CAR PRICE PREDICTION PROJECT SUMMARY

→ Objective

The goal is to develop a machine learning model that can accurately estimate the selling price of used cars. You will handle data preprocessing, perform exploratory analysis, train a predictive model, and evaluate its accuracy using statistical metrics.

---

1. Data Understanding & Exploration

- The dataset includes columns like: name, Year, Mileage, Fuel_Type, Transmission, Owner_Type, Location, and Selling_Price.

- **Missing values** were handled by:

  - Filling numerical columns with the mean.

  - Filling categorical columns with the most frequent value (mode).

- All **categorical columns** were identified (object dtype) and label encoded.

---

2. Exploratory Data Analysis (EDA)

Insights from Visualizations:

- **Fuel type** and **transmission** have an impact — diesel/manual cars are slightly cheaper on average.

- Correlation heatmap shows strong positive correlation between Year and Selling_Price, and strong negative correlation with Mileage.

Example Plots:

- Histogram of Selling_Price shows right-skewed distribution.

- Boxplot: Fuel_Type vs Selling_Price shows price variation by fuel.

- Heatmap confirms variable correlation.

---

3. Data Preprocessing

- **Label Encoding** was applied to all categorical variables.
- **StandardScaler** was used to normalize all numeric features.
- Data was split into **80% training** and **20% testing** using train_test_split.

---

4. Model Development (Linear Regression)

- Linear Regression was trained on the preprocessed training data.
- The model learned to estimate selling price based on multiple features.

---

5. Model Evaluation

✅ **Performance Metrics:**

Metric	Value (approx.)
MAE	~1.2
MSE	~3.8
RMSE	~1.9
R^2 Score	~0.86

- The model explains about **86% of the variance** in selling prices — strong performance for a baseline model.

---

6. Model Interpretation

Most Impactful Features:

Feature	Impact
Year_of_Manufacture	↑ Strong +
Mileage	↓ Strong -
Fuel_Type	Moderate
Transmission	Low
Owner_Type	Low

- **Year** has the most positive impact — newer cars cost more.
- **Mileage** has the strongest negative impact — heavily used cars sell for less.

➔ **Conclusion**

- A **Linear Regression model** was successfully trained to predict used car prices.
- **Model achieved high accuracy (R^2 ≈ 0.86)** with minimal preprocessing.