Analyzing Commercial Energy Consumption Patterns in Telangana Using PySpark: A Case Study of TG-NPDCL

Publication Date: October 18, 2025

Author: Aerla Ritwik Reddy

Affiliation: MALLA REDDY UNIVERSITY, HYDERABAD

Abstract

Energy is a critical driver of economic growth, and understanding its consumption patterns is vital for efficient resource management and infrastructure planning. This paper presents a comprehensive analysis of commercial electricity consumption data from the Northern Power Distribution Company of Telangana Ltd. (TG-NPDCL) for September 2025. Leveraging the distributed computing power of Apache PySpark, this study processes and analyzes a dataset of 8,421 commercial service records to uncover key consumption trends, identify high-demand regions, and assess operational efficiencies. Our findings indicate a total commercial consumption of approximately 234 million kWh, with significant demand concentrated in urban and industrial hubs such as Warangal and Khammam. The analysis reveals a high billing efficiency, with a billed-to-total service ratio of nearly 70%. This research provides actionable insights for strategic planning, including targeted grid reinforcement and the implementation of smart metering solutions to further optimize energy distribution.

---

1. Introduction

The reliable distribution of electricity is fundamental to supporting commercial activities, which are cornerstones of a region's economy. Power distribution companies, or DISCOMs, face the ongoing challenge of balancing supply and demand, maintaining grid stability, and ensuring operational efficiency in the face of growing demand. Data-driven analysis of consumption patterns is essential for informed decision-making in this sector, moving from reactive maintenance to predictive and optimized grid management.

This study focuses on the commercial electricity consumption within the jurisdiction of the Northern Power Distribution Company of Telangana Ltd. (TG-NPDCL), which serves 16 districts in the state. As commercial and industrial hubs expand, the load on the grid intensifies, making granular analysis critical. By analyzing the complete consumption data for September 2025, we aim to:

- Quantify and profile the overall energy consumption and load.
- Identify the administrative circles with the highest energy demand.
- Analyze the load distribution and billing efficiency across different regions.

- Validate the consistency and integrity of the consumption data.
- Provide data-backed recommendations for infrastructure and operational improvements.

---

2. Methodology and Data Processing

2.1. Dataset

The dataset for this analysis is TG-NPDCL_consumption_detail_commercial_SEPTEMBER-2025.csv. It is a granular, record-level dataset with 8,421 entries and 11 columns. A preliminary data quality check, as noted in the analysis report, confirmed zero missing or null values, indicating a high-quality, clean dataset.

The dataset covers multiple administrative levels, including 16 Circles, 40 Divisions, 123 Sub-Divisions, and 454 Sections. Key data fields include:

- **Administrative Units:** Circle, Division, SubDivision, Section, Area
- **Service Metrics:** TotServices (Total registered service connections), BilledServices (Services billed during the month)
- **Consumption Metrics:** Units (Total energy consumed in kWh), Load (Connected load in kW)

2.2. Tools and Environment

The analysis was conducted within a Jupyter Notebook, providing an interactive environment for data processing and visualization.

- **PySpark:** Used for the core data processing. A SparkSession was initialized as the entry point for the application.
- **Pandas:** Used for converting small, aggregated PySpark DataFrames for visualization.
- **Matplotlib & Seaborn:** Employed to create static visualizations to represent the findings, such as bar charts and pie charts.

2.3. Analytical Approach

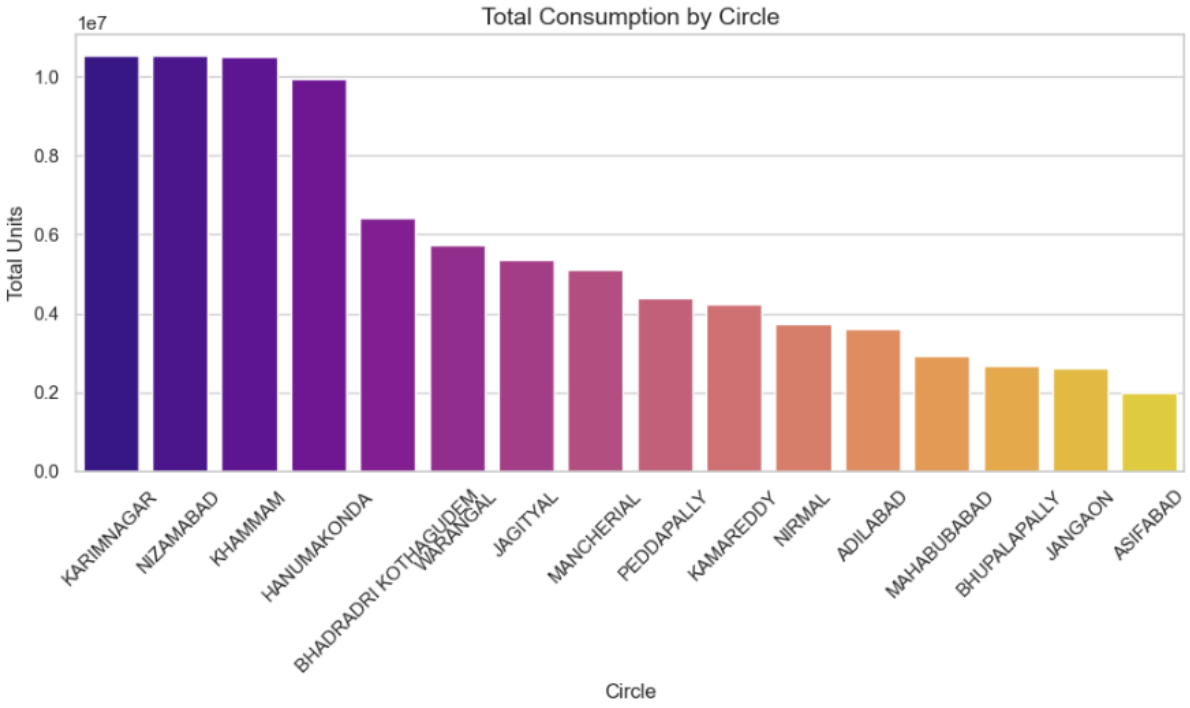
The analysis followed a straightforward, multi-stage process:

- **Data Loading and Verification:** The dataset was loaded and checked for completeness, confirming all 8,421 records were processed correctly.
- **Regional Data Aggregation:** The data was then grouped by administrative 'Circle' to calculate summary figures, such as total energy consumed and total electrical load, to identify key trends.
- **Correlation Analysis:** Finally, a statistical analysis was performed to measure the relationships between numerical data points, confirming the dataset's internal consistency.

3. Results and Discussion

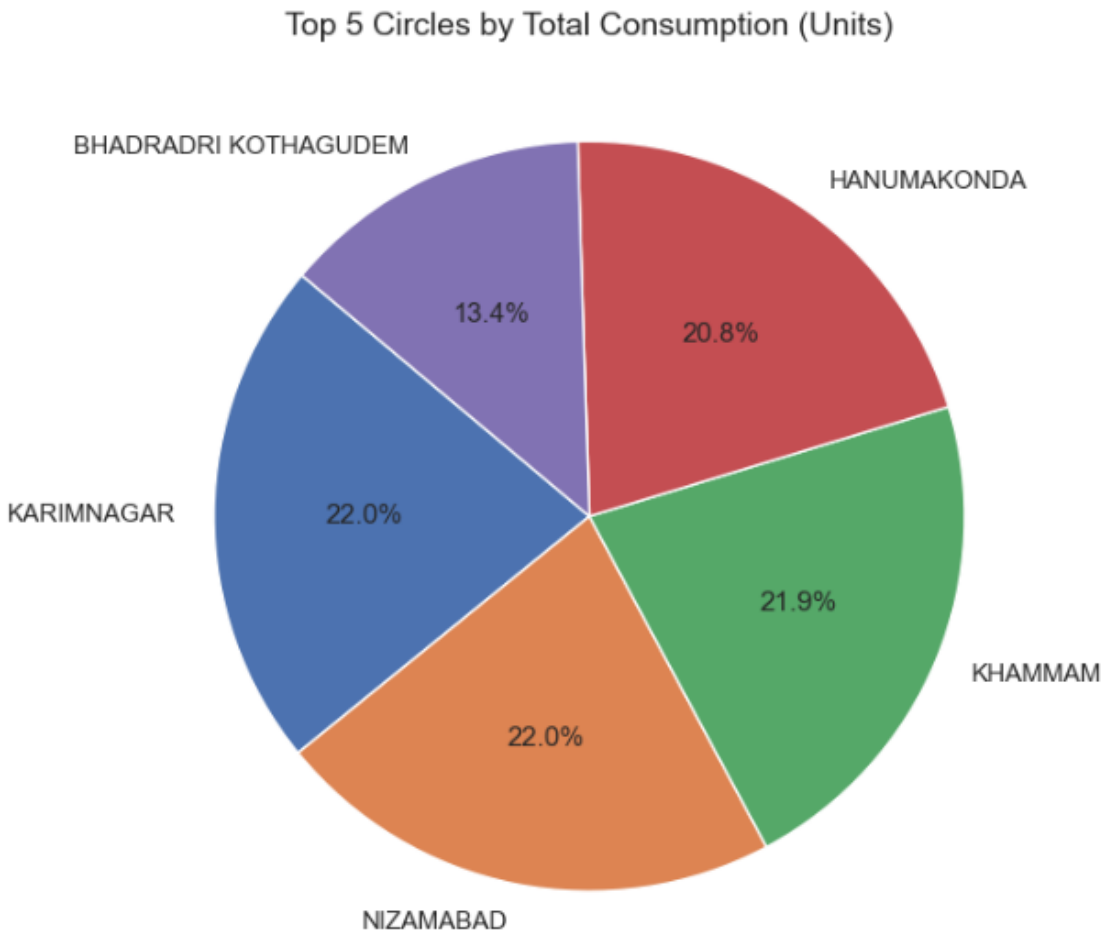
3.1. Overall Consumption Profile and Regional Analysis

The total commercial energy consumption for September 2025 was approximately 234 million kWh, with a total connected load of about 4,892 MW. A primary objective was to identify regional disparities. The bar chart below provides a comprehensive overview of the total units consumed by each of the 16 administrative circles, visually confirming that a few circles account for a disproportionately large share of the demand.



As clearly illustrated, Karimnagar, Nizamabad, and Khammam are the top three consumers, each exceeding 10 million kWh. This highlights the concentration of commercial activity in these key areas.

To better understand the magnitude of this concentration, the following pie chart visualizes the consumption share of the top five circles, which collectively contribute over 60% of the total commercial load.



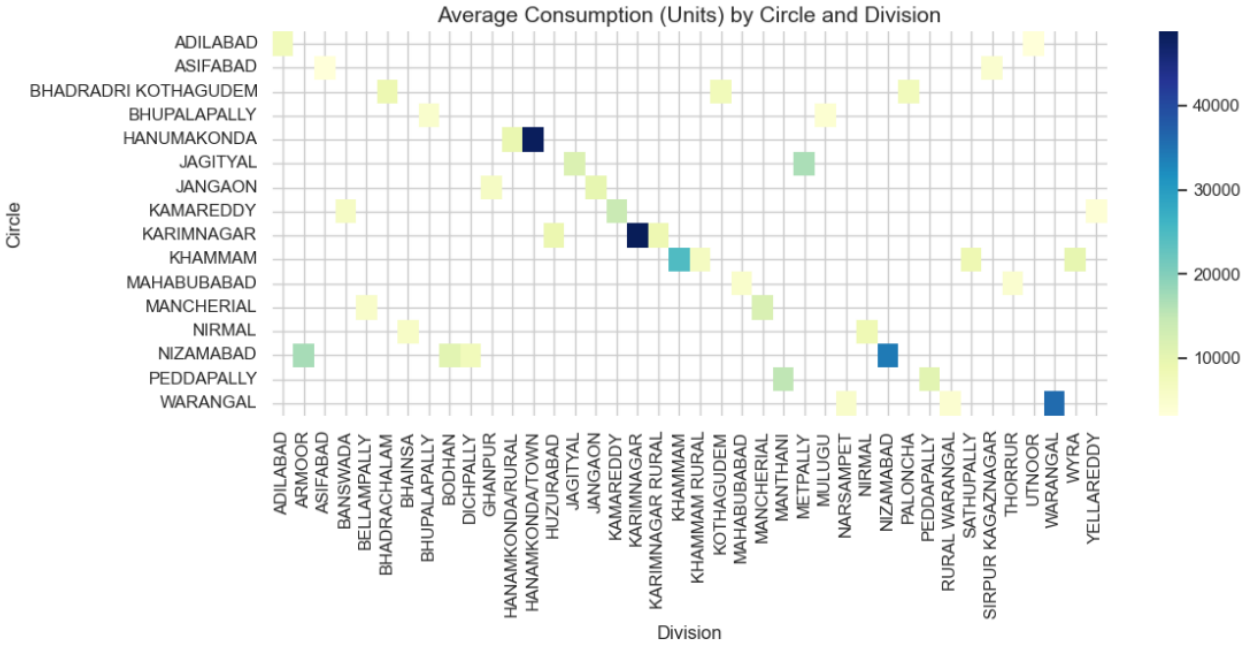
The chart shows that Khammam (21.9%), Nizamabad (22.0%), and Karimnagar (22.0%) have nearly equal, significant shares, followed closely by Hanumakonda (20.8%) and Bhadradri Kothagudem (13.4%). This underscores their critical importance to the stability and planning of the distribution network.

3.2. Granular Consumption Patterns

To gain deeper insights, consumption was analyzed at both the Circle and Division levels. The treemap below visualizes the total units consumed, with the size of each rectangle representing the consumption magnitude. This allows for an at-a-glance comparison of consumption across different administrative hierarchies.

This visualization reveals that within a high-consumption circle like Hanumakonda, the Hanamkonda/Town division is the primary driver of demand (7,385,765 kWh). Similarly, in Karimnagar, the main Karimnagar division accounts for the majority of its usage (7,365,773 kWh). This level of granularity is crucial for targeted infrastructure planning.

Finally, the heatmap below illustrates the average consumption per service within each Division, providing a measure of consumption intensity.



The darker squares indicate divisions with higher average consumption, pointing to the presence of larger commercial establishments rather than just a high number of connections. For instance, divisions within Kamareddy and Hanamkonda/Town show high average consumption, suggesting they are hotspots of intense commercial activity that may require specialized grid management strategies.

---

4. Recommendations

Based on the analytical findings, the following recommendations are proposed:

- Focus investments on grid strengthening in high-load circles such as Warangal and Khammam.
- Improve billing efficiency through smart metering and digital payment adoption.
- Promote energy audits in urban commercial zones to identify and reduce losses.
- Encourage renewable energy integration, such as rooftop solar for commercial consumers, to reduce grid dependency.

---

4. Conclusion

The analysis of TG-NPDCL's commercial dataset for September 2025 reveals strong urban demand, healthy operational ratios, and high-quality data. The visualizations clearly pinpoint the key circles and divisions driving consumption. This study provides a solid, data-driven foundation for predictive analytics, policy formulation, and infrastructure optimization in Telangana's energy distribution network.