A

Data Analytics

Project on

Analyzing the Impact of Car Features on Price and Profitability

File Links:
This File: [Click Here!]

Working File: [Analysis Files Folder]

[Car_data.xlsx]

By:

Prince Kumar
prince22495@gmail.com

CONTENT

1. Project Description
2. Approach
3. Tech-Stack Used
4. Insights
5. Summary of Results

Project Description:

Overview:

The automotive industry is undergoing tremendous transformations as a result of a focus on fuel efficiency, environmental sustainability, and technological innovation. Understanding consumer demand for cars is critical for car manufacturers to stay competitive in this dynamic landscape. This initiative attempts to help a car manufacturer maximize profitability by improving pricing and product development decisions. The goal is to give relevant insights that align with consumer preferences and create corporate success by leveraging the power of data analysis.

Business Problem and Data Sources:

Understanding the factors that impact consumer demand for cars and their relationship with pricing and profitability is the primary business concern. To address this problem, we use a large and detailed dataset titled "**Car Features and MSRP**", which was collected and made available on **Kaggle** by **Cooper Union,** a private college located in New York City. This dataset contains details on approximately **11,000 car models**, including make, model, year, engine specifications, gearbox type, fuel efficiency, market category, size, and manufacturer's suggested retail price (MSRP).

Data Cleaning and Preprocessing Steps:

Before beginning the data analysis journey, we prioritize extensive data cleaning and preprocessing activities to ensure accurate and dependable outcomes. This critical phase includes addressing missing data, deleting duplicates, converting data types, finding and treating outliers, scaling numerical variables, and transforming the data as needed. Furthermore, we standardize market categories to reduce complexity and improve analysis results. By properly preparing the data, we lay the groundwork for effective exploratory data analysis and model creation, allowing the automobile manufacturer to make educated decisions.

Interactive Dashboard Building:

The data analysis culminates in the creation of an interactive dashboard using advanced Excel skills. The dashboard addresses the client's specific questions and visualizes the distribution of car prices by brand and body style, the average MSRPs across different

car brands and body styles, and the impact of different features on MSRP, among other insights. I employed filters and slicers to ensure user-friendly interactivity, enabling the car manufacturer to make strategic decisions in real-time.

Conclusion:

This data-driven approach provides the car manufacturer with a competitive edge in understanding consumer preferences, optimizing pricing strategies, and guiding product development efforts. By leveraging the power of data analysis and an interactive dashboard, we empower the manufacturer to navigate the rapidly changing automotive landscape successfully. The project's comprehensive insights and actionable recommendations drive long-term profitability and cater to the evolving demands of the automotive market.

Approach

Data Cleaning Process:

1. **Imputations:**

- **Engine Cylinder (Blanks) for Electric**: In the "Engine Cylinder" column, I observed blank entries for electric cars since they do not have cylinders. To maintain consistency and accuracy, I replaced these blank entries with "0" to indicate that electric cars have zero cylinders. A total of 10 blank entries were imputed.

- **Engine Fuel Type Imputation:** Some rows in the dataset had missing values in the "Engine Fuel Type" column. However, the corresponding MPG city and highway averages matched those of "flex-fuel (unleaded/natural gas)" vehicles. Therefore, I imputed the missing values with "flex-fuel (unleaded/natural gas)" for these rows to ensure data consistency. A total of 3 missing values were imputed.

2. **Deletion:**

- **Removal of Rows with Blank Engine HP:** I removed 69 rows from the dataset that had blank entries in the "Engine HP" column. Engine horsepower is a crucial attribute for analysis, and to ensure accurate results, I decided to remove these rows.

- **Removal of Rows with Blank Engine Cylinder:** Additionally, I removed 20 rows from the dataset that had blank entries in the "Engine Cylinder" column. As the number of cylinders impacts various car performance aspects, removing these rows is essential for accurate analysis.

After performing data cleaning, the dataset was refined, resulting in 11,826 rows and 16 columns. Initially, the dataset had 11,915 rows.

Data analysis tool pack was also used to find out regression values and to identify most important variables in determining the MSRP of a car.

Apart from this my approach was also to include dashboards wherever necessary to make the analysis interactive and easily understandable to the user.

Tech Stack Used: MS Excel

For this project, the primary tech stack utilized is Microsoft Excel, a powerful spreadsheet software that offers a wide range of data analysis and visualization capabilities. Excel is a versatile tool that facilitates data manipulation, cleaning, and exploration, making it ideal for handling the dataset and performing various analytical tasks. The following features and functionalities of MS Excel were leveraged in the project:

- **Pivot Tables and Pivot Charts**
- **Data Visualization**
- **Regression Analysis**
- **Slicers and Filters**

Insights:

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.

Solution:

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
Crossover,Performance	69	2586
Crossover,Hybrid	42	2563
Luxury,Performance,Hybrid	11	2333
Diesel,Luxury	51	2275
Flex Fuel	872	2217
Hatchback,Factory Tuner,Performance	22	2159
Crossover,Luxury,Diesel	34	2149
Factory Tuner,Luxury,High-Performance	215	2133
Hatchback,Hybrid	72	2121
Hybrid	123	2106
Crossover,Flex Fuel	64	2074
Crossover,Hatchback,Performance	6	2009
Crossover,Hatchback,Factory Tuner,Performance	6	2009
Factory Tuner,High-Performance	106	1941
Crossover,Factory Tuner,Luxury,High-Performance	26	1823
High-Performance	199	1821
Diesel	84	1731
Factory Tuner,Performance	92	1696
Flex Fuel,Performance	87	1680
Crossover,Hatchback	72	1676
Luxury,High-Performance	334	1668
Crossover,Flex Fuel,Luxury,Performance	6	1624
Hatchback,Luxury,Performance	38	1566
Crossover	1110	1545
Factory Tuner,Luxury,Performance	31	1413
Exotic,Performance	10	1391
Flex Fuel,Luxury,Performance	28	1380
Hatchback,Luxury	46	1380

Performance	601	1349
Crossover,Luxury,Performance	113	1345
Hatchback	641	1319
Luxury,Performance	673	1293
Exotic,High-Performance	261	1271
Hatchback,Factory Tuner,High-Performance	13	1205
Crossover,Flex Fuel,Luxury	10	1173
Luxury	855	1103
Exotic,Factory Tuner,High-Performance	21	1046
Hatchback,Performance	252	1040
Crossover,Luxury,High-Performance	9	1037
Hatchback,Factory Tuner,Luxury,Performance	9	887
Crossover,Luxury	410	885
Flex Fuel,Luxury,High-Performance	33	879
Crossover,Diesel	7	873
Hatchback,Diesel	14	873
Flex Fuel,Luxury	39	747
Luxury,Hybrid	52	674
Crossover,Luxury,Hybrid	24	631
Factory Tuner,Luxury	2	617
Luxury,High-Performance,Hybrid	12	569
Exotic,Factory Tuner,Luxury,Performance	3	520
Exotic,Flex Fuel,Luxury,High-Performance	11	520
Exotic,Flex Fuel,Factory Tuner,Luxury,High-Performance	13	520
Exotic,Factory Tuner,Luxury,High-Performance	52	518
Exotic,Luxury,High-Performance	79	467
Hatchback,Luxury,Hybrid	3	454
Flex Fuel,Factory Tuner,Luxury,High-Performance	1	258
Crossover,Exotic,Luxury,Performance	1	238
Crossover,Exotic,Luxury,High-Performance	1	238
Exotic,Luxury,Performance	36	217
Crossover,Factory Tuner,Performance	4	210
Exotic,Luxury,High-Performance,Hybrid	1	204
Crossover,Hatchback,Luxury	7	204
Flex Fuel,Performance,Hybrid	2	155
Performance,Hybrid	1	155
Flex Fuel,Hybrid	2	155
Exotic,Luxury	12	113

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

Solution:
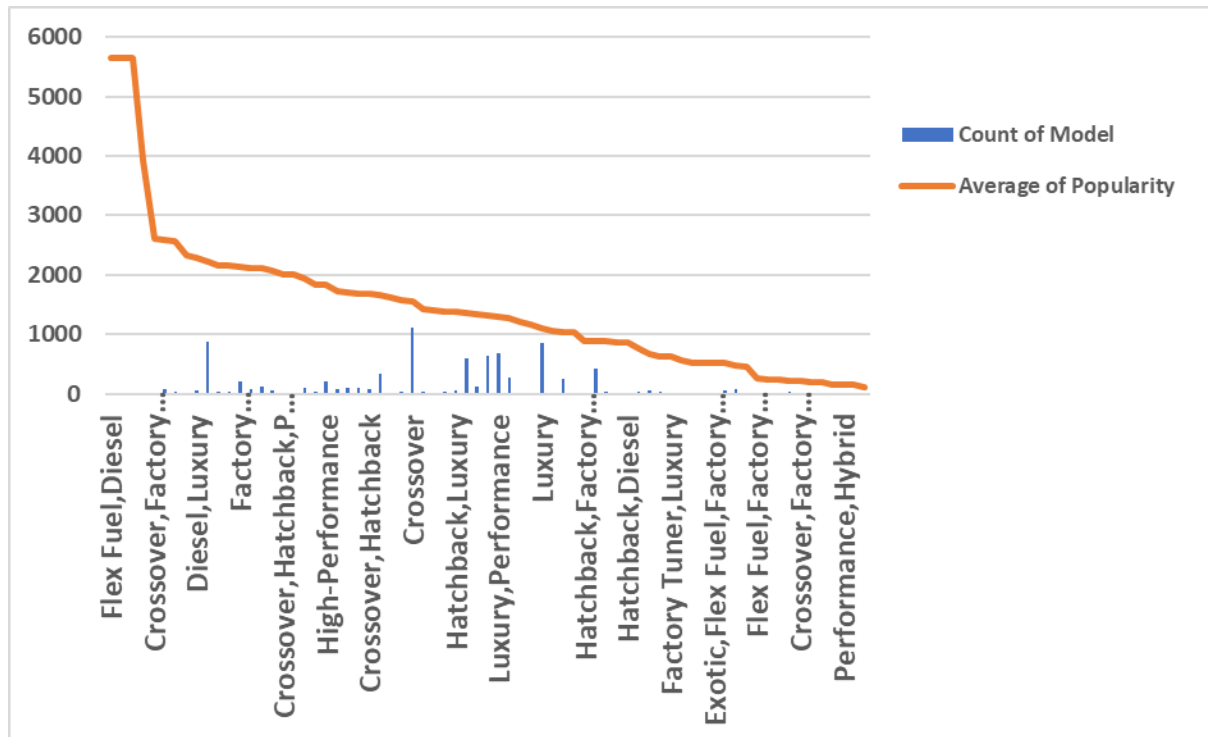
Insight:

- It could be found from the above results that the most popular market category belongs to (**Flex Fuel, Diesel), (Hatchback, Flex Fuel)** and (**Crossover, Flex Fuel, Performance)** with an average popularity score of **5657**.
- **Crossover** has the greatest number of car models which is **1110**.
- **(Exotic, Luxury)** market category has scored the least average popularity score of **113**.

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.

Solution:
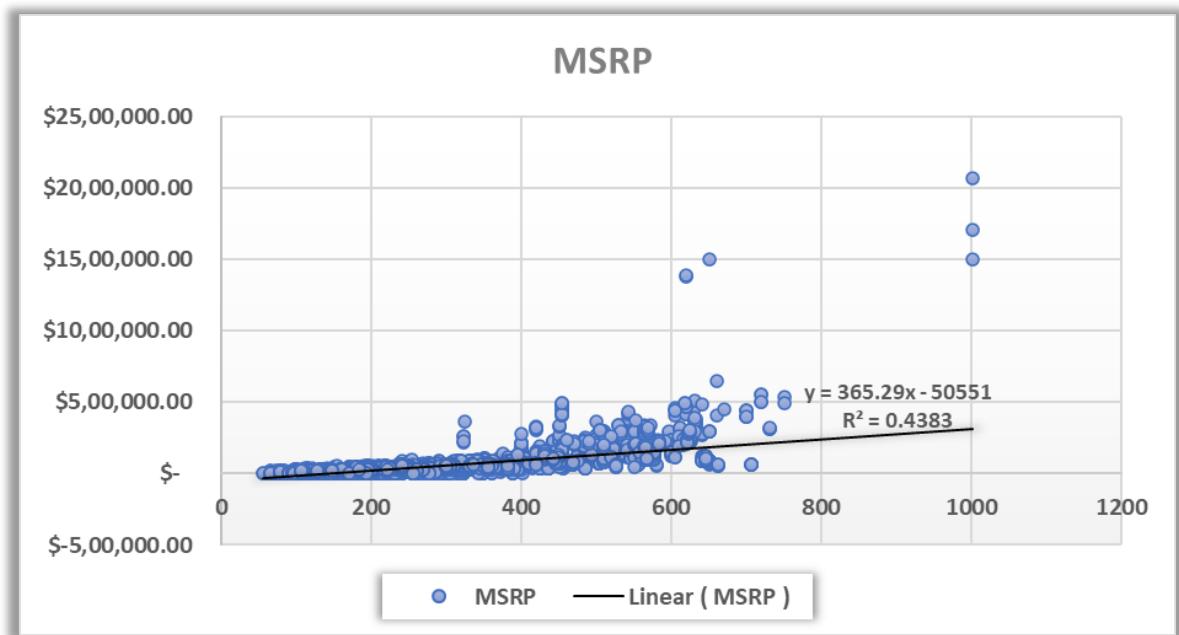
Insight:

The trendline of this scatter plot between engine power and price of cars shows the positive relation between the two variables. The trendline is in black colour and it is sloped upwards indicating a positive relation between the engine power and the price of cars. This means that as the engine power increases, the price also increases accordingly.

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.

Solution:

Importance of car features in price determination:

Variables and their coefficient values calculated by doing regression analysis.

Year	101.0670754
Engine HP	315.433388
Engine Cylinders	7097.586063
Number of Doors	-4885.17824
highway MPG	572.4894366
city mpg	963.8382761
Popularity	-3.333442435
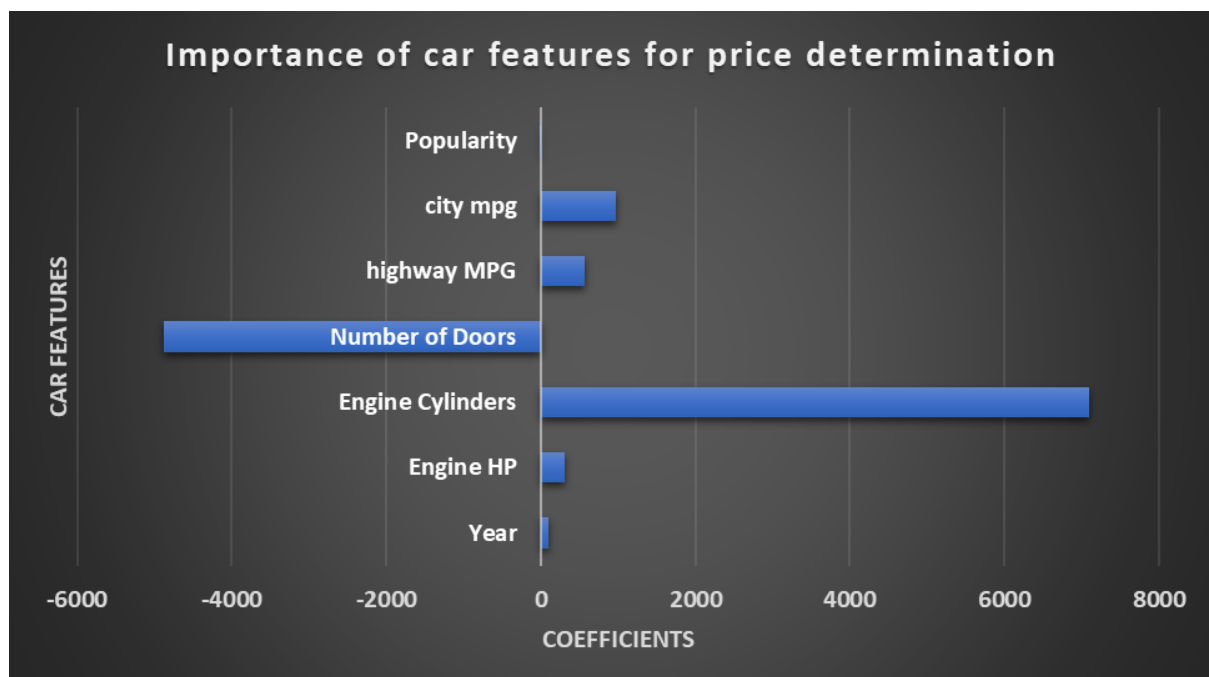
Insight:

- Through regression analysis and the bar chart shown above, it is found that **Engine Cylinders** play the most significant role in determining the Price of a car. Other important variable is the city **mileage**.
- Also, it could be seen that the **Number of doors** a car offers doesn't have any impact in determining the price of a car.

**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.

Solution:

Row Labels	Average of MSRP
Acura	$ 34,887.59
Alfa Romeo	$ 61,600.00
Aston Martin	$ 1,97,910.38
Audi	$ 53,452.11
Bentley	$ 2,47,169.32
BMW	$ 61,546.76
Bugatti	$ 17,57,223.67
Buick	$ 28,206.61
Cadillac	$ 56,231.32
Chevrolet	$ 28,350.39
Chrysler	$ 26,722.96
Dodge	$ 22,390.06
Ferrari	$ 2,38,218.84
FIAT	$ 22,670.24
Ford	$ 27,399.27
Genesis	$ 46,616.67
GMC	$ 30,493.30
Honda	$ 26,674.34
HUMMER	$ 36,464.41
Hyundai	$ 24,597.04
Infiniti	$ 42,394.21
Kia	$ 25,310.17
Lamborghini	$ 3,31,567.31
Land Rover	$ 67,823.22
Lexus	$ 47,549.07
Lincoln	$ 42,839.83
Lotus	$ 69,188.28
Maserati	$ 1,14,207.71

Maybach	$ 5,46,221.88
Mazda	$ 20,039.38
McLaren	$ 2,39,805.00
Mercedes-Benz	$ 71,476.23
Mitsubishi	$ 21,240.54
Nissan	$ 28,583.43
Oldsmobile	$ 11,542.54
Plymouth	$ 3,122.90
Pontiac	$ 19,321.55
Porsche	$ 1,01,622.40
Rolls-Royce	$ 3,51,130.65
Saab	$ 27,413.50
Scion	$ 19,932.50
Spyker	$ 2,13,323.33
Subaru	$ 24,827.50
Suzuki	$ 17,907.21
Tesla	$ 85,255.56
Toyota	$ 29,030.02
Volkswagen	$ 28,102.38
Volvo	$ 28,541.16

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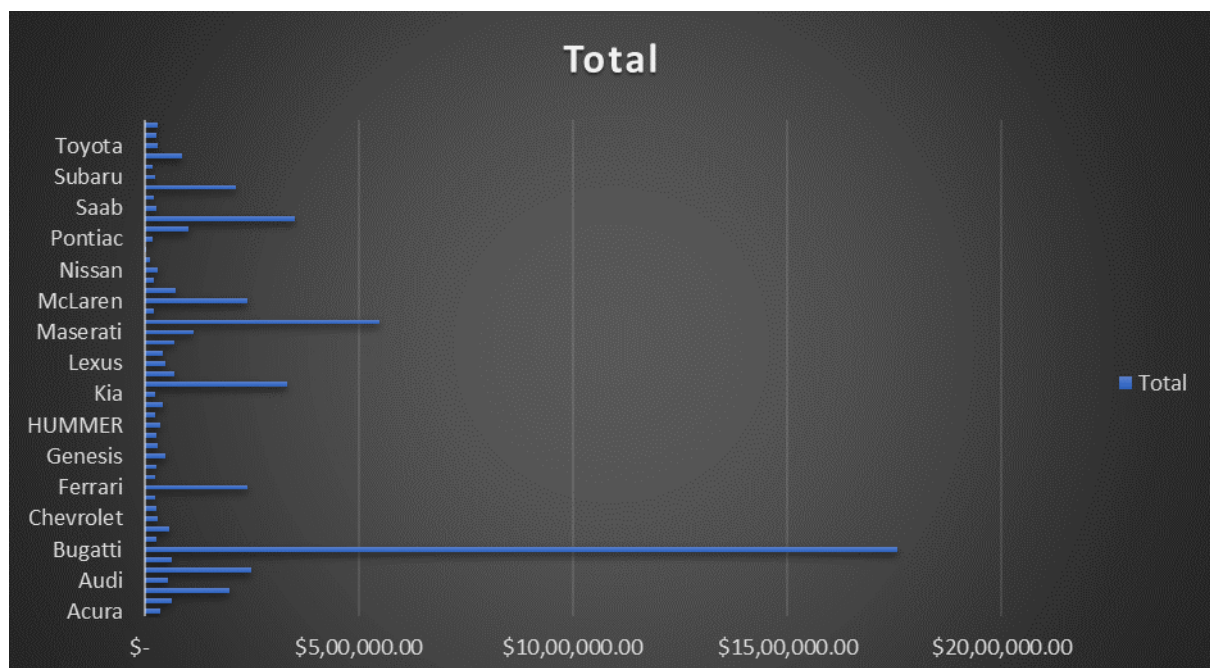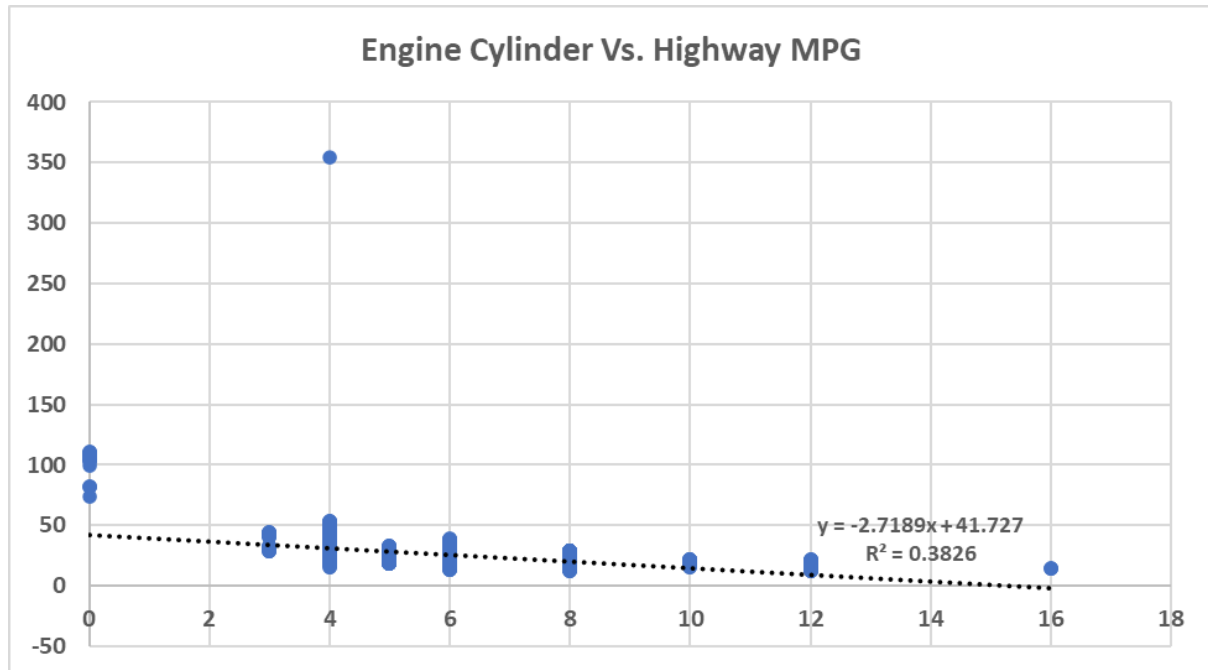

Insight:

Bugatti stands at the top with the highest average price among different manufacturers. Followed by **Maybach** at the second position.

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.

Solution:

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

Solution:

| CORREL COFFIECIENT	-0.618546383

Insight:

- The correlation coefficient **(-0.6185)** indicates a **moderate negative** relationship between **engine cylinder** count and **highway miles per gallon (MPG)**. As the number of cylinders **increases**, highway MPG tends to **decrease**.
- This correlation suggests that vehicles with **larger engines (more cylinders)** tend to have **lower fuel efficiency** on highways. Conversely, **smaller engines (fewer cylinders)** tend to have **better highway fuel efficiency.**

Building the Dashboard:

Task 1: How does the distribution of car prices vary by brand and body style?

Solution:

Distribution of car total prices varying with brands and vehicle style.
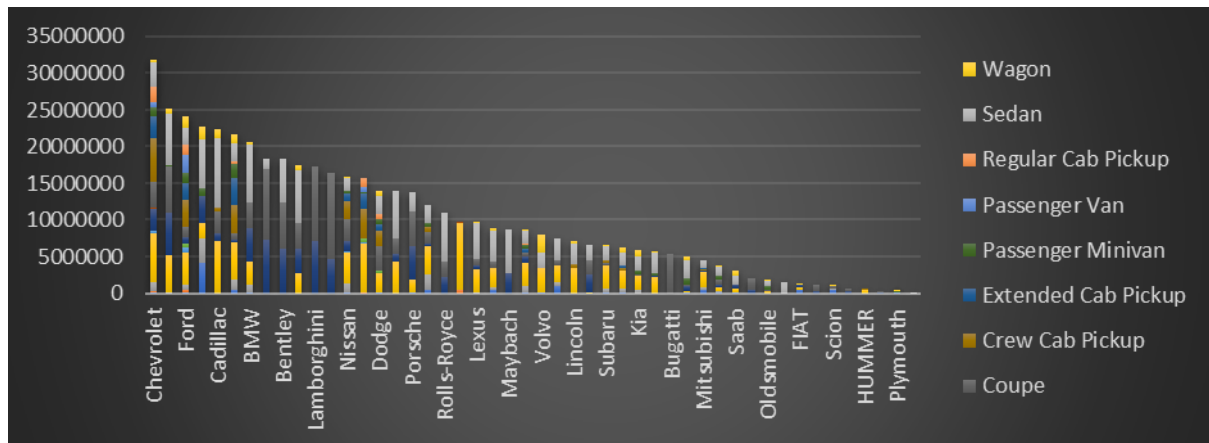
This stacked chart comes with slicers for showing results as per the Make and the vehicle style of cars. (Please look in the excel sheet provided.) Pivot table was used to create this stacked chart.
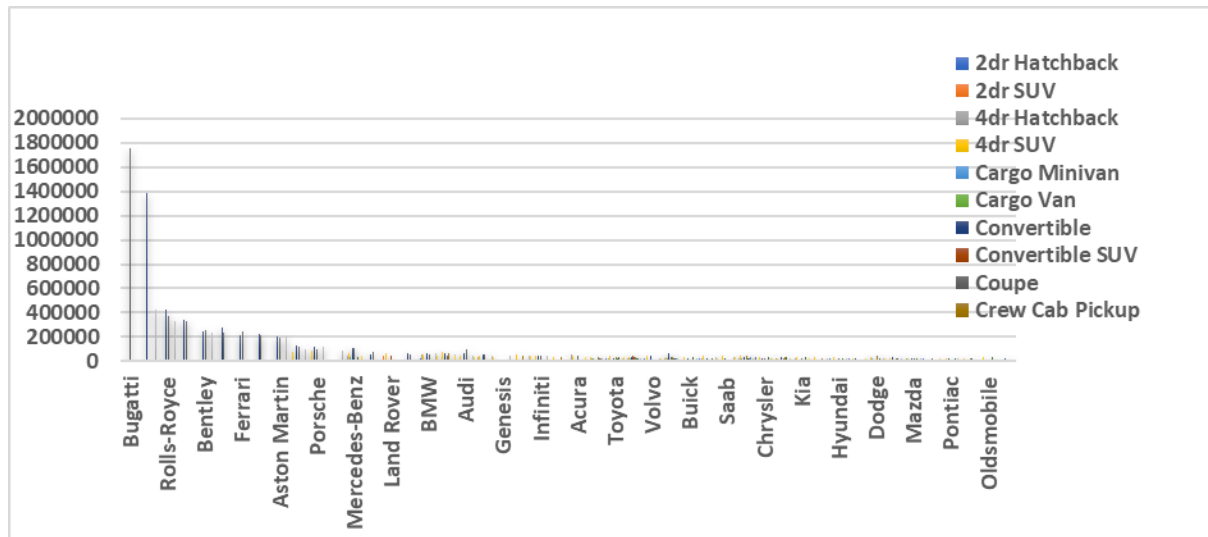
Insight:
Chevrolet has the **maximum** total car price among all the car brands. Followed by **Mercedes-Benz** and **Ford.**

Task 2: Which car brands have the highest and lowest average MSRPs, and how does this

vary by body style?

Solution:

Distribution Car average prices and its variation with brand and vehicle style.

(For better look at the dashboard with slicers, kindly refer to the excel sheet provided along with this document.)
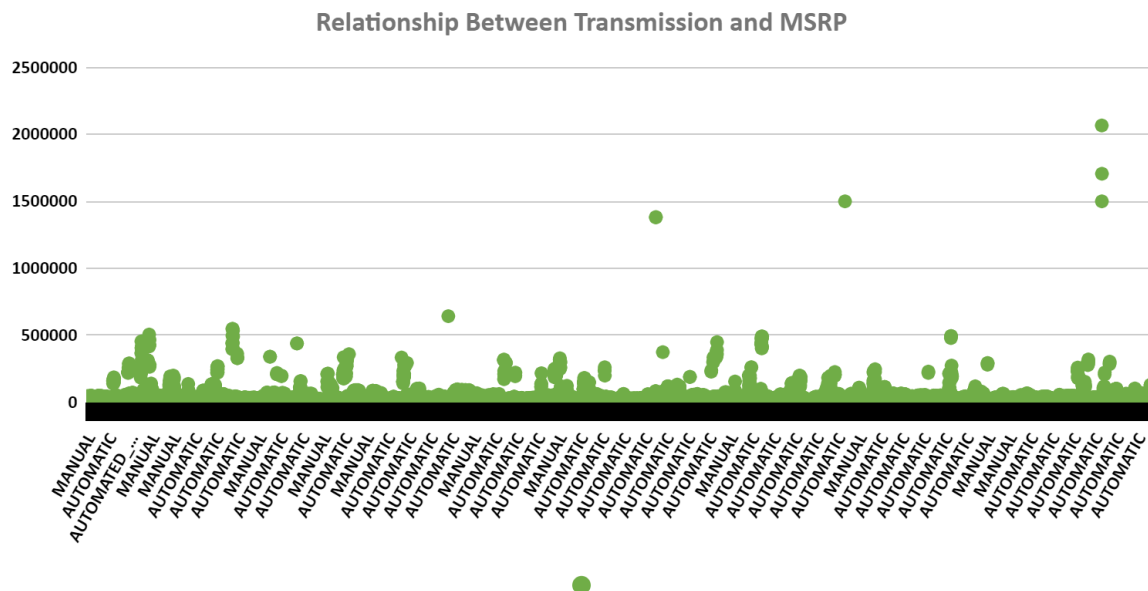
Insight:

Bugatti has the **highest** average MSRP and **Plymouth** has the lowest average **MSRP**.

Task 3: How do the different feature such as transmission type affect the MSRP, and how does this vary by body style?
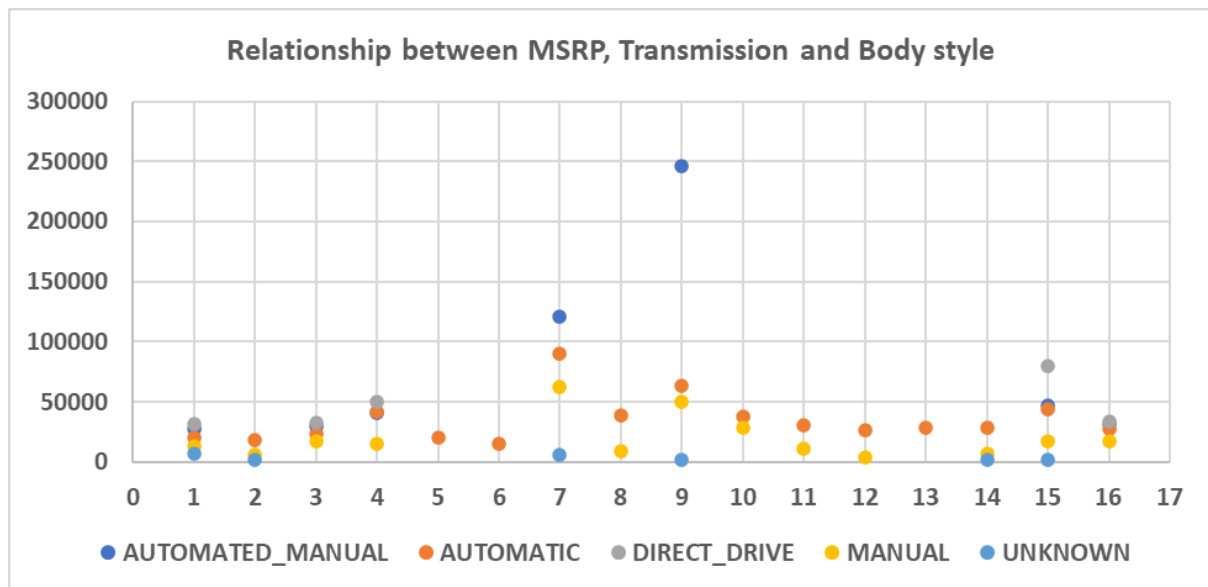
Solution:

Relationship Between Transmission and MSRP

(Kindly refer to the excel sheet for better view. Please Note that I made this chart on Google sheets since my excel sheet chart was not showing horizontal axis labels properly.)

Pivot table having vehicle style in rows, transmission in columns, and MSRP in values:

S. NO.	Row Labels	AUTOMATED_MANUAL	AUTOMATIC	DIRECT_DRIVE	MANUAL	UNKNOWN
1	2dr Hatchback	27180.96491	20926.464	31800	13353.65831	7361.5
2	2dr SUV		18615.20455		6303.811111	2371
3	4dr Hatchback	29249.07407	23833.67898	32799.72973	17594.41313	
4	4dr SUV	40451.15385	41535.60646	49800	15426.46226	
5	Cargo Minivan		20920.98592			
6	Cargo Van		15280.22105			
7	Convertible	121256.6444	90637.3869		62357.75625	5783.5
8	Convertible SUV		38925.5		9233.142857	
9	Coupe	245977.4252	63371.81076		50484.37241	2000
10	Crew Cab Pickup		37744.07154		28360.52632	
11	Extended Cab Pickup		30637.34973		10884.19455	
12	Passenger Minivan		26412.68159		4405.333333	
13	Passenger Van		29015.20313			
14	Regular Cab Pickup		28536.8239		7557.773333	2000
15	Sedan	47498.70813	43769.1165	79512.25	17119.23374	2000
16	Wagon	31985.27778	27613.19169	34250	17844.13971	

Relationship between MSRP, Transmission and Body style

● AUTOMATED_MANUAL ● AUTOMATIC ● DIRECT_DRIVE ● MANUAL ● UNKNOWN
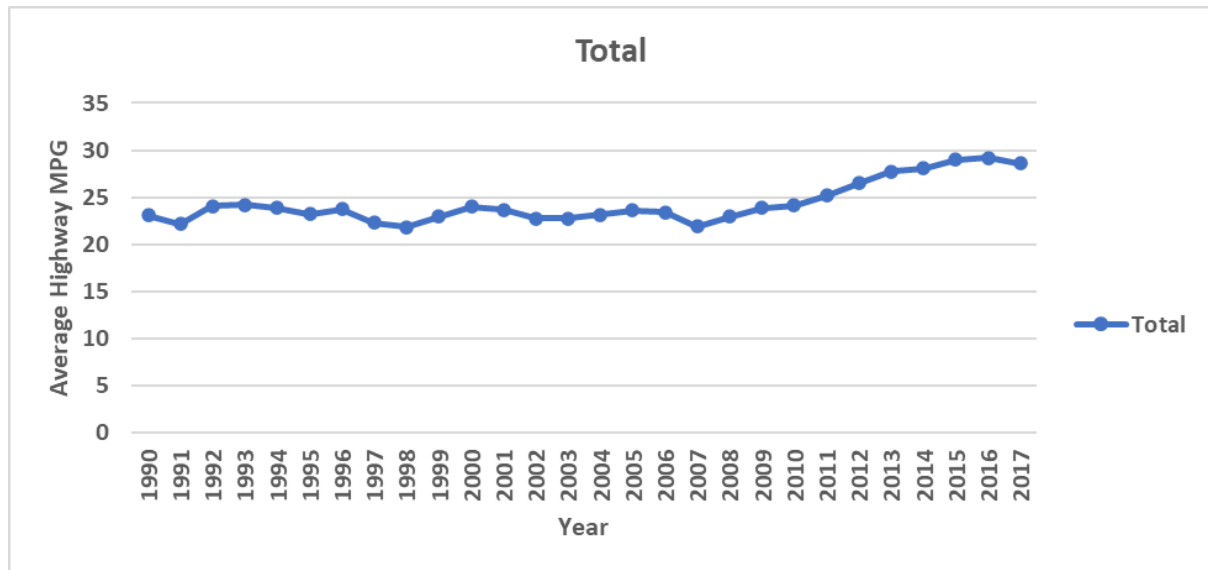
Insight:

- **Automatic_manual** transmission is mostly present at **higher** MSRP. For e.g. In **convertible** and **coupe** vehicle styles.
- Most cars have **manual** or **automatic transmission** which includes cars with **lower** MSRP as well.
- **Automatic** transmission comes in all vehicle types.

Task 4: How does the fuel efficiency of cars vary across different body styles and model years?

Solution:

Relationship between Year and average Highway MPG, Filtered with a Vehicle style Slicer:

(For better view of the dashboard kindly refer to the attached excel file along with this document.)

**Pivot table was used to arrive at following insights.**

Insight:

- Highway Mileage gradually increases from year **2007** till **2017** where it reaches from **22 MPG** to **29 MPG**.
- **2dr Hatchback** and **4dr hatchback** are the **most efficient** vehicle styles out of all the styles.
- **Cargo Van** and **Passenger Van** vehicle styles are the **least efficient** of the whole list of vehicle styles.

Task 5: How does the car's horsepower, MPG, and price vary across different Brands?

Solution:

Car's horsepower, MPG, and price variation across different brands:
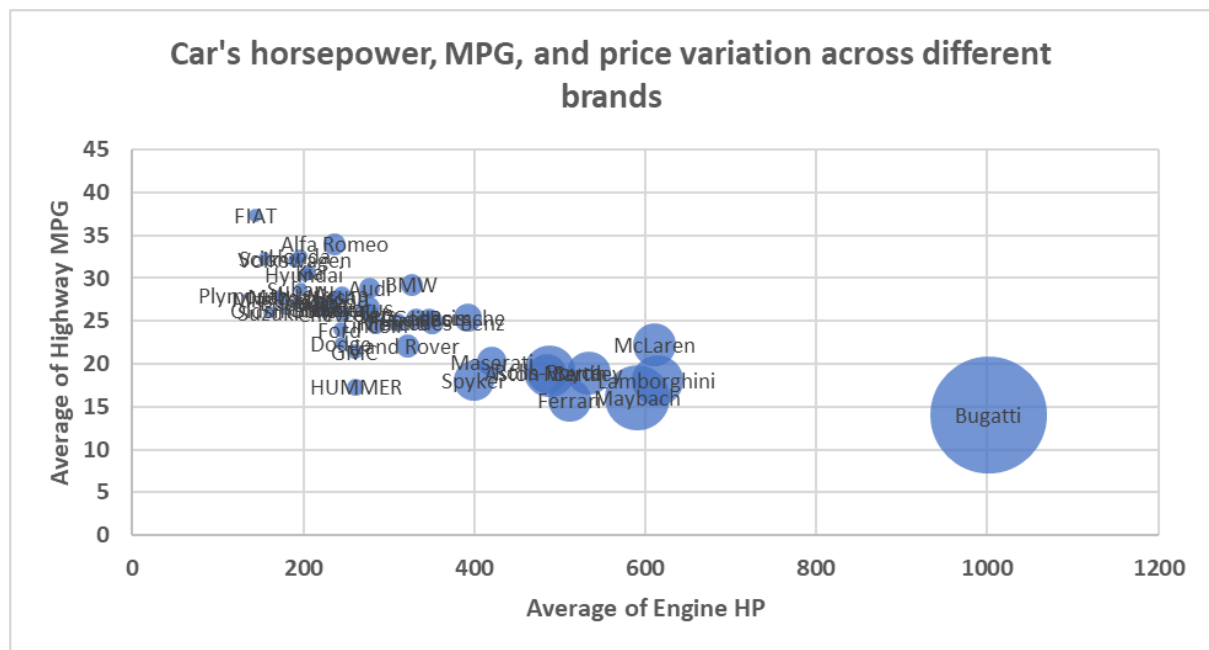
Bubble chart where size of the bubble represents the average MSRP values of the brand. Brand in labelled on the circle.

Insight:

● It can be seen here that **Bugatti** has the highest average MSRP value. Followed by **Maybach** and **Rolls-Royce.**

● Most number of cars lie in the **200 - 400 HP** range and have the highway MPG values between **20 and 35.**

Summary of results:

All the questions were answered with visualizations for better understanding. Here is the summarized list of all the insights from the all the tasks:

Task 1:

- The most popular market category belongs to (**Flex Fuel, Diesel), (Hatchback, Flex Fuel)** and (**Crossover, Flex Fuel, Performance)** with an average popularity score of **5657**.
- **Crossover** has the greatest number of car models which is **1110**.
- **(Exotic, Luxury)** market category has scored the least average popularity score of **113.**

Task 2:

- The trendline of this scatter plot between engine power and price of cars shows the positive relation between the two variables. The trendline is sloped upwards indicating a **positive** relation between the **engine power** and the **price** of cars. This means that as the **engine power increases, the price also increases accordingly.**

Task 3:

- Through regression analysis, it is found that **Engine Cylinders** play the most significant role in determining the Price of a car. Other important variable is the **city mileage**.
- The **Number of doors** a car offers, doesn't have any impact in determining the price of a car.

Task 4:

- **Bugatti** stands at the top with the highest average price among different manufacturers. Followed by **Maybach** at the second position.

Task 5:

- The correlation coefficient (-0.6185) indicates a **moderate negative** relationship between **engine cylinder** and **highway miles** per gallon (MPG). As the number of **cylinders increases**, **highway MPG** tends to **decrease**.

Building the Dashboard:

Task 1:

- **Chevrolet** has the **maximum** total car price among all the car brands. Followed by **Mercedes-Benz** and **Ford.**

Task 2:

- **Bugatti** has the **highest** average MSRP and **Plymouth** has the lowest average **MSRP**.

Task 3:

- **Automatic_manual** transmission is mostly present at **higher** MSRP. For e.g. In **convertible** and **coupe** vehicle styles.

- Most cars have **manual** or **automatic transmission** which includes cars with **lower** MSRP as well.

- **Automatic** transmission comes in all vehicle types.

Task 4:

- Highway Mileage gradually increases from year **2007** till **2017** where it reaches from **22 MPG** to **29 MPG**.

- **2dr Hatchback** and **4dr hatchback** are the **most efficient** vehicle styles out of all the styles.

- **Cargo Van** and **Passenger Van** vehicle styles are the **least efficient** of the whole list of vehicle styles.

Task 5:

- **Bugatti** has the highest average MSRP value. Followed by **Maybach** and **Rolls-Royce.**

- Most number of cars lie in the **200 - 400 HP** range and have the highway MPG values between **20 and 35.**