

# Determining The Relationship Between Used Car Prices and Their Attributes

---

Joey Billet

## Objective

To identify used car attributes in effort to determine pricing before placing a bid

## Utilization

Utilizing data from 2010-2020 will give the opportunity to buyers to make informed decisions based on what price point best fits their goals



Bidding on cars can be stressful



Determining the optimal price to purchase a vehicle

# How much of a factor do car features play in regards to pricing?

By looking at data, can we determine which features have a high correlation with price?

## Price

---

Utilizing Price as our target variable

## Mileage

---

Determine whether high or low mileage effects price.

## \*Condition

---

Does the “time remaining” on a used car bid directly effect the price?

## Age

---

Are the age and price of a used car independent of each other?

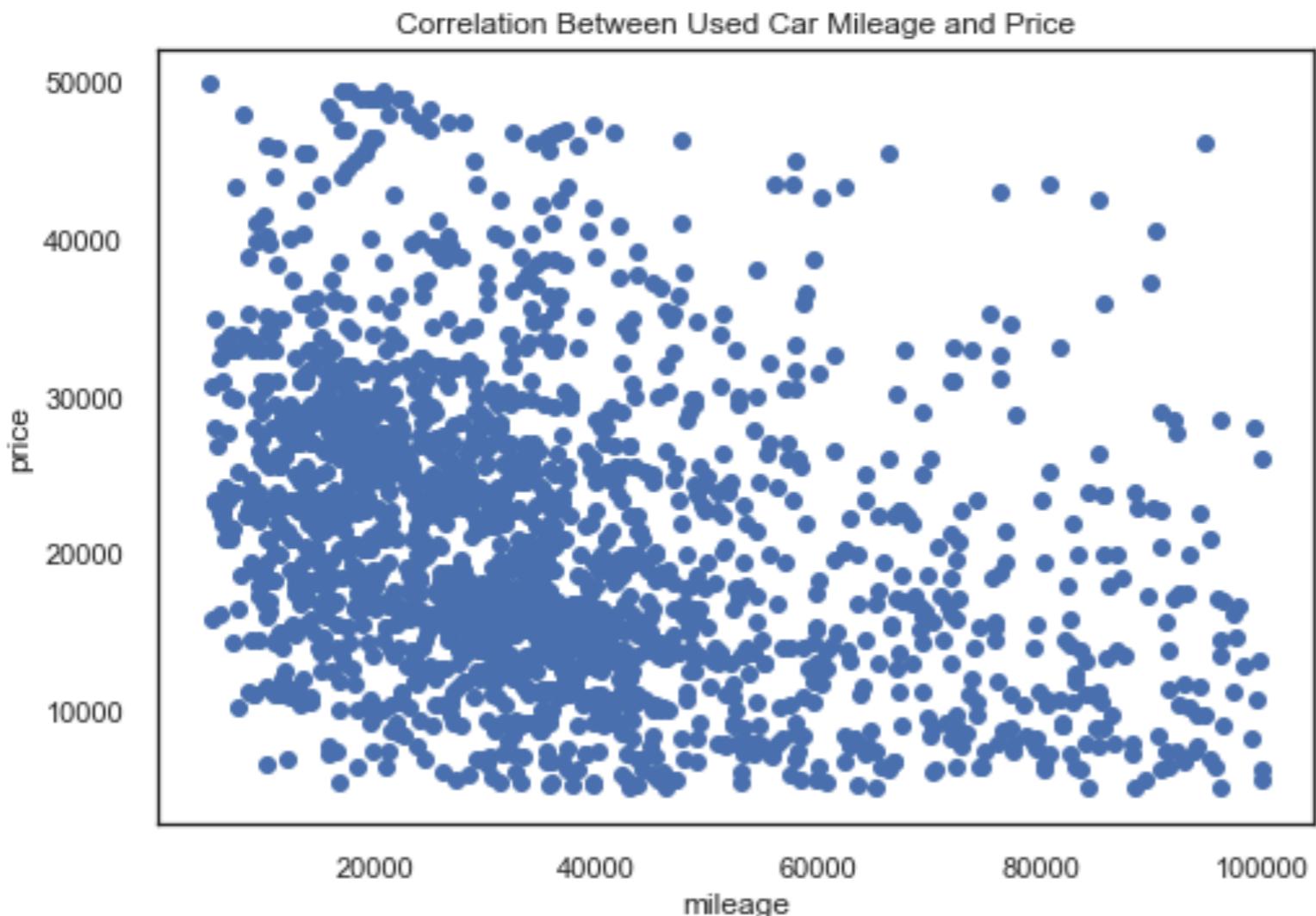
\*Condition is the time remaining of the auction for a used car

## Workflow

Utilizing the many libraries on the Jupyter Notebook created a seamless integration of data



Looking at the below chart, there is a **negative** correlation between price and mileage when bidding on a used car



## The Mileage Effect

ANOVA Test

-

H<sub>0</sub>: There is no significant difference in used car prices and mileage

-

H<sub>a</sub>: There is a significant difference in used car prices and mileage

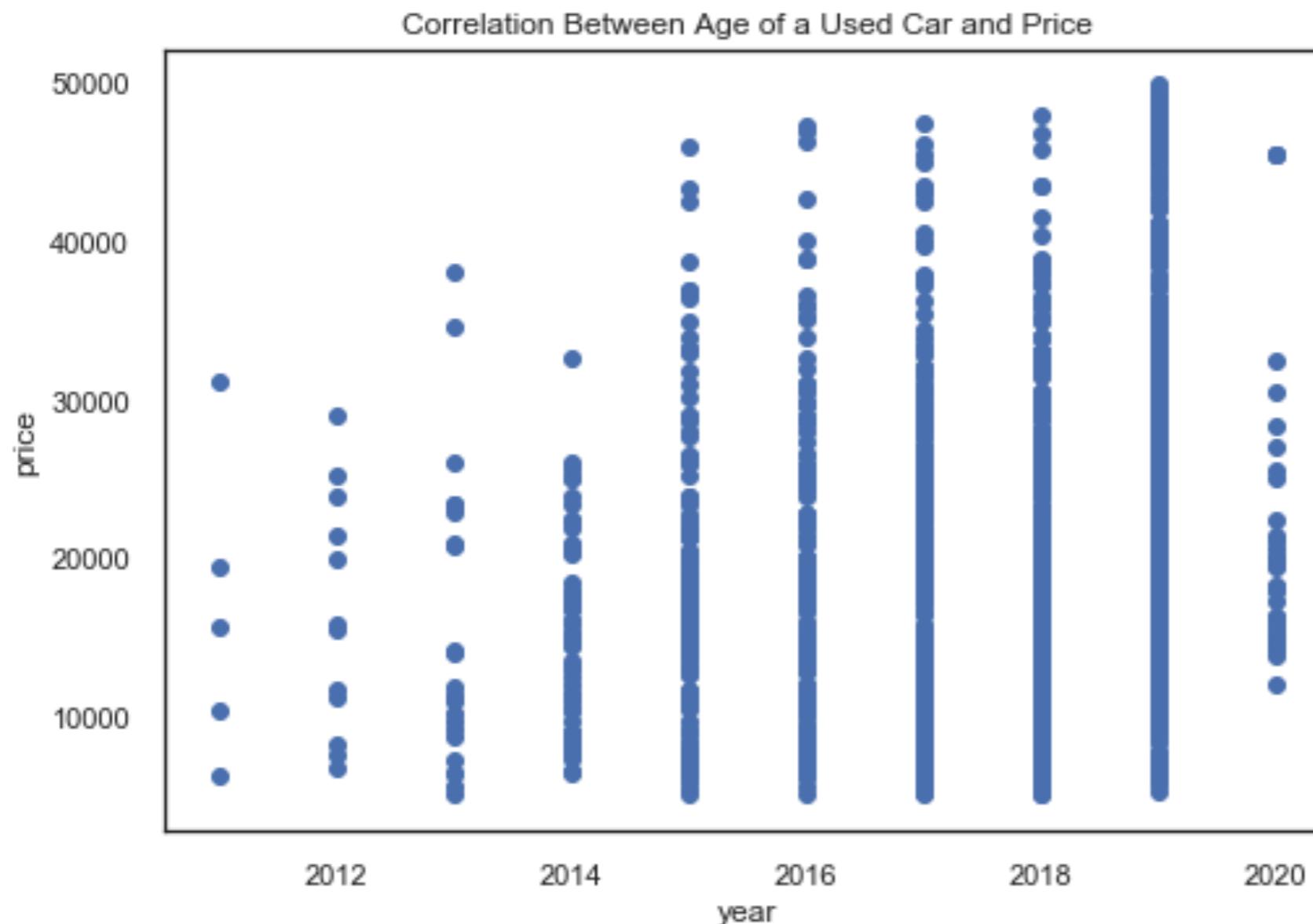
-

We can reject the null that the means are equal for these groups

```
F_onewayResult(statistic=56.28195170657985, pvalue=1.0374327074104347e-24)
```

The older the car is, the *less* expensive it is

Both of these features are dependent on each other



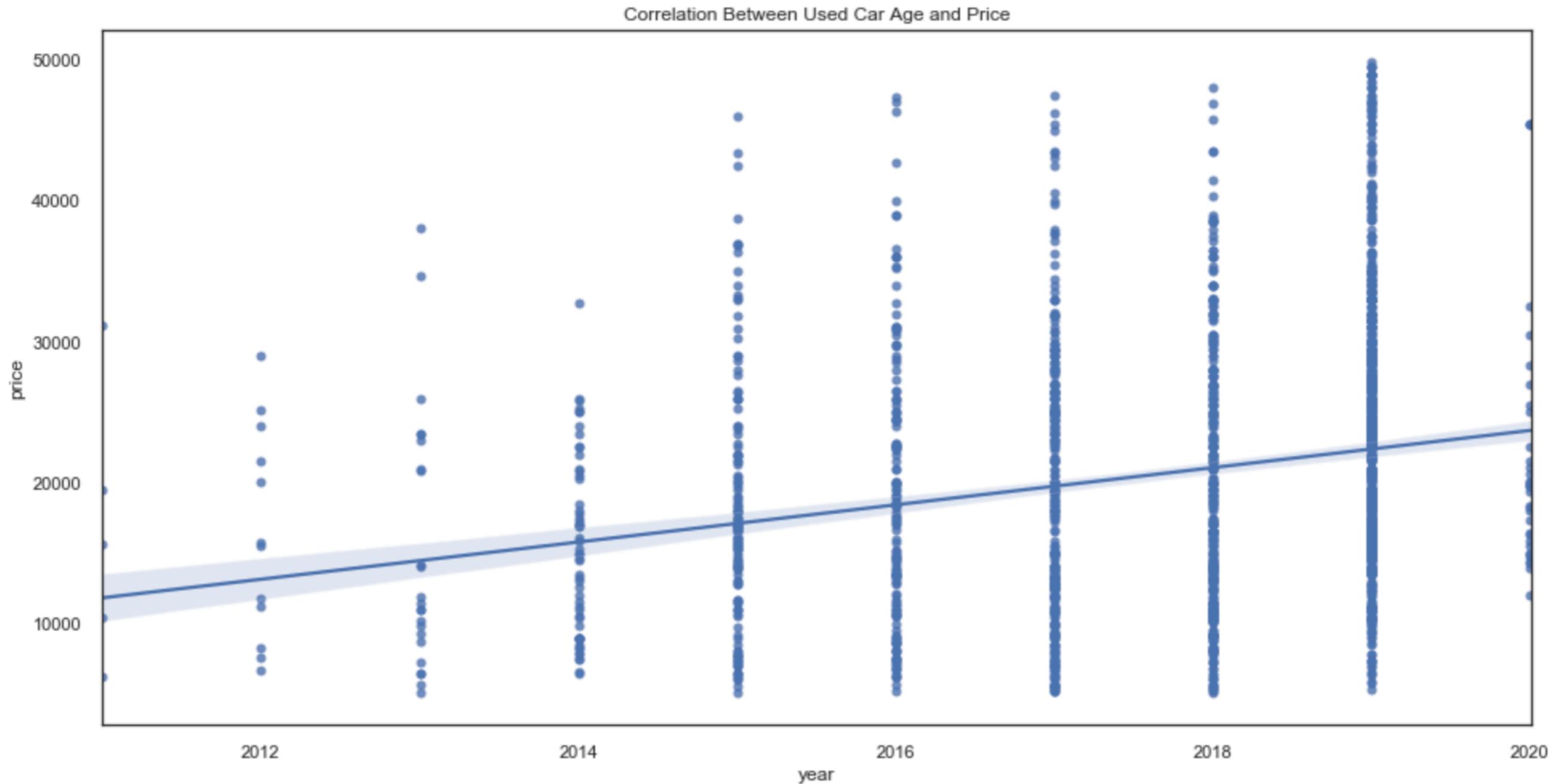
### Chi-Square Test

H0: There is no significant difference in used car prices and age

H0: There is no significant difference in used car prices and age

We can reject the null hypothesis that the two features are independent of each other

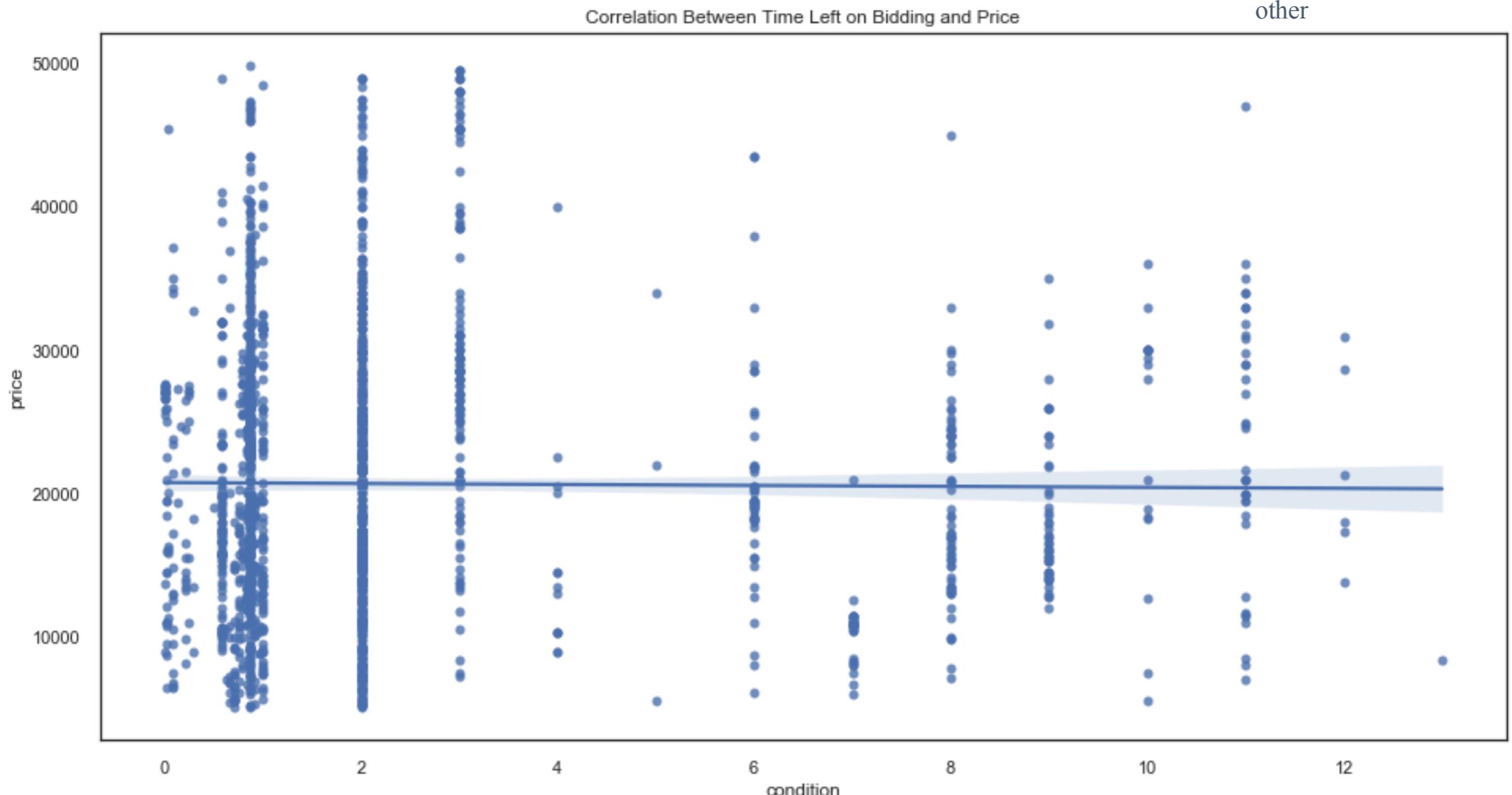
## Car Age $\diamond$ Price



Bid strategically as the price consistently rises while time elapses

## Chi-Square Test

- H<sub>0</sub>: There is no significant difference in used car prices and age
- H<sub>0</sub>: There is no significant difference in used car prices and age
- We can reject the null hypothesis that the two features are independent of each other



Determining the best **best** model to use:

### Models Used

Linear Regression  
Lasso Model  
Second Degree Poly  
Third Degree Poly

### Best Model

Lasso Model

### Lasso Model

Training RMSE: 5033.116070372421  
Testing RMSE: 5631.757613282632

The Lasso Model was the best fitting model.

After assessing the used car data, through the data the following **following** has been determined:



- There is a negative correlation between price and mileage. The more mileage, the lower the price.



- The age of a car and the price of a car are dependent on each other. The older the car, the lower the price.



- As time winds down on bidding, prices tend to be higher. Place bids strategically for the best chance to win.

Based on the data, we see that cars depreciate in value the moment they are considered used

Thank you

Thank you!

---

Joey Billet