

Assignment 8

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.

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 linear model (`lm(y~x)`, with covariates of your choosing from the training dataset).
4. Make predictions from the linear model.
5. Now, predict the probability of being a lemon using a logistic regression (`glm(y~x,family=binomial(link="logit"))`), again using covariates of your choosing.
6. Make predictions from the logit model. Make sure these are probabilities.
7. Create a confusion matrix from your linear model and your logit model.