Project - 7

Analyzing the Impact of Car Features on Price and Profitability

Final Project-3

Project description:

This project focuses on helping a car manufacturer optimize pricing and product development strategies to maximize profitability while meeting evolving consumer demands. With the increasing popularity of electric and hybrid vehicles, along with ongoing demand for traditional gasoline-powered cars, it has become essential to understand which car features and market categories drive consumer preferences and profitability.

The primary business question we aim to address is: *How can the manufacturer balance pricing and product development to ensure profitability while satisfying consumer demand?*

For this analysis, we are using a dataset that includes information on car features, market categories, fuel types, and pricing. Data cleaning and preprocessing involved handling missing values, standardizing categories, and ensuring data consistency across variables. Assumptions made during the project include that the dataset reflects the current market landscape and that factors like car features and market category significantly influence pricing decisions.

By applying data analysis techniques such as regression and market segmentation, this project will identify trends that can guide future pricing strategies and product development, ultimately enhancing the manufacturer's competitiveness and profitability.

Approach:

- Cleaned and preprocessed the dataset using Excel functions to handle missing data and standardize variables.
- Applied descriptive statistics, pivot tables, scatter plots, regression analysis, and interactive dashboards to extract insights on pricing, features, and market categories for optimizing profitability.
- Descriptive statistics and visualizations provide a straightforward way to explore the data and reveal initial insights. Excel's built-in functions and charting tools make it easy to summarize and interpret key patterns.

Tech-Stack Used:

- For this project, **Microsoft Excel 2019** was the primary tool used for data analysis, visualization, and modeling. Excel was chosen for its user-friendly interface, built-in functions, and versatility in handling various types of data analysis tasks.

Tasks: Analysis:

Task 1: Popularity of Car Models Across Market Categories

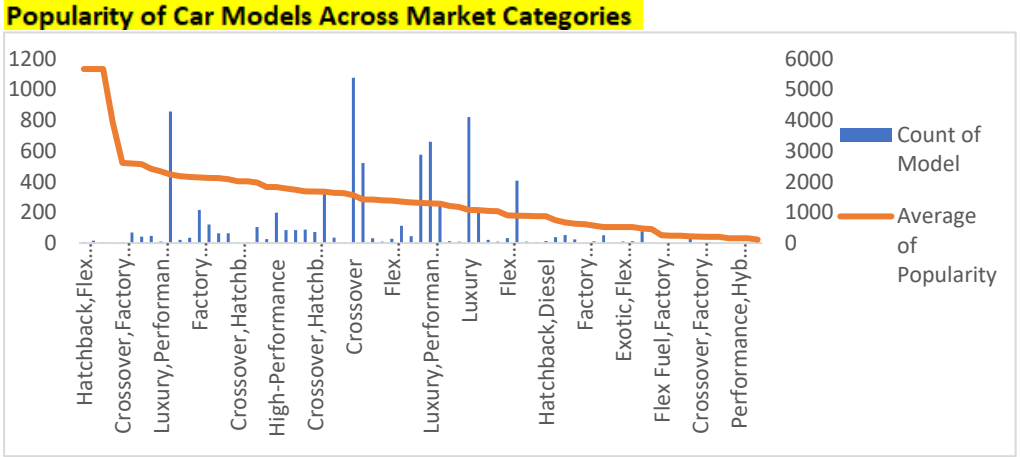
- **Task 1.A**: Create a pivot table that shows the number of car models in each market category and their corresponding popularity scores.

Output:

Row Labels	Count of Model	Average of Popularity
Hatchback,Flex Fuel	7	5657
Flex Fuel,Diesel	16	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
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,Factory Tuner,Luxury,Performance	3	520

The pivot table shows the count of models in each category and their average popularity.

- **Task 1.B:** Create a combo chart that visualizes the relationship between market category and popularity.



Insights:

1. Consumer Preference by Category:

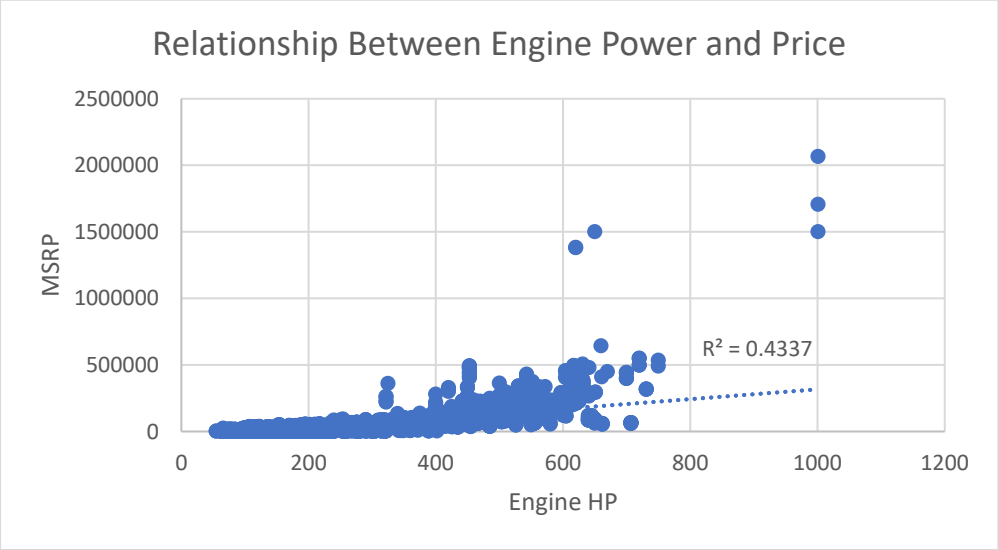
- Higher average popularity in certain market categories like hatchback,flexfuel or flexfuel,diesel indicate that consumers tend to prefer car models within these segments, reflecting strong demand.
- Lower average popularity in other categories flexfuel,hybrid or exotic,luxury suggest lower consumer interest.

2. Market Segments with High Potential: Categories with a consistently high average popularity across models could be identified as high-potential segments for manufacturers to focus on for future product development and marketing.

Task 2: Relationship Between Engine Power and Price

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.

Output:



Insights:

An R^2 value of 0.4337 suggests a moderate positive correlation between engine power and price.

This means that about 43.37% of the variation in car prices can be explained by differences in engine power.

While there is a relationship but not very strong, indicating that other factors also significantly influence the price.

Task 3: Car Features Impacting Price

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.

Output:

For this we take numerical columns and some categorical columns like fuel type, transmission type but they need to be assessed as they are also important car features so we assign them numbers to make them numerical column for further analysis.

Engine Fuel Type	
assign	**as**
electric	0
premium unleaded(required)	1
premium unleaded (recommended)	2
regular unleaded	3
flex-fuel (unleaded/natural gas)	4
flex-fuel (unleaded/E85)	5
flex-fuel (premium unleaded required/E85)	6
flex-fuel (premium unleaded recommended/E85)	7
diesel	8
natural gas	9

transmission type	
assign	**as**
AUTOMATED_MANUAL	0
AUTOMATIC	1
DIRECT_DRIVE	2
MANUAL	3
UNKNOWN	4

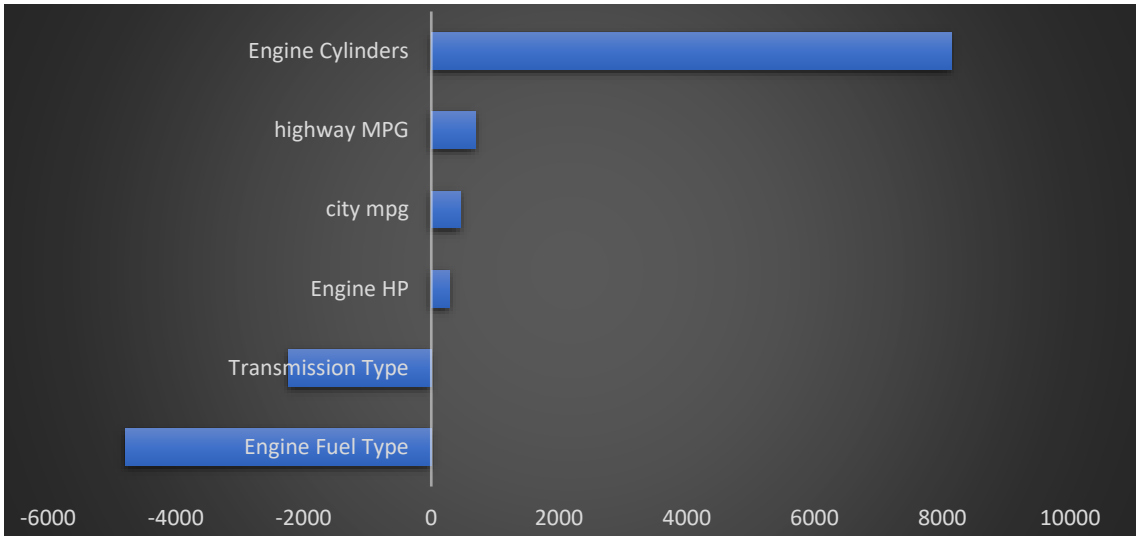
Regression analysis

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.679358454
R Square	0.461527909
Adjusted R Square	0.461239236
Standard Error	45166.92479
Observations	11199

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-85955.67802	3494.163642	-24.5998	3.5048E-130	-92804.85362	-79106.50242	-92804.85362	-79106.50242
Engine Fuel Type	-4780.786894	375.2990523	-12.7386	6.5068E-37	-5516.439078	-4045.134711	-5516.439078	-4045.134711
Engine HP	279.532205	6.814786898	41.01848	0	266.1740235	292.8903865	266.1740235	292.8903865
Engine Cylinders	8139.33074	457.1028186	17.80635	5.79805E-70	7243.32878	9035.3327	7243.32878	9035.3327
Transmission Type	-2230.969668	488.1143522	-4.57059	4.91566E-06	-3187.759691	-1274.179645	-3187.759691	-1274.179645
highway MPG	689.6090577	106.9158254	6.450019	1.16447E-10	480.0352261	899.1828893	480.0352261	899.1828893
city mpg	461.5416294	101.5428298	4.54529	5.54347E-06	262.4998146	660.5834441	262.4998146	660.5834441

Coefficient vs variables



Insights:

The positive coefficient of variables like engine cylinders, highway MPG, City MPG, engine HP shows that these features increase the price while the negative ones tend to decrease the price.

Bar chart also shows the Variables with larger coefficients (both positive and negative) have a greater impact on the price.

Positive coefficients indicate that the variable increases the price, while negative coefficients indicate a decrease.

Task 4: Average Price by Manufacturer

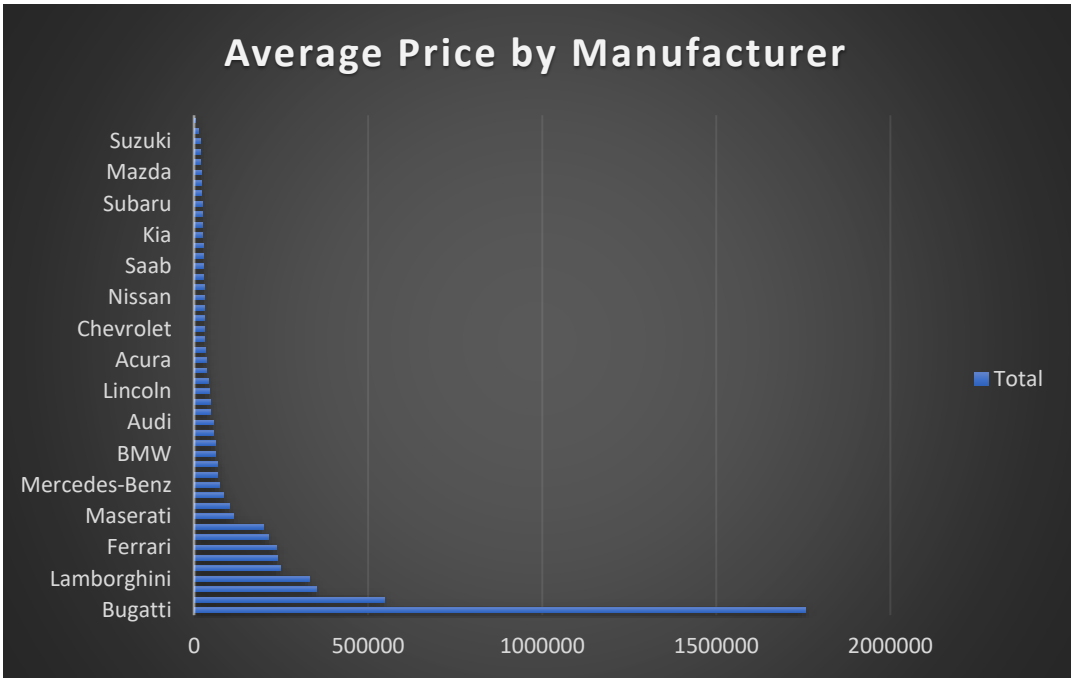
- **Task 4.A:** Create a pivot table that shows the average price of cars for each manufacturer.

Output:

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
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
Grand Total	**41925.92714**

The table shows the highest and lowest average prices based on the manufacturer.

- **Task 4.B:** Create a bar chart or a horizontal stacked bar chart that visualizes the relationship between manufacturer and average price.



Insights:

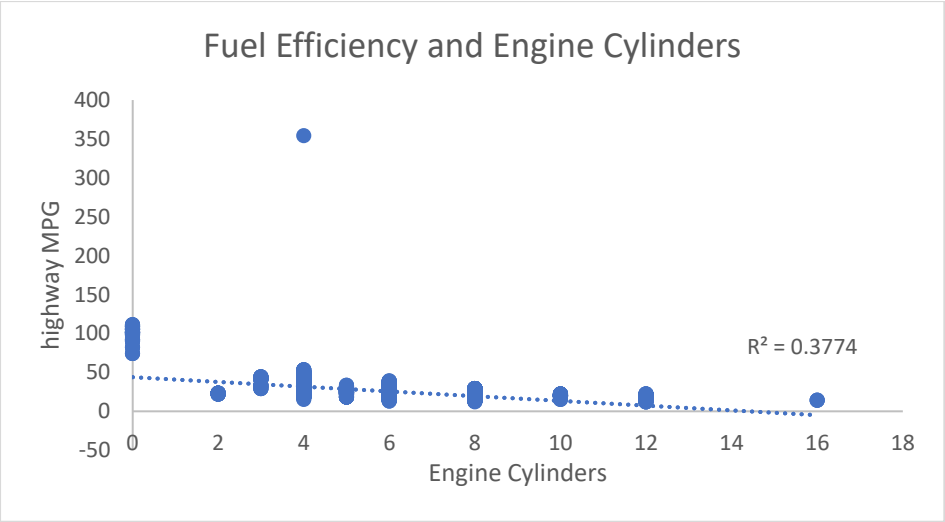
Variation in Car Prices by Manufacturer:

- we can observe how different manufacturers position their cars in the market based on average price.
- Luxury brands like Bugatti, Lamborghini, Rolls-Royce may have significantly higher average prices compared to economy brands like Suzuki, Honda, Ford.

Task 5: Fuel Efficiency and Engine Cylinders

- **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.

Output:



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

Output:

correlation coefficient	-0.614333

Insights:

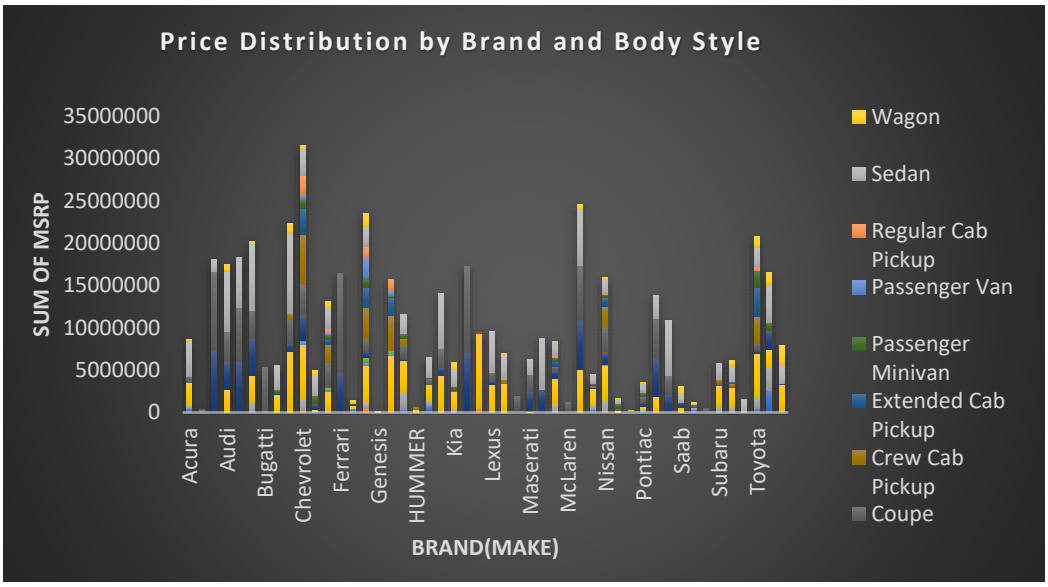
A negative correlation suggests that as the number of cylinders increases, the highway MPG decreases. This is a common finding, as vehicles with more cylinders tend to consume more fuel and therefore have lower fuel efficiency.

Building the Dashboard:

Task 1: Price Distribution by Brand and Body Style

Hints: Stacked column chart to show the distribution of car prices by brand and body style. Use filters and slicers to make the chart interactive. Calculate the total MSRP for each brand and body style using SUMIF or Pivot Tables.

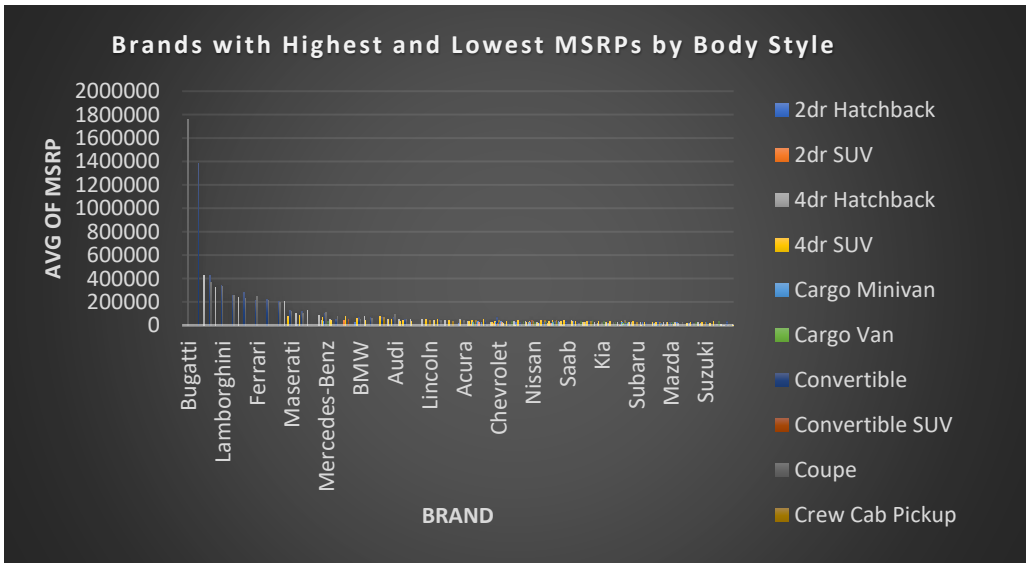
Output:



Task 2: Brands with Highest and Lowest MSRPs by Body Style

Hints: Clustered column chart to compare the average MSRPs across different car brands and body styles. Calculate the average MSRP for each brand and body style using AVERAGEIF or Pivot Tables.

Output:



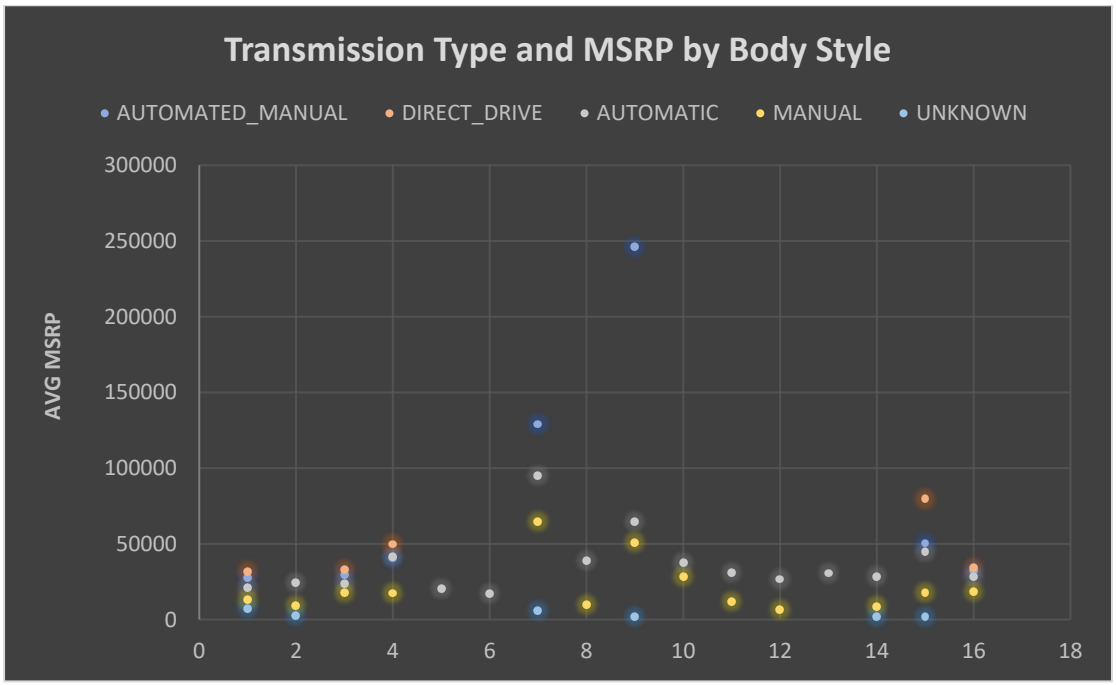
Insights:

- The brands like Bugatti, Maybach have the highest average MSRP and Bugatti having only coupe body style cars at highest average. The brands like Plymouth, Oldsmobile have the lowest average MSRP with having body styles like hatchback, sedan, wagon, and coupe.
- This analysis helps understand how different brands position themselves in the market, whether they focus on budget-friendly models or luxury vehicles.
- We can observe certain body styles coupe tend to have higher average prices compared to others like hatchbacks.

Task 3: Transmission Type and MSRP by Body Style

Hints: Scatter plot chart to visualize the relationship between MSRP and transmission type, with different symbols for each body style. Calculate the average MSRP for each combination of transmission type and body style using AVERAGEIFS or Pivot Tables.

Output:



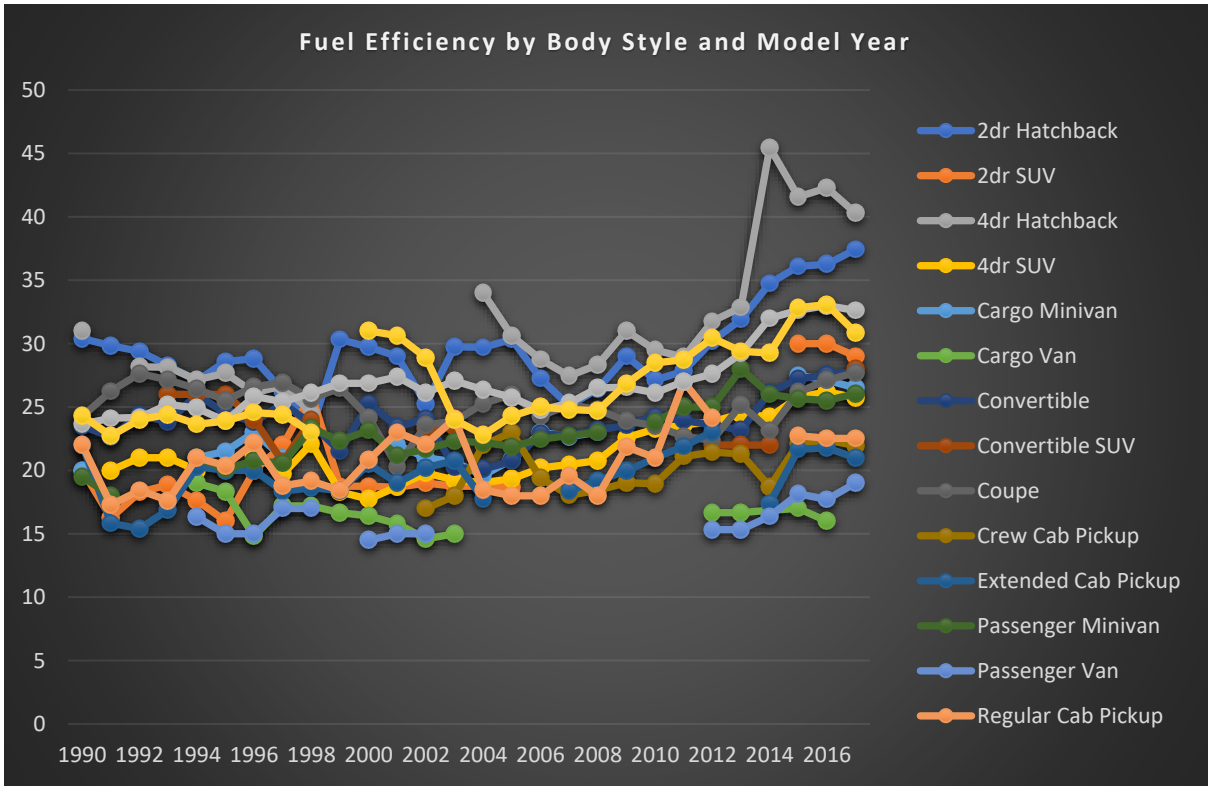
Insights:

You might find that certain body styles like convertible and coupe have higher average MSRPs with automatic transmissions compared to manual ones, indicating consumer preference.

Task 4: Fuel Efficiency by Body Style and Model Year

Hints: Line chart to show the trend of fuel efficiency (MPG) over time for each body style. Calculate the average MPG for each combination of body style and model year using AVERAGEIFS or Pivot Tables.

Output:



Insights:

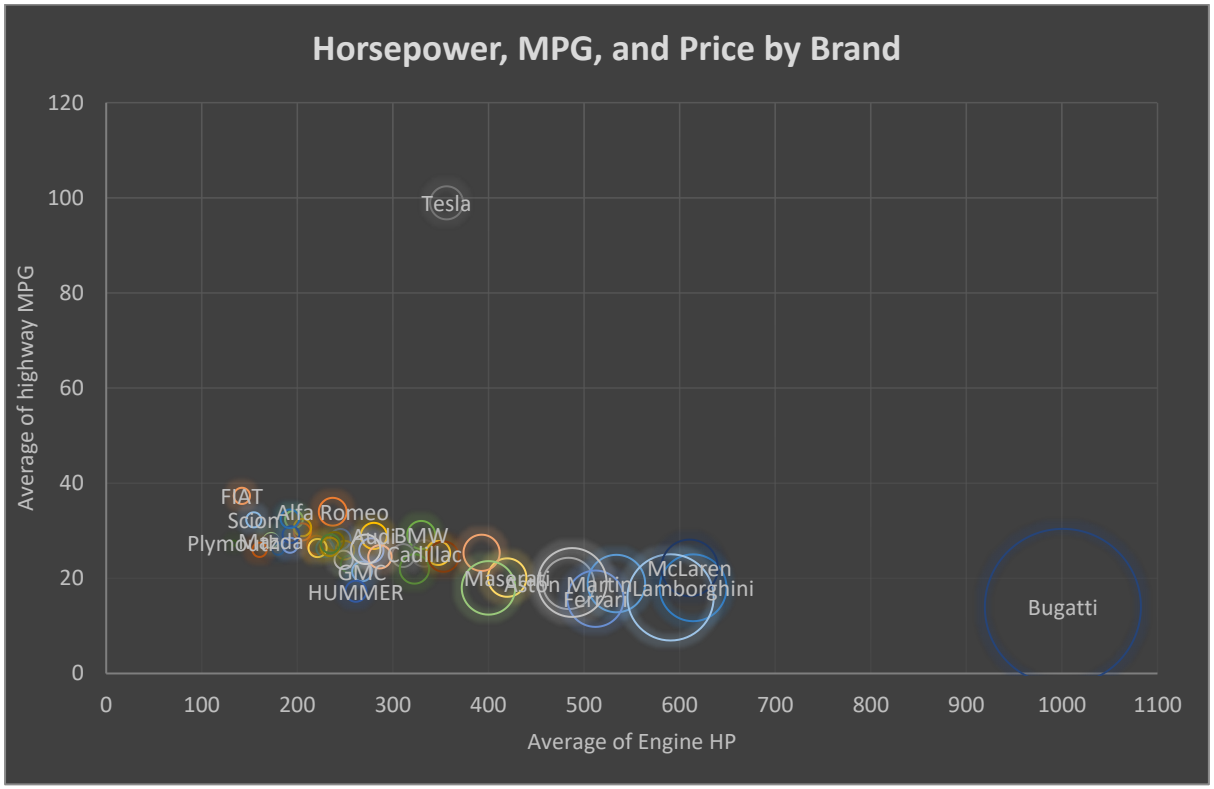
The fuel efficiency of 4dr Hatchback has increased over the years but cargo van, coupe, passenger van and few more fuel efficiency has been stagnant over the years.

Body styles like hatchback, sedan generally have better fuel efficiency than larger models like SUV, cargo van.

Task 5: Horsepower, MPG, and Price by Brand

Hints: Bubble chart to visualize the relationship between horsepower, MPG, and price across different car brands. Assign different colors to each brand and label the bubbles with the car model name. Calculate the average horsepower, MPG, and MSRP for each car brand using AVERAGEIFS or Pivot Tables.

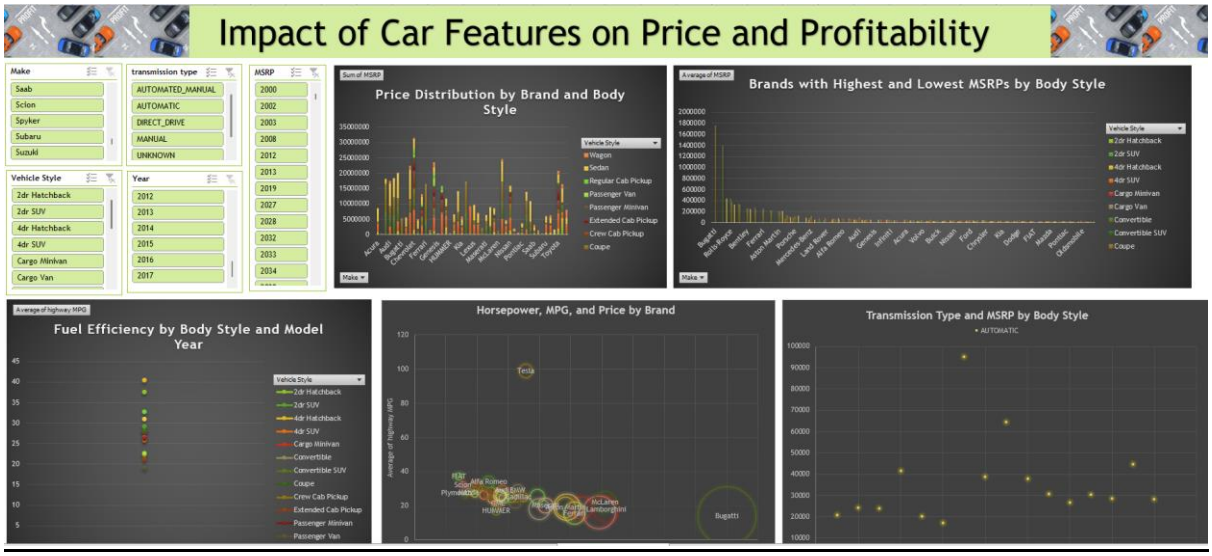
Output:



Insights:

- Cars with higher horsepower like Bugatti have higher prices.
- The chart can reveal higher horsepower comes at the expense of fuel efficiency lower MPG for certain brands.
- Brands cluster in certain areas of the chart. For example, Fiat and Honda have smaller bubbles (lower price), higher MPG, and moderate horsepower, while some brands like Ferrari, Bentley have higher horsepower but lower MPG and larger bubbles (higher price).

FINAL DASHBOARD



Conclusion:

- Higher average popularity in certain market categories like hatchback,flexfuel or flexfuel,diesel indicate that consumers tend to prefer car models within these segments, reflecting strong demand.
- Lower average popularity in other categories flexfuel,hybrid or exotic,luxury suggest lower consumer interest.
- Categories with a consistently high average popularity across models could be identified as high-potential segments for manufacturers to focus on for future product development and marketing.
- An R^2 value of 0.4337 suggests a moderate positive correlation between engine power and price. This means that about 43.37% of the variation in car prices can be explained by differences in engine power.
- While there is a relationship but not very strong, indicating that other factors also significantly influence the price.
- The positive coefficient of variables like engine cylinders, highway MPG, City MPG, engine HP shows that these features increase the price while the negative ones tend to decrease the price.
- Bar chart also shows the Variables with larger coefficients (both positive and negative) have a greater impact on the price.Positive coefficients indicate that the variable increases the price, while negative coefficients indicate a decrease.
- we can observe how different manufacturers position their cars in the market based on average price.Luxury brands like Bugatti, Lamborghini, Rolls-Royce may have significantly higher average prices compared to economy brands like Suzuki, Honda, Ford.
- A negative correlation suggests that as the number of cylinders increases, the highway MPG decreases. This is a common finding, as vehicles with more cylinders tend to consume more fuel and therefore have lower fuel efficiency.

- You might find that certain body styles like convertible and coupe have higher average MSRPs with automatic transmissions compared to manual ones, indicating consumer preference.

Result:

From working on this project, I gained valuable insights into the relationship between a car's features and its market performance, as well as how these factors influence pricing decisions.

Through tasks such as analyzing the popularity of car models across market categories and examining the relationship between engine power and price, I developed a deeper understanding of how consumer preferences and technical specifications play a role in shaping the automotive market.

This project improved my skills in data analysis using Excel, particularly in areas like pivot tables, regression analysis, and visualizations through scatter plots and trendlines. Additionally, working with real-world automotive data helped me appreciate the importance of normalizing comparisons and understanding how various factors contribute to overall trends. This experience strengthened my ability to extract actionable insights and make data-driven recommendations, which will be essential for future projects in the field of data analysis.

The excel sheet: linked here