# Car Price Prediction - Report

## Project Title:

Building a Car Price Prediction Model - A Machine Learning Approach

## 1. Overview

This project focuses on predicting the selling price of used cars using linear regression. Real-world factors like brand, fuel type, mileage, and more were considered. The dataset was cleaned, analyzed, and used to train a model to forecast car prices with statistical accuracy.

## 2. Exploratory Data Analysis (EDA)

- Top brands: Maruti, Hyundai, Honda dominate the dataset.
- Price trends:
  * Older cars typically have lower resale prices.
  * Diesel cars slightly higher-priced on average.
- Outliers: Cars priced above Rs. 25L or driven over 200,000 km were rare and excluded.
- Visuals:
  * Heatmap showed strong negative correlation between age and price.
  * Boxplots revealed manual transmission cars are more frequent but automatic variants are more expensive.

## 3. Data Preprocessing

- Categorical Encoding: Label Encoding (Transmission, Fuel, Owner), OneHot (Brand, Location).
- Scaling: StandardScaler applied to mileage and year.
- Train-Test Split: 80:20 ratio.

## 4. Model Development

- Algorithm: Linear Regression
- Training R2 Score: 0.84
- Test R2 Score: 0.81

## 5. Model Evaluation

# Car Price Prediction - Report

| Metric | Value |
|---|---|
| MAE | Rs. 71,250 |
| MSE | Rs. 9.8 x 10^9 |
| RMSE | Rs. 99,000 |
| R2 Score | 0.81 |

## 6. Conclusion

The linear regression model accurately predicts used car prices based on key attributes. Car age, mileage, and brand were most influential. While good for initial estimates, the model could be improved by removing more outliers or using tree-based models like Random Forest for better accuracy.