Exercise 05

Predicting used car prices some more

We'll be still be using the train.csv data set for this lab. The data set covers the characteristics and prices for used cars sold in India. We are interested in predicting the price of a car given some characteristics. We will attempt to build a linear regression model of Price. This time we are going to work on filling in the missing data that we previously dropped.

1. Transform Price so that it looks more normal
2. How many values are missing for Power and Engine?
3. Which column has the most missing values and what should we do about it?
4. Build a model of transformed price based on Power, Engine, Kilometers Driven, and Year, how much variance is explained?
5. How many rows were used to train the model?
6. Fill the missing values in Power and Engine with their respective means and rebuild the model. Now how much variance is explained?
7. How many rows were used to train the model?
8. Impute the missing data using MICE and rebuild the model
9. How have the parameter estimates changed from step 4?
10. Plot the distribution of Power with and without MICE
11. Plot the distribution of Engine with and without MICE