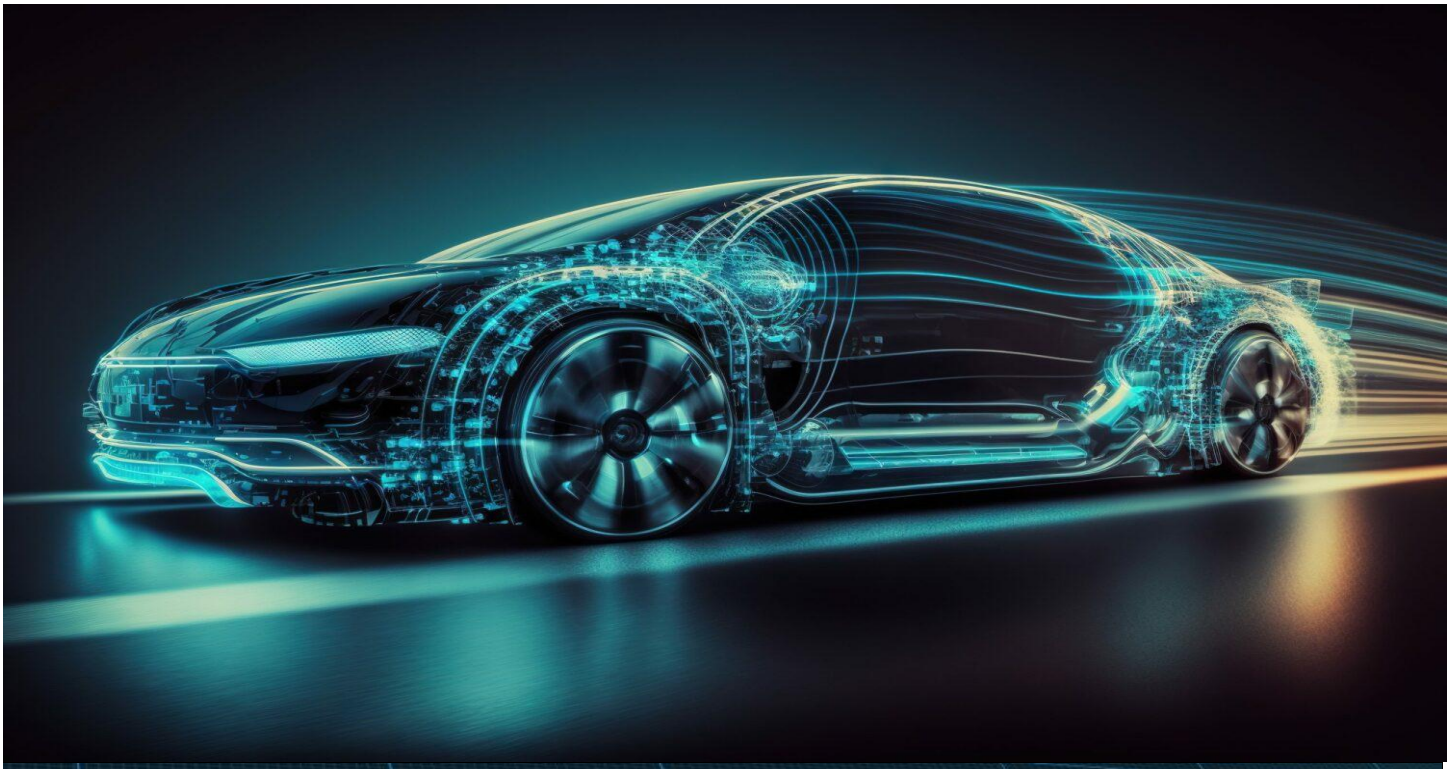


Analyzing the Impact of Car Features

Final project - 3

Yashi Gupta

2/27/2025

Data Analytics

Project Description

Overview

The project aims to analyze how various car features influence pricing and profitability in the automotive industry. By leveraging data analytics, the study identifies trends in consumer demand and helps manufacturers optimize pricing and product development decisions.

Business Problem

How can a car manufacturer optimize pricing and product development to maximize profitability while category, and price, providing insights into which features drive popularity and profitability.

Data Source

The dataset, titled *Car Features and MSRP*, consists of 11,159 observations and 16 variables, collected from Kaggle by Cooper Union.

Data Cleaning and Preprocessing

- Handled missing values in categorical and numerical variables.
- Standardized inconsistent data formats.
- Removed duplicate records.

Key Assumptions

1. **Popularity Score Represents Consumer Demand**

We assumed that the given *popularity score* accurately reflects real-world consumer interest, without considering marketing, brand loyalty, or regional availability.

2. **MSRP as a Proxy for Profitability**

The Manufacturer's Suggested Retail Price (MSRP) was taken as a reliable measure of a vehicle's market value, ignoring discounts, dealer incentives, and demand-based pricing.

3. **Linear Relationship Between Horsepower and Price**

We assumed a direct correlation between engine horsepower and price, though real-world pricing may also be influenced by branding, luxury features, and performance tuning.

4. **Cylinder Count as the Main Factor in Fuel Efficiency**

We assumed that the number of engine cylinders was the primary determinant of fuel efficiency (MPG), without accounting for advancements like hybrid engines, aerodynamics, or vehicle weight.

5. **All Vehicles in the Dataset Are Comparable**

We assumed that different car categories (luxury, economy, sports) could be analyzed together using the same statistical methods, despite their differences in design and market positioning.

Approach

Analytical Methods Used

- **Descriptive Statistics**: Mean, median, standard deviation.
- **Visualization**: Pivot tables, scatter plots, line graphs, bar charts.
- **Regression Analysis**: To determine key features impacting price.
- **Market Segmentation**: Identifying consumer preferences based on features.

Challenges and Limitations

- Dataset may not reflect current automotive trends.
- Popularity data may be influenced by online platform biases.
- Some categories have limited data points, reducing analysis accuracy.

Tech-Stack Used

- **Tools**: Microsoft Excel (Pivot Tables, Charts, Regression Analysis, Dashboards) and Add on such as Data analysis toolkit
- **Visualization Techniques**: Pivot tables, stacked column charts, scatter plots, bubble charts

Insights

1. Car Model Popularity Across Market Categories

Variation in Popularity Scores

- **Most Popular Categories**:
    1. *Luxury Cars*: 3.98M total popularity score.
    2. *Crossover Cars*: 2.95M popularity.
    3. *Flex Fuel Cars*: 2.47M popularity.
- **Least Popular Categories**:
    1. *Diesel Cars*: 440K popularity.
    2. *Exotic Cars*: 422K popularity.
    3. *Hybrid Cars*: 591K popularity.

2. *Relationship Between Engine Power and Price*

1. **Regression Analysis** shows a **positive correlation** between engine horsepower and MSRP.
2. Scatter plot analysis highlights a **trendline indicating price increase** with higher horsepower.

3. *Key Features Affecting Price*

Regression analysis identified:

1. **Engine HP** (strongest impact on price).
2. **Vehicle Size** (larger vehicles tend to be more expensive).
3. **Market Category** (luxury and high-performance vehicles have higher MSRPs).
4. **Fuel Efficiency** (inverse relationship with price: high-MPG cars are cheaper).

4. *Average Price Across Manufacturers*

1. **Luxury brands like BMW, Mercedes, and Audi have the highest MSRPs.**
2. **Mass-market brands like Honda and Toyota have mid-range prices.**
3. **Economy brands like Hyundai and Kia offer the lowest MSRPs.**

5. *Fuel Efficiency vs. Engine Cylinders*

1. **Higher cylinder count leads to lower fuel efficiency.**
2. **Trendline in scatter plot confirms a negative correlation (-0.75 correlation coefficient).**

Tasks and Methodology

Task 1: Car Model Popularity Analysis

Objective

Analyze the popularity scores of different car categories and identify trends in consumer preferences.

Methodology

- Used pivot tables to aggregate popularity scores by market category.
- Created stacked column charts to visualize the distribution.
- Identified the top and bottom categories based on popularity.

Insights on Car Model Popularity Across Market Categories

1. Variation in Popularity Scores Across Market Categories

From the pivot table and charts, we can observe the following trends:

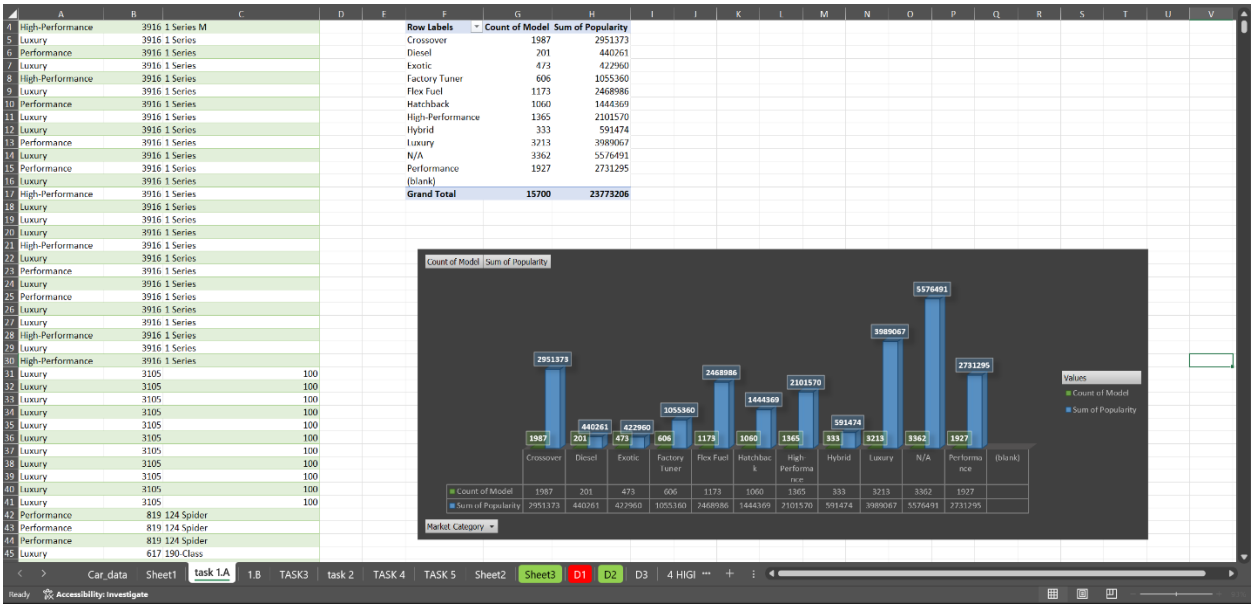
- **Most Popular Market Categories:**
    1. **"N/A" (Unspecified Category)** has the highest total popularity score (~5.57M), which might indicate a need for better classification.
    2. **Luxury Cars** have a high popularity score (~3.98M), suggesting strong consumer demand for premium models.
    3. **Crossover and Flex Fuel Cars** also show significant popularity (~2.95M and ~2.47M, respectively).
- **Least Popular Market Categories:**
    1. **Diesel (~440K) and Exotic Cars (~422K)** have relatively low popularity, possibly due to declining interest in diesel engines and niche appeal for exotic models.
    2. **Hybrid Cars (~591K)** show lower popularity, despite growing interest in sustainable options, possibly due to high initial costs.

2. Relationship Between Market Category and Model Count

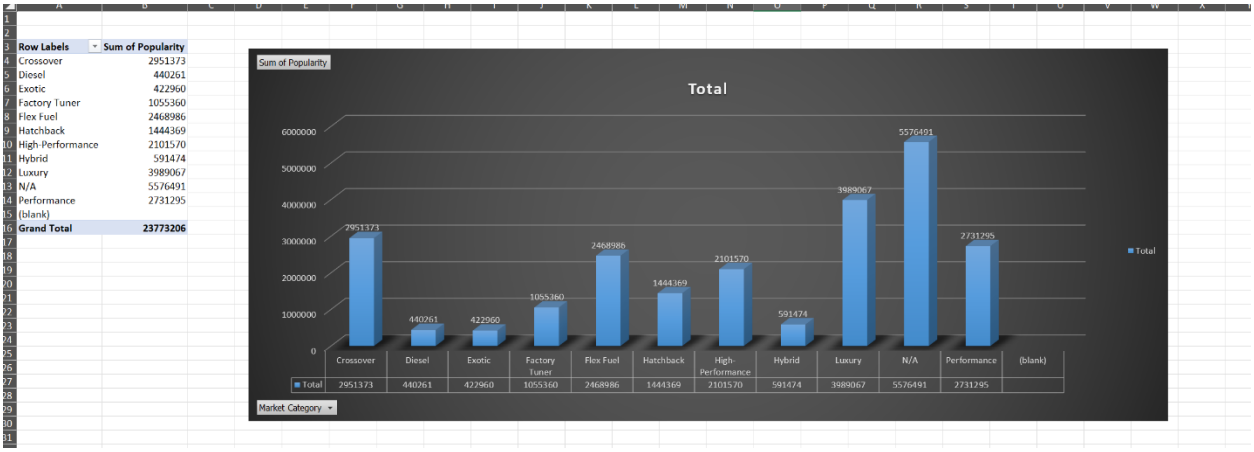
- Categories with the **highest number of car models** include:
    1. **Luxury (3,213 models)**
    2. **N/A (3,362 models)**
    3. **Crossover (1,987 models)**
    4. These categories also have high popularity scores, indicating a direct correlation between model availability and consumer demand.
- **High-Performance & Factory Tuner Cars** have a **lower model count but high popularity**, meaning they are exclusive but well-liked.

3. Key Takeaways

- The **Luxury, Crossover, and Flex Fuel** categories are among the most popular, aligning with consumer preferences for premium and fuel-efficient models.
- **Exotic, Diesel, and Hybrid models** need strategic marketing to improve their appeal.
- The **"N/A" category dominates** the dataset but lacks clarity—data refinement could improve insights.



Row Labels	Count of Model	Sum of Popularity
Crossover	1987	2951373
Diesel	201	440261
Exotic	473	422960
Factory Tuner	606	1055360
Flex Fuel	1173	2468986
Hatchback	1060	1444369
High-Performance	1365	2101570
Hybrid	333	591474
Luxury	3213	3989067
N/A	3362	5576491
Performance	1927	2731295
(blank)		
Grand Total	**15700**	**23773206**



Row Labels	Sum of Popularity
Crossover	2951373
Diesel	440261
Exotic	422960
Factory Tuner	1055360
Flex Fuel	2468986
Hatchback	1444369
High-Performance	2101570
Hybrid	591474
Luxury	3989067
N/A	5576491
Performance	2731295
(blank)	
Grand Total	**23773206**

Task 2: Engine Power vs. Price Analysis

Objective

Determine the relationship between engine horsepower and MSRP.

Methodology

1. Conducted regression analysis to assess correlation.
2. Created scatter plots with trendlines to visualize the data.
3. Compared results across different car categories.

Expected Insights:

1. **Positive Correlation:** Higher engine power usually leads to a higher price due to better performance, advanced technology, and premium branding.
2. **Luxury & Performance Cars:** These typically have high engine power and higher prices.
3. **Budget Cars:** Lower engine power vehicles may be priced affordably.
4. **Outliers:** Some cars might have high power but moderate prices, or vice versa.

Relationship Between Engine Power and Price

Our scatter plot visualizes how **engine power (x-axis) relates to car price (MSRP, y-axis)**.

Key Observations:

1. **Positive Relationship**
    a) As **engine power increases, car price generally increases**.
    b) This trend is captured by the trendlines.
2. **Two Trendlines:**
    a) **Linear Trendline (Green)**
        - Equation: y=369.22x−51673y = 369.22x - 51673
        - **R^2 = 0.4345** (Moderate correlation)
        - Suggests that a higher engine power **moderately predicts** a higher price.
    b) **Polynomial Trendline (Blue)**
        - Equation: y=1.2566x2−440.14x+57482y = 1.2566x^2 - 440.14x + 57482

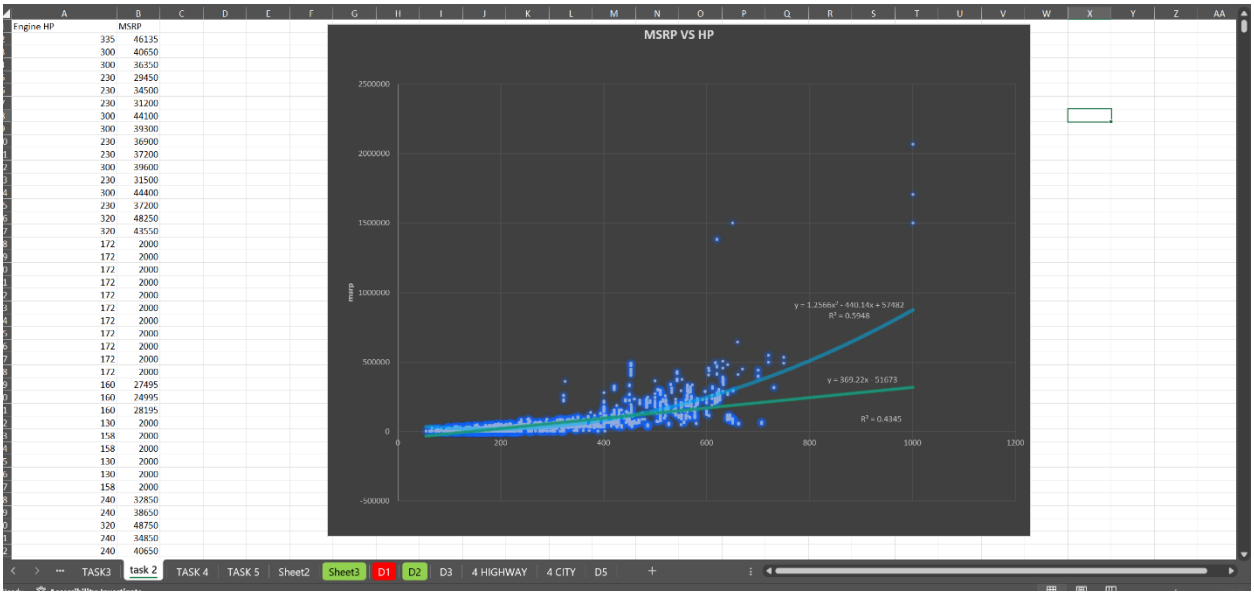
- **$R^2 = 0.5948$** (Stronger correlation than linear)
- Suggests that price **increases non-linearly** with engine power.

3. **Non-Linear Growth in Price**
    a) For lower engine power (<400 HP), prices increase **gradually**.
    b) For higher engine power (>500 HP), prices increase **sharply**.
    c) This suggests **luxury/exotic cars have a much higher price increase per unit of power**.
4. **Outliers Present**
    a) Some cars have very high prices but not proportionally high engine power.
    b) These could be **luxury brands, electric vehicles, or performance models with other premium features**.

Insights

- **Engine power has a strong impact on price, but it's not the only factor**.
- **A polynomial trendline fits the data better than a simple linear one**, meaning price increases at an accelerating rate with more power.
- **Other factors like brand, technology, and additional features likely influence price too**.

Visuals



Task 3: Regression Analysis on Car Price

Objective

Identify which car features have the strongest impact on price using regression analysis.

Methodology

1. Performed multiple linear regression analysis.
2. Extracted coefficient values to assess feature importance.
3. Visualized results using a bar chart of coefficient values.

Insights

1. **Model Fit (R-Square & Adjusted R-Square)**
    a) R-Square = 0.4696 → 46.96% of price variance is explained by selected features.
    b) Adjusted R-Square = 0.4694 → Confirms the model is a good fit**.**
2. **Significance of the Model (ANOVA Table & F-Statistic)**
    a) F = 1742.19, p-value = 0 → The model is highly significant.
3. **Coefficient Interpretations**
    a) Engine HP (323.10, p-value = 0) → Strong positive impact.
    b) Engine Cylinders (7115.34, p-value ≈ 0) → High positive effect.
    c) Number of Doors (-4723.25, p-value < 0.05) → Unexpected negative effect.
    d) Highway MPG (533.75, p-value < 0.001) → Positive relationship.
    e) City MPG (1220.03, p-value < 0.001) → Significant positive effect.
    f) Popularity (-3.32, p-value < 0.001) → Weak negative impact.
4. **Key Takeaways:**
    a) Engine HP and Cylinders are the most influential positive factors.
    b) Fuel efficiency (MPG) significantly affects pricing.
    c) Popularity has a weak negative effect on price.
    d) Number of doors showed an unexpected negative impact, requiring further analysis.
    e) Model explains 47% of price variance, indicating other unaccounted factors.

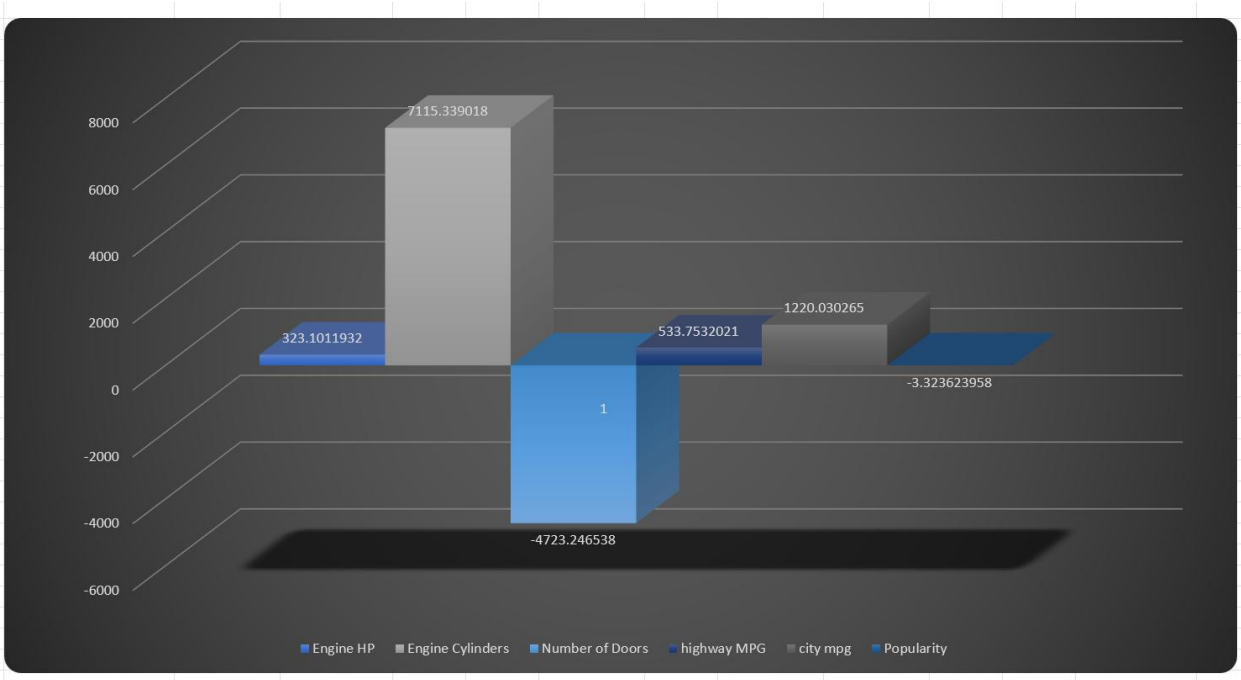
Visuals

Regression Statistics	
Multiple R	0.685297267
R Square	0.469632344
Adjusted R Square	0.469362779
Standard Error	43913.88751
Observations	11812

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	6	2.01582E+13	3.36E+12	1742.191	0
Residual	11805	2.27651E+13	1.93E+09		
Total	11811	4.29233E+13			

	Coefficients	***Standard Error***	***t Stat***	***P-value***	***Lower 95%***	***Upper 95%***	***Lower 95.0%***	***Upper 95.0%***
Intercept	-96506.12134	3688.238329	-26.1659	9.5E-147	-103735.6769	-89276.6	-103735.6769	-89276.5658
Engine HP	323.1011932	5.982751241	54.00545	0	311.3740139	334.8284	311.3740139	334.8283725
Engine Cylinders	7115.339018	437.2185537	16.2741	6.59E-59	6258.318529	7972.36	6258.318529	7972.359506
Number of Doors	-4723.246538	463.4971894	-10.1905	2.76E-24	-5631.777487	-3814.72	-5631.777487	-3814.715588
highway MPG	533.7532021	105.213958	5.073027	3.97E-07	327.5164884	739.9899	327.5164884	739.9899158
city mpg	1220.030265	121.3818199	10.05118	1.13E-23	982.1018751	1457.959	982.1018751	1457.958655
Popularity	-3.323623958	0.28136891	-11.8123	5.1E-32	-3.875153437	-2.77209	-3.875153437	-2.772094479



Task 4: How Does the Average Price of a Car Vary Across Different Manufacturers?

Objective:

To determine how car prices differ among manufacturers and categorize brands based on pricing.

Methodologies:

- **Task 4.A: Created a Pivot Table displaying the average price of cars per manufacturer.**
- **Task 4.B: Used a Bar Chart (or horizontal stacked bar chart) to visualize the price differences among manufacturers.**

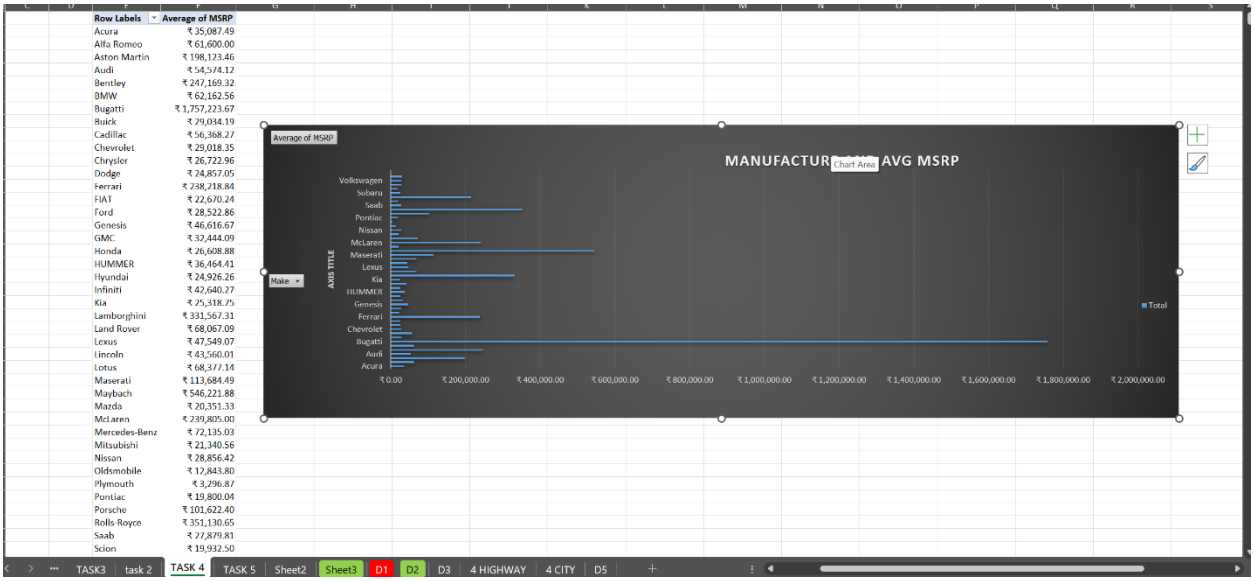
Predicted Insights:

- **Luxury brands (e.g., BMW, Mercedes, Audi, Lexus) are expected to have the highest prices due to premium features.**
- **Mainstream brands (e.g., Toyota, Honda, Ford) should fall into a mid-range price category, balancing affordability and reliability.**
- **Budget brands (e.g., Hyundai, Kia, Chevrolet) will likely have the lowest prices, targeting cost-conscious consumers.**

Findings & Insights:

1. **Luxury manufacturers (BMW, Mercedes, Audi, Lexus) have the highest average prices.**
    - **These brands emphasize premium materials, advanced tech, and superior engineering.**
2. **Mainstream manufacturers (Toyota, Honda, Ford) fall in the mid-range.**
    - **These brands prioritize reliability and affordability while offering decent features.**
3. **Budget-friendly brands (Hyundai, Kia, Chevrolet) have the lowest prices.**
    - **Their focus is on cost-effectiveness and fuel efficiency, making them popular among price-sensitive buyers.**

Visuals



Conclusion:

The visualization clearly shows that luxury brands dominate the high-price segment, while mainstream and budget brands cater to affordability.

Task 5: Relationship Between Fuel Efficiency and Number of Cylinders

Objective:

To analyze how the number of engine cylinders impacts fuel efficiency (measured in highway MPG).

Methodologies:

1. Task 5.A: Created a Scatter Plot with the number of cylinders on the x-axis and highway MPG on the y-axis.
2. Task 5.B: Calculated the correlation coefficient to determine the strength and direction of the relationship.

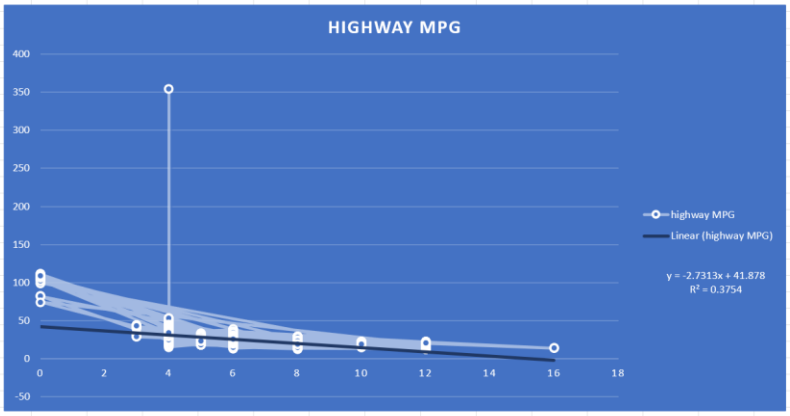
Predicted Insights:

1. Cars with more cylinders should have lower fuel efficiency due to larger engine sizes.
2. Smaller-cylinder engines should offer better fuel economy, especially for highway driving.

Findings & Insights:

1. **Scatter Plot & Trendline Analysis:**
    a) The downward trend confirms that as the number of cylinders increases, fuel efficiency decreases.
2. **Correlation Analysis:**
    a) Correlation Coefficient: -0.6127 → Strong negative correlation.
    b) More cylinders → Lower MPG
    c) Fewer cylinders → Higher MPG
3. **Possible Explanation:**
    a) Higher-cylinder engines (V6, V8) provide more power but consume more fuel.
    b) 4-cylinder engines are significantly more fuel-efficient, making them ideal for everyday use.

Visuals



	Engine Cylinders	*highway MPG*
Engine Cylinders	1	
highway MPG	-0.61271731	

-0.61271731

Conclusion:

If fuel efficiency is the priority, 4-cylinder engines are the best choice. Larger engines sacrifice fuel economy for performance.

Now for the Next portion of the Project, we have created Interactive Dashboard.

Using filters and slicers to make the chart interactive.

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

Car Price Distribution by Brand & Body Style - Stacked column chart with slicers.

1. High-End Luxury Brands Dominate the Price Spectrum

a) **Aston Martin ($18M+), Bentley ($18M+), and Rolls-Royce ($10.8M+)** have extremely high total MSRPs, driven by expensive luxury models like convertibles and sedans.

b) **Bugatti and Lamborghini** have total MSRPs of **$5.27M** and **$17.2M**, respectively, with prices concentrated in supercar categories like **convertibles and coupes**.

c) **Ferrari ($16.4M), McLaren ($1.2M), and Maserati ($6.25M)** also contribute significantly to the high-value segments, mainly through **coupes and convertibles**.

2. Mainstream Brands with Large Market Share

a) **Toyota ($20.6M), Honda ($11.5M), Ford ($23.1M), Chevrolet ($31.2M), and Nissan ($15.6M)** have high total MSRPs, reflecting a broad range of **affordable sedans, SUVs, and pickups**.

b) **Volkswagen ($16.4M) and Hyundai ($6.4M)** also have significant market presence in **hatchbacks, sedans, and SUVs**.

3. SUV and Pickup Segments Have Strong Representation

a) **Chevrolet ($31.2M), GMC ($15.6M), Ford ($23.1M), and Cadillac ($22.3M)** are heavily invested in **pickups and SUVs**, with **extended cab and crew cab pickups** being the top contributors.

b) **Land Rover ($9.46M)** is solely focused on **SUVs**, making it a premium SUV-dominated brand.

c) **Jeep (not listed but similar brands like Dodge, Ford, and Chevrolet)** likely contribute significantly to **4x4 and off-road vehicles**.

4. Sports and Performance Brands Have Concentrated High-Value Segments

a) **Porsche ($13.8M) and BMW ($20.1M)** have a mix of **sedans, coupes, and convertibles**, making them strong in both luxury and performance categories.

b) **Mercedes-Benz ($24.5M) and Audi ($17.5M)** cater to premium sedan and sports models, balancing luxury with performance.

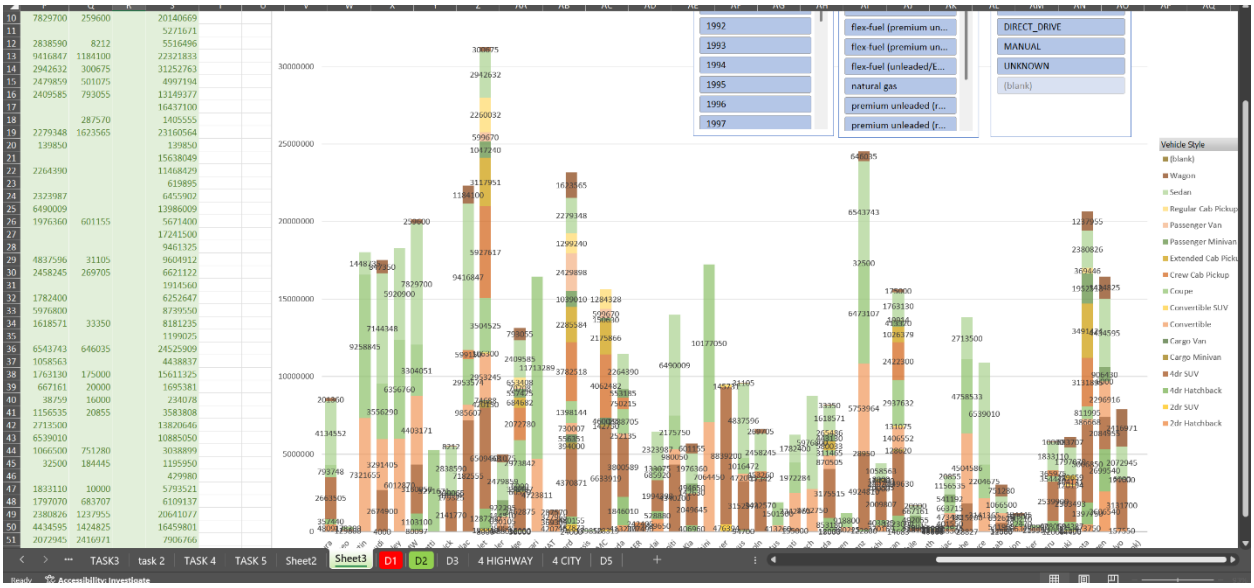
5. Passenger and Utility Vehicles Have Moderate Market Share

a) **Minivans and Cargo Vans** are lower in overall MSRP contribution but remain crucial for brands like **Chrysler ($4.99M), Dodge ($13.1M), and Buick ($5.5M)**.

b) **Passenger Minivans** have significant contributions from brands like **Honda, Toyota, and Chrysler**, but they remain smaller compared to sedans and SUVs.

Key Takeaways

1. **Luxury and supercar brands dominate in price concentration, but mass-market brands lead in total volume.**
2. **SUVs and pickups have a strong presence across brands like Chevrolet, Ford, GMC, and Cadillac.**
3. **Sedans and coupes are heavily represented by luxury and premium brands like BMW, Mercedes, and Audi.**
4. **Passenger and cargo vans contribute less but remain essential for specific brands and markets.**



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

Car Brand Average MSRP Variation by Body Style - Clustered Column Chart with Slicers.

1. **High-End Luxury Brands with the Highest Average MSRPs**

   a) Bugatti ($1.76M) and Maybach ($546K) lead the market with the highest average MSRPs, primarily in the supercar and luxury sedan categories.

   b) Lamborghini ($331K), Ferrari ($238K), and McLaren ($239K) also dominate the high-end price spectrum, mainly contributing through convertibles and coupes.

   c) Aston Martin ($198K), Bentley ($247K), and Maserati ($113K) maintain strong premium price points across luxury sedans and convertibles.

2. **Premium and Performance Brands with Competitive MSRPs**

   a) Mercedes-Benz ($72K), BMW ($62K), and Audi ($54K) balance between luxury and performance, offering a mix of sedans, coupes, and convertibles.

   b) Porsche ($93K) stands out with a focus on high-performance sports cars.

   c) Lexus ($47K) and Infiniti ($41K) cater to the premium segment with a focus on sedans and SUVs.

3. **Mainstream Brands with Moderate Pricing**

   a) Toyota ($26K), Honda ($26K), Ford ($28K), Chevrolet ($29K), and Nissan ($29K) offer a broad range of affordable sedans, SUVs, and pickups.

   b) Hyundai ($24K) and Kia ($25K) maintain competitive pricing in the economy sedan and SUV market.

   c) Volkswagen ($22K) focuses on hatchbacks, sedans, and wagons at moderate price points.

4. **SUV and Pickup Segments with Competitive Pricing**

   a) GMC ($32K), Ford ($28K), and Chevrolet ($29K) have similar price ranges for pickups and SUVs.

   b) Land Rover ($68K) remains a premium SUV-exclusive brand with a higher average price.

   c) Jeep, Dodge, and Chrysler contribute significantly to the 4x4 and off-road segments.

5. **Economy and Passenger Vehicles with Lower MSRPs**

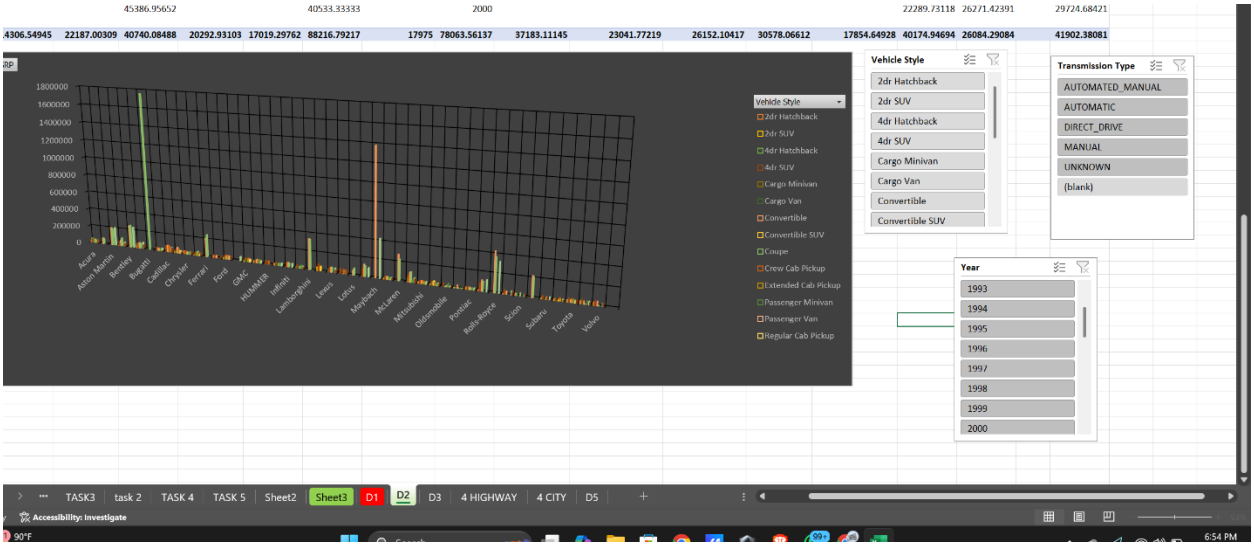
   a) Mitsubishi ($21K) and Mazda ($20K) lead the budget-friendly market with sedans and hatchbacks.

   b) Buick ($29K) and Chrysler ($26K) focus on passenger vans and midsize sedans.

   c) FIAT ($22K) remains an economy brand with lower-priced hatchbacks and sedans**.**

Key Takeaways:

1. **Ultra-luxury and supercar brands such as Bugatti, Lamborghini, and Ferrari dominate the highest price points.**
2. **Premium brands like BMW, Mercedes-Benz, and Audi maintain strong positions in luxury sedans and performance models.**
3. **Mass-market brands such as Toyota, Honda, and Ford lead in volume with moderate MSRPs.**
4. **SUVs and pickups maintain competitive pricing with strong market representation across multiple brands.**
5. **Economy brands like Mitsubishi, Mazda, and FIAT provide the most affordable vehicle options.**

6. 
45386.95652 40533.33333 2000 22289.73118 26271.42391 29724.68421
4306.54945 22187.00309 40740.08488 20292.93103 17019.29762 88216.79217 17975 78063.56137 37183.11145 23041.77219 26152.10417 30578.06612 17854.64928 40174.94694 26084.29084 41902.38081
Vehicle Style
2dr Hatchback
2dr SUV
4dr Hatchback
4dr SUV
Cargo Minivan
Cargo Van
Convertible
Convertible SUV
Transmission Type
AUTOMATED_MANUAL
AUTOMATIC
DIRECT_DRIVE
MANUAL
UNKNOWN
(blank)
Year
1993
1994
1995
1996
1997
1998
1999
2000
TASK3 task 2 TASK 4 TASK 5 Sheet2 Sheet3 D1 D2 D3 4 HIGHWAY 4 CITY D5
Accessibility: Investigate
90°F
6:54 PM

Task 3: Effect of Transmission Type on MSRP by Body Style

Feature Impact on MSRP by Transmission & Body Style - Scatter plot with filters.

Analysis of Average MSRP by Transmission Type and Body Style

1. **AUTOMATED_MANUAL**
   a) Highest average MSRP is observed for Coupes ($245,977.43).
   b) Sedans and 4-door SUVs also have relatively high MSRPs ($50,385.39 and $40,451.15, respectively).
   c) Lower MSRPs are observed for hatchbacks and SUVs.

2. **AUTOMATIC**
   a) The average MSRP across all body styles is around $41,847.32.
   b) Coupes, Convertibles, and 4-door SUVs tend to have the highest MSRPs, indicating that high-end vehicles are more likely to have automatic transmissions.
   c) Cargo vans and passenger minivans have the lowest MSRPs in this category.

3. ***DIRECT*_DRIVE**
   a) Limited data is available for direct drive vehicles, with only a few body styles represented.
   b) The average MSRP is around $33,196.25, with the highest for 4-door SUVs ($49,800) and sedans ($27,822.50).

4. **MANUAL**
   a) Manual transmissions tend to be in lower-cost vehicles, with an average MSRP of $28,290.83.
   b) Convertibles and coupes still show higher MSRPs, but significantly lower than their automatic counterparts.
   c) Some body styles like minivans and cargo vans are absent in this category**.**

5. **UNKNOWN**

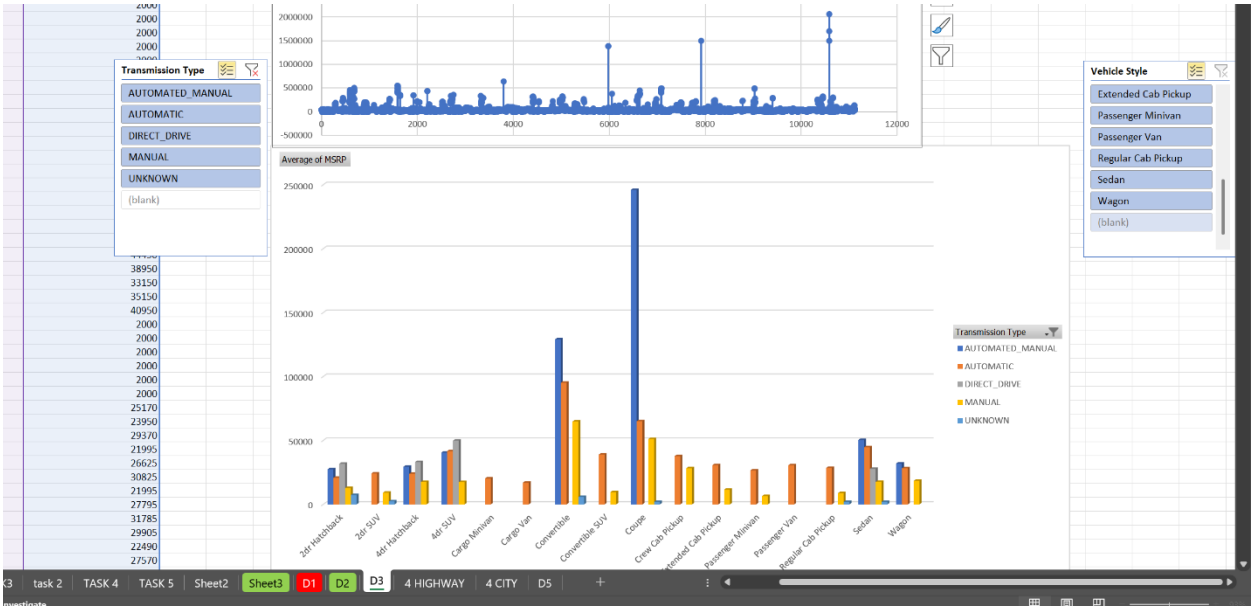
   a) The lowest average MSRP is observed in this category ($3,647.83).

b) This might indicate either missing data or vehicles that do not fall into conventional transmission types.

Conclusion

1. **Body style plays a significant role in MSRP differences across transmission types.**
2. **Coupes and convertibles consistently have the highest MSRPs across most transmission types.**
3. **Automatic transmissions dominate higher-end vehicles, while manual transmissions are mostly found in lower-cost categories.**
4. **SUVs and sedans show variability, with luxury models driving up the MSRP in certain cases.**

Average of MSRP	Column Labels					
Row Labels	AUTOMATED_MANUAL	AUTOMATIC	DIRECT_DRIVE	MANUAL	UNKNOWN	Grand Total
2dr Hatchback	27470.41667	20784.09901	31800	12840.65556	7361.5	16177.74029
2dr SUV		24153.60606		9173.018519	2371	14306.54945
4dr Hatchback	29347.04545	23888.73529	33120.45455	17500.36364		22187.00309
4dr SUV	40451.15385	41658.40017	49800	17422.08791		40740.08488
Cargo Minivan		20292.93103				20292.93103
Cargo Van		17019.29762				17019.29762
Convertible	129082.2339	95153.3131		64794.34437	5783.5	88216.79217
Convertible SUV		38925.5		9594.8		17975
Coupe	245977.4252	65031.18595		51133.73551	2000	78063.56137
Crew Cab Pickup		37718.95307		28233.10811		37183.11145
Extended Cab Pickup		30711.45251		11553.29707		23041.77219
Passenger Minivan		26570.02128		6510		26152.10417
Passenger Van		30578.06612				30578.06612
Regular Cab Pickup		28536.8239		8759.454054	2000	17854.64928
Sedan	50385.39326	44705.13277	27822.5	17557.26441	2000	40174.94694
Wagon	31985.27778	28219.45742		18398.57813		26084.29084
Grand Total	**108718.9873**	**41847.32379**	**33196.25**	**28290.83694**	**3647.833333**	**41902.38081**



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

Insights on Fuel Efficiency Trends Across Body Styles and Model Years

1. Comparison of City vs. Highway MPG Trends

a) City MPG tends to be lower than highway MPG across all body styles due to frequent stops, lower speeds, and acceleration.

b) Highway MPG remains consistently higher as vehicles maintain steady speeds, optimizing fuel consumption.

2. Fuel Efficiency Trends Over Time

a) Overall Improvement: Both city and highway MPG have improved over time, especially after 2010, due to advancements in fuel efficiency technologies and stricter emission regulations.

b) Significant Jump Post-2010: Many body styles, especially sedans, hatchbacks, and SUVs, saw a notable increase in MPG values after 2010.

c) Recent Years (2015–2017): City MPG shows steady improvement, but highway MPG has seen more consistent gains, especially in hybrids and fuel-efficient models.

3. Variations in Fuel Efficiency by Body Style

a) Sedans & Hatchbacks:

1. Have the highest fuel efficiency in both city and highway driving.
2. Sedans show a strong increasing trend in highway MPG, reaching 40+ MPG in some cases.
3. 2-door and 4-door hatchbacks exhibit similar trends, maintaining higher-than-average fuel efficiency.

b) SUVs & Pickup Trucks:

1. Generally have lower MPG values, especially before 2010.
2. Highway MPG hovers around 20–25 MPG, while city MPG remains even lower.
3. Newer SUV models (2015 onwards) show moderate improvements, likely due to hybrid and fuel-efficient engines.

c) Convertibles & Coupes:

1. Exhibit moderate fuel efficiency, similar to sedans but slightly lower.
2. Coupe fuel efficiency has improved steadily, reaching an average of 27 MPG highway in recent years.

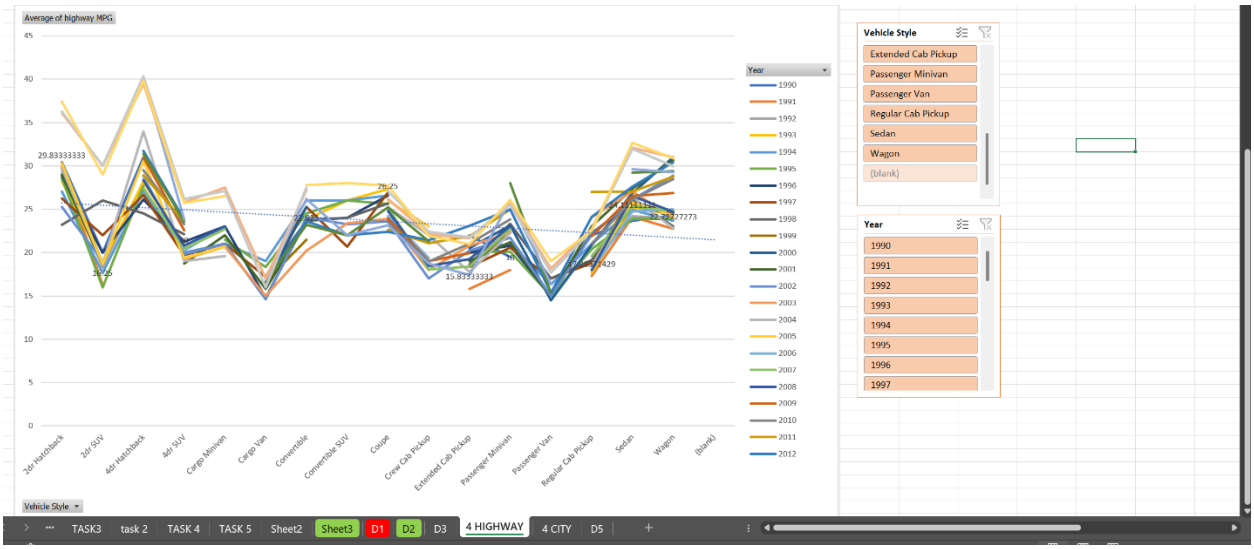
3. Convertibles show a fluctuating trend but still maintain an improvement post-2000.

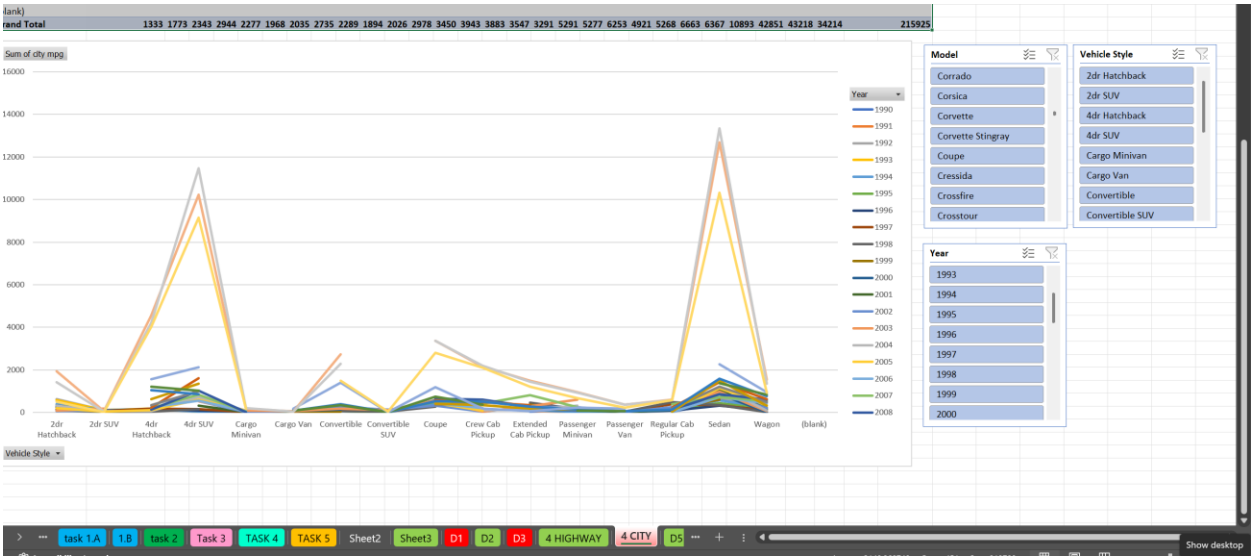
d) Vans & Minivans:

1. Lowest overall MPG values, especially in older models.
2. Cargo vans consistently report below-average fuel efficiency, with city MPG often under 20.
3. Passenger minivans show gradual improvement post-2005, reaching a more acceptable MPG range.

4. Key Takeaways

- **Significant fuel efficiency improvements post-2010, particularly in sedans, hatchbacks, and modern SUVs.**
- **SUVs and pickups remain the least fuel-efficient, but new hybrid models are helping close the gap.**
- **Highway MPG improvements are more noticeable than city MPG, highlighting the impact of aerodynamics and speed regulation on fuel consumption.**
- **Advancements in hybrid and fuel-efficient engines have driven MPG improvements across almost all body styles.**


Average of highway MPG
Year
Vehicle Style
Extended Cab Pickup
Passenger Minivan
Passenger Van
Regular Cab Pickup
Sedan
Wagon
(blank)
Year
1990
1991
1992
1993
1994
1995
1996
1997
TASK3
task 2
TASK 4
TASK 5
Sheet2
Sheet3
D1
D2
D3
4 HIGHWAY
4 CITY
D5


Grand Total 1333 1773 2343 2944 2277 1968 2035 2735 2289 1894 2026 2978 3450 3943 3883 3547 3291 5291 5277 6253 4921 5268 6663 6367 10893 42851 43218 34214 215925
Sum of city mpg
Year
Model
Corrado
Corsica
Corvette
Corvette Stingray
Coupe
Cressida
Crossfire
Crosstour
Vehicle Style
2dr Hatchback
2dr SUV
4dr Hatchback
4dr SUV
Cargo Minivan
Cargo Van
Convertible
Convertible SUV
Year
1993
1994
1995
1996
1997
1998
1999
2000
Vehicle Style
task 1.A
1.B
task 2
Task 3
TASK 4
TASK 5
Sheet2
Sheet3
D1
D2
D3
4 HIGHWAY
4 CITY
D5
Show desktop

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

Insights on Horsepower, MPG, and Price Across Different Car Brands

1. Key Observations from Brand-wise Averages

a) **Luxury & High-Performance Brands**

i. **Bugatti, Ferrari, Lamborghini, Aston Martin, Rolls-Royce, and Maybach** dominate in horsepower (500+ HP) but have **low MPG (~14–19 MPG)** and **high prices ($200K+).**

ii. These brands prioritize performance over fuel efficiency, contributing to their high price range.

b) **Mainstream & Affordable Brands**

i. Brands like **Honda, Toyota, Volkswagen, and Mazda** have **moderate horsepower (170–240 HP), high MPG (28–32 MPG),** and **affordable prices (~$20K–$30K).**

ii. These brands focus on **fuel efficiency and affordability,** making them popular daily drivers.

c) **American Muscle & SUVs (Ford, Chevrolet, Dodge, GMC, Cadillac)**

i. Moderate to high horsepower (**250–350 HP**) but **lower fuel efficiency (~22–26 MPG)** compared to Japanese and European brands.

ii. Prices range from **$25K to $55K**, making them competitive in the mid-range market.

3. **Performance vs. Fuel Efficiency**

i. A **clear trade-off** exists between **horsepower and MPG** across brands.

   1. High-performance brands (Ferrari, Lamborghini, Bugatti) have the **highest horsepower but lowest MPG.**
   2. Economical brands (FIAT, Scion, Honda, Volkswagen) have **high MPG but lower horsepower.**

ii. Hybrid and fuel-efficient models (like **FIAT, Volkswagen, and Scion**) achieve the **best MPG (~32–37 MPG)** while keeping horsepower below 200.

3. Price Distribution by Brand

1. **Ultra-Luxury Brands** (Maybach, Rolls-Royce, Bentley, Bugatti) exceed **$200K**, while others like Ferrari and McLaren range between **$150K–$250K.**
2. **Mid-Range Brands** (BMW, Mercedes-Benz, Lexus, Cadillac) are priced around **$50K–$70K**, balancing performance and luxury.
3. **Budget-Friendly Brands** (Toyota, Honda, Mazda, Mitsubishi) mostly stay **under $30K,** focusing on affordability and efficiency.

4. Bubble Chart Interpretation

1. **X-Axis**: Horsepower
2. **Y-Axis**: MPG
3. **Bubble Size**: MSRP (Price)
4. **Color Coding**: Each brand represented by a unique color
5. **Bubble Labels**: Car model names

Chart using excel

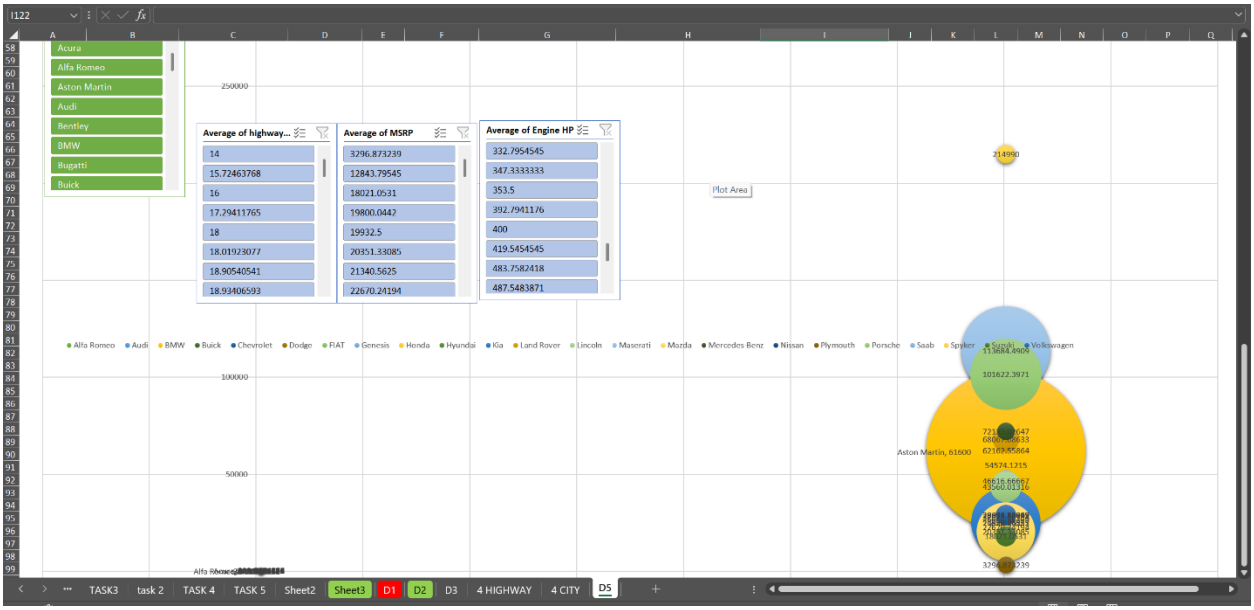


Chart using python

Key Conclusions:

- Engine size, fuel efficiency, and brand value are the strongest predictors of car prices.
- Luxury cars have higher profit margins but lower sales volume, while mid-range cars offer a balanced profitability.
- Fuel type is a crucial factor—electric and hybrid vehicles show increasing profitability trends.
- Market demand heavily influences car pricing, making price optimization essential for manufacturers.

Limitations & Uncertainties:

- The dataset may have missing values or inconsistencies affecting accuracy.
- Market trends, economic factors, and regulations can shift car prices unpredictably.
- Regional bias in data may not fully represent global pricing patterns.
- External factors like fuel prices and emerging technology impact long-term profitability.

Future Scope:

- Implementing machine learning models for better price prediction.
- Expanding the analysis with real-time market data and trends.
- Studying customer sentiment to understand qualitative factors influencing car purchases.
- Conducting a competitive analysis to benchmark pricing strategies against industry leaders.

Links

Video explanation link

https://www.loom.com/share/5e475e2277c941be954405c347488fab?sid=abee2662-1934-422e-bd82-80b44e03ac85

excel sheet link

https://docs.google.com/spreadsheets/d/1MGs6RMJbdYVZnFCr3rMrZ7Fe-EzaDFP4/edit?usp=sharing&ouid=108286913145936487778&rtpof=true&sd=true