CUSTOMER Base for Personal Loan

- An Analysis at THERA BANK.

Pradeep Panicker 3/1/20 DATA MINING

TABLE OF CONTENTS

PROJECT OBJECTIVE

DATA INTERPRETATION -

A DATA-SET CONTAINING 'THERA Bank's CUSTOMER Profile ' IS COLLECTED WITH SPECIFIC INFORMATION ON 'PERSONAL LOAN'; 'CREDIT CARD USAGE'; 'DEPOSIT CERTIFICATES'; 'INTERNET BANKING'.

THE DATA IS COMPILED FOR FIVE-THOUSAND CUSTOMERS.

PROJECT OBJECTIVE –

- TO PERFORM EXPLORATORY DATA ANALYSIS.
- UNIVARIATE AND MULTI-VARIATE ANALYSIS.
- CREATE A CLASSIFICATION TREE BASED ON 'CART' TECHNIQUE, INTERPRETATIONS THEREOFF – for CUSTOMER CLASSIFICATION
- CREATE A CLASSIFICATION TREE BASED ON 'RANDOM FOREST' TECHNOQUE, INTERPRETATIONS THEREOFF – for CUSTOMER CLASSIFICATION.
- EVALUATE THE CONFUSSION MATRIX AND FURTHER INTERPRETATION.
- MODEL PERFORMANCE MATRIX WITH SELF-INTERPRETATION – TO TARGET INCREASE IN PERSONAL LOAN BASE.

WORKING ENVIORNMENT SETUP

I. SOFTWARE TOOL 'R STUDIO' IS ESTABLISHED .

II. WORKING DIRECTORY IS SET TO ACCESS WORKING FILE – 'Thera Bank_Personal_Loan_Modelling-dataset-1'.

   a. setwd (C:/Users/prade/OneDrive/Desktop/BABI/DATA MINING/PROJECT 1- THERA')

III. Working file imported using 'Import Dataset' option in Global Environment.

IV. The file is assigned as 'bank' in R Studio. Dimensions are verified -

   a. dim(bank) – gives a result of 5000 Elements with 14 variables have been successfully imported.

V. The Structure of Data is checked for class of each variables.

   a. ALL 14 VARIABLES are found to be in 'NUMERIC' observations
   b. 02 Variable – 'ID' & 'ZIP CODE', are found to be in-significant w.r.t. analysis as they have been identified as 'IDENDITY NUMBERS'. Hence these variables have been 'factored'
   c. str(bank) – revealed the presence of missing Values in the variable 'FAMILY MEMBERS'. These were represented as 'NA' – counted to 18 records. These indicated the absence of family members for the respective customers. Hence for avoiding error in analysis, these 'NA' were converted to 'Zeros'.

NOW THE DATA FILE - 'bank' - IS READY FOR DATA MANUPULATIONS AND STATISTICAL ANALYSIS.

UNIVARIATE ANALYSIS – SUMMARY.

SUMMARY of data –

The SUMMARY of Data is called for and verified for any missing values.

Command - [summary(bank)]

NOTE – FOR SUMMARY TABLE , PLEASE REFER TO ANNEXURE -I ; TABLE-I

The FIVE Basic Statistical elements of analysis namely – Minimum Value (Min.)of the observed 5000 Values, First Inter Quartile Range (1st Qu.), Median of the 5000 Observations (Median), Average / Mean of the 5000 Observations (Mean), Third Inter Quartile Range (3rd Qu.) and Maximum Value of the observed 5000 Values – for EACH OF THE TWELVE VARIABLES is listed.

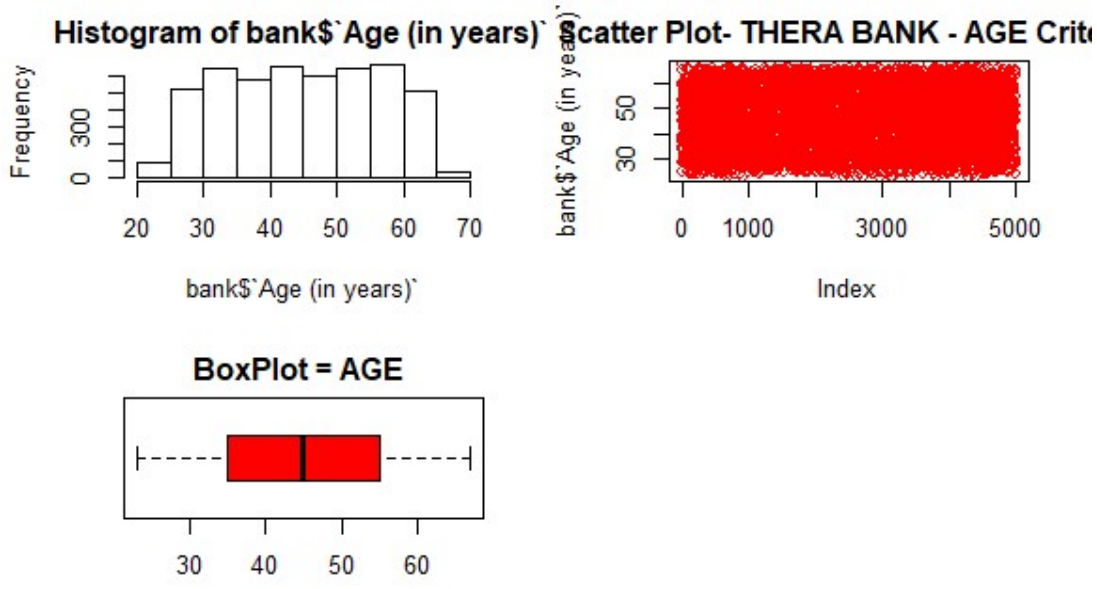
INSIGHTS -

I. **IDENTIFIED Missing value has been treated. All data points are complete in nature.**
II. **The vide GAP between the 3rd Quartile figure and the MAX. Value in Variables which account for monetary flow indicate clearly that OUTLIERS are present.**

UNIVARIATE ANALYSIS – 'bank' Data File.

CUSTOMER'S AGE in Years [AGE (in Years)] –

Graph -1



INFERENCES -

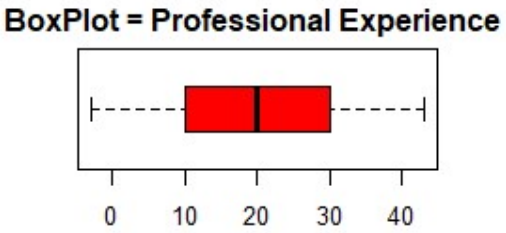
Data Distribution – Histogram [Graph I] shows data is evenly distributed. **Interesting to note that CUSTOMER RANGE falls in the YOUNG & ADULT Category, thereby a continuous income group is being targeted.**

Scatter Plot indicates NO LINEAR Relationship.

BOX PLOT Suggests no skewness in the data. NO OUTLIERS HAVE BEEN IDENTIFIED.

YEARS of Professional Experience [Experience] –

Graph-2



INFERENCES -

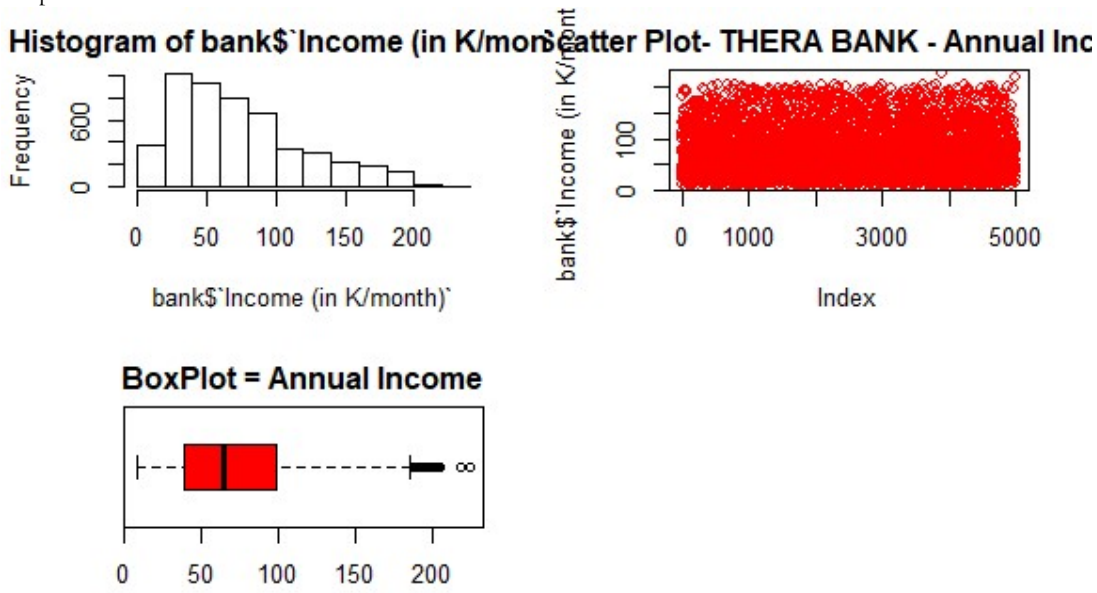
Data Distribution – Histogram [Graph 2] shows data is evenly distributed. **Do take a note that negative data is present – indicating CUSTOMERS with NO JOB for those many years prior to the date of collection of Data/ Profiling.**

Scatter Plot indicates NO LINEAR Relationship.

BOX PLOT Suggests no skewness in the data. NO OUTLIERS HAVE BEEN IDENTIFIED.

Annual Income of the Customer(in $000)[Income] –

Graph 3-



INFERENCES -

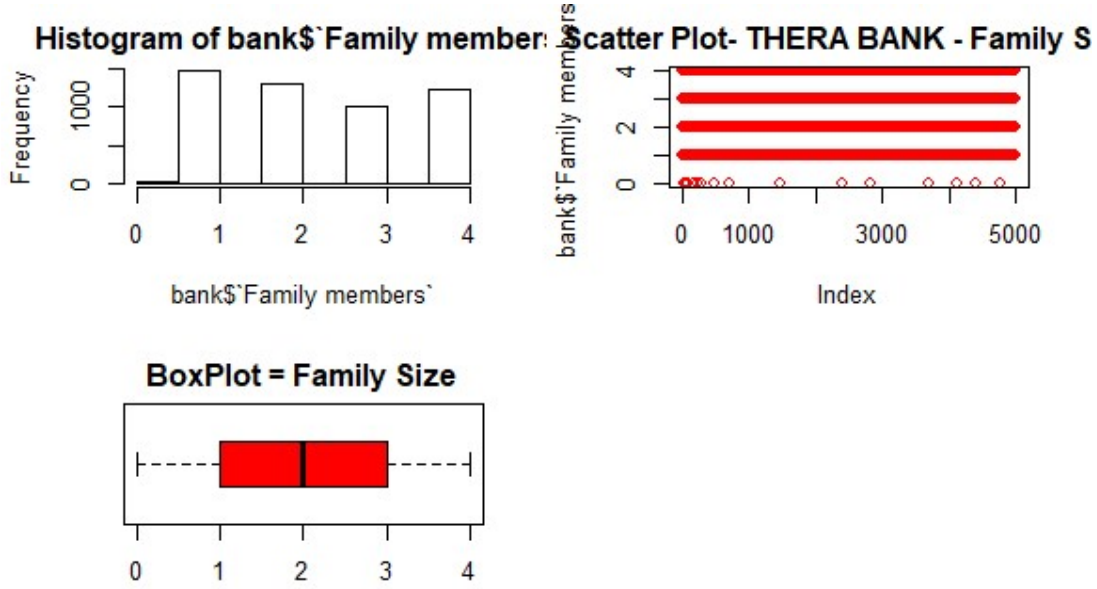
Data Distribution –Histogram[Graph 3] shows a RIGHT SKEWEDNESS of the data. Maximum Customers do fall in the range of Income $20,000 ~ $100,000.00

Scatter plot suggests high density of Customers present below the $100K range.

BOXPLOT shows RIGHT SKEWNESS of DATA**. Outliers are present , highly-significant at the 'Above the 1.5IQR + Q3 range'.**

FAMILY SIZE of the Customers[Family members] –

Graph 4 -



INFERENCES -

Data Distribution –Histogram[Graph 4] shows that data is QUITE DISCREET in NATURE. FIVE values have been assigned distinctively between the FIVE THOUSAND Customers.

0 – Indicates Customer with NIL Family Members.(classified as NA in the original data)

1 – Indicates Customer with ONE Family Members.

2 – Indicates Customer with TWO Family Members.

3 – Indicates Customer with THREE Family Members.

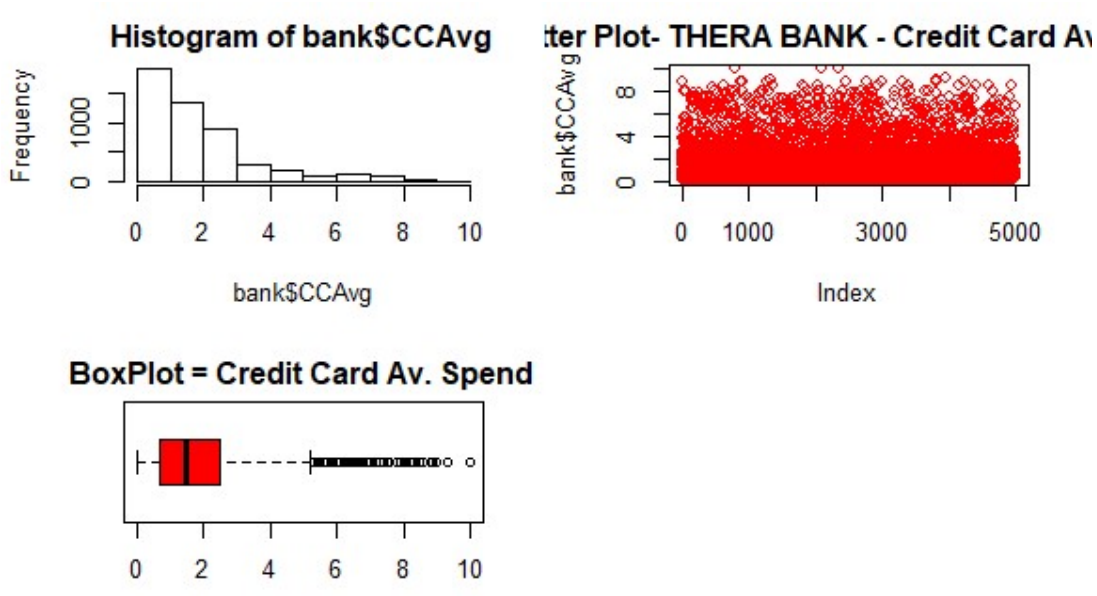
4 – Indicates Customer with FOUR Family Members.

Scatter Plot does not show any distribution pattern since the data being DISCREET in nature.

BOXPLOT indicates – NO OUTLIERS Present.

CREDIT CARD – Average Spending[CCAvg] –

Graph 5 -



INFERENCES -

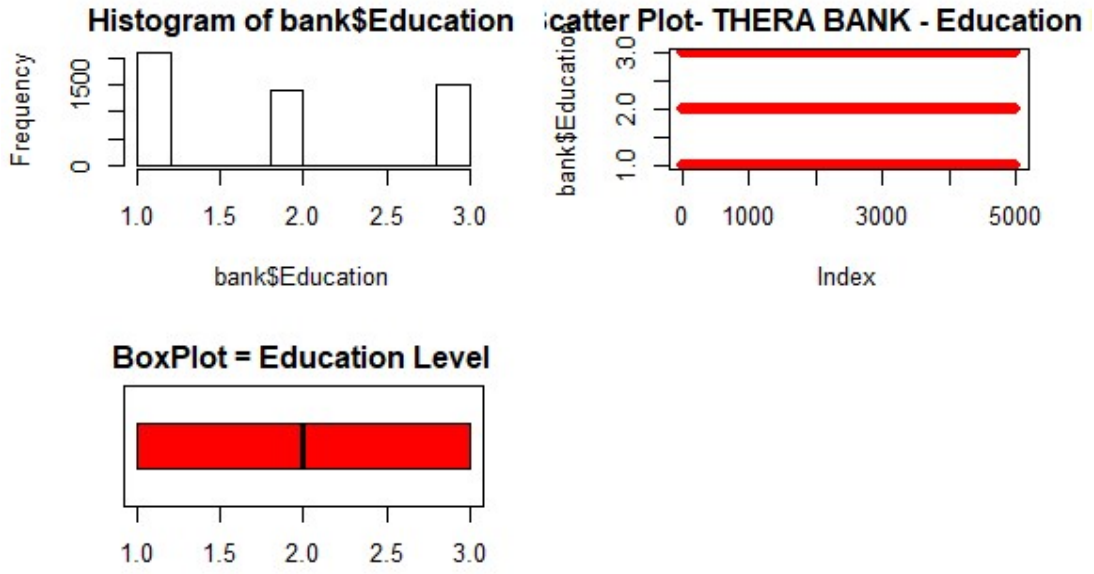
Data Distribution –Histogram[Graph 5] shows that data is completely RIGHT SKEWED. Maximum Customers do fall in the range of AVERAGE SPEND PER MONTH by CREDIT CARDS as $500 ~ $3,000.00

Scatter plot suggests high density of Customers present below the $3000 range.

BOXPLOT shows RIGHT SKEWNESS of DATA**. Outliers are present , highly-significant at the 'Above the 1.5IQR + Q3 range'.**

EDUCATION LEVEL [Education] –

Graph 6 –



INFERENCES -

Data Distribution –Histogram[Graph 6] shows that data is QUITE DISCREET in NATURE. THREE Observed values have been assigned distinctively between the FIVE THOUSAND Customers.

1.0 – Indicates Level 1 of education classified as 'Undergraduate'.

2.0 - Indicates Level 2 of education classified as 'Graduate'.

3.0 - Indicates Level 3 of education classified as 'Advanced/Professional'.

Scatter Plot does not show any distribution pattern since the data being DISCREET in nature.

BOXPLOT indicates – NO OUTLIERS Present.

MORTGAGE Taken from Bank [Mortgage] –

Graph 7 –

INFERENCES -

Data Distribution –Histogram[Graph 7] shows that data is completely RIGHT SKEWED. Maximum Customers have taken MORTGAGE below $150,000.00

Scatter plot suggests high density of Customers present below the $150,000 range.

BOXPLOT shows RIGHT SKEWNESS of DATA**. Outliers are present , highly-significant at the 'Above the 1.5IQR + Q3 range'. Indicates certain customers have availed HUGE MORTGAGE.**

PERSONAL Loan: Securities Account, CD Account, Online Banking & Credit Card Availed from Thera Bank – A consolidated View

Graph 8

INFERENCES -

All the above listed variables having Binary Values, the above plotting depicts clearly the trend of the customers.

Personal Loan - MAXIMUM Customers have NOT-AVAILED the Personal Loan, during the campaign held by the bank.

Securities Account held with the Bank - - MAXIMUM Customers ARE NOT HOLDING Securities at the bank.

Certificate of Deposit – Only Three Hundred & TWO customers of the FIVE Thousand data collected DO HOLD Certificate of Deposit at the Bank.

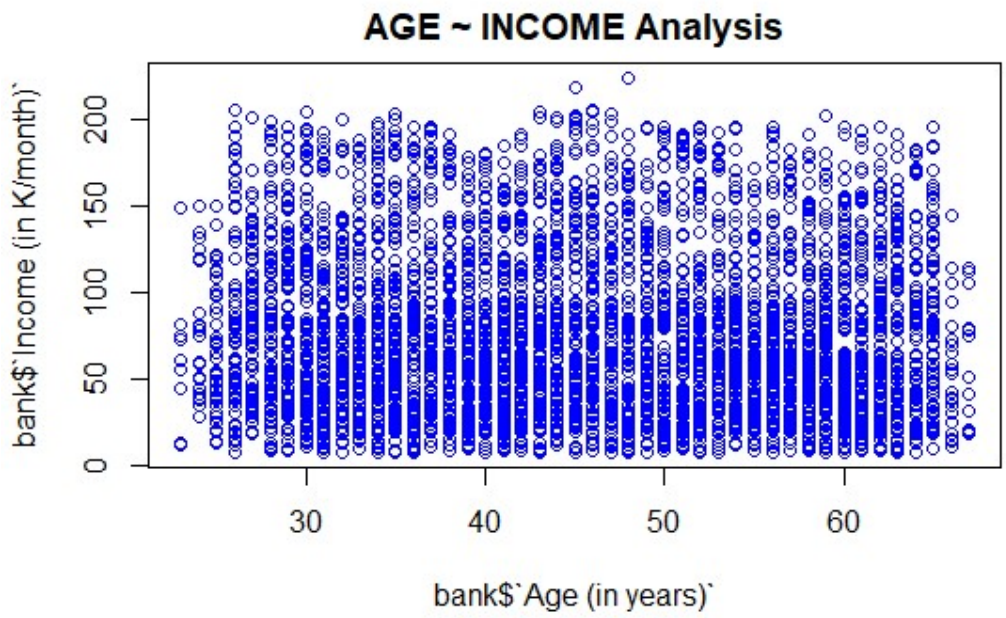
ONLINE Banking – Data shows a MAJORITY of the CUSTOMERS utilize the Online Banking facility, provided by the bank.

THERA Bank CREDIT Card – Majority of the customers scanned, DO NOT USE the credit card issued by the bank.

MULTIVARIATE ANALYSIS – 'bank' Data File.

AGE Vs. INCOME Analysis ➔

Graph 9



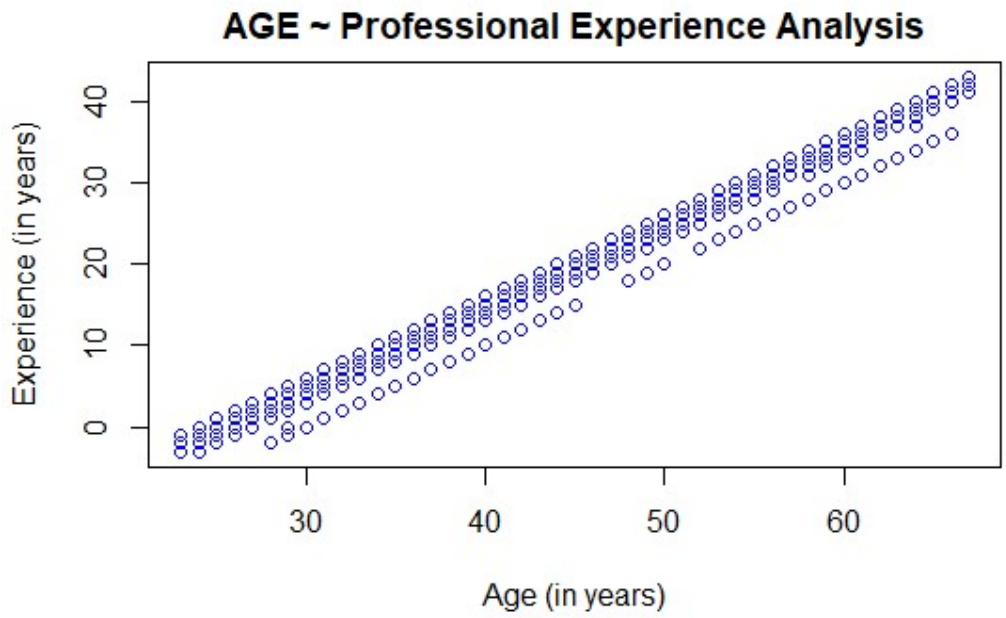
INFERENCES -

Plotting as above shows NO correlation between AGE & INCOME of customers.

The correlation factor analysis reveals – 5.53%[negative] which indicates almost NO CORRELATION.

AGE Vs. Professional Experience Analysis ➔

Graph 10



INFERENCES -

A VERY POSITIVE CORRELATION EXISTS BETWEEN THE AGE AND THE PROFESSIONAL EXPERIENCE OF THE CUSTOMERS – INDICATING THAT AS AGE INCREASES => EMPLOYEMENT OPPURTUNITY INCREASES AND THEREBY PROFESSIONAL EXPERIENCE INCREASES.

THE CORRELATION FACTOR, WHEN ANALYSED, STAND AT 99.42 % - ABSOLUTLY POSITIVE INDICATING PERFECT CORRELATION.

INCOME Vs. Professional Experience Analysis ➔

Graph 11

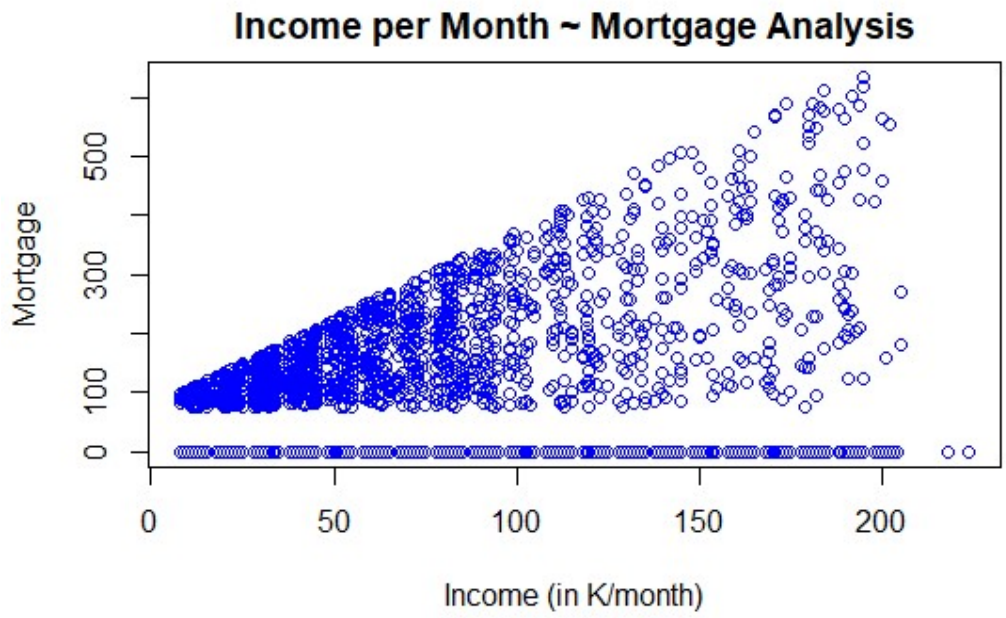
INFERENCES -

Plotting as above shows NO correlation between Professional Experience & INCOME of customers.

The correlation factor analysis reveals – 4.66%[negative] which indicates almost NO CORRELATION.

INCOME per Month Vs. MORTGAGE Analysis ➔

Graph 12



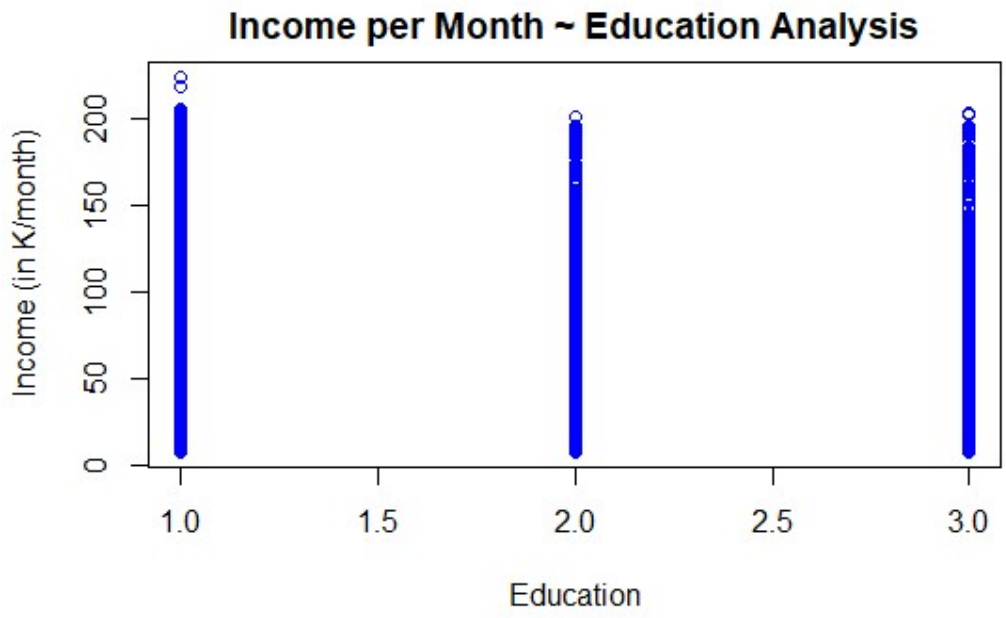
INFERENCES -

Plotting indicates NO STRONG correlation between Income of a Customer & Mortgage the customers avail.
MORTGAGE remains a facility which is taken as required. The plot shows High Income group customers availing LOW Mortgage.

The correlation factor analysis reveals 20.68% which indicates slight positive CORRELATION.

INCOME per Month Vs. Education Analysis ➔

Graph 13



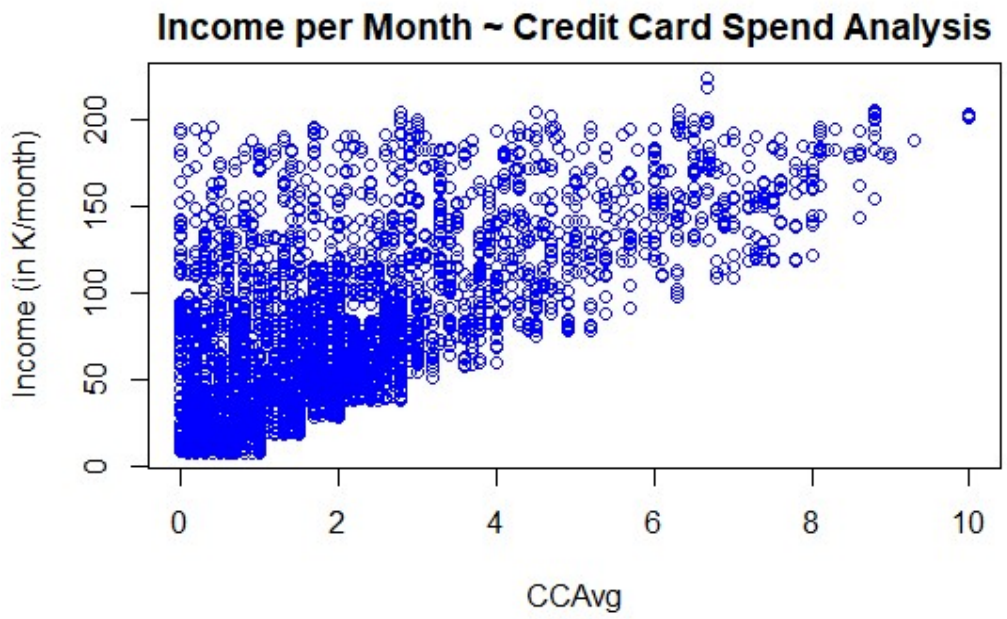
INFERENCES -

Above Plotting indicates that Education at all levels provide INCOME of similar nature. NO positive correlation is derived to say that Higher education yields Higher Income.

The correlation factor analysis reveals -18.75%[negative] which indicates a negative correlation to the effect that HIGHER INCOME exists with professions in LOWER Education grades.

INCOME Vs. Credit Card Spend Analysis ➔

Graph 13



INFERENCES -

Above Plotting assures that the CREDIT CARD Spend has a positive Correlation to the INCOME . This means that customers with High Income tend to spend more with Credit Cards.
This trend is good since it reveals that a control of issuance of Credit card to the right income group is prevailing.

The correlation factor analysis reveals 64.60% which indicates a high positive correlation.

CART USAGE FOR 'THERA BANK- Customer data Analysis & Prediction'

Data Set Preparation –

Alternate data set is created for evaluation – 'bank'.

Column 'ID' & 'ZIP CODE' is removed since being non-significant w.r.t. the 'Personal Loan' factor.

The dataset is purified to omit the 'NAs' from the 'Family Members' variable.

All variables, other than target Variable & Boolean Variables , are SCALED – to normalize the data and increase accuracy.

Target Variable – 'Personal Loan' is converted to an Integer Variable and the rest FOUR as Factors Variables.

Structure of the dataset is checked.

SEED is set and data set is split as TRAIN Dataset with 70% of 'bank5'.

Proportion of split is checked for the target variable – to check accuracy of SPLIT function.

Accuracy found abnormal. Hence splitting is performed after changing SEED NUMBER.

Third iteration gave most accurate split, as below –

```
> prop.table(table(bank5$Personal.Loan))
        0         1
0.9040546 0.0959454
> prop.table(table(banktrain$Personal.Loan))
        0         1
0.9012643 0.0987357
> prop.table(table(banktest$Personal.Loan))
         0          1
0.90963855 0.09036145
```

A CART MODEL is generated – 'model', with following rules ➔

Having 3322 data in the train dataset, the model is generated with pre-pruning parameters as –

Minsplit – 320(1/10th of observations); Minbucket -110(1/3rd of Minsplit); cp = ZERO, xval =10.

A Classification tree corresponding to 'model' is generated – as 'tree'.

'tree' is appended as separate 'pdf' attachement to this report.

Pruning of 'tree' –

Since the 'tree' was generated with full split – means a full-grown tree was made, we now need to check for OVERFITTING / UNDERFITTING.

For this certain pruning parameters such as 'Complexity Parameter(cp)' and Cross-Validation Error(xerror) needs to be evaluated and applied to the full-grown tree 'tree'

A complexity parameter table is generated for the 'model' –

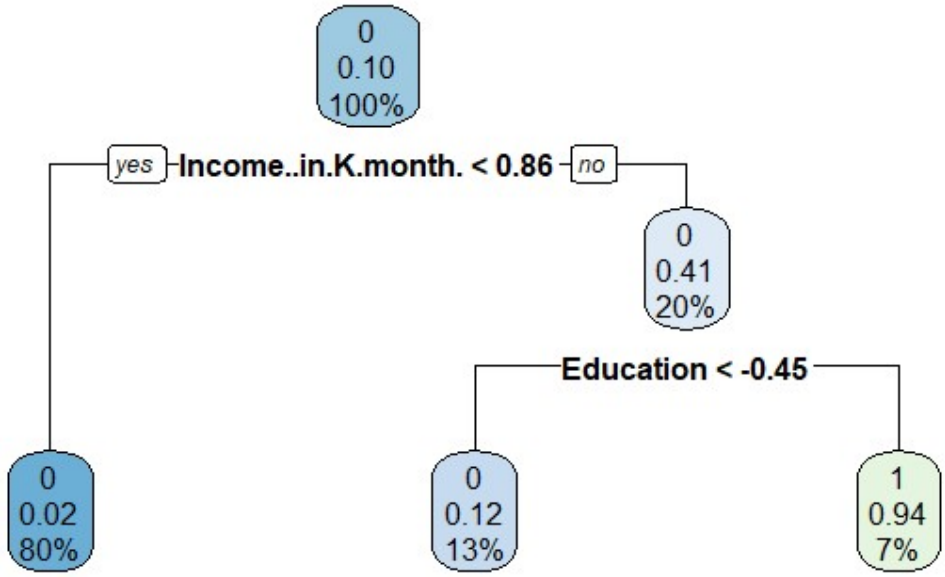
```
    CP nsplit rel error  xerror     xstd
1 0.31953      0   1.00000 1.00000 0.051659
2 0.01000      2   0.36095 0.36686 0.032348
```

Best 'cp' value is derived as 0.01

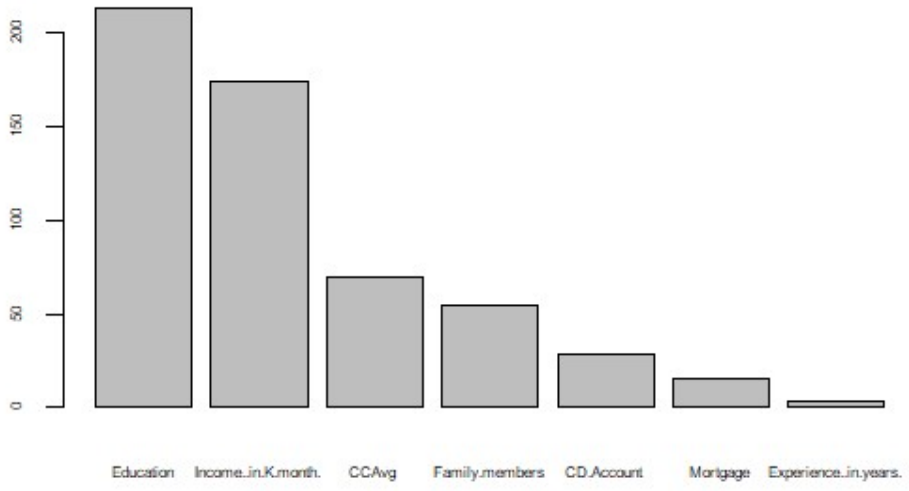
```
bestcp= model$cptable[which.min(model$cptable[,'xerror']), 'CP']
> bestcp
[1] 0.01
```

'tree' is now pruned, to avoid overfitting, applying the Best 'cp' value – 'prunedtree'

```
prunedtree= prune(tree= model,cp = bestcp)
```



Important Variables used for the above classification were –



Having the best classification of variables in hand, we now proceed to predict the values for target variable – both in the Train dataset & Test dataset.

```
banktrain$predict= predict(prunedtree,data= banktrain, type = 'class')
banktest$predict= predict(prunedtree,newdata= banktest, type= 'class')
```

The PREDICTED Values are now added to the dataset – 'banktrain' & 'banktest'

Accuracy of 'Predicted Values' ~ 'Actual Values' for the target variable 'Personal Loan' is evaluated, and tabled as below –

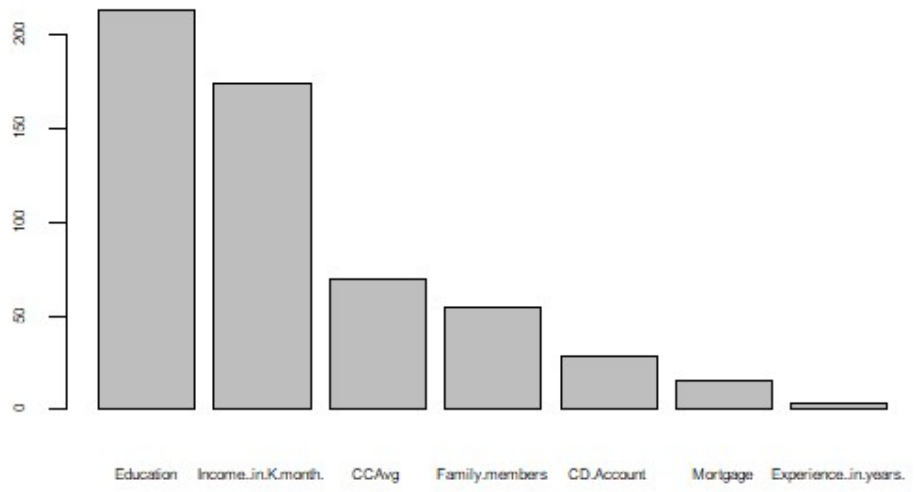
```
accuracytrain
[1] 0.9662854
> accuracytest
[1] 0.9650602
```

INFERENCES -

THE TEST DATASET ACCURACY IS ALMOST EQUIVALENT TO THE TRAIN DATASET ACCURACY – INDICATING THAT THE CART MODEL IS ROBUST.

Important Note –

An alternative MODEL was created by including 'ZIP CODE' as a variable.
Below is the VARIABLE IMPORTANCE PLOT – indicating that ZIP CODE IS NOT AN IMPORTANT VARIABLE.
The Accuracy of the Dataset was found to remain same as mentioned above -



'RANDOM FOREST' Technique FOR 'THERA BANK- Customer data Analysis & Prediction'

Alternate data set is created for evaluation – 'bank6'.
The data set is 'SCALED' for normalization of values, to give optimum accuracy in prediction.
Variable 'ID' & 'ZIP CODE' is nullified from the data set, since these are non-contributory variables.
'Personal Loan' variable is the Dependent Variable.
Structure of the dataset is checked.
SEED is set and data set is split as TRAIN Dataset with 70% of 'bank6'.
Proportion of split is checked for the target variable – to check accuracy of SPLIT function.
Accuracy found abnormal. Hence splitting is performed after changing SEED NUMBER.

```
prop.table(table(bank6$`Personal Loan`))
        0         1
0.9040546 0.0959454
> prop.table(table(bank6train$`Personal Loan`))
         0          1
0.90484794 0.09515206
> prop.table(table(bank6test$`Personal Loan`))
         0          1
0.90246839 0.09753161
```

A different seed is set for Random Forest tree evaluation.
Random forest Tree cluster is generated on the train dataset using the algorithm –
rforesttrain= randomForest(`Personal Loan`~.,data = bank6train, ntree= 501, mtry= floor(sqrt(ncol(bank6train))),nodesize= 10,importance= TRUE)

A full-blown tree growth was permitted by keeping 'ntree' @ 501, 501 Trees were allowed to be grown.

The print data of the RF Model reveals the parameters as –

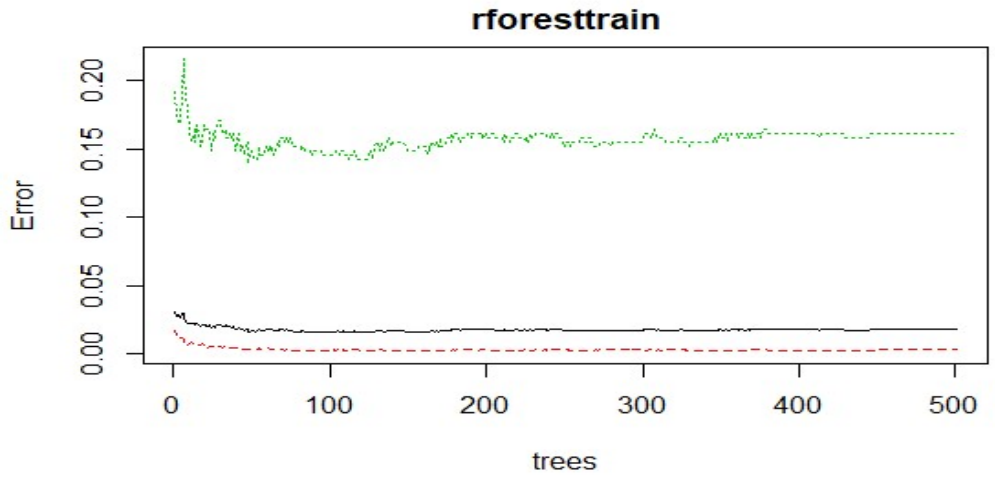
```
randomForest(formula = `Personal Loan` ~ ., data = bank6train,      ntree
= 501, mtry = floor(sqrt(ncol(bank6train))), nodesize = 10,
importance = TRUE)
               Type of random forest: classification
                     Number of trees: 501
No. of variables tried at each split: 3

        OOB estimate of  error rate: 1.78%
Confusion matrix:
     0   1 class.error
0 2997   8  0.00266223
1   51 265  0.16139241
```

The confusion Matrix shows that the ERROR factor in prediction is very less – 16%.
The 'Out Of Bag' error rate is also very low @ 1.78%.

Further, for TUNNING the Random Forest model, we will be analyzing the OPTIMUM TREES that need to be built by the Model as well as the Optimum VARIABLES that need to be included for analysis.

OPTIMUM NUMBER of TREES –

```
plot(rforesttrain)
```



Analysis of the above graph reveals that the 'OOB' Error rate has stabilized beyond 90 Trees. Thereby, we can prune the MODEL with 'ntree' = 91.

IMPORTANCE OF VARIABLES-

Contribution of each variable w.r.t. its Importance in Attaining Accuracy of the MODEL created – is evaluated as below –`importance(rforesttrain)`

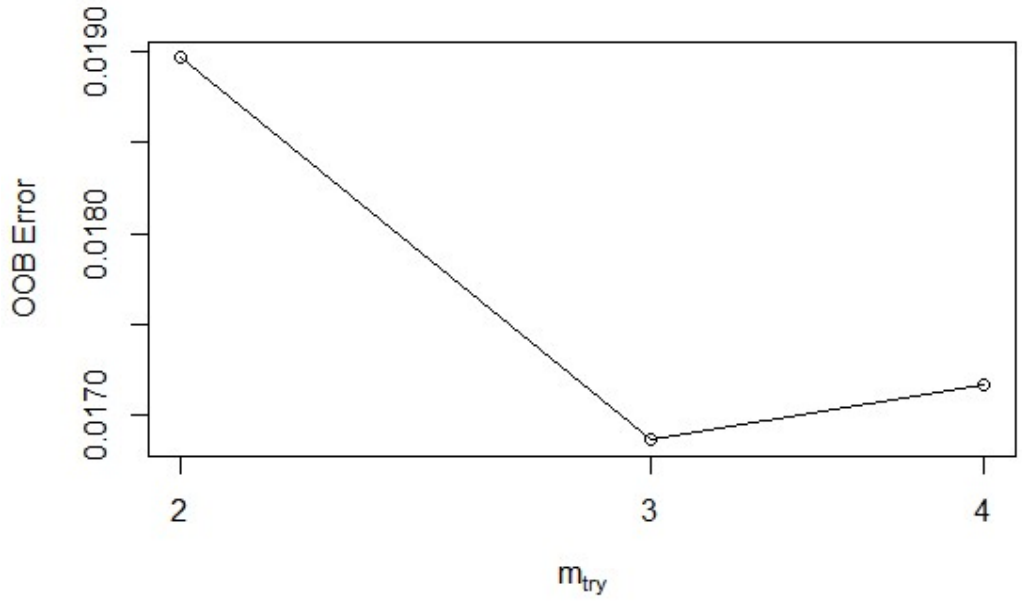
```
                             0          1 MeanDecreaseAccuracy
Age (in years)        13.599141  1.1014637            13.522359
Experience (in years) 12.787784  0.3913977            12.843585
Income (in K/month)   87.666828 75.5275085            95.127393
Family members        62.401495 31.8511101            63.049234
CCAvg                 29.694494 28.1089994            33.648511
Education             95.860091 61.7079291            97.926000
Mortgage               9.533724  0.8047892             8.934975
Securities Account     1.381382  4.7749848             4.483411
CD Account            11.264727 10.9992668            15.009302
Online                 2.766780  3.2111281             4.375709
CreditCard             8.459298  2.1695829             8.829368
                      MeanDecreaseGini
Age (in years)               12.718619
Experience (in years)        13.032218
Income (in K/month)         170.897822
Family members               54.751029
CCAvg                        87.204169
Education                   113.507210
Mortgage                     17.292020
Securities Account            1.391286
CD Account                   26.859780
Online                        2.093027
CreditCard                    3.280585
```

'EDUCATION'; 'INCOME'; 'FAMILY MEMBERS' are variables that contribute most for the accuracy of the model.

With the above factored parameters , we will now attempt to TUNE the Model –

```
set.seed(1500)
trforest= tuneRF(x=bank6train[,-c(8)],y= bank6train$`Personal Loan`,mytrystart =5, stepFactor = 1.5, ntree= 91,improve = 0.001,nodesize= 10,trace = TRUE, plot = TRUE,doBest = TRUE, importance= TRUE)
```

Plot for OOB ~ mTry =



Below is the method adopted by the algorithm to check the Best number of 'mTry' for the least 'OOB' error rate –

```
mtry = 3  OOB error = 1.69%
Searching left ...
mtry = 2       OOB error = 1.9%
-0.125 0.001
Searching right ...
mtry = 4       OOB error = 1.72%
-0.01785714 0.001
```

Out of the trial attempted by the algorithm, with mTry = 3, the OOB error rate reduced from 1.78%[earlier evaluated with 501 trees] to 1.69%. The above plot also reveals the same.

FINE-TUNED ; FINAL Model of Random Forest -

Applying the above parameters, the best random forest model is generated –

```
rforesttrain= randomForest(`Personal Loan`~.,data = bank6train, ntree= 91, mtry=3,nodesize= 10,importance= TRUE)
```

The Evaluations are tabled below ➔

```
randomForest(formula = `Personal Loan` ~ ., data = bank6train,      ntree
= 91, mtry = 3, nodesize = 10, importance = TRUE)
               Type of random forest: classification
                     Number of trees: 91
No. of variables tried at each split: 3

        OOB estimate of  error rate: 1.6%
Confusion matrix:
     0   1 class.error
0 3001   4 0.001331115
1   49 267 0.155063291
```

It is obvious that-

A. **The 'Out-Of-Bag' error rate is reduced to its lowest - @ 1.6%**
B. **The Confusion Matrix estimates only 15.50% error on the predicted values over the actual values.**

Having generated the best Classification tree using the random Forest Technique, now we move on for target variable prediction. This will help us derive the accuracy of the model ➔ comparing the Predicted values to the Target actual values – BOTH ON THE TRAIN Dataset and the TEST Dataset.

A column is created in both TEST & TRAIN Dataset for predicted values-
bank6train$predict= predict(rforesttrain, data= bank6train, type= 'class')
bank6test$predict= predict(rforesttrain, newdata= bank6test, type = 'class')

A **CONFUSION MATRIX** is generated for comparing error rate Predicted Values ~ Actual Values- both for TRAIN Dataset & TEST Dataset. The same is tabulated below –

```
tabtrainrf
             predict
Personal Loan    0    1
            0 3001    4
            1   49  267
> tabtestrf
             predict
Personal Loan    0    1
            0 1496    3
            1   17  145
```

The Accuracy parameters are derived from this Confusion Matrix , tabled as below –

```
accuracytrainrf
[1] 0.984041
> accuracytestrf
[1] 0.9879591
```

INFERENCES -

THE TEST DATASET ACCURACY IS ALMOST EQUIVALENT TO THE TRAIN DATASET ACCURACY – INDICATING THAT THE RANDOM FOREST MODEL IS ROBUST.

CONFUSION MATRIX INTERPRETATION – CART MODEL & RANDOM FOREST MODEL

The data as copied from the code file is tabulated and presented below for analysis -

TRAIN DATA SET =>				TEST DATA SET =>			
		PREDICTED				PREDICTED	
RANDOM FOREST		0	1	RANDOM FOREST		0	1
ACTUAL	0	2997	8	ACTUAL	0	1496	3
	1	51	265		1	17	145
CART MODEL				CART MODEL			
	0	2981	13		0	1506	4
	1	99	229		1	54	96

The above data is the CONFUSION MATRIX for each technique of classification used on a SAME DATA SET – target variable being ' Personal Loan'.

The Confusion Matrix results of the Random Forest model IS BETTER THAN the CART model – both in the TRAIN Dataset & TEST Dataset.
i.e. IN BOTH THE MODELS DEFINED BY THE RANDOM FOREST TECHNIQUE – TRUE POSITIVE + TRUE NEGETIVE DATA POINTS ARE MORE THAN AS DEFINED BY THE CART TECHNIQUE.

Additional Performance parameters check is also tabulated as below, which evidences that MODEL built with the RANDOM Forest technique is BETTER THAN the MODEL built with CART Technique –

TRAIN Data-Set

	Random Forest	CART
Classification Error	98.223%	96.629%
Sensitivity	83.860%	69.810%
Specificity	99.730%	99.560%

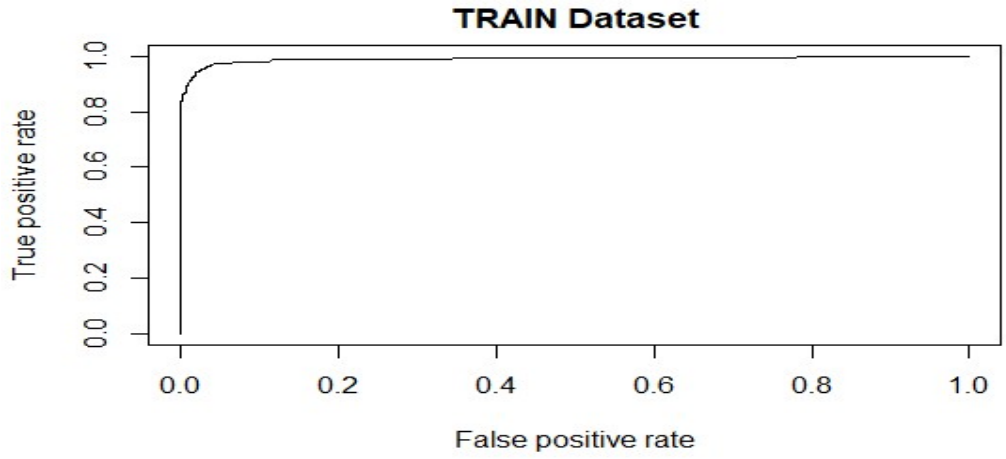
TEST Data-Set

	Random Forest	CART
Classification Error	98.796%	96.506%
Sensitivity	89.506%	64.000%
Specificity	99.800%	99.735%

MODEL PERFORMANCE – AOC & ROC Techniques

Using the library(ROCR), we are now evaluating the ROC Curve for the TRAIN Data Set & TEST Dataset of BOTH Models – The Model derived from CART Technique & RANDOM Forest Technique.

ROC Curve for MODEL from Random Forest Technique –

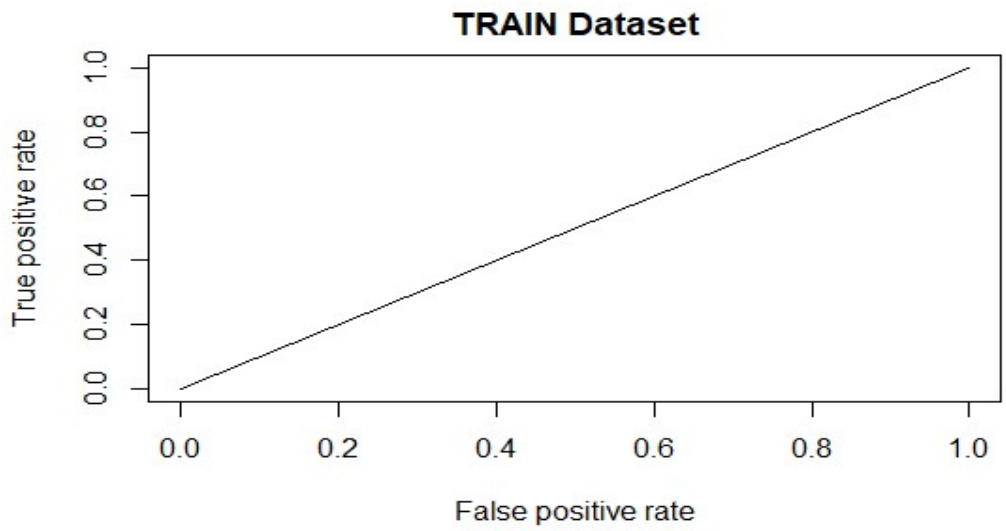


The Curve height in the TEST Data Set is much higher than the TRAIN Dataset. The higher the curve, LESSER the FPR Rate & GREATER the TPR Rate – which indicates the BEST MODEL.

AUC Values as tabulated as –

```
[1] "TRAIN AUC(RF) 0.990894395416921"
[1] "TEST AUC(RF) 0.998700779943831"
```

ROC Curve for MODEL from CART Technique –



The ROC Curve plotting as above DOES NOT INDICATE A CORELATION BETWEEN 'TPR' & 'FPR' values. This means that the MODEL does not produce a STABLE performance for the predicted values.

AUC Values as tabulated as –

```
[1] "TRAIN AUC(CART) 0.5"
[1] "TEST AUC(CART) 0.5"
```

MODEL VALIDATION EXERCISE

The MODEL Performance is being graded on the evaluation of FIVE BASIC Parameters –

1. SENSITIVITY – being the ratio of TRUE POSITIVE RATE to the TOTAL POSITIVE Outcomes correctly identified.
2. KS Values – The Percentage figure at which the maximum response is received.
3. LIFT – Value indicates how many times the Positive Response we got w.r.t. the random base line response.
4. CONCORDANCE – gives the percentage of Occurrence of the event whenever the actual occurrence was NIL.
5. GINI Coefficient – represents the inequality between objects and takes a value between 0 & 1.

The MODEL PERFORMANCE , of TWO Techniques used for the same data set, is tabulated below –

	RANDOM FOREST Technique		CART Technique	
	TRAIN Dataset	**TEST Dataset**	**TRAIN Dataset**	**TEST Dataset**
KS Value	92.90%	95.97%	0	0
LIFT	10.51	10.25	NaN	NaN
AUC	99.08%	99.87%	0.5	0.5
Sensitivity	1.00	1.00	1	1
Concordance	98.50%	99.86%	0	0
GINI	90.04%	89.51%	0	0

THE MODEL GENERATED WITH RANDOM FOREST TECHNIQUE GIVES MORE ROBUST STATUS OF PERFORMANCE.