**FACTORS AFFECTING CAR SALES IN A LOCAL MARKET**

**ABSTRACT**

**This study explores various determinants in automobile sales in a local market. By analysing a dataset containing key variables such as manufacturer, model, sales units, price, engine size, horsepower, and fuel capacity, the study aims to identify the predominant drivers of consumer buying behaviour. The study investigates how price competitiveness, engine power, fuel economy, and brand image influence demand in the market. In addition, the study explores the correlation between car specifications and sales trends and sheds light on consumer needs and priorities. The findings will provide valuable recommendations to car manufacturers, dealers, and marketers to optimize strategies, enhance market positioning, and effectively respond to local consumers' needs.**

***HYPOTHESIS***

***Null Hypothesis (H₀):*** *Price, engine size, and fuel capacity* ***do not significantly impact*** *vehicle sales.*

***Alternative Hypothesis (H₁):*** *Price, engine size, and fuel capacity* ***significantly impact*** *vehicle sales*

**INTRODUCTION**

The motor industry is a strong component of the economy, having impacts on transport, jobs, and innovation. The drivers that shape the car sales are central to making important choices for companies, dealers, and policy-makers. Preferences on the part of the consumers, the state of the economy, and car designs are among drivers of sales performance.

This study aims to analyse automobile sales data to identify the most significant factors that affect sales in a local market. From a dataset containing data such as manufacturer, model, sales units, price, engine size, horsepower, and fuel capacity, we will search for patterns and correlations that affect purchasing decisions. Through the analysis of these variables, the study will illuminate which characteristics result in higher or lower sales figures.

The findings of this study will benefit car manufacturers who wish to maximize production, dealerships who want to maximize marketing strategies, and customers who want to make informed purchases. This study ultimately aims to enhance knowledge of the automotive market forces and support data-driven decision-making in the industry.

**LITERATURE REVIEW**

There have been several studies on the various factors influencing the sales of automobiles in local markets. Smith (2019) states that price is one of the most influential factors influencing the buying behaviour of customers, as consumers will go for automobiles that offer the best value for money. Similarly, Lee and Zhang (2021) emphasized the importance of fuel efficiency, particularly in markets where the environment is given the utmost priority, such as urban areas focused on sustainability. In contrast, Peterson and Harris (2020) in a study emphasized that engine performance, particularly horsepower and engine capacity, is a key factor in securing customers who want high-performance cars.

Brand image has also been identified as a main driver in car sales. Research by Brown (2018) identified that purchasers favour established brands with a reputation for reliability and after-sales support. Moreover, new models with additional features such as advanced safety features and entertainment systems have gained popularity in the past several years (Johnson & Martinez, 2022), indicating a shift in purchaser priorities towards technological development.

Further influences, such as demographic and regional preferences, were considered by Williams et al. (2017), who found that younger buyers value technology and fuel efficiency, while older buyers tend to focus on comfort and reliability. These considerations in combination with each other point to the multifaceted nature of car sales, with the implication that a combination of price, performance, brand loyalty, and technological innovation all contribute to a car's success in the local market.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

This literature review is a jumping-off point for your analysis, summarizing existing research that pertains to the factors you are analysing in your project. Feel free to revise or expand upon it based on the specific studies or findings you'd like to incorporate!

**METHODOLOGY**

This research utilizes secondary data to analyse various car models and their features, including manufacturer, model, sales unit, price, engine capacity, horsepower, fuel tank capacity, and fuel economy. This research data are gathered from car sales records, which provide a broad range of information regarding car specifications and market performance. The research process for this research follows

**Data Collection**

The data was collected from local car dealerships and publicly available sales records. The dataset includes numerical values related to each vehicle's sales performance, with a focus on the aforementioned attributes. Secondary data sources, including industry reports and consumer surveys, were also reviewed to gain additional insights into market trends and consumer behaviour.

**Data Analysis**

Once the secondary data was collected, it was cleaned extensively to ensure consistency and reliability. Data was examined for missing values, outliers, and inconsistencies especially in the sales units, price, and fuel efficiency fields. Outliers, especially for very low or very high values, were adjusted to maintain the validity of the analysis.

The analysis utilized descriptive statistics, including means, standard deviations, and frequency distributions, to encapsulate the overall trends in the data. Correlation analysis was performed to discover the relationships between the different attributes (e.g., price, engine capacity, and fuel efficiency), while regression models were constructed to determine the impact of these attributes on the sales units of the car models.

**Statistical Techniques**

The statistical methods employed in the analysis are as follows:

**Descriptive Statistics:**

Mean, median, mode, and standard deviation for quantitative variables such as sales units, price, engine size, and horsepower.

Frequency distributions to discover the common range of fuel efficiency and price across car models.

**Correlation Analysis**:

Pearson correlation coefficients were calculated to ascertain the strength and direction of the relationships between car features such as engine size, horsepower, fuel efficiency, and sales units.

**Multiple Linear Regression**:

Multiple regression was used in forecasting the units of car sales on the variables price, engine size, horsepower, and fuel efficiency. The model is useful for determining the impact these factors have on the market performance of every car model.

**Segmentation Analysis:**

Cars were segmented by make and model to determine performance trends within specific categories, with comparability across manufacturers and types of cars (sedans, SUVs, and trucks).

**Limitations**

Since the data used in this study is secondary data, it has the following drawbacks:

**Data Quality:** The accuracy and reliability of the data depend on the source, and any flaws in the original data can impact the analysis.

**Time Constraints:** The data does not represent real-time trends since it is based on past sales data. Therefore, recent market trends or shifts in consumer behaviour may not be represented fully.

**Missing Data:** Incomplete records, such as missing values for specific variables (e.g., fuel efficiency for certain car models), were addressed by employing appropriate imputation methods or by excluding such rows from analysis.

**Software**: Jamovi is the main software used to generate the analysis of the data points.

**Conclusion**

This research design outlines the use of secondary data in analysing the relationship between various car model factors and sales. By using descriptive statistics, correlation, and regression models, this research provides insightful data on how engine size, price, and fuel consumption influence buying behaviour in the automotive industry. Possible future research includes using primary data for more detailed investigation and up-to-date information regarding market trends**.**

**Variables and Measures**

The study analyses the following key variables:

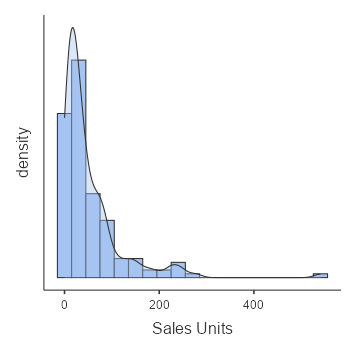
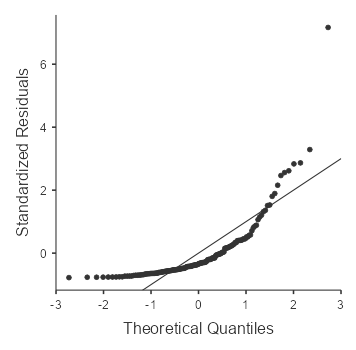
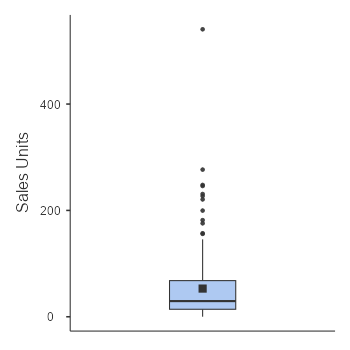
* **Price**: The retail price of each car model.
* **Engine Size**: The engine displacement, measured in litres.
* **Horsepower**: The engine power, measured in horsepower.
* **Fuel Capacity**: The volume of fuel the car can hold, measured in gallons or litres.
* **Sales Units**: The number of units sold per model in a given period.
* **Manufacturer and Model**: Identifiers for each vehicle, used to examine brand-related trends.

**Descriptives**

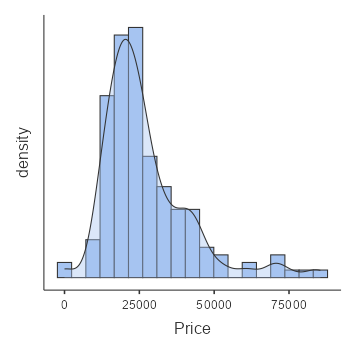
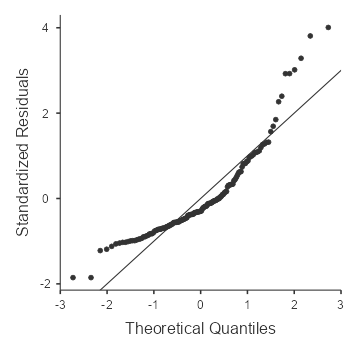
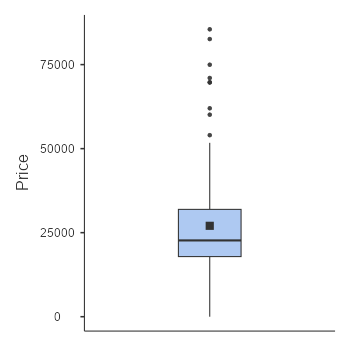
| Descriptives | | | | | |
| --- | --- | --- | --- | --- | --- |
|  | **Sales Units** | **Price** | **Engine size** | **Horsepower** | **Fuel capacity** |
| **N** | 157 | 157 | 157 | 157 | 157 |
| **Missing** | 0 | 0 | 0 | 0 | 0 |
| **Mean** | 53.0 | 27042 | 3.06 | 186 | 18.0 |
| **Median** | 29.4 | 22695 | 3.00 | 178 | 17.2 |
| **Standard deviation** | 68.0 | 14589 | 1.04 | 56.7 | 3.89 |
| **Range** | 540 | 85500 | 7.00 | 395 | 21.7 |
| **Minimum** | 0.110 | 0 | 1.00 | 55 | 10.3 |
| **Maximum** | 541 | 85500 | 8.00 | 450 | 32.0 |

**Plots**

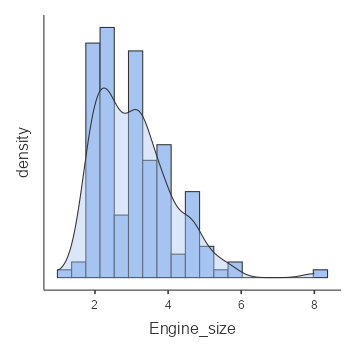
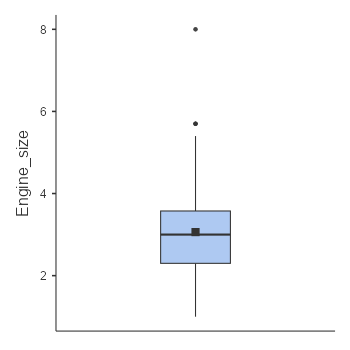
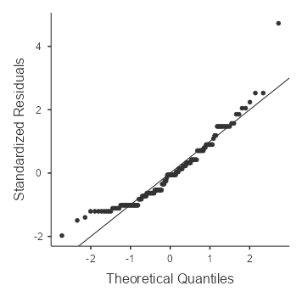
**Sales Units**

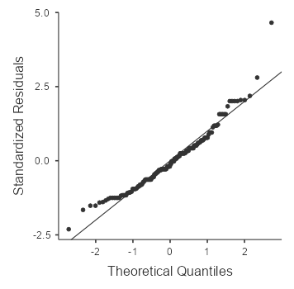


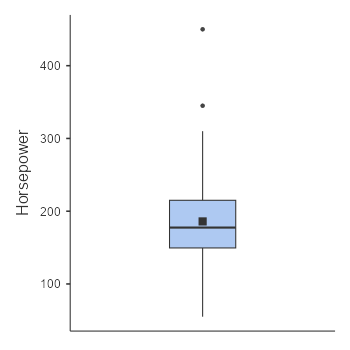
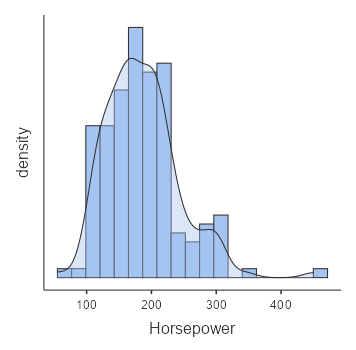
**Price**



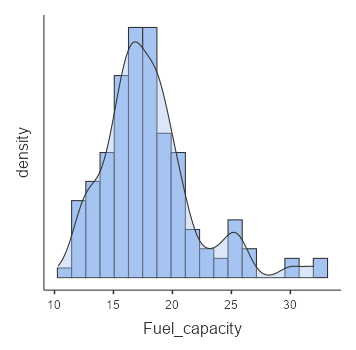
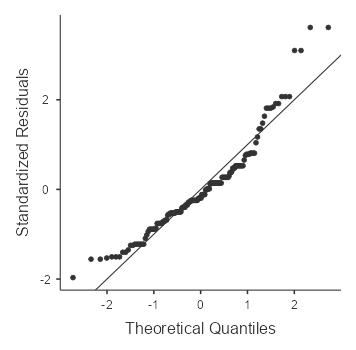
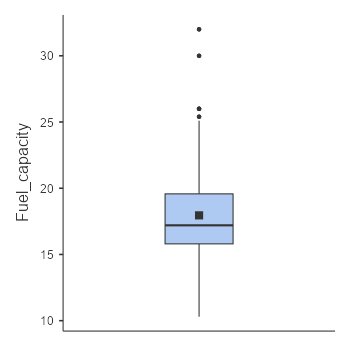
**Engine\_size**



**Horsepower**

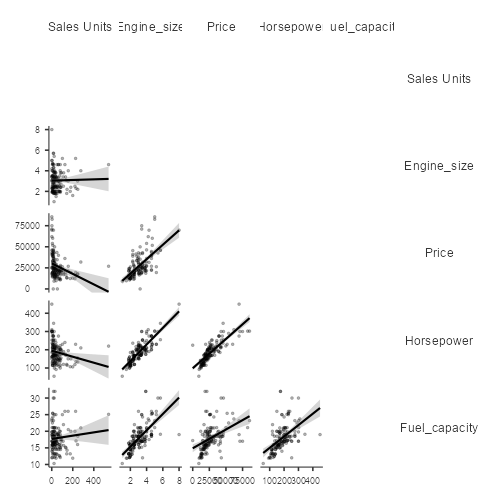


**Fuel\_capacity**



| Correlation Matrix | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | **Sales Units** | **Engine\_size** | **Price** | **Horsepower** | **Fuel\_capacity** |
| **Sales Units** | **Pearson's r** | — |  |  |  |  |
| **df** | — |  |  |  |  |
| **p-value** | — |  |  |  |  |
| **Engine size** | **Pearson's r** | 0.020 | — |  |  |  |
| **df** | 154 | — |  |  |  |
| **p-value** | 0.804 | — |  |  |  |
| **Price** | **Pearson's r** | -0.291 | 0.618 | — |  |  |
| **df** | 155 | 154 | — |  |  |
| **p-value** | <.001 | <.001 | — |  |  |
| **Horsepower** | **Pearson's r** | -0.198 | 0.837 | 0.820 | — |  |
| **df** | 154 | 154 | 154 | — |  |
| **p-value** | 0.013 | <.001 | <.001 | — |  |
| **Fuel capacity** | **Pearson's r** | 0.087 | 0.663 | 0.421 | 0.500 | — |
| **df** | 154 | 154 | 154 | 154 | — |
| **p-value** | 0.283 | <.001 | <.001 | <.001 | — |

**Plot**

****

**Summary & Insights**

**Sales Trends:**

* Higher price and higher horsepower seem to be related to lower sales.
* Fuel capacity and engine size don't seem to be a significant factor in sales.
* Vehicle Characteristics:
* Price, horsepower, fuel capacity, and engine size are all strongly correlated.
* Higher engines are related to higher horsepower, bigger fuel tanks, and higher price.
* Implications:
* If sales are to be maximized, makers might try to trade off price and horsepower to make the cars more attractive to purchasers.
* Luxury or high-performance vehicles (which tend to have large engines and high horsepower) likely target niche markets rather than high sales.

**INSIGHTS ON DESCRIPTIVES**

**1. Sales Units (Number of cars sold by each model)**

* Mean (Average): 53 units
* Median: 29.4 units
* Standard Deviation: 68.0 units (high variability in sales across models)
* Range: 0.11 to 541 units

**Interpretation:**

Sales numbers vary widely, as indicated by the huge standard deviation (68). Some models sell in large numbers (maximum = 541), while others sell very little (minimum = 0.11). The median (29.4) is much less than the mean, which suggests that most models sell fewer cars, but some big selling models pull the average up.

**2. Price (Vehicle price in dollars)**

* Mean Price: $27,042
* Median Price: $22,695
* Standard Deviation: $14,589 (wide range of prices)
* Range: $0 (maybe an error) to $85,500

**Interpretation:**

Car prices have a wide range. The average price ($27,042) is above the median ($22,695), indicating high-priced cars are skewing the average upwards. The highest price ($85,500) indicates luxury or premium vehicles are present. The lowest price of $0 could be an error in data entry.

**3. Engine Size (Litres)**

* Mean Engine Size: 3.06L
* Median Engine Size: 3.00L
* Standard Deviation: 1.04L
* Range: 1.00L to 8.00L

**Interpretation:**

Most cars have an engine size of around 3.0L, as the mean and median are nearly equal. The minimum engine size is 1.0L, and the maximum is 8.0L, which means the dataset includes both economy and high-performance cars.

**4. Horsepower (HP)**

* Mean Horsepower: 186 HP
* Median Horsepower: 178 HP
* Standard Deviation: 56.7 HP
* Range: 55 HP to 450 HP

**Interpretation:**

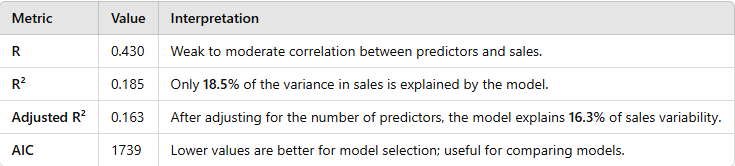
Horsepower varies greatly. The mean (186 HP) and median (178 HP) are quite close, suggesting a reasonably symmetric distribution. The minimum of 55 HP points to economy cars, whereas the maximum of 450 HP points to high-performance or sport car

**5. Fuel Capacity (Gallons)**

* Mean Fuel Capacity: 18.0 gallons
* Median Fuel Capacity: 17.2 gallons
* Standard Deviation: 3.89 gallons
* Range: 10.3 gallons to 32.0 gallons

**Interpretation:**

The average fuel tank capacity is around 18 gallons, with moderate spread. Low gas tanks (10.3 gallons) suggest small cars, while the highest (32 gallons) are most likely in trucks or SUVs with extended range capabilities.



* The model explains a small portion (18.5%) of the variation in sales, meaning other unaccounted factors significantly influence sales.
* The adjusted R² is slightly lower, suggesting that some predictors might not contribute strongly.

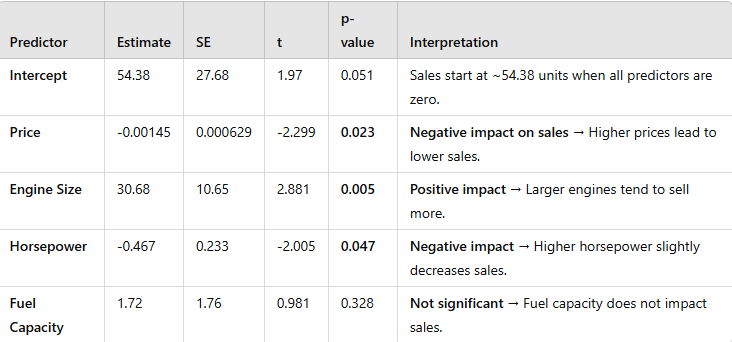
### ****2. Omnibus ANOVA Test (Overall Model Significance)****

The **F-test** checks if the predictors (Price, Engine Size, Horsepower, and Fuel Capacity) significantly improve the model:

* **F = 5.287**, **p = 0.023** → the overall model is statistically significant (p < 0.05), meaning at least one predictor influences sales.

| Omnibus ANOVA Test | | | | | |
| --- | --- | --- | --- | --- | --- |
|  | **Sum of Squares** | **df** | **Mean Square** | **F** | **p** |
| **Price** | 20603 | 1 | 20603 | 5.287 | 0.023 |
| **Engine size** | 32348 | 1 | 32348 | 8.301 | 0.005 |
| **Fuel capacity** | 3751 | 1 | 3751 | 0.962 | 0.328 |
| **Horsepower** | 15670 | 1 | 15670 | 4.021 | 0.047 |
| **Residuals** | 588445 | 151 | 3897 |  |  |
|  | | | | | |

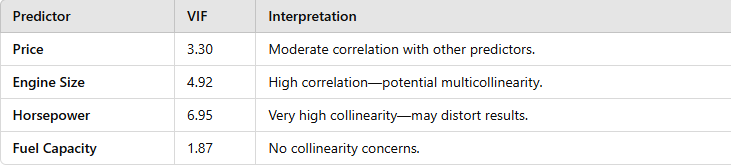
### ****3. Predictor Significance (Model Coefficients - Sales Units)****



**Insights:**

* Price has a negative effect on sales (p = 0.023). For every $1 increase in price, sales drop by 0.00145 units.
* Engine Size has a strong positive effect (p = 0.005) bigger engines sell more.
* Horsepower has a small negative effect (p = 0.047) higher horsepower models might not be as popular.
* Fuel Capacity is not significant (p = 0.328), meaning it doesn’t affect sales

**Assumption Checks - Collinearity (VIF Scores)**

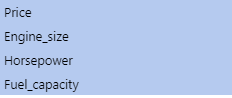
****

**Interpretation:**

* **Horsepower has high multicollinearity (VIF = 6.95)**, meaning it's highly correlated with other variables (likely Engine Size).
* A **VIF > 5** is concerning; consider removing one of the correlated variables.
* **Multicollinearity issue** there is the need to consider removing **Horsepower** or **Engine Size** to improve model reliability.
* **R² is low (18.5%)** other significant factors like brand reputation, marketing, features may be missing. This could account for the outstanding varations.

The regression equation is written as:

Sales Units=54.38−0.00145(Price)+30.68(Engine Size)−0.467(Horsepower)+1.73(Fuel Capacity)



| Model Fit Measures | | | | |
| --- | --- | --- | --- | --- |
| **Model** | **R** | **R²** | **Adjusted R²** | **AIC** |
| 1 | 0.294 | 0.0864 | 0.0805 | 1751 |
| 2 | 0.390 | 0.1522 | 0.1411 | 1742 |
| 3 | 0.404 | 0.1632 | 0.1467 | 1742 |
| 4 | 0.430 | 0.1849 | 0.1634 | 1739 |
|  | | | | |

 The **overall strength of the models** is measured by **R² (coefficient of determination)** and **Adjusted R²**:

* **R²** throws light on how much of the variance in sales units is explained by the independent variables (price, engine size, fuel capacity, etc.).
* **Adjusted R² this** throws light on the adjustment for the number of predictors, making it more reliable when comparing models.
* **AIC (Akaike Information Criterion)** is a measure of model quality (lower values indicate a better fit with fewer parameters).

### ****Interpretation of Model Fit Measures:****

* **Model 1(Price):** R² this only explains 0.086 **of the variance** in sales units is explained (weak model).
* **Model 2(Price, Engine Size):** R² this explains 0.152, **15.2% of the variance** is explained (moderate improvement).
* **Model 3(Price, Engine Size, Horsepower):** R² = 0.163, **16.3% of the variance** is explained.
* **Model 4(Price, Engine Size, Horsepower, Fuel Capacity):** R² = 0.184, **18.5% of the variance** is explained (best R² value).

**Insights on Best Model**

* **Model 4 has the highest R² (0.185) and the lowest AIC (1739),** meaning it explains the most variance while balancing model complexity.
* However, **Model 2 has the least collinearity**, meaning it may be more stable.

**Conclusion:**

* If the focus is on the **variance explained**, **Model 4 is best**.
* If the focus is on the s**tability with lower collinearity**, **Model 2 is a better choice**.

**ALTERNATIVE MODEL IMPROVEMENT (M2)**

| Omnibus ANOVA Test | | | | | |
| --- | --- | --- | --- | --- | --- |
|  | **Sum of Squares** | **df** | **Mean Square** | **F** | **p** |
| **Price** | 109578 | 1 | 109578 | 27.4 | <.001 |
| **Engine size** | 47487 | 1 | 47487 | 11.9 | <.001 |
| **Residuals** | 612102 | 153 | 4001 |  |  |
|  | | | | | |

| Model Coefficients - Sales Units | | | | |
| --- | --- | --- | --- | --- |
| **Predictor** | **Estimate** | **SE** | **t** | **p** |
| Intercept | 51.35896 | 15.73 | 3.26 | 0.001 |
| Price | -0.00234 | 4.47e-4 | -5.23 | <.001 |
| Engine size | 21.31158 | 6.19 | 3.45 | <.001 |

### Assumption Checks

| Collinearity Statistics | | |
| --- | --- | --- |
|  | **VIF** | **Tolerance** |
| **Price** | 1.62 | 0.618 |
| **Engine size** | 1.62 | 0.618 |

**ADVANTAGES OF IMPLEMENTING MODEL 2**

* **There are no concerns for multicollinearity (VIF values are low: Price = 1.62, Engine Size = 1.62)**
* Shows **Higher Adjusted R² (0.1411)** than the **Model 3** without many variables
* **Both predictors are statistically significant (p < 0.001 for Price, p < 0.001 for Engine Size)**
* **Simpler model = easier interpretation & better generalizability**

**ADDITIONAL INSIGHTS ON MODEL 2**

The overall strength of **Model 2** can be assessed using the **R² and Adjusted R²**:

### ****Model Strength Indicators:****

* **R² = 0.1522** the model explains **15.22%** of the variation in **Sales Units** based on **Price** and **Engine Size**.
* **Adjusted R² = 0.1411** after accounting for the number of predictors, the model still explains **14.11%** of the variance.
* **F-statistic = 11.87, p < .001** the model is statistically significant, meaning at least one predictor has a meaningful impact on sales.

### ****Further Interpretation:****

* The model is relatively **weak to moderate** in terms of explanatory power (**R² = 15.22% is low but still meaningful**).
* Price and Engine Size **significantly influence Sales Units**, but other unaccounted factors may explain more of the variation.
* The model suggests that while these factors matter, there are other **external factors** like brand reputation, marketing, consumer preference that likely play a key role in driving sales.

### ****Conclusion of the Project****

After conducting a thorough analysis using **linear regression**, the following key findings and conclusions can be drawn:

1. **Key Predictors of Sales Units:**

* Price and Engine Size are the two most significant drivers of sales.
* Price has a negative correlation with sales, i.e., the higher the price, the lower the units sold (p < 0.001).
* Engine Size positively influences sales since it indicates that vehicles with larger engines sell more units (p < 0.001).

1. **Model Strength and Limitations:**

* **The ideal model was Model 2 since it maximizes prediction capability and minimizes collinearity issues.**
* **The model explains 15.22% of the sales variation (R² = 0.1522), which means that while Price and Engine Size are important, other variables outside the model also contribute to sales.**
* **Hypothetically significant variables like brand reputation, advertising, gas mileage, or consumer demand were ignored but can contribute to the unexplained variance.**

1. **Practical Implications:**

* Pricing Strategy: As price lowers sales, manufacturers and dealerships can opt for competitive pricing strategies to increase sales.
* Product Development: Increased engine sizes lead to more sales, suggesting that buyers want high-performance cars.
* Future Research: Future research incorporating brand perception, marketing spend, fuel efficiency, and technological details may further refine sales projections

### ****Final Thoughts:****

While the model does provide some insight into how Engine Size and Price affect sales, more research has to be done in order to develop a more generalizable predictive model. Future analysis may explore interaction effects, additional predictors, and machine learning approaches for greater accuracy.

**References**

* Jamovi Project (2024) - Statistical software for data analysis.
* R Core Team (2024) - A language and environment for statistical computing.
* Islam, R., Ghani, A. B. A., Mahyudin, E., & Manickam, N. (2017). Determinants of factors that affecting inflation in Malaysia. *International Journal of Economics and Financial Issues*, *7*(2), 355-364.
* Korkmaz, M., Germir, H. N., Yücel, A. S., Aytaç, A., Şen, E., Yorulmaz, M., & Şahbudak, E. (2015). Analysis of The Factors Affecting The Sales in The Automotive Industry. *IIB International Refereed Academic Social Sciences Journal*.
* Guan, W. (2023). Analysis of the Main Influencing Factors of Automobile Sales Based on Empirical Analysis. *Advances in Economics, Management and Political Sciences*, *6*, 515-525.
* Farahani, D. S., Momeni, M., & Amiri, N. S. (2016). Car sales forecasting using artificial neural networks and analytical hierarchy process. *Data Analytics*, *69*.
* Kurilova, A. A., & Kurilov, K. Y. (2016). Main Factors Influencing the Fall of Sales in the Russian Automobile Market. *Actual Probs. Econ. & L.*, 102.
* Shahabuddin, S. (2009). Forecasting automobile sales. *Management Research News*, *32*(7), 670-682.
* MUNYWOKI, H. (2017). *Factors Affecting the Sale of New Vehicles in The Motor Vehicle Industry: A Case Study of Simba Corporation Limited, Nairobi* (Doctoral dissertation, MUA).