Assignment 6 - Classification

Lacey Hartigan

Due 7/20/21

"When life gives you lemons, don't make lemonade. Make life take the lemons back! Get mad!" - Cave Johnson

For this assignment, you'll be using the lemons dataset, which is a subset of the dataset used for a Kaggle competition described here: https://www.kaggle.com/c/DontGetKicked/data. Your job is to predict which cars are most likely to be lemons. Note: you'll need to import both the training and testing datasets (and identify when to use which for the steps below).

Complete the following steps.

1. Calculate the proportion of lemons in the training dataset using the `IsBadBuy` variable.

2. Calculate the proportion of lemons by Make.

3. Now, predict the probability of being a lemon using a logistic regression, using covariates of your choosing.

4. Make predictions from the logit model. Make sure these are probabilities.

5. Calculate the accuracy, sensitivity and specificity of your model using a threshold of .5.

6. Calculate the AUC for the predictions from the ROC based on the logit model.