# Analyzing the Impact of Car

# Features on Price and

# Profitability

**A Data-Driven Approach to**

**Optimize**

**Product Development**

**& Pricing Decisions**

Deepthy A

Trainity Data Analyst Trainee

Contents

[Project Description 2](#_Toc204201340)

[Approach 3](#_Toc204201341)

[Tech Stack Used 4](#_Toc204201342)

[Task 1: Market Category Popularity Analysis 5](#_Toc204201343)

[Task 2: Relationship Between Horsepower and MSRP 7](#_Toc204201344)

[Task 3: Identifying Key Car Features Affecting Price 9](#_Toc204201345)

[Task 4: Analysis of Average Car Price by Manufacturer 11](#_Toc204201346)

[Task 5: Relationship Between Fuel Efficiency and Engine Cylinders 13](#_Toc204201347)

[Interactive Dashboard Analysis 15](#_Toc204201348)

[Dashboard Filters & Interactivity 22](#_Toc204201349)

[Conclusion 23](#_Toc204201350)

# Project Description

Overview of the Project

The automotive industry is undergoing a significant transformation driven by rising consumer expectations, environmental sustainability, and fierce market competition. This project aims to uncover key insights into how car features influence pricing and profitability using a dataset of over 11,000 car models and specifications.

Business Problem

The key business question addressed in this project is:  
**“How can a car manufacturer optimize pricing and product development decisions to maximize profitability while meeting consumer demand?”**  
This involves identifying:

* Which car features drive up prices
* What consumers value most (popularity)
* Which product categories and brands are most profitable

Data Source

The dataset, titled **“Car Features and MSRP”**, is publicly available on Kaggle and was originally compiled by Cooper Union, a private college in New York.  
It contains **11,159 car entries** with **16 variables**, including:

* **Car make & model**
* **Fuel type**, **engine specs**, **transmission**, **drivetrain**
* **Market category**, **vehicle size/style**
* **City & highway MPG**, **popularity**, **MSRP**

Data Cleaning & Preprocessing

To ensure accuracy and analytical relevance:

* **Removed duplicates** across Make, Model, and Year
* **Handled missing values** in Engine HP, Cylinders, and Market Category
* **Converted columns** like MSRP, Engine HP, and Popularity to proper numeric format
* **Standardized categorical columns** (e.g., Transmission Type, Fuel Type)
* **Split multivalued fields** like Market Category using delimiter logic where necessary

Assumptions

* MSRP represents the base model price and doesn’t reflect customizations
* Popularity scores are used as a **proxy for consumer interest** (based on views on Edmunds.com)
* Data is considered valid for trend analysis despite being **last updated in 2017**

# Approach

This project follows a structured analysis pipeline to ensure insights align with the business objective of **maximizing profitability through data**. The steps include:

**1. Descriptive Analytics**

* Frequency and distribution analysis using **Pivot Tables**
* Visualizations such as **bar charts**, **combo charts**, and **scatter plots**

**2. Diagnostic Analytics**

* **Correlation analysis** between features like engine specs and fuel efficiency
* **Market segmentation** by vehicle category, body style, and brand

**3. Predictive Analysis**

* **Linear regression modeling** using the **Data Analysis ToolPak** to identify the most influential car features affecting price

**4. Interactive Dashboarding**

* Created a comprehensive **Excel dashboard** with slicers, stacked/clustered charts, and trend lines
* Enabled **brand, year, and body style filters** for end-user exploration

Reasoning for Approach

The analytical methods were chosen to balance:

* **Interpretability** for stakeholders
* **Ease of implementation** in Excel
* **Depth of insight** into feature-price relationships
* **Dynamic exploration** of data using dashboards

Challenges & Limitations

* Missing or ambiguous values in multivalued fields (e.g., Market Category)
* Dataset reflects trends up to 2017; newer models and electric vehicles are underrepresented
* Popularity metric may be **biased by marketing or visibility**, not pure consumer preference

# Tech Stack Used

|  |  |
| --- | --- |
| Tool | Purpose |
| Microsoft Excel (Office 365) | Primary tool for data cleaning, transformation, pivoting, regression, visualization, and dashboarding |
| Excel Add-ins (Data Analysis ToolPak) | Used for regression modeling and correlation analysis |
| Formulas Used | IF(), SUMIF(), AVERAGEIF(), COUNTIF(), CORREL(), TEXT(), and others |
| Interactive Tools | Pivot Charts, Slicers, Combo Charts, Scatter Plots, Bubble Charts |

**Why Excel?**

Excel was chosen as the core analysis tool due to its:

* Business-friendliness for stakeholders
* Built-in visualization features
* Support for pivoting, regression, and interactivity without third-party software

# Task 1: Market Category Popularity Analysis

1A. Pivot Table: Market Category vs Popularity Share & Count of Models

To understand consumer interest across different car types, we analyzed the **Market Category** against two key metrics:

* **Sum of Popularity Share** – total popularity score for all models in each category
* **Count of Models** – total number of car models within that market category

**Key Data Highlights (from Pivot Table):**

|  |  |  |
| --- | --- | --- |
| **Row Labels** | **Sum of Popularity Share** | **Count of Model** |
| Crossover | 130333.3858 | 1994 |
| Diesel | 10474.62895 | 202 |
| Exotic | 40903.89335 | 491 |
| Factory Tuner | 73272.64555 | 609 |
| Flex Fuel | 67210.1191 | 1179 |
| Hatchback | 95807.66747 | 1075 |
| High-Performance | 112726.2815 | 1270 |
| Hybrid | 49324.62519 | 334 |
| Luxury | 217571.5196 | 3209 |
| N/A | 339361.7448 | 3376 |
| Performance | 139270.6992 | 1923 |
| **Grand Total** | **1276257.21** | **15662** |

**Note**: The "N/A" category represents cars not explicitly classified under a specific market segment but still scored high in popularity, possibly indicating general-purpose or versatile appeal.

1B. Visual Analysis: Market Category vs Popularity (Combo Chart)

The chart visualizes:

* **Blue Line**: Total Popularity Share (with values labeled)
* **Orange Line**: Model Count per category
* **Trendline**: Linear regression on popularity scores

**Insights:**

* **Luxury** and **N/A** categories dominate both in model count and popularity, showing high market coverage and consumer interest.
* **Performance** and **High-Performance** cars, though fewer in count, still attract significant attention - suggesting niche but enthusiastic demand.
* **Hybrid**, **Diesel**, and **Exotic** categories show **lower overall popularity**, possibly due to limited availability or appeal during the dataset's timeline.
* The upward sloping **trendline** confirms a general positive correlation between market category and consumer popularity.

**Interpretation**: Not all popular segments are the most common. A few highly appealing models in niche categories (like Performance or Factory Tuner) punch above their weight in popularity, suggesting opportunities in focused innovation.

# Task 2: Relationship Between Horsepower and MSRP

The chart titled **"Horsepower vs. MSRP"** displays:

* A **positive linear trendline**, suggesting a **direct relationship** between MSRP and Horsepower.
* The equation of the trendline is:

where **y = Horsepower** and **x = MSRP**.

* The **R² value is 0.4183**, which indicates a **moderate correlation**:
  + About **41.8% of the variation in Horsepower can be explained by the MSRP**.
  + The rest (~58.2%) is influenced by other factors (e.g., vehicle type, technology, brand).

**Conclusion**:

Higher-priced vehicles generally tend to have higher horsepower, but the correlation isn't very strong—luxury/performance models may drive exceptions.

From the chart:

* Several **outliers** are present:
  + Some vehicles have **extraordinarily high horsepower** (e.g., over **1,000,000 HP**), which is **physically unrealistic** and likely a **data entry error** or an artifact of calculation.
* These **outliers distort the scale** of the chart, especially vertically, stretching the Y-axis (up to 2,500,000 HP).
* **A few data points have negative or zero MSRP or Horsepower**, which may be invalid entries.

**Recommendation**:

* **Clean the data** to remove or correct unrealistic values (like horsepower > 2000 HP or MSRP ≤ 0).
* Consider **log-scaling** MSRP or HP, or filtering to a realistic range for better visual interpretation.

# Task 3: Identifying Key Car Features Affecting Price

Objective

The goal of this task is to determine which car features have the most significant impact on the vehicle's **selling price**. This insight will help stakeholders understand what factors most strongly influence profitability and pricing strategy.

Methodology: Regression Analysis

To analyze the relationship between car features and price, we performed a **multiple linear regression** using Microsoft Excel. The model treats **Price** as the dependent variable and includes several independent variables (features) such as:

* Highway MPG
* City MPG
* Popularity
* Year
* Engine Horsepower (HP)
* Engine Cylinders
* Transmission Type (encoded)
* Make (encoded)
* Number of Doors

Each feature's **regression coefficient** indicates the **magnitude and direction** of its impact on price. A higher absolute coefficient suggests a stronger influence.

Regression Output

Below is a summary of the regression coefficients for each feature:

|  |  |
| --- | --- |
| Feature | Coefficient |
| Popularity | 3916 |
| Year | 2011 |
| Engine Horsepower (HP) | 335 |
| Highway MPG | 26 |
| City MPG | 19 |
| Engine Cylinders | 6 |
| Transmission Type (Encoded) | 1(Manual) |
| Make (Encoded) | 1(BMW) |
| Number of Doors | 2 |

To better visualize their relative importance, a **bar chart** was created based on these coefficient values.

# Task 4: Analysis of Average Car Price by Manufacturer

Objective

To identify how the **average price of cars (MSRP)** varies across **different manufacturers**.

4A. Pivot Table Insight

A pivot table was created to calculate the **average MSRP (Manufacturer's Suggested Retail Price)** for each car brand. The key findings are:

|  |  |
| --- | --- |
| Manufacturer | Average MSRP (USD) |
| Bugatti | **1,757,223.67** |
| Maybach | 546,221.88 |
| Rolls-Royce | 351,130.65 |
| Lamborghini | 331,567.31 |
| Bentley | 247,169.32 |
| McLaren | 239,805.00 |
| Ferrari | 238,218.84 |
| Maserati | 113,684.49 |
| Aston Martin | 198,123.46 |
| Spyker | 214,990.00 |

**Average MSRP across all manufacturers:** 258,231.88

4B. Visualization

A **horizontal bar chart** was created to represent the average MSRP by each manufacturer visually.

It clearly shows that:

* **Bugatti** has the highest average car price by a large margin.
* **Luxury brands** like Rolls-Royce, Maybach, and Lamborghini also command significantly high prices.
* Other premium brands like Ferrari, Bentley, and McLaren follow with moderately high MSRPs.

**Insight:**

There’s a **wide variation in pricing across manufacturers**, and a small group of ultra-luxury brands dominates the high end of the price spectrum. Manufacturers significantly impact car pricing. Ultra-luxury brands like **Bugatti and Maybach** skew the average MSRP upward.

# Task 5: Relationship Between Fuel Efficiency and Engine Cylinders

Objective

To evaluate how **fuel efficiency** (measured by highway MPG) is affected by the **number of engine cylinders**.

5A. Scatter Plot & Trendline

A scatter plot was created with:

* **X-axis:** Number of Cylinders
* **Y-axis:** Highway MPG

A **trendline** was added with the equation:

and

This indicates a **negative linear relationship** between cylinders and highway MPG:

* As the **number of cylinders increases**, **fuel efficiency decreases**.
* The downward trend is consistent, though not perfectly linear.

5B. Correlation Coefficient

The **correlation coefficient** between the number of cylinders and highway MPG is:

|  |  |
| --- | --- |
| Correlation | -0.61034 |

**Insight:**  
This is a **moderately strong negative correlation**, indicating that cars with more cylinders are typically **less fuel efficient**. Fuel efficiency **declines** with more engine cylinders, reinforcing the trade-off between **power and economy**.

# Interactive Dashboard Analysis

An interactive Excel dashboard was created to provide a consolidated and visually driven understanding of car pricing, performance, and profitability across various features. This dashboard includes slicers for **Car Brand**, **Body Style**, and **Model Year**, which allow users to dynamically explore how different factors influence MSRP (Manufacturer’s Suggested Retail Price), fuel efficiency, horsepower, and profit.

Below is a breakdown of the key dashboard components and their insights:

1. MSRP by Car Brand and Body Style

To understand consumer preferences and market distribution, we analyzed how different **body styles** are represented in the dataset using a **bar chart visualization**.

**Key Observations:**

* **Sedans** emerged as the most prevalent body style, significantly outnumbering other categories. This suggests that sedans continue to be a widely preferred choice among consumers due to their balance of comfort, fuel efficiency, and affordability.
* **SUVs**, **hatchbacks**, and **convertibles** followed, indicating moderate representation. Their popularity often reflects lifestyle choices, such as off-road capabilities, urban mobility, or luxury appeal.
* **Coupes** and **wagons** accounted for a smaller share, possibly due to niche market appeal or evolving consumer trends.

**Approach:**

* We utilized **pivot tables** to aggregate the number of vehicles under each body style category.
* The summarized data was visualized through a **bar chart** for a clear and comparative overview.

This analysis helps stakeholders gauge which segments dominate the market and where there may be opportunities for targeting underrepresented niches.

2. Top and Bottom 5 Average MSRP by Brand

From the analysis using a pivot table of **Brand vs. Body Style**, the following insights were obtained:

**Top Brands by Highest Average MSRP:**

* **Bugatti** leads all brands with an exceptionally high average MSRP — **over $1.7 million**, attributed to its ultra-luxury sports lineup.
* **Luxury brands** like **Rolls-Royce**, **Maybach**, **Bentley**, and **Ferrari** follow, with average MSRPs exceeding **$250,000** across body styles such as sedans, coupes, and convertibles.
* These brands cater to a niche high-net-worth customer segment and consistently position themselves in the ultra-premium market.

**Bottom Brands by Lowest Average MSRP:**

* **Mainstream brands** like **Hyundai**, **Kia**, **Chevrolet**, **Toyota**, and **Ford** have some of the lowest average MSRPs — typically ranging from **$15,000 to $35,000**.
* These brands focus on affordability and volume, offering economy sedans, compact cars, and SUVs with competitive pricing.

The contrast between the top and bottom brands underscores how pricing strategy reflects brand positioning and target audience. High-end luxury brands focus on exclusivity, while economy brands prioritize accessibility.

**Result:**

A **column chart** was created using a pivot table to show average MSRP per brand across different body styles.

* The chart visually identifies outliers like **Bugatti** and **Maybach** with extremely tall bars.
* Brands like **Hyundai** and **Kia** appear on the lower end of the chart, with consistently short bars across all styles.
* This chart allows easy comparison of **top and bottom brands** across segments such as **SUV**, **Sedan**, **Coupe**, and **Convertible**.

**Interactive slicers** were added for Body Style, allowing users to isolate trends within specific categories and observe which brands dominate or trail in each.

3. Transmission Type Impact on MSRP

**Approach:**To analyze how vehicle transmission types influence MSRP across different body styles, a line graph was used to visualize average MSRP values. Data was summarized using Pivot Tables with AVERAGEIFS, simplifying comparison across categories.

**Findings:**

* Luxury Pricing Trends:  
  Vehicles with Automatic transmissions in Convertible and Coupe body styles recorded the highest average MSRP, often surpassing $90,000. These segments cater to luxury and performance markets.
* Mid-Range Pricing:  
  Sedans and Wagons, available in both transmission types, exhibited moderate MSRPs. Among these, Automatic variants generally cost more than Manual ones.
* Budget Segment:  
  Hatchbacks and Cargo Vans with Manual transmissions showed the lowest MSRP, reflecting their utility-driven design and economic positioning.

**Conclusion:**  
Automatic transmission vehicles in premium body styles (like Coupes and Convertibles) command higher prices, while Manual transmissions are more prevalent in cost-effective models like Vans and Hatchbacks.

4. Fuel Efficiency (MPG) Trends Over Years

**Approach:**To assess trends in fuel efficiency (Highway MPG) over time, a Pivot Table was created to calculate the average MPG for each body style and model year. A line chart was then used to visualize the yearly progression of fuel efficiency by body style.

**Findings:**

* **General Trend:**There is a clear upward trend in highway MPG across most vehicle body styles from 1990 to 2017, reflecting industry-wide improvements in fuel efficiency due to technological advancements and stricter emission standards.
* **Highest Efficiency:**
  + 4dr Hatchbacks recorded the highest average MPG over the years, with values peaking above 45 MPG in later years like 2014 and beyond.
  + 2dr Hatchbacks also consistently performed well, averaging 31.3 MPG over the entire period.
* **Lowest Efficiency:**
  + Cargo Vans and Passenger Vans had the lowest average MPG, around 16.5 MPG and 17.3 MPG respectively. These body styles prioritize cargo/passenger capacity over efficiency.
* **Steady Improvements:**
  + Sedans and Wagons showed steady gains, reaching over 33 MPG in recent years.
  + SUVs (both 2dr and 4dr) started below 20 MPG in the 1990s but improved to mid-20s or higher in later years.

**Conclusion:**Your analysis demonstrates that fuel efficiency has improved substantially over the years for almost all vehicle body styles. Compact and hatchback designs lead in efficiency, while larger utility vehicles still lag behind despite moderate gains. These insights are valuable for understanding how body style impacts environmental performance over time**.**

5. Brand-wise Comparison of Horsepower, MPG, and Price

To evaluate how car performance, fuel efficiency, and pricing vary across different brands, we developed a **bubble chart** where:

* The **X-axis** shows the **average horsepower** per brand.
* The **Y-axis** shows the **average MSRP**.
* **Bubble size** represents the **average city MPG**.
* Each **bubble is color-coded and labeled by brand name**, making the insights visually intuitive.

**Key Observations:**

* **Luxury & Performance Brands**:
  + Brands such as **Bugatti**, **Rolls-Royce**, **Ferrari**, and **McLaren** occupy the upper-right quadrant, indicating **very high horsepower and extremely high prices**.
  + These brands also have **smaller bubble sizes**, reflecting **lower fuel efficiency**, a common trade-off for performance.
* **Top Performer in Price & Power**:
  + **Bugatti** is a clear outlier with **over 1000 HP** and an MSRP exceeding **$1.75 million**, standing far apart from the rest of the market.
* **Fuel-Efficient Premium Options**:
  + **Tesla** distinguishes itself with **large bubble size**, signifying **very high MPG (due to electric engines)**, while still maintaining a **premium price point** and **moderate horsepower**. This underscores the evolving trend toward efficient luxury.
* **Balanced Value Brands**:
  + Brands like **Honda**, **Toyota**, and **Mazda** are found in the lower-left region of the chart, with **moderate horsepower**, **high MPG**, and **affordable prices**, making them strong contenders in the mid-market value segment.

**Approach:**

* We calculated **average horsepower, MSRP, and MPG** for each brand using **Pivot Tables**.
* The resulting summary was used to create a **bubble chart** that visually encapsulates the brand-level differences in car features.
* This visual aids in **comparing performance-to-price ratios and fuel economy trade-offs** across different segments of the automobile industry.

# Dashboard Filters & Interactivity

To enhance user experience, the dashboard incorporates slicers for:

* **Brand**
* **Body Style**
* **Model Year**
* **Transmission type**

These filters allow for real-time visual adjustments across all charts, making the dashboard highly interactive and ideal for business decision-making.

**Overall Insights from Dashboard:**

* **Premium brands** clearly charge more due to performance and branding.
* **Body style** and **transmission type** significantly influence MSRP.
* **Fuel efficiency** has improved over time across the industry.
* There is a **clear trade-off** between horsepower and MPG, with price often reflecting this balance.

A screenshot of a computer

AI-generated content may be incorrect.

# Conclusion

This project provided an in-depth analysis of car features and their impact on pricing and profitability using Excel-based data analytics techniques. Through tasks involving pivot tables, charts, and conditional formulas, we derived meaningful business insights relevant to marketing, product development, and strategic pricing decisions in the automotive industry.

Key findings from the analysis include:

* **Feature Impact on Price**: Transmission type and body style significantly influence the Manufacturer’s Suggested Retail Price (MSRP). Manual transmissions tend to be priced lower on average, with body styles such as coupes and convertibles skewing towards higher price points regardless of transmission.
* **Fuel Efficiency Trends**: Cars with higher MPG (Miles per Gallon) are not necessarily more expensive, indicating a potential opportunity for eco-friendly models in mid-range pricing tiers.
* **Transmission Type Preferences**: Automatic transmissions dominate across most body styles, correlating with higher average MSRPs—suggesting consumer preference for convenience features.
* **Brand Performance**: Pivot-based analysis revealed stark differences in average horsepower, fuel economy, and pricing across brands. Premium brands demonstrated higher MSRPs but lower MPG, consistent with performance-focused offerings.
* **Data-Driven Recommendations**:
  + Consider bundling premium features with automatic transmissions for compact SUVs and sedans, where customer demand and pricing flexibility coexist.
  + Encourage competitive pricing for fuel-efficient models to target value-conscious consumers.
  + Monitor brands with extreme variance in horsepower vs. MPG to better segment marketing efforts.

This analytical approach not only showcased how individual features correlate with price but also demonstrated the power of Excel in driving structured business intelligence. By using pivot tables, line graphs, bar charts, and aggregations, we successfully transformed raw data into actionable insights.

<https://docs.google.com/spreadsheets/d/1_SaT3nLd_TILczyhDz_PgZiuUQTdHxs5/edit?usp=sharing&ouid=112959782025131466050&rtpof=true&sd=true>