USED CARS PRICE ANALYSIS:

A Comprehensive Data Analytics Study of Cars24 Platform

PRESENTED BY:

GOVINDULA ACHYUTH

ACKNOWLEDGMENT

CHAPTER 1

INTRODUCTION

India's booming used car market presents a range of challenges and opportunities. Buyers often face difficulties in accessing structured and reliable information regarding price, condition, and vehicle popularity. This project aims to solve that issue by leveraging data analytics to provide clarity on car pricing trends, brand preferences, and vehicle conditions. The study involves web scraping, cleaning, analyzing, and visualizing data from Cars24 to offer structured insights that assist consumers, sellers, and platforms alike.

1.1 Overview of Project

This project analyzes trends in the Indian used car market using data from Cars24. It addresses the challenge of unstructured listings by extracting, cleaning, and visualizing data on brands, prices, usage, and features.

Key insights are presented through Power BI dashboards, supporting informed decisions for buyers, sellers, and platforms.

The project consists of several phases:

Web Scraping: Using Selenium and BeautifulSoup to collect listing data from Cars24 for cities like Pune, Mumbai, Delhi, etc.

Data Cleaning: Utilizing Python libraries like Pandas and NumPy to handle missing values, duplicates, and inconsistent formats.

Visualization: Creating interactive Power BI dashboards to illustrate trends and comparisons based on price, brand, city, fuel type, etc.

1.2 Problem Statement

- Unstructured listings make it difficult to compare used cars across different platforms.
- Multiple decision criteria such as price, brand, fuel type, city, and vehicle condition can overwhelm users.

- Lack of standardized data representation across listings.
- No unified platform to aggregate and filter used car listings by user preferences.
- Users often rely on manual comparisons which are time-consuming and error-prone.
- Absence of intelligent tools or dashboards that provide comparative visual insights.
- Limited access to location-based or feature-specific price analysis.
- Transparency and decision confidence are significantly reduced for buyers and sellers.

1.3 Project Objectives

This project aims to provide a comprehensive analysis of the Indian used car market using data from the Cars24 platform. The key objectives include:

- Collecting detailed car listings from major Indian cities.

- Automating data scraping using Python tools.

- Cleaning and structuring data for accurate analysis.

- Uncovering trends in pricing, brand popularity, and vehicle features.

- Creating interactive dashboards in Power BI for visual insights.

- Supporting buyers, sellers, and analysts with actionable data.

- Preparing the dataset for potential future predictive modeling.

1.4 Scope

This project analyzes used car listings from the Cars24 platform, focusing on eight major Indian cities—Delhi, Mumbai, Pune, Ahmedabad, Bengaluru, Hyderabad, Chennai, and Kolkata. It examines key vehicle attributes such as brand, model, year, kilometers driven, fuel type, transmission, price, and location.

The scope includes:

- **Market trends analysis** across cities.

- **Consumer preferences** by fuel, brand, and transmission.

- **Price variability** based on age and usage.

Data is visualized using interactive Power BI dashboards to support buyers, sellers, and analysts. While limited to Cars24 listings and static data snapshots, the project is scalable for future inclusion of additional platforms, attributes, and predictive modeling.

CHAPTER 2
DATA COLLECTION AND SOURCES

2.1 Overview

The foundation of this project lies in the robust and accurate collection of used car listings from the Cars24 online marketplace. Given the dynamic nature of such websites and the variability in vehicle information, a comprehensive web scraping strategy was developed to ensure consistency, quality, and breadth of data across multiple urban regions in India.

2.2 Data Source

The primary dataset for this project was collected from **Cars24** (https://www.cars24.com), one of India's most prominent online platforms for buying and selling used vehicles. Cars24 offers a wide inventory of second-hand vehicles, spanning various brands, fuel types, models, and price ranges. As a marketplace that operates across India, it provides an ideal foundation for analyzing regional trends and buyer preferences in the used car ecosystem.

Why Cars24?

- **Market Leadership:** Cars24 has a significant footprint in the Indian second-hand car industry, catering to millions of users annually.

- **Comprehensive Listings:** Each vehicle entry typically includes a range of attributes including model year, kilometers driven, fuel type, location, and pricing.

- **City-Specific Portals:** URLs are structured by city (e.g., /buy-used-cars-mumbai/, /buy-used-cars-bangalore/), allowing for easy segmentation of regional data.

- **Diverse Inventory:** Listings range from budget hatchbacks to luxury sedans and SUVs, providing insight into consumer preferences across various price segments.

Data Source Highlights

- **Platform:** Cars24 (https://www.cars24.com)

- **Scope of Listings:** Used cars listed for resale by individual owners and dealerships.

- **City Segmentation:** Each city has a dedicated landing page, which simplifies filtering of geographic data.

- **Listing Format:** Listings include rich metadata (e.g., brand, model, year, location, price), making them ideal for structured extraction and analysis.

Targeted Cities for Data Collection

The following cities were selected based on their population size, economic activity, and volume of used car transactions on the platform:

City	Region	Reason for Inclusion
Delhi	North	High listing volume, broad brand variety
Mumbai	West	Diverse car segments, high demand zone
Bengaluru	South	Tech hub, growing interest in newer models
Hyderabad	South	Mix of mid-range and premium vehicles
Chennai	South	Strong secondary car market
Kolkata	East	Price-sensitive buyers, older car listings
Pune	West	High student and professional population
Ahmedabad	West	Emerging market with affordable segment focus

Types of Vehicles Included

The dataset encompasses a broad spectrum of cars:

- Entry-Level Hatchbacks (e.g., Maruti Alto, Hyundai Eon)

- Mid-Range Sedans (e.g., Honda City, Hyundai Verna)

- Premium SUVs (e.g., Toyota Fortuner, Mahindra XUV500)

- Luxury Cars (e.g., BMW, Audi, Mercedes-Benz)

Limitations of the Data Source

While Cars24 is a valuable source, some limitations were noted:

- **No historical pricing data:** Listings represent only the current asking price.

- **Missing condition metrics:** No detailed vehicle condition reports (e.g., dents, service history).

- **Dynamic content:** Car listings change frequently, so data is only a snapshot in time.

Despite these limitations, the platform offers a rich and credible dataset for understanding consumer behavior and market dynamics in India's used car segment.

2.3 Scraping Methodology

To effectively capture real-time vehicle listings, the scraping process included the following components:

1. Browser Automation with Selenium

- Used for loading dynamically generated JavaScript content.

- Automated browser scroll-down actions were implemented to load hidden listings on long pages.

2. HTML Parsing with BeautifulSoup

- Parsed individual car listing cards for specific data fields.

- Ensured tags and class attributes were accurately located despite nested or dynamic layouts.

3. Data Fields Extracted

Each listing was scraped for the following attributes:

- Brand and Model Name

- Year of Manufacture

- Kilometers Driven

- Fuel Type (Petrol, Diesel, CNG, Electric)

- Transmission Type (Manual, Automatic)

- Price (INR)

- City (based on URL path structure)

Each record was stored in a structured Python dictionary, later converted into a tabular format using pandas.

2.4 Challenges

The data collection process faced several technical and structural challenges:

- **Dynamic Web Content:** Cars24 loads listings via JavaScript, requiring the use of Selenium for full-page rendering.

- **Rate Limiting:** Frequent requests triggered temporary blocks, mitigated using request delays and user-agent rotation.

- **Inconsistent Data Entries:** Some listings lacked complete information (e.g., missing year or price), requiring filtering and validation.

- **HTML Structure Changes:** Variations in listing card layouts necessitated flexible parsing logic.

These challenges were addressed through robust scraping strategies, conditional parsing, and thorough data validation to ensure a clean, usable dataset.

CHAPTER 3

DATA CLEANING AND PREPROCESSING

The raw datasets obtained from both web scraping and Cars24's internal exports required significant cleaning and preprocessing to ensure reliability and consistency for analysis. This stage is crucial as it transforms unstructured, messy data into a refined format suitable for visualization and decision-making.

3.1 Data Sources

Three primary data files were used:

- cars24_multicity_data_final_1.csv: Scraped data from eight major Indian cities.

- Cars24_Train_Data.csv: Historical training data with pricing trends and standard formats.

These sources varied in structure and completeness, requiring consolidation and normalization before integration.

3.2 Preprocessing Techniques

Removing Duplicates

- Duplicate listings were common due to repeated car posts across city pages.

- Duplicates were identified using combinations of brand, model, year, and kilometers driven.

- A total of **278 duplicate rows** were removed to preserve data integrity.

Handling Missing Values

- Missing values in Kilometers Driven, Fuel Type, and Transmission were handled using:

  o **Median Imputation** for numerical fields (e.g., Kilometers Driven).

- o **Mode Imputation** or "Unknown" tag for categorical variables.

- Rows with missing Price or Model Year were dropped if no accurate imputation was feasible.

3.3 Standardization of Columns

- **Price Column**:

  - o Raw data had inconsistent formats like "₹5.2L", "Rs. 3.5 lakhs", etc.

  - o Regular expressions were used to remove currency symbols and convert all values to integers (INR).

  - o Converted prices to uniform numeric scale in rupees (e.g., ₹5.2L → 520000).

- **Kilometers Driven**:

  - o Cleaned entries like "56,000 km", "85k", and "30K KM" were standardized using regex.

  - o Values were converted to integers, stripping commas and non-numeric characters.

- **Brand and Model Normalization**:

  - o Brands like "Maruti", "Maruti Suzuki", and "MS" were unified under a single label: **"Maruti Suzuki"**.

  - o Model names were also cleaned for consistency, ensuring uniform representation.

3.4 Encoding & Transformation

- **Transmission & Fuel Type**:

  - o Encoded into categorical values (e.g., "Manual", "Automatic", "Petrol", "Diesel", "CNG").

  - o Typos and alternate terms like "Manual Trans.", "Auto", "Petro" were mapped appropriately.

- **Year of Manufacture**:

  - Converted to datetime-compatible integers for temporal analysis.

  - Removed entries with invalid or future years (e.g., 2027, 2030).

- **Location Mapping**:

  - Standardized city names across all datasets.

  - Ensured consistency in records across Mumbai, Delhi, Chennai, etc.

3.5 Merging and Validation

- Merged the three data sources using a **common schema** of brand, model, year, and city.

- Validated merged records through:

  - Null value checks post-merge.

  - Manual sampling to ensure no mismatch in city-car allocations.

- Exported the final cleaned dataset to cleaned_cars24_data.csv for use in Power BI.

CHAPTER 4

DATA ANALYSIS AND VISUALIZATION

4.1 Overview of Analysis

Data analysis serves as the core of this project, translating raw values into actionable insights. Using the cleaned dataset, multiple dimensions of the used car market were explored. Power BI dashboards enabled dynamic interaction with the data, allowing for intuitive comparisons and pattern recognition.

This chapter highlights the analytical methods used and the insights derived from key variables such as city, brand, model year, transmission, fuel type, and price.

4.1 Objectives of the Analysis

The analysis was designed to answer the following key questions:

- What are the most popular car brands and models?

- How do car prices vary across different cities and years?

- What is the impact of fuel type and transmission on pricing?

- Which cities have the widest variety of used cars?

- How does mileage (kilometers driven) affect resale value?

4.2 Dimensions of Analysis

Brand and Model Trends

- Maruti Suzuki, Hyundai, and Honda are the most listed brands.

- Popular models include Swift, i20, and Honda City, especially in metro cities.

- Bar charts and pie charts were used to visualize brand distribution.

Price Patterns

- Price ranges from ₹50,000 to ₹25 lakhs, with metro cities like Mumbai and Delhi showing higher medians.

- Entry-level cars are most affordable; luxury cars are outliers.

- Visualized using box plots, scatter plots, and histograms.

Fuel and Transmission Insights

- Petrol and manual transmission dominate the listings.

- Automatic cars are fewer but priced higher.

- Stacked bar charts and heatmaps revealed city-wise variations.

Vehicle Age and Mileage

- Most cars listed are 3–10 years old.

- Cars with under 50,000 km driven fetch significantly higher resale value.

- Depreciation trends and mileage impact were analyzed using line and bubble charts.

4.3 Dashboard Highlights

To provide users with a rich and interactive analytical experience, a comprehensive **Power BI dashboard** was developed. The key features include:

Dashboard Component	Description
Brand Explorer	Dynamic slicers to compare average prices, mileage, and availability by brand
Price Visualizer	Box and violin plots for price comparison across years and cities
Fuel & Transmission Matrix	Cross-tab view of price and availability based on fuel and gearbox type

Depreciation Trendline	Line graphs to analyze price drop across years and mileage
City-wise Insights Panel	A heatmap and ranking chart to compare cities in terms of variety and value

4.4 Key Findings

- **Maruti Suzuki** is the most prevalent brand, followed by Hyundai and Honda.

- **Petrol + Manual** is the most common car configuration.

- **Delhi, Mumbai, and Bangalore** lead in listings and price variation.

- Cars aged **3–5 years** offer the best value in terms of price-to-feature ratio.

- Vehicles driven **under 50,000 km** tend to command premium resale prices.

CHAPTER 5

TOOLS AND TECHNOLOGIES

This project leverages a modern stack of tools and technologies to efficiently manage every stage of the analytics pipeline—from data collection and transformation to exploration and visualization. The selection of tools was based on their reliability, community support, and capability to handle large-scale data processes and rich visual outputs.

5.1 Software and Platforms Used

Python (Jupyter Notebooks)

- Used extensively for web scraping, data cleaning, transformation, and analysis.

- Provided an interactive and modular environment for iterative code development and testing.

- Allowed integration of visual inspection using libraries like matplotlib and seaborn during exploration.

Power BI

- Deployed for building dynamic dashboards and producing impactful visualizations.

- Enabled stakeholders to slice and filter data interactively using dashboards.

- Ideal for non-technical users to explore data intuitively via maps, drill-through reports, and dynamic visuals.

Microsoft Excel

- Used for quick data inspection, validation, and cross-verification of data structure and integrity.

- Assisted in initial format checks before importing data into Power BI.

5.2 Python Libraries

The project utilized several key Python libraries for automation, data handling, and cleaning tasks:

Library	Purpose
Pandas	Data manipulation, cleaning, joining datasets, handling missing values
NumPy	Numerical computations, data transformation, and working with arrays
Selenium	Browser automation for dynamic page scraping from Cars24
BeautifulSoup	Parsing and extracting content from the HTML structure
Regex (re)	Cleaning and extracting text-based fields (e.g., prices, mileage)
Matplotlib/Seaborn	Exploratory data visualization (used during notebook-level analysis)

5.3 Web Scraping Technologies

- **Selenium WebDriver**:
    - Employed to interact with dynamically loaded content on Cars24's web pages.
    - Enabled scrolling simulation and waiting for content loads through WebDriverWait.

- **BeautifulSoup**:
    - Parsed HTML content retrieved by Selenium.
    - Extracted specific tags containing car listings, brand names, prices, and specifications.

- **Chrome Headless Mode**:

  - Configured using ChromeOptions for lightweight, GUI-less scraping.

  - Reduced processing time and memory usage during multi-city scraping loops.

- **Error Handling Mechanisms**:

  - Try-except blocks and custom logging were implemented to ensure robustness and retry failed scrapes.

5.4 Power BI Features Utilized

Power BI served as the core visualization platform for this project, offering real-time interactive insights. Key features used include:

| Feature	Description
Data Modeling	Created relationships between cleaned tables for multidimensional analysis
DAX (Data Analysis Expressions)	Used to compute custom metrics such as average price by brand or price-to-year ratios
Slicers and Filters	Enabled user-driven exploration across cities, brands, years, and fuel types
Drill-through Reports	Allowed navigation from high-level metrics to detailed vehicle data
Custom Visuals	Integrated charts: bar, box, scatter, heatmaps, and stacked visualizations for storytelling

CHAPTER 6

DASH BOARDS

The Power BI dashboard developed for this project provides a comprehensive, interactive overview of India's used-car market on Cars24. It integrates multiple visual components to deliver actionable insights across key vehicle parameters such as price, location, brand, fuel type, transmission, and time.

6.1 Overview

The dashboard is titled **"CARS24 USED-CAR PRICE ANALYSIS"** and is designed for both granular and comparative analysis. It enables users to filter by city, brand, transmission type, and fuel category while observing metrics such as average price, count of listings, and market share distribution.
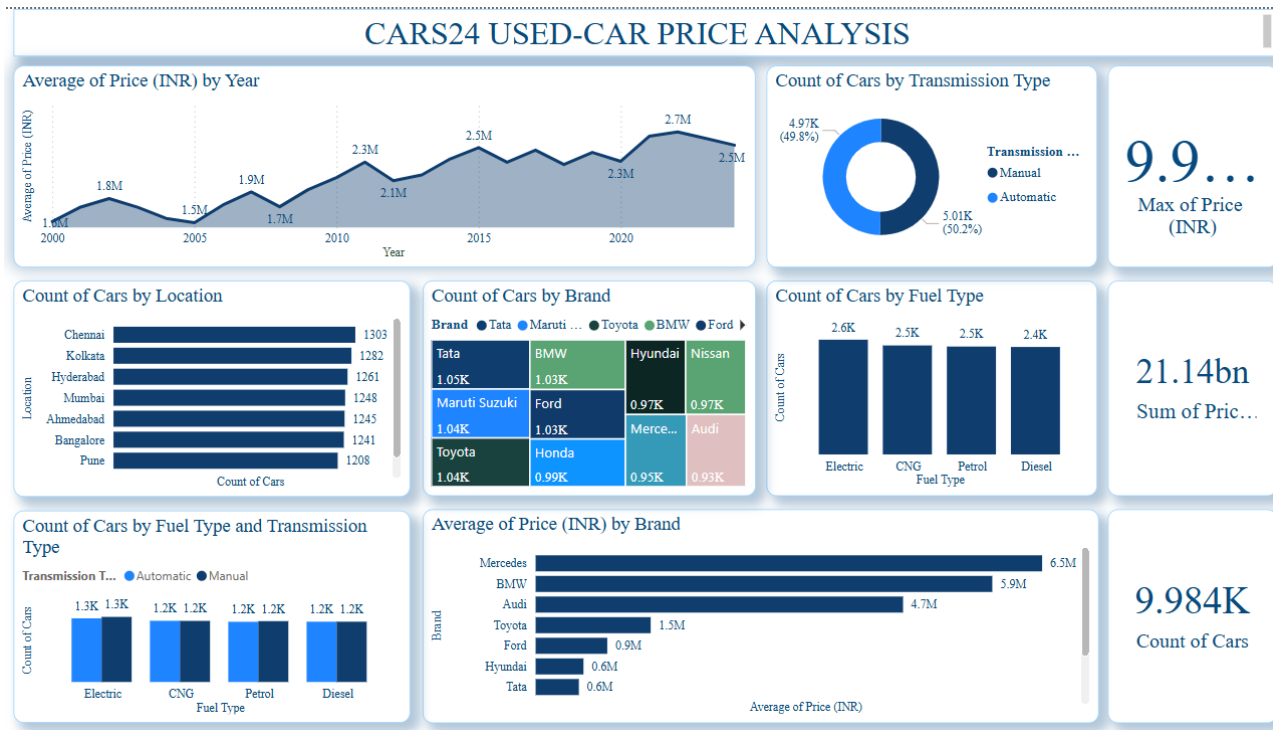
6.2 Key Dashboard Components

- **Price Trend by Year:** Shows average car prices from 2000 to 2023 with peaks in 2011, 2016, and 2021.

- **Transmission Type Split:** Manual and automatic cars are almost evenly split (~50% each).
- **Highlights (Cards):**
  - Max Price: ₹9.9M
  - Total Listings: ~9.98K
  - Total Price Sum: ₹21.14B
- **Top Locations:** Chennai, Kolkata, Hyderabad have the most listings.
- **Popular Brands:** Tata, Maruti, Toyota, and BMW lead in volume.
- **Fuel Type Count:** Electric, CNG, Petrol, and Diesel are nearly equal in listings.
- **Fuel vs. Transmission:** Balanced distribution across fuel types.
- **Brand-wise Pricing:** Mercedes, BMW, and Audi are highest in average price.

6.3 Design and Usability Features

- **Interactive Visuals**: All charts are interconnected—clicking one element filters all others.
- **Color Palette**: Blue gradient tones ensure visual consistency and clarity.
- **Responsive Cards**: Real-time updating of total cars, sum of prices, and max price values.
- **Hierarchical Filters**: Enables drill-down by city, fuel type, or brand instantly.

6.4 Insights Enabled by the Dashboard

- Compare used-car supply and pricing across India's major cities.
- Identify pricing patterns by year to understand depreciation.
- Understand the rising popularity of electric vehicles.
- Spot brand-wise pricing disparities and market dominance.
- Analyze the shift in consumer preference toward automatic transmission.

CHAPTER 7

CONCLUSION

This project aimed to explore and analyze the Indian used-car market using a rich dataset sourced from Cars24. Through a comprehensive process involving data cleaning, preprocessing, visualization, and interactive dashboard development, we gained critical insights into consumer behavior, vehicle trends, and pricing dynamics.

The analysis revealed valuable patterns—such as the steady increase in average car prices over the years, the near-equal popularity of manual and automatic transmissions, and the growing share of electric vehicles in the resale market. City-level comparisons highlighted regional variations in inventory and pricing, while brand-level analysis identified both premium and budget segments of the market.

The interactive Power BI dashboards played a pivotal role in making the data accessible and interpretable. These tools allowed users to explore the dataset across multiple dimensions, empowering decision-makers to compare cities, brands, fuel types, and more, with ease and clarity.

Key Takeaways:

- **Data-driven storytelling** enhances understanding of large-scale vehicle marketplaces.

- **Visualization tools** like Power BI are essential for translating raw data into insights.

- The **resale value of cars** varies significantly based on brand, location, and condition indicators such as mileage and age.

- **Electric vehicles** are emerging strongly, reflecting changing buyer preferences and market adaptation.

Executive Summary

This project presents an analytical overview of the Indian used-car market using real-world data from Cars24. The primary objective was to uncover insights from the dataset through data cleaning, preprocessing, and interactive visualizations using Power BI.

The dataset included thousands of vehicle listings with attributes like brand, location, fuel type, transmission, manufacturing year, and price. After a thorough data cleaning and transformation process, key trends and relationships were explored.

Key Insights:

- **Average car prices** have shown an overall increasing trend from 2000 to 2023.

- **Manual and automatic transmissions** are almost equally popular.

- **Electric vehicles** are gaining traction, competing closely with petrol and diesel cars.

- **Top contributing cities** include Chennai, Kolkata, and Hyderabad in terms of listing volume.

- **Luxury brands** like Mercedes and BMW command the highest average resale prices.

A dynamic **Power BI dashboard** was developed to allow stakeholders to explore these insights interactively. The dashboard enables filtering and comparison by year, location, brand, fuel type, and transmission, making it a valuable tool for both consumers and businesses.

This project highlights the effectiveness of data-driven analysis and visualization in understanding complex market trends and supports better decision-making in the automotive resale industry.