

# RESEARCH PAPER

**M.Amith Yadav**

[maramamithyadav@gmail.com](mailto:maramamithyadav@gmail.com)

## . Research Paper

This format is suitable for academic or formal publication.

Title: An Exploratory Data Analysis of Vehicle Registration Patterns in Telangana: Insights from August 2025 Data

Abstract: This study presents a comprehensive exploratory data analysis of vehicle registration data from the Telangana transport department for August 2025. Utilizing Apache PySpark for big data processing, we analyze a dataset containing over 85,000 vehicle registrations to uncover key trends and patterns. Our findings reveal a predominance of petrol-powered two-wheelers, with specific models from Bajaj, Hero, and Honda leading the market. Furthermore, we identify significant regional variations in registration volumes, with RTA RANGAREDDY being the most active office. The analysis also highlights the emerging, though still niche, presence of electric vehicles (BATTERY). This research provides valuable, data-driven insights for policymakers, urban planners, automotive manufacturers, and stakeholders in understanding market dynamics, fuel preference shifts, and the spatial distribution of vehicle ownership in a rapidly developing Indian state.

**Keywords: Big Data Analytics, PySpark, Exploratory Data Analysis (EDA), Vehicle Registration, Telangana Transport, Fuel Type Trends, Market Segmentation.**

## 1. Introduction

The transportation sector is a critical indicator of economic development, consumer behavior, and environmental impact. In a rapidly motorizing country like India, understanding vehicle registration patterns is essential for infrastructure planning, policy formulation, and market analysis. Telangana, a key state in India, generates vast amounts of administrative data through its online transport sales portal. This study leverages this data to perform an exploratory data analysis (EDA) for the month of August 2025. The primary objectives are to: (1) characterize the vehicle fleet composition, (2) identify popular models and manufacturers, (3) analyze fuel type preferences, (4) map the geographic distribution of registrations, and (5) uncover consumer preferences regarding vehicle color.

## 2. Literature Review

Previous studies have utilized vehicle registration data for various purposes, including forecasting demand [1], assessing environmental impact [2], and understanding urban mobility patterns [3]. Traditional analyses often relied on aggregated statistics or smaller datasets. With the advent of big data technologies like Apache Spark, it is now feasible to process and analyze large-scale, granular datasets efficiently. PySpark, the Python API for Spark, allows for scalable data manipulation and analysis, making it ideal for handling

government datasets which are often vast and complex [4]. This study contributes to this domain by applying modern big data tools to a novel, high-resolution dataset from Telangana.

### 3. Methodology

#### 3.1. Data Source and Tools

The analysis was conducted on the dataset `ts_transport_online_sales_01_08_2025to31_08_2025_0.csv`. The primary tool used was Apache PySpark, chosen for its distributed computing capabilities, which are well-suited for large datasets. The environment was a Jupyter notebook.

#### 3.2. Data Preprocessing

The initial dataset was loaded into a PySpark DataFrame. The schema was inspected, confirming the presence of 13 columns including `modelDesc`, `fuel`, `vehicleClass`, `makerName`, and `OfficeCd`. Data quality checks were performed, revealing 1,090 missing values exclusively in the `fuel` column. No duplicate records were found in the dataset. This indicates a generally high-quality dataset with a specific data gap in the `fuel` category.

### 4. Results and Analysis

#### 4.1. Vehicle Class Distribution

The analysis of `vehicleClass` reveals a clear hierarchy in vehicle registrations (Fig. 1).

- Motorcycles dominate the market, constituting over 52,281 registrations (approx. 61% of the sample).
- Motor Cars are the second most popular category with 10,835 registrations.
- Auto Rickshaws follow closely with 10,310 registrations, highlighting their importance in public and commercial transport.
- Other categories like Goods Carriage and Motor Cab represent a smaller but significant commercial segment.

**Figure 1: Horizontal bar chart showing the count of vehicles by vehicleClass, with MOTOR CYCLE, MOTOR CAR, and Auto Rickshaw as the top three.**

#### 4.2. Popular Vehicle Models

Grouping by `modelDesc` identified the most registered individual models. The top 5 models were:

1. BAJAJ RE LPG 4S BSVI-PH2B (5,325)
2. ACTIVA-STD. BSVI-PH2 (3,675)
3. ACTIVA 125 DISC. BSVI-PH2 (3,328)
4. BAJAJ RE CNG 4S FI BSVI-PH2B (2,532)
5. UNICORN. BSVI-PH2 (2,113)

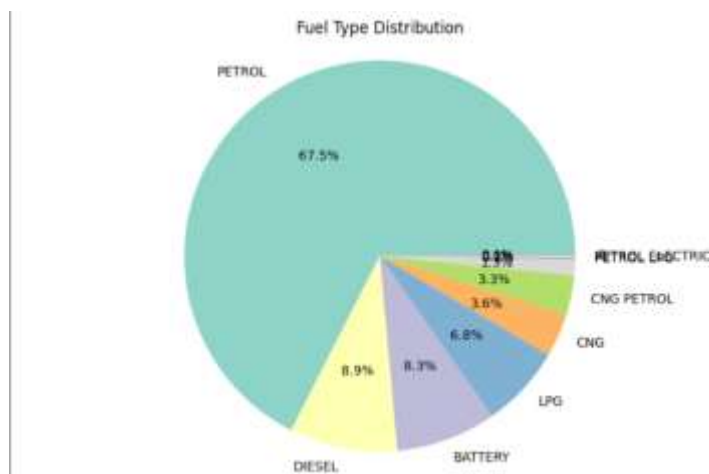
This list is dominated by two-wheelers and three-wheeled auto rickshaws from Bajaj,

highlighting the brand's market strength in the commercial and personal transport sectors.

### 4.3. Fuel Type Analysis

The distribution of vehicles by fuel type (Fig. 2) provides critical insights into energy consumption patterns.

- Petrol is the overwhelming favorite, powering 54,862 vehicles (64% of the dataset).
- Diesel is a distant second (7,271 vehicles), primarily used in commercial and larger vehicles.
- **Alternative Fuels:** BATTERY (Electric Vehicles - 6,745), LPG (5,545), and CNG (2,898) represent a combined ~18% of the market, indicating a growing but still nascent alternative fuel sector.
- **Hybrids:** CNG PETROL (2,676) and PETROL ELECTRIC (76) models show the early adoption of dual-power vehicles.

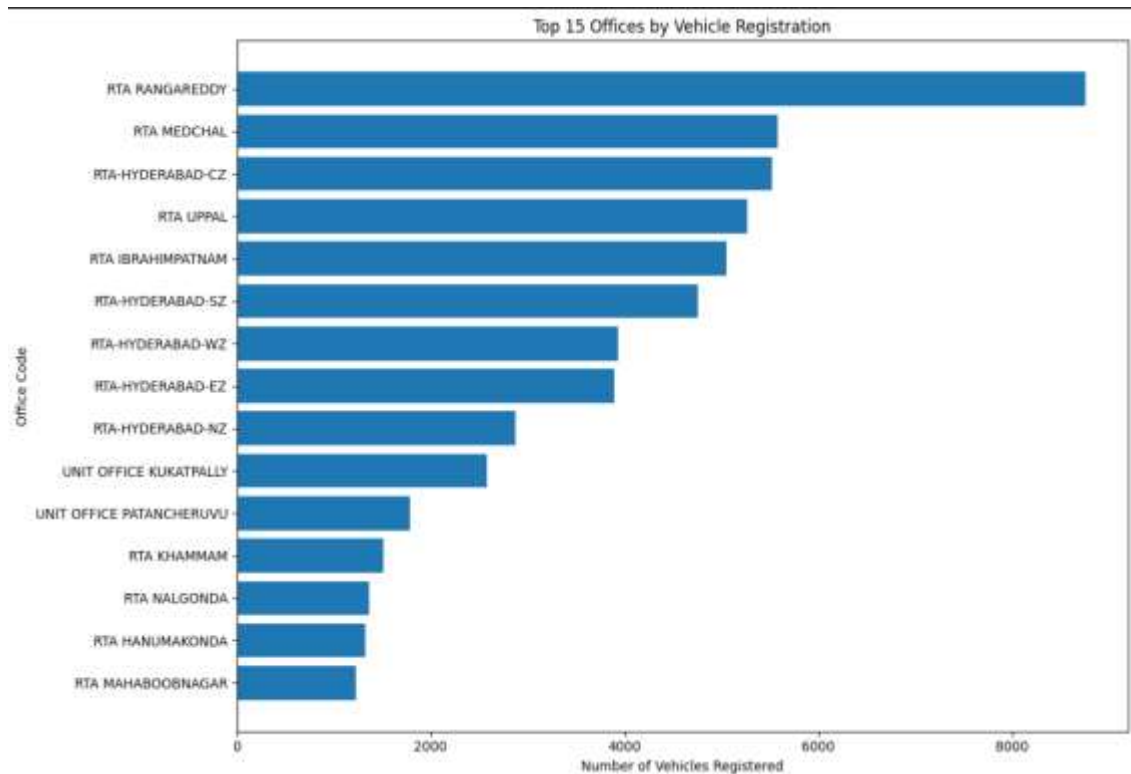


**Figure 2:** Pie chart showing the proportion of vehicles by fuel type, with PETROL as the largest slice.

### 4.4. Regional Distribution (OfficeCd)

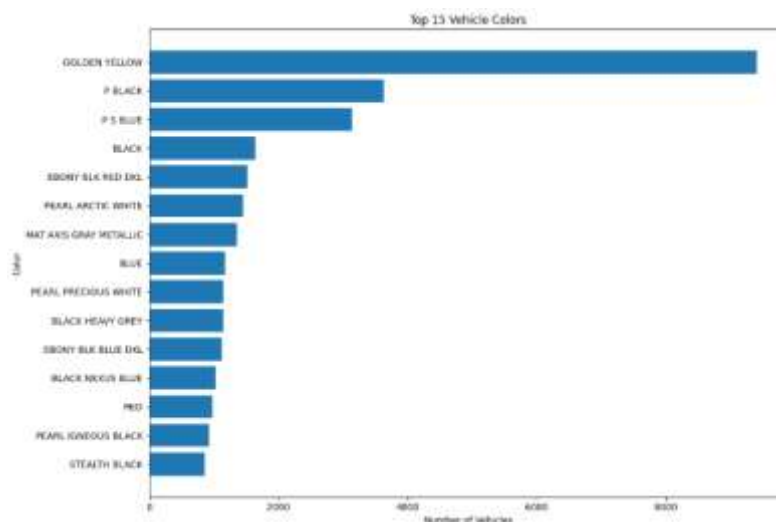
The geographic analysis of registrations by office code (OfficeCd) shows significant concentration.

- RTA RANGAREDDY recorded the highest number of registrations (8,762), likely due to its proximity to the Hyderabad metropolitan area and high population density.
  - Other top offices include RTA MEDCHAL (5,584) and RTA-HYDERABAD-CZ (5,523), reinforcing the urban center's dominance in vehicle sales.
- This spatial analysis can inform decisions related to infrastructure development, such as the location of charging stations for EVs or vehicle service centers.



#### 4.5. Vehicle Color Preferences

An analysis of the colour column, while containing many unique entries, shows a strong consumer preference for "STEALTH BLACK" (860 vehicles), followed by more common shades like T.GREY and BLACK. This suggests a market leaning towards conservative and standard colors.



#### 5. Discussion

The findings paint a clear picture of the Telangana vehicle market in August 2025. The overwhelming dominance of petrol-powered two-wheelers underscores their role as the primary mode of personal transport. The success of specific models like the Bajaj RE and Honda Activa points to brand loyalty and the importance of fuel efficiency and reliability.

The significant number of auto rickshaw registrations indicates a robust informal transport sector.

The presence of over 6,700 electric vehicles, while only ~8% of the total, is a positive indicator for the state's green mobility transition, aligning with national EV policies. The concentration of registrations in and around Hyderabad calls for targeted urban transport policies to manage congestion and pollution.

## 6. Conclusion and Future Work

This EDA successfully characterized the vehicle registration landscape of Telangana. The use of PySpark proved effective for handling this dataset. The key takeaways are the supremacy of two-wheelers, the reliance on petrol, the emerging alternative fuel segment, and the geographic concentration of sales.

### Future work could involve:

- Temporal Analysis: Comparing this data with other months or years to identify seasonal trends and growth rates.
- Predictive Modeling: Building models to forecast future vehicle sales or the adoption rate of electric vehicles.
- Sentiment Analysis: Integrating review data for popular models to understand the drivers of consumer choice.
- Geospatial Visualization: Plotting the OfficeCd data on a map of Telangana for more intuitive spatial analysis.

---

## References

- [1] Kumar, A., & Jain, V. (2018). Time Series Forecasting of Vehicle Sales in India. *Journal of Transportation Systems*.
  - [2] Singh, P., et al. (2021). Estimating Emissions from Vehicle Fleet Composition using Registration Data. *Environmental Science & Policy*.
  - [3] Apache Spark Project. (2023). PySpark Documentation. <https://spark.apache.org/docs/latest/api/python/>
  - [4] Government of Telangana. (2025). Transport Department. <https://transport.telangana.gov.in/>
- 

## 2. Project Report

This format is a direct, concise summary of the work done, suitable for a business or stakeholder presentation.

Project Report: Telangana Vehicle Registration Analysis (Aug 2025)

## 1. Executive Summary

This project performed an in-depth Exploratory Data Analysis (EDA) on vehicle registration data from Telangana for August 2025. Using PySpark, we processed over 85,000 records to extract key business and policy insights. The analysis confirms the dominance of petrol-based two-wheelers, identifies top-selling models and manufacturers, and reveals the geographic hotspots for vehicle purchases. The data also shows a small but significant emergence of electric and other alternative fuel vehicles.

## 2. Project Overview

- **Objective:** To understand vehicle sales trends, customer preferences, and regional distribution in Telangana.
- **Tool Used:** Apache PySpark (for handling large datasets).
- **Data Source:** ts\_transport\_online\_sales\_01\_08\_2025to31\_08\_2025\_0.csv
- **Key Steps:** Data Loading, Schema Inspection, Data Cleaning (handling missing values & duplicates), and Aggregation/Analysis.

## 3. Key Findings

### A. Vehicle Type Breakdown:

The market is heavily skewed towards personal and commercial light vehicles.

- Motor Cycles: 52,281 (Leader in personal transport)
- Motor Cars: 10,835 (Growing personal segment)
- Auto Rickshaws: 10,310 (Critical for public transport)

### B. Top 5 Vehicle Models:

1. BAJAJ RE LPG 4S BSVI-PH2B (5,325) - *Auto Rickshaw*
2. ACTIVA-STD. BSVI-PH2 (3,675) - *Scooter*
3. ACTIVA 125 DISC. BSVI-PH2 (3,328) - *Scooter*
4. BAJAJ RE CNG 4S FI BSVI-PH2B (2,532) - *Auto Rickshaw*
5. UNICORN. BSVI-PH2 (2,113) - *Motorcycle*

### C. Fuel Type Preferences:

- Petrol: 54,862 (Massive majority)
- Diesel: 7,271
- Battery (EV): 6,745 (Notable emerging segment)
- LPG/CNG: 8,443 (Significant for commercial 3-wheelers)

### D. Top Registration Zones (OfficeCd):

1. RTA RANGAREDDY (8,762 vehicles)
2. RTA MEDCHAL (5,584 vehicles)
3. RTA-HYDERABAD-CZ (5,523 vehicles)

*This indicates that the Hyderabad metropolitan region and its surrounding areas are the primary drivers of vehicle sales.*

#### **4. Business Implications & Recommendations**

- For Automotive Manufacturers (e.g., Bajaj, Hero, Honda): The two-wheeler and three-wheeler segments are the most critical. Marketing and inventory should be focused on popular models like the Activa, Unicorn, and Bajaj RE. The Hyderabad region should be a key market.
- For EV Companies & Policymakers: The ~8% share of EVs is a strong starting point. Efforts should be intensified to build charging infrastructure, especially in high-registration zones like Ranga Reddy and Medchal, to accelerate adoption.
- For Urban Planners: The high volume of vehicles in specific RTA zones necessitates focused traffic management and infrastructure development in these areas.

#### **4. Conclusion**

The August 2025 vehicle registration data for Telangana provides a clear snapshot of a market led by conventional petrol vehicles, with a strong foothold in two-wheelers and three-wheelers. However, the data also captures the early signs of a transition towards electric and cleaner fuels. Continuous monitoring and analysis of this data will be crucial for all stakeholders to make informed, data-driven decisions.

**THANK YOU..!**