Ysgol Cwm Brombil Energy Report

Contents

- Dataset overview
- Covariate data
- AI modeling

Dataset overview

Dataset overview

- Ysgol Cwm Brombil –Site electricity demand data (kWh)
- 24/06/2018 – 13/02/2021
- Sampled at 30 minute intervals
- 46321 data points after processing

Dataset overview – 14/07/2018 – 14/08/2018

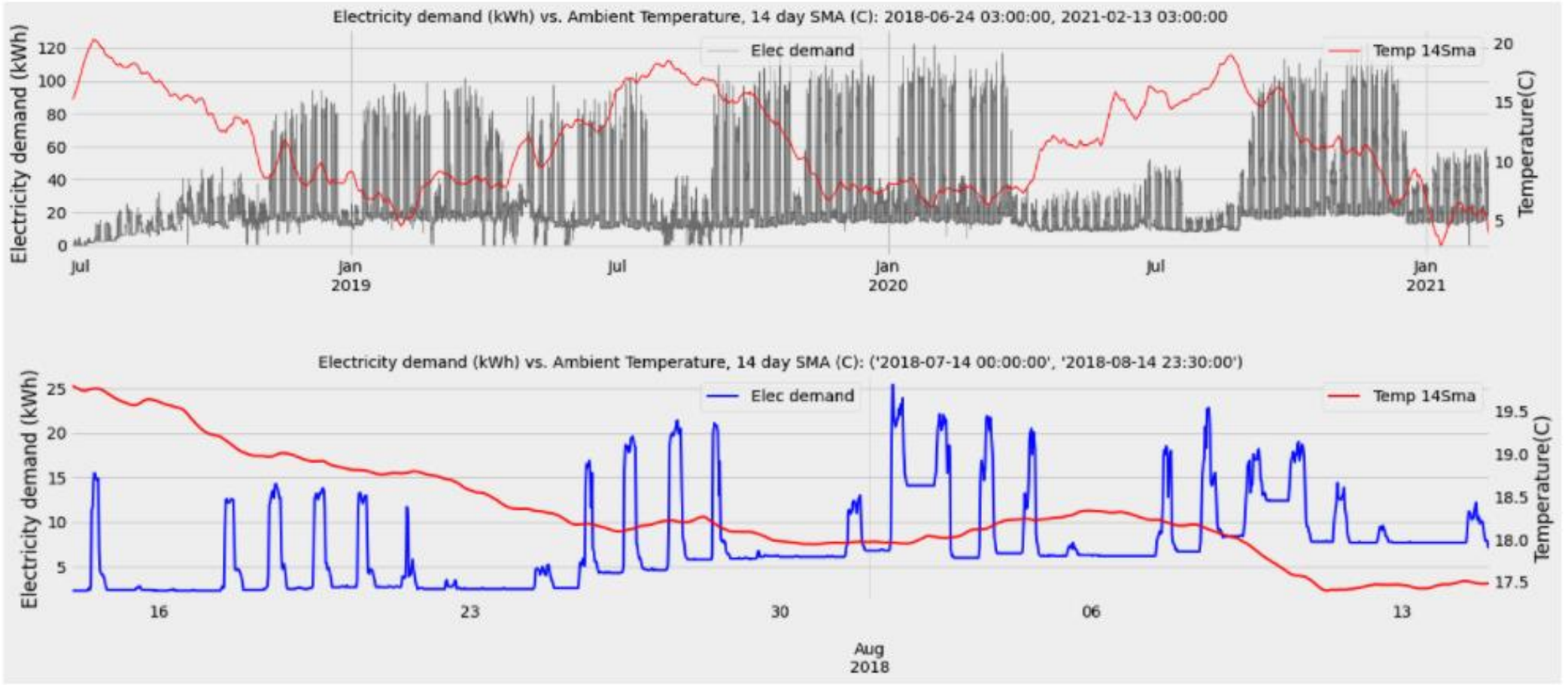


Figure 1: Electricity demand overview, June 2018 – February 2021 demand data with 14 day ambient temperature simple moving average (top). Demand data from 14/07/2018 – 14/08/2018 with 14 day ambient temperature simple moving average (bottom).

Dataset overview – 14/09/2020 – 14/10/2020

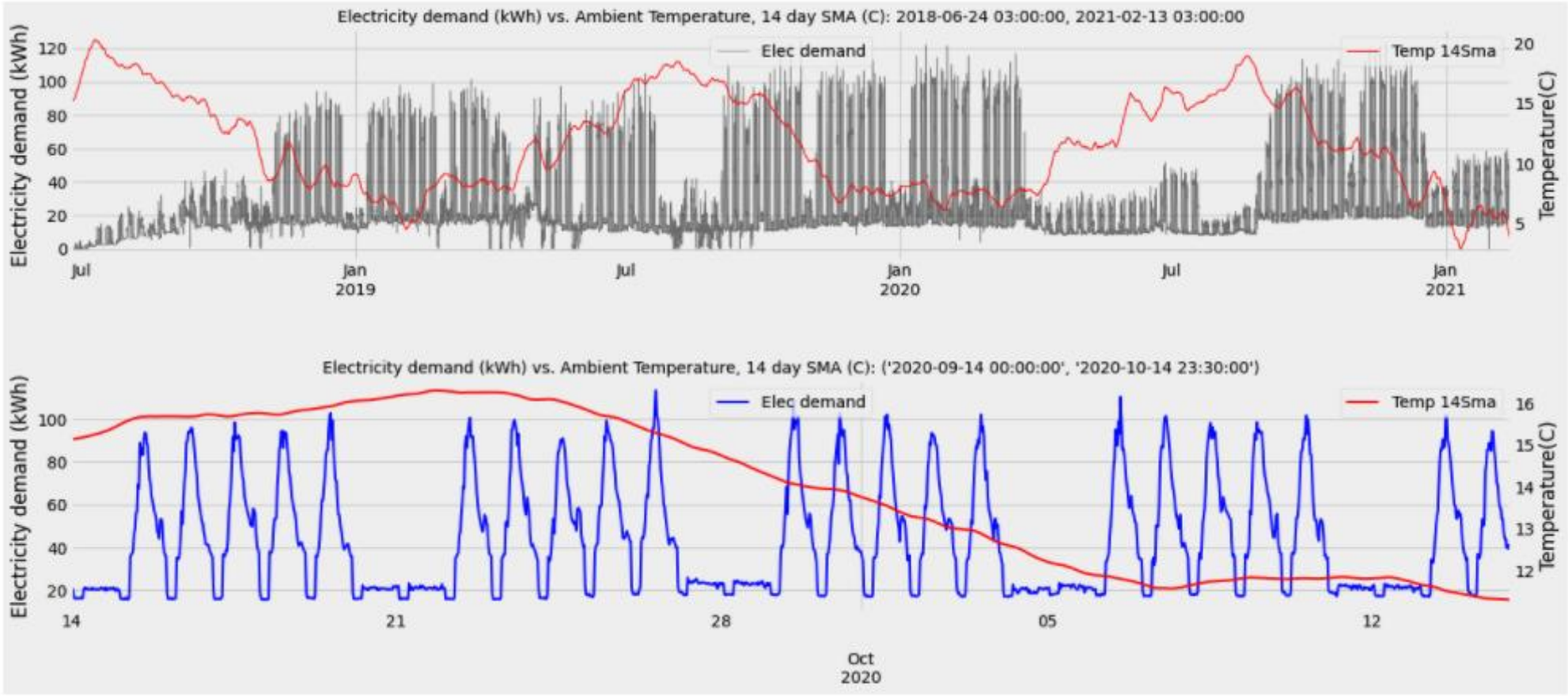


Figure 2: Electricity demand overview, June 2018 – February 2021 demand data with 14 day ambient temperature simple moving average (top). Demand data from 14/09/2020 – 14/10/2020 with 14 day ambient temperature simple moving average (bottom).

Dataset overview

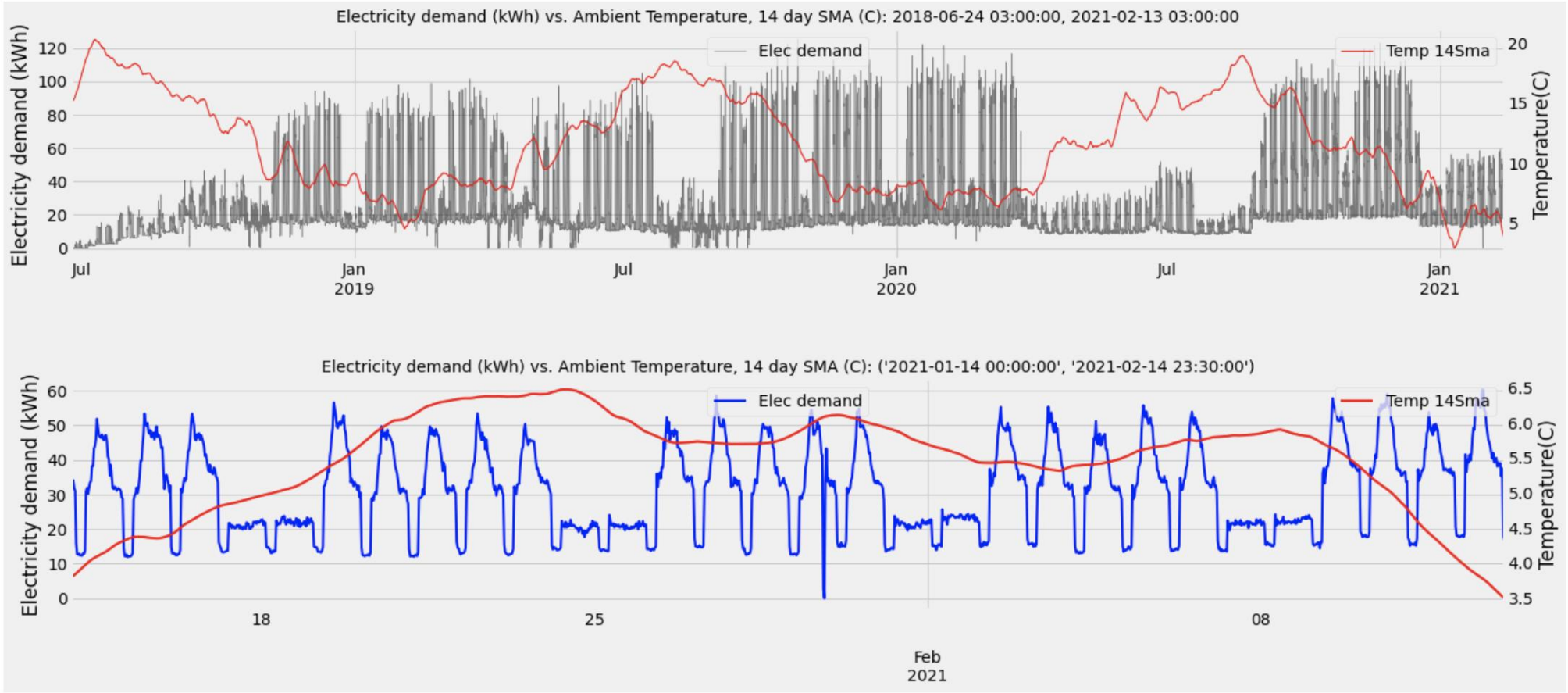


Figure : Electricity demand overview, June 2018 – February 2021 demand data with 14 day ambient temperature simple moving average (top). Demand data from 14/01/2021 – 14/02/2021 with 14 day ambient temperature simple moving average (bottom).

Dataset overview – School day vs. Non-school day

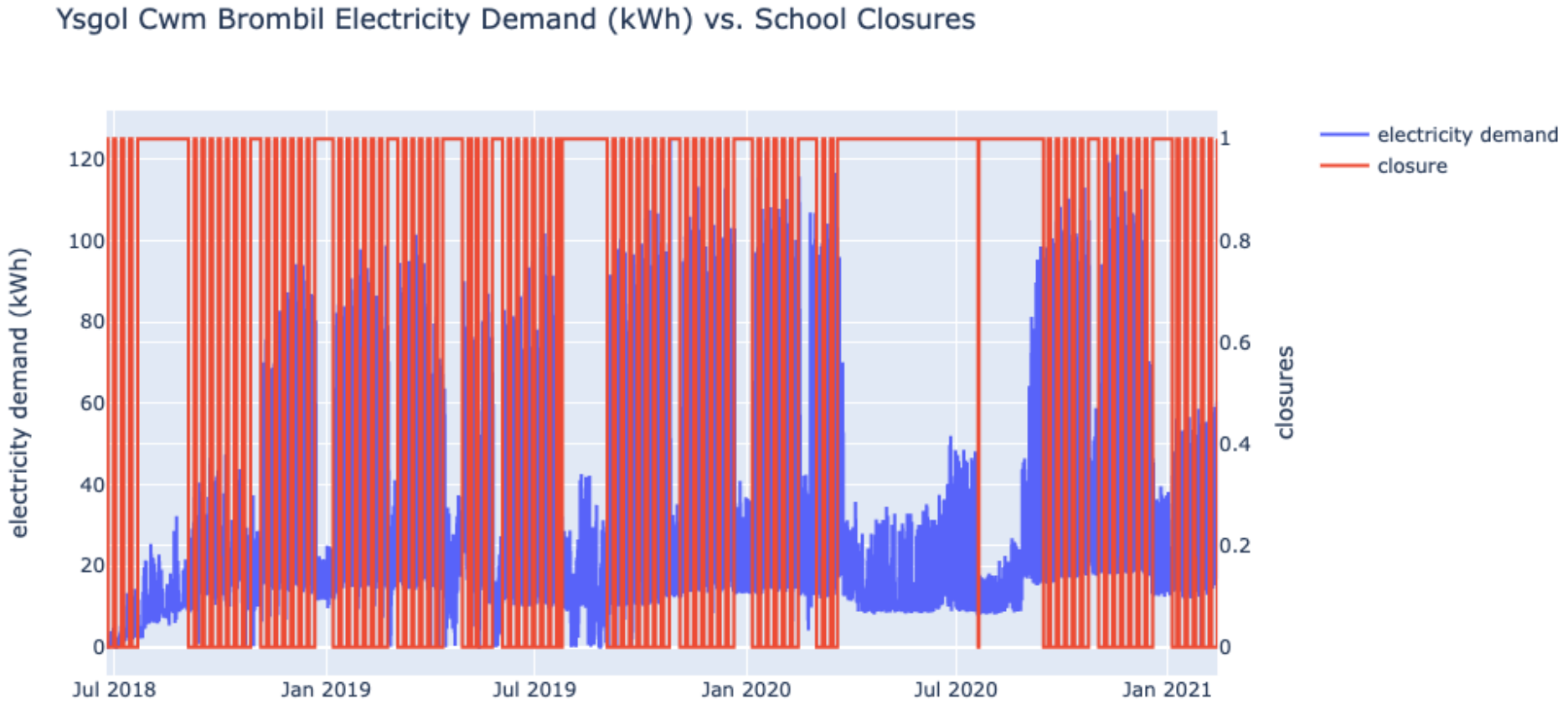


Figure : Demand chart from June 2018 – February 2021 with school closures including weekends, school holidays, bank holidays and closures due to Covid19

Dataset overview – Total demand by month

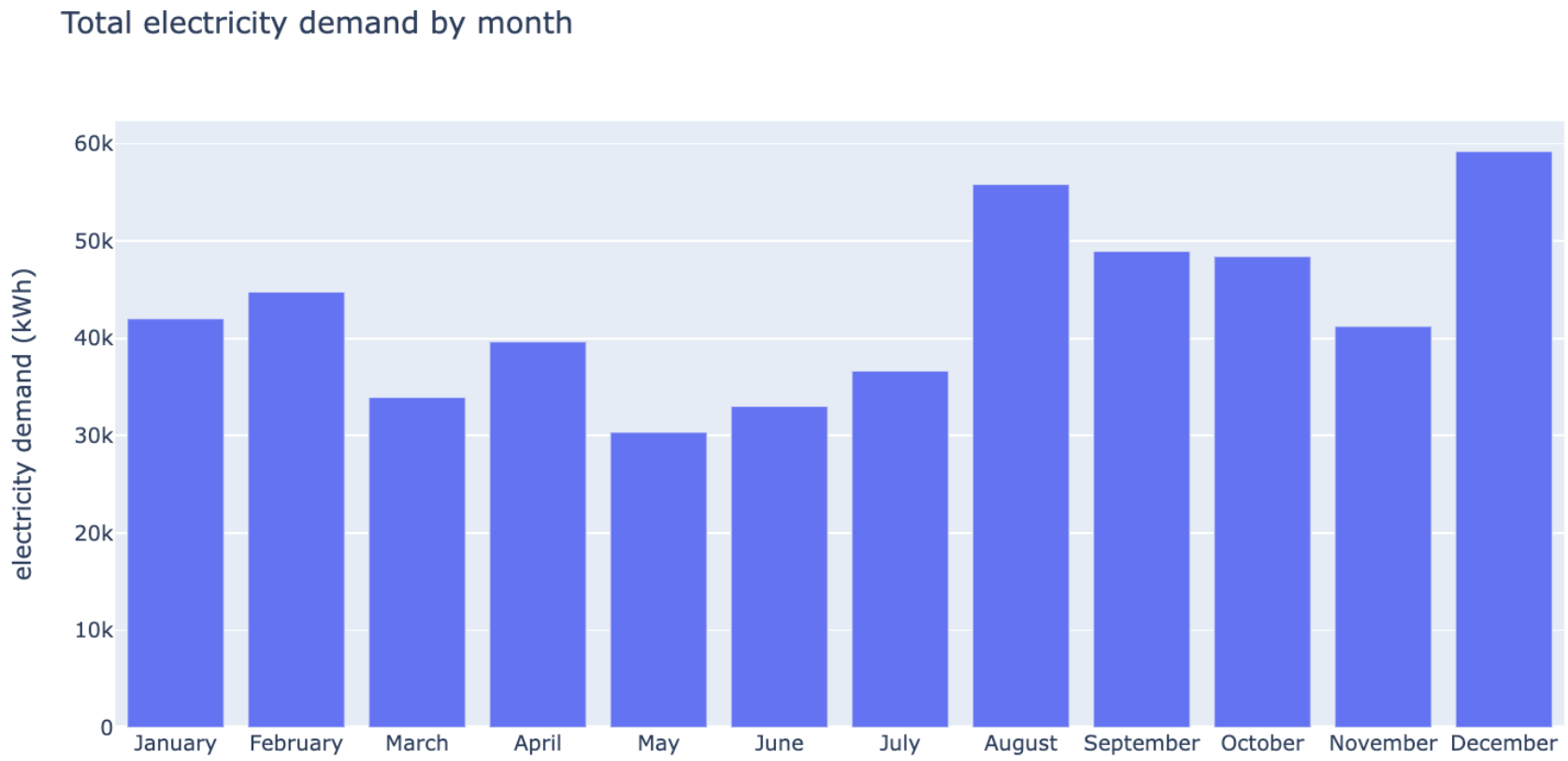


Figure #: Bar chart of demand aggregated by month from June 2018 – February 2021

Dataset overview – Total term time demand by month

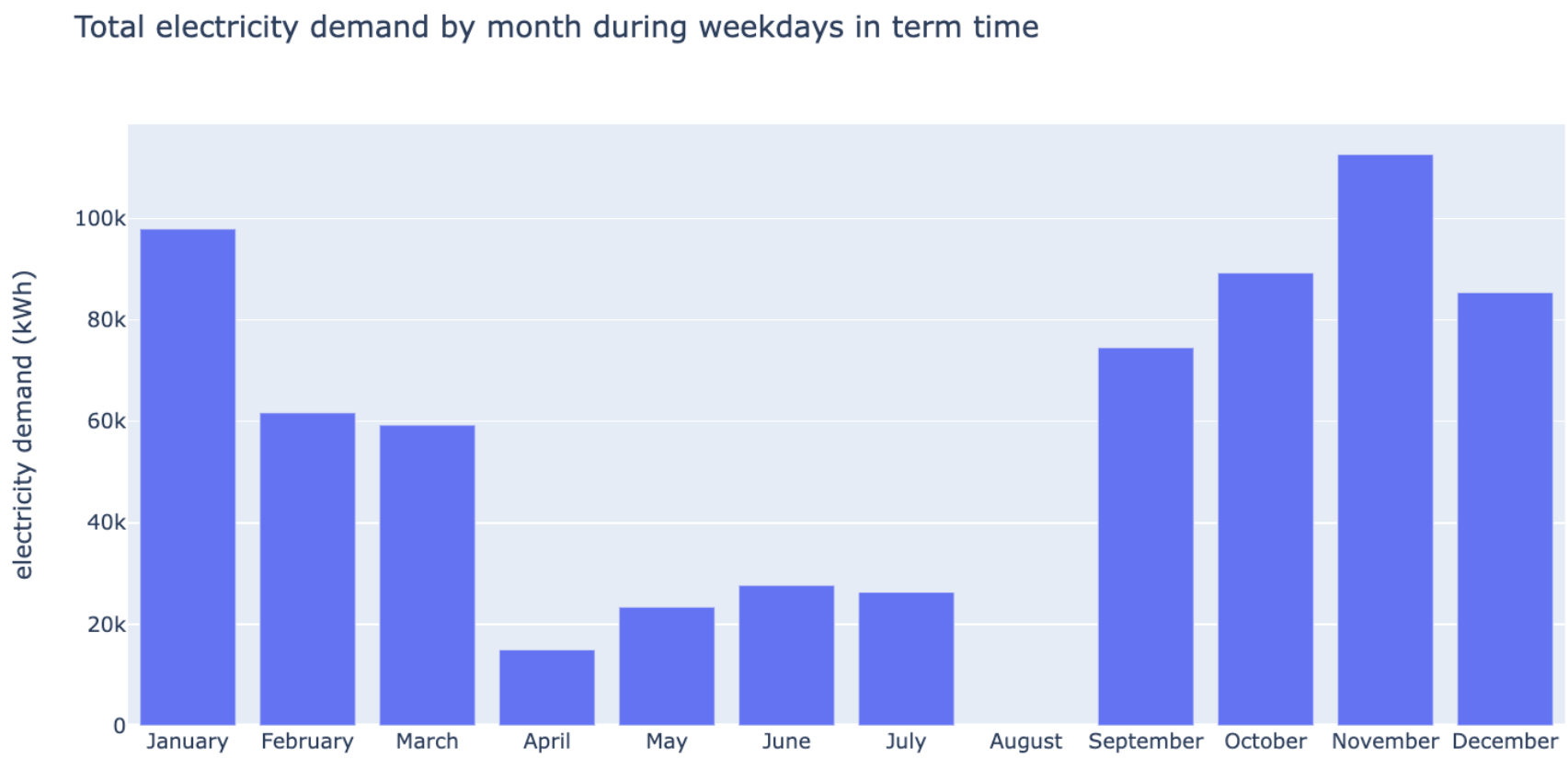


Figure #: Bar chart of aggregated demand by month from June 2018 – February 2021 during weekdays in term time

Dataset overview – Total demand during closures by month

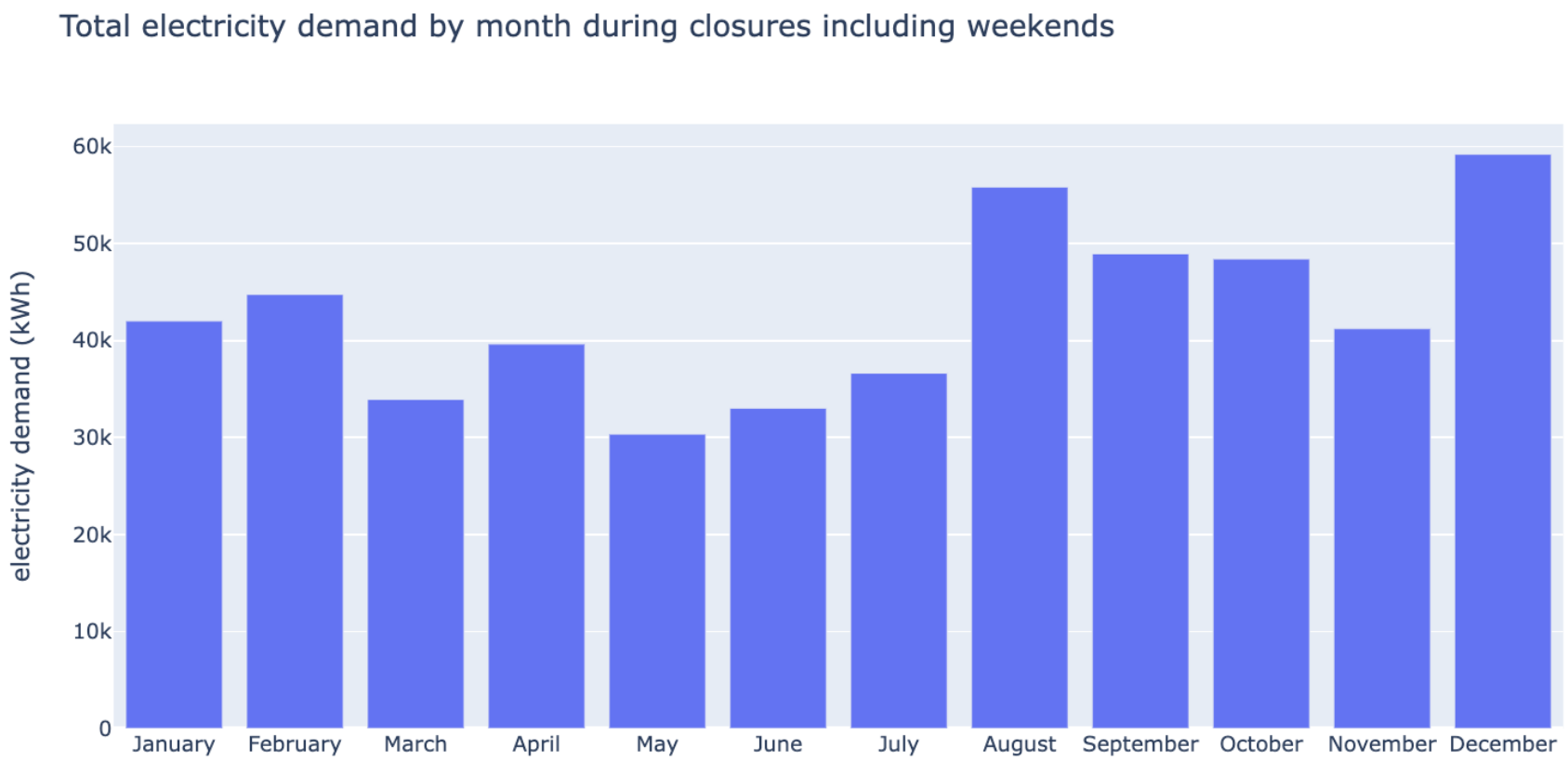


Figure #: Bar chart of aggregated demand by month from June 2018 – February 2021 during weekends, holidays and closures due to Covid19

Dataset overview – Demand distribution by season

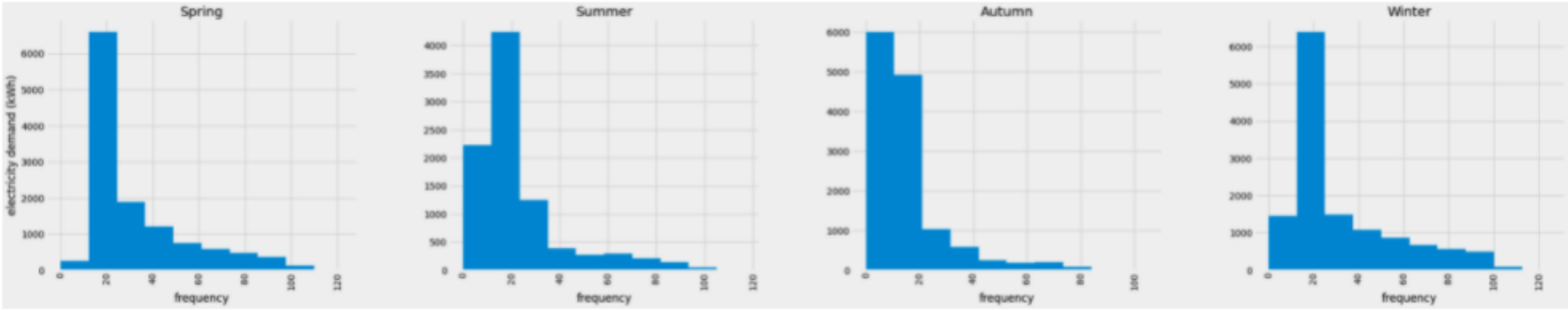
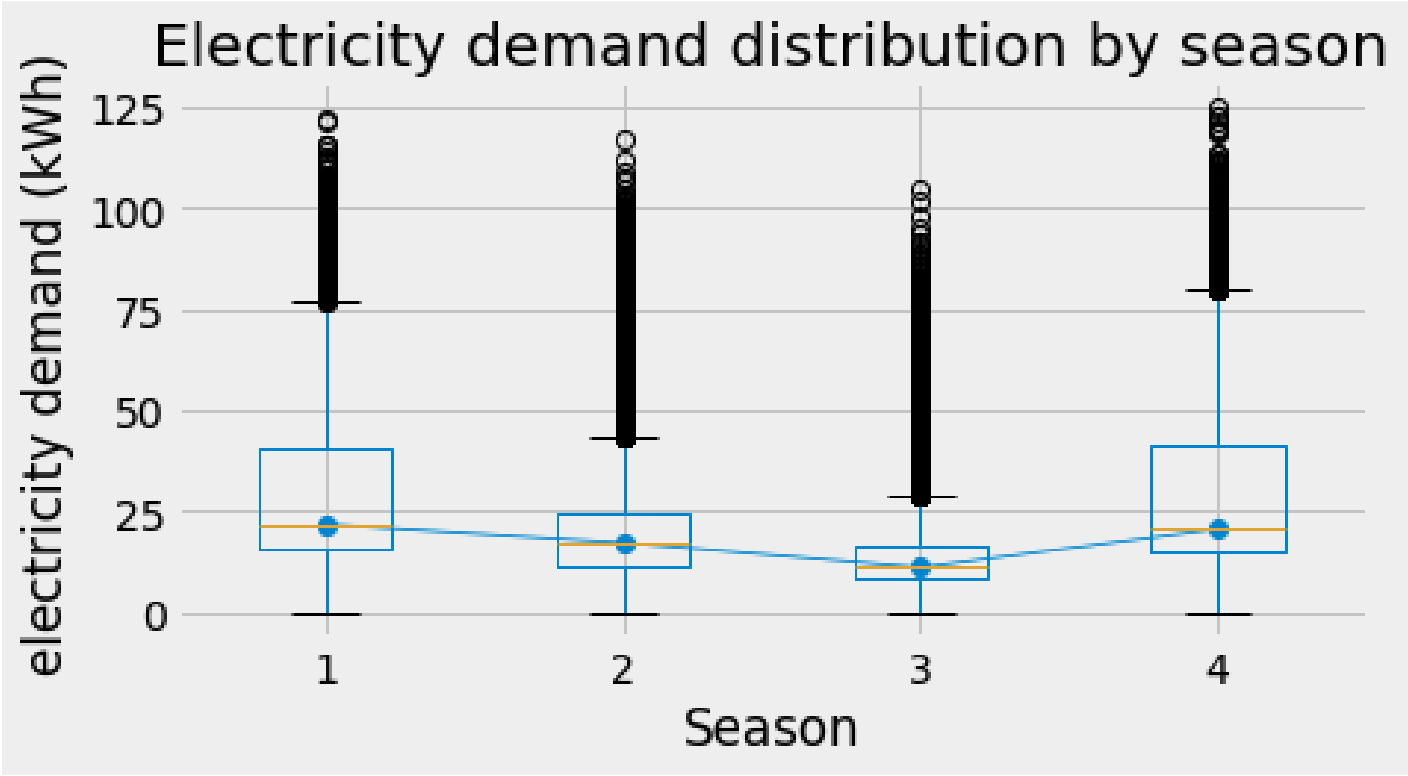


Figure #: Histograms showing demand distribution by season from June 2018 – February 2021

Dataset overview – Demand distribution by season



1 – Spring
2 – Summer
3 – Autumn
4 – Winter

Figure #: Box and whisker plots showing demand distribution by season from June 2018 – February 2021

Dataset overview – Demand distribution by month

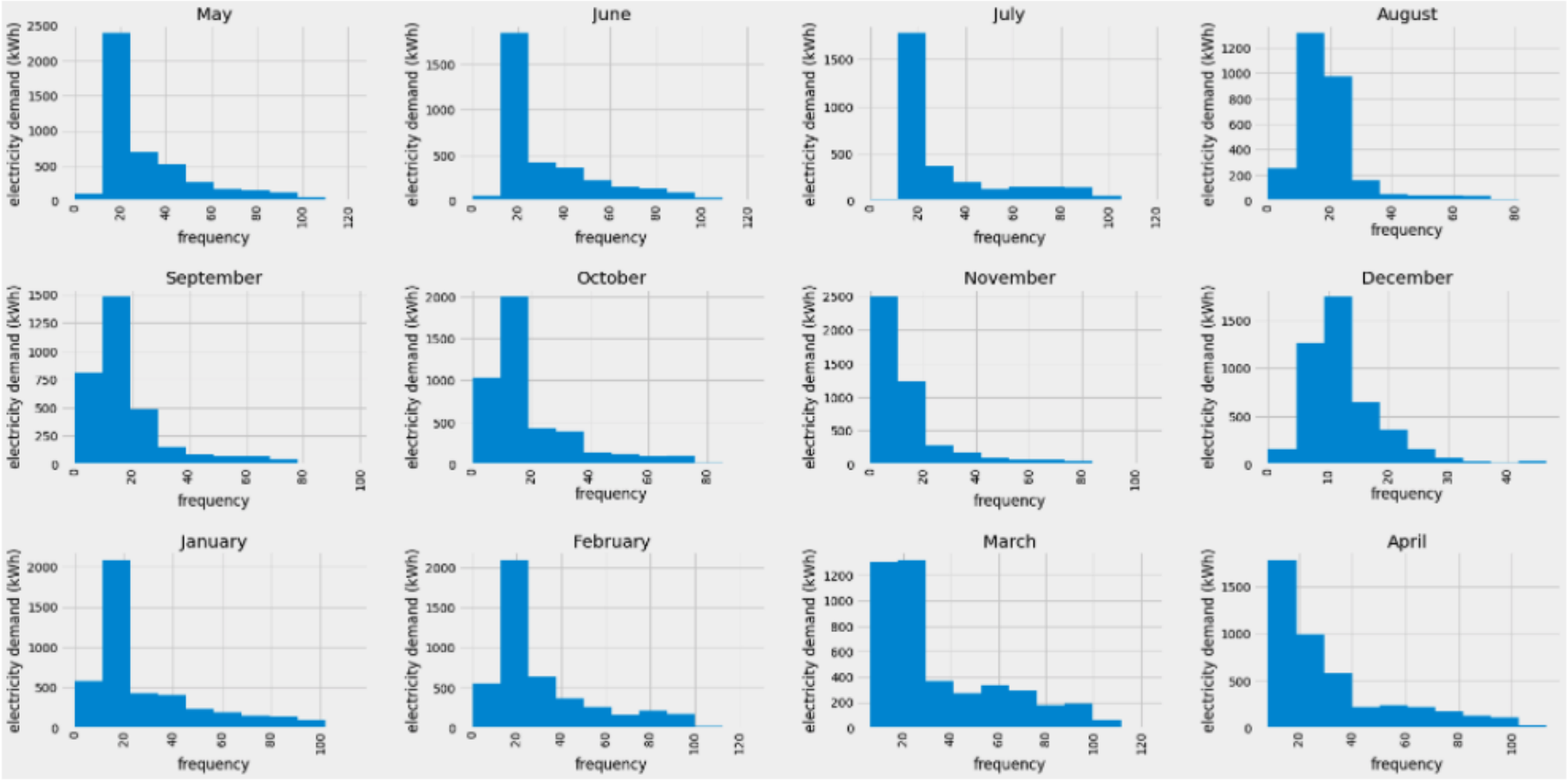
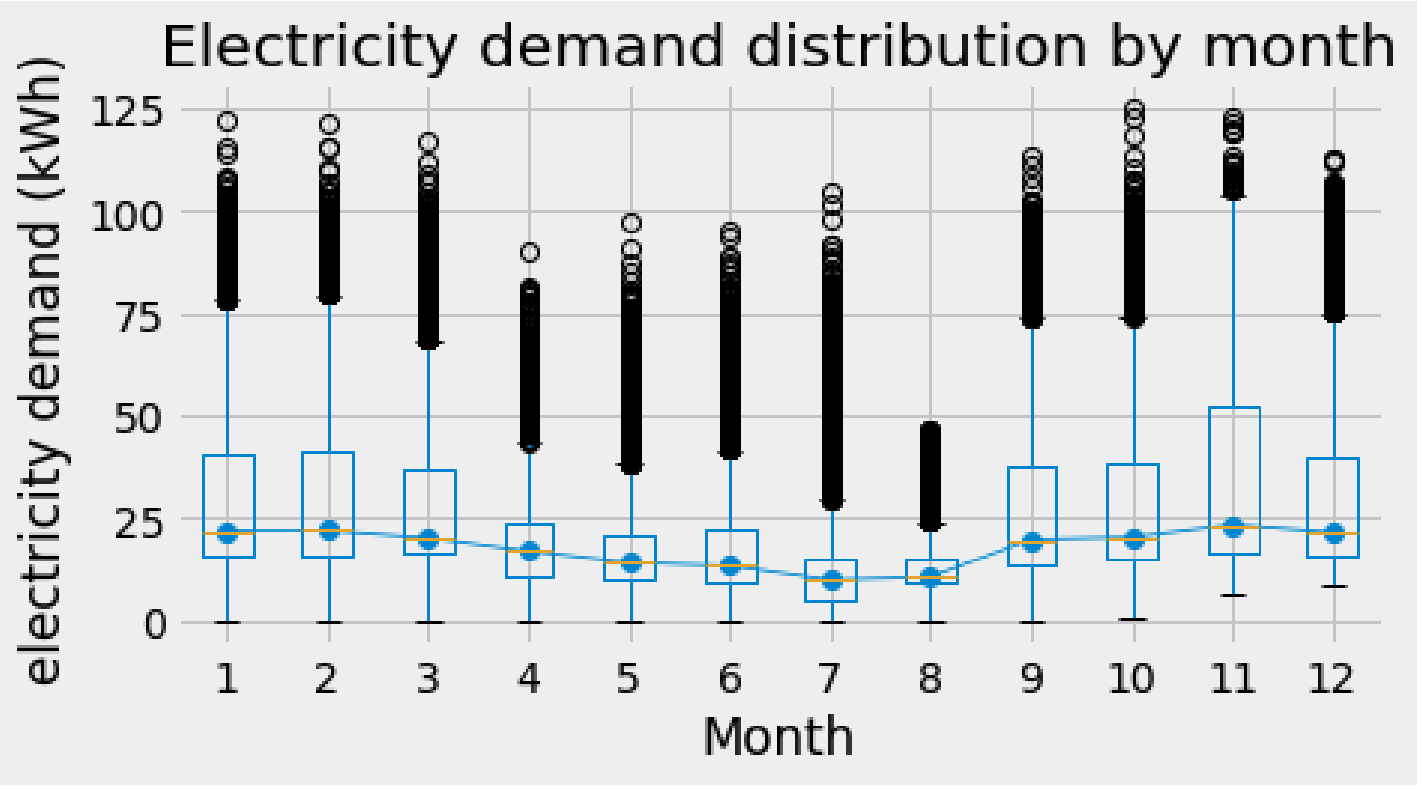


Figure : Histograms of demand distribution by month from June 2018 – February 2021

Dataset overview – Demand distribution by month



1 – May
2 – June
3 – July
4 – August
5 – September
6 – October
7 – November
8 – December
9 – January
10 – February
11 – March
12 – April

Figure #: Box and whisker plot showing demand by month of the year June 2018 – February 2021

Dataset overview – Demand distribution by weekday

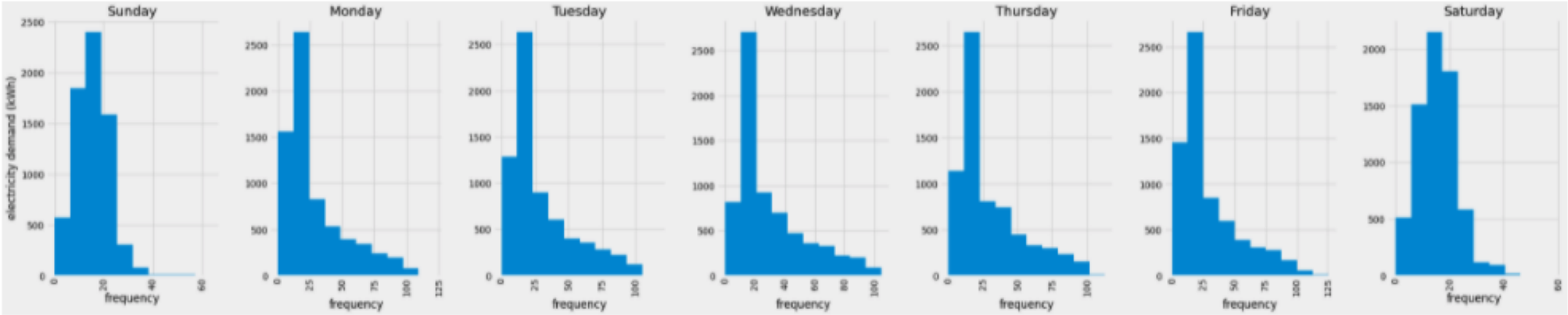
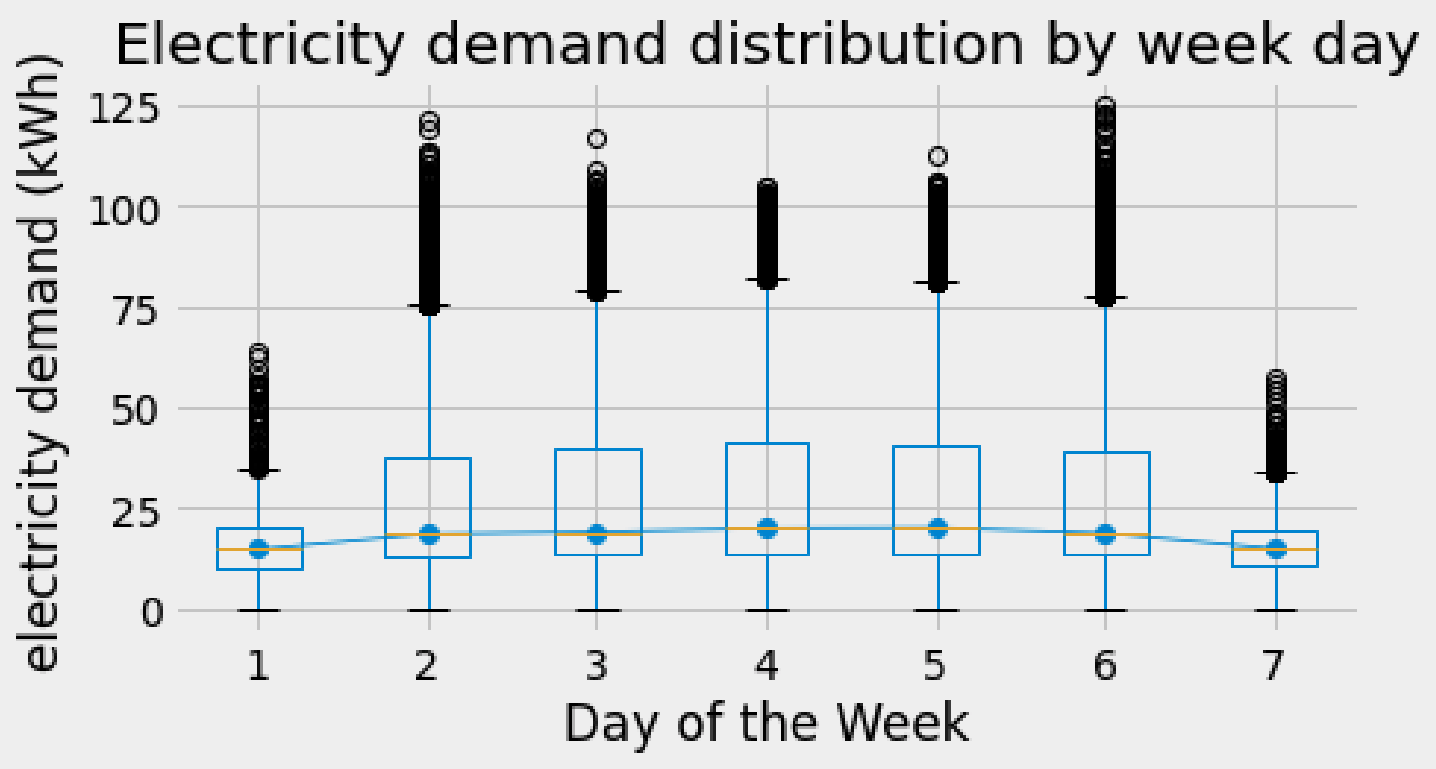


Figure #: Histograms of demand distribution by day of the week from June 2018 – February 2021

Dataset overview – Demand distribution by weekday



1 – Sunday
2 – Monday
3 – Tuesday
4 – Wednesday
5 – Thursday
6 – Friday
7 – Saturday

Figure #: Box and whisker of demand distribution by day of the week from June 2018 – February 2021

Dataset overview – Demand distribution by half hour of the day

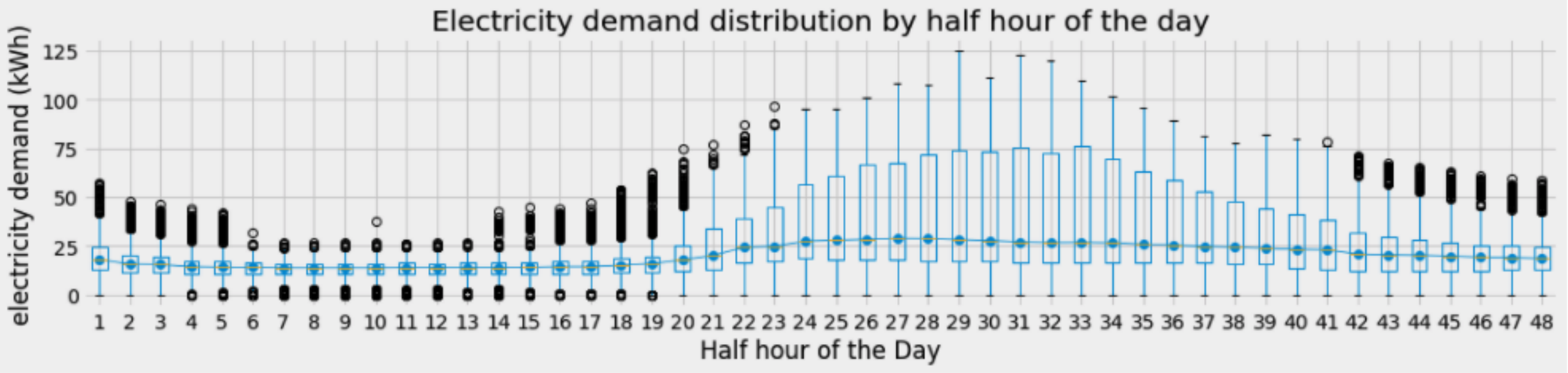


Figure : Box and whisker plot of demand aggregated by half hour of the day from June 2018 – February 2021

Covariate data

Covriate data – Correlation matrix

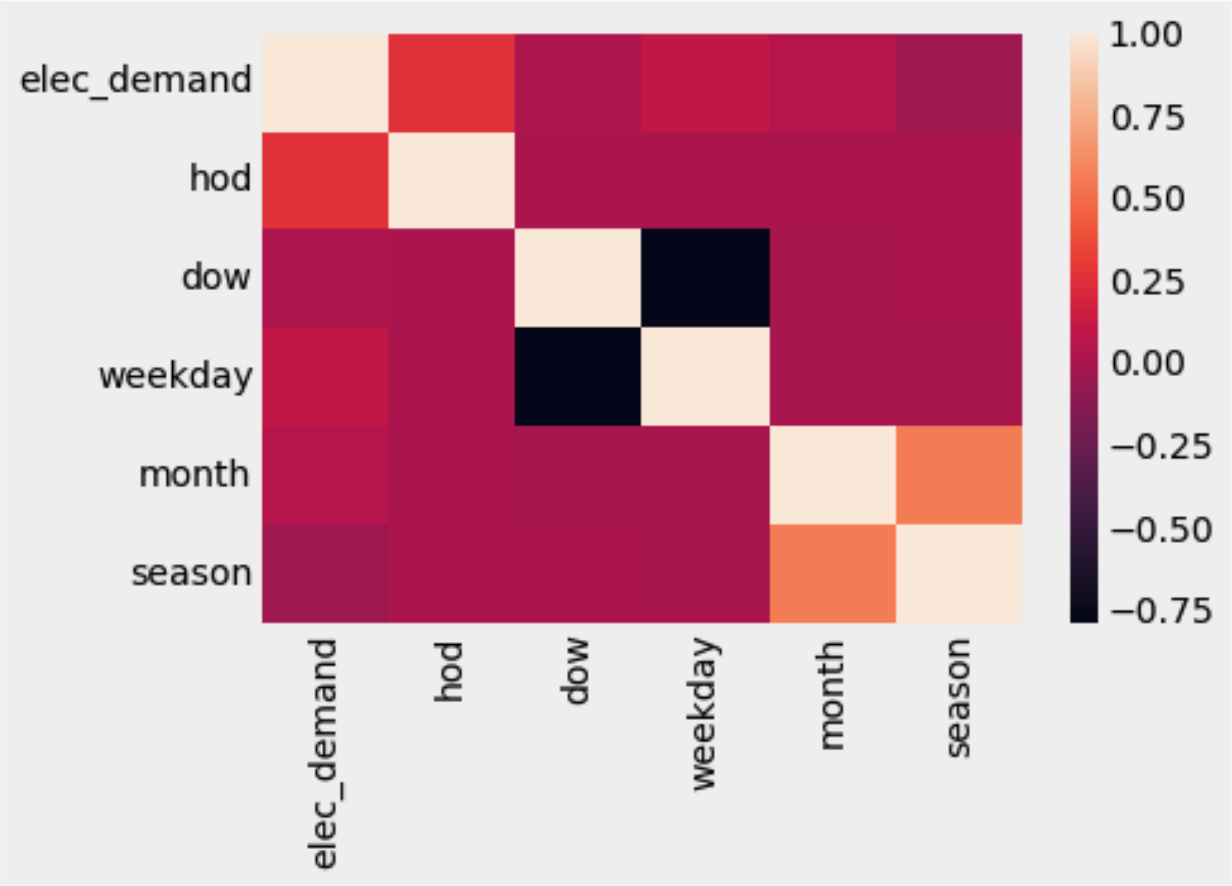


Figure #: Pearson correlation matrix heat map of demand vs. seasonal features

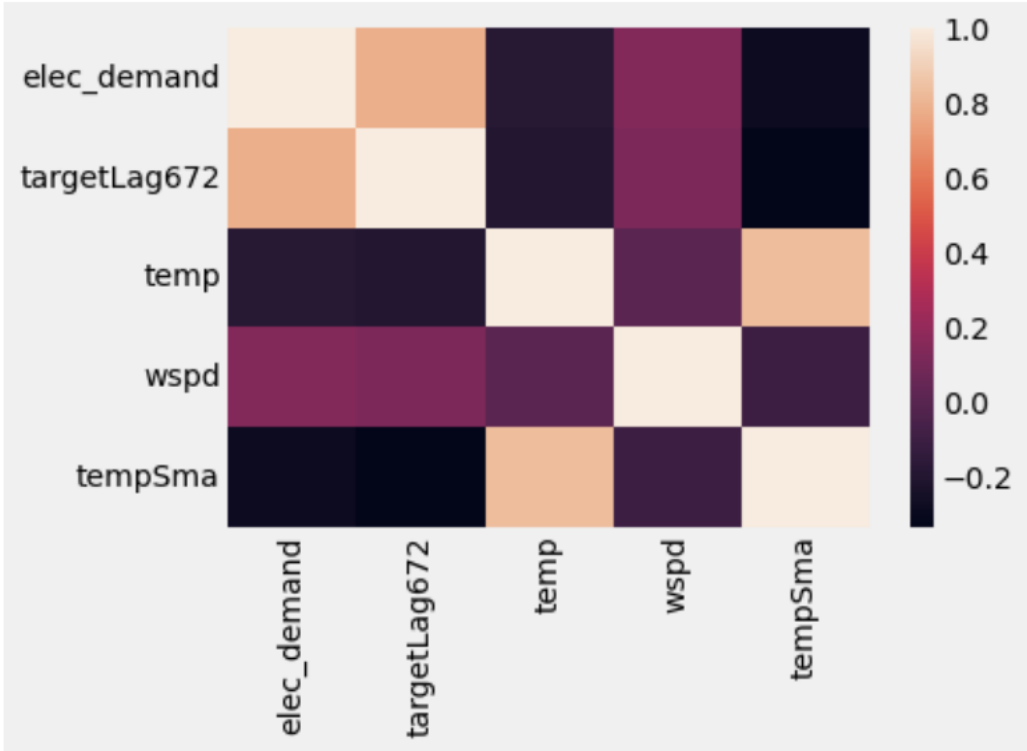
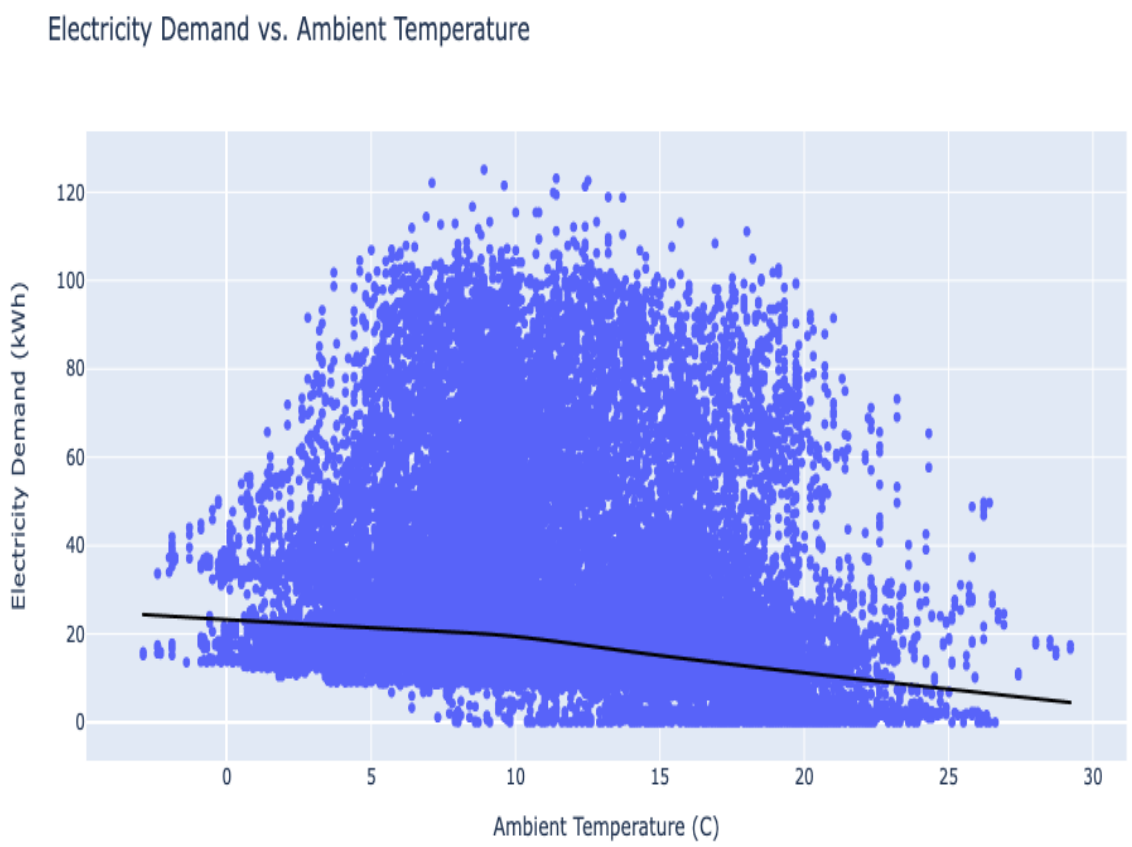


Figure #: Pearson correlation matrix heat map of demand vs. weather and lag features

Covriate data – Weather data relationship



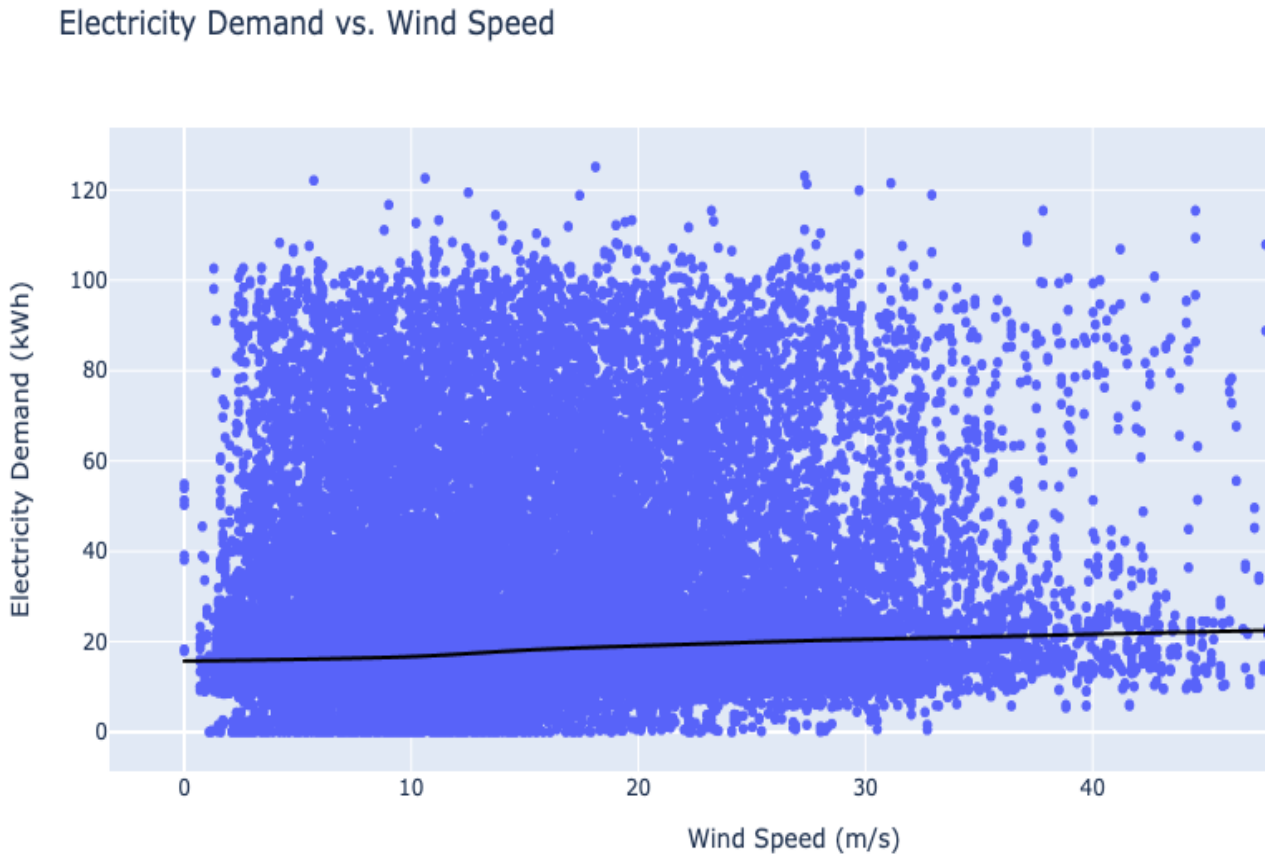


Figure #: Scatter plot of demand vs. ambient temperature (left). Scatter plot of demand vs. wind speed (right)

AI Modeling

AI modelling – Experiment setup

- Dataset: July 2018 – January 2020
- Train/test split: 80:20
- All features used for training
- No hyperparameter tuning performed

AI modelling – Training and testing data

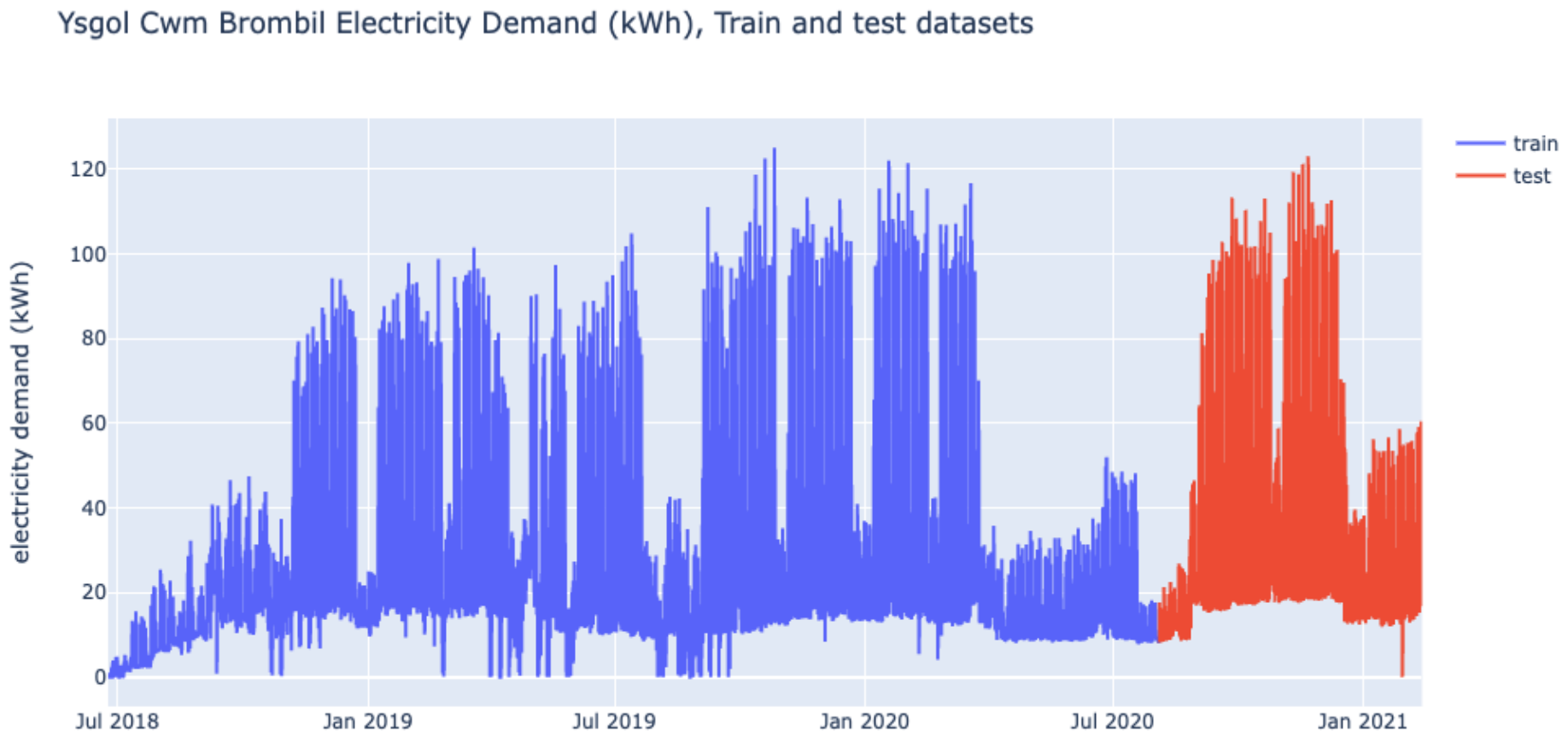


Figure #: Demand dataset train/test split from July 2018 – February 2021

AI modelling – Training and testing data

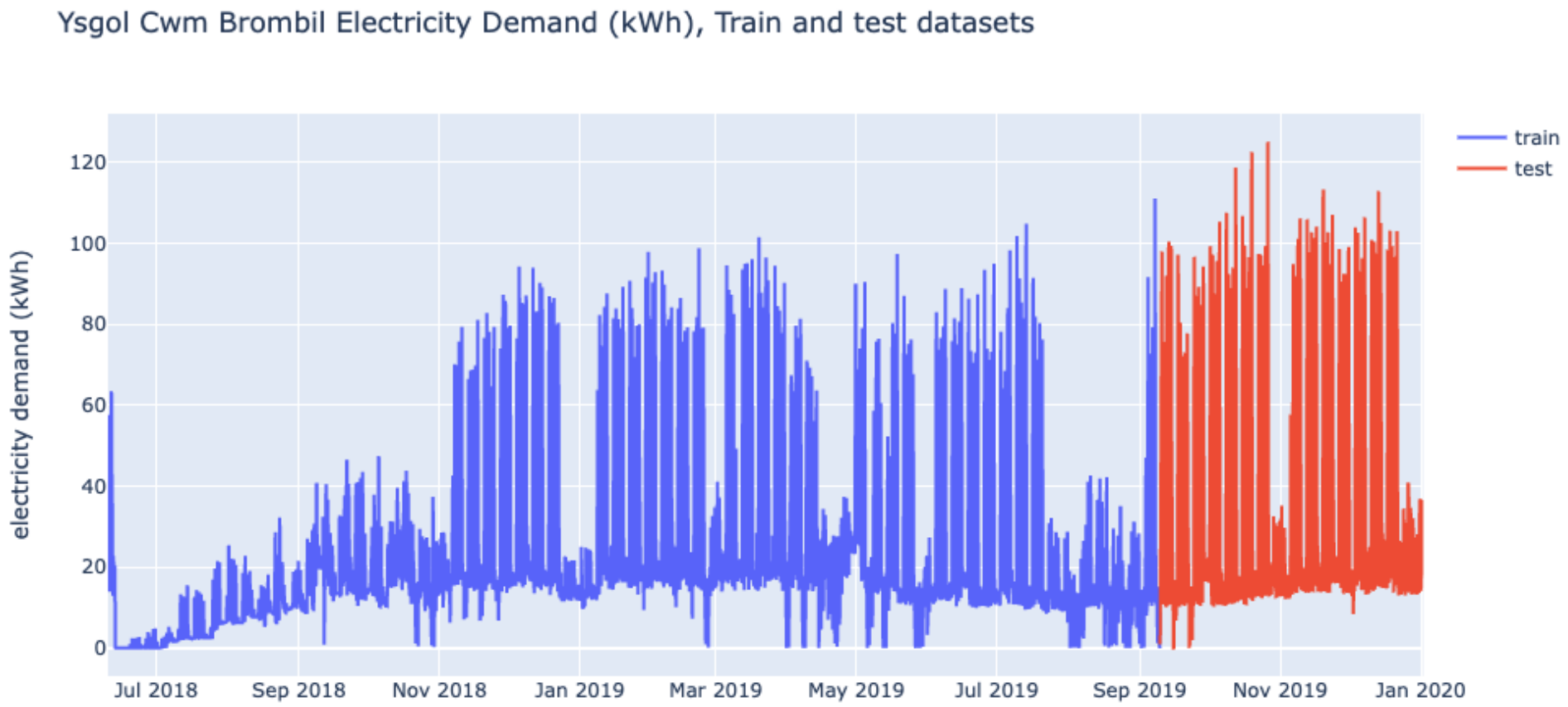


Figure #: Demand train/test split on subset of data from July 2018 – February 2021

AI modelling – Training features

Feature	Type	Description
closure	Other	Closure (Integer, 0-1)
month	Seasonal	Month of the year (Integer, 0-11)
dow	Seasonal	Day of the week (Integer, 0-6)
hod	Seasonal	Half hour of the day (Integer 0-47)
season	Seasonal	Season of the year (Integer, 0-11)
weekday	Seasonal	Weekday (Integer, 0-1)
temp	Weather	Ambient temperature (Floating point number (C))
wspd	Weather	Windspeed (Floating point number (m/s))
tempSma	Weather	14 day simple moving average of ambient temperature (Floating point number (C))
targetLag672	Other	Value of demand at same period 14 days prior (Floating point number, (kWh))

AI modeling – Initial results

	Traditional Forecasting models			Machine Learning models		Deep Learning models	
Metric/Model	SARIMA	ETS	Holt-Winters	Random Forest	XGBoost	DeepAR	MLP
R squared	-1.06	-0.69	-0.74	0.79	0.72	0.67	0.6
RMSE	32.6	29.6	30	12	13.8	14.7	112.6
MAPE	inf	inf	inf	19.5%	23.7%	27.5%	8.7%
MAE	23.7	20.3	20.8	7.5	9	9.3	84.7

All models were evaluated using the 20% testing data and machine and deep learning models were trained using default hyperparameters

R squared – Shows the correlation between the predictions and the actual values.
RMSE – Root mean squared error, shows the squared root of the average squared error.
MAPE – Mean absolute percentage error, shows the average absolute error as a percentage at any point in the test dataset.
MAE – Mean absolute error, Shows the average absolute error at any point in the test dataset.

AI modeling – Best model

- Random Forest Regression model – Sci-kit Learn (Python)
- Features: Half hour of the day, day of the week, month, weekday, season, closure, 14 day temperature simple moving average, 14 day past lag value at same period, ambient temperature and wind speed.
- Hyperparameters – Default parameters, {random state: 42}

AI modeling – Test dataset predictions

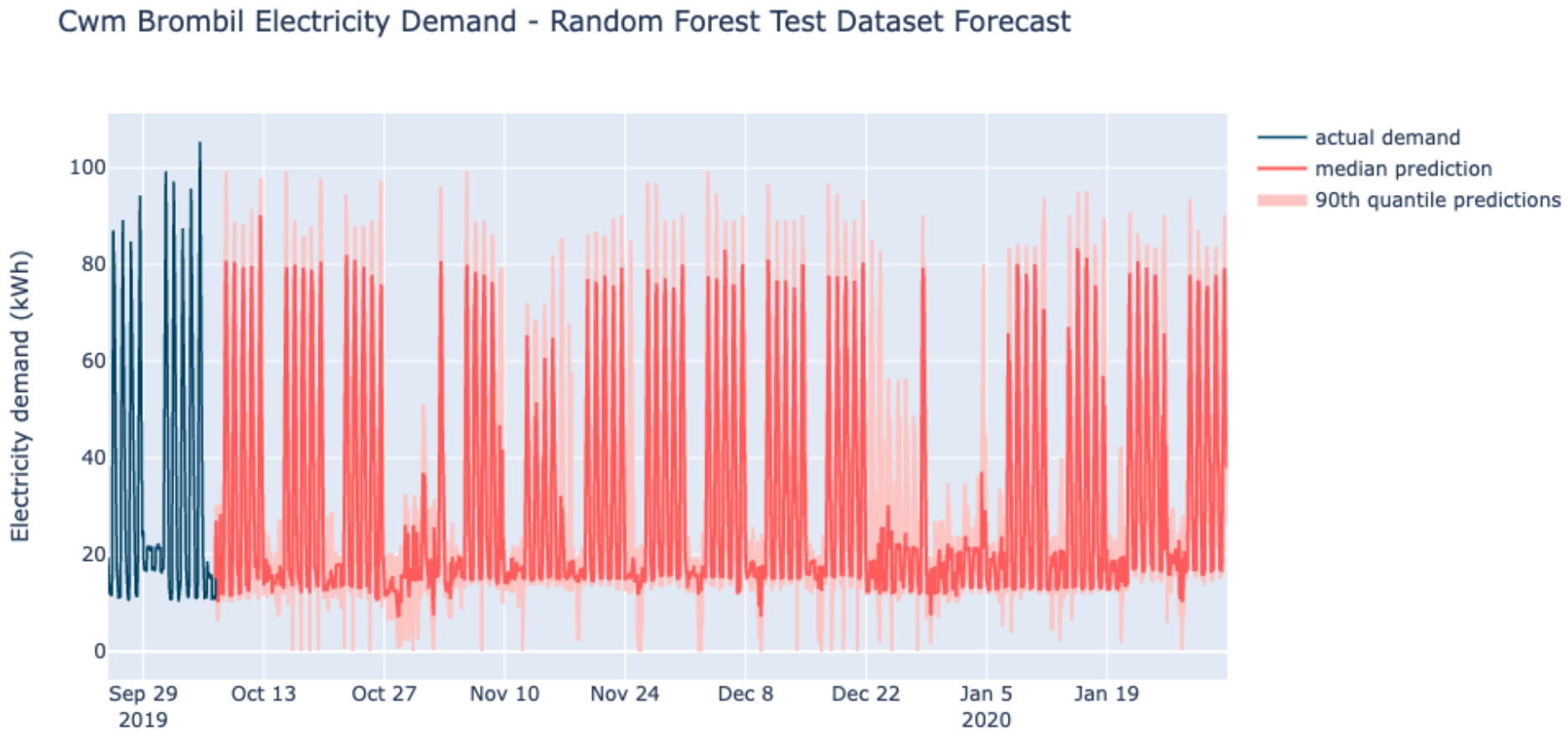


Figure #: Test dataset predictions using Random Forest regression model for entire test period

AI modeling – Test dataset predictions

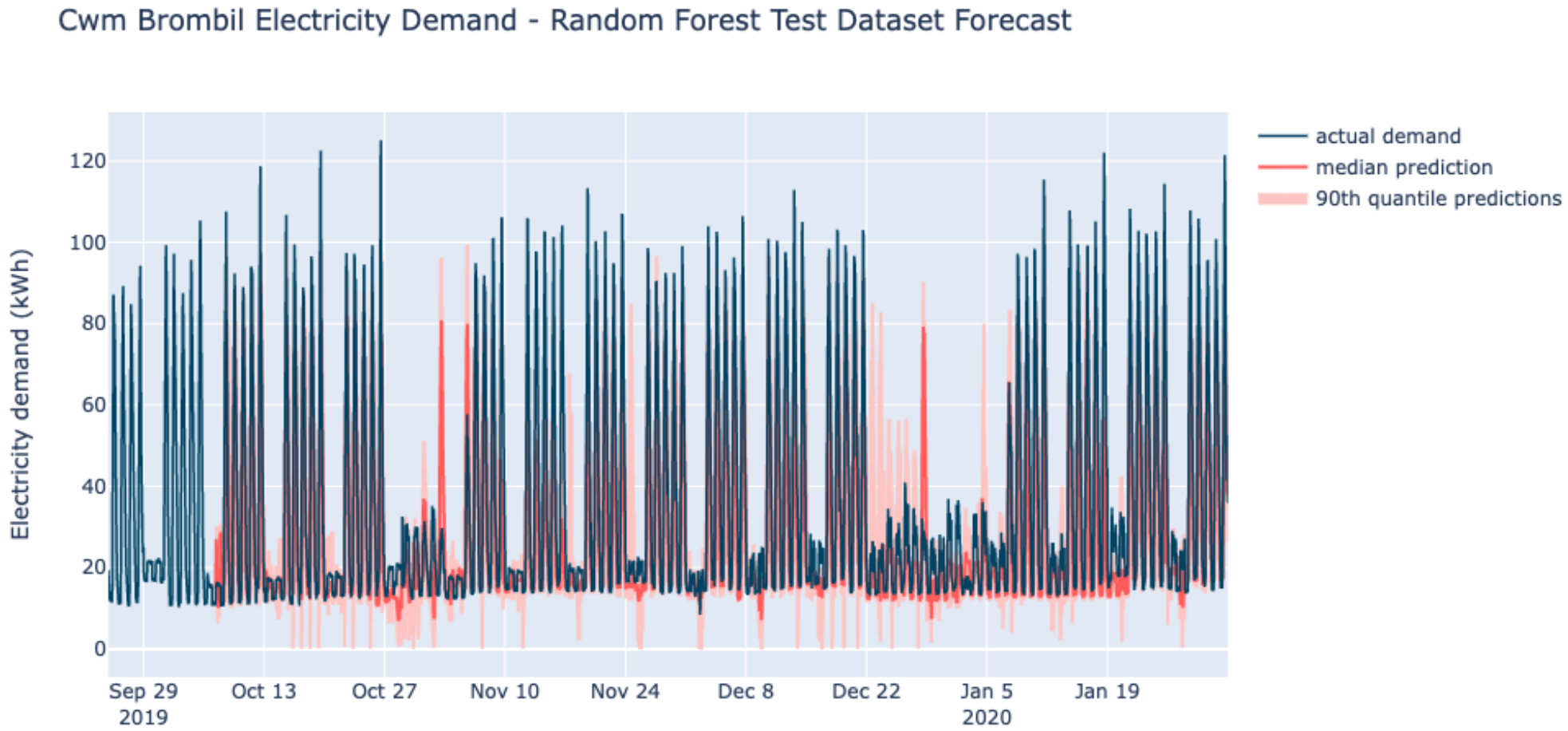


Figure #: Test dataset predictions using Random Forest regression model for entire test period

AI modeling – Test dataset predictions

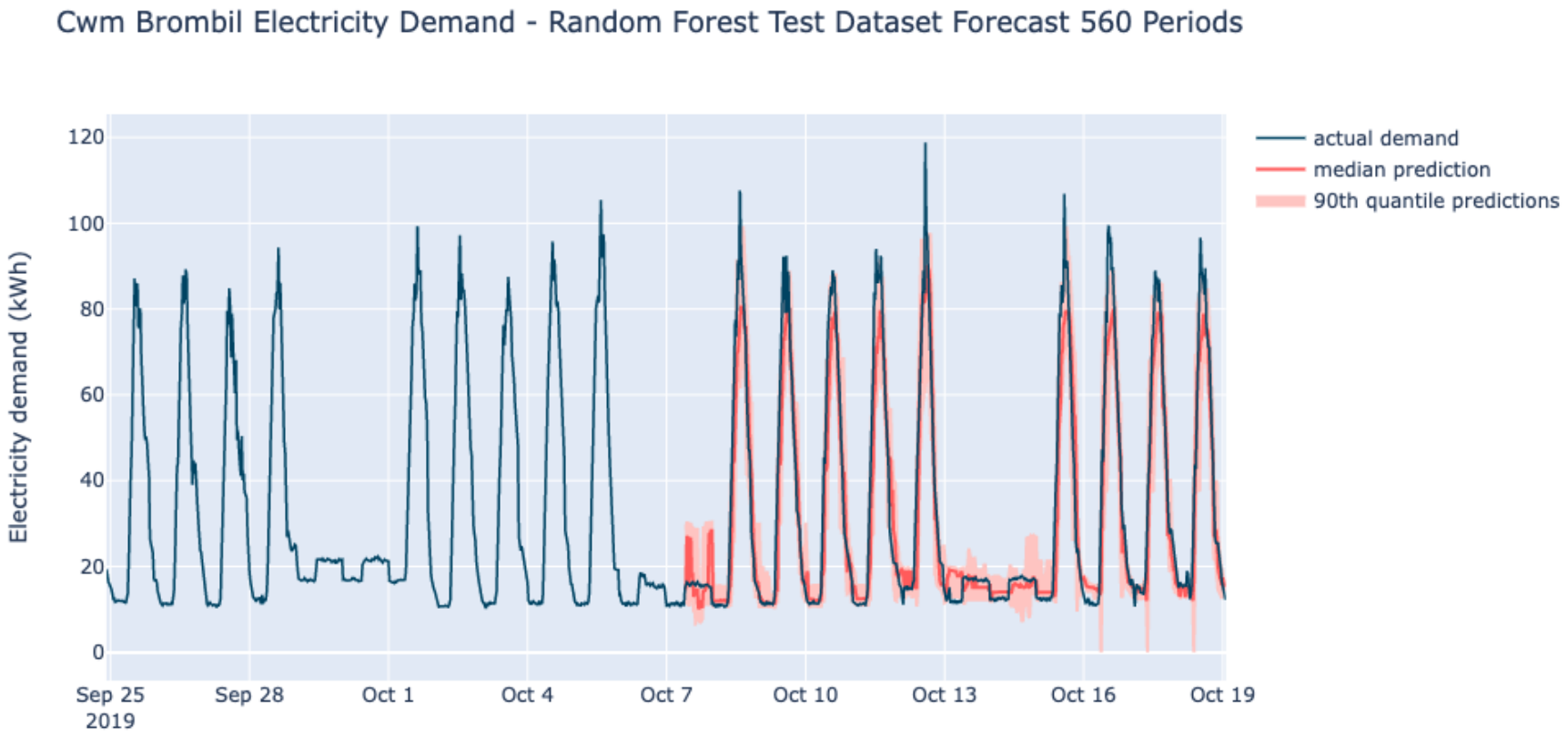


Figure : Test dataset predictions using Random Forest regression model on first 10 days of test data

AI modeling – Test dataset predictions

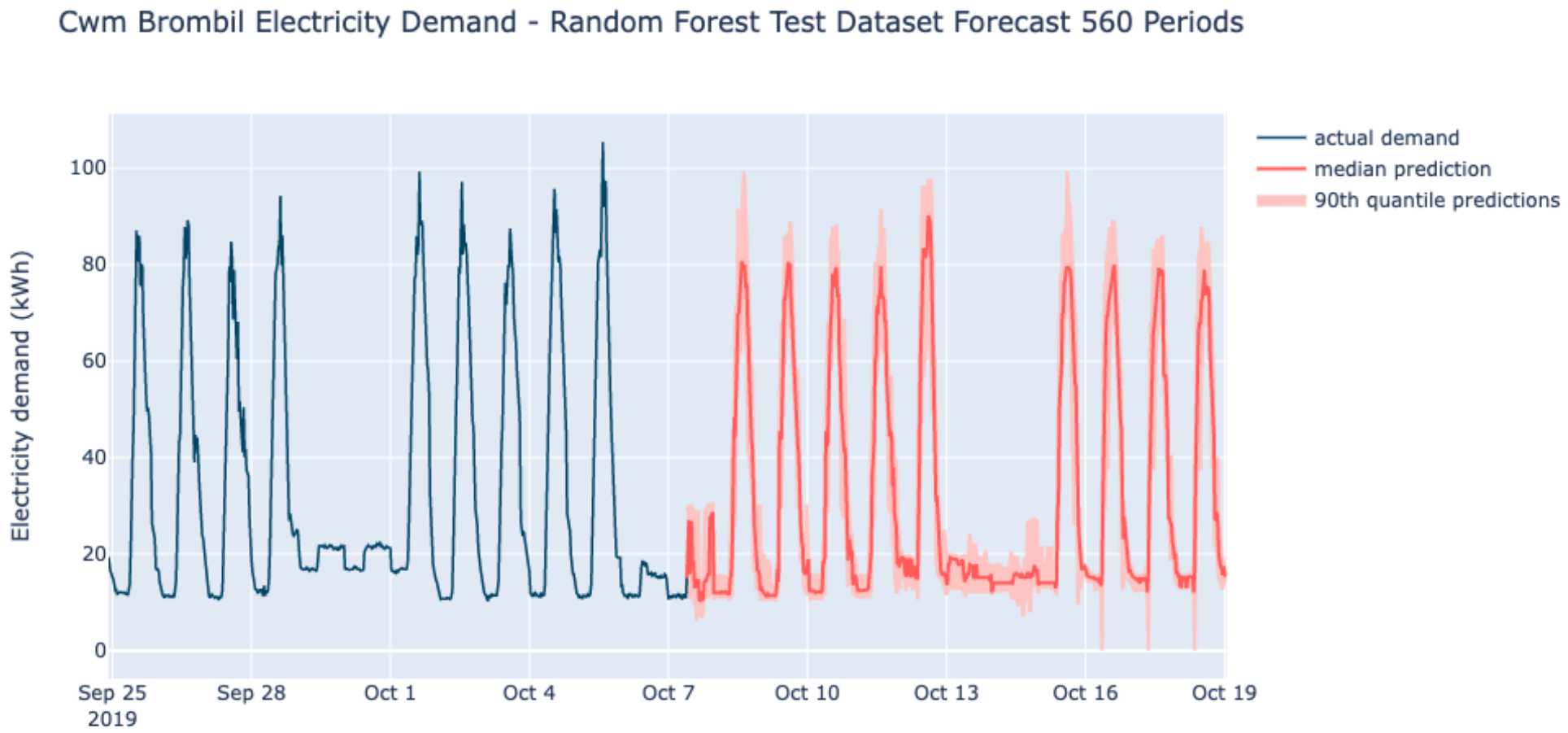


Figure #: Test dataset predictions using Random Forest regression model on first 10 days of test data

AI modeling – Best model feature importances

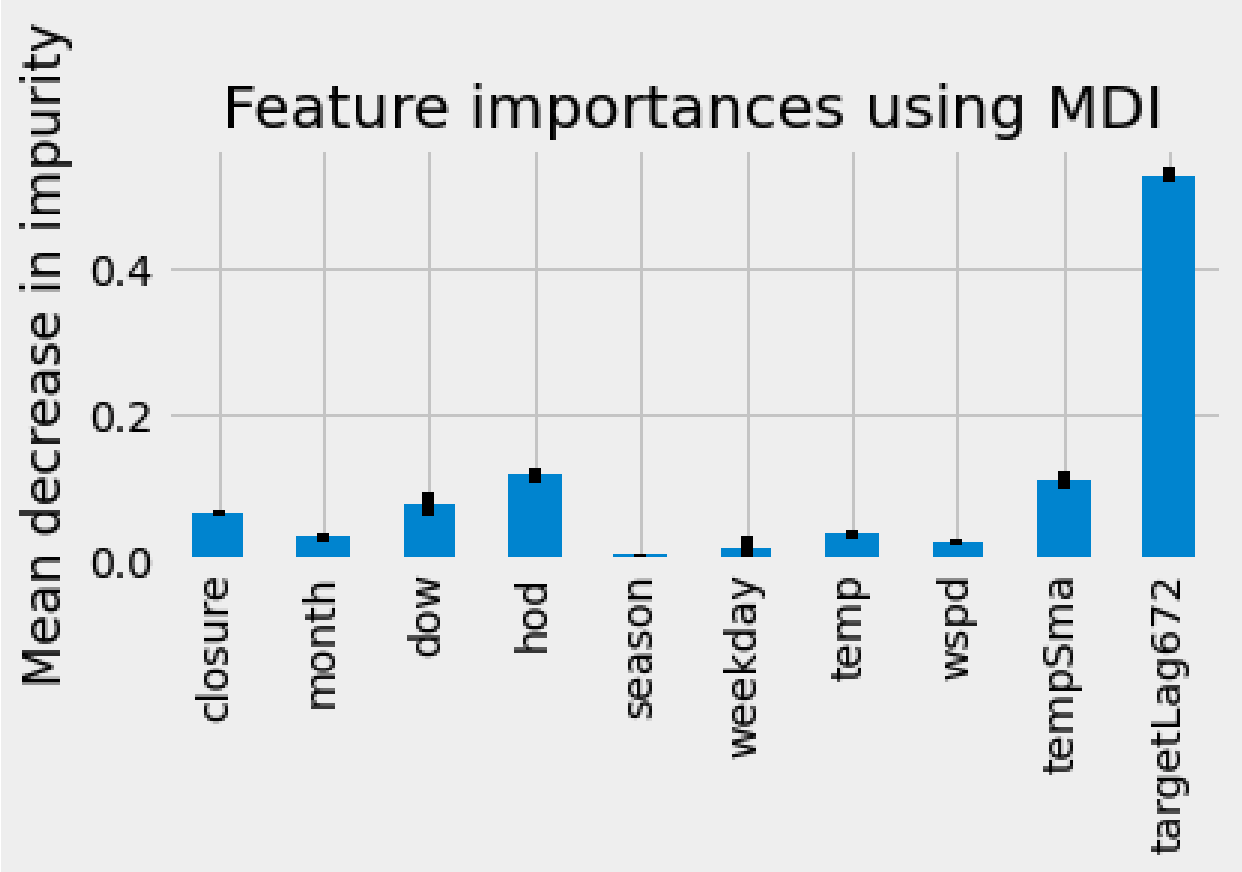
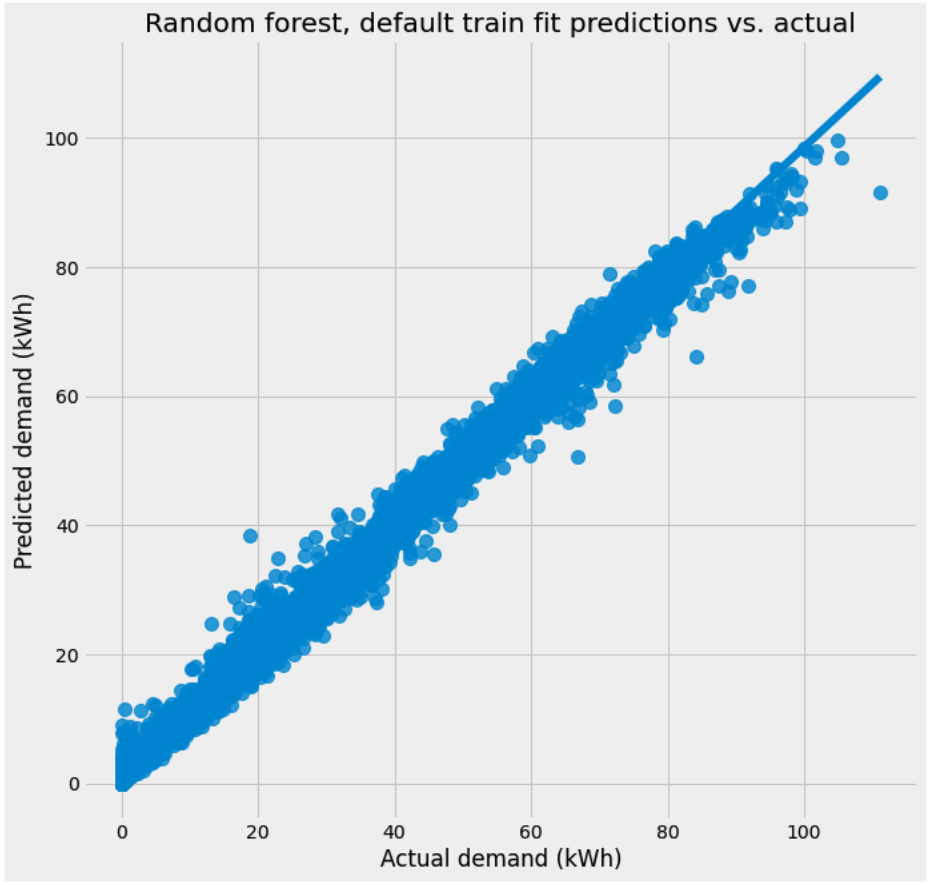


Figure #: Seasonal features Pearson correlation matrix heat map

AI modeling – Best model train/test fit



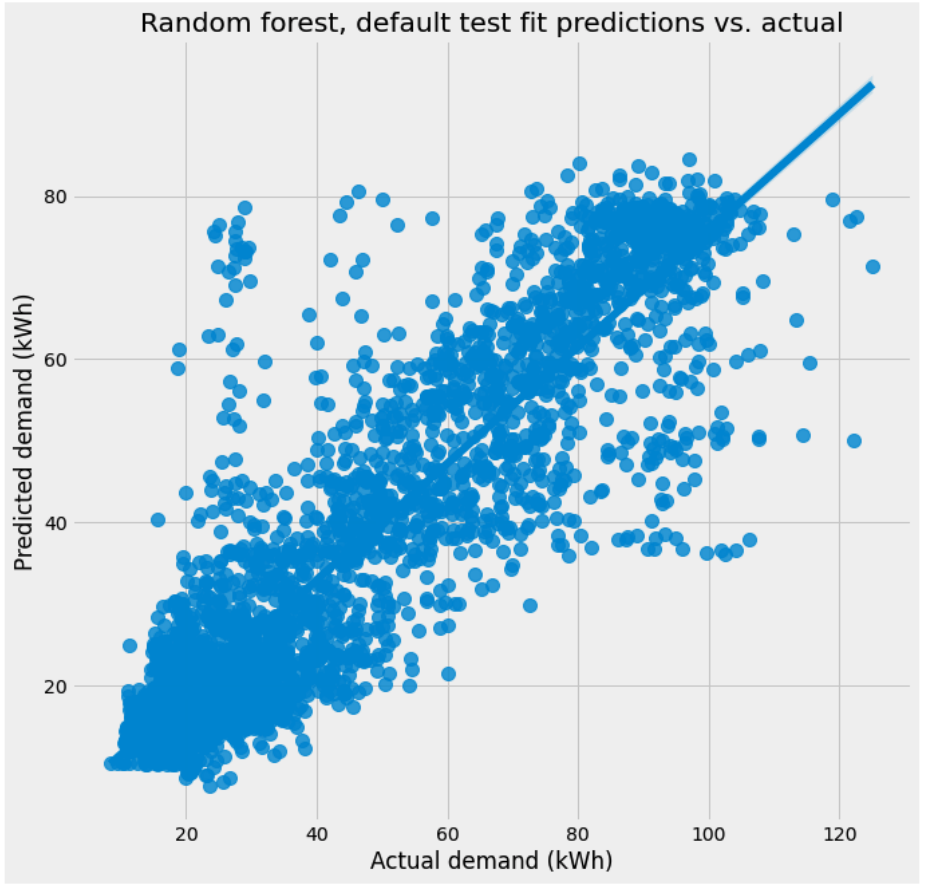


Figure #: Scatter plot showing predictions vs. observed demand for the training data (left) and the test dataset (left)

Questions