

Statistics and Trends

Exploratory Data Analysis of Cars Dataset

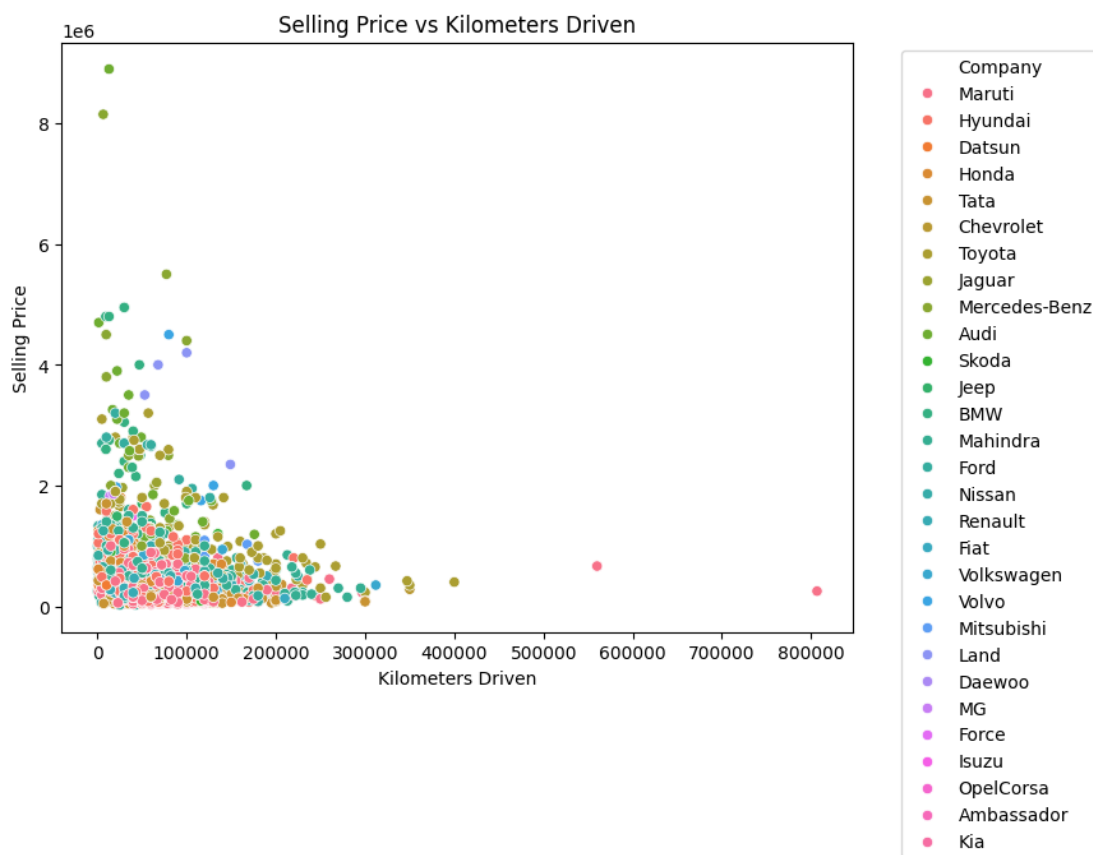
Introduction:

In this report, we explore a dataset containing information about car prices. We aim to uncover insights about the relationship between numerous factors such as the car's age, mileage, fuel type, and selling price. Through visualizations and statistical analysis, we aim to comprehensively understand the dataset.

[\[https://www.kaggle.com/datasets/akshaydattatraykhare/car-details-dataset\]](https://www.kaggle.com/datasets/akshaydattatraykhare/car-details-dataset)

Exploratory Data Analysis:

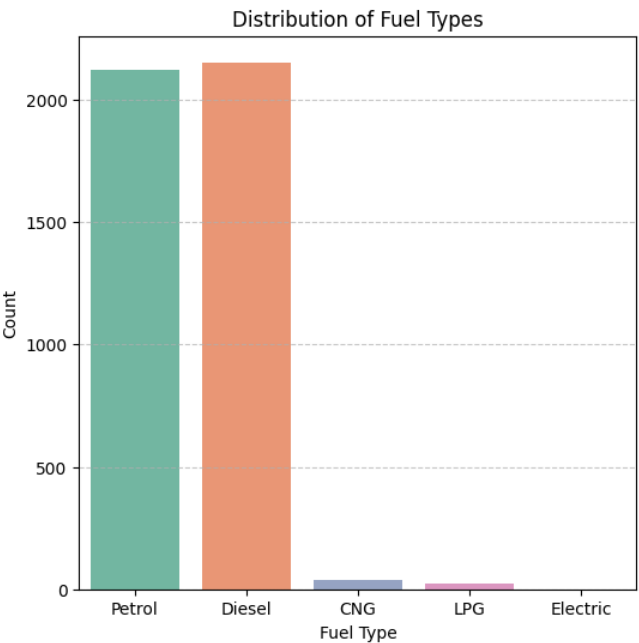
1. Scatterplot: Selling Price vs Kilometres Driven



We observe a negative correlation between these two variables, indicating that as the kilometres driven increase, the selling price tends to decrease. This is intuitive, as higher mileage typically implies more wear and tear on the car, leading to a decrease in its value.

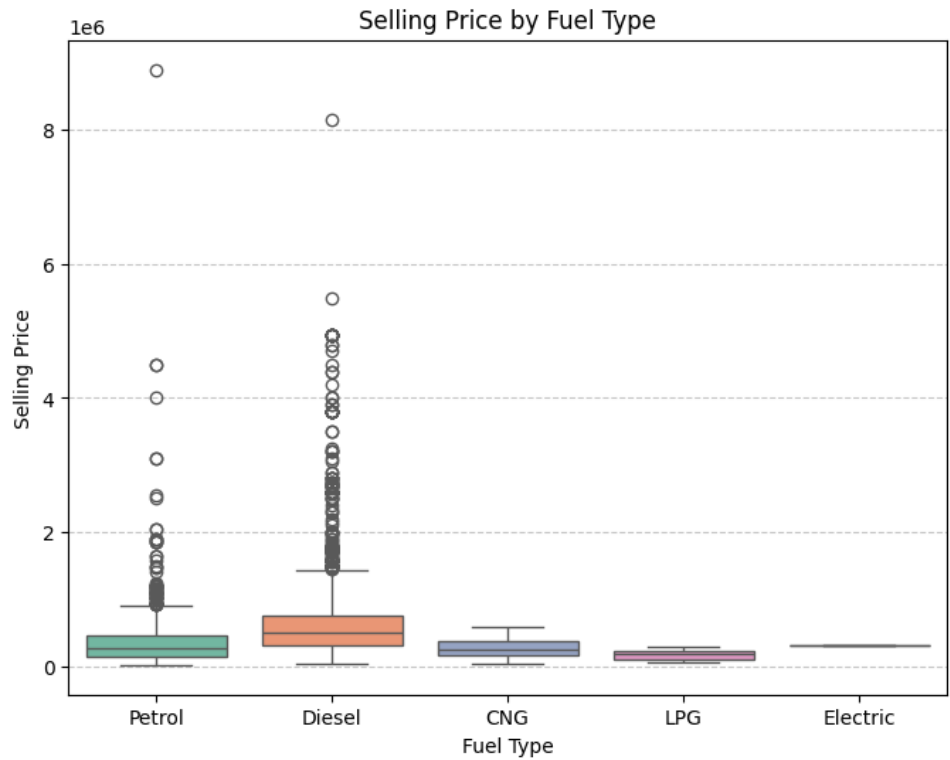
2. Histogram: Distribution of Fuel Types

We observe that petrol cars are the most prevalent in the dataset, followed by diesel and CNG. This information could be valuable for understanding market trends and consumer preferences.



3. Boxplot: Selling Price by Fuel Type

We notice variations in the selling prices among different fuel types, with diesel cars generally having higher median selling prices compared to petrol and CNG cars. This indicates that fuel type is a significant factor influencing the selling price of a car.



Statistical Analysis:

The dataset spans a range of years from 1992 to 2020. The mean year of the cars listed for sale is around 2013, indicating that the majority of cars in the dataset are relatively modern. The standard deviation of approximately 4.22 suggests moderate variability in the years of the cars. The selling prices of cars vary widely, as indicated by the mean of around 504,127 and the standard deviation of approximately 578,548. The minimum selling price is 20,000, while the maximum is a substantial 8,900,000, showcasing the diverse price range within the dataset. The median (50th percentile) selling price is 350,000, providing a measure of central tendency that is less influenced by extreme values. The interquartile range (IQR) between the 25th and 75th percentiles is substantial, indicating significant variability in selling prices. The dataset includes cars with a wide range of mileage, as suggested by the mean of approximately 66,216 and a standard deviation of around 46,644. The minimum recorded kilometres driven is 1, while the maximum is a high 806,599, highlighting the diversity in the mileage of the listed cars. The median (50th percentile) mileage is 60,000, indicating that half of the cars have mileage below this value.

Conclusion:

In conclusion, this analysis provides valuable insights into the factors influencing car prices. We have explored the relationship between selling price and kilometres driven, analysed the distribution of fuel types, and investigated the impact of fuel type on selling prices. By combining visualizations and statistical analysis, we have gained a deeper understanding of the dataset, which can inform decision-making processes in the automotive industry.

Name: Sravanth Baratam

Student ID: 23001152

Code: https://github.com/sravanth-space/statistics_and_trends