Jason Melnik

10/09/2021

Homework 6

* 1. This graph helps show us that most cars are within the price range of 6000 to 14000. The most common price range is 8000 to 10000, the lowest price range is 0 to 5000, and the highest price range is 30000 to 35000. This is a symmetric graph due to the price distribution.
  2. It appears that the higher the age the lower the price is for a corolla. This will help with linear regression since there is a strong correlation between how age affects price.
  3. It seems like fuel type does not affect price due to the averages being around the same price range. Fuel type will have a small impact in determining price since they are around the same range with some outliers in diesel.
  4. The regression model seems to be doing a decent job but being too high for corolla’s older than 30 but being to low for corollas age 35 and younger.
  5. It’s not surprising since the data doesn’t include data of cars that are 120 months old thus will have an inaccurate guess of what the price of the car will be. It predicted a negative price because of the linear regression line that is descending and goes negative in the 120-month time frame.
  6. The 95% prediction interval for a 12-month-old corolla would be in the range of $14,816 to $21,660. So, if we found a corolla that costs $13,000 that would be surprising since it is not in the 95% predicted range making the $13,000 corolla within the 5% error.
  7. The value of the RMSE for this regression is $1768.
  8. The new predictor with Age\_sqrt seems to have a much more linear relationship to price with Price than Age without being square rooted.
  9. I would trust this prediction more than the prediction from the original regression since the range is smaller than the original range. The difference of the original range is 6844 while the difference of the new range is 6353 which means that there is a higher change that the predicted value is more accurate than the old regression. The new range is between $16,194 to $22,547 while the old was between $14,816 to $21,660 making the new more accurate.
  10. The RMSE of the new regression “Age\_sqrt” is 1603 which is lower than the old regression “Age” of 1768 making the new regression a better predictor.
  11. The regression equation we obtain is y = -.05584x + 14577. With the slop being -.05584 and y-intercept 14577.
  12. The prediction interval is between $3129 to $14857 which tells us basically nothing since that is a large range of prices. Since the most common price for the corolla car is between 6000 to 14000 this prediction doesn’t help us in reality.
  13. The RMSE of the KM linear regression is $2967 which is much worse than the “Age” and “Age\_sqrt” RMSE. This regression is less reliant than our Age regression prediction.
  14. I would say that the square root KM graph is roughly the same as the original KM linear regression graph.
  15. The linear regression function is y = -31.13x + 18565.41
  16. The RMSE for square root KM is $2729 which makes the square root more successful than the original KM.