

FINAL EXAM

Amritha Menon

August 8, 2022

Contents

INTRODUCTION	1
DATA PREPROCESSING	2
HEAD OF THE DATA	2
FINDINGS	3
STATES WITH HIGHEST AND LOWEST VEHICLE LISTINGS	3
RELATIONSHIP BETWEEN THE PRICE AND ODOMETER READINGS ON THE BASIS OF CONDITION	4
AVERAGE MILES DRIVEN FOR UPTO 60 YEAR OLD VEHICLES	5
MOST POPULAR CYLINDERS AND THEIR PRICES	6
POPULAR FUEL TYPE IN THE TOP 5 CALIFORNIA REGIONS	7
TOP 5 TYPES OF VEHICLES AND THEIR DRIVE TYPE	8
TREND OF VEHICLE LISTING FROM 2005 TO 2012	9
AVERAGE PRICE OF VINTAGE VEHICLES ACCORDING TO MANUFACTURER	10
CHANGE IN CONDITION OF VEHICLE BASED ON MILES DRIVEN	11
MICHIGAN'S TOP MANUFCATURERS IN ITS TOP REGIONS	12
CONCLUSION	12
APPENDIX	13
DATA DICTIONARY	13

INTRODUCTION

The exploratory data analysis of the **Vehicle Listings** from Craigslist contains data of used cars listings posted in the US. The data provides an insight into the different types of vehicles, the quality of its life, and various other trends one can predict. This analysis could help people who are looking to buy or sell vehicles. The dataset contains various information like the type of vehicle, the manufacturer, the model, color, fuel, drives, and cylinders. One can also say how reliable craigslist is when it comes to vehicle listings.

One advantage of using Craigslist is that the seller does not need to deal with any middle man and can interact with buyer. This is beneficial for both the parties as the buyer can negotiate over the price with the seller itself. And the buyer can confirm if the opposite party is a fraud or if he is genuine. The dataset is focused only on listings posted on May and April of 2021 and 2022.

To thoroughly analyze the data, I seek answers to the following questions:

- Which state has the highest and the lowest number of vehicles listed ?
- What is the relationship between price and the odometer readings ?
- Does the average miles driven increase as the age of the vehicle increases ?
- What is the most popular cylinder used and what is its price?
- What is the popular fuel type in the California region ?
- What are top vehicle types according to its drive ?
- What is the trend of vehicle listing from 2005 to 2012 ?
- What are the top manufacturers and average price of vintage vehicles ?
- How does the miles traveled affect the condition of the vehicle ?
- Where are cars most manufactured in Michigan and who are the top manufacturers ?

From the answers to these questions, an analysis can be made on the data.

DATA PREPROCESSING

The data is read and then pre processed. To preprocess this data, I first removed columns I found irrelevant like id, Vehicle Identification Number, image, county, and region url, size, title_status, transmission, paint color, model and the description. The dataset contains certain values of type integer64 which was changed to type double for further use. All the duplicate values have also been removed and the empty strings were changed to null values to be removed later as each visualization is being made.

A new column Years_used and posting_year was also created to calculate the number of years the vehicle was used and the year posted.

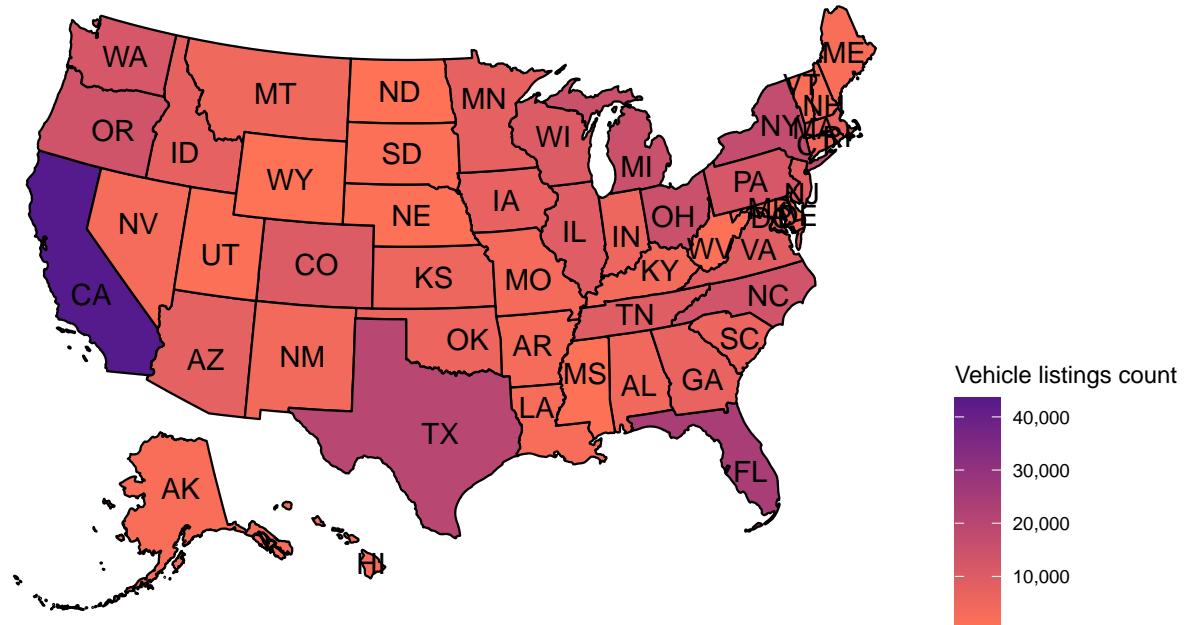
HEAD OF THE DATA

region	price	year	manufacturer	condition	cylinders	fuel	odometer	drive	type	state	lat	lon
auburn	15000	2013	ford	excellent	6 cylinders	gas	128000	rwd	truck	al	32.5920	-82.0000
auburn	27990	2012	gmc	good	8 cylinders	gas	68696	4wd	pickup	al	32.5900	-82.0000
auburn	34590	2016	chevrolet	good	6 cylinders	gas	29499	4wd	pickup	al	32.5900	-82.0000
auburn	35000	2019	toyota	excellent	6 cylinders	gas	43000	4wd	truck	al	32.6013	-82.0000
auburn	29990	2016	chevrolet	good	6 cylinders	gas	17302	4wd	pickup	al	32.5900	-82.0000
auburn	38590	2011	chevrolet	good	8 cylinders	gas	30237	rwd	other	al	32.5900	-82.0000

FINDINGS

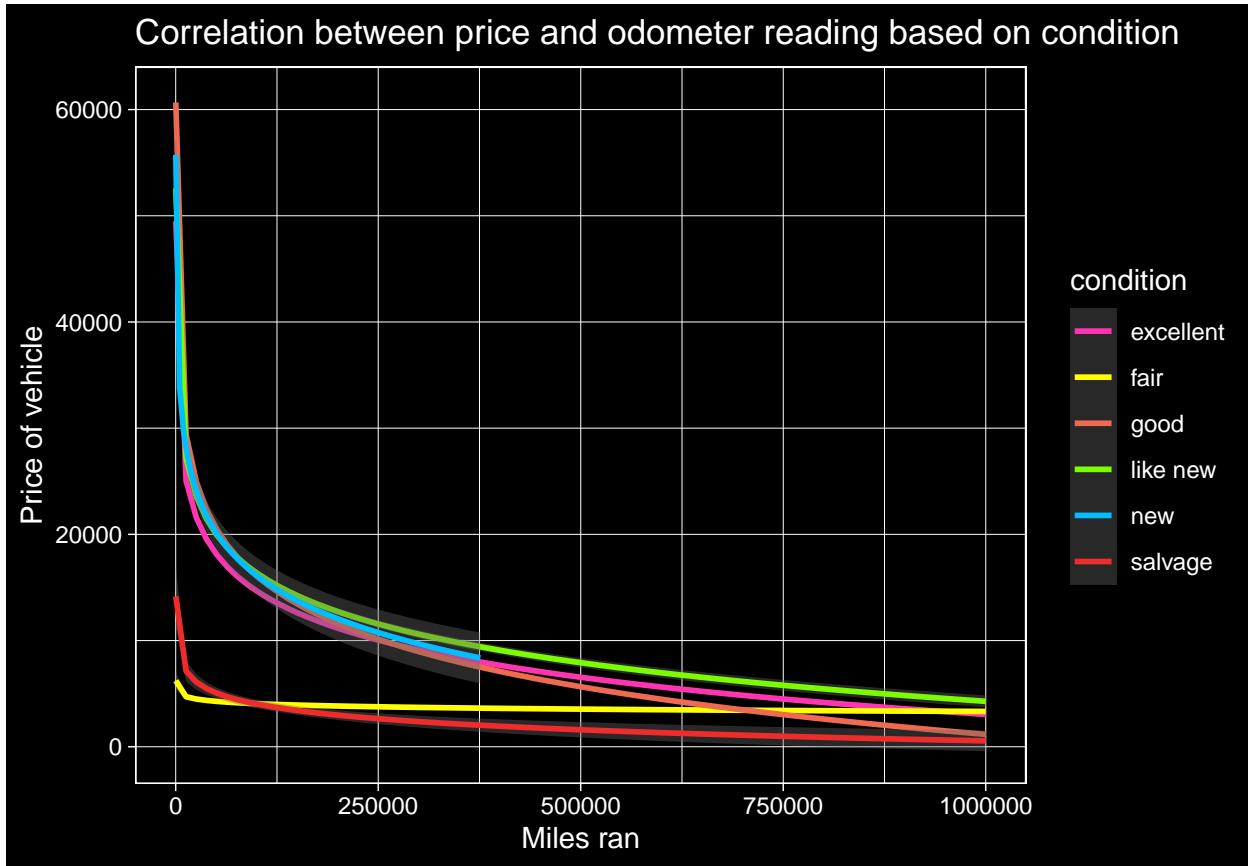
STATES WITH HIGHEST AND LOWEST VEHICLE LISTINGS

Vehicle listing population across the country



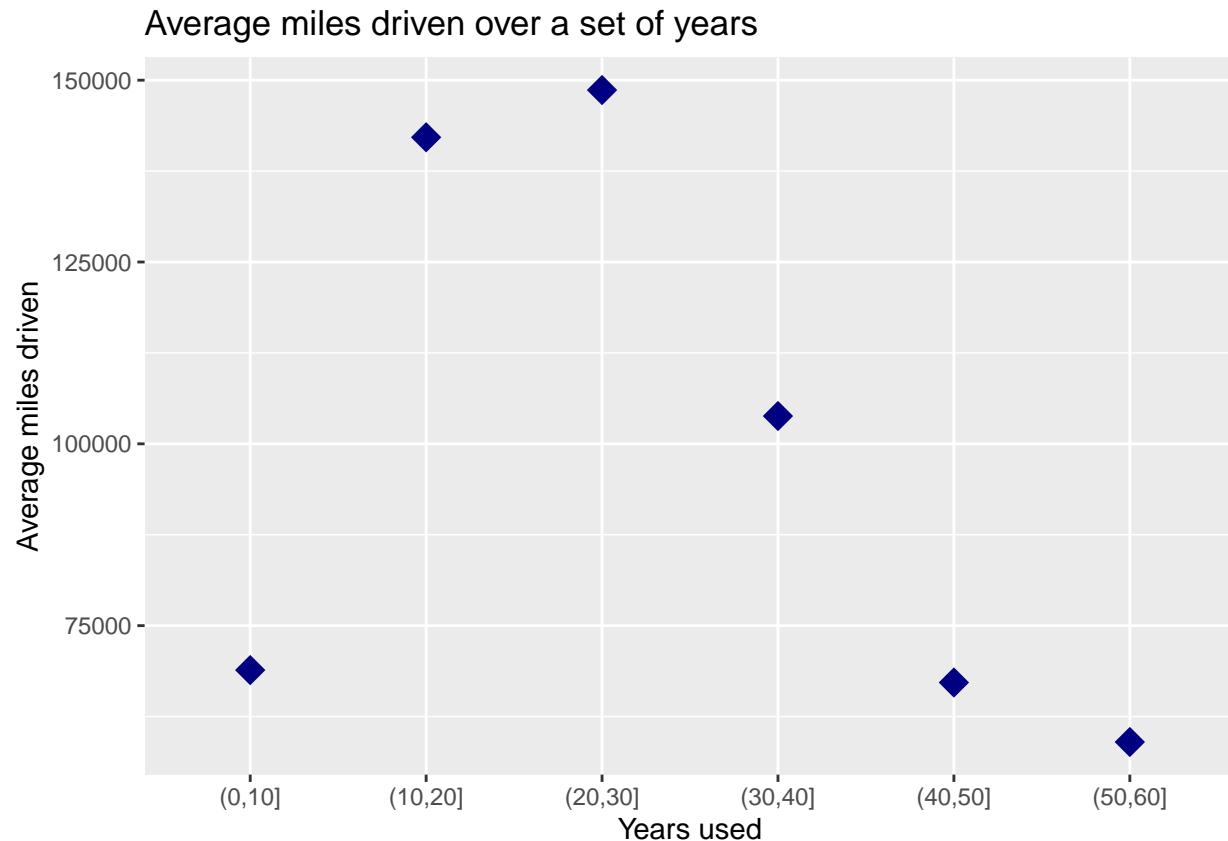
The visualization represents the number of listings from each state across the country. The highest number of listings is from California and the lowest number of listings is from North Dakota.

RELATIONSHIP BETWEEN THE PRICE AND ODOMETER READINGS ON THE BASIS OF CONDITION



The above visualization depicts the relationship between the price and odometer readings. The price of the vehicle decreases as the miles driven increases. This is based on the condition of the vehicle. Fair condition vehicles all have relatively lower price, even lower than salvage condition vehicles.

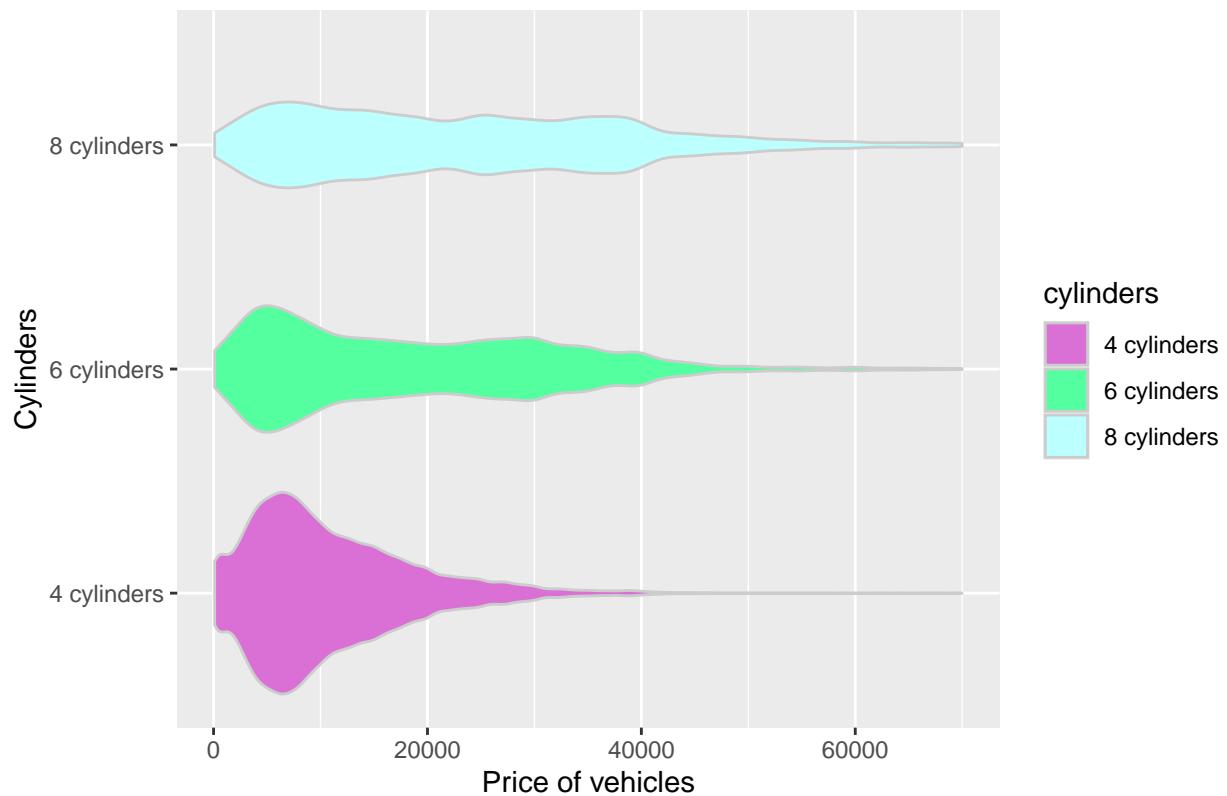
AVERAGE MILES DRIVEN FOR UPTO 60 YEAR OLD VEHICLES



The analysis on the years the vehicle was used and its average miles says that the average miles increases upto 30 years, but after that point it decreases. It is surprising to see that cars with the age 50 - 60 has driven under 70000 miles. The average miles driven do not increase as the age of the vehicle increases.

MOST POPULAR CYLINDERS AND THEIR PRICES

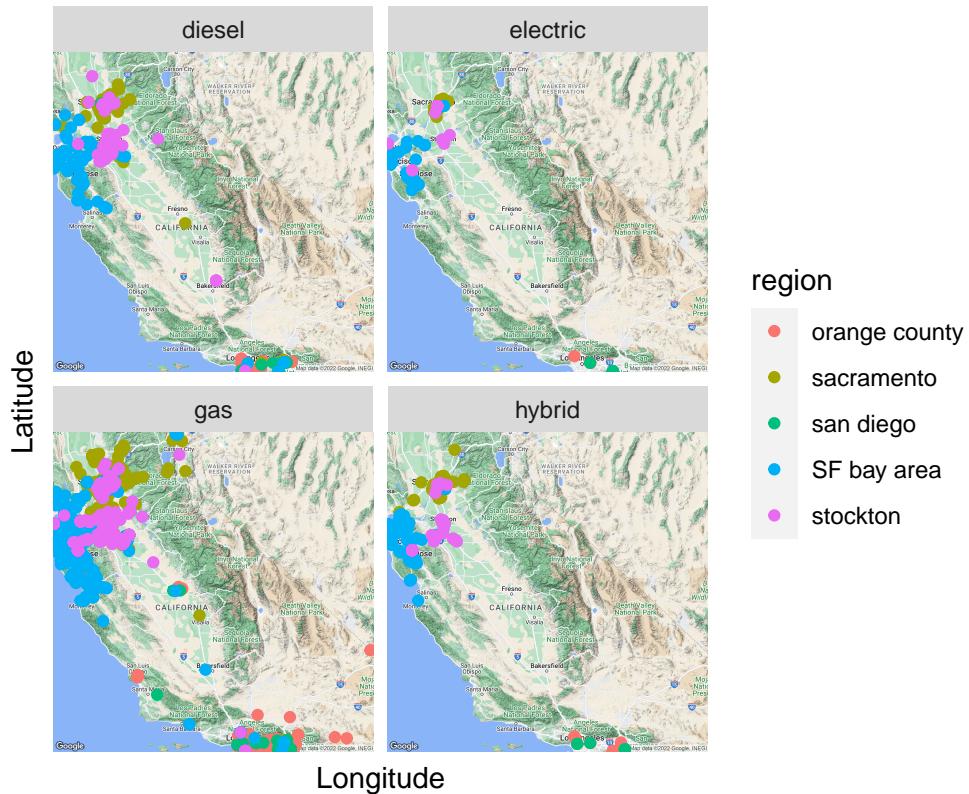
Distribution of top 3 cylinders based on price



The top 3 cylinders are 4, 6, and 8 cylinders. The price of 4 cylinders are cheaper than the other two which would make it most popular. 8 cylinder vehicles are the most expensive.

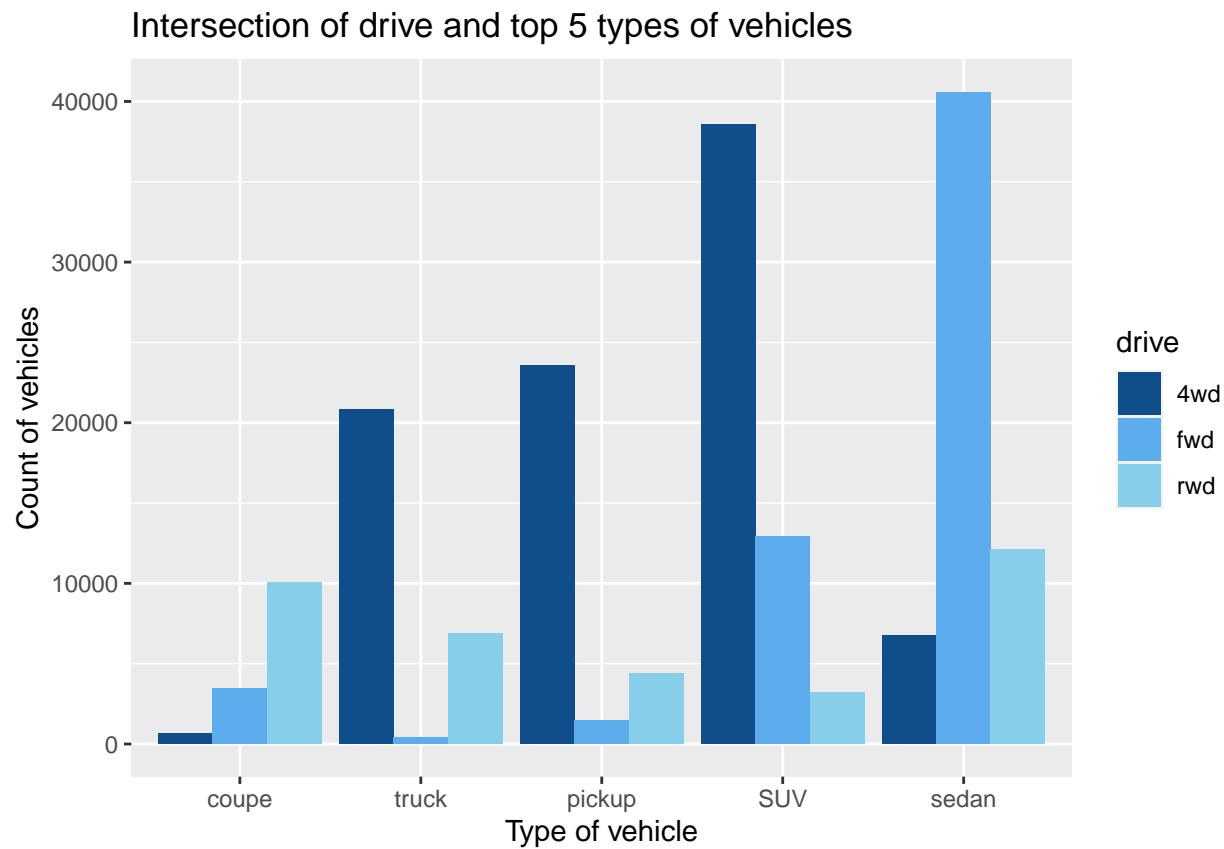
POPULAR FUEL TYPE IN THE TOP 5 CALIFORNIA REGIONS

Top 5 regions in California based on fuel



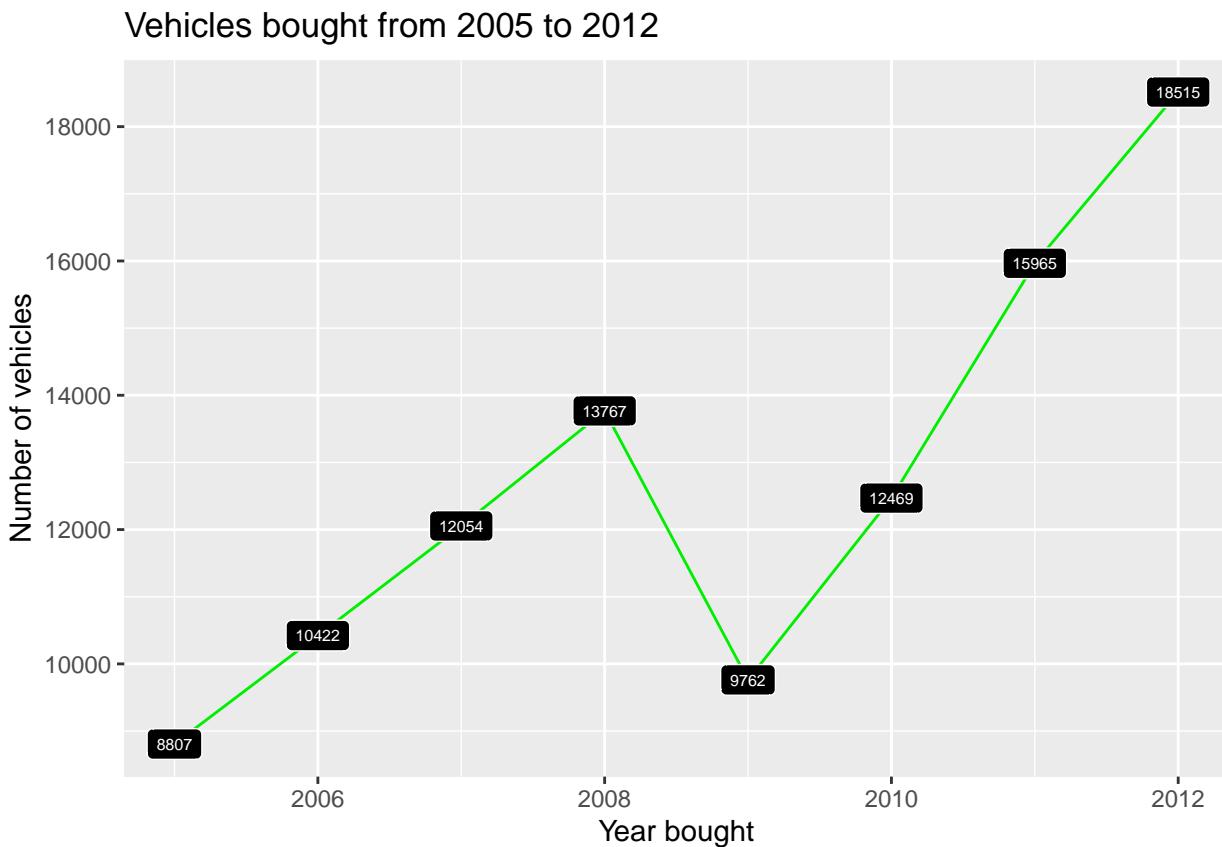
Gas type vehicles are the most popular across all the top 5 regions of California. The San Fransico bay area is also the most popular area with regards to vehicle listings. Electric cars are the least popular in California and is seen mostly in the Bay area only since it is one of the wealthiest area in California.

TOP 5 TYPES OF VEHICLES AND THEIR DRIVE TYPE



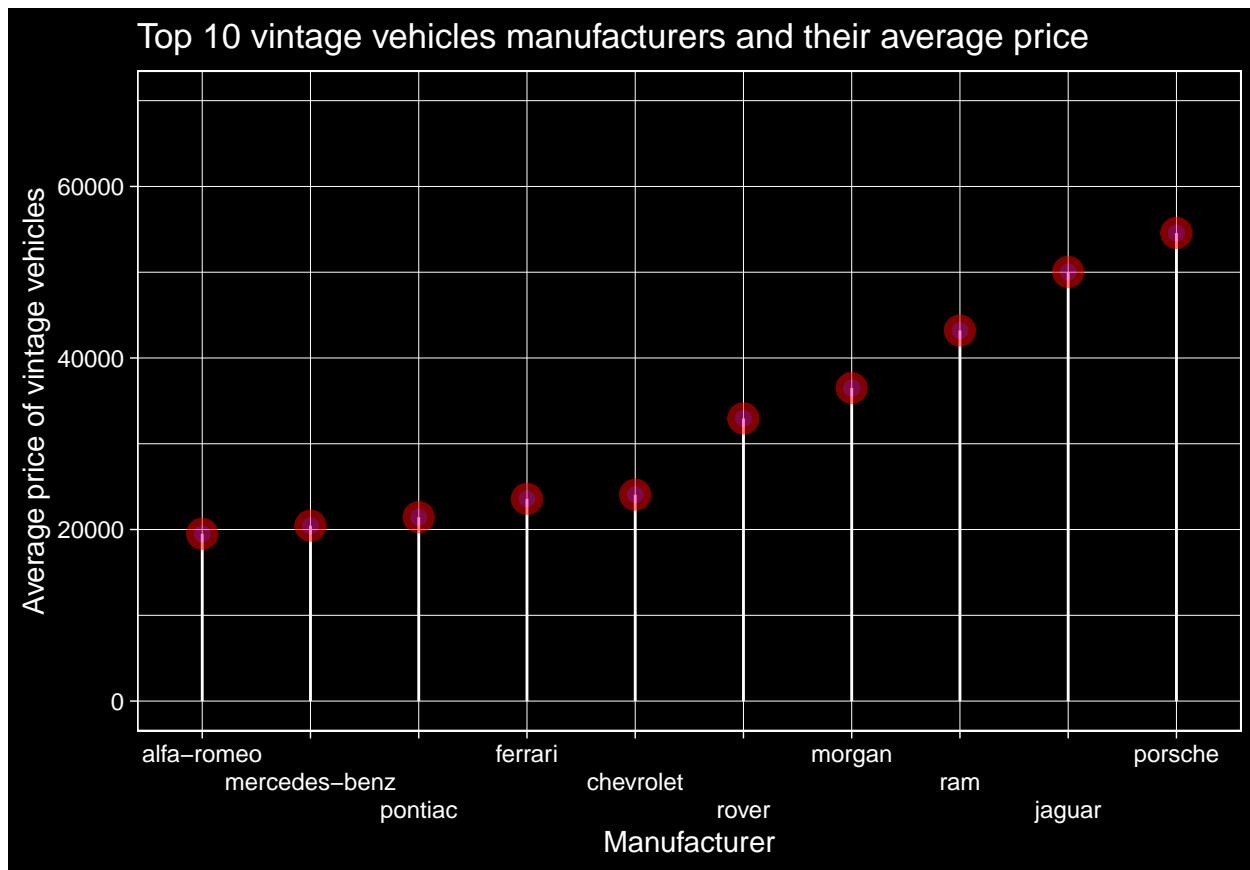
The most popular drive is four wheel drive which is popular in trucks, pickups and SUV but Sedans' most popular drive is forward wheel drive. The reverse wheel drive is most popular in a coupe.

TREND OF VEHICLE LISTING FROM 2005 TO 2012



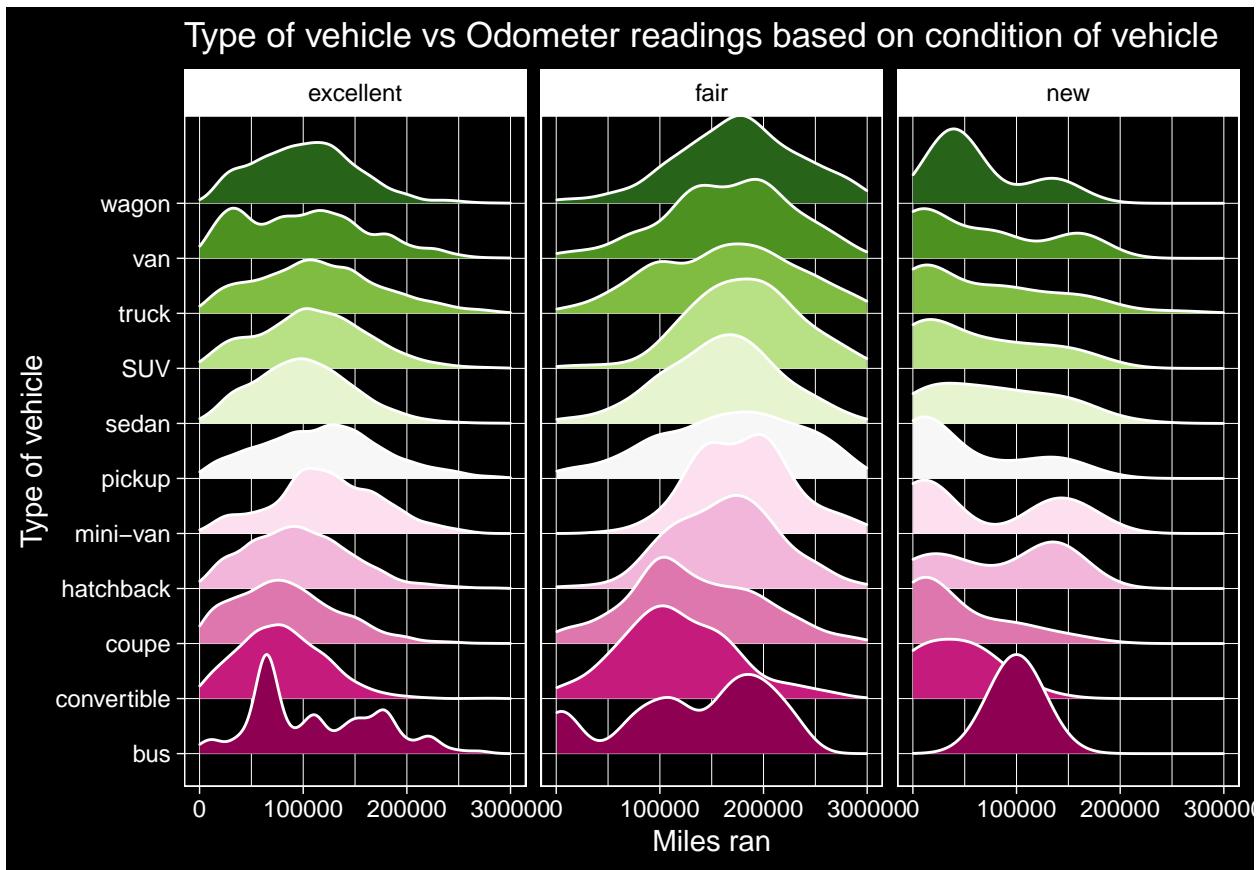
The number of vehicle listings increase slowly from 2005 but there is a huge dip in 2009 and then there was a huge steady increase. The slow increase from 2005 and the dip in 2009 could indicate the great recession period in the US where all industries including automobile industry were hit hard. After the great recession period, we can see that the vehicle listings are increasing in high numbers.

AVERAGE PRICE OF VINTAGE VEHICLES ACCORDING TO MANUFACTURER



When it comes to vintage vehicles, other than the most popular manufacturers, the average price is below \$20,000. Porsche, Jaguar, Ram, Morgan are the top vintage vehicle manufacturers.

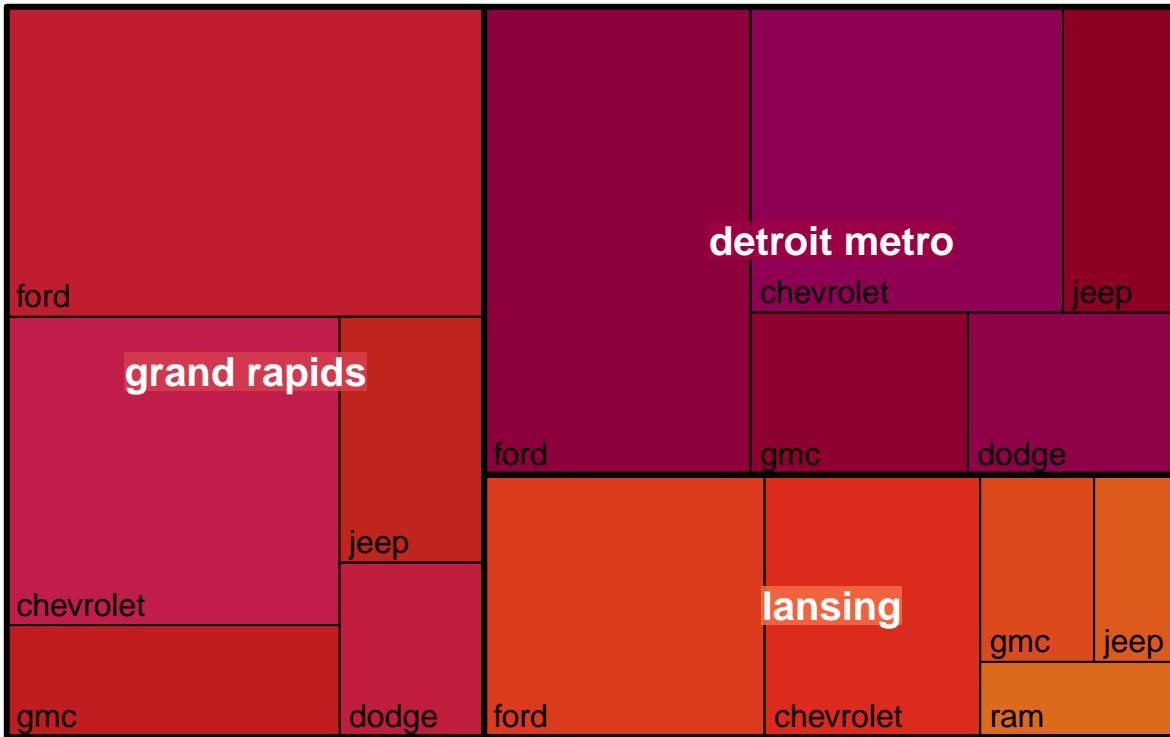
CHANGE IN CONDITION OF VEHICLE BASED ON MILES DRIVEN



The visualization represents how the miles driven by the vehicle affects the condition of the vehicle based on the type of vehicle. The new vehicles have driven the lowest miles all over and the fair vehicles have driven the most. Bus is the vehicle is that has traveled more miles across all three conditions.

MICHIGAN'S TOP MANUFACTURERS IN ITS TOP REGIONS

Top 5 manufacturers in top 5 regions of Michigan



Michigan is known for the automobile industry. The top manufacturer in Michigan are Ford, Chevrolet, Jeep, Gmc, Dodge, and Ram. The top regions in Michigan where vehicle manufacturing is done are the Grand Rapids, Detroit, and Lansing.

CONCLUSION

The Exploratory Data Analysis of the vehicle listings dataset proved efficient in understanding all the various factors that affect the price of a vehicle. This analysis can also be helpful in predicting future trends regarding the data. The data had a lot of outliers and missing values which was removed. I also found some listings where the values were randomly added by the user which made me doubt the credibility of the listing and so those values were also removed.

Looking at the number of listings state wise, **California** tops the list. This makes sense because California is the most populous state in the US. While just considering the state of California, it is evident that most popular fuel type used is **gas vehicles** especially in the **San Francisco bay area**. **Electric vehicles** are the least popular here, especially in the **Sacramento region** even though the region's population is pretty high.

While comparing the **price and odometer readings** for the all the listings, it can be seen that they are **inversely proportional** each other. When looking at this relationship on the basis of the condition of the vehicle, **Fair and Salvage** condition vehicles are very low in price and the **New** vehicles are very high in price which makes sense.

When observing the average miles driven, if we consider the age of the vehicle in groups of 10, upto **30 years** the average miles driven increase but there is a decrease after. A reason for the decrease could be

that since the vehicle is very old, in those times people did not drive much which has been reflected in this data. The miles driven by the vehicle also affects the condition of the vehicle. It can be seen that new and good condition vehicles have driven very few miles when compared to vehicles in the fair condition.

Going through the other features of a vehicle like cylinders and drives, it is seen that **4 cylinder** vehicles are the **cheapest** and most preferred among people. The **8 cylinder** vehicles are not preferred as much and they are more **expensive** than the 4 cylinder vehicles. The different drives of a vehicle also depends on the type of the vehicle. **Sedan**'s most popular drive is **forward wheel drive** which could be a reason why it is preferred most in the snowy states.

There were a variety of manufacturers in the listings where **Ford**, **Chevrolet** and **Toyota** tops the list. When considering **Vintage vehicles**, vehicles that are 45 years and older, the top manufacturers are **Porsche**, **Ferrari**, **Mercedes**, **Pontiac**, and **Jaguar** and their **average price** is high being a vintage classic vehicle. It is surprising that the top three manufacturers are not in the top vintage vehicle manufacturers.

An analysis was also made on **Michigan**. Michigan is number one automobile manufacturer state in the US. **Detroit** and **Grand Rapids** are the top regions in Michigan where vehicles are manufactured. Looking at the top here, it is interesting that **Toyota** is not in the top among the popular regions.

Overall, the analysis tells us that some key factors in buying a vehicle would be its drive, cylinder, price, manufacturer and the mileage it would give. This could help people looking to buy a vehicle off craigslist since most vehicles are second handed.

APPENDIX

DATA DICTIONARY

The data regarding all the relevant columns are listed here:

Region

The region column specifies the Craigslist region in each state.

Example: Sacramento, Detroit, Florida Keys

Price

The price at which the vehicle is being sold. Value varies from 3736928711.

Year

The year in which the vehicle was bought. Values vary from 1990 to 2022.

Manufacturer

The manufacturer of the vehicle listed.

Example : Ford, Tesla, Toyota

Condition

The condition of the vehicle currently.

Example: Fair, New, Excellent, Like new.

Cylinders

The number of cylinders that the vehicle possesses.

Example: 4 cylinders, 12 cylinders, 8 cylinders

Fuel

The type of fuel the vehicle uses.

Example : Gas, Electric, Diesel, and Hybrid

Odometer

The miles traveled by the vehicle. The value varies from 0 to 10000000.

Drive

The type of drive of the vehicle.

Example: fwd, 4wd, and rwd

Type

The type of the vehicle.

Example : Bus, Truck, SUV, Sedan

State

The state in which the vehicle was listed.

Example: California, Michigan, Florida

Lat

The latitude of the region.

Example : 25.97170

Long

The longitude of the region.

Example : -97.46990

posting_date

The date and time at which the vehicle listing was posted.

Example : 05/24/2021 5:10:10

NEW COLUMNS**Years_Used**

The number of years the car was used. It varies from 0 to 121.

Posted_Year

The year the listing was posted. It is between 2021 and 2022.