R-Programming-Project

Data Analysis on Super Cars

Lovely School of Computer Science

Lovely Professional University, Phagwara

Under the Guidance of Dr. Amritpal Singh: 17673

Discipline of CSE/IT

Submitted by

Name: Viswaprasad S

(Project Semester January - May 2023)

DECLARATION

I, Viswaprasad S, student of Computer science engineering (Program name) under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: Viswaprasad S

Registration No. 12016097 Name of the student

CERTIFICATE

This is to certify that Viswaprasad S bearing Registration no. 12016097 has completed INT232 project titled, **"Data Analysis on Super Cars"** under my guidance and supervision. To the best of my knowledge, the present work is the result of his/her original development, effort, and study.

Signature and Name of the Supervisor
Designation of the Supervisor
School of Computer Science and engineering
Lovely Professional University
Phagwara, Punjab.

Date: 20.04.2023

Acknowledgement

I have made efforts in this project. However, it would not have been possible without the kind support and help of many individuals. I would like to extend my sincere thanks to all of them.

I am highly indebted to our teacher Dr. Amritpal Singh: 17673 for their guidance and constant supervision as well as for providing necessary information regarding the project and for their support in completing the project. His constant guidance and willingness to share his vast knowledge made us understand this project and its manifestations in great depth and helped us to complete the assigned tasks on time.

My thanks and appreciations also go to my mentors in developing the project and people who have willingly helped me out with their abilities.

Table of Contents

Introduction

This project was developed to analyze the Super car price and other aspects of super car, in this project I have analyzed the dataset and made some conclusion according to the data. This project will help us to keep track of increasing price and quality of super cars in the market, and this project will show the most important details that can be derived from the dataset.

Dashboard includes:

The dashboard contains three pages with different types of charts and conclusions according to those charts.

Page 1:

- Chart 1: Top 5 Super Cars w.r.t price($USD)
- Chart 2: No. of Supercars w.r.t Years
- Chart 3: Top 10 Supercars w.r.t Torque
- Chart 4: Horsepower vs Torque of Top 10 Supercars

Page 2:

- Chart 5: Count of Car models according to Car brands
- Chart 6: Count of Car models according to Top 5 Car brands vs other brands
- Chart 7: Top 5 Supercars who can go fast(0-60MPH_insec)
- Conclusion

Page 3:

- Summary of dataset and Details of author

Objectives / Scope of analysis

1. To analyze which top 5 cars are most expensive among all other supercars.
2. No. of Supercars launched per Years from 2019 - 2022.
3. To analyze which top 10 cars has highest torque among all other supercars.
4. To measure the top 10 cars, have the highest Horsepower and torque among all other supercars.
5. To measure the Count of Car models of Top 10 Car brands who launched in this span year of 2019 – 2022.
6. To analyze the percentage of Top 5 Car brand launches vs other car brands.
7. To analyze which car can go fast among all other cars - Top 5 Supercars who can go fast(0-60MPH_insec).

Source of dataset

Link:

https://www.kaggle.com/datasets/rkiattisak/sports-car-prices-dataset?resource=download

Reference:

https://www.kaggle.com

Dataset includes:

Number of columns: 8

Number of Rows: 178(After removing duplicates and Null/NA Values)

The Original Dataset had 1007 rows with repetition of entries.

Columns includes (important columns):

1. Car_Brand
2. Car_Model
3. Year
4. Engine_Size
5. Horsepower
6. Torque
7. 0-60mph (in sec)
8. Price_USD

Analysis on dataset (for each analysis)

Introduction:

This project is completely about Super car price and features analysis which has various Car brands data and Car models data and by analyzing that we can come to know that which is the costliest car and which car is fastest one. By knowing the count of the cars launched and value of the car we can predict and analyze the company's improvement and failure.

General Introduction of Dataset:

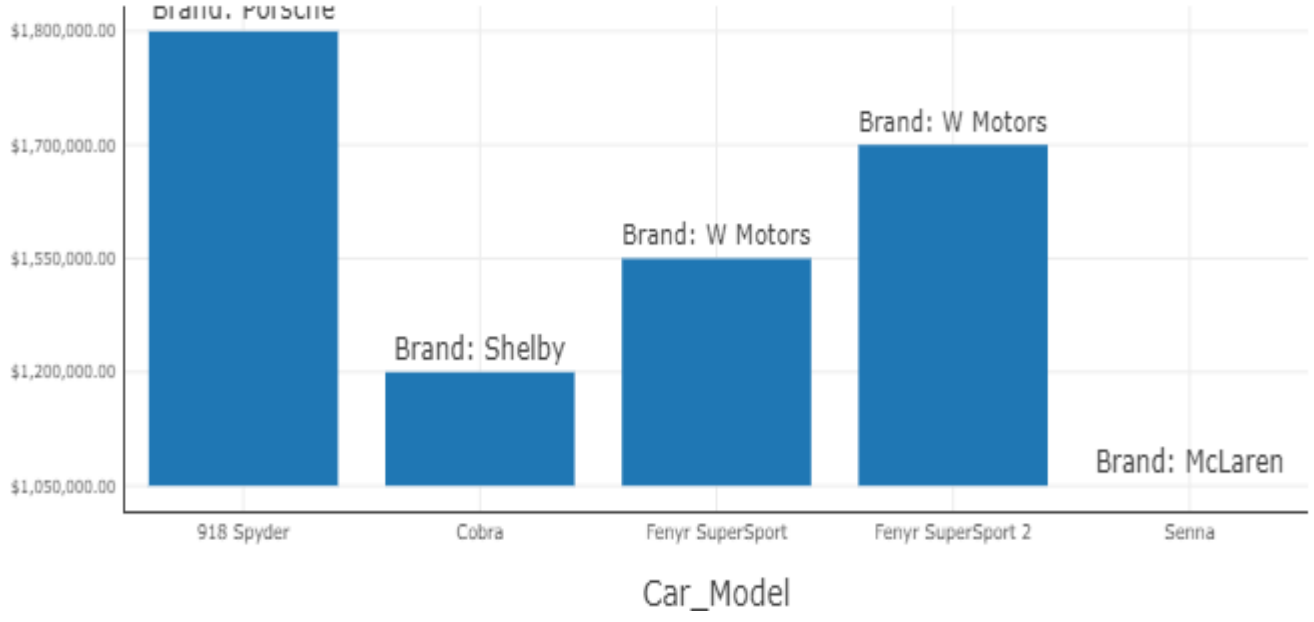
1. Car_Brand: This column consists of Brands of the car.
2. Car_Model: This column consists of Model of the Car Brand.
3. Year: The year in which car was launched.
4. Engine_Size: It has the size of the engine (ex: 4 cylinders, electric etc...).
5. Horsepower: This column consists of HP of the car model.
6. Torque: This column consists of Torque the car model.
7. 0-60mph (in sec): Time taken by the car to reach 0-60mph.
8. Price_USD: The price of the car.

Specific Requirements

1. R studio: The RStudio is an integrated development environment for R, a programming language for statistical computing and graphics.
2. library(ggplot2): ggplot2 is an open-source data visualization package for the statistical programming language R. ggplot2 is a plotting package that provides helpful commands to create complex plots from data in a data frame.
3. library(plotly): plotly is an R package for creating interactive web-based graphs via the open-source JavaScript graphing library plotly. js.
4. library(plyr): The plyr package is a set of clean and consistent tools that implement the split-apply-combine pattern in R. This is an extremely common pattern in data analysis: you solve a complex problem by breaking it down into small pieces, doing something to each piece and then combining the results back together again.
5. library(dplyr): Dplyr is a data manipulation package that is part of the tidy verse universe, a collection of libraries that has the goal of making R faster, simpler and easier. Other than the cool functions you can access by installing the package, Dplyr leverages the pipe (%>%) structure, a better way to encapsulate functions.
6. library(flexdashboard): R Markdown Format for Flexible Dashboards Format for converting an R Markdown document to a grid-oriented dashboard. The dashboard flexibly adapts the size of its components to the containing web page.
7. library(sqldf): sqldf() transparently sets up a database, imports the data frames into that database, performs the SQL select or other statement and returns the result using a heuristic to determine which class to assign to each column of the returned data frame.

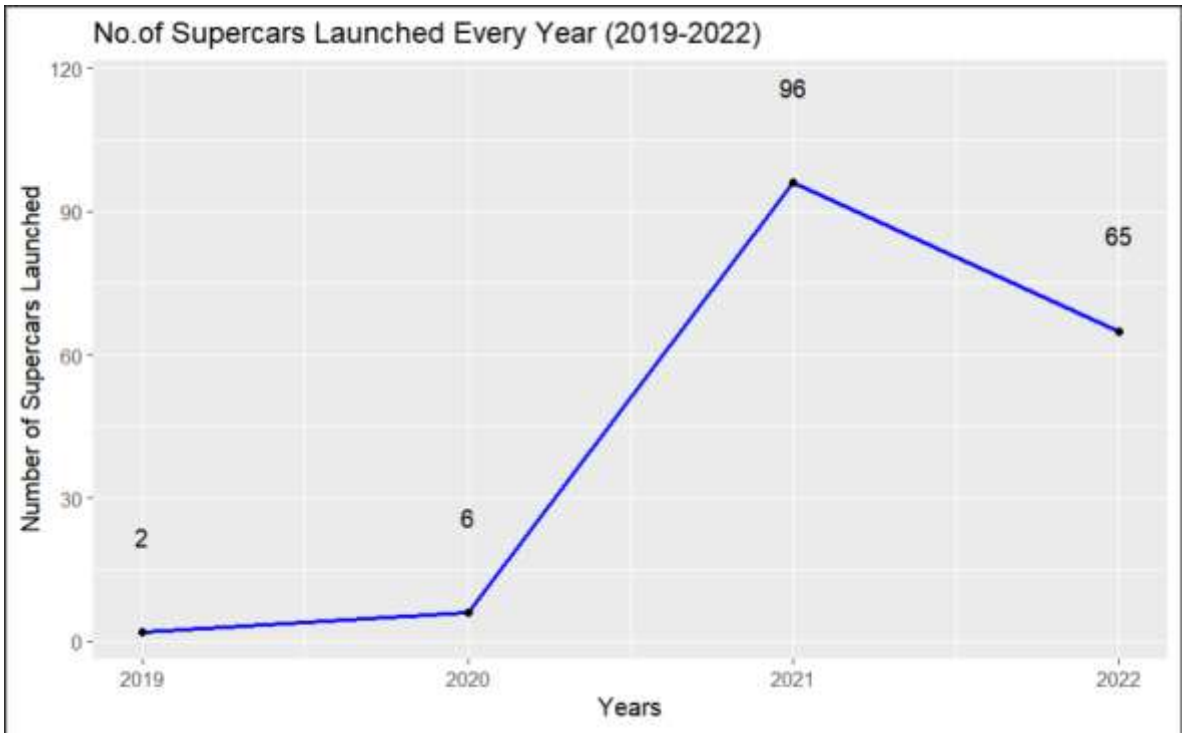
To analyze which top 5 cars are most expensive among all other supercars:

```
 dat <- read.csv("C:\\Users\\viswa\\OneDrive\\Desktop\\Sport car price.csv")
 dat
 s <- sqldf("Select [Car_Brand],[Car_Model],[Price_USD] from dat")
 price <- s%>% arrange(Price_USD)%>%slice(1:5)
plot_ly(
  data = price,
  x = ~Car_Model,
  y = ~Price_USD,
  type = "bar",
  text = ~paste0("Brand: ",(Car_Brand)),
  textposition = "outside"
) %>%

    layout(xaxis = data.frame(titel = "Model",showgrid = T, zeroline = T,
                   showline = T,range = c(0,4,000,000)),
           yaxis = data.frame(title = 'Price',
                   showgrid = T, zeroline = T,
                   showline = T,range = c(0,4,000,000)))%>%
layout(xaxis = list(tickfont = list(size = 8)),
         yaxis = list(tickfont = list(size = 7)))
```



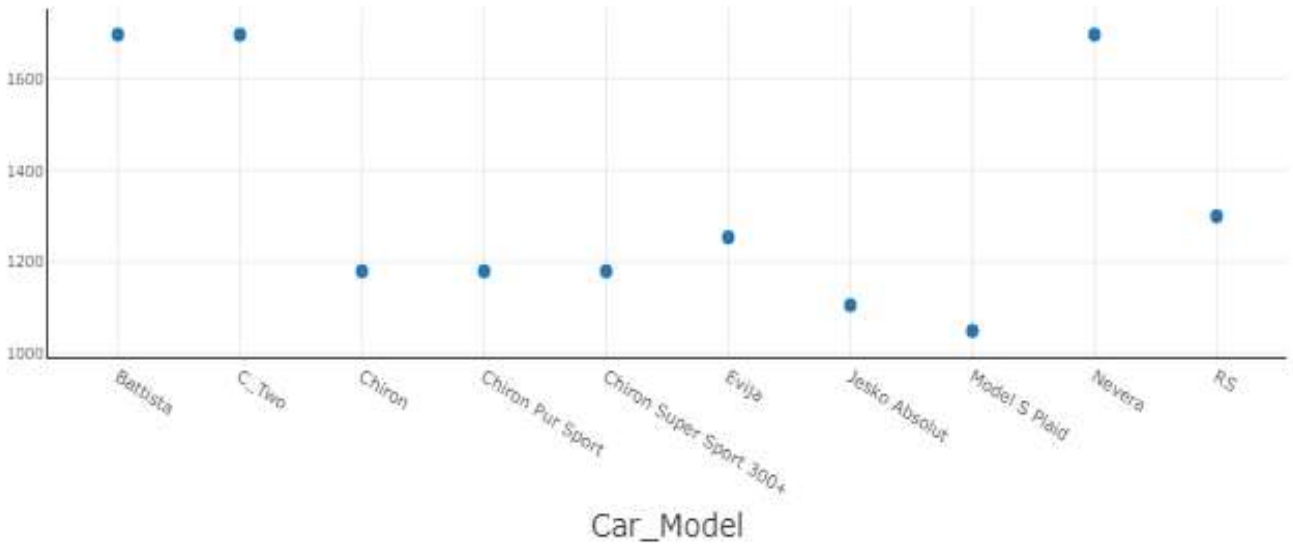
No. of Supercars launched per Years from 2019 – 2022:

```
Year <- sqldf("Select [Car_Brand],[Year] from dat where Year = 2019")
Year1 <- sqldf("Select [Car_Brand],[Year] from dat where Year = 2020")
Year2 <- sqldf("Select [Car_Brand],[Year] from dat where Year = 2021")
Year3 <- sqldf("Select [Car_Brand],[Year] from dat where Year = 2022")
merging <- sqldf("SELECT count([Car_Brand]),Year FROM Year UNION ALL SELECT count([Car_Brand]),Year FROM Year1 UNION ALL SELECT count([Car_Brand]),Year FROM Year2 UNION ALL SELECT count([Car_Brand]),Year FROM Year3")
df <- merging %>%
   rename( "Supercars" = "count([Car_Brand])")

line<-ggplot(data=df, aes(x=Year, y=Supercars, group=1,label = Supercars)) +
  geom_line(color="blue",size=1.0)+  geom_text(nudge_y = 20)+
  geom_point()

line+ggtitle("No.of Supercars Launched Every Year (2019-2022)")+
     labs(x="Years",y="Number of Supercars Launched")
```



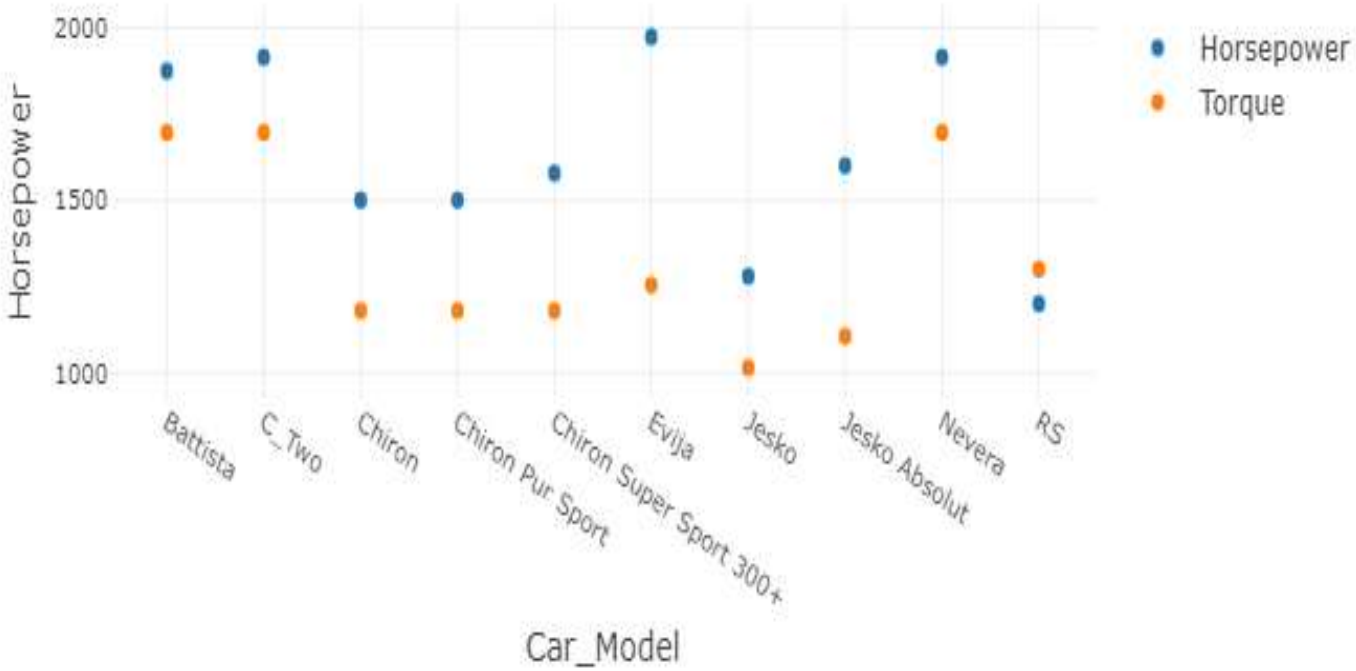
To analyze which top 10 cars has highest torque among all other supercars:

```
chart3 <- sqldf("Select [Car_Brand],[Car_Model],[Torque],[Horsepower] from dat")
torque <- chart3%>% arrange(desc(Torque)) %>% slice(1:10)
torque

plot_ly(
  data = torque,
  x = ~Car_Model,
  y = ~Torque,
  z = ~Horsepower,
  type = "scatter",
  text = ~paste0("Brand: ",(Car_Brand)),
  textposition = "outside"
) %>%

    layout(
        xaxis = data.frame(titel = "Model",showgrid = T, zeroline = T,
                    showline = T,tickfont = list(size = 2)),
        yaxis = data.frame(title = 'Torque',
                    showgrid = T, zeroline = T,
                    showline = T,range = c(0,4,000,000)))%>%
layout(xaxis = list(tickfont = list(size = 8)),
       yaxis = list(tickfont = list(size = 7)))
```



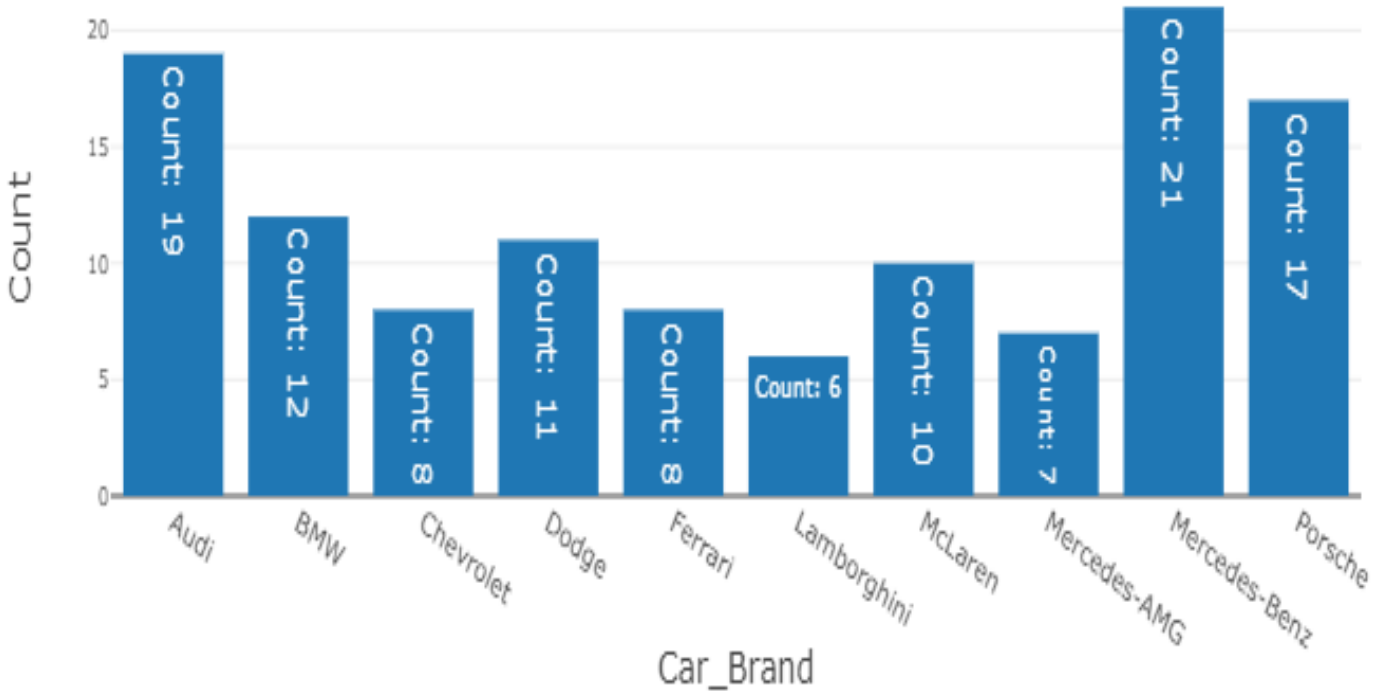
To measure the top 10 cars, have the highest Horsepower and torque among all other supercars:

```
hp <- chart3%>% arrange(desc(Horsepower)) %>% slice(1:10)
hp

plot_ly(
  data = hp,
  x = ~Car_Model,
  y = ~Horsepower,
  name = "Horsepower",
  text = ~paste0("Brand: ",(Car_Brand)),
  type = "scatter"
) %>%
add_trace(y = ~Torque, name = "Torque")%>%
  layout(xaxis = list(tickfont = list(size = 10)),
         yaxis = list(tickfont = list(size = 10)))
```



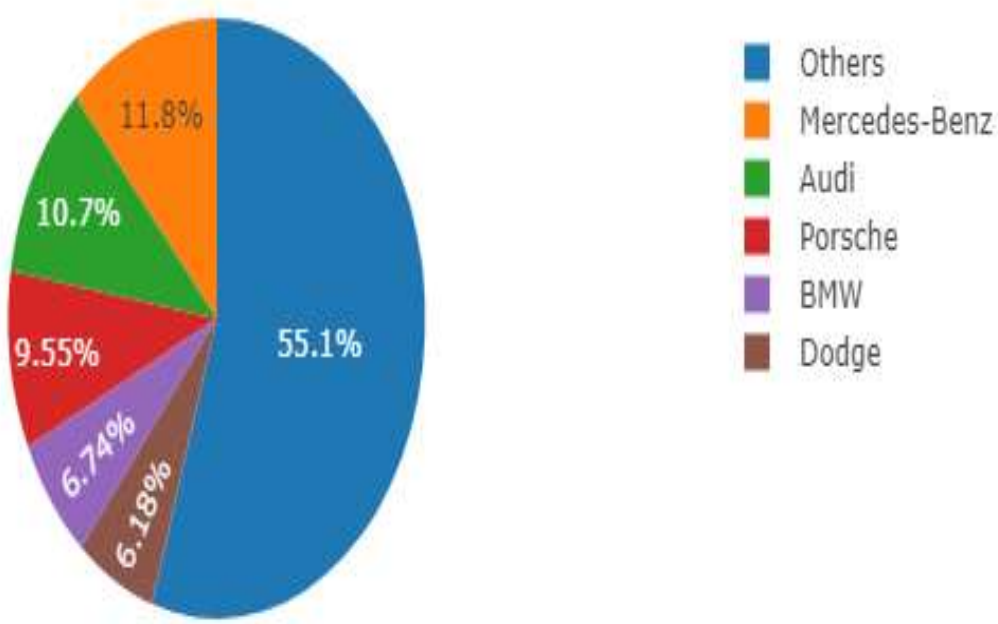
To measure the Count of Car models of Top 10 Car brands who launched in this span year of 2019 – 2022:

```
chart5<-dat%>%  count(Car_Brand, sort = T) %>% slice(1:10)
chart5

attach(chart5)
hist <- plot_ly(x = Car_Brand,y = n, type = "bar",  text = ~paste0("Count: ",(n)))
layout(hist,
xaxis = list(title = "Car_Brand"),
yaxis = list(title = "Count"))%>%
layout(xaxis = list(tickfont = list(size = 10)),
       yaxis = list(tickfont = list(size = 8)))
```



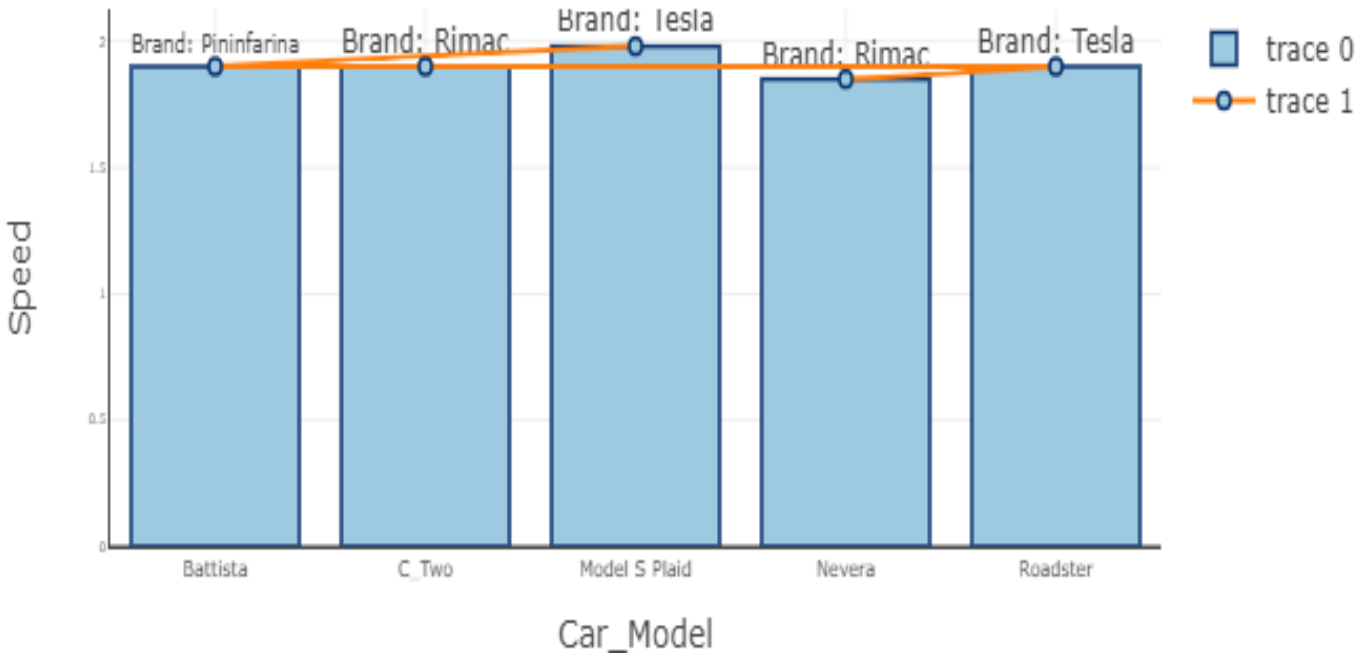
To analyze the percentage of top 5 Car brand launches vs other car brands:

```
chart7<-dat%>%  count(Car_Brand, sort = T) %>% slice(1:5)
chart7
chart72 <- dat %>% count(Car_Brand,sort = T) %>% slice(6:38) %>%
bind_rows(summarise(., across(where(is.numeric), sum),across(where(is.character),
~'Others')))
chart72
other <- sqldf("Select [Car_Brand],[n] from chart72 where n = 98")
other
merging2 <- sqldf("SELECT [Car_Brand],n FROM chart7 UNION ALL SELECT
[Car_Brand],n FROM other")
merging2

fig <- plot_ly(merging2, labels = ~Car_Brand, values = ~n, type = 'pie')
fig <- fig %>% layout(
       xaxis = list(showgrid = FALSE, zeroline = FALSE, showticklabels = FALSE),
       yaxis = list(showgrid = FALSE, zeroline = FALSE, showticklabels = FALSE))

fig
```



To analyze which car can go fast among all other cars - Top 5 Supercars who can go fast(0-60MPH_insec):

```
plot_ly(
  data = speed,
  x = ~Car_Model,
  y = ~X0.60MPH_insec,
  type = "bar",
  text = ~paste0("Brand: ",(Car_Brand)),
  textposition = "outside",marker = list(color = 'rgb(158,202,225)',
  line = list(color = 'rgb(8,48,107)', width = 1.5))
) %>%add_trace(data , x = ~Car_Model, y = ~X0.60MPH_insec, type = 'scatter',  mode = 'lines') %>% layout(xaxis = data.frame(titel = "Model",showgrid = T, zeroline = T,
                    showline = T),
         yaxis = data.frame(title = 'Speed',
                    showgrid = T, zeroline = T,
                    showline = T))%>%
layout(xaxis = list(tickfont = list(size = 8)),
       yaxis = list(tickfont = list(size = 5)))
```



Analysis Result

1. Rimac Nevera is fastest among all other super cars who can reach 0-60mph in 1.85 seconds.

2. Porche 918 Spyper is most expensive car among all other super cars with the price tag of $1,800,000 USD.

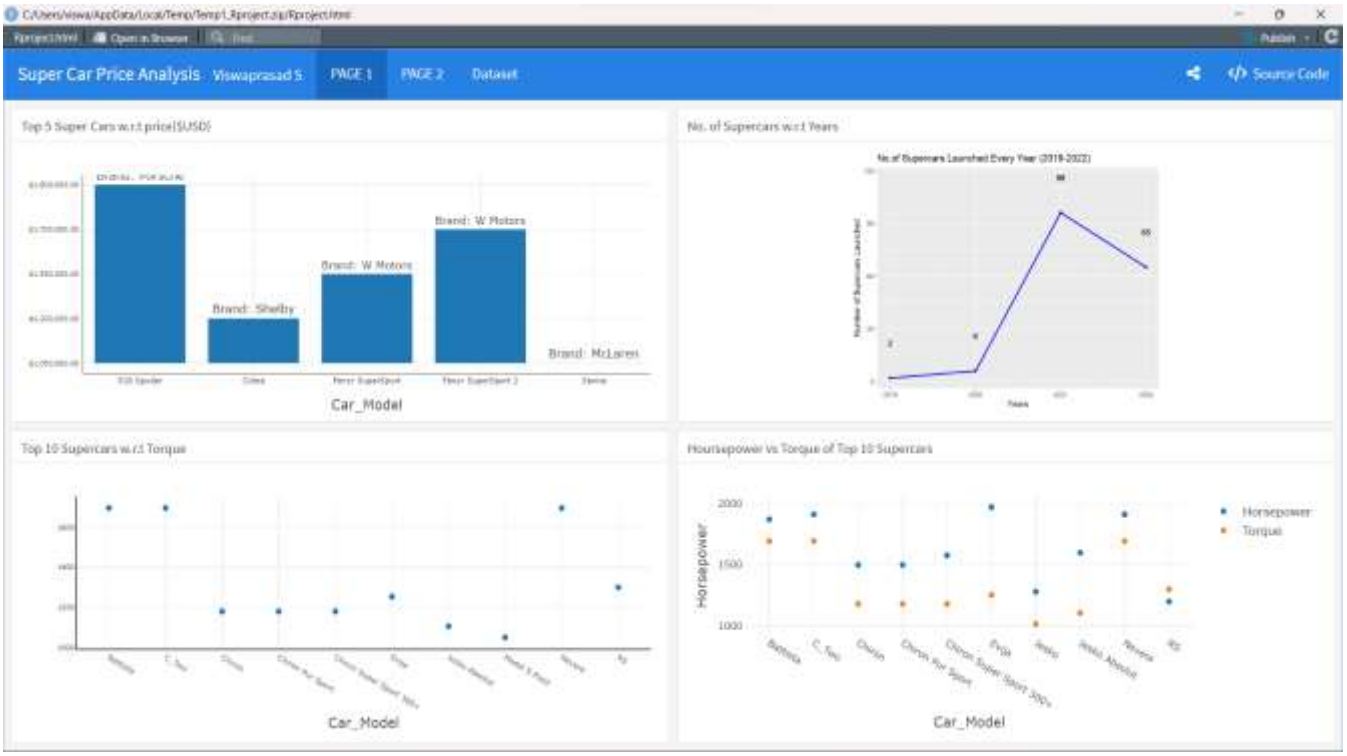
3. Lotus evija has highest HP of 1973 and Pininfarina Battista with 2 other cars has torque of 1696 among all other super cars.

4. Mercedes-Benz launched the greatest number of supercars (21) in the recent 5 Years.

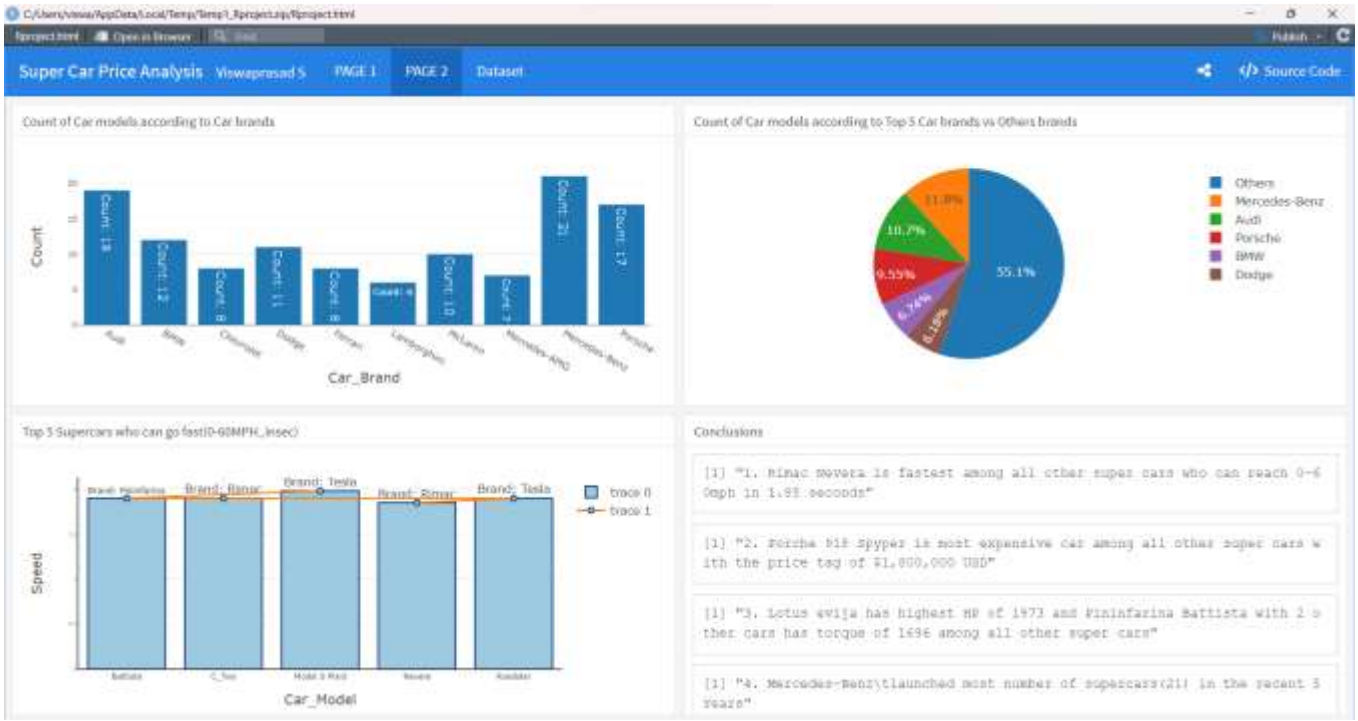
The above mentioned are only some of the conclusions. By the visualization shown above we can get many conclusions to get understanding of Super Cars Data.

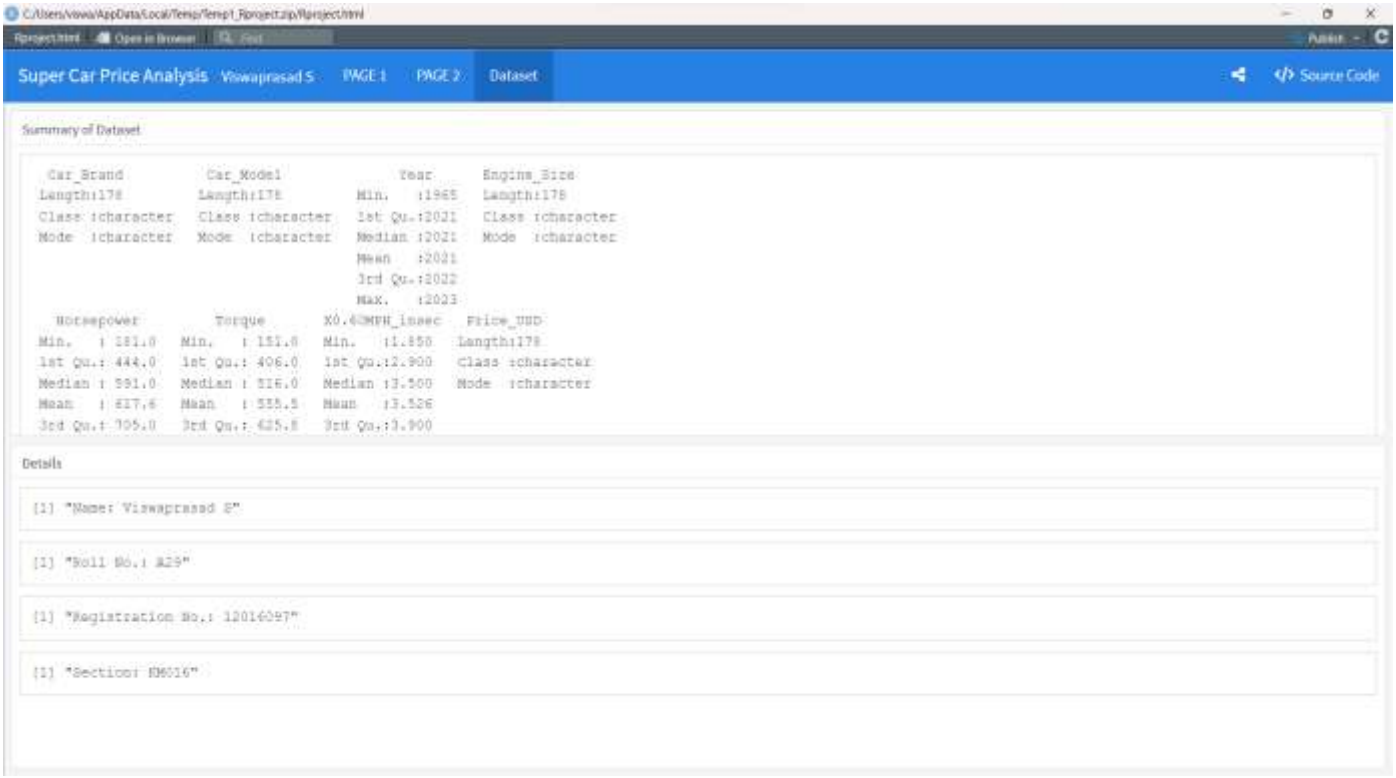
The Dashboard

Page 1



Page 2





Future Scope

Analysis like this can be essential to develop the business and the future scope of data analysis is vast and promising. With the explosion of data in recent years, there is a growing need for individuals and organizations to extract insights from the data they collect.

The data analysis can help the car companies to know about their peer competition and they can improve the organization through overcoming the need and knowledge gained by their peer competition.

Reference

Dataset:

https://www.kaggle.com/datasets/rkiattisak/sports-car-prices-dataset?resource=download

Other than the dataset no reference where took by me.

To gain graph knowledge I referred:

https://community.plotly.com/

And took some other websites to gain knowledge about libraries used.

Thank You