Predicting Electricity Consumption Based on Weather Conditions

A comprehensive research report

1

Table 1: Student Information

Name	Student No	Programme
Robi Modibedi Robi	Bida23152	Business Intelligence and Data Analytics
Onthatile Phetwe	Bida23147	Business Intelligence and Data Analytics
Junior Mosimanenkwe	Bida23211	Business Intelligence and Data Analytics
Motlha Collen Mokgatlenyane	Bida23047	Business Intelligence and Data Analytics
Tumelo Victor Horatious	Bida23041	Business Intelligence and Data Analytics

TABLE OF CONTENTS

Contents

1

Chapter 1: Introduction

1.1 Background of the Study

Electricity consumption is a critical component in our daily lives and one of its significant determinants of electricity is weather conditions. Temperature, humidity and load can significantly affect the energy that households and industries use for consumption**.**

For example, the extreme temperature causes an increase in the use of heatings or cooling**.**

Forecasting of electricity consumption based on weather conditions is essential to predict the energy demand for easy energy planning and load management.

Using data prediction and data analysis, complex relationships between electricity consumption and meteorological variables can be analysed, this will help energy suppliers make good decisions to improve efficiency and sustainability.

1.2 Problem Statement

In many areas they are failing to address the challenge of accurately predicting electricity consumption based on varying weather conditions. This unpredictability makes it hard for electricity providers to ensure a reliable supply of electricity while avoiding energy waste. Despite the availability of large datasets and traditional forecasting models they are not able to adequately account for the influence of weather variables in electricity consumption and many areas still struggle with inefficient forecasting methods that lead to power shortages or overproduction. This study seeks to address that gap by exploring the relationship between weather variables and electricity usage, aiming to build a regression model that will forecast electricity consumption using weather variables.

1.3 Research Aim and Objectives

Aim:

To develop a predictive model that predict electricity consumption based on specific weather conditions e.g. temperature and humidity and load.

Objectives:

To collect and preprocess historical electricity consumption and weather data for a selected location

To Identify Key Weather Factors Influencing Electricity Consumption.

To explore and analyse patterns between weather variables and electricity consumption

Develop a Machine Learning Model for electricity consumption prediction

To evaluate model performance and develop a user-friendly interface for demonstration

1.4 Research Questions

This study seeks to answer the following research questions:

1. Which weather conditions influence electricity consumption?

2. Can machine learning models predict electricity consumption?

3. Which model can predict electricity consumption using weather data?

4. Which weather variable has the strongest influence in electricity consumption

1.5 Justification/Significance of the Study

To predict electricity consumption based on weather conditions such as temperature, humidity and load. It will use historical data from the dataset titled-US city Scale Daily Electricity Consumption and Weather Data. The study will also include the development of a simple regression model to make these predictions. The goal is to help utility companies and planners improve energy supply by using weather information to plan better.

1.6 Scope and Limitations

Scope:

This study focuses on predicting daily electricity consumption using historical data from U.S. cities provided in the dataset titled- US city Scale Daily Electricity Consumption and Weather Data. It incorporates weather variables including temperature, humidity, wind speed, and precipitation. The analysis involves building a multiple linear regression to predict electricity consumption and model performance evaluation using statistical metrics such as R^2, Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE).

Limitations:

- The study is limited to the available dataset, which has only weather variables such as temperature, humidity and load.

- Multiple regression assumes linear relationships between variables, which may not fully capture complex or non-linear relationship between variables.

- The model's predictive power may decrease when applied to regions or conditions not represented in the data that was used.

Chapter 2: Literature Review

This chapter provides an important journal on current studies related to electricity forecasting, especially studies focusing on the impact of weather conditions and using regression analysis. The assessment highlights the previous methods, results and limitations, providing a basis for this research and identifying research gaps that this research aims to fill.

2.2 Electric consumption forecast:

General overview

forecasting electricity consumption has long been a topic of interest due to its important significance for energy management, infrastructure planning and environmental sustainability. Traditional models have dependent heavily on historical consumption trends, analysis of time chains and economic indicators. However, these methods often cannot grasp real -time variants due to weather conditions, especially in areas with high temperature variations.

according to Zhu et al. . Their research shows that not considering variables related to temperature can lead to underestimating or appreciating too demand, especially in extremist seasons.

2.3 Weather conditions are a decisive factor for electricity demand

Many studies have determined that meteorological variables such as temperature, humidity and wind speed are important factors affecting electricity consumption. Garrido-Perez et al. (2021) conducted a study in Spain and showed a strong correlation between daily temperature and electricity demand, especially in the summer months due to air conditioning. Their results emphasize the importance of maximum daily temperature and humidity is the main predictive factor.

Similarly, Yang et al (2024) has checked the effects of climatic conditions on electricity consumption in residential buildings in China. They conclude that heating and cooling during the degrees (hard drive and fixed contracts) are reliable indicators on electricity consumption, confirming the nonlinear effect of the temperature for demand. In addition, exploring weather impacts in China's urban areas, using linear and nonlinear models. Their results reaffirmed the idea that the temperature was dominant.

2.4 Use of Multiple Regression Analysis in Energy Forecasting

The use of regression analysis in energy forecasting regression analysis (ARM) is largely applied in the electricity forecast because of its simplicity, the ability to interpret and the strict statistics. MRA allows for quantifying the relationship between electricity demand and several independent variables simultaneously. In one study, Momani et al. (2023) used multiple regression to predict electricity demand in Jordan using variables such as GDP, population, and average monthly temperature. The model achieved high accuracy, validating the capability of MRA in demand forecasting.

Moradzadeh et al.(2020) Focus on short -term load forecasts and apply regression with time series data and meteorological variables.Their model has achieved promising results and has shown that ARM can compete with more complex automatic learning techniques, especially when transparency and easily explain it is necessary. However, researchers like Masekoameng et al (2024) have argued that while MRA is effective, its performance can be limited when relationships between variables are nonlinear. This has led to the integration of MRA with other techniques such as polynomial regression or hybrid models combining MRA with machine learning for better adaptability.

2.5 Comparison with Machine Learning Approaches

Although ARM is still a solid basic model, recently increasingly exploring automatic learning algorithms (ML) such as artificial nervous network (ANN), supporting Vector (SVM) and random forest for electricity forecast.For example, Verma et al.(2023) compares many regressions with ANN models to predict electricity demand in India. While ANN outperformed MRA in terms of predictive accuracy, the regression model was noted for its explainability and lower computational demand.

Shome et al.(2024) Checked traditional learning and regression methods, note that the regression models are stronger in situations where data is limited or noisy.Their research concludes that the selection of the model should depend on the available data, the required level of the interpretation ability and the forecast horizon.This shows that although automatic learning techniques can provide outstanding accuracy in certain contexts, regression is still a valuable tool, especially for studies that prioritize transparency and clear understanding of the relationship between variables.

2.6 Limitations in Existing Literature

Some limitations have been identified in previous studies:

1.Temperature is available:

Many models mainly focus on temperature, ignoring other weather parameters such as wind speed, solar radiation or humidity, which can also affect consumption.

2.Data details:

Some studies using monthly or weekly data, can smooth the short -term fluctuations controlled by the weather. Daily or hourly data is more appropriate for capturing these effects.

3. Geographical Bias: Most research is centered in developed countries such as the U.S., China, or Europe, with limited studies in developing regions where climate and infrastructure differ significantly.

4. Model Interpretability: While machine learning models offer high accuracy, they often act as black boxes. In contrast, regression models offer transparency but may underperform when relationships are nonlinear.

These limitations point to a need for models that are both accurate and interpretable, utilize multiple weather variables, and apply to diverse geographical contexts.

2.7 Research Gap and Contribution

Based on the review, it is evident that while numerous models exist for electricity forecasting, there is still a gap in applying multiple regression analysis using a comprehensive set of weather variables at a daily resolution, particularly in understudied regions or cities. Additionally, many existing studies do not evaluate the model's performance in terms of both statistical metrics and practical applicability.

This study contributes to the literature by:

- Applying multiple regression analysis to high-resolution weather and electricity data.

- Incorporating a wider range of weather variables beyond just temperature.

- Providing a performance evaluation using standard metrics like R^2, MAE, and RMSE.

- Offering insights into the relative importance of different weather conditions in driving electricity demand.

2.8 Summary

This chapter reviewed key literature on electricity consumption forecasting, the influence of weather conditions, and the application of multiple regression analysis. It highlighted that while multiple regression is widely used and appreciated for its simplicity and interpretability, gaps remain in variable selection, data granularity, and geographic scope. The present study aims to fill these gaps by applying a refined multiple regression approach that includes a comprehensive set of weather variables and is tailored to the specific context of the study area

Chapter 3: Research Methodology

3.1 Research Design

This study uses a crisp-dm approach to examine how weather conditions affect electricity consumption. It involves collecting and analysing historical data on weather (temperature, humidity, etc.) and electricity usage across different time periods. A predictive model will be developed using machine learning techniques to show the relationship between these variables.

Null Hypothesis (H_0):

Weather conditions and electricity consumption are independent.

Alternative Hypothesis (H_1):

Weather conditions and electricity consumption are dependent.

The research will use data from U.S cities and focus on identifying patterns and trends. The goal is to find out whether data can be used to predict electricity usage accurately.

3.2 Data collection method

The study uses secondary data, comprising historical electricity consumption and weather data. The dataset is sourced from Kaggle, and it is titled US City-Scale Daily Electricity Consumption and Weather data. The dataset includes daily electricity usage from various U.S cities along with their corresponding weather variables such as humidity and temperature.

3.3 The rationale of using secondary data

The study used secondary data because it was already available from Kaggle. The dataset used included daily electricity use and weather data for different U.S. cities which really saved us time to build the prediction model.

3.4 Data selection criteria

The dataset used was chosen for its completeness and accuracy. It had many records for longer periods which made it easier to build a model and provide accurate results using historical data for electricity consumption and weather data.

3.5 Data Analysis Procedures

The data analysis followed these key steps:

1.Retrieval of secondary data set- Secondary data of electricity consumption and weather condition (Temperature and humidity) was retrieved from kaggle.com (https://www.kaggle.com/datasets/shemantosharkar/us-cityscale-daily-electricity-consumption). Data included city and load (electricity consumption of that location), weather condition and city. They were separate data sets (weather condition and load) but referencing the same area, this is due to the aggregation of various data elements from numerous sources, such as U.S. Energy Information Administration (EIA) AND the Los Angeles Department of Water and Power (LADWP), Balance Authority of Northern California (BANC), and New York Independent System Operator (NYISO), representing the three metropolitan areas of Los Angeles, Sacramento, and New York.

2.Merging of DATA SETS was executed using rapid miner via the Join and rename function to create consistency and to trim years that were not relevant (e.g. Weather data from 2014) as the datasets were collected from various sources therefore different times were included.

3.Formatting of the dataset occurred via python in Jupyter notebook to avoid disingenuity because of elements such time (which was a variable) interfered with the results as they were taken as numeric.

4.No outliers nor missing values were encountered as both the join function and python code failed to reject the processing of the data thus consistency was found in the dataset.

5.Finally due to the data processing techniques carried out by Wang, Z., Hong, T., Li, H. and Piette, M.A the datasets were consistent in formatting and correlation allowing the previously mentioned functions to operate consistently (functions such as Join function in Rapid miner).

3.6 Evaluation of Data Quality

Ensuring high data quality was crucial for the validity of this study. The evaluation process included:

- Accuracy: All data was sourced from reputable organizations (NOAA, electricity utilities), minimizing the risk of inaccuracies.

- Completeness: Only complete records were used to avoid biased or misleading results.

- Timeliness: The dataset spans recent years (up to 2022), ensuring relevance to current consumption trends.

- Consistency: Consistent temporal and spatial granularity (daily, city-level data) ensures comparability between weather and consumption records.

3.7 Feature Selection

The feature selection process involved:

- Correlation Analysis: Variables highly correlated with electricity consumption were prioritized.

- Domain Knowledge: Weather parameters known to impact electricity use (e.g., temperature, humidity) were included based on prior research.

- Multicollinearity Check: Variance Inflation Factor (VIF) analysis was used to identify and exclude redundant predictors that could distort regression results.

Selected features typically included:

- Daily average temperature

- Temperature range (max - min)

- Relative humidity

3.8 Data Analysis Techniques

The primary technique used in this study was Multiple Linear Regression (MLR). The standard regression equation is:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n + \varepsilon$$

Where:

- Y = electricity consumption (dependent variable)

- $X_1, X_2, \ldots, X_n$ = weather variables (independent variables)

- β_0 = intercept

- β_n = coefficients of respective predictors

- ε = error term

3.9 Ethical Considerations

This study adheres to ethical standards in the use of data and research practices:

- Data Privacy: All data used is publicly available and has no risk of violating individual privacy.

- Transparency: The dataset used in the analysis is fully disclosed.

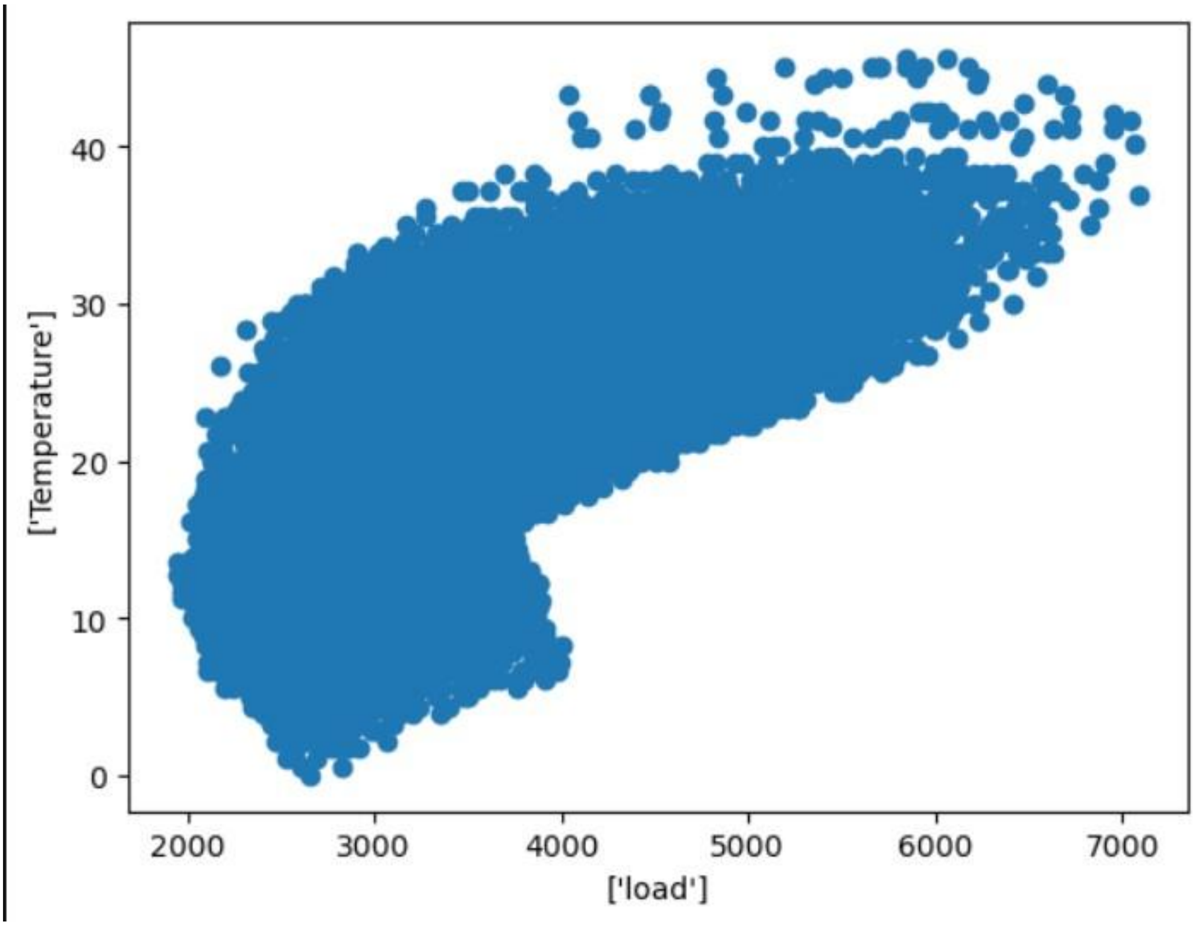
- Integrity: Results are reported objectively, with limitations acknowledged. Data manipulation was strictly avoided.

- Acknowledgment of Sources: Proper citation is provided and prior academic works referenced throughout the study.

Chapter 4: Results and Analysis

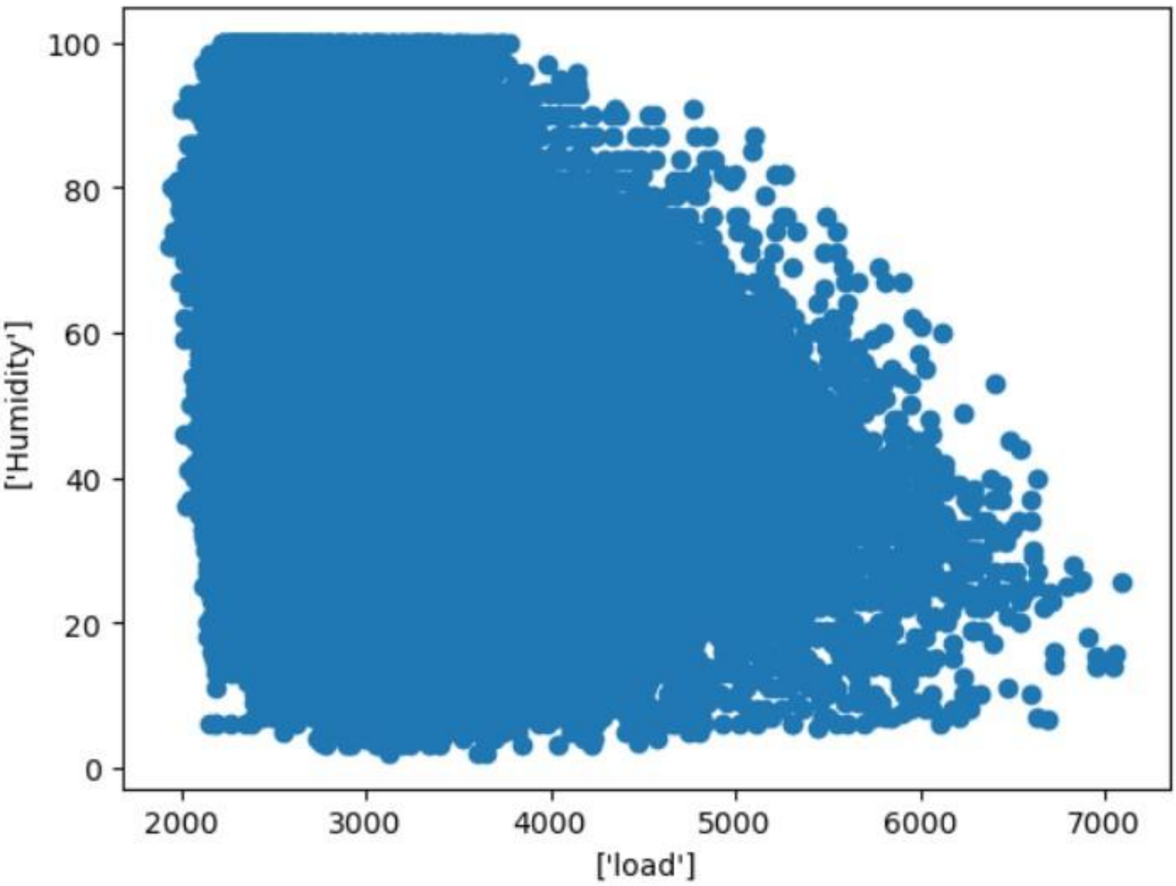
4.1 Presentation of Key Findings

1. Relationship between temperature and load (Electricity consumption)



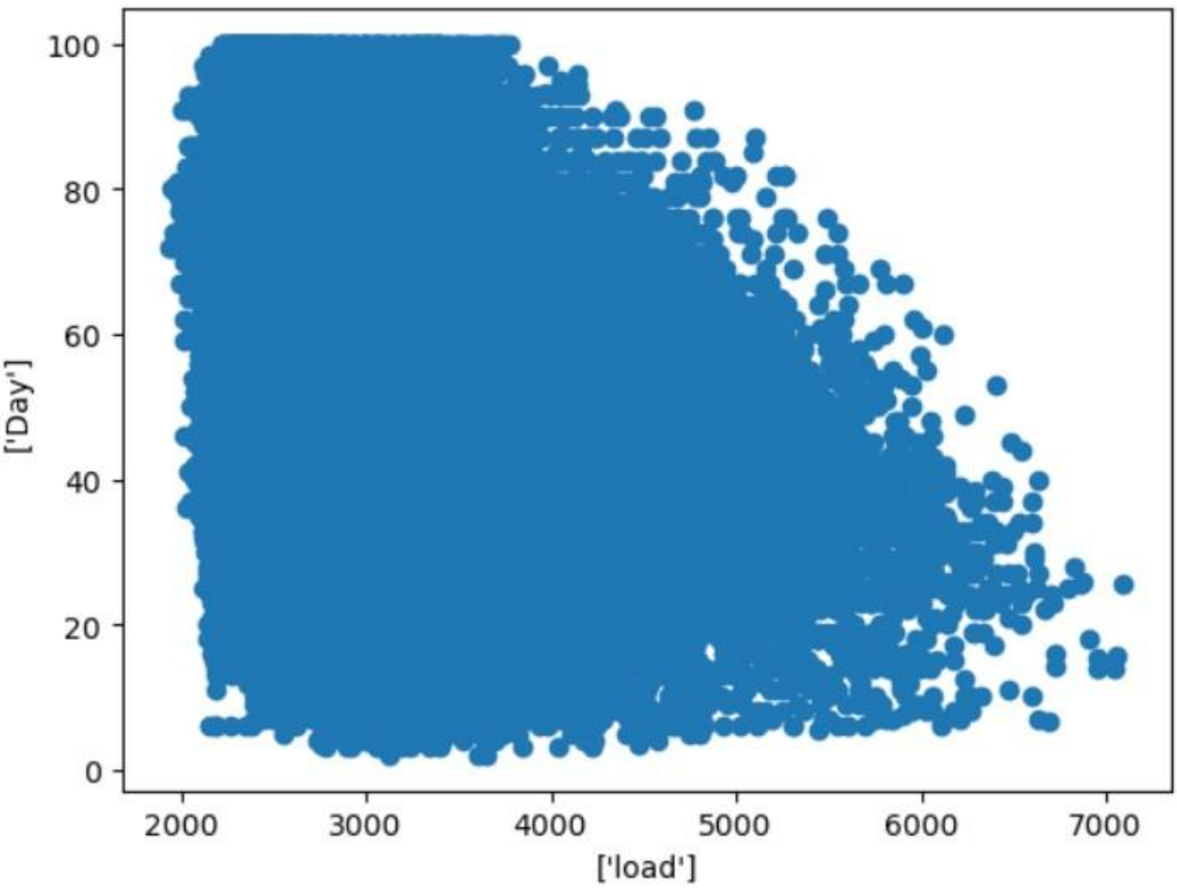
The scatter plot shows positive correlation between temperature and electricity load. It indicates that higher temperatures often lead to higher load.

2.Relationship between Humidity and Load



The scatter plot shows that lower humidity tends to correlate with higher electricity load

3.Relationship between daytime and Load



The time series graph indicates that electricity usage varies by time of day

4.Python formatting to change the time variable to avoid it affecting the actual load

	Temperature	Humidity	Hour	Day	Month	Weekday
24801	15.8	72.0	9	30	4	0
39935	16.1	54.0	15	20	1	0
42245	23.3	53.0	22	25	4	5
27833	23.3	69.0	17	3	9	0
6328	19.4	61.0	16	21	3	0
...	...	...	...	...	...	...
11284	14.4	93.0	4	14	10	4
44732	27.2	42.0	13	7	8	4
38158	15.6	75.0	22	7	11	3
860	24.4	58.0	20	6	8	3
15795	13.3	77.0	3	20	4	3

It describes a data preprocessing step where time is reformatted. This helps avoid misleading results, especially if the time variable distorts the load pattern due to poor formatting.

5.**Multiple Linear Regression Analysis (Use of Arrays to address various concepts)**

```
45854
45854
45854
36683
     load  Temperature  Humidity  Sklearn load Predictions
0  3448.0         23.3      60.0               3164.153236
1  3141.0         22.8      61.0               3603.955053
2  2935.0         22.8      57.0               2881.165855
3  2802.0         22.8      59.0               3758.306676
4  2704.0         22.2      61.0               2387.567715
```

It shows how both temperature and humidity together influence load

6.Use of Multiple Linear regression to establish a prediction based on Humidity and temperature

	Actual Load	Temperature	Humidity	Predicted Load
0	2906.0	18.900000	73.000000	3164.153236
1	3251.0	19.400000	68.000000	3603.955053
2	2268.0	16.700000	86.000000	2881.165855
3	3538.0	22.200000	27.000000	3758.306676
4	2493.0	13.966667	90.333333	2387.567715

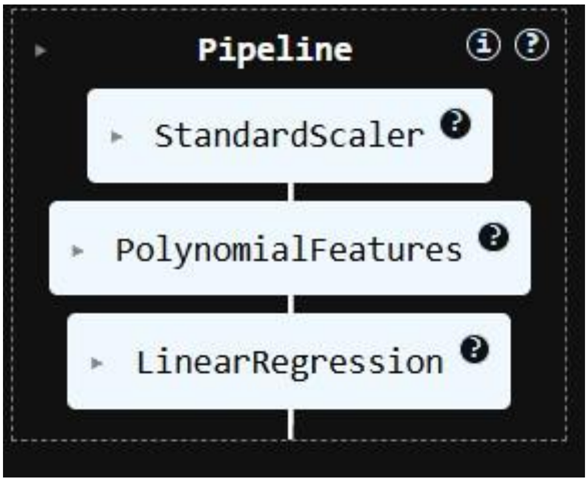
It describes the integration between multiple linear regression and polynomial regression models.

7.Multiple Linear regression Model and polynomial regression model to develop the absolute prediction (FINAL MODEL)

	Actual Load	Temperature	Humidity	Predicted Load	Polynomial Predictions	Best Predictions
43696	2906.0	18.900000	73.000000	134.050000	3027.682218	3164.153236
44286	3251.0	19.400000	68.000000	133.050000	3444.625801	3603.955053
34756	2268.0	16.700000	86.000000	137.900000	2769.483754	2881.165855
28916	3538.0	22.200000	27.000000	97.900000	3715.382371	3758.306676
6458	2493.0	13.966667	90.333333	139.266667	2410.088430	2387.567715
...	...	...	...	...	...	...
11308	2312.0	15.600000	83.000000	136.200000	2590.984840	2652.682675
44864	3101.0	20.600000	61.000000	119.450000	2850.128311	2920.066932
38264	2750.0	17.200000	60.000000	114.900000	2880.889679	3097.553869
860	4356.0	24.400000	58.000000	125.300000	4318.963871	3961.492431
15836	4009.0	26.700000	20.000000	95.900000	3835.758120	3967.846520

The table shows the actual electricity load values with different prediction models

8.Pipeline between multiple linear regression and polynomial regression model established (Prior to the final model)



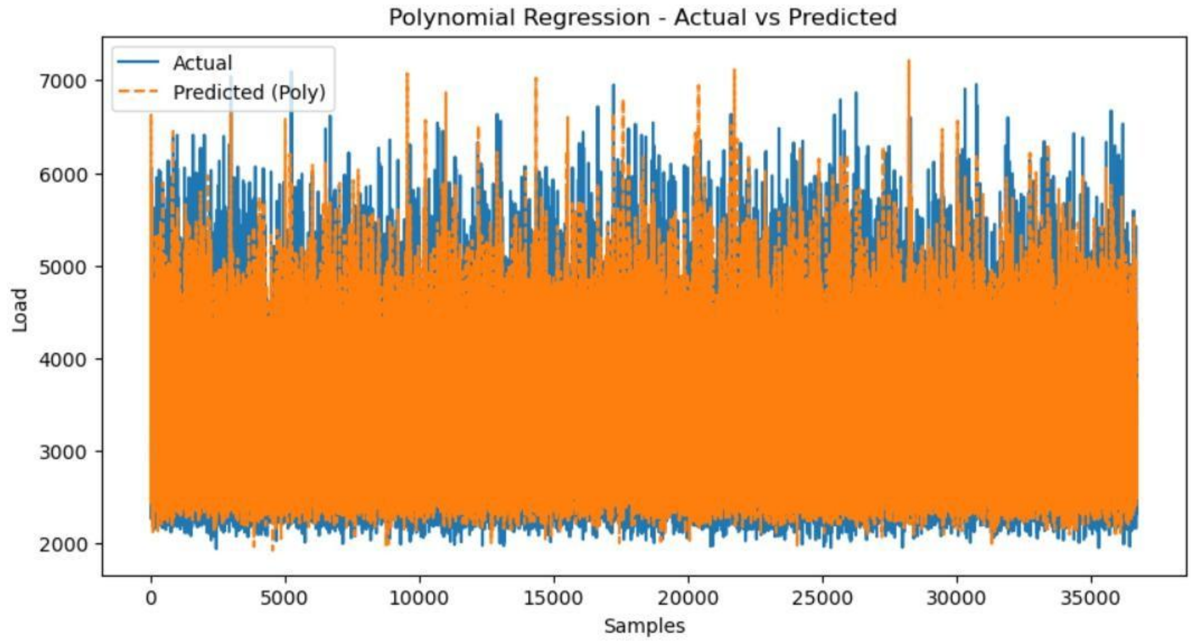
The diagram shows a machine learning pipeline used for building the final prediction model for predicting electricity consumption.

9.Prediction establishment of polynomial model (Establishment)

	Actual Load	Temperature	Humidity	Predicted Load (Poly)
0	2906.0	18.900000	73.000000	3027.682218
1	3251.0	19.400000	68.000000	3444.625801
2	2268.0	16.700000	86.000000	2769.483754
3	3538.0	22.200000	27.000000	3715.382371
4	2493.0	13.966667	90.333333	2410.088430

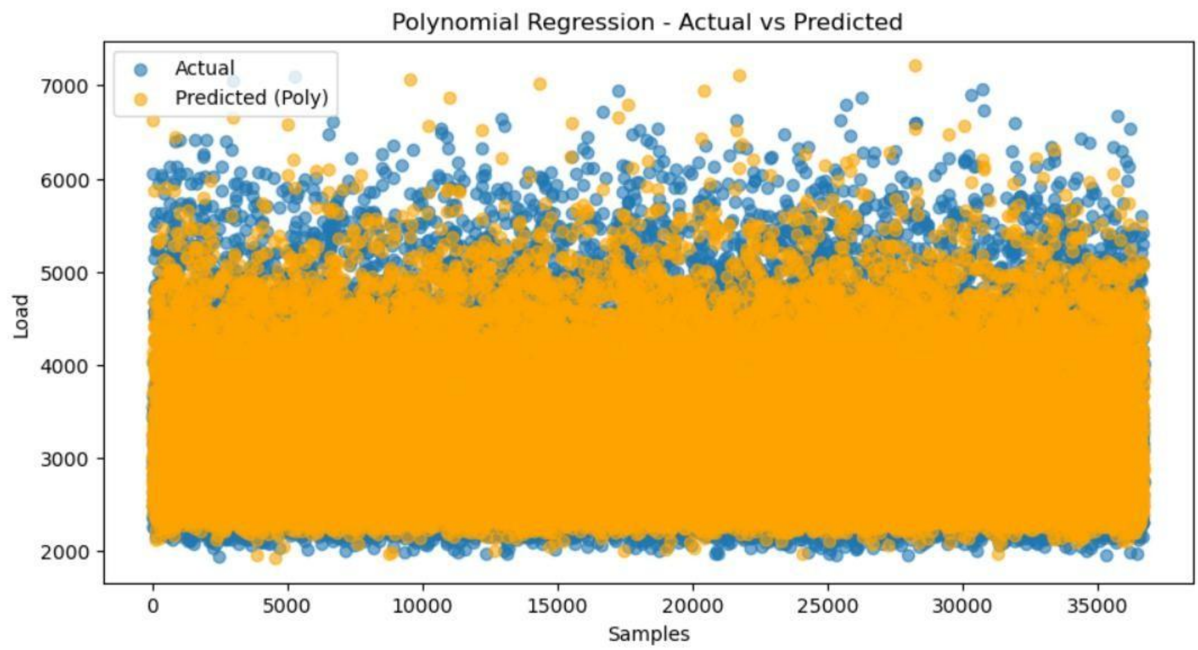
This table compares actual electricity load with predicted load using polynomial regression model alongside temperature and humidity values.

10. Show casing success of polynomial regression in prediction accuracy



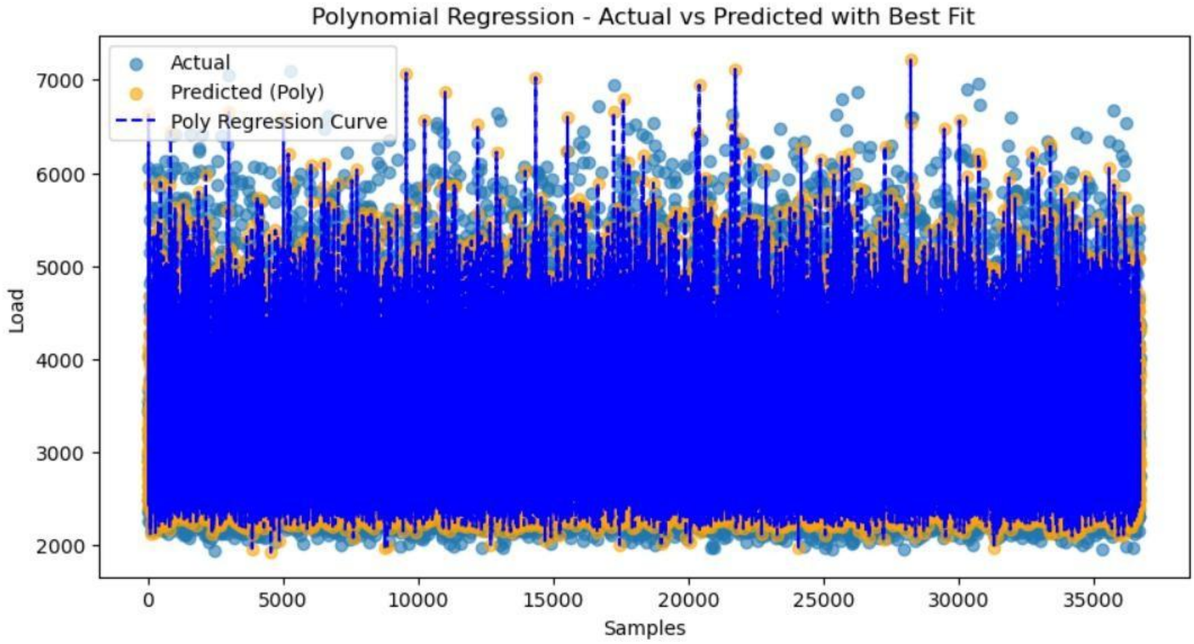
The graph shows a comparison of actual load values with predicted load values over many samples. The blue line represents the actual load data while the orange dashed line represents the load predicted by the polynomial regression model.

11. Polynomial regression prediction accuracy



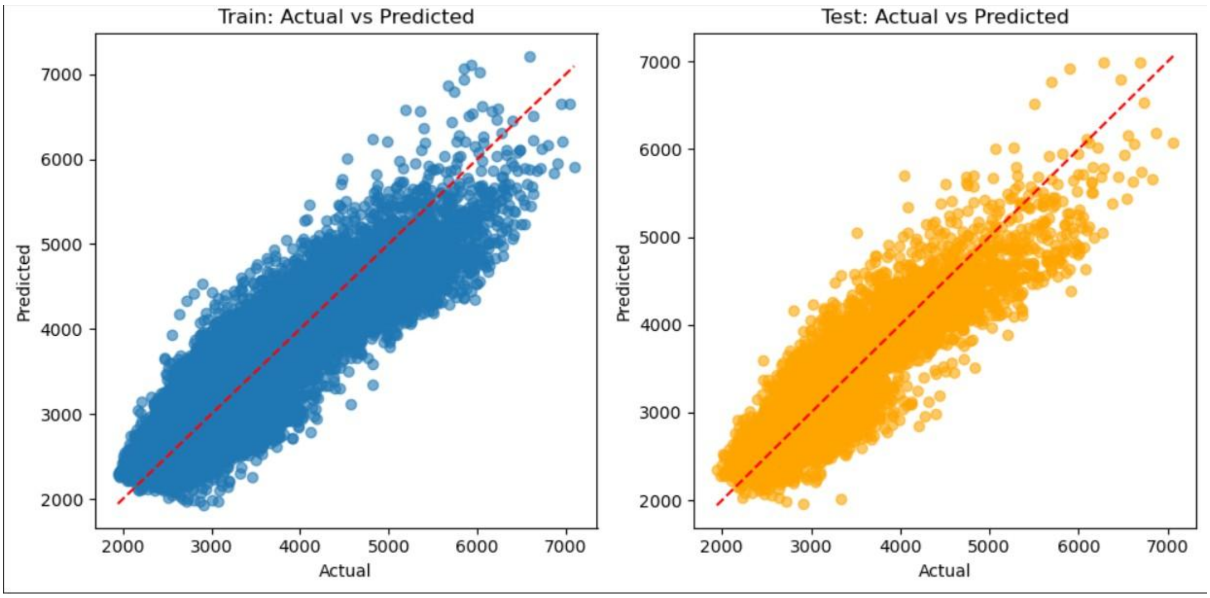
The scatter plot shows a comparison of actual load and predicted load using Polynomial Regression

12.Final regression model prediction accuracy verified



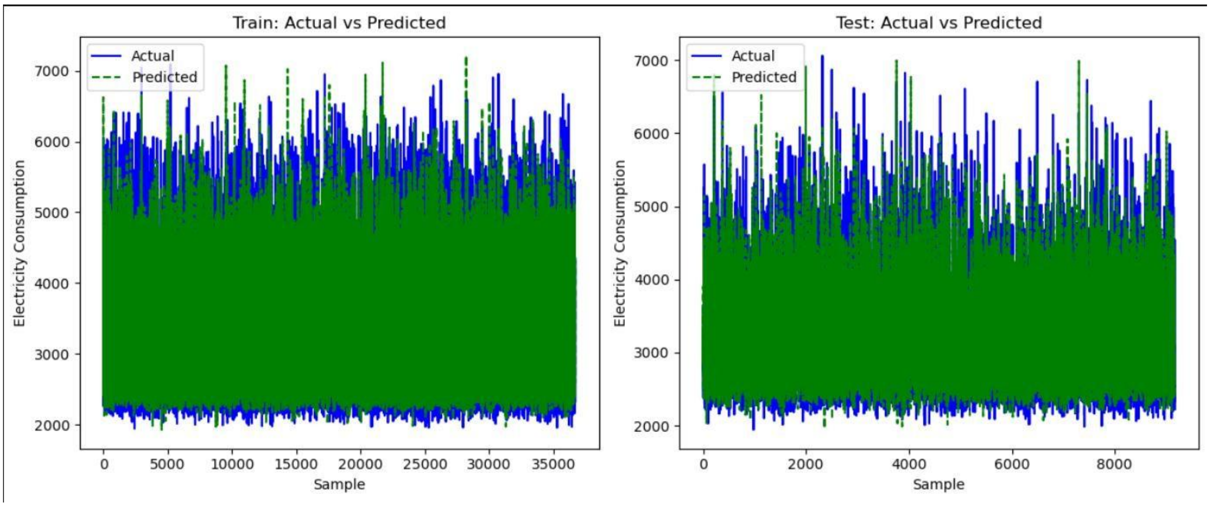
The plot shows that the model prediction is accurate.

13. Actual load consumption vs predicted load consumption



The plot shows that the model prediction is accurate.

14. Actual load consumption and predicted load consumption accuracy insuring trust in the model.



4.2 Linking findings to research questions

1. Which weather conditions influence electricity consumption?

Temperature and humidity influences electricity consumption

2. Can machine learning models predict electricity consumption?

Other models like polynomial regression model can predict electricity consumption

3. Which model can predict electricity consumption using weather data?

Polynomial regression model can predict electricity consumption

4. Which weather variable has the strongest influence in electricity consumption

Temperature has the strongest influence in electricity load as compared to humidity

Chapter 5: Discussion

5.1 Critical Discussion of Findings

Since temperature has a high positive correlation with electrical load, it is the primary driver of energy demand. Lower temperatures result in the usage of heaters, while higher temperatures frequently lead to an increase in the use of cooling devices, especially air conditioners. As such, energy demand is highly sensitive to heat waves and warmer periods. Humidity also showed a statistically significant positive relationship but not as much as temperature. Humidity still plays a role in how much electricity people use. This supports the idea that weather's impact on electricity consumption is not limited to temperature alone.

5.2 Comparison with Previous Studies

The results of this study are largely consistent with findings from prior research on the relationship between weather and electricity demand:

- Meng et al. (2020), in their study of energy usage in residential buildings, found that temperature was the most critical weather-related predictor of electricity consumption. Our findings mirror this result, reinforcing temperature's central role in energy demand forecasting.

- Yang et al. (2024) similarly observed a strong dependence of electricity demand on ambient temperature, especially during summer seasons in Mediterranean cities. This matches the findings in our study, where higher temperatures correlated strongly with increased electricity usage.

- Emenekwe (2022) used panel data analysis and found that both heating and cooling degree days were significantly associated with electricity consumption. While our study focused on average temperature, the concept is comparable, and the results similarly point to a weather-driven energy demand dynamic.

- Woods (2022) analyzed weather and energy consumption in developing countries and noted the compounding role of humidity in driving energy use. This aligns well with our discovery that humidity was a statistically significant predictor, though it is often underrepresented in simpler energy models.

However, unlike many studies that find a role for precipitation, our results showed it to be statistically insignificant. For example, Elsaid et al. (2021) observed a minor but noticeable impact of rainfall on electricity demand, particularly in regions with limited indoor ventilation. The divergence may be due to differences in study location, infrastructure, or cultural behavior around weather events.

In terms of methodology, many earlier studies have also used multiple regression and found it effective, though some have supplemented it with machine learning techniques such as neural networks and support vector machines for improved prediction accuracy. While our study demonstrates high predictive power with linear regression alone, more advanced models could be explored in future work to test for non-linear relationships or complex interactions between variables.

5.3 Explanation of Unexpected Results

While the general findings were aligned with expectations, there were a few results that were either surprising or less impactful than anticipated:

5.3.1 Absence of Interaction Effects

This study assumed a purely additive model, meaning each weather variable independently influences electricity consumption. However, in real-world conditions, interactions between variables (e.g., high temperature and high humidity together) may have amplified or non-linear effects.

Chapter 6: Conclusion and Recommendations

6.1 Summary of Key Findings

This study aimed to predict electricity consumption based on weather conditions using a regression model. The investigation, which was based on secondary data that included daily power use and climatic factors (temperature and humidity), produced a number of significant findings:

- With a substantial positive association, temperature was the most significant predictor of power use.
- **Humidity** also significantly contributed to increased energy demand. These findings confirm that weather conditions, particularly temperature and humidity, are critical determinants of electricity consumption and can be reliably used in forecasting models

6.2 Contributions to Knowledge

This research contributes to both the academic and applied understanding of energy consumption patterns in the following ways:

- **Model validation**: It affirms that polynomial regression is a viable and interpretable method for predicting electricity demand based on environmental factors.
- **Variable emphasis**: It underscores the importance of including temperature and humidity as predictors

6.3 Practical Implications

The results of this study have several practical applications, especially for stakeholders in the energy and urban planning sectors:

- **Energy forecasting and planning**: Utility companies can use weather-based prediction models to anticipate demand spikes and manage supply proactively.
- **Smart grid integration**: The findings support integrating predictive weather models into smart grids for real-time demand-response adjustments.

6.4 Limitations of the Study

While the study yielded significant findings, certain limitations must be acknowledged:

- Geographical scope: analysis focuses on a city data in the United States, limiting the generalization of other areas with different climates, infrastructure and behaviours.

- Simplifying model structure:The use of linear model suggests that additives and linear relationships, unable to grasp complex and nonlinear interactions (for example, extreme temperature threshold).

- Limitations:variables that are likely to cause other influence, such as solar radiation, rainfall and wind speed have been introduced.

- Secondary data constraints:Research entirely based on secondary data, may be limited to accuracy, complete or consistent. These limitations show that the results must be explained within the scope and context of the study.

These limitations suggest the findings should be interpreted within the defined scope and context of the study.

6.5 Recommendations for Future Research

To rely on the results of this study and meet its limits, the following recommendations are given for future research:

1. Developing geographical scope: Conducting similar studies in different cities

2. Willer variables, such as seasonal holidays, the occupation of development instead of focusing on temperature and humidity.

3. Use automatic learning methods: Discover nonlinear models such as regression of the support vector (SVR), random forest or neurological network to grasp complex interactions between variables and improve performance.

4. Implicated with real -time applications: cooperates with public service providers to implement and evaluate predictable models in the real world of energy management systems.

6.6 Final Thoughts

This study has shown that electricity consumption is closely related to weather conditions, especially temperature and humidity. By taking advantage of statistical models such as polynomial regression, stakeholders can plan more efficiently for sustainable, efficient and resistant energy systems.

References

Garrido-Perez, J.M., Barriopedro, D., Garcí a-Herrera, R. and Ordo n ez, C., 2021. Impact of climate change on Spanish electricity demand. *Climatic Change*, *165*, pp.1-18.

Zhu, J., Dong, H., Zheng, W., Li, S., Huang, Y. and Xi, L., 2022. Review and prospect of data-driven techniques for load forecasting in integrated energy systems. *Applied Energy*, *321*, p.119269.

Yang, S., Jahanger, A. and Awan, A., 2024. Temperature variation and urban electricity consumption in China: Implications for demand management and planning. *Utilities Policy*, *90*, p.101782.

Momani, M.A. and Alhmoud, L., 2023. Statistical modeling of the determinants driving the electricity demand in Jordan. *IEEE Open Access Journal of Power and Energy*, *10*, pp.294-303.

Moradzadeh, A., Zakeri, S., Shoaran, M., Mohammadi-Ivatloo, B. and Mohammadi, F., 2020. Short-term load forecasting of microgrid via hybrid support vector regression and long short-term memory algorithms. *Sustainability*, *12*(17), p.7076.

Masekoameng, J.L., 2024. Assessing Multiple Regression Analysis (MRA) model fit for forecasting air traffic movements using log transformation: a case study on ATNS air traffic movement dataset during COVID-19 pandemic

Verma, M., Ghritlahre, H.K. and Chandrakar, G., 2023. Wind speed prediction of central region of Chhattisgarh (India) using artificial neural network and multiple linear regression technique: a comparative study. *Annals of Data Science*, *10*(4), pp.851-873

Shome, A., Mukherjee, G., Chatterjee, A. and Tudu, B., 2024. Study of Different Regression Methods, Models and Application in Deep Learning Paradigm. In *Deep Learning Concepts in Operations Research* (pp. 130-152). Auerbach Publications.

Meng, Q., Xiong, C., Mourshed, M., Wu, M., Ren, X., Wang, W., Li, Y. and Song, H., 2020. Change-point multivariable quantile regression to explore effect of weather variables on building energy consumption and estimate base temperature range. *Sustainable Cities and Society*, *53*, p.101900

Yang, S., Jahanger, A. and Awan, A., 2024. Temperature variation and urban electricity consumption in China: Implications for demand management and planning. *Utilities Policy*, *90*, p.101782

Emenekwe, C.C. and Emodi, N.V., 2022. Temperature and residential electricity demand for heating and cooling in G7 economies: A method of moments panel quantile regression approach. *Climate*, *10*(10), p.142

Woods, J., James, N., Kozubal, E., Bonnema, E., Brief, K., Voeller, L. and Rivest, J., 2022. Humidity's impact on greenhouse gas emissions from air conditioning. *Joule*, *6*(4), pp.726-741.

Elsaid, A.M. and Ahmed, M.S., 2021. Indoor air quality strategies for airconditioning and ventilation systems with the spread of the global coronavirus (COVID-19) epidemic: Improvements and recommendations. *Environmental Research*, *199*, p.111314.

Appendix A

A.1 VISUALIZATIONS VIA POWERBI

As the focus was to build a predictive model for electricity consumption via weather condition (Humidity and Temperature).

Data set used to train and test the model was US **US City-Scale Daily Electricity Consumption and WD** (https://www.kaggle.com/datasets/shemantosharkar/us-city-scaledaily-electricity-consumption) specifically LA weather condition data and the load.

After the launch of power BI (N.B. The predictive model was built in Python-jupyter notebook-). Loaded the original dataset LA Merged dataset and the predicted dataset i.e. electricity_predictions_test.

- **Editing and Preparation**

PowerQuery added RowID in both the datasets.

RowID	Date/Time	load	Temperature	Humidity
1	02/07/2015 00:00:00	3448	23.3	60
2	02/07/2015 01:00:00	3141	22.8	61
3	02/07/2015 02:00:00	2935	22.8	57
4	02/07/2015 03:00:00	2802	22.8	59
5	02/07/2015 04:00:00	2704	22.2	61
6	02/07/2015 05:00:00	2700	22.2	66
7	02/07/2015 06:00:00	2820	23.3	57
8	02/07/2015 07:00:00	3041	24.4	54
9	02/07/2015 08:00:00	3368	25	52
10	02/07/2015 09:00:00	3653	25.6	50
11	02/07/2015 10:00:00	3863	26.1	50
12	02/07/2015 11:00:00	4107	28.9	41
13	02/07/2015 12:00:00	4373	29.4	37
14	02/07/2015 13:00:00	4564	31.1	31
15	02/07/2015 14:00:00	4749	30	36
16	02/07/2015 15:00:00	4878	31.1	22
17	02/07/2015 16:00:00	4968	31.7	29
18	02/07/2015 17:00:00	5012	29.4	35
19	02/07/2015 18:00:00	4898	26.7	47
20	02/07/2015 19:00:00	4634	26.1	49
21	02/07/2015 20:00:00	4354	24.4	58
22	02/07/2015 21:00:00	4255	23.3	64
23	02/07/2015 22:00:00	4102	21.7	70
24	02/07/2015 23:00:00	3766	21.1	71
25	03/07/2015 00:00:00	3428	20.6	73
26	03/07/2015 01:00:00	3112	20.6	73
27	03/07/2015 02:00:00	2886	20	76
28	03/07/2015 03:00:00	2709	19.4	79
29	03/07/2015 04:00:00	2606	18.9	81

	RowID	Temperature	Humidity	Actual_Consumption	Predicted_Consumption
1	1	11.7	52	2766	2922.355236
2	2	27.8	29	3254	3906.639419
3	3	9.4	44	3162	2676.287247
4	4	13.15	92	2382	2485.074556
5	5	18.9	75	3652	3598.073229
6	6	20.6	57	3106	2679.714248
7	7	14.4	60	2961	3105.369246
8	8	20	84	3273	3298.922959
9	9	27.2	8.5	3221	2960.141757
10	10	19.4	18	3312	3191.955922
11	11	10	93	3255	2727.214615
12	12	27.8	47	5251	4527.098713
13	13	11.7	71	2573	2338.817023
14	14	13.9	72	2413	2637.463194
15	15	3.9	82	2951	2842.210879
16	16	30	40	3647	4285.688059
17	17	11.1	77	3388	3057.019046
18	18	20	39	2896	2958.221048
19	19	12.1	93.5	2657	2512.612041
20	20	10.6	61	3705	3277.246318
21	21	32.2	27	3874	4259.243994
22	22	15.6	75	3147	3420.538425
23	23	30.6	48	5573	5138.189709
24	24	22.2	35	3618	3799.83332
25	25	16.7	84	2848	2648.473849
26	26	6.1	79	2585	2514.608332
27	27	17.8	84	3658	3533.950846
28	28	13.9	81	2309	2365.73883
29	29	16.1	46	3318	2866.997485

profiling based on top 1000 rows

LA merged dataset

electricity_predictions_te...

This was to allow the creation of a separate table with aligned values.

```
= Table.AddColumn(#"Removed Columns", "Hour", each Time.Hour([#"Date/Time"]), Int64.Type)
```

	RowID	Date/Time	load	Temperature	Humidity	Temperature.1	Humidity.1
1	1	02/07/2015 00:00:00	3448	23.3	60	11.7	
2	2	02/07/2015 01:00:00	3141	22.8	61	27.8	
3	3	02/07/2015 02:00:00	2935	22.8	57	9.4	
4	4	02/07/2015 03:00:00	2802	22.8	59	13.15	
5	5	02/07/2015 04:00:00	2704	22.2	61	18.9	
6	6	02/07/2015 05:00:00	2700	22.2	66	20.6	
7	7	02/07/2015 06:00:00	2820	23.3	57	14.4	
8	8	02/07/2015 07:00:00	3041	24.4	54	20	
9	9	02/07/2015 08:00:00	3368	25	52	27.2	
10	10	02/07/2015 09:00:00	3653	25.6	50	19.4	
11	11	02/07/2015 10:00:00	3863	26.1	50	10	
12	12	02/07/2015 11:00:00	4107	28.9	41	27.8	
13	13	02/07/2015 12:00:00	4373	29.4	37	11.7	
14	14	02/07/2015 13:00:00	4564	31.1	31	13.9	
15	15	02/07/2015 14:00:00	4749	30	36	3.9	
16	16	02/07/2015 15:00:00	4878	31.1	22	30	
17	17	02/07/2015 16:00:00	4968	31.7	29	11.1	
18	18	02/07/2015 17:00:00	5012	29.4	35	20	
19	19	02/07/2015 18:00:00	4898	26.7	47	12.1	
20	20	02/07/2015 19:00:00	4634	26.1	49	10.6	
21	21	02/07/2015 20:00:00	4354	24.4	58	32.2	
22	22	02/07/2015 21:00:00	4255	23.3	64	15.6	
23	23	02/07/2015 22:00:00	4102	21.7	70	30.6	
24	24	02/07/2015 23:00:00	3766	21.1	71	22.2	
25	25	03/07/2015 00:00:00	3428	20.6	73	16.7	
26	26	03/07/2015 01:00:00	3112	20.6	73	6.1	
27	27	03/07/2015 02:00:00	2886	20	76	17.8	
28	28	03/07/2015 03:00:00	2709	19.4	79	13.9	

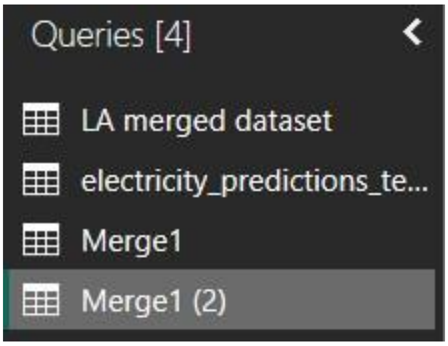
Separated the time into hour, month, day this done by selecting date/time column in the new merged dataset (Merged1 (name))

Day Name	Month Name	Hour
Thursday	July	0
Thursday	July	1
Thursday	July	2
Thursday	July	3
Thursday	July	4
Thursday	July	5
Thursday	July	6
Thursday	July	7
Thursday	July	8
Thursday	July	9
Thursday	July	10
Thursday	July	11
Thursday	July	12
Thursday	July	13
Thursday	July	14
Thursday	July	15
Thursday	July	16
Thursday	July	17
Thursday	July	18
Thursday	July	19
Thursday	July	20
Thursday	July	21
Thursday	July	22
Thursday	July	23
Friday	July	0
Friday	July	1
Friday	July	2
Friday	July	3

And created a new time handling table (Calendar[date/time])

RowID	Calendar[Date/Time]
1	02/07/2015 00:00:00
2	02/07/2015 01:00:00
5	02/07/2015 04:00:00
8	02/07/2015 07:00:00
9	02/07/2015 08:00:00
10	02/07/2015 09:00:00
11	02/07/2015 10:00:00
12	02/07/2015 11:00:00
13	02/07/2015 12:00:00
14	02/07/2015 13:00:00
15	02/07/2015 14:00:00
17	02/07/2015 16:00:00
19	02/07/2015 18:00:00
23	02/07/2015 22:00:00
24	02/07/2015 23:00:00
25	03/07/2015 00:00:00
27	03/07/2015 02:00:00
28	03/07/2015 03:00:00
29	03/07/2015 04:00:00
31	03/07/2015 06:00:00
38	03/07/2015 13:00:00

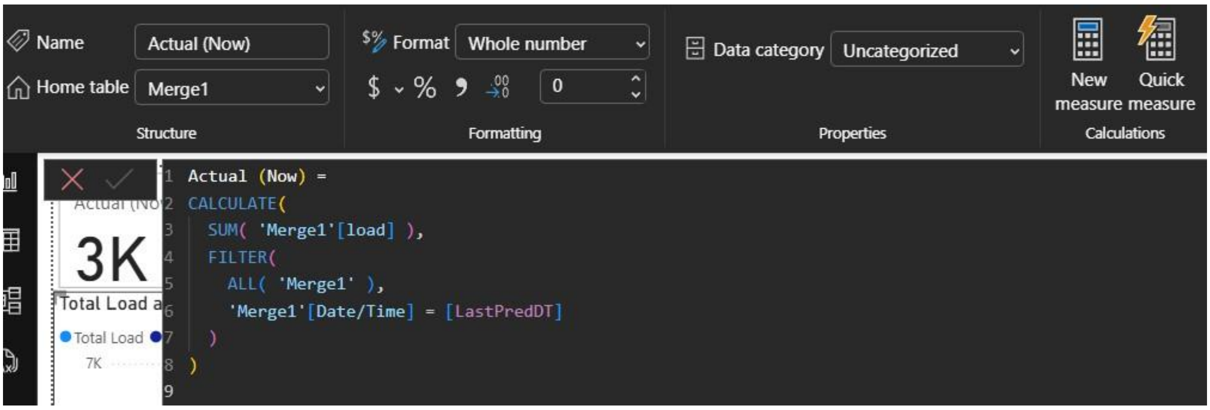
This is for handling time related analytics (in displaying that is).



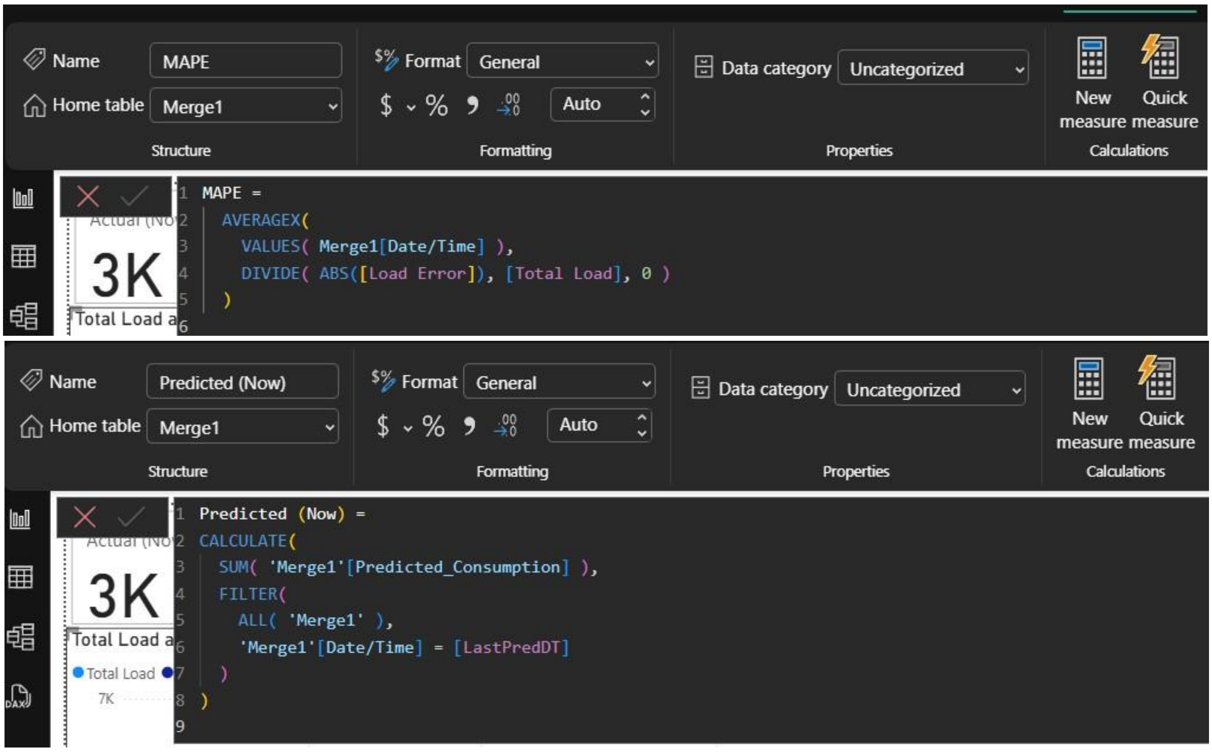
- **Dashboard Visualization**

Added KPI cards to visualise the most important value
Added new measures and inserted a code to allow view if what should be communicated.

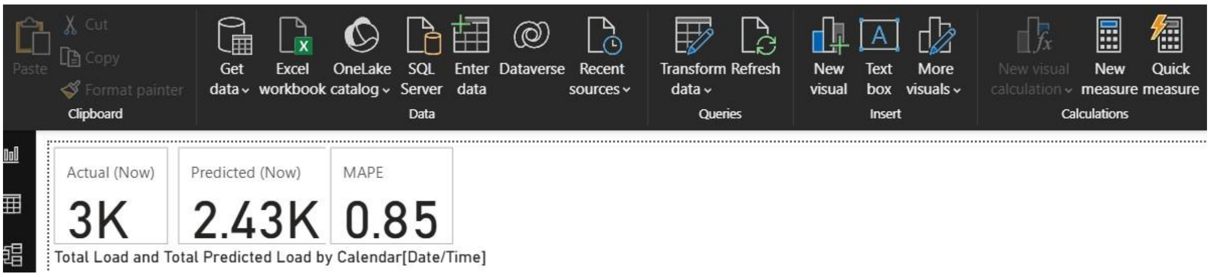
The measure was known as actual now.



Repeated for Predicted now and mean absolute error.



Visual



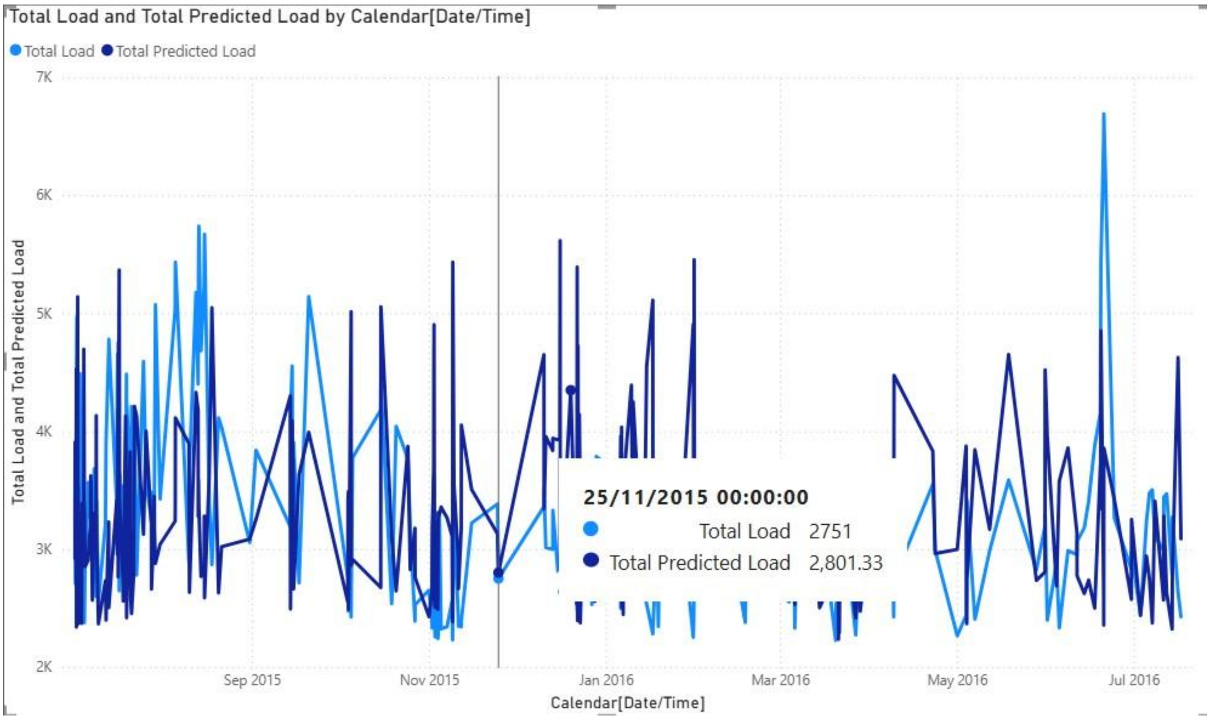
Emphasizing the common value (sample against the predicted value of the model)

3000 and 2430 in electricity consumption in a given day in LA

While the 0.85 is the test for Mean Absolute error, proving all values accurate.

- **Trend Line**

Developed the Line graph to show model prediction from September 2015 till July 2016 against actual electricity consumption.

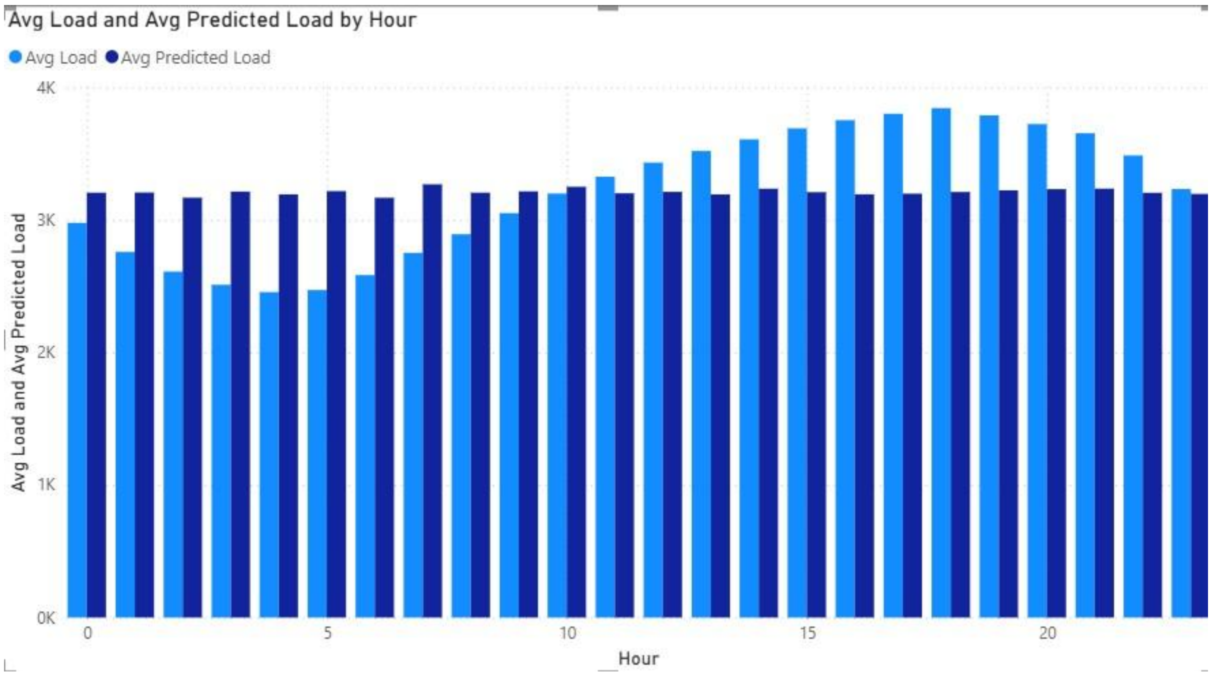


Hovering over a specific timestamp highlights the hour and load predicted vs load recorded (also specific date). Showing the margin of the model's prediction is near accurate.

Specifically, the predictions were narrowed to that period between 2015 and 2016 to see model performance on a modularised state.

The accuracy hinges itself in the alignment of the predicted load (navy blue) and actual load (sky blue).

Hour-of-Day Bar Chart

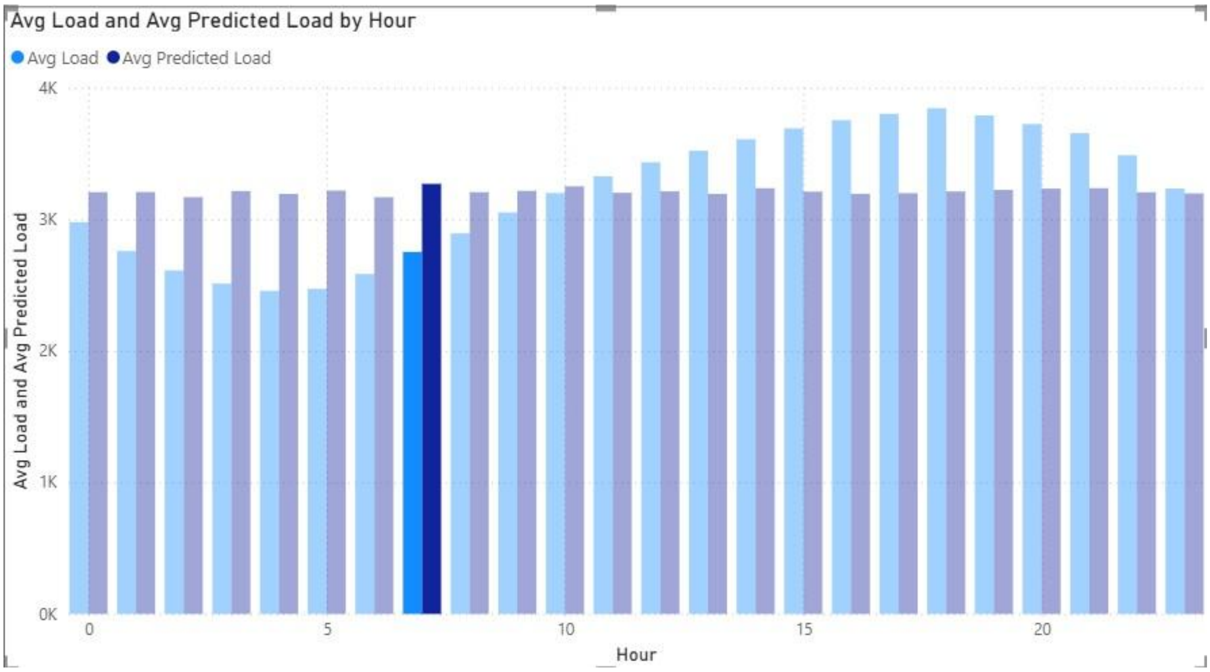


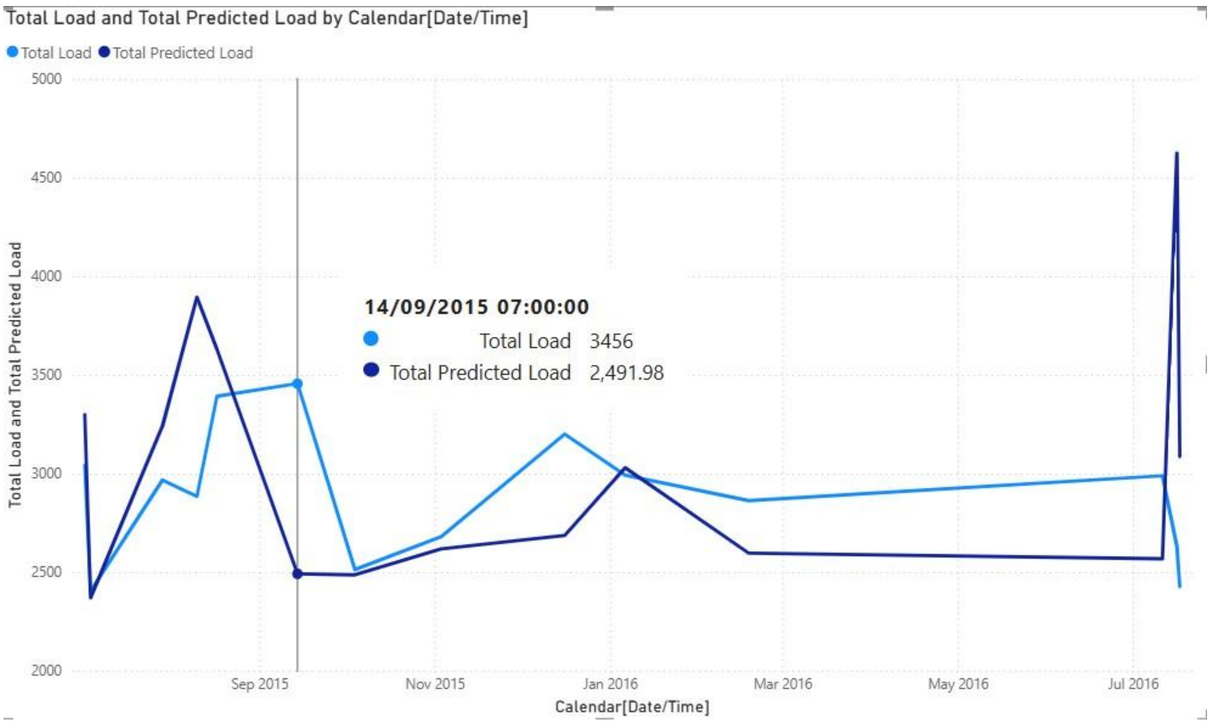
The visualization for hourly prediction against actual hourly consumption for the day.

This helps ensure day to day effects (change in weather condition) and how much consumption is being recorded at which hour.

It also shows how the interaction between the bar (responsible for hourly electricity consumption) corresponds to the line and donut graph (coming next).

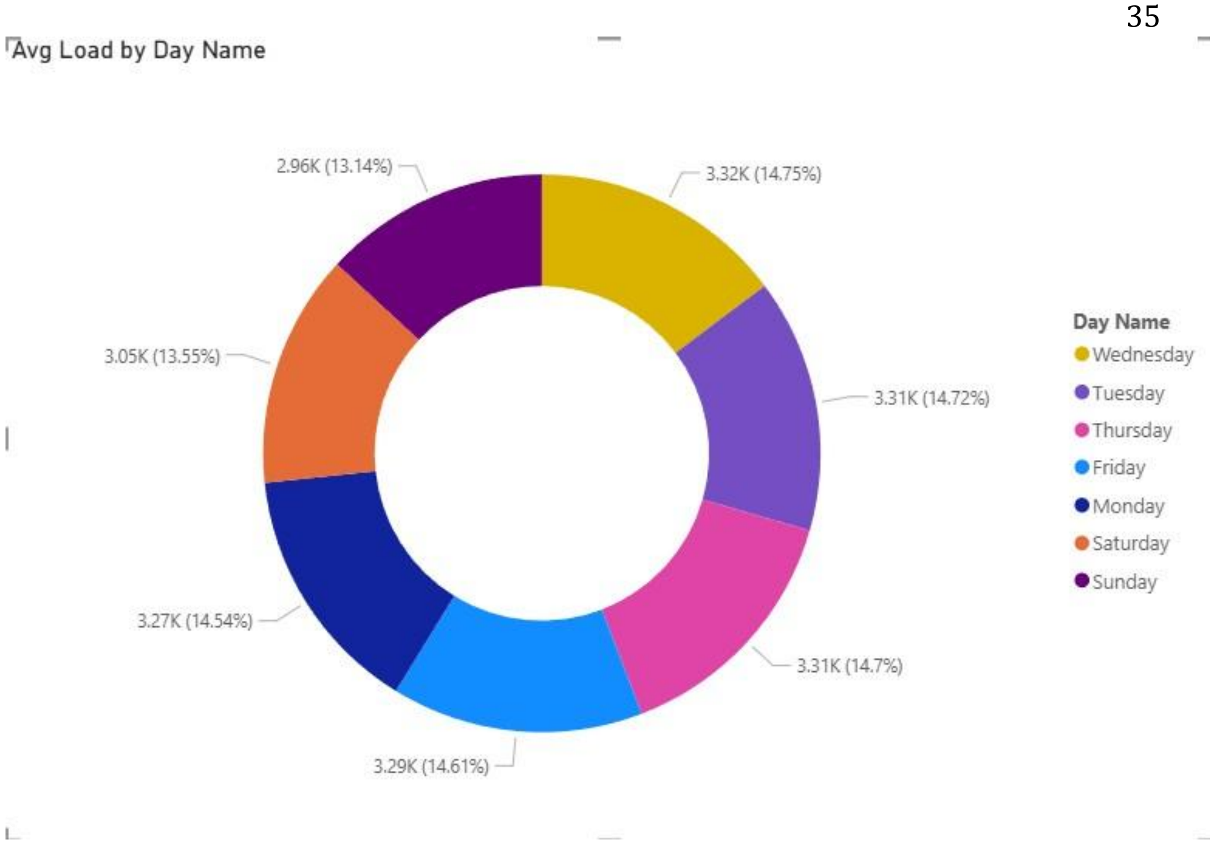
Example:





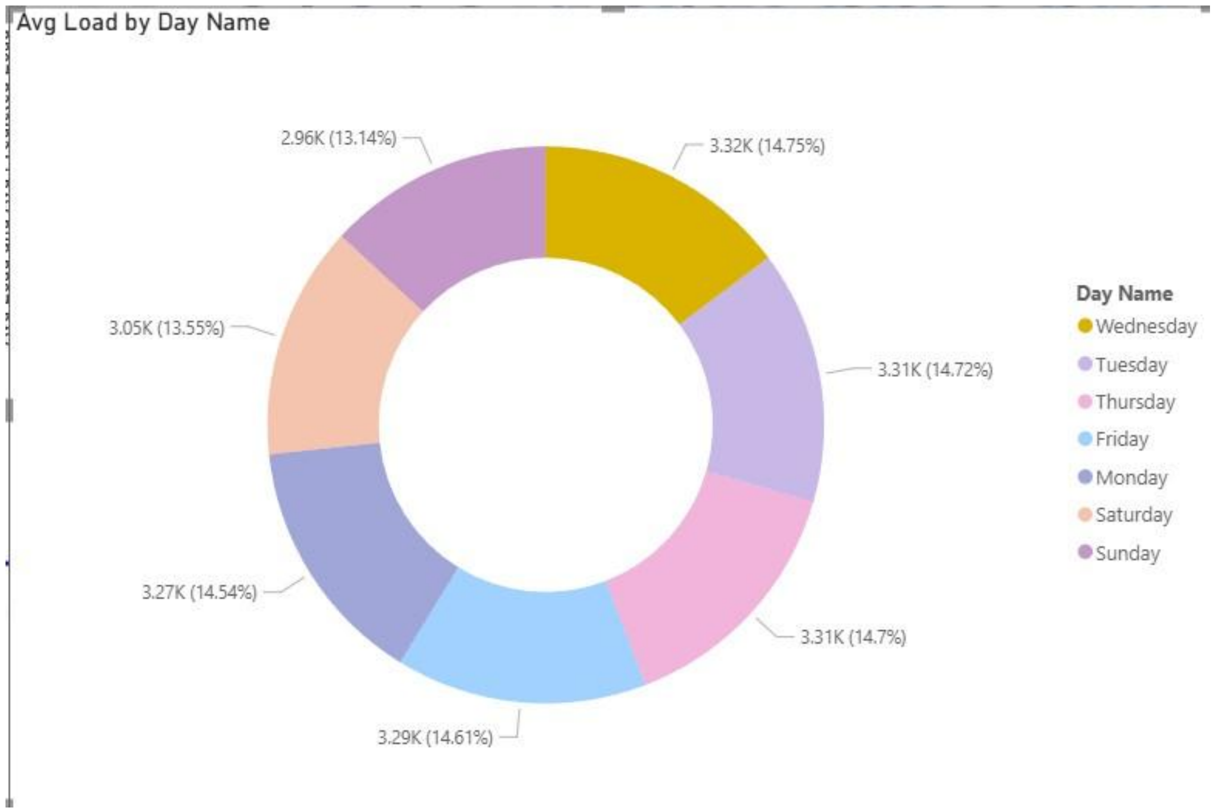
This shows interaction and response of the visualization as the model can pick specific periods in time (e.g. 0700 as observed in the two visuals above).

- **Day-of-Week Donut**

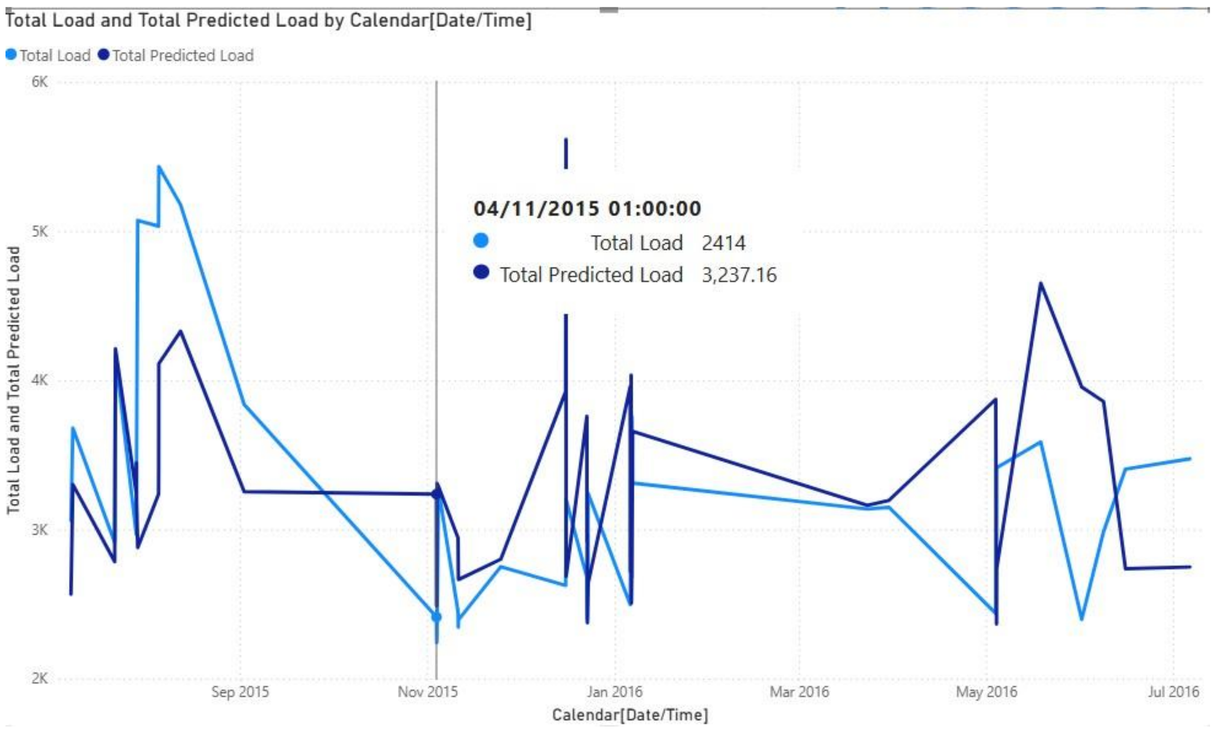


The donut chart depicts the consumption rate of electricity as per weekday. As seen in the chart it denotes percentages of average consumption rates as per weekday and one can compare outside factors such as working days (weekdays) vs weekends.

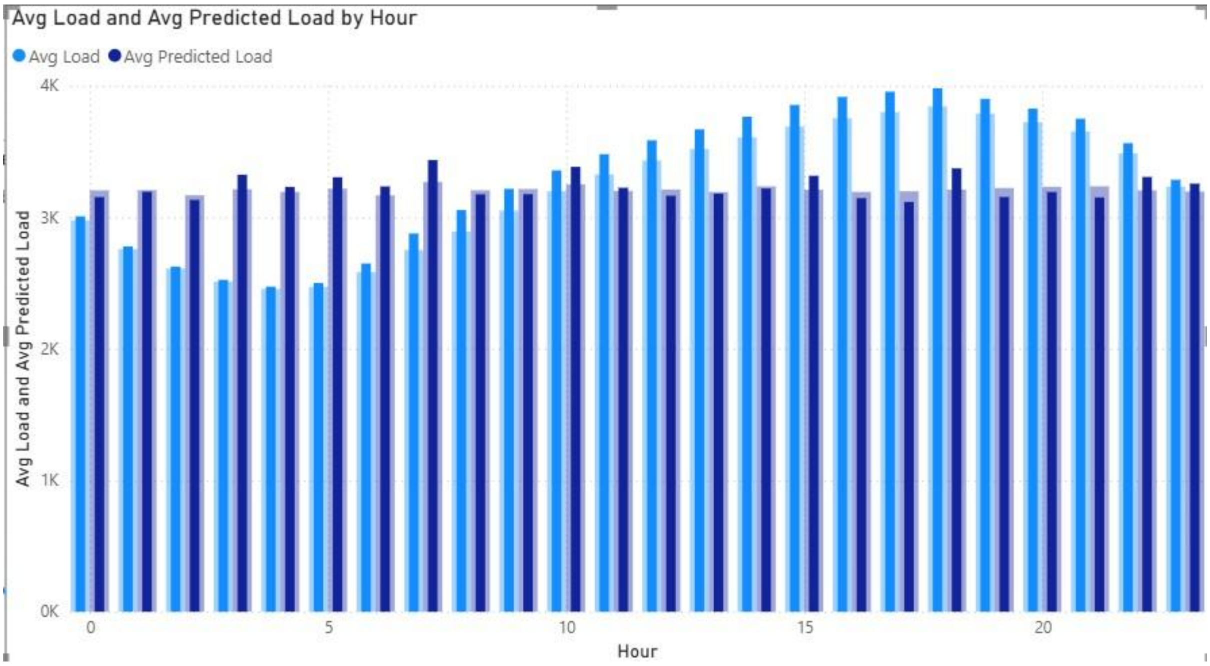
This is very useful in resource utilization and just like the histogram, line graph it directly affects the other to show different utilization of electricity during the same time phases.



Selected Wednesday to observe different consumptions of electricity in the same day across different iterations of time.



Note. The predictive model is being compared to actual load.



Above shows different consumptions of electricity by the hour. Predicitve model against Actual load.

- **Key Findings & Observations**

The model seems to struggle against spikes and unusual spikes. This might be due to unknown factors such as population, supply and demand and economic shift (e.g. stock market and trade).

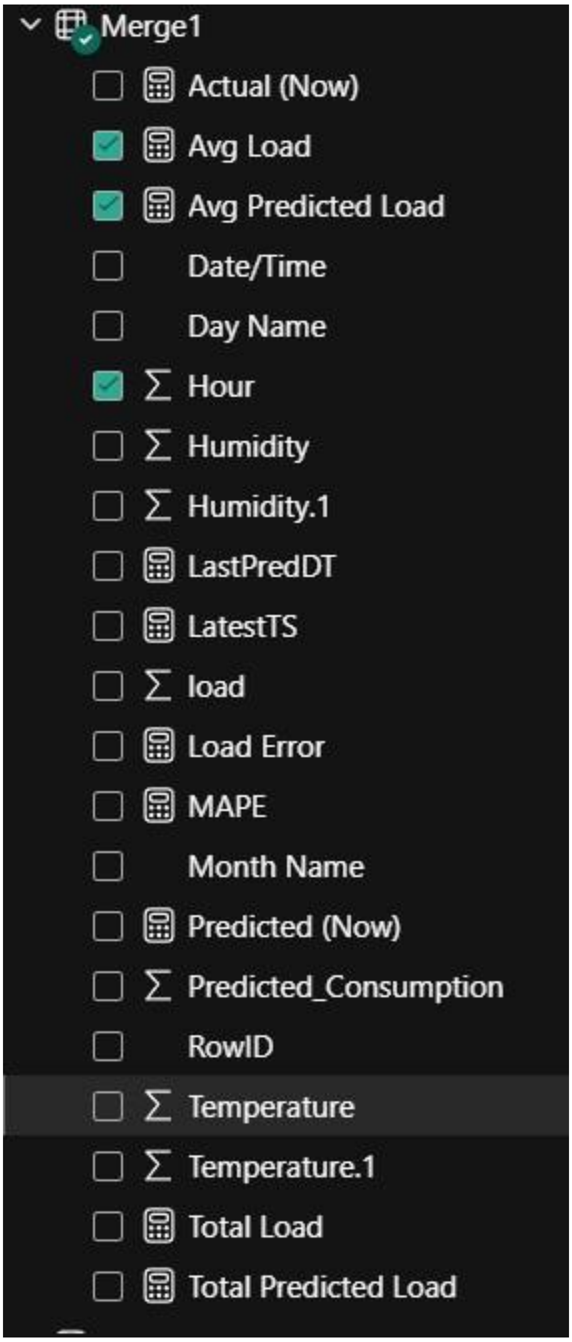
Most day of the week captures seem to be accurate but there are cases were there are shifts.

- **Challenges & Solution**

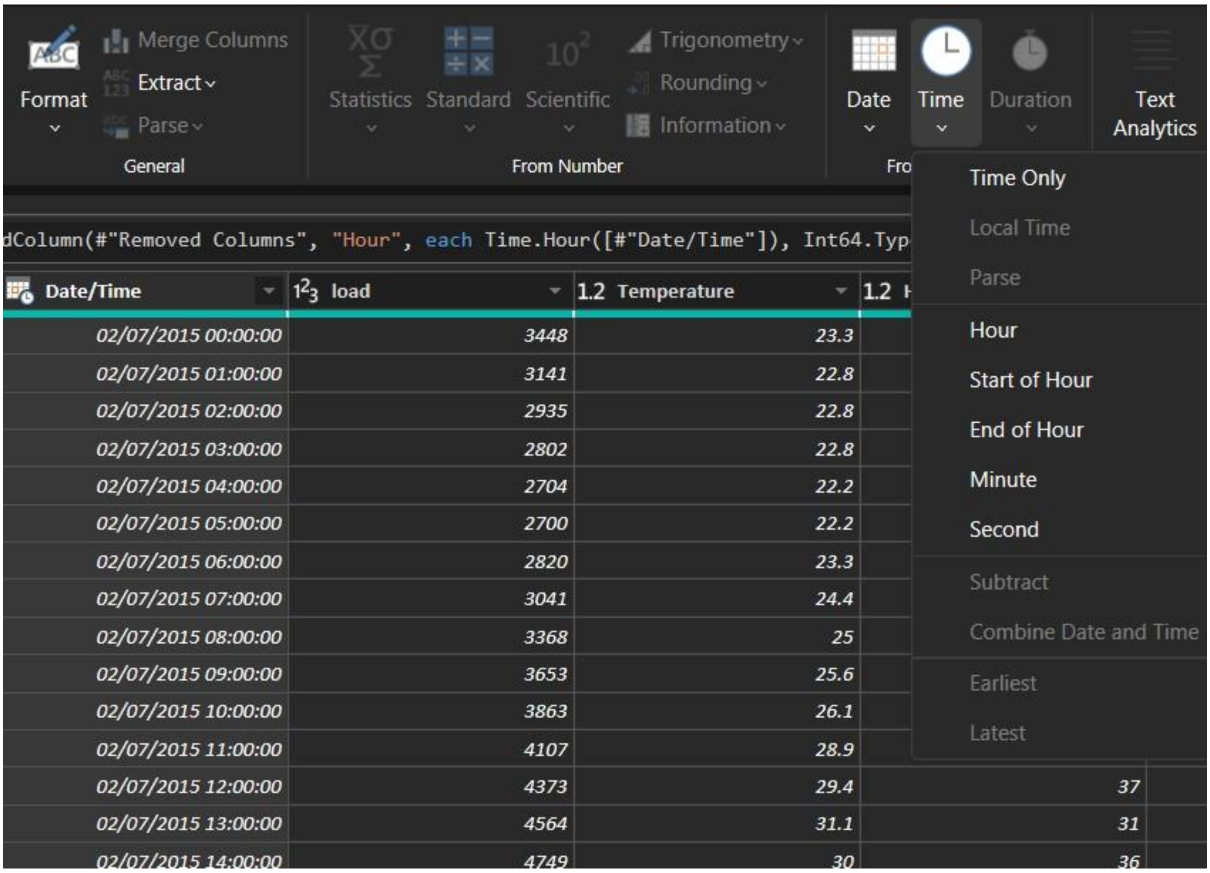
The High Mean Absolute percentage error suggests that we might add more variables to improve the model proficiency e.g. considering holidays, events etc.

Use of further enhancements such as random forest and fuzzy logics to handle unknown unnormalizes.

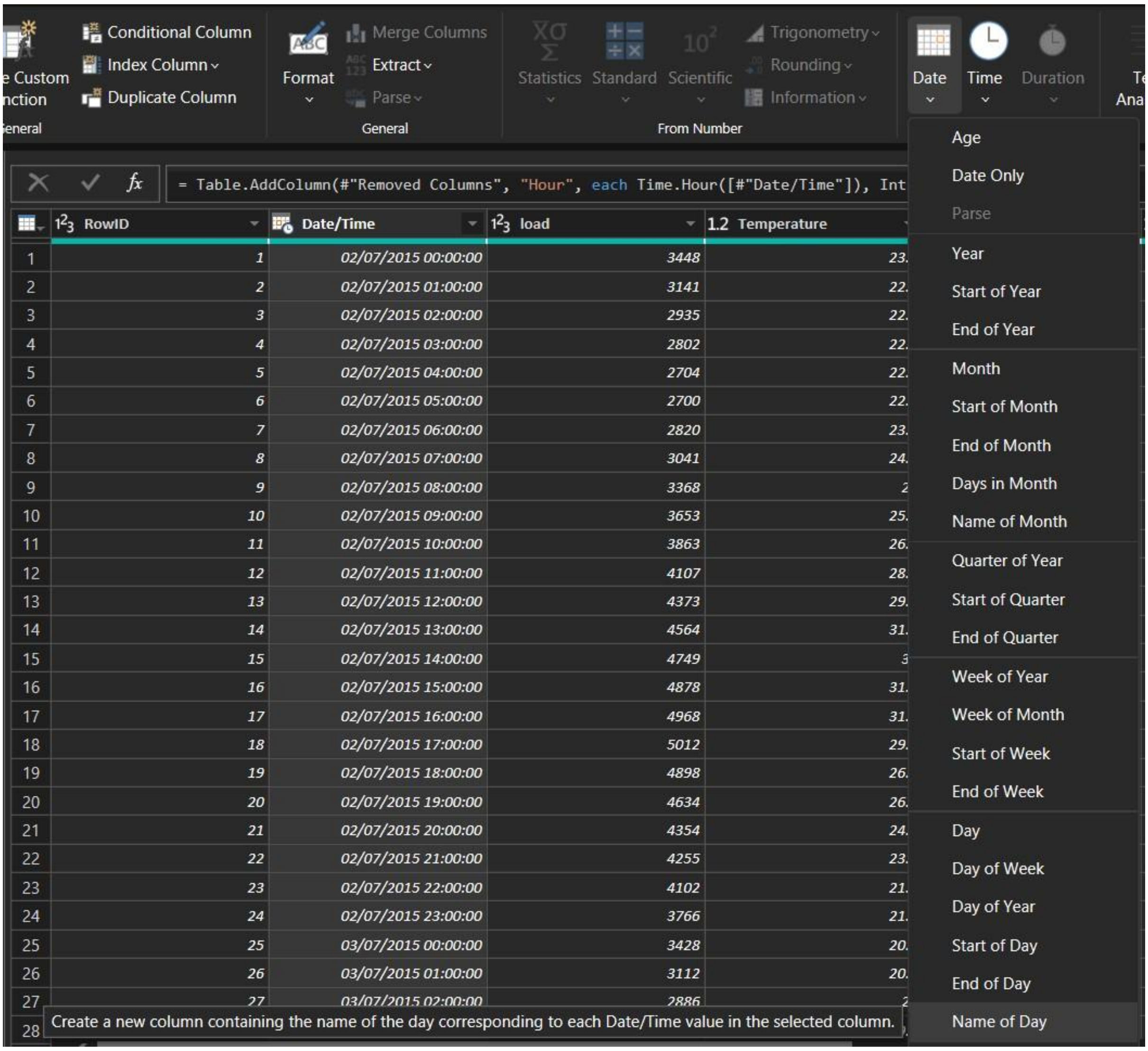
- **Appendix**



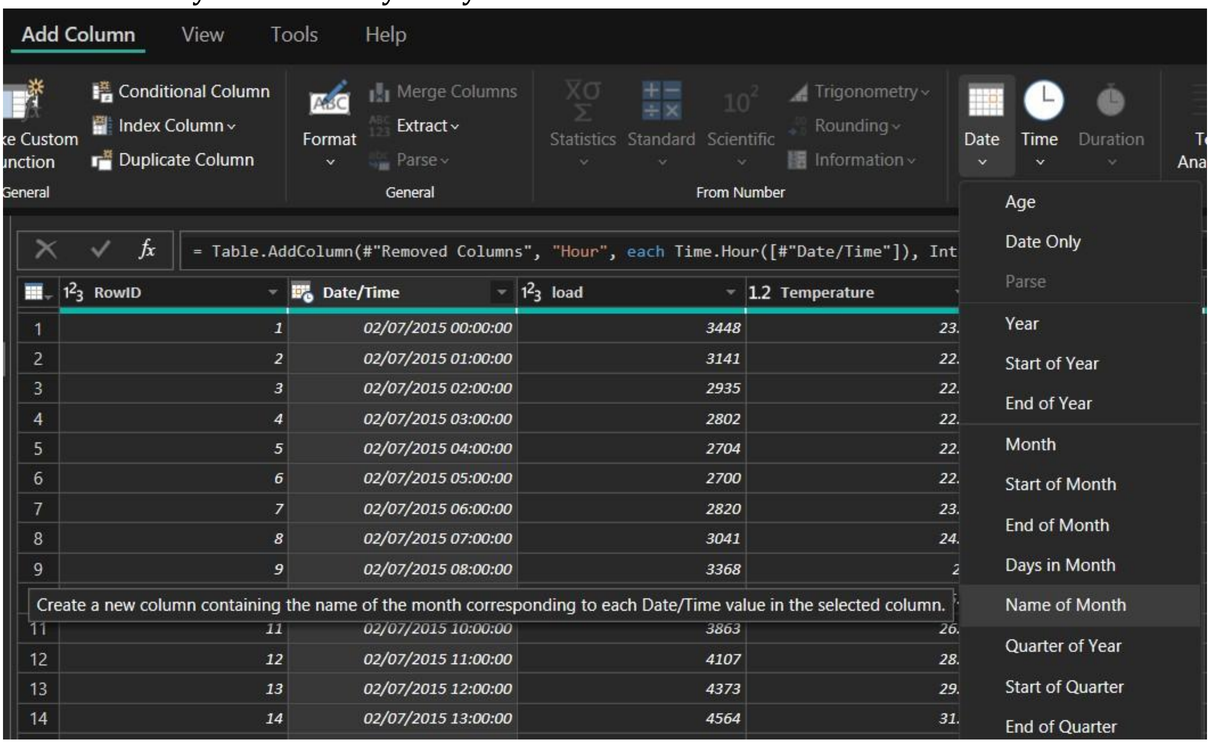
Other variables used to build the final product.



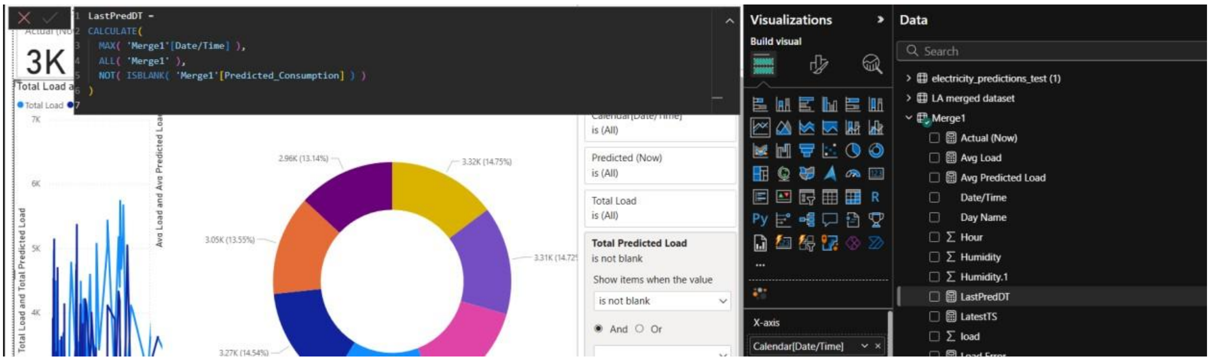
Above is selection by hour to allow hour analysis



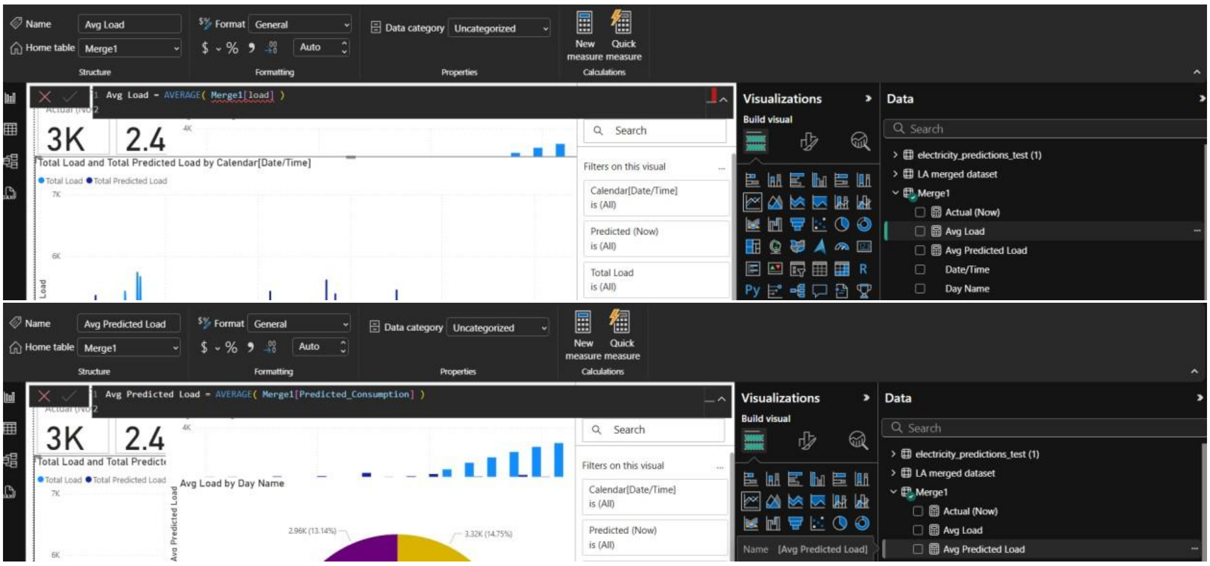
Selection of day to allow daily analysis.



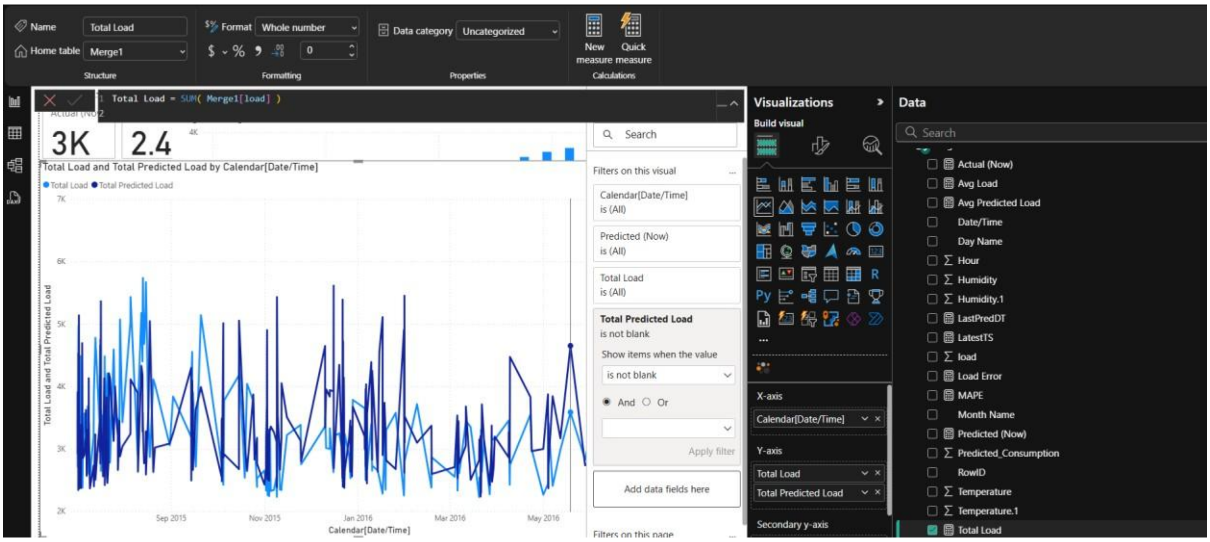
Selection of month to allow monthly analysis.

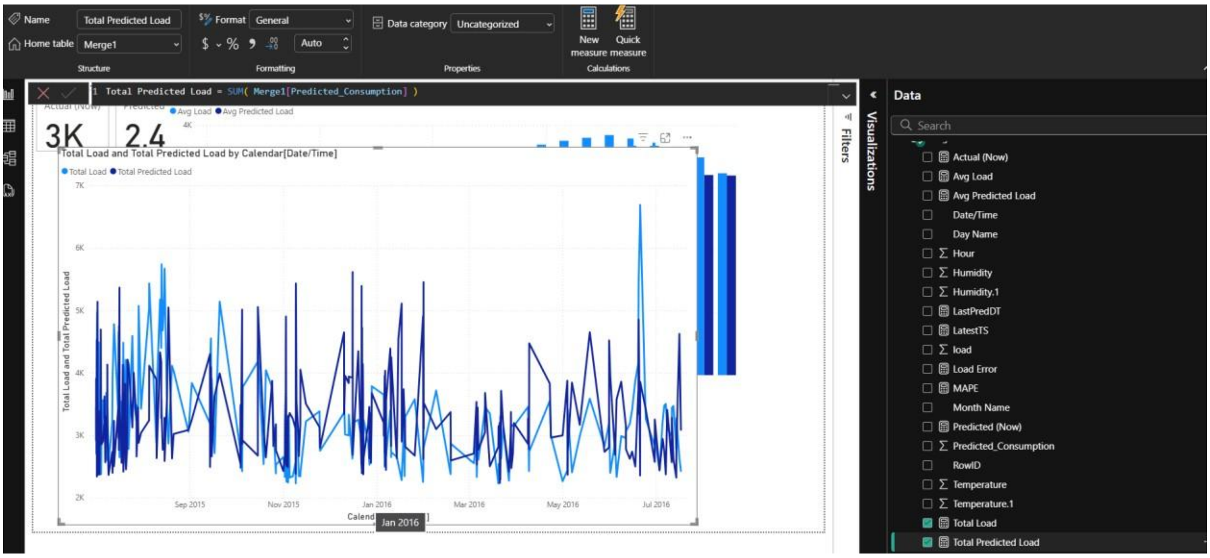


Measure such as Last predicted date and time added to increase precision.



Average load measure was added to increase precision of the histogram and donut chart.





Total load and total predicted load measures were added to create of the line graph.

1.2 A.2 Project Schedule