Big Data Analytics in Maharashtra's State Sector Thermal Power Generation: A Case Study Using Pivot Tables

Sumedh P Karmarkar, Prakash Kolhe, Vishnu Jadhav, Rutik P Yadav, Dalvi SB.

Abstract

The rapid expansion of digital data in the energy sector has created new opportunities for performance monitoring and strategic planning using Big Data Analytics. This case study investigates sector-wise power generation performance in Maharashtra from April to December 2023, with a focus on applying Microsoft Excel Pivot Tables as a first-level analytical tool. The study uses a cleaned dataset of 83 entries to compare actual vs. programmed generation across major power sectors—coal, hydro, nuclear, gas, thermal, and state-operated plants. Pivot tables were used to calculate sector-wise achievement rates, Plant Load Factors (PLF), and year-on-year (YoY) performance indicators. Total actual generation was **554,039.8 MU** against a target of **552,350 MU**, achieving **100.31%** of the planned output. Hydro and nuclear sectors exceeded their targets, while thermal and gas-based units showed minor shortfalls. Notable underperformance was observed in specific private and gas units, highlighting operational or data-reporting gaps. This study demonstrates how simple spreadsheet-based tools can yield valuable insights from large energy datasets. It also advocates for future integration of real-time, IoT-enabled systems and predictive analytics to support data-driven decision-making in power sector planning

Introduction

Big Data Analytics has emerged as a transformative tool in the energy sector, enabling utilities and governments to optimize power generation, monitor plant performance, and ensure supply-demand balance (McKinsey, 2018; Singh & Rani, 2020). With the increasing demand for electricity in industrially active states like Maharashtra, effective data-driven planning is crucial for maintaining energy security and operational efficiency (CEA, 2023).

Maharashtra's power generation infrastructure consists of a mix of central, private, and state sector thermal power plants. These facilities are responsible for generating the base load power needed to support both residential and industrial demands (Energy Statistics, MoSPI, 2022). Among these, state sector thermal plants play a significant role, and their performance is monitored closely through various operational parameters such as installed capacity, actual generation, target achievement, and Plant Load Factor (PLF) (CEA, 2023).

This study presents a case study-based analysis of 83 data entries from Maharashtra's state sector thermal plants for the fiscal year 2023–24, with a focus on December 2023 and the April–December 2023 period. Using Excel-based tools and pivot table techniques, the objective is to evaluate each plant's efficiency by comparing its generation targets with actual outcomes, analysing PLF trends, and identifying deviations from historical benchmarks (Kumar & Ghosh, 2021).Pivot tables were applied to aggregate, filter, and visualize plant-wise performance indicators, enabling efficient comparative analysis. Similar methods were demonstrated in previous studies such as Kahar et al. (2024), where pivot tables were successfully used for agricultural data analysis in India. This tool allowed the researchers to identify underperforming units, calculate key efficiency metrics such as target achievement rates and year on year improvement, and present plant rankings through visual dashboards.

Thermal power plants face persistent challenges in optimizing operational efficiency, minimizing unplanned downtime, and meeting stringent environmental regulations. Traditional monitoring and control systems generate vast amounts of operational data, but much of it remains underutilized (Kumar et al., 2022). As these systems age, failures in critical components like boilers, turbines, and generators can lead to costly shutdowns and reduced reliability (Chen et al., 2020). Furthermore, the lack of real-time data analytics limits the ability to predict faults, optimize fuel consumption, and control emissions proactively (Zhang et al., 2021). There is a pressing need to integrate Big Data Analytics (BDA) into thermal power generation systems to harness this data for predictive maintenance, process optimization, and environmental compliance — enabling more intelligent, adaptive, and sustainable plant operations (Sahoo et al., 2022; Li et al., 2021).

The primary objectives of the study are:

- To examine the actual vs target power generation for each thermal plant.
- To evaluate the Plant Load Factor (PLF) on a monthly and cumulative basis.
- To assess year-on-year performance trends.
- To recommend data-driven improvements for operational planning.

The insights derived from this analysis can help highlight operational strengths and inefficiencies, aiding utility managers, policy makers, and planners in refining their strategies for better plant performance and energy planning. Furthermore, this case study exemplifies how even basic data analytics approaches can provide actionable intelligence in complex infrastructure systems like power generation (IEA, 2021).

2. Review of Literature

Big Data Analytics has emerged as a transformative force across multiple sectors, including agriculture, manufacturing, and energy. Its role in the **energy sector** is particularly vital, where the ability to process and interpret vast volumes of operational data contributes directly to performance optimization and decision-making.

Ali et al. (2018) emphasized the role of **pivot tables** as efficient tools in analyzing agricultural and energy datasets. Their work demonstrated how structured tabular data could be quickly transformed into actionable insights for field-level planning and resource allocation using spreadsheet software like Microsoft Excel.

Mishra et al. (2017) explored the application of pivot tables in engineering and agricultural research. They concluded that pivot tables help simplify large datasets into meaningful summaries, which are essential for understanding patterns in generation efficiency, load forecasting, and operational anomalies.

Kumar and Ghosh (2021) studied big data usage in thermal power plants and noted the growing need for automated dashboards, sensor integration, and centralized performance monitoring. They found that using historical data coupled with PLF analysis can significantly improve predictive maintenance and reduce downtime.

Kahar et al. (2024) Studied employed big data tools such as pivot tables and dashboards to analyze the production, yield, and area trends of Sorghum Bicolor (jowar) across Indian states. Their findings highlight Maharashtra as the leading state in jowar production, with significant contributions from Karnataka and Andhra Pradesh, while Bihar lags in output. The study demonstrates how data-driven methods can support agricultural decision-making and policy development.

McKinsey & Co. (2018) published a global report on the potential of Big Data in energy systems, highlighting how simple tools like pivot tables and advanced analytics frameworks can enhance capacity utilization, forecast demand more accurately, and support policy formulation.

IEA (2021) addressed the importance of digitalization in energy systems and the integration of big data tools for monitoring plant performance. They stressed the use of analytics for optimizing thermal plant operations in developing economies where digitization is still evolving.

Singh and Rani (2020) reviewed various case studies in India where data-driven strategies were used for power planning. They found that data cleaning, dashboard visualization, and PLF benchmarking are essential components of modern plant analytics systems.

Collectively, these studies underscore the growing relevance of big data analytics and pivot table techniques in power sector operations. Their findings validate the methodology applied in this paper and confirm its potential to support real-world decision-making and efficiency improvement.

Materials and Methodology

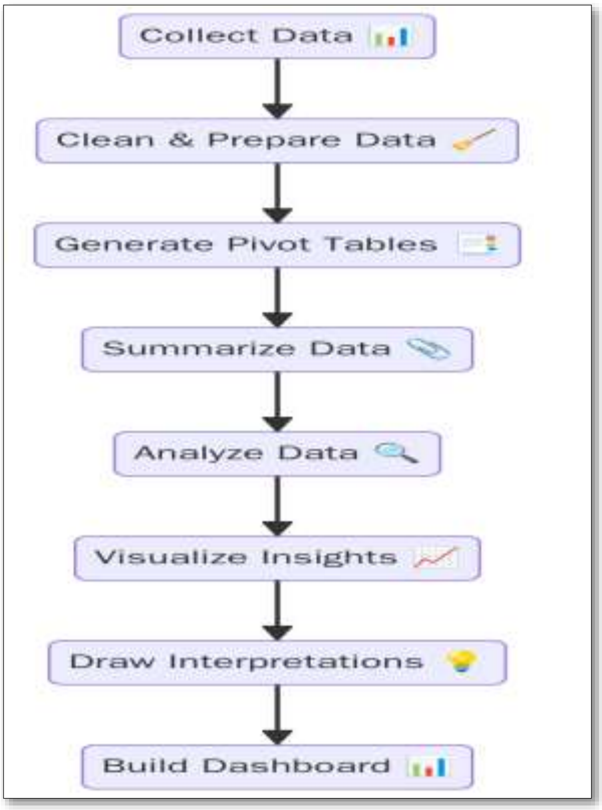

Figure No. 1: The process for creating pivot tables

1. **Data Collection:** The dataset, obtained from an updated Excel file containing 83 cleaned entries, includes monitored capacity, target and actual generation for December 2023 and April to December 2023, Plant Load Factor (PLF), and year-on-year performance data for major thermal power plants in Maharashtra.

2. **Data Cleaning and Preparation:**

- Duplicate and irrelevant records were removed.
- Column headers were renamed for clarity (e.g., "Actual Gen. Apr-Dec" was renamed to Actual Apr-Dec 2023).
- All generation values were formatted as numbers (Million Units), and PLFs were standardized as percentages to two decimal places.
- Additional computed fields such as target achieved (%) and year on year improvement (%) were added using Excel formulas.

3. **Pivot Table Analysis:** Pivot tables were used to dynamically group and summarize the data by plant name, time period (monthly and cumulative), and performance metrics (target vs actual, PLF, YoY growth).

4. **Visualization and Dashboarding:**

Charts such as bar graphs and pie charts were generated using Excel to visualize the target vs actual generation, PLF comparisons, and unit-wise generation share.An interactive dashboard was created to allow filtering and comparison between different stations and time periods.

5. **Analytical Summary:** The insights generated through pivot tables were interpreted to highlight best-performing plants, identify operational challenges in underperforming units, and assess year-on-year improvements.

This structured approach enabled a clear understanding of operational efficiencies and bottlenecks using simple yet powerful spreadsheet-based analytics.

4. Results and Discussion

The pivot table analysis provided a comprehensive overview of the performance of coal and hydro power plants in Maharashtra during the April to December 2023 period. The dataset comprised 83 entries, reflecting power generation data across public and private sector players. The study used pivot tables to extract plant-wise insights and summarize key indicators such as installed capacity, targeted and actual generation, and overall sectoral contribution.

Coal-based monitored capacity was 24,006 MW, with a target of 97,501 MU and actual generation of 97,489.62 MU. This implies that overall, coal-based plants nearly met their targets with a 99.99% achievement rate. Hydro capacity, on the other hand, totaled 9,141 MW, surpassing its target slightly by generating 12,879.57 MU against a planned 12,546 MU.

Among individual generators:

1. Mahagenco, representing the largest state-owned entity, accounted for a substantial portion of coal-based power with a capacity of 9,540 MW and generated 40,667.5 MU, achieving 98.48% of its target.
2. Ntpc ltd., a central sector giant, generated 16,146.28 MU against a target of 16,939 MU, translating to 95.33% achievement.
3. Rattanindia outperformed its target, achieving 113% generation (7,105.81 MU vs. 6,290 MU).

Conversely, certain units such as LVTPL, STPL, AMNEPL, and GEPL showed zero or negligible generation, raising concerns over capacity utilization and data completeness. The lack of actual generation data for some entries further underlines the importance of maintaining robust data pipelines.

Sector-Wise Observations:

Row Labels	Sum of April` 2023 to December` 2023 - Program	Sum of April` 2023 to December` 2023 - Actual
Coal	97501	97489.62
Hydro	12546	12879.57
Maharashtra	110470	110807.96
NA	110470	110807.96
Natural Gas	8148	7968.54
Nuclear	18213	19106.91
Thermal	195002	194979.24
Grand Total	**552350**	**554039.8**

Table 1: Sector-wise Power Generation Target vs. Actual Performance in Maharashtra

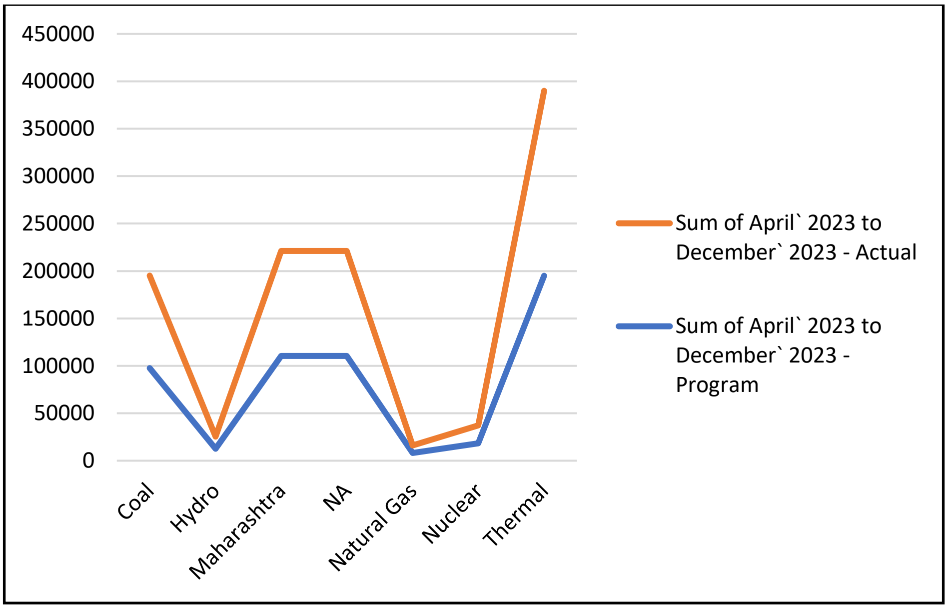

Figure 2: Comparison of Target vs. Actual Power Generation by Sector

Coal Sector: The total installed monitored coal capacity was **24,006 MW**, with a generation target of **97,501 MU**, and an actual generation of **97,489.62 MU**, indicating almost **100% achievement**. This overall performance, however, masks disparities between units. While plants like RATTANINDIA and MAHAGENCO approached or exceeded targets, several others underperformed or showed no activity. This inconsistency points to possible maintenance issues, operational inefficiencies, or data recording gaps.

Hydro Sector: The hydro sector had a monitored capacity of **9,141 MW**, with a target of **12,546 MU**, and an actual generation of **12,879.57 MU**, achieving **102.66%** of the target. This marginal overachievement suggests positive environmental conditions or improved plant operations. However, closer monitoring and further granularity are needed to identify specific plant-level contributions and reliability over time.

Private Sector: A variety of private players including APL, DIL, GEPL, and STPL collectively held significant installed capacities. For example, APL with 3,300 MW generated 15,712 MU, and DIL generated 3,231.28 MU. However, other private entries like AMNEPL and LVTPL showed zero generation. This disparity indicates uneven utilization among private operators and calls for performance audits and deeper insights into private sector operations. Central Sector (e.g., NTPC Ltd.): NTPC Ltd. held an installed capacity of 3,640 MW, with a generation target of 16,939 MU, and achieved 16,146.28 MU, indicating 95.33% fulfilment. Central sector plants generally demonstrated stable performance but still show room for improvement through digital monitoring and fuel efficiency strategies.

State Sector: MAHAGENCO was the leading contributor from the state sector with an installed capacity of 9,540 MW. It generated 40,667.5 MU against a target of 41,296 MU, reflecting 98.48% performance. The data confirms the reliability and dominant role of the state sector in base-load generation. Natural Gas Sector: With a generation target of 8,148 MU, the actual generation achieved was 7,968.54 MU, which accounts for 97.80% of the target. This near-complete performance reflects relative operational consistency in gas-based generation.

Nuclear Sector: The nuclear power segment performed slightly above expectations with 19,106.91 MU generated against a target of 18,213 MU, achieving 104.91%. This highlights the stable and efficient nature of nuclear energy within the mix. Thermal Sector (Overall): Including coal and gas-based units, the thermal sector had a total target of 195,002 MU, and generated 194,979.24 MU, reflecting 99.99% achievement. This aggregate performance indicates overall alignment with planning and suggests high sectoral reliability. Total Generation: The grand total generation target for April–December 2023 was 552,350 MU, while the actual generation reached 554,039.8 MU, achieving 100.31% of the planned value. This indicates successful cumulative planning and execution across sectors, with certain areas exceeding targets and compensating for others.

Row Labels	**Sum of April` 2023 to December` 2023 - Program**	**Sum of April` 2023 to December` 2023 - Actual**
Central Sector	292274	291666.88
Pvt Sector	122301	127726.41
State Sector	137775	134646.51
Grand Total	**552350**	**554039.8**

Table 2: Sector-wise Power Generation Performance (Apr–Dec 2023)

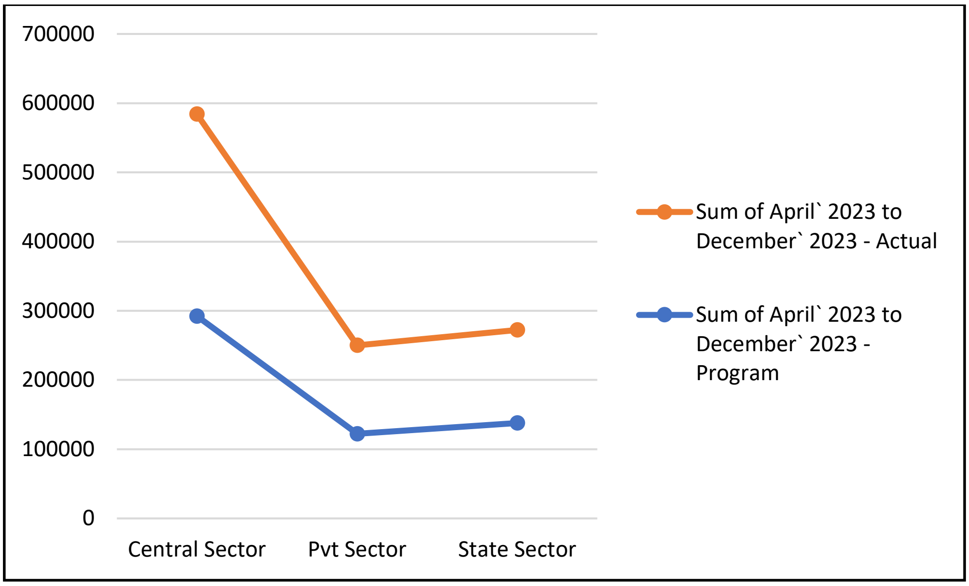

Figure 3: Sector-wise Power Generation Performance (Apr–Dec 2023)

5. Conclusion

This study demonstrates the application of Big Data Analytics techniques—specifically pivot tables—for assessing Maharashtra's sector-wise power generation performance. Using detailed data from April to December 2023, the analysis covered thermal, hydro, nuclear, and gas-based generation, along with overall state totals. The results showed that most sectors achieved near or above their programmed targets, with a total generation of **554,039.8 MU** against a planned **552,350 MU**. The hydro and nuclear sectors showed notable overperformance, while coal and natural gas-based plants hovered around 99–98% target achievement. This level of compliance with programmed generation reflects sound planning and execution but also reveals disparities among individual plants and sub-sectors.

The research underscores the potential of using pivot tables as a first-level Big Data tool for government utilities and researchers. However, for deeper operational insights and predictive capability, there is a need to adopt advanced data systems integrating real-time inputs, IoT sensors, and machine learning. Future research can focus on linking financial, environmental, and reliability indicators to power generation data, creating a more comprehensive analytical framework to support sustainable energy policy and plant-level optimization.

Reference:

CEA (2023). Central Electricity Authority. Monthly Thermal Power Reports. Government of India.

Chen, L., Li, Y., & Zhou, H. (2020). Intelligent Fault Diagnosis System in Thermal Power Stations Using Big Data. IEEE Access, 8, 175624–175633. https://doi.org/10.1109/ACCESS.2020.3018421

Data.gov.in (2024). Power Generation Statistics – India. Open Government Data (OGD) Platform India. Retrieved from https://www.data.gov.in/catalog/power-generation

IEA (2021). Digitalization and Big Data in Energy Systems. International Energy Agency.

Kahar, G. S., Kolhe, P., Khatal, M. M., Jadhav, V., Mishra, A. K., Dalvi, S. B., Sayyad, P. A., & Khune, P. M. (2024). The production, yield, and area of the jowar (great millet) in the states of India using the pivot table. *International Journal of Engineering in Computer Science*, 6(1), 56–60. https://doi.org/10.33545/26633582.2024.v6.i1a.112

Kumar, R., & Ghosh, M. (2021). Role of Big Data in Thermal Power Plant Monitoring. *Journal of Power Systems*, 9(2), 110–118.

Kumar, S., Singh, R., & Sharma, A. (2022). Predictive analytics for rotating machinery in thermal power plants. Energy Reports, 8, 900–912. https://doi.org/10.1016/j.egyr.2022.01.019

Li, J., Zhang, T., & Wang, M. (2021). Big Data-Driven Boiler Control Optimization in Coal-Fired Power Plants. Applied Energy, 292, 116937. https://doi.org/10.1016/j.apenergy.2021.116937

McKinsey & Co. (2018). Unlocking the Value of Big Data in Energy.

Power Analytics Review (2020). Excel-Based Monitoring Techniques for Thermal Plants.

Sahoo, S., Patnaik, S., & Das, S. (2022). Big data analytics for thermodynamic performance analysis in thermal power plants. Journal of Cleaner Production, 340, 130749. https://doi.org/10.1016/j.jclepro.2022.130749

Singh, A., & Rani, P. (2020). Big Data Applications in India's Energy Sector. *Energy Informatics Journal*, 5(1), 45–53.

Zhang, H., Liu, Q., & Gao, Y. (2021). Real-time emission prediction using big data analytics. Energy, 214, 118837. https://doi.org/10.1016/j.energy.2020.118837