Final Project Presentation

By: Juan Morales
11/21/2024

Problem Statement:

- I will use a database of car sales to identify which cars provide the best value for the money based on the most considered specifications of a car (horsepower and miles per gallon).
- I will also train a model to predict the price of the car based on its specifications.

Why is this problem important:

- Buying a car is a big investment, and the buyer wants to get the most value out of their purchase.
- Predicting the most valuable car will help the buyer to make an informed decision on which car to buy.

Dataset Overview:

Link to dataset:
Car Features and MSRP

Description:

> A car feature and MSRP dataset that includes features like make, model, year, horsepower, number of cylinder, and MPG.

Source: Kaggle

Data Size: 11,915

Key Features used: Make, Model, Horsepower, Price, and MPG

Total Features: 16

Values Removed: Engine Fuel Type, Popularity, Number of Doors, Drive type, Vehicle Size, Vehicle Style, and year.

Data cleaning:

- Dropped features that I will not use.
- Renamed features for ease of use.
- Removed duplicate rows.
- Dropped all missing values.
- Dropped any misleading information:
  - Horsepower at 0.
  - Price at 0.
- Focused on new cars.

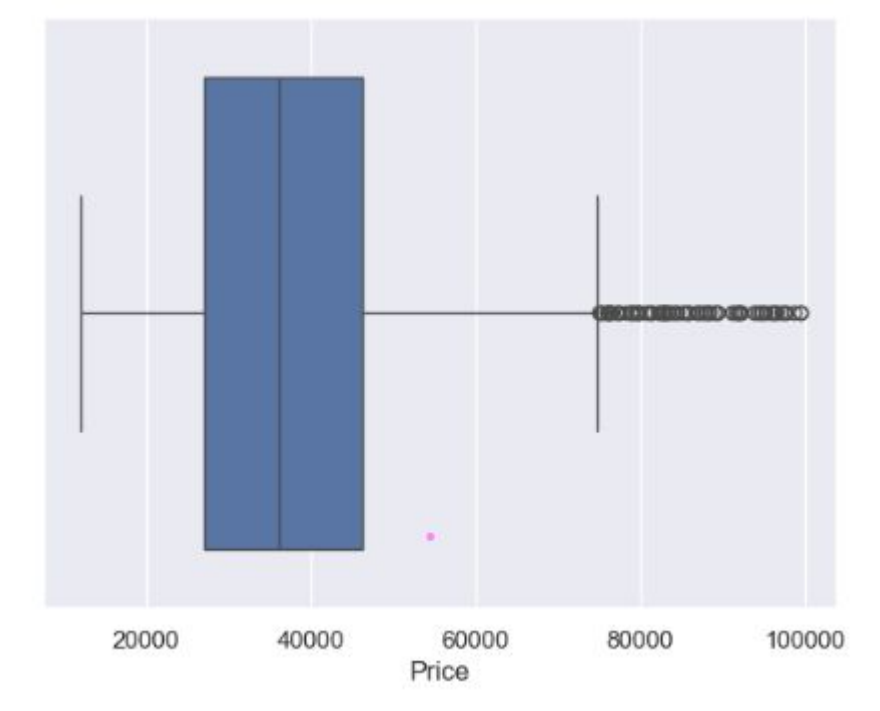
Tools used:

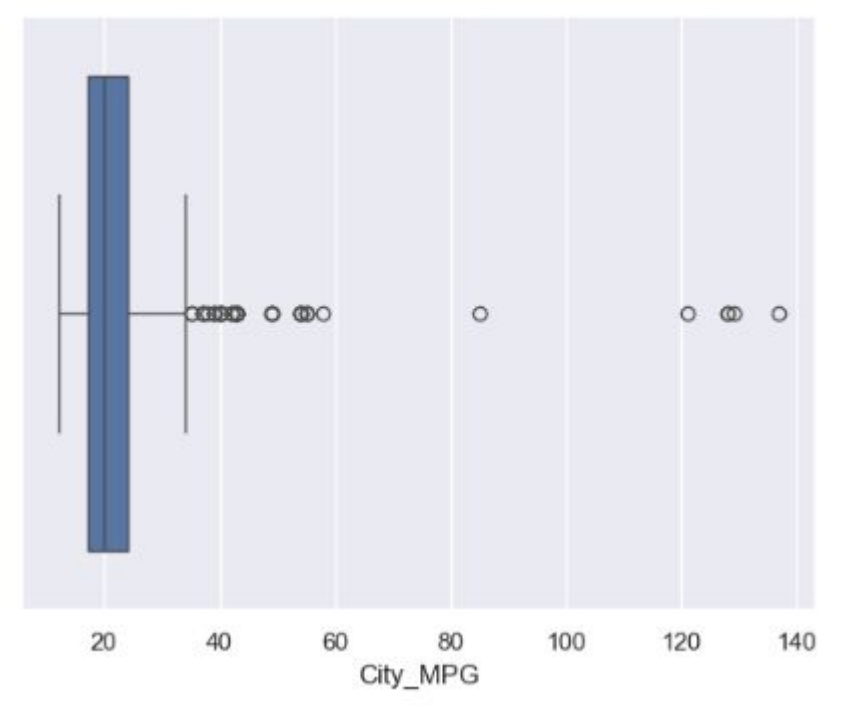
- Pandas for data handling
- Matplotlib and seaborn for visualizations.
- Scikit-learn for algorithms.

Data set info

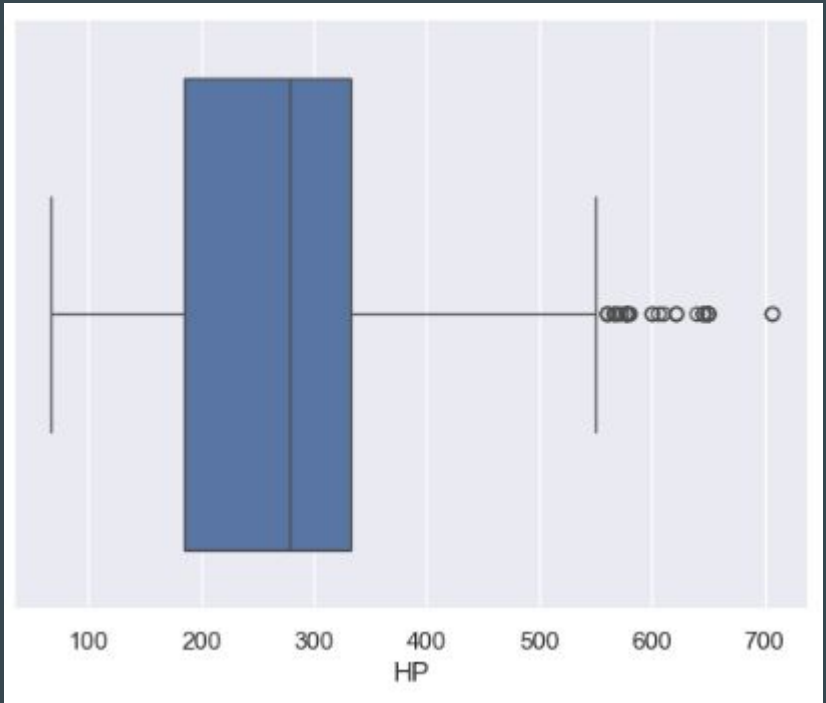
	Year	HP	Cylinders	Highway_MPG	City_MPG	Price
count	1,608.00	1,608.00	1,608.00	1,608.00	1,608.00	1,608.00
mean	2,017.00	275.02	5.43	28.28	21.19	42,196.06
std	0.00	105.57	1.59	7.68	8.64	24,206.36
min	2,017.00	66.00	0.00	17.00	12.00	11,990.00
25%	2,017.00	185.00	4.00	23.00	17.00	27,390.00
50%	2,017.00	278.00	6.00	27.00	20.00	36,675.00
75%	2,017.00	333.00	6.00	32.00	24.00	47,820.00
max	2,017.00	707.00	12.00	111.00	137.00	247,900.00

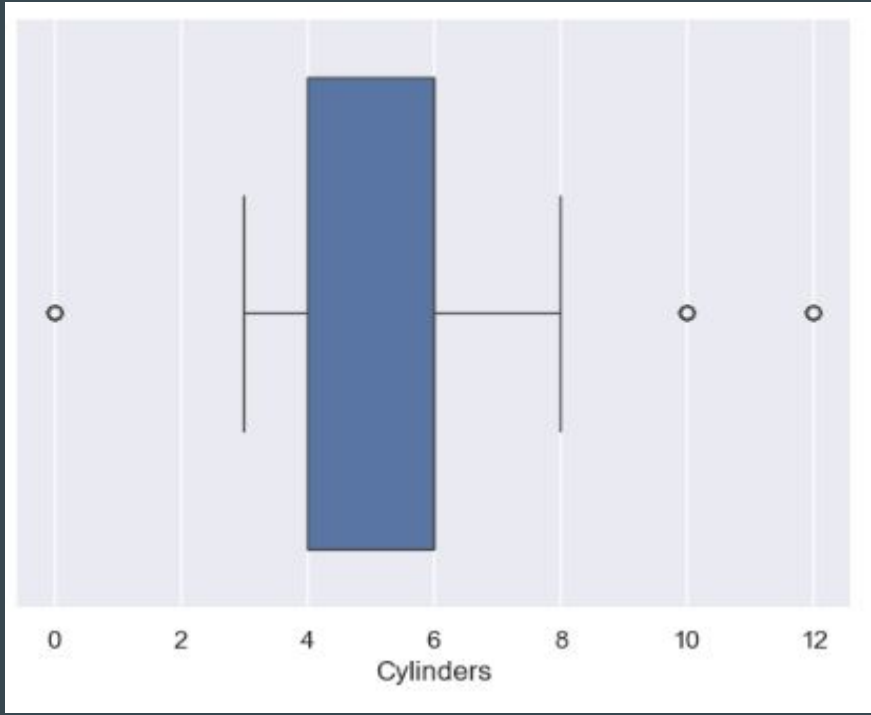
Data Box Plot



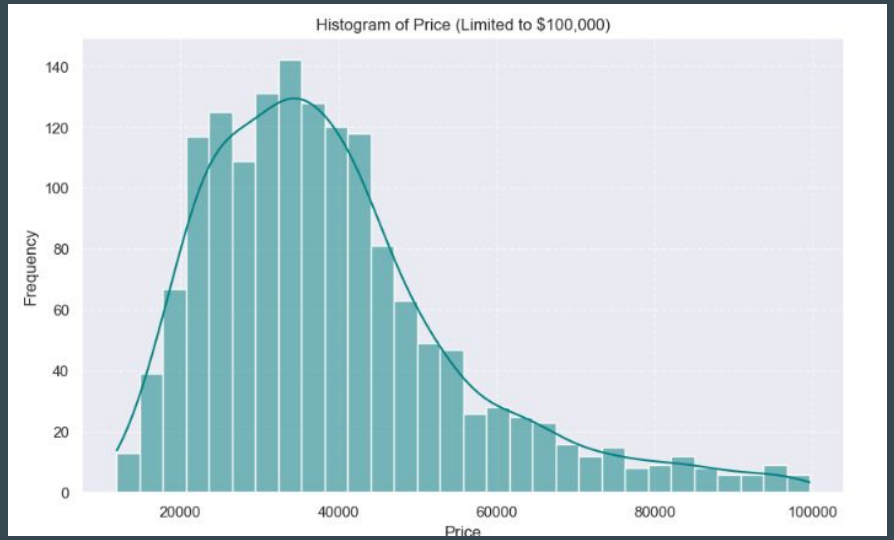


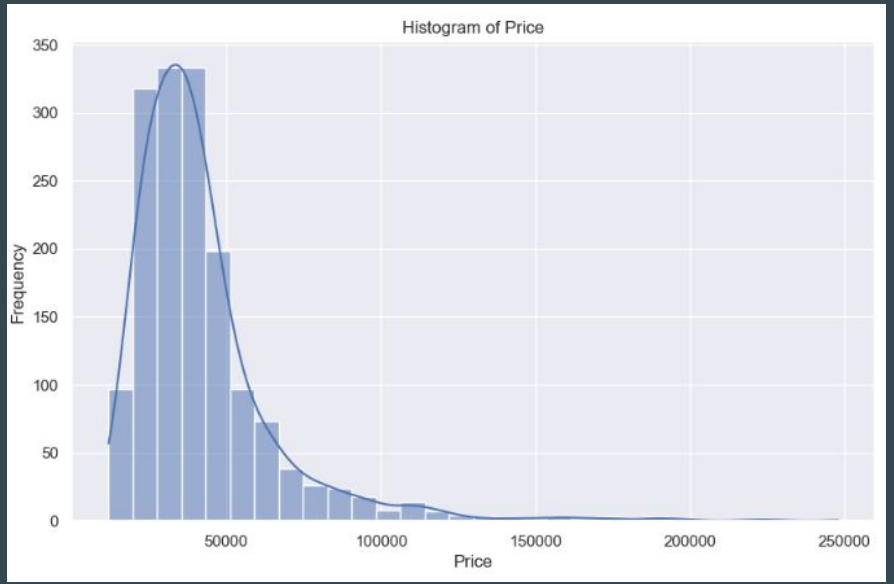
Data Box Plot



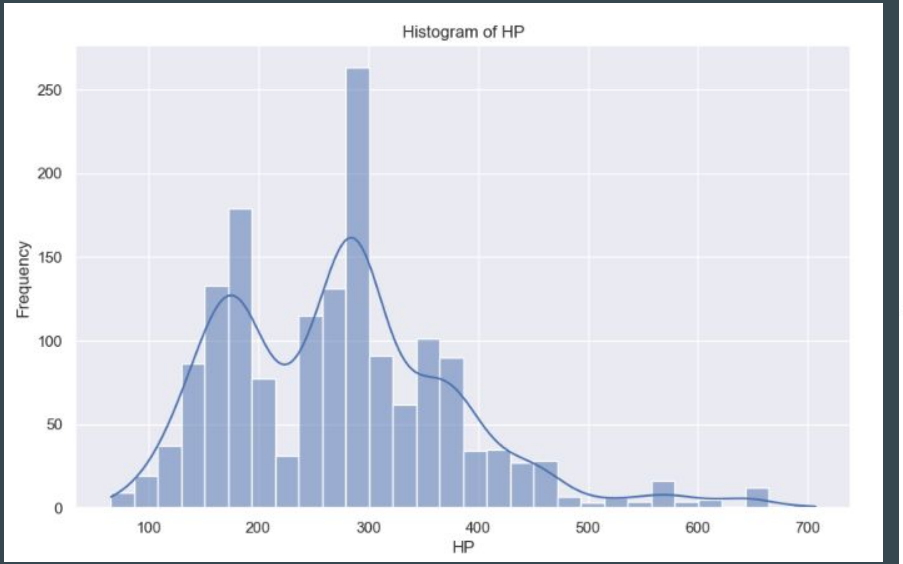


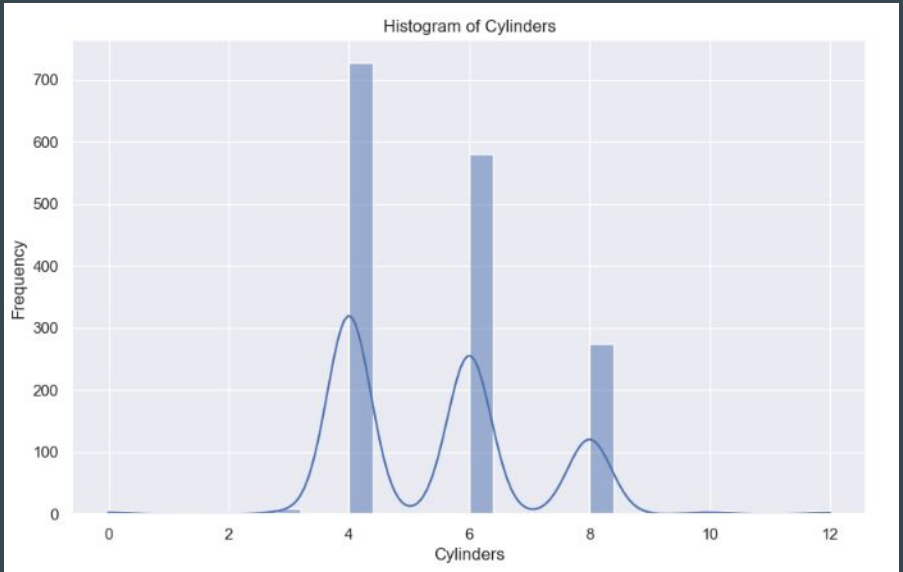
Histogram Plot



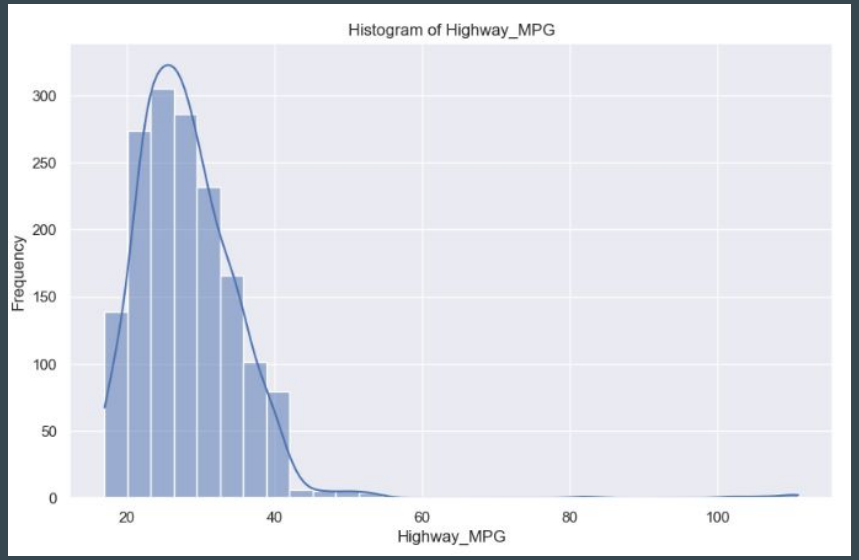


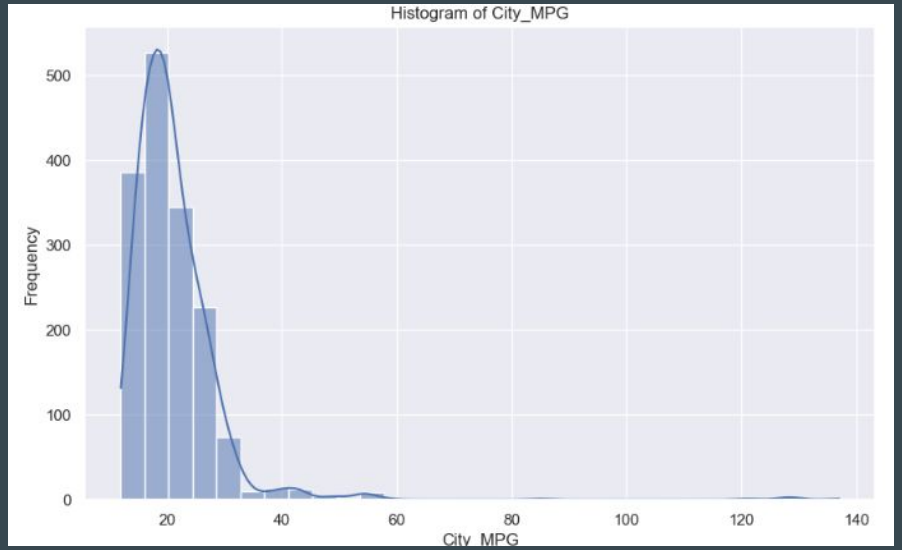
Histogram Plot





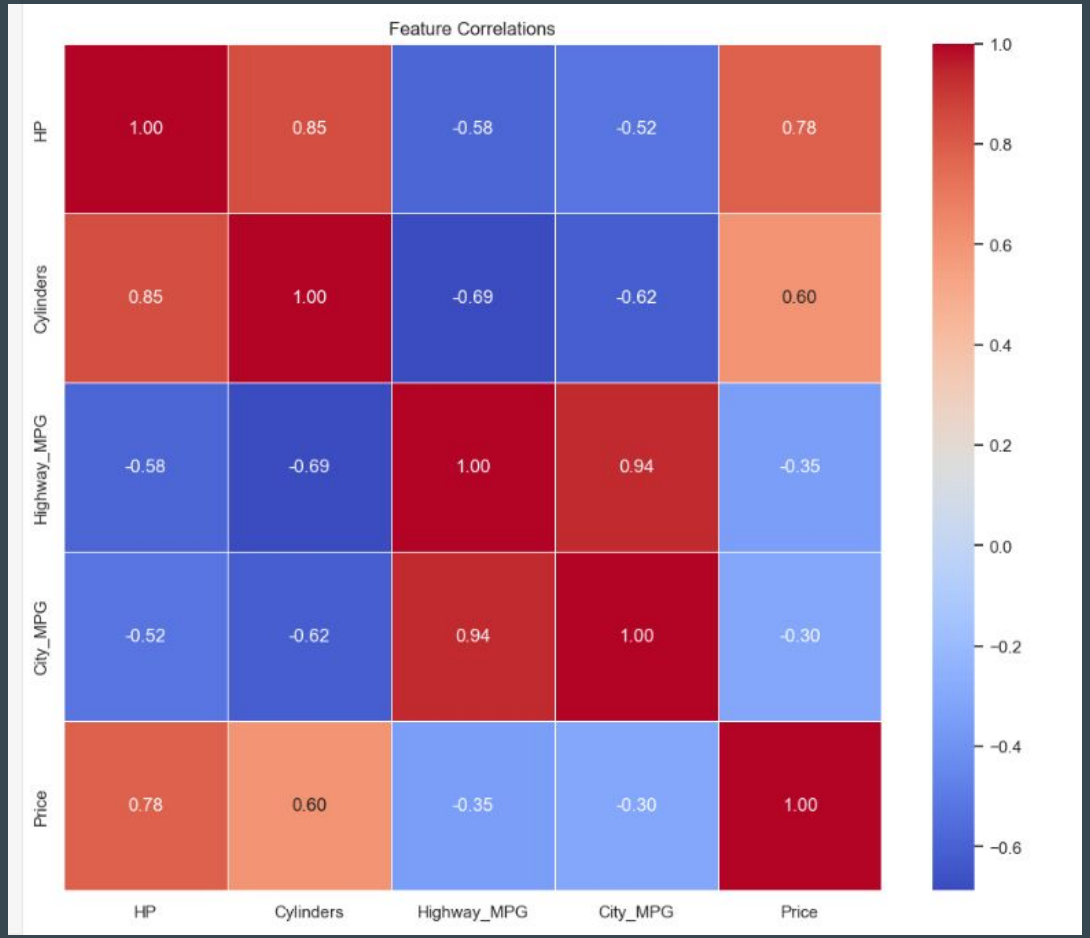
Histogram Plot



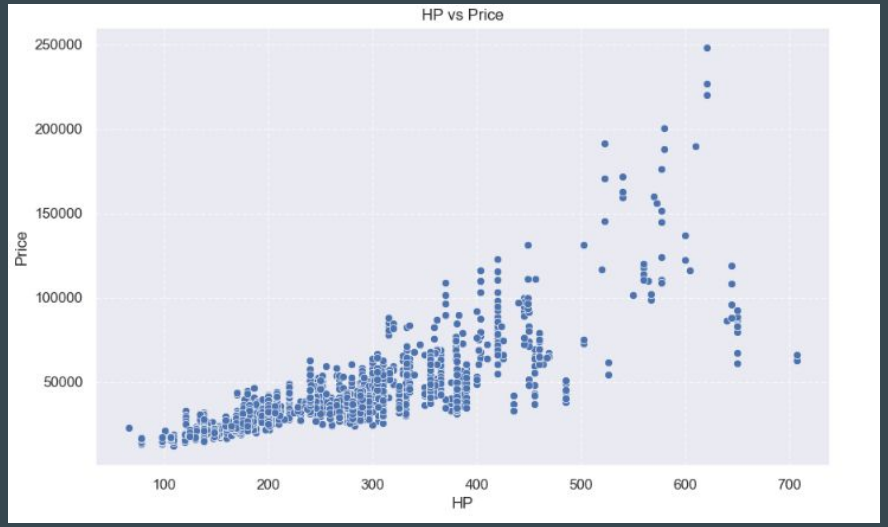


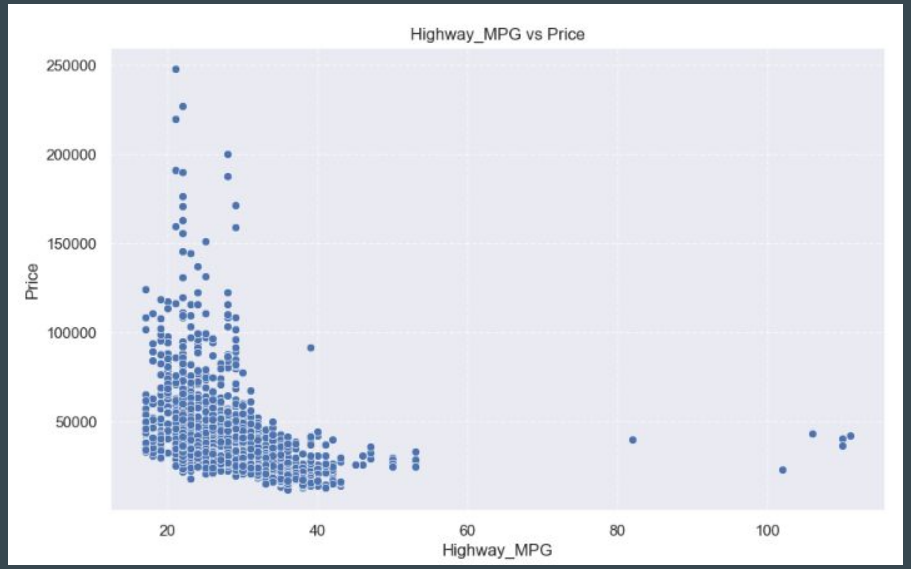
Heat map:

- Notice the correlation between the price, hp, and the amount of cylinders.
- The mpg suffers with more hp and cylinders.
- The price is slightly lower for cars with higher mpg overall.



Scatter plot:





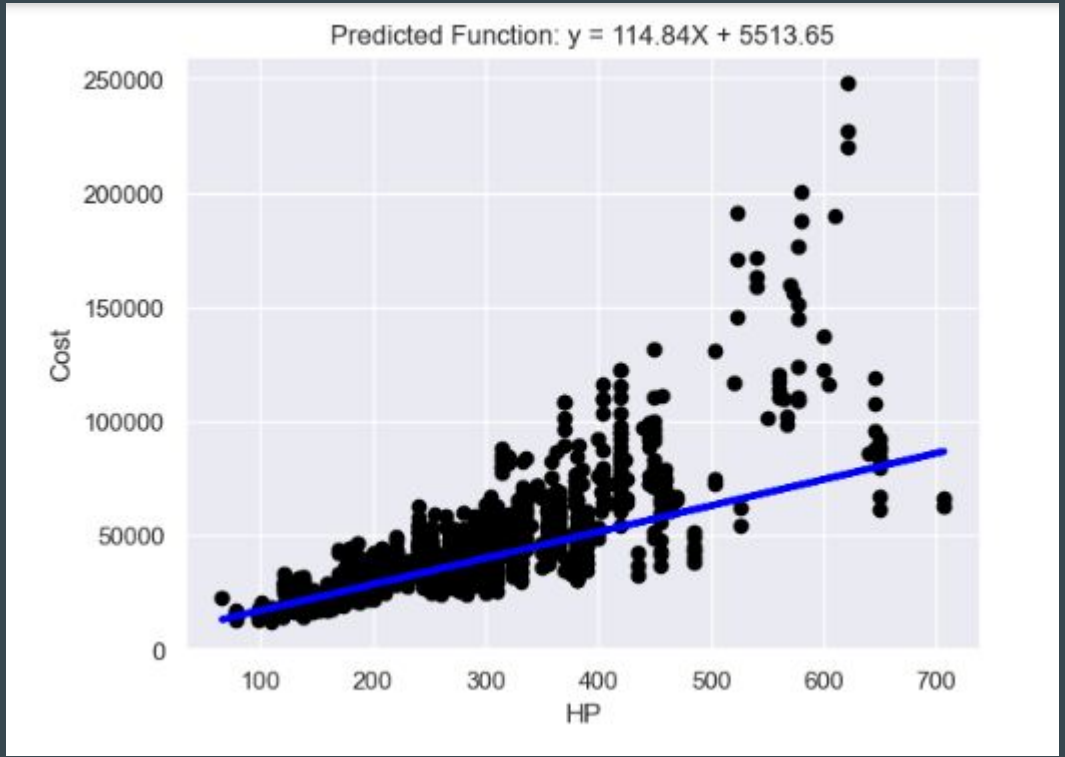
What does this information tell us:

- A higher price usually comes with more power.
- A higher price does not means higher miles per gallon.
- Most of the cars that reach over 30mpg are under 50k.
- Most of the cars under 50k are less than 400hp.
- This means that the car that has the most value will be under 50k and it will have less than 400hp.
- This also shows us that the buyers with more purchasing power do not care to much about mpg and worry more about hp. Therefore, it is expected that there will be no expensive cars on our value list.

Predictions with linear regression model (cost and hp)

Predictions

- Predicted cost of a car with 200hp = $28481.66
- Predicted cost of a car with 300hp = $39965.66
- Predicted cost of a car with 400hp = $51449.66



Ridge Regression model predictions:

Value was found by finding the predicted price and dividing by the actual price. Was found with the MPG, HP, and the price.

Ridge Regression model predictions:

Top 3 value cars:

1. Ford Mustang (v8) ($32.645, 25 mpg, 435hp)
2. Dodge Challenger ($37.995, 23 mpg, 485hp)
3. Chevrolet Camaro ($36.905, 25 mpg, 455hp)

Top 3 best value cars under $20000:

1. Honda Civic ($19.700, 39mpg, 174hp)
2. Chevrolet Cruze ($16.975, 39mpg, 153hp)
3. Ford Focus ($16.775, 36mpg, 160hp)

Top 3 best value cars under $25000:

1. Ford Mustang (v6) ($24.645, 27mpg, 300hp)
2. Mitsubishi i-MiEV ($22.995, 102mpg, 66hp)
3. Dodge Grand Caravan ($23.995, 25mpg, 283hp)

Top 3 best value cars under $35000:

1. Ford Mustang (v8) ($32.645, 25mpg, 435hp)
2. Ford Mustang (i4t) ($25645, 30mpg, 310hp)
3. Toyota Tundra ($30.400, 18mpg, 381hp)

Ridge Regression model predictions:

Top 10 best value cars under $45,000:

1. Ford Mustang (v8) ($32.645, 25 mpg, 435hp)
2. Dodge Challenger ($37.995, 23 mpg, 485hp)
3. Chevrolet Camaro ($36.905, 25 mpg, 455hp)

Ridge Regression analysis:

- As expected the top cars where under 50k and around 400hp.
- As you may have noticed the top cars where all muscle cars, this is because they offer plenty of hp for a great price relative to the competition.

Ridge Regression analysis:

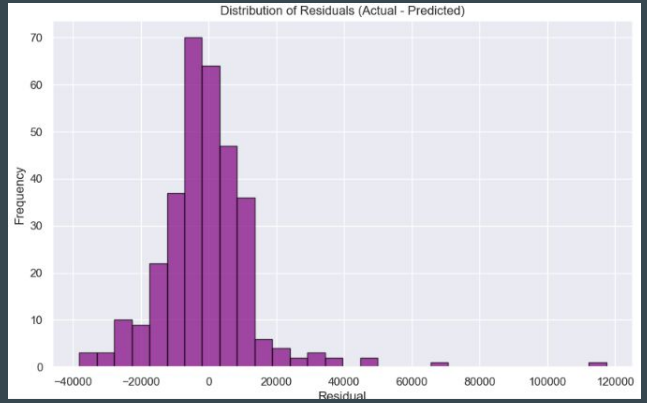
- The predictions for all of the cars where a bit to optimistic.
- Since the data for all of the cars very too much the accuracy suffered.
- Perhaps with a different algorithm it will predict better.

Linear Regression Metrics:

MSE: 202831235.8033986

RMSE: 14241.883155095698

R-squared: 0.6091517708183793



Decision Tree model predictions:

Top 3 value cars:

1. Ford Mustang (v8) ($32.645, 25 mpg, 435hp)
2. Dodge Challenger ($37.995, 23 mpg, 485hp)
3. Ford Mustang (v6) ($24.645, 27mpg, 300hp)

Top 3 best value cars under $20000:

1. Kia Rio ($14.165, 36mpg, 138hp)
2. Nissan Versa ($16.975, 36mpg, 153hp)
3. Ford Focus ($16.775, 36mpg, 160hp)

Top 3 best value cars under $25000:

1. Ford Mustang (v6) ($24.645, 27mpg, 300hp)
2. Dodge Grand Caravan ($23.995, 25mpg, 283hp)
3. Kia Rio ($14.165, 36mpg, 138hp)

Top 3 best value cars under $35000:

1. Ford Mustang (v8) ($32.645, 25mpg, 435hp)
2. Ford Mustang (v6) ($24.645, 27mpg, 300hp)
3. Ford Mustang (i4t) ($25645, 30mpg, 310hp)

Decision Tree model predictions:

Top 10 best value cars under $45000:

1. Ford Mustang (v8) ($32.645, 25 mpg, 435hp)
2. Dodge Challenger ($37.995, 23 mpg, 485hp)
3. Ford Mustang (v6) ($24.645, 27mpg, 300hp)

Decision Tree analysis:

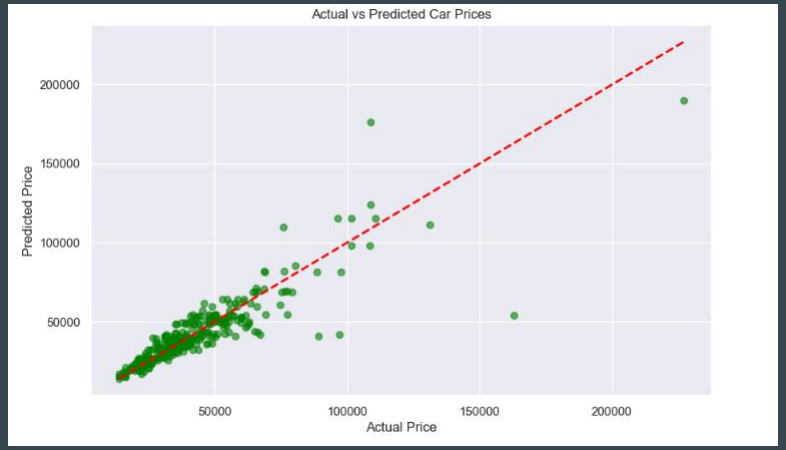
- Compared to the Ridge Regression algorithm this one did better.
- It was able to more accurately predict the value of the car.

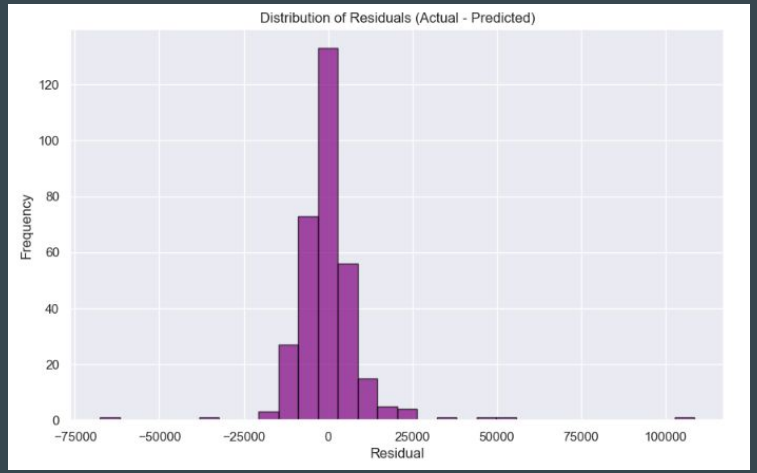
Decision Tree Metrics:

MSE: 11797149.784453844

RMSE: 10861.4684939256

R-squared: 0.7726733220166643





Results:

- Best Model: Decision tree since it out performed the ridge regression.
- Best overall value car according to the models: Ford Mustang (v8)
- Unexpected finds: Mitsubishi i-MiEV was a unique pick (funny looking) and only ev.

Conclusion

- Goals Achieved: Successfully identified best value cars and predicted car prices.
- Key Takeaways:
  - Ford Mustang offer the most value.
  - Decision Tree was the best model for value identification.
- Practical Applications:
  - Helps car buyers focus on value focused decisions. Car buying is a personal decision and numbers should not be the only deciding factor.

Questions

Thank You