Assignment Module 8 + 9

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.

8. Plot the probability of a car being a bad buy by make.

9. Create a table that shows the probability of a car being a bad buy by make.

Bonus: Create a heatmap of the probability of a car being a bad buy by make and acquisition type.

Bonus: Create a decision tree to predict whether a car is a bad buy. Compare the results with logistic model.