# What drives price of used car

A basic report by Chang Liu

## Summary of findings:

1. I find the car price is a function of many things, including cylinder count, drive type, manufacturer, size, odometer, and age (current year – manufactured year). I focus on these six features.
2. I did data cleaning to remove NaN, wrong inputs, and apparent outliers (in price, odometer, and age)
3. I used one hot encoding of cylinder, drive, and size, since the count is not too high.
4. I split data sets and build multi-variate linear regression model to predict the car price.
5. Overall prediction and correlation for all the data is poor, since there are too many car models and conditions.
6. At the end I did a mini dataset with only used Toyota cars, and build a second order parabolic regression fit of price vs. odometer.

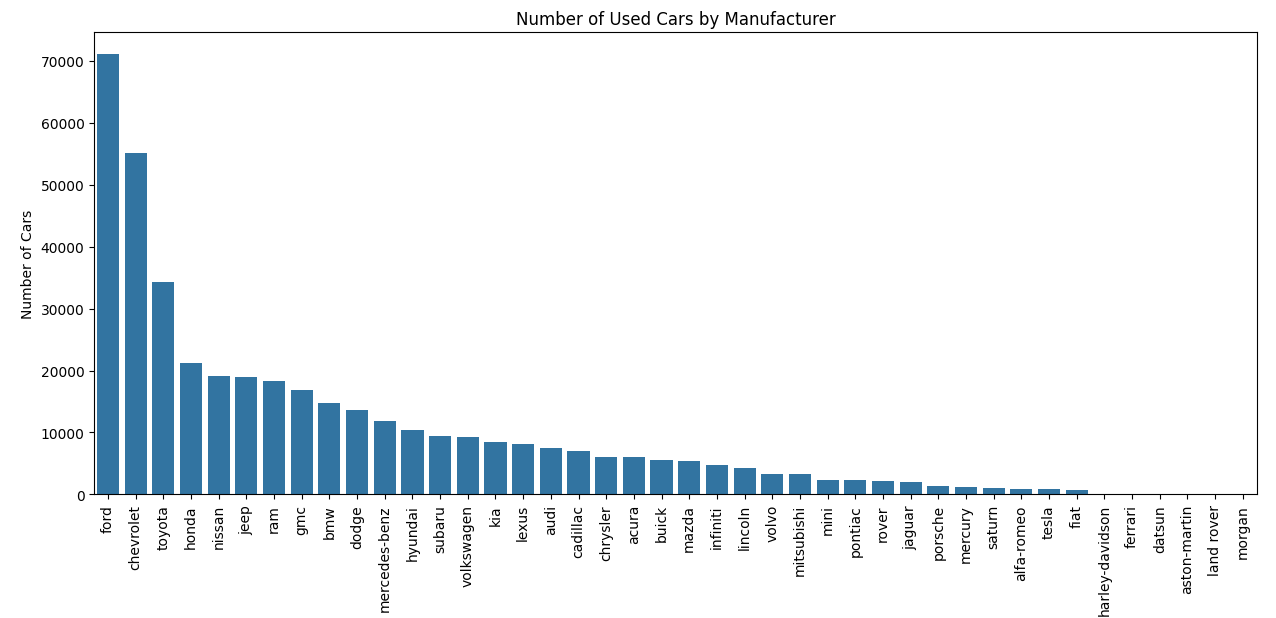
## Link to notebook:

<https://github.com/changliu8888/usedcarprice/blob/91d0b2c5bd91360da8b7266a472e248f485885a9/price_of_car_Chang_Liu.ipynb>

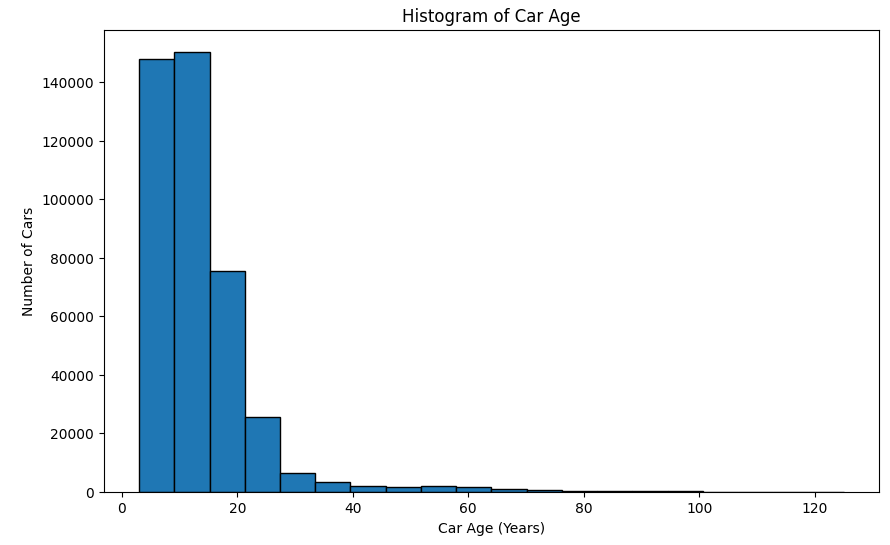
## **A list of business suggestions to the car dealer:**

1. Stock on 3 best volume cars – Ford, Chevy, and Toyota.
2. Used car price drops monotonically over age – except for exotic cars.
3. Second hand cars of four years command a higher price than “younger” used cars.
4. Tesla cars have good price but the history of retail is not proven.
5. Toyota is a better second hand car to have in inventory than Honda, Hyundai, and Volvo. It has longer retail cycle and higher price.

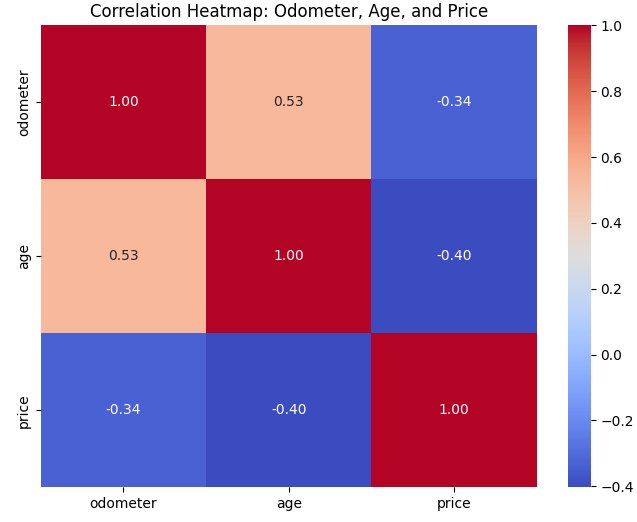
## Below are some of the most important sample graphs I would use in a presentation to them.



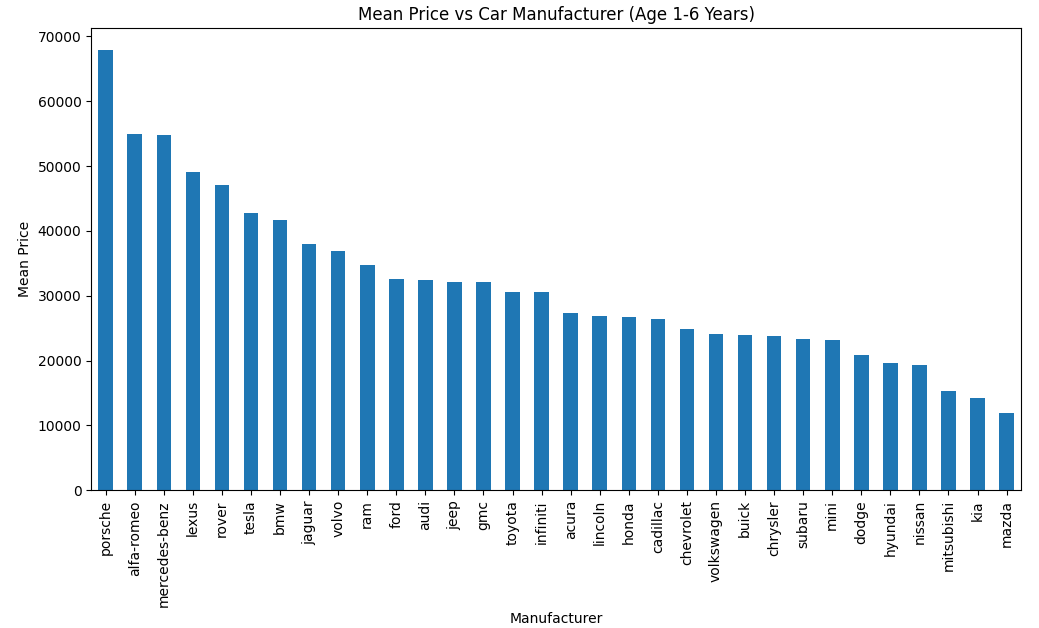
Ford, Chevy and Toyota are top selling by volume.



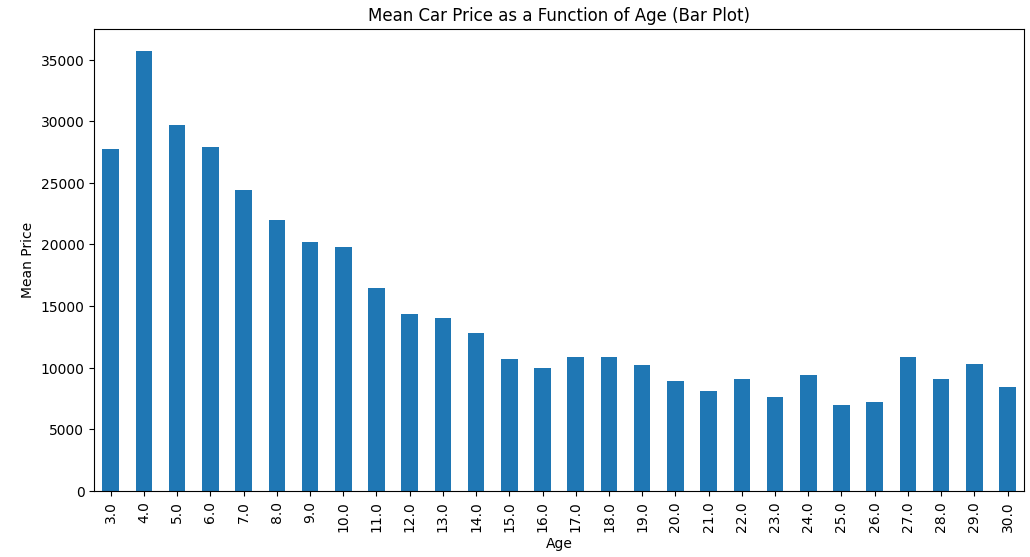
Medium price of car by age. There is a sharp drop off after 15 years. Before that the used cars seem to retain good price.



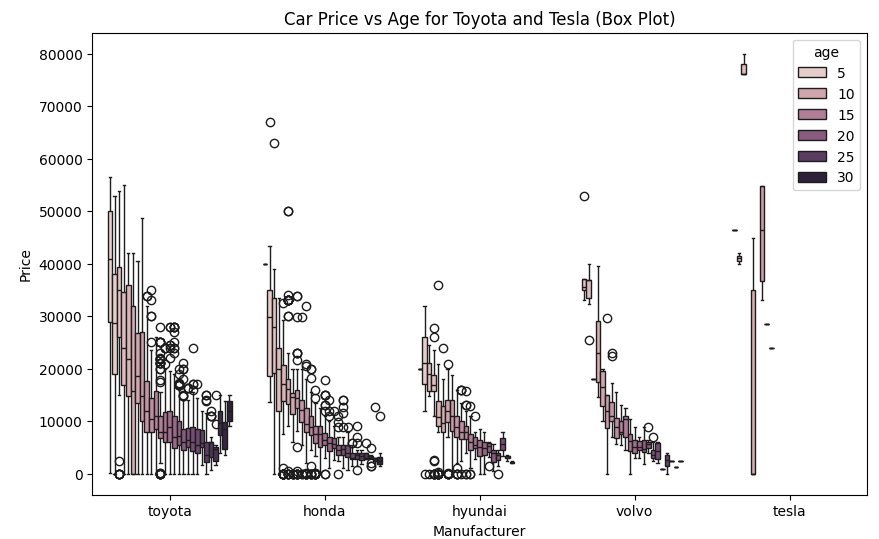
Correlation between odometer, price, and age is rather poor as a general database with all different models, drives, and manufacturers mixed.



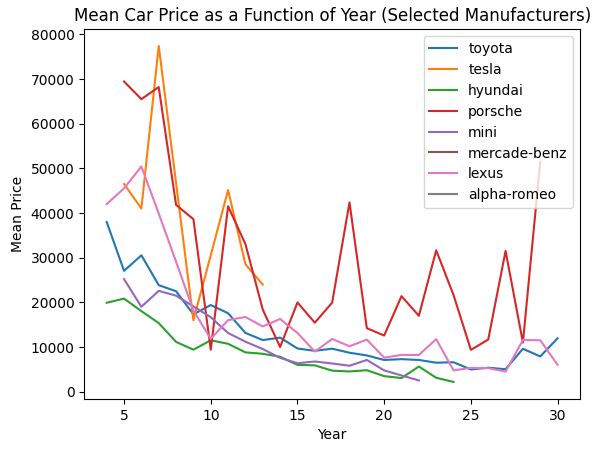
Within six years of age, luxury cars have the highest price.



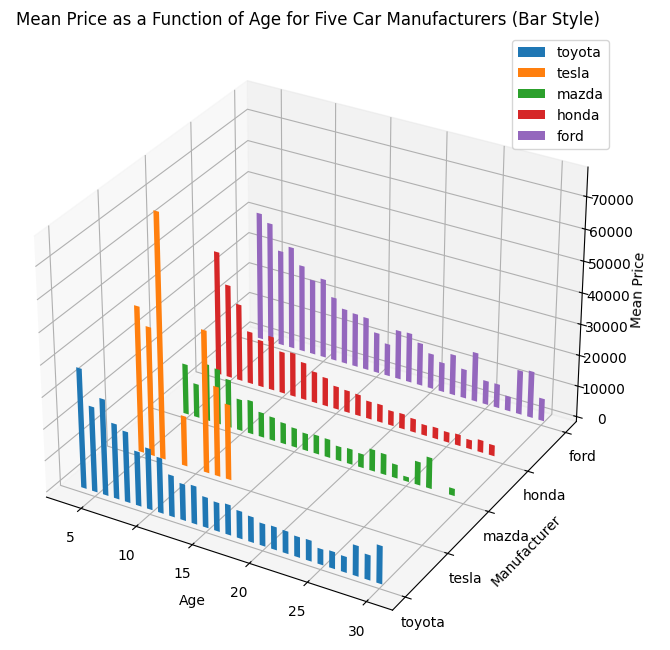
The best price of a used car is year 4.



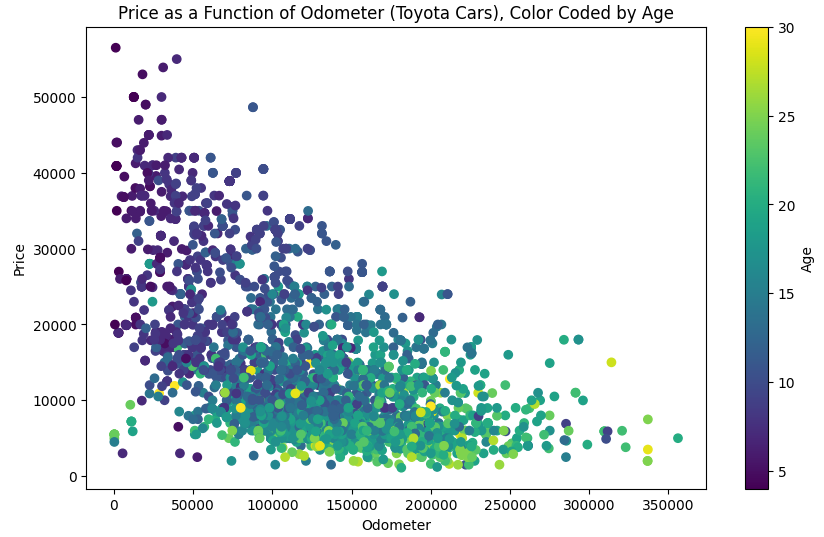
Comparison of median price and spread aa a function of age, for different manufacturers.



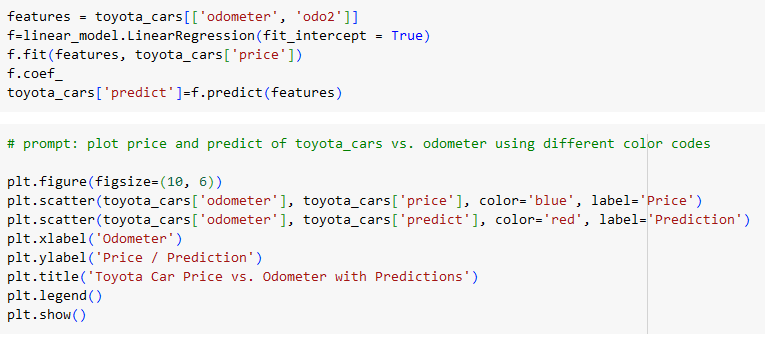
Comparison of median price and spread aa a function of age, for some models I suspect are popular.

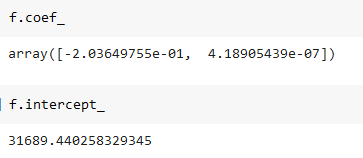


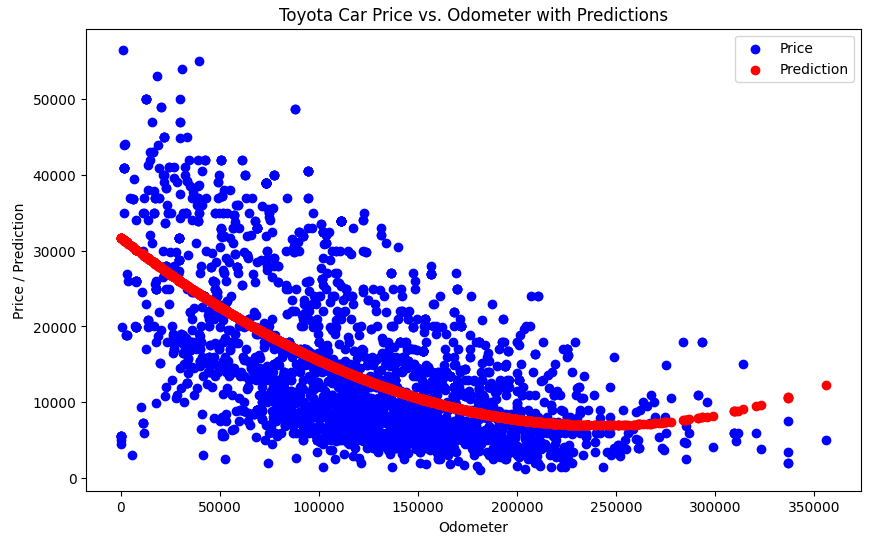
Tesla cars have high second hand pricing, but the data is sparse. Ford, Honda, Toyota are all good inventory items in my opinion.



There seems to be two clusters of used cars (Toyota only). The ones with low miles and the ones with higher miles. Low mile Toyota cars are of course the best to have, in terms of pricing.







Predicted price of Toyota cars as a second order function of odometer. Second order seems a bit over fit at high miles. Overall the price drops linearly for the first 150,000 miles and then levels off.