Used Cars Price Prediction

By: Matthew Leiser, Samantha Lemley, Ryoichi Nakayama, and Shahla Ahmed

Topic and Reason

- the automotive industry has faced a shortage in the Semiconductor Integrate Chips globally
- Semicoductor IC is a critical component for controlling several electronic devices in the vehicle
- car sales industry is making up for the shortage by raising their APR and prices
- increasing demand of used cars which is making the prices of used cars higher as well
- limit our studies and findings for US market only

Data Sources

- https://www.kaggle.com/code/maciejautuch/car-price-prediction/data

- dataset collected in Kaggle is mainly from craiglist.org (used item selling website) from all over US
- cars are from different manufacturers and of different years
- Price as our target variable and rest will pass as features
- dropping off null values and some columns that are not needed as they dont impact the price of the used cars much

Questions to Answer

1. How does the mileage affect the price of the used car?
2. How does size of the car impact the price?
3. How does the age of the car, condition and fuel type affect the price of the car?
4. Will this affect the overall demand for a used car in place of a new car for consumers?

Data Analysis & Exploration

- creating box plots, subplots and graphs
- Oregon had an average price of car at 1.6 million and California had an average price 139,000. Based on what we know about markets, California's average price should be way higher than Oregon's
- actual prices and the predicted prices were shown through the scatterplot
- tableau to get a better understanding of the prices and regions

What we used

- Decision Tree Regressor
- Linear Regression Model
- Lasso Model
- Data Cleaning: that includes getting rid of all undesired columns
- Creating table in SQL
- Utilize Tableau
- Machine Learning Model: this includes choosing X variable as a collection of features and Y as a target variable which will be Price
- Model for Regression
- Linear Regression Logistic Regression K&N Algorithm Decission Tree etc