

# S364 FINAL PROJECT

Zach Hahn, Sunny Sivarathri, Avery Lopez, and Myles  
Baer

# RESEARCH QUESTION:

## WHAT IS THE BEST USED CAR OF THE 21ST CENTURY?

- We agree that "best" is a very subjective criteria, which is why we completed a very thorough analysis to find out what "best" truly means. There are a few sub-questions that we ask and answer in our analysis to help clarify the answer to our main research question.



# EXAMINING AND CLEANING THE DATA

## ORIGINAL DATA SET

- 426880 rows × 26 columns
- Columns: {id, url, region, region\_url, price, year, manufacturer, model, condition, cylinders, fuel, odometer, title\_status, transmission, VIN, drive, size, type, paint\_color, image\_url, description, county, state, lat, long, posting\_date}
- Columns containing outliers: {price, year, odometer}
- Numerical Variables:{price, year, odometer, lat, long}
- Categorical Variables:{url, region, region\_url, manufacturer, model, condition, cylinders, fuel, title\_status, transmission, VIN, drive, size, type, paint\_color, image\_url, description, county, state, posting\_date}

## CLEANSING DATA

- Dropped rows with 13 or more "NaN" values
  - We deemed 13 to be our threshold
- Columns dropped from original data frame: {region\_url,url,posting\_date,lat,long,image\_url,county,VIN,description}
  - These columns were removed because we didn't see any use for them in our analysis and visualizations. We didn't feel they would be good uses of categorical variables since they have multiple values and wouldn't make sense to categorize

## OUTLIERS AND EXTRA DATA REMOVAL

- Price: Between 500 and 350,000
  - We set this range of price to eliminate outliers. Anything lower than 500 seemed to be very old cars or false entries into the system. Anything greater than 350,000 we declared to be either too rare, antique, or false entries.
- Odometer: Between 200 and 250,000
  - Any number lower than 200 was declared to be too new, since we decided to focus on used cars, or also a false entry. Any number over 250,000 was declared to be too old because it is unlikely someone will want a car that used.
- Year: Later than 2000
  - We desired to focus in on vehicles in the 21<sup>st</sup> century since most of the cars that are on the market are likely within this timeframe.
- Type: No busses or miscellaneous
  - We eliminated busses because it would have slightly skewed our data. Busses tend to have a high odometer number and are likely priced at various price points. We decided to eliminate the “other” category of types because we don’t necessarily know what may be in other and it may skew our data. “Other” could consist of cars, trucks, busses, or any other vehicle that may not be in the data.
- Condition: No new vehicles
  - We simply wanted to focus on used vehicles of any kind, so we eliminated any “new” vehicles

## NEW DATA SET

- 283865 rows × 17 columns
- Columns: {id, region, price, year, manufacturer, model, condition, cylinders, fuel, odometer, title\_status, transmission, drive, size, type, paint\_color, state}
- Numerical Columns used: {price, year, odometer}
- Categorical Columns used: {manufacturer, state, condition, type, cylinder}



# WHICH USED CAR ATTRIBUTES GIVE YOU THE "BEST BANG FOR YOUR BUCK?"

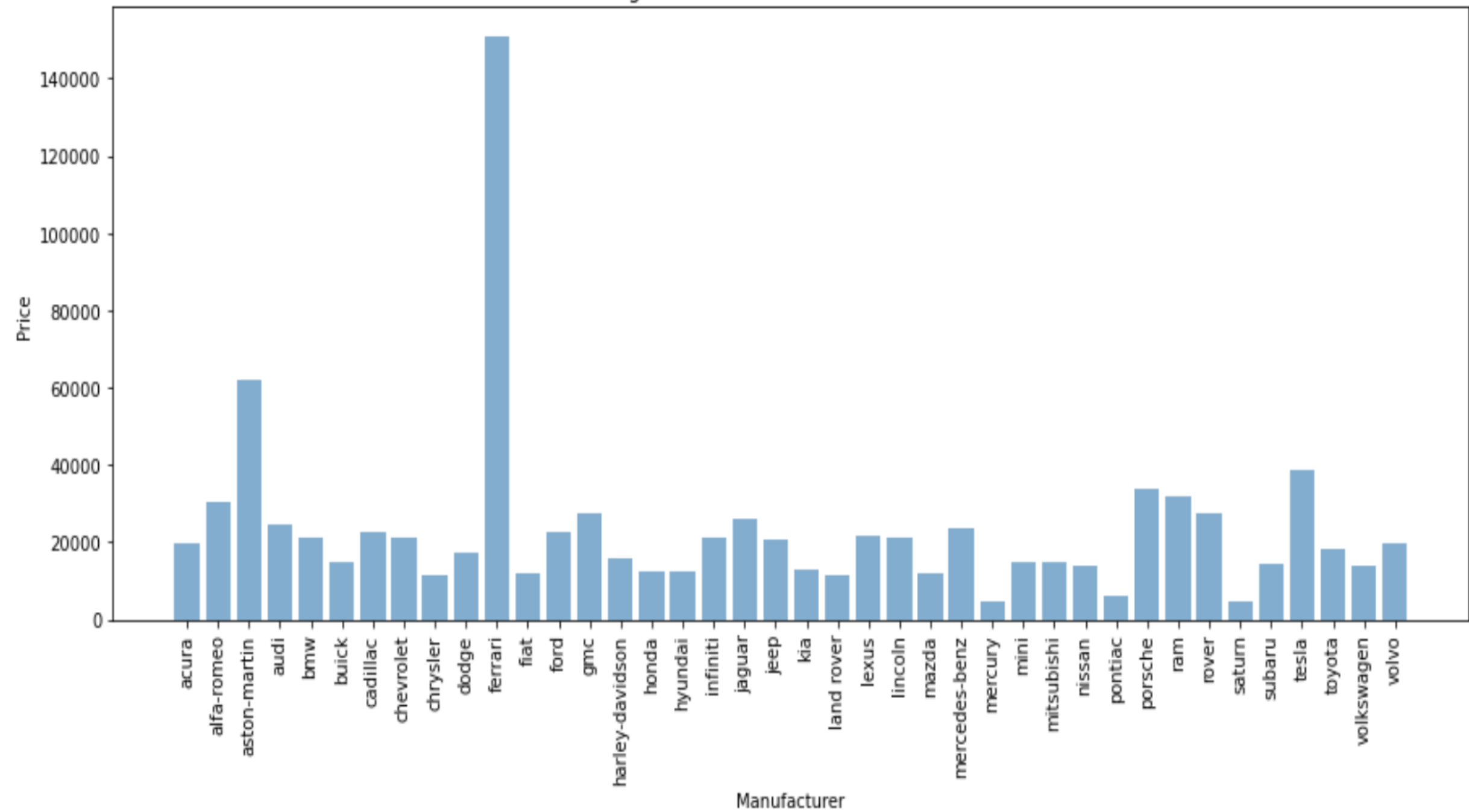
In this section, we discuss used car prices vs certain car attributes. We compare average prices to each attributes' popularity.



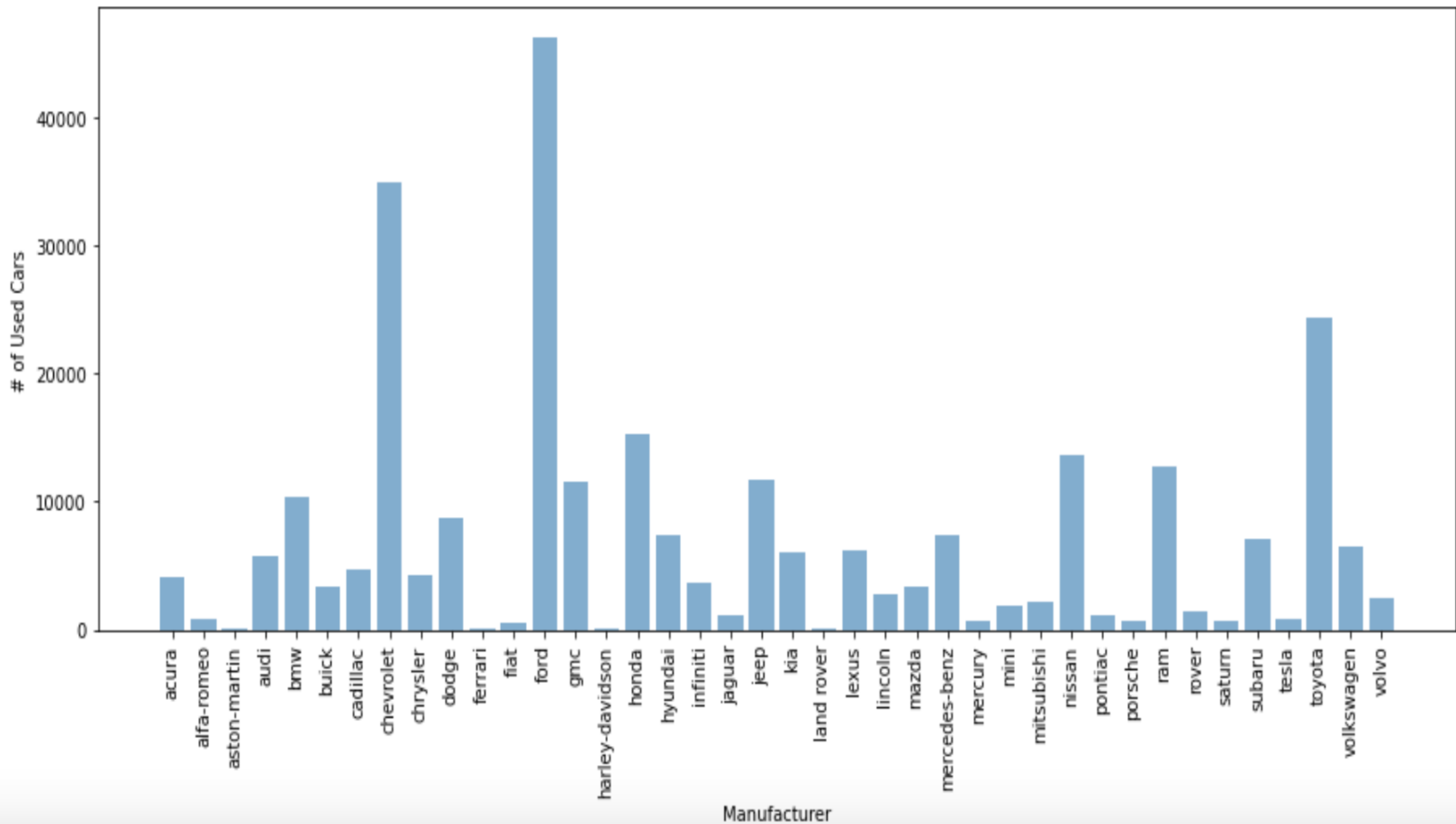


**MANUFACTURER**

Average Price of Used Cars Per Manufacturer



Total # of Used Cars Per Manufacturer

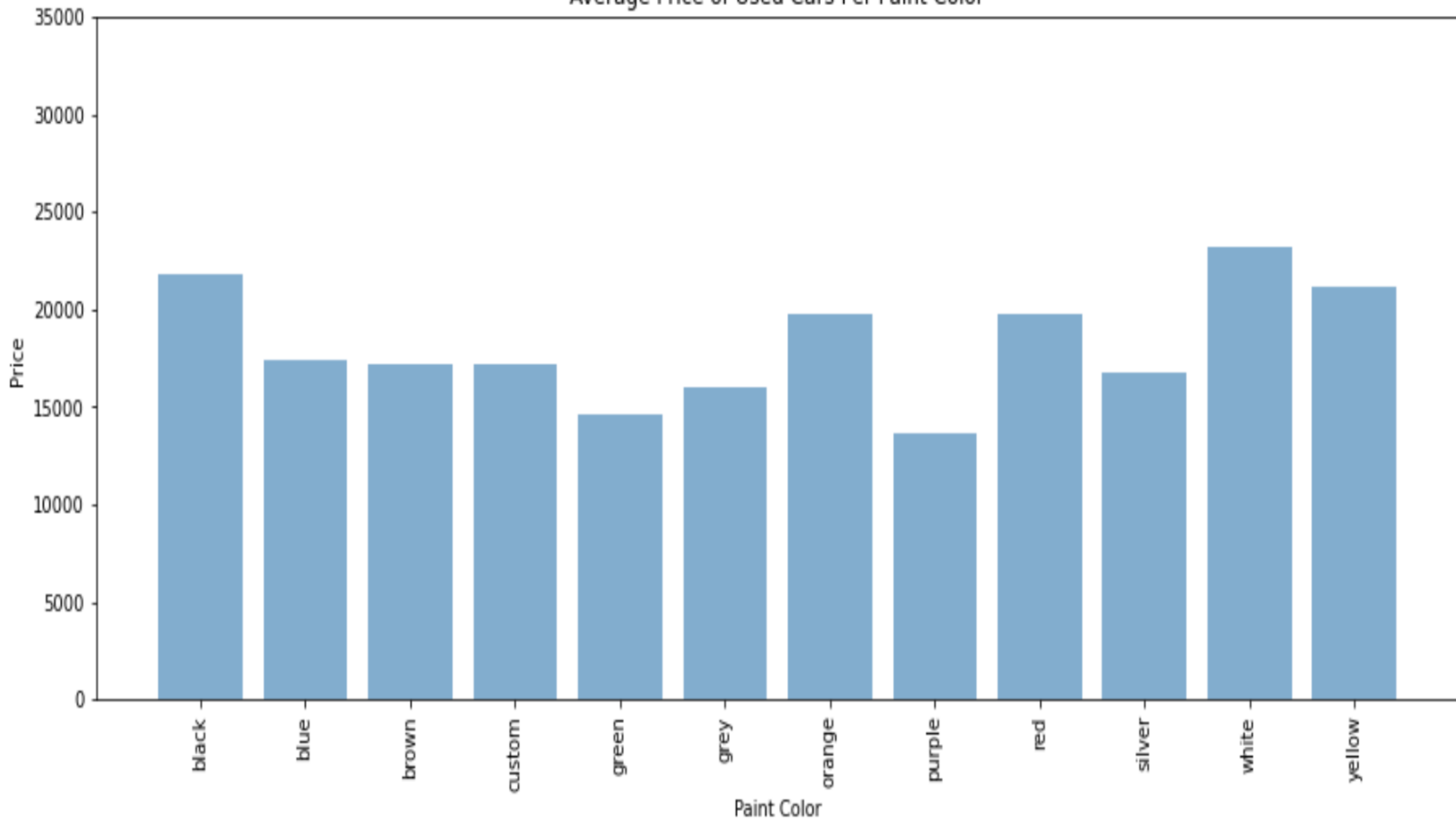


# ANALYSIS

- In the first chart, it describes the average price of a used car per manufacturer. As you can see, Ferrari is clearly the most expensive used car. Other expensive manufacturers on this chart include Aston-Martin, Porsche, and Tesla. The cheapest manufacturers on this list include Mercury, Pontiac, and Saturn.
- In the second chart, it describes the total # of used cars per manufacturer. In this chart, the most popular manufacturers are Chevrolet, Ford, and Toyota. The least popular manufacturers are Aston-Martin, Ferrari, Harley-Davidson, and Land Rover.
- Conclusion: It's tough to tell what the "best bang for your buck" used car is based on manufacturer. We dive deeper into this later to get a better answer.

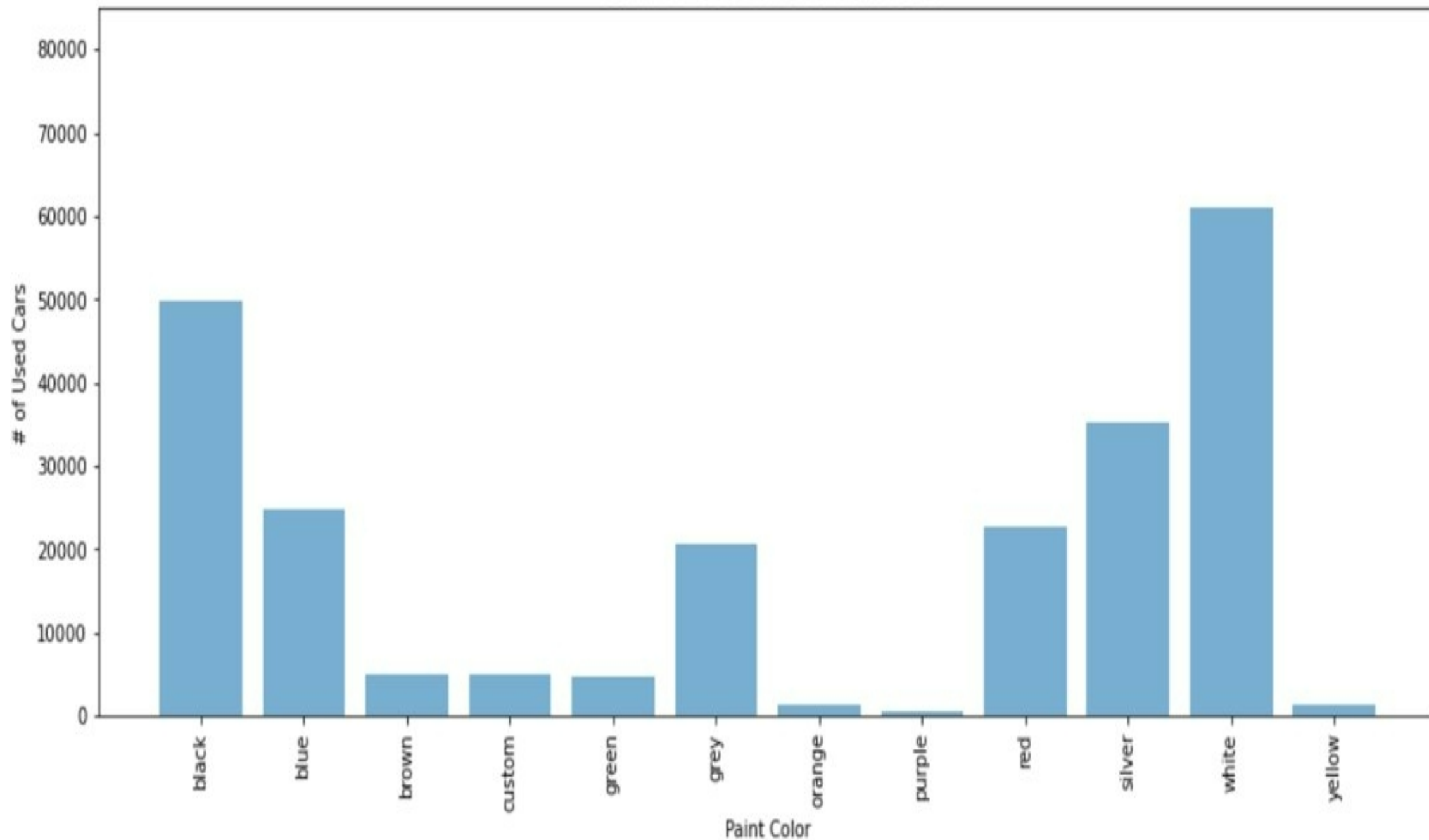
PAINT COLOR

Average Price of Used Cars Per Paint Color





Total # of Used Cars Per Paint Color

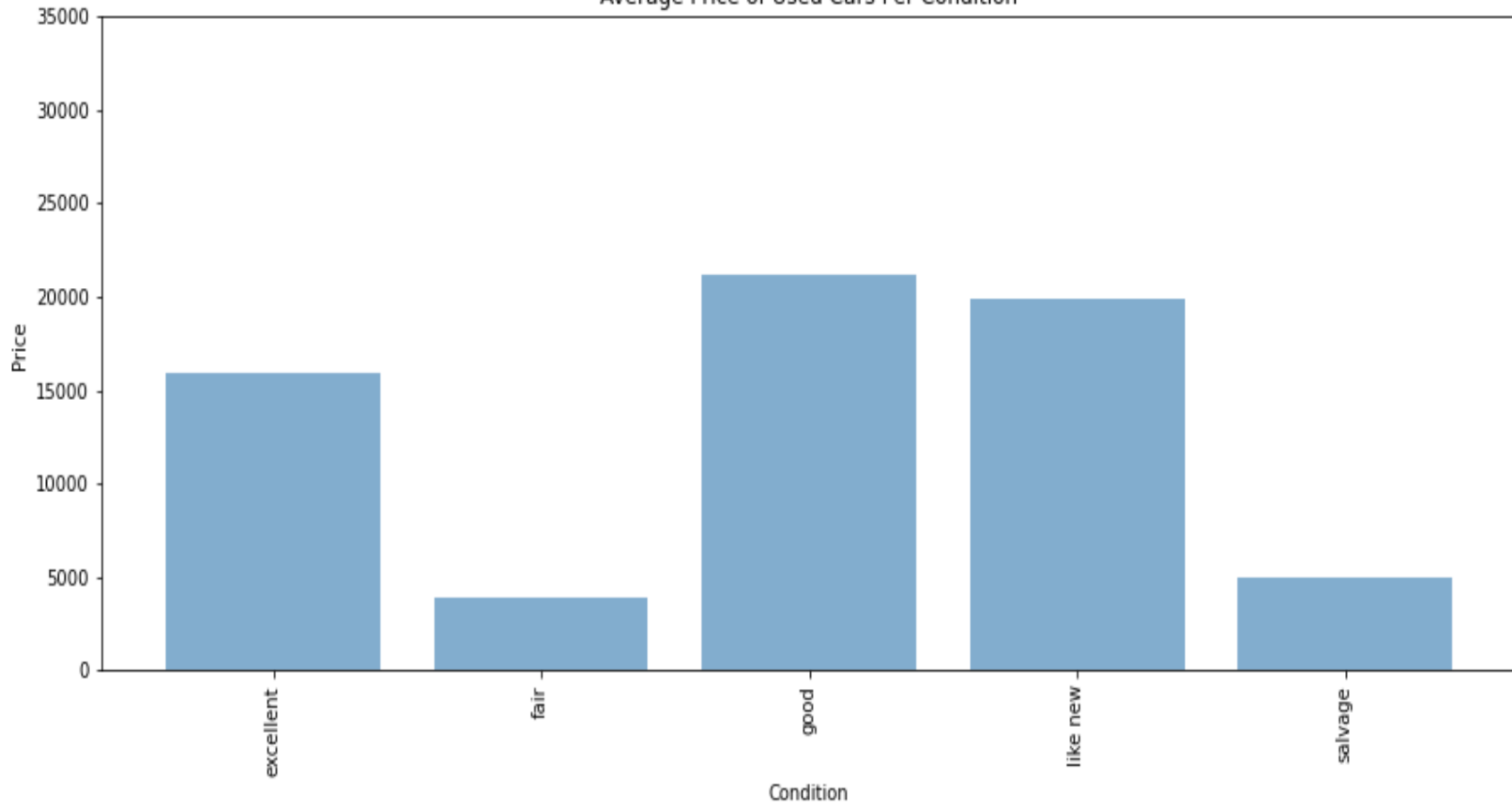


# ANALYSIS

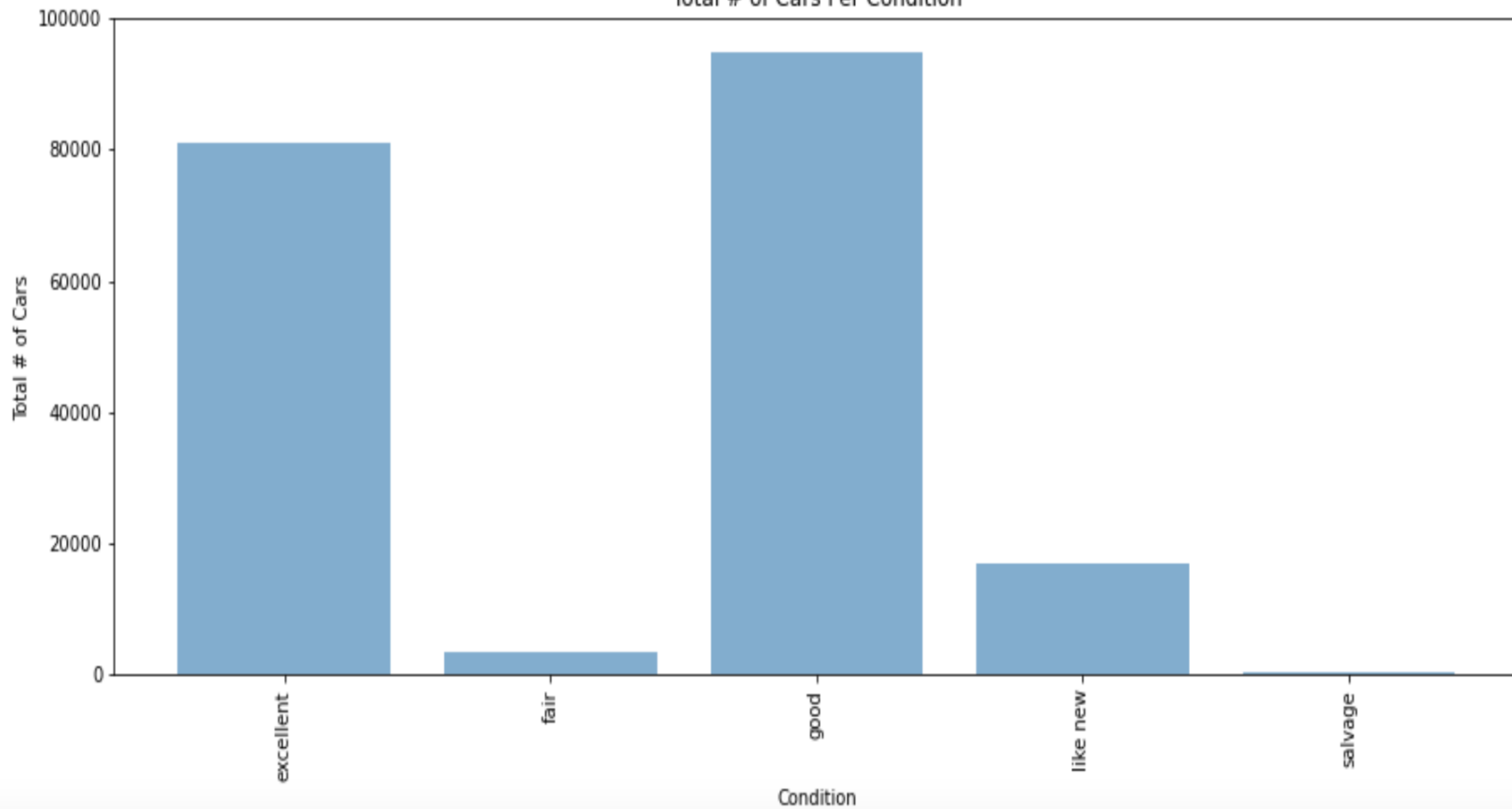
- In the first graph, it shows the average price of a used car per paint color. On average, the most expensive used car colors are black, white, and yellow.
- In the second graph, it shows the total # of used cars per paint color. The most popular used car colors are black, white, and silver.
- Conclusion: Because silver is so popular and it is NOT one of the most expensive used car colors, then we can confidently say the "best" used car color is a grey car.

CONDITION

Average Price of Used Cars Per Condition



Total # of Cars Per Condition



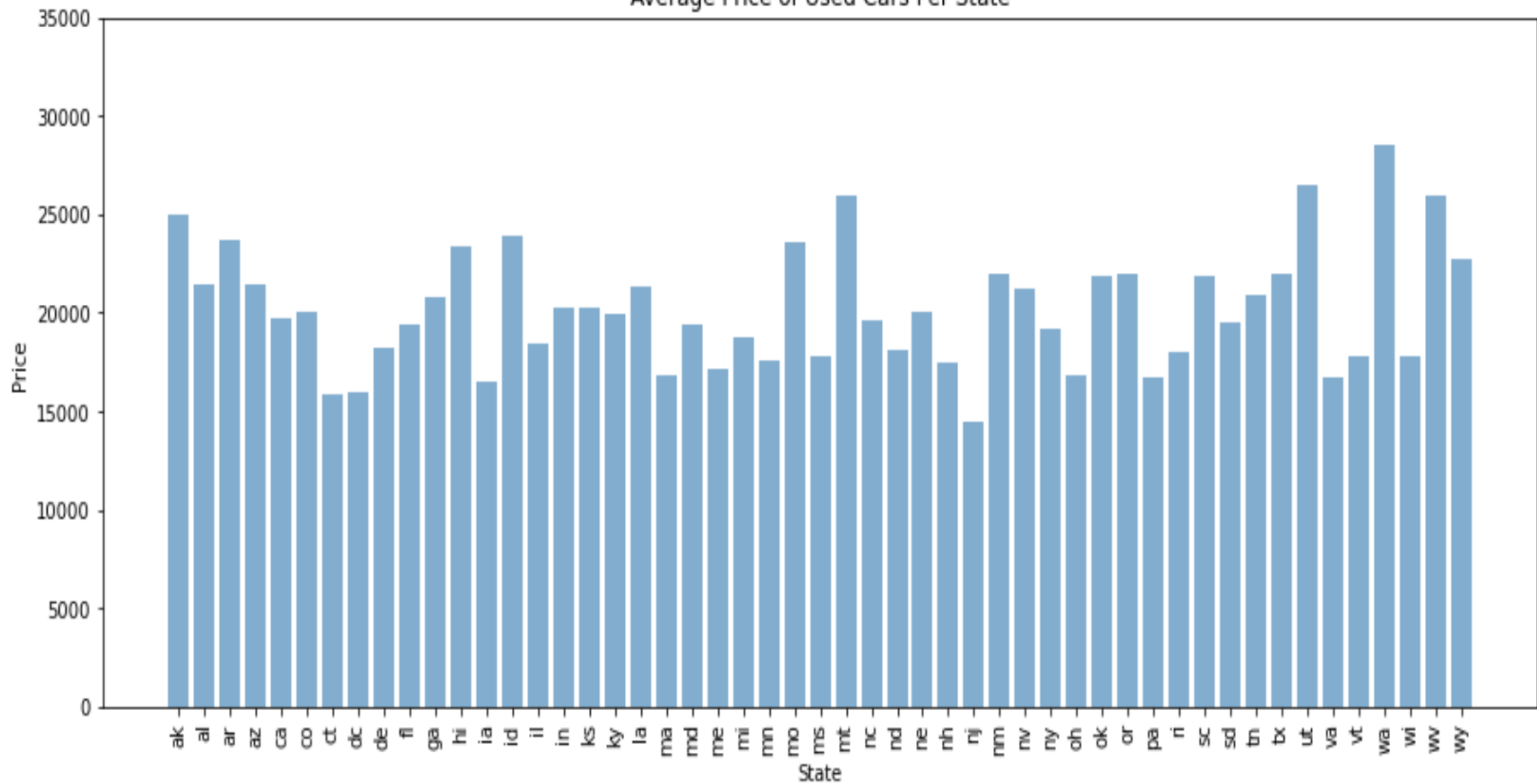
# ANALYSIS

- In the first graph, it shows the average price of a used car per paint condition. On average, the most expensive used car conditions are excellent, good, and like new. It was certainly interesting to see that a "good" conditioned used car was, on average, more expensive than both "excellent" and "like new" conditioned used cars.
- In the second graph, it shows the total # of used cars per condition. The most popular used car conditions are also excellent, good, and like new.
- Conclusion: Because "excellent" is the second most popular used car condition and it is ranked 3 out of 5 for most expensive used cars, then we can confidently say the "best" used car condition is an "excellent" car.

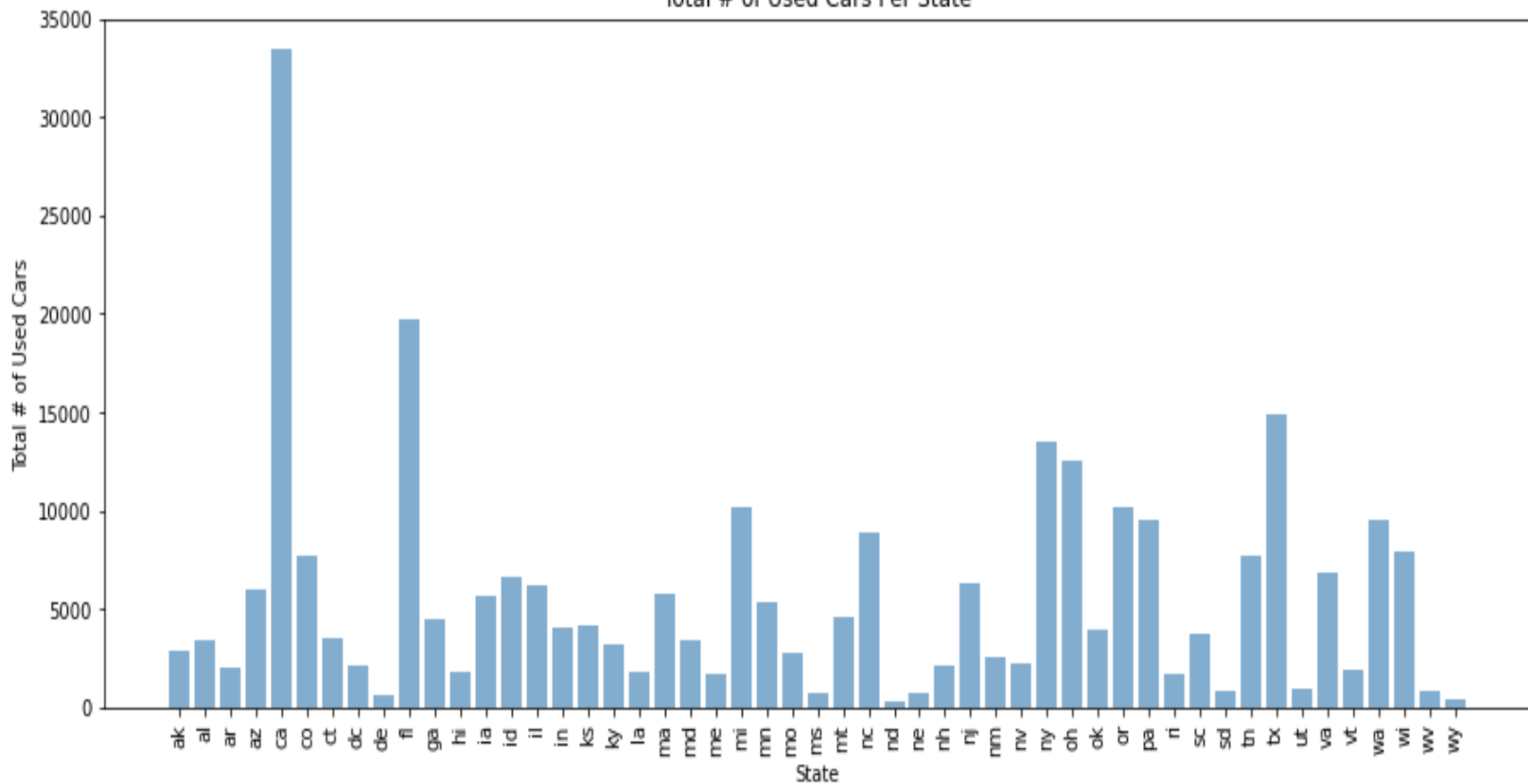


STATE

Average Price of Used Cars Per State



Total # of Used Cars Per State

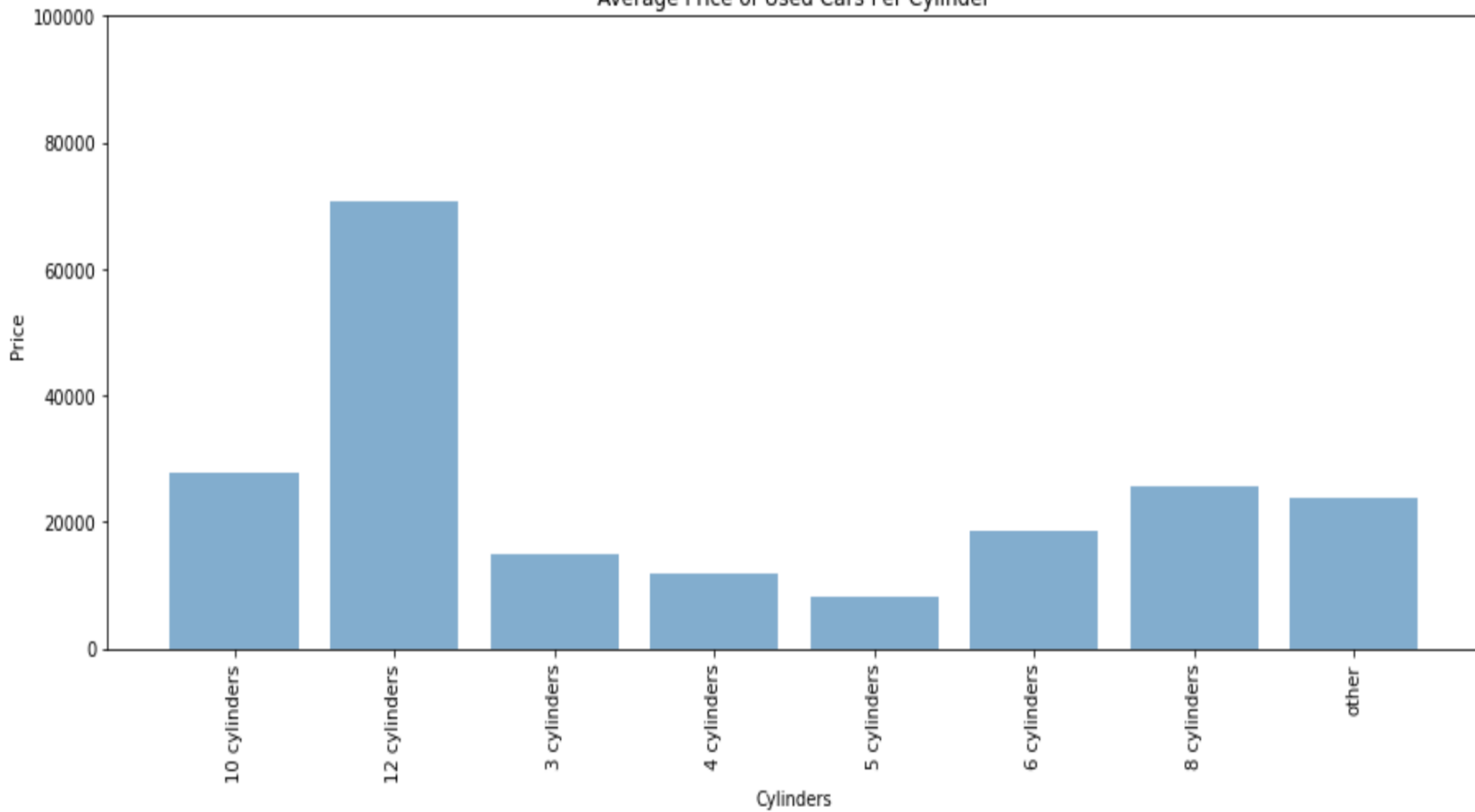


# ANALYSIS

- In the first graph, it shows the average price of a used car per state. On average, the most expensive used cars purchased in the US are Montana, Utah, Washington, and West Virginia.
- In the second graph, it shows the total # of used cars per state. The most popular used cars purchased in the US are California, Florida, New York, and Texas.
- Conclusion: Because California is the most popular state a used car is purchased and it is NOT in the top 50% of most expensive states to purchase a used car, then we can confidently say the "best" state to purchase a used car is in California.

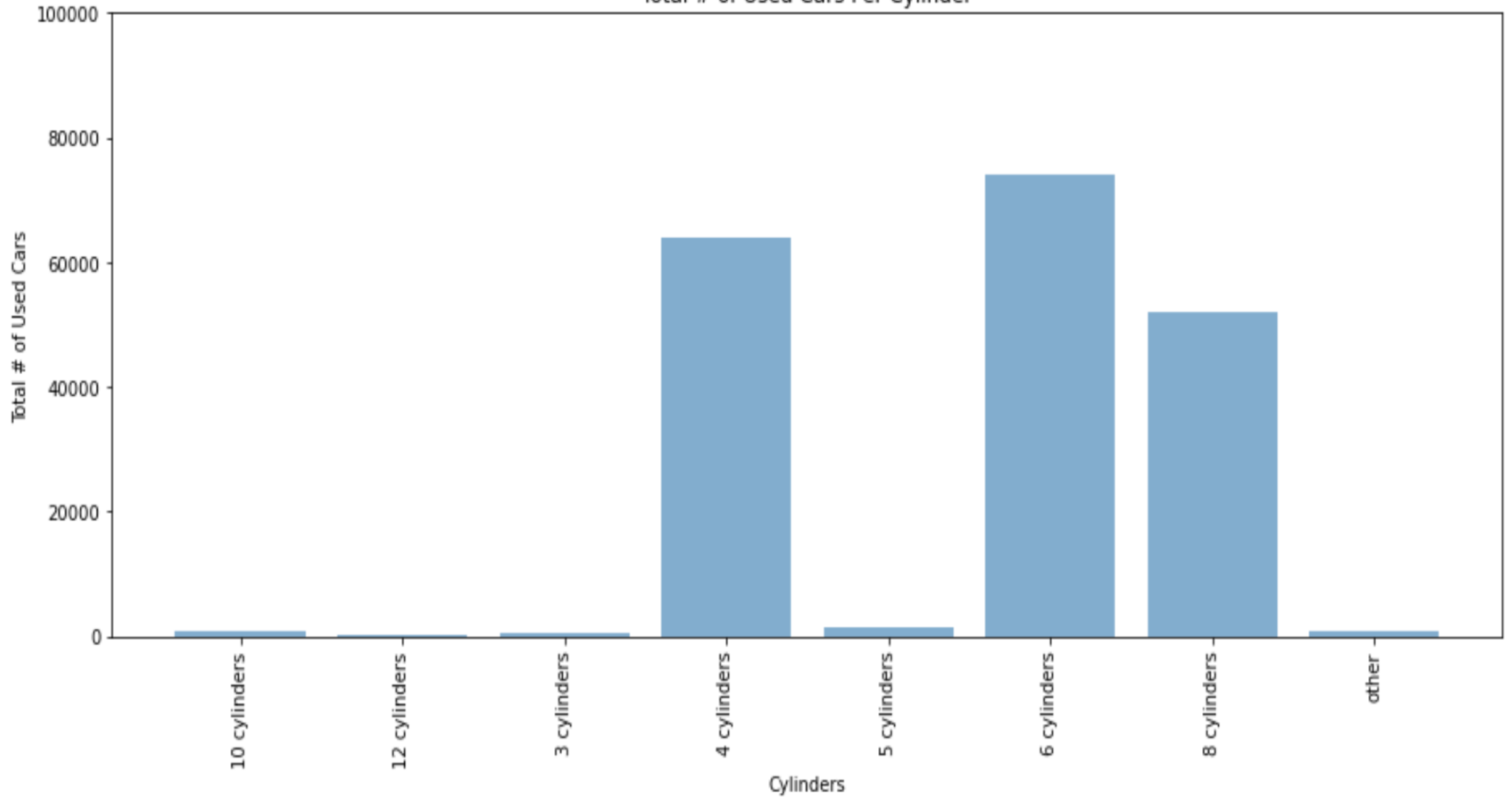
CYLINDER

Average Price of Used Cars Per Cylinder





Total # of Used Cars Per Cylinder

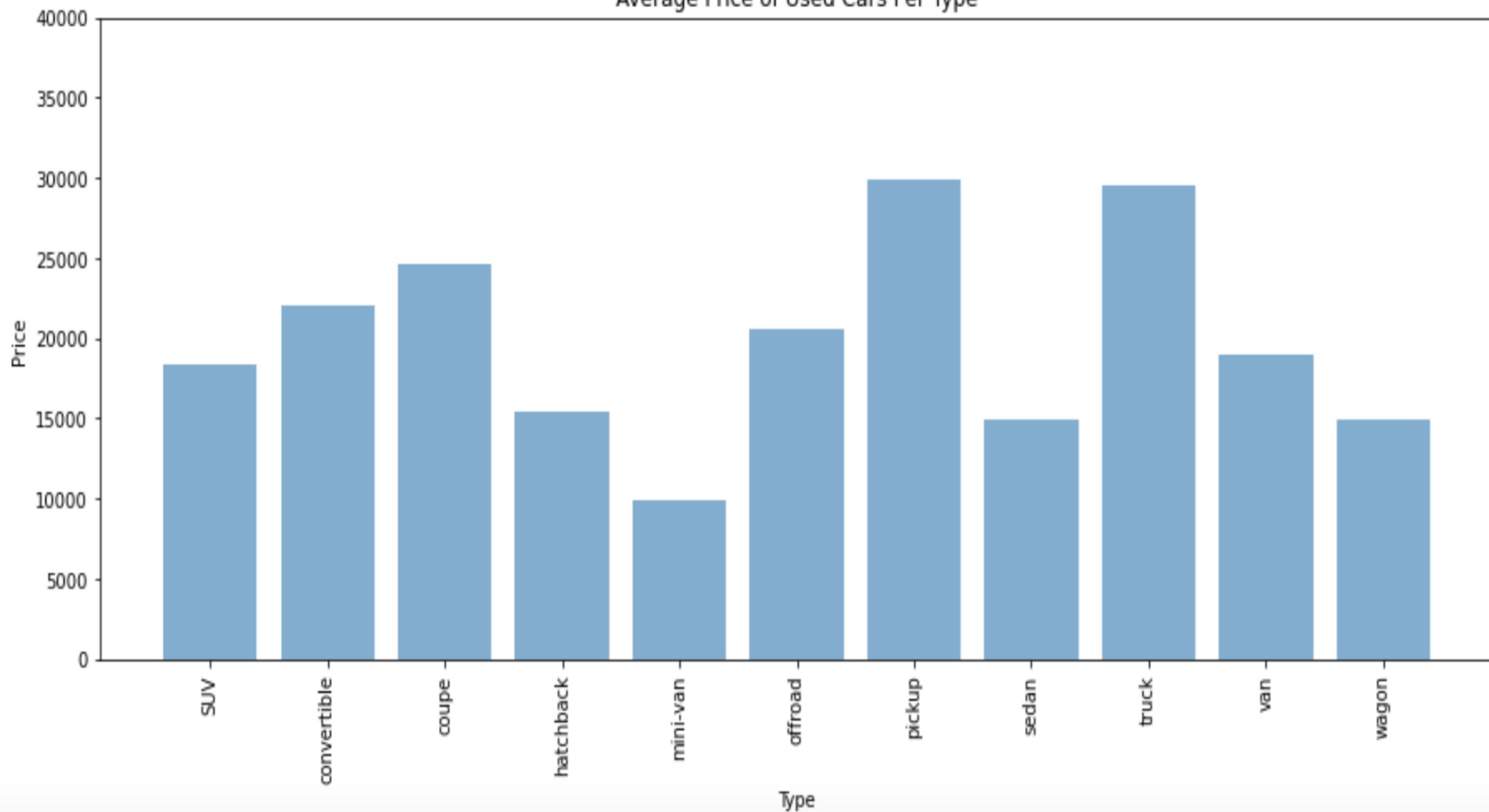


# ANALYSIS

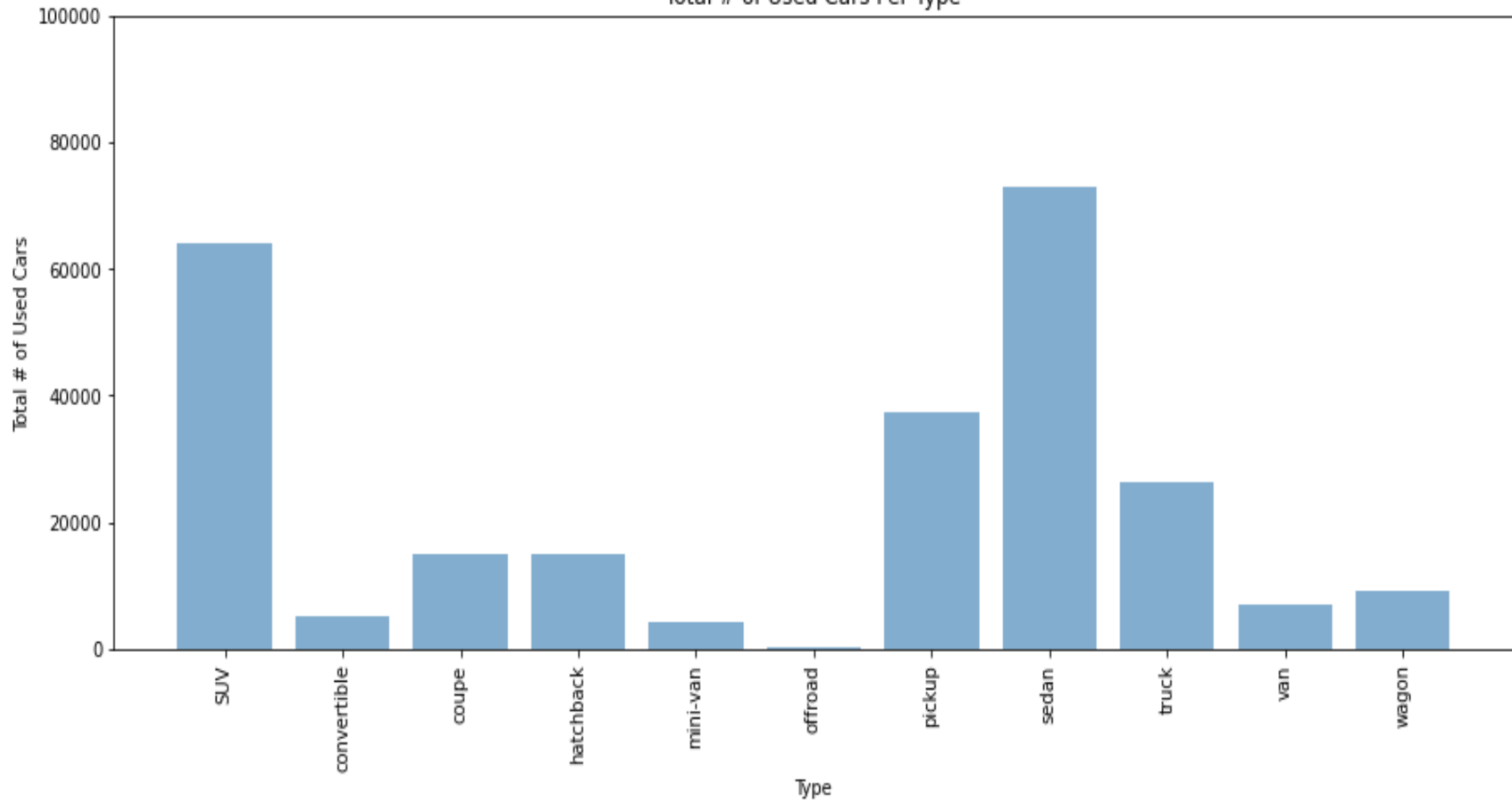
- In the first graph, it shows the average price of a used car per cylinder. On average, the most expensive used cars per cylinder are 6-, 8-, 10-, and 12-cylinder cars.
- In the second graph, it shows the total # of used cars per cylinder. The most popular used cars per cylinder are 4-, 6-, and 8-cylinder cars.
- Conclusion: Because a 6-cylinder car is the most popular car per cylinder and it is ranked 4 out of 8 in terms of price, we can confidently say the "best" used car cylinder is 6-cylinder car.

TYPE

Average Price of Used Cars Per Type



Total # of Used Cars Per Type



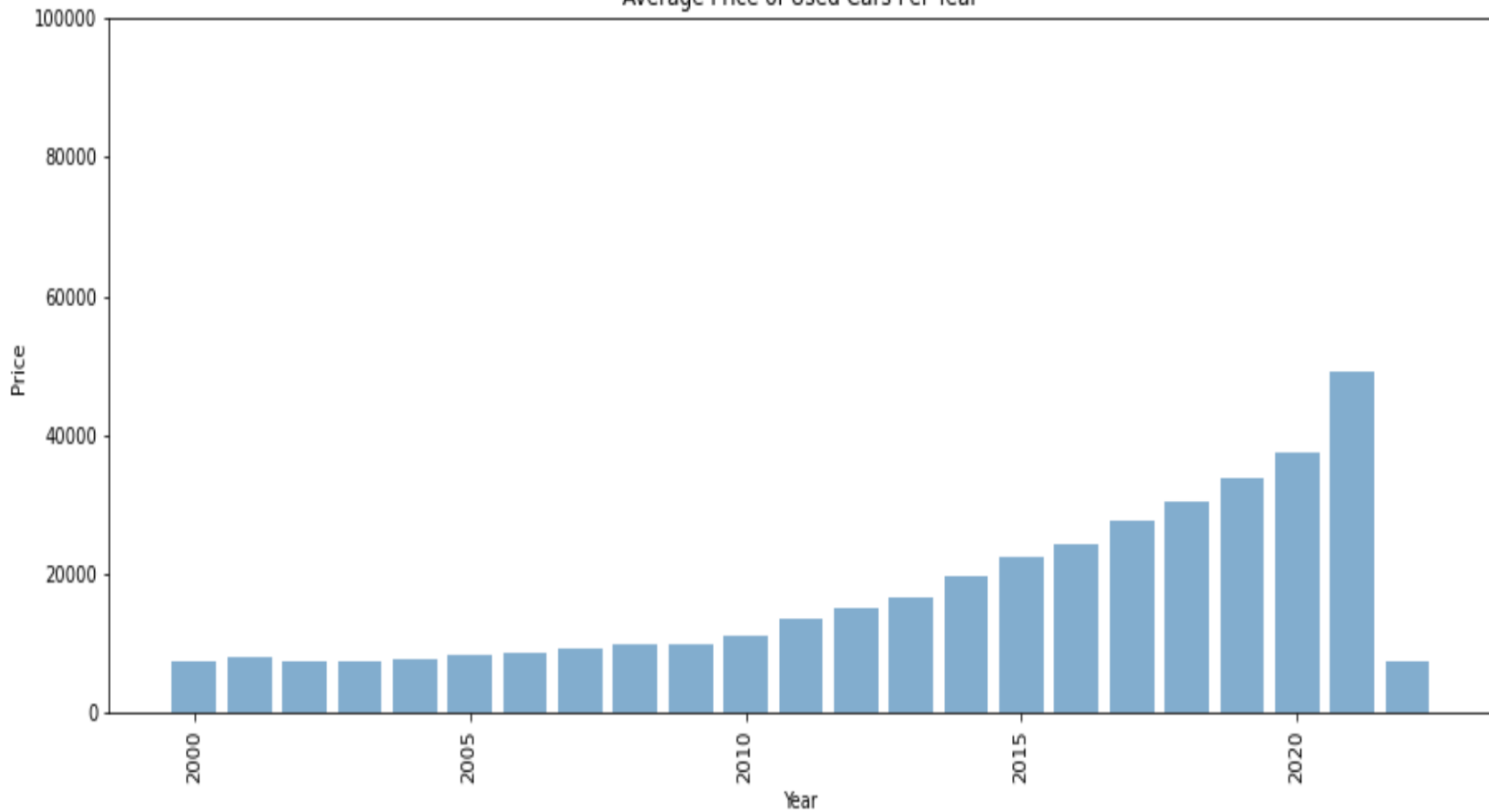
# ANALYSIS

- In the first graph, it shows the average price of a used car per type. On average, the most expensive used cars per type are coupe, truck, and pickup cars. The least expensive used cars per type are hatchback, mini-van, and sedan cars.
- In the second graph, it shows the total # of used cars per type. The most popular used cars per type are SUVs, pickup, and sedan cars.
- Conclusion: Because Sedans and SUVs are one of the least expensive used car types and are one of the most popular used car types, then we can confidently say the "best" used car type are SUVs and sedans. We dig further into this to get a more definitive answer as to the best car types.

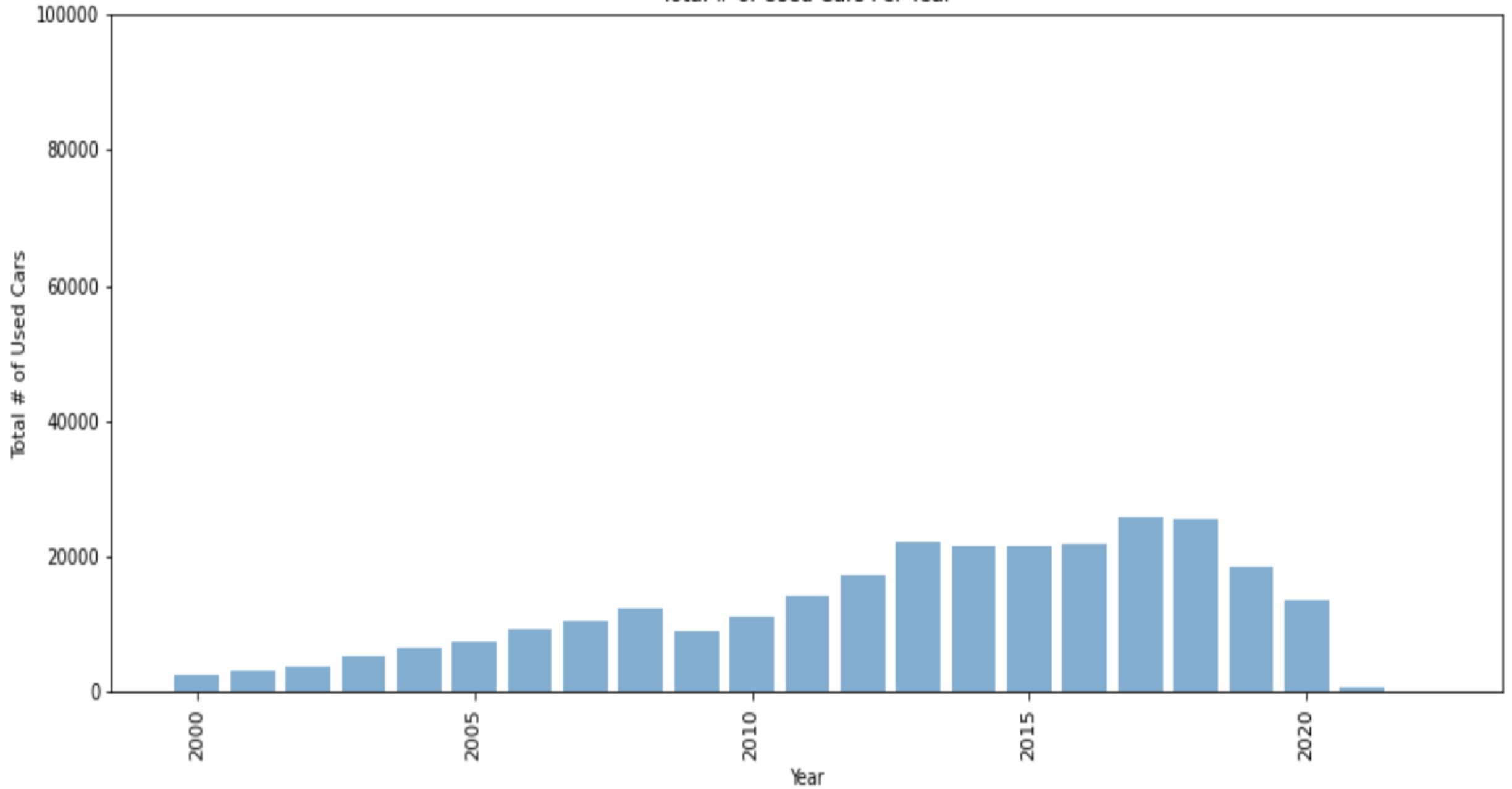


YEAR

Average Price of Used Cars Per Year



Total # of Used Cars Per Year

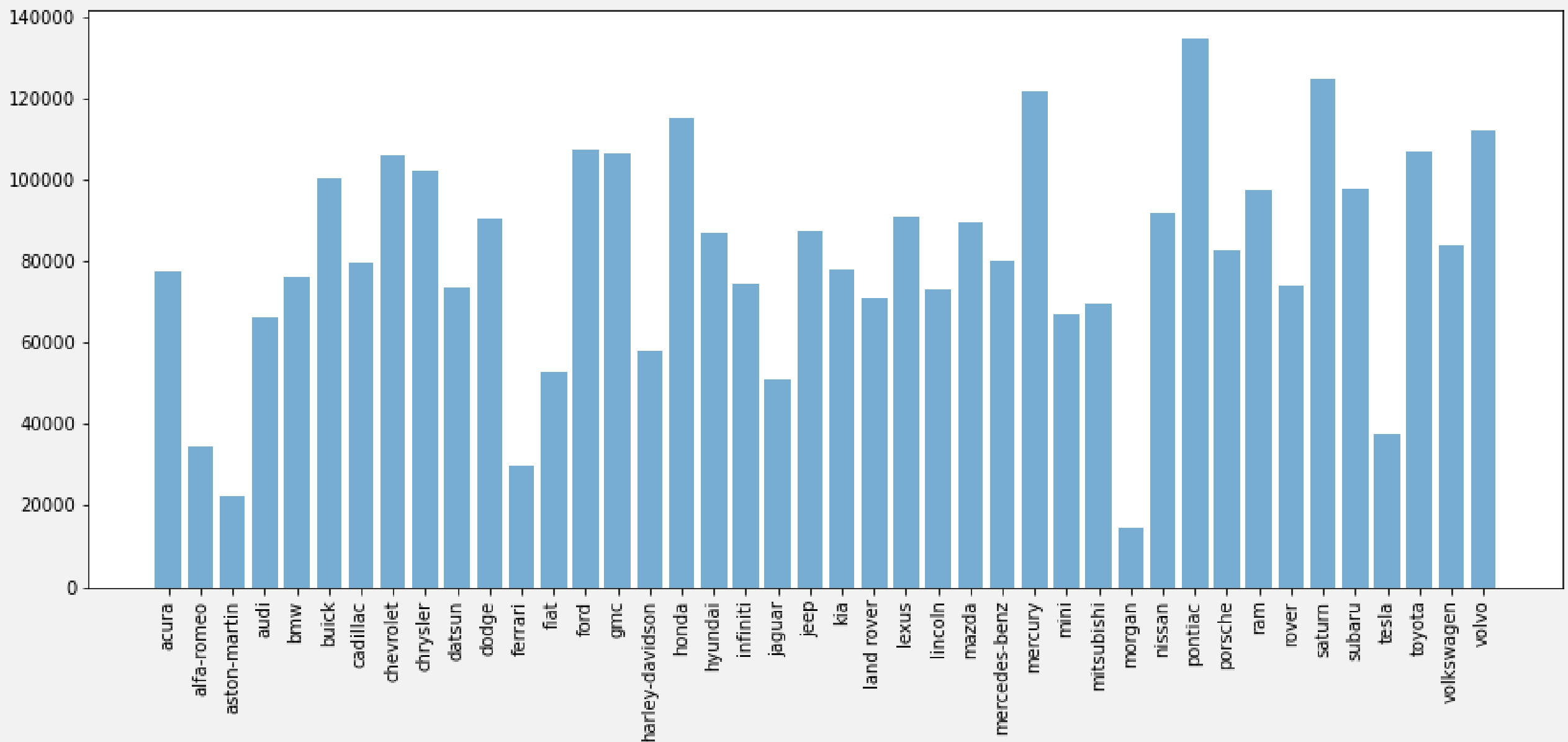


# ANALYSIS

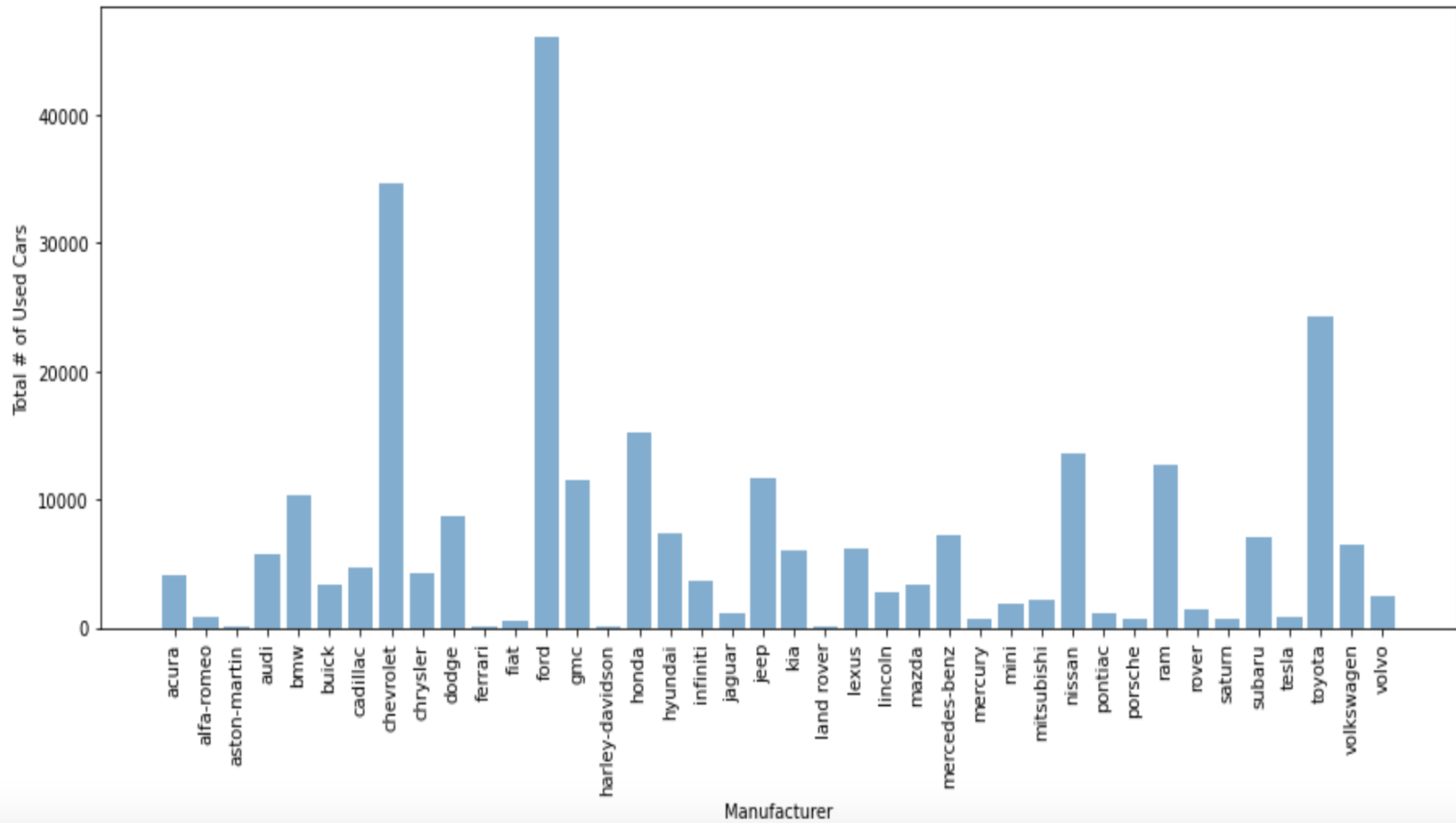
- In the first graph, it shows the average price of a used car per year. On average, the most expensive used cars per year are 2018, 2019, 2020, and 2021. This is self-explanatory.
- In the second graph, it shows the total # of used cars per year. The most popular used cars per year are 2013, 2017, and 2018.
- Conclusion: Because 2013 is the third most popular used car and it is NOT one of the most expensive cars per year, we can confidently say the "best" used car year is a 2013 car.

**MILEAGE**

Average Mileage of Used Cars Per Manufacturer



Total # of Used Cars Per Manufacturer



# ANALYSIS

- In the first graph, it shows the average mileage of a used car per manufacturer. The top four manufacturers with the highest average mileage are Honda, Mercury, Pontiac, and Saturn.
- In the second graph, it shows the total # of used cars per manufacturer. The top four are Chevrolet, Ford, Honda, and Toyota.
- Conclusion: Because Honda is one of the most popular used cars per manufacturer and it has one of the top average mileages of a used car per manufacturer, we can confidently say the "best" used car based on mileage per manufacturer is a Honda. The average Honda's mileage is around 115,000 miles.



## WHAT ARE THE MOST DRIVEN USED CAR TYPES BY MANUFACTURER?

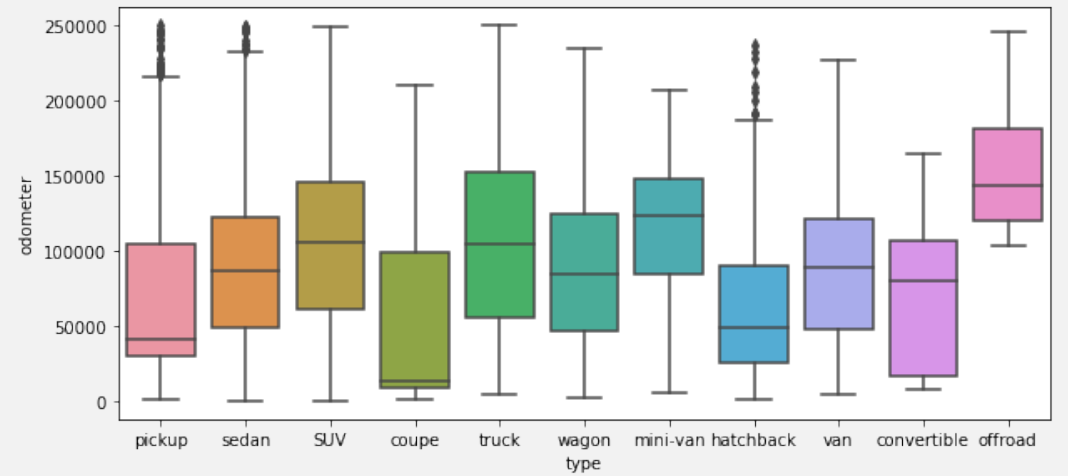
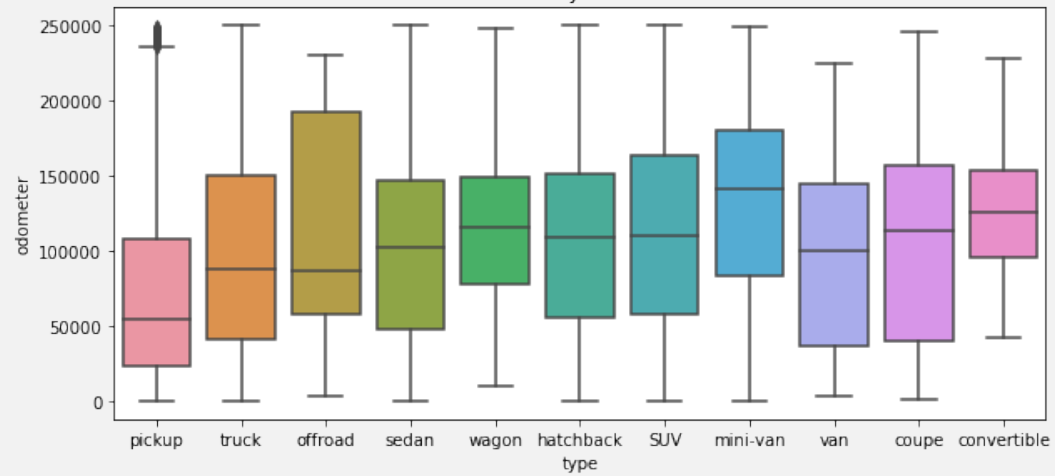
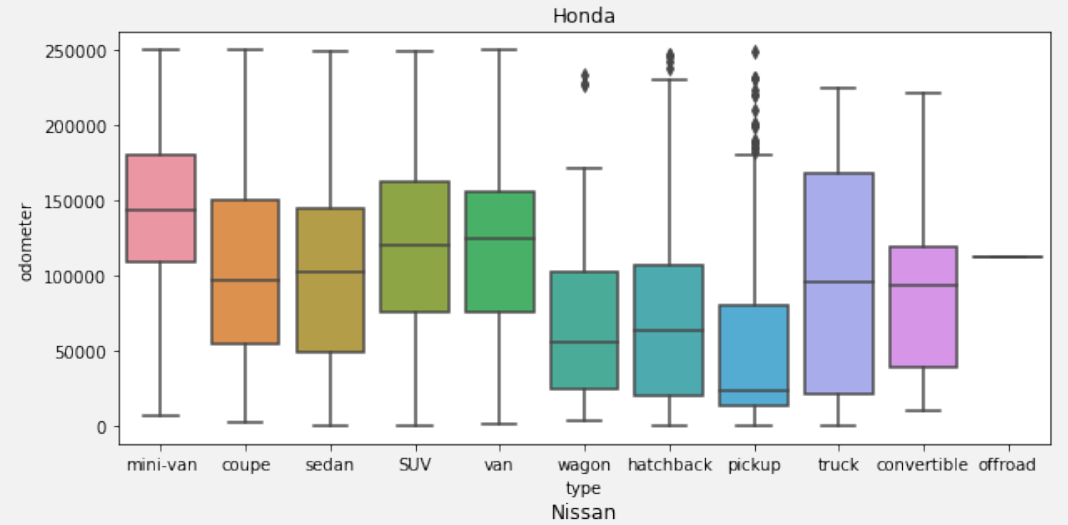
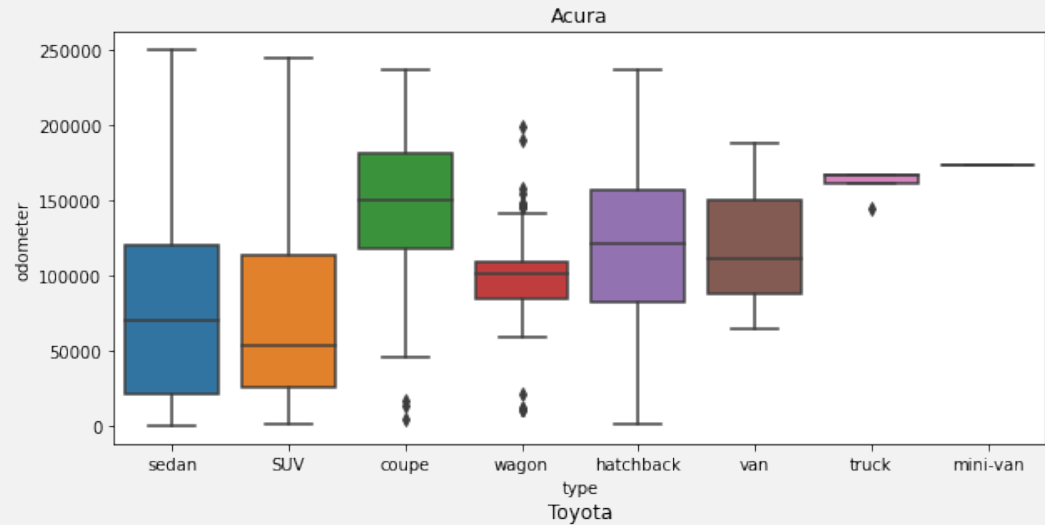
In this section of our analysis, we wanted to focus on what types of cars were driven the most for each manufacturer by comparing type of car vs odometer readings using a boxplot to capture median, IQR, and outliers. This is because different manufacturers have different specialties. For instance, Subaru is known for their hatchbacks while companies like Ram and Dodge are notorious for their pickups. In this analysis we used the assumption that the higher the median, the more popular a car was. It is also important to take sample size into account.

Additionally, we grouped manufacturers by country of origin, as similar companies have similar expectations from the consumer (i.e., Japanese cars are known to hold better value).

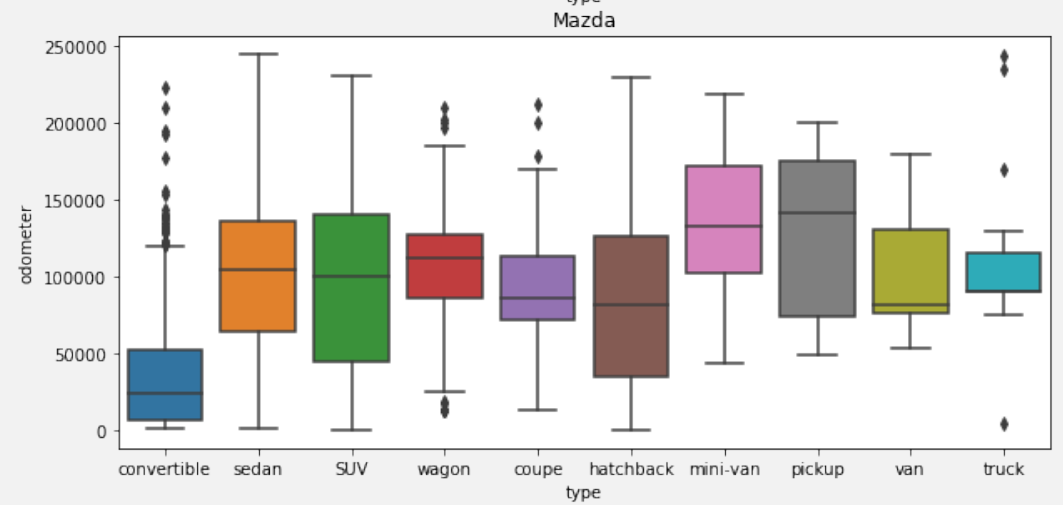
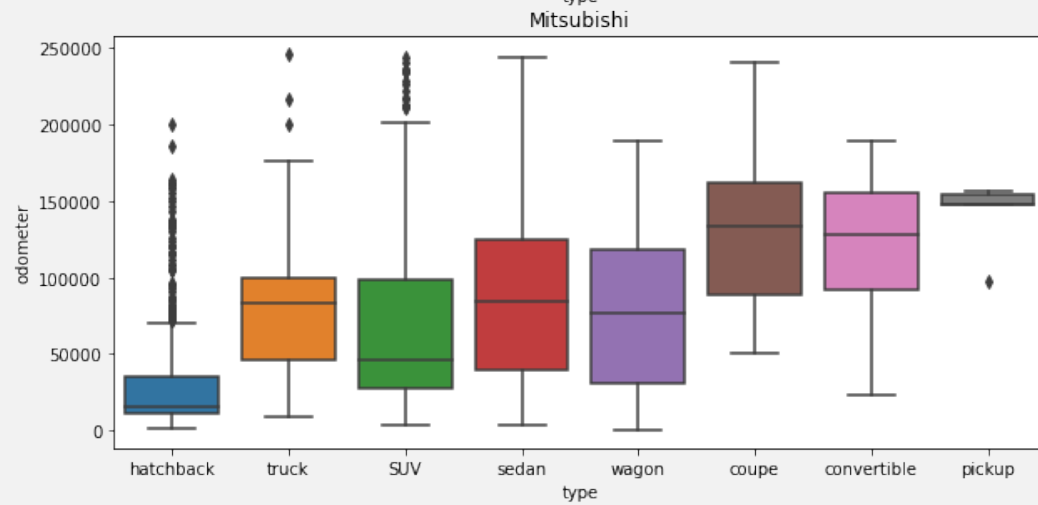
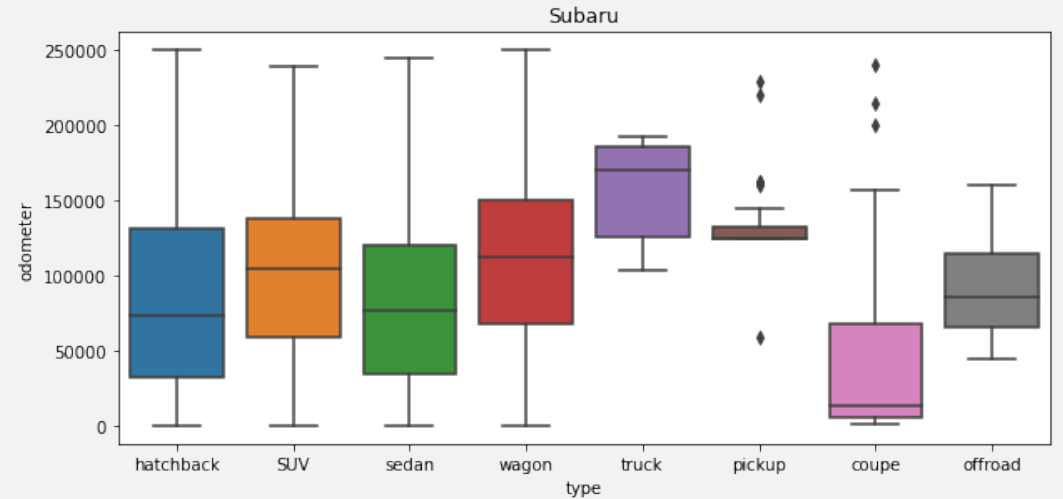
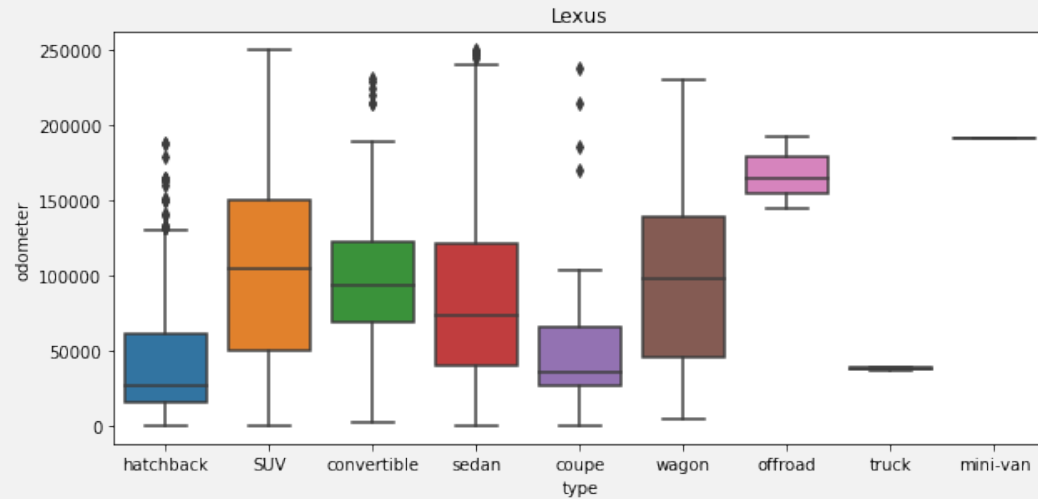


# JAPANESE CARS

Japanese Car Brands Mileage by Type



Japanese Car Brands Mileage by Type

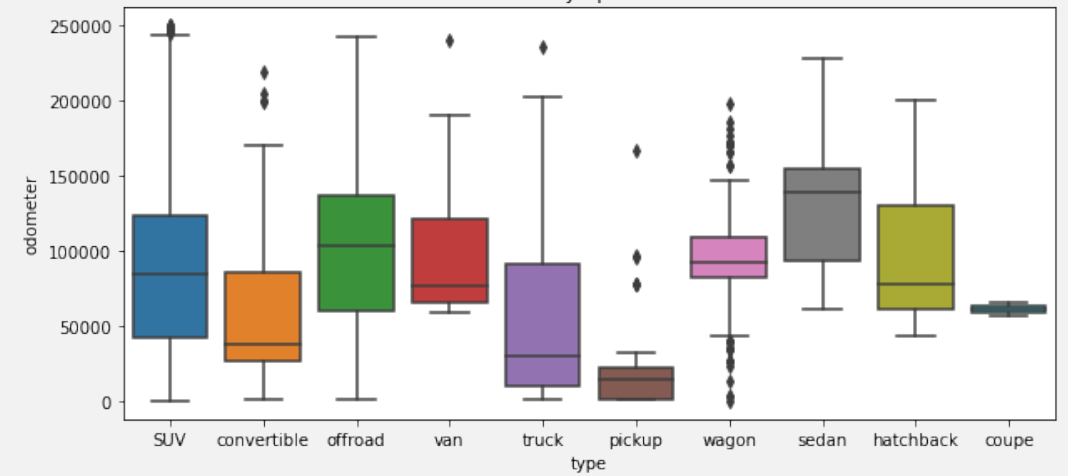
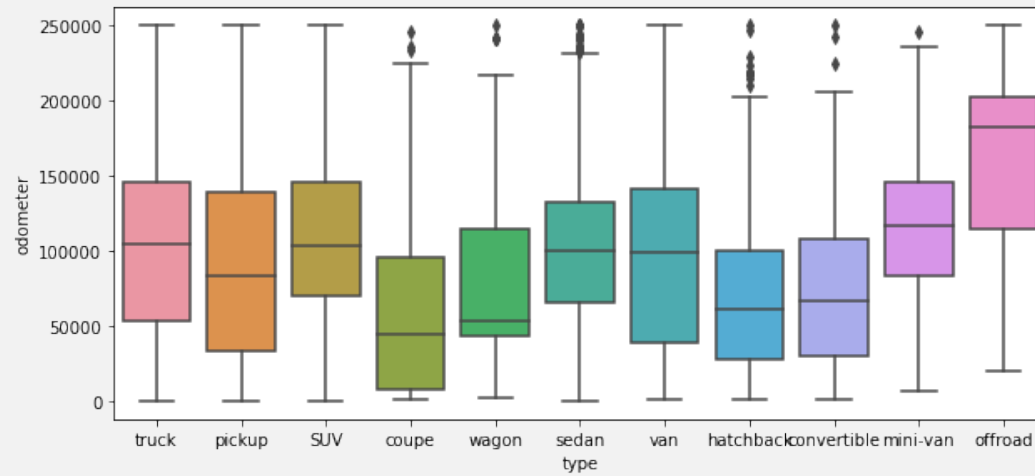
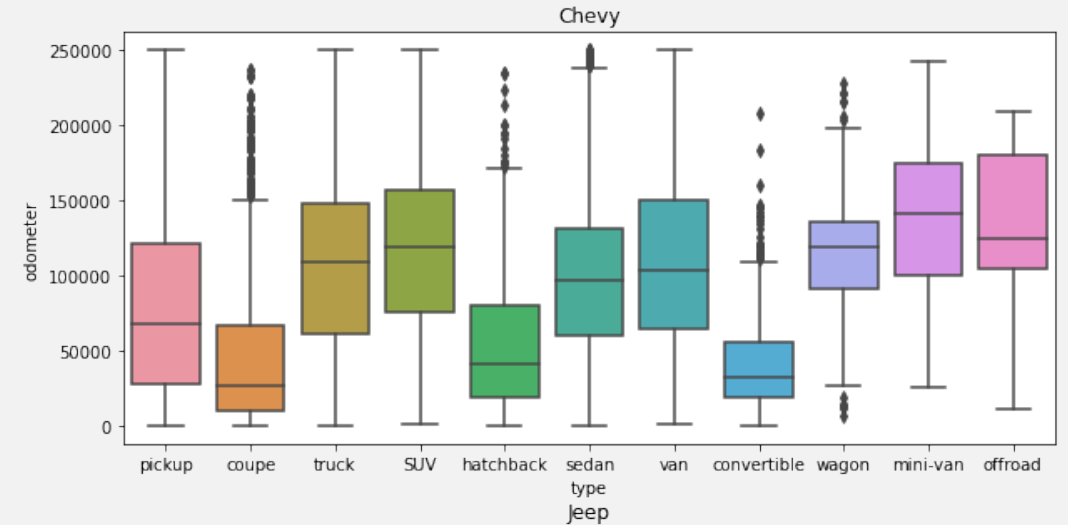
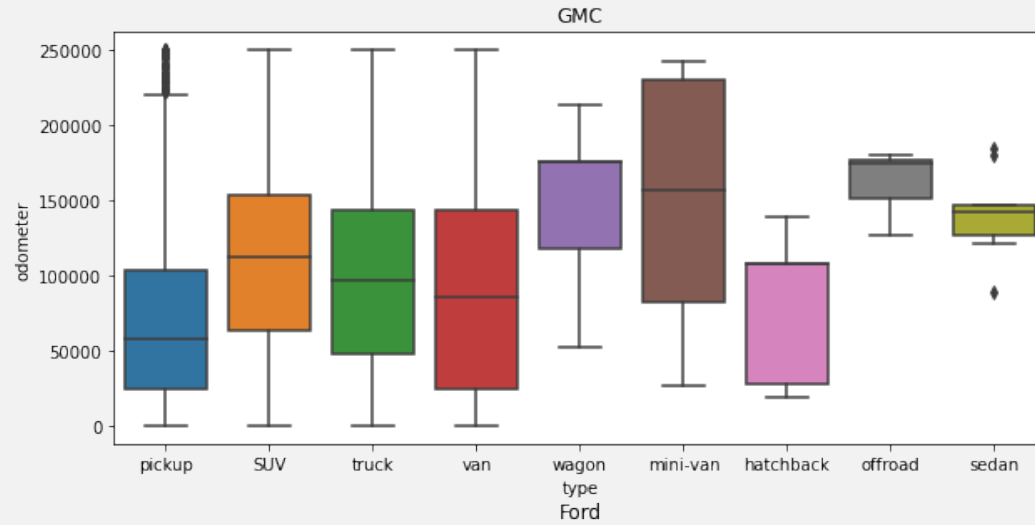


# ANALYSIS

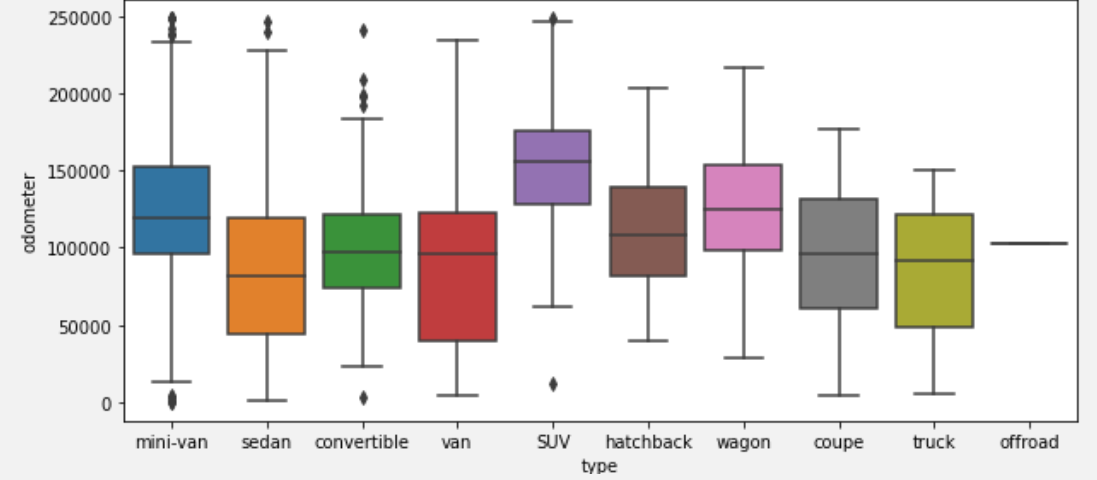
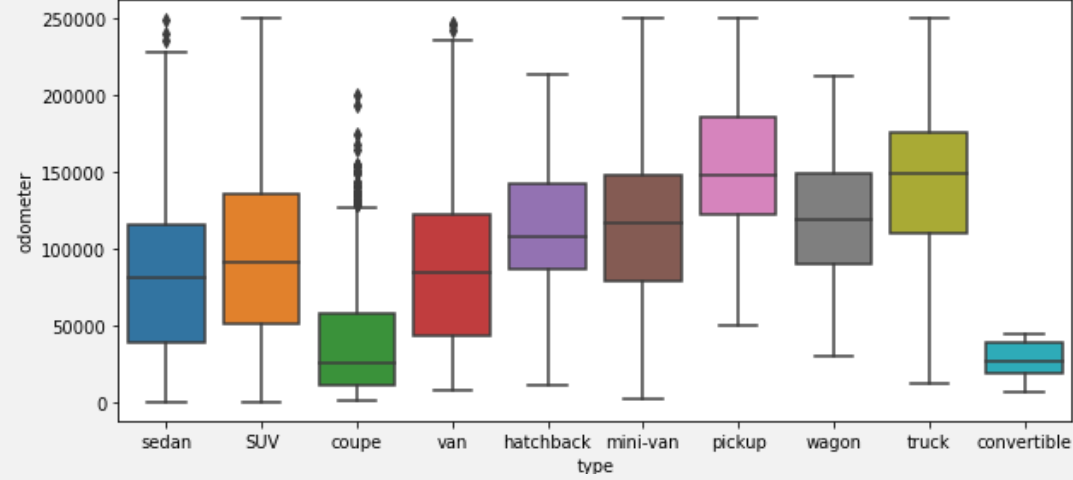
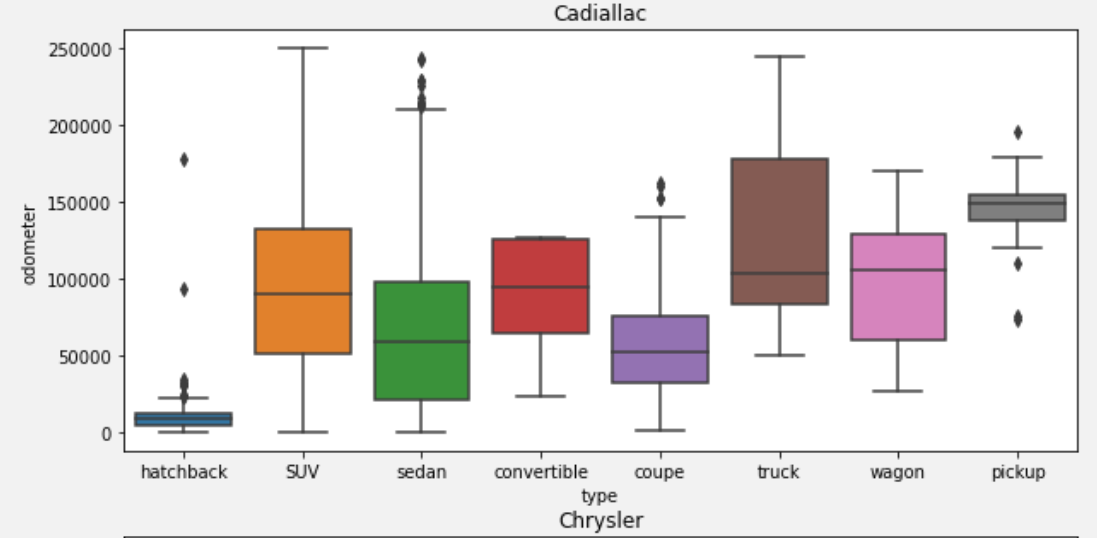
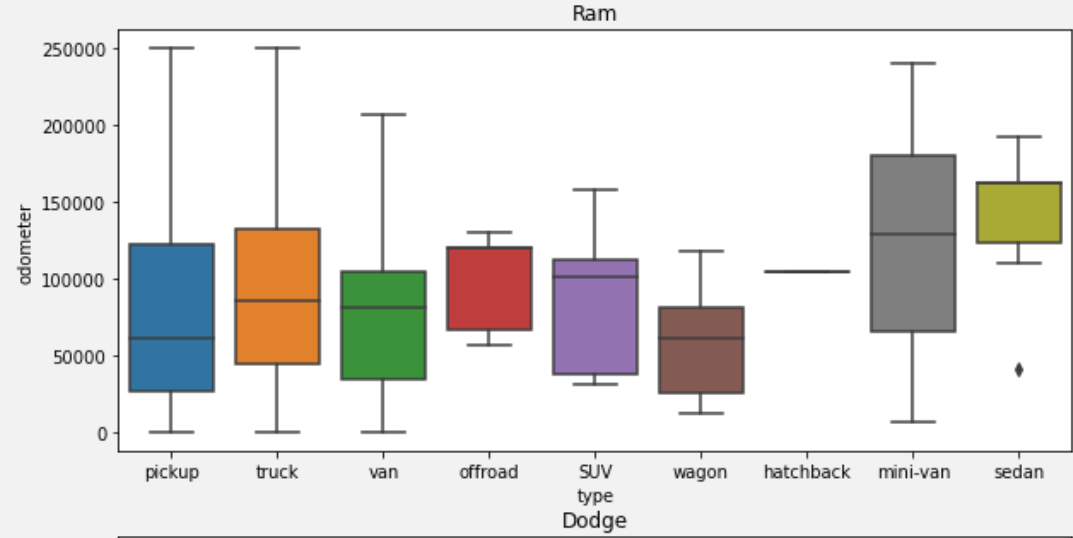
- For Japanese cars, it appears that more family-oriented such as coupes, sedans, SUVs, and mini vans are the most driven. It is interesting to note that Subaru's hatchbacks were not their most driven car as the median miles were less than both SUVs and Wagons, and that they are only the 3rd highest listed Subaru in the dataset. It appears that the most driven Japanese cars are Honda mini vans, as their median miles is approximately 150,000.

# AMERICAN CARS

American Car Brands Mileage by Type

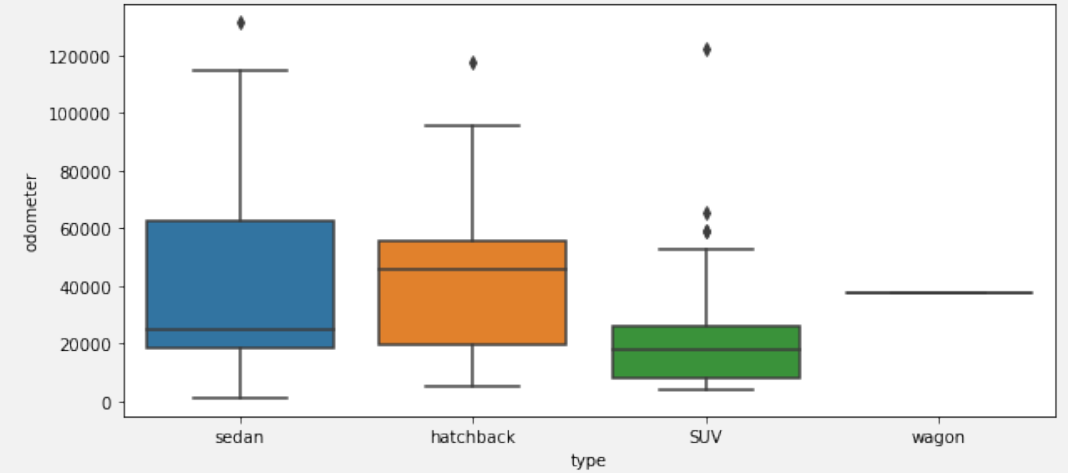
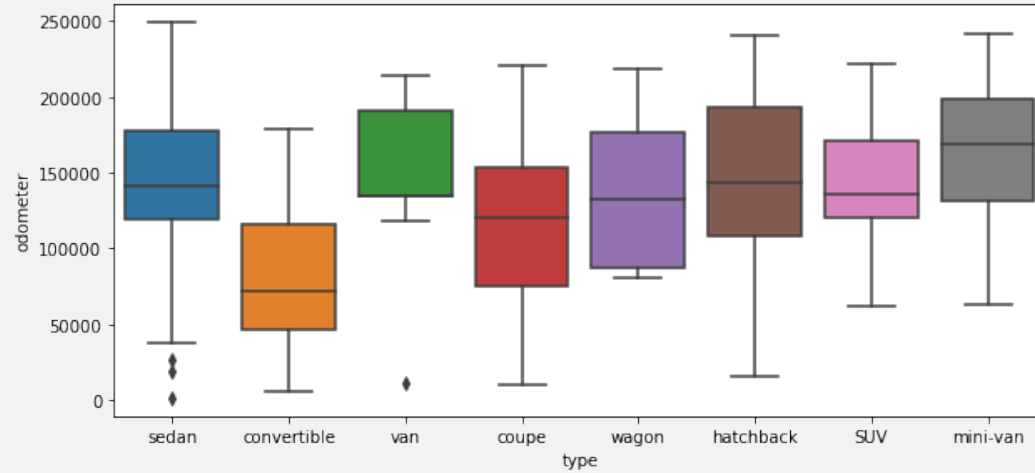
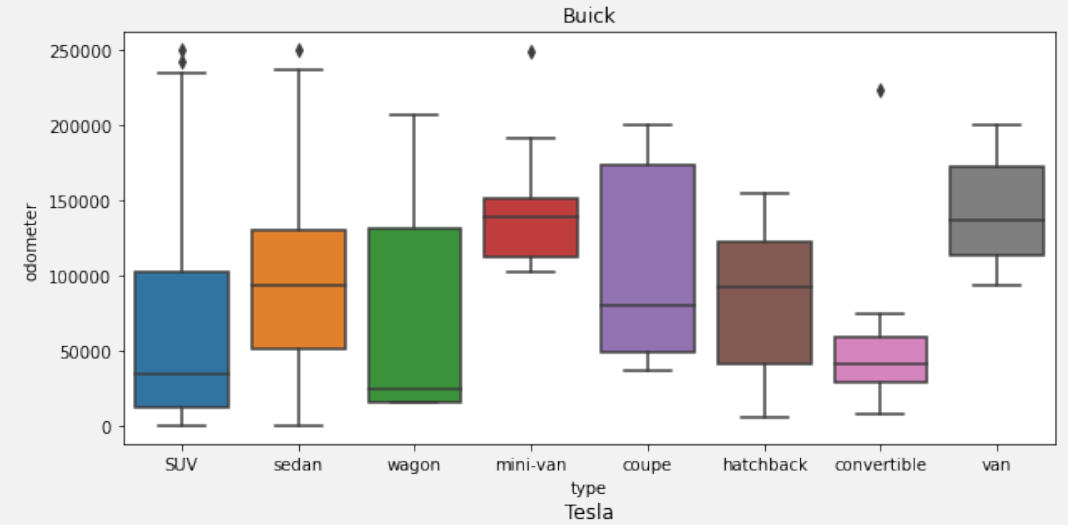
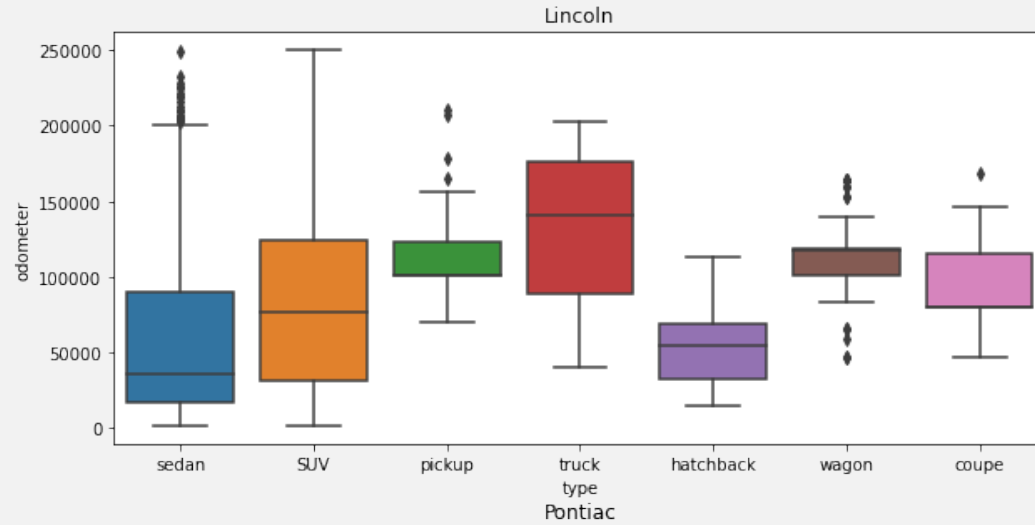


American Car Brands Mileage by Type

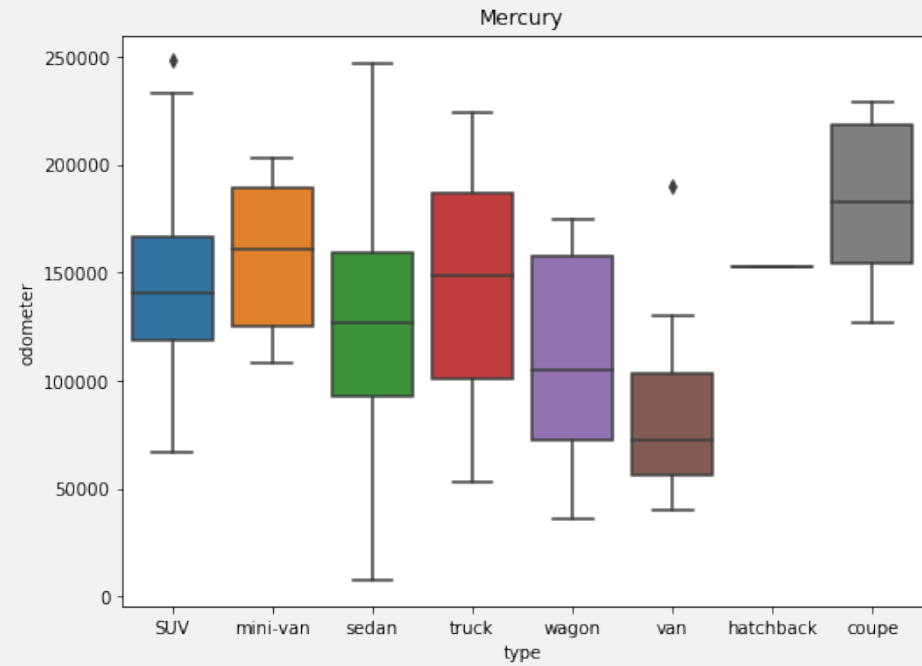
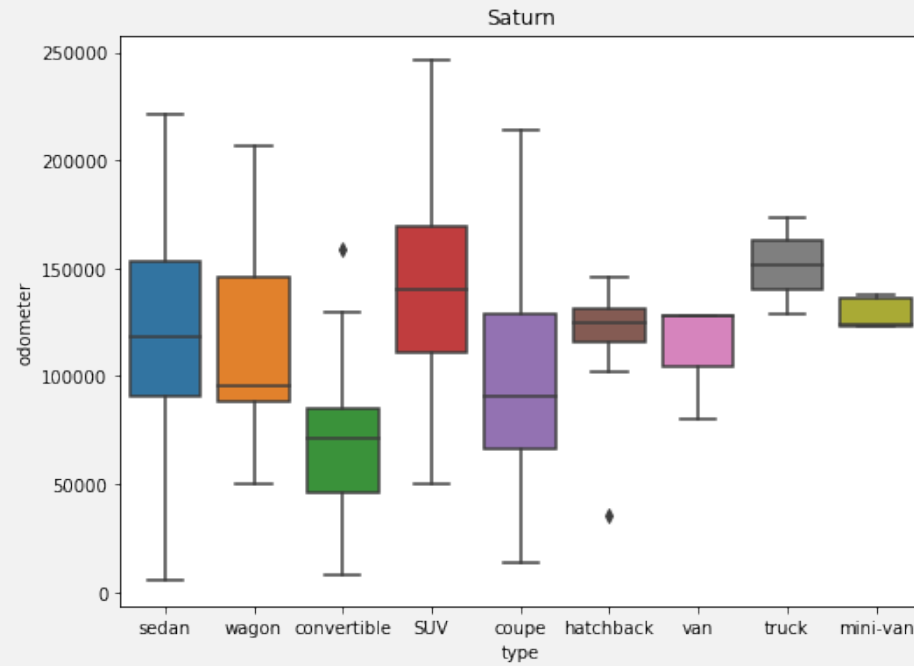




American Car Brands Mileage by Type



American Car Brands Mileage by Type

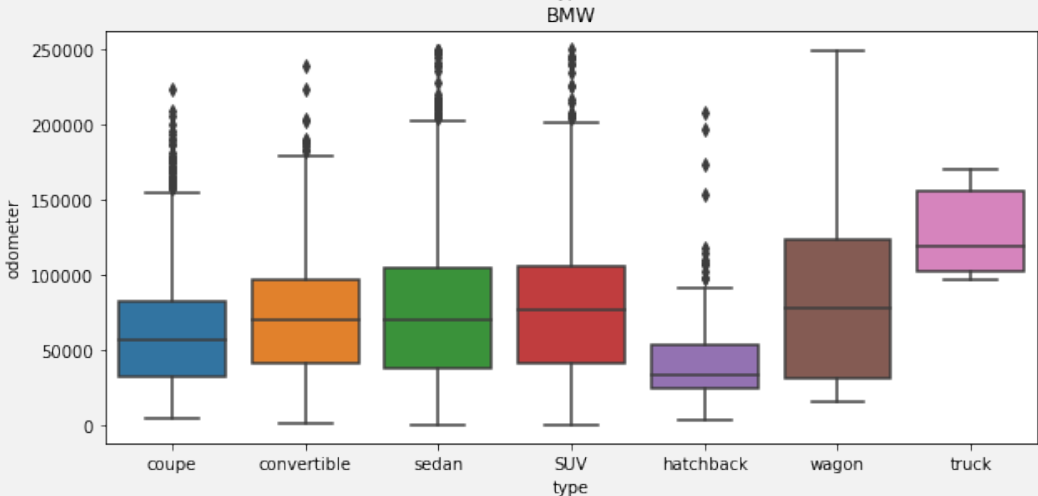
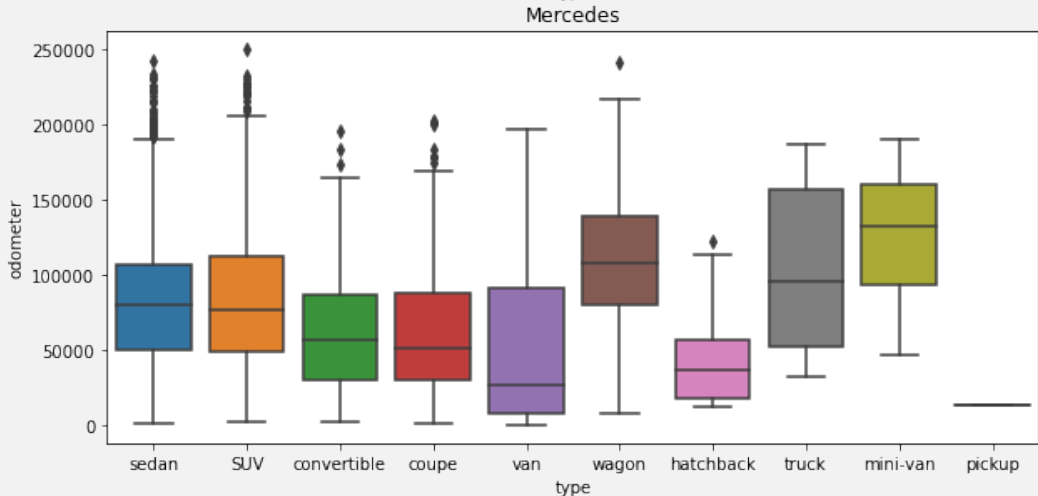
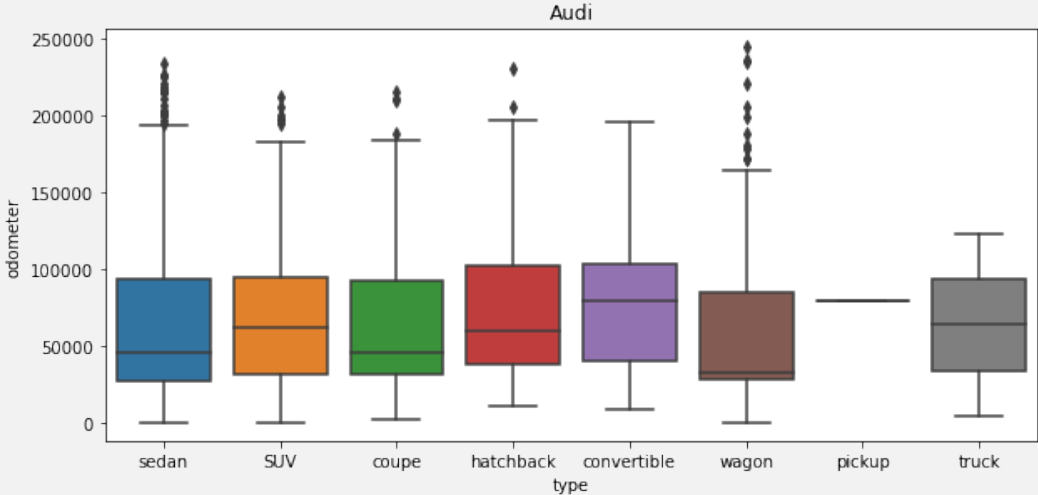
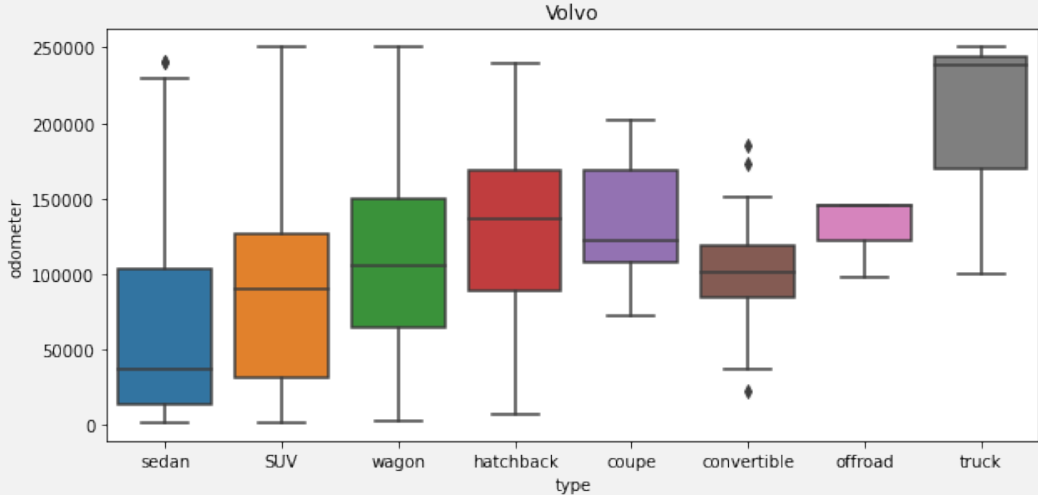


# ANALYSIS

- For American cars, it appears that trucks and SUVs have the highest median mileage. It is interesting to note that many of the median miles were lower than those of Japanese counterparts. This goes along with the common theory that Japanese hold better value long term. It is also interesting to note that there are 16 American manufacturers, which is double that of the next largest, Japanese manufacturers.

# GERMAN CARS

German Car Brands Mileage by Type

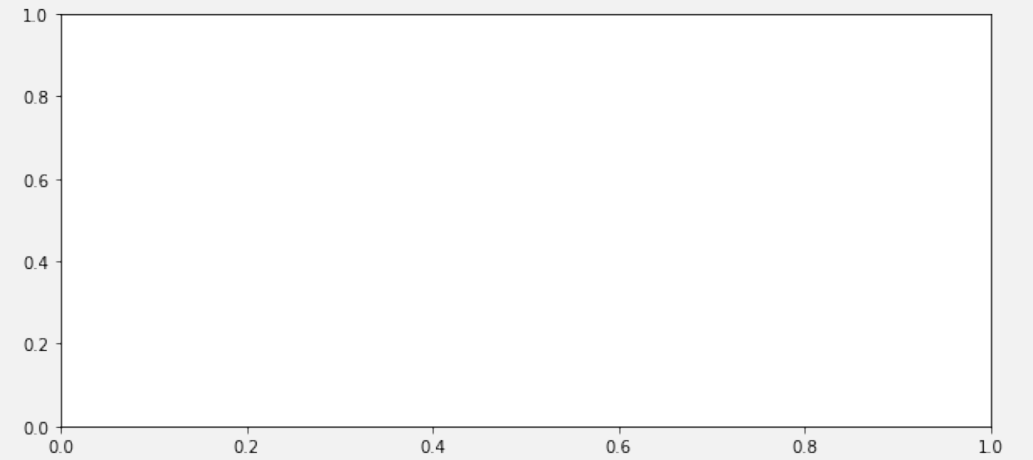
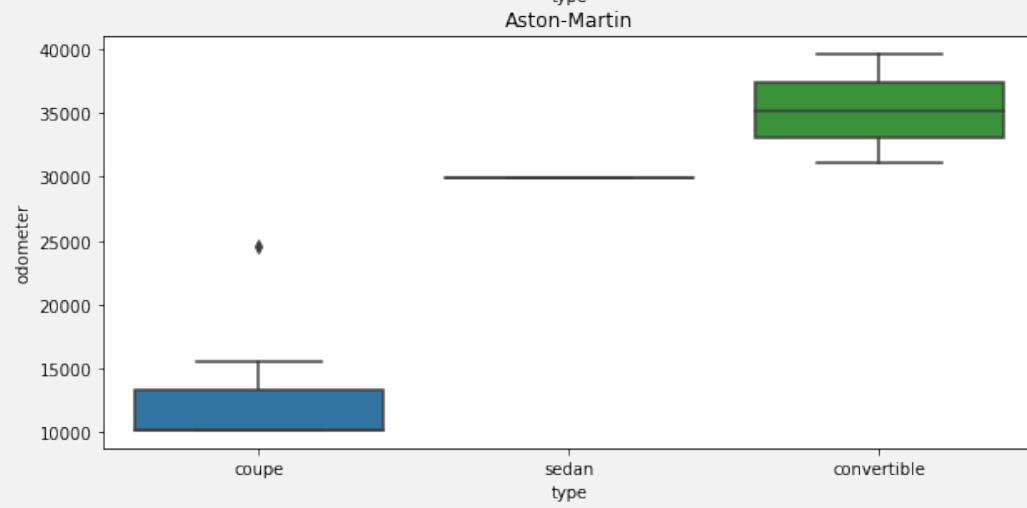
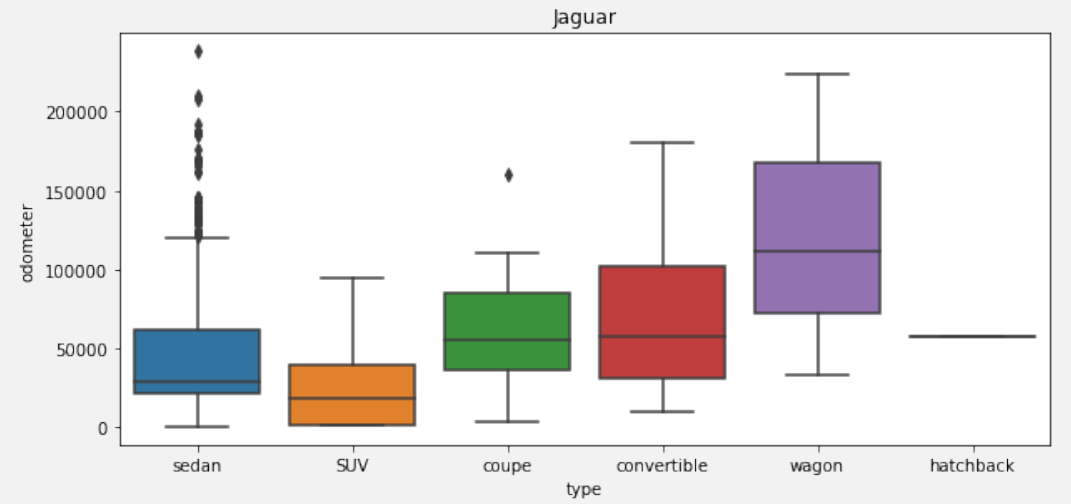
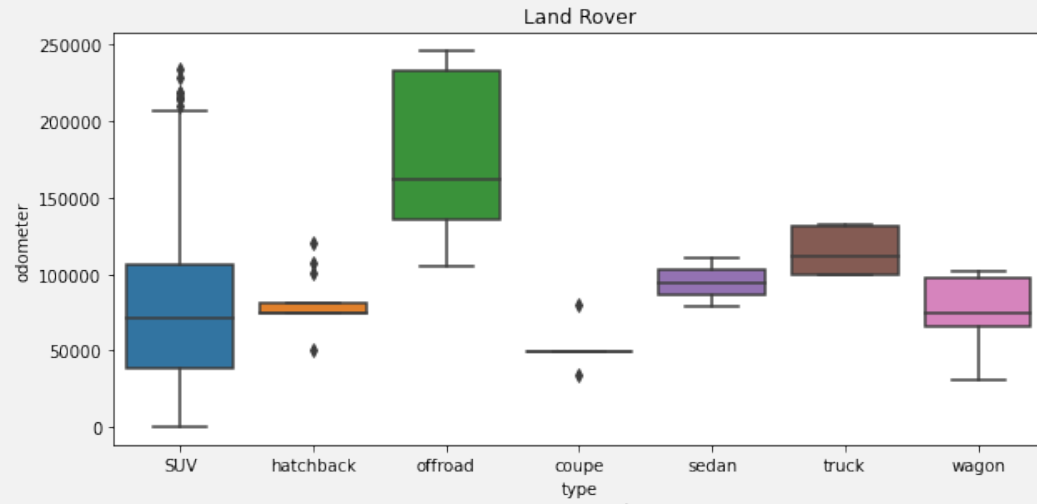


# ANALYSIS

- For German cars, it appears that SUVs have the most median miles for all manufacturers (Volvo only has 3 trucks and 34 hatchbacks available). Additionally, all four companies SUVs have a median mileage of less than 100,00 miles. This follows a common theme of German cars being a hassle to deal with because of their expensive maintenance, especially once approaching 100,000 miles. It also appears that there are a lot more outliers in mileage for German cars as opposed to others.

# BRITISH CARS

British Car Brands Mileage by Type



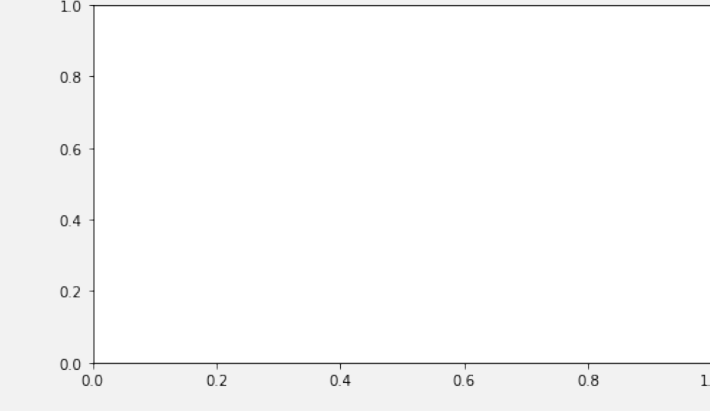
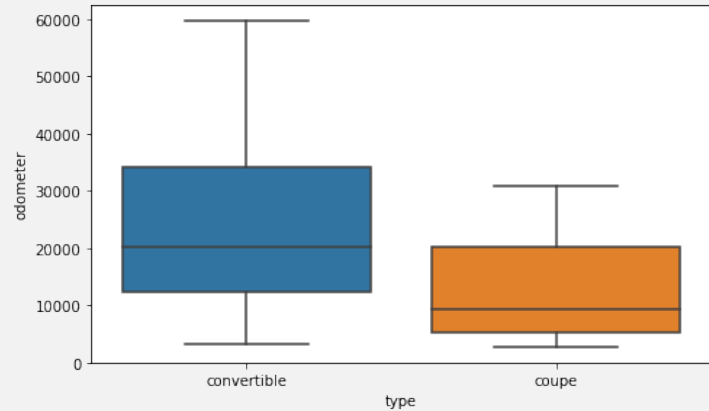
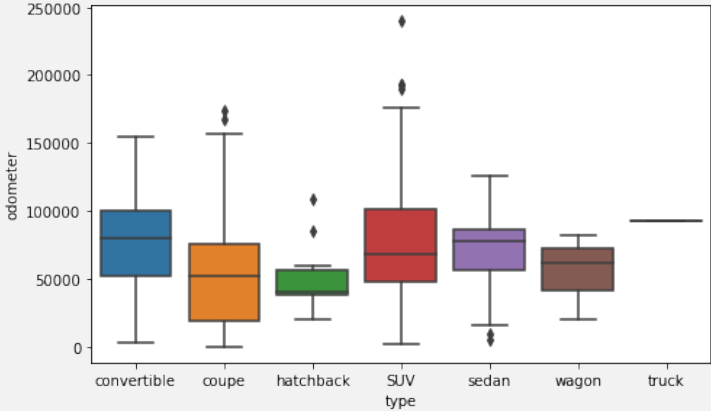
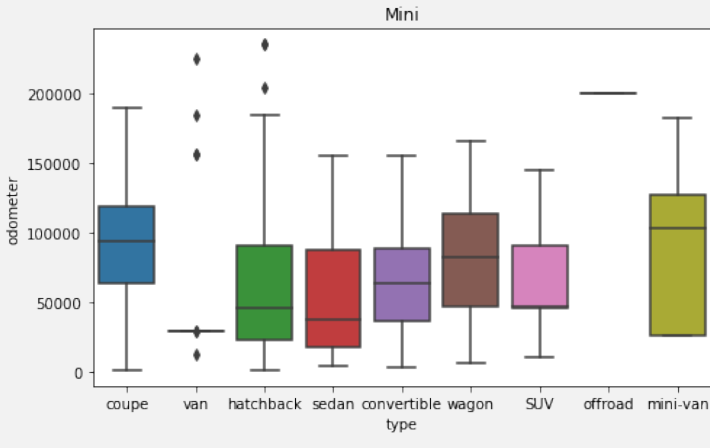
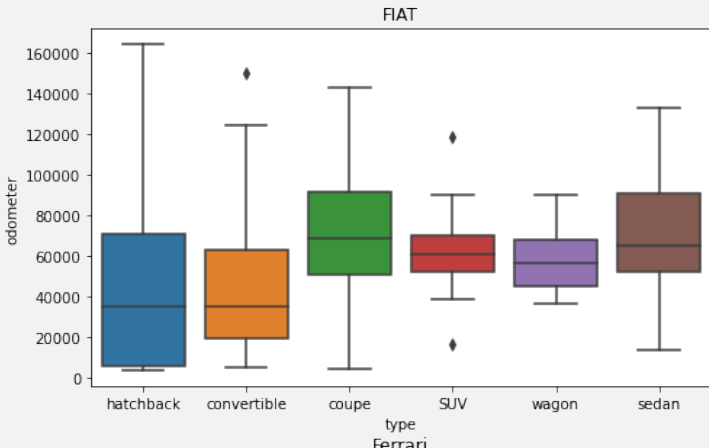
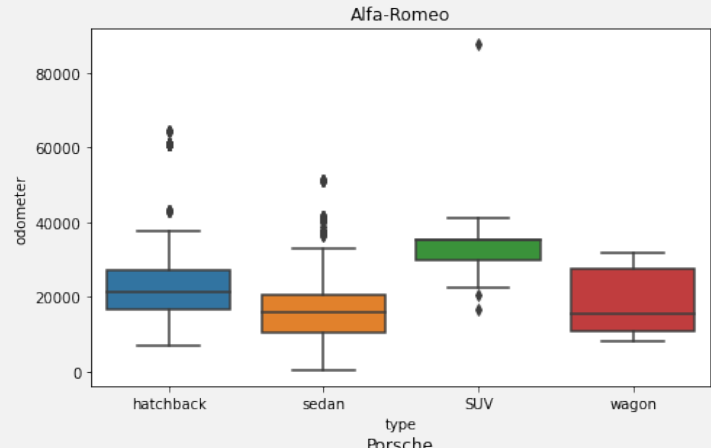


# ANALYSIS

- There are far fewer British manufacturers and cars available than other types. Furthermore, Aston Martin only has 13 cars available in the dataset. Therefore, this data was omitted from analysis as the sample size is too small. SUVs had the most cars listed for Land Rover and Sedans for Jaguar. The lower median mileage makes sense as both manufacturers are deemed as high luxury brands. It is interesting to note the large amount of Jaguar sedans that are outliers way above the median.

# ITALIAN CARS

Italian Car Brands Mileage by Type

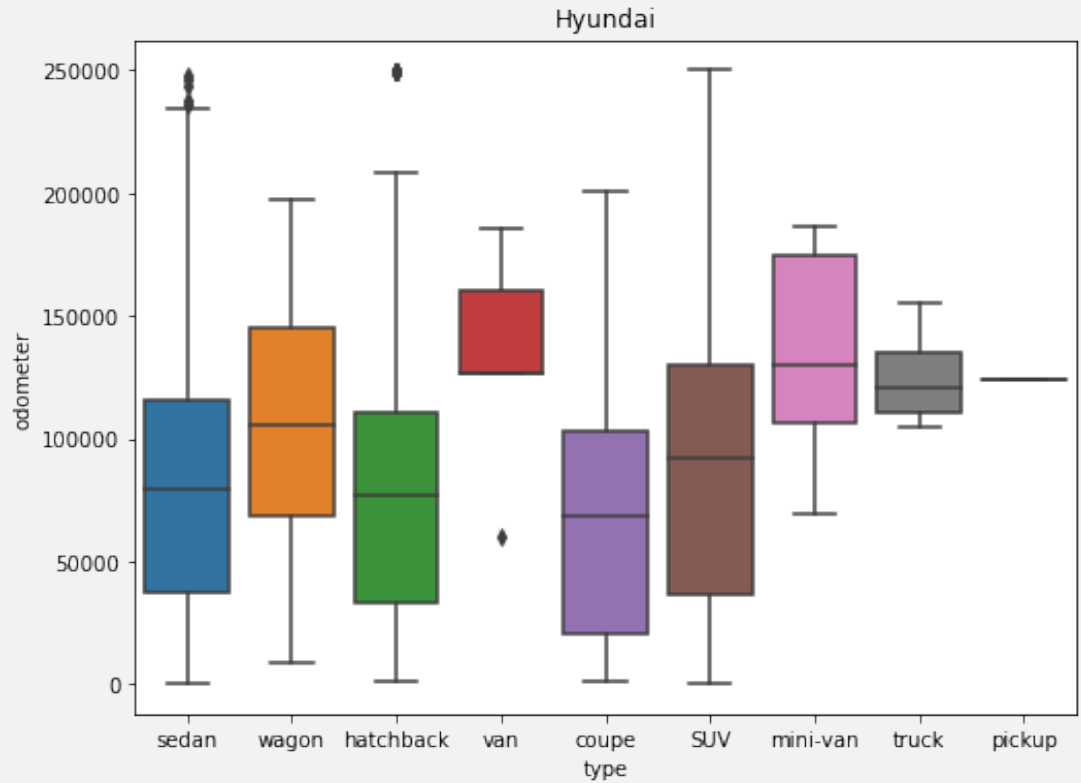
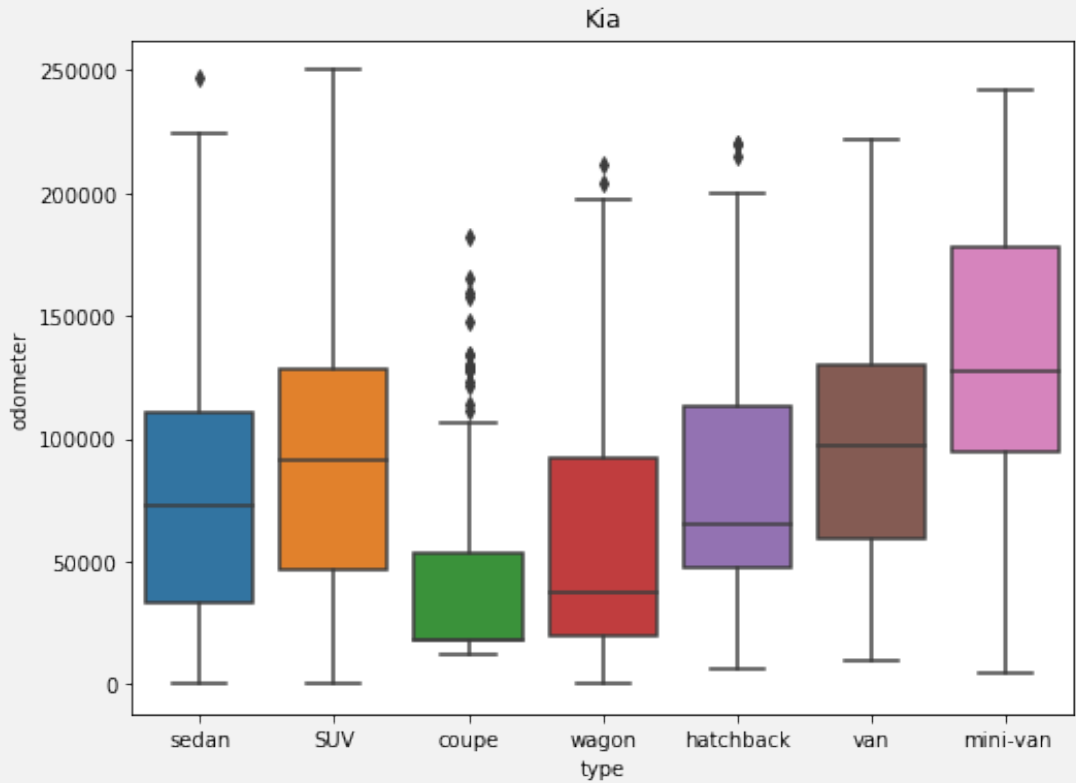


# ANALYSIS

- There are no cars in this dataset with a median of 100,000 miles. This may be because these cars are very expensive and luxurious, meaning they would not be driven as much as something like a Japanese car. It is interesting to note that FIAT, although not a luxury brand, has low median miles on hatchbacks and convertibles, which are their most common cars. This follows a historical trend that FIAT cars were not a very worthwhile car as they tend to break down easily.

# KOREAN CARS

Korean Car Brands Mileage by Type



# ANALYSIS

- These cars have the highest range of miles in the dataset, with most of them containing at least one car with 200,000 miles in the range. SUVs appear to have the most median miles with a wide upper bound that shows these cars can be driven for a while.

# OVERALL ANALYSIS

- SUVs are one of the most popular cars listed and have the highest median miles. We can now conclude they are the "best" type of car.
- Compared to other countries, Japanese cars do seem to have more median miles, which helps the theory that they hold more value and may be a more popular car. We can now conclude they are the "best" engineered car.
- Best Paint Color: Grey
- Best Condition: Excellent
- Best Cylinder: 6-Cylinder
- Best State: California
- Best Year: 2013
- Best Mileage: 115,000
- Best Manufacturer: Honda



# RESEARCH QUESTION ANSWERED

- Research Question: **What is the best used car of the 21st century?**
- Research Answer: **The best used car of the 21st Century is a 2013, grey, excellent conditioned, 6-cylinder Honda CR-V LX Sport with 115,000 miles and it should be purchased in California to get the best "bang for your buck."**

