Automotive Market Analytics- Price Optimization Study

Project Description

Overview:
This project analyses a car dataset to understand key factors influencing car pricing, popularity, and fuel efficiency. The goal is to provide actionable insights to help automotive manufacturers and marketers optimize pricing, feature development, and market segmentation.

Business Problem:
The project aims to answer critical questions such as:

- What drives the popularity of different car models?
- How do performance features like engine power and cylinders impact car price and fuel efficiency?
- Which manufacturers lead in pricing tiers, and how does market category influence consumer appeal?

Data Sources:
The analysis uses a comprehensive car dataset containing information on vehicle specifications (e.g., engine horsepower, cylinders, transmission), pricing (MSRP), popularity metrics, fuel efficiency, and manufacturer details.

Data Cleaning & Preprocessing:

- Missing values were identified in columns like Market Category, Engine HP, and Transmission.
- 715 duplicate records based on Make, Model, and Year were removed.
- "Unknown" values in Transmission were replaced by the most frequent mode value.
- "N/A" entries in Market Category were filled using a brand-wise most frequent category via Excel's INDEX + MATCH formulas to preserve brand-specific nuances.
- Ensured consistency and correctness across all categorical and numeric variables.

Assumptions:

- Missing transmission values can be reasonably imputed by the mode without biasing results.
- Brand-wise popular market categories are representative of missing entries.
- The dataset is assumed to be an accurate snapshot of the car market for the period analyzed.

Approach

Analytical Methods Used:

- Descriptive statistics to summarize central tendencies and variability of key variables.
- Visualization through histograms, scatter plots, and dashboards to identify trends and relationships.
- Correlation and regression analysis to quantify relationships between features and pricing.
- Pivot tables and Excel formulas for categorical data imputation and segmentation.

Reasoning Behind Choices:
Excel was chosen for its accessibility and powerful data manipulation capabilities suitable for exploratory data analysis (EDA). Regression analysis allowed us to understand how multiple features collectively influence car prices.

Modeling Techniques:

- Linear regression was applied to identify significant predictors of car price with an R-squared of 0.46, indicating moderate explanatory power.
- Correlation analysis assessed relationships such as cylinders versus fuel efficiency.

Challenges & Limitations:

- Some missing values required assumptions for imputation which may introduce minor bias.
- The model explains less than half of the price variability, suggesting other unmeasured factors affect pricing.
- Non-linear relationships and interaction effects could be explored further in future work.

Tech-Stack Used

Tools and Software:

- Microsoft Excel was the primary tool for data cleaning, analysis, and dashboard creation due to its wide availability and flexibility.

Reasoning for Tech Stack:

- Excel's user-friendly interface and powerful functions (INDEX, MATCH, pivot tables) made it ideal for managing categorical imputation and exploratory analysis without programming overhead.
- Visualization and dashboards in Excel helped communicate insights clearly.

Additional Libraries/Packages:

- No external libraries were used; all work was conducted within Excel.

Insight

Key Findings:

- Market Category & Popularity: Niche segments (e.g., Crossover + Performance) show higher popularity despite fewer models, indicating targeted market appeal.
- Engine Power & Price: A strong exponential relationship exists; higher horsepower corresponds to premium pricing, especially in luxury segments.
- Regression Analysis: Engine HP and number of cylinders significantly drive price, while more doors tend to decrease price.
- Manufacturer Pricing: Luxury brands (Bugatti, Rolls-Royce) command extremely high prices, while mainstream brands cluster at lower tiers. Tesla is notable for premium pricing within an electric vehicle context.
- Cylinder Count & Fuel Efficiency: Negative correlation (-0.60) shows more cylinders reduce highway MPG, aligning with fuel-conscious consumer priorities.

Business Relevance:

- Manufacturers can focus on niche market segments to boost model popularity.

- Pricing strategies should emphasize performance features for luxury market segments.
- Fuel efficiency considerations favor smaller engines for mainstream markets.

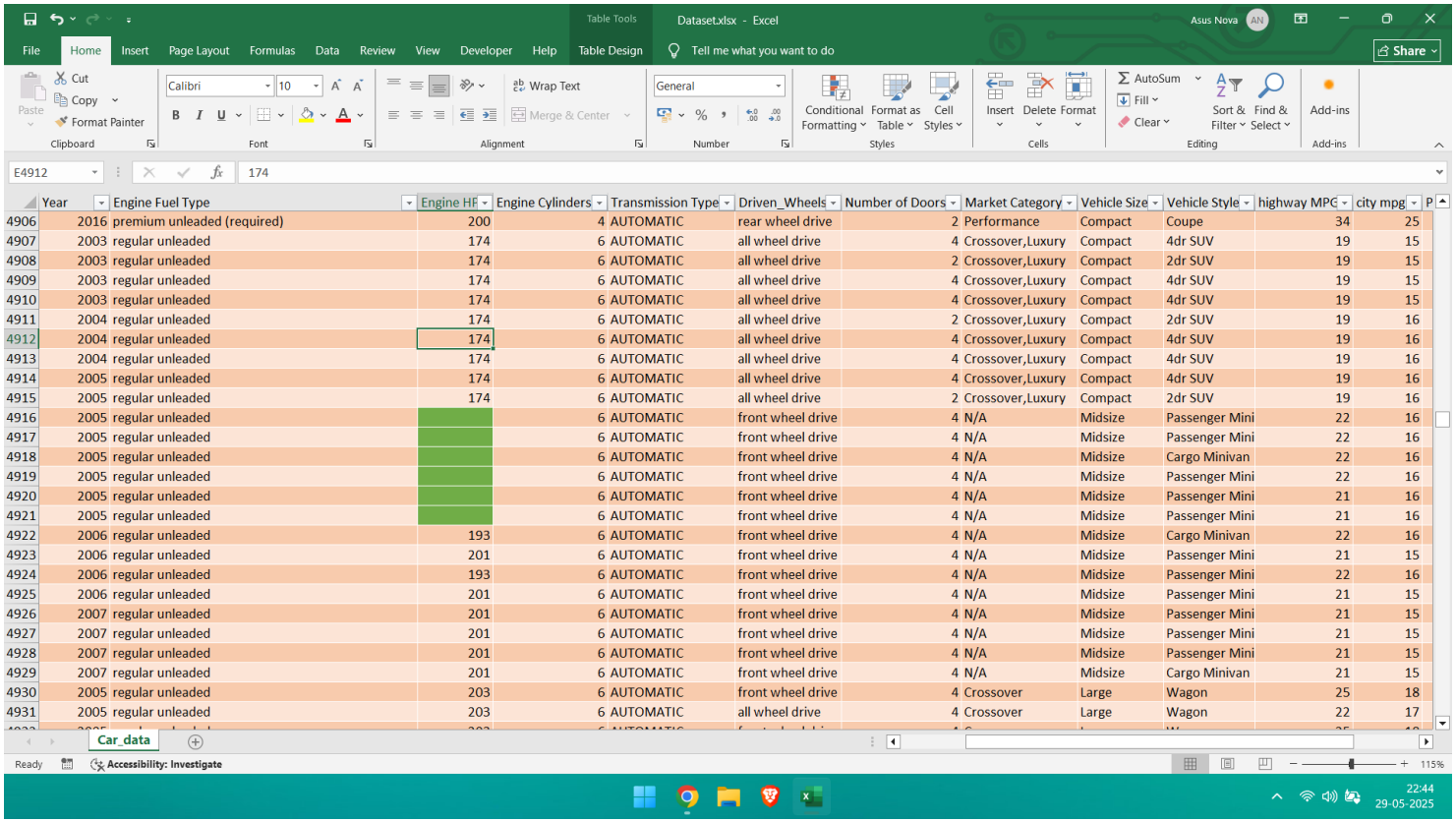
Recommendations:

- Prioritize feature development that enhances engine performance for premium pricing.
- Target marketing efforts towards niche segments that show higher popularity.
- Balance engine size and cylinders to optimize fuel efficiency and appeal to eco-conscious consumers.

Data Cleaning and Preparation

Overview

The raw car dataset underwent comprehensive preprocessing in Excel to address missing values, duplicates, and inconsistencies, ensuring data integrity for reliable analysis.



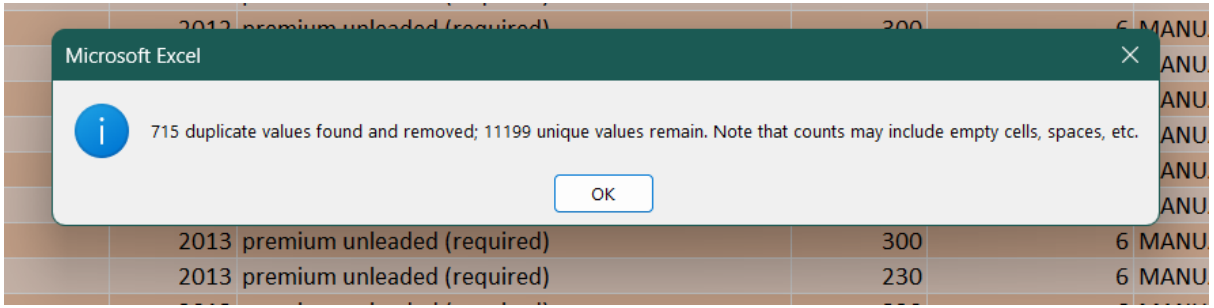
Key Steps

Missing Values:

- Identified missing values in several columns: 69 in Market Category, 30 in Engine HP, and 6 in Transmission.

Duplicate Records:

- Removed 715 duplicate rows based on the combination of Make, Model, and Year to avoid skewing results.



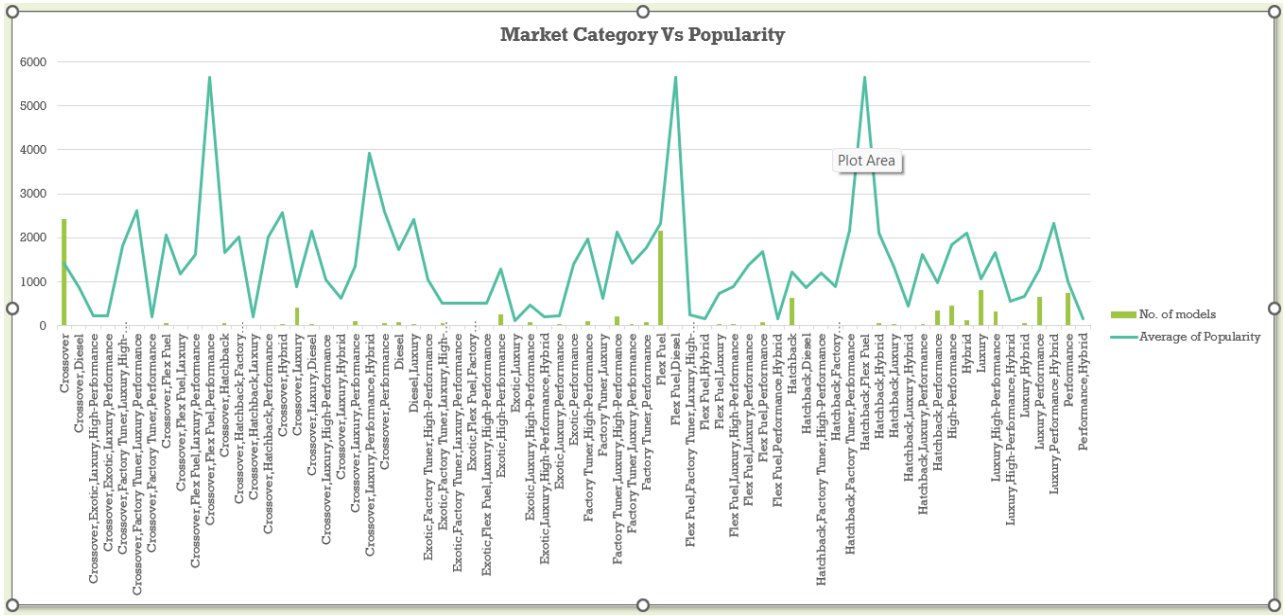
Categorical Data Issues:

- Transmission: 11 "unknown" entries replaced with the column mode due to lack of model-level matches.

- Market Category: 3,376 "N/A" entries replaced using brand-wise pivot tables to assign the most frequent category per brand (using an INDEX + MATCH formula).

Data Analytics Tasks

Task 1: Market Category vs Popularity



Key Insights

Categories such as Crossover, Luxury, and Factory Tuner have the largest number of car models but vary widely in average popularity.

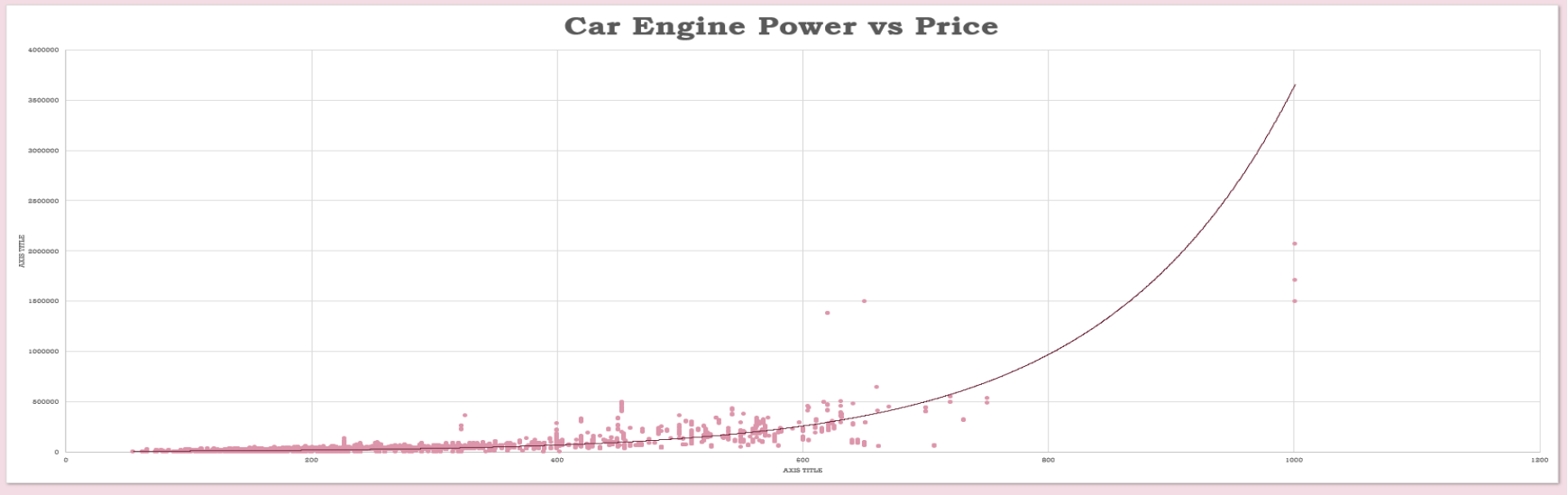
Niche segments like Crossover + Performance and Factory Tuner + Hybrid achieve high popularity despite having fewer models.

Popularity is influenced more by specific category combinations than by market category alone.

Conclusion

Targeted marketing towards niche market segments can yield higher consumer appeal and outperform broader categories.

Task 2: Car Engine Power vs Price



Key Insights

A non-linear, exponential relationship exists between engine power and car price.

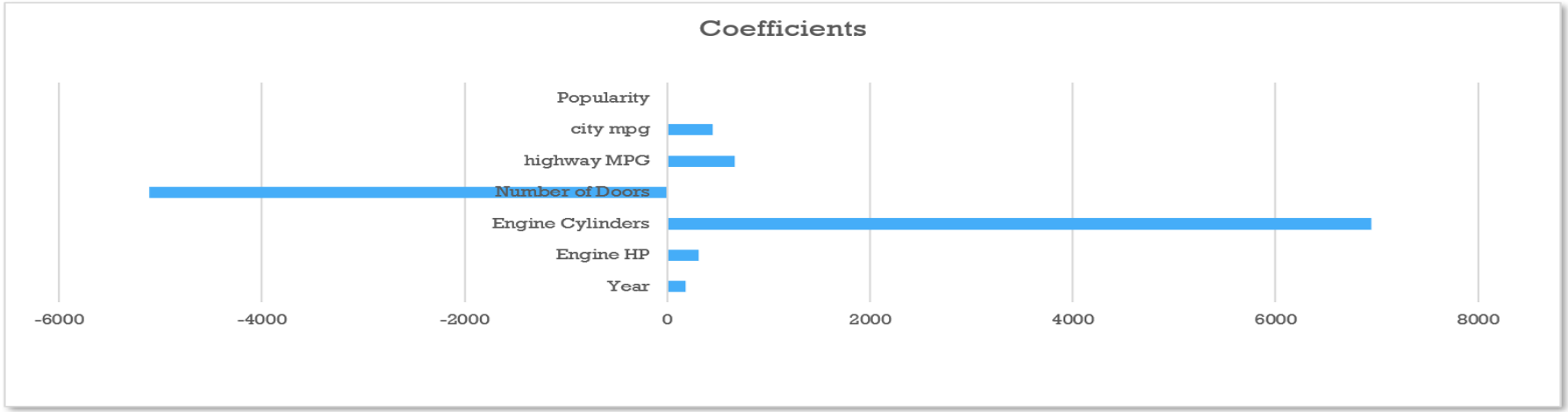
Most vehicles have engine power under 400 HP, correlating with moderate prices; higher HP models command premium prices.

Trendline analysis confirms engine power as a major driver of price, especially in luxury and performance segments.

Conclusion

Performance specifications like engine horsepower significantly affect pricing, essential for targeting premium market segments.

Task 3: Regression Analysis – Key Drivers of Car Price



Regression Statistics	
Multiple R	0.679670053
R Square	0.461951381
Adjusted R Square	0.46161483
Standard Error	45151.1781
Observations	11199

Feature	*Coefficients*
Year	185.0370743
Engine HP	307.0775282
Engine Cylinders	6941.616188
Number of Doors	-5107.906175
highway MPG	664.5009962
city mpg	450.1768297
Popularity	-3.544762442

Key Insights

The regression model explains 46% of the variance in car price ($R^2 = 0.46$).

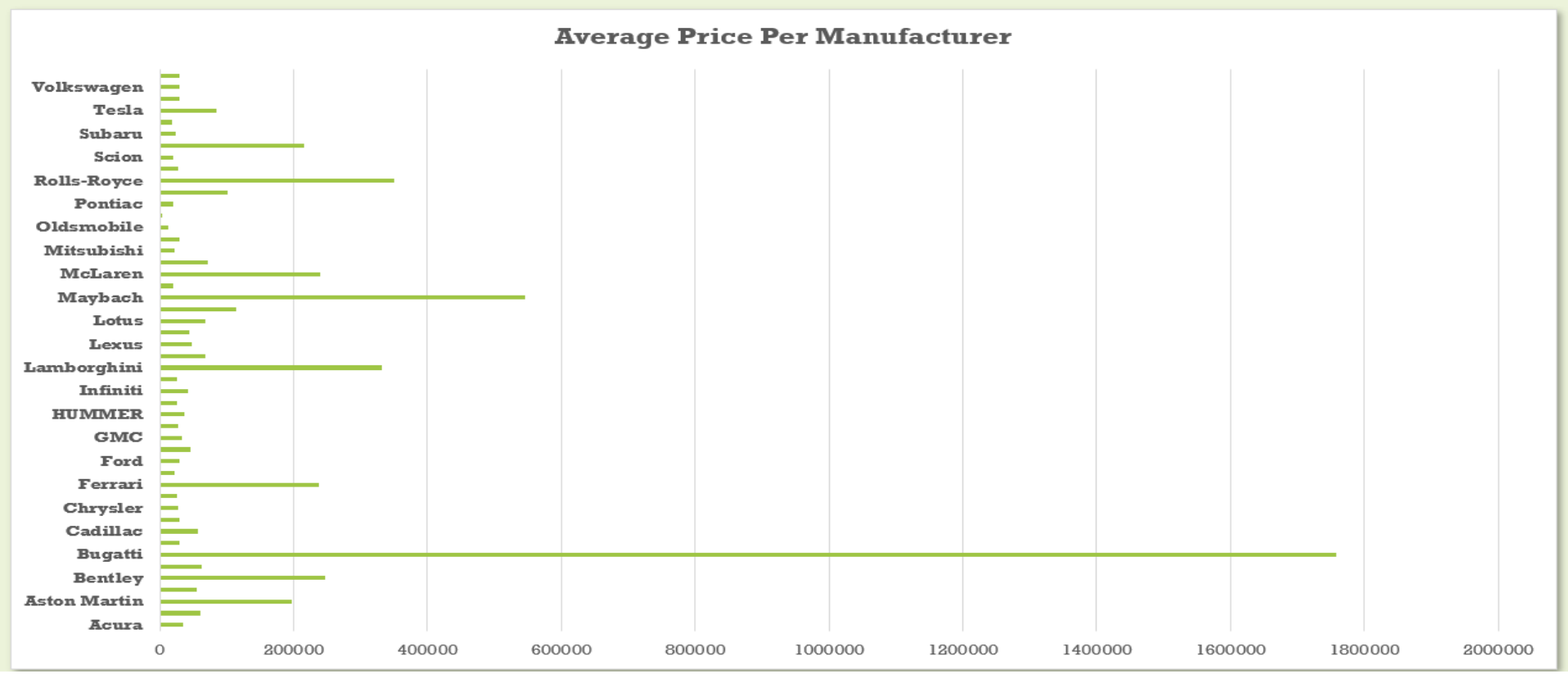
All included features (Engine HP, Cylinders, Doors, MPG, Year, Popularity) are statistically significant ($p < 0.05$).

Engine HP and Cylinders are the strongest positive contributors to price, while more doors negatively affect price.

Conclusion

Performance-related features heavily influence car pricing, providing actionable insights for feature prioritization and pricing strategies.

Task 4: Average Car Price by Manufacturer



Key Insights

Top Luxury Brands: Bugatti leads with an average price of $1.76M, followed by Rolls-Royce ($351K), Lamborghini ($332K), McLaren ($240K), and Ferrari ($238K).

Mid-Range Luxury: Brands like Maserati, Porsche, and Mercedes-Benz fall in the $68K–$114K range.

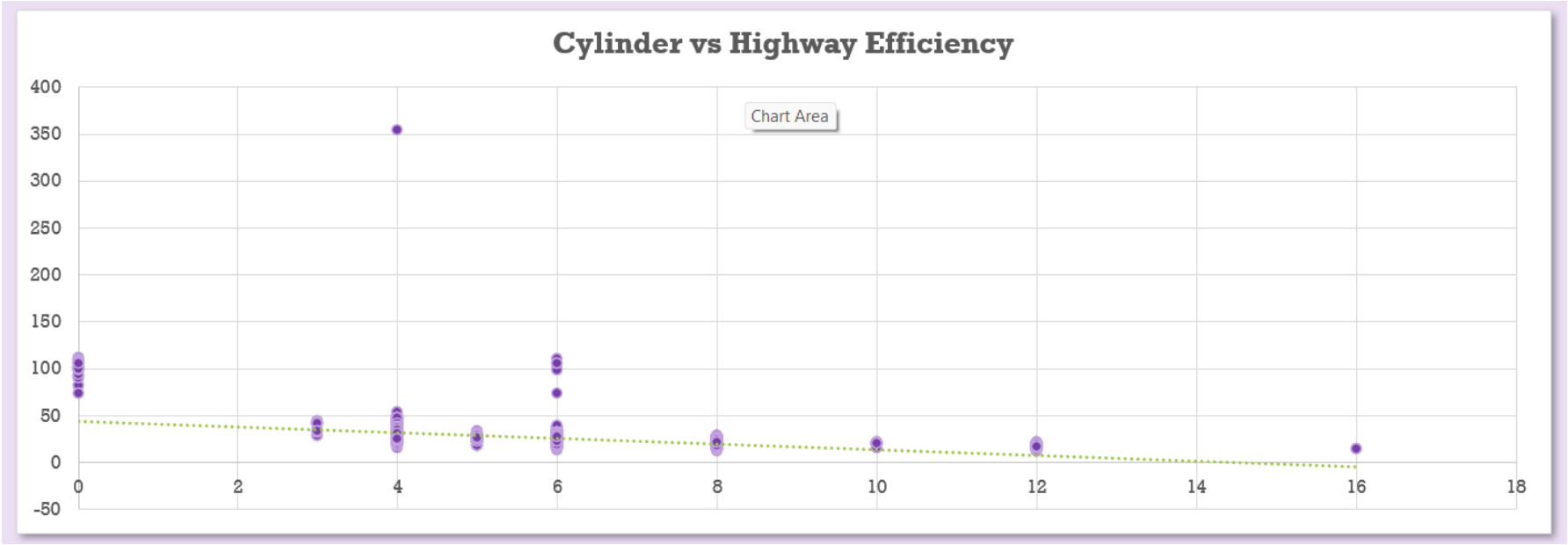
Mainstream Brands: Toyota and Ford average around $29K; Honda averages $27K.

Notable: Tesla's average price ($85K) surpasses many luxury brands, while discontinued brands like Pontiac are on the lower end.

Conclusion

Manufacturer branding plays a critical role in pricing tiers, enabling precise market segmentation.

Task 5: Cylinder Count vs Highway Efficiency



Key Insights

A moderate negative correlation (-0.60) exists between cylinder count and highway MPG.

More cylinders correspond with reduced fuel efficiency, while fewer cylinders improve highway MPG.

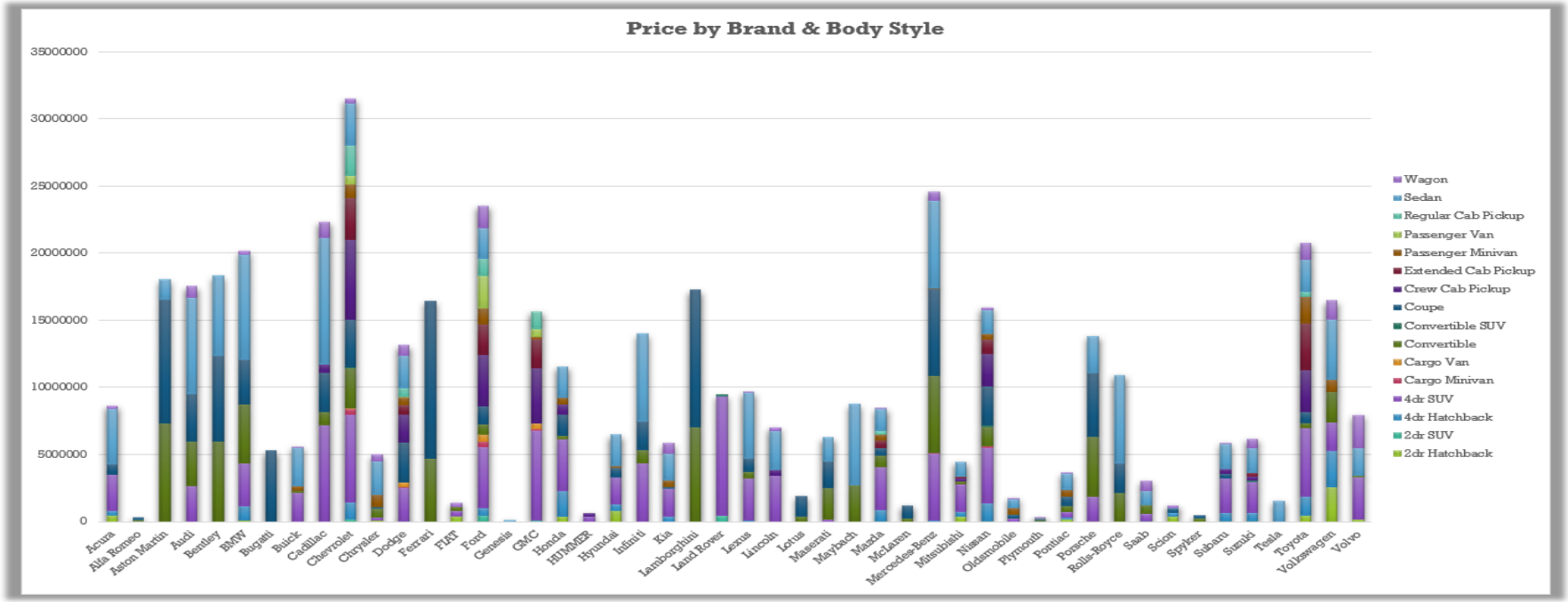
Although some outliers exist (e.g., efficient 4-cylinder models), the general trend remains consistent.

Conclusion

Engine size (cylinder count) significantly impacts fuel efficiency, an important consideration for fuel-conscious consumers.

Dashboard Report: Car Dataset Analysis

Dashboard Task 1: Price by Brand & Body Style



Key Insights

Bugatti dominates with total MSRP near $3.5M, followed by Rolls-Royce and Ferrari, each exceeding $2M.

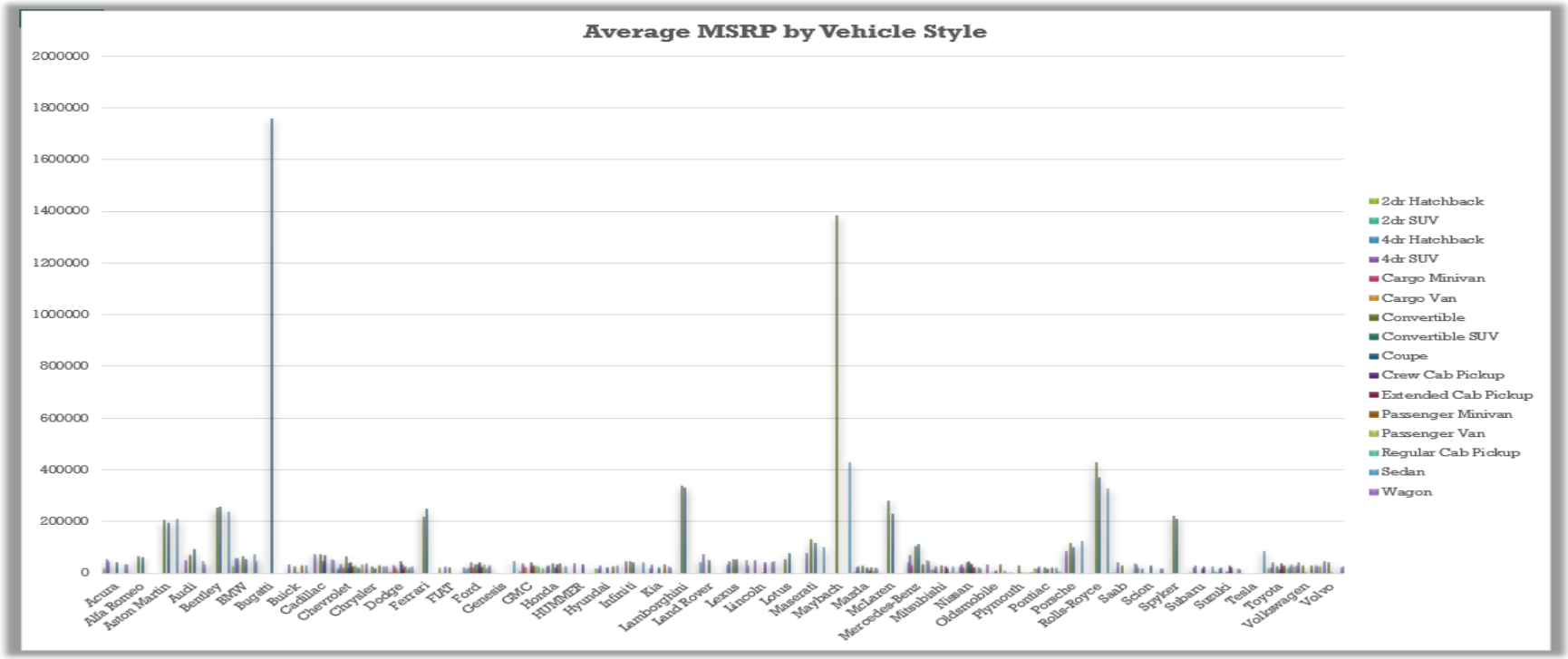
Mainstream brands like Toyota and Honda have total MSRPs below $1M.

Luxury brands focus on Coupes and Convertibles, whereas mainstream brands emphasize Sedans and SUVs.

Conclusion

Body style strongly influences pricing, with luxury body styles commanding significant premiums.

Dashboard Task 2: Average MSRP by Vehicle Style



Key Insights

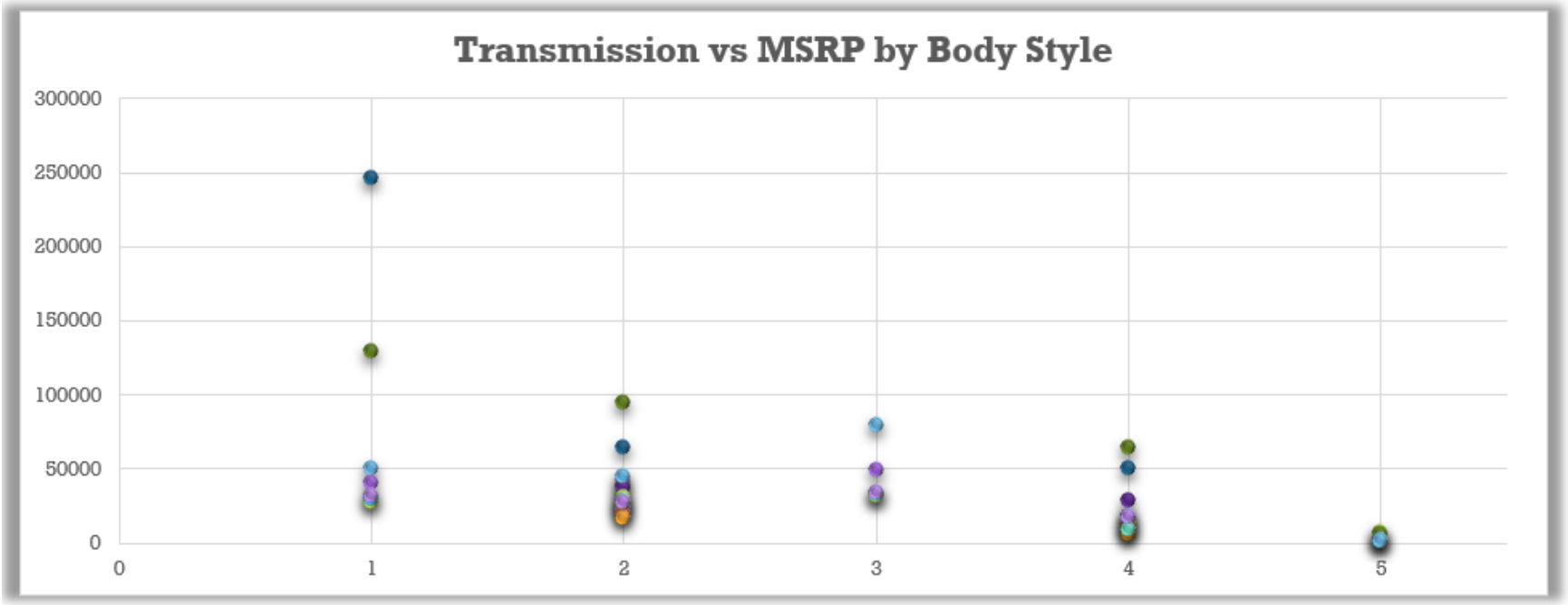
Bugatti Convertibles average $1.4M MSRP, the highest among all vehicle styles.

Mainstream vehicles generally have MSRPs below $50K; luxury vehicles range between $200K and $400K.

Conclusion

Vehicle style impacts average pricing, reinforcing the value of luxury designs.

Dashboard Task 3: Transmission vs MSRP by Body Style



Category	Transmission code
AUTOMATED_MANUAL	1
AUTOMATIC	2
DIRECT_DRIVE	3
MANUAL	4
UNKNOWN	5

Key Insights

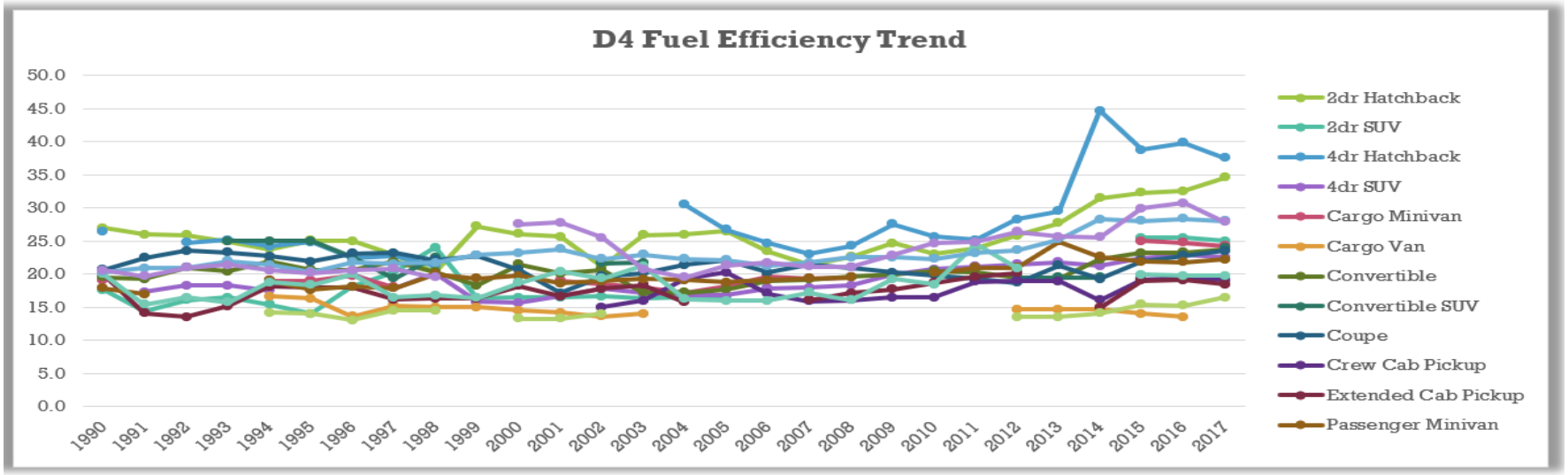
Automated Manual Coupes have the highest average MSRP (~$129K).

Manual transmission vehicles, such as Passenger Minivans (~$6.5K), are the most affordable.

Conclusion

Transmission types correlate with MSRP, with advanced transmission systems linked to higher prices.

Dashboard Task 4: Fuel Efficiency Trend



Key Insights

Fuel efficiency improves over time, especially post-2010.

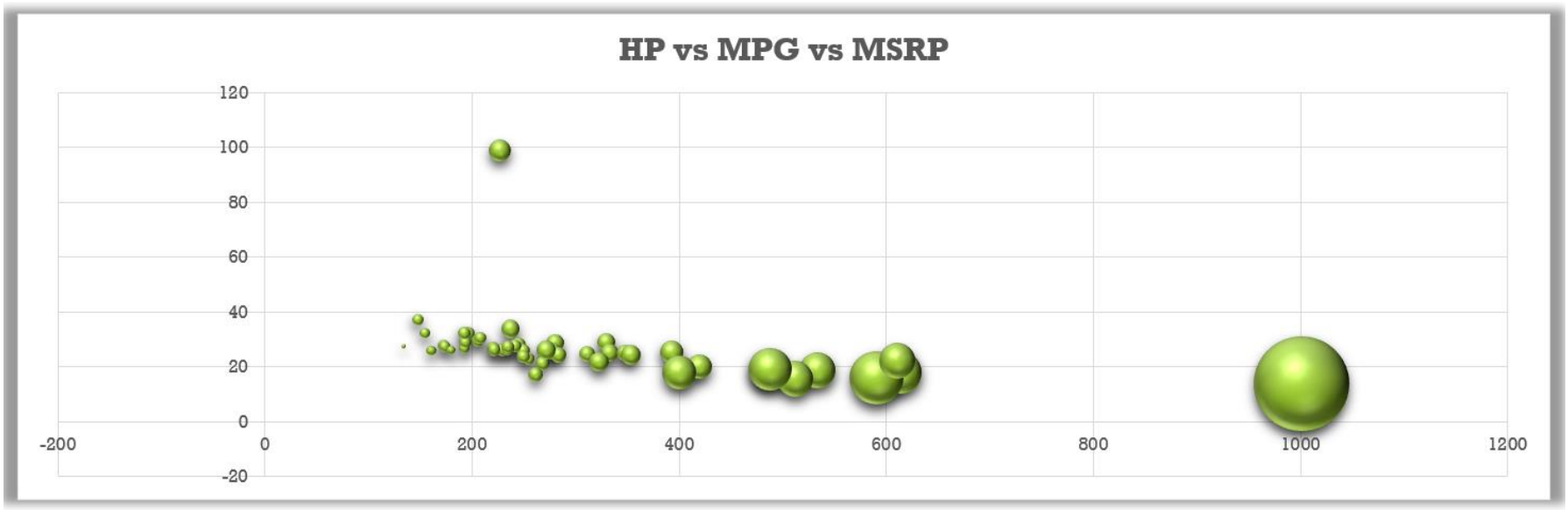
4-door Hatchbacks peak at 40–45 MPG by 2017.

Cargo Vans and Pickups remain below 20 MPG; Sedans and SUVs average 30–35 MPG.

Conclusion

Compact vehicle styles lead in fuel efficiency improvements, important for environmentally friendly design.

Dashboard Task 5: Horsepower vs MPG vs MSRP



Key Insights

High HP (600–1000) cars exhibit low MPG (20–30) and high MSRP.

Low HP (100–200) vehicles achieve higher MPG (40–60) and have lower MSRP.

Outliers with ~100 MPG and ~100 HP suggest the presence of niche efficient vehicles (e.g., hybrids).

Conclusion

High-performance vehicles trade fuel efficiency for premium pricing, whereas efficient

Results and Insights

Visualizations:

- Dashboards illustrated pricing distribution by brand and body style, transmission vs MSRP, and fuel efficiency trends over time.
- Scatter plots confirmed the exponential price increase with horsepower.
- Pivot tables summarized average prices and popularity by market category and manufacturer.

Discussion:

- The results demonstrate clear segmentation in car markets based on performance, brand, and vehicle style.
- Fuel efficiency improvements over time emphasize shifting consumer and regulatory priorities.
- Transmission type and body style also influence pricing strategies.

Limitations:

- The regression model has moderate explanatory power, suggesting other variables (e.g., technology features, brand perception) were not captured.
- Missing data imputation, while methodical, introduces potential biases.
- The static dataset reflects a snapshot in time; market trends may evolve.

Future Directions:

- Incorporate additional variables such as safety ratings, technology packages, and resale value for enhanced models.
- Explore non-linear and interaction effects in predictive modeling.
- Extend analysis to include temporal trends and market shifts over multiple years.

Drive Link [Open In Excel]

https://docs.google.com/spreadsheets/d/1IMQkDf1oB0_mkIkUuTYFh1xFsB1r3fRg/edit?usp=sharing&ouid=115912486417524737249&rtpof=true&sd=true