CarDekho Dataset - Preprocessing & Exploratory Data Analysis

Prepared for: Assignment Submission

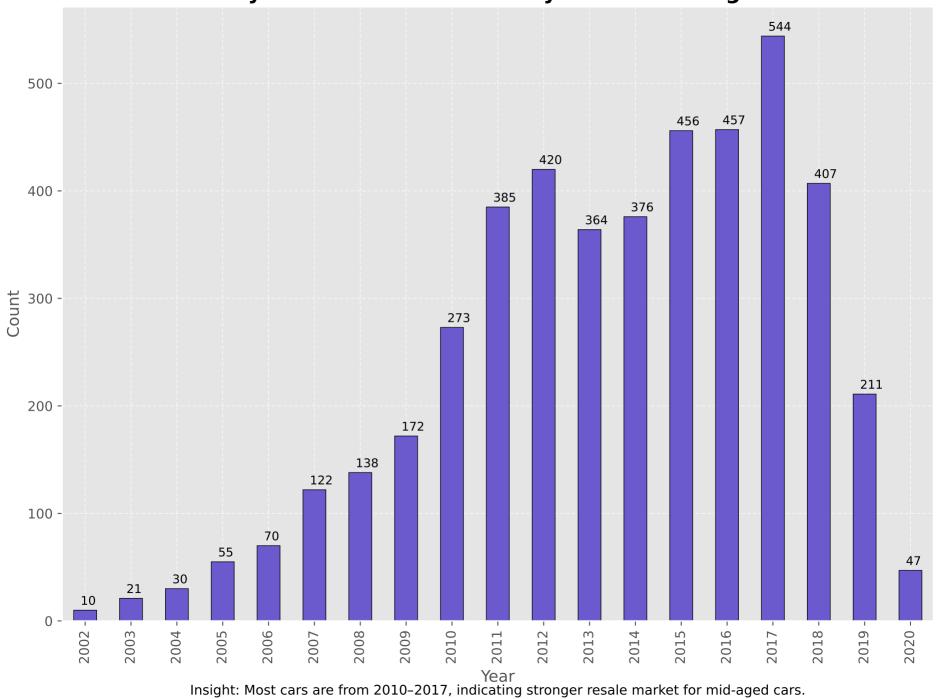
Contents:

- 1. Data Overview & Preprocessing Summary
- 2. Analysis 1: Year-wise Distribution of Cars
- 3. Analysis 2: Average Selling Price by Fuel Type
- 4. Analysis 3: Engine Size vs Selling Price
- 5. Analysis 4: Seller Type vs Average Selling Price
- 6. Analysis 5: Transmission Type Impact on Price
- 7. Correlation Heatmap

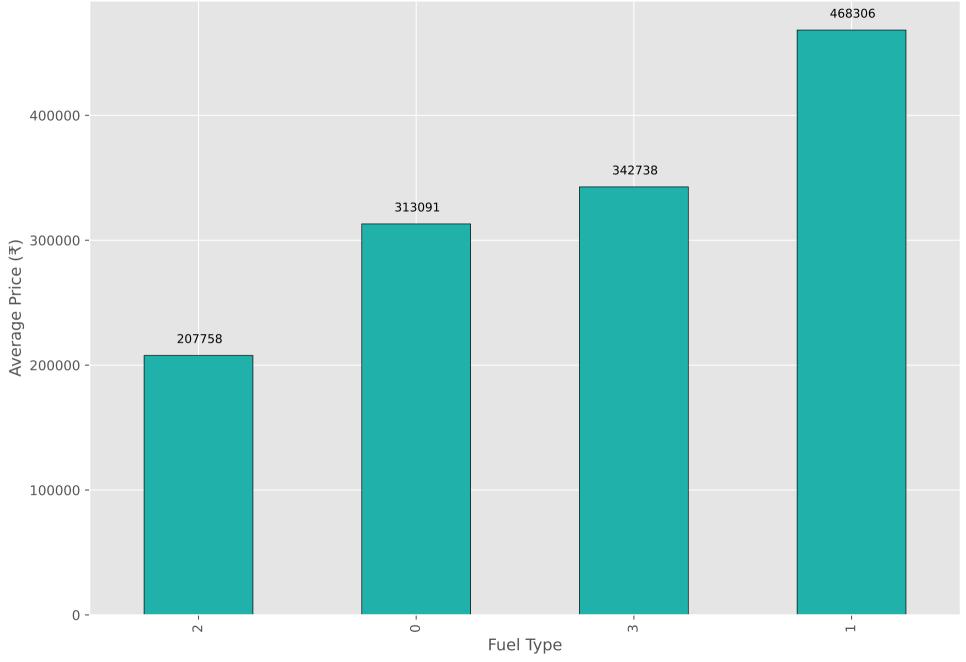
Preprocessing Steps:

- Missing value handling (Mean/Mode)
- Label Encoding for categorical fields
- Outlier removal using IQR
- Feature Scaling using StandardScaler

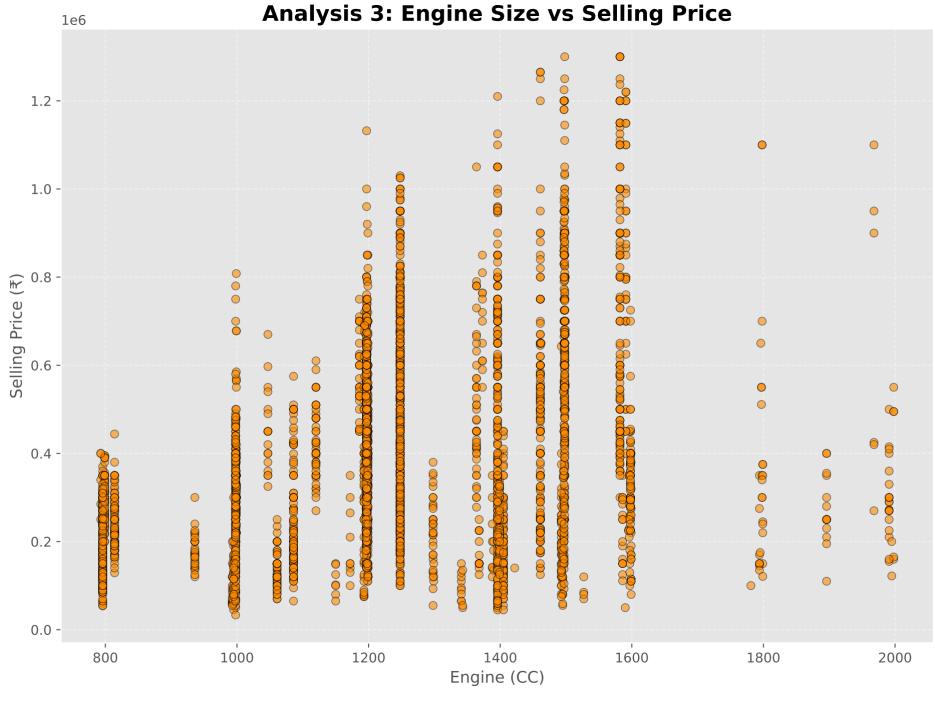
Analysis 1: Number of Cars by Manufacturing Year



Analysis 2: Average Selling Price by Fuel Type

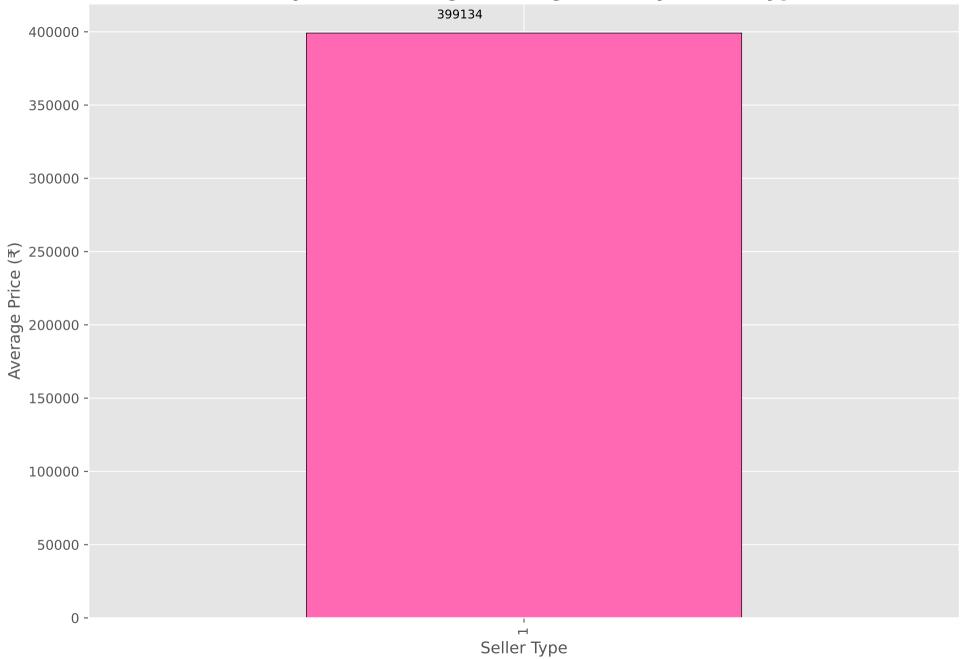


Insight: Diesel cars show higher resale prices compared to Petrol or CNG cars.



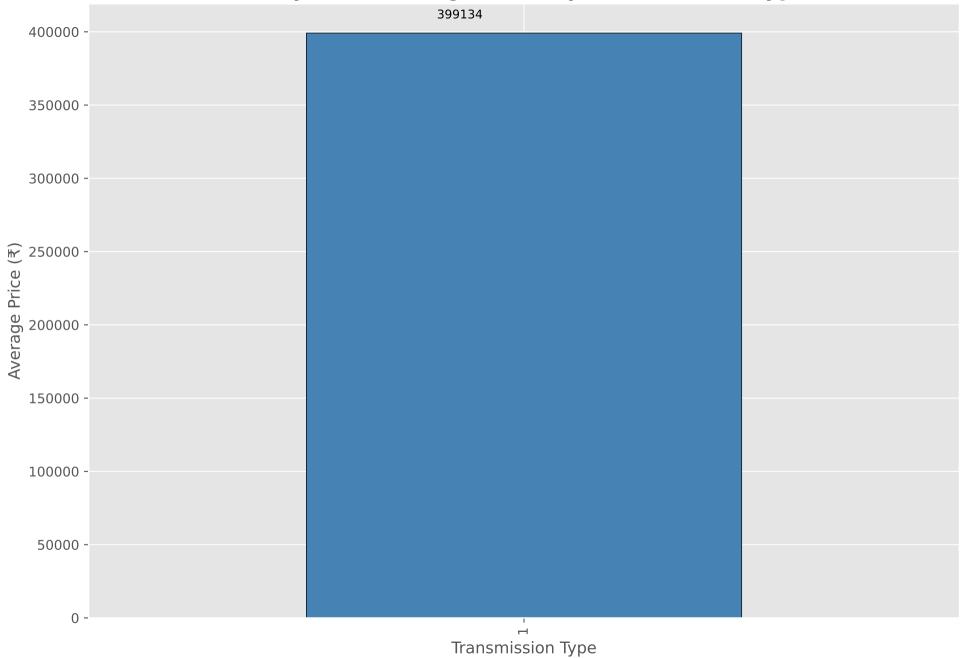
Insight: Correlation = 0.347. Larger engines tend to command higher prices.

Analysis 4: Average Selling Price by Seller Type



Insight: Dealers list higher prices due to added service and warranty value.

Analysis 5: Average Price by Transmission Type



Insight: Automatic cars attract higher resale prices due to comfort demand.

Correlation Heatmap of Numerical Features - 1.0 -0.01 -0.06 0.08 -0.01 -0.12 nan nan 0.10 -0.05 -0.06 nan name --0.01 -0.03 0.49 -0.06 0.16 0.70 nan nan nan - 0.8 selling price --0.06 0.34 0.35 0.02 0.70 -0.26 -0.38 nan nan nan - 0.6 -0.13 km driven -0.35 0.27 -0.08 0.08 nan nan nan fuel - -0.12 -0.03 -0.26 nan nan -0.03 -0.03 nan - 0.4 seller type nan - 0.2 transmission nan -0.01 0.35 -0.03 0.08 -0.12 nan nan -0.24 nan - 0.0 mileage(km/ltr/kg) -0.10 0.34 -0.13 -0.17 0.25 0.49 nan nan -0.24 nan - -0.2 engine --0.05 -0.06 0.35 0.27 0.08 -0.17 -0.18 nan nan nan -0.06 0.02 0.25 max power -0.16 -0.08 -0.03 -0.12 -0.18 nan nan nan - -0.4 nan nan nan nan seats nan nan nan nan nan nan nan nan Insight: Selling price correlates strongly with engine capacity and power features. year