

## **Car Price Estimation Analysis**

In this analysis, we aimed to estimate the selling price of cars based on various predictors such as the year of manufacture, kilometers driven, fuel type, and transmission type. The dataset used for this analysis was sourced from Car Dekho.

### **Outcome of the Exploratory Data Analysis (EDA):**

The EDA involved examining histograms for the selected variables, calculating descriptive statistics, comparing probability mass functions (PMFs), creating a cumulative distribution function (CDF) for the selling price, and plotting scatter plots to analyze correlations. From the histograms, we observed the distribution of car years, kilometers driven, and selling prices. The PMF comparison highlighted the difference in car years before and after 2010. The CDF plot provided insights into the distribution of selling prices, showing the proportion of cars sold at or below a given price. Scatter plots revealed a positive correlation between the year of manufacture and selling price, and a negative correlation between kilometers driven and selling price.

### **Missed Aspects During the Analysis:**

One aspect that might have been missed during the analysis is the exploration of potential non-linear relationships between the predictors and the selling price. Additionally, interactions between variables, such as the combined effect of the year and kilometers driven on the selling price, were not investigated.

### **Potential Additional Variables:**

Including variables such as the car's make and model, engine size, horsepower, and the number of previous owners could have provided a more comprehensive analysis. These factors are likely to influence the selling price of a car significantly.

### **Assumptions and Incorrect Assumptions:**

The analysis assumed linear relationships in the regression models, which might not be entirely accurate for real-world data. It was also assumed that the selected variables were independent of each other, which may not be the case, as certain variables like year and kilometers driven could be correlated.

### **Challenges and Areas of Confusion:**

One challenge faced during the analysis was determining the appropriate statistical methods and visualizations to use for the given dataset. Additionally, interpreting the results of the regression analysis and understanding the implications of the coefficients in the context of the data required careful consideration.

## **Conclusion**

In conclusion, while the analysis provided valuable insights into the factors affecting car selling prices, there is room for improvement by incorporating additional variables, exploring non-linear relationships, and addressing the assumptions made during the analysis. Understanding these aspects could lead to a more accurate estimation of car selling prices and a deeper understanding of the automotive market.