Questions: Compare and contrast the output of the Logistic Regression Model and the CART Model. Pick one model that you would use to make predictions. Why do you select this model? What performance metric is most important to you? How do you make the choice between higher specificity vs higher sensitivity? Which is more important for this business application?

How we figure out each model: We tried to figure out the optimized threshold (threshold = 0.62) for logistic model to get the highest accuracy, since in the case, they want a more accurate model, didn't talk about information of buget limit. We get the best leaf note = 8 for the tree model in the same purpose.

The model we choose and why: Since the AUC for Decision Tree model (0.8209) is higher than the AUC for the logistic regression model (0.8016), we prefer to use the tree model with 8 leaf nodes.

What performance metric is most important to you?: Accuracy and AUC (ROC Curve)

Specificity vs sensitivity: The choice between higher specificity and higher sensitivity in a predictive model depends on the company's budget and target. If the goal is to maximize revenue by retaining as many schools as possible in 2012, prioritizing higher sensitivity is a better strategy. This ensures accurate identification of schools likely to participate again, enabling targeted retention strategies. Conversely, if resources are limited and prioritization is needed, higher specificity becomes more important. This ensures accurate identification of schools unlikely to participate, allowing focused resource allocation for retaining schools with the highest potential for participation.

Which is more important for this business application?:

In the context of the Scholastic Travel Company (STC), the decision tree models aim at enhancing the prediction accuracy for customer retention, specifically whether a group will return the next year. Achieving higher sensitivity in the model is crucial for this objective, as it implies a better ability to correctly identify true positives, and increase revenue, i.e., groups that will indeed rebook.

By accurately predicting customer rebooking behavior, these models directly contribute to cost reduction in several ways:

Targeted Marketing: By identifying characteristics of groups most likely to return, STC can design more focused marketing campaigns, allocating resources to prospects with the highest probability of rebooking, thus saving on broader, less effective marketing efforts.

Resource Optimization: Understanding which groups are more likely to return allows for more efficient allocation of resources, both in terms of customer service and trip planning, ensuring that efforts are concentrated on maintaining relationships with high-potential groups.

Enhanced Customer Experience: By predicting the likelihood of return, STC can tailor the travel experience to meet the needs and preferences of repeat groups, potentially increasing customer satisfaction and loyalty, which in turn can lead to word-of-mouth referrals and reduced acquisition costs.

Strategic Planning: Accurate predictions assist in forecasting revenue and making informed decisions regarding capacity, staffing, and other operational aspects, leading to more efficient management and reduced overheads.