

Used Cars Price Analysis – Project Report

(Cardekho.com)

Name: Ashay Wanjari

Tools Used :

Python | BeautifulSoup | Pandas | Power BI | Jupyter Notebook

Project Duration: Jan 2026 – Feb 2026

Contents

1. Executive Summary

2. Project Objective

3. Data Source and Collection

3.1 Data Source

3.2 Collection Method

3.3 Tool Used

3.4 Process

3.5 Data Extracted

4. Data Overview

5. Data Cleaning and Preprocessing

6. Feature Engineering (Additional Columns)

7. Data Analysis Approach

8. Dashboard Design and Visualization

9. Key Insights and Findings

10. Challenges Faced

11. Conclusion

1. Executive Summary

This project focuses on analysing used car prices collected from CarDekho.com using web scraping techniques. The main aim is to study price variation based on brand, fuel type, city, year of manufacture, and kilometres driven.

The analysis was performed using Python for data processing and Power BI for visualization. Interactive dashboards were created to provide insights into market trends, customer preferences, and pricing behaviour. The project helps buyers and sellers understand the used car market and supports better decision-making.

2. Project Objective

The main objectives of this project are:

- To collect used car data from CarDekho.com
- To clean and prepare raw data for analysis
- To analyse price patterns and trends
- To identify best value-for-money cars
- To develop interactive dashboards using Power BI

3. Data Source and Collection

3.1 Data Source

- Website: CarDekho.com

3.2 Collection Method

Data was collected using web scraping with BeautifulSoup library in Python.

3.3 Tools Used

- Python
- Requests
- BeautifulSoup

3.4 Process

- Sent HTTP requests to web pages
- Parsed HTML using BeautifulSoup
- Extracted required fields
- Stored data in CSV format

3.5 Data Extracted

- Car Name
 - Brand
 - Price
 - Year
 - Fuel Type
 - Transmission
 - KM Driven
 - City
-

4. Data Overview

• Total Records: 24,600+

- Multiple brands and cities
- Mixed data types
- Missing and inconsistent values

5. Data Cleaning and Preprocessing

Data cleaning was performed using Pandas in Python.

Cleaning Steps

- **Removed null values**
 - **Removed duplicate records**
 - **Converted price and KM to numeric format**
 - **Removed special characters**
 - **Standardized categorical values**
 - **Handled incorrect entries**
 - **Removing Outliers**
-

6. Feature Engineering.(Additional Columns)

- **Car Age = Current Year – Manufacturing Year**
- **KM per Year = KM Driven / Car Age**
- **Price Segment (Budget, Mid, Premium, Luxury)**
- **Brand Category**

7. Data Analysis Approach

Exploratory Data Analysis (EDA) was performed using Python and Power BI.

Analysis Methods

- Descriptive statistics
- Group-by operation
- Aggregation functions
- Correlation analysis
- Trend analysis

Focus Areas

- Brand-wise pricing
- Fuel preference
- City-wise demand
- Usage impact on price
- Manufacturing year trends

8. Dashboard Design and Visualization

Includes:

- **Top Selling Brand**
- **Displays the most listed and highest-selling car brand in the dataset, helping identify market leaders.**
- **Average Car Price**
- **Shows the overall average price of used cars to understand the general market value.**
- **Average KM Driven**
- **Represents the average distance driven by cars, indicating overall vehicle usage.**
- **Total Cars Listed**
- **Displays the total number of used cars available in the dataset.**
- **Maximum Car Price**
- **Highlights the highest-priced car, representing the premium segment.**
- **Brand Average Price**
- **Price Segment Distribution**

- Visualizes the average price of cars for each brand, allowing comparison between premium budget brands.
- Categorizes cars into Budget, Mid, Premium, and Luxury segments to understand customer affordability levels.
- Fuel Type Distribution
 - Shows the percentage of Petrol, Diesel, CNG, Electric, and LPG cars, reflecting customer fuel preferences.
- Cars by City
 - Displays the number of cars listed in different cities to analyze regional demand.
- Top Brands by Listings
 - Represents the number of cars listed for each brand, identifying high-demand manufacturers.
- Price vs KM per Year Analysis
 - Uses a scatter plot to study the relationship between car usage and resale price.
- Price Trend by Year of Manufacture
 - Shows how average car prices vary by manufacturing year, indicating depreciation market trends.

- Visualizes the average price of cars for each brand, allowing comparison between premium budget brands.
- Categorizes cars into Budget, Mid, Premium, and Luxury segments to understand customer affordability levels.
- Fuel Type Distribution
 - Shows the percentage of Petrol, Diesel, CNG, Electric, and LPG cars, reflecting customer fuel preferences.
- Cars by City
 - Displays the number of cars listed in different cities to analyze regional demand.
- Top Brands by Listings
 - Represents the number of cars listed for each brand, identifying high-demand manufacturers.
- Price vs KM per Year Analysis
 - Uses a scatter plot to study the relationship between car usage and resale price.
- Price Trend by Year of Manufacture
 - Shows how average car prices vary by manufacturing year, indicating depreciation market trends.

- **Interactive Filters (Slicers)**

Allows users to filter data by:

- **City**
 - **Transmission**
 - **Car Name**
 - **Fuel Type**
-

9. Key Insights and Findings

1. Brand Analysis

- **Maruti is the top-selling brand.**
- **Premium brands have higher average prices.**

2. Price Analysis

- **Newer cars have higher resale value.**
- **Older cars depreciate faster.**

3. Fuel Type

- **Petrol cars dominate the market.**
- **Diesel is second most preferred.**

4. City Analysis

- **Metro cities have more listings.**
- **Higher demand in major cities.**

5. KM Impact

- Higher KM driven reduces resale value.
- Lower usage cars are more expensive.

6. Best Value Segment

- Mid-range cars provide best balance between price and quality.
-

10. Challenges Faced

During the project, the following challenges were faced:

- Website structure changes
 - Missing and inconsistent data
 - Large dataset handling
 - Data formatting issues
 - Performance optimization in Power BI
-

11. Conclusion

This project successfully analysed used car pricing using real-world data from CarDekho.com. By combining web scraping, data cleaning, analysis, and visualization, meaningful insights were generated.

The Power BI dashboard provides an easy-to-use platform for exploring the used car market. The project enhanced practical skills in data analytics and business intelligence.