namely May, June, July and August see higher subscriptions than other months of the year. Our recommendation, based on this bar chart would be to increase telemarketing efforts in December.



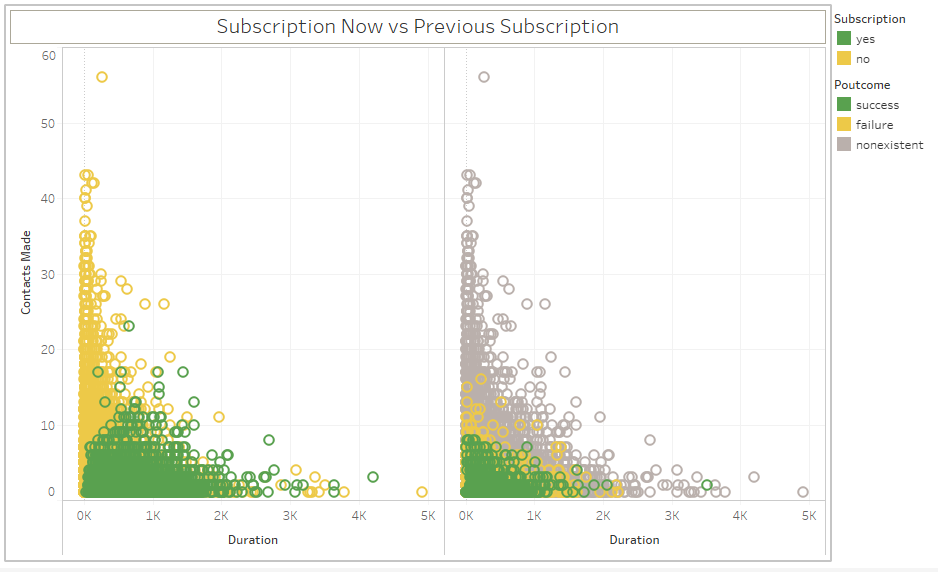
A higher percentage of customers of Vertex Bank Corp. are married and have housing loans followed by personal loan and term deposits. Customers who are divorced are potentials for applying for loans of for subscribing for term deposits.



Exploring the demographics that subscribe to term deposits, we found that people with administrative jobs and tertiary level education subscribe.



The scatter plot shows that a higher number of calls have been made to customers with lowers of negative annual account balance. The calls have not targeted customer that have a higher balance and that might be one of the reasons for customers not subscribing to term deposits. We recommend the bank to contact customers with higher annual balance.



The bank has gained a lot of new customers which were not present in the previous marketing campaign. The more number of calls made to a customer reduces chances of him/her subscribing to term deposits. The greater the duration between two calls also reduces the chances of customer subscribing to term deposits.