

# **Analysing the Impact of Car Features on Price and Profitability**

---

**By Anjali Gupta**

# Project Description

---

The automotive industry has rapidly evolved over the past few decades, with a growing focus on fuel efficiency, environmental sustainability, and technological innovation. With increasing competition among manufacturers and a changing consumer landscape, it has become more critical to understand the factors driving consumer demand for cars.

For the given dataset, the client has asked How can a car manufacturer optimise pricing and product development decisions to maximise profitability while meeting consumer demand?

Car\_Data provided has been used for the analysis.

For the given dataset, the duplicates were removed during the data cleaning stage.



# Approach

---

Various analytical techniques are used in this project which includes machine learning(linear regression),visualization, optimization and statistics

Modelling techniques such as segmentation is also used in this project.





# Tech-Stack Used

---

- Microsoft Excel – It is used for performing analysis on dataset and even for dashboard creation.
- Microsoft Power Point Presentation – It is used for providing a report of the given task



**Insight Required:** How does the popularity of a car model vary across different market categories?

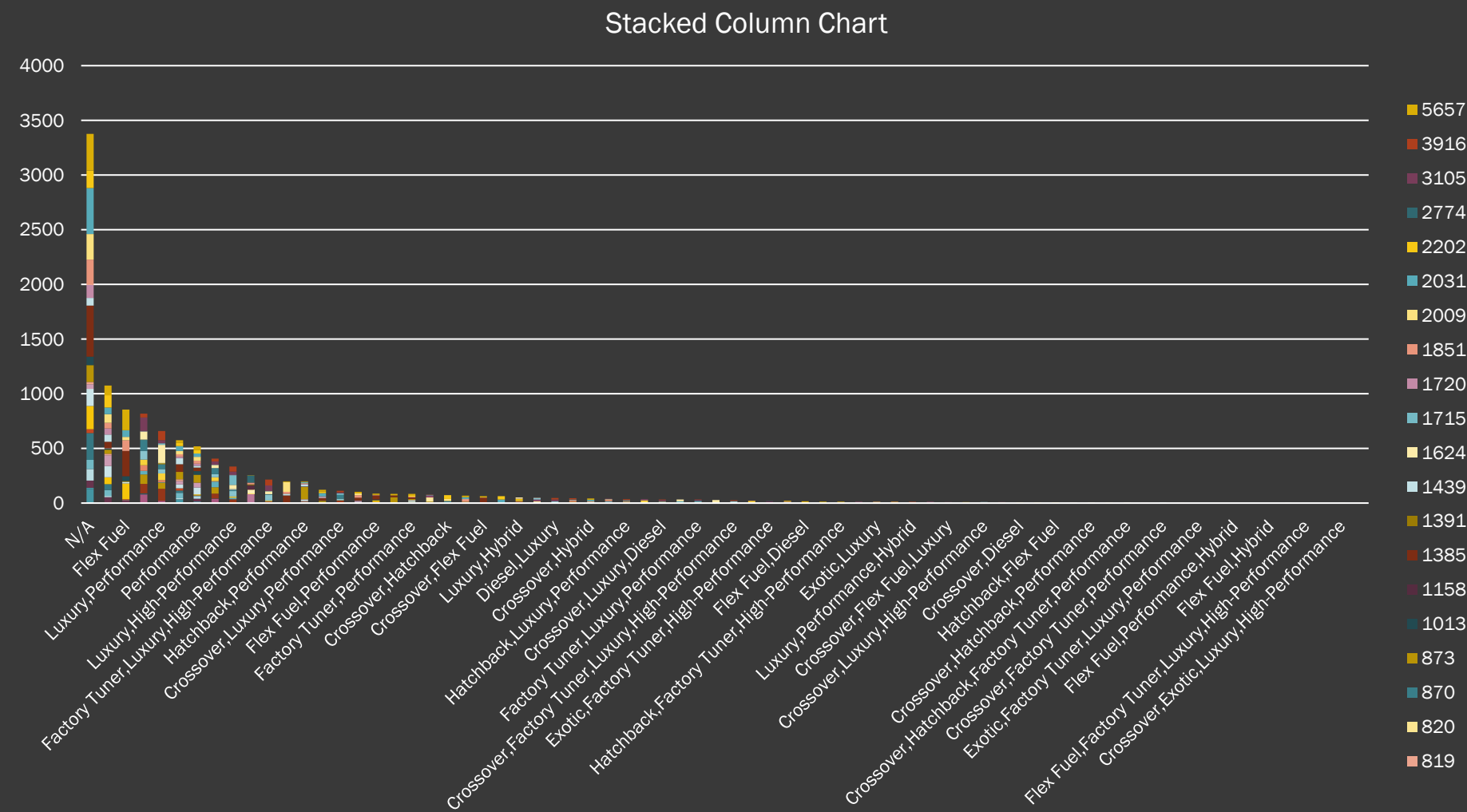
# Task 1.A:

Create a pivot table that shows the number of car models in each market category and their corresponding popularity scores.

Row Labels	Count of Model	Average of Popularity
Flex Fuel,Diesel	16	5657
Hatchback,Flex Fuel	7	5657
Crossover,Flex Fuel,Performance	6	5657
Crossover,Luxury,Performance,Hybrid	2	3916
Crossover,Factory Tuner,Luxury,Performance	5	2607.4
Crossover,Performance	69	2585.956522
Crossover,Hybrid	42	2563.380952
Diesel,Luxury	47	2416.106383
Luxury,Performance,Hybrid	11	2333.181818
Flex Fuel	855	2225.71345
Hatchback,Factory Tuner,Performance	21	2173.714286
Crossover,Luxury,Diesel	34	2149.411765
Factory Tuner,Luxury,High-Performance	215	2133.367442
Hybrid	121	2116.586777
Hatchback,Hybrid	64	2111.15625
Crossover,Flex Fuel	64	2073.75
Crossover,Hatchback,Factory Tuner,Performance	6	2009
Crossover,Hatchback,Performance	6	2009
Factory Tuner,High-Performance	104	1966.442308
Crossover,Factory Tuner,Luxury,High-Performance	26	1823.461538
High-Performance	198	1823.378788
Factory Tuner,Performance	84	1774.047619
Diesel	84	1730.904762
Flex Fuel,Performance	87	1680.471264
Crossover,Hatchback	72	1675.694444
Luxury,High-Performance	334	1668.017964
N/A	3376	1664.832938
Hatchback,Luxury,Performance	36	1632.25
Crossover,Flex Fuel,Luxury,Performance	6	1624
Crossover	1075	1556.168372
Performance	520	1415.209615
Factory Tuner,Luxury,Performance	31	1413.419355
Exotic,Performance	10	1391
Flex Fuel,Luxury,Performance	28	1380.071429
Crossover,Luxury,Performance	112	1349.089286
Hatchback,Luxury	45	1323.133333
Hatchback	574	1308.65331

Luxury,Performance	659	1293.062215
Exotic,High-Performance	254	1280.047244
Hatchback,Factory Tuner,High-Performance	13	1205.153846
Crossover,Flex Fuel,Luxury	10	1173.2
Luxury	819	1079.214896
Hatchback,Performance	198	1073.661616
Exotic,Factory Tuner,High-Performance	21	1046.380952
Crossover,Luxury,High-Performance	9	1037.222222
Flex Fuel,Luxury,High-Performance	32	898.3125
Crossover,Luxury	406	889.2142857
Hatchback,Factory Tuner,Luxury,Performance	9	886.8888889
Crossover,Diesel	7	873
Hatchback,Diesel	14	873
Flex Fuel,Luxury	39	746.5384615
Luxury,Hybrid	52	673.6346154
Crossover,Luxury,Hybrid	24	630.9166667
Factory Tuner,Luxury	2	617
Luxury,High-Performance,Hybrid	12	568.8333333
Exotic,Factory Tuner,Luxury,High-Performance	51	523.0196078
Exotic,Flex Fuel,Luxury,High-Performance	11	520
Exotic,Factory Tuner,Luxury,Performance	3	520
Exotic,Flex Fuel,Factory Tuner,Luxury,High-Performance	13	520
Exotic,Luxury,High-Performance	77	473.025974
Hatchback,Luxury,Hybrid	3	454
Flex Fuel,Factory Tuner,Luxury,High-Performance	1	258
Crossover,Exotic,Luxury,High-Performance	1	238
Crossover,Exotic,Luxury,Performance	1	238
Exotic,Luxury,Performance	36	217.0277778
Crossover,Factory Tuner,Performance	4	210
Crossover,Hatchback,Luxury	7	204
Exotic,Luxury,High-Performance,Hybrid	1	204
Performance,Hybrid	1	155
Flex Fuel,Performance,Hybrid	2	155
Flex Fuel,Hybrid	2	155
Exotic,Luxury	12	112.6666667
<b>Grand Total</b>	<b>11199</b>	<b>1558.483347</b>

**Task 1.B:** Create a stacked column chart that visualises the relationship between market category and popularity.





“

The following insights were found from the pivot table:

The most famous market category are:

Flex Fuel, Diesel has 855 models, and average popularity of 5657.

Hatchback flex fuel has 7 models and an average popularity of 5657.

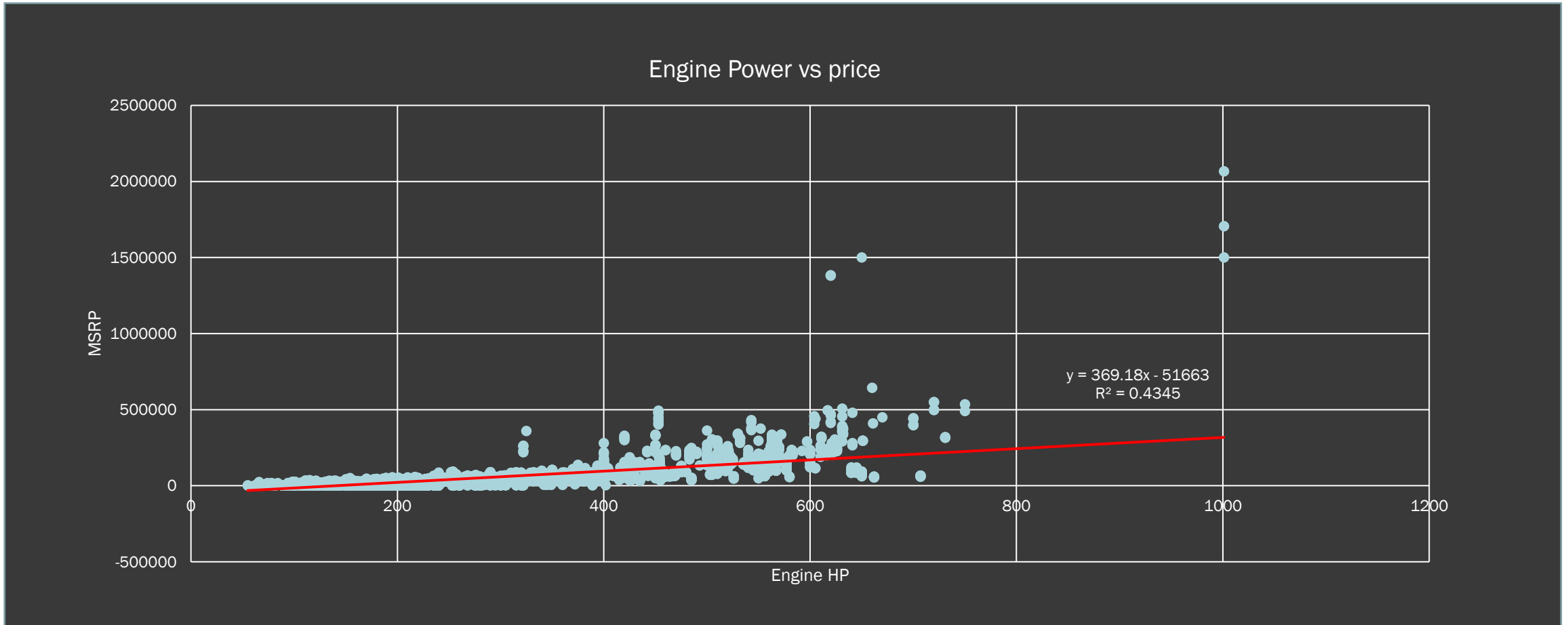
Crossover flex fuel, a performance which has 6 models and an average popularity 5657

The least popular market category is Exotic luxury, with 12 models but an average popularity of just 112.6667



**Insight Required:** What is the relationship between a car's engine power and its price?

**Task 2:** Create a scatter chart that plots engine power on the x-axis and price on the y-axis. Add a trendline to the chart to visualize the relationship between these variables.





The following insights were found from the scatter plot:  
The following plot has an r-squared value of 0.4345  
This means the variable has a moderate effect on each other  
Therefore we can say that they are not strongly related  
The trendline equation is, however given as  
 $y = 369.18x - 51663$

**Insight Required:** Which car features are most important in determining a car's price?

### Task 3:

Use regression analysis to identify the variables that have the strongest relationship with a car's price. Then create a bar chart that shows the coefficient values for each variable to visualize their relative importance.

Engine HP

SUMMARY OUTPUT

Regression Statistics

Multiple R

R Square

Adjusted R Square

Standard Error

Observations

ANOVA

df

SS

MS

F

Significance F

Regression

Residual

Total

Coefficients

Standard Error

t Stat

P-value

over 35

lower 35

lower 35

Upper 35

Upper 35

0%

Intercept

335

372.6487

4.06446879

91.68448174

0

-54448

-50045

-54448

-50045

23

364.68

380.62

364.68

380.61581

highway MPG

SUMMARY OUTPUT

Regression Statistics

Multiple R

R Square

Adjusted R Square

Standard Error

Observations

ANOVA

df

SS

MS

F

Significance F

Regression

Residual

Total

Coefficients

Standard Error

t Stat

P-value

over 35

lower 35

lower 35

Upper 35

Upper 35

0%

Intercept

26

-1162.79

64.60817007

-17.997528

2.2E-71

-1289

-1036

-1289

-1036

144

Engine Cylinders

SUMMARY OUTPUT

Regression Statistics

Multiple R

R Square

Adjusted R Square

Standard Error

Observations

ANOVA

df

SS

MS

F

Significance F

Regression

Residual

Total

Coefficients

Standard Error

t Stat

P-value

over 35

lower 35

lower 35

Upper 35

Upper 35

0%

Intercept

6

18694.198

276.988591

67.49085926

0

-66653

-60214

-66653

-60214

18151

19237

18151

19237

city mpg

SUMMARY OUTPUT

Regression Statistics

Multiple R

R Square

Adjusted R Square

Standard Error

Observations

ANOVA

df

SS

MS

F

Significance F

Regression

Residual

Total

Coefficients

Standard Error

t Stat

P-value

over 35

lower 35

lower 35

Upper 35

Upper 35

0%

Intercept

19

-1099.78143

63.08743583

-17.433

4E-67

-1223

-976.1

-1223

-976.1

Number of Doors

SUMMARY OUTPUT

Regression Statistics

Multiple R

R Square

Adjusted R Square

Standard Error

Observations

ANOVA

df

SS

MS

F

Significance F

Regression

Residual

Total

Coefficients

Standard Error

t Stat

P-value

over 35

lower 35

lower 35

Upper 35

Upper 35

0%

Intercept

2

-10537

668.94

-15.75

3E-55

-11848

-9226

-11848

-9226

Popularity

SUMMARY OUTPUT

Regression Statistics

Multiple R

R Square

Adjusted R Square

Standard Error

Observations

ANOVA

df

SS

MS

F

Significance F

Regression

Residual

Total

Coefficients

Standard Error

t Stat

P-value

over 35

lower 35

lower 35

Upper 35

Upper 35

0%

Intercept

3916

-2.102

0.4075

-5.153

3E-07

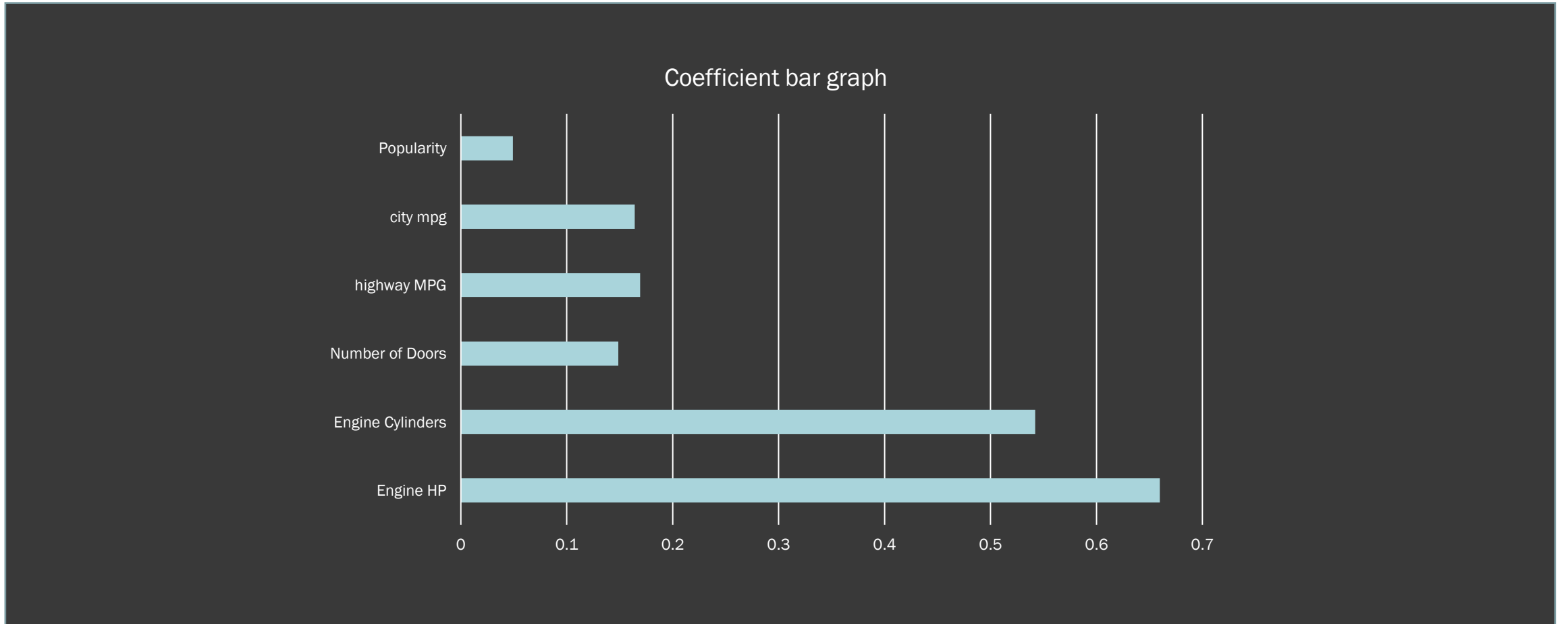
-2.901

-1.303

-2.901

-1.303







“

From the bar chart it can be clearly seen that Engine HP has the highest coefficient value followed by Engine cylinder followed by highway MPG which means that MSRP has the strongest relationship with Engine HP followed by Engine cylinder

**Insight Required:** How does the average price of a car vary across different manufacturers?

# Task 4.A:

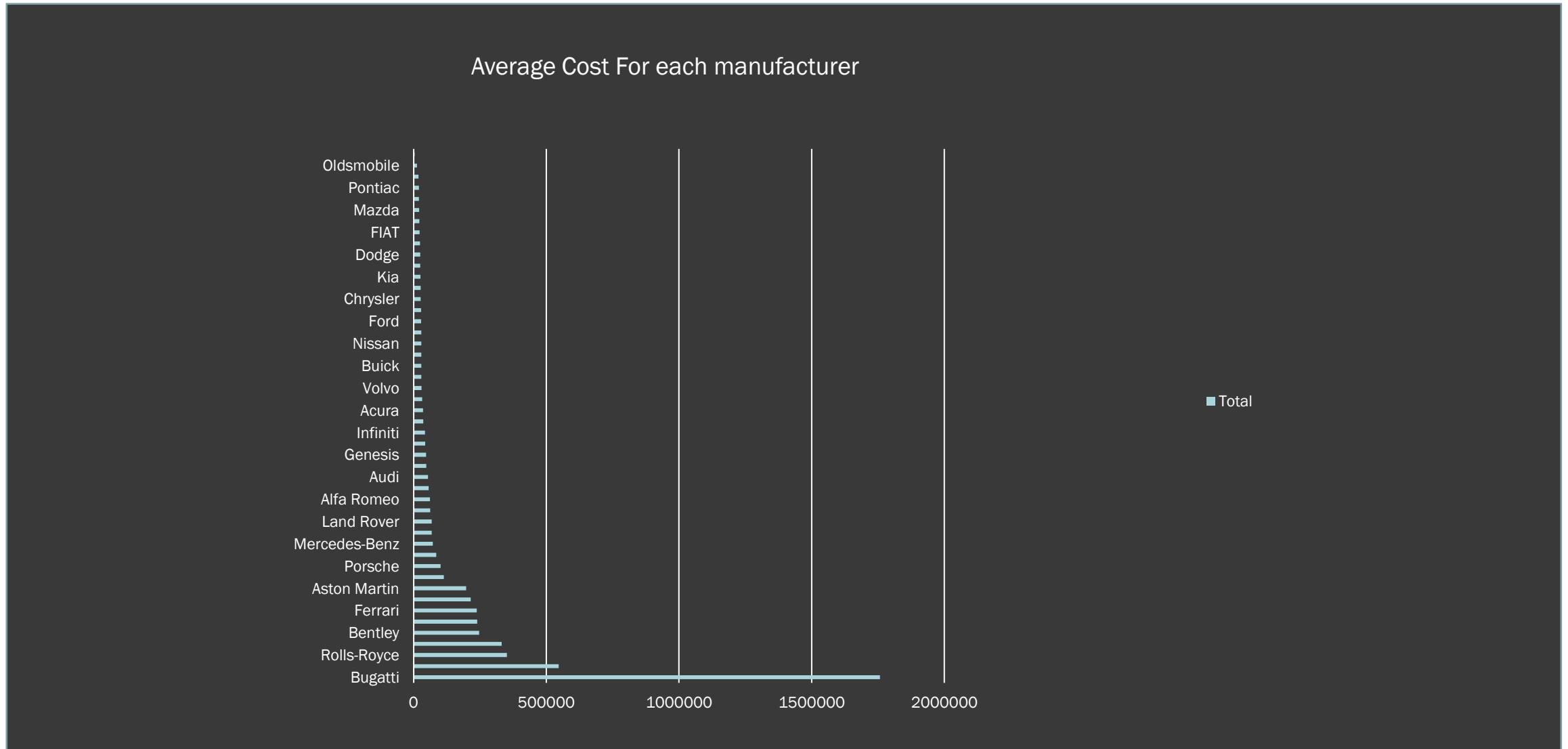
Create a pivot table that shows the average price of cars for each manufacturer.

Row Labels	Average of MSRP
Bugatti	1757223.667
Maybach	546221.875
Rolls-Royce	351130.6452
Lamborghini	331567.3077
Bentley	247169.3243
McLaren	239805
Ferrari	238218.8406
Spyker	214990
Aston Martin	198123.4615
Maserati	113684.4909
Porsche	101622.3971
Tesla	85255.55556
Mercedes-Benz	72069.52786
Lotus	68377.14286
Land Rover	68067.08633
BMW	62162.55864
Alfa Romeo	61600
Cadillac	56368.26515
Audi	54574.1215
Lexus	47549.06931
Genesis	46616.66667
Lincoln	43860.825
Infiniti	42640.27134

HUMMER	36464.41176
Acura	35087.4878
GMC	32444.08506
Volvo	29724.68421
Chevrolet	29074.72576
Buick	29034.18947
Volkswagen	28978.52289
Nissan	28921.15245
Toyota	28846.5605
Ford	28511.30788
Saab	27879.80734
Chrysler	26722.96257
Honda	26655.14781
Kia	25513.75546
Hyundai	24926.26255
Dodge	24857.04537
Subaru	24240.67364
FIAT	22670.24194
Mitsubishi	21340.5625
Mazda	20416.62379
Scion	19932.5
Pontiac	19800.0442
Suzuki	18026.4152
Oldsmobile	12843.79545
Plymouth	3296.873239
<b>Grand Total</b>	<b>41925.92714</b>

# Task 4.B:

Create a bar chart or a horizontal stacked bar chart that visualizes the relationship between manufacturer and average price.





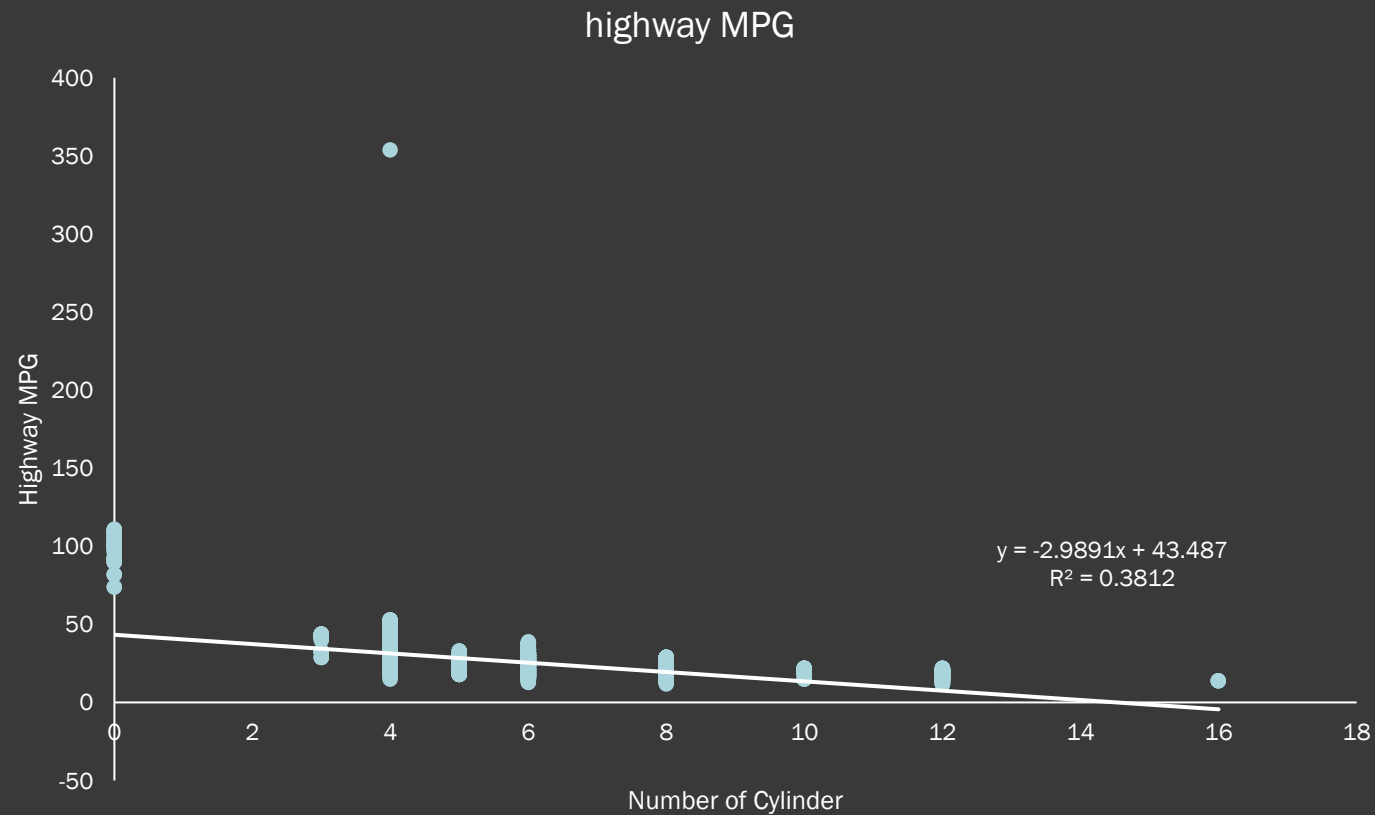


“

The following insights were found from the bar chart:  
The average price of the manufacturer Bugatti is maximum.  
followed by Maybach and Rolls-Royce  
For the given dataset, Plymouth has the lowest average  
price for car.

**Insight Required:** What is the relationship between fuel efficiency and the number of cylinders in a car's engine?

**Task 5.A:** Create a scatter plot with the number of cylinders on the x-axis and highway MPG on the y-axis. Then create a trendline on the scatter plot to visually estimate the slope of the relationship and assess its significance.





The following insights were found from the scatter plot:  
The trendline equation as can be observed from the graph  
is  $y = -2.9891x + 43.487$   
which gives slope,  $m = -2.9891$   
It means that for each one-unit increase in the number of  
cylinder, highway mpg decreases by 2.9891 unit from  
average

**Task 5.B:** Calculate the correlation coefficient between the number of cylinders and highway MPG to quantify the strength and direction of the relationship.

	Engine Cylinder	Highway MPG
Engine Cylinder	1	
Highway MPG	-0.617419591	1





“

It can be seen from the correlation matrix that the correlation coefficient is negative therefore, we can say that the relationship between the number of cylinders and highway mpg is such that when one increases, the other decreases since the value of correlation is 0.617, therefore we can say that it has a strong negative correlation

# Result

---

This project provided us with significant insights into how car features impact the price of the car

All the questions asked are answered in this ppt

For the regression analysis only those rows were included that didn't have blank value as data analysis add in required the data with no blank columns

As regression analysis can only be applied on numerical data so it is applied on numerical data analysis

For future analysis classification analysis can also be applied





“

The link of the analysis folder is attached here:

[https://drive.google.com/drive/folders/1bZsg\\_zS7\\_U0qrsK0QjDyopmp0oT2tD25?usp=share\\_link](https://drive.google.com/drive/folders/1bZsg_zS7_U0qrsK0QjDyopmp0oT2tD25?usp=share_link)

Direct Link to Excel file is attached here:

[Car\\_data \(1\).xlsx](#)



**Thank you**

---

