

Data Cleaning was performed after understanding the business requirement after which we have performed univariate and bivariate analysis on categorical and non-categorical data

The first step was to split the dataset into features and target variable. The features were all the columns in the dataset except for the "Converted" column. The target variable was the "Converted" column.

The next step was to split the data into training and testing sets. The features of the training and testing sets using the StandardScaler class.

The training set was used to train the model, and the testing set was used to evaluate the model's performance.

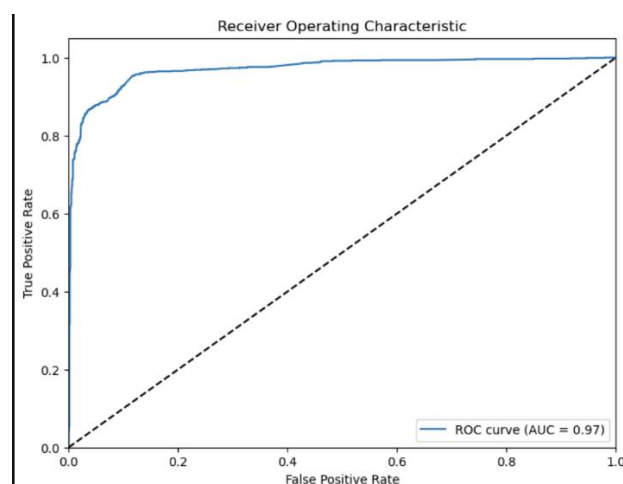
The next step was to build and train the logistic regression model. The logistic regression model is a supervised learning algorithm that can be used to predict binary outcomes. The model was trained using the training set.

The next step was to evaluate the model's performance on the testing set. The model's performance was evaluated using the following metrics:

- Accuracy
- Sensitivity
- Specificity

The accuracy of the model was 92.73%. This means that the model correctly predicted 92.73% of the instances in the testing set. The sensitivity of the model was 86.88%. This means that the model correctly predicted 86.88% of the instances that were actually churned. The specificity of the model was 96.39%. This means that the model correctly predicted 96.39% of the instances that were not actually churned.

The next step was to plot the ROC curve and AUC of the model. The ROC curve is a graphical representation of the model's performance. The AUC is the area under the ROC curve. The ROC curve and AUC of the model are shown below:



The AUC of the model is 0.97. This means that the model is very good at distinguishing between instances that will churn and instances that will not churn.

The final step was to evaluate the model's performance at different probability cutoffs. The probability cutoff is the threshold at which a prediction is made. For example, if the probability cutoff is 0.5, then any instance with a predicted probability of churn greater than or equal to 0.5 will be predicted to churn.

The model's performance at different probability cutoffs is shown below:

Probability Cutoff	Accuracy	Sensitivity	Specificity
0.2	90.95%	95.45%	88.08%
0.4	91.96%	88.39%	94.22%
0.6	92.35%	84.87%	97.11%
0.8	90.37%	77.81%	98.35%

The model's performance is best at the probability cutoff of 0.6. This is because the model has the highest accuracy and sensitivity at this probability cutoff.

Overall, the logistic regression model performed well on the dataset. The model was able to accurately predict 92.73% of the instances in the testing set. The model is also very good at distinguishing between instances that will churn and instances that will not churn.