Interim Project Report on

Bank Marketing (Long term deposit subscription)

Submitted by

Group No. 4 [Batch: Sep 2019, Location: Chennai]

Group Members

1. Balaaje V S - WFPBXFVJ7A
2. Govindha Raju - ZWTCM7P0RN
3. Hari krishna B - G55O0WR054
4. Rajasekaran R - B6CJWAGHCM
5. Santhosh G S – F1MTKDTJGM
6. Vignesh B – F4CU6UC8TX

Research Supervisor

Mr.Subramanian P V

Great Lakes Institute of Management

greatlearning

Problem Statement:

A Portuguese bank wants their customers to subscribe to their new term deposits. A marketing campaign is launched, and customers are contacted through phone calls. Results for these previous campaigns have been provided to the manager to use the same in making the next marketing campaign more effective.

Challenges of Manager:

- Customer complaints about irrelevant product calls.
- No prior framework to choose which customer to target.

Objectives:

- How to predict whether a customer will subscribe to a term deposit or not?
- Which determinants would indicate a customer is ready to subscribe to a term deposit through direct marketing?
- How to segment term deposit market?
- Are there any common features of customers who have subscribed to a term deposit?

Data Set Source:

http://archive.ics.uci.edu/ml/datasets/Bank+Marketing

Dataset attributes information:

bank client data:

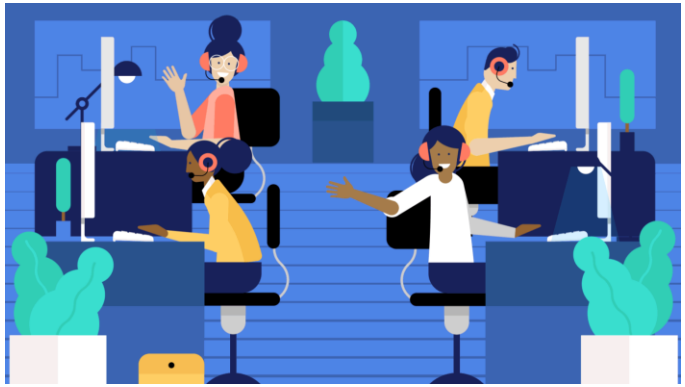
1 - age: age of the person (*numeric*)
2 - job : type of job (categorical: "admin.","blue-collar","entrepreneur","housemaid","management","retired", "self-employed","services","student","technician","unemployed","unknown")
3 - marital : marital status (categorical: "divorced", "married", "single", "unknown“ ; note: "divorced" means divorced or widowed)
4 - education (categorical: "basic.4y","basic.6y","basic.9y","high.school","illiterate","professional.course","university.degree","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")

Related with the last contact of the current campaign:

8 - contact: contact communication type (categorical: "cellular", "telephone")
9 - month: last contact month of year (categorical: “Jan", “Feb", "mar", ..., “Nov", “Dec")
10 - day_of_week: last contact day of the week (categorical: “Mon",“Tue",“Wed",“Thu",“Fri")
11 - duration: last contact duration, in seconds (numeric).

other attributes:

12 - campaign: number of contacts performed during this campaign and for this client (numeric, includes last contact)
13 - 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)
14 - previous: number of contacts performed before this campaign and for this client (numeric)
15 - poutcome: outcome of the previous marketing campaign (categorical: "failure","nonexistent","success")



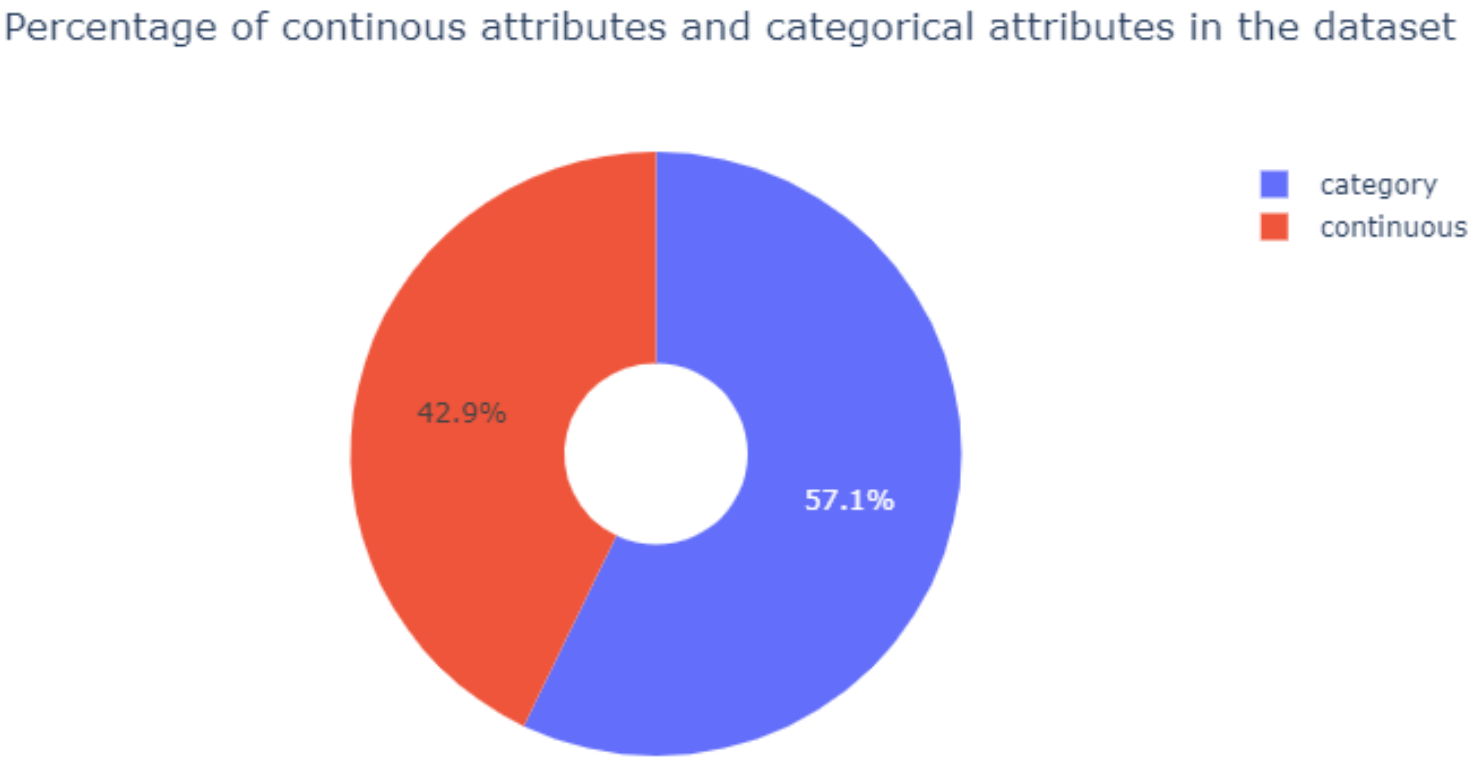
Social and economic context attributes:

16 - emp.var.rate: employment variation rate - quarterly indicator (numeric)
17 - cons.price.idx: consumer price index - monthly indicator (numeric)
18 - cons.conf.idx: consumer confidence index - monthly indicator (numeric)
19 - euribor3m: euribor 3 month rate - daily indicator (numeric)
20 - nr.employed: number of employees - quarterly indicator (numeric)

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

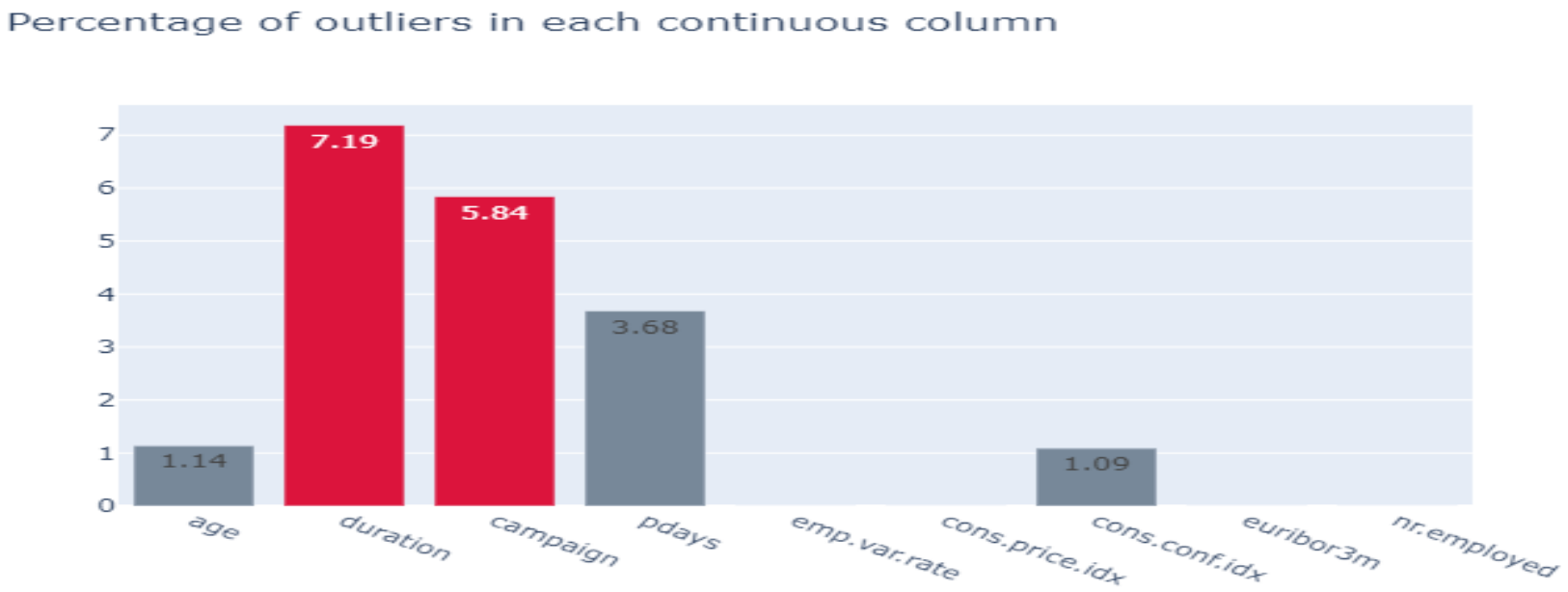
Data Set Shape:

Rows	Columns
41,188	21

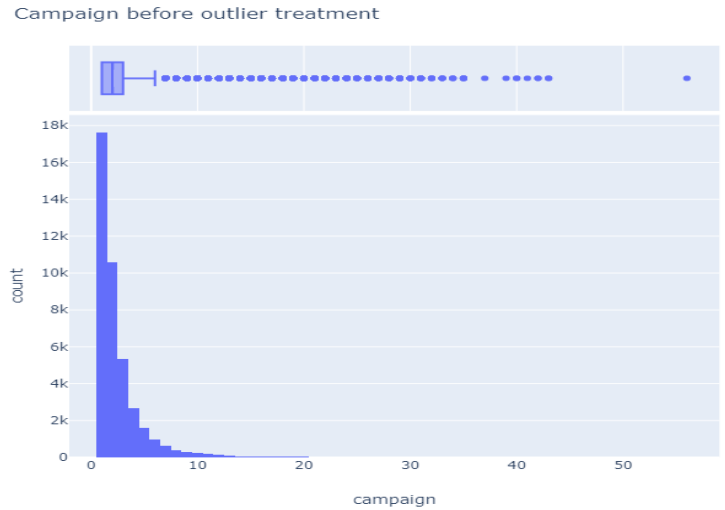


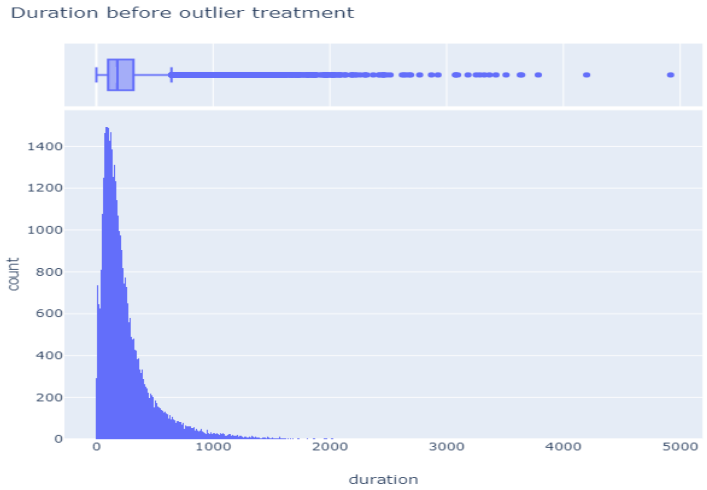
Exploratory Data Analysis

greatlearning

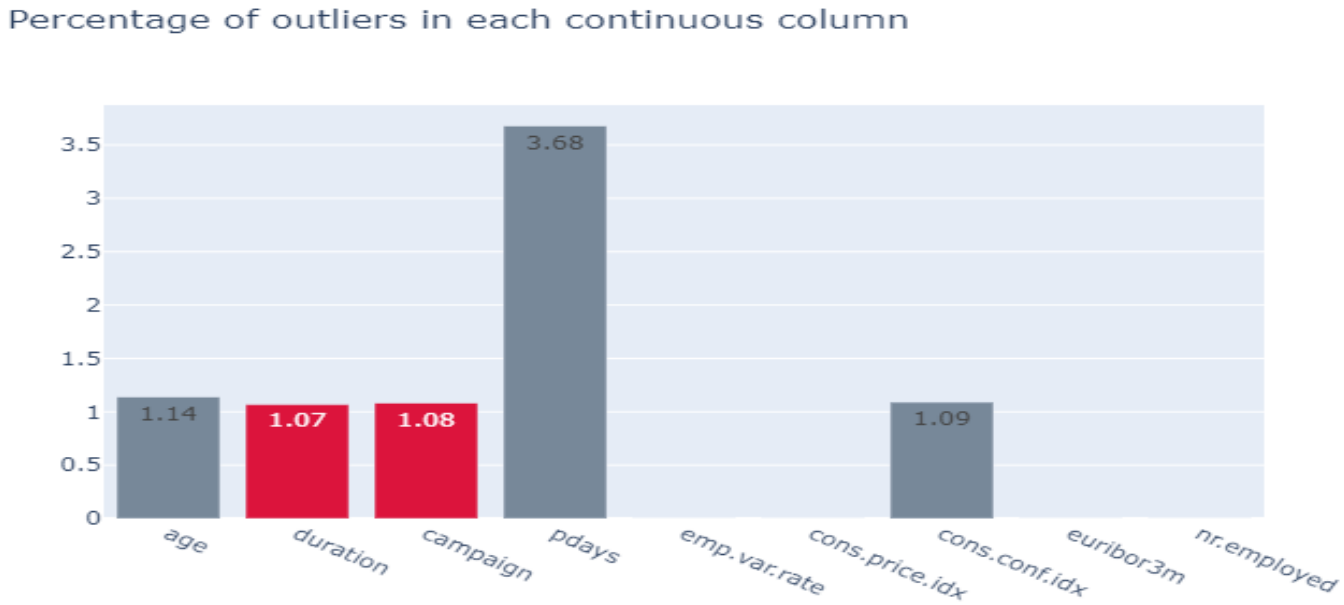


Duration and campaign column before Treatment:

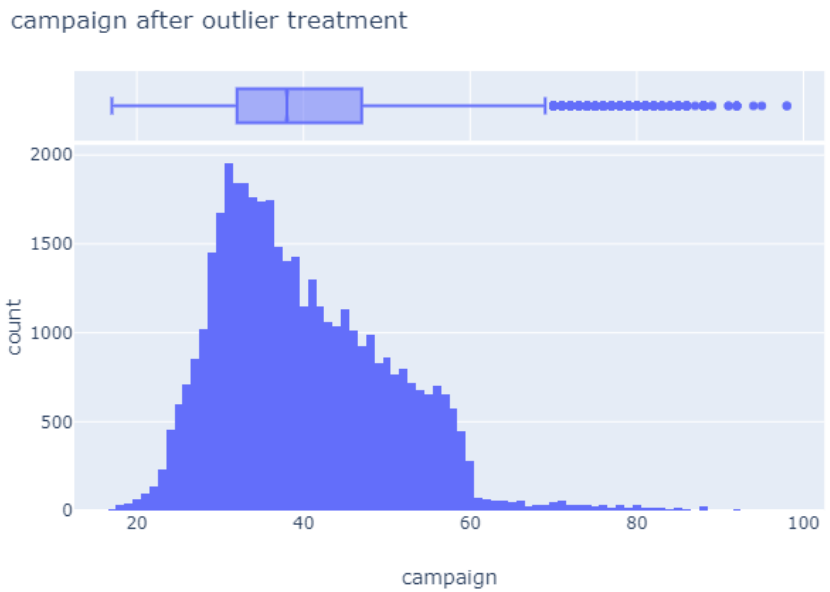


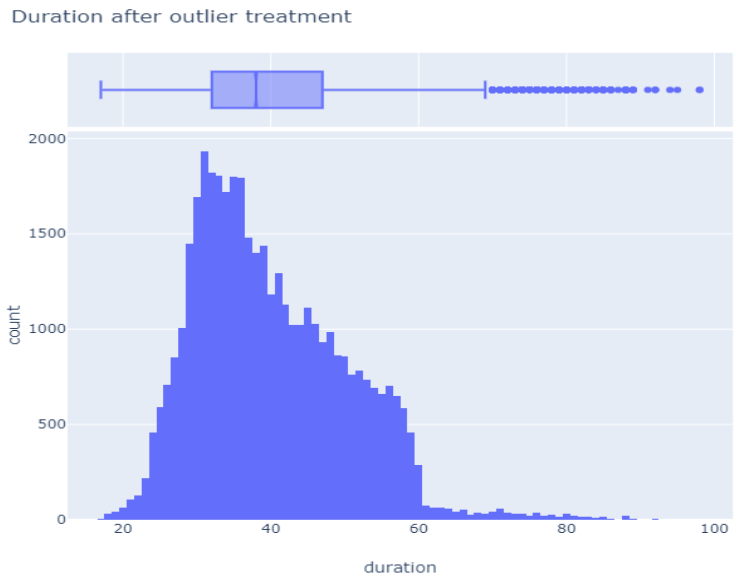


After Outlier Treatment:

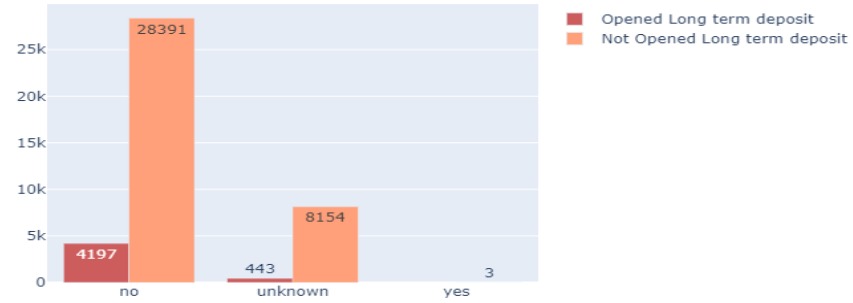


Duration and campaign column After outlier Treatment:



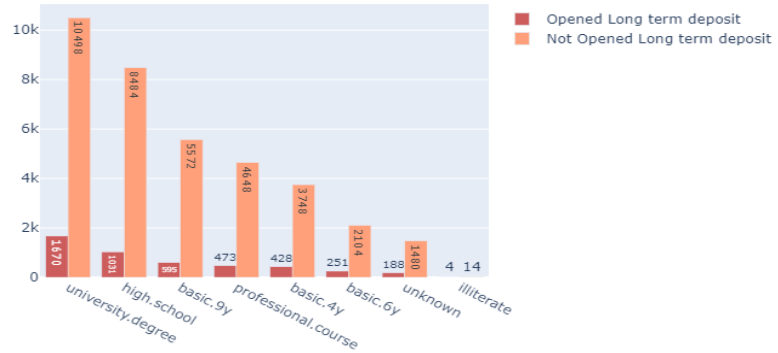


Default Status details on Longterm Subscription



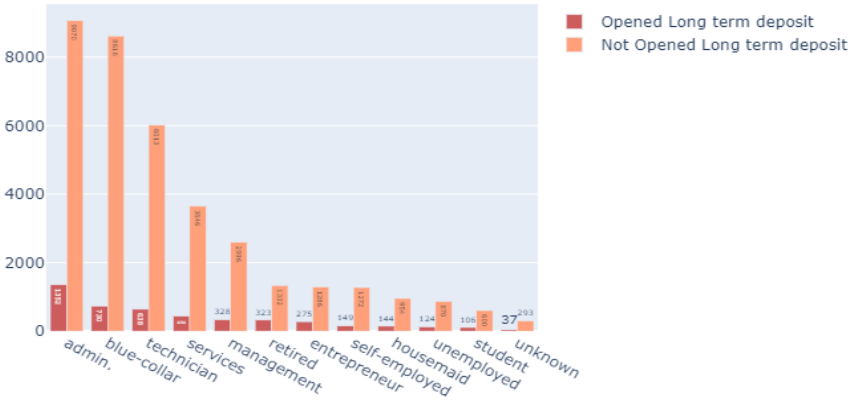
- Hence, lesser the default it is good for proceeding long-term Subscription process.

Education details on Longterm Subscription

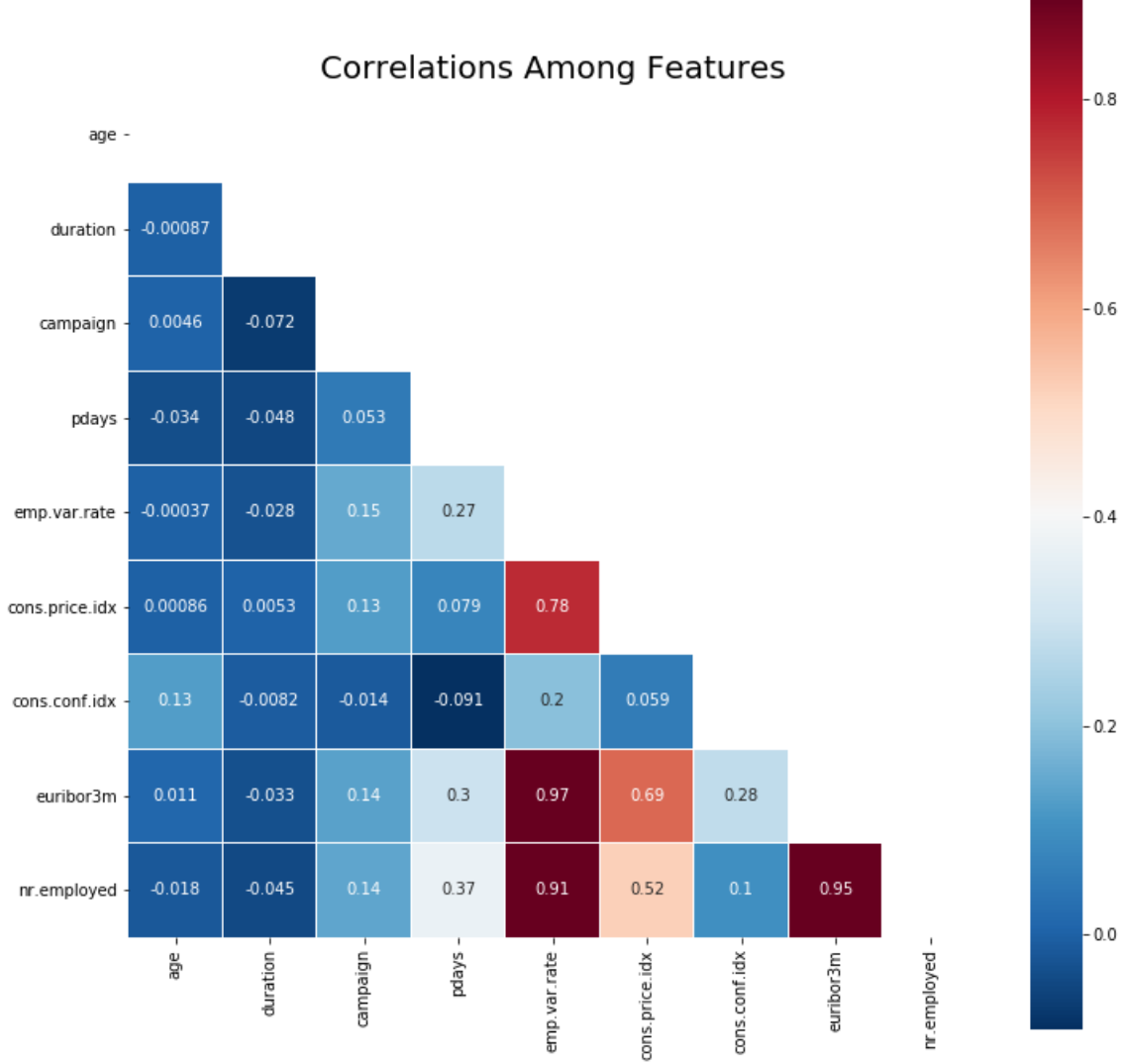


- Hence, education is one of the key factors to decide whether the customer will be able to pay the term subscription more the university degree is good for Long-term Subscription progress because university degree is more funded.

Job details on Longterm Subscription

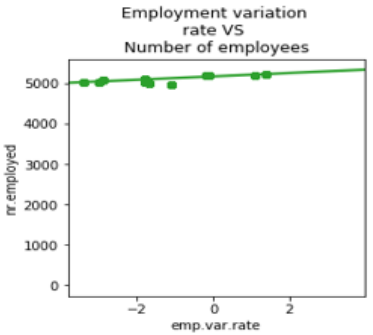


- The conversion Rate of employees is very less . we can see that only 14 people out of 100 called people are subscribing to the long term deposits.
- new strategy should be employed to improve the conversion rate . and they should target the retired and management people too .
- The probability of housemaid /unemployed/student subscribing to long term deposits is very low .hence new scheme should be targeted for these type of people.



- r = 0.97.

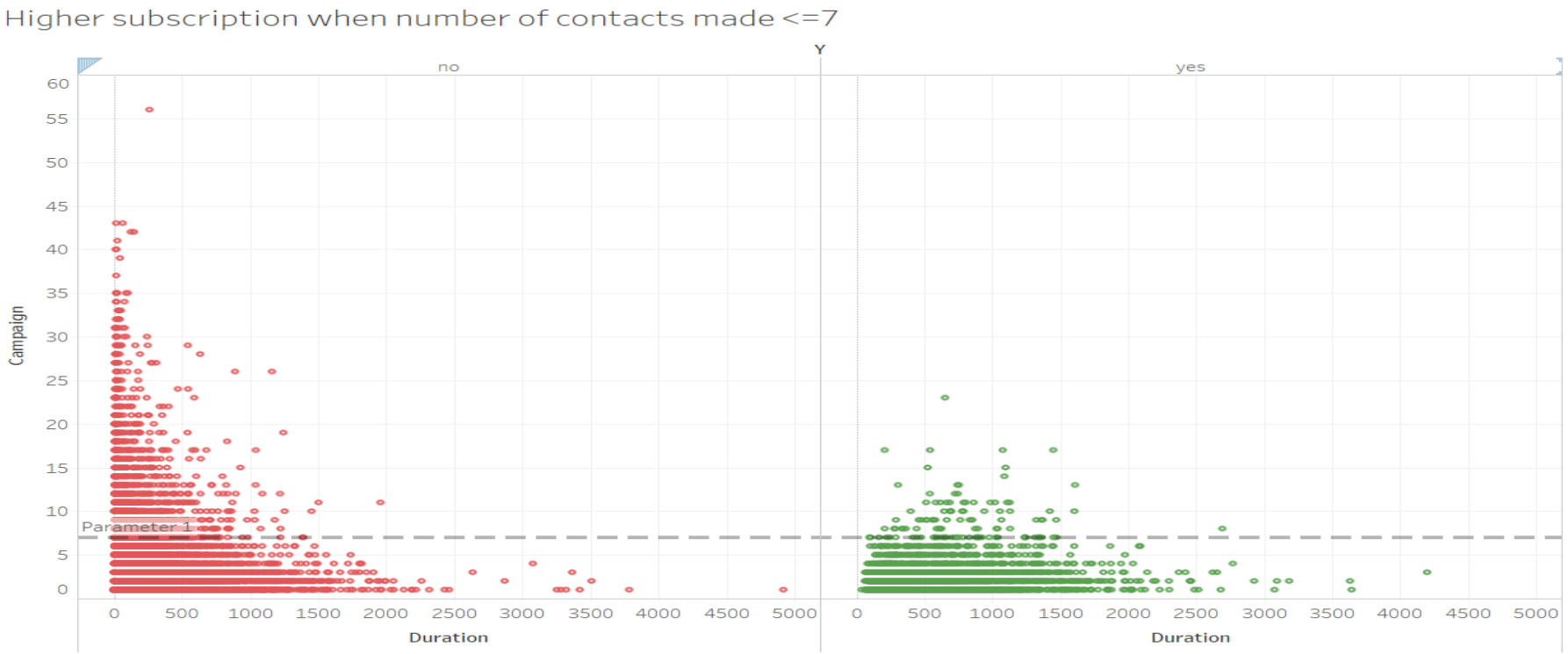
- r = 0.95.



r = 0.91

- EMPLOYEE VARIATION RATE,NUMBER OF EMPLOYEES,Euribor3m,cons.price.idx have high correlation among each other .

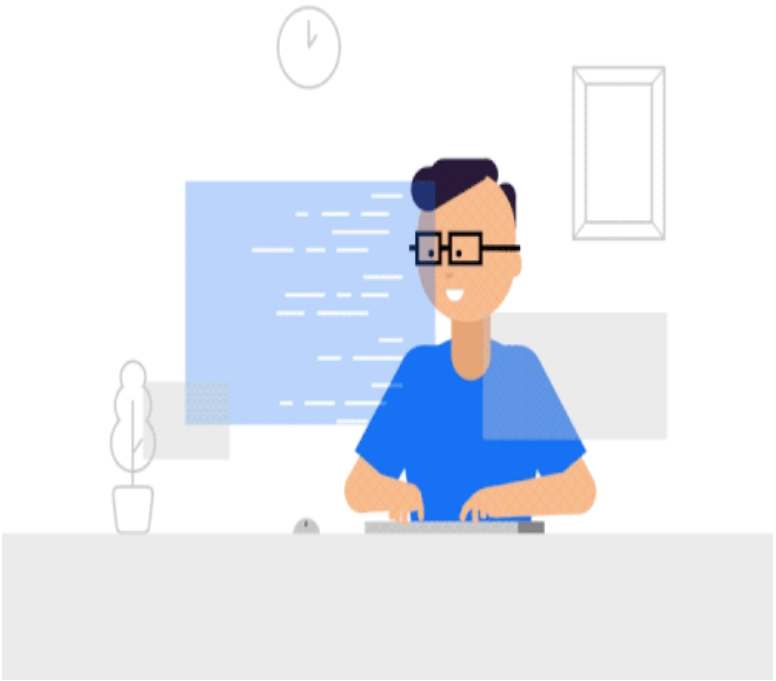
Subscription Rate based on contacts made



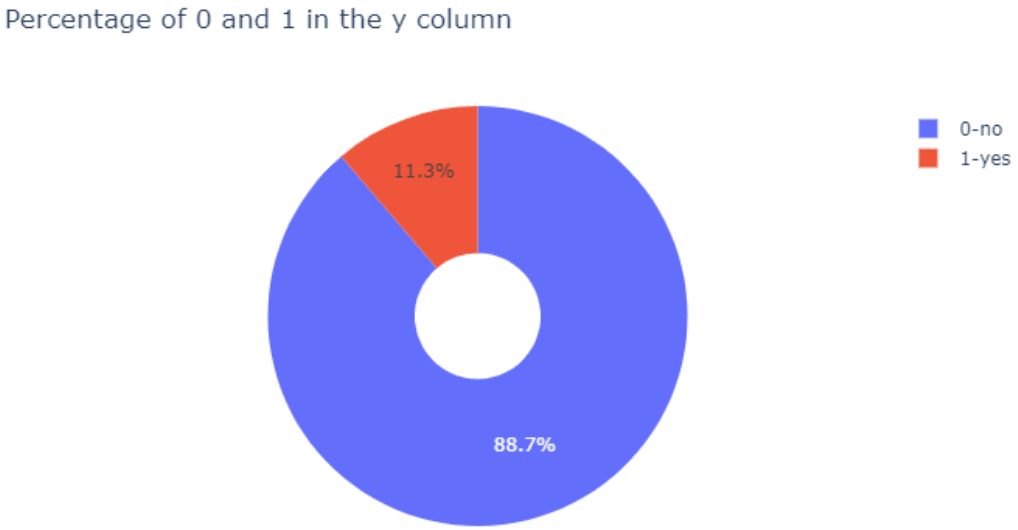
- If the targeted person receives a call for more than 7+ that subscription will be cancelled automatically due to customer patience level will be disturbed/testifies. The number of calls for communication will be the key factor for the probability of opening the long-term deposit by using the chance wisely.

Data Pre-processing

- Removed outliers Using IQR Method
- Imputed NULL values with Bfill method.
- Statistical significance Test performed .
- Scaled Data for Machine Learning Models.
- Checked for Dataset Balance(Target) Without Balancing the Target . computed the Machine Learning Models
- Dataset Balance done using Smote and computed the machine learning models
- Dataset Balance done using UP Sampling Technique and computed the machine learning models



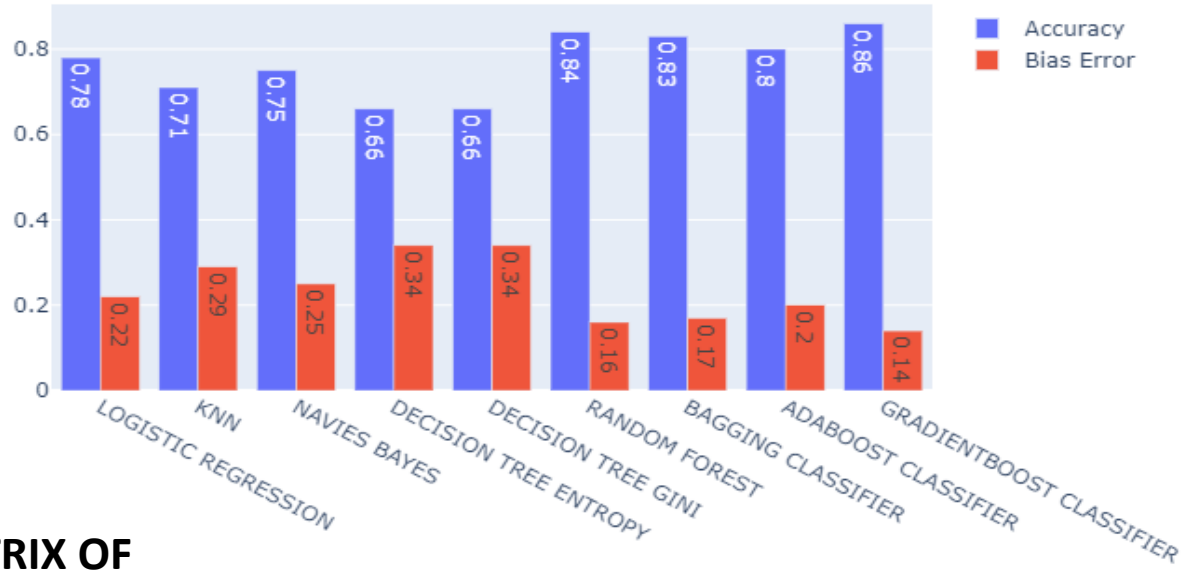
Before Balancing the Dataset :



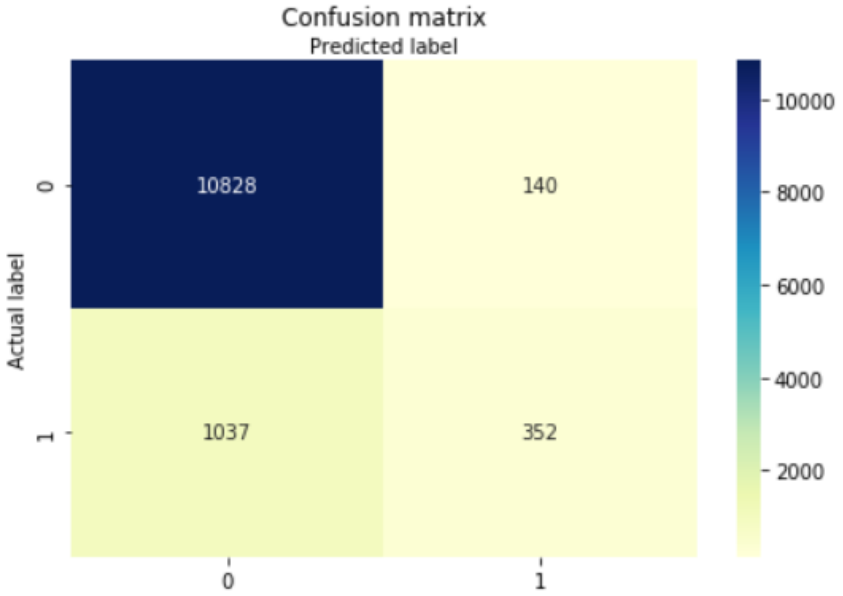
EVALUATION METRICS FROM TRAIN TEST SPLIT (70/30)

NAME	ACCURACY_TRAIN	ACCURACY_TEST	PRECISION	RECALL	ROC_AUC	F1-SCORE
LOGISTIC REGRESSION	0.9	0.9	0.67	0.2	0.62	0.31
KNN	0.91	0.89	0.56	0.25	0.62	0.34
NAVIES BAYES	0.48	0.49	0.16	0.83	0.62	0.27
DECISION TREE ENTROPY	1	0.87	0.42	0.42	0.62	0.42
DECISION TREE GINI	1	0.86	0.39	0.41	0.62	0.4
RANDOM FOREST	1	0.89	0.56	0.29	0.62	0.38
BAGGING CLASSIFIER	0.99	0.89	0.53	0.34	0.62	0.41
ADABOOST CLASSIFIER	0.9	0.9	0.69	0.2	0.62	0.32
GRADIENTBOOST CLASSIFIER	0.9	0.9	0.68	0.26	0.62	0.38

K FOLD CROSS VALIDATION (N=5) ACCURACY,BIAS AND VARIANCE



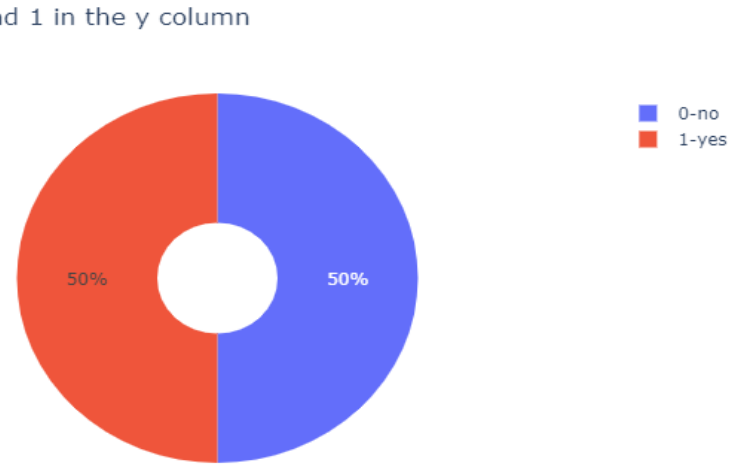
CONFUSION MATRIX OF GRADIENT BOOST CLASSIFIER



EVALUATION METRICS :

ACCURACY = 0.90
PRECISION = 0.71
RECALL = 0.25
F1 SCORE = 0.37
ROC_AUC = 0.62

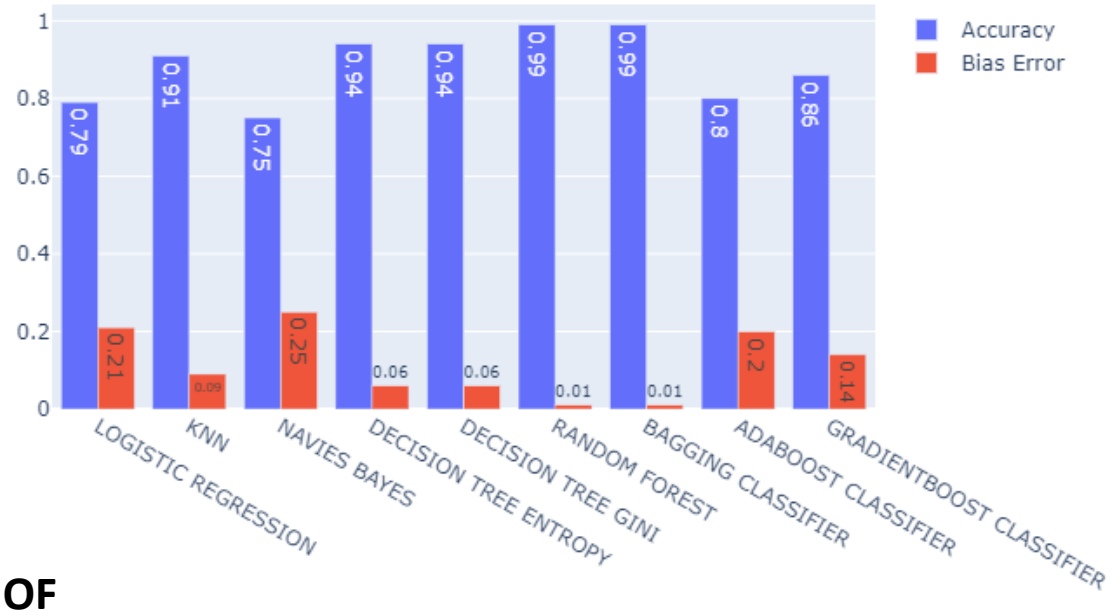
After Balancing the Dataset using Over sampling technique :



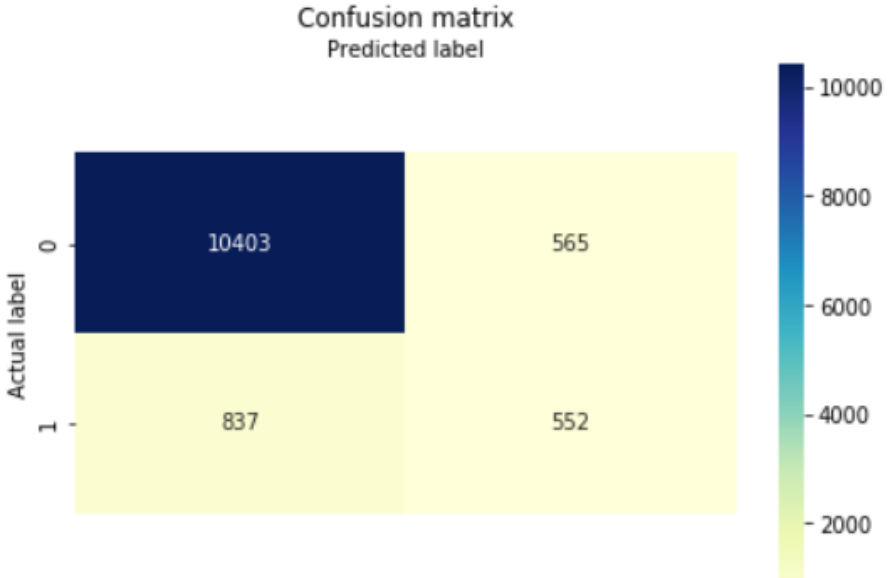
EVALUATION METRICS FROM TRAIN TEST SPLIT (70/30)

NAME	ACCURACY_TRAIN	ACCURACY_TEST	PRECISION	RECALL	ROC_AC	F1-SCORE
LOGISTIC REGRESSION	0.73	0.74	0.78	0.66	0.8	0.72
KNN	0.91	0.87	0.81	0.97	0.8	0.88
NAVIES BAYES	0.58	0.59	0.55	0.91	0.8	0.69
DECISION TREE ENTROPY	1	0.95	0.92	0.99	0.8	0.96
DECISION TREE GINI	1	0.95	0.91	0.99	0.8	0.95
RANDOM FOREST	1	0.97	0.94	0.99	0.8	0.97
BAGGING CLASSIFIER	1	0.96	0.93	0.99	0.8	0.96
ADABOOST CLASSIFIER	0.74	0.74	0.82	0.62	0.8	0.7
GRADIENTBOOST CLASSIFIER	0.8	0.8	0.85	0.72	0.8	0.78

K FOLD CROSS VALIDATION (N=5)ACCURACY,BIAS AND VARIANCE



CONFUSION MATRIX OF RANDOM FOREST CLASSIFIER



EVALUATION METRICS :

ACCURACY = 0.88
PRECISION = 0.49
RECALL = 0.39
F1 SCORE = 0.44
ROC_AUC = 0.67

Project outcome:

- We have decided to use past data to automate this decision, instead of manually choosing through each and every customer. Previous campaign data which has been made available to us; contains customer characteristics, campaign characteristics, previous campaign information as well as whether customer ended up subscribing to the product as a result of that campaign or not.
- Using this we plan to develop a statistical model which given this information predicts whether customer in question will subscribe to the product or not. A successful model which is able to do this, will make our campaign efficiently targeted and less bothering to uninterested customers.

Business outcome:

- Lets Assume the current bank uses 100 people as Tele Marketing executives for Long term deposit
- Salary for each person costs 20,000 INR
- By using this model we can Target the potential customers and can reduce the Marketing Executives by Half
- So we can save up to 10Lakh by using our current model.

Thanks!