Used Car Market on eBay and Craigslist Data Analysis Report

STA 141B

Data & Web Technologies for Data Analysis

Kevin Wang

Yulun Lu

Hui Zou

Table of Contents

[I. Introduction 3](#_Toc4005502)

[1. Motivation 3](#_Toc4005503)

[2. Introduction of Datasets 3](#_Toc4005504)

[a. eBay 3](#_Toc4005505)

[b. Craigslist 3](#_Toc4005506)

[c. Kaggle 3](#_Toc4005507)

[II. Visualization and Analysis 4](#_Toc4005508)

[1. Common Car Brands and Price Distribution 4](#_Toc4005509)

[2. Common Car Types and Price Distribution 6](#_Toc4005510)

[3. Fuel Types and Price Distribution 8](#_Toc4005511)

[4. Price & Years 10](#_Toc4005512)

[III. Interactive Visualization and Analysis 12](#_Toc4005513)

[IV. Limitations 13](#_Toc4005514)

[V. Summary 13](#_Toc4005515)

# I. Introduction

## 1. Motivation

For college students like us, having a car is important when we are not living in the campus. Because most students can’t afford a new car, a used car is a good choice because it’s cheaper and easier to find. However, second-hand car market is different from new car market because all used cars are in different conditions, and prices varies a lot based on the conditions like year and mileage. Most people won’t know whether the car they chose was in a reasonable price. To solve this problem, we looked into different websites for second hand cars information and chose eBay and Craigslist as the research platforms to make comparisons under different categories. By showing the overview of used car markets’ features, people who are interested in buying second hand cars would be able to obtain better and clearer understanding of this market, which can further help them make appropriate decisions, such as avoiding spending unnecessary time and money before they purchase cars.

## 2. Introduction of Datasets

### a. eBay

We choose eBay because it has a comprehensive used car market information across the nation and at the same time very friendly to scrape. After we scrape and clean the 21254 posts on the eBay, we got the eBay car dataset that contains 10 columns and 21254 rows. The ten columns include price, year, mileage, brand, model, body type, engine, fuel type, transmission and color.

### b. Craigslist

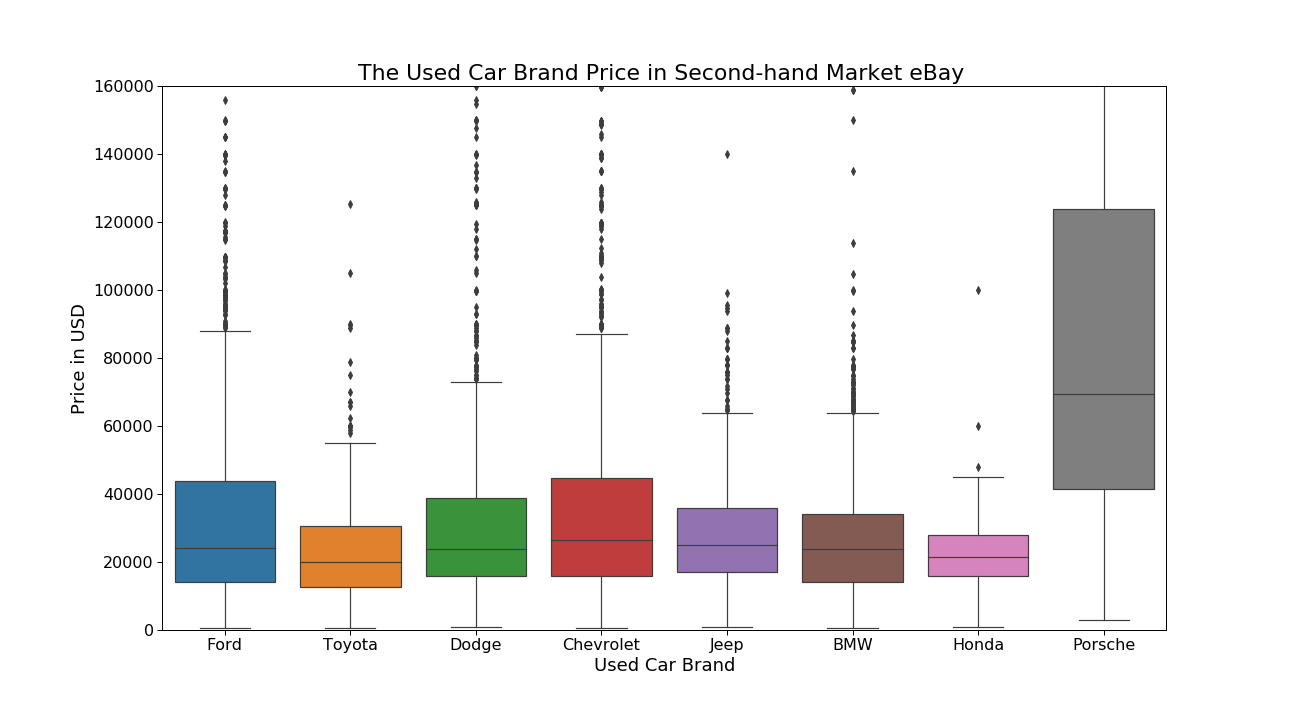
We choose Craigslist because it contains a huge amount of car information available. After we scrape and clean the data, we got the Craigslist car dataset that contains 9 columns and 25048 rows. The ten columns also include price, year, mileage, brand, body type, engine, fuel type, transmission and color.

### c. Kaggle

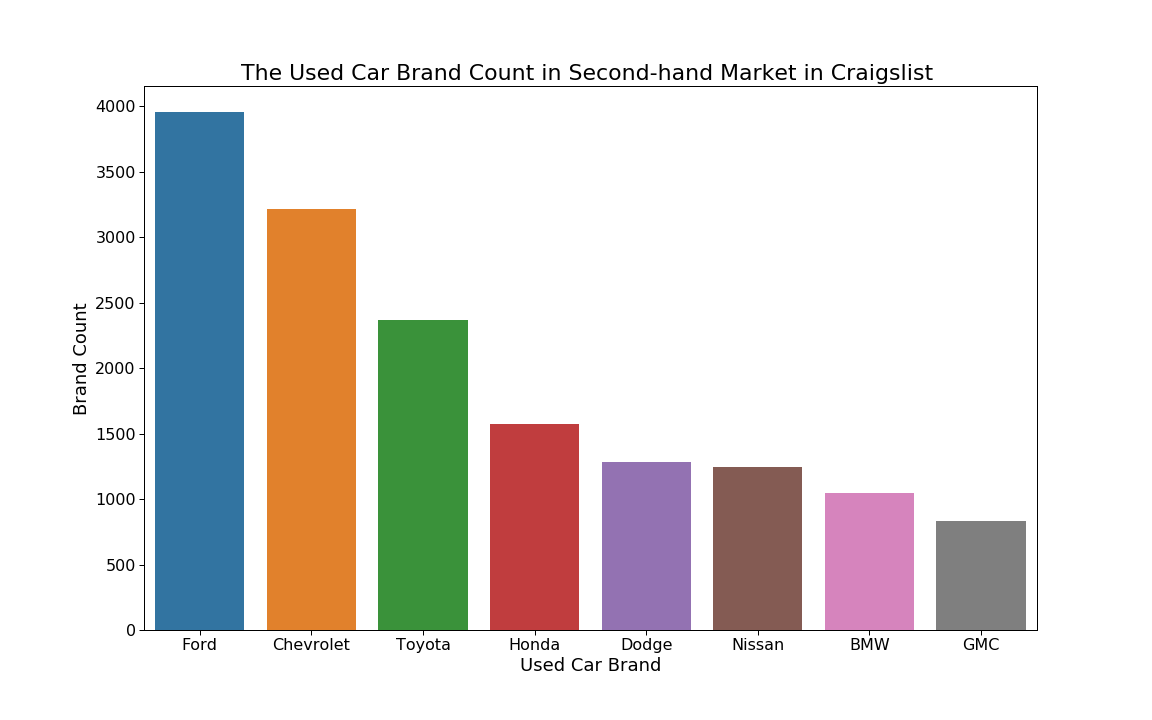
We choose a car Kaggle dataset because we want to investigate in how each car body type different in speed, acceleration, fuel efficiency, CO2 emission. The dataset contains 32848 models.

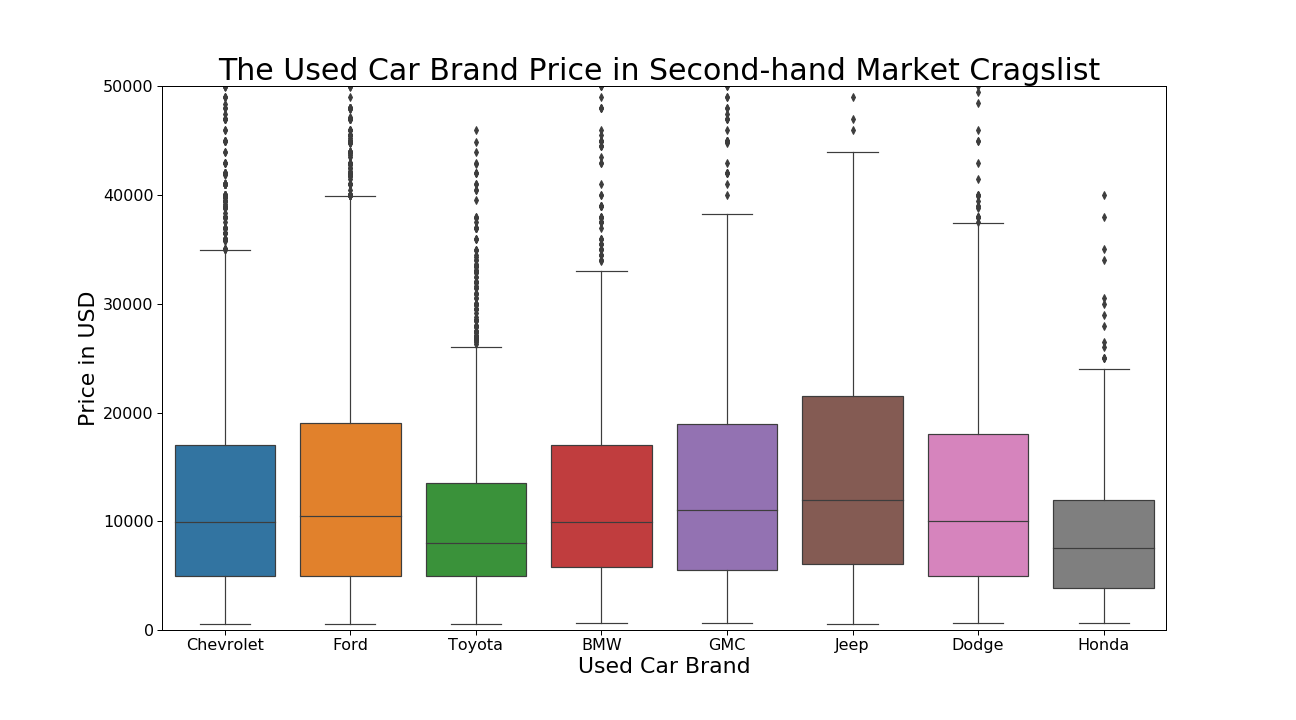
# II. Visualization and Analysis

## STA%20141B/1.png1. Common Car Brands and Price Distribution



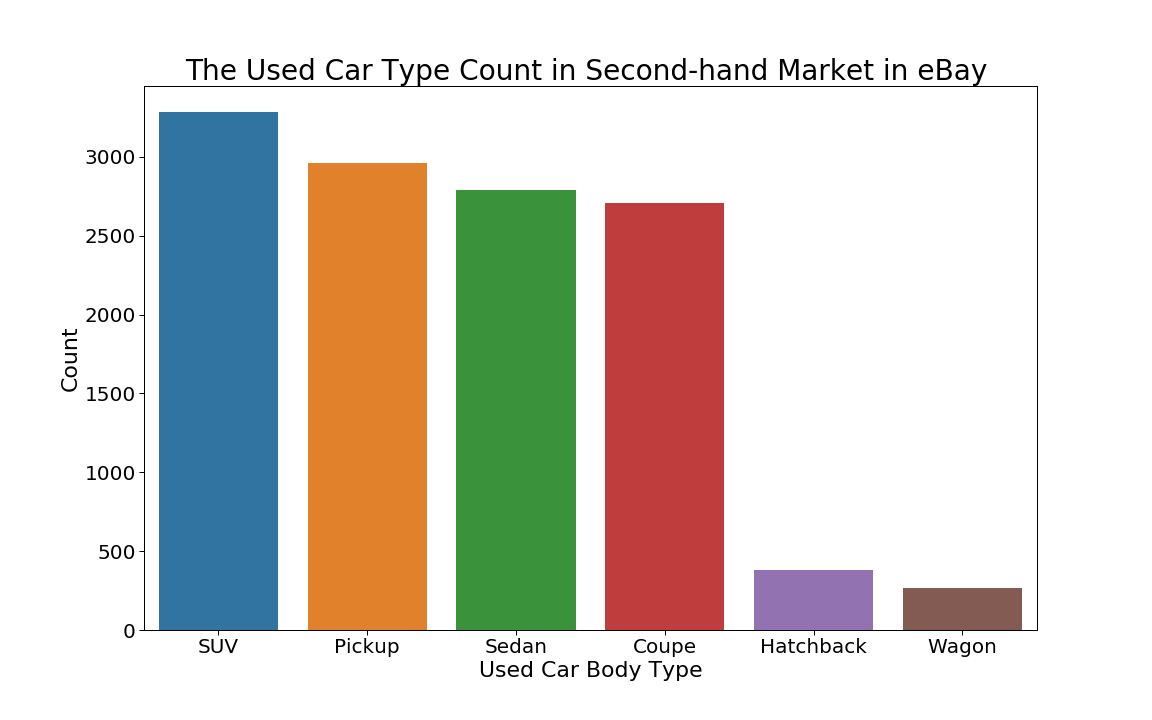
From the two plots above, we can see the top eight most common car brands in the eBay market. Based on the box plots, except for the Porsche’s price is much higher than others at around $70,000 (because it is a luxury car brand), all other brands’ price distributions are very similar at around $20,000. Therefore, I would suggest people who have limited budget to consider more of Ford and Chevrolet on eBay website because these two brands have lots of suppliers in the market, the sellers have to compete and give out relatively lower price. From the price distribution graph, we can see that the prices of Ford and Chevrolet used cars are also price friendly. Thus, buyers who choose Ford or Chevrolet would not only have more choices and reasonable price but also enjoy potential consumer surplus from competition. **(Consumer surplus: the difference between total amount that consumers are willing and able to pay for a good or service and the total amount that they actually do pay.)**

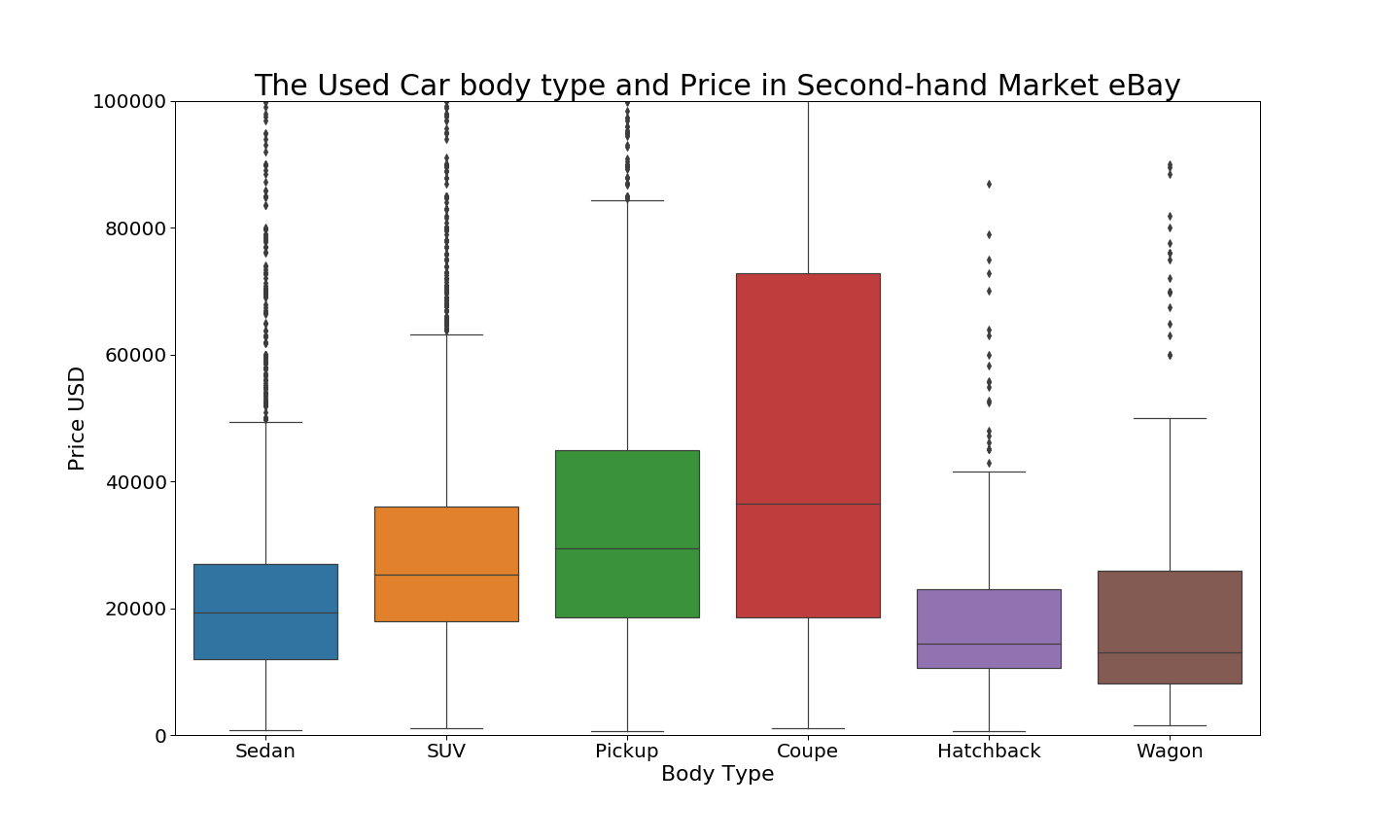




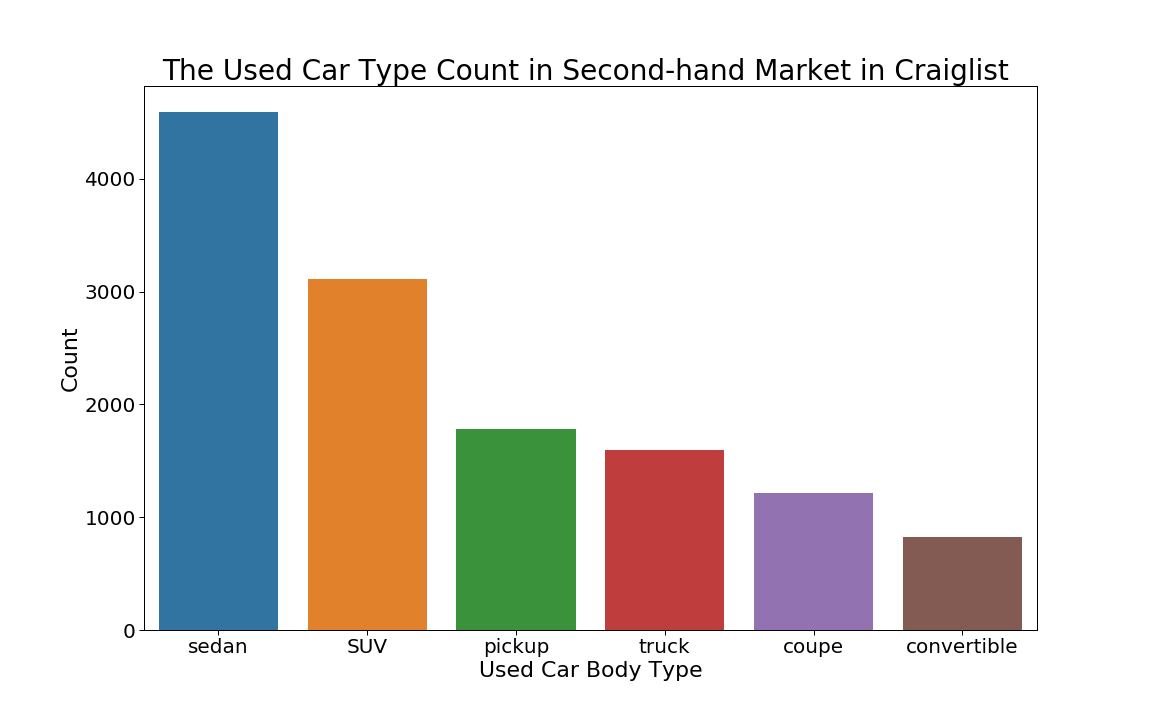
Then, we looked at the Craigslist used car market, the most common car brands are still Ford and Chevrolet, which are 3954 and 3219 respectively. However, Mercedes-Benz is out of the most common car brand list, instead Toyota and Honda became very popular. Also, from the price distribution graph, we can see that the prices of Ford and Chevrolet used cars are also price friendly at around $10,000. Thus, we would still suggest students who have limited budget to consider more of Ford and Chevrolet on Craigslist due to large amount of supplies and low price. Moreover, for some students who have even lower budget, we would suggest them to consider Honda because its first quartile price and median price are the lowest. Compared the two car markets, we can see the price distribution in craigslist is a lot lower than the price distribution of eBay markets. Thus, for people who have low budget, craigslist might be a better car market to look into.

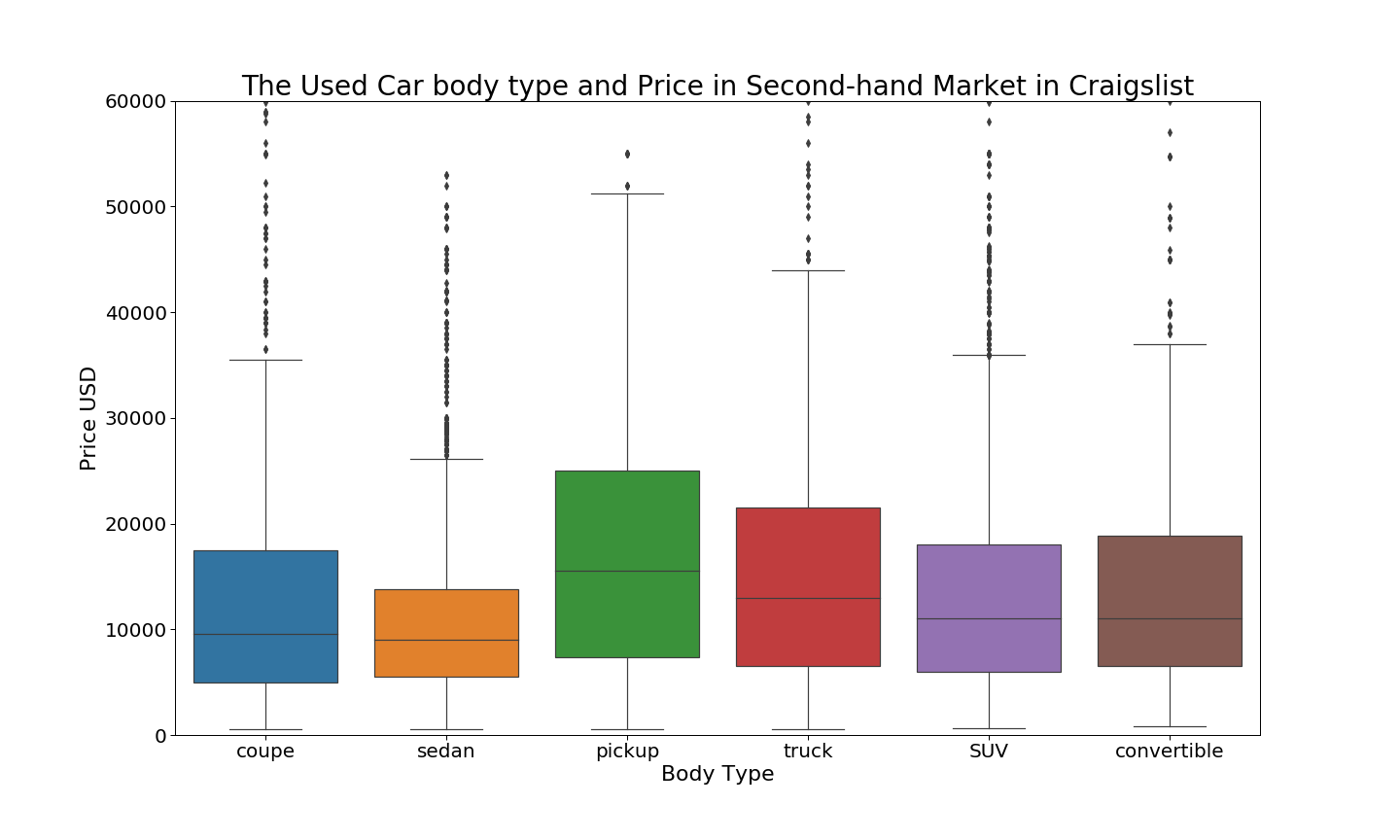
## 2. Common Car Types and Price Distribution

****

****

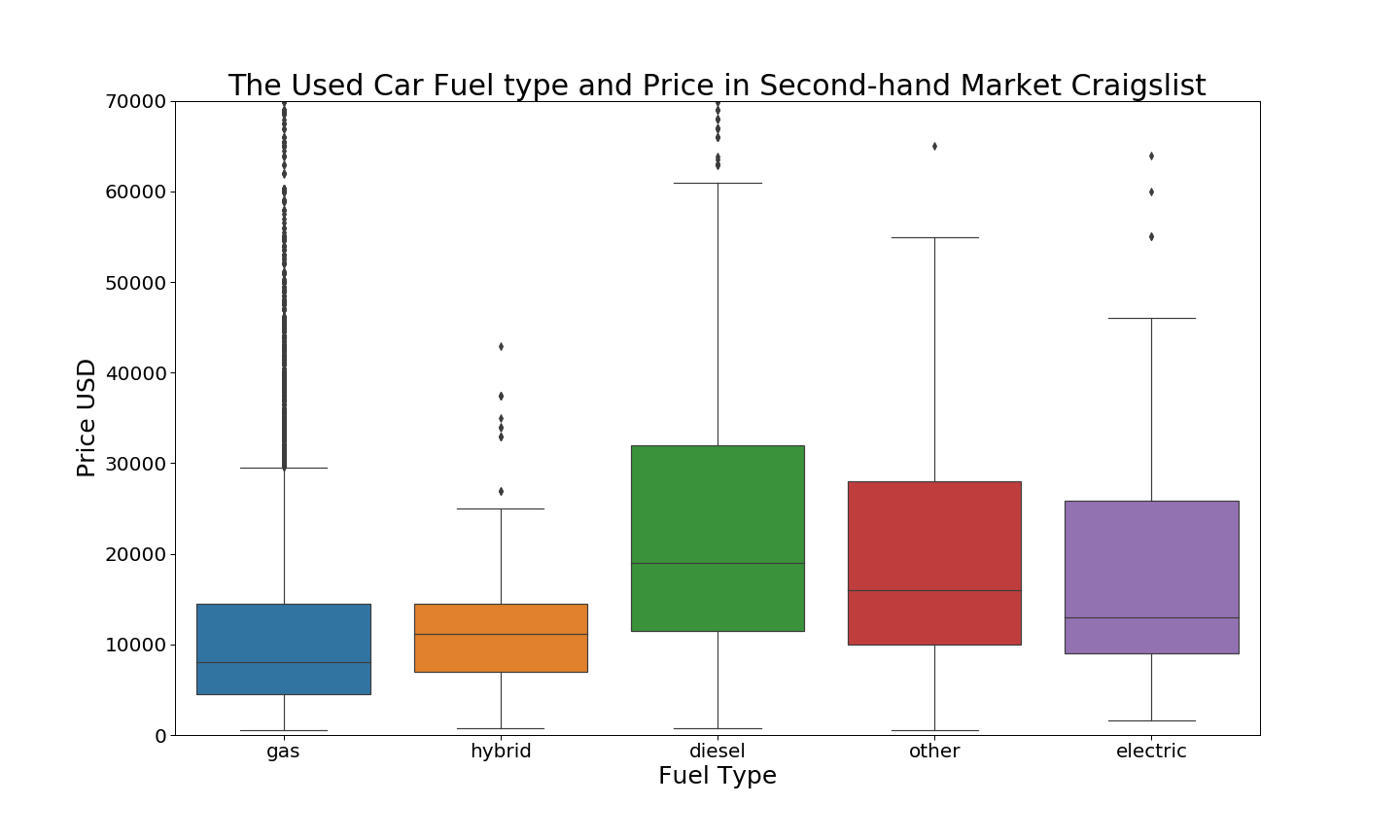
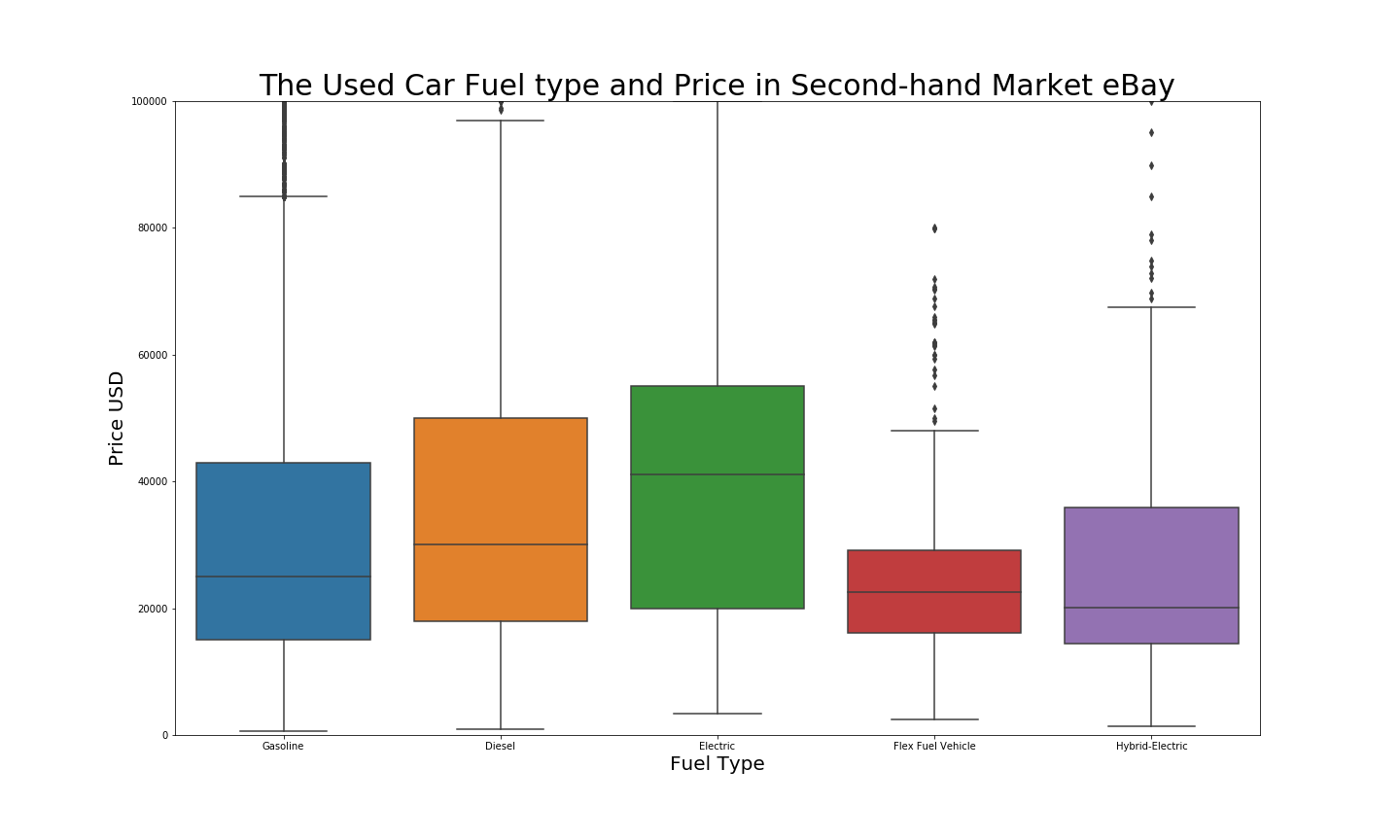
From the two plots above, we can see that the top six most common body type in the eBay market are SUV, Pickup, Sedan, Coupe, Hatchback and Wagon. Based on the box plots, Coupe has the highest median price and rather high price distribution. The Sedan and Hatchback have rather low price distribution. Therefore, after combining and analyzing these two plots’ information, we would suggest people who have limited budget to consider more of Sedan and Hatchback on eBay website because these car body types have high supply and rather low price.



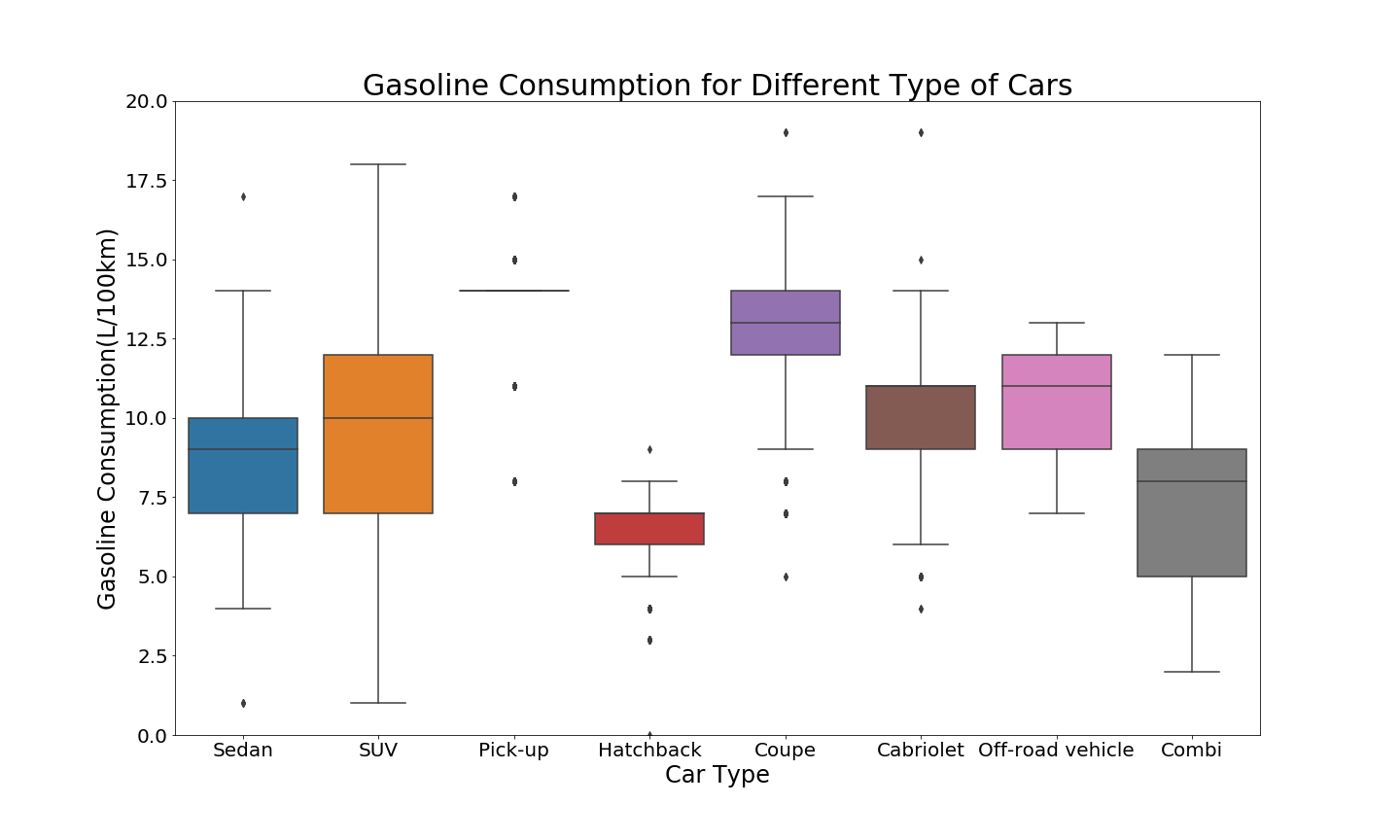
****

In the Craigslist used car market, we can see that the top six most common car types include truck and convertible instead of wagon and hatchback. From the price distributions based on different car types in the box plot, truck has the highest median price and first quartile price, and sedan has the lowest median price and fourth quartile price. Therefore, after combining two plots here, we would suggest people who have limit budgets to consider more on sedan because it has lots of supplies and relatively low price.

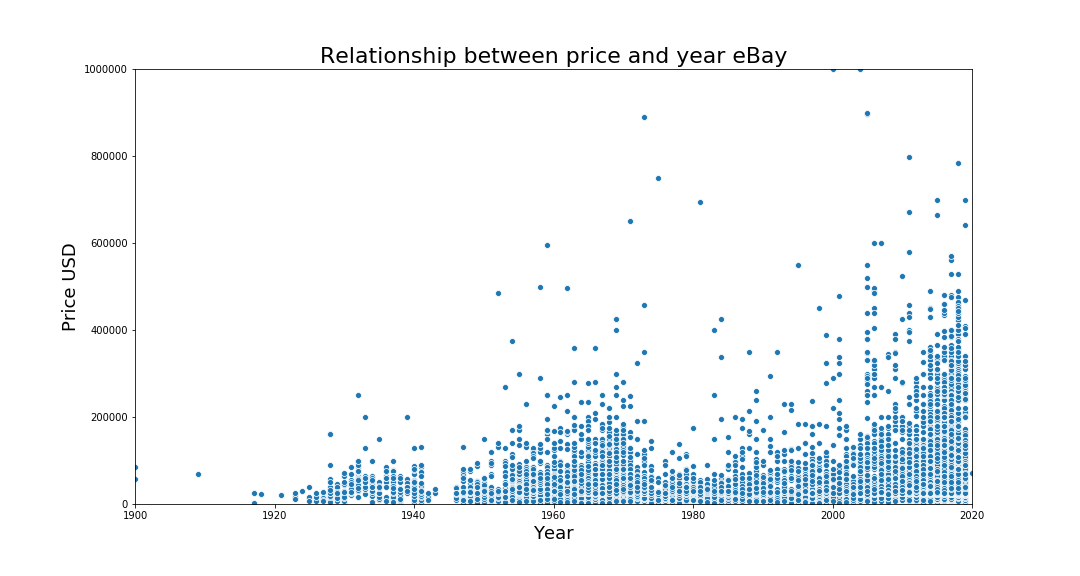
## STA%20141B/3.pngSTA%20141B/c.png3. Fuel Types and Price Distribution

We also look at the cars’ fuel types. From the plots we can see that the gasoline is the most common fuel type which weighs about 88.7% in the whole market, and the least common fuel type is electric which weighs only about 0.4%. However, from the box plot, the electric cars have high median prices. Though the current electric cars’ demand is gradually increasing, such as the rapid expansion from Tesla Company, the eBay second hand car market hasn’t been influenced by the expansion. Since they are relatively new, the reasonably higher overall price, limited driving range, high price and long charging time compared to other type of cars, we recommend avoiding purchasing electric car in the second-hand car, instead people can look more on gasoline type cars because of its friendly price distribution—low first quartile price and median price in both eBay and Craigslist market.

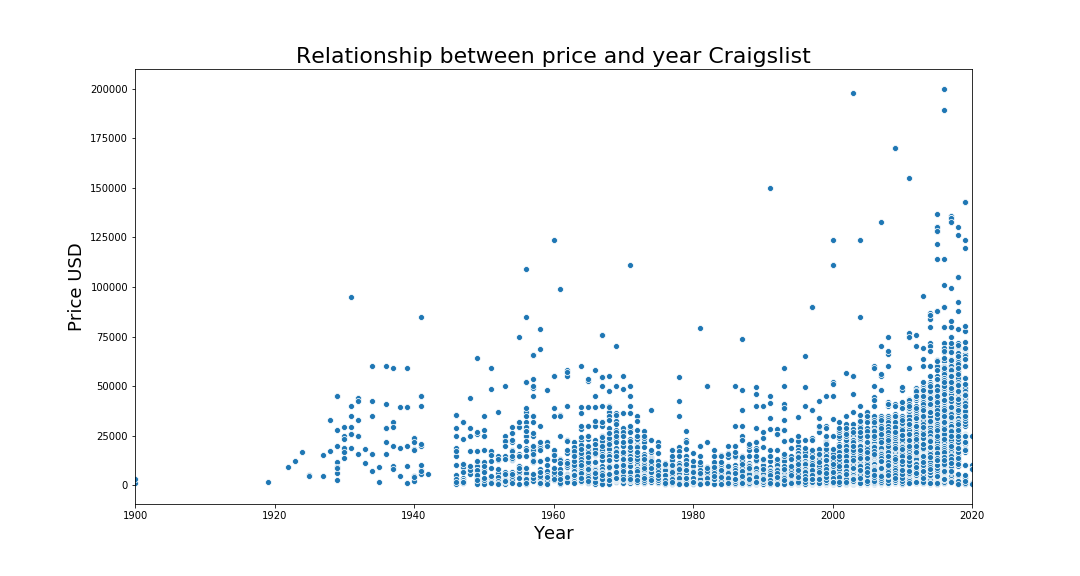
As we discussed above, the most common fuel type is gasoline, therefore next we will focus on the gasoline consumption for different car types. From the plot below, we can see that the all hatchbacks’ gasoline consumptions are less than 10L/100km, and sedan and Combi (Combi coupe) are also having a lower fuel consumption when 75% in both of them have a gasoline consumption less than 10L/100km. Some SUV are also having good gasoline consumption while Pick-up, Coupe, and Off-road Vehicle are having the highest average gasoline consumption. Therefore, we will recommend people to buy Hatchback, Combi, or Sedan in which fuel type is gasoline on eBay market. For students, we suggest them to buy Hatchback since students don’t need a lot of space. For big family, we suggest them to buy Sedan since Sedan cars have larger space.



## 4. Price & Years



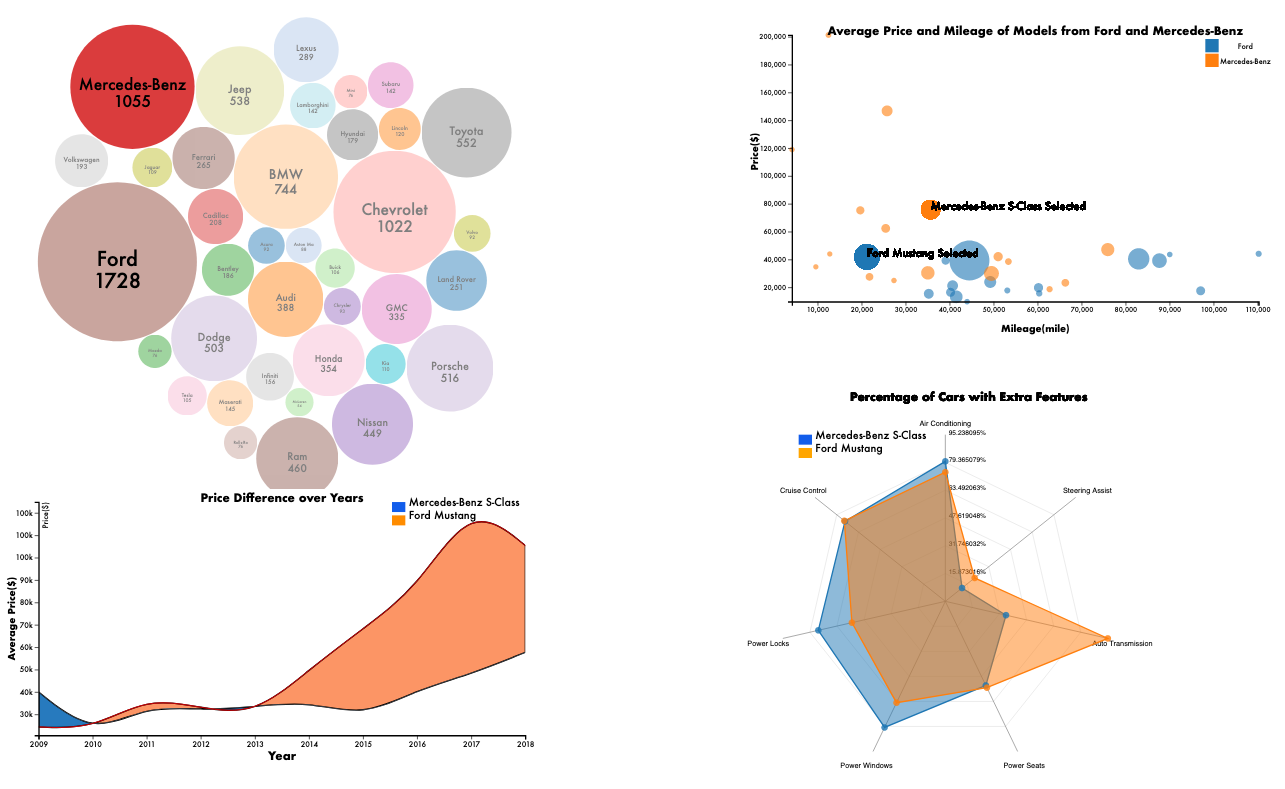
The last category we will explicate here is the relationship between used car prices and years of manufacture. From the plot above, we found a general increasing pattern as the year getting close to 2020 which makes sense because new cars are supposed to be more expensive. We also found an interesting fact that the graph shows a wave pattern. There are two peaks in the graph. One in 1960, and another one in 2020. For the peak in 2020, we can see that from 2000 to 2020, the used-car prices keep increasing at an exponential rate and approaching a high record in 2020, the reason could be the years of manufacture are newer.However, the peak in 1960 is special. We infer that around 1960 is the year range many old fashion cars produced so many people want to buy them as collection. The resource we found online confirms our inference. “Without doubt, the best-looking cars ever to grace the Earth were made in 1960s, and the best-looking automobiles of today are only trying to copy them.” Hence, if you are not interested in collecting cars, it’s not a good choice to buy cars at year around 1960. They are expensive but mostly for collection purpose rather than daily transportation purpose.



In Craigslist used-car market, we can see that there is a similar pattern to eBay market. The exponential increasing rate from 2000 to 2020 which is similar as in the eBay market (except some very high points around 2008). We also found the peak around 1960, which is similar as in the eBay market, too—infer that around 1960 is the year range many vintage cars produced so many people want to buy them for collection purposes. Thus, our recommendation is to avoid buying car around 1960 and cars after 2000. The car in the period 1980 to 2000 will be good choice for people with tight budget. One noticeable difference between the two markets is that almost all used cars’ prices in Craigslist are under $75,000 which is much cheaper than the one in eBay which is around $500,000. Compared to Craigslist market, there are a lot of expensive cars in eBay market. Thus, we recommend people spend more time on Craigslist market for cheaper price second hand cars.

# III. Interactive Visualization and Analysis

Because our project goal is helping customers save money on purchasing used cars, we created an interactive visualization in order to let users find some relationships between each car model and its prices and compare different car models. Users will have a overview of their preferred car models and make a better decision by comparing other available models. The interactive visualization is composed of 4 parts: bubble chart, scatter plot, line difference chart, and radar chart. The bubble chart and scatter plot are used for data selection. In the bubble chart, each bubble represents a car manufacturer when the size of bubbles encoding the amount of car made by the manufacturer. By selecting a manufacturer in the bubble chart, the dots in scatter plot will be updated to show each model of the selected manufacturers, the size of dots also encoding the amount of that model on eBay. After selecting two models to be compared, the other two plots will be updated to show the data of two selected models. The line difference is showing the price and price difference between two models over years. The radar chart shows the percentage of cars having features like “Cruise Control”. The map is at Interactive Visualization/Final Project.html and instructions are included in the same directory. After selecting car makes and models in the first two the map, we can see



# IV. Limitations

For this Project, we only did research on eBay and Craigslist used-car websites because there are about 40,000 posts available on these two websites, and we assume that the large number of posts can be representatives of the whole used-car markets. However, both of them are not real Used Car Website, and they may not provide some important about the cars like horsepower and fuel consumption, so we have to download another dataset to match the car information with the car post data by model. For the eBay market, there might be bias because some sellers set extremely high price, so it would be easier to keep track of their inventory. For the Craigslist market, the possible bias might be caused by the fact that the market is only for cars in California.

# V. Summary

After we compared different car features in eBay and Craigslist used car market, we decided that Craigslist is better website to look for cheap cars because it’s overall price is much lower than the used car in eBay. We also came up with suggestions regarded to different car features. For second hand car brand, we would recommend Ford, Chevrolet and Toyota because there are more choices in the market and the price is low. For car types, we would suggest Sedan and Hatchback for the same reason, more choices and low price. For car fuel type, we would suggest buying gasoline cars for their abundant supply and low price. After comparing gasoline consumption for different car types, we would once again recommend Sedan, Hatchback and Combi for their low fuel consumption. For cars’ year, we would recommend buying car from 1980 to 2000 between the two price peaks. Overall, for people who have lower budget like students, we would suggest them to buy Ford, Chevrolet and Toyota’s Sedan or Hatchback with gasoline fuel between year 1980 to 2000. If they want to get further help on model selection, our interactive visualization would be a great help because it clearly compared price and additional car features.

**Reference links:**

<https://www.kaggle.com/charliewu0729/dataset-for-car-models/version/1>?

<https://www.tutor2u.net/economics/reference/consumer-surplus>

<https://www.quora.com/Why-are-electric-cars-not-becoming-popular-if-they-are-superior-to-gasoline-cars-in-every-respect>

<https://www.ranker.com/crowdranked-list/the-best-1960_s-cars>