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
- Different manufacturers and years
- Price as our target variable and rest will pass as features

Questions to Answer

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

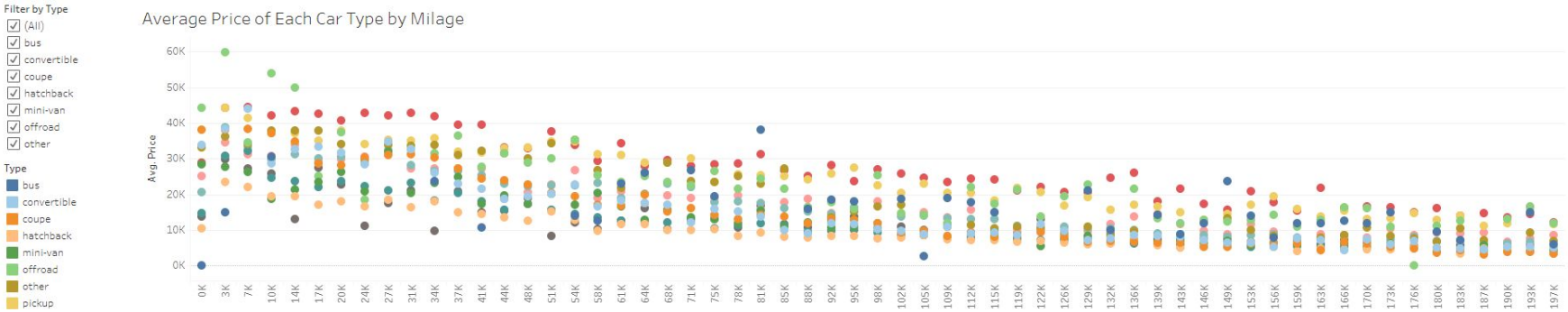
Data Analysis & Exploration

- Creating box plots, subplots and graphs
- Diesel was significantly more than other fuels
- Tableau to get a better understanding of the prices and regions

What we used

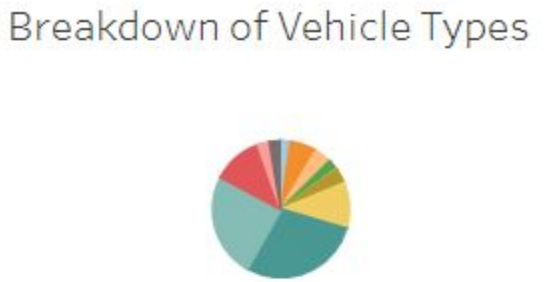
- Decision Tree Regressor
- Linear Regression Model
- Data Cleaning: that includes getting rid of all undesired columns
- Creating table in SQL
- Utilize Tableau
- Machine Learning Model: this includes choosing X variable as a collection of features and Y as a target variable which will be Price
- Model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

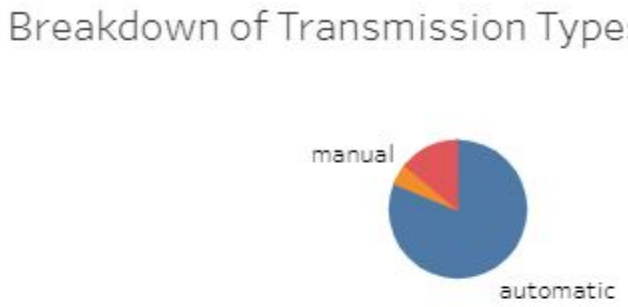
Dashboard

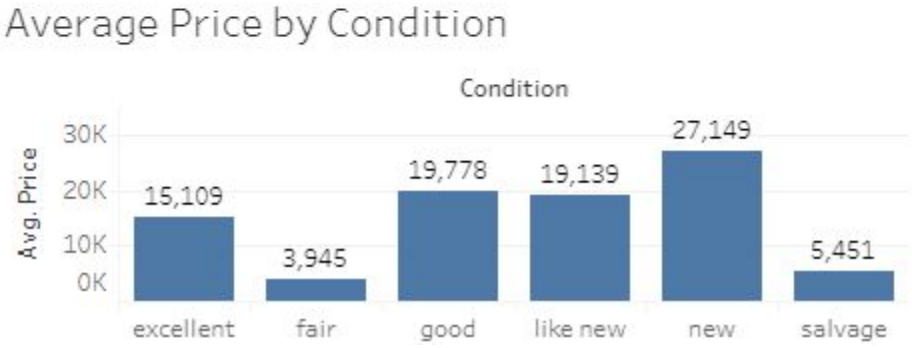


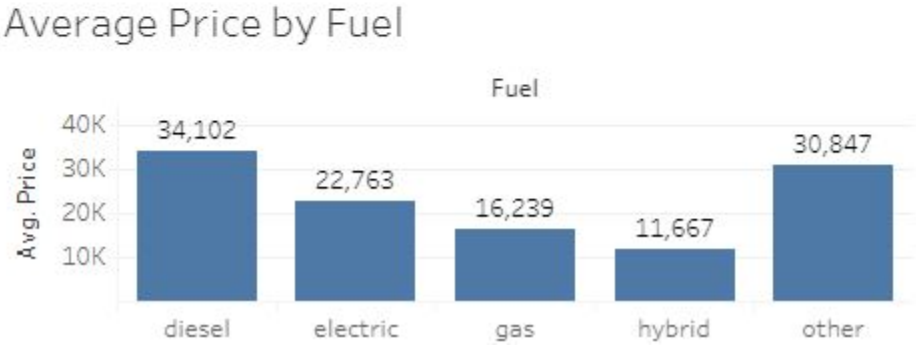
https://public.tableau.com/app/profile/matt.leiser/viz/CarPricesDashboard_16680540059890/Dashboard1?publish=yes

Breakdown of Types

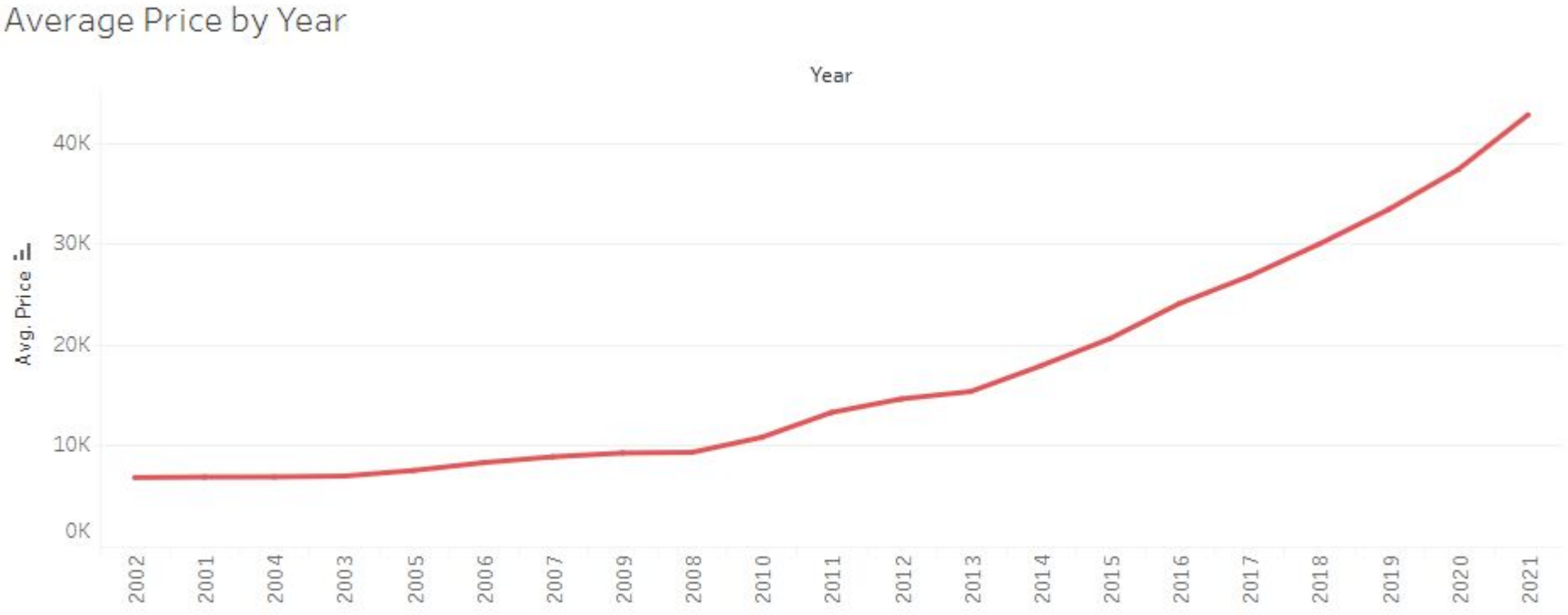








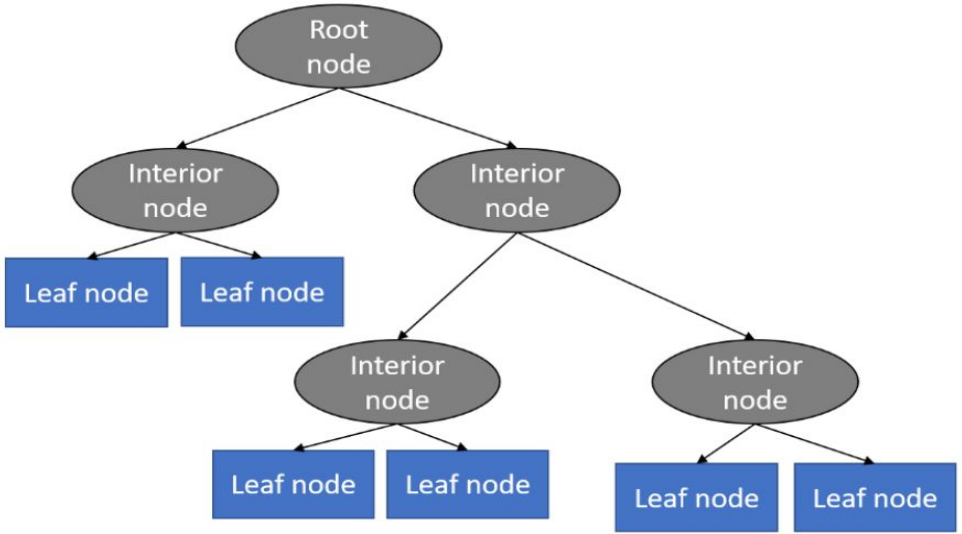
Average Price by Year



Condition vs. Price

Machine Learning Model

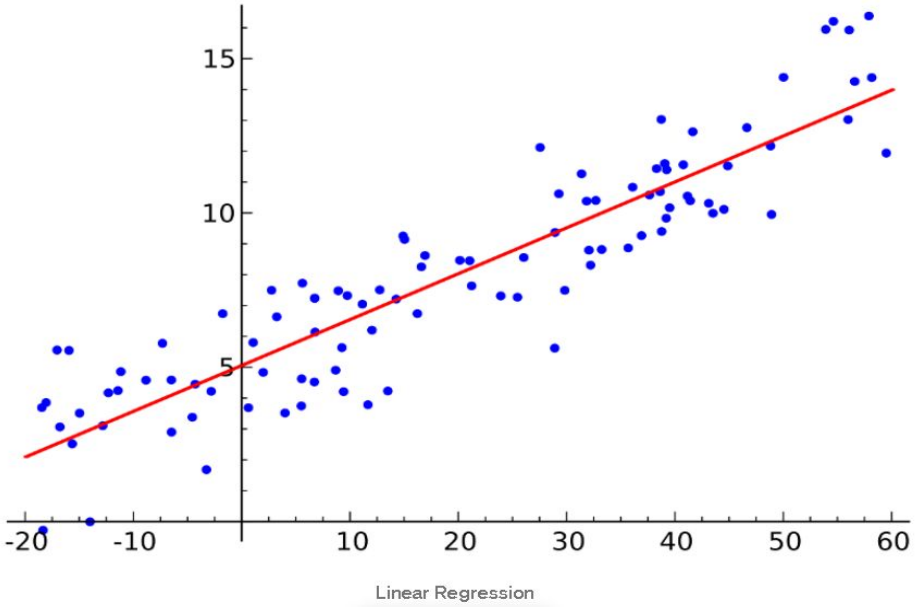
- Decision Tree Regressor
- Benefits/Limitations



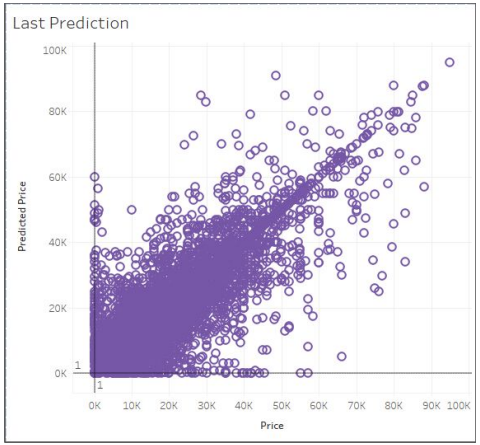
Source

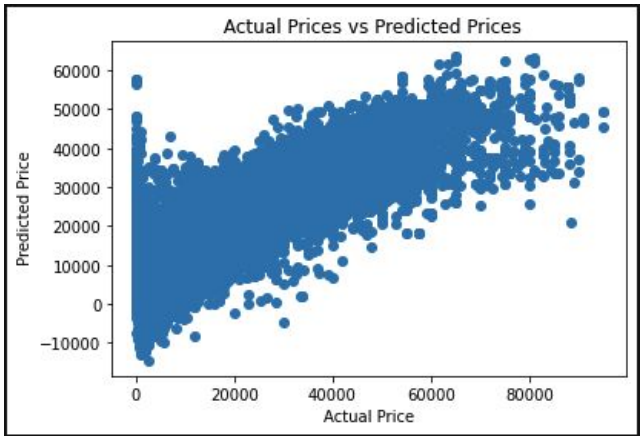
Linear Regression

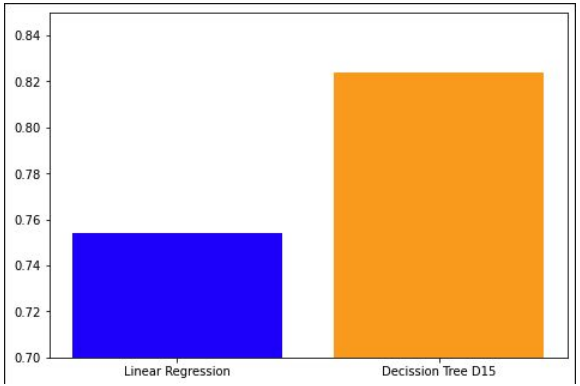
- Advantages/Disadvantages



Results







R^2 train: 0.882, test: 0.824

Recommendations/Final Thoughts

Questions?