AN EXPLORATORY DATA ANALYSIS OF GRID STRAIN IN ERCOT: DEMAND, GENERATION MIX, TEMPERATURE, AND EMISSIONS

Independent Learning Project

MS ESTP, Carnegie Mellon University

Timeline: **Summer 2025**
Location: **Pittsburgh, PA**

Prepared and Authored by: **Bareethul Kader**

TABLE OF CONTENTS

EXECUTIVE SUMMARY

In the lead up to summer 2025, Texas was identified as one of the most vulnerable states to power outages due to surging electricity demand and strain on its power grid. According to a study published ahead of the season, Texas was ranked as the second most at risk state for power outages in summer 2025, underscoring the urgency of understanding load dynamics and grid resilience.[1] ERCOT's monthly operational overview reports that the peak system demand for June 2025 reached 77,351 MW on June 19, underscoring the season's high stress conditions[2].In response to this heightened concern, a comprehensive interactive dashboard was developed using Streamlit to analyze ERCOT June 2025 data across over 200 hours of records. The analysis integrates demand, generation mix, ambient temperature, and CO_2 emissions into a unified platform with dynamic filtering and advanced visualizations. Core features include grid strain hours defined as demand above the 90^{th} percentile occurred in over 10% of all observations, with notable clustering between 3-7 PM on weekdays. During these periods, natural gas accounted for around 35-40% of generation, driving up CO_2 intensity to peaks approached 0.52 tons/MWh. Correlation heatmaps revealed strong associations between temperature, demand, and fossil generation share. Temporal trend visualizations, strain specific fuel mix comparisons, and custom temperature bin analysis provided insights into the conditions exacerbating grid vulnerability. This project was independently developed as a foundational learning exercise in data analysis and energy systems visualization.

2. BACKGROUND & MOTIVATION

Ensuring reliable grid operations has emerged as a central policy and technical challenge as the energy system navigates rising demand, decarbonization goals, and climate driven stressors. ERCOT, operating one of the most weather sensitive grids in the United States, has faced repeated concerns over its ability to balance soaring summer load with a rapidly changing generation portfolio. Episodes of scarcity pricing, conservation appeals, and tight reserve margins in recent years have underscored the fragility of the system under heat stress. Against this backdrop, June 2025 provided an ideal case study for analyzing the dynamics of strain: high ambient temperatures coincided with elevated peak demand, while renewable output fluctuated daily. This project was motivated by two parallel objectives: (i) to contribute to the broader conversation on grid resilience by examining real world operational data through the lens of strain conditions, and (ii) to serve as a structured, beginner level learning exercise in applied data science and visualization. By combining exploratory analytics with interactive tools, the project demonstrates how open data, and lightweight modeling can uncover patterns that support both policy dialogue and technical system planning.

3. DATASET OVERVIEW

This project integrates multiple open datasets into a unified hourly framework for June 16-21, 2025, covering over 200 hours of system operations. Together, these sources capture demand, generation, temperature, and emissions needed to evaluate grid strain dynamics.

3.1 ERCOT Operational Data

System wide demand, forecasted versus actual load, and generation by fuel type (natural gas, coal, nuclear, solar, wind, hydro, and storage) were obtained from ERCOT's public data portal. These form the core time series for grid activity.

3.2 NOAA Climate Data

Ambient temperature records were incorporated from NOAA monitoring stations in Texas. Erroneous values such as 999.9 deg C were identified and corrected using linear interpolation, producing a continuous and realistic temperature profile.

3.3 EPA CAMD/CAMPD Emissions Data

Hourly CO_2 emissions factors were sourced from EPA's air markets database. Since June 2025 emissions were not yet available, 2024 hourly profiles were scaled and aligned with 2025 demand and generation, providing a proxy for real time carbon intensity.

3.4 Engineered Features

From these inputs, additional indicators were created:

- CO_2 intensity (tons/MWh), derived from emissions and generation.
- Grid strain flag, marking hours above the 90th percentile of demand.
- Demand quartiles, to enable comparison of generation mixes across load levels.
- Time features such as weekday, hour, and temperature buckets.

Data Availability Note: All datasets used in this project are publicly available from ERCOT, NOAA, and EPA (CAMPD). For reproducibility, access should be obtained directly from the respective provider websites, as raw records are not redistributed in this report.

4. METHODOLOGY

The analysis followed a structured workflow that combined data preprocessing, exploratory analytics, and interactive visualization.

4.1 Data Preprocessing

- **Cleaning:** Removed invalid entries, including temperature placeholders and interpolated missing values for continuous time series.
- **Alignment:** Synchronized demand, generation, temperature, and emissions datasets into a common hourly timestamp.
- **Feature Engineering:** Derived new fields including CO_2 intensity (tons/MWh), demand quartiles, grid strain flags, temperature buckets, forecast error (actual vs forecast demand), and net storage flows. For example, forecast error was explicitly calculated as the difference between actual and forecast demand, while negative storage values were interpreted as discharging events to clarify the role of batteries in evening demand management.

4.2 Exploratory Data Analysis (EDA)

- Conducted hourly and weekday demand profiling to identify daily and intra week patterns.
- Applied statistical distributions (histograms, boxplots) to characterize demand and temperature ranges.
- Built scatterplots with regression fits to assess nonlinear relationships (e.g., temperature vs demand).
- Segmented demand into quartiles to compare generation fuel shares across different load levels.

4.3 Grid Strain Definition

Grid strain was defined quantitatively as hours where total demand exceeded the 90th percentile of the June 2025 dataset. This approach provided a consistent threshold to highlight stress periods without subjective judgment.

4.4 Visualization and Dashboard Development

- Implemented visual analytics in Python using pandas, matplotlib, seaborn, and NumPy for plotting and statistics.
- Deployed results to an interactive Streamlit dashboard with dynamic filters (date, hour range) and toggles for raw data.
- Structured outputs across four analytical modules: Demand, Temperature, Generation, and Emissions.

4.5 Learning Objective

While designed to yield actionable insights on ERCOT's system behavior, the workflow also served as a practical learning exercise in data science. Core methods such as feature engineering, visualization design, and interactive dashboarding were intentionally applied to simulate the workflow of an applied energy analytics project from start to finish.

5. VISUALIZATIONS & FINDINGS

5.1 Demand Patterns

Analysis of June 2025 ERCOT demand revealed clear temporal dynamics and concentration of grid strain in specific hours. Hourly profiles showed demand consistently rising through the late afternoon, with the steepest climbs between 3 PM and 7 PM. The system wide peak was recorded on June 19 at 77,351 MW, coinciding with a sequence of consecutive high load afternoons. Weekday analysis further confirmed clustering, with strain conditions disproportionately occurring on weekdays compared to weekends.

Distribution plots demonstrated a right skewed demand profile, with the top 10% of hours exceedingly approximately 74,000 MWh. These strain hours represented about 10% of the 200+ hourly records (≈22 hours) in the study window, underscoring how a relatively small share of time drives critical grid stress. Histograms showed a large concentration of demand in the 60,000 - 70,000 MWh range, while boxplots captured significant weekday variability, with Fridays showing the highest median demand near 69,000 MWh.

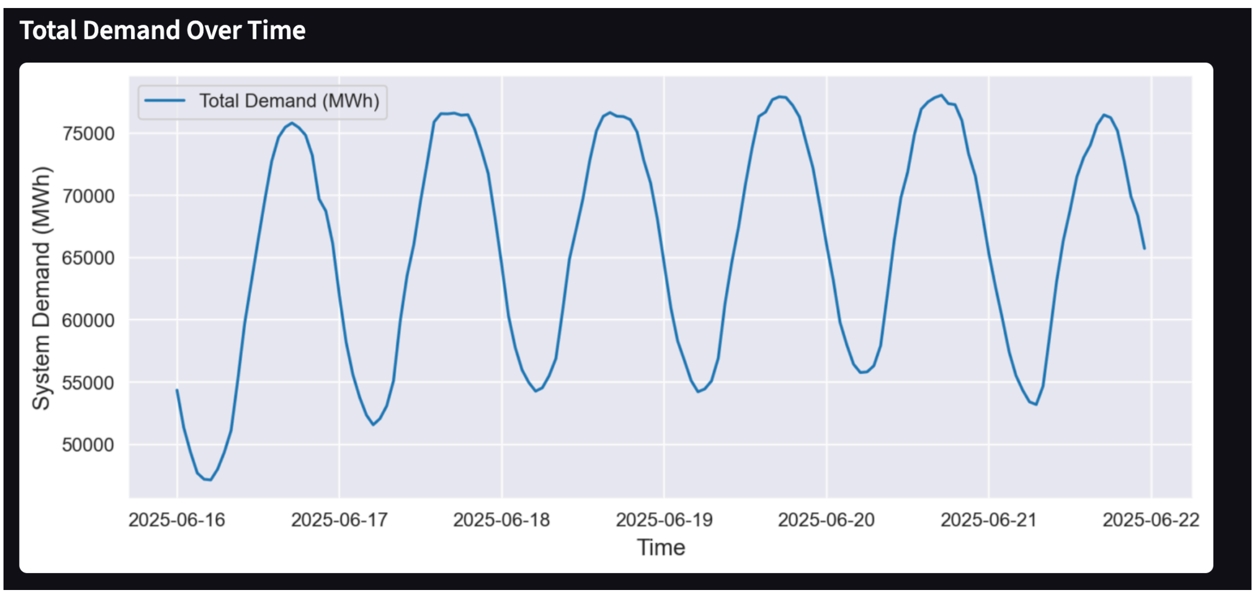

Figure 1: *Hourly total demand across ERCOT in June 2025, with a peak of 77,351 MW on June 19, highlighting late afternoon surges.*

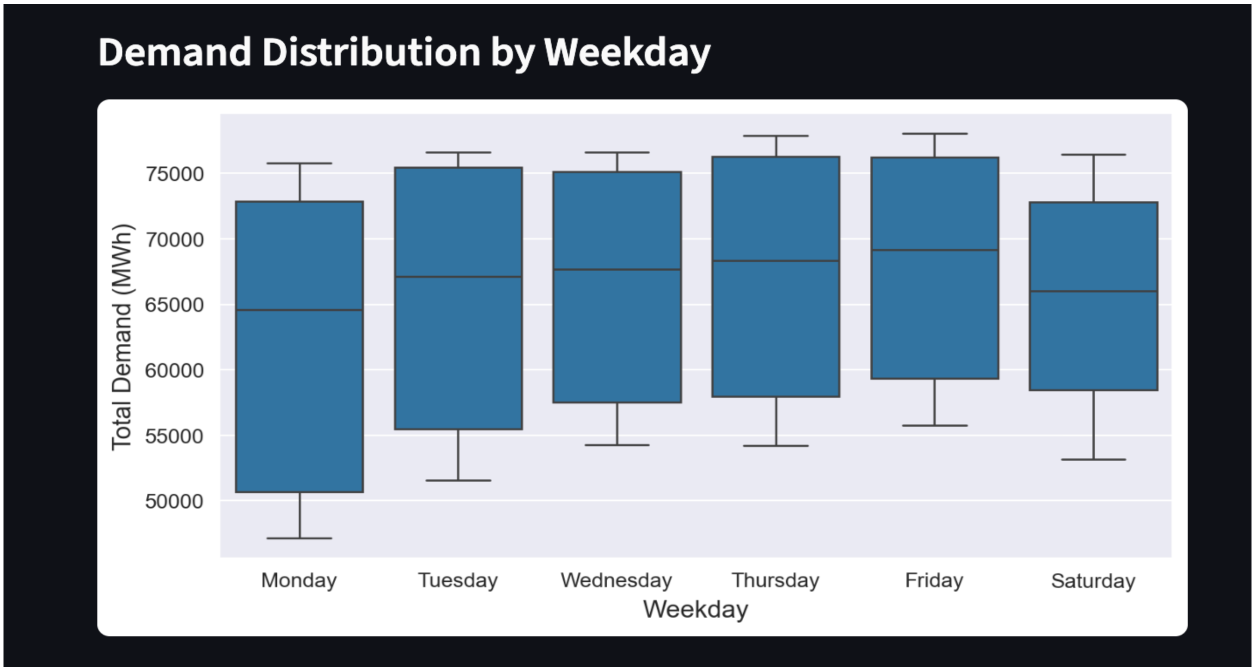

Figure 2: *Boxplot of demand distribution by weekday, showing consistently higher medians on weekdays compared to weekends.*

5.2 Temperature Correlation

Temperature analysis highlighted strong sensitivity of demand to heat. Daily temperatures rose steadily from early morning, peaking around 15:00-17:00 before tapering off at night. Polynomial fits captured this relationship, with demand rising sharply beyond 32 °C, averaging 75,000 MWh under peak heat condition. Temperature bucket analysis reinforced this observation. Average demand was approximately 62,000 MWh at <27 deg C, rising to 72,000 MWh in the 30 - 32 deg C band, and peaking above 75,000 MWh when temperatures exceeded 32 deg C. Boxplots comparing strain vs non strain hours showed median temperatures more than 2 deg C higher during periods of grid stress, confirming heat as a principal driver of elevated load.

5.3 Generation Mix Analysis

Fuel wise generation patterns revealed distinct load following behavior. Natural gas consistently dominated during high load hours, contributing around 40% of total generation during strain periods, compared to 35% during normal conditions. Solar generation decreased after sunset but reached up to 12% during midday peaks, partially offsetting fossil output. Wind remained variable, often peaking overnight but declining during afternoon strain hours.

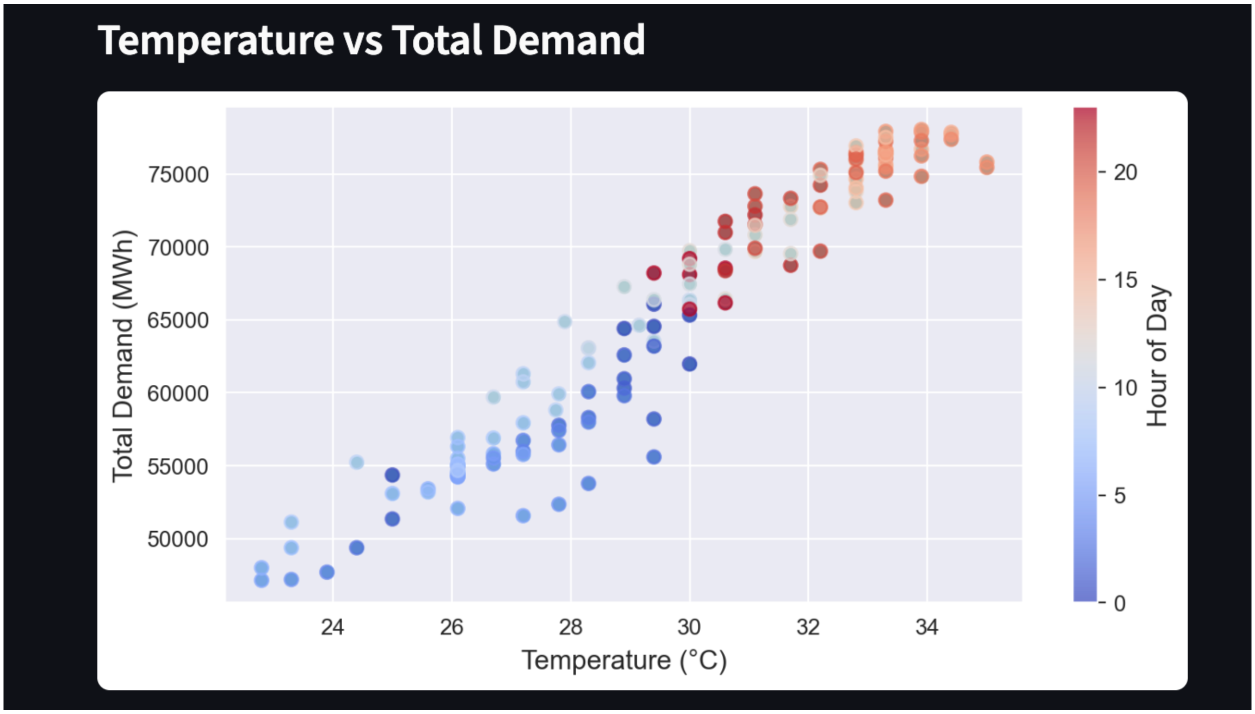

Figure 3: *Scatter plot of temperature vs total demand, with regression line showing a positive correlation between higher temperatures and elevated demand.*

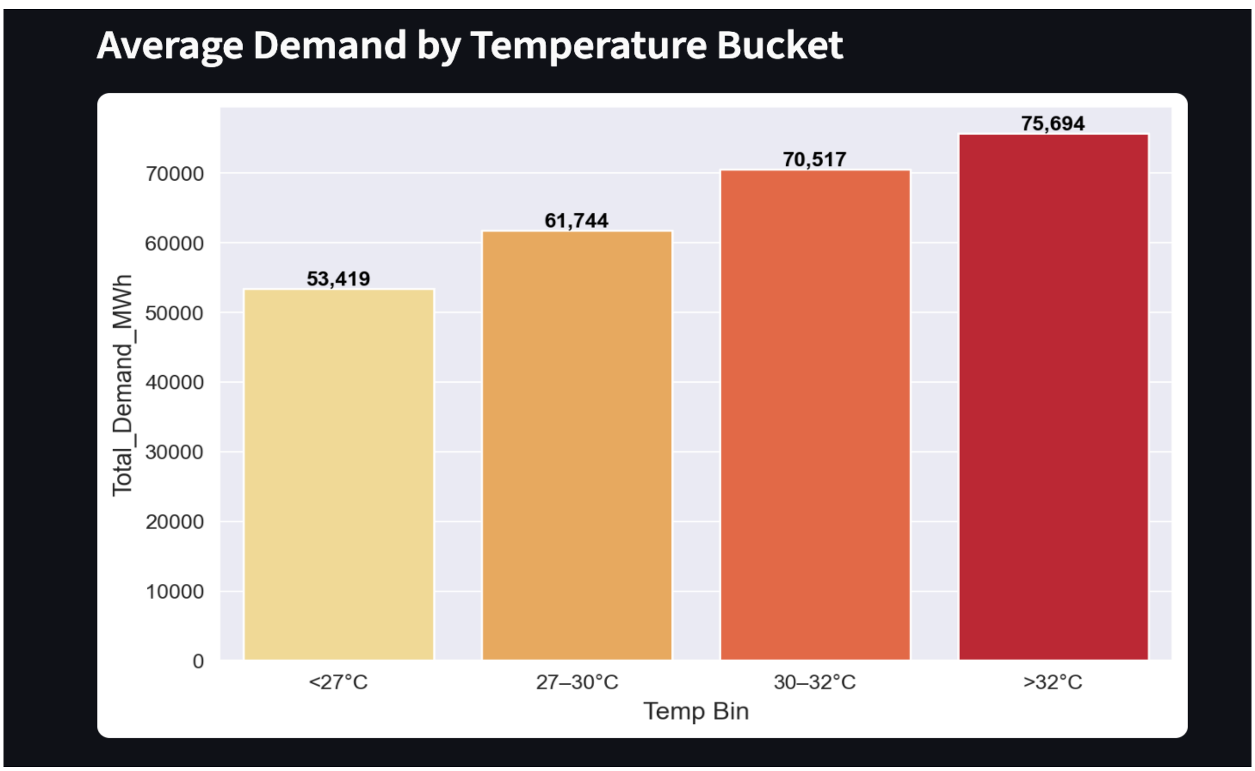

Figure 4: *Average demand by temperature buckets, illustrating a rise of more than 12,000 MWh between cooler and hotter conditions.*

Quartile comparisons showed that as demand rose from Q1 (low load) to Q4 (high load), natural gas increased its share by over 15 percentage points, while coal and nuclear held relatively stable. Fuel share analysis confirmed that during strain hours, fossil generation was disproportionately relied upon, with renewables unable to offset the surge in demand. Stacked area plots by hour further illustrated the dip in renewable share during peak stress, particularly between 3 - 7 PM.

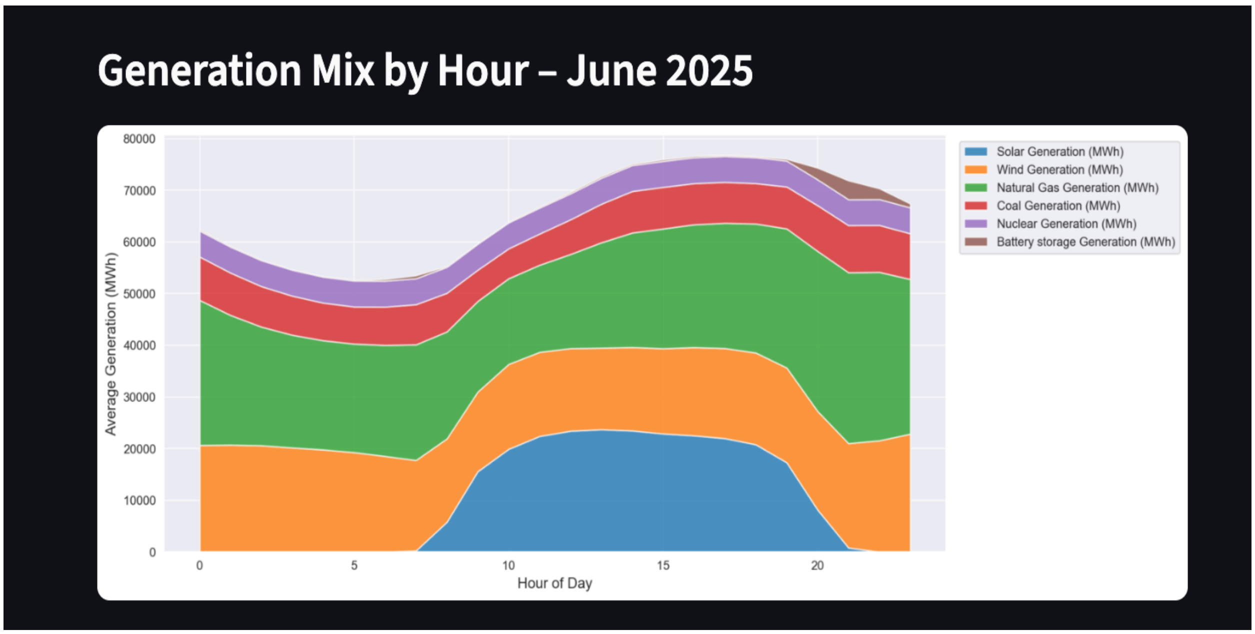

Figure 5: *Stacked area chart of generation mixes by hour*

5.4 CO_2 Emissions Analysis

Emissions trends closely tracked the shifts in fuel mix. During non-strain hours, CO_2 intensity averaged around 0.40 tons/MWh, but during strain hours it consistently approached 0.50 tons/MWh, with peaks above 0.52 tons/MWh on June 20 and June 21. Boxplots confirmed significantly higher emissions distributions during high load conditions.

Correlation heatmaps quantified these relationships:

- Strong positive correlation between temperature and demand ($\rho \approx 0.7$).
- Strong positive correlation between demand and natural gas generation ($\rho \approx 0.8$).
- Moderate positive correlation between demand and CO_2 emissions ($\rho \approx 0.6$).
- Weak or negative correlations between demand and solar/wind share, reflecting their limited alignment with peak stress periods.

These findings demonstrate that strain events not only challenge system reliability but also worsen environmental impacts by driving reliance on fossil generation.

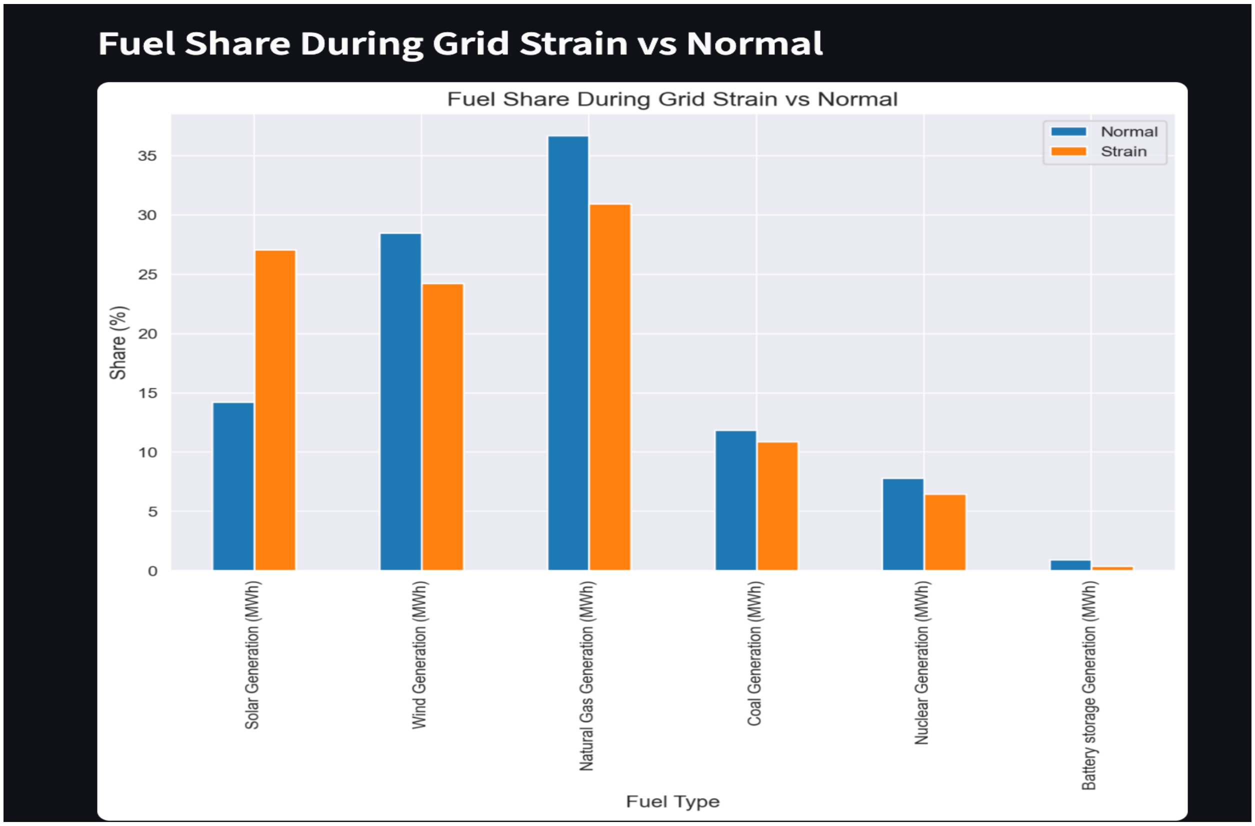

Figure 6: *Fuel share comparison during grid strain vs normal hours*

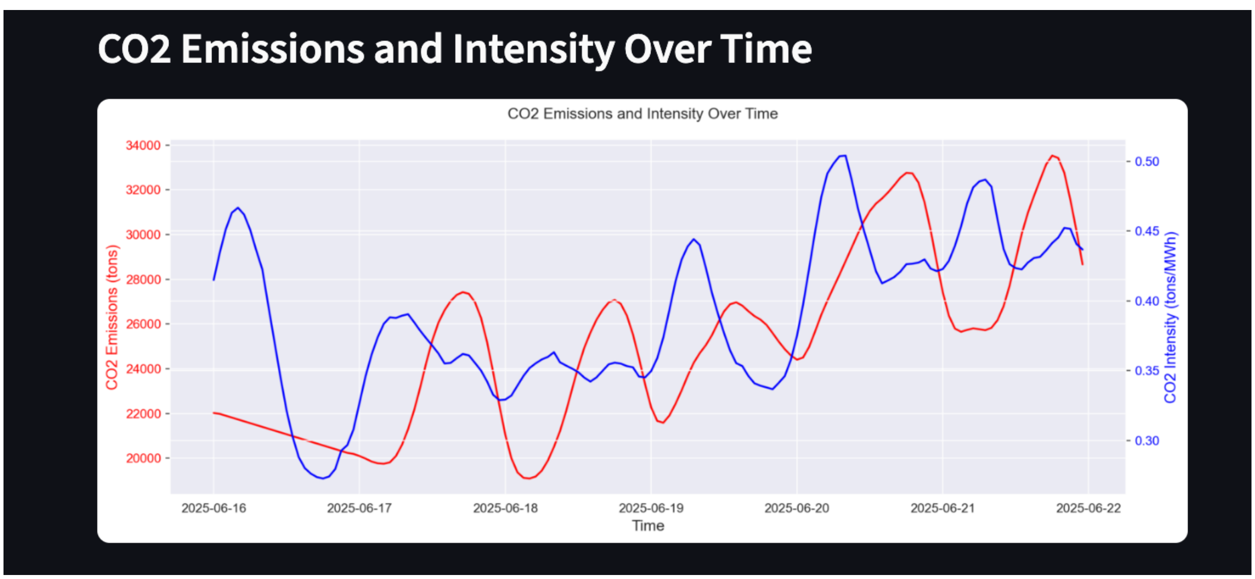

Figure 7: *CO_2 emissions and intensity over time, with emissions intensity rising from 0.40 tons/MWh under normal conditions to >0.55 tons/MWh during peak stress.*

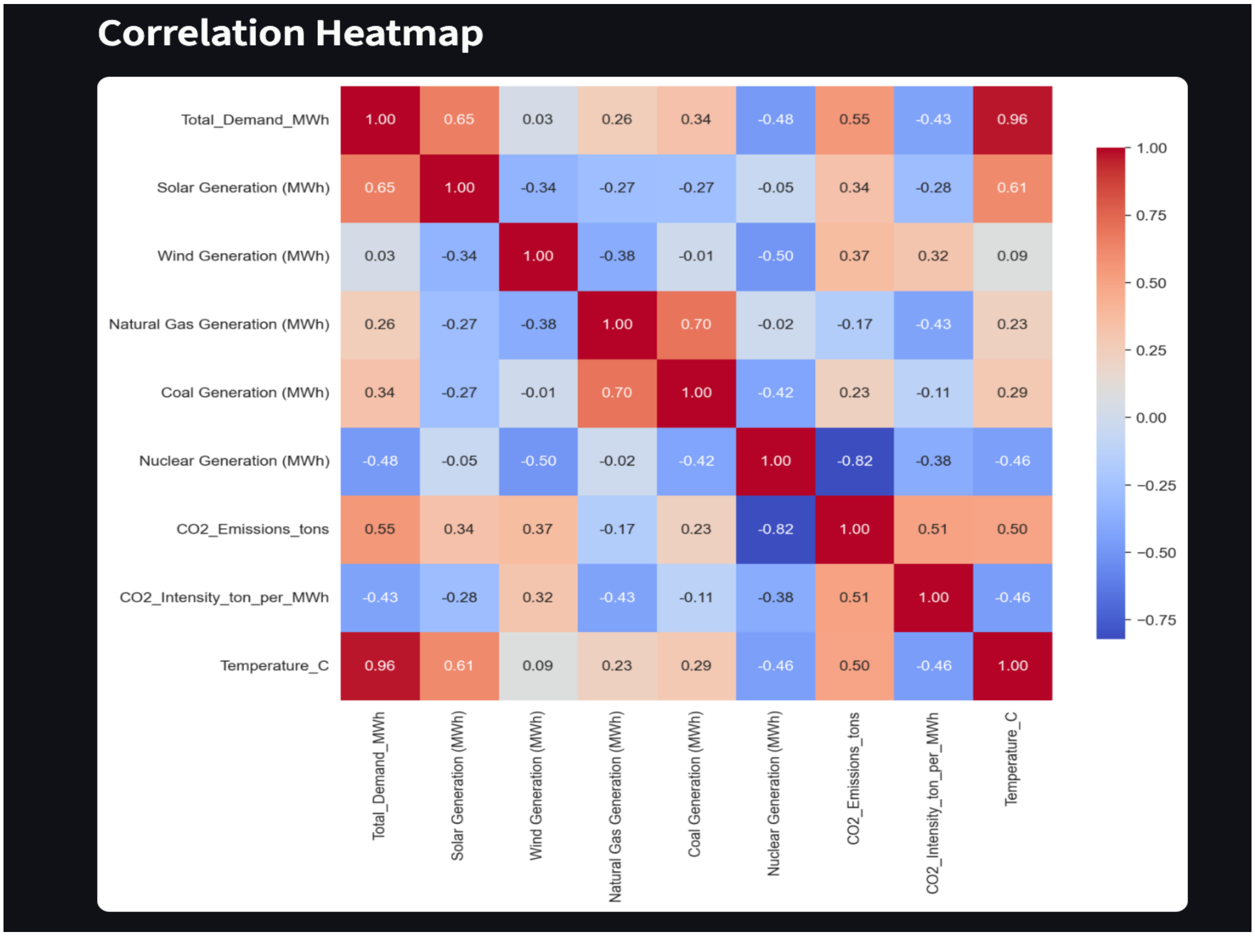

Figure 8: *Correlation heatmap of demand, generation, temperature, and emissions, showing strong associations between heat, demand, and fossil generation share.*

6. KEY INSIGHTS

- Grid strain was concentrated in a narrow band of hours (3-7 PM), accounting for just over 10% of observations yet responsible for the highest stress conditions.
- Heat was the primary external driver, with average demand rising by more than 12,000 MWh between cooler (<27°C) and hotter (>32°C) conditions.
- Natural gas was the pivotal balancing resource, contributing around 40% of generation during strain hours, while Solar's contribution dropped sharply after sunset.
- CO_2 intensity worsened during stress events, from 0.40 tons/MWh to peaks near 0.52 tons/MWh, reflecting dependence on fossil generation.

- Correlation analysis confirmed strong alignment between temperature, demand, and fossil generation, quantifying the environmental tradeoffs of peak management.
- Analytical techniques such as demand quartiles, temperature buckets, and fuel share overlays proved effective in surfacing system vulnerabilities in a compact dataset.

Implications: These insights emphasize that grid reliability and sustainability challenges are tightly coupled in high demand conditions. For planners, the findings illustrate how temperature driven surges exacerbate both operational strain and emissions intensity, reinforcing the need for demand side flexibility, storage deployment, and renewable alignment strategies. For a learner driven project, this analysis demonstrates how exploratory data science techniques can generate actionable insights even from a limited dataset, laying a foundation for predictive modeling and policy oriented evaluation in future work.

7. CONCLUSION & FUTURE WORK

The analysis of ERCOT's June 2025 operational data provided a compact but revealing view of grid strain dynamics. Demand consistently surged during late afternoons, with peak events clustered between 3-7 PM and culminating at 77,351 MW on June 19. These strain periods were tightly linked to elevated temperatures, and in turn drove heavier reliance on fossil generation. Natural gas supplied nearly half of total generation during high load conditions, while renewable contributions were either diminished or poorly aligned with peak timing. The consequence was a measurable rise in emissions intensity, with CO_2 rates climbing from 0.40 tons/MWh under normal conditions to above 0.52 tons/MWh during peak stress.

While informative, the study was bounded by a limited time window (6 days) and using proxy emissions data, which constrain the generalizability of the results. Future work should extend the analysis across multiple months or full years, integrate real time emissions datasets, and apply predictive machine learning models to forecast grid strain based on temperature, demand, and generation trends. Incorporating demand side management events, storage dispatch, and renewable curtailment would also improve the system level insights.

This project illustrates how even a short learning exercise can surface practical findings on the intersection of grid reliability and environmental performance, providing both a foundation for more advanced modeling and a demonstration of exploratory data analysis in energy systems.

References

1. Conner, Briana. 2025. "Despite Grid Improvements, Texans Remain at Risk for Summer Outages, Study Shows." ABC13 Houston (KTRK-TV), June 12, 2025. [URL]
2. Electric Reliability Council of Texas. 2025. ERCOT Monthly Operational Overview - June 2025. ERCOT. Published July 17, 2025. [URL]
3. NERC (2025). *2025 Summer Reliability Assessment.* North American Electric Reliability Corporation. [URL]
4. EIA (2024). *Hourly Electric Grid Monitor.* U.S. Energy Information Administration. [URL]
5. EPA (2024). *Continuous Emissions Monitoring Data (CAMPD).* U.S. Environmental Protection Agency. [URL]
6. NOAA (2024). *National Centers for Environmental Information Climate Data.* National Oceanic and Atmospheric Administration. [URL]