Car Price Prediction Project Summary

1. Introduction

In the competitive used car market, accurately estimating the price of pre-owned vehicles is critical for dealers and individual sellers.
This project focuses on building a Car Price Prediction Model using Linear Regression, leveraging real-world features such as car brand,
mileage, fuel type, and ownership history to predict the selling price of used cars.

2. Exploratory Data Analysis (EDA) Insights

- Selling Price: Right-skewed distribution; luxury car outliers exist.
- Year of Manufacture: Newer cars tend to be priced higher.
- Mileage: More mileage reduces car value.
- Fuel Type: Diesel cars often have higher resale values than petrol.
- Transmission: Automatic cars typically cost more.
- Owner Type: First-owner cars sell at a premium.

Correlation heatmap showed significant relationships between features and price, with no multicollinearity.

3. Data Preprocessing

- One-Hot Encoding for categorical features.
- StandardScaler for numerical features.
- Split data into 80% training and 20% testing sets.

4. Model Development

- Model: Linear Regression
- Training on processed dataset.

5. Model Results

Metrics:

- MAE: ~1.25 Lakhs
- MSE: ~4.3 Lakhs²
- RMSE: ~2.07 Lakhs
- R^2 Score: ~0.86

Interpretation:

The model explains 86% of the variance in car prices, performing well for a baseline linear model.

6. Conclusion

Strengths:

- Good predictive accuracy for mid-range cars.
- Key factors: Year, Mileage, Owner Type.

Limitations:

- Struggles with luxury car outliers.
- Linear assumptions limit high-end car predictions.

Future Work:

- Use Random Forest or XGBoost.
- Apply feature selection and outlier removal.
- Explore non-linear modeling techniques.

7. Deliverables

- Jupyter Notebook with code and outputs.
- This PDF report summarizing the project.