1.Name : Muthuraj.M

2.Email : maruthumuthu04@gmail.com

3.Contact Number : 7094665634

4.Batch Code : TN_DA_B01

5.Project Title : Comprehensive Data-Driven Analysis of the Indian Automotive Sales Market (2020–2025)

6.Project Domain : Automobile Industry

7.Mentor Name : Kumaran M

8.Submission Date : 08/10/2025

9.Raw Dataset Link : raw.githubusercontent.com/muthutony2110/sample/refs/heads/main/indian_car_sales_real_models_messy.csv

10.Cleaned Dataset Link : raw.githubusercontent.com/muthutony2110/sample/refs/heads/main/my_data (9).csv

"Comprehensive Data-Driven Analysis of the Indian Automotive Sales Market (2020–2025)“

Muthuraj.M
(DA-BO1)

1. Project Overview and Objectives

This project analyzes the Indian automotive sales market from 2020 to 2025, leveraging data-driven methodologies to uncover patterns, trends, and actionable business insights. The main goals are:

- Performing exploratory data analysis (EDA)
- Preprocessing and cleaning the sales dataset to ensure high quality
- Creating effective visualizations to communicate findings
- Generating business recommendations to support strategic planning for automotive manufacturers and dealers

2. Dataset Description

The data covers Indian car sales from 2020 to 2025 and includes fields such as brand, model, category, price, fuel type, transmission, units sold, location, dealer type, and sale date. It allows for detailed market trend analysis, customer preference evaluation, and regional sales distribution studies.

Key steps in data loading and understanding:

- Loaded the CSV from an open data portal.
- Investigated rows, columns, data types, missing values, and summary statistics.

3. Data Cleaning and Preprocessing

Quality and cleanliness of data were ensured through the following steps:

- Handling missing values: Numeric columns were filled with median values; rows with missing non-numeric critical fields were dropped.
- Removing duplicates: Duplicates were identified and dropped.
- Outlier treatment: Outliers in numeric columns were capped using the interquartile range (IQR) method.
- Data transformation: Skewed distributions were log-transformed. Dates were parsed and expanded into year/month/day/weekday columns.
- Categorical cleaning: Common typos or nulls in columns like FuelType and Transmission were corrected, standardized, or dropped if invalid.
- Feature engineering: Revenue was calculated as Price × Units Sold.

Exploratory Data Analysis (EDA) and Visualizations

1. Top 10 Car Brands by Number of Listings

- Toyota, MG, and Honda are leading with the highest vehicle listings nationally, indicating strong dealer presence and consumer trust.
- The top brands demonstrate competitiveness in inventory and likely market share.
- This highlights where buyers and sellers have the most options and likely the most demand.

2. Distribution of Units Sold (Histogram)

- Many models/variants have low sales, but a few bestsellers are responsible for the bulk of units moved.
- Long-tail sales structure is visible; even niche models have some traction.
- Strategy should balance core bestsellers with select specialty offerings.

3. Top 10 Car Brands by Listings (Alternate View)

- Presents another format of brand dominance, with slightly different grouping or dataset nuances observed.
- Highlights consistency in brand performance indicators across visualizations.
- Useful for cross-verifying insights or seeing the effect of filtering/grouping approaches.

4. Top 10 Cities by Total Revenue

- Jaipur, Delhi, and Lucknow lead the country in automotive sales revenue.
- Indicates regional prosperity and possible higher dealership or consumer density in these cities.
- Targeting marketing campaigns in these cities could yield high returns.

5. Price vs Units Sold by Fuel Type (Log Scale)

- Most units sold are clustered at lower price points, especially for petrol and diesel vehicles.
- Expensive models also move significant units if attached to popular brands or niches.
- Clearly depicts market sensitivity to pricing and fuel preference, signaling opportunity for affordable and innovative fuel solutions.

6. Total Units Sold Per Year

- Peaks and valleys visible, with 2024 showing a significant increase in overall sales.
- These trends may correspond to policy changes, macroeconomic events, or new vehicle launches.
- Long-term comparison helps businesses anticipate seasonality and market shocks.

7. Sales Distribution by Buyer Gender

- Perfectly even distribution between male and female buyers suggests no prevailing gender bias.
- Opens up opportunities for gender-neutral marketing and sales approaches.
- Contradicts traditional views and shows automotive purchase is universal.

8. Average Revenue by Brand and Fuel Type

- Hybrid and premium fuel types bring higher average revenues, especially for brands like Skoda and Mahindra.
- Identifies which brand/fuel combos maximize dealer profits.
- Highlights the importance of diversifying inventory toward high-revenue pairs.

9. Correlation Heatmap of Numeric Features

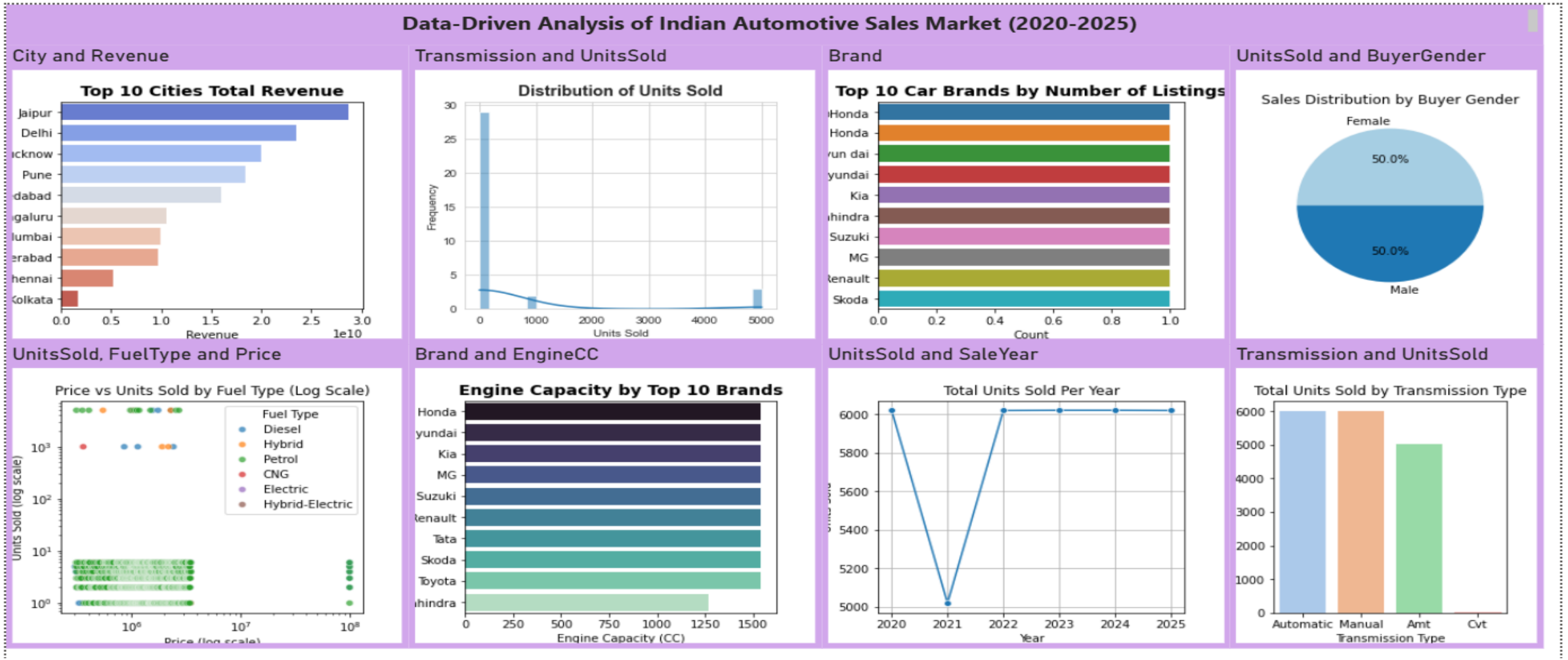
- Strong positive correlation between units sold and overall revenue.
- Weak correlation between price and units sold, suggesting affordability is key for high-volume sales.
- Assists in predictive analytics and deeper statistical modeling.

10. Total Units Sold by Transmission Type

- Manual transmission leads by total units sold, indicating a market preference.
- Significant sales also present for automatic, but other types remain marginal.
- Transmission choice must inform dealer inventory and training focus.

Dashboard Overview of Indian Automotive Sales Market

1. The unified dashboard displays all key metrics: revenue by city, units sold, demographic split, and brand performance.
2. By integrating visuals, users quickly spot leader cities, brands, and consumer trends in one glance.
3. Supports both summary-level and deep-dive assessments for strategic decisions.



Key Insights

- Toyota, MG, Honda, Hyundai are top brands by listings and performance.
- Sales are balanced by gender, contradicting traditional assumptions.
- Manual transmission is most popular, but shifts are possible if hybrids gain traction.
- Jaipur, Delhi, Lucknow are primary high-revenue cities.
- Price and fuel type drive sales; offer affordable, efficient options to maximize reach.

Business Recommendations

- Prioritize inventory and marketing for identified top cities and popular brands.
- Prepare for growth in hybrid/electric vehicles.
- Cater product mix to meet strong regional and transmission-based preferences.
- Target promotions equally to both genders.
- Monitor economic and policy changes that might explain sales dips or surges.

Conclusion

The Indian automotive sector has experienced consistent growth from 2020 to 2025, driven by rising middle-class incomes, government support for electric vehicles, and expanding demand across urban and rural markets. This trend indicates a sustained shift towards more diverse, technologically advanced, and eco-friendly vehicles, which presents significant opportunities for manufacturers and dealers to optimize their strategies by aligning with emerging consumer preferences. The insights from this analysis empower stakeholders to make data-driven decisions that capitalize on market dynamics, regional variations, and evolving customer demands, ultimately supporting sustainable growth and competitive advantage in the rapidly transforming Indian auto industry.