Exam

Neha Manish Vanjara

2022-08-05

INTRODUCTION

Americans rely on their automobiles for mobility; 91 percent of working adults commute to work in their cars, and cars offer a level of convenience that is frequently unsurpassed by other modes of transit. Purchasing a secondhand car may be exciting, whether you're 16 or 60. The average cost of a used car, according to some industry statistics, is close to $28,000, so it may also be a significant financial commitment. The overall number of registered vehicles in the United States is consistently rising year over year. Over six years, from 2014 to 2019, growth was constant. Approximately 260 million automobiles were registered in 2014, but by 2019, that figure had risen to over 276 million.

Do your research before visiting an auto dealer to look at used cars. You might actually save a lot of money. Take into account your budget, the type of car you require, and how you plan to utilize it. Don't forget about other expenses like registration, insurance, gas, and upkeep. Examine models, features, maintenance history, safety evaluations, and mileage. The used automobile listing utilized in the vehicle data set was found on Craigslist and includes columns for price, condition, manufacturer, latitude/longitude, and 18 other characteristics. We may also find information about each data set column in the data profiling report. Numerous uses for this data exist, including price prediction.

ABOUT CRAIGSLIST

Car purchase on Craigslist could result in significant savings on a pricey investment. On Craigslist, several vehicles are for sale. People can narrow down their search results to make it easier to identify the type of automobile they desire and the price range they can afford. They can look at the pictures if a car listing gets their attention. The listing includes information on the vehicles' year, make, model, and trim.

In this report I'm going to visualize these listings based on the above mentioned parameters.

AIM

1. Which car(manufacturer) has the highest number of listing on Craigslist?
2. What is the highest listing of different cars based on state?
3. Exploring top 5 States Manufacturer wise.
4. Figuring the kind of car somebody needs.
5. What is the price for different brand cars?

CONTEXT

Tools and techniques for data visualization assist uncover patterns and trends that are hidden beneath mountains of complex data, and data visualization results in better and quicker judgments.

This data set can be used by people around the US to study and predict future car sales. It explains the actual price, how to choose an automobile, and how to assess the car's condition. It even displays the most typical automobile model per state. The common characteristics and status of autos can be discovered by researchers.

THE DATA SET

The vehicle data set has 26 columns and 426880 rows. Each row represents an entry for a used car.The Data Set also has many missing values in each row.

TIDYING THE DATA SET

No data set is flawless, and missing values are frequently found in data sets. Since Craigslist ads allow for a lot of flexibility, several fields—including those for the condition, odometer, size, paint color, and cylinders—have blank values. When we look more closely, we see that most of the columns have multiple null entries. Either we can proceed by learning how to estimate these numbers, or we may do away with the rows altogether. There are missing values for several factors, including condition, odometer, size, paint color, and cylinders. Since none of the columns have any values and there are zero rows as a result, we are unable to use omit.na(). I've taken out any missing rows from the particular visualization to avoid this.

LOADING THE PACKAGE

```
packages <-
  c(
    "tidyverse",
    "janitor",
    "data.table",
    "devtools",
    "ggmap",
    "viridis",
    "lubridate",
    "reshape",
    "ggplot2",
    "wordcloud",
    "knitr",
    "rworldmap",
    "rworldxtra",
    "dplyr",
    "tidyr",
    "kableExtra"
  )
for (package in packages) {
  library(package, character.only = TRUE)
}
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.6     v purrr   0.3.4
## v tibble  3.1.7     v dplyr   1.0.9
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.1.2     v forcats 0.5.1

## -- Conflicts --------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

## Warning: package 'janitor' was built under R version 4.2.1

##
## Attaching package: 'janitor'

## The following objects are masked from 'package:stats':
##
##     chisq.test, fisher.test

##
## Attaching package: 'data.table'

## The following objects are masked from 'package:dplyr':
##
##     between, first, last

## The following object is masked from 'package:purrr':
##
##     transpose

## Loading required package: usethis

## Google's Terms of Service: https://cloud.google.com/maps-platform/terms/.

## Please cite ggmap if you use it! See citation("ggmap") for details.

## Loading required package: viridisLite

##
## Attaching package: 'lubridate'

## The following objects are masked from 'package:data.table':
##
##     hour, isoweek, mday, minute, month, quarter, second, wday, week,
##     yday, year

## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union

##
## Attaching package: 'reshape'
```

```
## The following object is masked from 'package:lubridate':
##
##     stamp

## The following object is masked from 'package:data.table':
##
##     melt

## The following object is masked from 'package:dplyr':
##
##     rename

## The following objects are masked from 'package:tidyr':
##
##     expand, smiths

## Loading required package: RColorBrewer

## Warning: package 'rworldmap' was built under R version 4.2.1

## Loading required package: sp

## ### Welcome to rworldmap ###

## For a short introduction type :   vignette('rworldmap')

## Warning: package 'rworldxtra' was built under R version 4.2.1

## Warning: package 'kableExtra' was built under R version 4.2.1

##
## Attaching package: 'kableExtra'

## The following object is masked from 'package:dplyr':
##
##     group_rows
```

LOADING THE DATA SET

```
vehicle<-fread("vehicles.csv") # data set
dim(vehicle) #calculating rows and columns
```

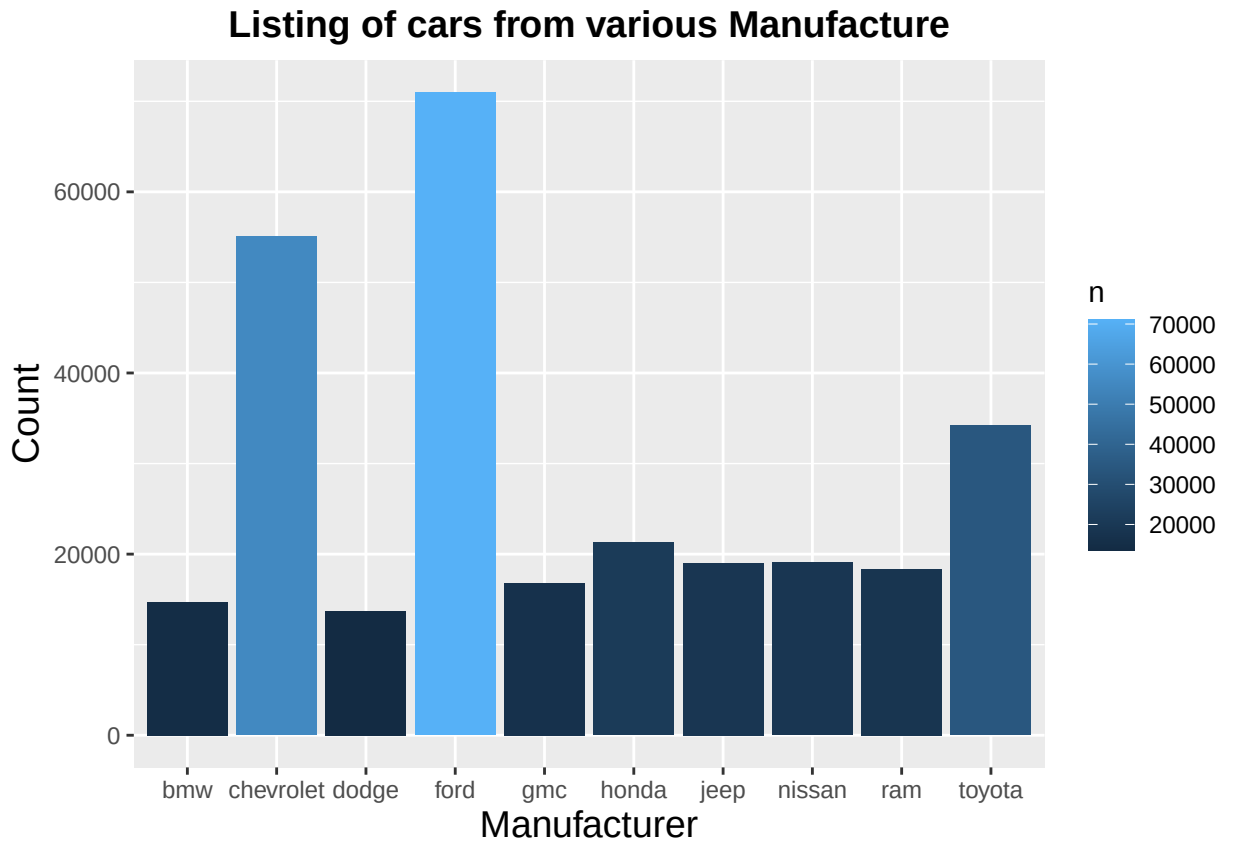
```
## [1] 426880     26
```

VISUALIZING THE DATA SET

```r
#1 Counting number of vehicles using manufacture

Manufacture_count <- vehicle %>% group_by(manufacturer) %>% count(sort = TRUE)
Manufacture_count <- Manufacture_count[-8, ] # omitting row 8 as it has missing value
Manufacture_count
```

```
## # A tibble: 42 x 2
## # Groups:   manufacturer [42]
##    manufacturer     n
##    <chr>        <int>
##  1 ford         70985
##  2 chevrolet    55064
##  3 toyota       34202
##  4 honda        21269
##  5 nissan       19067
##  6 jeep         19014
##  7 ram          18342
##  8 gmc          16785
##  9 bmw          14699
## 10 dodge        13707
## # ... with 32 more rows
```

```r
Manufacture_count1 <- Manufacture_count[1:10,]
manufacture <- ggplot(Manufacture_count1, aes(x=manufacturer ,y=n, fill = n))+
  geom_bar(stat = "identity") + labs(x="Manufacturer",y="Count") +
  ggtitle("Listing of cars from various Manufacture")+
  theme(plot.title = element_text(color="black", size=14, face="bold"),
    axis.title.x = element_text(color="black", size=14),
    axis.title.y = element_text(color="black", size=14)) +
  theme(plot.title=element_text(hjust=0.5))
manufacture
```



We can see from the above bar graph that Ford has the most car listings in the United States. There are more than 70000 Ford vehicles listed on Craigslist. This indicates that the majority of Americans favor Ford automobiles. Toyota is listed second on the list after Chevrolet, which has almost 50000 listings.

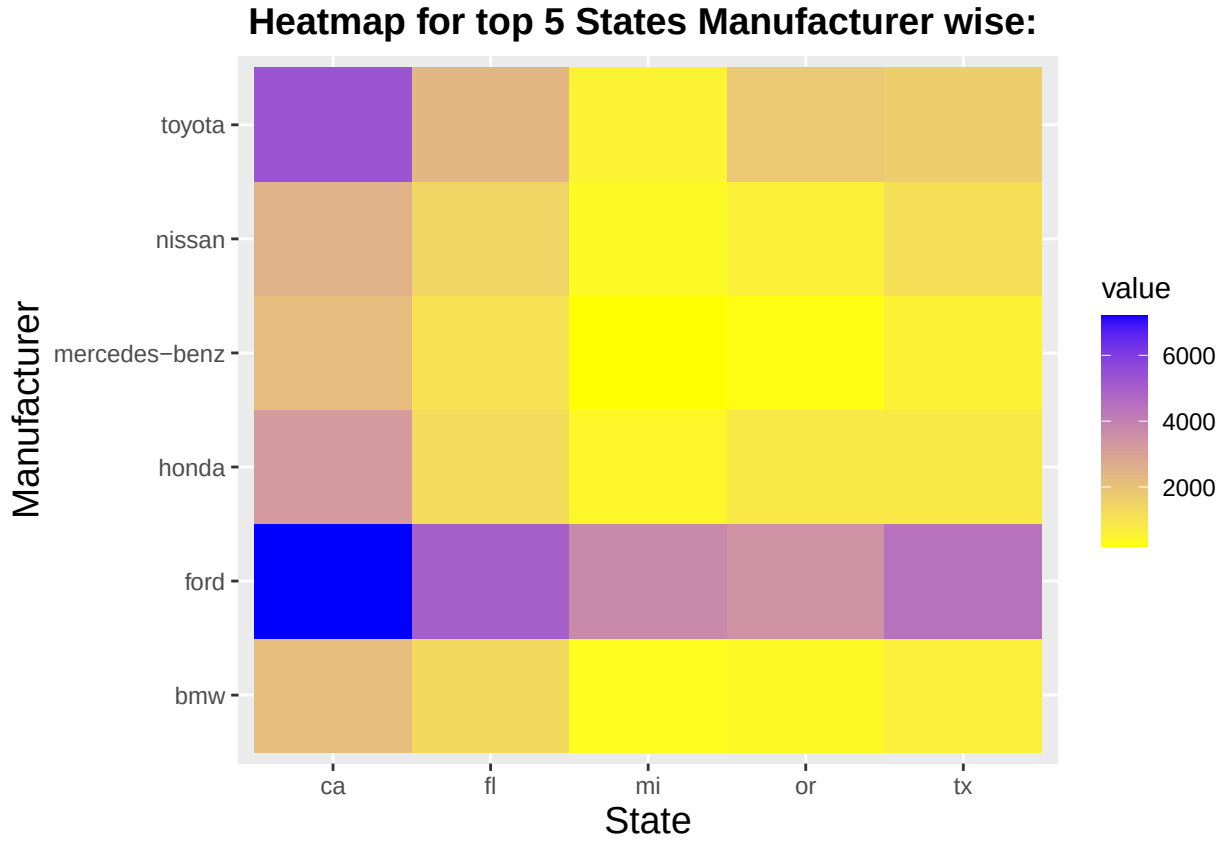
```
#2 Listings of manufacturer based on various state

State_Manufacturer_count <- vehicle %>% group_by(state , manufacturer) %>%
  count(sort = TRUE)
State_Manufacturer_count
```

```
## # A tibble: 1,977 x 3
## # Groups:   state, manufacturer [1,977]
##    state manufacturer     n
##    <chr> <chr>        <int>
##  1 ca    ford          7189
##  2 ca    chevrolet     5607
##  3 ca    toyota        5327
##  4 fl    ford          5033
##  5 tx    ford          4456
##  6 mi    ford          3753
##  7 or    ford          3470
##  8 ca    honda         3213
##  9 tx    chevrolet     3133
## 10 oh    ford          3093
## # ... with 1,967 more rows
```

```
#3 Heatmap for top 5 States Manufacturer wise

heatmap<- select(vehicle, state , manufacturer)
xtabA<- filter(heatmap,
                    manufacturer %in%c('ford','chevroet','toyota','honda',
                                       'nissan',   'mercedes-benz','bmw'))
xtabA<- filter(xtabA, state %in%c('ca','fl','tx','mi','or'))
p<-table(xtabA$state,xtabA$manufacturer)
data_melt <- melt(p)
```

```
## Warning in type.convert.default(X[[i]], ...): 'as.is' should be specified by the
## caller; using TRUE

## Warning in type.convert.default(X[[i]], ...): 'as.is' should be specified by the
## caller; using TRUE
```

```
xc<-ggplot(data_melt,aes(x=Var.1,y=Var.2))+geom_tile(aes(fill=value))+
  scale_fill_gradient(low="yellow",high="blue")
xc+ggtitle("Heatmap for top 5 States Manufacturer wise: ")+xlab("State")+
  ylab("Manufacturer")+theme(
  plot.title = element_text(color="black", size=14, face="bold"),
  axis.title.x = element_text(color="black", size=14),
  axis.title.y = element_text(color="black", size=14)
) + theme(plot.title=element_text(hjust=0.5))
```



The heatmap and table illustrate the various US states and manufacturers. Most postings are in the blue-represented areas, while the fewest are in the yellow-represented ones. It provides information on the number of cars from various manufacturers in various states. We can see from the data that California has a significant number of Ford vehicles listed. Ford is the market leader in Florida, Texas, and Michigan. Compared to other states, California has a tremendous number of cars, which may be a result of its large population. Since California has a population that is 24.9% higher than Texas, there are more listings for cars there.

```
#4 Cloud Image of Manufacturer in CA

cloud<-filter(vehicle,state%in%c('ca'))
cloud<-select(cloud,state,type)
cloud <- cloud %>% group_by(type)%>% summarise(Count = n()) %>% arrange(desc(Count))
wordcloud(words=cloud$type,freq = cloud$Count,scale=c(4,.5),
          max.words=100,rot.per=.5,colors=palette())
```

According to the cloud maps, California is the state where sedans are most frequently listed. Drivers in Southern California who frequently face challenging commutes and challenging parking situations should consider sedans as a practical option. This demonstrates the necessity of owning a sedan.

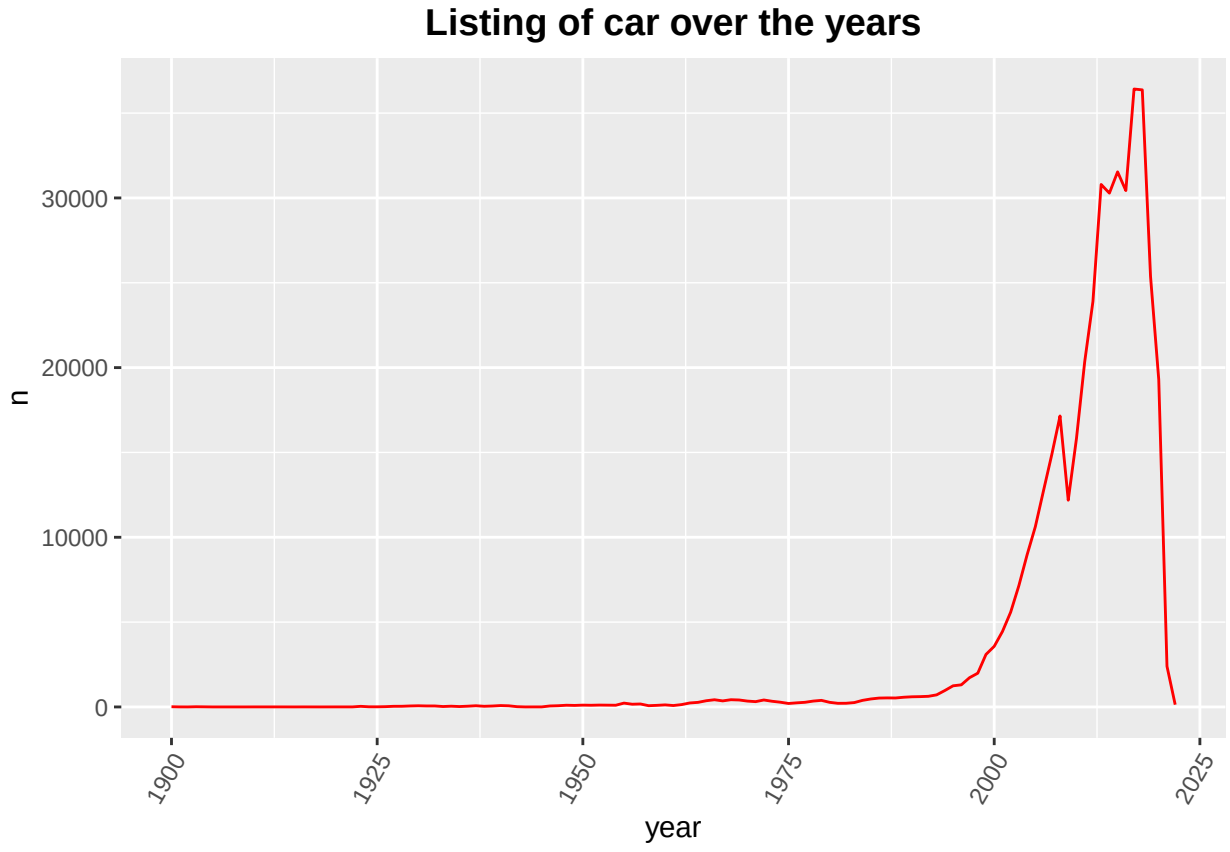
```
#5 Listing of car over the years
Year_count <- vehicle %>% group_by(year) %>% count(sort = TRUE)
Year_count
```

```
## # A tibble: 115 x 2
## # Groups:   year [115]
##     year     n
```

```
##    <int> <int>
##  1  2017 36420
##  2  2018 36369
##  3  2015 31538
##  4  2013 30794
##  5  2016 30434
##  6  2014 30283
##  7  2019 25375
##  8  2012 23898
##  9  2011 20341
## 10  2020 19298
## # ... with 105 more rows
```

```
b<- ggplot(Year_count, aes(x=year)) +
  geom_line( mapping = aes(y=n, group =1),color="red") +
  ggtitle("Listing of car over the years")+theme(
    plot.title = element_text(color="black", size=14, face="bold")) +
  theme(axis.text.x = element_text(angle = 60, hjust = 1)) +
  theme(plot.title=element_text(hjust=0.5))
b
```

```
## Warning: Removed 1 row(s) containing missing values (geom_path).
```



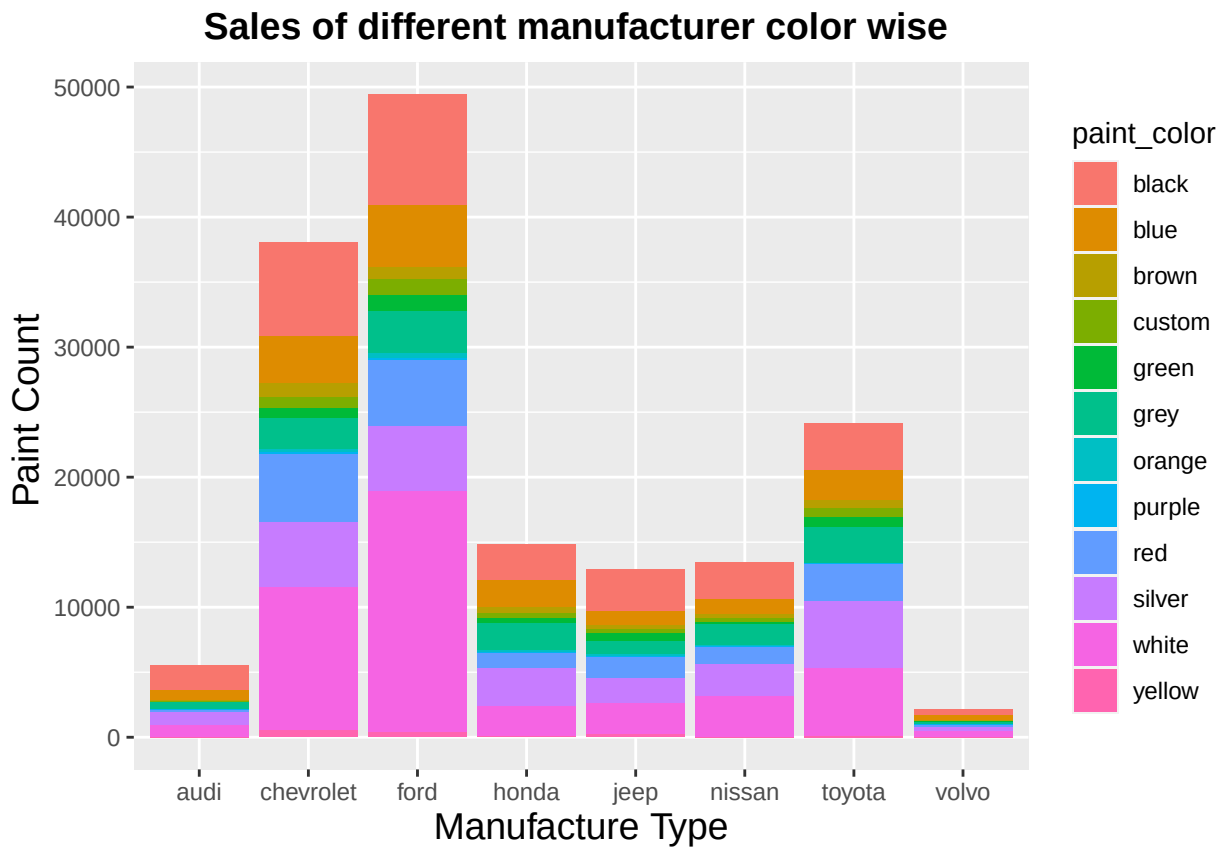
This line graph shows that the number of automobile listings has grown over time. This is closely related to the production and use of automobiles. Even the decline of 2007–2008 is visible. The economic disruption

that resulted in trade imbalances was accelerated by the recession of 2007–2008. One-fifth of the fall in auto sales can be attributed to the over 50 percent increase in oil costs in the years 2007–2008. People also believed that wealth was declining as a result of the reduction in housing prices. The graph above demonstrates the fact that 2012 saw another recession. Once more, we observe a decline around 2012.

```
#6 Number of manufacturer color wise

vehicle1 <- vehicle %>% filter(manufacturer %in% c("ford","chevrolet",
                      "toyota","honda","nissan",
                      "jeep","volvo","audi"))

c_count <- vehicle1 %>% group_by(manufacturer, paint_color) %>% count(sort = TRUE)
c_count <- c_count[!(c_count$paint_color == ""), ]
c_count
```

```
## # A tibble: 96 x 3
## # Groups:   manufacturer, paint_color [96]
##    manufacturer paint_color     n
##    <chr>        <chr>       <int>
##  1 ford         white       18609
##  2 chevrolet    white       11075
##  3 ford         black        8573
##  4 chevrolet    black        7225
##  5 chevrolet    red          5281
##  6 toyota       white        5267
##  7 ford         red          5167
##  8 toyota       silver       5120
##  9 ford         silver       4929
## 10 chevrolet    silver       4927
## # ... with 86 more rows
```

```
c <- ggplot(c_count, aes(x=manufacturer ,y=n, fill = paint_color))+
  geom_bar(stat = "identity") + labs(x="Manufacture Type",y=" Paint Count") +
  ggtitle("Sales of different manufacturer color wise ")+
  theme(plot.title = element_text(color="black", size=14, face="bold"),
        axis.title.x = element_text(color="black", size=14),
        axis.title.y = element_text(color="black", size=14)) +
  theme(plot.title=element_text(hjust=0.5))
c
```

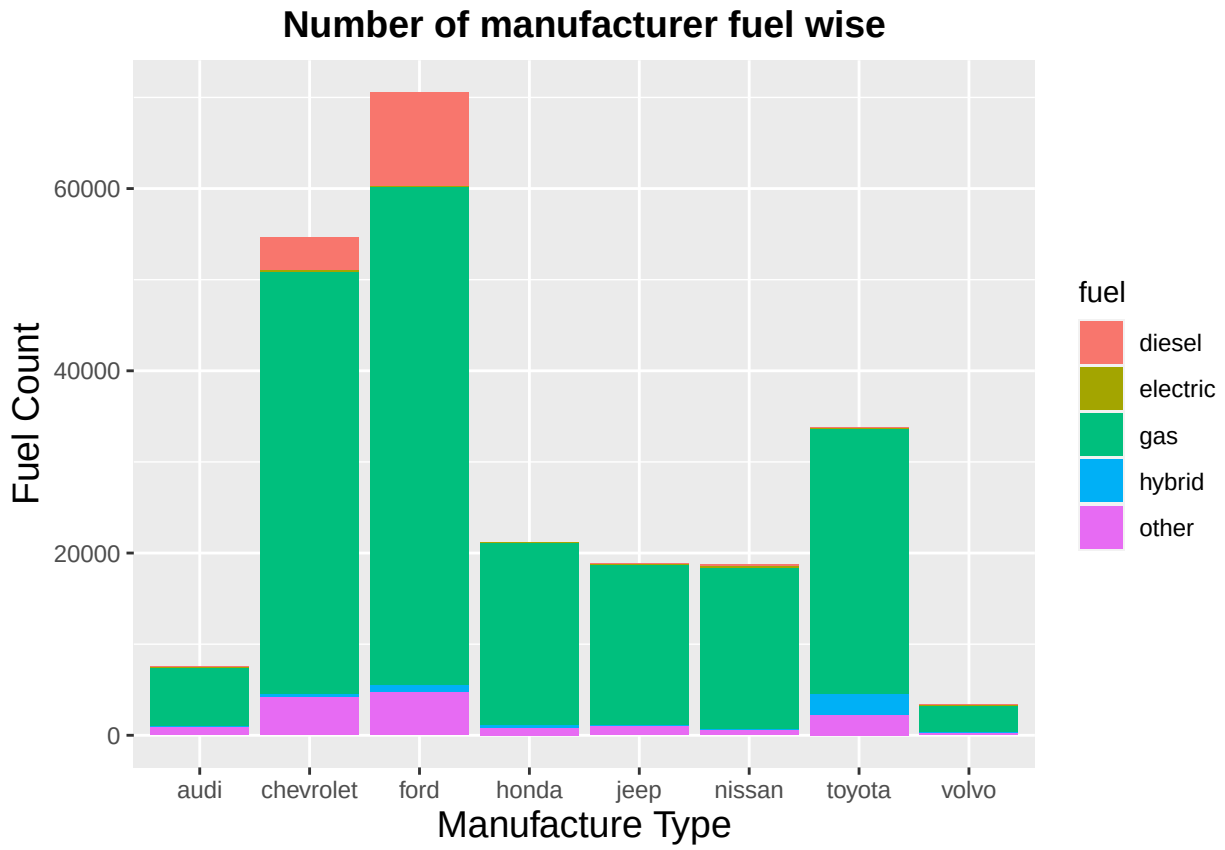


Regardless of the manufacturer type, the Statistic Plot in this case reveals that white is the most popular paint color (aside from missing values). This indicates that individuals prefer to drive white automobiles. Black is the second most popular listing color. We are aware that these two colors of cars are the most popular and basic.

```
#7 Number of manufacturer fuel wise

f_count <- vehicle1 %>% group_by(manufacturer, fuel) %>% count(sort = TRUE)
f_count <- f_count[!(f_count$fuel == ""), ]
f_count
```

```
## # A tibble: 40 x 3
## # Groups:   manufacturer, fuel [40]
##    manufacturer fuel       n
##    <chr>        <chr>  <int>
##  1 ford         gas    54669
##  2 chevrolet    gas    46386
##  3 toyota       gas    29168
##  4 honda        gas    19996
##  5 nissan       gas    17736
##  6 jeep         gas    17633
##  7 ford         diesel 10282
##  8 audi         gas     6380
##  9 ford         other   4829
## 10 chevrolet    other   4240
## # ... with 30 more rows
```

```
d <- ggplot(f_count, aes(x=manufacturer ,y=n, fill = fuel))+
  geom_bar(stat = "identity") + labs(x="Manufacture Type",y="Fuel Count") +
  ggtitle(" Number of manufacturer fuel wise ")+
  theme(plot.title = element_text(color="black", size=14, face="bold"),
        axis.title.x = element_text(color="black", size=14),
        axis.title.y = element_text(color="black", size=14)) +
  theme(plot.title=element_text(hjust=0.5))
d
```



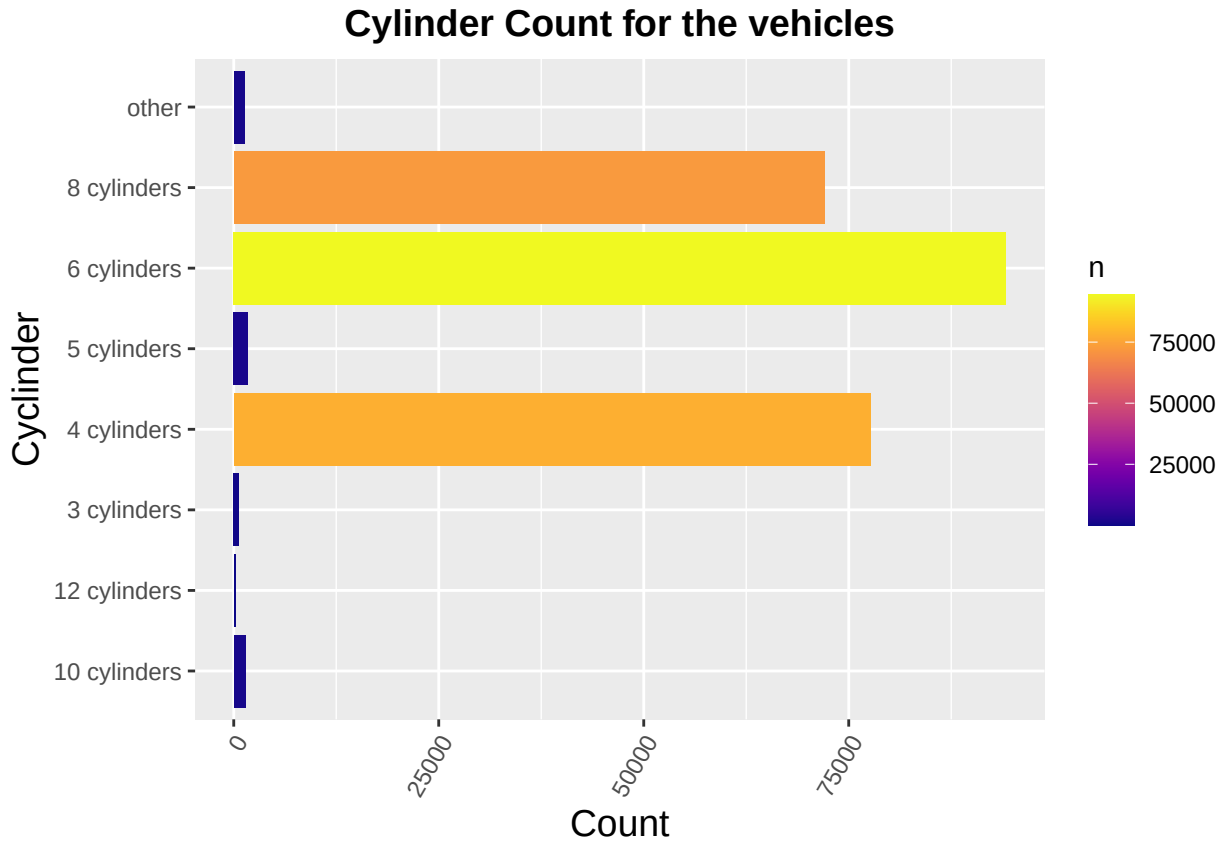
The graph above shows that gas is the most popular fuel type. All car models use gas as their main fuel. The cost of gas is clearly to blame for this. The cost of gas is significantly lower than that of other fuels.

```
#8 Cylinder Count for the vehicles

Cylinder_count <- vehicle %>% group_by(cylinders) %>% count(sort = TRUE)

Cylinder_count <- Cylinder_count[!(Cylinder_count$cylinders == ""), ]
Cylinder_count
```

```
## # A tibble: 8 x 2
## # Groups:   cylinders [8]
##   cylinders        n
##   <chr>        <int>
## 1 6 cylinders  94169
## 2 4 cylinders  77642
```

```
## 3 8 cylinders  72062
## 4 5 cylinders   1712
## 5 10 cylinders  1455
## 6 other         1298
## 7 3 cylinders    655
## 8 12 cylinders   209
```

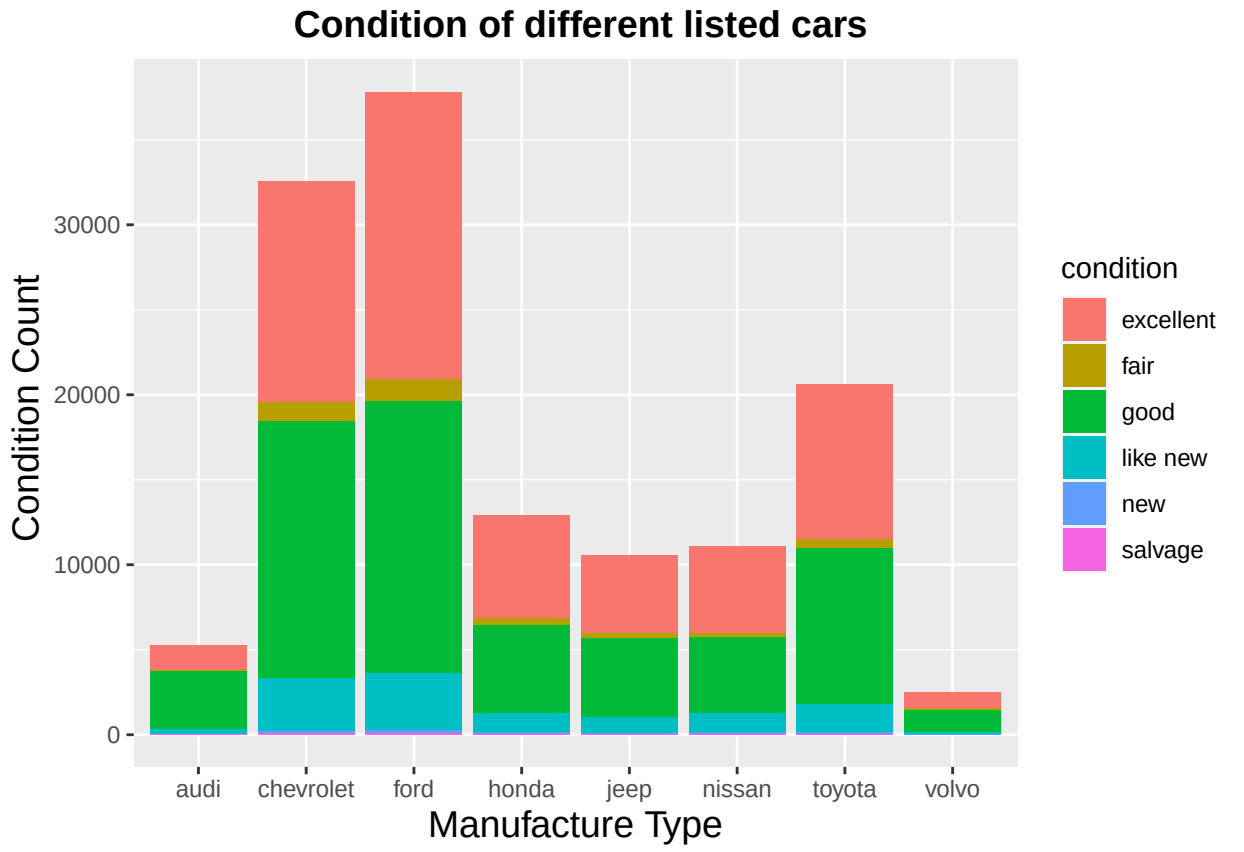
```
cylinder <- ggplot(Cylinder_count, aes(x=cylinders ,y=n, fill = n))+
  geom_bar(stat = "identity") + labs(x="Cyclinder",y="Count") +  coord_flip() +
  ggtitle("Cylinder Count for the vehicles")+
  theme(plot.title = element_text(color="black", size=14, face="bold"),
        axis.title.x = element_text(color="black", size=14),
        axis.title.y = element_text(color="black", size=14)) +
  theme(axis.text.x = element_text(angle = 60, hjust = 1)) +
  theme(plot.title=element_text(hjust=0.5))
cylinder + scale_fill_viridis(option = "C")
```



The most prevalent type of cylinder, according to this graph, is type 6. Type 4 and type 8 come after this type. We can observe from the bar plot above that the bulk of modern cars is powered by 4, 5, 6, or 8-cylinder engines.

```
#9 Condition of different listed cars

con_count <- vehicle1 %>% group_by(manufacturer, condition) %>% count(sort = TRUE)
con_count <- con_count[!(con_count$condition == ""), ]
```

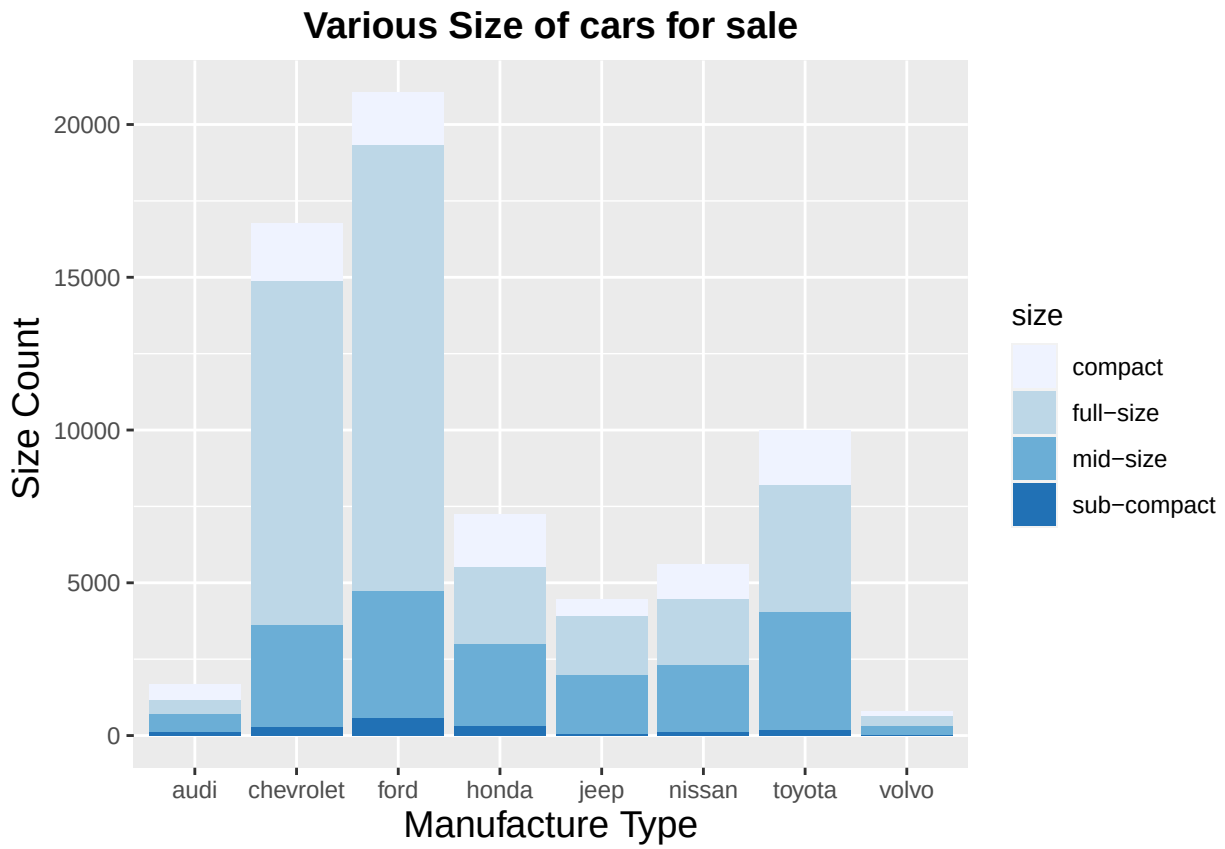
```
ab <- ggplot(con_count, aes(x=manufacturer ,y=n, fill = condition))+
  geom_bar(stat = "identity") + labs(x="Manufacture Type",y=" Condition Count") +
  ggtitle(" Condition of different listed cars")+
  theme(plot.title = element_text(color="black", size=14, face="bold"),
        axis.title.x = element_text(color="black", size=14),
        axis.title.y = element_text(color="black", size=14)) +
  theme(plot.title=element_text(hjust=0.5))
ab
```



The vast majority of the vehicles that are offered on Craigslist are in good shape. As a result, it is a convenient and safe website for used automobile buyers. We can observe that very few automobiles are in salvageable condition. Even now, there are relatively few new cars on the road, and most individuals will keep the vehicles they recently purchased.

```
#10 Various Size of cars for sale

size_count <- vehicle1 %>% group_by(manufacturer, size) %>% count(sort = TRUE)
size_count <- size_count[!(size_count$size == ""), ]

abc <- ggplot(size_count, aes(x=manufacturer ,y=n, fill = size))+
  geom_bar(stat = "identity") + labs(x="Manufacture Type",y=" Size Count") +
  ggtitle(" Various Size of cars for sale ")+
  theme(plot.title = element_text(color="black", size=14, face="bold"),
        axis.title.x = element_text(color="black", size=14),
        axis.title.y = element_text(color="black", size=14)) +
  theme(plot.title=element_text(hjust=0.5))
```

```
abc + scale_fill_brewer()
```



In this instance, we can see that Americans prefer full-size cars. Mid-size cars are the second most popular. Many people, notably contractors, who need larger pickup trucks for their jobs own them. However, just as many people also utilize these larger cars to transport their families. Because the nation's roads and transportation system have expanded to accommodate them, Americans now purchase larger cars.

```
#11 Price Chart of number of cars

p_count <- vehicle %>% group_by(price) %>% count(sort = TRUE)
p_count <- p_count[!(p_count$price == ""), ]
p_count
```

```
## # A tibble: 15,654 x 2
## # Groups:   price [15,654]
##      price     n
##    <int64> <int>
##  1    6995  3169
##  2    7995  3129
##  3    9995  2867
##  4    8995  2837
##  5    4500  2778
##  6    5995  2727
##  7    3500  2716
##  8   29990  2705
```

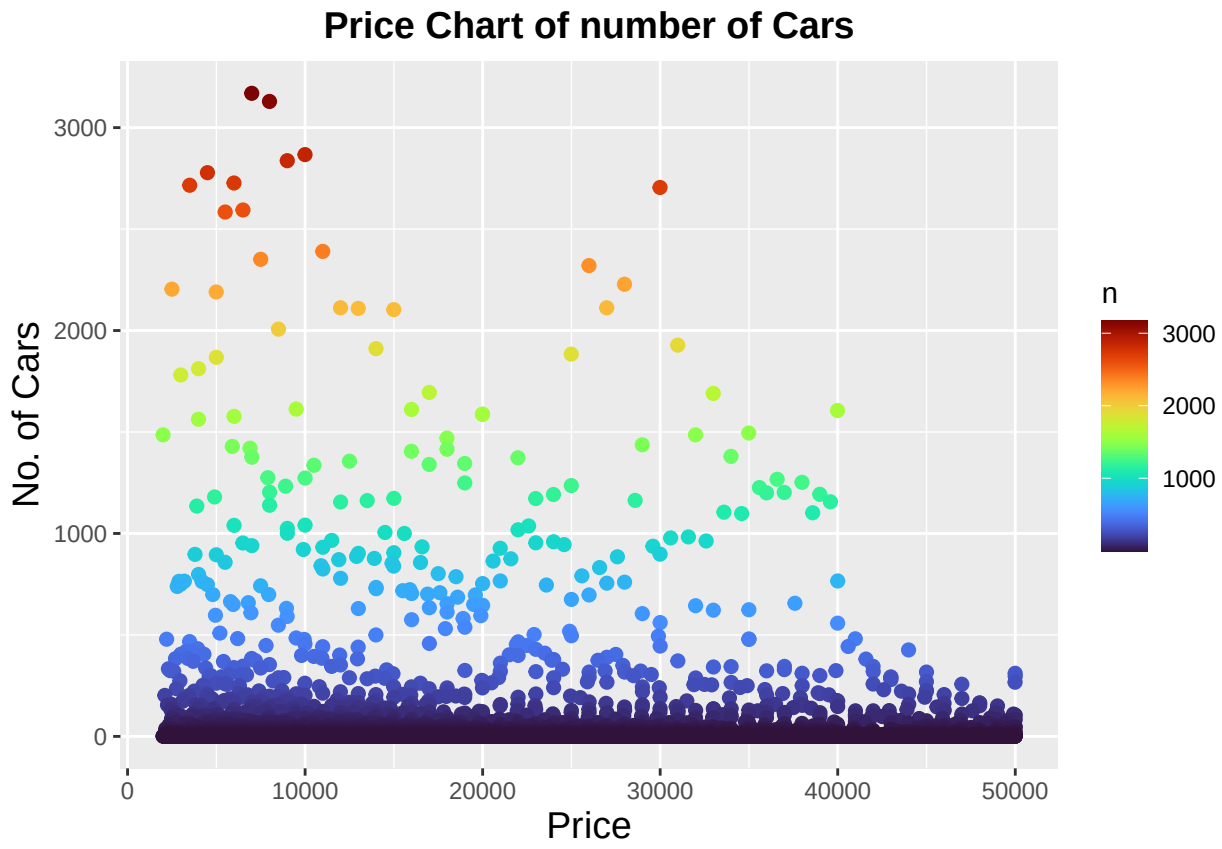
```
##  9    6500  2594
## 10    5500  2584
## # ... with 15,644 more rows
```

```
p_count$price  <- as.integer(p_count$price)
```

```
## Warning in as.integer.integer64(p_count$price): NAs produced by integer overflow
```

```
e <- ggplot(p_count, aes(x=price ,y=n))+
  geom_point(size=2, aes(colour= n )) + labs(x="Price",y="No. of Cars") +
  ggtitle("Price Chart of number of Cars")+
  theme(plot.title = element_text(color="black", size=14, face="bold"),
        axis.title.x = element_text(color="black", size=14),
        axis.title.y = element_text(color="black", size=14)) + xlim(2000,50000) +
  theme(plot.title=element_text(hjust=0.5))
e + scale_color_viridis(option = "H")
```

```
## Warning: Removed 2694 rows containing missing values (geom_point).
```



Here, I have used a scatter plot to try to visualize. This demonstrates a correlation between pricing and the number of cars listed. More cars are being sold for less, as can be seen. This clearly demonstrates that fewer autos are listed as prices rise. People like to purchase cars at a lower cost, thus they go for used vehicles. People won't opt to spend more money on used autos.

```r
#12 Price Chart of number of Car

mp_count <- vehicle1 %>% group_by(manufacturer, price) %>% count(sort = TRUE)
mp_count
```

```
## # A tibble: 23,015 x 3
## # Groups:   manufacturer, price [23,015]
##    manufacturer   price     n
##    <chr>        <int64> <int>
##  1 ford               0  6130
##  2 chevrolet          0  4421
##  3 toyota             0  2602
##  4 nissan             0  1672
##  5 jeep               0  1553
##  6 honda              0  1402
##  7 chevrolet      29990   554
##  8 ford            6995   554
##  9 ford            9995   502
## 10 ford            7995   501
## # ... with 23,005 more rows
```

```r
mp_count <- mp_count[!(mp_count$price == "0"), ]
mp_count$price  <- as.integer(mp_count$price)
```

```
## Warning in as.integer.integer64(mp_count$price): NAs produced by integer
## overflow
```

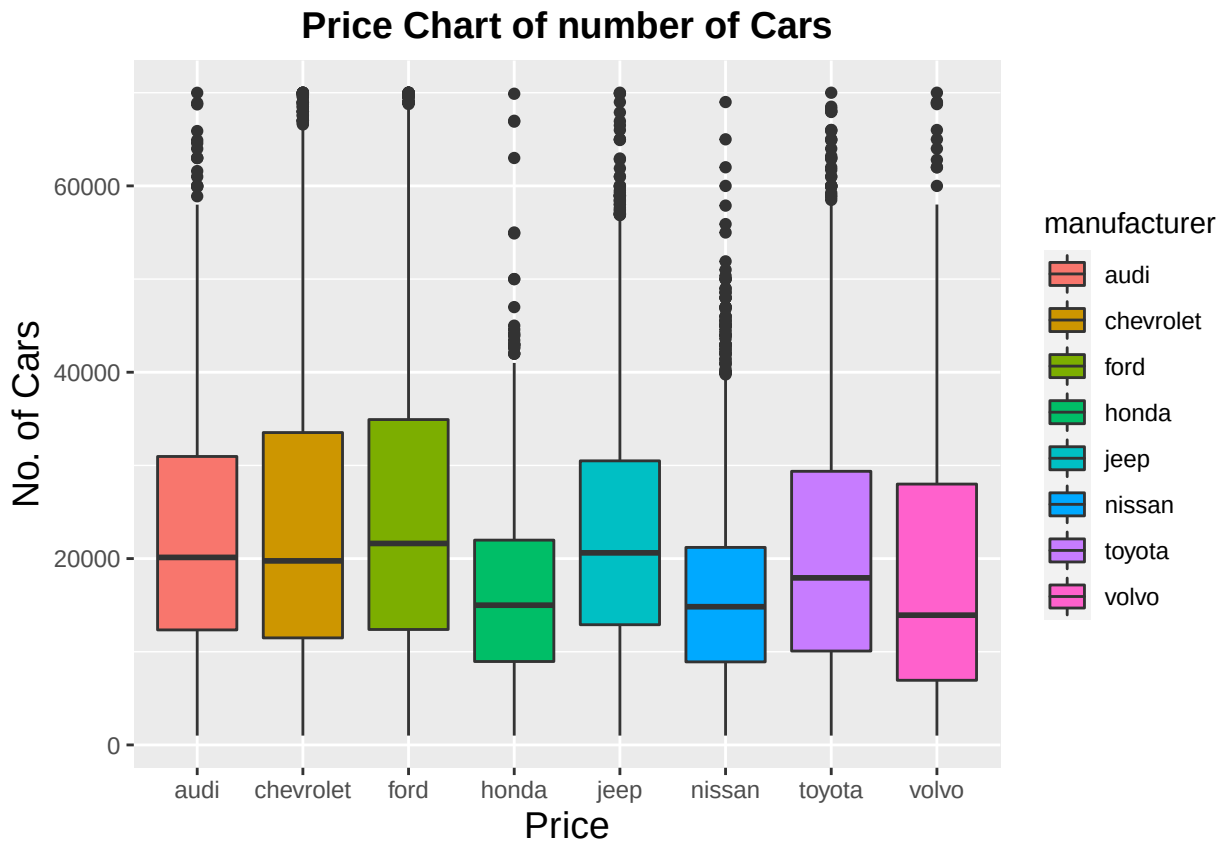
```r
mp_count
```

```
## # A tibble: 23,007 x 3
## # Groups:   manufacturer, price [23,007]
##    manufacturer price     n
##    <chr>        <int> <int>
##  1 chevrolet    29990   554
##  2 ford          6995   554
##  3 ford          9995   502
##  4 ford          7995   501
##  5 ford          8995   474
##  6 ford          5995   450
##  7 chevrolet     6995   427
##  8 ford         14995   411
##  9 ford         10995   409
## 10 ford          5500   408
## # ... with 22,997 more rows
```

```r
f <- ggplot(mp_count, aes(x=manufacturer, y=price , fill= manufacturer)) +
  geom_boxplot() + labs(x="Price",y="No. of Cars") +
  ggtitle("Price Chart of number of Cars")+
  theme(plot.title = element_text(color="black", size=14, face="bold"),
        axis.title.x = element_text(color="black", size=14),
        axis.title.y = element_text(color="black", size=14)) + ylim(1000,70000) +
```

```
  theme(plot.title=element_text(hjust=0.5))

f
```

```
## Warning: Removed 2172 rows containing non-finite values (stat_boxplot).
```



A boxplot of a few cars with a variety in price is shown here. Even while we are aware that Ford sells more automobiles overall, the average number of cars mentioned is practically identical to that of Audi, despite the latter's lower listing. It is based solely on the cost as shown in the graph above. Nissan and Honda vehicles are very scarce, despite the fact that their price range is not very high.

CONCLUSION

After thorough analysis of the data set, we were successful in making reasonable conclusions regarding the sale trends in used cars and automobile industry generally. From the above visualizations we observe that on the Craigslist listings, the highest number of car listed is from the brand Ford. This could be due to various reasons which we see later in the report. Ford is among the companies looking to offer a range of transportation services, rather than simply producing and selling cars. Then from the heatmap analysis of 5 highest manufacturer states we can see that California has the highest number of listings. Moreover, from our above analysis we can even see that California has Ford listings in the top followed by Toyota. We can see that from the table as well. A fair reason would be that inn California, a state that is generally more expensive to live in, the purchase of a car has to be made a lot more thoughtfully than many mid-western states where car ownership is a lot cheaper. All the cars in California's top 10 are fuel efficient sedans or hatchbacks and this is stated by the cloud image from our visualization.

We notice a fall in the listings in the year 2007-2008. Automobiles faced one of the hardest-hit during the recession. It was affected by two things: 1. Increase in the oil price(negative relationship between oil prices and auto sales). 2. Impact of the decline in home prices on auto sales.

From the next 5 bar graphs we can gather alot of information about buying a used car from Craigslist. These graphs gives us details for each many column in the dataset.For example the most popular paint colour (apart from missing values), is white. I was able to predict price of a car given the state it is being sold in , its type , title, drive, mileage and condition. From those graphs a person looking for a used car can see that they should go for a Ford car and specifically a white one from the graphs. Statistic shows that most of the people go for gas type fuel and 6 litre cyclinder. Majority of the modern cars have 4,6 and 8 litre cyclinder installed in them. Craigslist is a trsutworthy website and hence most of the available cars is in excellent condition.

The price chart for cars shows us that most of the listed cars are in a reasonable range. The number of cars listed and their prices are directly proportional. There are very few cars in high range and it is because people are trying to find cheaper options and that is the reason they're looking for second hand cars. Different brand cars have different prices. Even though Ford and Chevrolet are the most listed cars, the mean price of other top end cars such as audi, volvo is almost same.

I was somewhat successful in giving a general idea but because of limitations of time and data set we still have a long way to go.

SUMMARY

1. Downloaded the data set and performed cleaning operations.
2. Transformed the data set as required while answering certain questions.
3. Analysed data set at various levels and tried to identify some interesting trends.
4. Looked into the reasons for the trends observed.

FUTURE WORK

1. Perform a similar analysis using a similar dataset of another country.
2. Draw a comparison between the trends seen in the two countries.
3. Make use of the few columns left unused in this analysis.

DATA DICTIONARY

Mentioned below are the columns essential to our analysis.

1. manufacturer: Make of the vehicle listed.
2. price: Asking price for the vehicle in the listing.
3. state: State code of where the listing is made.
4. year :Year of registration of the vehicle listed.
5. condition: Condition of the vehicle listed.
6. cylinders: Engine size, based on the number of cylinders it has.
7. size: Which size category the vehicle falls in.
8. lat: Latitude of from where the listing is made.
9. long: Longitude of from where the listing is made.
10. region: Region from where the listing is made.
11. model: Model name of the vehicle listed.
12. odometer: The number of miles on the odometer of the vehicle.

13. type: Separates the vehicles on the basis of their type, eg. Hatchback, Pickup, Sedan etc.
14. transmission: Varied types of transmission are given.

TYPES OF GRAPHS VISUALIZED

1. Bar Plot
2. Scatter Plot
3. Box Plot
4. Line Graph
5. Heatmaps
6. cloud Image
7. Statistic Plot

```
sessionInfo()
```

```
## R version 4.2.0 (2022-04-22 ucrt)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 22000)
##
## Matrix products: default
##
## locale:
## [1] LC_COLLATE=English_United States.utf8
## [2] LC_CTYPE=English_United States.utf8
## [3] LC_MONETARY=English_United States.utf8
## [4] LC_NUMERIC=C
## [5] LC_TIME=English_United States.utf8
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
##  [1] kableExtra_1.3.4   rworldxtra_1.01    rworldmap_1.3-6    sp_1.5-0
##  [5] knitr_1.39         wordcloud_2.6      RColorBrewer_1.1-3 reshape_0.8.9
##  [9] lubridate_1.8.0    viridis_0.6.2      viridisLite_0.4.0  ggmap_3.0.0
## [13] devtools_2.4.3     usethis_2.1.6      data.table_1.14.2  janitor_2.1.0
## [17] forcats_0.5.1      stringr_1.4.0      dplyr_1.0.9        purrr_0.3.4
## [21] readr_2.1.2        tidyr_1.2.0        tibble_3.1.7       ggplot2_3.3.6
## [25] tidyverse_1.3.1
##
## loaded via a namespace (and not attached):
##  [1] colorspace_2.0-3    rjson_0.2.21        ellipsis_0.3.2
##  [4] rprojroot_2.0.3     snakecase_0.11.0    fs_1.5.2
##  [7] rstudioapi_0.13     farver_2.1.0        remotes_2.4.2
## [10] bit64_4.0.5         fansi_1.0.3         xml2_1.3.3
## [13] cachem_1.0.6        pkgload_1.2.4       spam_2.9-0
## [16] jsonlite_1.8.0      broom_0.8.0         dbplyr_2.2.0
## [19] png_0.1-7           compiler_4.2.0      httr_1.4.3
## [22] backports_1.4.1     assertthat_0.2.1    fastmap_1.1.0
## [25] cli_3.3.0           htmltools_0.5.2     prettyunits_1.1.1
## [28] tools_4.2.0         dotCall64_1.0-1     gtable_0.3.0
## [31] glue_1.6.2          maps_3.4.0          Rcpp_1.0.8.3
```

```
## [34] cellranger_1.1.0    vctrs_0.4.1         svglite_2.1.0
## [37] xfun_0.31           ps_1.7.0            brio_1.1.3
## [40] testthat_3.1.4      rvest_1.0.2         lifecycle_1.0.1
## [43] scales_1.2.0        hms_1.1.1           fields_14.0
## [46] yaml_2.3.5          memoise_2.0.1       gridExtra_2.3
## [49] stringi_1.7.6       highr_0.9           maptools_1.1-4
## [52] desc_1.4.1          pkgbuild_1.3.1      RgoogleMaps_1.4.5.3
## [55] rlang_1.0.2         pkgconfig_2.0.3     systemfonts_1.0.4
## [58] bitops_1.0-7        evaluate_0.15       lattice_0.20-45
## [61] labeling_0.4.2      bit_4.0.4           processx_3.5.3
## [64] tidyselect_1.1.2    plyr_1.8.7          magrittr_2.0.3
## [67] R6_2.5.1            generics_0.1.2      DBI_1.1.2
## [70] pillar_1.7.0        haven_2.5.0         foreign_0.8-82
## [73] withr_2.5.0         modelr_0.1.8        crayon_1.5.1
## [76] utf8_1.2.2          tzdb_0.3.0          rmarkdown_2.14
## [79] jpeg_0.1-9          grid_4.2.0          readxl_1.4.0
## [82] callr_3.7.0         reprex_2.0.1        digest_0.6.29
## [85] webshot_0.5.3       munsell_0.5.0       sessioninfo_1.2.2
```