1/23/2022

Abhishek Barve

Great Learning

Data Mining Project

**LIST OF FIGURES**

|  |  |
| --- | --- |
| Graph No | Graph Name |
| Fig 1.1 | Histplot-Spending |
| Fig 1.2 | Histplot-Advance payments |
| Fig 1.3 | Histplot-Probability of full payment |
| Fig 1.4 | Histplot-Current balance |
| Fig 1.5 | Histplot-Credit limit |
| Fig 1.6 | Histplot-Mini payment amt |
| Fig 1.7 | Histplot-Max spent in single shopping |
| Fig 2.1 | Boxplot- All Variables |
| Fig 2.2 | Boxplot- Mini payment amt |
| Fig 2.3 | Boxplot- Probability of full payment |
| Fig 3 | Pairplot to show multicollinearity |
| Fig 4 | Correlation Heatmap |
| Fig 6.1 | Dendrogram (Ward’s method) |
| Fig 6.2 | Truncate mode of Dendrogram |
| Fig 7.5 | Elbow curve using Plot model |
| Fig 8.6 | Boxplot to refer outliers |
| Fig 8.7 | Univariate & Bivariate analysis |
| Fig 8.9 | Countplot on Agency Code |
| Fig 8.15 | Pairplot to show multivariate analysis |
| Fig 8.16 | Heatmap to check correlation among multi variables. |
| Fig 9 | Boxplot to check presence of outliers |
| Fig 10.1 | AUC & ROC Curve on Training Data (DT) |
| Fig 10.2 | AUC & ROC Curve on Test Data (DT) |
| Fig 10.3 | AUC and ROC for the training data (RF) |
| Fig 10.4 | AUC and ROC for the test data (RF) |
| Fig 10.5 | AUC and ROC for the training data (ANN) |
| Fig 10.6 | AUC and ROC for the test data (ANN) |

**Problem 1**

1.1 Read the data and do exploratory data analysis (3 pts). Describe the data briefly. Interpret the inferences for each (3 pts). Initial steps like head() .info(), Data Types, etc. . Null value check. Distribution plots(histogram) or similar plots for the continuous columns. Box plots, Correlation plots. Appropriate plots for categorical variables. Inferences on each plot. Summary stats, Skewness, Outliers proportion should be discussed, and inferences from above used plots should be there. There is no restriction on how the learner wishes to implement this but the code should be able to represent the correct output and inferences should be logical and correct.……………………………………………………….................................... 6

1.2. Do you think scaling is necessary for clustering in this case? Justify The learner is expected to check and comment about the difference in scale of different features on the bases of appropriate measure for example std dev, variance, etc. Should justify whether there is a necessity for scaling and which method is he/she using to do the scaling. Can also comment on how that method works…………………………………...................................................................................................................2

1.3 Apply hierarchical clustering to scaled data (3 pts). Identify the number of optimum clusters using Dendrogram and briefly describe them (4). Students are expected to apply hierarchical clustering. It can be obtained via Fclusters or Agglomerative Clustering. Report should talk about the used criterion, affinity and linkage. Report must contain a Dendrogram and a logical reason behind choosing the optimum number of clusters and Inferences on the dendrogram. Customer segmentation can be visualized using limited features or whole data but it should be clear, correct and logical. Use appropriate plots to visualize the clusters………………………………………………………………………………………………………………………………………………………7

1.4 Apply K-Means clustering on scaled data and determine optimum clusters (2 pts). Apply elbow curve and silhouette score (3 pts). Interpret the inferences from the model (2.5 pts). K-means clustering code application with different number of clusters. Calculation of WSS(inertia for each value of k) Elbow Method must be applied and visualized with different values of K. Reasoning behind the selection of the optimal value of K must be explained properly. Silhouette Score must be calculated for the same values of K taken above and commented on. Report must contain logical and correct explanations for choosing the optimum clusters using both elbow method and silhouette scores. Append cluster labels obtained from K-means clustering into the original data frame. Customer Segmentation can be visualized using appropriate graphs. 7

1.5 Describe cluster profiles for the clusters defined (2.5 pts). Recommend different promotional strategies for different clusters in context to the business problem in-hand (2.5 pts). After adding the final clusters to the original dataframe, do the cluster profiling. Divide the data in the finalized groups and check their means. Explain each of the group briefly. There should be at least 3-4 Recommendations. Recommendations should be easily understandable and business specific, students should not give any technical suggestions. Full marks will only be allotted if the recommendations are correct and business specific. Variable means. Students to explain the profiles and suggest a mechanism to approach each cluster. Any logical explanation is acceptable...............................................................................................5

**Problem 2**

2.1 Read the data and do exploratory data analysis (4 pts). Describe the data briefly. Interpret the inferences for each (2 pts). Initial steps like head () .info (), Data Types, etc. . . Null value check. Distribution plots (histogram) or similar plots for the continuous columns. Box plots, Correlation plots. Appropriate plots for categorical variables. Inferences on each plot. Summary stats, Skewness, Outliers proportion should be discussed, and inferences from above used plots should be there. There is no restriction on how the learner wishes to implement this but the code should be able to represent the correct output and inferences should be logical and correct?....................................................................................................................................................6

2.2 Data Split: Split the data into test and train (1 pts), build classification model CART (1.5 pts), Random Forest (1.5 pts), Artificial Neural Network (1.5 pts). Object data should be converted into categorical/numerical data to fit in the models. (pd.categorical().codes(), pd.get\_dummies(drop\_first=True)) Data split, ratio defined for the split, train-test split should be discussed. Any reasonable split is acceptable. Use of random state is mandatory. Successful implementation of each model. Logical reason behind the selection of different values for the parameters involved in each model. Apply grid search for each model and make models on Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution 2.3 best\_params. Feature importance for each model.……………………………………………………………………………………………………………………………………………….5.5

2.4 Performance Metrics: Check the performance of Predictions on Train and Test sets using Accuracy (1 pts), Confusion Matrix (2 pts), Plot ROC curve and get ROC\_AUC score for each model (2 pts), Make classification reports for each model. Write inferences on each model (2 pts). Calculate Train and Test Accuracies for each model. Comment on the validness of models (overfitting or underfitting) Build confusion matrix for each model. Comment on the positive class in hand. Must clearly show obs/pred in row/col Plot roc\_curve for each model. Calculate roc\_auc\_score for each model. Comment on the above calculated scores and plots. Build classification reports for each model. Comment on f1 score, precision and recall, which one is important here……………………………..……………………………………………………………..7

2.5 Final Model - Compare all models on the basis of the performance metrics in a structured tabular manner (2.5 pts). Describe on which model is best/optimized (1.5 pts). A table containing all the values of accuracies, precision, recall, auc\_roc\_score, f1 score. Comparison between the different models (final) on the basis of above table values. After comparison which model suits the best for the problem in hand on the basis of different measures? Comment on the final model.......................................................................................................................................................4

2.6 Based on your analysis and working on the business problem, detail out appropriate insights and recommendations to help the management solve the business objective. There should be at least 3-4 Recommendations and insights in total. Recommendations should be easily understandable and business specific, students should not give any technical suggestions. Full marks should only be allotted if the recommendations are correct and business specific…………………………………………………………………4

5 - Quality of Business Report (Please refer to the Evaluation Guidelines for Business report checklist. Marks in this criteria are at the moderator's discretion)…………………………………………………………………………………..6 - Please reflect on all that you learnt and fill this reflection report:https://docs.google.com/forms/d/e/1FAIpQLSd7e\_bJVCiFpZAYbBTMtKrr9TLRnl8kuvZT7IsZ5MSjRt fjcQ/viewform?usp=sf\_link........................................................0

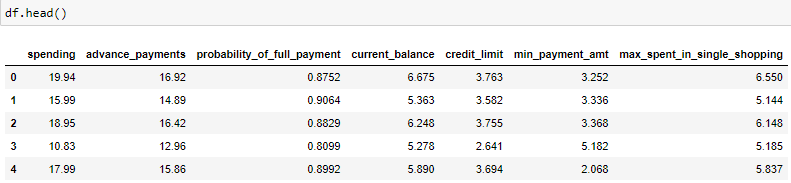
**Clustering**

A leading bank wants to develop a customer segmentation to give promotional offers to its customers. They collected a sample that summarizes the activities of users during the past few months. You are given the task to identify the segments based on credit card usage.

* 1. **Read the data, do the necessary initial steps, and exploratory data analysis (Univariate, Bi-variate, and multivariate analysis)**

**Initial Steps:**

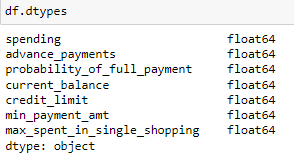
**Head**



Shape

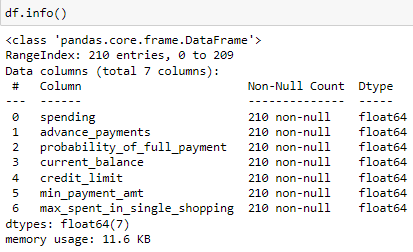
210 rows and 7 columns are present in data set.

**Data types**



All Variables are present in object data type in data set.

**Info**

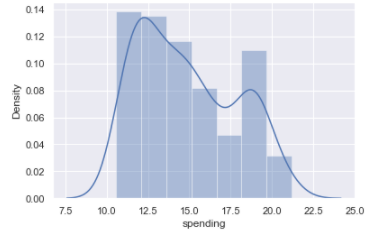
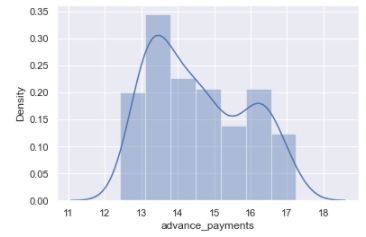


**Duplicate**

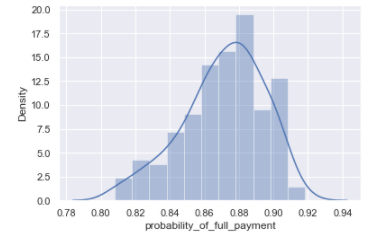
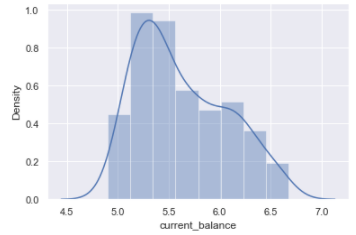


There is no missing value and duplicates in data set.

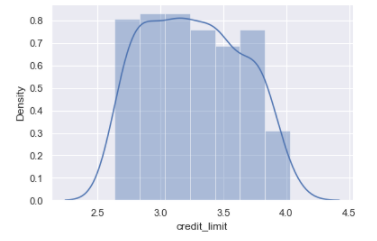
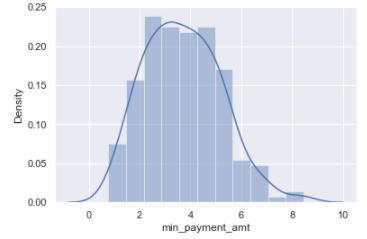
**Distribution Plots:**

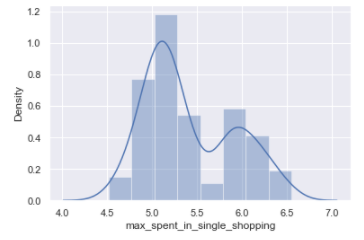
Fig(1.1) Fig(1.2)

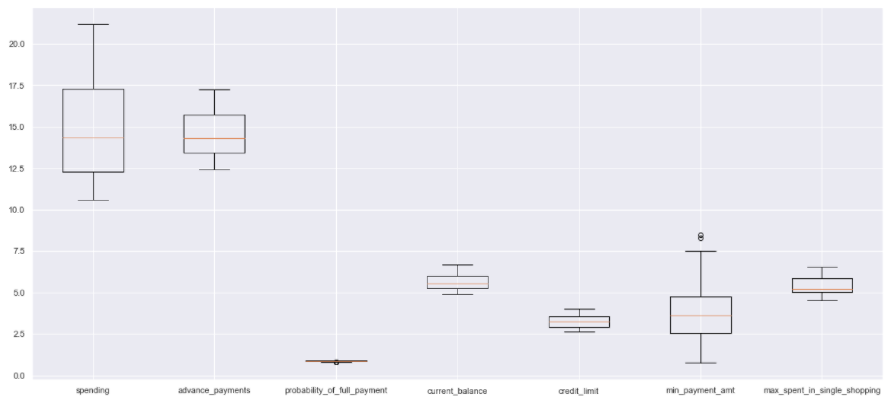
Fig(1.3) Fig(1.4)

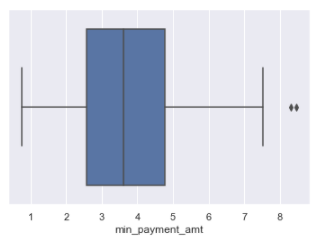
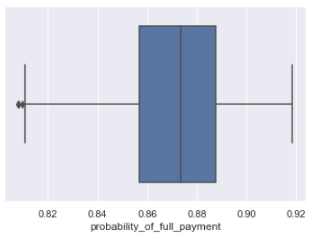
Fig(1.5) Fig(1.6)



Fig(1.7)



Fig(2.1)

Fig(2.2) Fig(2.3)

**Inferences:**

**Spending-**

* The box plot of the spending variable shows no outliers.
* Spending is positively skewed - 0.399889.
* Spending dist plot shows the distribution of data from 10 to 22.
* The box plot of the advance payments variable shows no outliers.
* Advance payments is positively skewed.
* The dist plot shows the distribution of data from 12 to 17.

**Probability of Full Payment-**

* The box plot of the probability of full payment variable shows few outliers.
* Probability of full payment is negatively skewed - -0.537954
* The dist plot shows the distribution of data from 0.80 to 0.92.

**Current Balance-**

* The box plot of the current balance variable shows no outliers.
* Current balance is positively skewed.
* The dist plot shows the distribution of data from 5.0 to 6.5.

**Credit Limit-**

* The box plot of the credit limit variable shows no outliers.
* Credit limit is positively skewed.
* The dist plot shows the distribution of data from 2.5 to 4.0.

**Min Payment Amount-**

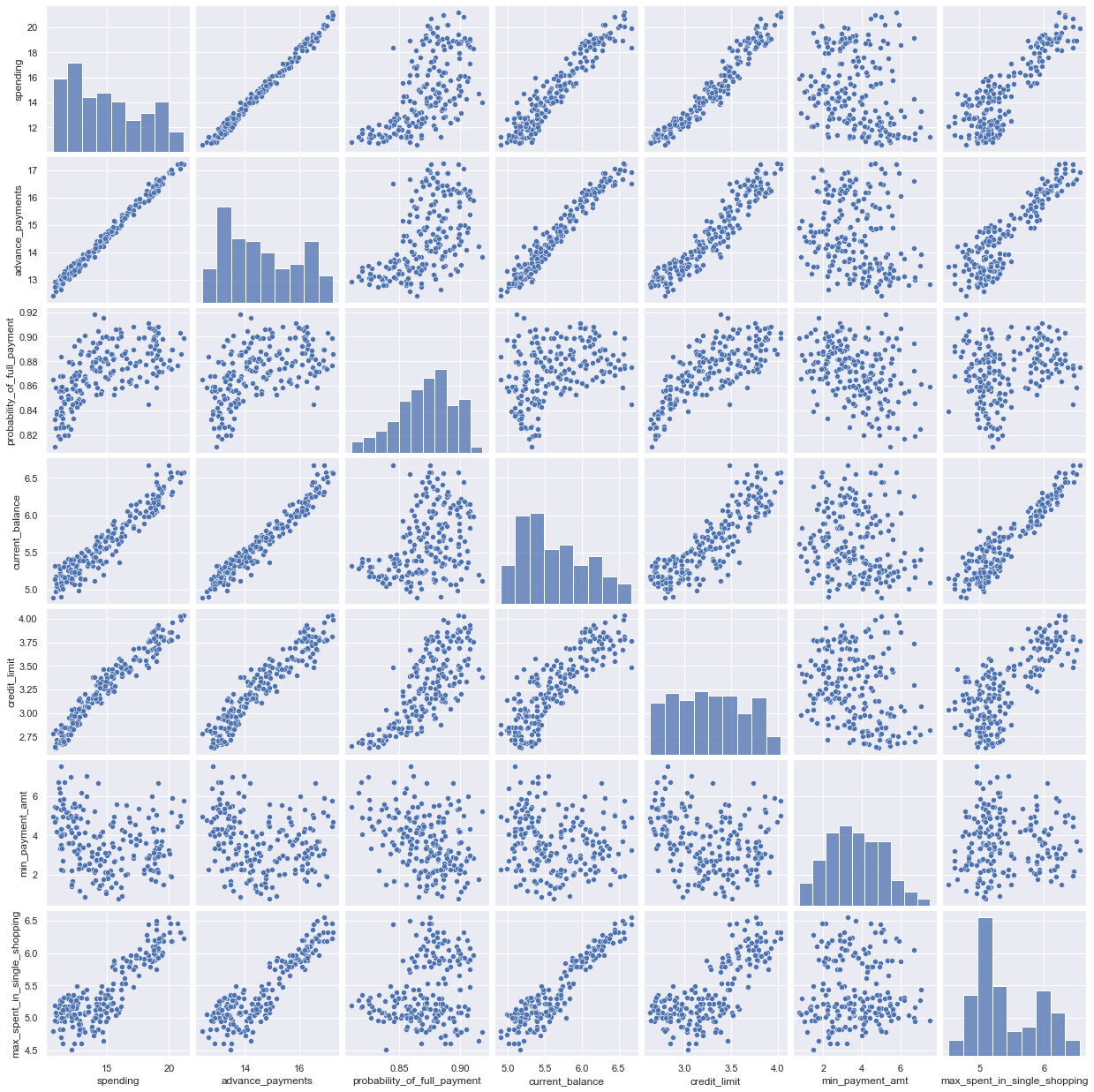
* The box plot of the min payment amount variable shows few outliers.
* Min payment amount is positively skewed.
* The dist plot shows the distribution of data from 2 to 8.

**Max Spent In Single Shopping-**

* The box plot of the max spent in single shopping variable shows no outliers.
* Max spent in single shopping is positively skewed.
* The dist plot shows the distribution of data from 4.5 to 6.5.

**Multivariate analysis-**

**Check for multicollinearity: Pair plot**

****

Fig(3)

**Heat map for correlation**

Fig(4)

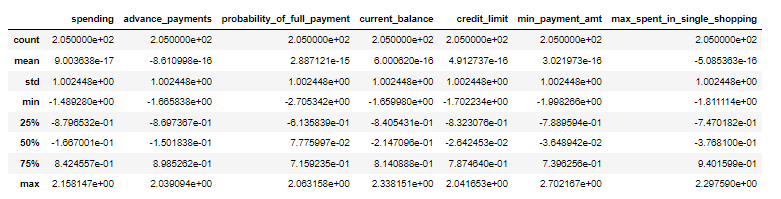
**Inferences:**

* Strong positive correlation between spending & advance payments.
* Strong positive correlation between advance payments & current balance.
* Strong positive correlation between Credit limit & spending.
* Spending & current balance.
* credit limit & advance payments.
* Max\_spent\_in\_single\_shopping current balance.
  1. **Do you think scaling is necessary for clustering in this case? Justify**

Yes, scaling is very important as the model works based on the distance based computations scaling is necessary for unscaled data.

Scaling needs to be done as the values of the variables are in different scales. Spending, advance payments are in different values and this may get more weightage. Scaling will have all the values in the relative same range. I have used standard scalar for scaling

Below is the snapshot of scaled data.



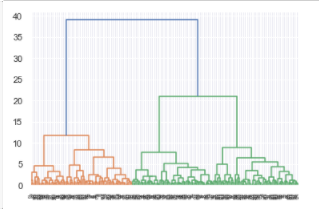
Fig(5)

* 1. **Apply hierarchical clustering to scaled data. Identify the number of optimum clusters using Dendrogram and briefly describe them.**

**Hierarchical clustering – ward’s method**

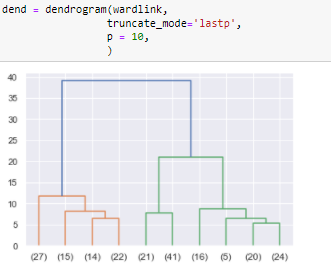
By choosing ward’s method to the scaled data,

For visualization purposes I have used Dendrogram.



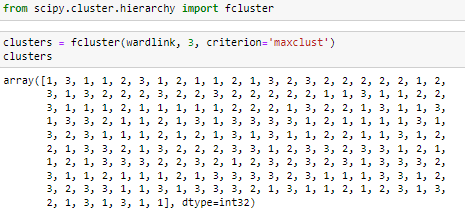
Fig(6.1)

The above dendrogram indicates all the data points have clustered to different clusters by wards method. To find the optimal number cluster through which we can solve our business objective we use truncate mode = lastp. Wherein we can give last p = 10 according to industry set base value.

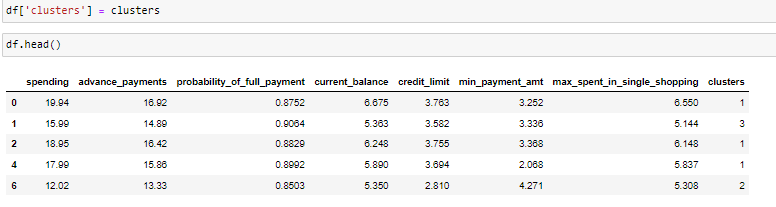


Fig(6.2)

Now, we can understand all the data points have clustered into 3 clusters. Next to map these clusters to our dataset we can use fclusters Criterion we can give “maxclust”

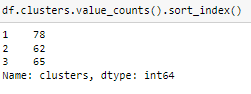


Fig(6.3)



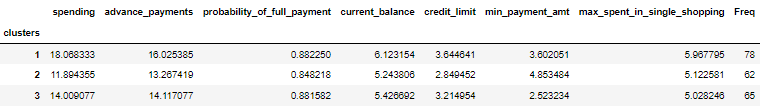
Fig(6.4)

Now, we can look at the cluster frequency in our dataset.

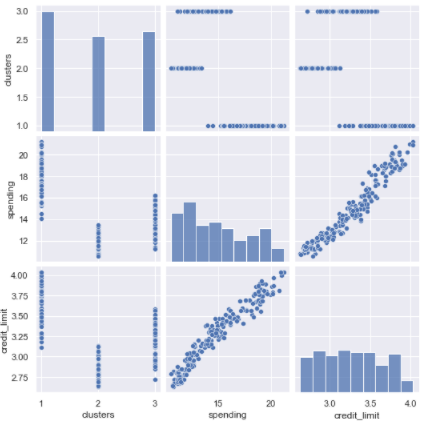


Fig(6.5)

Cluster profiling to understand the business problem.



Fig(6.6)



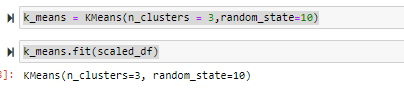
Fig(6.7)

**Observation**

After doing further analysis we have grouped the dataset in 3 clusters (1, 2 and 3. From the above figure we can clearly see cluster 1 seems to be high in spending with advance in payments. Whereas cluster 2 are spending less hence their credit limit is also at lower side among all clusters. We can rollout some attractive offers for cluster 1 customers so they can get more benefits.

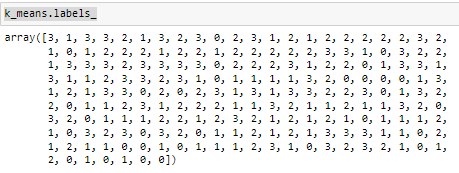
* 1. **Apply K-Means clustering on scaled data and determine optimum clusters. Apply elbow curve and silhouette score. Explain the results properly. Interpret and write inferences on the finalized clusters.**

K-means clustering, Randomly we decide to give n\_clusters = 3 and we look at the distribution of clusters according to the n\_clusters. We apply K-means technique to the scaled data.



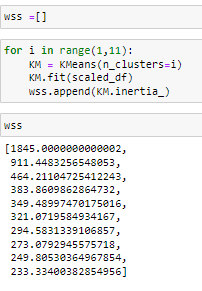
Fig(7.1)

Cluster output for all the observations in the dataset.



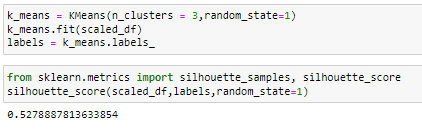
Fig(7.2)

We have 3 clusters 0,1,2 To find the optimal number of clusters, we can use k-elbow method.



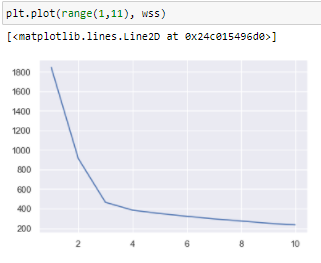
Fig(7.3)

To find the inertia value for all the clusters from 1 to 11, I used a for loop to find the optimal number of clusters. The silhouette score for 3 clusters is good.



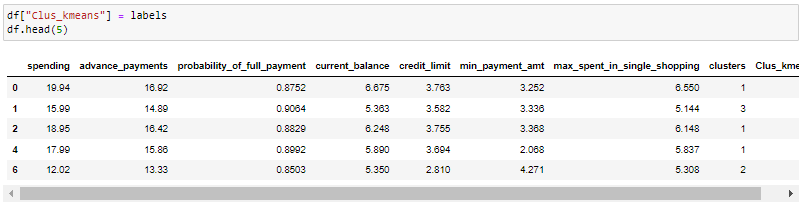
Fig(7.4)

The elbow curve seen here also shows us after 3 clusters there is no huge drop in the values, so we select 3 clusters.



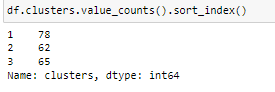
Fig(7.5)

So adding the cluster results to our dataset to solve our business objective.

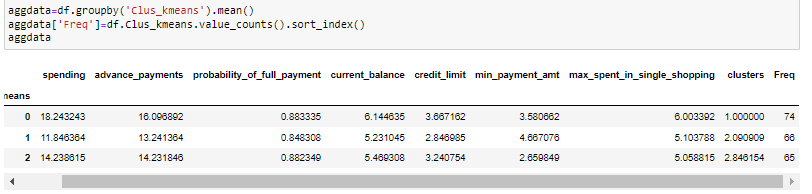


Fig(7.6)

This table shows the clusters to the dataset and also individual sil\_width score. Cluster frequency.



Fig(7.7)



Fig(7.8)

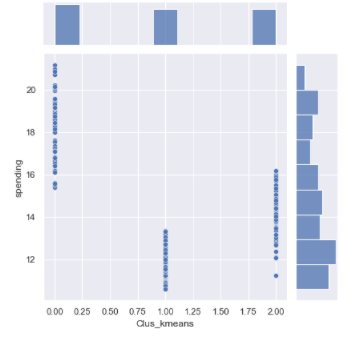
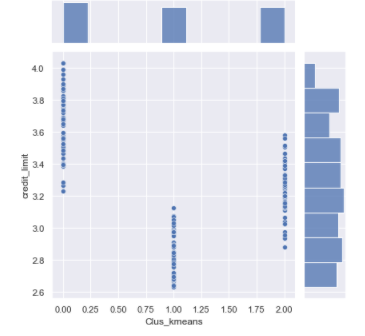
3-Group clusters via K- Means has equal split of percentage of results.

Cluster 0 – Medium

Cluster 1 – low

Cluster 2 – High

**Joint Plot-** Visualize a relationship between cluster, spending and credit limit.

**Observation**

By K- Mean’s method we can at cluster 3 we find it optimal after there is no huge drop in inertia values. Also the elbow curve seems to show similar results. The silhouette width score of the K – means also seems to very less value that indicates all the data points are properly clustered to the cluster. There is no mismatch in the data points with regards to clustering Cluster grouping based on the dendrogram, 3 or 4 looks good. Did the further analysis, and based on the dataset had gone for 3 group cluster And three group cluster solution gives a pattern based on high/medium/low spending with max\_spent\_in\_single\_shopping (high value item) and probability\_of\_full\_payment (payment made).

* 1. Describe cluster profiles for the clusters defined. Recommend different promotional strategies for different clusters.

**Group 1: High Spending Group –**

Giving any reward points might increase their purchases. – Maximum max\_spent\_in\_single\_shopping is high for this group, so can be offered discount/offer on next transactions upon full payment – Increase their credit limit and – Increase spending habits – Give loan against the credit card, as they are customers with good repayment record. – Tie up with luxury brands, which will drive more one\_time\_maximun spending

**Group 2: Low Spending Group –**

Customers should be given remainders for payments. Offers can be provided on early payments to improve their payment rate. - Increase their spending habits by tying up with grocery stores, utilities (electricity, phone, gas, others)

**Group 3: Medium Spending Group –**

They are potential target customers who are paying bills and doing purchases and maintaining comparatively good credit score. So we can increase credit limit or can lower down interest rate. - Promote premium cards/loyalty cars to increase transactions. - Increase spending habits by trying with premium ecommerce sites, travel portal, travel airlines/hotel, as this will encourage them to spend more.

**CART-RF-ANN**

An Insurance firm providing tour insurance is facing higher claim frequency. The management decides to collect data from the past few years. You are assigned the task to make a model which predicts the claim status and provide recommendations to management. Use CART, RF & ANN and compare the models' performances in train and test sets.

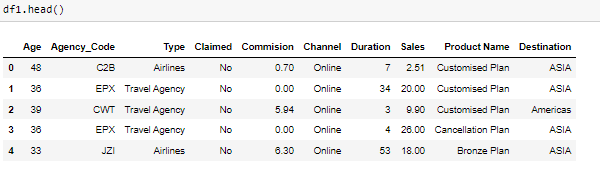
**Attribute Information:**

1. Target: Claim Status (Claimed)  
   2. Code of tour firm (Agency\_Code)  
   3. Type of tour insurance firms (Type)  
   4. Distribution channel of tour insurance agencies (Channel)  
   5. Name of the tour insurance products (Product)  
   6. Duration of the tour (Duration in days)  
   7. Destination of the tour (Destination)  
   8. Amount worth of sales per customer in procuring tour insurance policies in rupees (in 100’s)  
   9. The commission received for tour insurance firm (Commission is in percentage of sales)  
   10.Age of insured (Age)

**2.1 Read the data, do the necessary initial steps, and exploratory data analysis (Univariate, Bi-variate, and multivariate analysis).**

**Initial steps:** head() .info(), Data Types, etc .

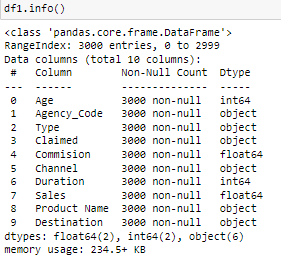
Reading the dataset-



Fig(8.1)

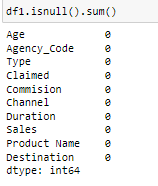
Info of the dataset-

The shape of the dataset is (3000, 10) Info function clearly indicates the dataset has object, integer and float so we have to change the object data type to numeric value.



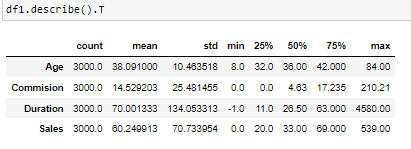
Fig(8.2)

Check missing value-



Fig(8.3)

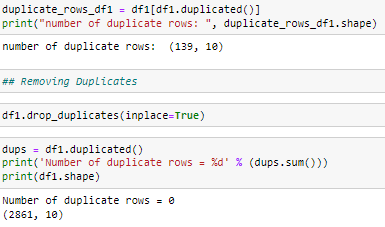
Summary of the dataset-



Fig(8.4)

We have 4 numeric values and 6 categorical values, Agency code EPX has a frequency of 1365, the most preferred type seems to be travel agency Channel is online Customized plan is the most sought plan by customers Destination ASIA seems to be most sought destination place by customers. We will further look at the distribution of dataset in univariate and bivariate analysis.

Checking for duplicates in the dataset-



Fig(8.5)

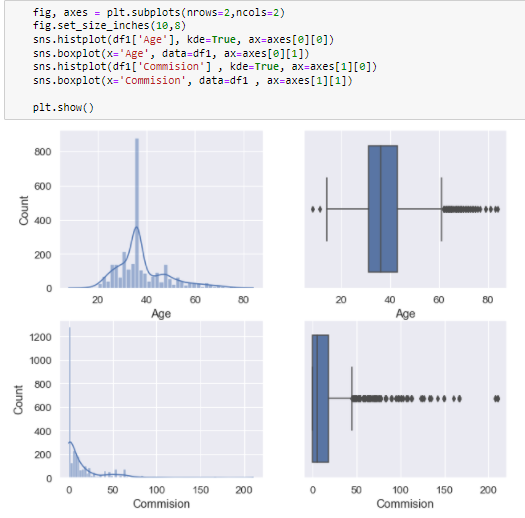
Checking for Outliers –



Fig(8.6)

Outliers exist in almost all the numeric values. We can treat outliers in random forest classification.

**Univariate & Bivariate Analysis –**



Fig(8.7)

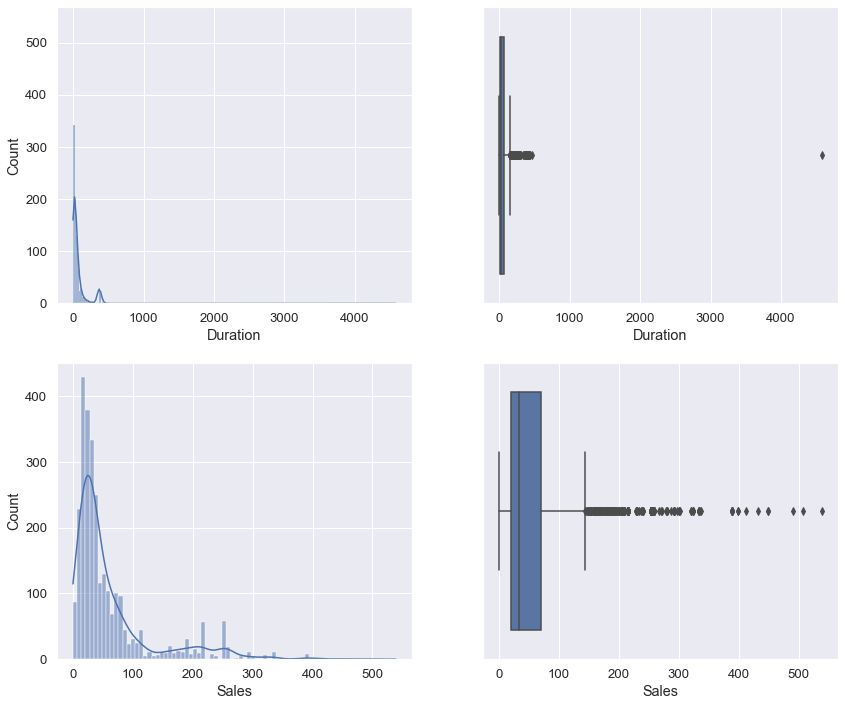
Below analysis is for Age variable –

The box plot of the age variable shows outliers. Spending is positively skewed.

The dist plot shows the distribution of data from 20 to 80 In the range of 30 to 40 is where the majority of the distribution lies.

Below analysis is for Commision variable –

The box plot of the commission variable shows outliers. Spending is positively skewed The dist plot shows the distribution of data from 0 to 30 majorly.

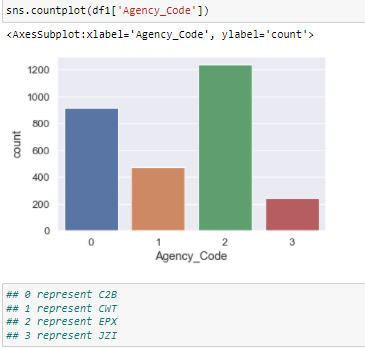


Fig(8.8)

The box plot of the duration variable shows outliers. Spending is positively skewed The dist plot shows the distribution of data from 0 to 100 The box plot of the sales variable shows outliers. Spending is positively skewed. The dist plot shows the distribution of data from 0 to 300.

**Categorical Variables-**

**Agency Code**



Fig(8.9)

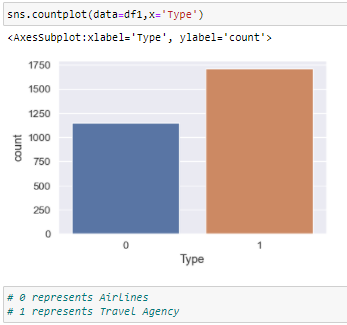
The distribution of the agency code, shows us EPX with maximum frequency.



Fig(8.10)

The box plot shows the split of sales with different agency code and also hue having claimed column. It seems that C2B have claimed more claims than other agency.

**Type**



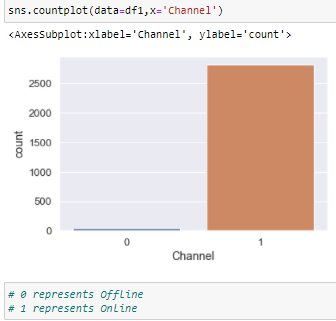
Fig(8.11)



Fig(8.12)

The box plot shows the split of sales with different type and also hue having claimed column. We could understand airlines type has more claims.

**Channel –**



Fig(8.13)

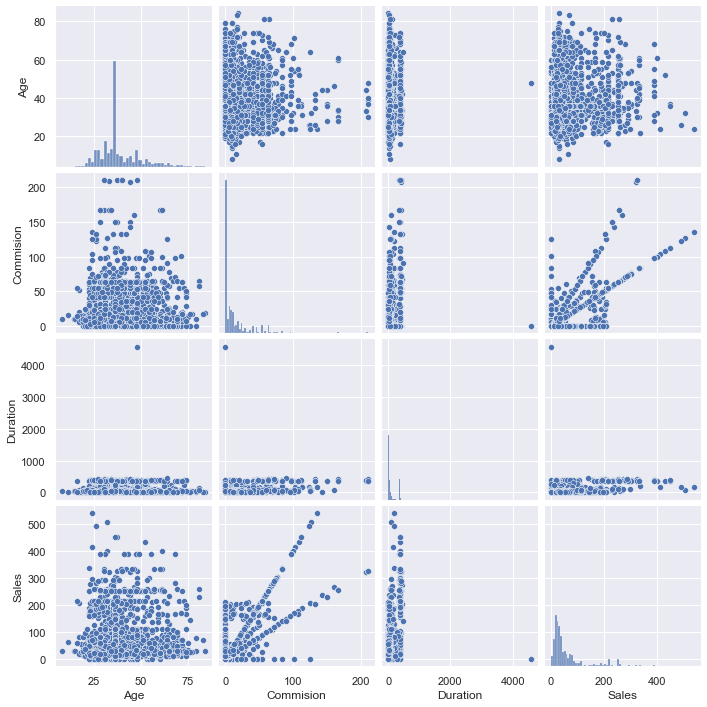
The majority of customers have used online medium, very less with offline medium.

**Product Name –**



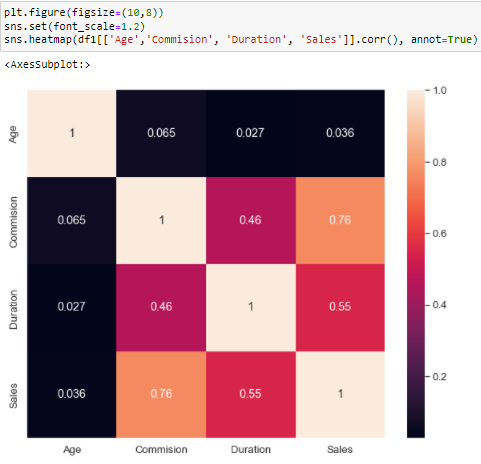
Customized plan seems to be most liked plan by customers when compared to all other plans.

**Checking pairwise distribution of the continuous variables-**

****

Fig(8.15)

**Checking for Correlations-**

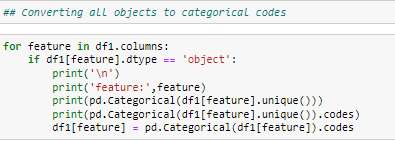


Fig(8.16)

**Inferences-**

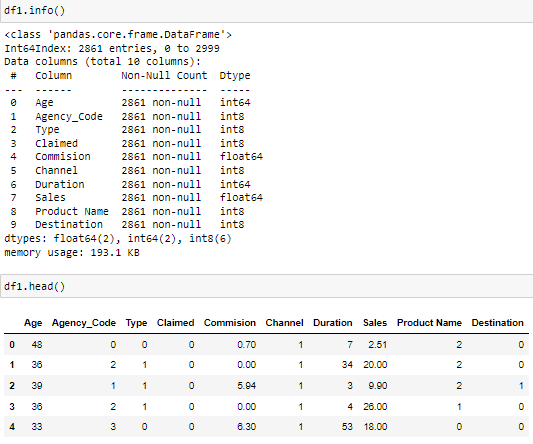
* Not much of multi collinearity observed.
* No negative correlation.
* Only positive correlation.

**Converting all objects to Categorical Codes-**



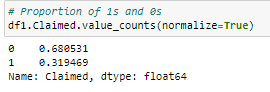
Fig(8.17)

To build our model we have changed the object data types into numeric values.



Fig(8.18)

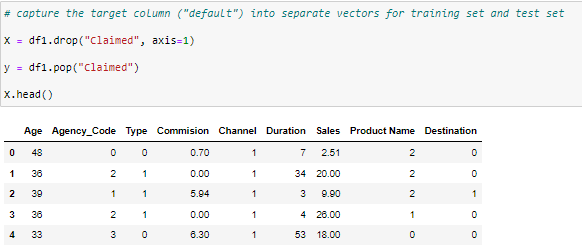
**Checking the proportion of 1s and 2s in the dataset. That is our target column.**



Fig(8.19)

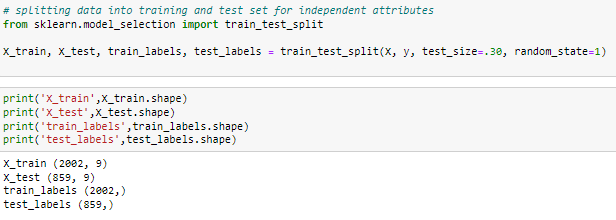
**2.2 Data Split: Split the data into test and train, build classification model CART, Random Forest, Artificial Neural Network**

**Extracting the target column into separate vectors for training set and test set**



Fig(9.1)

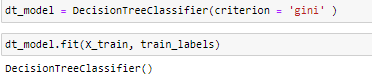
For training and testing purpose we are splitting the dataset into train and test data in the ratio 70:30.



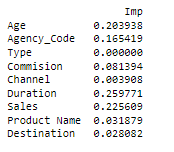
Fig(9.2)

We have bifurcated the dataset into train and test. We have also taken out the target column out of train and test data into separate vector for evaluation purposes.

**CART Model-**

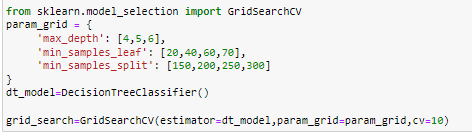


Checking the feature

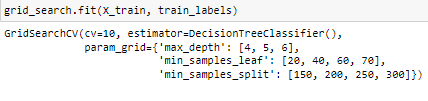


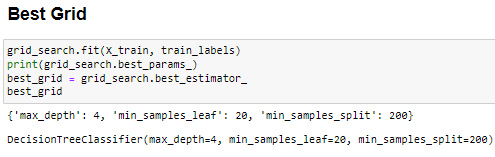
OPTIMAL VALUES FOR DECISSION TREE, GRID SEARCH FOR FINDING

Grid Search for finding out the optimal values for the hyper parameters



FITTING THE OPTMAL VALUES TO THE TRAINING DATASET





Generate New Tree

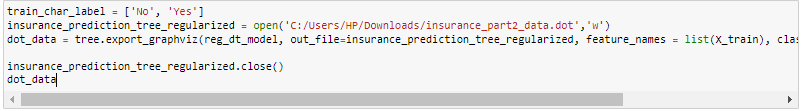
train\_char\_label = ['No', 'Yes']

insurance\_prediction\_tree\_regularized = open('C:/Users/HP/Downloads/insurance\_part2\_data.dot','w')

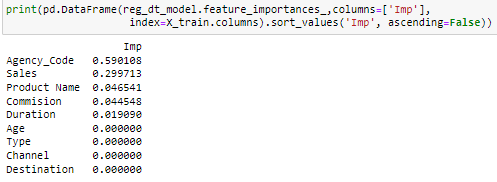
dot\_data = tree.export\_graphviz(reg\_dt\_model, out\_file=insurance\_prediction\_tree\_regularized, feature\_names = list(X\_train), class\_names = list(train\_char\_label))

insurance\_prediction\_tree\_regularized.close()

dot\_data



Important Variables-



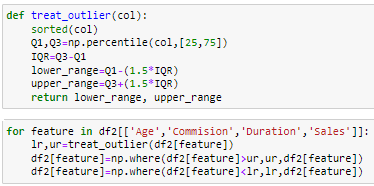
Predicting on Training dataset for Decision Tree-



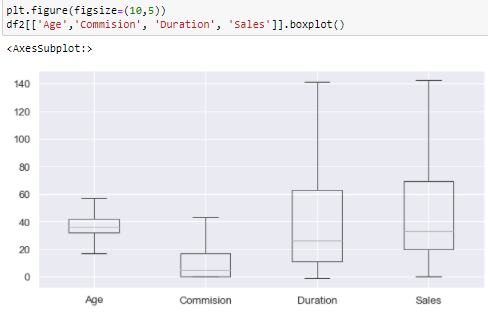
**Model 2**

Building Ensemble RandomForest Classifier

**TREATING OUTLIERS FOR RANDOM FOREST**

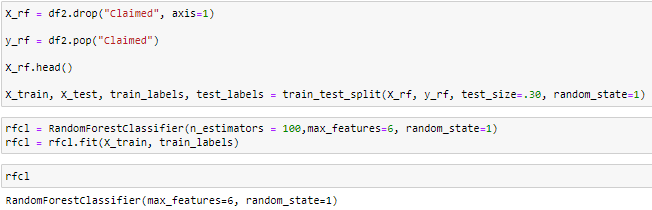


**BOX PLOT TO CHECK PRESENCE OF OUTLIERS**

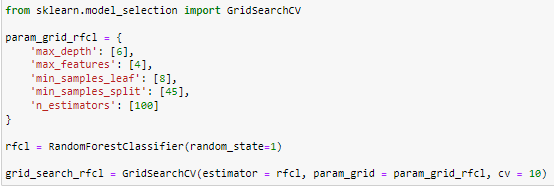


Fig(9)

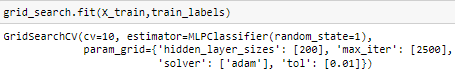
**RANDOM FOREST CLASSIFIER**



To find Optimal numbers using Grid Search Grid Search for finding out the optimal values for the hyper parameters



FIFTING THE MODEL TO RFCL VALUES OBTAINED BY OPTIMAL GRID SEARCH METHOD



**BEST GRID PARAM VALUES**

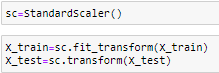




**MODEL 3**

**Building a Neural Network Classifier**

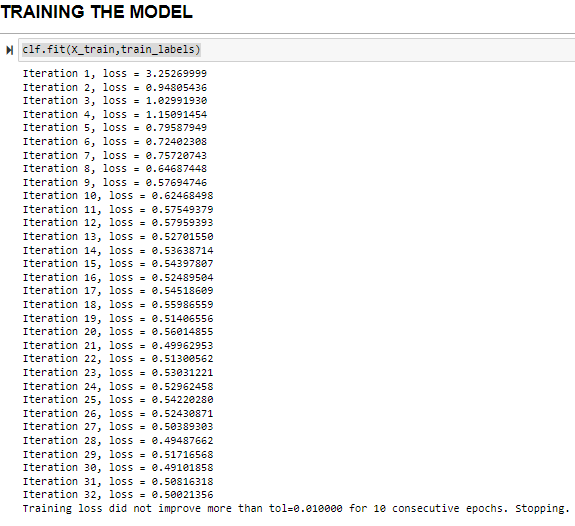
**BEFORE BUILDING THE MODEL WE SCALE THE VALUES, TO STANDARD SCALE USING MINMAXSCALER**



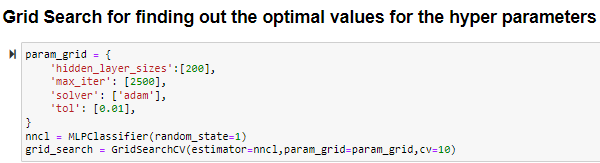
**MLP CLASSIFIER**

clf=MLPClassifier(hidden\_layer\_sizes=(100),solver='sgd',random\_state=21,tol=0.01,max\_iter=5000, verbose=True)

**Training the Model**

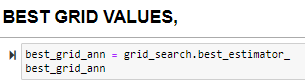


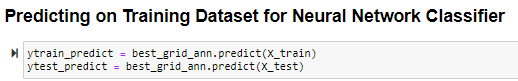
**GRID Search –**



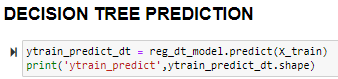
FITTING THE MODEL USING THE OPTIMAL VALUES FROM GRID SEARCH



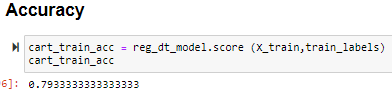




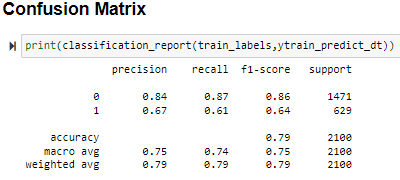
**2.3 Performance Metrics: Check the performance of Predictions on Train and Test sets using Accuracy, Confusion Matrix, Plot ROC curve and get ROC\_AUC score for each model**

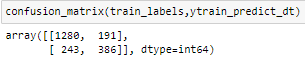


**Accuracy**



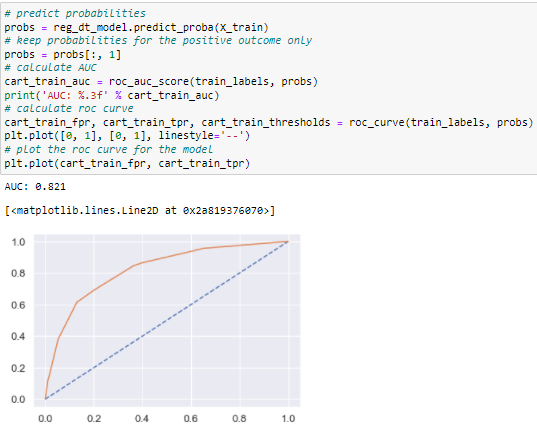
**Confusion Matrix**





**Model Evaluation for Decision Tree**

**AUC and ROC for the training data for Decision Tree**



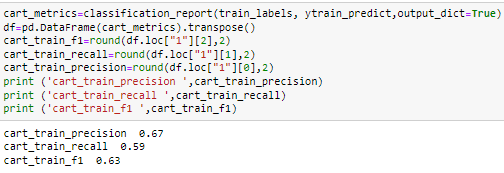
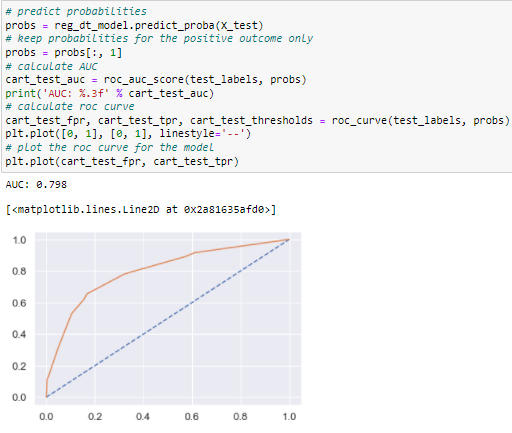


Fig 10.1

**AUC & ROC for the test data for Decision Tree**



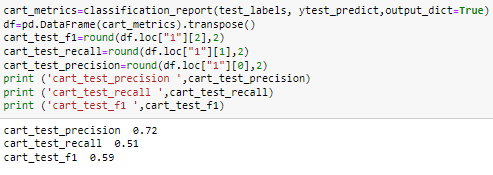
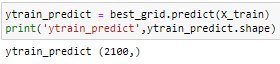


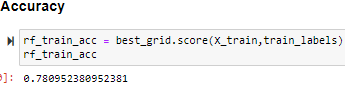
Fig 10.2

**MODEL 2 PREDICTION RANDOM FOREST**

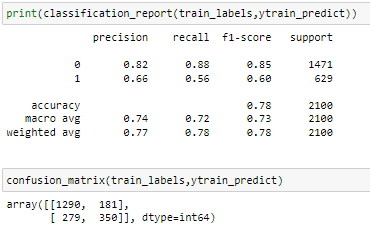
**Predicting on Training Dataset for Random Forest**



**Accuracy**

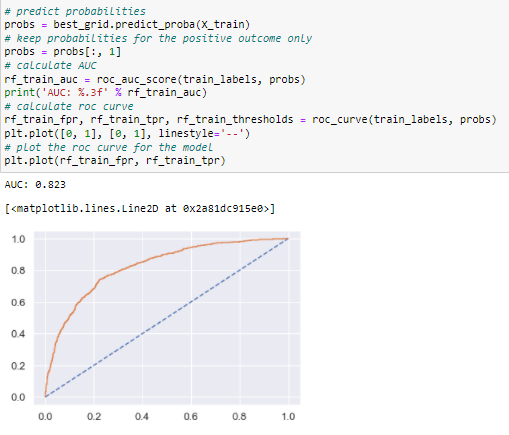


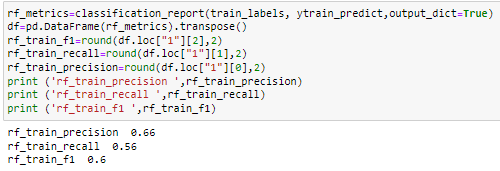
**Confusion Matrix**



**Model Evaluation for Random Forest**

**AUC and ROC for the training data for Random Forest**





Fig(10.3)

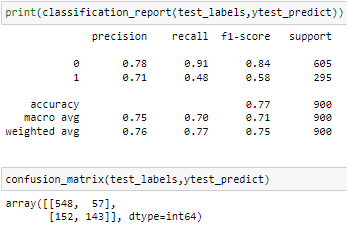
**Predicting on Test Dataset for Random Forest**



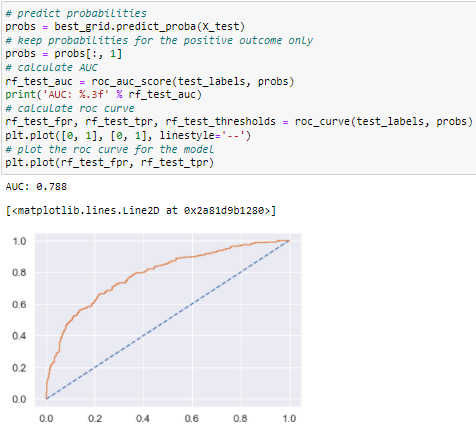
**Accuracy**



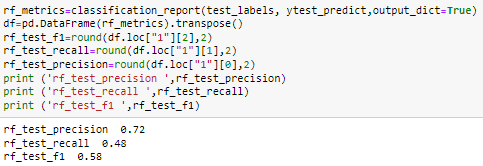
**Confusion Matrix**



**AUC & ROC for the test data for Random Forest**



Fig(10.4)

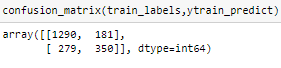


**Model 3**

**ANN**

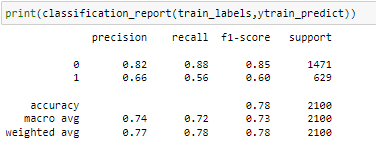


**Confusion Matrix**



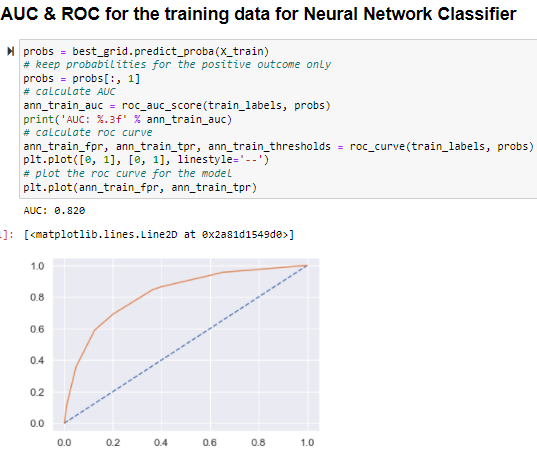
**Accuracy**





**Model Evaluation for Neural Network Classifier**

**AUC & ROC for the training data for Neural Network Classifier**



Fig(10.5)

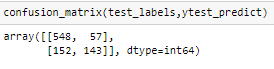
**Predicting on Test dataset for Neural Network Classifier**



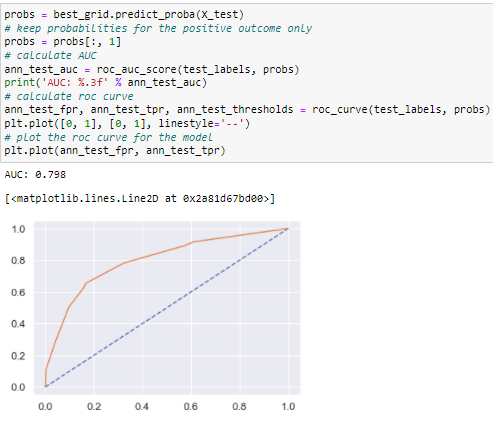
**Accuracy**



**Confusion Matrix**

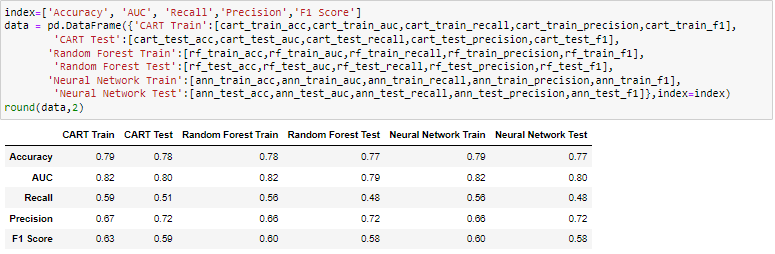


**AUC & ROC for the Test data for Neural Network Classifier**



Fig(10.6)

**2.4 Final Model: Compare all the model and write an inference which model is best/optimized.**



**CONCLUSION:**

I am selecting the CART model, as it has better accuracy, precision, recall, and f1 score better than other two RF & NN.

**2.6. Inference: Based on the whole Analysis, what are the business insights and recommendations?**

Looking at the model, more data will help us understand and predict models better. Streamlining online experiences benefitted customers, leading to an increase in conversions, which subsequently raised profits.

* As per the data 90% of insurance is done by online channel.
* Other interesting fact, is almost all the offline business has a claimed associated
* Need to train the JZI agency resources to pick up sales as they are in bottom, need to run promotional marketing campaign or evaluate if we need to tie up with alternate agency
* Also based on the model we are getting 80%accuracy, so we need customer books airline tickets or plans, cross sell the insurance based on the claim data pattern.
* Other interesting fact is more sales happen via Agency than Airlines and the trend shows the claim are processed more at Airline. So we may need to deep dive into the process to understand the workflow and why? Key performance indicators (KPI) The KPI’s of insurance claims are
* Increase customer satisfaction which in fact will give more revenue
* Combat fraud transactions, deploy measures to avoid fraudulent transactions at earliest
* Optimize claims recovery method
* Reduce claim handling costs.