# Car Price Prediction Model Report

1. Introduction

## This report walks through how we built a simple machine learning model to predict used car prices.

The goal was to understand which factors influence a car's selling price and to build a model that can make fairly accurate predictions.

2. Understanding the Data

## We used a dataset with 4,340 car listings and 8 columns, which included:

- Year of manufacture

- Selling price

- Kilometers driven

- Fuel type, Seller type, Transmission, and Ownership history

## Clean Data:

- No missing values

- Proper data types: numerical and categorical

## What We Noticed:

- Newer cars tend to sell for more

- Higher mileage generally lowers the price

- Most cars were recent with low mileage

- Few high-end luxury outliers

## Visual Insights:

- Scatter plots: Newer cars = higher price, More km = lower price

- Box plots: Diesel > Petrol, Dealers > Individuals, Automatic > Manual, First-owner cars get higher resale

3. Cleaning and Preparing the Data

## We cleaned and prepped the data:

- Dropped 'name' column due to too many unique values

- Removed outliers (luxury/high-mileage cars)

- Encoded categorical features

- Scaled numerical features

- Split data into 80% training and 20% testing

4. Building the Model

We used a Linear Regression model - simple, fast, and easy to interpret.

5. Model Performance

Before and After Cleaning:

| Metric | Before | After |
|---|---|---|
| Mean Absolute Error | Rs.2,19,541 | Rs.1,25,693 |
| Root Mean Square Error | Rs.4,26,536 | Rs.1,68,867 |
| R-squared Score | 0.40 | 0.54 |

## Improvements:

- Better accuracy

- Lower errors

- Higher variance explained by the model

6. Takeaways & What's Next

## Key Takeaways:

- Clean data = better results

- Even simple models can perform well

- Outlier removal and encoding made a big impact