Report

M.Amith Yadav

maramamithyadav@gmail.com

PySpark Data Analysis Report: Telangana Transport Online Sales (August 2025)

Project Overview

This project analyzes vehicle registration data from Telangana's online transport sales system for August 2025 using PySpark. The dataset contains detailed information about various vehicle types registered during this period.

Dataset Information

- **Source File**: ts_transport_online_sales_01_08_2025to31_08_2025_0.csv
- **Time Period**: August 1-31, 2025
- **Total Records**: Approximately 78,000+ vehicles
- **Data Quality**: No duplicates found, minimal missing values (primarily in fuel type)

Data Structure

The dataset contains 13 columns:

- modelDesc: Vehicle model description
- fuel: Fuel type (PETROL, DIESEL, BATTERY, CNG, LPG, etc.)
- colour: Vehicle color
- vehicleClass: Type of vehicle
- makeYear: Manufacturing year
- seatCapacity: Number of seats
- secondVehicle: Whether it's a second vehicle (Y/N)
- tempRegistrationNumber: Temporary registration ID
- category: Transport/Non-Transport classification
- makerName: Manufacturer name
- OfficeCd: Registration office code
- fromdate & todate: Registration period

Key Insights

1. Vehicle Type Distribution

Motorcycles dominate the market:

- **MOTOR CYCLE**: 52,281 vehicles (66.8%)
- **MOTOR CAR**: 10,835 vehicles (13.8%)
- **Auto Rickshaw**: 10,310 vehicles (13.2%)
- **Goods Carriage**: 1,899 vehicles (2.4%)
- Other vehicle types make up the remaining 3.8%

2. Popular Vehicle Models

Top Motorcycle Models:

1. BAJAJ RE LPG 4S BSVI-PH2B: 5,325 units
2. ACTIVA-STD. BSVI-PH2: 3,675 units
3. ACTIVA 125 DISC. BSVI-PH2: 3,328 units

Diverse Car Models: Many car models have only 1-2 registrations, indicating a fragmented car market.

3. Fuel Type Analysis

Fuel Preference Distribution:

- **PETROL**: 54,862 vehicles (70.1%)
- **DIESEL**: 7,271 vehicles (9.3%)
- **BATTERY (Electric)**: 6,745 vehicles (8.6%)
- **LPG**: 5,545 vehicles (7.1%)
- **CNG**: 2,898 vehicles (3.7%)

Notable Observation: Electric vehicles show significant adoption at 8.6% of total registrations.

4. Regional Distribution

Top Registration Offices:

1. RTA RANGAREDDY: 8,762 vehicles
2. RTA MEDCHAL: 5,584 vehicles
3. RTA-HYDERABAD-CZ: 5,523 vehicles
4. RTA UPPAL: 5,265 vehicles

5. RTA IBRAHIMPATNAM: 5,048 vehicles

Insight: Hyderabad metropolitan area and surrounding regions show the highest registration volumes.

5. Color Preferences

Popular Colors:

- STEALTH BLACK: 860 vehicles
- Various shades of blue, gray, and white follow
- The dataset shows diverse color preferences with many unique color options

6. Vehicle Category

- **Non-Transport**: Personal vehicles (majority)
- **Transport**: Commercial vehicles

Technical Implementation

Data Processing Steps:

1. **Data Loading**: Successfully loaded CSV using PySpark with inferred schema
2. **Data Quality**: Checked for missing values and duplicates
3. **Schema Validation**: Verified data types and structure
4. **Exploratory Analysis**: Multiple aggregation queries to understand distribution patterns

Libraries Used:

- PySpark (SparkSession, DataFrame operations)
- Matplotlib/Seaborn (for planned visualizations)
- Standard Python data science stack

Business Implications

Market Trends:

1. **Two-wheeler Dominance**: Motorcycles constitute over 66% of registrations
2. **Electric Vehicle Growth**: Strong EV adoption at nearly 9% market share
3. **Urban Concentration**: Higher registration density in metropolitan areas
4. **Fuel Diversity**: Multiple fuel options gaining traction alongside traditional petrol

Recommendations:

1. **Infrastructure Planning**: Focus on EV charging infrastructure in high-registration areas
2. **Inventory Management**: Stock popular models like Bajaj RE and Activa series
3. **Regional Strategy**: Tailor marketing efforts to regional preferences
4. **Service Network**: Expand service centers in top registration regions

Data Quality Assessment

- **Excellent Data Integrity**: No duplicate records
- **Minimal Missing Data**: Only 1,090 missing fuel entries (1.4% of total)
- **Consistent Formatting**: Standardized vehicle classifications and categories

Potential Enhancements

1. **Temporal Analysis**: Compare with previous months to identify trends
2. **Geospatial Analysis**: Map registration patterns across Telangana
3. **Price Analysis**: Incorporate pricing data for revenue insights
4. **Seasonal Patterns**: Analyze registration fluctuations throughout the year

This analysis provides a comprehensive overview of vehicle registration patterns in Telangana, offering valuable insights for policymakers, automotive manufacturers, and infrastructure planners.