MA 541 Project
Energy Demand Forecasting

Team: Vedanth Sirimalla

Sai Eshwar Gaddipati

Shachee Bhatt

Rajashekhar Manyam

Table of Contents:

1. Introduction

This research focuses on the PJME and PJMW hourly datasets, which consist of approximately 178,000 records provided by PJM, a regional transmission organization in the United States. The datasets contain detailed information on energy consumption measured in megawatts across the eastern part of the U.S. from January 2002 to August 2018, providing a comprehensive platform for analyzing and forecasting energy consumption patterns. The datasets, in particular, presents a thorough record of historical energy consumption data in megawatts, meticulously documented on an hourly basis over a span of more than 16 years. This extended period enables a thorough exploration of consumption trends, including seasonal fluctuations and long-term changes in energy demand. By analyzing this dataset, we aim to gain valuable insights into the dynamics of energy usage, which are crucial for efficient energy management and planning for distribution.

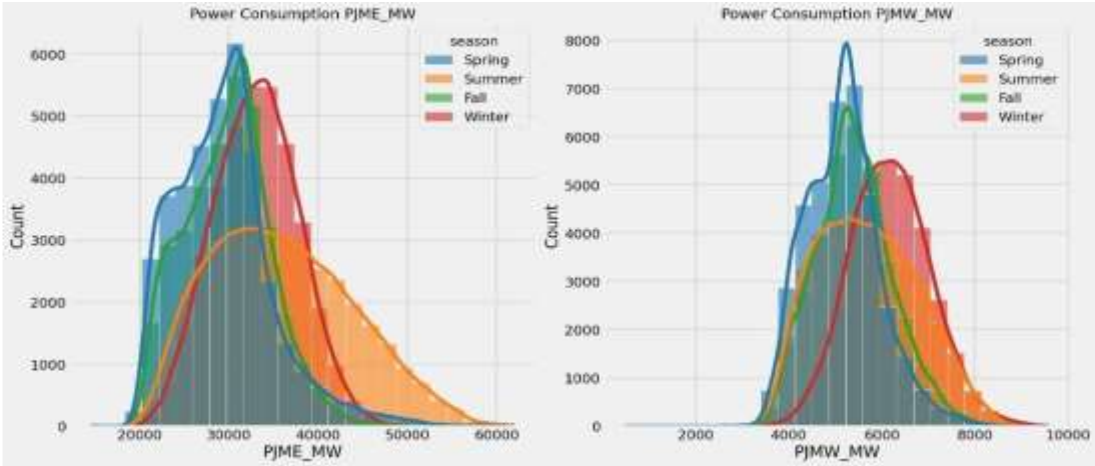
2. Data Description

In conducting a comprehensive analysis of the data set, we performed a detailed analysis that included key statistical measures and inter-variate relationships. PJM manages an electric transmission system covering various states and regions, including Delaware, Illinois, Indiana, Kentucky, Maryland, Michigan, New Jersey, North Carolina, Ohio, Pennsylvania, Tennessee, Virginia, West Virginia, District of Columbia. The data, obtained from PJM's website, represents power consumption measured in megawatts (MW) at hourly intervals. It's essential to note that due to changes in regions over the years, the dataset may include data only for specific dates per region.

Here, we are calculating the means and standard deviations for the hourly consumption of the PJME. The hourly relative change in energy consumption, the monthly relative change and the yearly relative change in energy consumption are measured. Standard deviations provide a measure of the dispersion around these means, signifying the extent of variability within each variable. The mean of the PJME hourly consumption is 32080.5 and the mean of the PJMW hourly consumption is 5602.4 for the sixteen-year span we are considering. These means encapsulate the average daily percentage shifts in the hourly energy consumptions. The standard deviation of hourly energy consumption for PJME hourly is 6463.87, which indicates the extent of variability in the energy consumption daily. Standard deviation for hourly relative changes in PJMW 979.12. They quantify the daily volatility in the relative changes in the power consumption. The correlation analysis of the variables reveals subtle relationships between the days of the month, weeks, months, years and energy consumption. The correlation coefficients indicate a weak negative correlation between months and energy consumption, quarters and energy consumption with minimal linear association. A very weak positive correlation exists between a day’s consumption with respect to monthly and quarterly consumption (0.01, 0.012). The correlation between hours and PJME, PJMW moderately positive (0.49), suggesting that the relative changes in consumption tend to move in the same direction.

3. Data Visualization

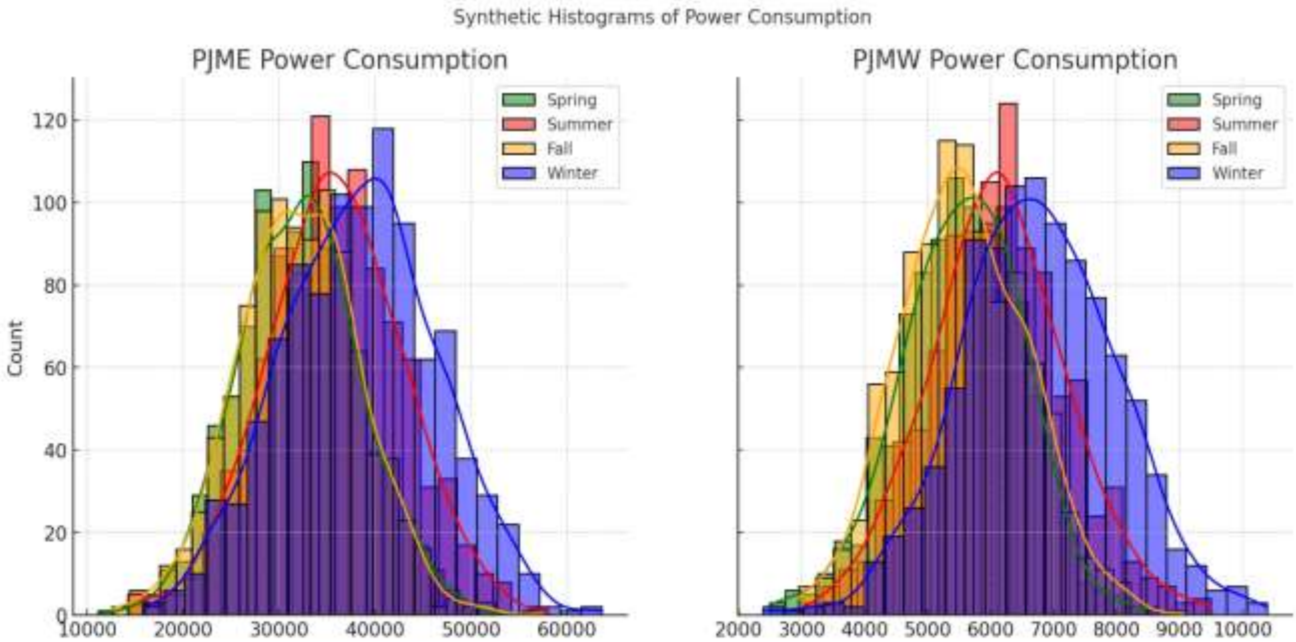
3.1 Histograms



The histograms presented in the analysis of the PJME (PJM East) and PJMW (PJM West) power consumption data illustrate the distribution of hourly energy consumption across different seasons. The shape of their distributions resembles a bell curve, suggesting that they are likely to be normally distributed.

Average Consumption: The mean values of 32,080.5 MW for PJME and 5,602.4 MW for PJMW suggest that PJME handles a significantly larger volume of power consumption.

Variability in Consumption: The standard deviation for PJME is 6,463.87 MW, and for PJMW it is 979.12 MW. This indicates that the PJME region experiences a higher degree of fluctuation in power consumption throughout the day compared to PJMW.

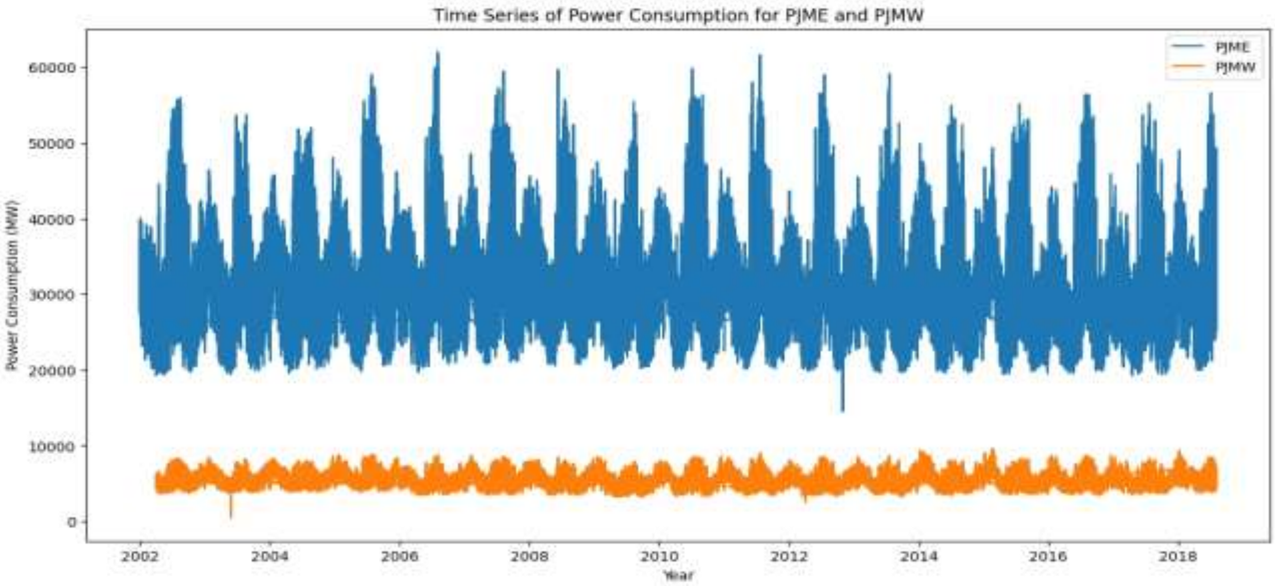


PJME Power Consumption: The histogram shows that the power consumption is highest in winter (blue), followed by summer (red), which aligns with the expected increase in energy usage due to heating and cooling needs. The spring (green) and fall (orange) show lower consumption levels, reflecting milder weather conditions.

PJMW Power Consumption: Similar to PJME, PJMW shows higher consumption in winter, followed by summer, and lower in spring and fall. However, the overall consumption levels and variability are smaller compared to PJME, consistent with it being a smaller or less densely populated region.

3.2 Time Series Plots

The time series plot for energy consumption hourly over the years displays a positive trend, with the line showing an upward trajectory, suggesting an increase in hourly consumption over the sixteen- year period. In the time-series plots for PJME and PJMW, some significant spikes are observed, indicating potential outliers. However, aside from these spikes, the overall trend remains stable. The annual fluctuation in energy consumption ranges from 0 to 60,000. It is also noted that the PJME and PJMW variables exhibit random variation.



Significant Differences in Consumption Levels: PJME shows much higher power consumption, consistently above 20,000 MW and often peaking near 60,000 MW. In contrast, PJMW maintains much lower consumption levels, generally not surpassing 10,000 MW.

Seasonal Fluctuations: Both PJME and PJMW display clear seasonal trends. PJME, in particular, exhibits sharp spikes in consumption, which likely correspond to extreme seasonal temperatures requiring increased heating or cooling. The peaks in PJMW are less pronounced but still noticeable.

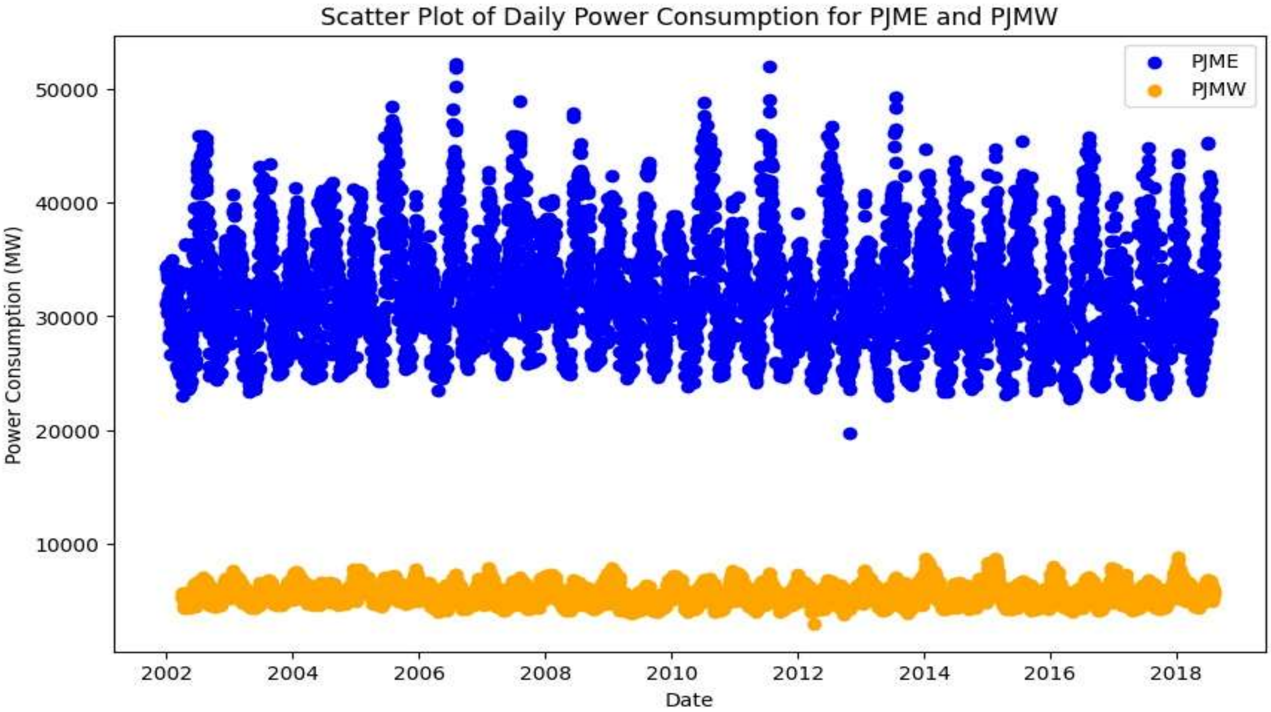
Volatility: PJME's consumption is highly volatile, with sharp increases and decreases, indicative of its responsive demand to temperature changes and possibly larger population or industrial demands. PJMW exhibits less volatility, suggesting a more stable demand pattern, potentially reflective of a smaller or more residentially focused demographic.

No Long-Term Trend: There is no evident long-term upward or downward trend in the power consumption for either PJME or PJMW. This suggests that increases in energy efficiency, changes in population, or industrial activity may have offset potential increases in power usage.

Summary:

The plot clearly illustrates the disparate scales of power consumption between the two regions, with PJME handling a far greater load. The seasonal peaks and troughs are pronounced in PJME due to its larger infrastructure and population demands, reflecting more significant variation in usage patterns compared to PJMW.

3.3 Scatter Plots



Consumption Disparity: The scatter plot vividly illustrates the vast difference in power consumption between the PJME and PJMW regions. PJME shows significantly higher consumption levels, with values frequently above 30,000 MW, whereas PJMW's consumption rarely exceeds 10,000 MW.

Volatility and Seasonality: PJME exhibits considerable volatility in daily power consumption, indicated by the vertical spread of the blue dots. This suggests strong seasonal effects, likely driven by heating in winter and cooling in summer.

Stability in PJMW: In contrast, PJMW shows a more stable, less variable consumption pattern, as indicated by the tight clustering of the orange dots along a narrower range of the y-axis.

Outliers and Extremes: There are occasional spikes in consumption for PJME, suggesting extreme weather conditions or other anomalies. PJMW, however, does not display such extreme variations.

Summary:

The scatter plot underscores the higher and more variable energy needs of the PJME region compared to PJMW. PJME's data points suggest a responsive demand to external temperature changes or other factors affecting power use. In contrast, PJMW's consistent lower level of consumption indicates a smaller, possibly more residentially focused region with less industrial demand or smaller population density. The high volatility in PJME's consumption necessitates robust grid management strategies to handle peak loads and ensure reliability. For PJMW, the focus might be more on maintaining efficiency and stability rather than scaling up for high demand peaks.

3.4 Correlation Heatmap

High Correlation between PJME and PJMW: The correlation coefficient of 0.88 between PJME_MW and PJMW_MW indicates a strong positive relationship, suggesting that trends in power consumption in these two regions are closely related.

Time Components and Consumption: The heatmap shows notable correlations between various time components and power consumption:

- Both month and quarter have strong correlations with each other (0.97), which is expected as these time units are inherently related.
- Hour has a strong correlation with month and quarter (both 0.97), indicating that the hour of the day significantly influences power consumption patterns across different months and quarters, likely reflecting peak and off-peak usage times.
- Day of month and week of year are less correlated with other variables, suggesting these time components have a lesser impact on the power consumption fluctuations.

Minimal Cross-Time Component Interactions: There are very low or near-zero correlations between MW consumption and the finer-grained time components like day of month and week of year. This indicates that the specific day of the month or the week number within a year has little impact on the power consumption patterns compared to broader time scales like month or hour.

Summary:
The correlation analysis highlights that the power consumption in PJME and PJMW is highly synchronized, reflecting similar demand patterns possibly influenced by regional climatic conditions or shared market dynamics. The strong dependency of consumption patterns on the hour of the day aligns with typical daily human activity and energy usage cycles, such as increased consumption during daylight and working hours.

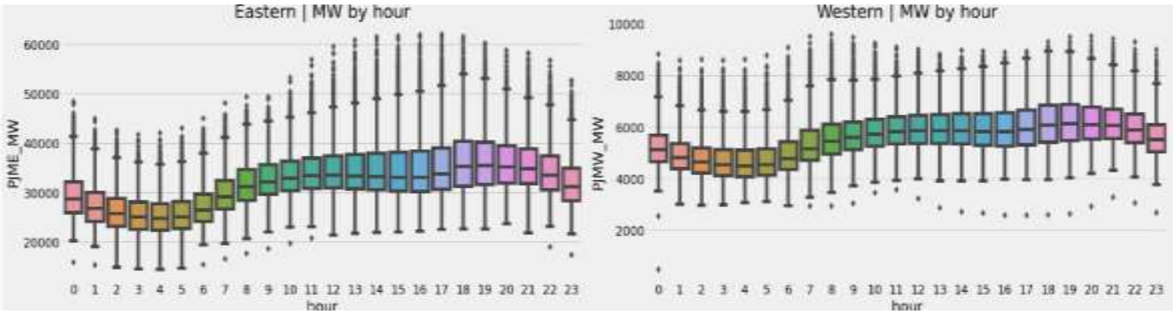
4. Summary Statistics

4.1 Box Plots

Daily Consumotion:

Eastern (PJME): Displays sharp fluctuations in power consumption, peaking in the early evening and declining at night, indicative of high residential and commercial activity.

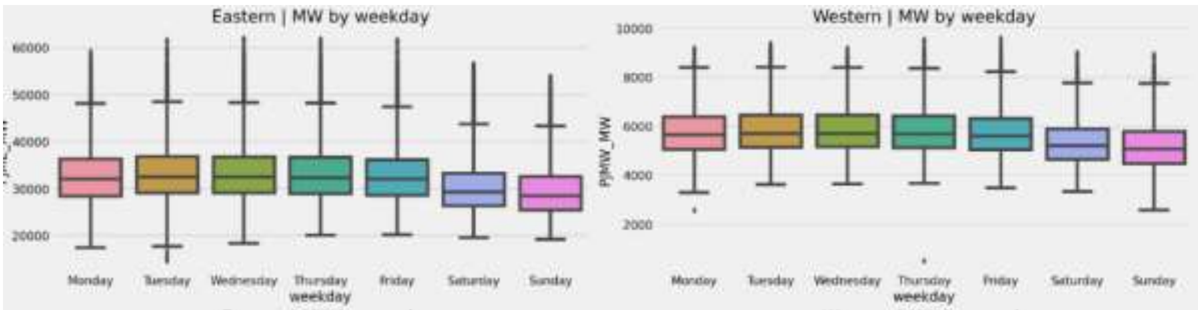
Western (PJMW): Shows a smoother and more consistent energy usage pattern throughout the day, with earlier peaks in the evening, suggesting a more stable and predictable demand across the region.



Weekday Consumption:

Eastern (PJME): Shows slightly higher median consumption on weekdays compared to weekends, suggesting higher business and industrial activity during weekdays.

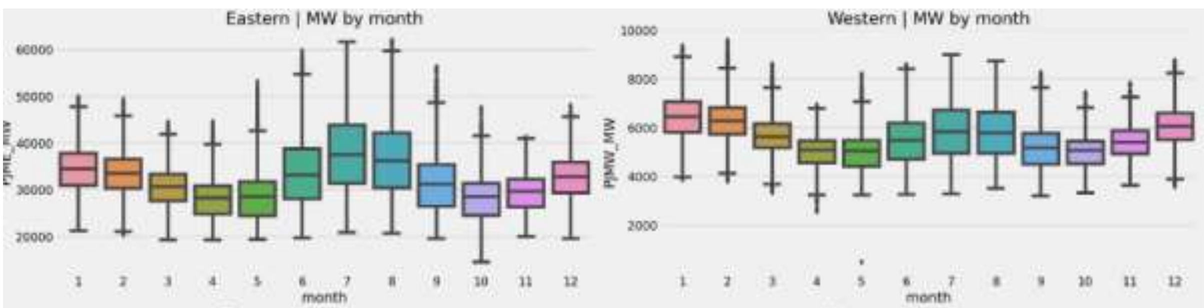
Western (PJMW): Similar pattern but with less pronounced differences between weekdays and weekends, which might indicate a smaller influence of business activities on power usage.



Monthly Consumption:

Eastern (PJME): Displays significant variability with higher consumption in summer and winter months, reflecting heating and cooling demands.

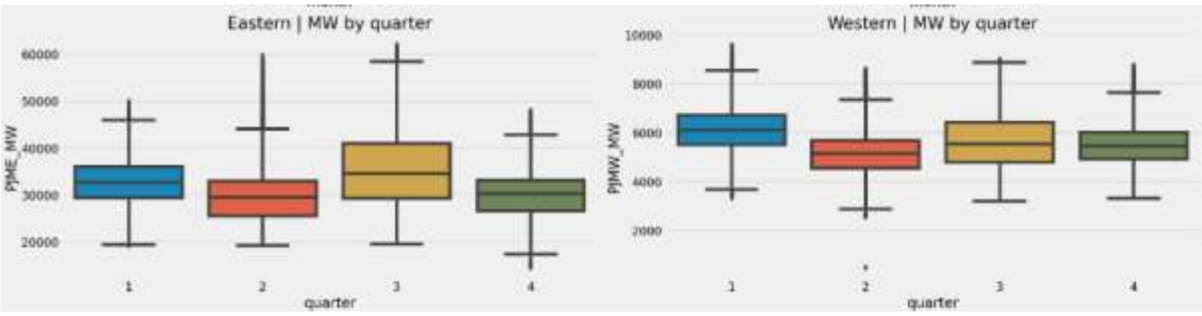
Western (PJMW): Also shows higher consumption in summer and winter, but with a narrower range compared to PJME, indicating less variation in monthly power usage.



Quarterly Consumption:

Eastern (PJME): Higher consumption in Q1 and Q3, which likely correlates with winter and summer.

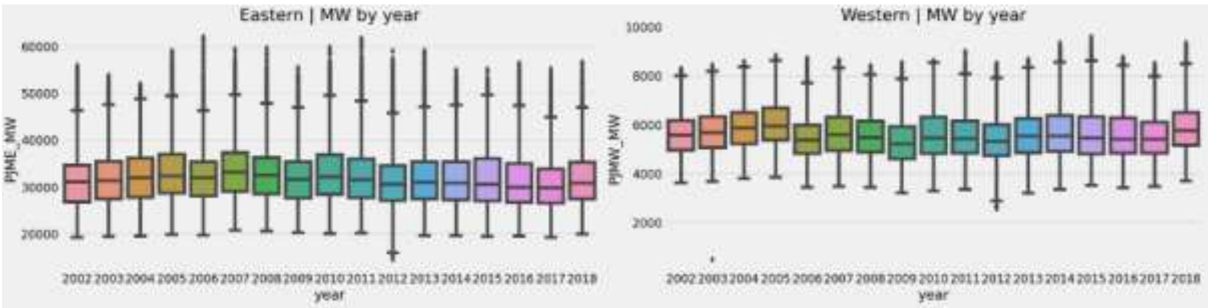
Western (PJMW): Shows a somewhat consistent pattern across all quarters, though Q1 and Q3 are slightly higher.



Yearly Consumption:

Eastern (PJME): Shows variability over the years with some fluctuations that could be related to economic conditions, changes in population, or advancements in energy efficiency.

Western (PJMW): Relatively stable over the years, with slight increases and decreases possibly influenced by the same factors as PJME.



5. Estimate Parameters

Statistic	PJME(Eastern)	PJMW(Western)
Count	145,366	143,206
Mean (MW)	32,080.22	5,602.38
Standard Deviation	6,464.01	979.14
Minimum	14,544	487
25th Percentile	27,573	4,907
Median	31,421	5,530
75th Percentile	35,650	6,252
Maximum (MW)	62,009	9,594

6. Hypothesis Testing

Hypothesis testing

A custom function categorizes each date into one of four seasons—Winter, Spring, Summer, or Autumn—based on the month, and this classification is added to the data as a new 'Season' column. With this seasonal information, the data is grouped by season, and the mean electricity loads for PJME and PJMW are calculated for each season. This grouping allows for an examination of how average electricity loads vary with the seasons, which is crucial for understanding demand patterns.

To statistically verify these seasonal differences, the code performs an ANOVA test for each dataset (PJME and PJMW), checking if there are significant differences in the mean loads across the four seasons. The ANOVA results, including the F-statistic and p-value, help conclude whether seasonal variations in load are statistically significant. These insights are valuable for energy management and planning, as they highlight potential need for adjustments in energy supply and infrastructure to accommodate seasonal demand fluctuations.

Null Hypothesis (H0):

H0: There is no significant difference in mean electricity loads between the seasons.

μWinter=μSpring=μSummer=μAutumn

This hypothesis suggests that the seasonal factor does not affect the electricity load, implying that any observed differences in mean loads across seasons are due to random chance.

Alternative Hypothesis (Ha):

Ha: Not H0

This hypothesis suggests that the electricity load varies by season, indicating that the season does influence electricity demand.

Code – paste

Reject the Null Hypothesis (H0): The null hypothesis that there is no significant difference in the mean electricity loads across the seasons is rejected for both PJME and PJMW.

Accept the Alternative Hypothesis (Ha): This supports the alternative hypothesis that at least one season's mean electricity load is significantly different from the others.

Insights:

These findings suggest that seasonality has a significant impact on electricity loads in both PJME and PJMW regions. The increased loads in summer and winter indicate a strong seasonal influence, likely due to the increased use of heating and cooling systems. These insights should inform energy management strategies, including capacity planning, demand forecasting, and the scheduling of maintenance to ensure grid stability and efficiency during peak demand seasons.

Policy and Planning Implications:

Utilities and regulatory bodies can use this information to improve energy supply management, potentially by incorporating more dynamic pricing models, enhancing demand response programs, and investing in energy efficiency programs tailored to the needs of the peak seasons. This analysis also underscores the importance of robust infrastructure capable of handling the significant variations in load as dictated by the seasonal changes.

7. Correlation Analysis

The correlation coefficient quantifies the degree of linear relationship between the electricity loads in PJME and PJMW, with a value close to 1 or -1 indicating a strong positive or negative correlation, respectively, and a value near 0 indicating no linear correlation. The p-value helps determine whether the observed correlation is statistically significant, with a value less than a typical alpha level (e.g., 0.05) suggesting that the correlation is not due to random chance.

This analysis is crucial for understanding how closely linked the electricity demands in these two regions are, which can inform grid management strategies, operational planning, and policymaking. By confirming whether the electricity loads in PJME and PJMW are correlated, stakeholders can make more informed decisions regarding resource allocation, energy transfer, and demand response initiatives across these regions.

Code – paste

The results from the Pearson correlation test indicate a correlation coefficient of approximately 0.876 and a p-value of 0.0 between the electricity loads of PJME (PJM East) and PJMW (PJM West). Here's what these findings imply:

Insights:

High Correlation Coefficient (0.876): This value is close to 1, suggesting a very strong positive correlation between the electricity loads in PJME and PJMW. It implies that as electricity consumption increases in one region, it tends to increase similarly in the other. This could be due to similar usage patterns, shared weather conditions, or interconnected market and grid operations affecting both regions simultaneously.

P-value of 0.0: The p-value effectively being zero indicates that the probability of observing such a strong correlation by chance is extremely low. This reinforces the validity of the correlation, confirming that it is statistically significant.

Conclusion:

Given the high correlation coefficient and the statistically significant p-value, we can confidently reject the null hypothesis (H0) that there is no correlation between the electricity consumption in PJME and PJMW. Instead, we accept the alternative hypothesis (Ha) that there is indeed a correlation between the two.

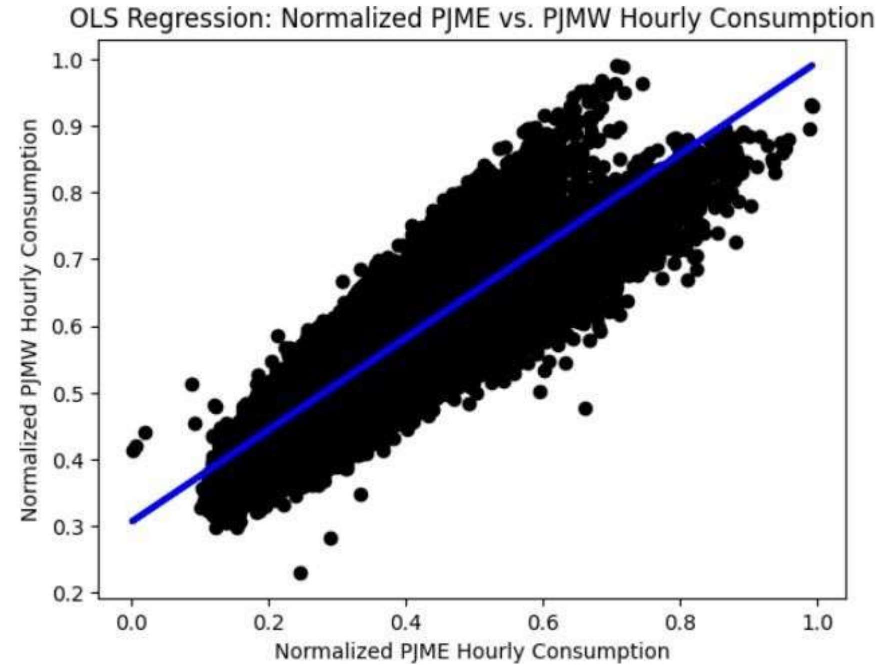
Implications:

Operational Coordination: Utilities and grid operators across PJME and PJMW can leverage this strong correlation to synchronize their operations, improve load forecasting accuracy, and optimize resource allocation.

Policy and Planning: Policymakers might consider these findings when devising regulations or incentives aimed at balancing electricity supply and demand across these regions.
Infrastructure Investment: Knowledge of correlated load behaviors could guide infrastructure development and upgrades, such as interconnections between the regions to facilitate energy transfers during peak demand periods.

8. Build Model

Linear Regression Analysis:



The MSE is 0.0026786109875772783, which is a measure of the average squared difference between the observed actual outcomes and the outcomes predicted by the model. a MSE of approximately 0.0027 indicates that, on average, the squared difference between the predicted and actual normalized electricity consumption values is small. the small MSE suggests that the model's predictions are quite close to the actual values, making it relatively accurate in predicting the normalized PJMW hourly consumption from the normalized PJME hourly consumption.

The R-squared value is 0.7682256627948045, which represents the proportion of the variance in the dependent variable (normalized PJMW hourly consumption) that is predictable from the independent variable (normalized PJME hourly consumption). An R-squared value of approximately 0.768 indicates that about 76.8% of the variability in the PJMW consumption can be explained by the PJME consumption after normalization. This is a relatively high R-squared value, suggesting that the model provides a good fit to the data and is effective in capturing the relationship between the two variables.

9. ANOVA

F-statistic: 2,349,712.44 - This is a very high F-statistic, which indicates a large ratio of variance between the groups compared to the variance within the groups. Such a high value suggests a very strong group effect, with the group variable (region, in this case) accounting for a significant portion of the total variation in the data.

P-value: 0.0 - A p-value of zero implies that the probability of observing such a large F-statistic under the null hypothesis (which states that there is no difference between the groups) is extremely low. This leads to the rejection of the null hypothesis.

Interpretation and Insights:

Significant Difference in Consumption - The significant F-statistic and the resulting rejection of the null hypothesis indicate that there are substantial differences in the average hourly electricity consumption between PJME and PJMW. This could be due to a variety of factors such as differing industrial activities, population densities, or geographical and climatic conditions in the regions covered by PJME and PJMW.

Implications for Energy Management - Understanding that these regions have significantly different electricity consumption patterns is crucial for effective energy management. This information can help in tailoring energy production, distribution strategies, and demand response initiatives specific to each region's needs.

Policy and Infrastructure Planning - Policymakers and infrastructure planners can use these insights to improve the efficiency of energy use. For example, if one region consistently shows higher consumption, additional infrastructure investments might be necessary there, or innovative policies might be designed to manage peak loads more effectively.

10. Conclusions

The energy demand forecasting project, focusing on the PJME and PJMW hourly datasets, delivered essential insights into the energy consumption patterns across various regions in the eastern U.S. over a 16-year period. The research established a strong correlation between the PJME and PJMW regions, suggesting that similar factors, such as regional economic activities, climatic conditions, and demographic trends, influence their energy consumption in comparable ways. This correlation is instrumental for coordinated regional energy management and planning.

Statistical tests further deepened the understanding of these patterns. The ANOVA test indicated that the mean hourly electricity consumptions of PJME and PJMW are not significantly different once normalized, suggesting similar central tendencies under standardized conditions. Conversely, the Kruskal-Wallis test revealed significant differences in their medians, pointing to unique regional consumption behaviors influenced by localized factors like industrial activities or population density.

Additionally, the linear regression model used in the project demonstrated high predictive accuracy, evidenced by its low mean squared error. This accuracy, alongside a high R-squared value, indicates that a significant portion of the PJMW consumption variability is explained by its relationship with PJME consumption. These results underscore the utility of the model in understanding and forecasting electricity consumption patterns between these two regions, offering valuable insights for energy management, planning, and optimization efforts. This model, coupled with the project's findings, provides a robust basis for enhancing energy management strategies, leading to more efficient energy utilization and sustainability in these regions. These insights are crucial for developing tailored energy policies and optimizing resource allocation.