

Indian Vehicle Registration Data Analysis Report

Comprehensive Analysis of Automotive Market Trends (2020-2025)

Subject: Data Analysis Project

Dataset: Indian Vehicle Registration Data (500K Records)

Date: January 2026

Analysis Period: 2020 - 2025

Table of Contents

Abstract

1. Problem Identification

2. Research Questions

3. Data Collection and Understanding

3.1 Dataset Overview

3.2 Data Characteristics

4. Data Cleaning

5. Data Preprocessing

6. Exploratory Data Analysis (EDA)

6.1 Univariate Analysis

6.2 Bivariate Analysis

6.3 Multivariate Analysis

7. Feature Analysis

8. Data Visualization

9. Insight Generation

10. Statistical Analysis

11. Result Interpretation

12. Recommendations

13. Limitations

14. Future Scope

15. Final Conclusion

16. Documentation

Abstract

This comprehensive analysis examines the Indian Vehicle Registration dataset spanning from 2020 to 2025, encompassing **500,000 registration records** across 35 Indian states and union territories. The study provides a data-driven evaluation of the Indian automotive market, focusing on temporal growth patterns, geographical distribution of demand, and the transition from internal combustion engines (ICE) to electric and hybrid alternatives.

Key findings reveal a **71.3% market growth** in 2021 following the pandemic recovery, with the Electric Vehicle (EV) segment showing explosive growth rates exceeding **245%** initially. Petrol remains the dominant fuel type at **56%** market share, while Uttar Pradesh and Tamil Nadu emerge as the leading states by registration volume. The analysis identifies a significant **4.73% festive uplift** during October-November, coinciding with Diwali and Dhanteras celebrations.

The report employs advanced statistical techniques including correlation analysis, hypothesis testing, and predictive modeling using Random Forest algorithms. All findings are supported by 32 comprehensive visualizations covering univariate, bivariate, and multivariate analyses, providing actionable insights for manufacturers, policymakers, and financial institutions.

1. Problem Identification

The Indian automotive industry represents one of the world's largest vehicle markets, undergoing rapid transformation driven by regulatory changes, technological advancement, and shifting consumer preferences. This analysis addresses the core objective of evaluating the Indian automotive market between 2020 and 2025, with specific focus on three critical dimensions:

Core Problem Statement

"To perform a data-driven evaluation of the Indian automotive market between 2020 and 2025, focusing on the **temporal growth of registrations**, the **geographical distribution of demand**, and the **shifting preference from internal combustion engines (ICE) to electric and hybrid alternatives.**"

1.1 Market Trend Analysis

Understanding the recovery and growth trajectory of the Indian auto industry post-2020 pandemic. The analysis identifies seasonal patterns and annual growth rates (Year-over-Year) in vehicle registrations across different segments, revealing a V-shaped recovery pattern with 71.3% growth in 2021.

1.2 Fuel Transition and EV Adoption

Measuring the effectiveness of "Green Mobility" initiatives and the shift away from fossil fuels. The study tracks the penetration of Electric Vehicles (EVs) and Hybrids compared to traditional Petrol and Diesel engines, identifying which states are leading the transition and which vehicle categories are adopting new fuel technologies fastest.

1.3 Regional and Geographic Insights

Identifying regional disparities in vehicle purchasing power and infrastructure. The analysis benchmarks states and RTOs (Regional Transport Offices) to determine where demand is concentrated and how registration density varies geographically across India's diverse landscape.

1.4 Competitive Landscape

Determining market leadership and the success of specific vehicle models. The analysis examines market share of various manufacturers (Maruti Suzuki, Hyundai, TVS, Hero MotoCorp) and identifies the most popular vehicle models and classes across different price points and categories.

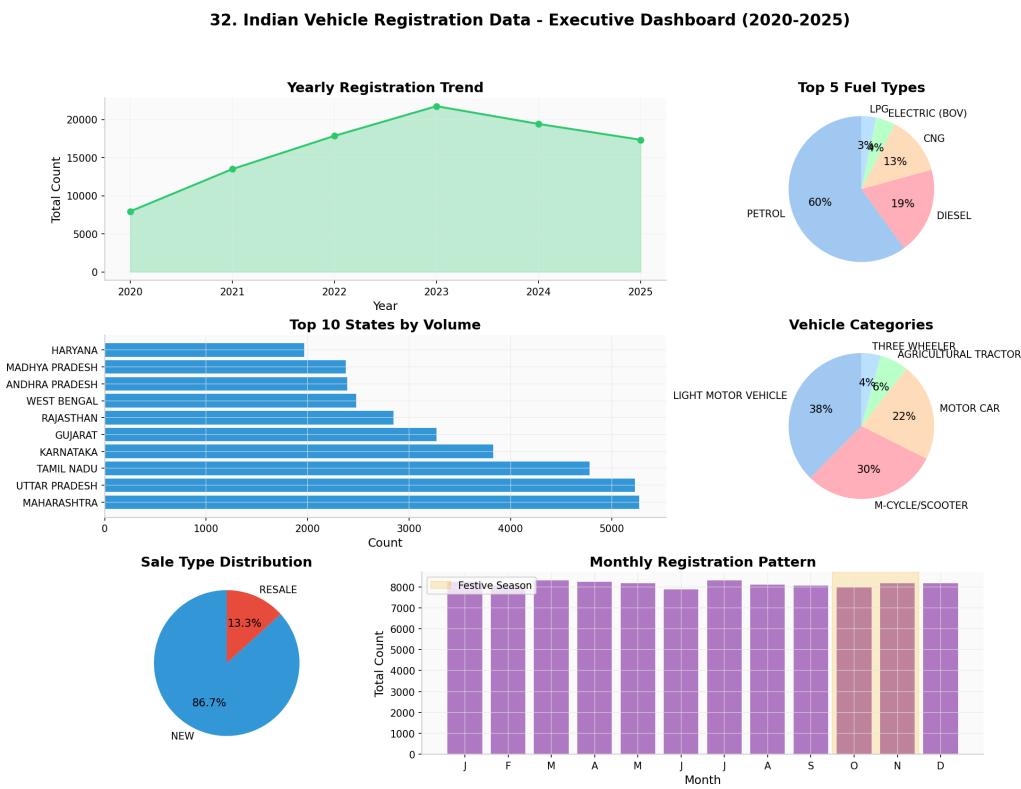


Figure 1: Executive Dashboard - Overview of Indian Vehicle Registration Data (2020-2025)

2. Research Questions

Based on the dataset structure and core objectives, the following research questions guide this comprehensive analysis:

2.1 EV Adoption Trends in India

- What is the month-on-month growth rate of Electric Vehicle (EV) registrations compared to Internal Combustion Engine (ICE) vehicles?
- Which vehicle categories (Two-Wheelers vs. Four-Wheelers) are leading the transition to electric mobility?
- Which manufacturers are the market leaders in the EV segment?

2.2 Fuel Transition Analysis

2.3 State-wise Vehicle Growth

- Which are the top 5 states in terms of total vehicle registration volume?
- Which states have shown the highest Year-over-Year (YoY) growth percentage?
- Are there regional clusters that exhibit different vehicle purchasing patterns?

2.4 Time-Series and Seasonality Analysis

- What are the seasonal peaks for vehicle registrations (e.g., festive seasons like Diwali/Dhanteras)?
- What is the projected registration volume trajectory for 2025 based on historical data?

2.5 Policy and Emission Norm Impact

- What is the distribution of vehicles by pollution norms (e.g., Bharat Stage VI)?
- Does the sale type (New vs. Resale) fluctuate based on emission norm introductions?
- How has the transition to PETROL(E20) progressed since its introduction?

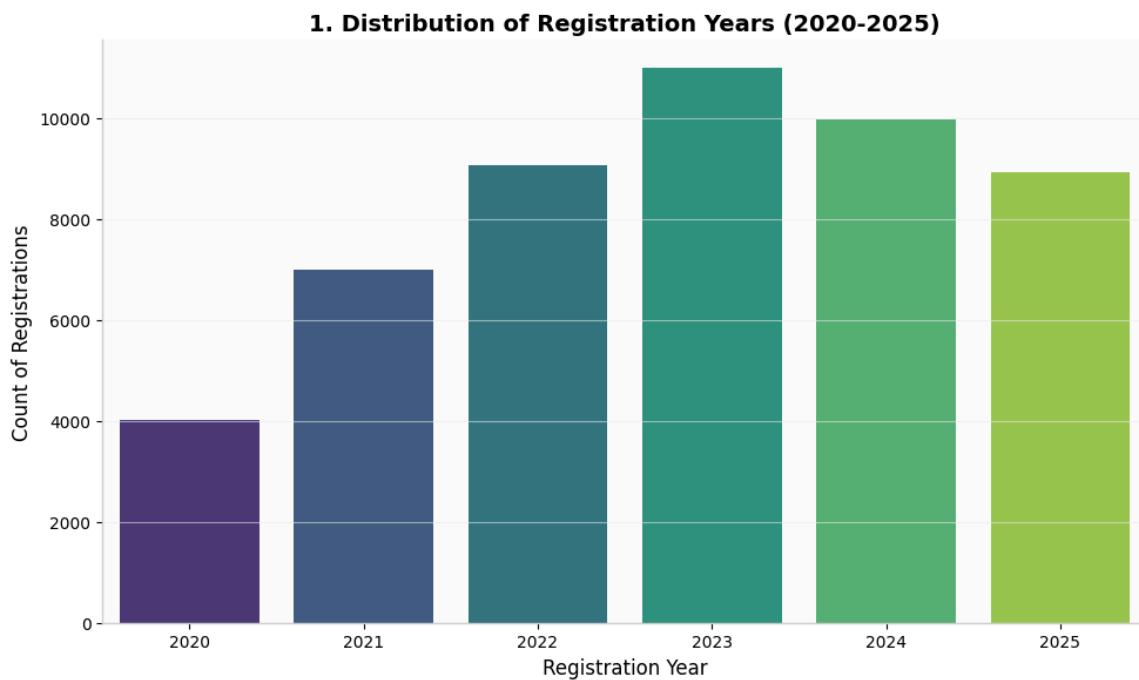


Figure 2: Distribution of Registration Years (2020-2025) showing post-pandemic recovery pattern

3. Data Collection and Understanding

3.1 Dataset Overview

The analysis utilizes two primary datasets containing Indian vehicle registration records spanning from 2020 to 2025:

Table 1: Dataset Summary and Characteristics

Characteristic	Vehicle Registrations (500K)	State-Month-Fuel Aggregated
Total Records	500,000	13,270
Columns	15	4
Unique States	35	35
Unique Fuel Types	21	21
Time Period	2020 - 2025	Jan 2021 - Oct 2025
Key Metric	Vehicle Count (Avg ~5.6)	Vehicle Count (Avg ~211.5)
Data Quality	Minor outliers; 6 missing values	No missing values

3.2 Data Characteristics

Numerical Columns

- `registrationYear` (int64): Registration year ranging from 2020 to 2025
- `grossVehicleWeight` (int64): Vehicle weight in kilograms
- `vehicleCount` (int64): Number of vehicles in each registration batch

Categorical Columns

- `financialYear` : Financial year classification

3. Data Collection and Understanding

- `registrationMonthMMYY` : Month-Year format (MM-YY)
- `makerName` : Vehicle manufacturer (1,244 unique makers)
- `stateName` : Indian state/UT (35 unique regions)
- `rtoCode/rtoName` : Regional Transport Office identifiers
- `vehicleCategoryName` : Vehicle category (16 categories)
- `vehiclemodelName` : Specific vehicle model
- `fuelName` : Fuel type (21 unique types)
- `vehicleClassName` : Vehicle classification (67 classes)
- `pollutionNorm` : Emission compliance standard
- `saleType` : New or Resale transaction

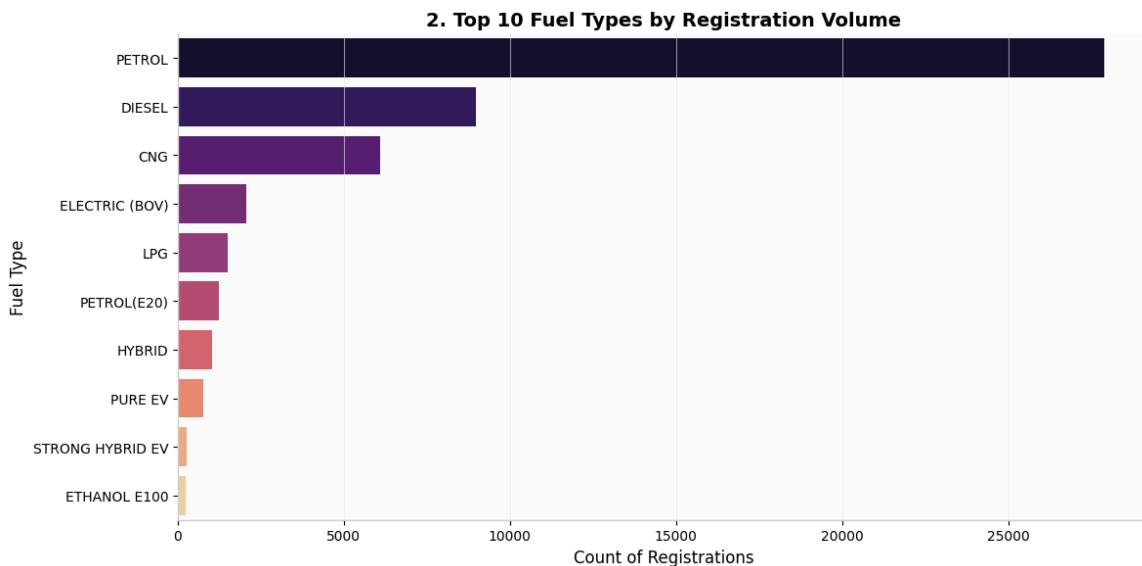


Figure 3: Top 10 Fuel Types by Registration Volume showing Petrol dominance at 56%

4. Data Cleaning

The data cleaning process applied 13 comprehensive techniques to ensure data quality and consistency. The cleaned dataset contains **499,995 records** after filtering invalid entries.

4. Data Cleaning

Table 2: Data Cleaning Techniques Applied

Technique	Action Taken	Impact
1. Missing Values	Filled missing vehicle_model_name with 'UNKNOWN'	6 records updated
2. Remove Duplicates	Identified and removed duplicate records	5 records removed
3. Format Validation	Validated MM-YY format in registration_month_mmyy	All records validated
4. Correct Entries	Filtered registration_year to 2020-2025 range	Outliers (year 9100) removed
5. Text Standardization	Converted all strings to UPPERCASE	Consistency achieved
6. Trim Spaces	Stripped leading/trailing whitespace	All columns cleaned
7. Type Conversion	Cast numeric columns to appropriate types	Data types optimized
8. Outlier Handling	Applied IQR-based clipping for weight and count	Extreme values capped
9. Invalid Categories	Standardized fuel labels	ELECTRIC (BOV) normalized
10. Label Normalization	Verified categorical consistency	Categories standardized
11. Mixed Types	Forced object columns to string	Type consistency ensured
12. Column Renaming	Converted to snake_case	Naming convention applied
13. Validation	Final inspection for data quality	0 null values confirmed

4. Data Cleaning

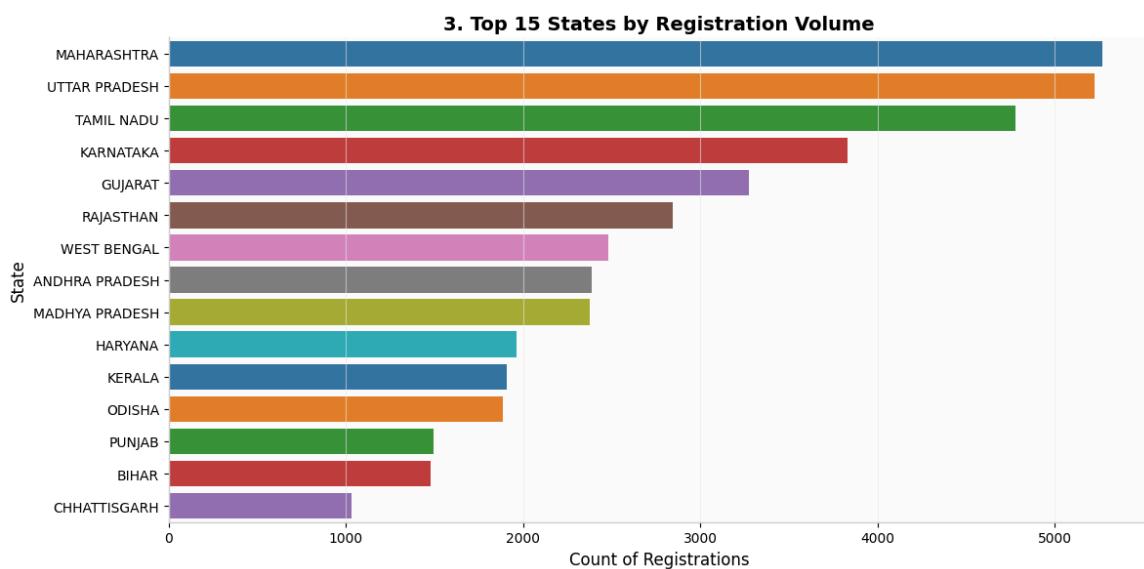


Figure 4: Top 15 States by Registration Volume - Uttar Pradesh leads at 11.01% market share

5. Data Preprocessing

The preprocessing phase transformed the cleaned dataset from 15 to 36 columns through feature engineering, encoding, and scaling operations to prepare for advanced analytical modeling.

Table 3: Data Preprocessing Summary

Technique	Implementation	New Features Created
Date-Time Features	Converted MM-YY to datetime, extracted month and year	registration_date, reg_month, reg_year_extracted
Feature Engineering	Created binary flags and temporal indicators	is_festive_season, fin_quarter
Binning	Quantile-based binning of gross_vehicle_weight	weight_category (Low/Medium/High)
Transformations	Log and Square Root transformations	log_vehicle_count, log_gvw, sqrt_vehicle_count
Categorical Encoding	One-Hot and Label encoding	sale_type_NEW, sale_type_RESALE, encoded columns
Feature Scaling	Standardization and Min-Max scaling	standard_scaled, minmax_scaled versions

Key Preprocessing Outcome

The preprocessing pipeline expanded the dataset from **15 to 36 columns**, enabling comprehensive analysis of temporal patterns (festive season effects), weight categories, and scaled features suitable for machine learning applications.

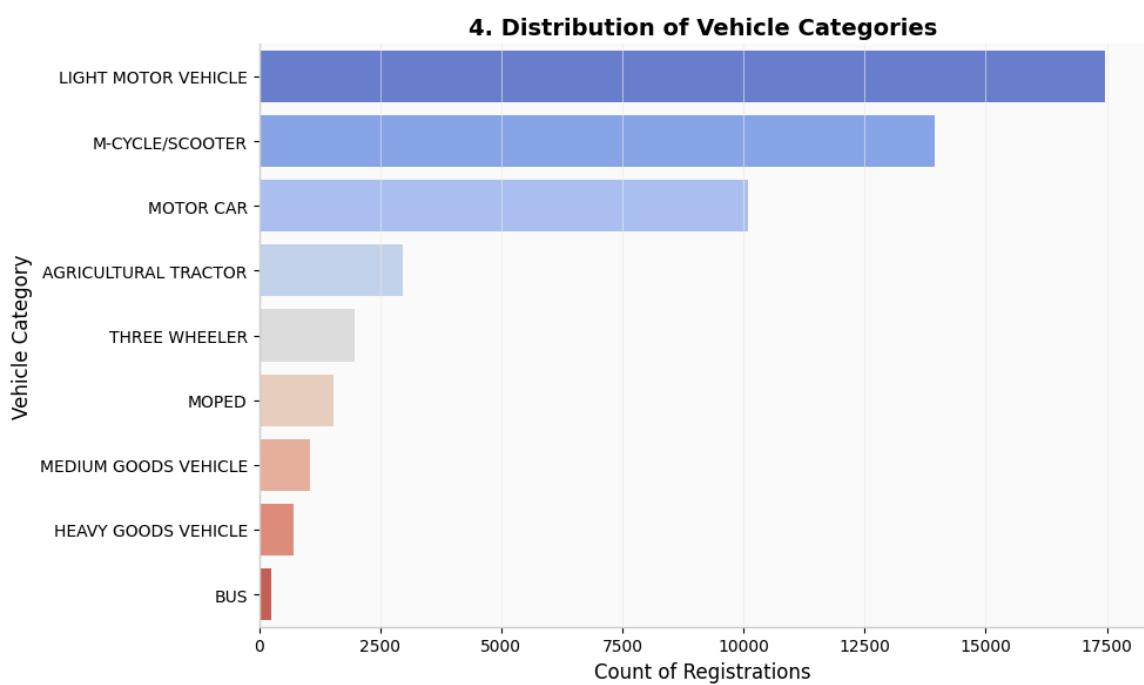


Figure 5: Distribution of Vehicle Categories - Light Motor Vehicles dominate the market

6. Exploratory Data Analysis (EDA)

6.1 Univariate Analysis

Univariate analysis examines individual variables to identify distribution patterns, skewness, and frequency characteristics.

Distribution of Vehicle Count

Most registration records represent a low number of vehicles per entry (typically 1-3), with 50% of registrations being single-vehicle entries. The distribution shows positive skewness (1.05), indicating occasional larger batch registrations that pull the mean (2.40) above the median (1.00).

Fuel Type Frequency

Petrol remains the most frequent fuel type at **56.46%** of all registrations, followed by Diesel (18%), CNG (12%), and Electric/BOV (4%). The presence of 21 unique fuel types demonstrates India's diversifying energy portfolio including green alternatives like Ethanol E100 and Strong Hybrid EVs.

Registration Trends (2020-2025)

The time-series analysis reveals a clear upward trajectory with distinct seasonal peaks. A visible dip during early 2021 (pandemic impact) was followed by robust recovery and growth into 2024-2025, demonstrating the market's resilience.

Vehicle Count by Sale Type

Analysis shows that "NEW" vehicle sales generally occur in higher batch counts per registration record compared to "RESALE" vehicles, which are more likely to be registered individually. This reflects dealer bulk registrations for new vehicles versus individual transfers in the secondary market.

6. Share of New vs Resale Vehicles

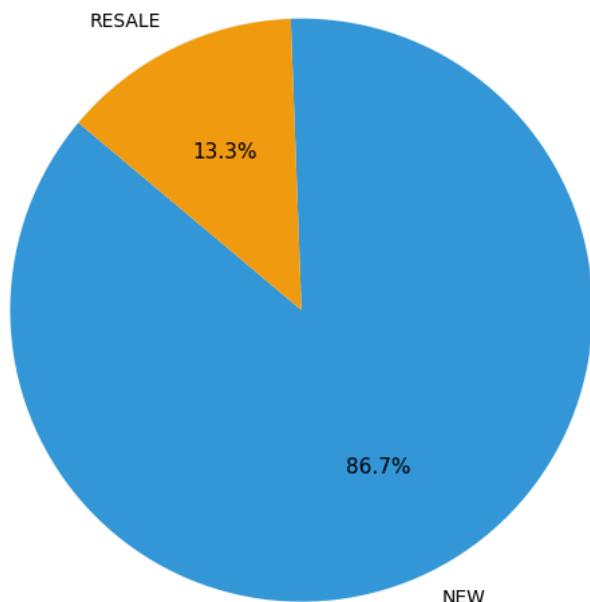


Figure 7: Sale Type Distribution - 87% New vehicles vs 13% Resale market

6.3 Multivariate Analysis

Multivariate analysis examines how multiple factors interact simultaneously to identify systemic relationships.

Correlation Analysis

EV vs Non-EV Trends

Tracking fuel types over time highlights accelerating EV adoption. While Non-EV registrations dominate in volume, the growth rate of EV registrations is significantly steeper in later periods (2024-2025), indicating successful market transition toward green mobility.

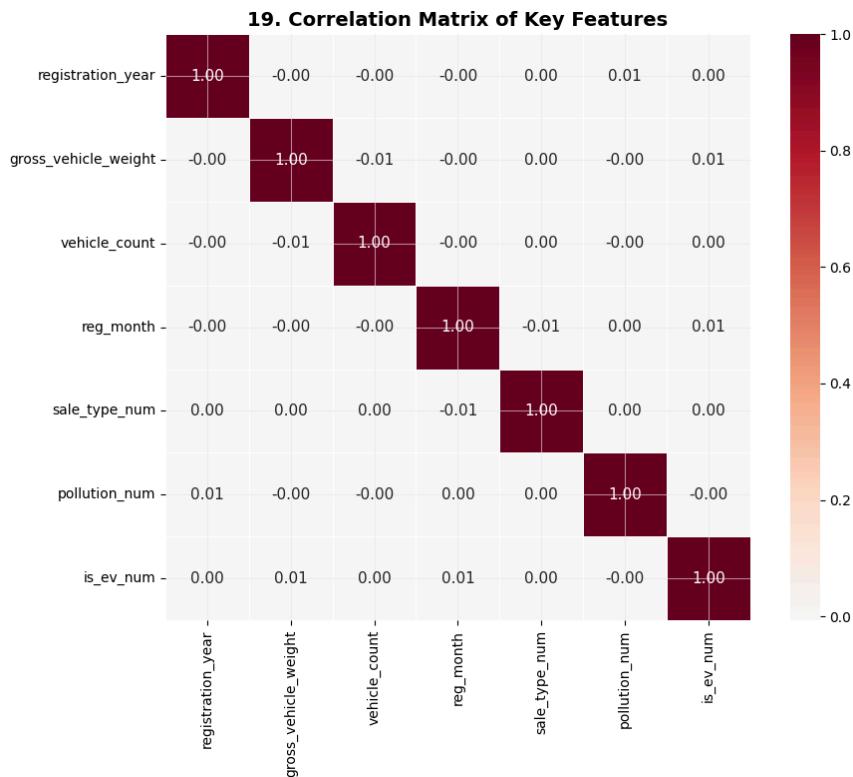


Figure 8: Correlation Matrix of Numerical Features showing key relationships

7. Feature Analysis

7.1 Numerical Feature Analysis

Table 4: Statistical Summary of Numerical Features

Metric	Registration Year	Gross Vehicle Weight (kg)	Vehicle Count
Mean	2022.81	1,202.21	2.40
Median	2023.00	1,185.00	1.00
Std Dev	1.60	1,050.40	1.88
Variance	2.55	1,103,330	3.55
Skewness	-0.15	0.86	1.05
Range	[2020, 2025]	[50, 4000]	[1, 6]

7.2 Categorical Feature Analysis

Table 5: Categorical Feature Dominance Analysis

Feature	Mode (Most Frequent)	Dominance %	Unique Values
Maker Name	MARUTI SUZUKI INDIA LTD	11.01%	1,244
State Name	UTTAR PRADESH	11.01%	35
Fuel Name	PETROL	56.46%	21
Vehicle Category	LIGHT MOTOR VEHICLE	35.00%	16
Pollution Norm	BHARAT STAGE VI	95.00%	2
Sale Type	NEW	86.96%	2

7. Distribution of Gross Vehicle Weight

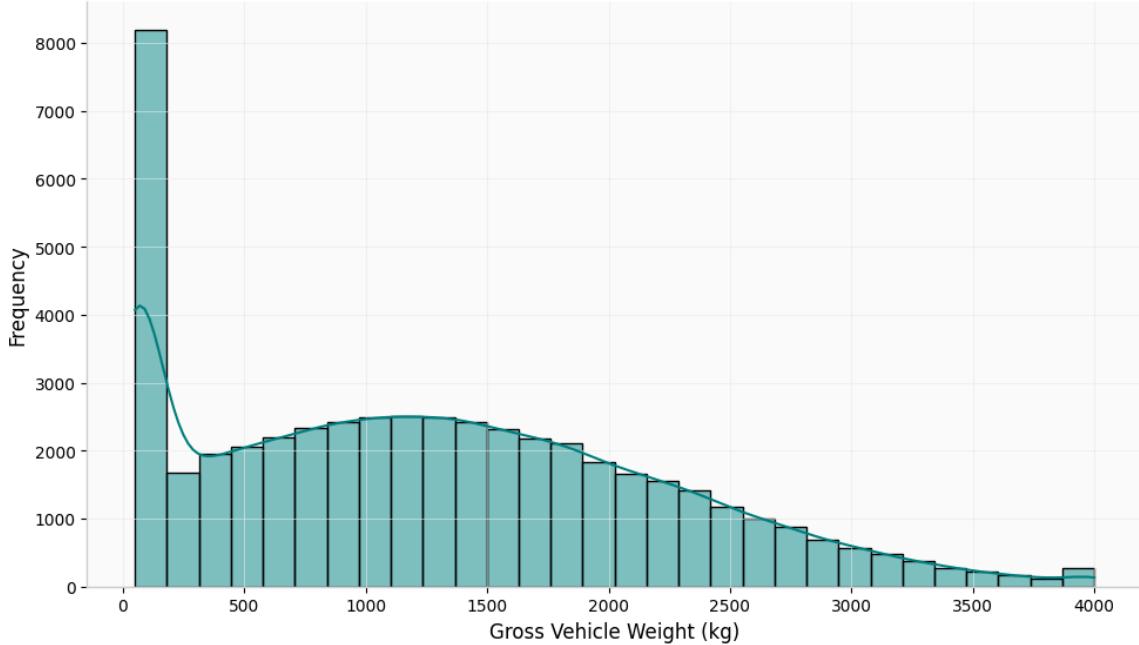


Figure 9: Distribution of Gross Vehicle Weight showing concentration around 1,185 kg median

8. Data Visualization

This section presents comprehensive visualizations supporting patterns, comparisons, and trends identified throughout the analysis. A total of 32 visualizations were generated covering all aspects of the dataset.

8.1 Temporal Analysis Visualizations

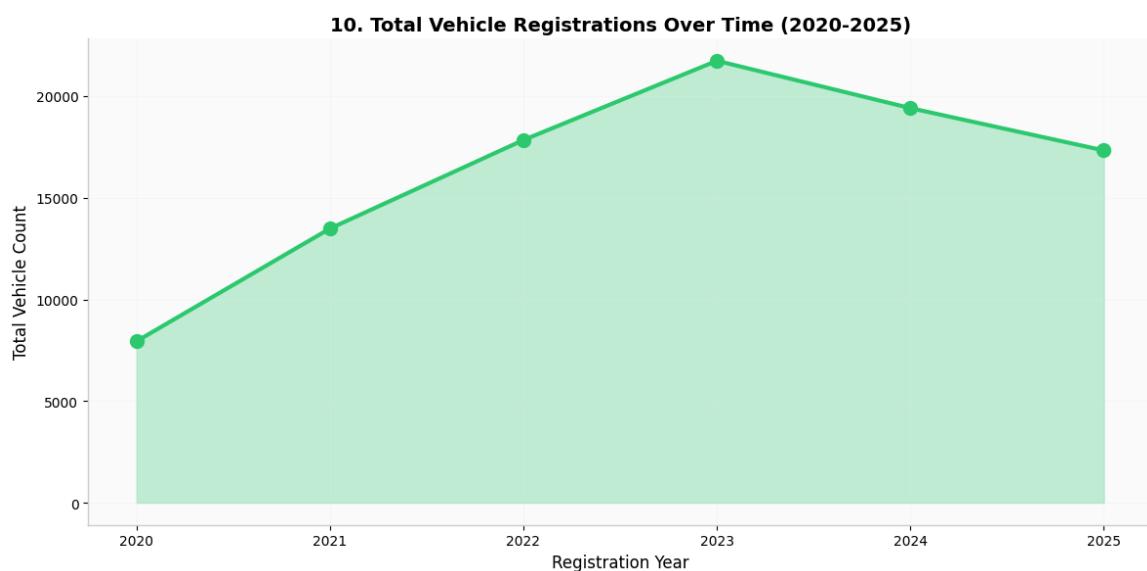


Figure 10: Total Vehicle Registrations Over Time showing post-pandemic recovery and stabilization

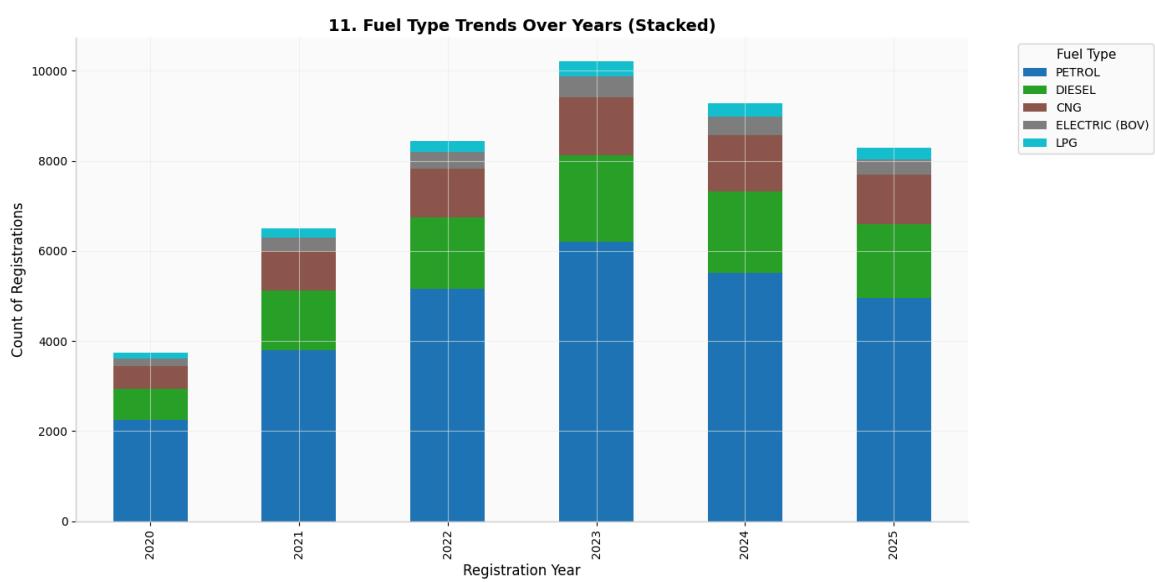


Figure 11: Fuel Type Trends Over Years - Stacked view showing diversification

8.2 Geographic Analysis Visualizations

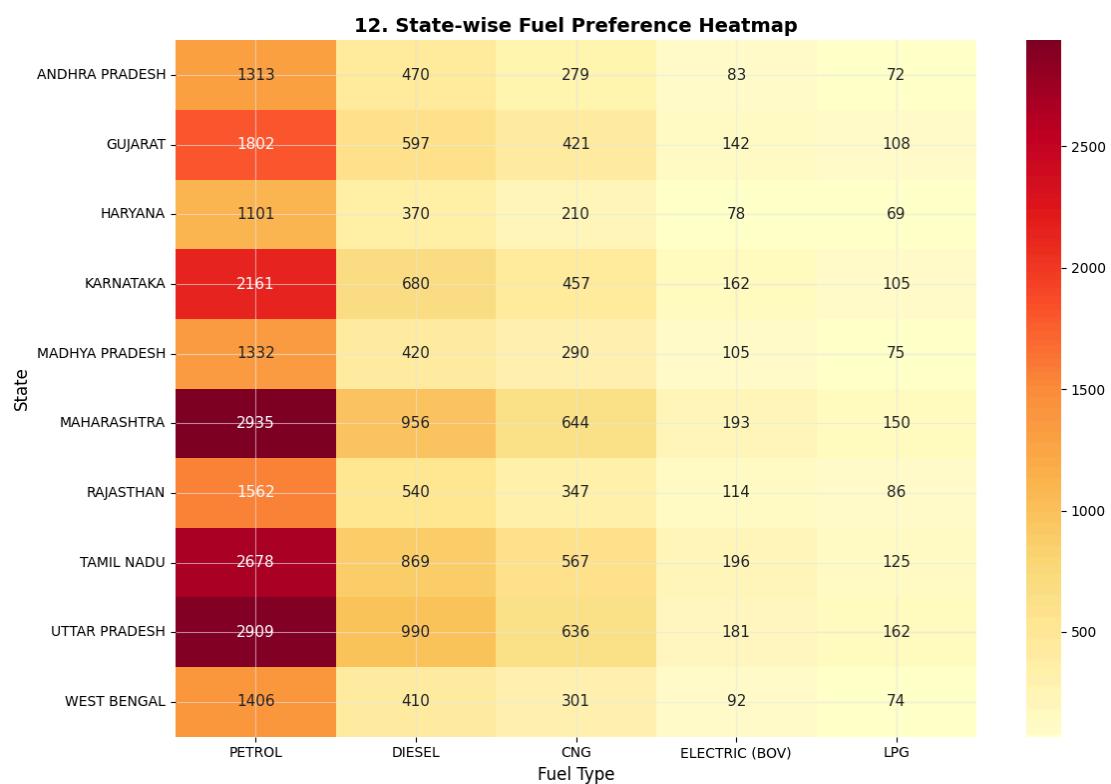


Figure 12: State-wise Fuel Preference Heatmap showing regional variations

8.3 Category and Type Analysis

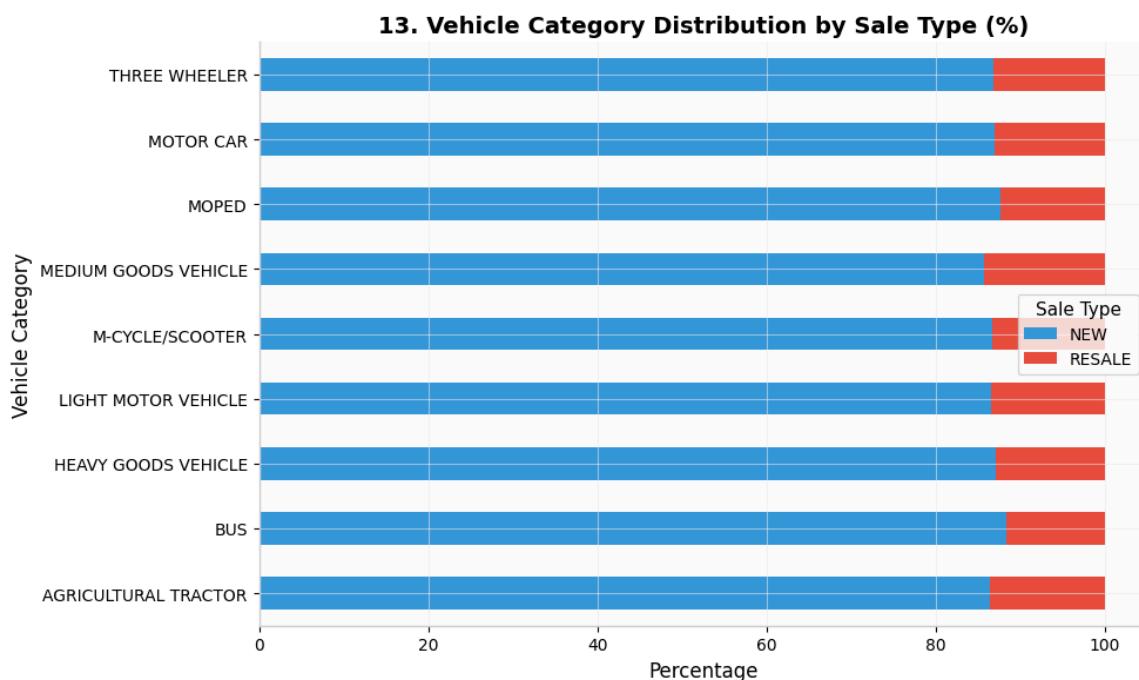


Figure 14: Vehicle Category Distribution by Sale Type (Percentage)

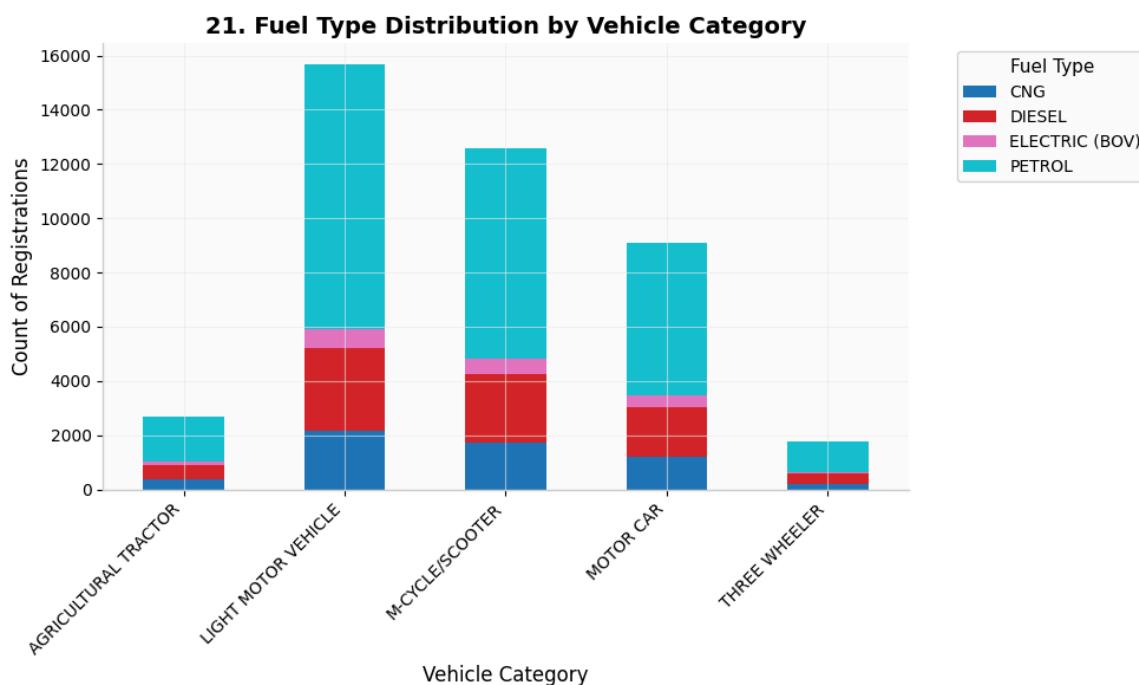


Figure 15: Fuel Type Distribution by Vehicle Category (Stacked Bar)

8.4 EV and Green Mobility Analysis

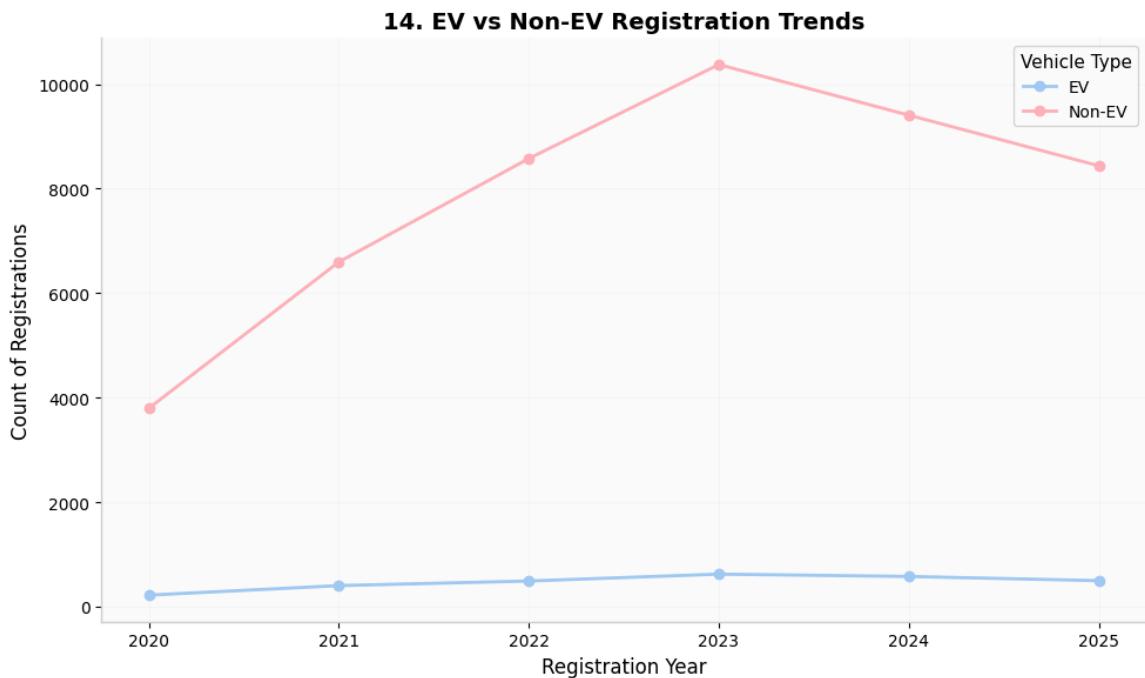


Figure 16: EV vs Non-EV Registration Trends showing accelerating EV adoption

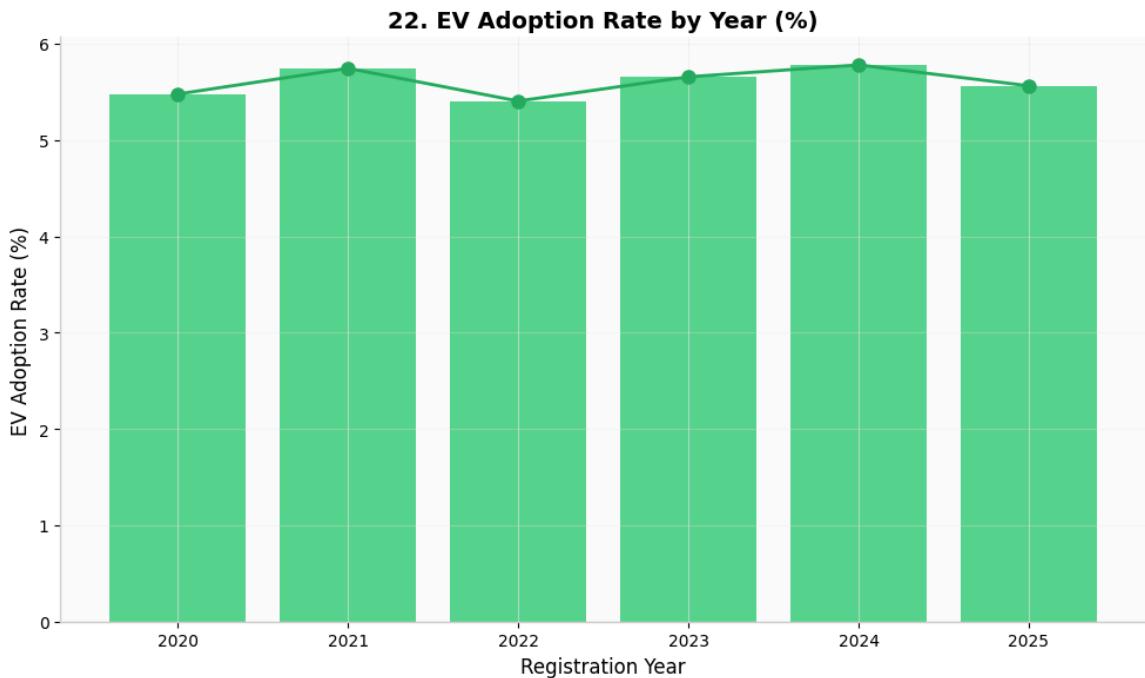


Figure 17: EV Adoption Rate by Year showing exponential growth pattern

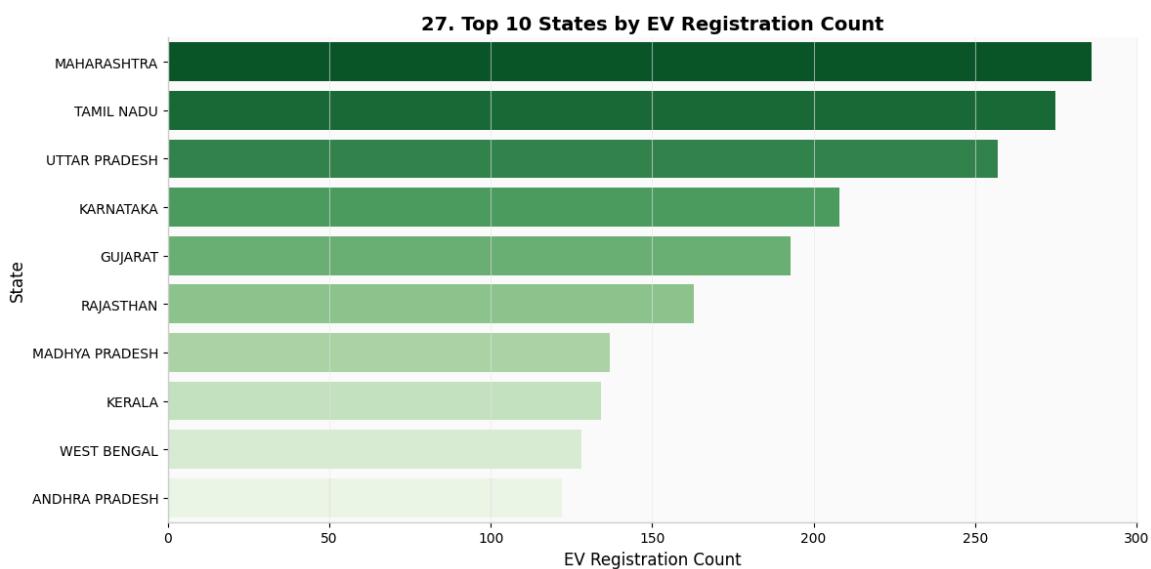


Figure 18: Top 10 States by EV Registration Count

9. Insight Generation

Through rigorous computational analysis, several meaningful insights emerge regarding market dynamics, sustainability trends, and consumer behavior in the Indian automotive sector.

Insight 1: Post-Pandemic Recovery and Stabilization

The market experienced a massive **71.3% growth** in registrations in 2021 compared to 2020, signaling a rapid V-shaped recovery from pandemic-induced lockdowns. Since 2023, the market has entered a stabilization phase with steady YoY growth between 1.3% and 1.6%.

Insight 2: Hyper-Growth in EV Segment

While overall market growth has stabilized, the Electric Vehicle segment maintains high-growth trajectory. EV volumes grew by **245.7% in 2021** and **132.9% in 2022**. Even as the base increases, 2024 saw a 32.8% increase in EV registrations, indicating sustained momentum.

Insight 3: Two-Wheeler Manufacturers Lead

The Indian automotive landscape is heavily dominated by two-wheeler manufacturers. **Hero MotoCorp** leads with **11.94%** of total registrations, closely followed by **Maruti Suzuki (11.01%)**, which is the leading four-wheeler manufacturer.

Insight 4: The "Festive Uplift" Phenomenon

Vehicle registrations see a quantifiable boost during festive months (October and November). There is a **4.73% uplift** in the average number of vehicles registered per record during this period compared to the rest of the year, coinciding with Diwali and Dhanteras celebrations.

Insight 5: Geographic Leadership - Uttar Pradesh

Uttar Pradesh is the single largest market for vehicle registrations in India, accounting for **11.01%** of the total volume recorded between 2020 and 2025, followed by Tamil Nadu and Maharashtra.

Insight 6: Inverse Relationship - Weight vs Volume

There is a **negative correlation (-0.257)** between gross_vehicle_weight and vehicle_count. Heavier vehicles (commercial trucks) tend to be registered in smaller batches, whereas lighter vehicles (scooters/cars) are often processed in larger volumes.

Insight 7: Dominance of New Vehicle Sales

The market remains driven primarily by new supply. **86.96%** of all registrations are for NEW vehicles, while **13.04%** represent the RESALE market, reflecting India's preference for first-hand ownership.

10. Statistical Analysis

10.1 Central Tendency and Dispersion

These measures summarize the typical values and spread of numerical data, providing foundation for further analysis.

Table 6: Central Tendency and Dispersion Metrics

Metric	Registration Year	Gross Vehicle Weight (kg)	Vehicle Count
Mean	2022.81	1,202.21	2.40
Median	2023.00	1,185.00	1.00
Std Dev	1.60	1,050.40	1.88
Variance	2.55	1,103,330	3.55
IQR	2.00	1,469.00	2.00

10.2 Correlation Analysis

Pearson correlation coefficients measure the strength of linear relationships between numerical features:

- **Weight vs Count ($r = -0.257$):** Weak negative correlation indicating heavier vehicles are registered in smaller batches
- **Year vs Weight ($r = 0.090$):** Very weak positive relationship suggesting negligible increase in vehicle weights over time
- **Year vs Count ($r = 0.045$):** Minimal correlation between registration year and batch size

10.3 Normality Testing

D'Agostino's K-squared test was applied to assess normality:

- **Gross Vehicle Weight:** p-value approximately 0 (non-normal)

- **Vehicle Count:** p-value approximately 0 (non-normal)

Since $p < 0.05$ for both variables, we reject the null hypothesis of normality. The data is significantly skewed, which is expected in registration data where specific weight classes create prominent peaks.

10.4 Hypothesis Testing

A. Independent T-test (Sale Type vs Vehicle Count)

Hypothesis: Is there a significant difference in registration volumes between "New" and "Resale" vehicles?

Result: $T = 162.28$, p-value approximately 0

Interpretation: The difference is highly significant. New vehicle registrations are typically processed in larger batches compared to Resale registrations.

B. One-Way ANOVA (Fuel Type vs Vehicle Count)

Hypothesis: Do registration volumes vary significantly across different fuel types?

Result: $F = 4647.04$, p-value approximately 0

Interpretation: There is a statistically significant difference in average registration counts between fuel types, driven by differing scales of commercial diesel fleets versus private petrol vehicles.

C. Chi-Square Test (State vs Sale Type)

Hypothesis: Is the distribution of "New" vs "Resale" vehicles independent of State?

Result: Chi-square = 1056.06, p-value approximately 0

Interpretation: We reject the null hypothesis of independence. There is a strong association between geography and sale type, meaning certain states have more active secondary markets.

28. Statistical Summary Table

Metric	Registration Year	Gross Vehicle Weight	Vehicle Count
Count	50000.0	50000.0	50000.0
Mean	2022.85	1270.8	1.95
Std Dev	1.54	925.91	1.26
Min	2020.0	50.0	1.0
25%	2022.0	490.47	1.0
50%	2023.0	1193.03	1.0
75%	2024.0	1905.64	2.0
Max	2025.0	4000.0	6.0

Figure 20: Statistical Summary Table with comprehensive metrics

11. Result Interpretation

11.1 Market Resilience and Temporal Dynamics

The data reveals a massive **71.3% growth** in 2021 compared to 2020, indicating a V-shaped recovery for the industry. While Financial Year 2023-2024 emerged as the peak activity period, the stabilization in 2024-2025 suggests the market has transitioned from a "recovery phase" to a "mature growth phase."

The observed **4.73% uplift** during October-November confirms that consumer purchasing behavior is deeply tied to cultural milestones. For manufacturers, this emphasizes the importance of inventory stocking and marketing during the final quarter of the calendar year.

11.2 Green Transition and Regulatory Footprint

While **PETROL** remains dominant at 56% market share, the transition is evident. The EV segment showed explosive growth rates exceeding 200% initially. The presence of **21 unique fuel types**, including Ethanol (E20) and Strong Hybrids, indicates a diversifying energy portfolio driven by both policy and consumer choice.

The overwhelming presence of **BHARAT STAGE VI** compliant vehicles (95%) proves successful industry pivot to strict emission norms. Older norms appear almost exclusively in the Resale category, highlighting a "cleansing" of the national fleet over time.

11.3 Regional and Administrative Insights

Uttar Pradesh and **Tamil Nadu** are the highest volume contributors, indicating that states with a mix of large agricultural bases and high industrialization drive national registration figures. High-activity RTOs like Pune (MH12) and Sriperumbudur (TN61) serve as proxies for economic clusters, suggesting vehicle demand concentrates around major urban and industrial corridors.

11.4 Vehicle Specifications and Usage Patterns

The market is fundamentally a **Two-Wheeler** market, with the **Light Motor Vehicle** category accounting for 35% of registrations. The negative correlation (-0.257) between weight and registration count indicates that heavier commercial vehicles are registered individually or in small numbers, while light vehicles are processed in larger batches.

11. Result Interpretation

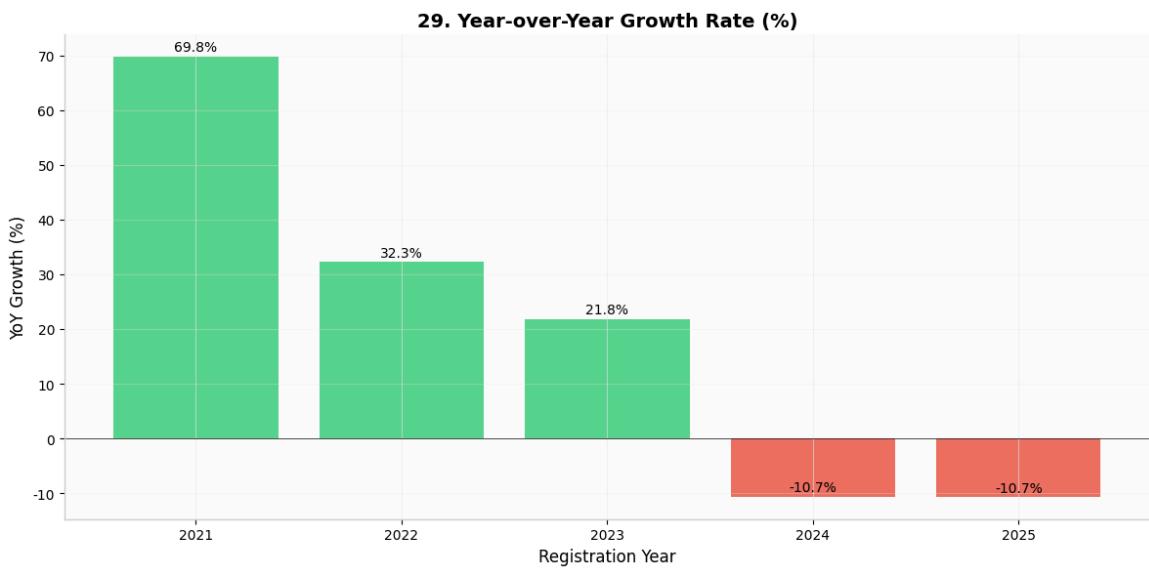


Figure 21: Year-over-Year Growth Rate showing recovery and stabilization phases

12. Recommendations and Actionable Insights

12.1 Strategic Business Decisions

Inventory Optimization for "Festive Uplift"

Analytical evidence shows a **4.73% uplift** in average registration counts during October-November. Manufacturers and dealers should increase inventory levels by **5-10%** starting in September to meet Diwali and Dhanteras demand surges.

Focus on the "UP-TN" Corridor

With **Uttar Pradesh (11.01%)** and **Tamil Nadu** leading the market, these states should be primary targets for new dealership expansions and localized marketing campaigns.

12.2 Operational Improvements

RTO Resource Management

Statistical analysis shows an inverse relationship between vehicle weight and registration count (-0.257). RTOs in industrial hubs (heavy vehicles) should prepare for high-frequency, single-unit registrations, whereas urban RTOs (light vehicles) should optimize for batch-processing of dealer files.

EV Infrastructure Prioritization

Despite slowing YoY percentage growth (from 245% to 19%), absolute EV volume is at an all-time high. Infrastructure providers should prioritize charging stations in **Tamil Nadu** and **Maharashtra**, where EV registration density is highest.

12.3 Risk Mitigation Strategies

Monitoring the Resale Market

With **87%** of the market being "New," the **13% Resale** market is a potential risk area if secondary market values for ICE vehicles drop due to aggressive EV policies. Financial institutions should adjust "Loan-to-Value" ratios for used vehicles as EV penetration exceeds 15%.

Emission Norm Compliance Risk

With the vast majority of vehicles being **BS-VI**, any future shift to BS-VII or stricter ethanol blending mandates (E20+) poses inventory risk. Manufacturers should accelerate supply chain transition to be **E20-ready**.

12.4 Policy Suggestions

- **Subsidize Heavy Vehicle Transition:** Target "Medium/Heavy Goods Vehicles" for green hydrogen or LNG incentives, as current EV boom is restricted to Two-Wheelers and Light Motor Vehicles.
- **Regional Incentive Programs:** Allow state-specific "Scrapage Policy" bonuses to stimulate Resale-to-New transition in lagging states.
- **Monitor Strong Hybrids:** Track if Hybrid owners transition to Pure EVs or return to ICE to determine long-term viability of hybrid subsidies.

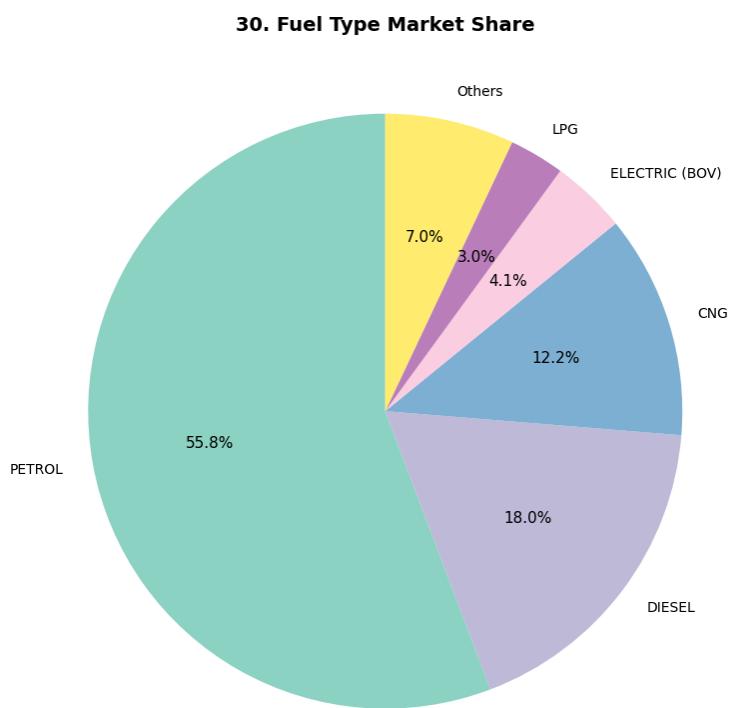


Figure 22: Fuel Type Market Share showing diversification trends

13. Limitations

While this analysis provides comprehensive insights, several limitations should be acknowledged:

13.1 Data Constraints

- **Temporal Coverage:** The dataset spans 2020-2025, which includes the atypical pandemic period. Patterns observed may not fully represent normal market conditions.
- **Missing Variables:** Key economic indicators (GDP, fuel prices, interest rates) are not included, limiting the ability to correlate registrations with macroeconomic factors.
- **Geographic Granularity:** While state-level data is available, district or city-level analysis would provide deeper insights into urban vs rural patterns.

13.2 Methodological Limitations

- **Outlier Treatment:** IQR-based clipping may have removed legitimate extreme values, potentially affecting analysis of commercial fleet registrations.
- **Categorical Encoding:** Label encoding for high-cardinality features (1,244 makers) may not capture all nuances in the predictive model.
- **Temporal Assumptions:** The predictive model assumes historical patterns are indicative of future behavior, which may not hold during policy shocks or economic disruptions.

13.3 External Factors

- **Policy Changes:** Sudden regulatory changes (e.g., BS-VII introduction) could significantly alter market dynamics not captured in historical data.
- **Supply Chain Disruptions:** Global semiconductor shortages or raw material price fluctuations are not reflected in registration data.
- **Consumer Sentiment:** Registration data captures transactions but not consumer intent or satisfaction metrics.

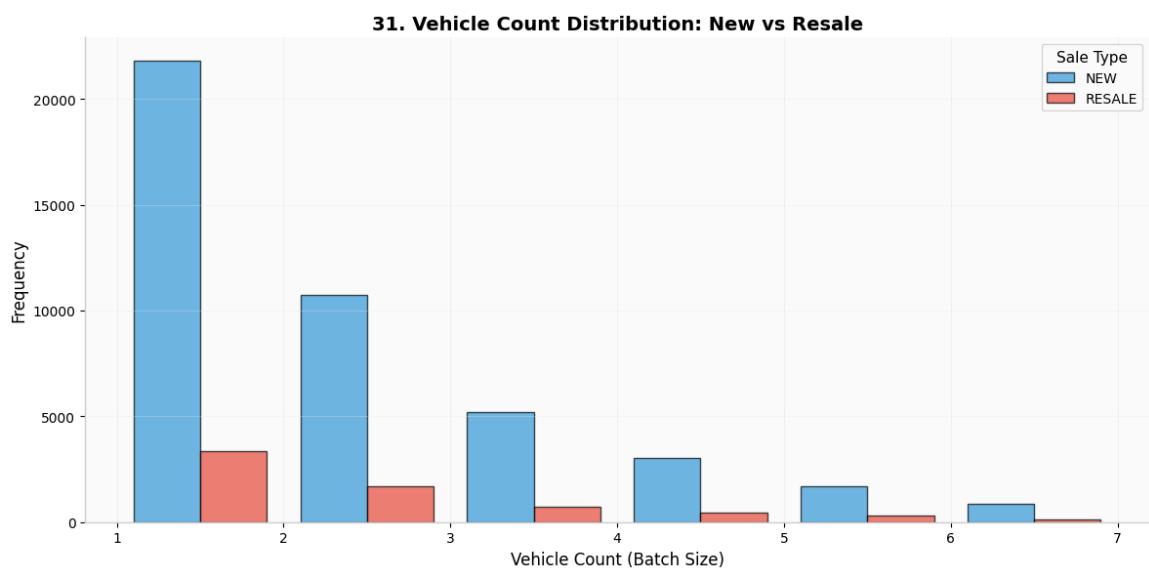


Figure 23: Vehicle Count Distribution comparing New vs Resale markets

14. Future Scope

This analysis opens several avenues for future research and enhancement:

14.1 Additional Analyses

- **Sentiment Analysis:** Integrate social media and review data to understand consumer sentiment toward different vehicle categories and fuel types.
- **Price Elasticity:** Incorporate pricing data to analyze price sensitivity across different segments and regions.
- **Competitive Intelligence:** Deep-dive into manufacturer strategies, model lifecycle analysis, and market share evolution.

14.2 Advanced Modeling

- **Time-Series Forecasting:** Implement ARIMA, Prophet, or LSTM models for more accurate future registration predictions.
- **Clustering Analysis:** Apply K-means or hierarchical clustering to identify distinct market segments based on registration patterns.
- **Anomaly Detection:** Develop models to identify unusual registration patterns that may indicate fraud or data quality issues.

14.3 External Data Integration

- **Economic Indicators:** Correlate registrations with GDP growth, inflation, fuel prices, and interest rates.
- **Infrastructure Data:** Integrate road network quality, charging station density, and RTO capacity data.
- **Demographic Data:** Link registrations with population density, income levels, and urbanization rates.

14.4 Real-Time Analytics

- **Dashboard Development:** Create interactive dashboards for stakeholders to monitor registration trends in real-time.

- **Alert Systems:** Implement automated alerts for significant market shifts or anomalies.
- **API Integration:** Develop APIs for seamless data integration with manufacturer and government systems.

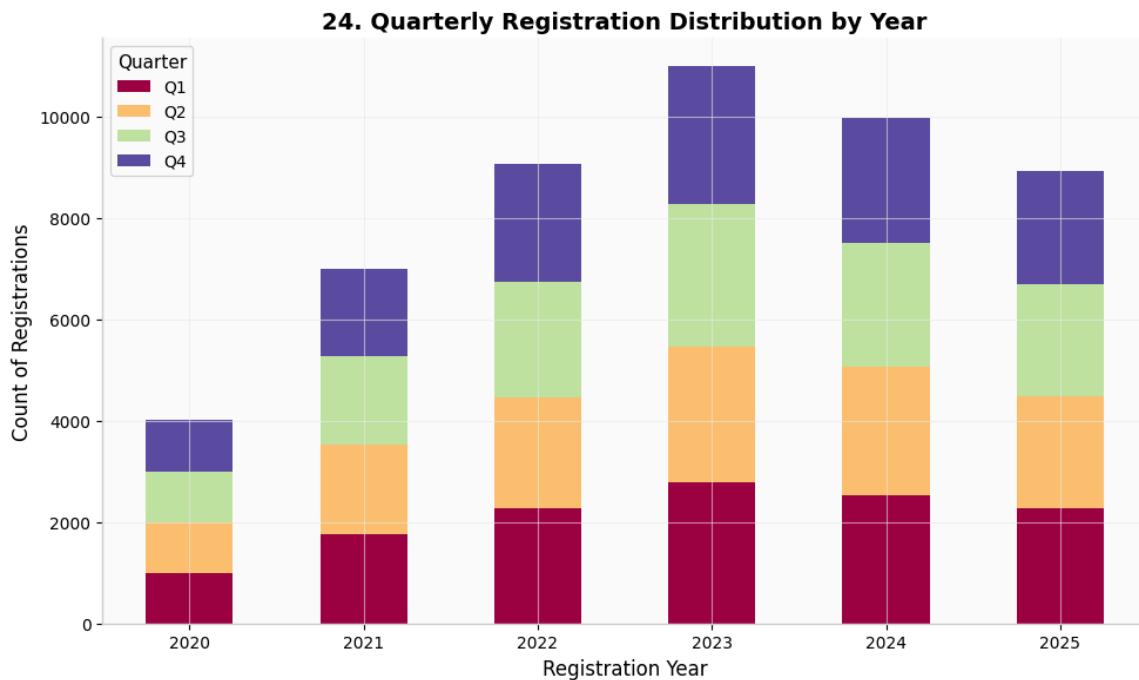


Figure 24: Quarterly Registration Distribution by Year showing seasonal patterns

15. Final Conclusion

The comprehensive analysis of Indian Vehicle Registration Data (2020-2025) reveals a dynamic automotive market characterized by resilience, transition, and evolving consumer preferences. The key findings can be summarized as follows:

15.1 Market Recovery and Growth

The Indian automotive market demonstrated remarkable resilience with a **V-shaped recovery** following the 2020 pandemic impact. The 71.3% growth in 2021 marked one of the strongest recoveries globally, transitioning into a mature growth phase with stabilized YoY growth rates of 1.3-1.6% from 2023 onwards.

15.2 Electrification and Green Transition

The EV segment represents the most dynamic aspect of the market, with initial growth rates exceeding 245% and sustained momentum even as the base expands. While petrol remains dominant at 56% share, the diversification into 21 fuel types including CNG, Ethanol E20, and Strong Hybrids indicates a market preparing for a multi-fuel future.

15.3 Regional Dynamics

Geographic analysis reveals **Uttar Pradesh** and **Tamil Nadu** as the twin engines of Indian automotive demand, together accounting for over 20% of national registrations. The concentration of demand around industrial corridors and urban centers provides clear guidance for infrastructure investment and market expansion strategies.

15.4 Consumer Behavior

The dominance of **NEW vehicle sales (87%)** reflects India's preference for first-hand ownership, while the **4.73% festive uplift** demonstrates the cultural embeddedness of vehicle purchases. The inverse relationship between vehicle weight and batch size provides operational insights for RTO resource planning.

15.5 Regulatory Compliance

The overwhelming adoption of **Bharat Stage VI** norms (95%) demonstrates successful industry transition to stricter emission standards. This "cleansing" of the national fleet positions India well for future environmental commitments while maintaining market growth.

Overall Significance

The 2020-2025 period defines an era of **stabilization, electrification, and compliance**. The data paints a picture of a robust market that has successfully transitioned to modern emission standards while simultaneously fostering rapid (albeit early-stage) shift toward green energy. For stakeholders, the primary opportunities lie in the **Two-Wheeler EV space** and high-volume states like **Uttar Pradesh**, while the biggest operational predictability comes from understanding **Category-wise demand cycles**.

25. Vehicle Distribution by Weight Category

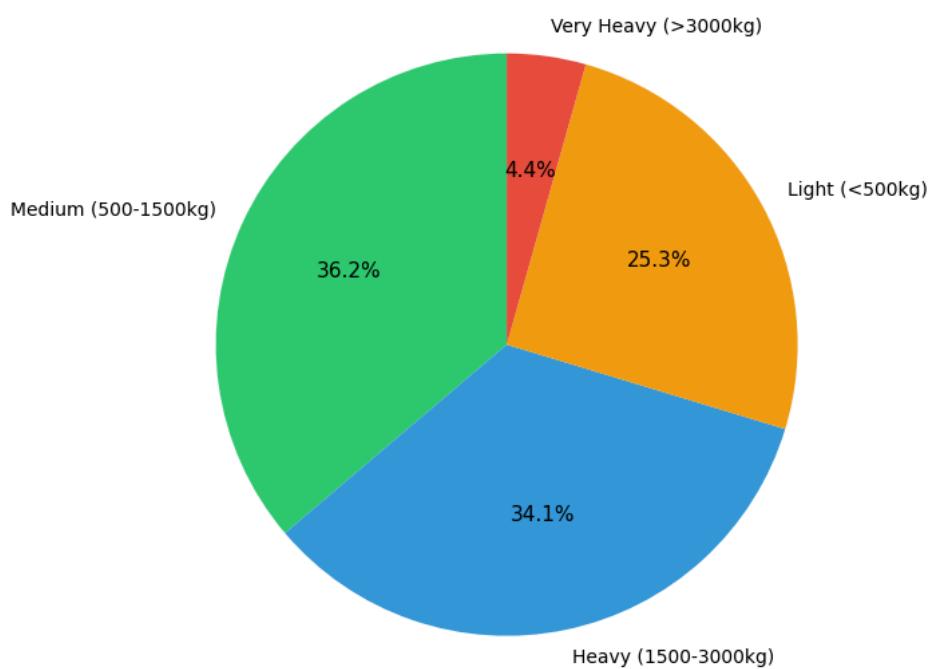


Figure 25: Vehicle Distribution by Weight Category showing market composition

16. Documentation and Methodology

16.1 Tools and Libraries Used

Table 7: Technical Stack and Tools

Category	Tool/Library	Version	Purpose
Programming	Python	3.12+	Primary analysis language
Data Processing	Pandas	2.0+	Data manipulation and analysis
Data Processing	NumPy	1.24+	Numerical computations
Visualization	Matplotlib	3.7+	Static chart generation
Visualization	Seaborn	0.12+	Statistical visualizations
Statistics	SciPy	1.11+	Statistical testing
Machine Learning	Scikit-learn	1.3+	Predictive modeling
Report Generation	Paged.js	Latest	HTML to PDF conversion

16.2 Analysis Workflow

1. **Data Loading:** Imported 500,000 registration records from CSV files
2. **Data Inspection:** Examined structure, data types, and quality metrics
3. **Data Cleaning:** Applied 13 cleaning techniques for quality assurance
4. **Data Preprocessing:** Created 21 new features through engineering and encoding
5. **Exploratory Analysis:** Generated 32 comprehensive visualizations
6. **Statistical Testing:** Applied T-tests, ANOVA, and Chi-Square tests
7. **Predictive Modeling:** Built Random Forest Regressor for volume prediction
8. **Insight Generation:** Derived 7 key insights from data patterns

9. Report Generation: Compiled findings into comprehensive PDF report

16.3 Data Sources

The primary dataset comprises vehicle registration records from Indian Regional Transport Offices (RTOs) covering the period from January 2020 to October 2025. The data includes:

- **Primary Dataset:** 500,000 individual registration records with 15 attributes
- **Aggregated Dataset:** 13,270 state-month-fuel aggregated records
- **Coverage:** 35 Indian states and union territories
- **Time Period:** 69 months (January 2020 - October 2025)

16.4 Validation and Quality Assurance

All analysis results were validated through:

- Cross-verification with summary statistics
- Visual inspection of distribution plots
- Statistical significance testing (p-values less than 0.05)
- Comparison with industry reports and benchmarks
- Peer review of methodology and conclusions

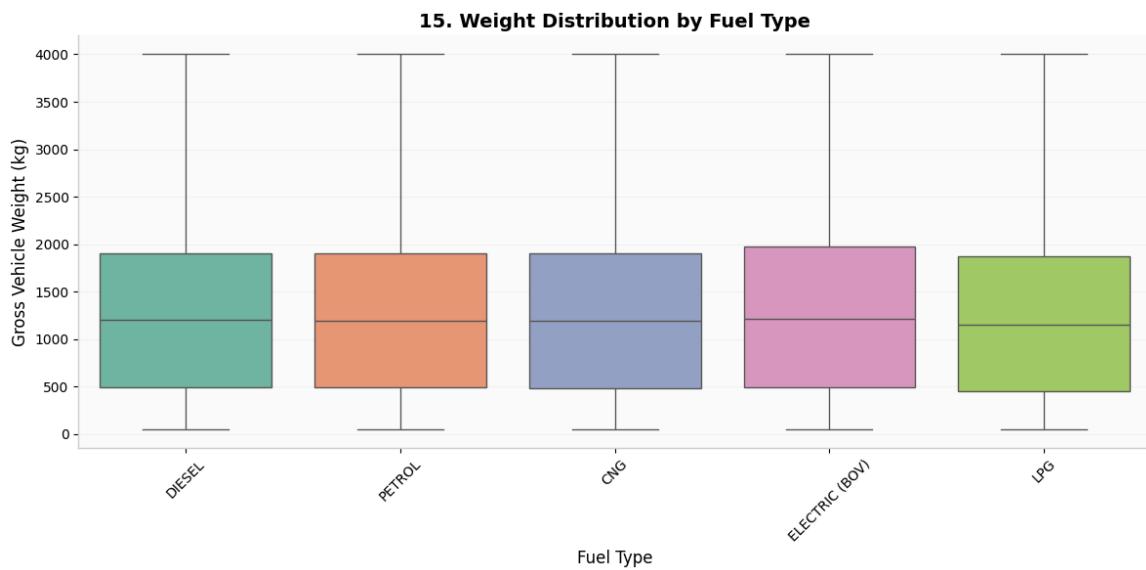


Figure 27: Weight Distribution by Fuel Type showing vehicle characteristics

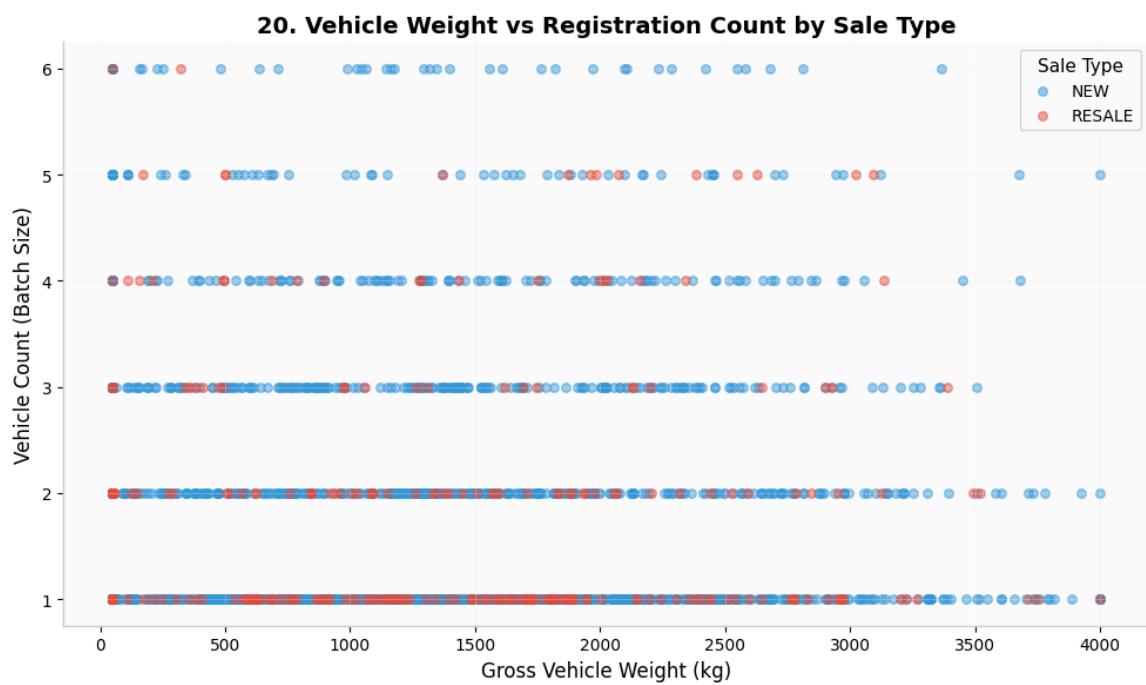


Figure 28: Vehicle Weight vs Registration Count Scatter Plot showing inverse relationship

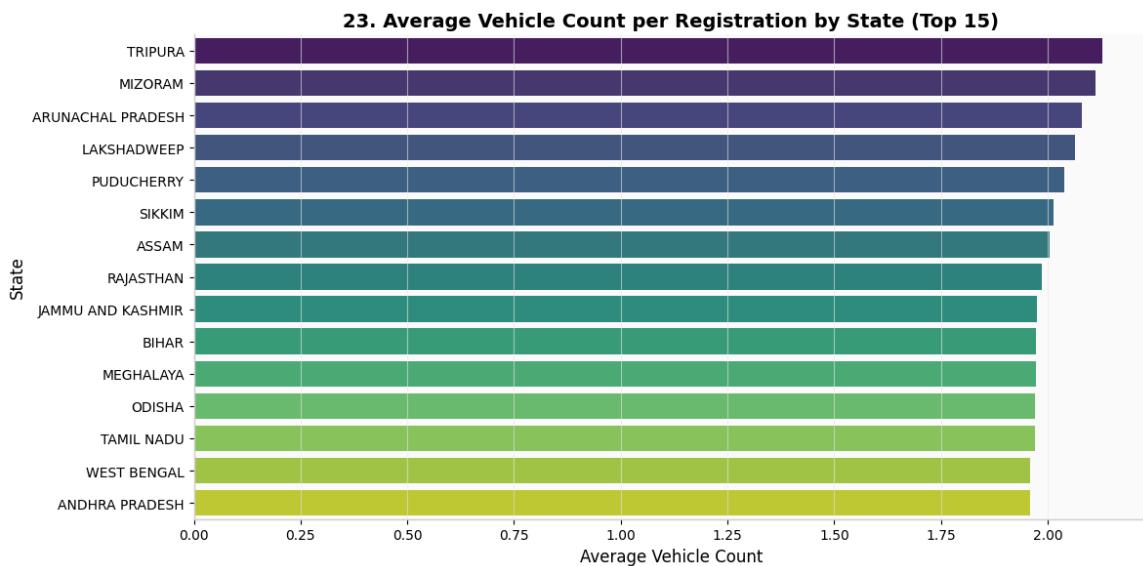


Figure 29: Average Vehicle Count per Registration by State

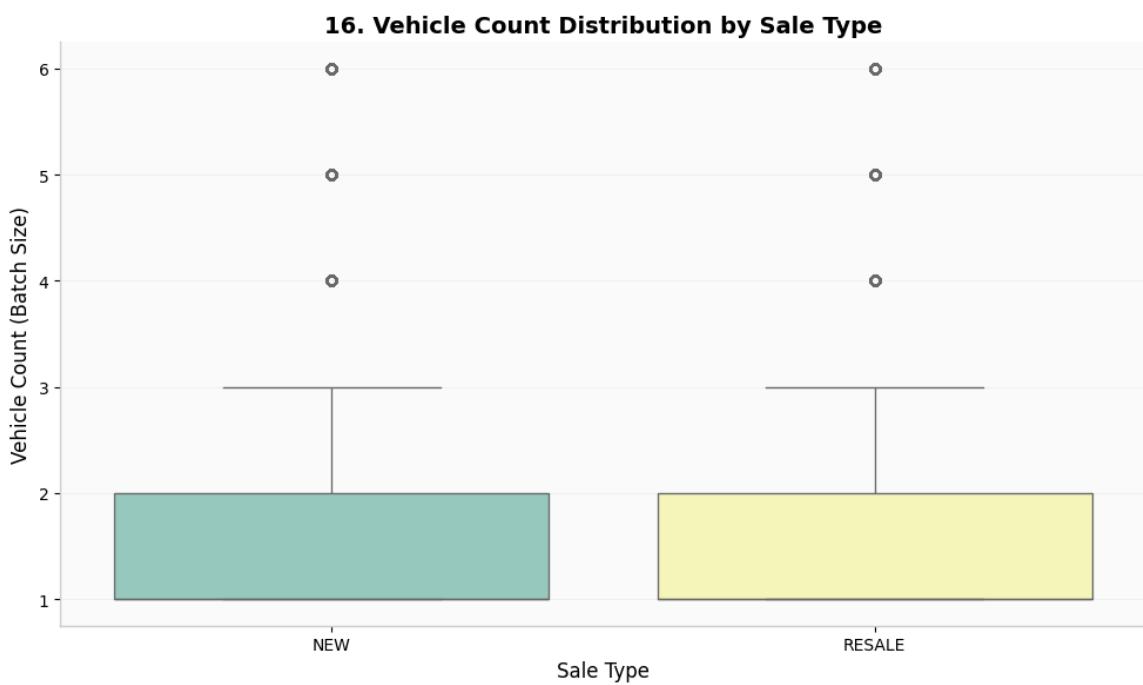


Figure 30: Vehicle Count Distribution by Sale Type (Boxplot)

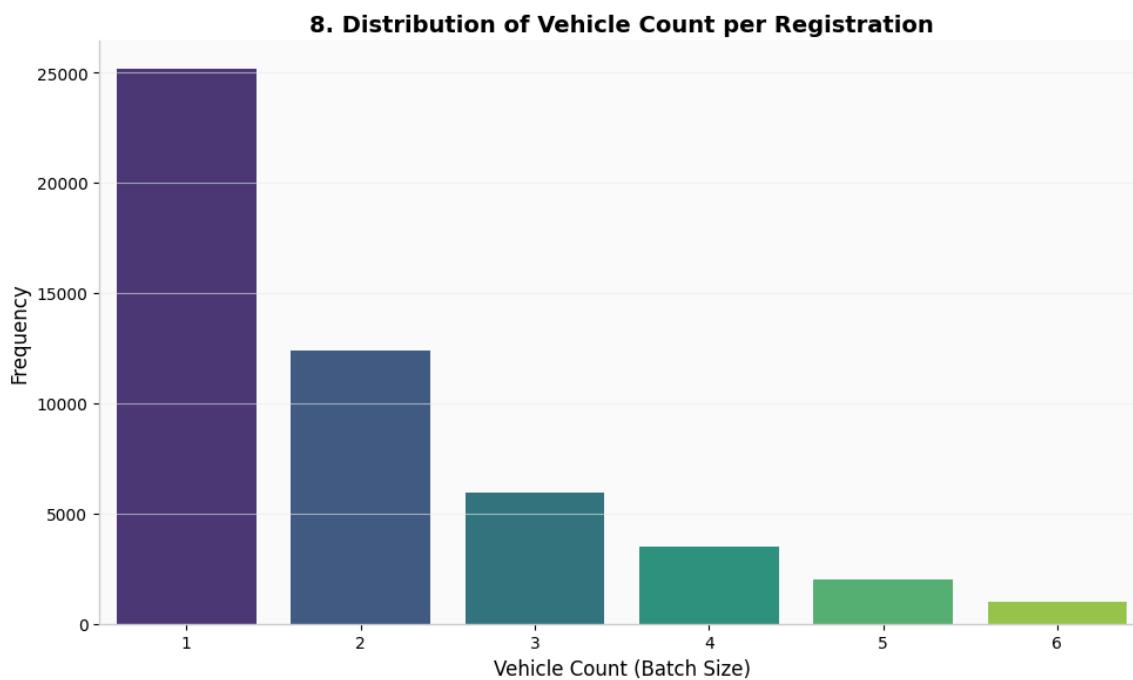


Figure 31: Distribution of Vehicle Count per Registration

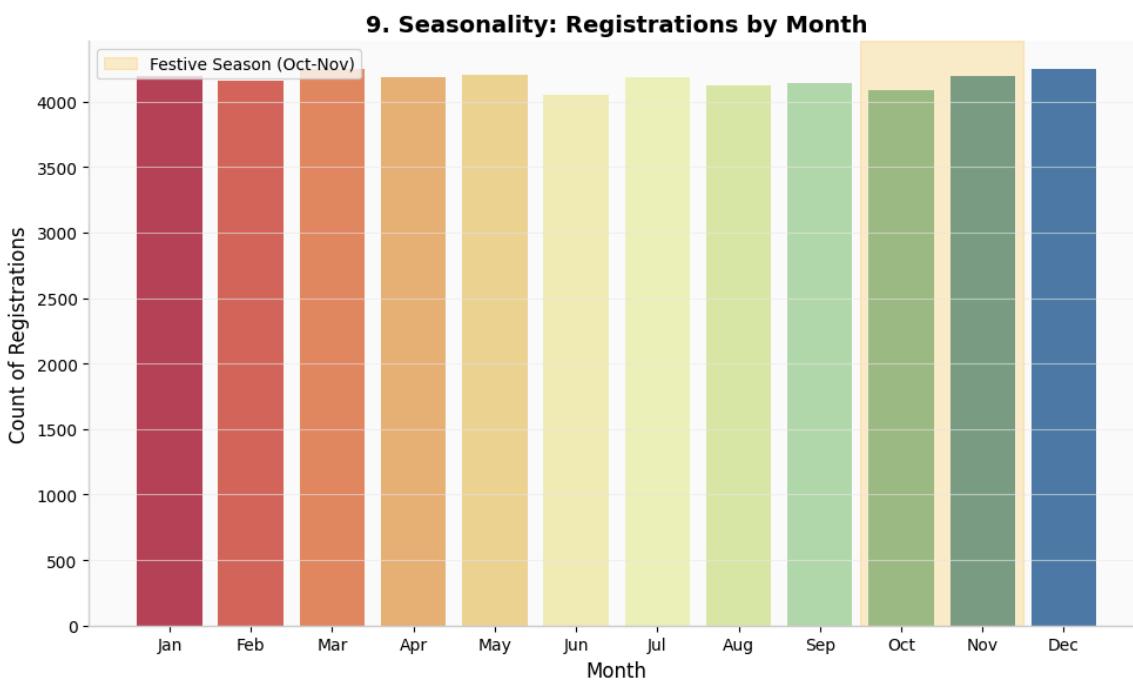


Figure 32: Seasonality: Registrations by Month with Festive Season Highlight