Analyzing the Impact of Car Features on Price and Profitability

Project Description

Overview:

This project investigates automotive market data to analyze pricing dynamics, performance factors, and design an interactive Excel dashboard that reveals how car features affect MSRP (Manufacturer's Suggested Retail Price).

Business Problem:

"What features most significantly influence car pricing, and how do brand, performance, and vehicle type interact in determining value?"

The goal is to empower analysts, dealerships, and auto manufacturers with a self-service dashboard to guide decisions on pricing, inventory, and marketing.

Data Source:

This dataset which was downloaded from **KAGGLE** website, consists of:

- 11,000+ records of cars from various brands and body styles
- Features like Engine HP, Transmission Type, Driven Wheels, Vehicle Style, MPG, MSRP, Year, etc.

Approach

The project was executed using a structured data analysis pipeline:

- **Data Cleaning & Preprocessing**
  - Performed using Python before import into Excel:
  - Python considers "**N/A**" in "Market Category" as **NaN**, so we changed that value by "**UNKNOWN**".
  - Imputed missing values using mode or mean whichever applicable.
  - Removed duplicates.
  - Exported to **.xlsx** format for Excel analysis
- **Analytical Techniques:**
  - **Descriptive Statistics** using PivotTables because Excel is accessible and perfect for non-technical users.
  - Visual formats suit executive insights hence we use **Excel** for **Visualization**
  - **Regression Analysis** using Excel's Data Analysis Toolpak because Regression helps quantify impact of variables like HP and MPG on pricing.
- **Modelling Technique:**
  - **Multiple Linear Regression** using:
    - X variables: Engine HP, Cylinders, Doors, MPG
    - Y variable: MSRP
    - Tools: Excel's Regression tool (Toolpak)
  - Interpreted **R^2**, **coefficients**, and **p-values** to rank predictor strength.

- **Assumptions**:
  - No inflation adjustment to MSRP across years
  - Categorical imputation defaults to most frequent class
  - Every row corresponds to one distinct car configuration
- **We have faced few Challenges like-**
  - Missing values in categorical columns
  - Excel's limitation on dynamic tooltips and conditional interactivity

Tech-Stack Used

In this project, we have used-

- **Python (Jupyter Notebook, Pandas, NumPy)** – Employed for data cleaning
- **Microsoft Excel 2024** – Used for initial data exploration, applying formulas (COUNTIF, AVERAGE, MEDIAN, STDEV, CORREL) for statistical calculations, using graphs for visualization, summarizing data and creating Dashboard.
- **Google Drive** – Hosting and sharing reports.

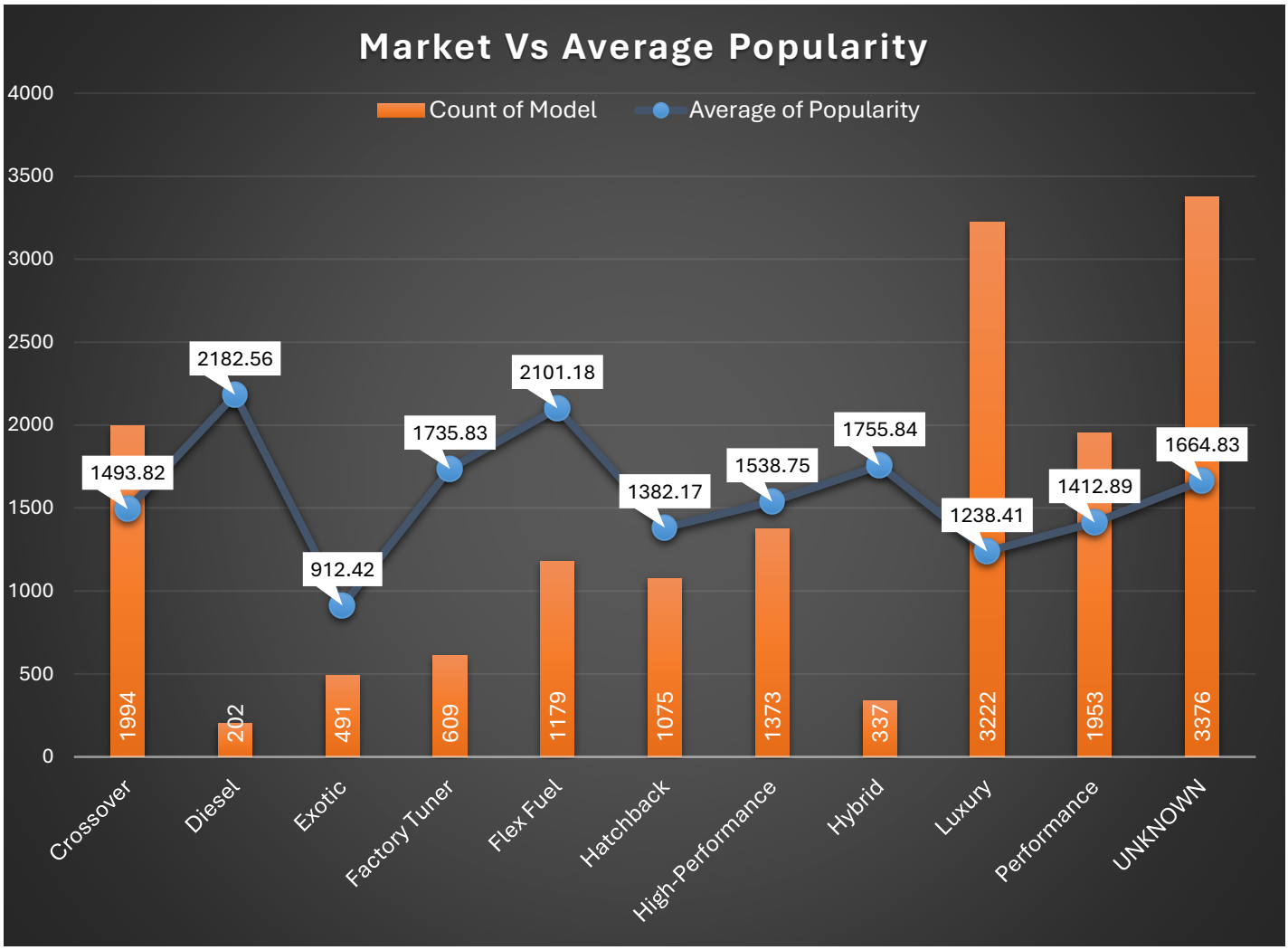
Data Analytics Tasks:

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.

Market Category	Average of Popularity	Count of Model
Crossover	1493.82	1994
Diesel	2182.56	202
Exotic	912.42	491
Factory Tuner	1735.83	609
Flex Fuel	2101.18	1179
Hatchback	1382.17	1075
High-Performance	1538.75	1373
Hybrid	1755.84	337
Luxury	1238.41	3222
Performance	1412.89	1953
UNKNOWN	1664.83	3376

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

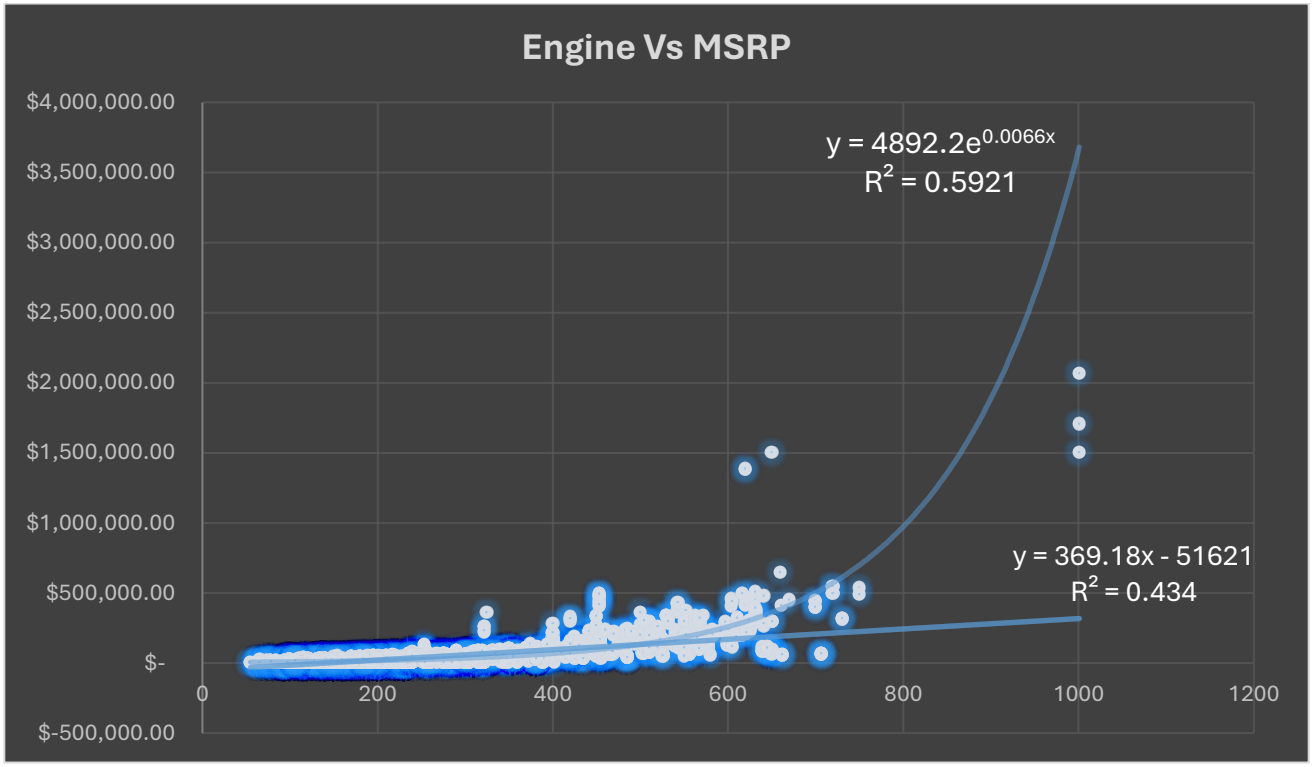


Insights-

- **Diesel** and **Flex Fuel** cars are small segments but have very loyal followings.
- **Luxury** and **Exotic** cars look flashy but attract fewer buyers on average.
- **Crossovers** and **Performance** have many models but only average popularity — too many options dilute attention.
- **Hybrids** and **Factory Tuners** stand out, showing strong appeal for innovation and customization.

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.

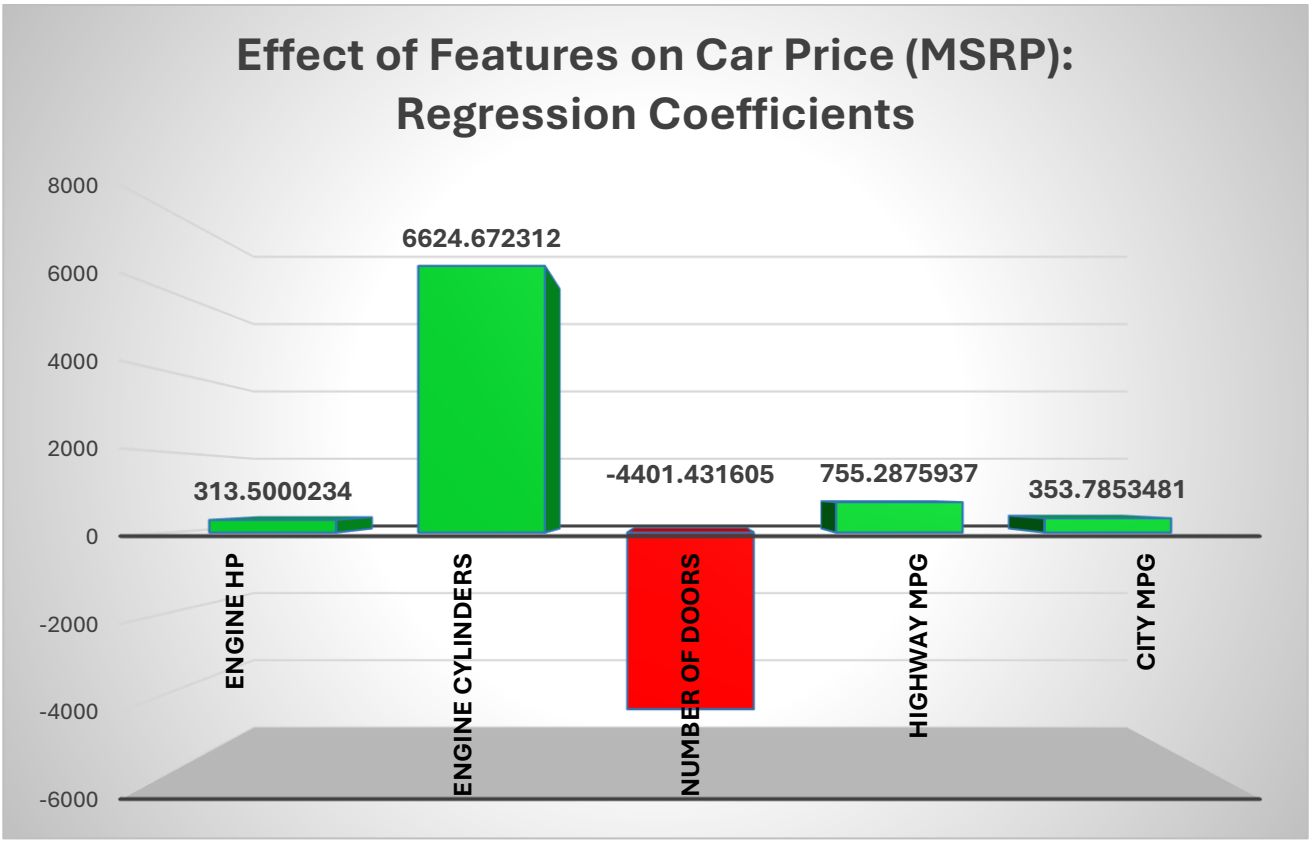


Insights-

- There's a clear positive relationship between **engine power** and **car price**. Bigger engines generally mean more expensive cars.
- While regular engines follow a steady price rise, very high engine powers (high-performance or exotic models) drive prices up sharply — showing a luxury/performance premium.
- The exponential model fits better **($R^2 = 0.592$)** than the linear one **($R^2 = 0.434$),** meaning price increases aren't just steady, they accelerate as engine power increases.
- This suggests manufacturers (and buyers) value extreme performance disproportionately, paying far more for small jumps in horsepower at the top end.

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.



Insights-

- **Engine Cylinders** are our strongest predictor. **More cylinders** result much **higher price.**
- More horsepower significantly raises price. This is because powerful engines cost more.
- **Highway MPG** & **City MPG** have slight positive impact i.e. fuel-efficient cars tend to price a bit higher, possibly due to advanced technology.
- Surprisingly, more doors are linked to lower prices, possibly because high-end sports and luxury cars often have fewer doors.

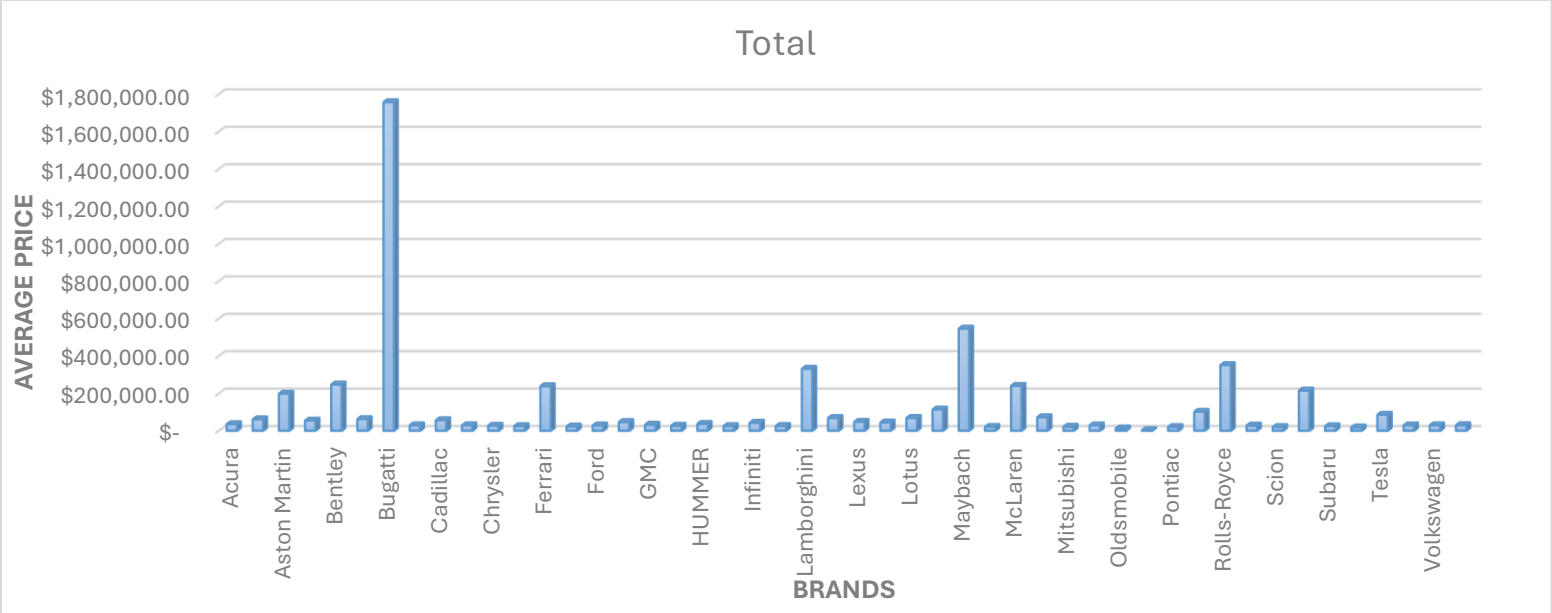
Insight Required: How does the average price of a car vary across different manufacturers?

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

Row Labels	Average of MSRP
Acura	$ 35,087.49
Alfa Romeo	$ 61,600.00
Aston Martin	$ 198,123.46

Audi	$	54,574.12
Bentley	$	247,169.32
BMW	$	62,162.56
Bugatti	$	1,757,223.67
Buick	$	29,034.19
Cadillac	$	56,368.27
Chevrolet	$	29,074.73
Chrysler	$	26,722.96
Dodge	$	24,857.05
Ferrari	$	238,218.84
FIAT	$	22,670.24
Ford	$	28,511.31
Genesis	$	46,616.67
GMC	$	32,444.09
Honda	$	26,655.15
HUMMER	$	36,464.41
Hyundai	$	24,926.26
Infiniti	$	42,640.27
Kia	$	25,513.76
Lamborghini	$	331,567.31
Land Rover	$	68,067.09
Lexus	$	47,549.07
Lincoln	$	43,860.83
Lotus	$	68,377.14
Maserati	$	113,684.49
Maybach	$	546,221.88
Mazda	$	20,416.62
McLaren	$	239,805.00
Mercedes-Benz	$	72,069.53
Mitsubishi	$	21,340.56
Nissan	$	28,921.15
Oldsmobile	$	12,843.80
Plymouth	$	3,296.87
Pontiac	$	19,800.04
Porsche	$	101,622.40
Rolls-Royce	$	351,130.65
Saab	$	27,879.81
Scion	$	19,932.50
Spyker	$	214,990.00
Subaru	$	24,240.67
Suzuki	$	18,026.42
Tesla	$	85,255.56
Toyota	$	28,846.56
Volkswagen	$	28,978.52
Volvo	$	29,724.68

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

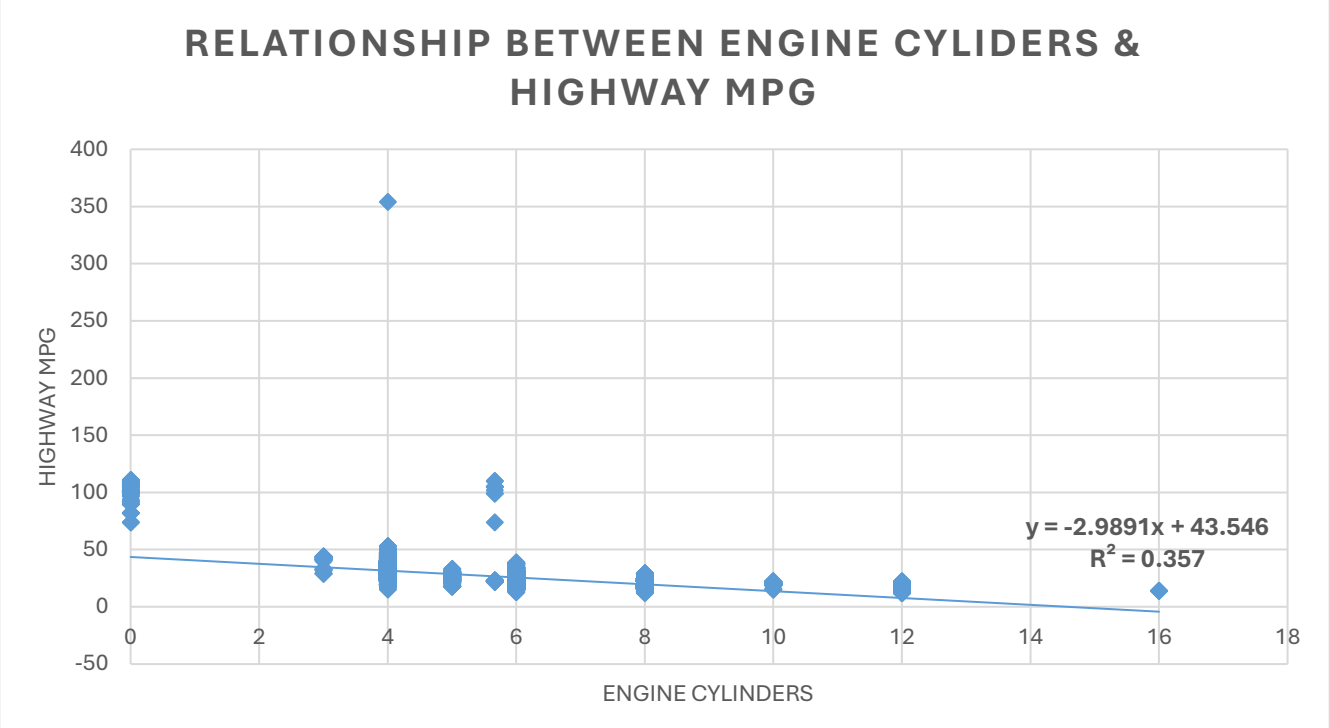


Insights-

- **Bugatti** leads with an ultra-high average price **($1.75M),** far above all others.
- **Maybach**, **Rolls-Royce**, **Lamborghini**, **McLaren**, **Ferrari**, **Bentley**, **Aston Martin** follow as top **luxury brands ($200K–$546K).**
- **Porsche**, **Tesla**, **Mercedes**, **Land Rover**, **Lexus**, **Lincoln** comes in the premium range **($43K–$101K)**.
- **Mainstream brands** like **Toyota**, **Honda**, **Ford**, **Hyundai**, **Chevrolet** average **$24K–$30K**, offering affordability.
- Budget brands (**Mazda**, **Mitsubishi**, **Scion**, **Suzuki**, **FIAT**) cluster **below $22K**.
- The market shows a huge price gap between **ultra-luxury** and **mainstream brands**, reflecting diverse positioning.

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.



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

Correlation coefficient between Engine cylinders & Highway MPG is -0.599853

- **No. of cylinders in engine** & **highway MPG** (miles per gallon) shows negative correlation **~0.6** which means, as the number of engine cylinders **increases**, highway MPG tends to **decrease**.
- From which we can conclude that bigger engines (more cylinders) generally consume more fuel per mile, making them less fuel-efficient on highways.

Insights-

- **Electric vehicles (0 cylinders)** top the chart with the highest average highway MPG i.e. around **99.7 MPG**, making them the most efficient by far.
- **3-cylinder cars** follow with about **39.4 MPG**, offering strong fuel economy for small, lightweight vehicles.
- **4-cylinder cars** are the most common, averaging **31.5 MPG**, providing a good balance of power and efficiency.
- As cylinders **increase**, fuel efficiency steadily **drops.**
- With every jump in cylinder count, fuel efficiency drops by **~3 MPG**, showing that bigger, more powerful engines demand more fuel on highways.
- If efficiency is our goal, fewer cylinders (or no cylinders / electric) are the clear winners. But if power and performance matter more, then we have to sacrifice miles per gallon.

Dashboard Tasks-

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

Observations-

- High total price in **Convertibles, Coupes, and 4dr SUVs** often overlaps with luxury brands, making these styles ideal for **premium positioning**.
- **Wagons, Minivans, and Cargo Vans** register lower totals — they serve utility needs, not luxury aspirations.

Insight-

- If our focus is **market share**, then we have to follow Chevrolet/Toyota model which is strong in volume-driven segments like Sedans and SUVs.
- If we're focusing in **luxury branding**, leverage high-value segments like **Convertibles** and **Coupes** but ensure there's enough product breadth to matter.
- For **competitive benchmarking**, map brand-body combinations not just on price points but on **total value contribution** — that's where business impact lies.

Task 2: Which car brands have the highest and lowest average MSRPs, and how does this vary by body style?

Observations-

- Luxury brands (e.g., **Rolls-Royce, McLaren, Maybach**) tend to focus on high-MSRP styles like **Convertibles, Coupes,** and **Sedans.**
- Budget brands (e.g., **Suzuki, Plymouth, Mitsubishi**) are found mostly in lower-cost body styles such as **Hatchbacks** and **Vans**.
- Some brands (e.g., **BMW**, **Audi**, **Mercedes**) span across mid-to-high MSRP styles, offering both sporty and executive options.

Insight-

- If we are aiming for premium market segments, then focus on brands like **Bugatti**, **Rolls-Royce**, **McLaren** with body styles like **Convertibles** and **Coupes**.
- For value-seeking consumers, targeting **Mazda**, **Mitsubishi**, or **Suzuki** models in **Hatchbacks** or **Pickups** aligns with budget preferences.
- Market positioning analysis should always consider both brand tier and body style — together, they shape perceived and actual vehicle value.

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

Observations-

- **Automated_manual** and **Direct_Drive** suggest luxury/performance or EVs.
- **Manual** points to budget or enthusiast performance.
- **Automatic** serves all tiers, from low-cost family vehicles to high-end models.
- **SUVs** and **Sedans** show the widest MSRP range across transmissions, reflecting market segmentation (economy to luxury).

- Meanwhile, styles like **Hatchbacks** and **Vans** remain tightly tied to low-cost transmissions (**Manual/Automatic** only).

Insight-

- If we are targeting premium or EV markets, then we have to favour **Automated_Manual** or **Direct_Drive** in **Coupes**, **Sedans**, and **SUVs**.
- Using **Manual** or **Automatic** in **Minivans**, **Hatchbacks**, and **Pickups** can help us to retain price accessibility.
- Offering **Manual** in **Convertibles** or **Coupes** creates niche appeal for performance enthusiasts at a slightly lower MSRP than automatic peers.

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

Observations-

- Vehicles like **2dr Hatchbacks**, **4dr Hatchbacks**, and **Sedans**, typically show **highway MPG > 30**, **city MPG ~24–27**, reflecting aerodynamic and lightweight design.
- **4dr SUVs** and **Minivans** tend to sit in the **25–29 MPG highway** and **18–22 MPG city** range. These styles trade off fuel efficiency for space and utility.
- styles such as **Crew Cab Pickups**, **Regular Cab Pickups**, and **Cargo Vans**, shows **highway MPG ~17–22** and **city MPG as low as ~13–15**, due to heavier weight and utility-focused engines.
- Vehicles during Early 90s (1990–1994) were already relatively fuel-efficient due to smaller average sizes and less power-demanding tech.
- In **Late 1990s to early 2000s** there is a slight dip in fuel efficiency (~27–29 highway, ~20–22 city), possibly due to the **SUV boom** and **engine upsizing** trends.
- **2000s onwards,** efficiency generally improves again, likely reflecting mission regulations, Hybrid technologies, Improved engine designs etc.

Insight-

- Smaller, lighter body styles like **Hatchbacks** and **Sedans** consistently offer better fuel economy, while **pickups**, **vans**, and **SUVs** sacrifice efficiency for size and utility.
- Fuel efficiency has fluctuated over the decades, dipping slightly during the SUV craze, but generally improving again due to technological advances and regulations.

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

Observations-

- **Bugatti, Maybach, Rolls-Royce, Lamborghini, McLaren, Ferrari, Aston Martin**, all are luxury-performance or exotic car brands. They show, Extreme HP, very low MPG, very high MSRP. These brands prioritize performance and prestige over efficiency or affordability.

- **BMW, Audi, Lexus, Mercedes-Benz, Cadillac, Land Rover, Infiniti**, all are positioned as luxury yet practical brands having reasonable **MPG (~25–30)**, **HP (~250–350).** Prices are high but not exotic **(typically $40K–$70K).**
- **Honda, Toyota, Hyundai, Subaru, Nissan, Mazda, Volkswagen**. These brands focus on affordable, fuel-efficient daily drivers, ideal for practical consumers.
- Tesla show **very high MPG** (due to EV efficiency ratings), **Moderate HP (~300),** but still high-performing due to instant torque. MSRP is relatively high but smaller than exotic brands.
- **Chevrolet, GMC, Ford, Dodge**. These brands cater to work-focused or utility-first consumers, where performance and towing trump fuel economy.

Insights-

- Luxury-performance brands dominate on power and price, but suffer on MPG.
- Mainstream brands win on fuel efficiency and affordability, especially in compact and sedan segments.
- EVs like **Tesla** break the traditional trade-offs by offering performance with unmatched efficiency.

Results-

We used pivot tables, scatter plots, regression bar charts, and interactive Excel dashboards to show:

- How popularity varies by market category.
- How engine power drives price.
- Which features (like cylinders, horsepower) most impact MSRP.
- Brand pricing gaps from budget to ultra-luxury.
- The strong negative link between cylinders and fuel efficiency.

❖ Discussion of Results

- More cylinders and horsepower → higher prices, lower MPG.
- Luxury brands dominate on power and price; mainstream brands win on affordability and efficiency.
- EVs stand out with unmatched fuel economy.
- Body style strongly influences price positioning (e.g., Convertibles vs. Minivans).

❖ Limitations

- No inflation adjustments over years.
- Some missing values imputed.
- Excel dashboards have limited interactivity.
- Dataset may not cover all markets or rare models.

❖ Future Directions

- Adjust prices for inflation.
- Expand data (add safety, resale value).
- Build advanced dashboards (Power BI, Tableau).

Google Drive Links

Python code used for cleaning dataset -
Data cleaning using Python

Cleaned Dataset (advised to open in excel) -
Cleaned Car Dataset