Operational Efficiency & Power Generation Performance Analysis of NTPC Stations

Sector: Energy & Power Utilities
Team details: Section-D/G-17

- Mansi Agarwal
- Aditya Samadhiya
- Kapish Rohilla
- Deepanshu
- Prashant Pandey
- Harshit Kudhial

Institute: Newton School of Technology

Faculty: Archit

2. Executive Summary

Problem

Thermal power generation is hindered by operational inefficiencies. The fleet operates at an average **Daily Capacity Factor (DCF) of 65.9%**, significantly below the potential demonstrated by top performers. A "Low Capacity" segment trails high performers by **11.8%**, and **607,137.78 MW** of cumulative capacity has been lost to outages.

Approach

We utilized a dataset of 7,500 records spanning 2017–2026. The raw data was cleaned and transformed in Google Sheets to build a 7-point KPI framework. The analysis focuses on the gap between **Total Monitored Capacity (3,033,314 MW)** and **Total Available Capacity (2,426,176 MW)** to quantify losses.

Key Insights

- **Generation Volume:** Total Power Generation over the period stands at **49,620.6 MU**.
- **Reliability Gap:** There is a massive gap between Monitored and Available capacity, driven by the **607k MW** outage volume.
- **Asset Reliance:** "High Capacity" units are efficient but only contribute **26.97%** of the total generation share.
- **Critical Failures: 70** specific "Zero Generation" events were detected, flagging units with chronic starting issues.

Key Recommendations

Launch a "Performance Turnaround" for the bottom 25% of units, deploy a technical task force to UP (the highest outage contributor), and align maintenance strictly with the August low-demand window.

3. Sector & Business Context

Sector Overview

The Indian power sector is transitioning towards higher efficiency to meet growing energy demands. Thermal power plants (NTPC) remain the backbone of this supply. Operational efficiency (PLF/PAF) is critical for profitability and grid stability.

Current Challenges

- Aging infrastructure leading to frequent forced outages.
- Coal availability issues affecting Plant Load Factors (PLF).
- Variability in demand requiring flexible operation, which stresses equipment.

Why this problem was chosen

Maximizing the utilization of existing assets is more cost-effective than building new capacity. Identifying specific underperforming units and outage trends offers immediate opportunities for operational improvement and cost saving.

4. Problem Statement & Objectives

Formal Problem Definition

"To analyze historical daily power generation data to identify factors contributing to suboptimal Capacity Utilization Factors (CUF) and quantify the impact of outages across NTPC stations."

Project Scope

Analysis of daily generation reports, outage data, and coal stock levels for monitored NTPC stations (2017-2026).

Success Criteria

- Accurate tracking of the 7 core KPIs
- Visual identification of state-wise bottlenecks
- Quantification of revenue loss.

5. Data Description

Exact Dataset Source

- Source: Government of India / NTPC Daily Generation Reports (https://ndap.niti.gov.in/dataset/7686)
- Access: Data provided via RawDataset.csv.

Data Structure

The dataset is a time-series record of daily operations for individual power station units.

Columns Explanation (Key Columns)

- **Name of the power station:** Name of the specific unit.
- **Monitored capacity:** Installed capacity of the unit (MW).
- **Available capacity:** Capacity actually available for generation (MW).
- **Power generation under todays actual:** Actual energy generated on the date (MU - Million Units).
- **Capacity under outage:** Capacity unavailable due to forced or planned maintenance (MW).
- **Coal Stock:** Days of coal stock available (critical for thermal plants).

Data Size

- 7,500 rows and 16 columns.

Data Limitations

- Inconsistent date formats in the raw source.
- Significant missing values in Coal Stock and Maintenance Start Date.
- Partial data for the year 2026.

6. Data Cleaning & Preparation

All primary cleaning and transformation steps were executed in Google Sheets prior to analysis, as evidenced by the transition from RawDataset.csv to Cleaned.csv.

Missing Values Handling

- Coal Stock: 2,042 missing values in raw data were imputed with 0 or treated as "Not Available" to prevent calculation errors.
- Capacity under outage: Null values (7 records) were filled with 0, assuming no outage if not reported.

Outlier Treatment

- Calculated Daily_Capacity_Factor. Values exceeding 100% or effectively 0% were flagged using the Zero_Gen_Flag for further inspection.

Transformations

- Date Standardization: Converted various text formats into a standard Date format.
- Text Cleaning: "Name of the power station" was cleaned to "Station_Clean" by removing underscores and standardizing spacing.

Feature Engineering

- Daily_Capacity_Factor: Calculated to normalize performance comparisons.
- Capacity_Category: Created to segment stations into "High" vs "Low" buckets based on performance thresholds.
- Zero_Gen_Flag: A binary/text flag created to instantly identify days with 0 generation.

Assumptions

- Blank Outage implies the station was fully operational.
- Monitored Capacity is constant for the day.

7. KPI & Metric Framework

KPI Name	Value	Why it matters

1. Total Power Generation	**49,620.6 MU**	The primary output metric; total energy delivered to the grid.
2. Avg Daily Capacity Factor	**65.9%**	The efficiency score. Measures actual generation vs. potential.
3. High Capacity Gen Share	**26.97%**	Measures reliance on top-performing assets vs. the rest of the fleet.
4. Total Outage	**607,137.78 MW**	The volume of lost capacity due to technical or maintenance issues.

8. Exploratory Data Analysis (EDA)

Trend Analysis

- Yearly Growth: Power generation shows a steady upward trend from 2017 to 2022, peaking at ~6,924 MU (in this sample dataset), followed by a slight stabilization.
- Seasonality: March and April are the peak performance months (Avg Gen > 7.3 MU), likely due to pre-monsoon demand. August sees the lowest average generation (6.46 MU), possibly due to monsoon maintenance or lower demand.

Comparison Analysis

- State-wise: Uttar Pradesh performs poorly on reliability metrics, contributing the highest volume of outages.
- Station-wise: Talcher (Old) TPS Unit 2 is the top performer with ~99.4% average DCF. In contrast, stations like Dadri (NCTPP) show high variability.

Distribution Analysis

- The Daily_Capacity_Factor is bimodal. Most stations operate either at high efficiency (>80%) or very low efficiency/shutdown (<20%), validating the need for the "High/Low" capacity segmentation.

Correlation

- Strong negative correlation observed between Outage (MW) and Daily_Capacity_Factor. High outage days directly cripple efficiency.

9. Advanced Analysis

Segmentation (Capacity Category Analysis)

- We segmented stations into "High Capacity" and "Low Capacity".
- Insight: The "High Capacity" segment operates at 79.3% DCF with minimal outages. The "Low Capacity" segment operates at 67.5% DCF but accounts for the vast majority of the recorded outage MW (~98% of outage volume comes from this segment).
- Conclusion: The problem is not systemic across all NTPC stations but concentrated in specific "problem assets."

Anomaly Detection (Zero Gen Analysis)

- 70 events were flagged where generation was exactly 0.
- These are not random; they cluster in specific stations (e.g., Faridabad CCPP, Dadri CCPP). This suggests specific units have chronic starting/technical issues.

10. Dashboard Design

Dashboard Implemented

A comprehensive dashboard was designed in Google Sheets utilizing the processed data.

Dashboard Top Line KPI Strip

Displays the 7 key metrics:

- High Cap Share (26.97%)
- Total Gen (49,620.6)
- Avg DCF (65.9%)
- Outage (607k)
- Available Cap (2.4M)
- Monitored Cap (3.0M)
- Zero Gen (70)

Visualizations

- **Year-wise Trend Analysis:** Line chart showing generation growth over time (Pivot 3).
- **Month-wise Seasonality:** Column chart highlighting the August dip (Pivot 6).
- **State-wise Performance:** Stacked Bar Chart comparing Monitored vs. Available Capacity per state (Pivot 2).
- **Capacity Category Analysis:** Donut/Pie chart showing the 26.97% High Capacity share (Pivot 4).
- **Zero Generation & Outage Analysis:** Chart tracking the breakdown of the 70 critical failure events (Pivot 5).
- **Station Performance Summary:** Bar chart/Table highlighting top and bottom performing stations (Pivot 1).

Interactive Elements

- Year
- Month
- State
- Station Name
- Capacity Category

11. Insights Summary

- **Efficiency Gap:** A significant **11.8% efficiency gap** exists between capacity categories. "High Capacity" units operate at an average Daily Capacity Factor (DCF) of **79.3%**, while "Low Capacity" units trail significantly at **67.5%**.
- **Massive Capacity Loss:** The fleet lost a staggering **607,137.78 MW** (cumulative daily) to outages over the analysis period, driving a wedge between the Total Monitored Capacity (3,033,314 MW) and Actual Available Capacity.
- **Regional Disparity: Uttar Pradesh** is the primary bottleneck, accounting for **~23%** of the total outage volume among the top 5 worst-performing states. The state's "Stacked" performance bar shows a disproportionately high ratio of Outages to Generation.
- **Operational Benchmark: Talcher (Old)** consistently sets the fleet standard, operating near **100% DCF**, proving that high efficiency is achievable even with older assets.
- **Asset Reliance Risk:** Despite their superior performance, "High Capacity" units only contribute **26.97%** of the total generation share, revealing an over-reliance on lower-efficiency units for the bulk of power production.
- **Seasonal Vulnerability:** Generation drops by **~12% in August** compared to the peak in March. This seasonal dip represents a predictable low-demand window that is currently underutilized for planned maintenance.

- **Chronic Failures: 70 specific instances** of "Zero Generation" were detected. These are not random but cluster in specific units (e.g., Faridabad, Dadri), indicating chronic technical failures rather than fleet-wide fuel shortages.
- **Coal Availability:** While coal stock data had gaps (27% missing), existing records show a clear correlation where periods of "0" stock align with sharp drops in generation, validating fuel logistics as a secondary performance driver.
- **Maintenance Misalignment:** Maintenance schedules do not consistently align with the August low-demand season, leading to avoidable generation losses during higher-demand months.

12. Recommendations

Insight	Recommendation	Impact & Feasibility
Low Capacity Underperformance	**Recommendation 1: Performance Turnaround Program** Initiate a targeted operational review for the bottom 25% of stations (specifically the "Low Capacity" category). Focus on bringing their Daily Capacity Factor (DCF) from the current average of 67.5% up to a target of 75%.	**High Impact:** Raising the baseline efficiency of these units would add massive generation volume without new infrastructure. **Medium Feasibility:** Requires process changes and potential retrofitting.
High Outages in UP	**Recommendation 2: Technical Task Force for UP** Deploy a specialized technical audit team to Uttar Pradesh stations (the highest outage contributor) to diagnose specific equipment failure	**High Impact:** Reducing UP's disproportionate outage volume by even 20% would significantly cut fleet-wide losses. **High Feasibility:** Can be executed immediately

	patterns (e.g., Boiler tube leaks vs. Turbine vibrations).	using existing engineering resources.
70 Zero Gen Events	**Recommendation 3: Automated "Zero-Gen" Protocol** Implement an automated "Zero-Gen Alert" system that flags any unit reporting 0 generation and requires a mandatory Root Cause Analysis (RCA) report within 24 hours of occurrence.	**Medium Impact:** Prevents repeated failures by enforcing accountability and rapid diagnosis. **High Feasibility:** purely a process and software implementation.
Seasonal Dip in August	**Recommendation 4: Strategic Maintenance Scheduling** Align planned annual maintenance (overhauls) strictly with the **August** low-generation window (observed ~12% dip) to ensure maximum availability during peak demand months (March/April).	**Medium Impact:** Optimizes availability when the grid needs it most. **High Feasibility:** Requires only planning and scheduling adjustments.

13. Impact Estimation

By implementing the targeted recommendations—specifically the "Performance Turnaround" for the Low Capacity segment and the technical audit in Uttar Pradesh—the potential value delivered is estimated as follows:

- **Generation Recovery**

    - **Current State:** The analysis recorded a cumulative daily outage volume of **~607,137 MW**.
    - **Target:** A conservative reduction of **10%** in outage volume through improved maintenance planning and RCA.
    - **Recovered Capacity:** 607,137 MW * 10% = **60,714 MW** (cumulative daily).
- **Energy Gain**
    - Converting this recovered capacity into actual energy units (Million Units - MU): **Energy** = 60,714 MW * 24 hours * 0.001 = **1,457 MU**
- **Financial Impact**
    - Assuming an average revenue realization (tariff) of **Rs 4.00 per unit**: **Revenue Recovered** = 1,457 MU * Rs 0.4 Crores/MU = **Rs 582.8 Crores**
- **Efficiency Lift**
    - Recovering this lost generation would improve the overall fleet Daily Capacity Factor (DCF) by approximately **1.5%**, shifting the fleet average from 65.9% to **~67.4%.**

14. Limitations

While the analysis provides robust insights, the following limitations in the source dataset and methodology should be noted:

- **Data Completeness (Fuel Correlation):** The **Coal Stock** column contained approximately **27% missing data**. This significant gap limits the ability to definitively correlate every instance of low generation with fuel shortages, although available data does show a strong positive correlation.
- **Assumption Risks (Outage Reporting):** In the absence of explicit "0" values, null (blank) entries in the **Outage** column were treated as "Zero Outage" (fully operational). If these blanks actually represented unreported data rather than zero outages, the total outage volume and associated financial loss may be under-reported.
- **Causality Constraints:** The dataset is descriptive rather than diagnostic. It effectively tracks **that** an outage occurred and **when**, but it lacks specific "Reason Codes" (e.g., Boiler Tube Leakage vs. Generator Protection Trip). Therefore, while we can quantify the loss, the specific mechanical root cause requires a secondary technical dataset.

15. Future Scope

To further enhance operational efficiency, we propose three advanced analytics initiatives for the next phase:

- **Predictive Maintenance (PbM):** Integrate IoT sensor data (vibration, temperature) with Machine Learning to predict equipment failures before they cause "Zero Generation" events.

- **Coal Logistics Optimization:** Incorporate external railway and supply chain data to model "Days of Stock" accurately, overcoming the current 27% data gap.
- **Financial Dashboarding:** Add real-time "Cost of Generation" and revenue data to the dashboard, transforming operational metrics into financial actionable insights.

16. Conclusion

This project successfully transformed raw operational data into actionable business intelligence. We identified that performance issues are not systemic across the fleet but are concentrated in a specific subset of "Low Capacity" stations and regional inefficiencies in Uttar Pradesh. By addressing the **11.8% efficiency gap** and the critical **607,137 MW outage volume**, NTPC stands to recover potentially **₹582+ Crores** in lost revenue opportunity while enhancing national grid reliability.

17. Appendix

- Data Dictionary: (Included in Section 5)
- Logic Used:
  Daily_Capacity_Factor = (Generation / (Monitored * 0.024)) * 100
  Outage = Cleaned version of Capacity under outage
- Files Reference:
  DVA Capstone - RawDataset.csv
  DVA Capstone - Cleaned.csv
  DVA Capstone - Calculations_Pivots.csv
  DVA Capstone - Dashboard.csv

18. Contribution Matrix

Team Member	Dataset & Sourcing	Cleaning	KPI & Analysis	Dashboard	Report Writing	PPT
Mansi Agarwal	✔️	✔️	✔️	—	—	—

Aditya Samadhiya	—	—	✔	✔	✔	✔
Deepanshu	✔	—	—	✔	✔	—
Kapish Rohilla	—	✔	—	—	✔	✔
Harshit Kudhial	✔	✔	—	—	—	—
Prashant Pandey	✔	✔	—	—	—	—