**Analysis of Used Cars Dataset**

**Muzammal Javed**  
**Student ID: 23073275**

**GitHub Repository:**

**Introduction**

This report is about analyzing a dataset of used cars to see what factors affect their selling price. The dataset includes details like the car brand, kilometers driven, fuel type, ownership history, and selling price. I wanted to figure out if there are any patterns or trends, especially to understand how things like the car’s mileage and brand impact its price.

**Descriptive Statistics**

First, I looked at some basic statistics of the dataset to get a general overview:

|  |  |  |
| --- | --- | --- |
| **Statistic** | **km\_driven** | **selling\_price** |
| **Count** | 8128 | 8128 |
| **Mean** | 69,819.51 | 638,271.80 |
| **Standard Dev** | 56,550.55 | 806,253.40 |
| **Minimum** | 1 | 29,999 |
| **25% Quartile** | 35,000 | 254,999 |
| **Median** | 60,000 | 450,000 |
| **75% Quartile** | 98,000 | 675,000 |
| **Maximum** | 2,360,457 | 10,000,000 |

**Visualizations and Analysis**

**1. Correlation Matrix Heatmap**

The correlation matrix shows a weak negative correlation (-0.23) between **km\_driven** and **selling\_price**.

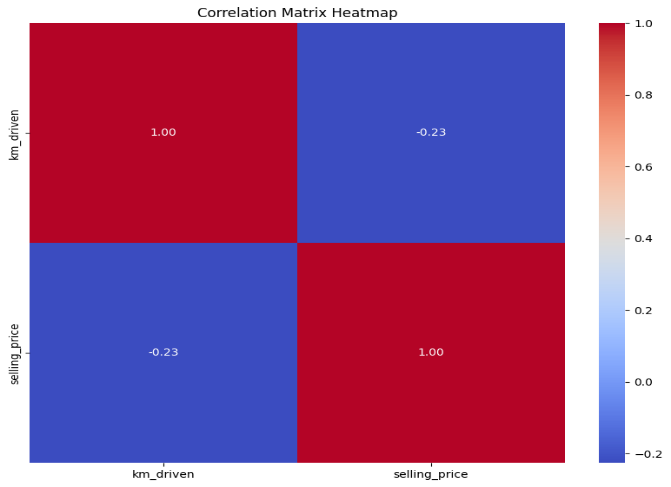


Figure 1Correlation Matrix Heatmap

**2. Average Selling Price by Brand (Bar Chart)**

The bar chart for average selling price by brand reveals some clear patterns:

* **Luxury brands** like **BMW**, **Mercedes-Benz**, and **Land Rover** have much higher average prices, often over 4 million.
* More common brands, like **Maruti**, **Hyundai**, and **Honda**, have much lower average prices, showing that they’re more affordable.

A graph of a number of blue bars

Description automatically generated

Figure 2Average Selling Price by Brand (Bar Chart)

**3. Selling Price vs Kilometers Driven (Scatter Plot)**

The scatter plot shows the relationship between **km\_driven** and **selling\_price**. There’s a slight downward trend, where cars with higher mileage tend to have lower prices. However, the points are pretty widespread , which means mileage alone doesn’t explain much of the price variation

A graph of a number of blue dots

Description automatically generated

Figure 3Selling Price vs Kilometers Driven (Scatter Plot)

**Conclusion**

To sum up, **brand** and **kilometer driven** seem to be important factors when it comes to the selling price of a used car. Luxury brands tend to hold their value better, while cars with more kilometers on them are usually cheaper. However, mileage only has a weak impact on price, so it’s clear that other factors (like brand reputation, fuel type, and ownership history) play a big role as well.