

## Car Price Prediction Model

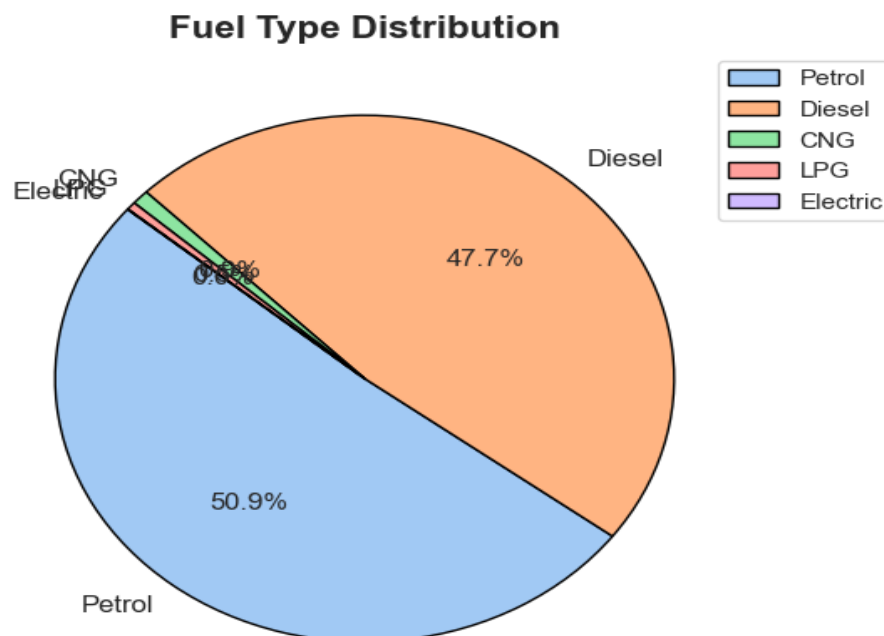
This project aims to predict the selling price of used cars using a comprehensive machine learning approach. The model is built to help both buyers and sellers in the used car market make data-driven decisions by providing an accurate estimation of a car's value.

## Methodology

The analysis and model development followed a robust data science pipeline:

- **Data Collection & Integration:** The project utilizes two distinct datasets sourced from popular car portals. These datasets were merged and cleaned to form a unified data source.
- **Data Preprocessing:** We handled missing values through a **median imputation strategy** for the Present\_Price column. Categorical variables, such as Owner and Fuel\_Type, were preprocessed to a numerical format suitable for machine learning algorithms.
- **Exploratory Data Analysis (EDA) :**

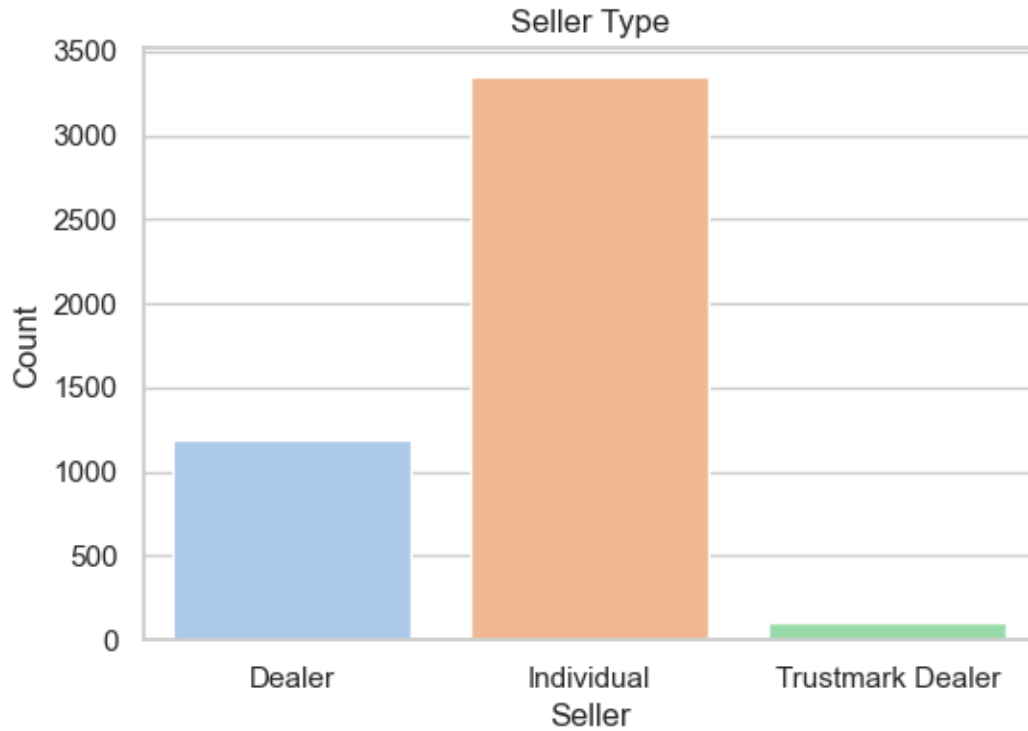
The Fuel\_Type column was analyzed and a Pie Chart was plotted to show the proportions.



Observation: Most vehicles use Petrol, followed by Diesel, while the percentage of electric or other fuel types is very low.

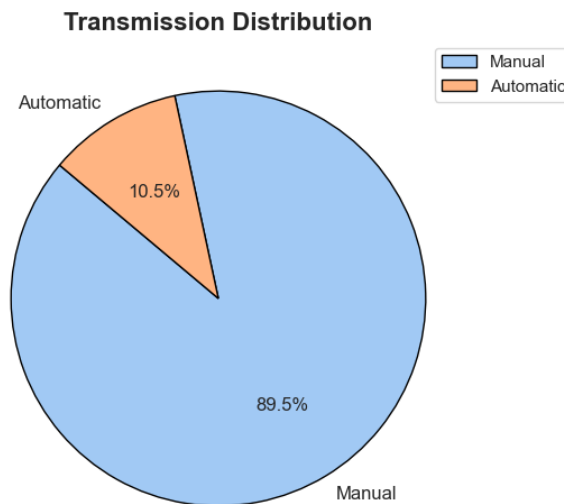
The Seller\_Type column was analyzed to show the distribution of seller categories.

**Observation:** The majority of vehicles are sold by **Individual sellers**, followed by **Dealers**, and then **Trustmark Dealers**.



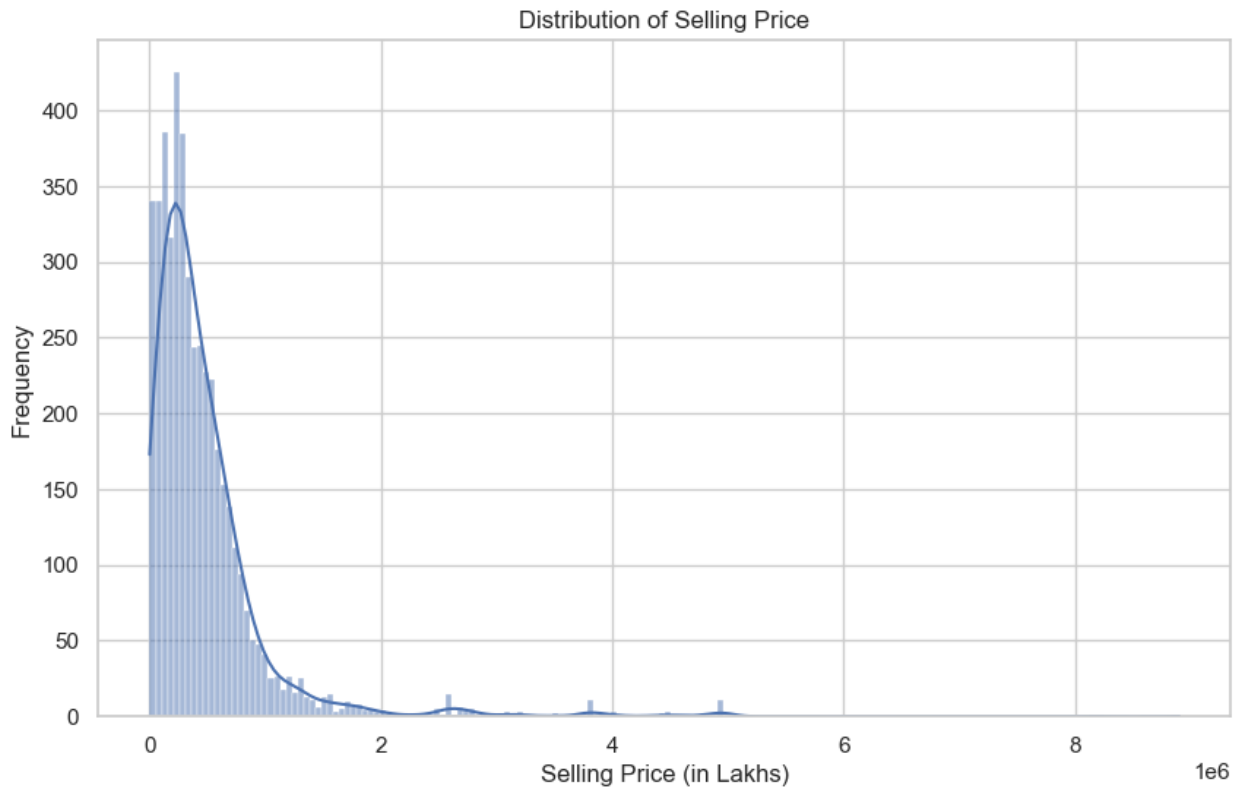
The Transmission column was analyzed to show the proportion of transmission types.

**Observation:** Most vehicles are **Manual (89.5%)**, while **Automatic** vehicles account for only **10.5%**.



The Selling\_Price column was analyzed using a **histogram with a KDE curve** to visualize its distribution.

**Observation:** The selling prices are **left-skewed**, with most vehicles priced in the higher range, and fewer vehicles in the lower range.



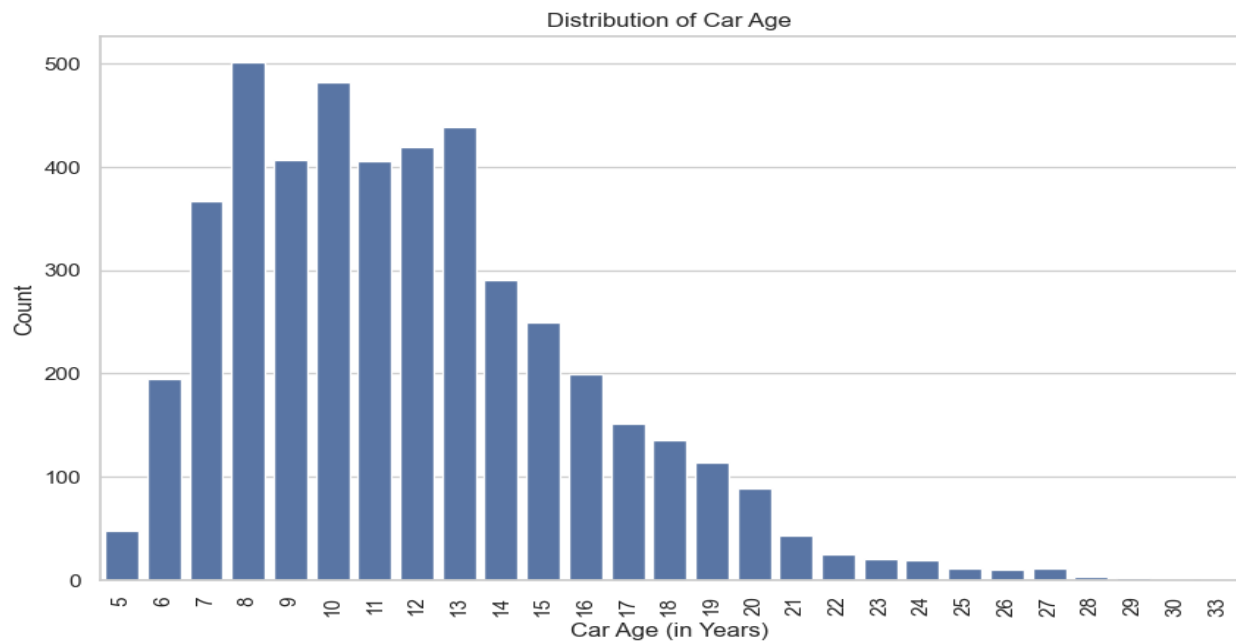
- **Feature Engineering:** New, highly predictive features were engineered from the raw data, including the **age of the car** and the **ratio of kilometers driven to the car's age**.
- **Model Building:** A **Gradient Boosting Regressor** was chosen as the core machine learning model due to its strong performance on regression tasks.
- **Hyperparameter Tuning:** The model was optimized using a **Grid Search Cross-Validation** to find the best combination of hyperparameters, ensuring peak performance and generalization.

## Results

The final model was evaluated on a held-out test set, achieving a strong **R-squared score of 0.878**. This indicates that the model can explain approximately 87.8% of the variance in the car's selling price, demonstrating its high predictive accuracy.

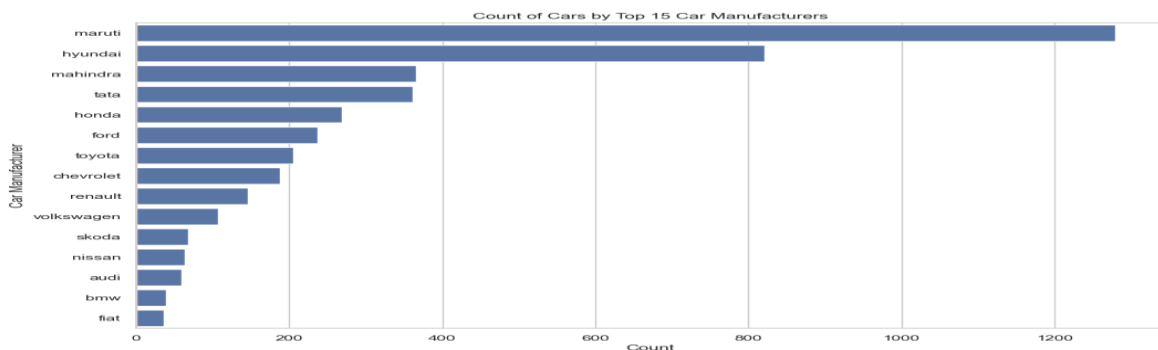
The Age column was analyzed using a **count plot** to visualize the distribution of vehicle ages.

**Observation:** Most cars in the dataset are **8 and 10 years old**, while very few cars are **29, 30, or 33 years old**.



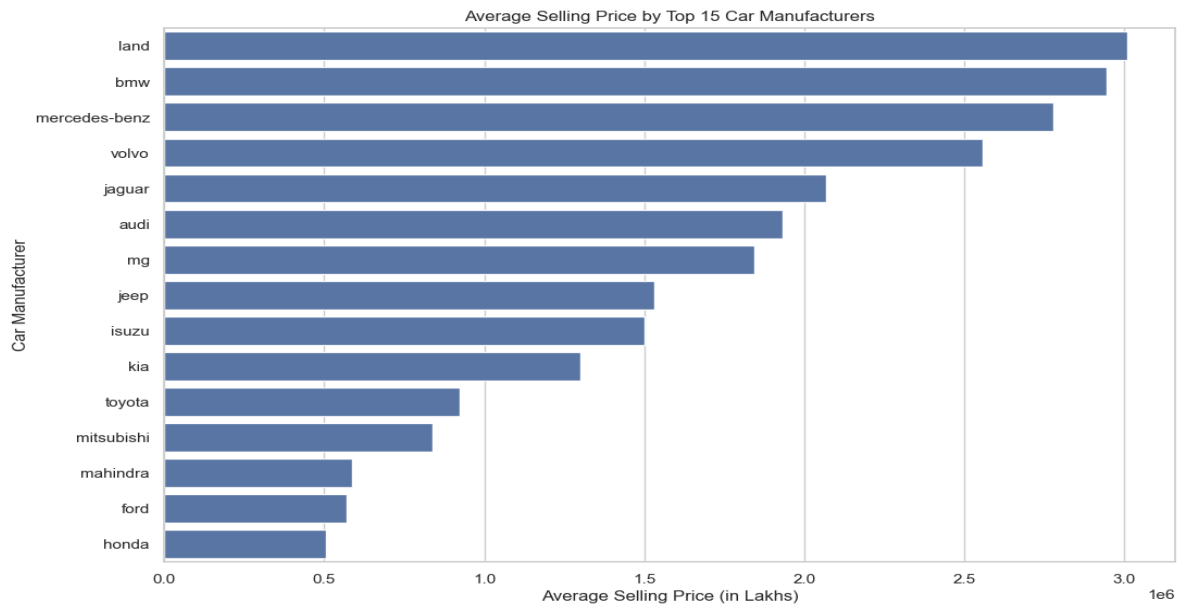
The Car\_Manufacturer column was analyzed using a **count plot** to visualize the top 15 car manufacturers by number of vehicles.

**Observation:** The most common manufacturers are **Maruti, Hyundai, Mahindra, Tata, Honda, Ford, and Toyota**, with Maruti leading the count. Other manufacturers appear much less frequently.



The Selling\_Price column was analyzed to calculate the **average selling price** for each car manufacturer, and the top 15 were visualized using a **bar plot**.

**Observation:** The most expensive brands on average are **Land Rover, BMW, Mercedes-Benz, Volvo, Jaguar, and Audi**, while other brands have lower average prices.



The dataset was analyzed to understand how **car age** affects selling price and depreciation. Three line plots were created:

### 1. Average Selling Price vs Car Age

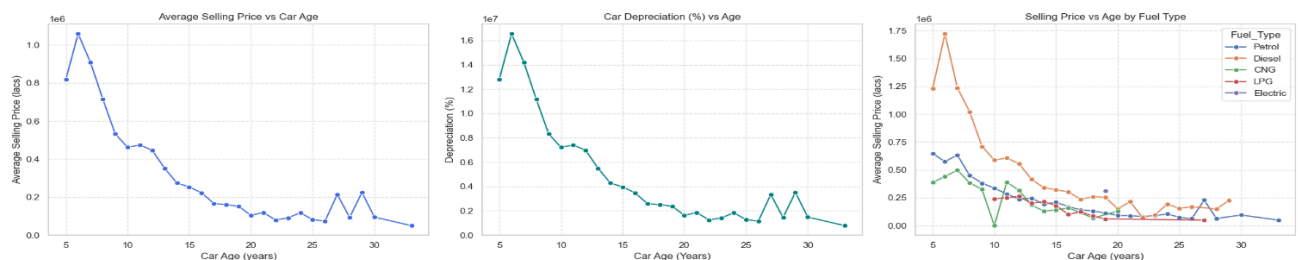
**Observation:** The average selling price **decreases as the car gets older**, showing a typical depreciation pattern.

### 2. Depreciation (%) vs Car Age

**Observation:** Depreciation **increases with car age**, indicating that older cars lose a higher percentage of their original value.

### 3. Selling Price vs Age by Fuel Type

**Observation:** When segmented by Fuel\_Type, **Petrol cars** generally maintain higher prices compared to Diesel or other fuel types at similar ages. The downward trend is visible across all fuel types.



**-- Team members --**

**Name: Hagar Mohamed Sayed**

**Name: Berehan AboElezz Mohamed**