Question12:

a). (5 points) How many observations are in the Training and the Testing partitions?

462 observations are in the Training partitions.
155 observations are in the Testing partitions.

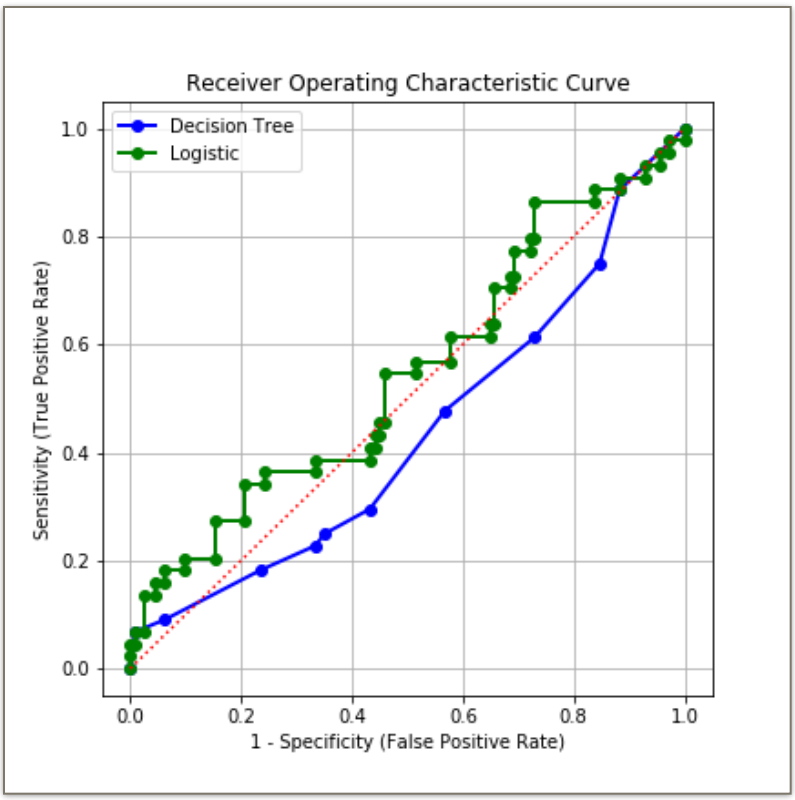
b). (5 points) What is the claim rate in the Training partition?

The claim rate in the Training partition is 0.2879.

c). (5 points) Use the claim rate in the Training partition as the probability threshold in the misclassification rate calculation. A claim is predicted if the predicted probability of filing a claim is greater than or equal to the probability threshold. Calculate the Area Under Curve metric, the Root Average Squared Error metric, and the Misclassification Rate for both models using the Testing partition. Present your results in a table, list the metrics in the column dimension and the models in the row dimension.

	Area Under Curve	Root Average Squared Error	Misclassification Rate
Classification Tree Model	0.4351	0.5089	0.4581
Logistic Model	0.5463	0.4539	0.4839

d). (5 points) Calculate (but no need to display) the coordinates of the Receiver Operating Characteristic curve for both models using the Testing partition. Plot both ROC curves in the same chart but uses a different color for each curve. The chart (including the axes, the title, and the curve legends) must be properly labeled.



e). (5 points) Based on the evaluation and the comparison results in (c) and (d), recommend your model. You must state your reasons for your recommendation.

We prefer a model with higher Area Under Curve metric, lower Root Average Squared Error metric and lower Misclassification Rate.

Compare two models, the Area Under Curve metric of the logistic model is 0.5463 which is higher than the Area Under Curve metric of the classification tree model: 0.4351.

And the Root Average Squared Error metric of the logistic model: 0.4539 is lower than the Root Average Squared Error metric of the classification tree model: 0.5089.

Even though the Misclassification Rate of the logistic model: 0.4839 is a little bit higher than the Misclassification Rate of the classification tree model: 0.4581, there is no significant difference.

Based on the three metrics, I recommend Logistic Model.