Cars24 Used Car Price Analysis

From web scraping to actionable market insights

Meet the Team

Our dedicated team of experts collaborated to deliver this comprehensive analysis.

Web Scraping

Anand Raj (Team Lead)

Atharv Kakade

Data Analysing & Cleaning

Siddhesh Mali (Co Lead)

Abhishek Mishra

Data Visualisation

Linda Shalini T (Co Lead 2)

Project Presentation

Siddhesh Mali (Co Lead)

Himanshu Shakya

Project Report

Anand Raj (Team Lead)

Project Aim and Objectives

Project Aim

To extract, clean, and analyze used car data from Cars24.com, focusing on Mumbai-based Mahindra listings. This will develop practical skills in web scraping, data processing, and Python analysis.

Key Objectives

- **Scrape key car details (Year, Kilometers, Fuel, Transmission, Price).**
- **Organize and clean data into a structured CSV format.**
- **Analyze and visualize data to derive market insights.**
- **Present findings via Python Notebook, Report, and Presentation.**

Scope & Coverage

Target Website: Cars24.com. Brand: Hyundai. Primary Location: Mumbai.

Technology Stack

Web Automation

Selenium WebDriver with ChromeDriver for dynamic page scraping

Data Processing

Pandas for manipulation, Regex for text parsing, CSV for storage

Visualisation

Matplotlib and Seaborn for statistical graphics

Environment

Google Colab and Jupyter Notebooks for development

Web Scraping Methodology

01

Automated Navigation

Selenium controlled browser instances to traverse Cars24 listing pages systematically

02

Data Extraction

Captured key attributes: Car Name, Year, Kilometres, Fuel Type, Transmission, Price,

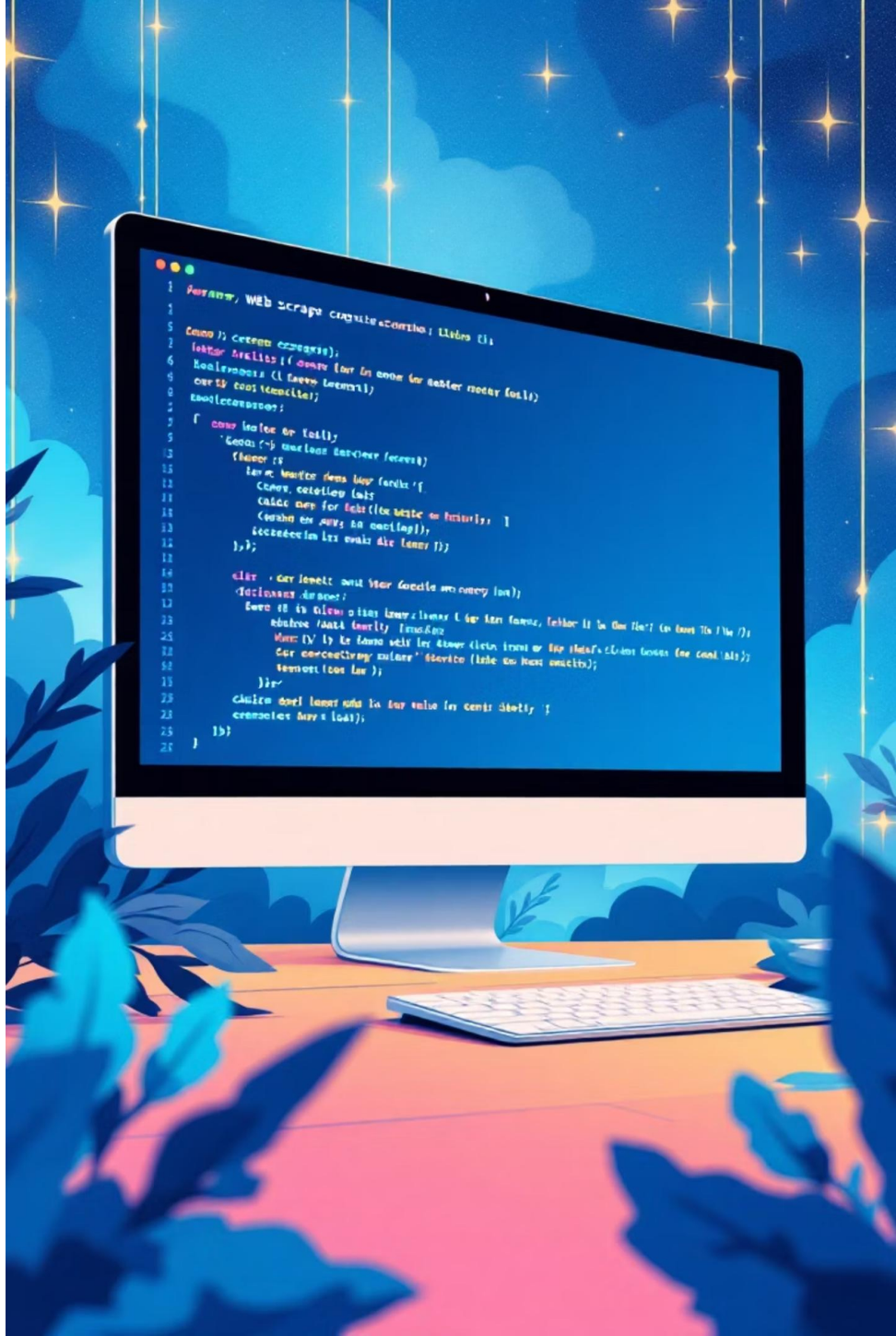
03

Pagination Handling

Implemented robust logic to navigate through multiple pages whilst managing dynamic content loading

Technical Challenges

Overcame issues with dynamically loaded content, fragile CSS selectors, and inconsistent page structures across city-specific listings.



Data Cleaning & Preparation

Cleaning Operations

- ### Duplicate Removal

  Identified and eliminated redundant entries

- ### Type Conversion

  Standardised numeric and categorical fields

- ### Missing Values

  Applied appropriate imputation strategies

- ### Categorical Consistency

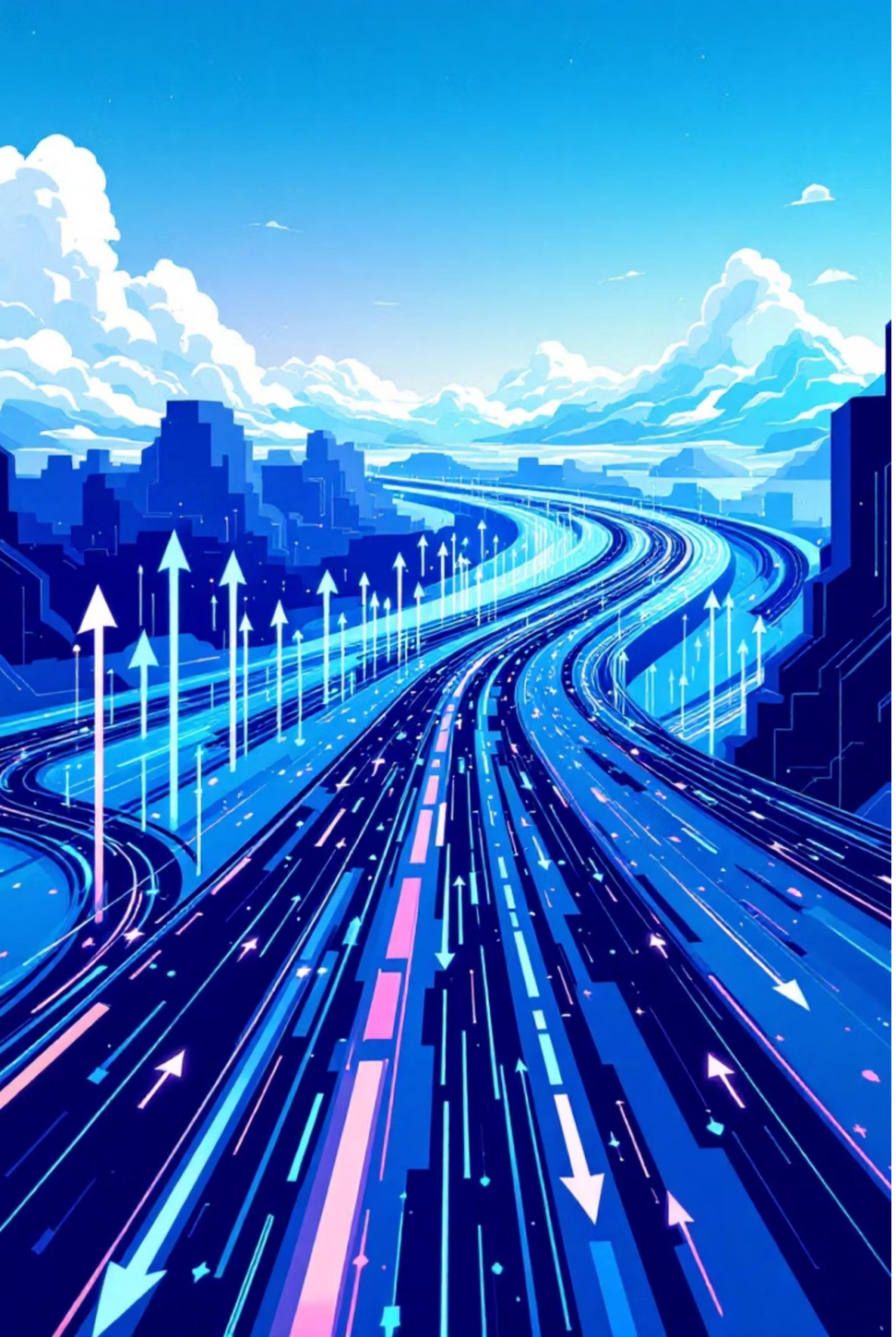
  Normalised fuel types and transmission values

Derived Features

Car_Age = 2025 − Year

Price_per_Km = Price ÷ Kilometres

These engineered features enabled more sophisticated analysis of depreciation patterns and value metrics across the dataset.



Feature Transformation

Transformed raw scraped data into analysis-ready features for deeper insight into pricing and usage patterns.

1

Cleaned Car Name

Removed year prefixes (e.g., 2016 Honda City → Honda City) for consistent brand and model grouping.

2

Standardised Kilometers Driven

Converted all formats (e.g., “25k km”, “2.5L km”) into integer kilometers for accurate numeric comparison.

3

Normalized Price (INR)

Removed currency symbols and standardised ₹, lakh, crore values into a uniform INR scale.

4

Derived Car Age

Calculated as `2025 - Year` to capture depreciation effects and support price–age relationship analysis.

5

Created Price per Kilometer

Derived from `Price(INR) / Kilometers_Driven` to identify value efficiency and pricing anomalies.

Outcome

- **Ensured uniform numeric types for robust modeling and visualization.**
- **Enabled comparative insights between car brands, fuel types, and usage patterns.**
- **Prepared data effectively for advanced statistical analysis and machine learning models.**

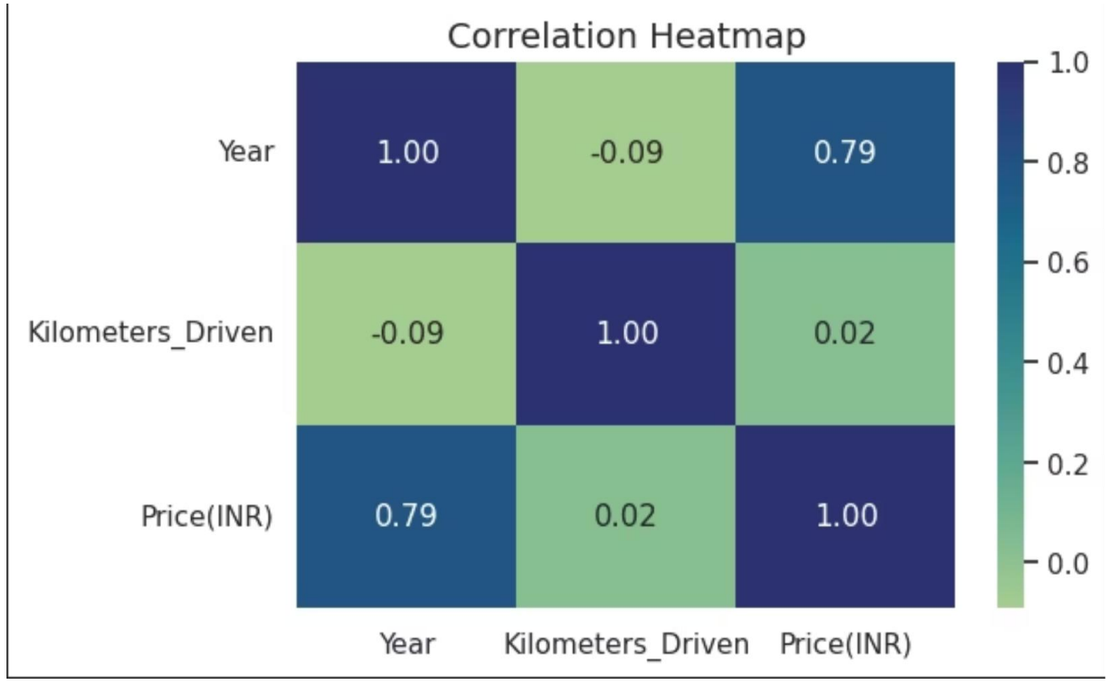
Exploratory Data Analysis

Univariate Analysis

- **Price distribution showed right-skewed pattern**
- **Car age concentrated in 3–7 year range**
- **Kilometre readings varied widely across listings**

Bivariate Relationships

- **Strong negative correlation between price and vehicle age**
- **Inverse relationship between kilometres driven and valuation**
- **Transmission type significantly influenced pricing**

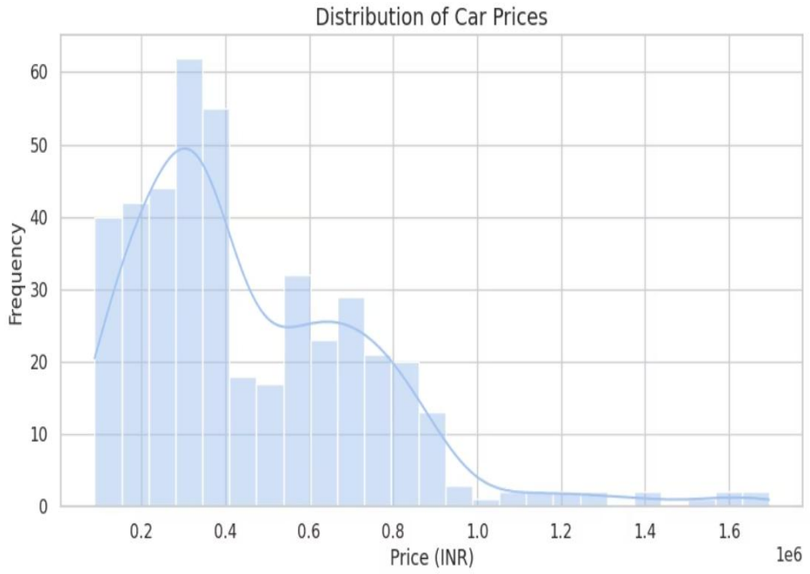


Correlation matrix revealed strongest predictive relationships, with age and mileage showing the highest negative correlation with price.

Visualisation Highlights

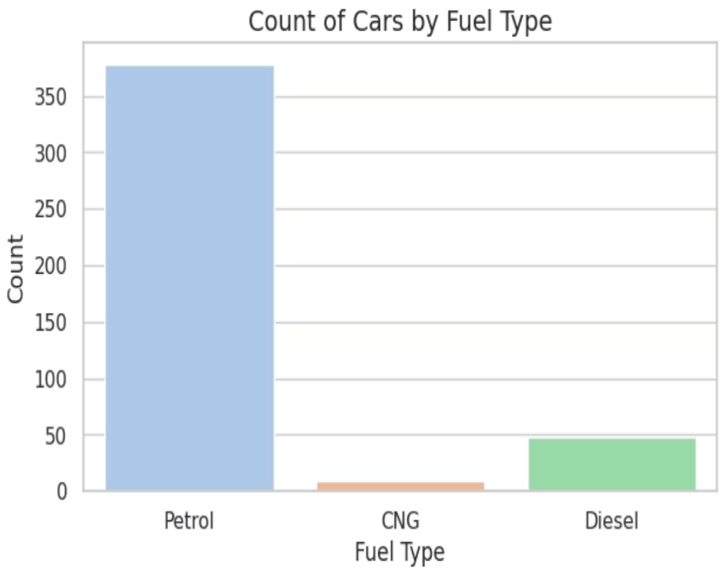
Price Distribution

Right-skewed histogram indicating concentration in mid-range segment



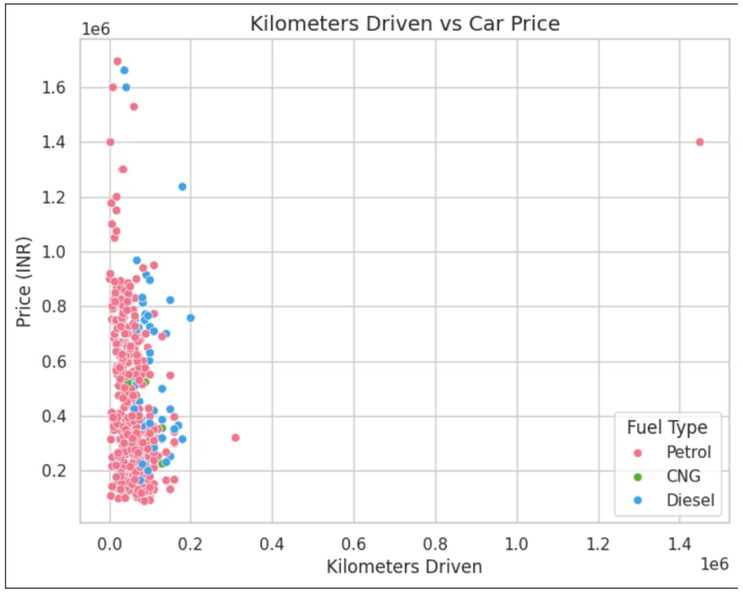
Fuel Type Analysis

Countplots revealed petrol dominance in the used car market



Price vs Kilometres

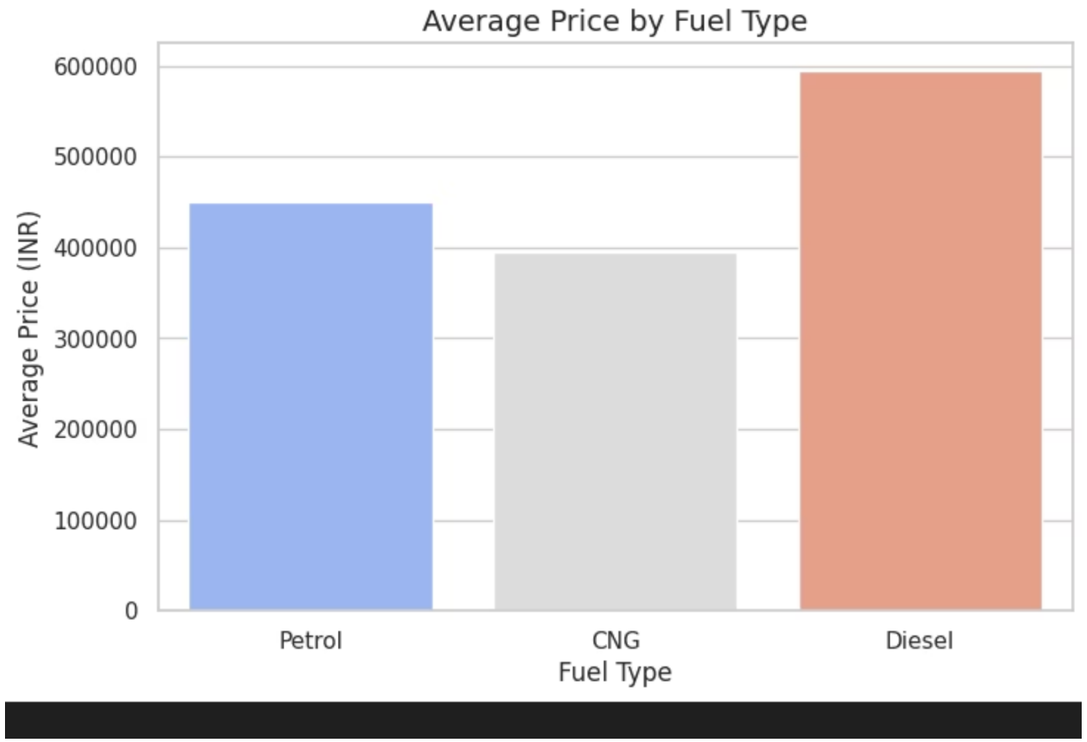
Scatterplot demonstrated depreciation with increased mileage by fuel type



Deeper Dive into Market Trends

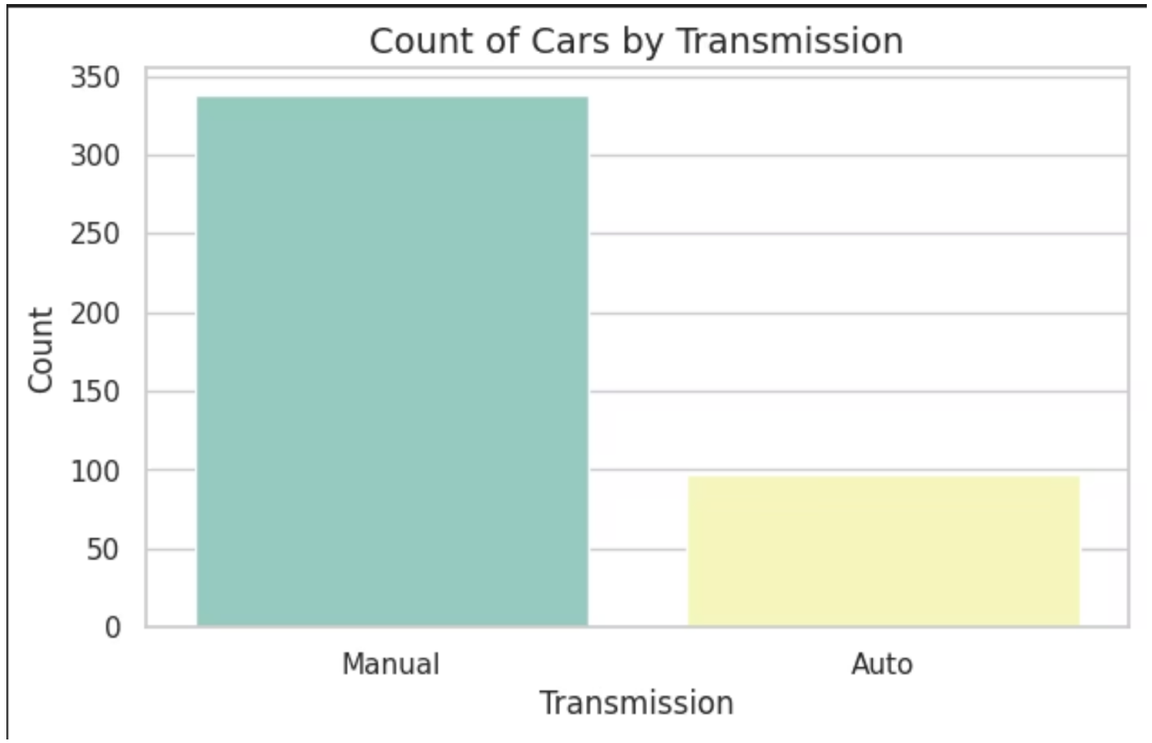
Average Price by Fuel Type

Diesel cars command the highest average prices (~₹6 lakh), reflecting their long-distance efficiency and perceived market value. Petrol models follow at ₹4–4.5 lakh, while CNG cars are the most affordable, catering to budget-conscious and eco-friendly buyers. This indicates fuel type is a strong predictor in pricing models.

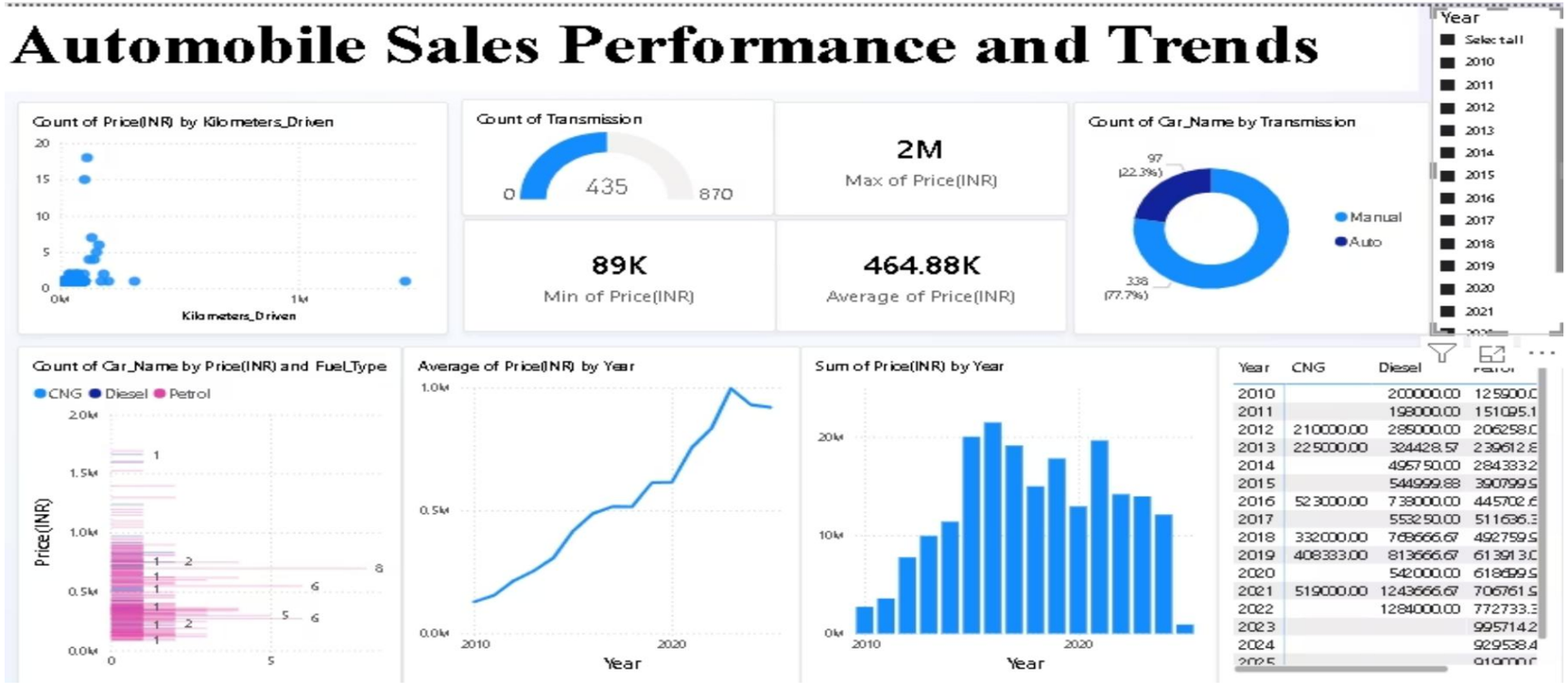


Count of Cars by Transmission

Manual transmission vehicles dominate the market with listings three times higher than automatics, suggesting a strong preference among budget-oriented consumers. Automatic cars form a smaller, premium segment, appealing to urban users prioritising comfort. Transmission type significantly influences both market availability and demand.



Automobile Sales Trends & Insights (Power BI Dashboard)



Key Observations from Dashboard

- **Price vs Kilometers Driven: As kilometers increase, car prices decrease, confirming mileage as a strong negative predictor of resale value.**
- **Transmission Trends: Automatic cars dominate (870 units) compared to Manual (435), indicating a shift toward convenience-focused buyers. This suggests a changing market preference.**
- **Fuel Type & Price Patterns: Diesel cars command the highest prices, followed by Petrol, while CNG remains the most budget-friendly. This confirms fuel type as a key factor in price modeling.**
- **Yearly Price Trends: Average and total prices show an upward trend from 2010–2020, reflecting market growth and rising consumer spending power.**

.

Strategic Insights

Factor	Observation	Impact
Mileage	Higher km → Lower price	Use in depreciation modeling
Transmission	Automatics rising	Market demand shift
Fuel Type	Diesel premium	Pricing & eco-strategy
Year Trend	Rising prices	Time-based forecasting

Conclusion & Future Directions

Project Accomplishments

1. Successfully scraped comprehensive vehicle data
2. Cleaned and engineered analytical features
3. Generated actionable market insights

Next Steps

Predictive Modelling

Develop ML algorithms for price estimation

Interactive Dashboard

Create Power BI visualisation platform

API Integration

Implement automated, scalable data collection

This project demonstrates the complete data science pipeline from raw data acquisition to meaningful business intelligence.

Thank You!

We appreciate your time and interest in our Cars24 Used Car Price Analysis project. We hope our insights provide valuable understanding of the market dynamics.