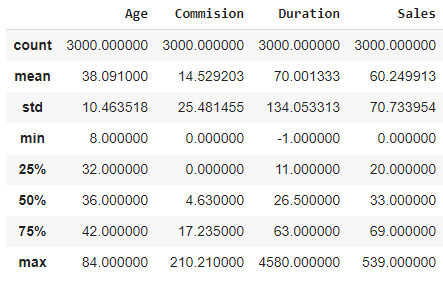
**Q2**

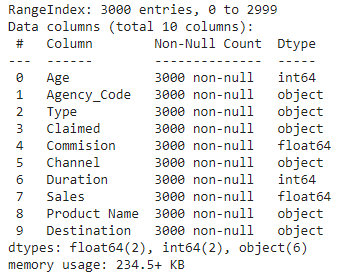
**Q 2.1**

The uni, bi and multi variate analysis are as follows:



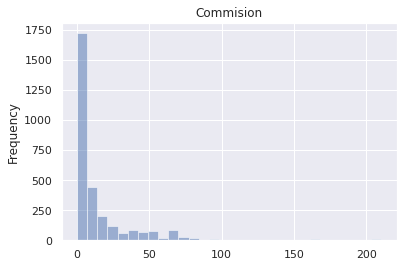
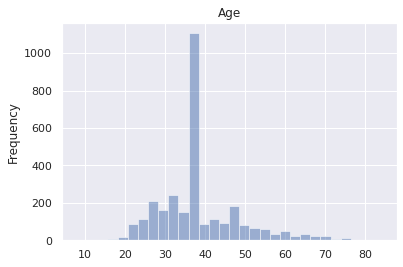
Above figures gives us the Statistics summary of the dataset. We see the duration has negative value (min=-1) which is not possible. This shows that few samples from the data needs treatment.

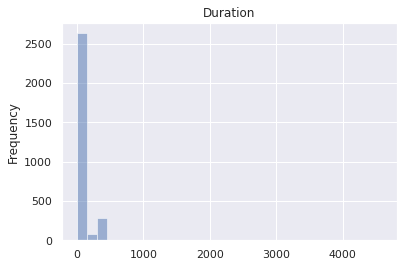
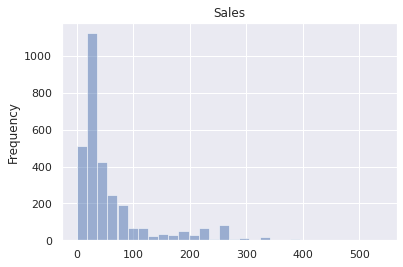
Next we will gain some information on the type of dataset and variables



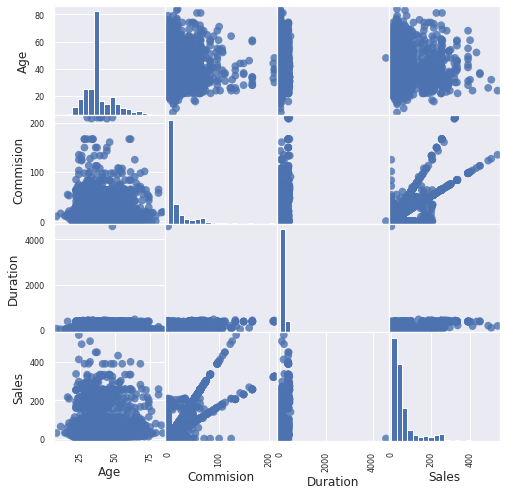
We see that all the data present are non-null and dtypes are int, object and float. Categorical values are also present here.

Histogram plot of all the numerical variables (single plot)

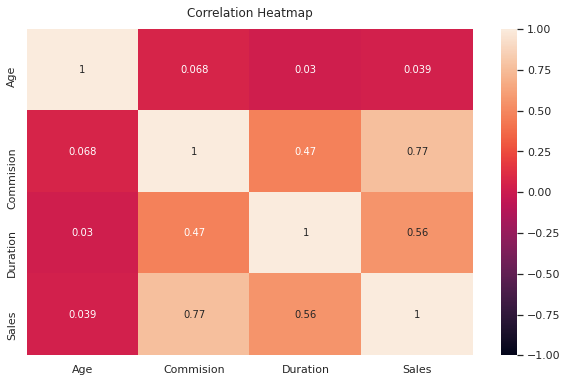


****

This pair plot for multivariate analysis shows us the complete relationship between different variables in terms of the distribution of the plot points.

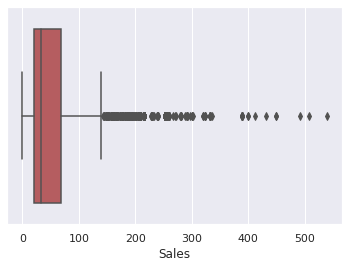
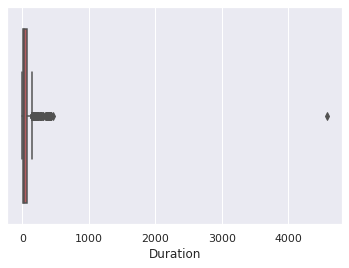
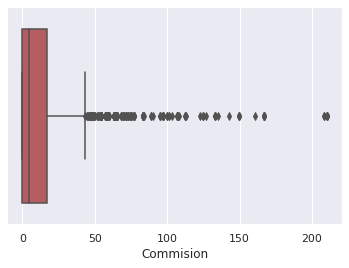
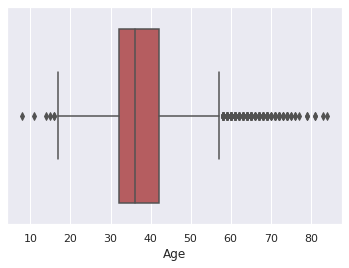


**Correlation Matrix**

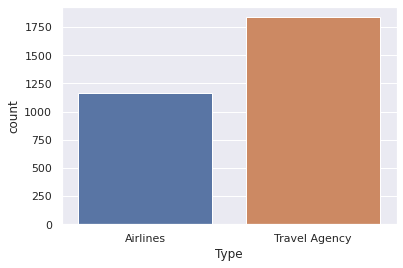
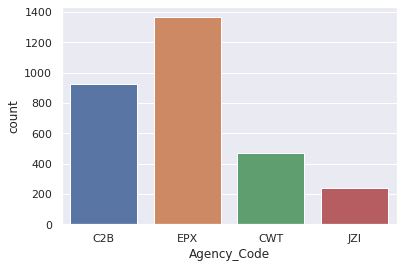
****

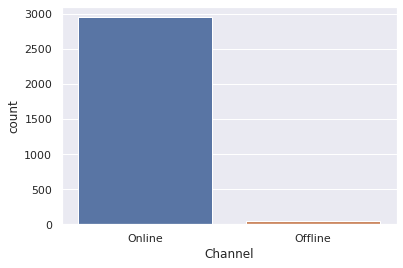
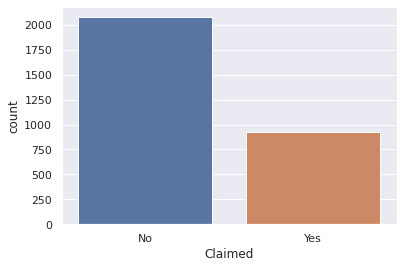
We see high correlation values between sales, commission and duration

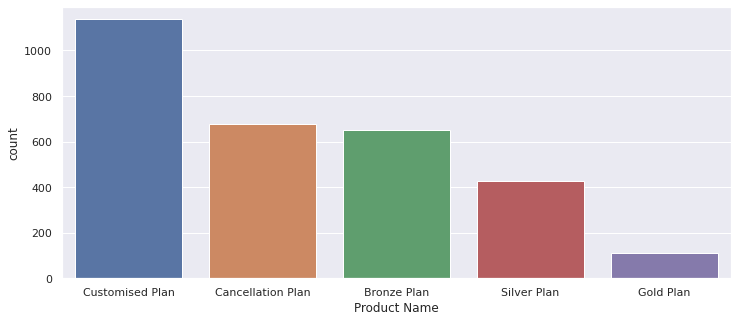
We do find many outliers in the variables (the black scatter points represents outliers before scaling)

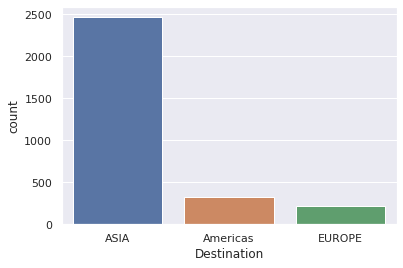


Now we plot the value count of the labels in categorical variables









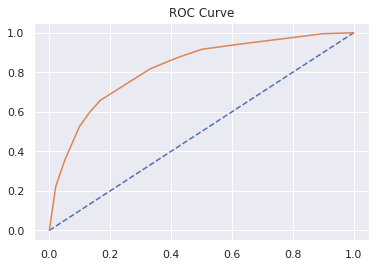
**Q 2.2**

Splitted data into training (70%) and test set (30%) and prepared the model!

**Q 2.3**

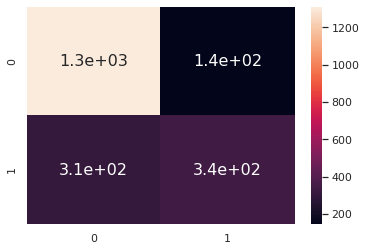
1. Accuracy, Confusion Matrix, Plot ROC curve, ROC\_AUC score and classification reports for **CART** on Train and Test dataset.
2. Train

AUC Train dataset: 0.823

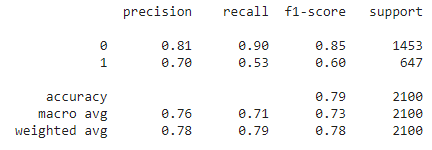


Train Accuracy: 0.785

Confusion matrix

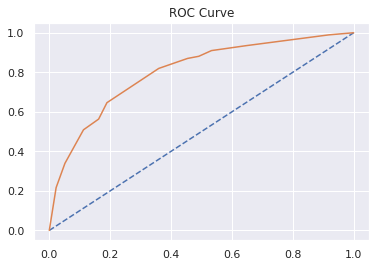


Classification report



1. Test

AUC Test dataset: 0.801

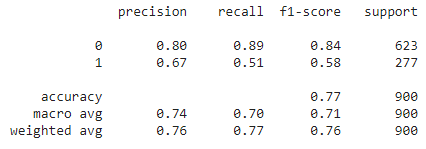


Test Accuracy: 0.771

Confusion Matrix



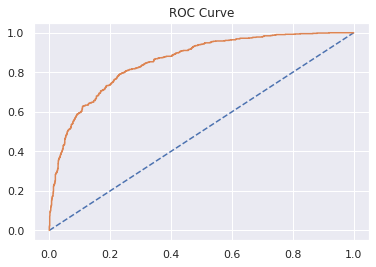
Classification Report



1. Accuracy, Confusion Matrix, Plot ROC curve, ROC\_AUC score and classification reports for **RF** on Train and Test dataset.

A.Train

AUC Train dataset: 0.856

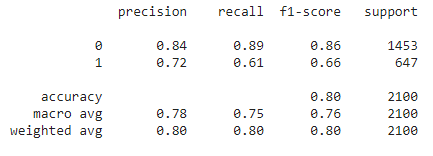


Train Accuracy: 0.804

Confusion Matrix

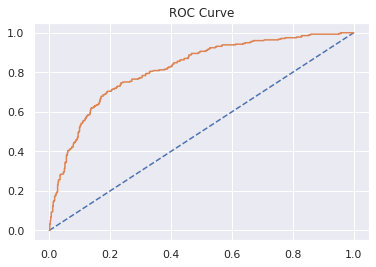


Classification Report



B.Test

AUC Test dataset: 0.818

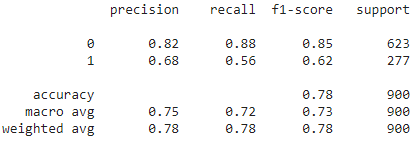


Test Accuracy: 0.784

Confusion Matrix



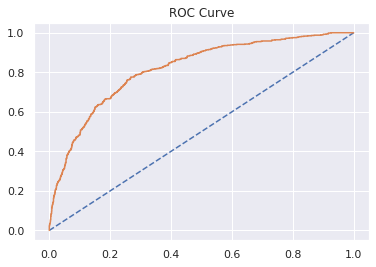
Classification Report



1. Accuracy, Confusion Matrix, Plot ROC curve, ROC\_AUC score and classification reports for **ANN** on Train and Test dataset.

A.Train

AUC Train dataset: 0.817

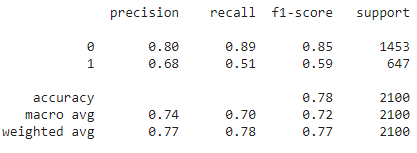


Train Accuracy: 0.776

Confusion Matrix

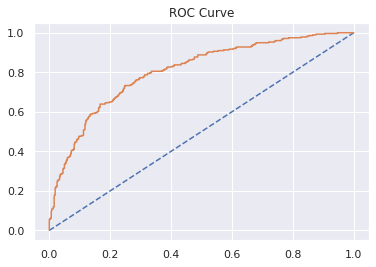


Classification Report



B.Test

AUC Test dataset: 0.804

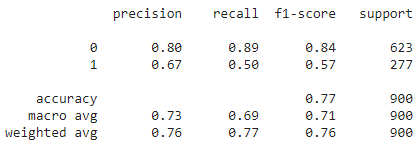


Test Accuracy: 0.769

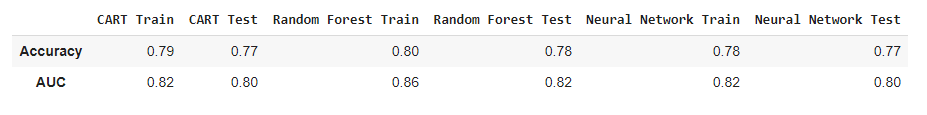
Confusion Matrix



Classification Report



**Q 2.4**



Therefore the best optimized model is **RF model**, as it has better accuracy and AUC than other two **CART & ANN**.

**Q 2.5**

Collect more real time data and past data if possible.

By looking at the insurance data and when we find correlation of different variables such as day of the incident, time, age group, etc, we can conclude that more data is required

• 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 and almost all the offline business has a claimed associated

• 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