



Innovation & Entrepreneurship Hub for Educated Rural Youth (SURE Trust – IERY)

Used Cars Sales Analysis

**The domain of the Project:
Data Science and Data Analytics**

**Mr. Purnangshu Nath Roy
AI Consultant @CSR BOX**

**Team Member:
Ms. Tanvi Gunwant Pohankar**

Period of the project

October 2025 to December 2025



Innovation & Entrepreneurship Hub for Educated Rural Youth (SURE Trust – IERY)

Declaration

The project titled **“Used Cars Data Analysis”** was mentored by **Mr. Purnangshu Nath Roy**, organized by **SURE Trust**, for the benefit of educated unemployed rural youth to gain hands-on experience in industry-relevant projects. I hereby declare that the team member mentioned below has successfully worked on this project and enhanced her practical knowledge in the domain of Data Science and Data Analytics.

Team Member

Ms. Tanvi Gunwant Pohankar

Mentor

Mr. Purnangshu Nath Roy
AI Consultant @CSR BOX

Prof. Radhakumari
Executive Director & Founder
SURE Trust



Innovation & Entrepreneurship Hub for Educated Rural Youth (SURE Trust – IERY)

Table of Contents

- 1. Executive Summary**
- 2. Introduction**
- 3. Project Objectives**
- 4. Methodology and Results**
- 5. Social / Industry Relevance**
- 6. Learning and Reflection**
- 7. Conclusion and Future Scope**



Executive Summary

The **Used Car Sales Analytics** project focuses on analyzing used car data to identify pricing trends and determine factors influencing vehicle resale value. The primary objective was to clean raw data, analyze key variables affecting car prices, and build predictive models to support data-driven business decisions.

An end-to-end analytics workflow was implemented using **Excel, SQL, Python, and Power BI**. Excel was used for data cleaning and preliminary analysis, SQL for structured querying and aggregations, Python for exploratory analysis and predictive modeling, and Power BI for creating interactive dashboards.

Key findings indicate that **brand, mileage, vehicle age, and fuel type** significantly impact resale prices. The predictive models demonstrated good accuracy in estimating prices and classifying high-value vehicles. Overall, the project showcases a complete real-world analytics pipeline from raw data to actionable business insights.



Introduction

Background and Context

The used car market is expanding rapidly and has become increasingly competitive. Vehicle prices are influenced by multiple factors such as brand, mileage, age, fuel type, and location. Accurate data analysis is essential for businesses to understand these factors and make informed decisions related to pricing, inventory management, and customer targeting.

Problem Statement and Goals

The key challenge addressed in this project is the absence of structured analysis and predictive insights in used car pricing. Raw datasets are often inconsistent and difficult to interpret. The goal of this project is to clean and analyze used car data, identify major price drivers, and build predictive models to estimate prices and classify high-value vehicles.

Scope:

- Analysis of used car pricing trends
- Brand, fuel type, mileage, and age-based comparisons
- Predictive modeling for price estimation and value classification
- Interactive dashboards for business users

Limitations:

- Analysis is restricted to the available dataset
- Real-time market fluctuations are not included
- External factors such as negotiation, vehicle condition reports, and demand dynamics are not considered

Innovation Component

The innovation of this project lies in its end-to-end integration of multiple tools. Instead of relying on a single platform, the project connects Excel, SQL, Python, and Power BI into a unified analytics pipeline, combining descriptive and predictive analytics for practical decision support.



Project Objectives

Objectives

- To clean, structure, and standardize used car sales data
- To analyze key factors influencing resale price and vehicle value
- To build predictive models for price estimation and high-value classification
- To develop an integrated analytics pipeline using Excel, SQL, Python, and Power BI

Expected Outcomes and Deliverables

- A clean and analysis-ready dataset
- Insights into pricing trends, brand performance, and mileage impact
- Predictive model outputs for price estimation and value classification
- An interactive Power BI dashboard for business decision-making



Methodology and Results

Methods / Technology Used

The project follows a structured, step-by-step data analytics workflow. Raw used car data is initially cleaned and prepared to ensure quality and consistency. The cleaned data is then structured and analyzed using SQL queries to derive meaningful summaries. Further analysis and basic modeling are performed using Python in Jupyter Notebook, and the final results are presented through interactive dashboards in Power BI. Each stage builds upon the previous one to maintain data accuracy and reliability throughout the project.

Tools / Software Used

- **Microsoft Excel**
Used for raw data import, data cleaning, removal of duplicates, creation of derived columns, pivot tables, dashboards, and macro-based automation.
- **SQL Server Management Studio (SSMS)**
Used for creating tables, importing cleaned data, performing joins, aggregations, subqueries, window functions, views, and stored procedures.
- **Jupyter Notebook (Python)**
Used for data loading, exploratory data analysis, visualizations, feature engineering, and building basic regression and classification models.
- **Power BI Desktop**
Used for data modeling, creating DAX measures, KPI cards, filters, slicers, drill-through reports, maps, and interactive dashboards.

Data Collection Approach

The project uses a pre-collected CSV dataset of used car sales. The data was imported into Excel for cleaning and enhancement. After cleaning, it was exported to SQL for structuring and querying, then

Project Architecture

The project follows a linear and integrated architecture:

Excel → SQL → Python → Power BI

- Excel prepares and cleans the raw data
- SQL structures the data and generates analytical outputs
- Python performs analysis and generates additional outputs
- Power BI visualizes results and insight.

This architecture ensures smooth data flow and tool-specific responsibility.



Innovation & Entrepreneurship Hub for Educated Rural Youth (SURE Trust – IERY)

Project GitHub Link

GitHubRepository:

[Final Capstone Repository](#)

The repository contains datasets, SQL scripts, Python notebooks, and supporting files used in the project.



Final Project Implementation – Tool-wise Outputs

This section presents the **final working screenshots** of the project arranged in the correct execution order, along with clear supporting explanations for each tool used.

Step 1: Microsoft Excel – Data Cleaning & Preparation

Purpose: Excel was used as the first layer of data processing to ensure the dataset was clean, consistent, and analysis-ready before loading into databases or analytical tools.

Key Activities Performed:

- Imported raw used car sales CSV data
- Removed duplicates and handled missing values
- Standardized columns such as brand names, fuel type, and city
- Created derived columns (Car Age, Price per KM, Category)
- Built basic pivot summaries for validation

Outcome: A clean and structured dataset suitable for SQL storage and further analytics.



Step 2: SQL Server – Data Storage & Querying

Purpose: SQL Server was used to store the cleaned data and perform structured querying for analytical summaries.

Key Activities Performed:

- Created database and tables for used car data
- Imported cleaned Excel dataset into SQL tables
- Executed SELECT queries for validation
- Performed aggregations such as average price, brand-wise counts, and city-wise analysis
- Used filtering and sorting for exploratory insights

Outcome: A structured relational dataset enabling fast, reliable querying and analytical extraction.



Innovation & Entrepreneurship Hub for Educated Rural Youth (SURE Trust – IERY)

Row	ID	Brand	Price	Model	Variant	FuelType	Color	Kilometer	BodyType	ManufactureDate	ModelYear	Owner	OwnerRole	OwnerName	City	State
1	536022	CHEVROLET	115000.00	SPARK	PS 1.0	PETROL	Blue	34522	HATCHBACK	2009-04-01	2008	1st Owner	West Bengal	Mazda Motors	Rohilkhand	0
2	525975	CHEVROLET	145000.00	SPARK	E 1.0	PETROL	Red	43800	HATCHBACK	2011-05-01	2011	1st Owner	Haryana	Tata Motors	Gujarat	0
3	506028	HARLEY DAVIDSON	120000.00	ALTO	1.0 BS6	PETROL	Red	26541	HATCHBACK	2006-01-01	2006	1st Owner	West Bengal	Suzuki Automobiles (P) Ltd.	Kolkata	0
4	661722	HARLEY DAVIDSON	170000.00	ALTO X10	1.0	PETROL	Black	38884	HATCHBACK	2010-02-01	2010	1st Owner	West Bengal	Suzuki Automobiles (P) Ltd.	Rohilkhand	0
5	540543	HARLEY DAVIDSON	170000.00	JAZZ ESTATE	1.0 BS6	PETROL	Black	38844	HATCHBACK	2010-01-01	2010	1st Owner	West Bengal	Ashok Leyland	Rohilkhand	0
6	544454	HARLEY DAVIDSON	180000.00	WAVE R 1.5	VIE	PETROL	White	22000	HATCHBACK	2010-10-01	2010	1st Owner	Uttar Pradesh	Hero Car Sales	Rohilkhand	0
7	544243	HARLEY DAVIDSON	1701000.00	ENTRA	SMART HATCHBACK-V0	PETROL	White	43733	MPV	2021-01-01	2021	1st Owner	Madhya Pradesh	Daewoo Motors	Indore	1
8	543876	HARLEY DAVIDSON	125000.00	VTITRA BREZZA	Z05	DIESEL	Red	44888	HATCHBACK	2018-06-01	2018	1st Owner	Telangana	BMW Motorcycles	Hyderabad	1
9	506029	HARLEY DAVIDSON	125500.00	VTITRA BREZZA	Z05 PLATINUM	PETROL	Grey	18202	SUV	2020-12-01	2020	1st Owner	Delhi	Car Chassis Exclusive	Delhi	1
10	572871	HARLEY DAVIDSON	185500.00	VTITRA BREZZA	Z05 AMT	DIESEL	White	52170	SUV	2020-01-01	2020	1st Owner	Delhi	Maruti Suzuki	Pune	1
11	585554	HYUNDAI	140000.00	ACCENT	EXECUTIVE	PETROL	White	73860	SEDAN	2010-01-01	2010	1st Owner	Tamil Nadu	MM Motors	Chennai	0
12	608081	HYUNDAI	180000.00	SANTAFE	2.2	PETROL	Black	98140	HATCHBACK	2019-11-01	2009	1st Owner	Karnataka	ACB MOTORS	Bengaluru	0
13	581232	HYUNDAI	160000.00	SANTAFE	2.0	PETROL	White	88200	HATCHBACK	2011-02-01	2011	1st Owner	West Bengal	Reed Motors (Prop. Auto Damage P) Ltd.	Kolkata	0
14	525989	HYUNDAI	170000.00	IX25	1.6	PETROL	White	50300	HATCHBACK	2010-06-01	2010	1st Owner	Haryana	Tata Motors	Gujarat	0
15	551895	HYUNDAI	180000.00	DS	SPORTS 1.6 85kW	PETROL	Red	72495	HATCHBACK	2010-01-01	2010	1st Owner	West Bengal	Suzuki Automobiles (P) Ltd.	Aligarh	0
16	566247	HYUNDAI	180000.00	CHEETA	1.0 CRDI 85kW	DIESEL	White	63000	HATCHBACK	2013-03-01	2013	1st Owner	Delhi	Car Chassis Exclusive	Delhi	1
17	552212	HYUNDAI	180000.00	CHEETA	1.0 CRDI 85kW	DIESEL	White	63004	HATCHBACK	2013-01-01	2013	1st Owner	Delhi	Car Chassis Exclusive	Delhi	1
18	58783	HYUNDAI	185000.00	VERNA	1.6 VTVT EX	PETROL	Metallic	34790	SUV	2020-02-01	2020	1st Owner	Tamil Nadu	MM Motors	Chennai	1
19	542071	FORD	160000.00	EXPLORER	3.0D	DIESEL	Grey	80000	HATCHBACK	2012-10-01	2012	1st Owner	West Bengal	Motor Hull	Rohilkhand	0
20	534236	FORD	180000.00	FESTA	EX DIAMANTO 1.4	DIESEL	White	78155	HATCHBACK	2011-06-01	2011	1st Owner	Tamil Nadu	VVD Motors	Hyderabad	0
21	530984	FORD	185000.00	CITY	1.5 VVT-i	PETROL	White	31311	HATCHBACK	2018-03-01	2018	1st Owner	Delhi	Car Chassis Exclusive	Delhi	1
22	530562	MAHINDRA	120000.00	XUV 300	1.6	DIESEL	White	80237	SUV	2017-12-01	2017	1st Owner	West Bengal	Reed Motors (Prop. Auto Damage P) Ltd.	Kolkata	1
23	562033	MAHINDRA	130000.00	SCORPIO	2.2 D	DIESEL	White	133328	SUV	2019-01-01	2018	1st Owner	Delhi	Car Chassis Exclusive	Delhi	1
24	552711	MAHINDRA	150000.00	E-OLARAV200	E300 CNG AUTOMAT	DIESEL	Orange	123965	HATCHBACK	2013-05-01	2013	1st Owner	Tamil Nadu	MM Motors	Chennai	1
25	520850	MAHINDRA	152000.00	MARIZAB	1.6 PLUB 8 8THR	DIESEL	White	28774	MPV	2020-10-01	2020	1st Owner	West Bengal	Reed Motors (Prop. Auto Damage P) Ltd.	Rohilkhand	1
26	550567	MAHINDRA	185000.00	CHEETA	1.1 PETROL	PETROL	White	28264	SUV	2021-07-01	2021	1st Owner	Haryana	Reed Motors (Prop. Auto Damage P) Ltd.	Gurugram	1
27	520564	TATA	185000.00	HEXA	8T 4x2 9 GTR	DIESEL	Grey	53814	SUV	2017-06-01	2017	1st Owner	Haryana	Daewoo Motors	Gujarat	1
28	550567	HYUNDAI	184500.00	VERNA	1.6 CRDI RX GL	DIESEL	White	64711	HATCHBACK	2016-03-01	2016	1st Owner	Madhya Pradesh	Daewoo Motors	Indore	1
29	570793	HYUNDAI	185000.00	VERNA	1.6 CRDI RX PLUS AT	DIESEL	White	62170	HATCHBACK	2018-02-01	2018	1st Owner	Delhi	Maruti Suzuki	Mumbai	1
30	587830	HYUNDAI	185000.00	VERNA	1.6 CRDI RX PLUS AT	DIESEL	White	63386	SUV	2019-06-01	2019	1st Owner	Uttar Pradesh	Reed Cars Pvt. Ltd.	Noida	1
31	587830	HYUNDAI	185000.00	VERNA	1.6 CRDI RX PLUS AT	DIESEL	White	63386	SUV	2019-06-01	2019	1st Owner	Uttar Pradesh	Reed Cars Pvt. Ltd.	Noida	1

Step 3: Python – Data Analysis & Modeling

Purpose: Python was used for exploratory data analysis (EDA), visualization, feature engineering, and predictive modeling.

Key Activities Performed:

- Loaded cleaned dataset using Pandas
- Conducted EDA to understand distributions and relationships
- Visualized insights such as Top Brands by Median Price
- Engineered features including encoded brand values and car age
- Built a machine learning model to classify high-value cars
- Evaluated model performance using accuracy and classification metrics

Outcome: Analytical insights and predictive outputs that quantify pricing drivers and identify high-value vehicles.





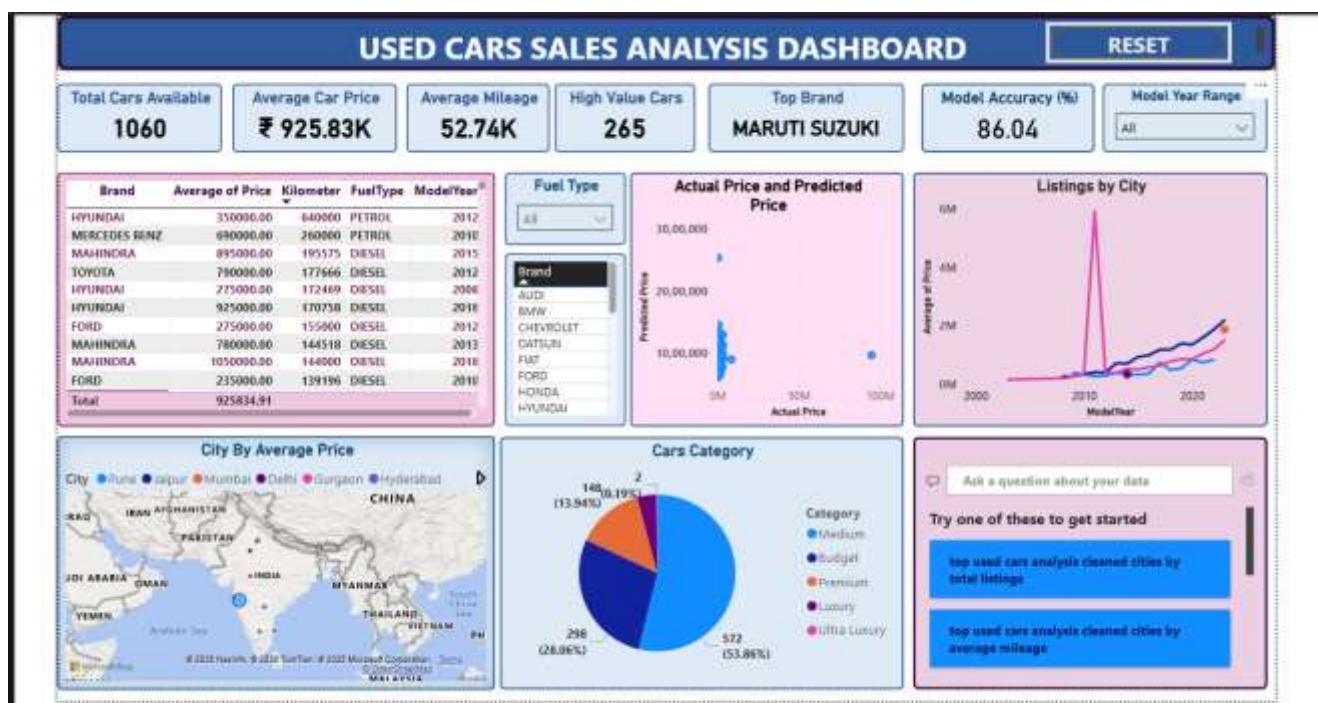
Step 4: Power BI – Interactive Dashboard & Business Insights

Purpose: Power BI was used to convert analytical results into interactive, decision-ready visual dashboards.

Key Activities Performed:

- Imported processed data into Power BI Desktop
- Built data models and relationships
- Created DAX measures for KPIs such as Average Price, High Value Cars, and Model Accuracy
- Designed slicers for brand, fuel type, and model year
- Developed interactive visuals including maps, trend charts, category distributions, and prediction comparison charts

Outcome: A professional dashboard enabling stakeholders to explore pricing trends, brand performance, and predictive insights interactively.





Innovation & Entrepreneurship Hub for Educated Rural Youth (SURE Trust – IERY)

Learning and Reflection

Key Learnings

- Applied Excel for structured data cleaning, feature creation, and dashboard preparation.
- Used SQL Server Management Studio to create tables, write joins, aggregations, views, and stored procedures.
- Performed data analysis and basic modeling using Python in Jupyter Notebook.
- Built interactive reports and KPIs using Power BI Desktop.
- Learned to maintain data consistency across tools in an end-to-end analytics workflow.

Overall Experience

This project provided practical exposure to working with data across multiple tools. It strengthened understanding of data preparation, querying, analysis, and visualization in a structured workflow. The experience improved technical execution, accuracy in analysis, and clarity in presenting results.



Conclusion and Future Scope

Conclusion

This project successfully achieved its objectives of cleaning and structuring used car sales data, analyzing key factors influencing car prices, and presenting insights through interactive dashboards. By integrating Excel, SQL Server Management Studio, Jupyter Notebook, and Power BI Desktop, the project demonstrates a complete and well-organized data analytics workflow. The final outputs provide clear insights into pricing trends, brand performance, and mileage impact, supporting informed business decision-making.

Future Scope

- Inclusion of real-time or regularly updated data sources
- Expansion of analysis with additional vehicle attributes
- Improvement of modeling techniques for more accurate price estimation
- Deployment of dashboards with automated refresh and wider user access



Innovation & Entrepreneurship Hub for Educated Rural Youth (SURE Trust – IERY)