

FINAL EXAM

Sidhartha Amperayani

August 8, 2022

Contents

Introduction	1
Parameters	2
Data Pre processing	2
Correlation between miles driven and price across vehicles whose age is less than 20	3
Average prices across the states	4
Manufacturers by the continent they were founded in	5
Time Series for Continent wise manufacturer's	6
Top 5 Manufacturers and their most selling 3 models	7
Average price analysis of top 5 Manufacturer's top selling model	8
Average Odometer readings of top manufacturers by their condition	9
Top 5 Manufacturer's listings across popular Types	10
Distribution of electric, and hybrid vehicles across the state of California	11
Top 3 manufacturers of Electric, and Hybrid vehicles in California	12
Average price comparison of top 2 Electric and Hybrid vehicle manufacturers in California by region	13
Conclusion	13
Appendix	15
Data Dictionary	15

Introduction

The dataset is about the vehicle listings posted up on the world's largest collection of used vehicles for sale, Craigslist during the months of April 2021 and May 2021.

The dataset features information related to the vehicles (odometer reading, the year the vehicle was bought in, price, manufacturer and model of the vehicle, condition of the vehicle, and, fuel used) and the location (latitude, longitude, region, and state) they were posted in.

Through this project, we can better understand the metrics of the automobile industry as a whole. It opens the window of performing analysis based on state, manufacturer, fuel used, etc.

- Which manufacturers and models are popular.
- Depreciation of a vehicles value over the years.
- States where the average price is least, hinting best state to buy a vehicle in.
- Number of vehicles that are being sold by the vehicle's age.
- Manufacturers and their best performing models and their average prices, and average odometer readings
- Most popular fuel type
- state and region wise popular manufacturers and their prices, and many more

With these metrics, a user of Craigslist or a person looking to buy a used vehicle/ new vehicle can make an informed decision as well as help the vehicle manufacturer to better understand their market and their competitor's market.

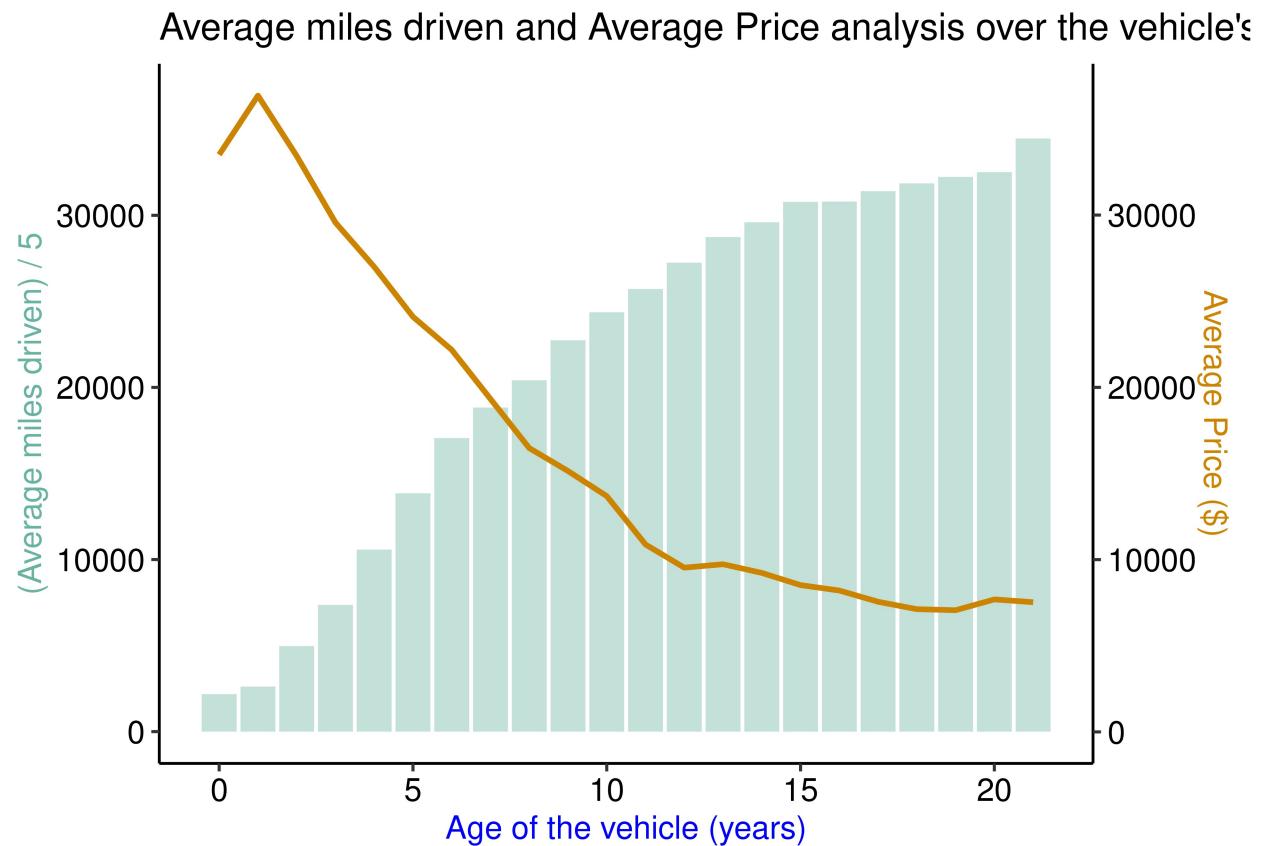
Parameters

All the parameters required for this file are initialized here.

Data Pre processing

- Removed id, url, region_url, VIN, paint_color, county, image_url, title_status, size, description, posting_date, transmission.
- converted integer64 to double.
- Considered only those listings whose prices are greater than \$100.
- Replaced empty values with NA.
- Removed listings whose model or manufacturer details are missing.
- Created another column continent based on the manufacturer's origin.
- Removed all the vehicles manufactured in 2022 and posted in 2021.
- Replaced model names which are dirty with NA.
- Considered *rover* vehicles as *Land Rover*.
- Selected all the distinct rows.

Correlation between miles driven and price across vehicles whose age is less than 20

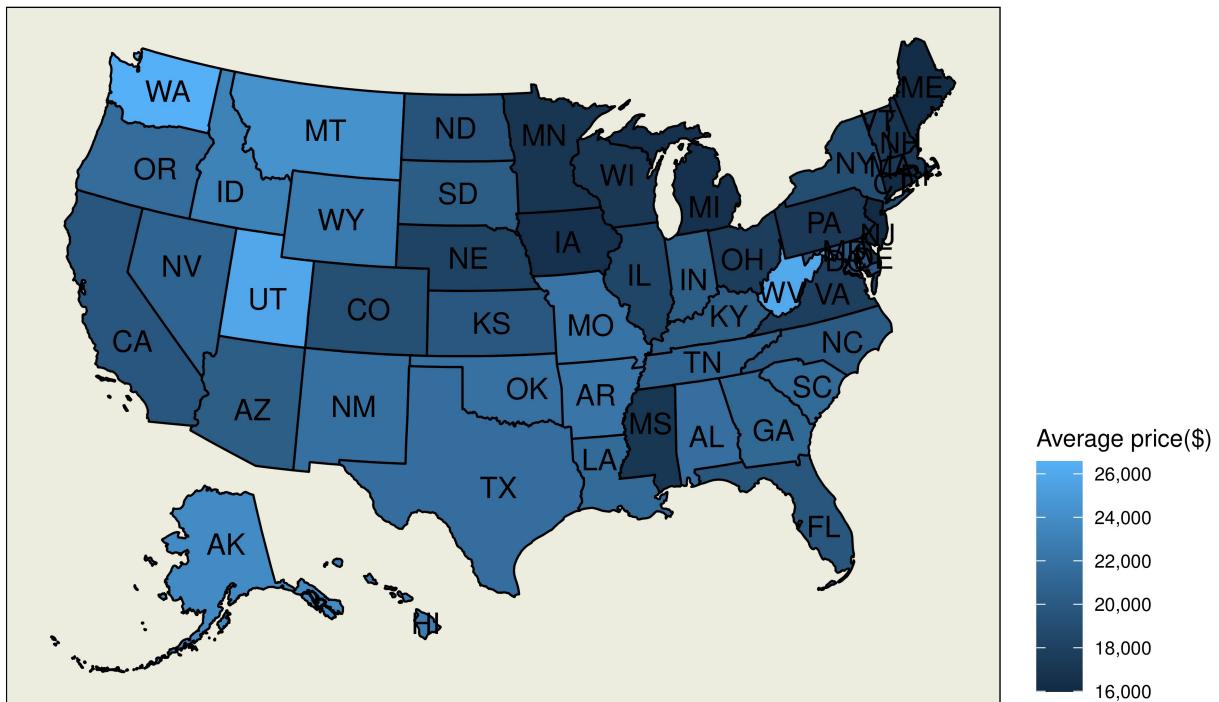


For the purpose of aesthetics and better understanding, I have decreased the scale of y-left by a factor of 5. It can be inferred that as the age of the vehicle increases the price keeps decreasing, and similarly with the age, average miles driven increases.

There is a hike in average price for vehicles that are 1 year old which happens to be the highest average price across all the ages. Price difference for vehicles between the ages of 15 and 20 is almost stagnant.

Average prices across the states

Average prices across states

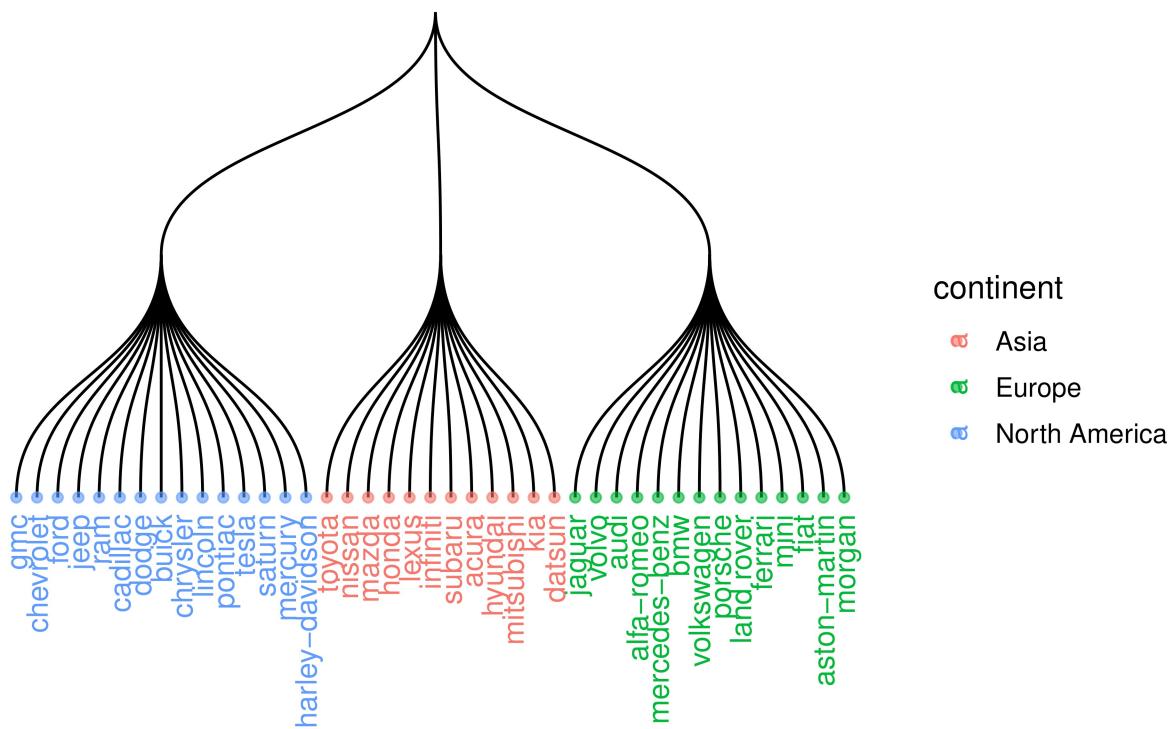


Map showing average prices across the states in the US. West Virginia, Utah, and Washington have the highest average prices. Although the number of postings on Craigslist for California is the highest, its average price is less.

The North-East region has the lowest average price across the country, while the North-West region has comparatively higher prices.

Manufacturers by the continent they were founded in

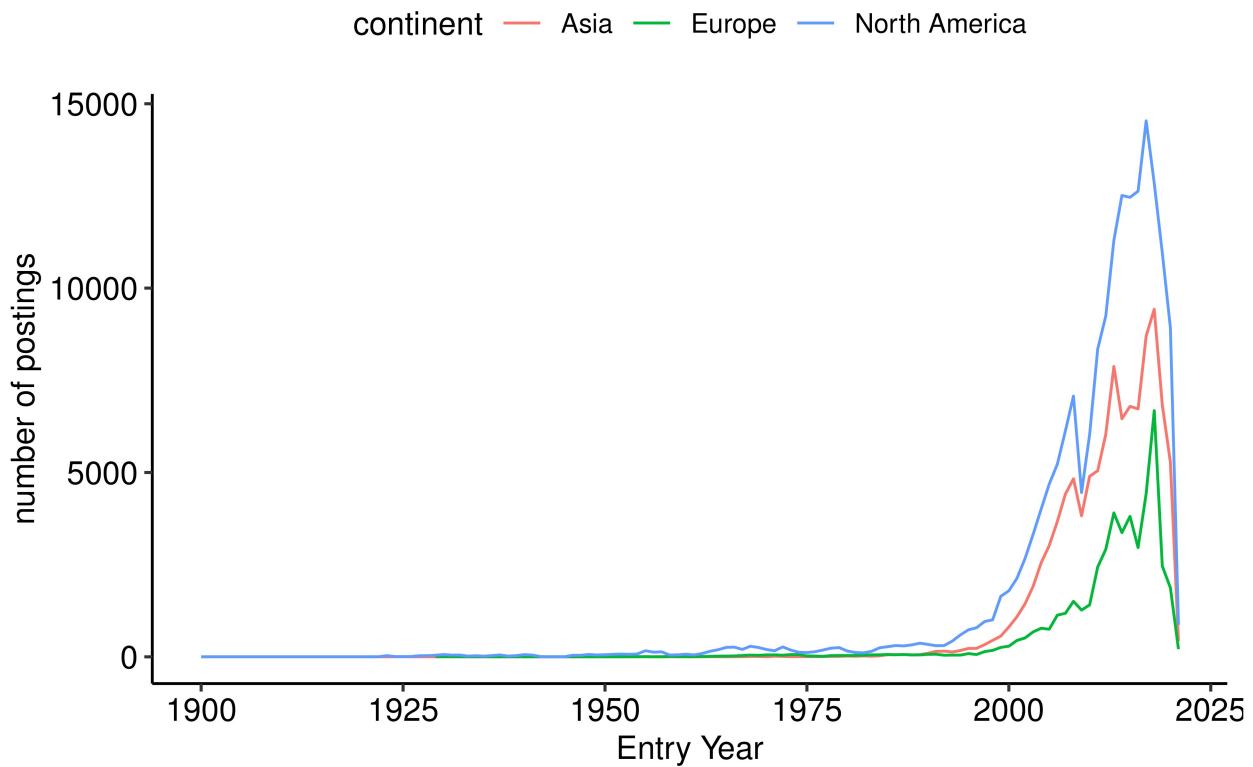
CONTINENT WISE MANUFACTURERS



All the manufacturers on Craigslist are based on three continents, Asia, North America, and Europe. Most manufacturers in Asia belong to Japan or South Korea, and in North America, it's just the USA. The number of manufacturers across the continents differs by at most 3.

Time Series for Continent wise manufacturer's

Popularity of continent-wise-makes over the years

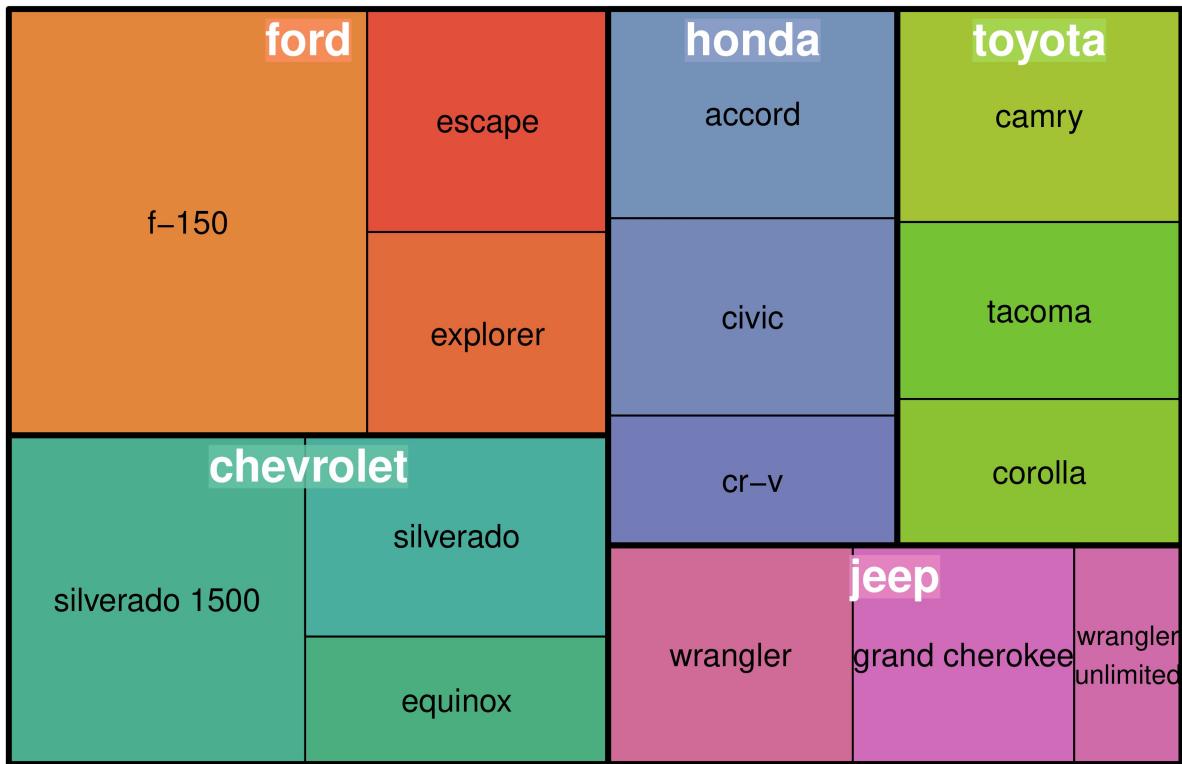


Between 2000 and 2021, the listings of all the three categories has increased gradually, except for the year 2007 when there was a massive dip for North American makes. Amongst the 3, North American brands are more popular, and Asian makes are the least.

It can be observed that around the entry year 1930, Craigslist has listed vehicles founded in Europe and Craigslist has Asian-based vehicles around the entry year 1970.

Top 5 Manufacturers and their most selling 3 models

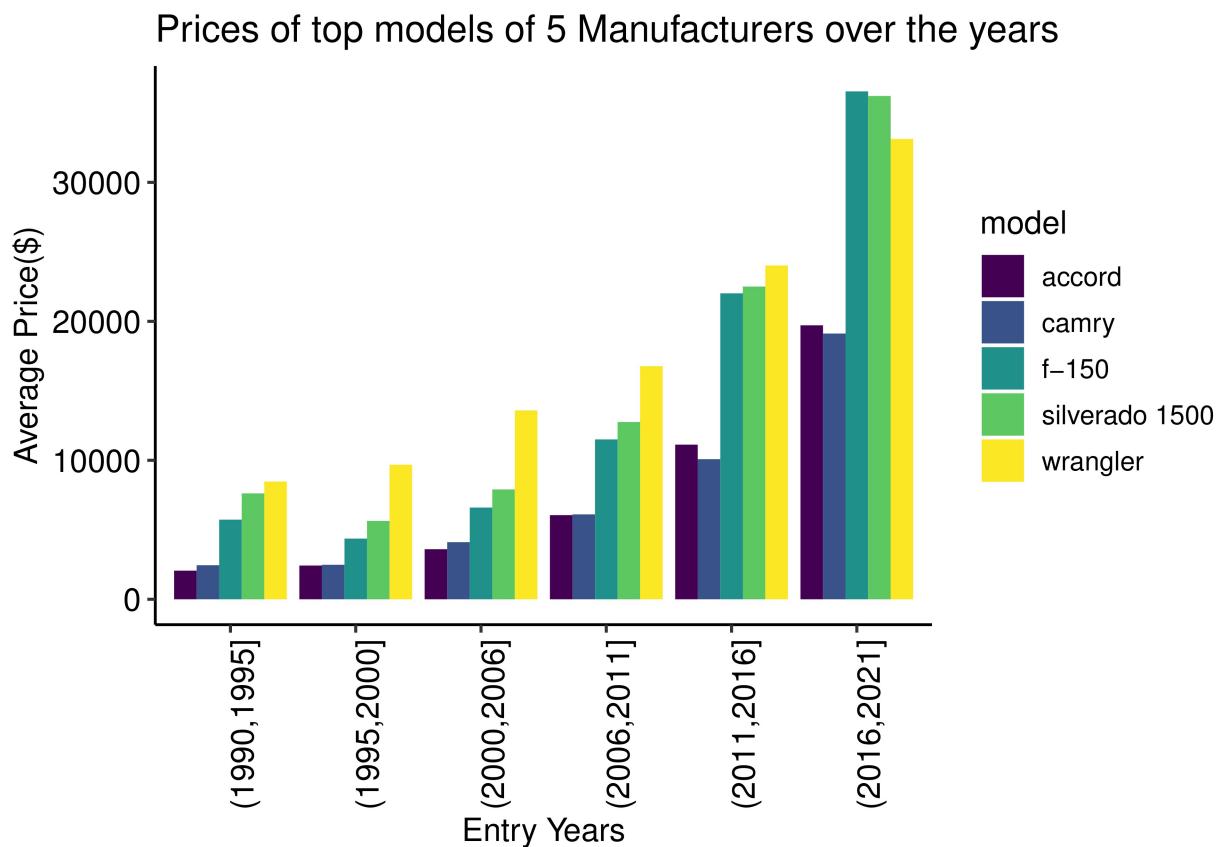
Top 3 Models of 5 most popular Manufacturers



3 vehicle manufacturers that sold most vehicles are Ford, Chevrolet, Honda, Toyota and Jeep. Their best models are F 150, Silverado 1500, Accord, Camry and Wrangler respectively.

Major share of the market belongs to Ford's F-150 followed by Silverado 1500. Listings of Accord and Camry are close to each other.

Average price analysis of top 5 Manufacturer's top selling model

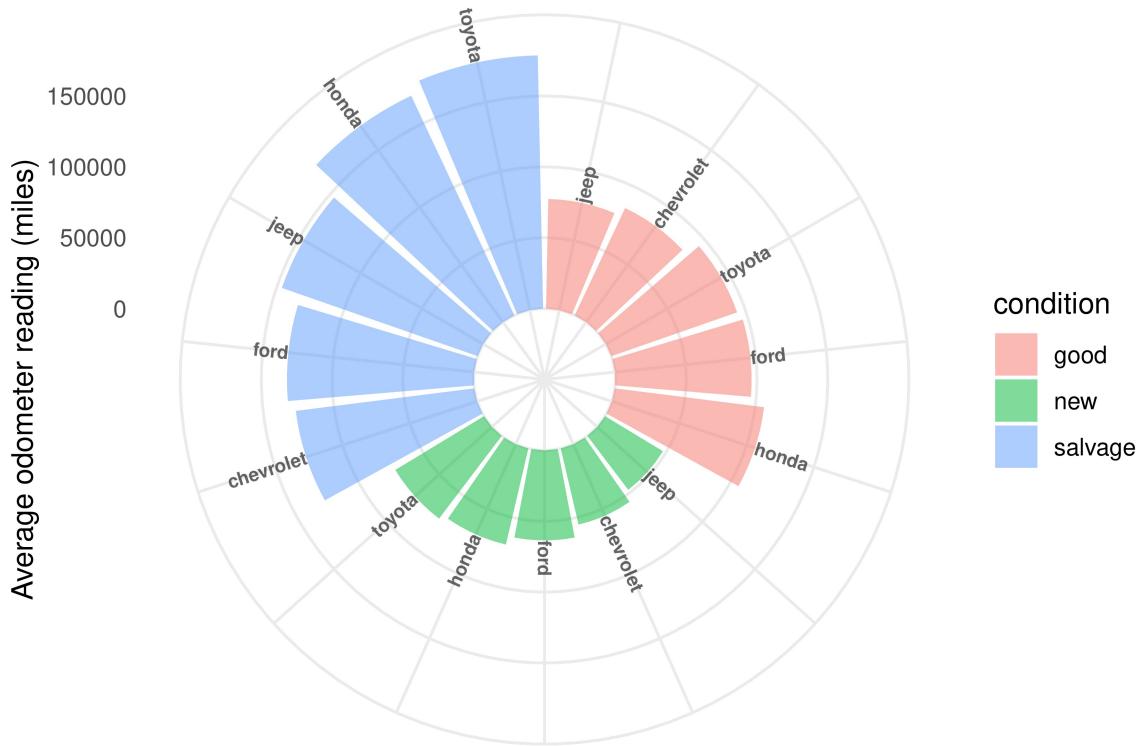


There was a steady increase in the prices of all 5 models over the years. In vehicles manufactured in 2016 and 2021, the price hike is the most. Although Ford's F-15 and Chevrolet's Silverado 1500 are top-selling models, their average prices are low compared to Jeep's Wrangler until 2016.

Honda Accord, Toyota Camry and Ford F-150, and Chevrolet's Silverado 1500 are 2 pairs of competitors across the years the vehicle was manufactured in. Prices of Camry and Accord which entered the market between 1995 and 2000 have dipped, and they picked up thereon.

Average Odometer readings of top manufacturers by their condition

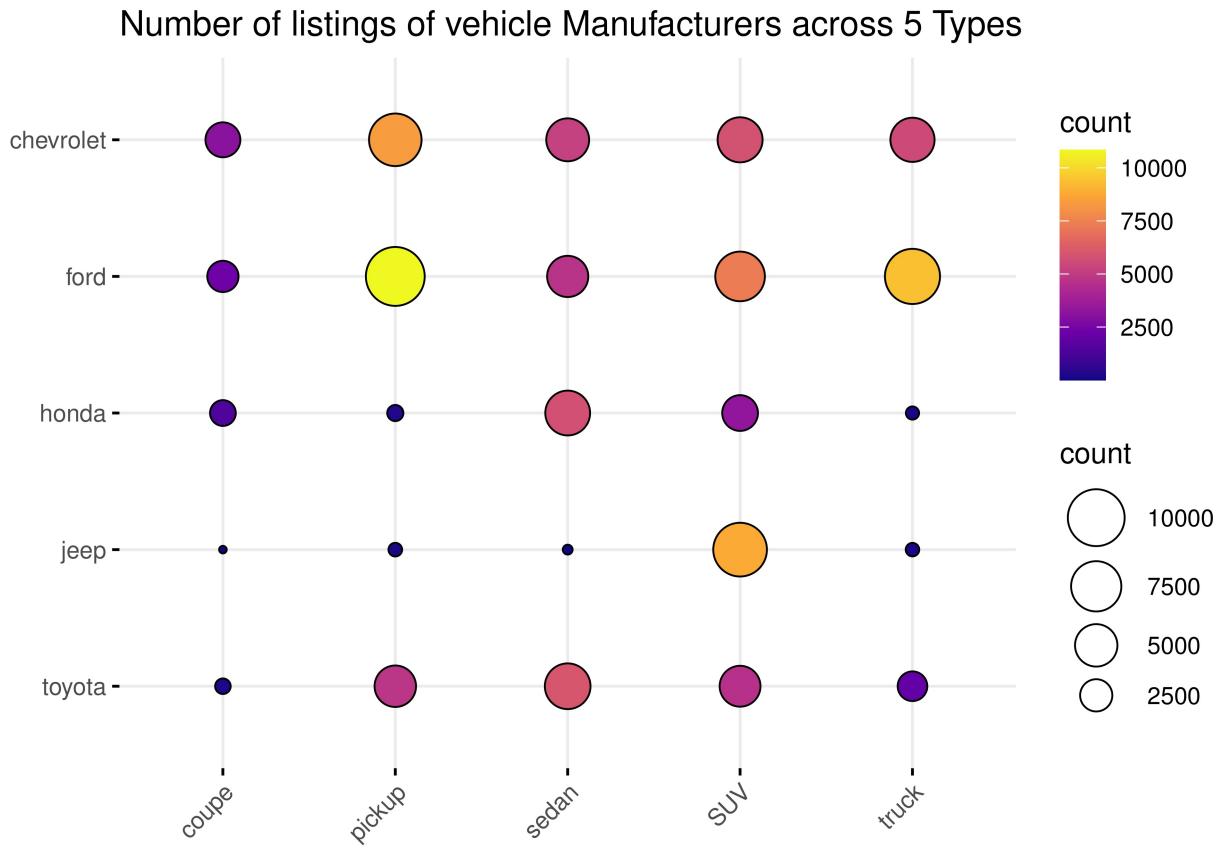
Average odometer readings of top 5 manufacturers



The average odometer readings of all the top 5 manufacturers are comparable in a given condition. Salvage vehicles had the highest odometer reading while the new ones had the least. Amongst the salvage, Toyota and Honda vehicles had the highest average odometer reading of 150,000 miles while Jeep and Chevrolet had the lowest odometer readings amongst the new vehicles of 50,000 miles.

Chevrolet's and Jeep's good condition vehicle's average odometer readings are comparable with that of Toyota and Honda's new condition average odometer readings.

Top 5 Manufacturer's listings across popular Types

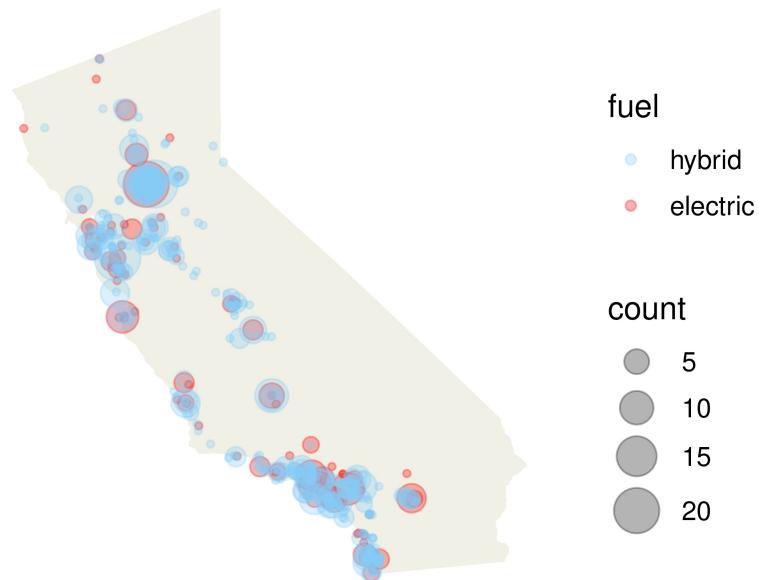


According to this balloon plot visualization, Chevrolet, Ford, and Toyota have a solid footprint across all 5 variants. Jeep is most popular among the top manufacturers of SUVs. Honda and Toyota have the most number of sedan listings on Craigslist.

Unlike Ford's other competitors, Ford has a clear dominance in the truck variant with a count of almost 10,000 listings. Both Ford and Chevrolet have the most pickup vehicles.

Distribution of electric, and hybrid vehicles across the state of California

Electric and Hybrid listings in California

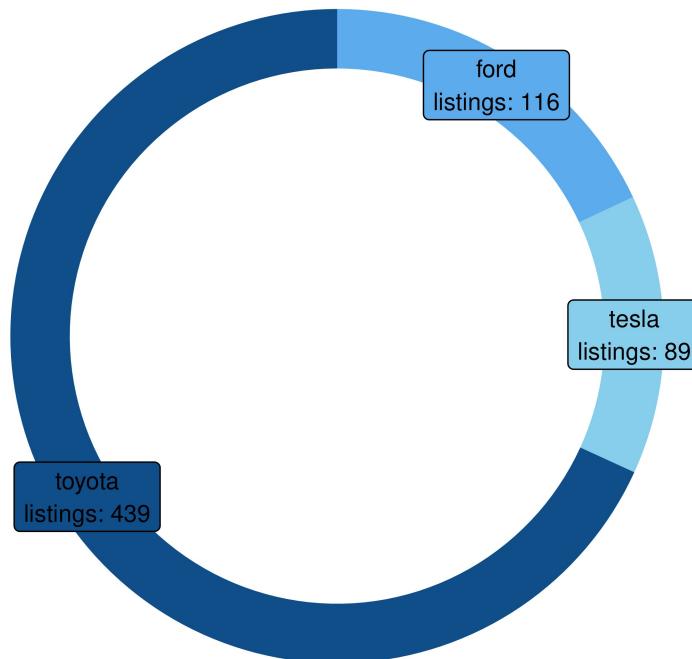


Here is a map of California featuring all the electric and hybrid vehicle listings on Craigslist across the state. Most electric and hybrid listings are in California. However, there are only about 1,166 listings for the same have been put up on Craigslist in the months of April and May, 2021.

Through this visualization, distribution and count of Hybrid vehicles are more than that of Electric's. San Francisco, and Riverside have more listings of Electric vehicles than Hybrid. Otherwise, it is mostly hybrid vehicles across the state. Most listings are in the Bay area, Santa Barbara, Ventura, Los Angeles, and Sacramento.

Top 3 manufacturers of Electric, and Hybrid vehicles in California

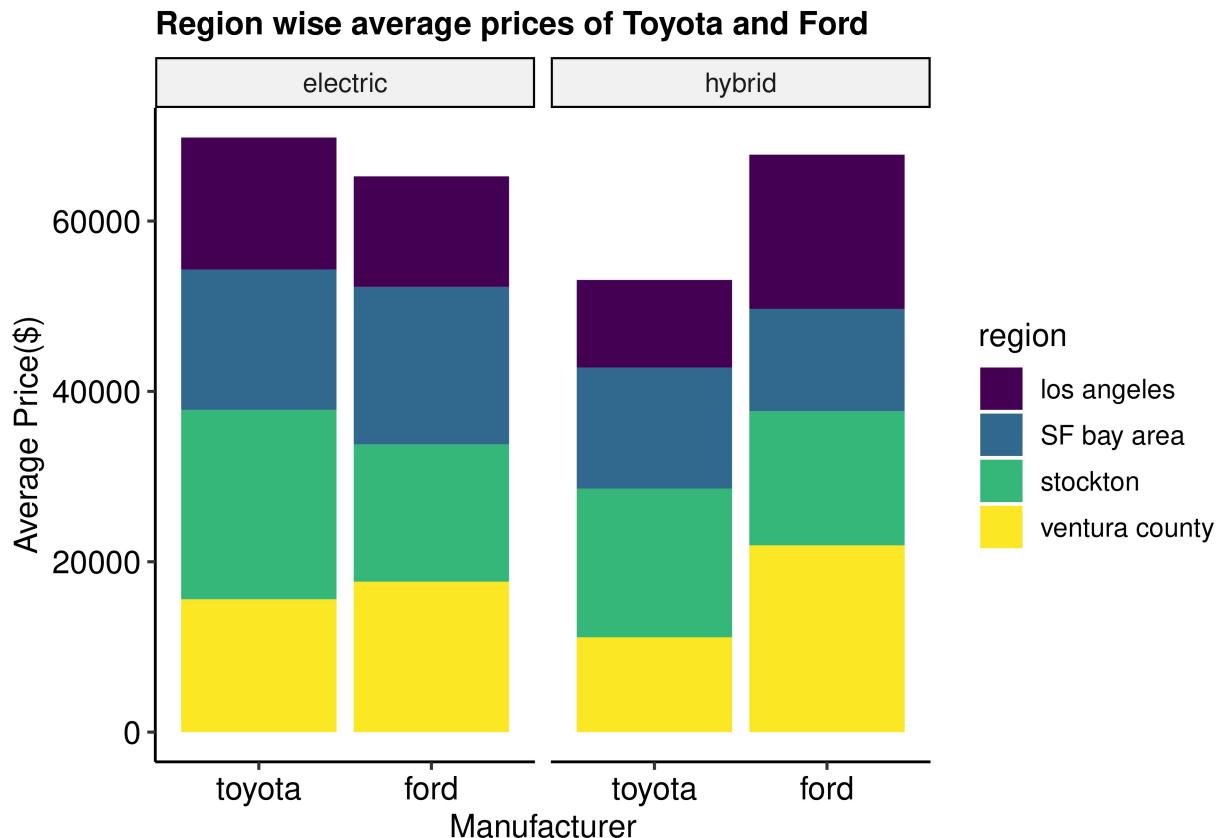
Top 3 Electric and Hybrid vehicle manufacturers by listings



In the above doughnut chart number of listings of the top 3 manufacturers of electric and hybrid vehicles in California state are represented. It can be observed that Toyota has a clear dominance, over 50%, with its competitors, Ford and Tesla.

This is probably because Tesla is relatively new in 2021 and not many have put up their vehicles for sale on Craigslist then. Toyota's hybrid vehicles are more in number than electric, hence the dominance.

Average price comparison of top 2 Electric and Hybrid vehicle manufacturers in California by region



It can be inferred that in most regions of California, electric variants are costlier than hybrids except for Ford in Ventura county and Los Angeles. Suggesting that the hybrid vehicles of Ford are far more superior/valuable than Toyota.

Toyota's Hybrid variants are priced higher than Ford's in the SF Bay area and Stockton, while coming to electric, Toyota has higher value in Los Angels and Stockton. The average prices of both vehicle manufacturers in all regions are under \$25,000

Conclusion

Through this dataset, an analysis was made on several aspects. I included another column that identifies the continent in which the vehicle was founded.

Through this, we can better understand the trust and popularity of brands concerning their origin.

After performing data cleaning and analysis, we can understand that amongst the vehicles manufactured in the last 20 years the average odometer reading increases gradually up to 150,000 miles.

With the age, contrary to the odometer reading, the price of the car decreased almost linearly and flattened for the vehicles whose age is between 15 and 20.

After performing an average price analysis by the state on the US map, we learn that West Virginia, Utah, and Washington have the highest average prices.

Although the number of postings on Craigslist for California is the highest its average price is comparatively low.

It is also observed that there is a stark difference between the North-West and North-East regions of the US concerning the price.

It was found that the distribution of manufacturers by their origin is almost uniform.

There are 15 from *North America, 12 from Asia, and 14 from Europe.

However, all the manufacturers based in Asia are either from Japan or South Korea and all the vehicles from North America are based in the US.

Their popularity across the different ages of the vehicle has been increasing except in 2007 when there was a massive dip for American-made manufacturers.

There are listings of European vehicles which are as old as almost 90 years and Asian vehicles which are almost 50 years old.

Amongst all the manufacturers, Ford, Chevrolet, Honda, Jeep and Toyota are most popular ones and their best models are F-150, Silverado 1500, Accord, Camry and Wrangler respectively.

F-150 holds a major share in the market as compared to any other model.

Amongst all the aforementioned models, the average price increased with decreasing age.

Wrangler consistently had higher prices except for the cars manufactured in the last 5 years, both Silverado 1500 and F-150 had overtaken Wrangler. Camry and Accord were valued almost the same across the years.

As per the mentioned condition (good, new, and salvage), the average odometer readings of all the top 5 manufacturers are comparable. Salvage vehicles had the highest odometer reading while the new ones had the least.

Amongst the salvage, Toyota and Honda vehicles had the highest average odometer reading of 150,000 miles while Jeep and Chevrolet had the lowest odometer readings amongst the new vehicles of 50,000 miles.

Chevrolet, Toyota, and Ford have a good market across the variants, coupe, pickup, sedan, SUV, and truck. Across the 5 manufacturers, Jeep has a stronghold for SUVs. Honda and Toyota have the most number of sedan listings on Craigslist.

Unlike Ford's other competitors, Ford has a clear dominance in the truck variant with a count of almost 10,000 listings. Both Ford and Chevrolet have the most pickup vehicles. The entire world is going towards electric vehicles due to increasing global warming.

To better understand the trends of electric vehicles in the US market I have made some visualizations, however, since the number of listings for electric is low, I included hybrid as well.

In the state of California, which has the most number of electric and hybrid listings, San Francisco, and Riverside have more electric vehicles for sale than any other region in the state when compared with the number of listings of hybrid vehicles.

Most listings are in the Bay area, Santa Barbara, Ventura, Los Angeles, and Sacramento. The top 3 manufacturers of electric and hybrid are Toyota, Tesla, and Ford. It can be observed that Toyota has a clear dominance, over 50%, over its competitors. Toyota's hybrid vehicles are more in number than electric, hence the dominance.

It can be inferred that in most regions of California, electric variants are costlier than hybrids except for Ford in Ventura county and Los Angeles. Suggesting that hybrid vehicles of Ford are far more superior/valuable than Toyota. Average prices of both vehicle manufacturers, Toyota and Ford in all regions are under \$25,000. Toyota's Hybrid variants are priced higher than Ford's in SF Bay area and Stockton, while coming to electric, Toyota has higher value in Los Angeles and Stockton.

Through this assignment I was able to analyze some of the things mentioned in the introduction and went beyond those by including the continent column.

Appendix

Data Dictionary

The data regarding all the relevant columns are listed here:

Region- City or Region in which the listing was made

Price- The price of the vehicle in USD

Year- Year in which the vehicle was bought

Manufacturer- manufacturer of the vehicle

Model- model of the vehicle

Condition- Condition of the vehicle it is in

Cylinders- Number of cylinders of the engine

Fuel- Type of the fuel used

Odometer- Odometer readings of the vehicle

Drive- type of drive

Type- Type of the vehicle

State- The state in which the listing was made

Lat- latitude of the region

Long- longitude of the region

Continent- Continent in which the vehicle was founded