DSCI521-001

POWER PRODUCTION DATA ANALYSIS

LAUREN MILLER
LUKE CHESLEY
CALEB MILLER
HASHIM AFZAL

CONTENT

OUR TEAM

Caleb Miller

- Background in Analytics
- Skills include:
  - Data Analysis & Visualization
  - Strong Python knowledge

Hashim Afzal

- Background in Biology
- Skills include:
  - Statistical Analysis
  - Study design using Scientific Method

Lauren Miller

- Background in Food Science & Biotechnology
- Skills include:
  - Analytic Thinking
  - Problem Solving
  - Determination

Luke Chesley

- Background in Instrumental Performance & Math
- Skills include:
  - Creativity
  - Strong Python knowledge

ABOUT THE DATA SOURCE

US Energy Information Administration: https://www.eia.gov/

This organization offers free and open data available through an Application Programming Interface (API) and its open data tools.

EIA's API is multi-facetted and contains the following time-series data sets organized by the main energy categories.

THE DATA

- Data was obtained through the EIA's API
- We are analyzing hourly energy consumption data:
  - 2018 - Present
  - We aim to examine patterns within the data, including:
    - trends over different time frames
    - shifts in types of energy consumption.

THE DATA

- Contains over 300,000 rows
  - Can be seamlessly integrated in various applications or investigative efforts
  - The mass volume of data allows for:
    - Machine Learning Applications
    - Expansion of knowledge through valuable insights
  - The depth of the dataset enhances the potential for informed decision making and strategic planning

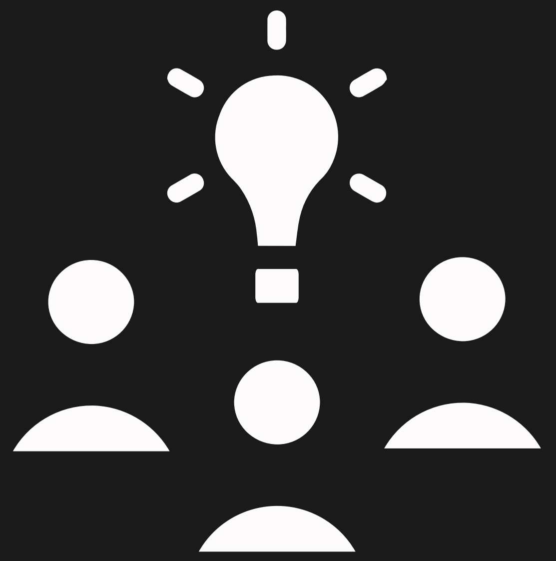
DATA PREVIEW

power_consumption_by_fuel_type

	period	respondent	respondent-name	fueltype	type-name	value	value-units
0	2018-07-01T05	PJM	PJM Interconnection, LLC	COL	Coal	34345.0	megawatthours
1	2018-07-01T05	PJM	PJM Interconnection, LLC	OTH	Other	934.0	megawatthours
2	2018-07-01T05	PJM	PJM Interconnection, LLC	NUC	Nuclear	33643.0	megawatthours
3	2018-07-01T05	PJM	PJM Interconnection, LLC	WAT	Hydro	538.0	megawatthours
4	2018-07-01T05	PJM	PJM Interconnection, LLC	OIL	Petroleum	183.0	megawatthours
5	2018-07-01T05	PJM	PJM Interconnection, LLC	NG	Natural gas	29070.0	megawatthours
6	2018-07-01T05	PJM	PJM Interconnection, LLC	SUN	Solar	1.0	megawatthours
7	2018-07-01T05	PJM	PJM Interconnection, LLC	WND	Wind	2545.0	megawatthours
8	2018-07-01T06	PJM	PJM Interconnection, LLC	SUN	Solar	1.0	megawatthours
9	2018-07-01T06	PJM	PJM Interconnection, LLC	WND	Wind	2401.0	megawatthours
10	2018-07-01T06	PJM	PJM Interconnection, LLC	NG	Natural gas	27542.0	megawatthours
11	2018-07-01T06	PJM	PJM Interconnection, LLC	NUC	Nuclear	33654.0	megawatthours
12	2018-07-01T06	PJM	PJM Interconnection, LLC	OIL	Petroleum	183.0	megawatthours
13	2018-07-01T06	PJM	PJM Interconnection, LLC	OTH	Other	730.0	megawatthours
14	2018-07-01T06	PJM	PJM Interconnection, LLC	WAT	Hydro	488.0	megawatthours
15	2018-07-01T06	PJM	PJM Interconnection, LLC	COL	Coal	32128.0	megawatthours
16	2018-07-01T07	PJM	PJM Interconnection, LLC	OIL	Petroleum	188.0	megawatthours
17	2018-07-01T07	PJM	PJM Interconnection, LLC	WND	Wind	2036.0	megawatthours
18	2018-07-01T07	PJM	PJM Interconnection, LLC	COL	Coal	30084.0	megawatthours
19	2018-07-01T07	PJM	PJM Interconnection, LLC	SUN	Solar	1.0	megawatthours
20	2018-07-01T07	PJM	PJM Interconnection, LLC	NG	Natural gas	25939.0	megawatthours

WHO MIGHT BE INTERESTED?

- Parties who may show interest in this data encompass, but are not confined to:
    - **Power producers:** enhance profitability, understanding energy sources that may pose challenges to sustained profit growth.
    - **Power consumers** (including businesses and individuals): cost-effective resources, monitor personal energy usage
    - **Government:** policy formulation, resource planning, and environmental impact.



POWER PRODUCERS

POWER CONSUMERS

GOVERNMENT

DATA PROCESSING

Accessing Data

The API was accessed through U.S. Energy Information Administration.

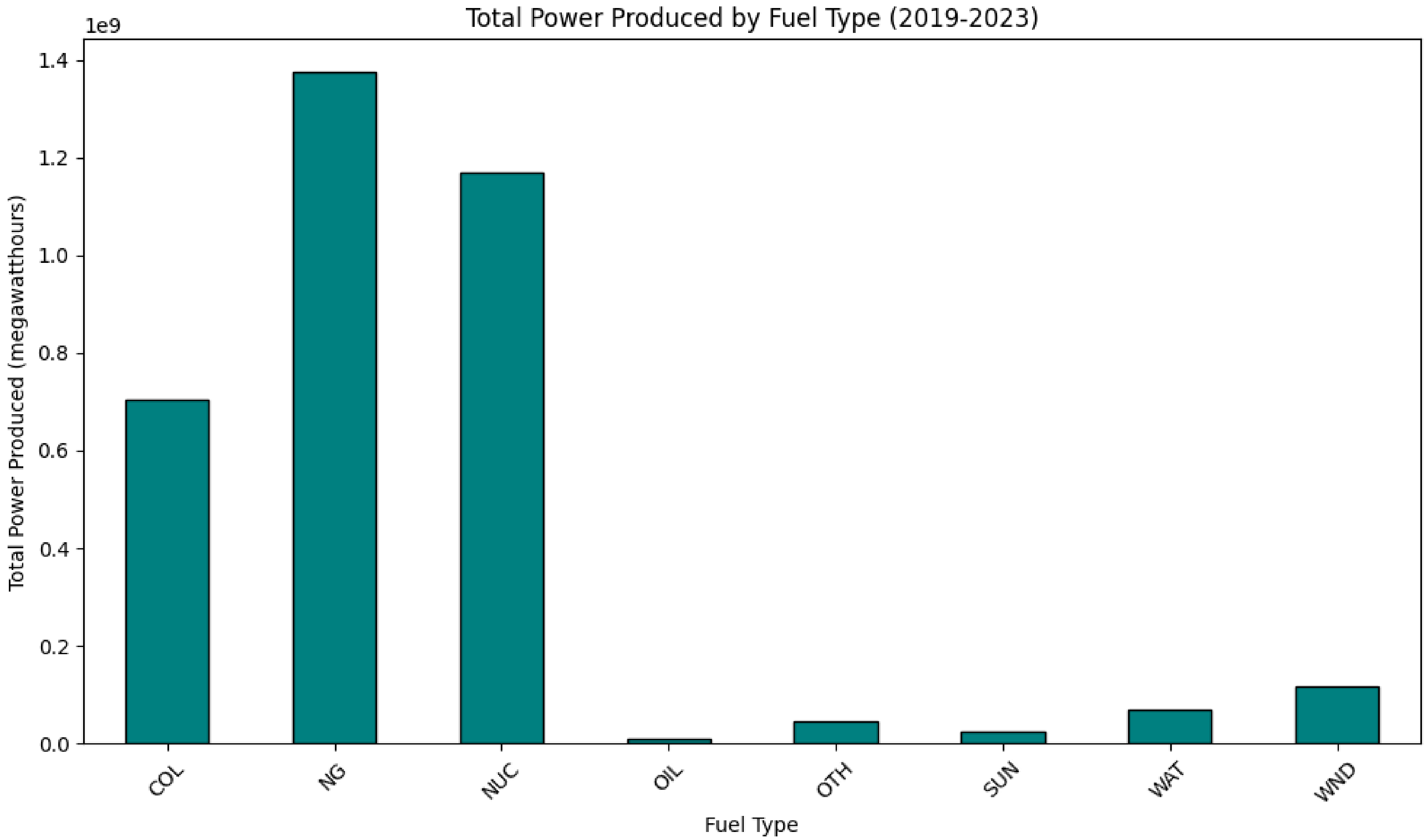
Model Building

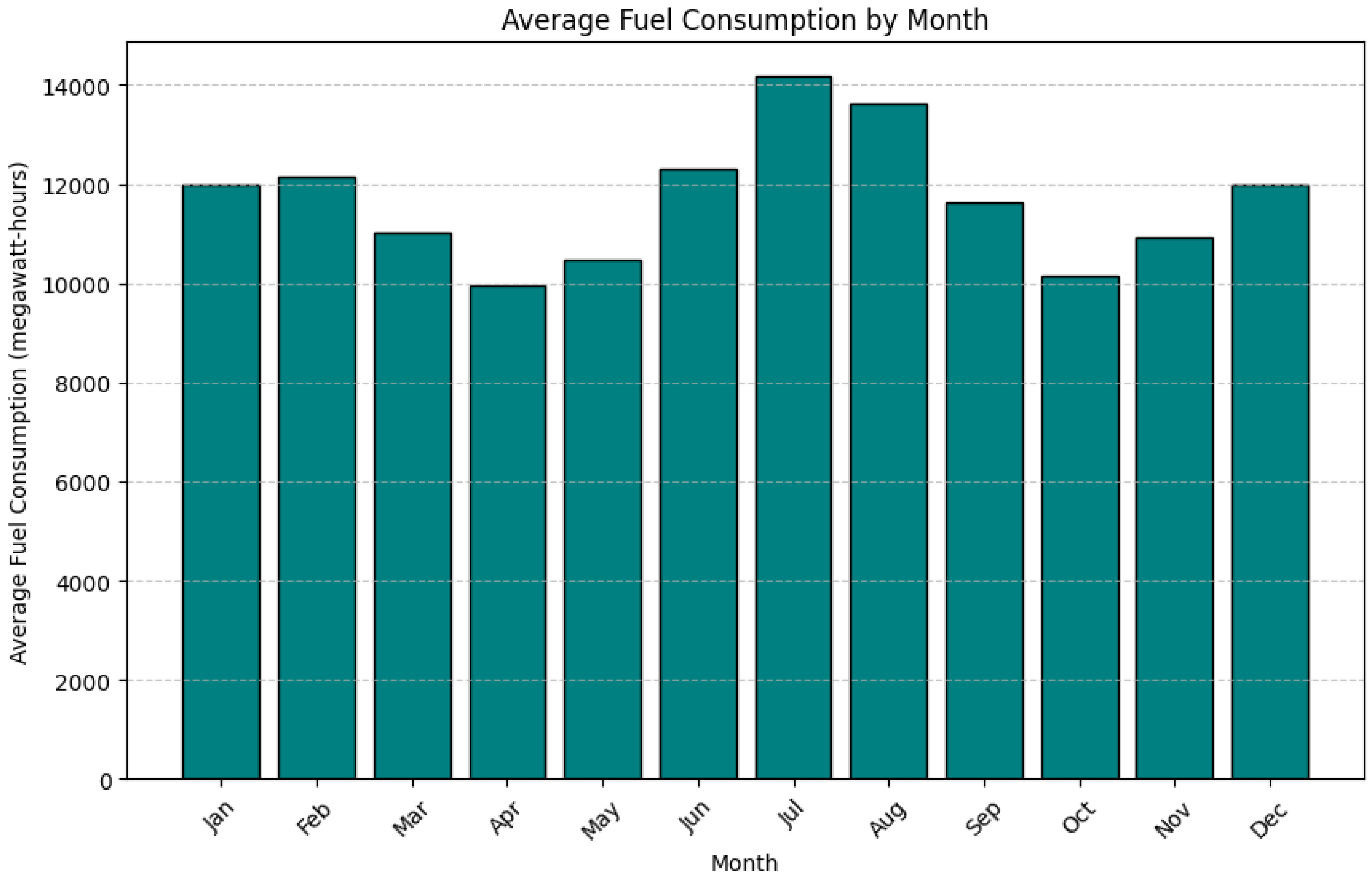
Temporal Fusion Transformer for Interpretable Multi-horizon Time Series Forecasting

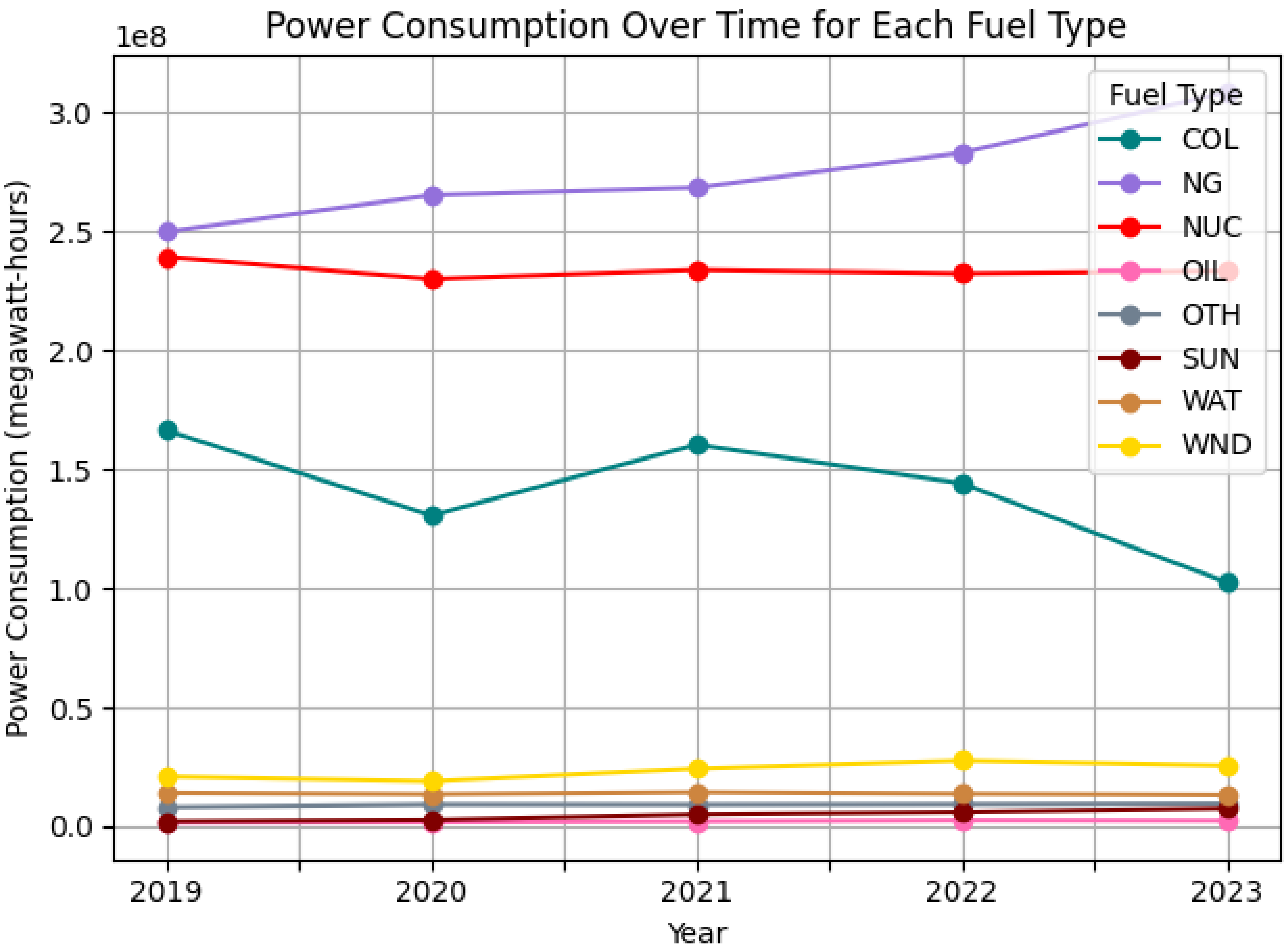
Visualizations

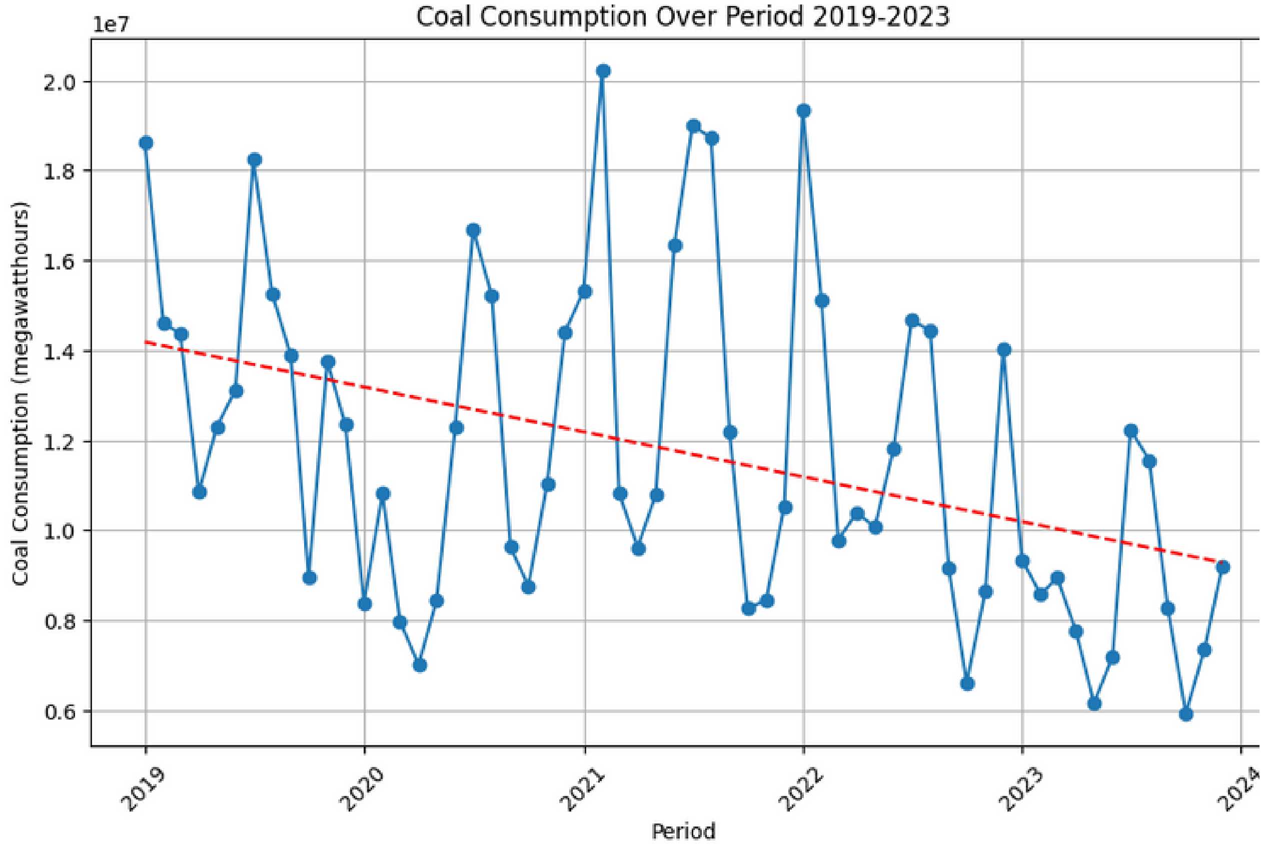
Visualizations created to interpret data and model predictions.

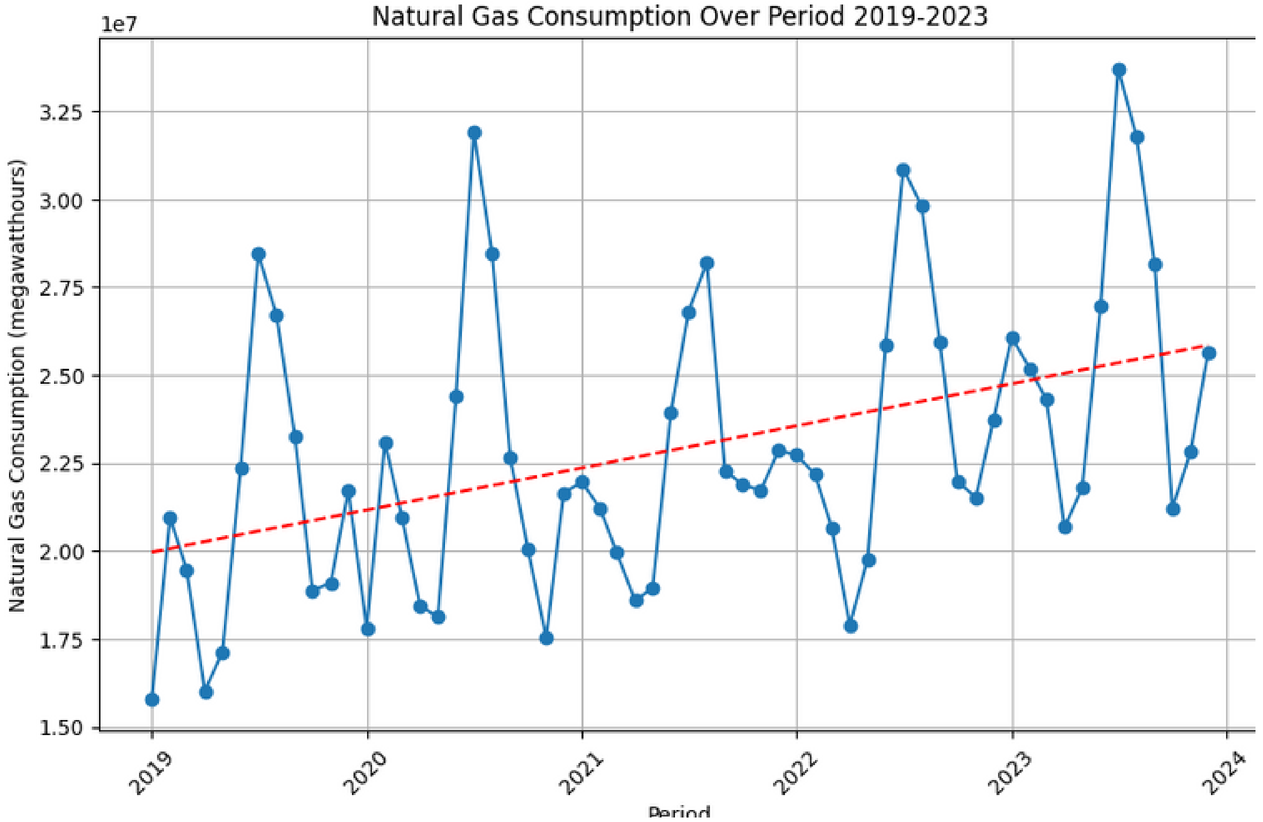
EXPLORATORY DATA ANALYSIS

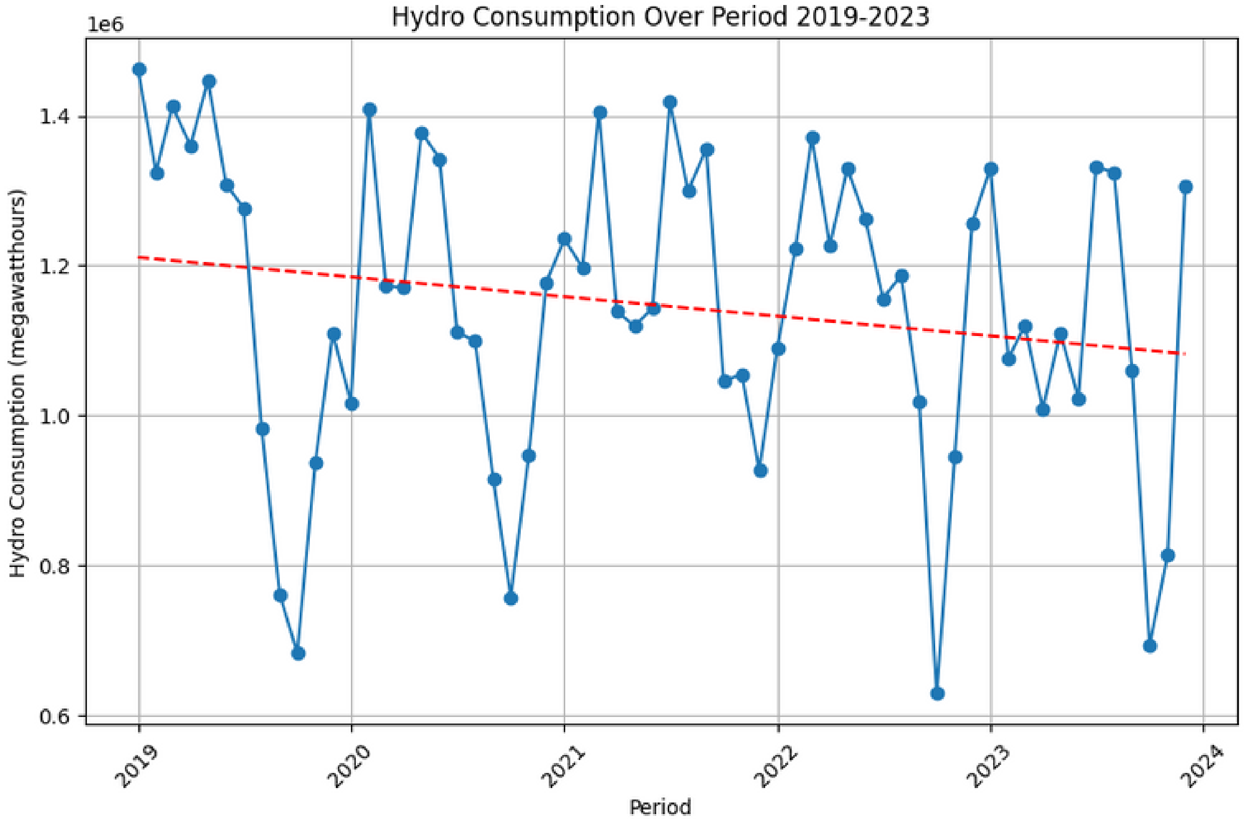


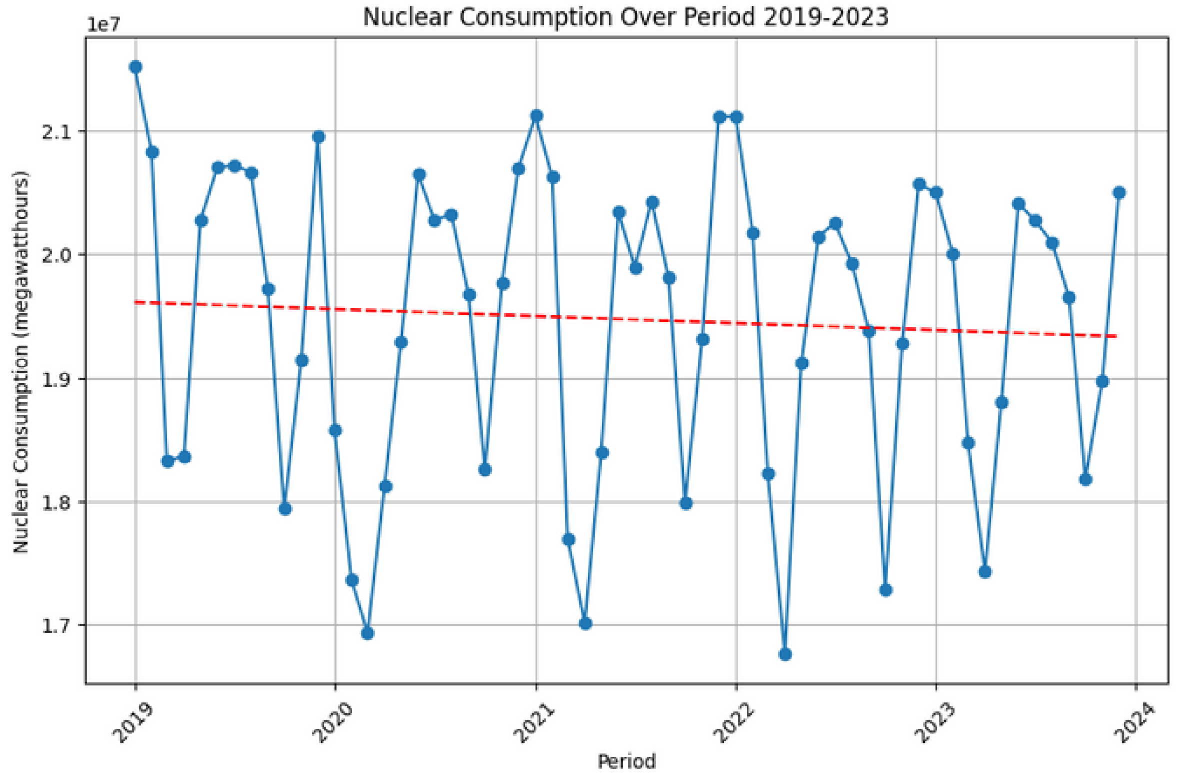


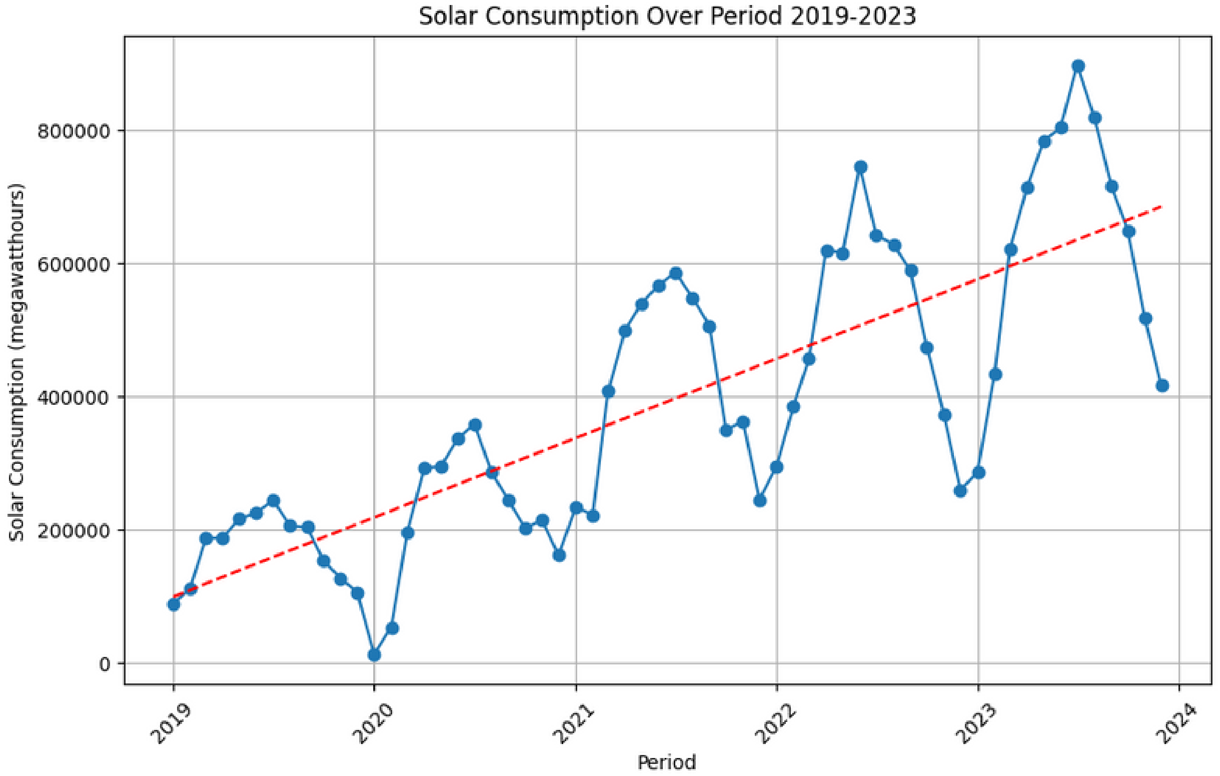

Power Consumption Over Time for Each Fuel Type
1e8
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Power Consumption (megawatt-hours)
2019
2020
2021
2022
2023
Year
Fuel Type
COL
NG
NUC
OIL
OTH
SUN
WAT
WND


Coal Consumption Over Period 2019-2023
1e7
2.0
1.8
1.6
1.4
1.2
1.0
0.8
0.6
Coal Consumption (megawatthours)
2019
2020
2021
2022
2023
2024
Period


Natural Gas Consumption Over Period 2019-2023
1e7
3.25
3.00
2.75
2.50
2.25
2.00
1.75
1.50
Natural Gas Consumption (megawatthours)
2019
2020
2021
2022
2023
2024
Period


Hydro Consumption Over Period 2019-2023
1e6
1.4
1.2
1.0
0.8
0.6
Hydro Consumption (megawatthours)
2019
2020
2021
2022
2023
2024
Period

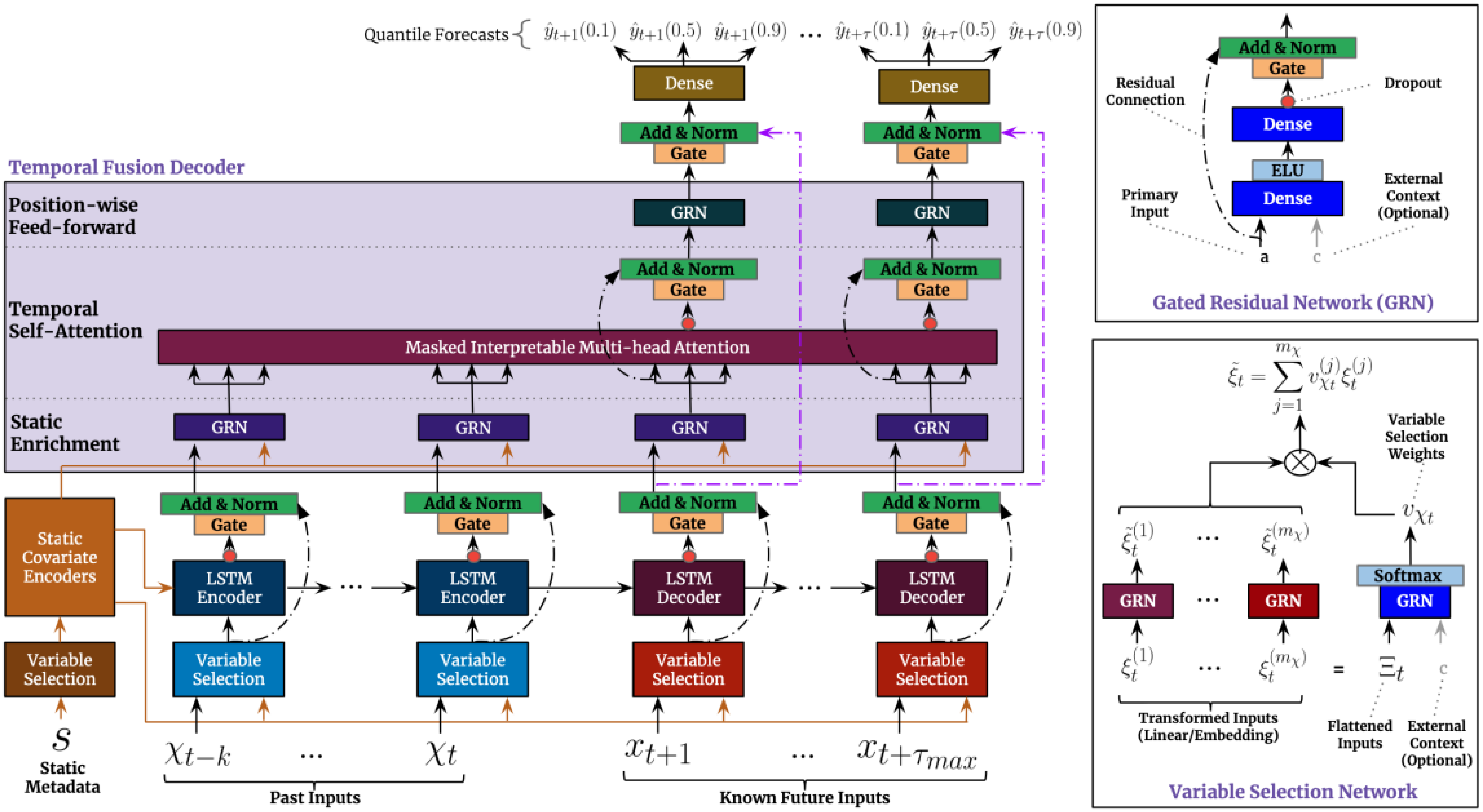

Nuclear Consumption Over Period 2019-2023
1e7
2.1
2.0
1.9
1.8
1.7
Nuclear Consumption (megawatthours)
2019
2020
2021
2022
2023
2024
Period


Solar Consumption Over Period 2019-2023
800000
600000
400000
200000
0
Solar Consumption (megawatthours)
2019
2020
2021
2022
2023
2024
Period

MODEL BUILDING:

TEMPORAL FUSION TRANSFORMER

Architecture



MODEL BUILDING:

TEMPORAL FUSION TRANSFORMER

Feature Creation

This project relies on the creation of new features from existing data. This allows the model to more easily understand data and create an output.

In this dataset, the datetime column is split hours, days of the week, day of year, week of year, month, and quarter. Each variable is transformed to sine and cosine. This created a more robust representation of the same information.

Additionally, the data is grouped by fuel type and the rolling mean and standard deviation for each time stamp and fuel type is calculated. There is one time stamp per hour of each fuel type.

Features before feature creation:
Date (2023-12-31 00:00:00
Fuel type (Oil, Solar, etc...)
Value (Megawatt-hours)

MODEL BUILDING:

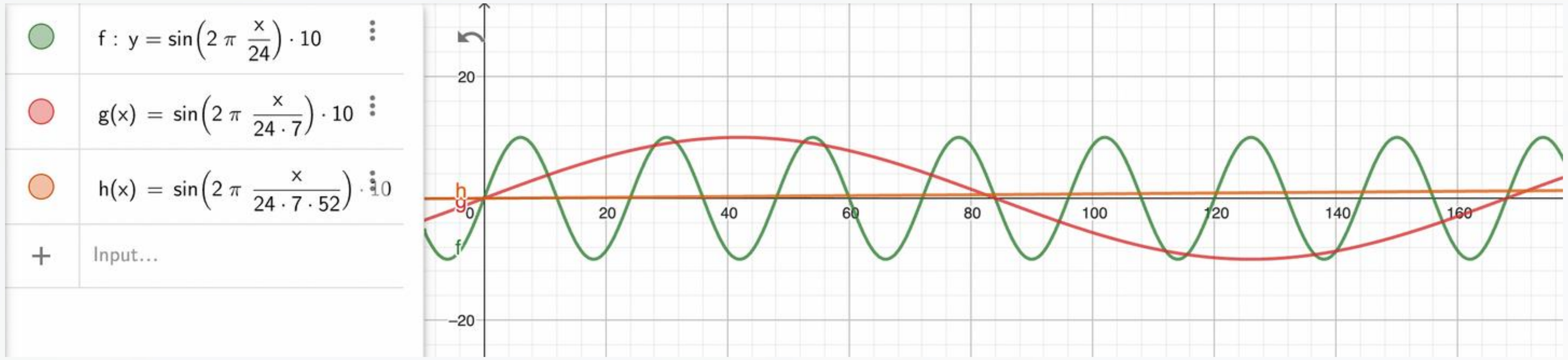
TEMPORAL FUSION TRANSFORMER

Feature Creation

Features after feature creation:

Value, Fuel type

'sin_hour', 'cos_hour', 'sin_day_week', 'cos_day_week', 'sin_day_of_year', 'cos_day_of_year', 'sin_week_year', 'cos_week_year', 'sin_month', 'cos_month', 'sin_quarter', 'cos_quarter', 'year', 'rolling_date_mean_2', 'rolling_date_std_2', 'rolling_date_mean_4', 'rolling_date_std_4', 'rolling_date_mean_12', 'rolling_date_std_12', 'rolling_date_mean_24', 'rolling_date_std_24', 'rolling_date_mean_48', 'rolling_date_std_48', 'rolling_type_mean_2', 'rolling_type_std_2', 'rolling_type_mean_4', 'rolling_type_std_4', 'rolling_type_mean_12', 'rolling_type_std_12', 'rolling_type_mean_24', 'rolling_type_std_24', 'rolling_type_mean_48', 'rolling_type_std_48', 'shifted_168', 'shifted_730', 'shifted_8760',



MODEL BUILDING:

TEMPORAL FUSION TRANSFORMER

Variable Selection Network (VSN)

Variable selection network to **learn** and select relevant input variables at each time step.While multiple variables may be available, their relevance and specific contribution to the output are typically unknown. TFT is designed to provide instance-wise variable selection through the use of variable selection networks applied to both static covariates and time-dependent covariates.

Beyond providing insights into which variables are most significant for the prediction problem, variable selection also allows TFT to remove any unnecessary noisy inputs which could negatively impact performance.

MODEL BUILDING:

TEMPORAL FUSION TRANSFORMER

Long short-term memory (LSTM)

The nature of LSTMs(RNN) limit the effectiveness of learning very long term patterns

LSTMs carry an internal state that is updated at each time step. Across very long sequences the signal can degrade

In the context of this model, it learns the short to medium term temporal dependencies on the scale of hours and days

Attention mechanism

Balances with the LSTM, able to directly model relationships over long sequences without signal degradation issues inherent to LSTMs

The attention mechanism allows the TFT to selectively and dynamically attend to specific parts of the input data that are more relevant for making predictions. This means it can identify and weigh more heavily the time steps or features that are most informative for the forecast.

MODEL BUILDING:

TEMPORAL FUSION TRANSFORMER

Loss and Output

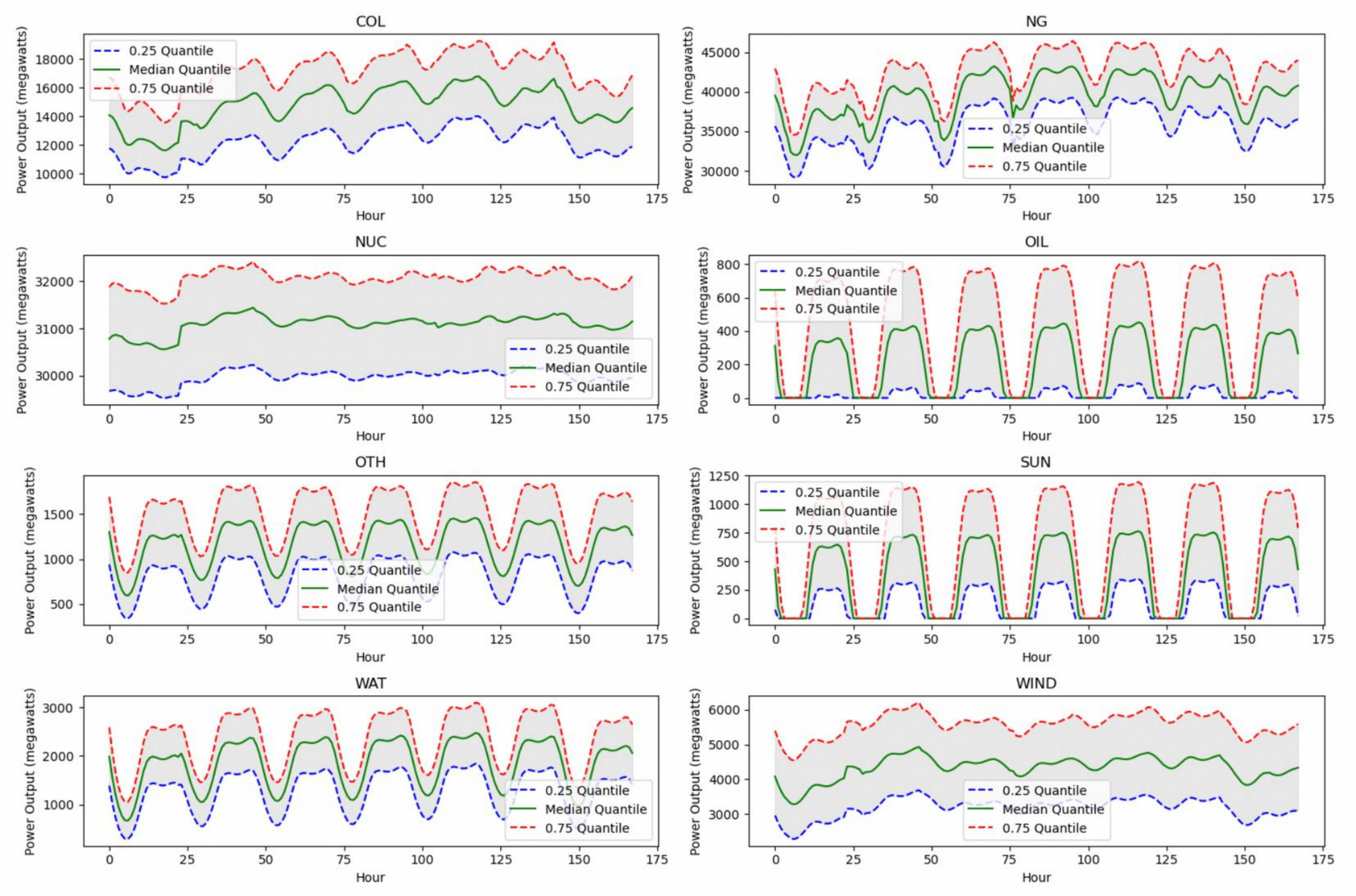
Quantile regression and loss allow the model to predict probabilistic forecasts.

Predictions are conditional quantiles of a distribution, rather than just the mean

Outputs can be interpreted as a confidence interval, allowing decision makers to consider the range of possible outcomes

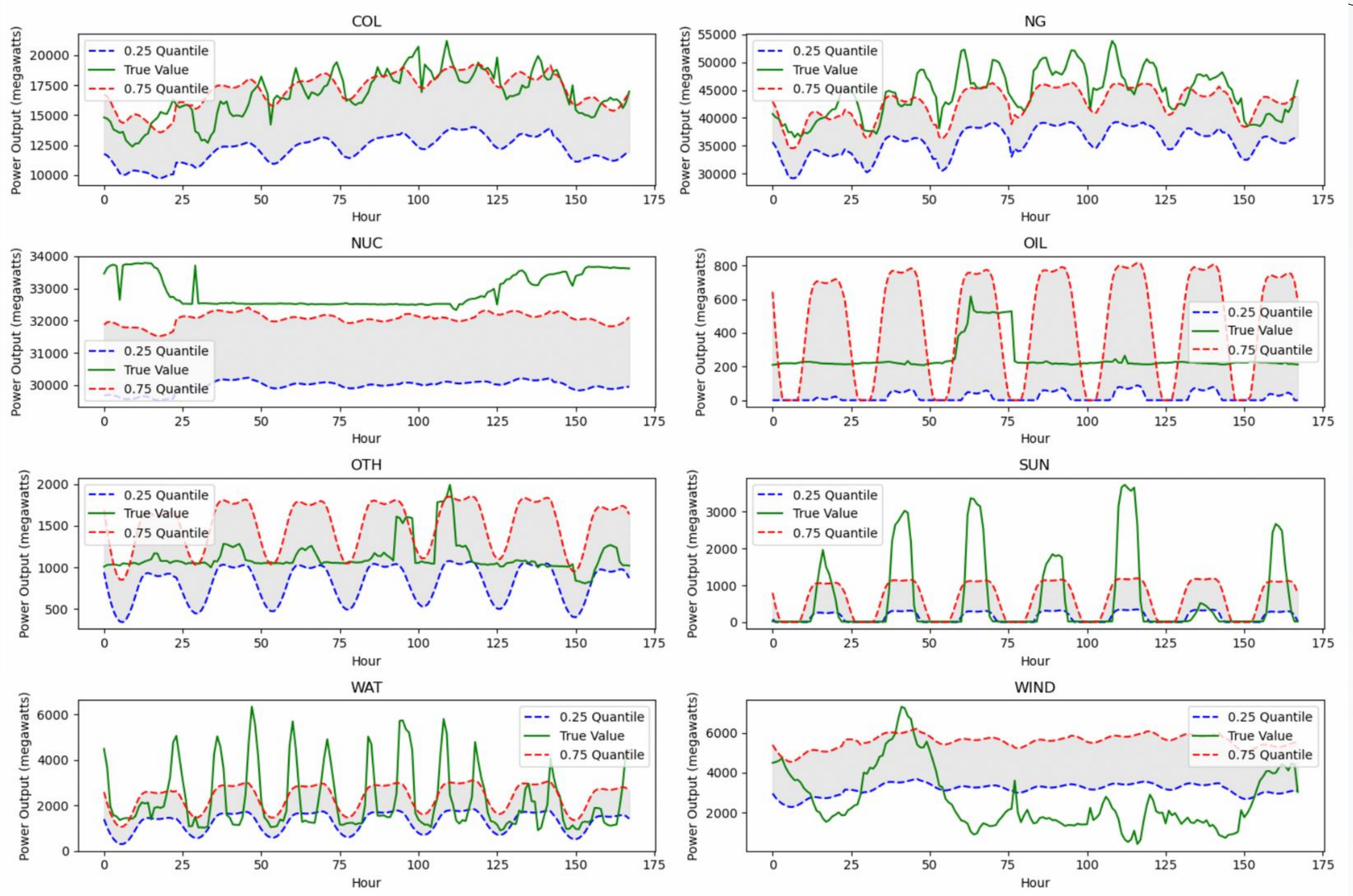
Quantile Output of TFT

VISUALIZATIONS



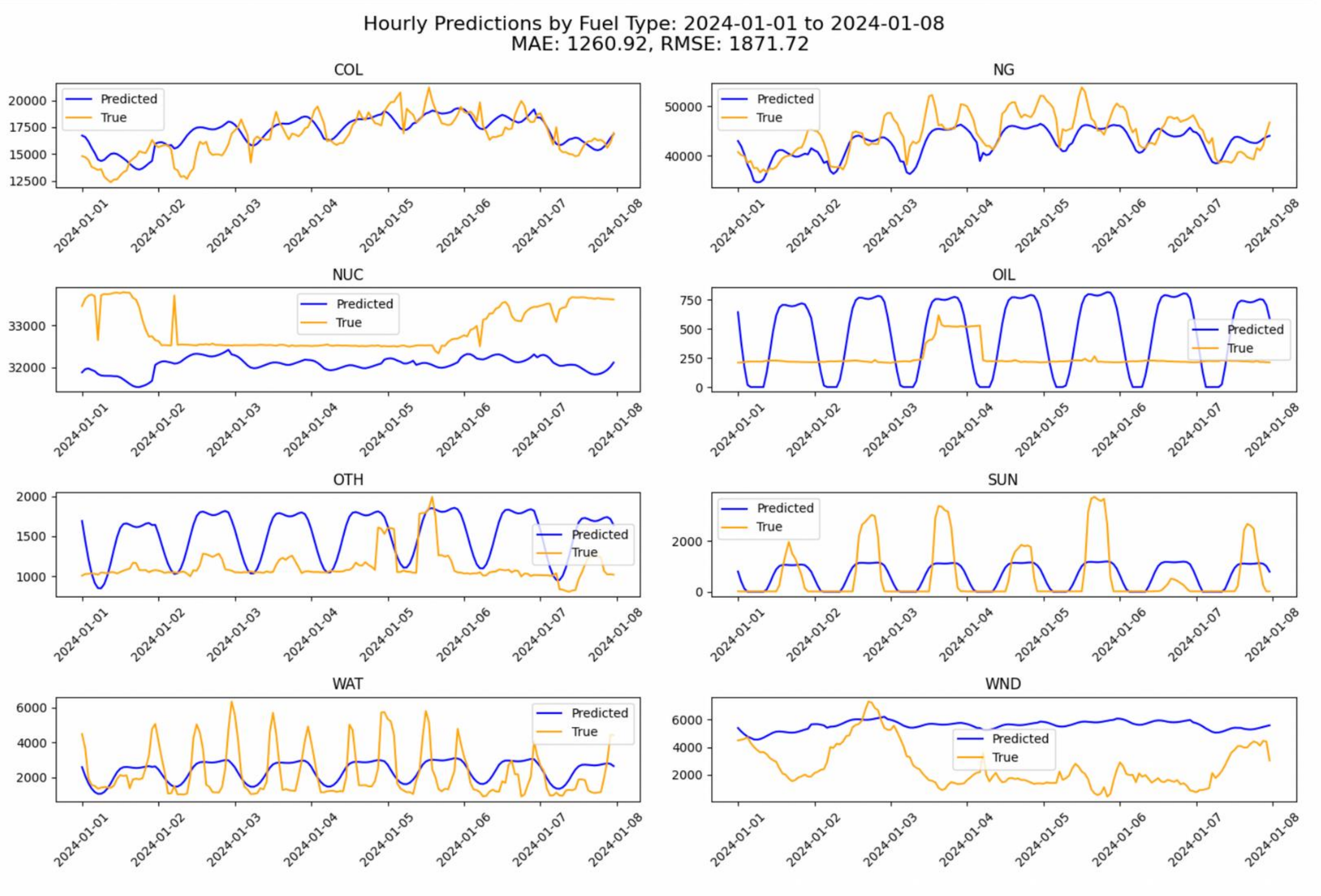
How did True Value compare to 25th and 75th Quantiles

VISUALIZATIONS



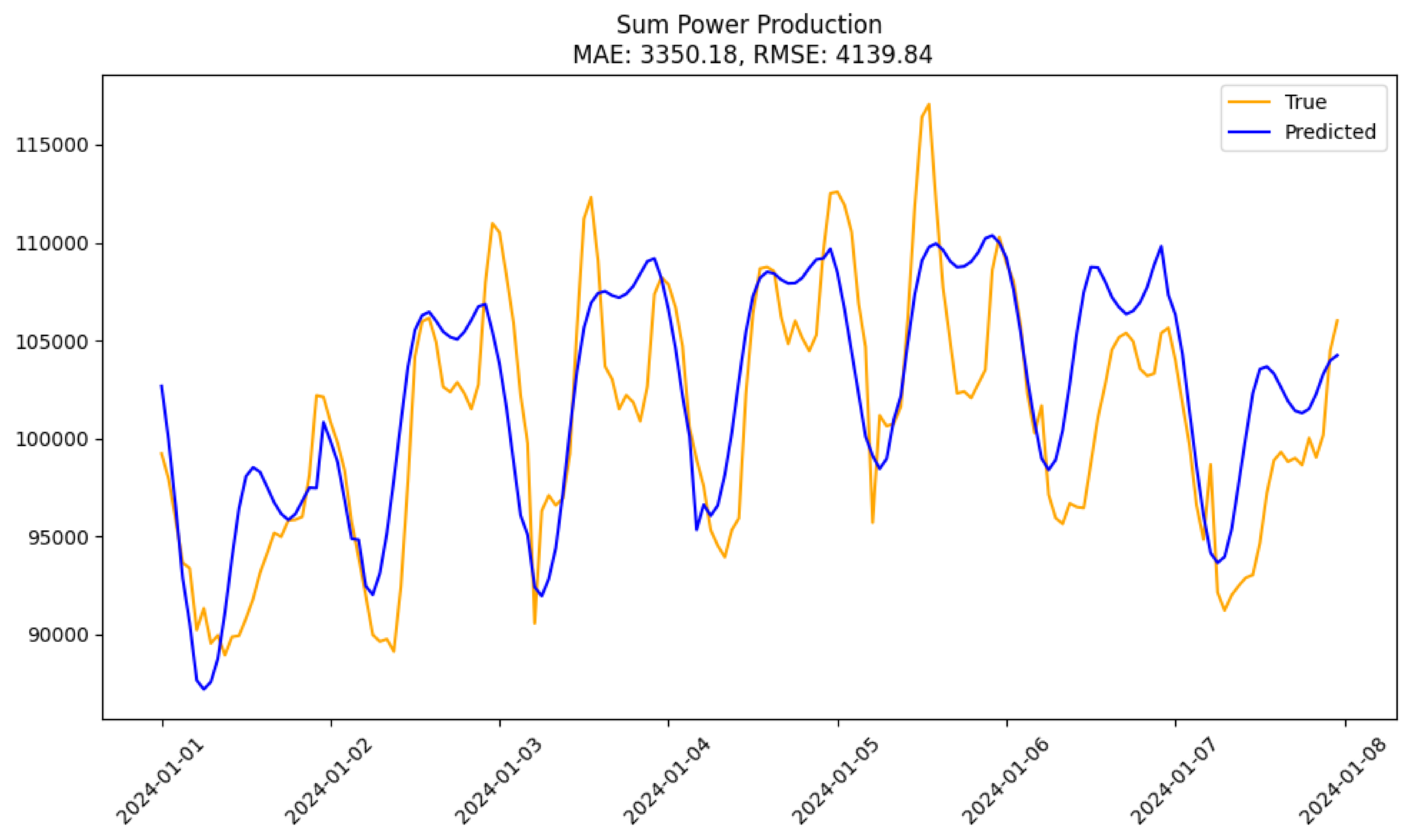
Predicted vs Actual Trend Lines

VISUALIZATIONS



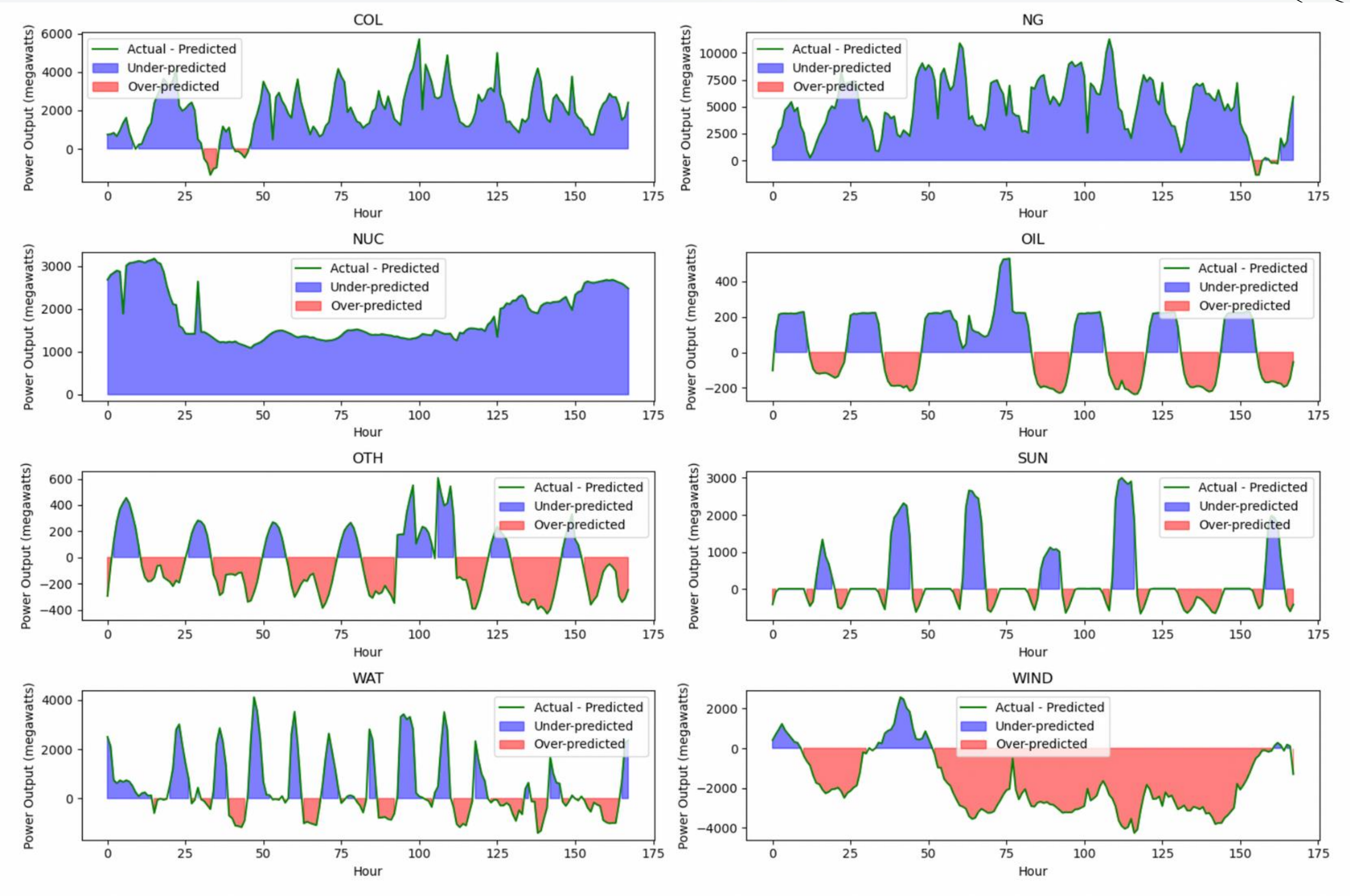
Sum of Predicted vs Actual Trend Lines

VISUALIZATIONS



Residual Analysis: Did the model over or under predict?

VISUALIZATIONS



FINDINGS/RESULTS

Model Accuracy Metrics:

Mean-Absolute Error: 1260.92
Sum Mean-Absolute Error: 3350.18

This analysis shows the relative power output of different sources of energy over time. As shown in the visualizations, some sources of power are more predictable in their outputs than others (i.e. wind is less predictable than natural gas)

LIMITATIONS

01

Lack of information on the cost of the fuel source in the data.

02

External factors such as natural disasters, weather conditions, or power outages can significantly impact fuel consumption but are not accounted for in the analysis.

03

The category 'Other' may introduce ambiguity

04

Analysis does not explicitly account for holidays

05

Analysis does not differentiate between residential and commercial fuel consumption

06

Missing data

CONTINUED ANALYSIS

01

Price data correlated with consumption data allows for comprehensive economic analysis and understanding how fuel prices correlate with consumption patterns and provides insights into the cost-effectiveness of various energy sources

02

Incorporating weather data such as temperature, precipitation, etc could reveal correlations between weather patterns and energy consumption trends

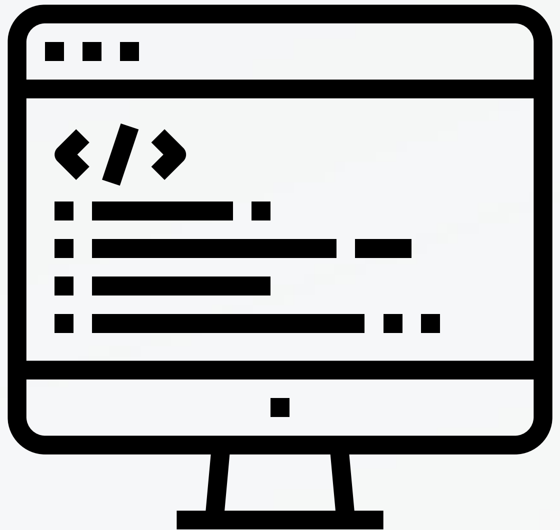
03

Incorporating holiday calendars allows for identification of holiday specific trends/anomalies

DISSEMINATION OF ANALYSIS

Analysis can be documented in a Jupyter Notebook or similar format with clear explanation and visuals. This analysis can be shared on GitHub and LinkedIn to reach a professional audience.

Further creation of an interactive dashboard using tools like Tableau or Dash allow for an engaging way the audience to interact with the analysis and draw their own conclusions. The target audience is fuel/power producers and consumers, which includes most people and businesses in the world.



DIVISION OF WORK

Luke Chesley
Project Scoping
Model Building
API access

Lauren Miller
Presentation
Project Scoping
Project Submission

Caleb Miller
Project Scoping
Data Visualizations
Presentation

Hashim Afzal
Project Scoping
Data Visualizations

THANK YOU

Questions?