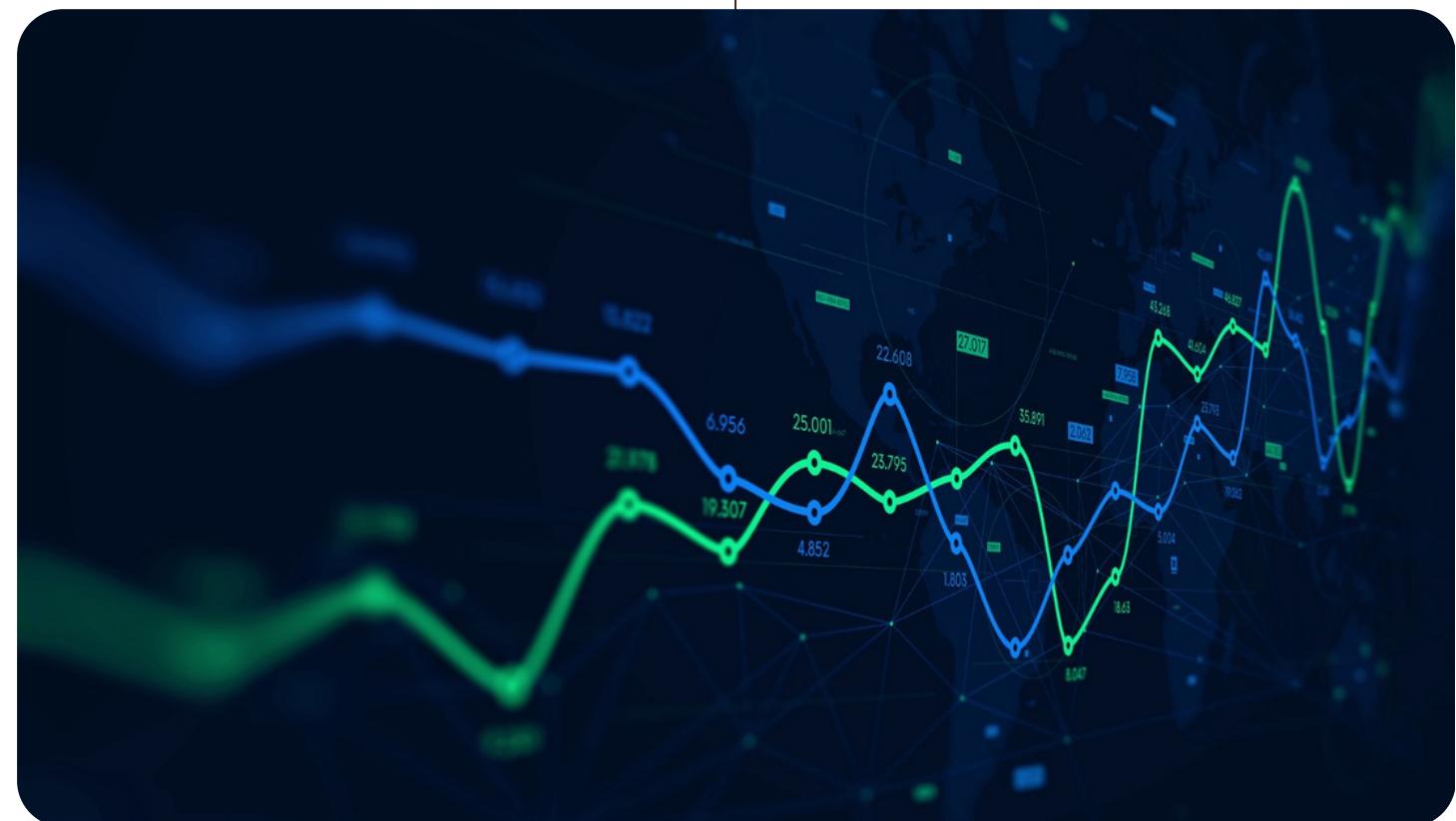


...

Time Series Project



Nicola Szwaja
Piotr Droś

...

Table of Contents

-
1. Introduction to dataset
 2. Exploring and understanding the data
 3. Data preprocessing
 4. **Models and results** 

...



Dataset overview

The dataset contains information about **how much electricity is used** in three distinct distribution networks of Tetouan city in northern Morocco.

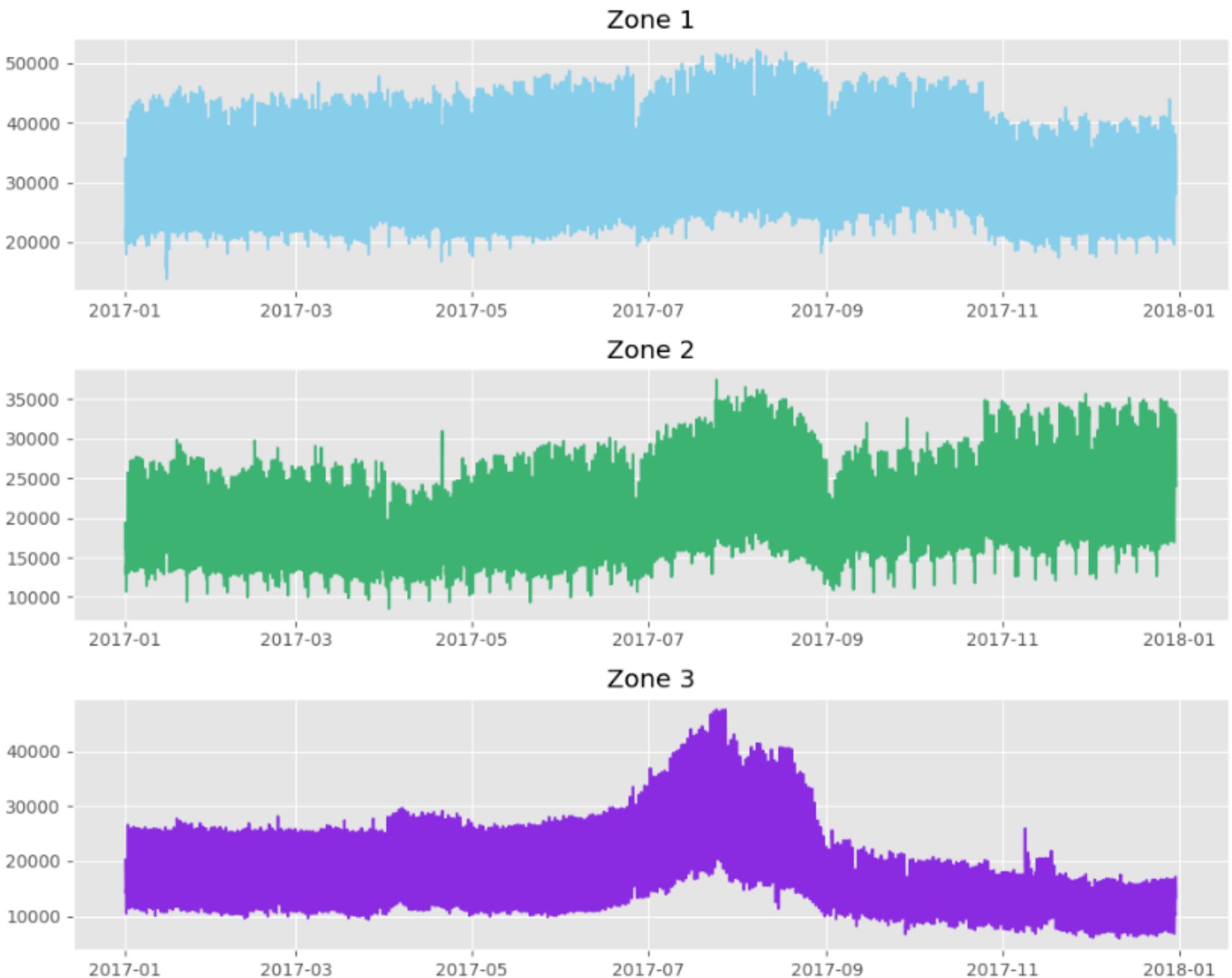


Dataset overview

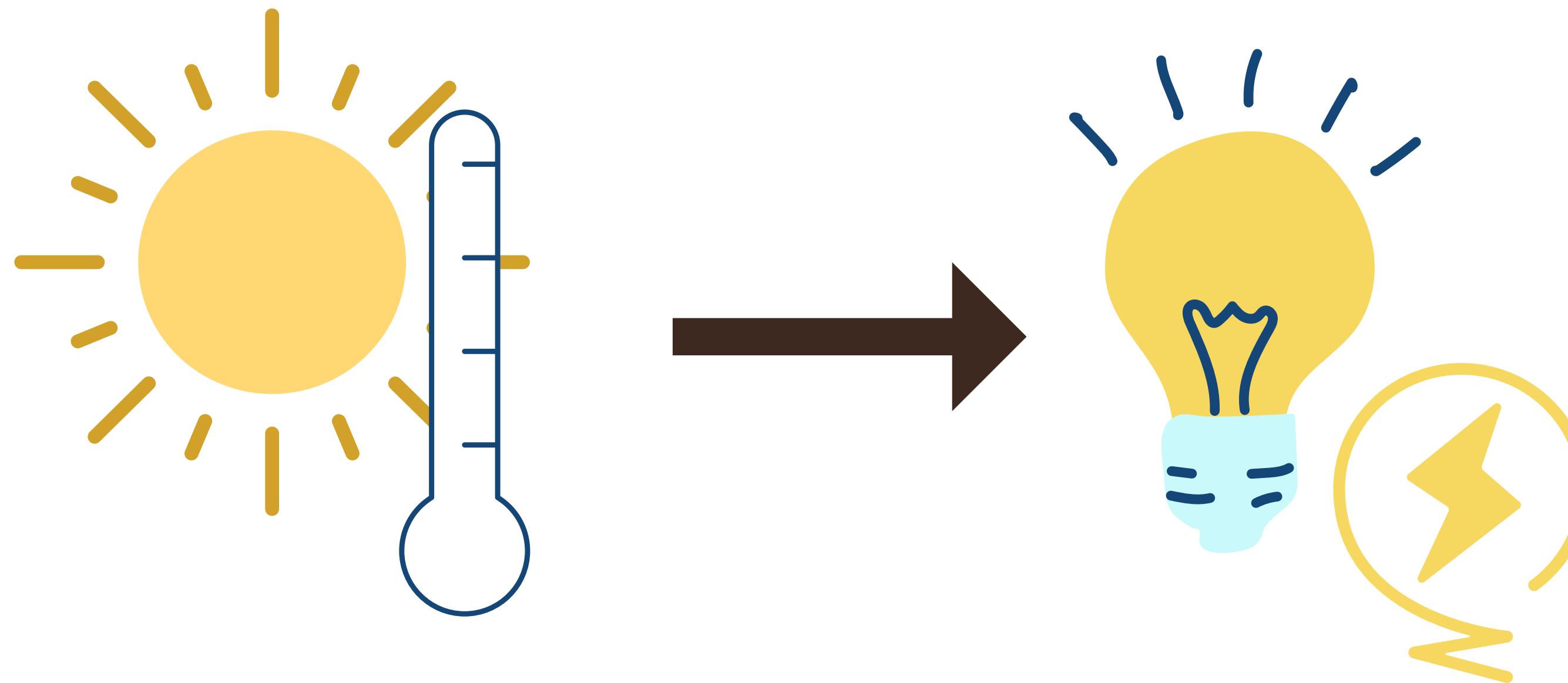
	Temperature	Humidity	Wind Speed	general diffuse flows	diffuse flows	Zone 1 Power Consumption	Zone 2 Power Consumption	Zone 3 Power Consumption
Datetime								
2017-01-01 00:00:00	6.559	73.8	0.083	0.051	0.119	34055.69620	16128.87538	20240.96386
2017-01-01 00:10:00	6.414	74.5	0.083	0.070	0.085	29814.68354	19375.07599	20131.08434
2017-01-01 00:20:00	6.313	74.5	0.080	0.062	0.100	29128.10127	19006.68693	19668.43373
2017-01-01 00:30:00	6.121	75.0	0.083	0.091	0.096	28228.86076	18361.09422	18899.27711
2017-01-01 00:40:00	5.921	75.7	0.081	0.048	0.085	27335.69620	17872.34043	18442.40964
...
2017-12-30 23:10:00	7.010	72.4	0.080	0.040	0.096	31160.45627	26857.31820	14780.31212
2017-12-30 23:20:00	6.947	72.6	0.082	0.051	0.093	30430.41825	26124.57809	14428.81152
2017-12-30 23:30:00	6.900	72.8	0.086	0.084	0.074	29590.87452	25277.69254	13806.48259
2017-12-30 23:40:00	6.758	73.0	0.080	0.066	0.089	28958.17490	24692.23688	13512.60504
2017-12-30 23:50:00	6.580	74.1	0.081	0.062	0.111	28349.80989	24055.23167	13345.49820

Feature Type	# Instances	# Features
Real, Integer	52417	6

Dataset overview



What will the model predict?



...

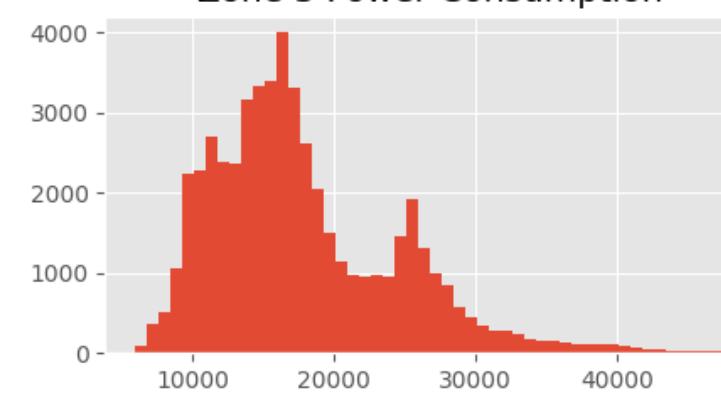
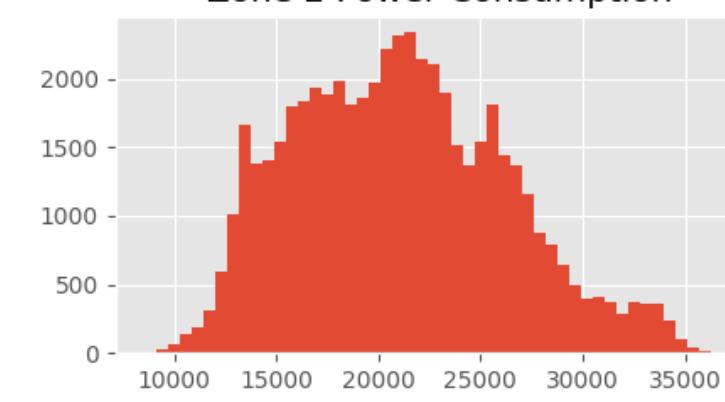
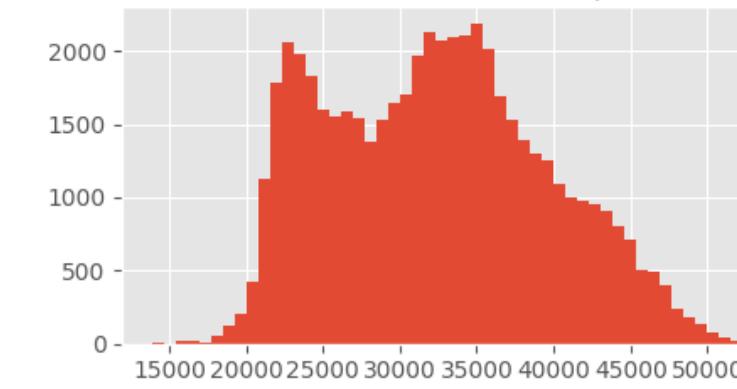
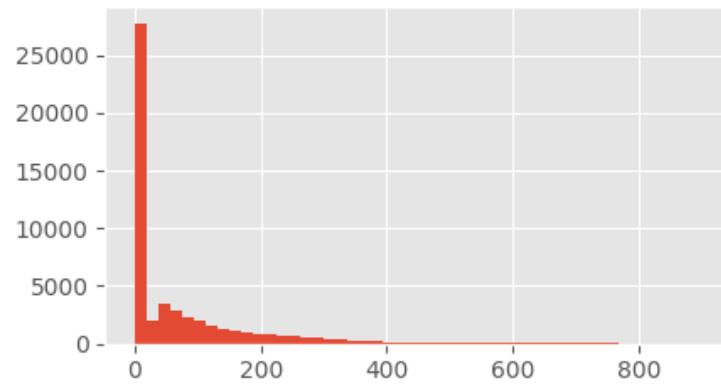
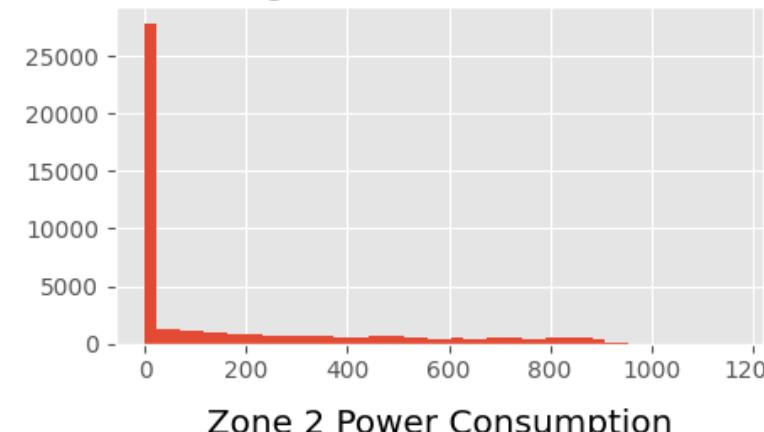
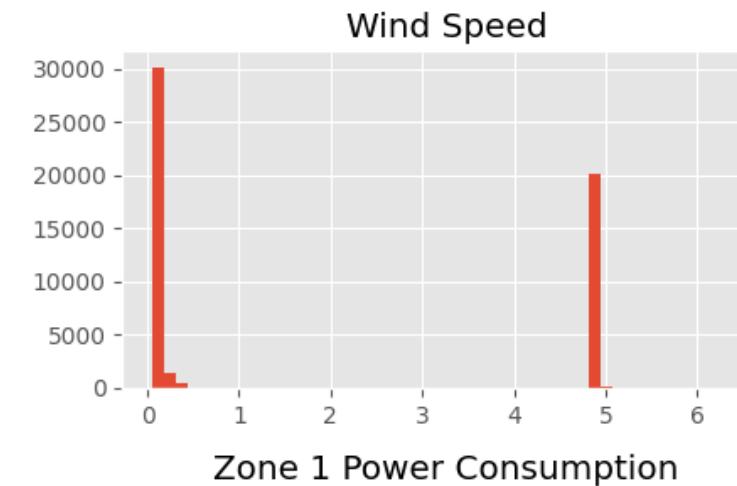
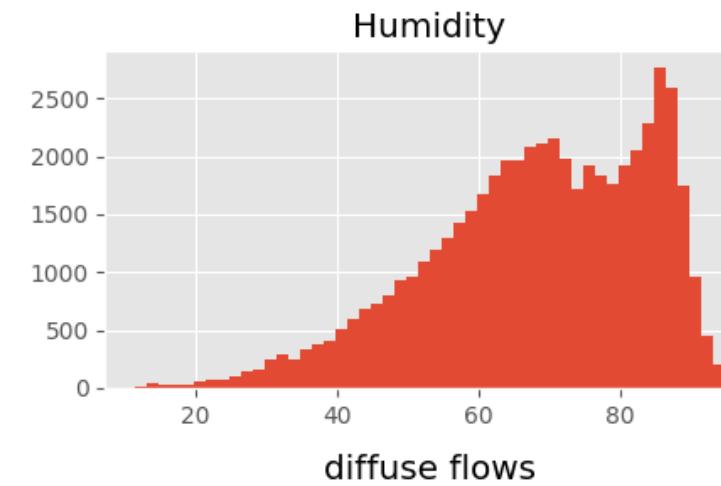
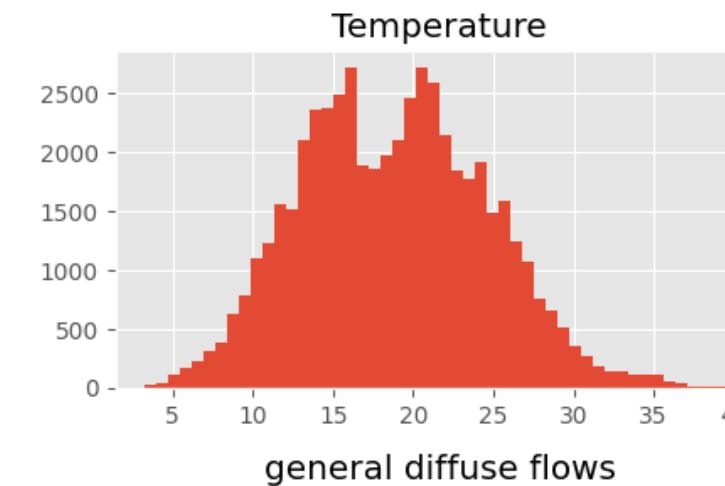
Data preprocessing



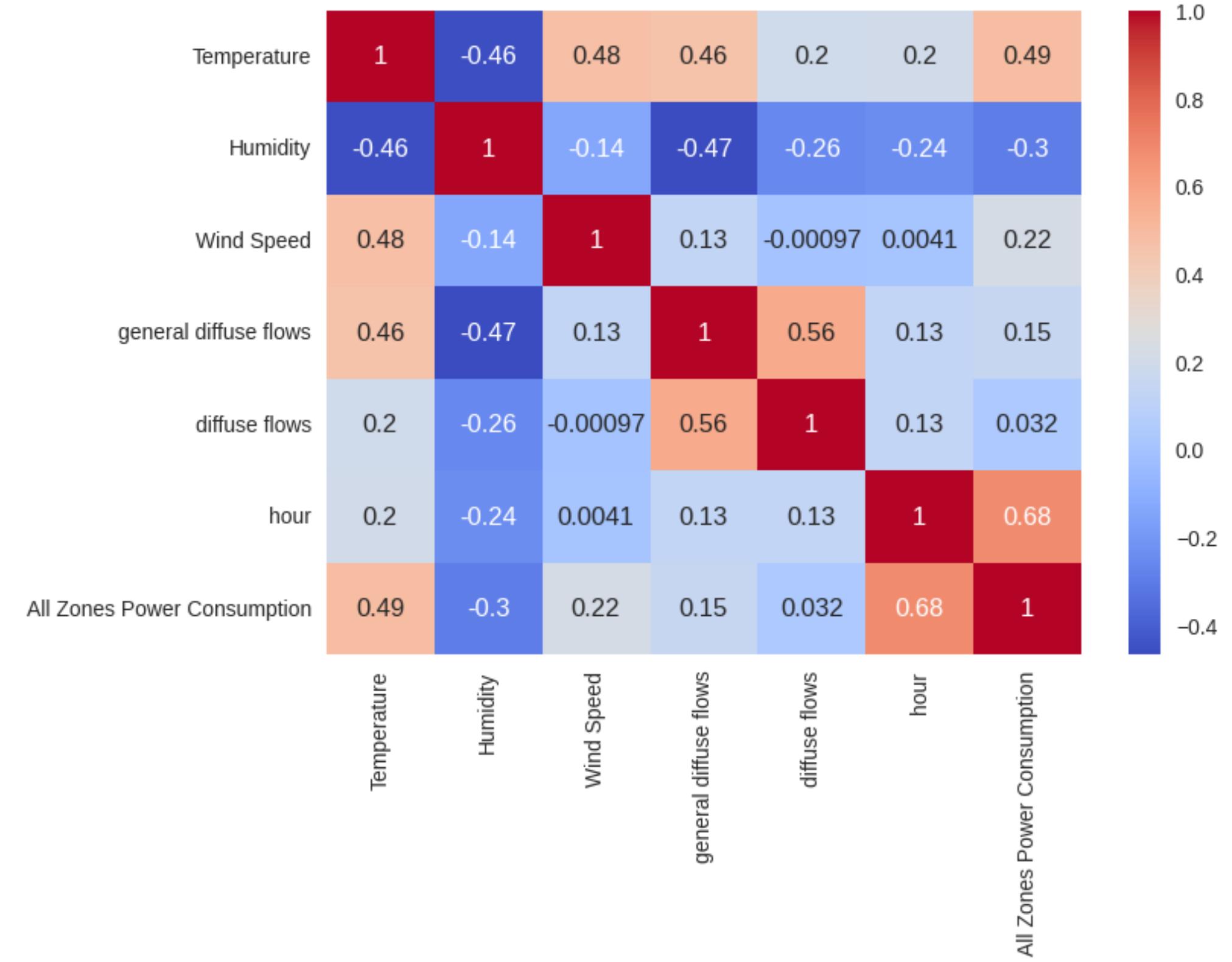
Feature Engineering

hour	dayofweek	month	All Zones Power Consumption
0	6	1	70425.53544
0	6	1	69320.84387
0	6	1	67803.22193
0	6	1	65489.23209
0	6	1	63650.44627
...
23	5	12	72798.08659
23	5	12	70983.80786
23	5	12	68675.04965
23	5	12	67163.01682
23	5	12	65750.53976

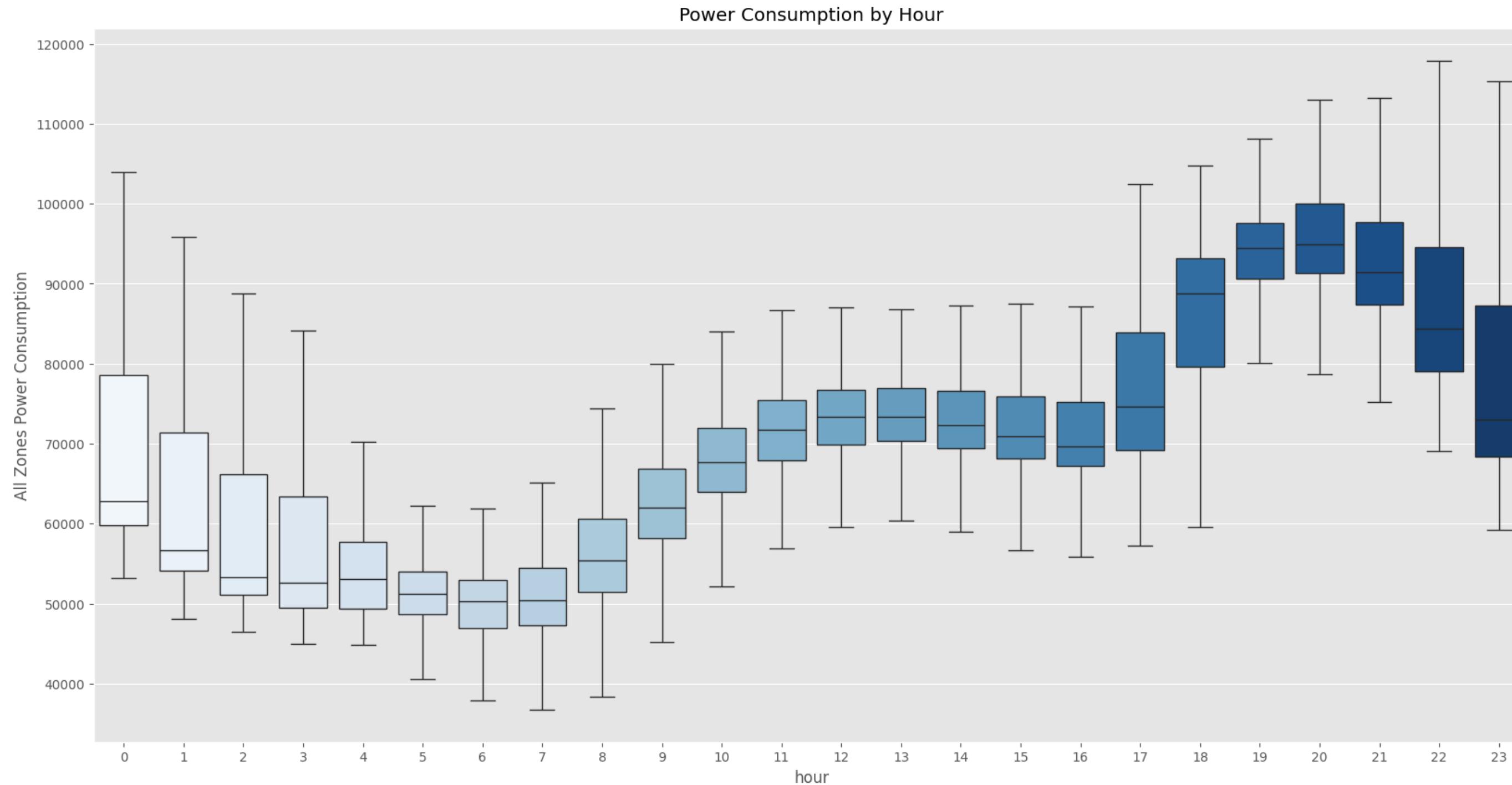
Distribution of data



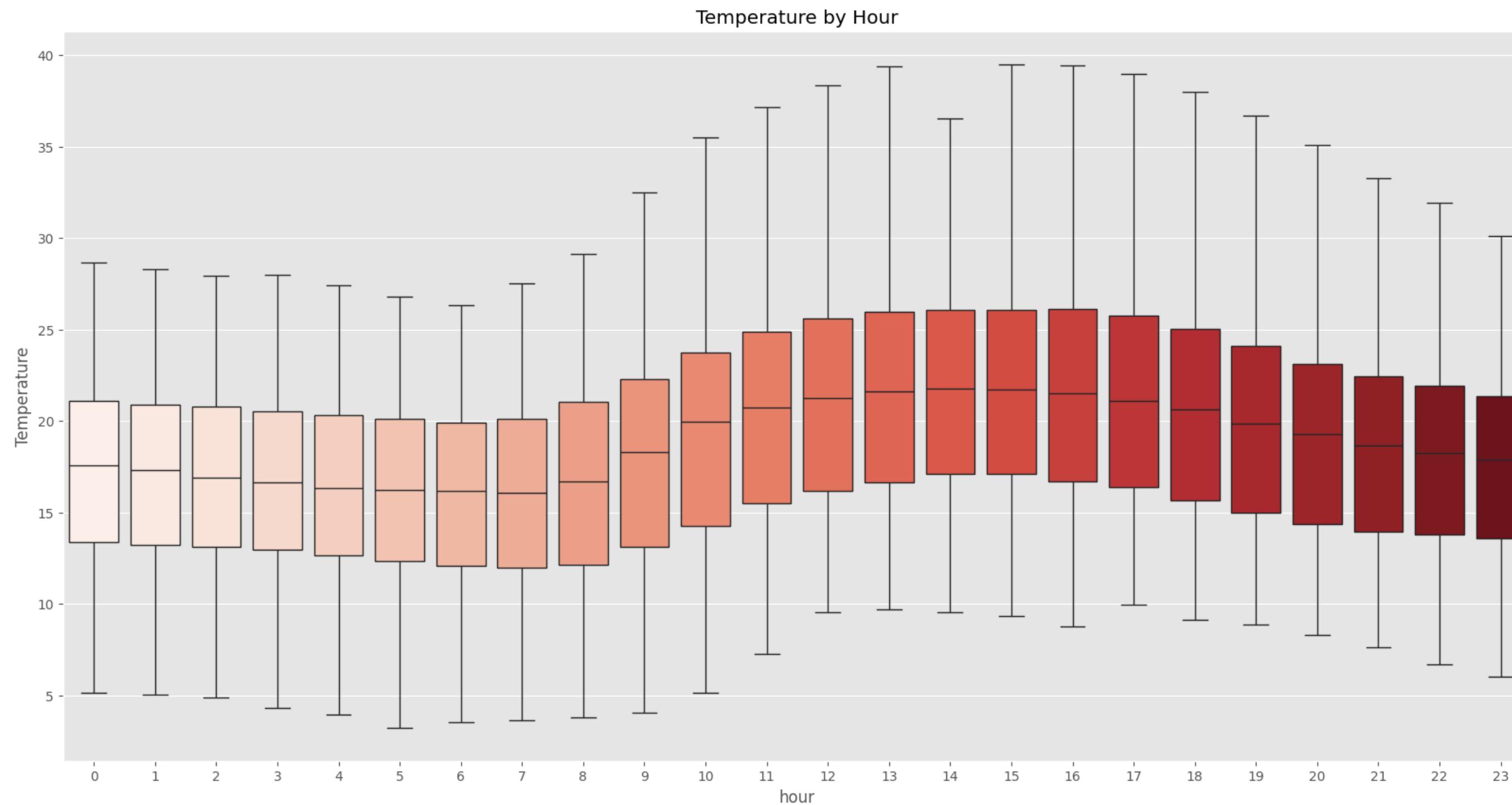
Feature Correlation



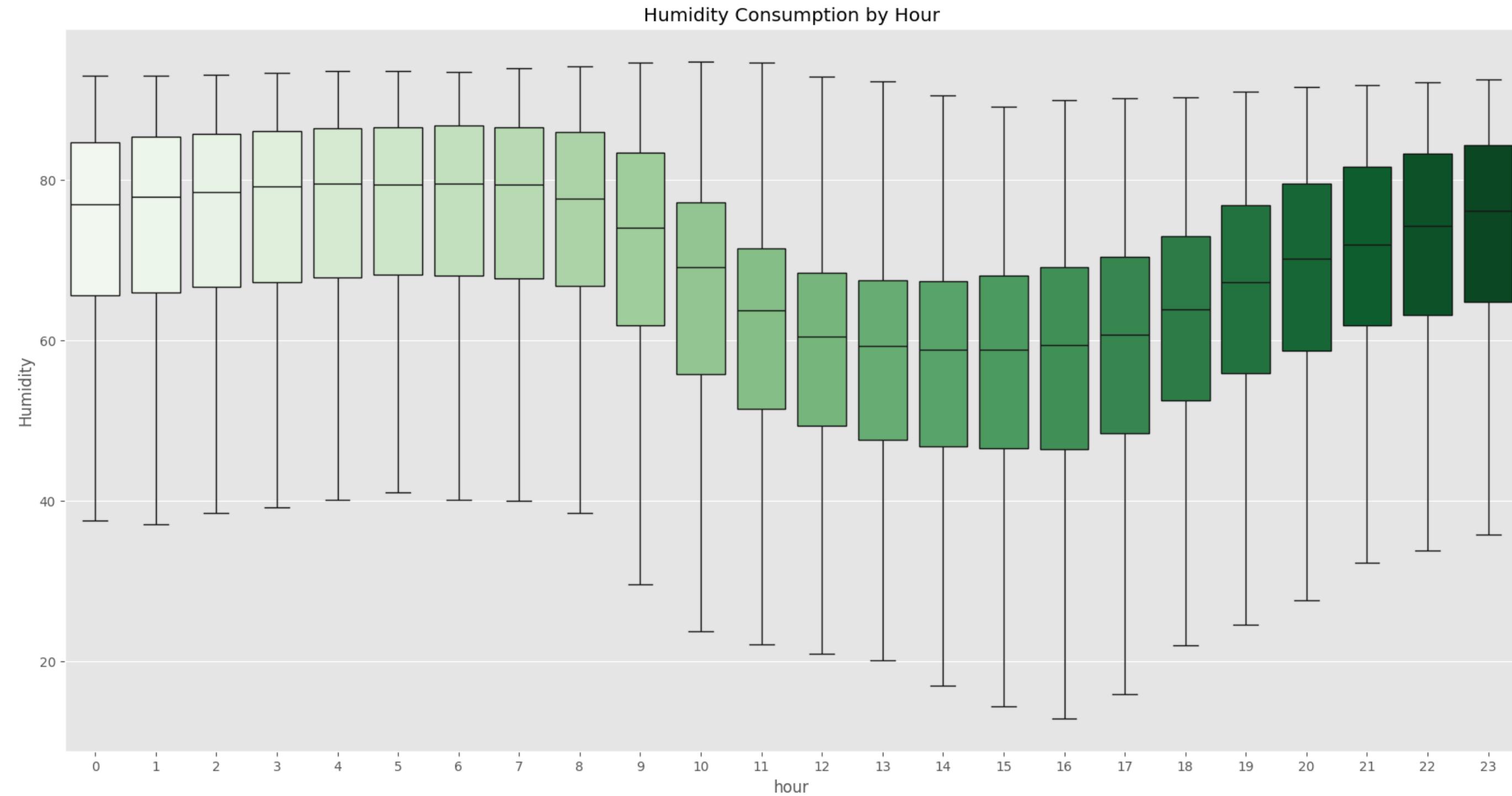
Power Consumption by Hour



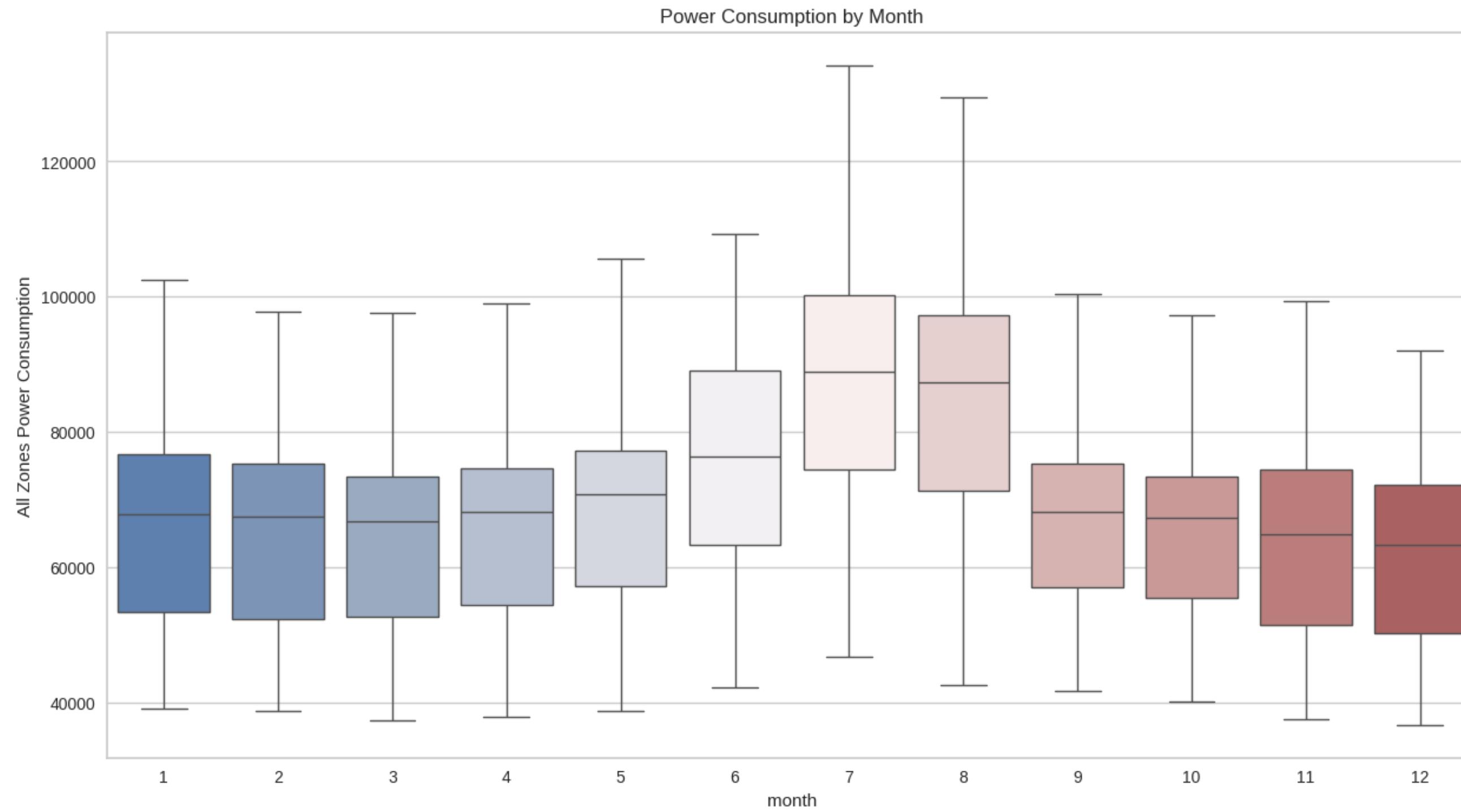
Temperature by Hour



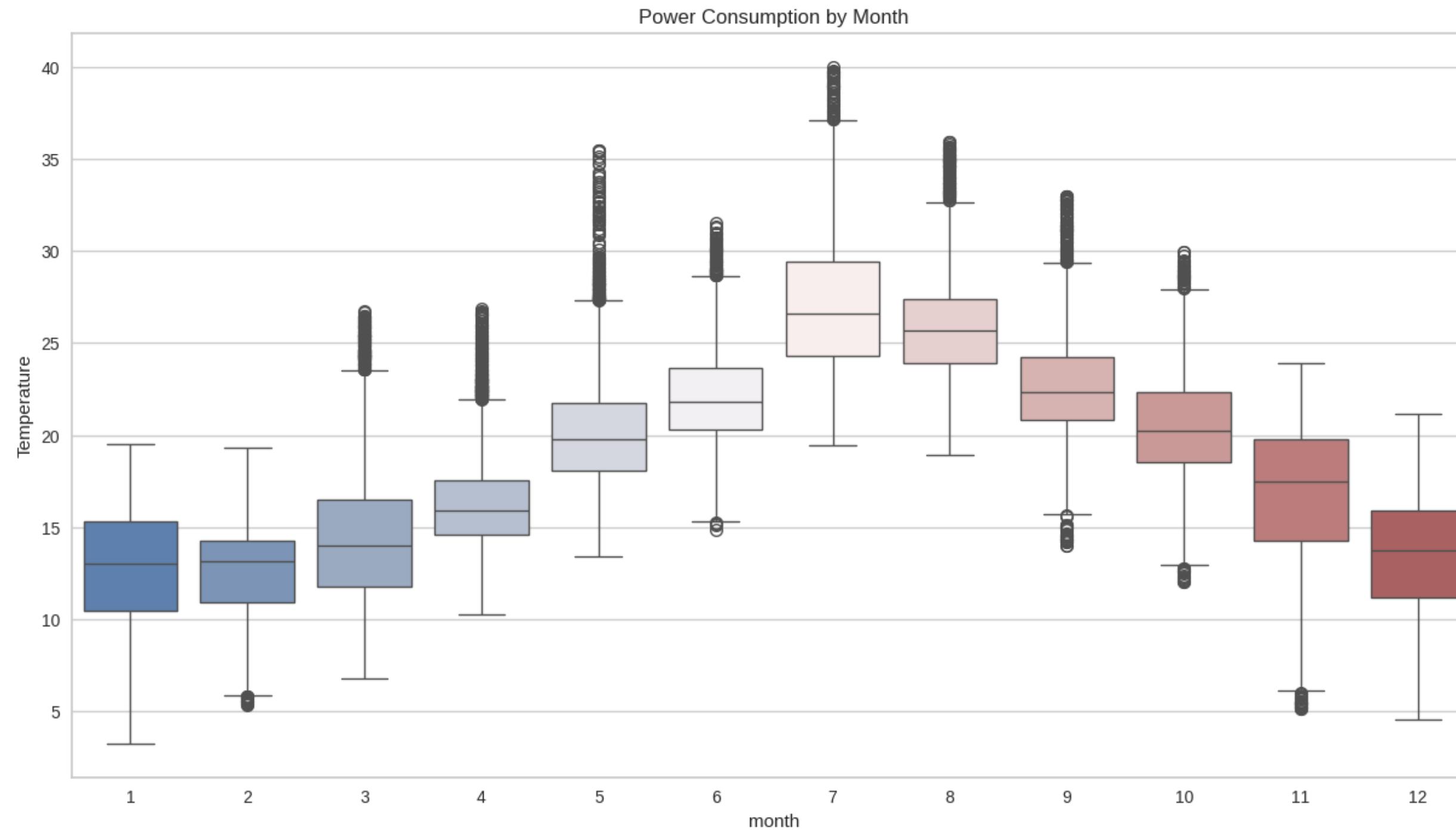
Humidity by Hour



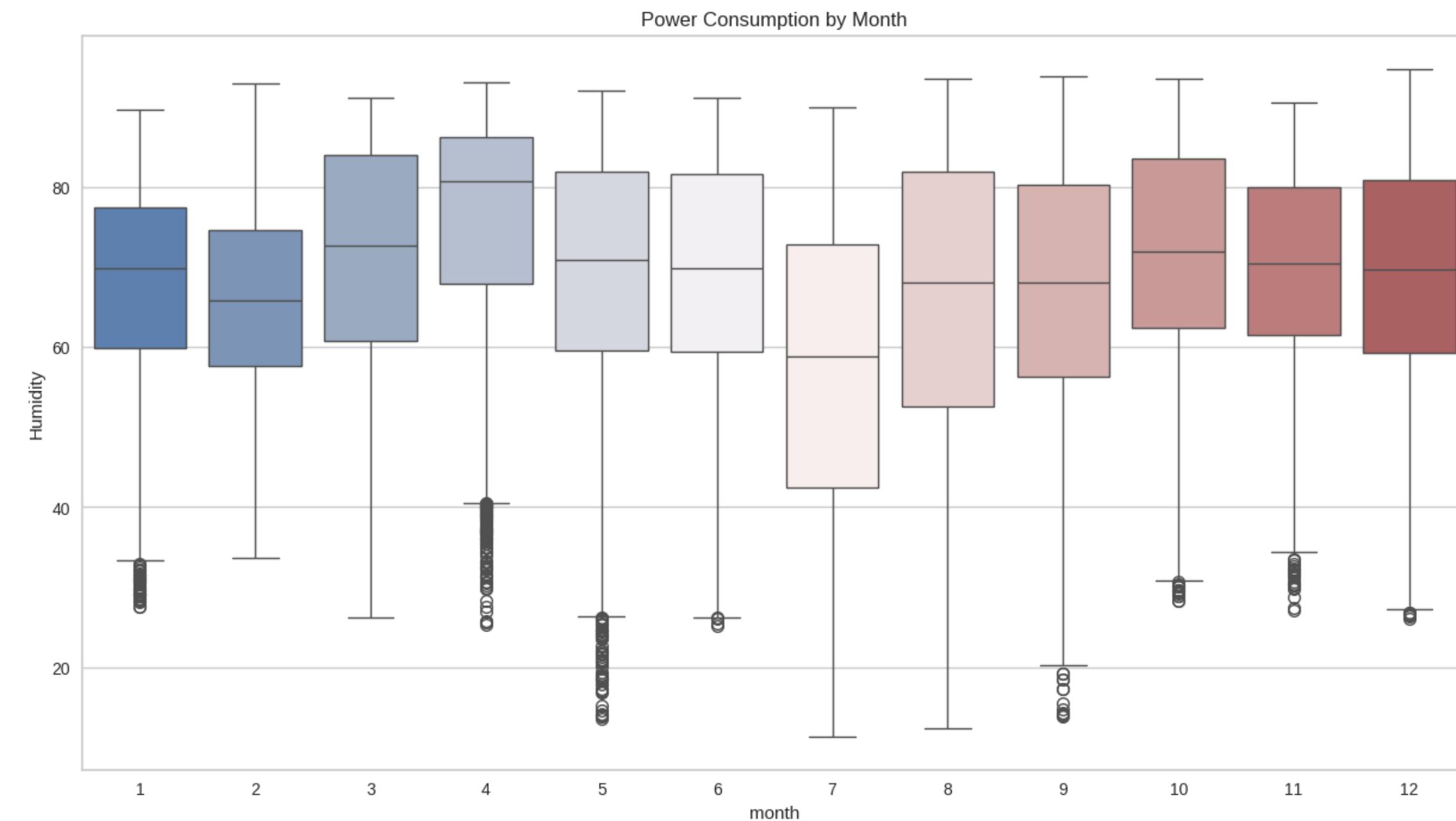
Power Consumption by Month



Temperature by Month



Humidity by Month



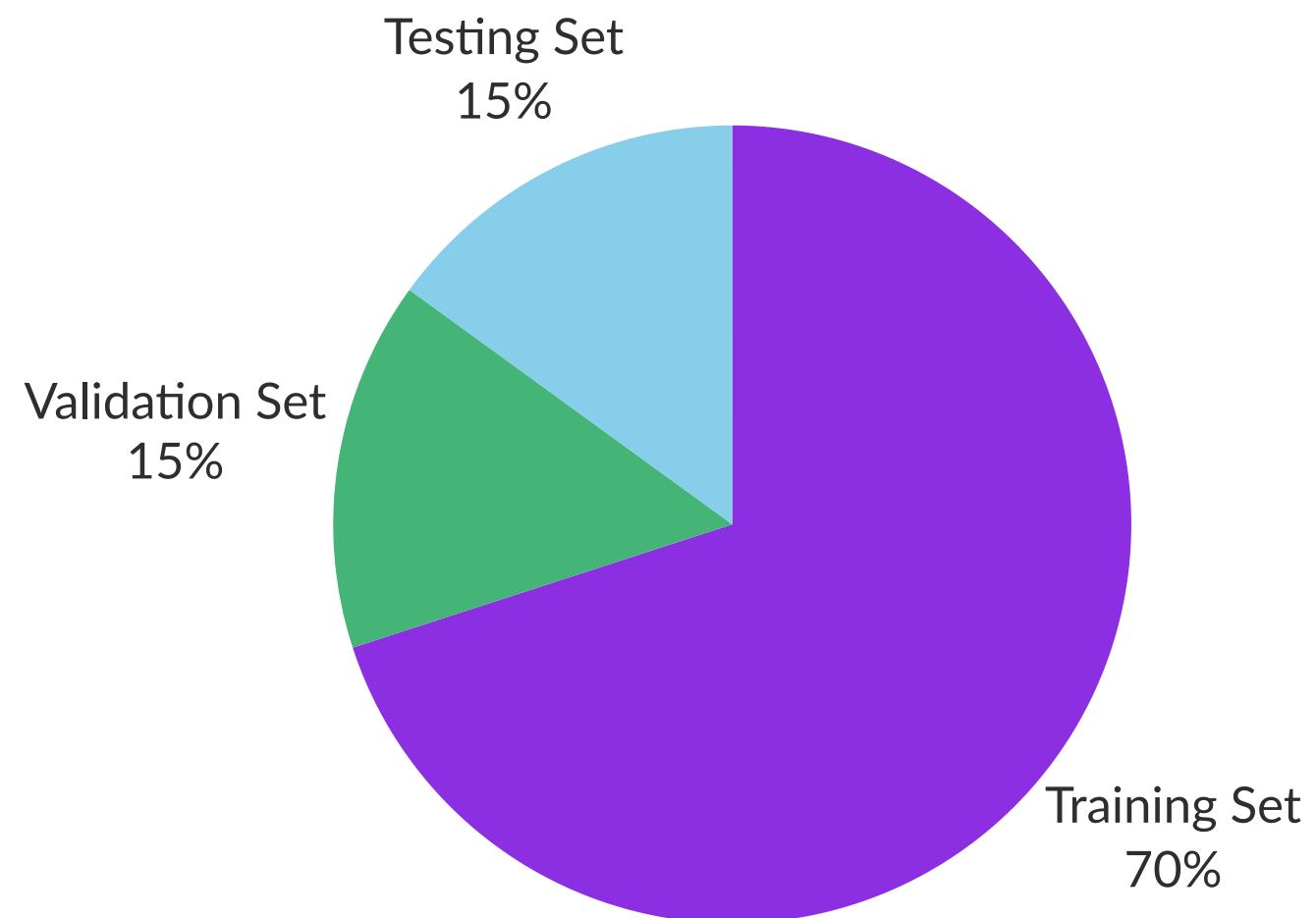
No missing values found in the dataset!

```
Data columns (total 8 columns):
```

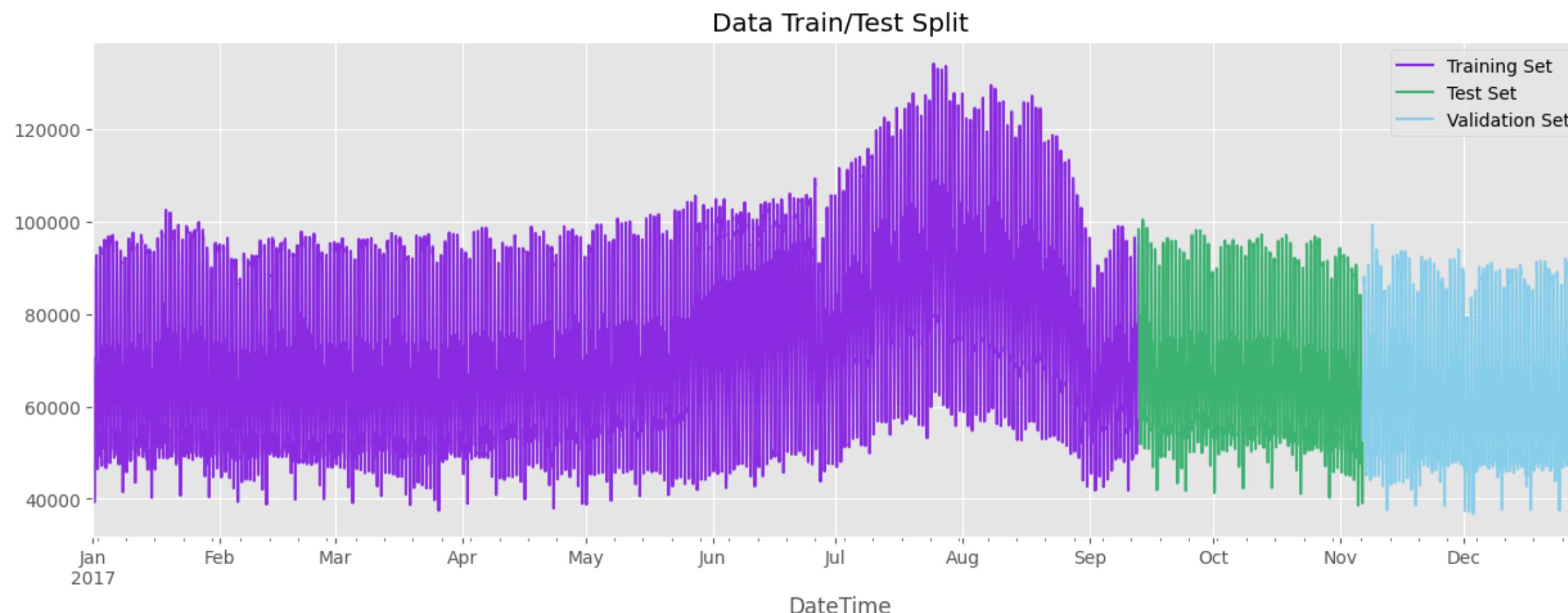
#	Column	Non-Null Count	Dtype
---	---	-----	----
0	Temperature	52416 non-null	float64
1	Humidity	52416 non-null	float64
2	Wind Speed	52416 non-null	float64
3	general diffuse flows	52416 non-null	float64
4	diffuse flows	52416 non-null	float64
5	Zone 1 Power Consumption	52416 non-null	float64
6	Zone 2 Power Consumption	52416 non-null	float64
7	Zone 3 Power Consumption	52416 non-null	float64



Data split



Data split



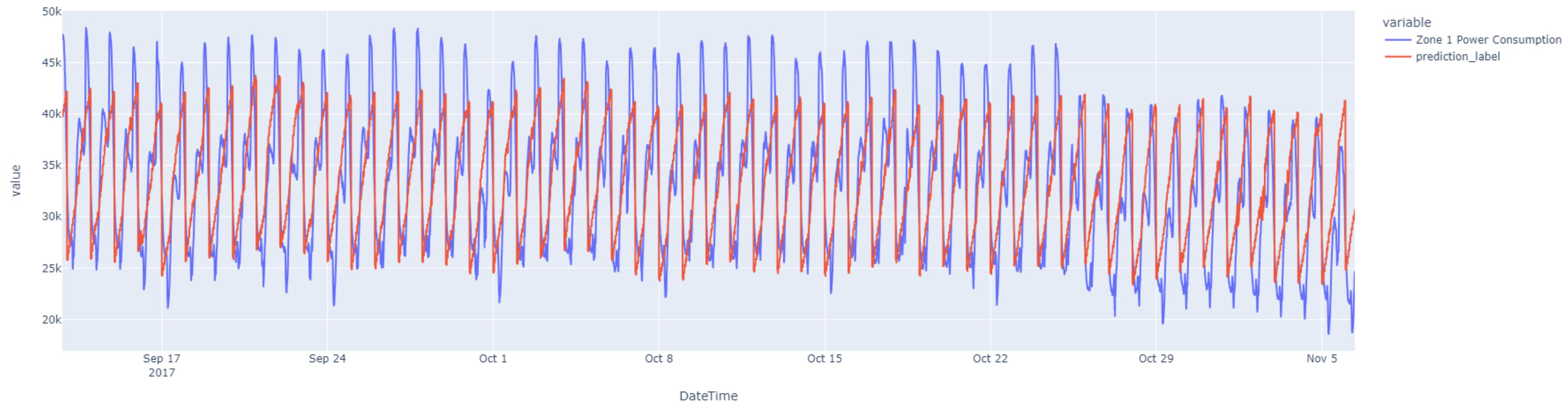
• • •

Models



•••

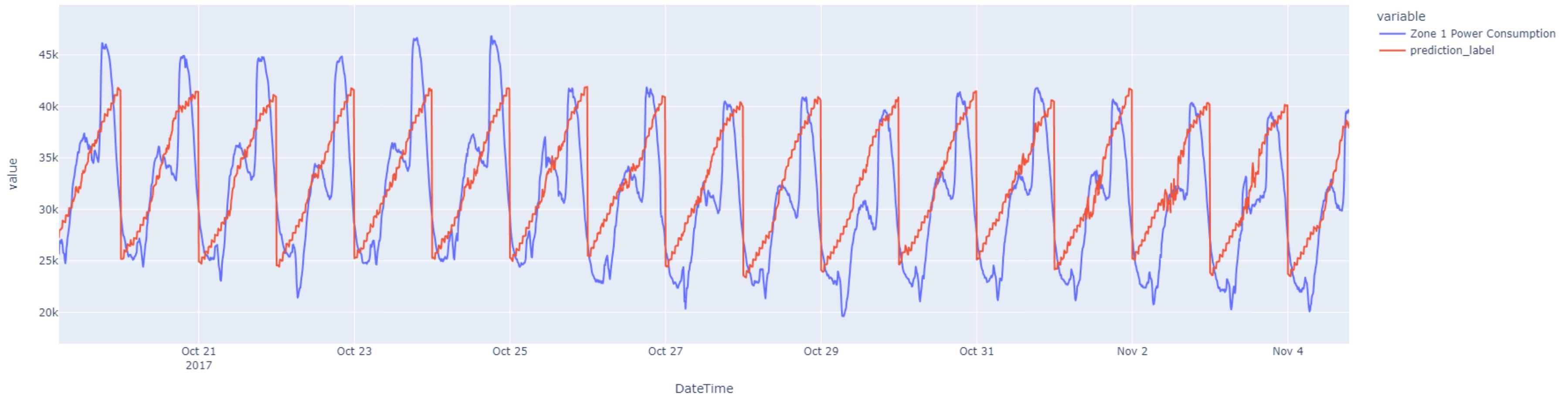
Linear Regression



Model	MAE	MSE	RMSE	R2	RMSLE	MAPE
Linear Regression	3183.1628	16310749.0000	4038.6567	0.6278	0.1216	0.1001

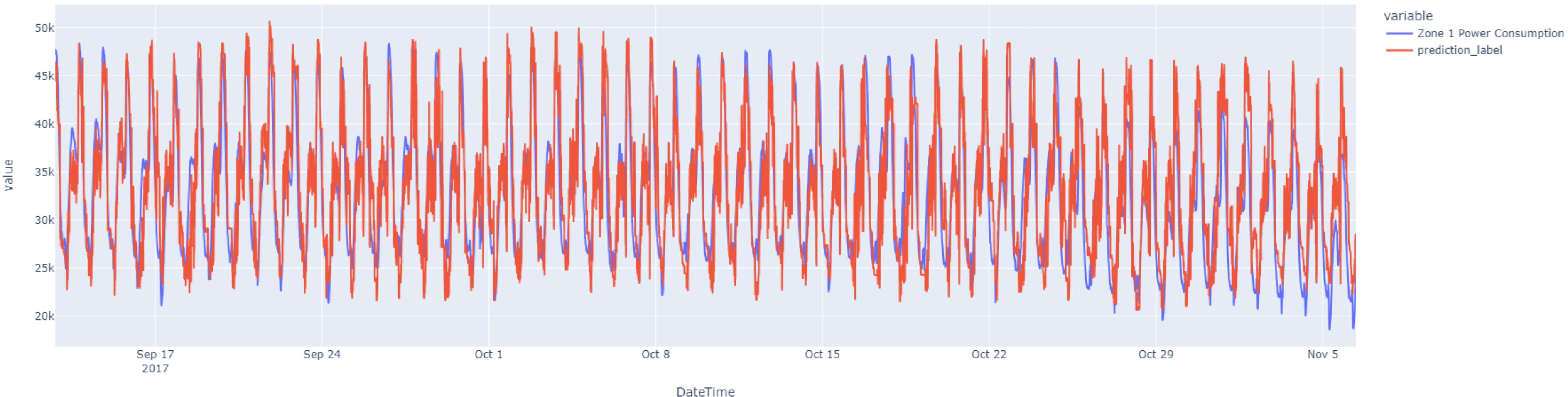
...

Linear regression - closeup



•••

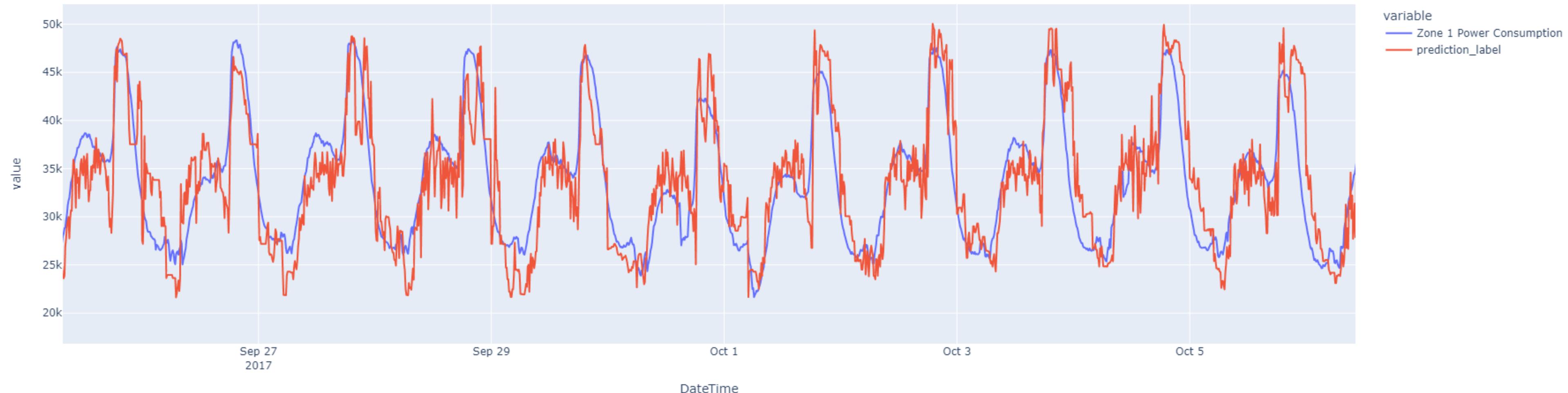
KNN



	Model	MAE	MSE	RMSE	R2	RMSLE	MAPE
0	K Neighbors Regressor	3056.5813	15741823.0000	3967.5964	0.6408	0.1217	0.0959

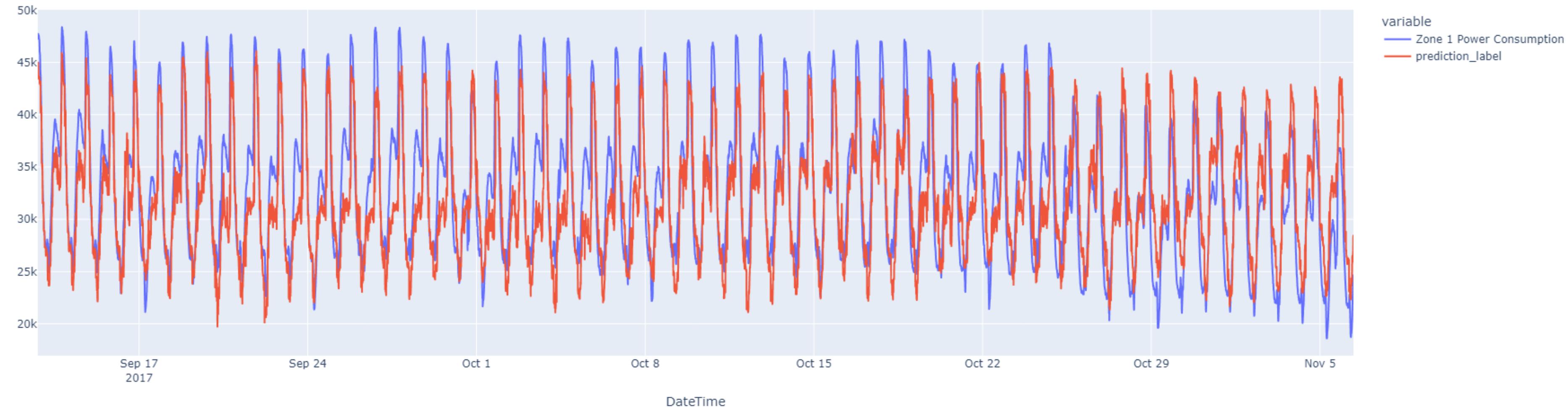
•••

KNN - closeup



•••

XGBoost



Model	MAE	MSE	RMSE	R2	RMSLE	MAPE
Extreme Gradient Boosting	3106.6379	14996891.0000	3872.5820	0.6578	0.1190	0.0954

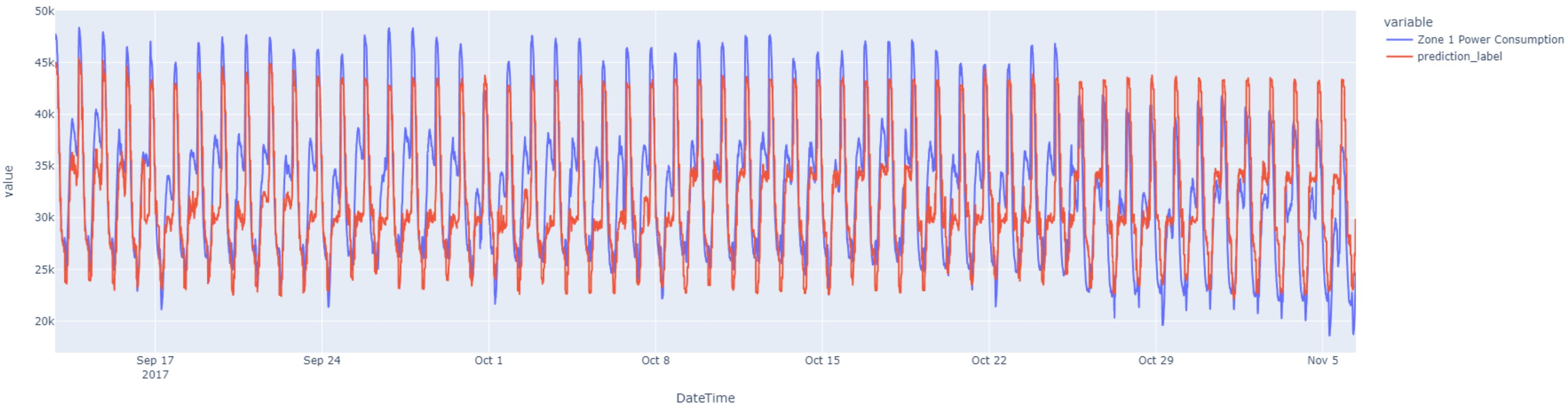
•••

XGBoost - closeup



...

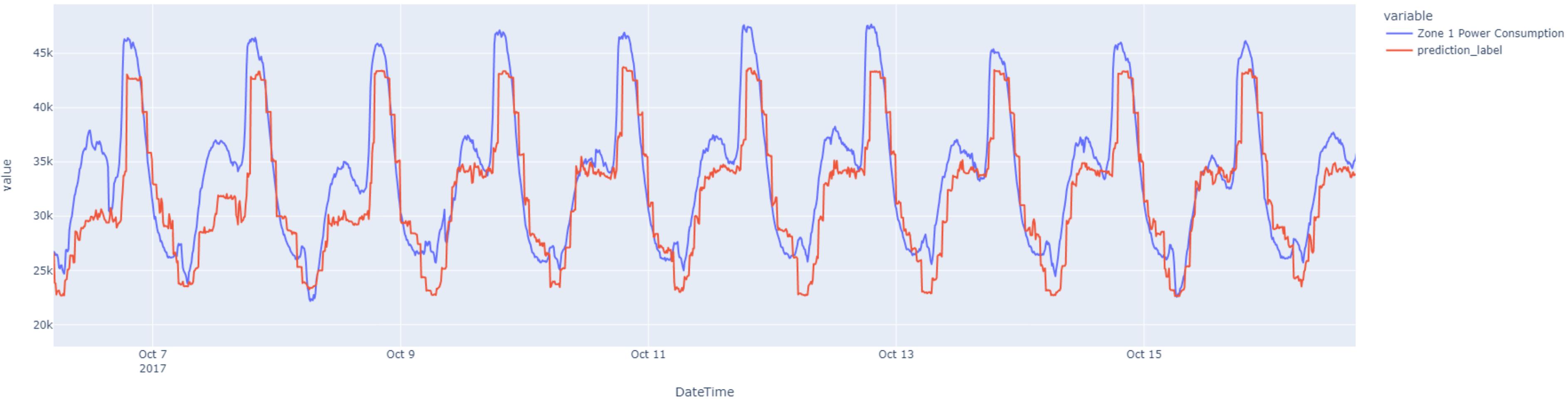
Light Gradient Boost



Model	MAE	MSE	RMSE	R2	RMSLE	MAPE
Light Gradient Boosting Machine	3185.2947	15801996.5946	3975.1725	0.6394	0.1214	0.0971

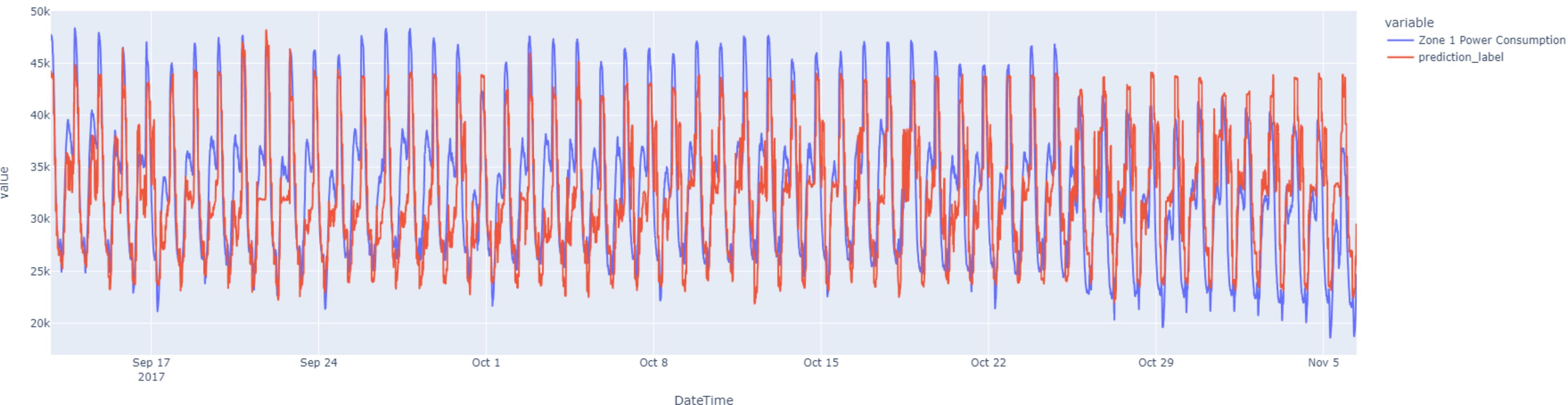
...

Light Gradient Boost - closeup



...

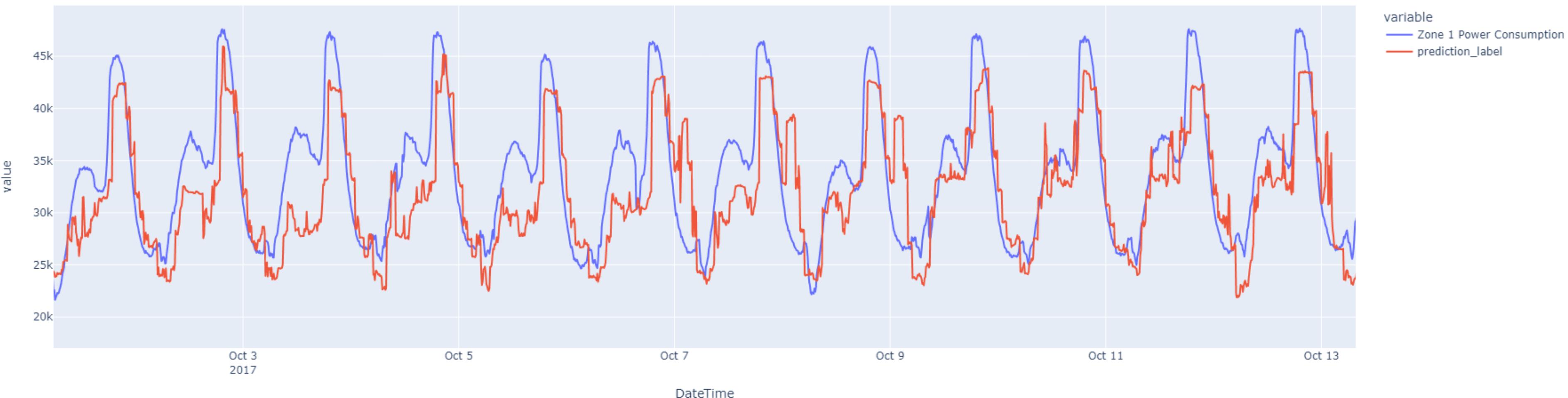
Random Forest



Model	MAE	MSE	RMSE	R2	RMSLE	MAPE
Random Forest Regressor	3484.6397	20436208.4304	4520.6425	0.5336	0.1398	0.1097

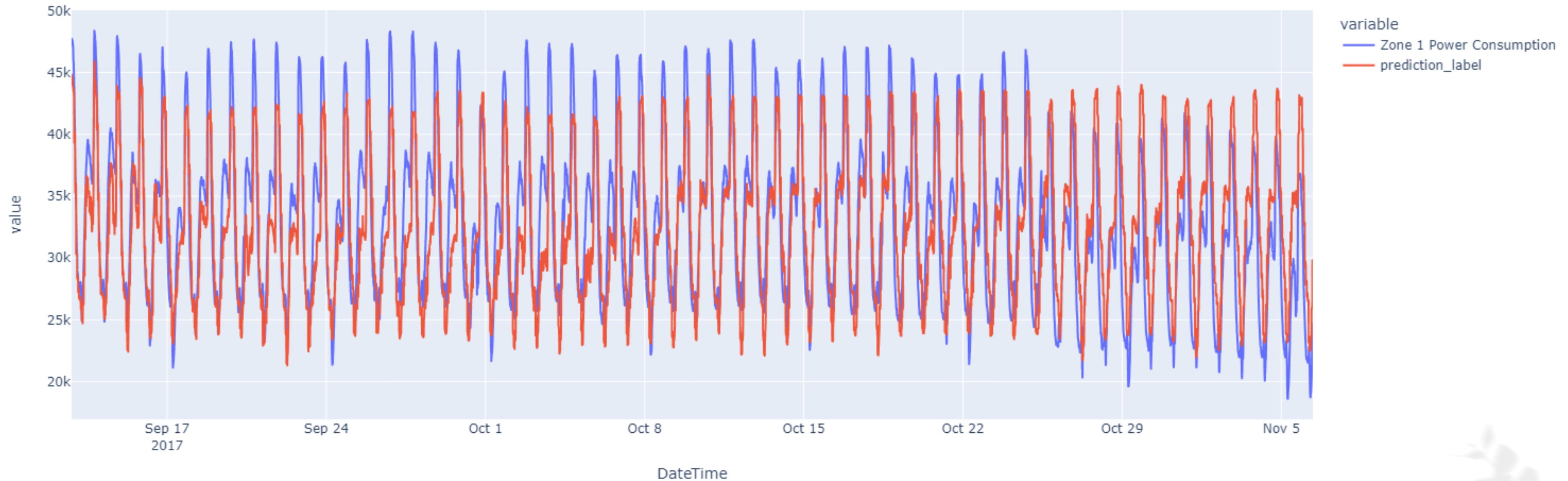
...

Random Forest - closeup



•••

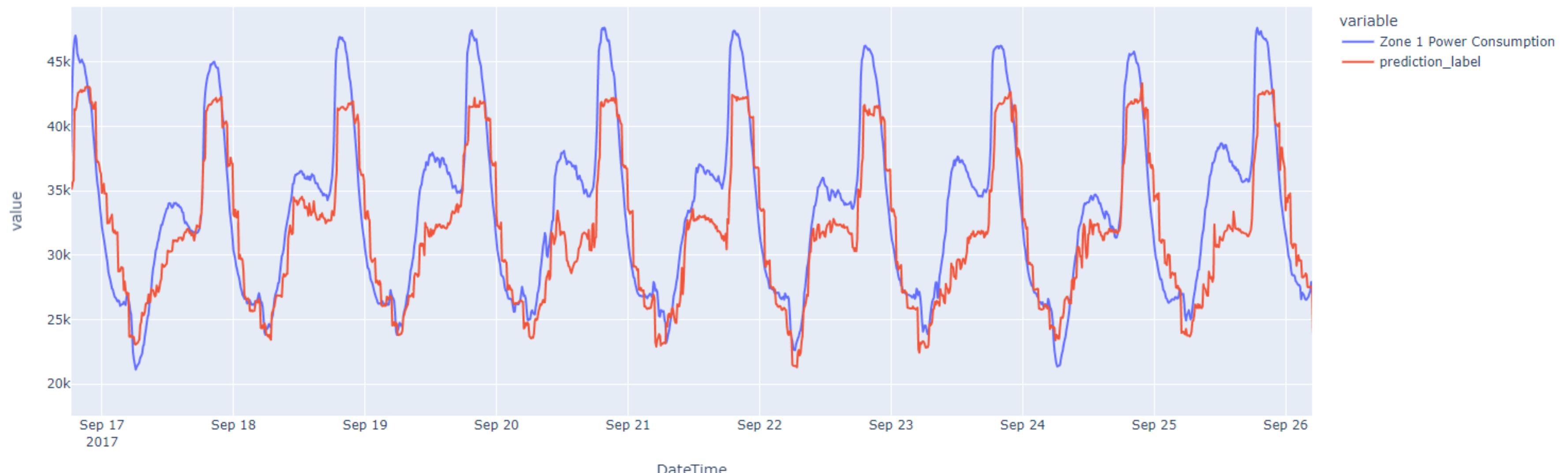
Extra Trees



Model	MAE	MSE	RMSE	R2	RMSLE	MAPE
Extra Trees Regressor	3095.5254	14226226.1840	3771.7670	0.6753	0.1168	0.0966

•••

Extra Trees - closeup



...

Testing models' performance



...

Effectiveness measures

1

MSE

Average **squared difference** between the **predicted values** and the **actual values**.

Squaring the differences emphasizes larger errors, making it sensitive to outliers.

2

MAE

Average **absolute difference** between the **predicted** and **actual values**. It is less sensitive to outliers than MSE. MAE is easier to interpret in the original units of the data

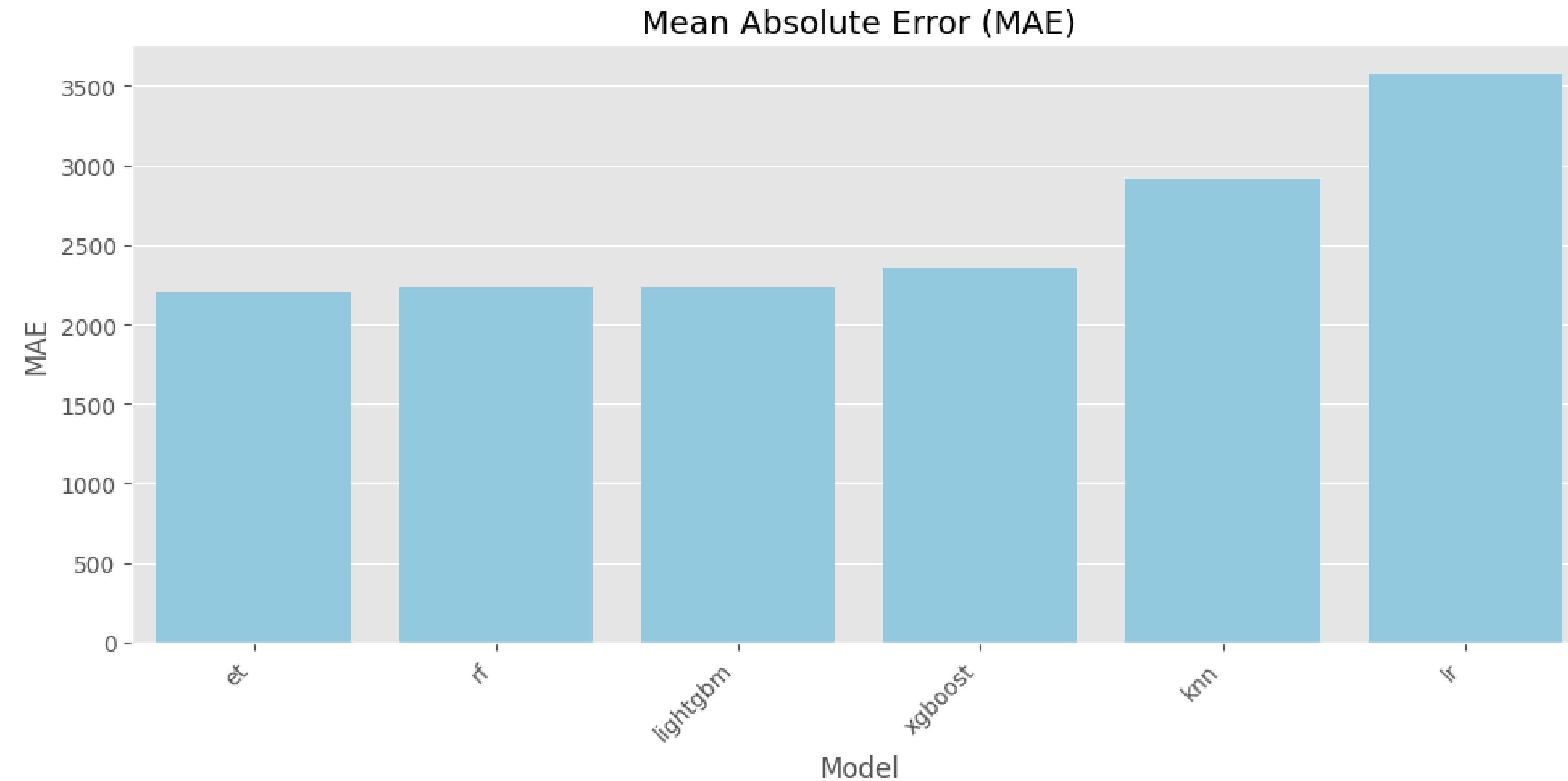
3

MAPE

Average **percentage difference** between the **predicted** and **actual values**. It is useful when you want to understand the size of errors relative to the actual values.

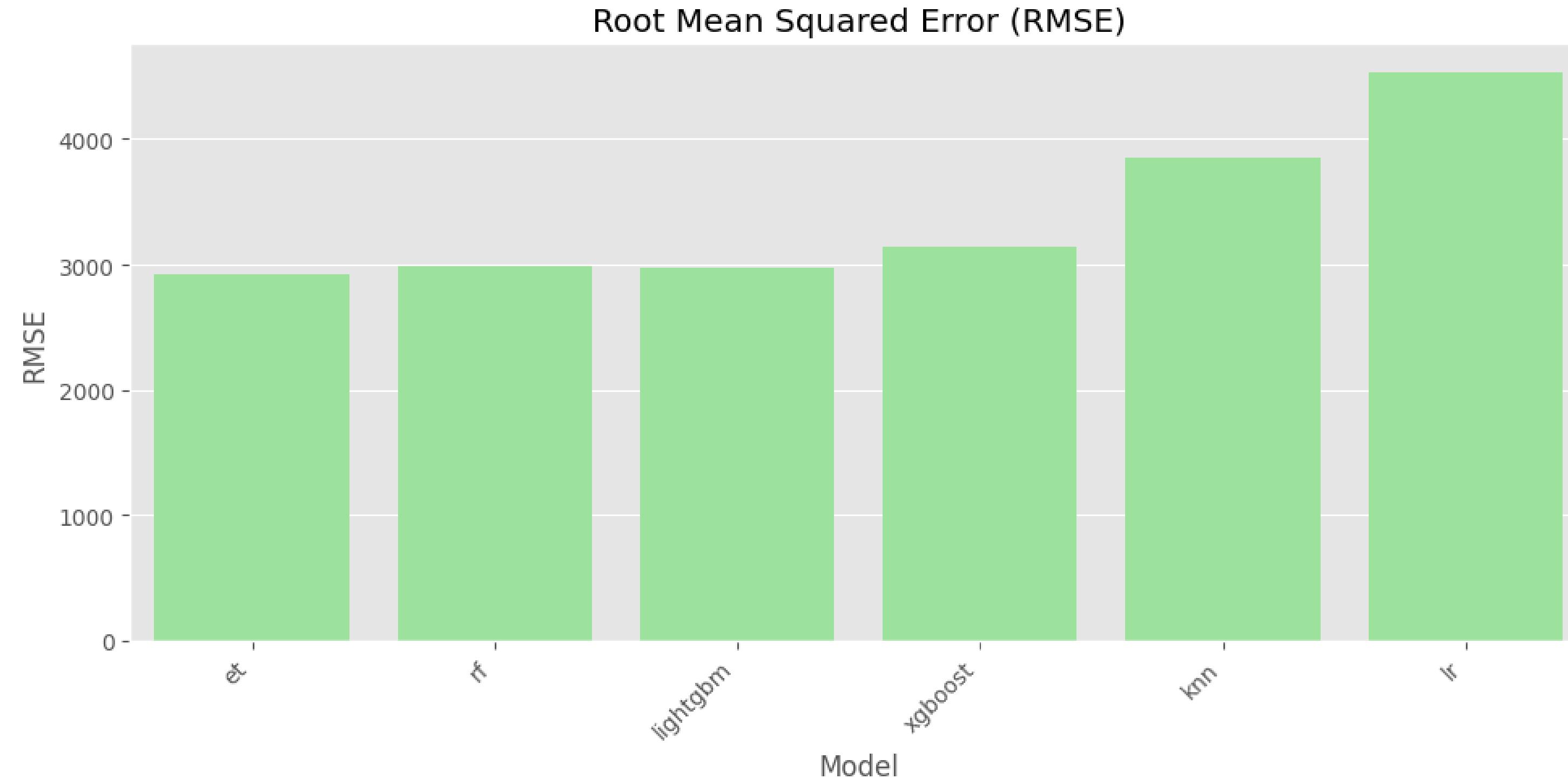
...

Testing models' performance



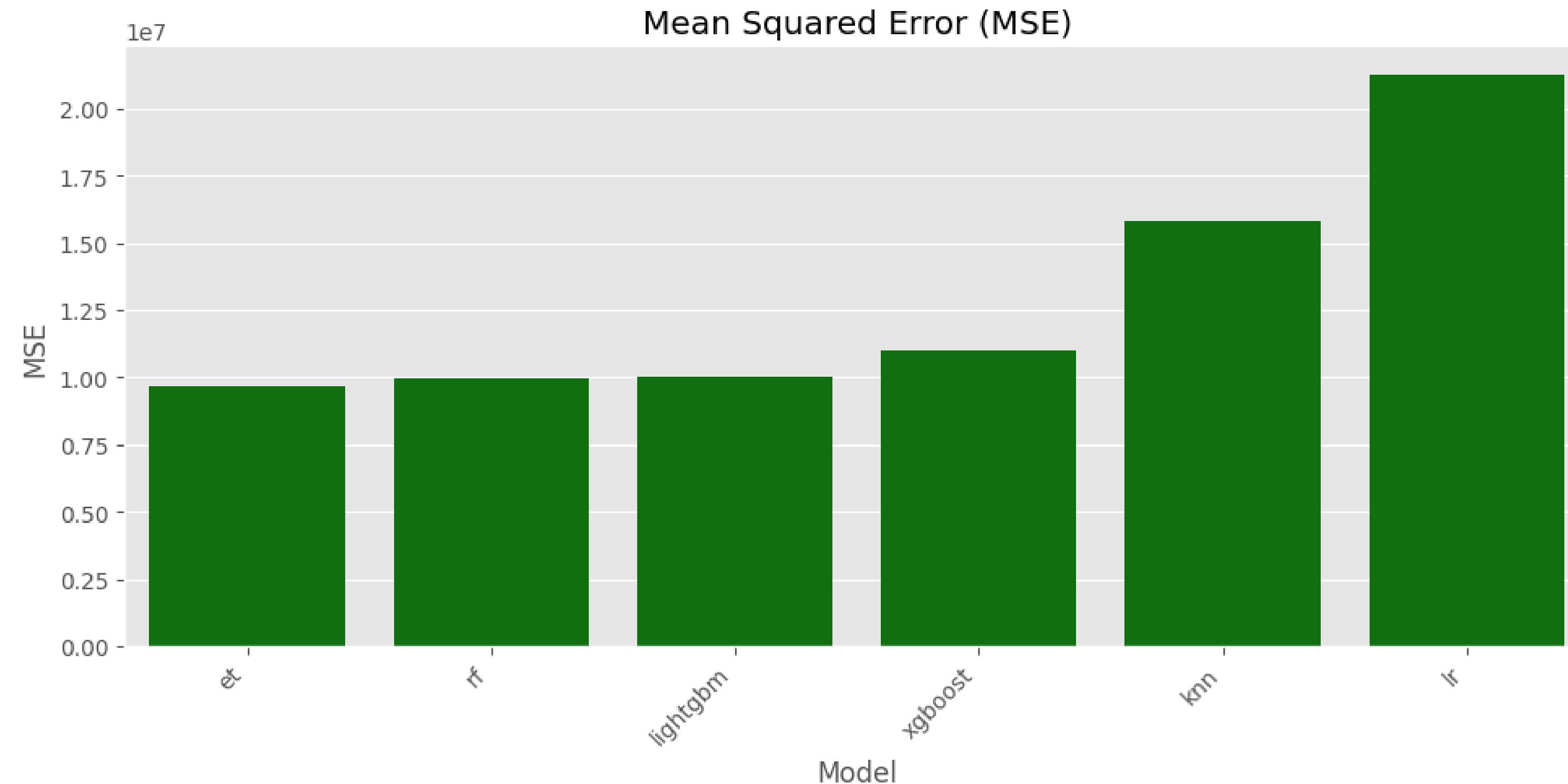
...

Testing models' performance



...

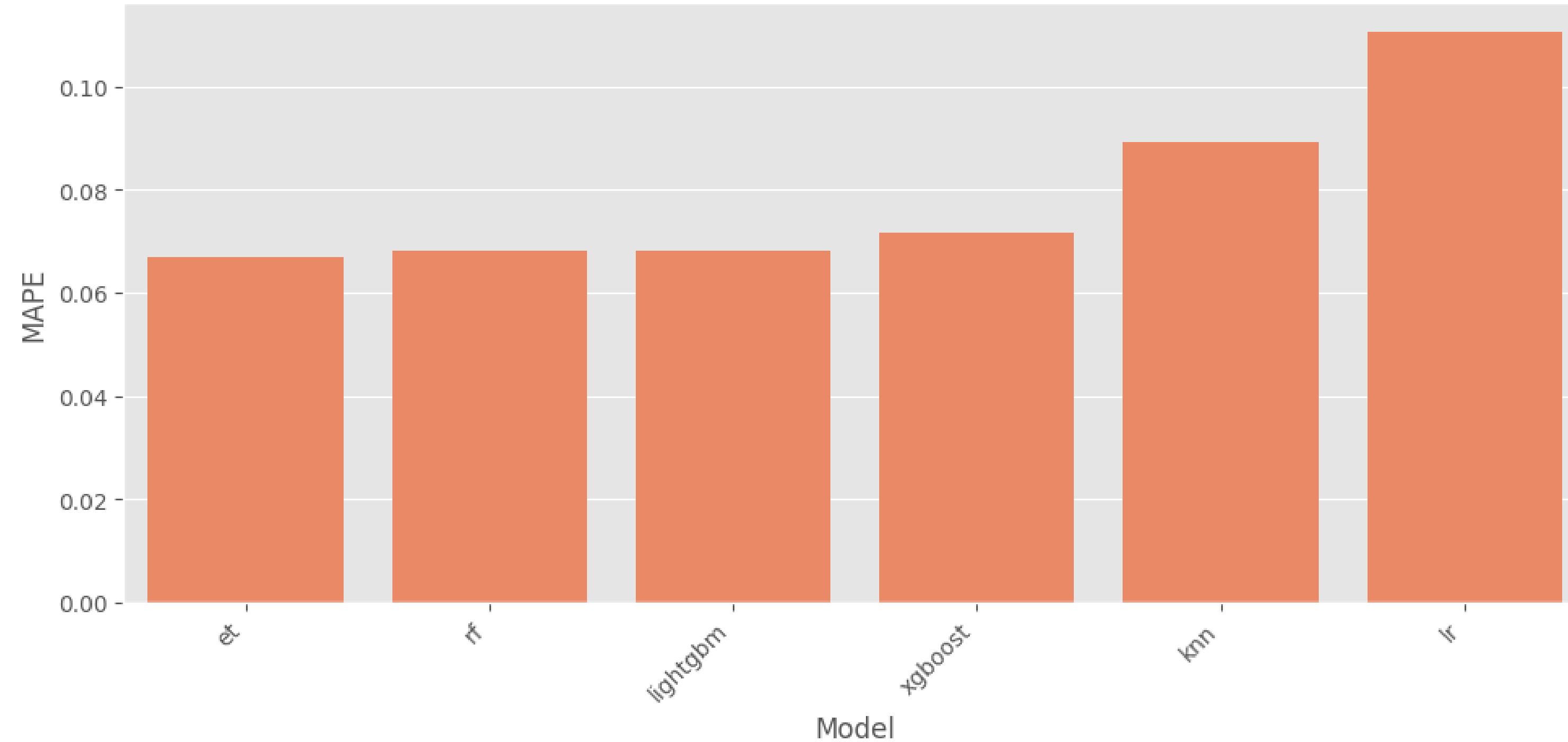
Testing models' performance



...

Testing models' performance

Mean Absolute Percentage Error (MAPE)



...

WINNER!



Extra
Trees

Tuning Extra Trees

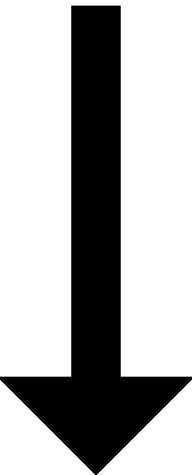
```
custom_grid = {
    'n_estimators': [50, 100, 150],
    'max_depth': [10, 20],
    'min_samples_split': [2, 5, 10],
    'min_samples_leaf': [1, 2, 4],
    'max_features': ['auto', 'sqrt', 'log2'],
    'bootstrap': [True, False]
}
```

```
ExtraTreesRegressor
ExtraTreesRegressor(bootstrap=True, max_depth=10, max_features='auto',
                   min_samples_leaf=2, min_samples_split=10, n_estimators=150,
                   n_jobs=-1, random_state=7681)
```

...

Final results after tuning

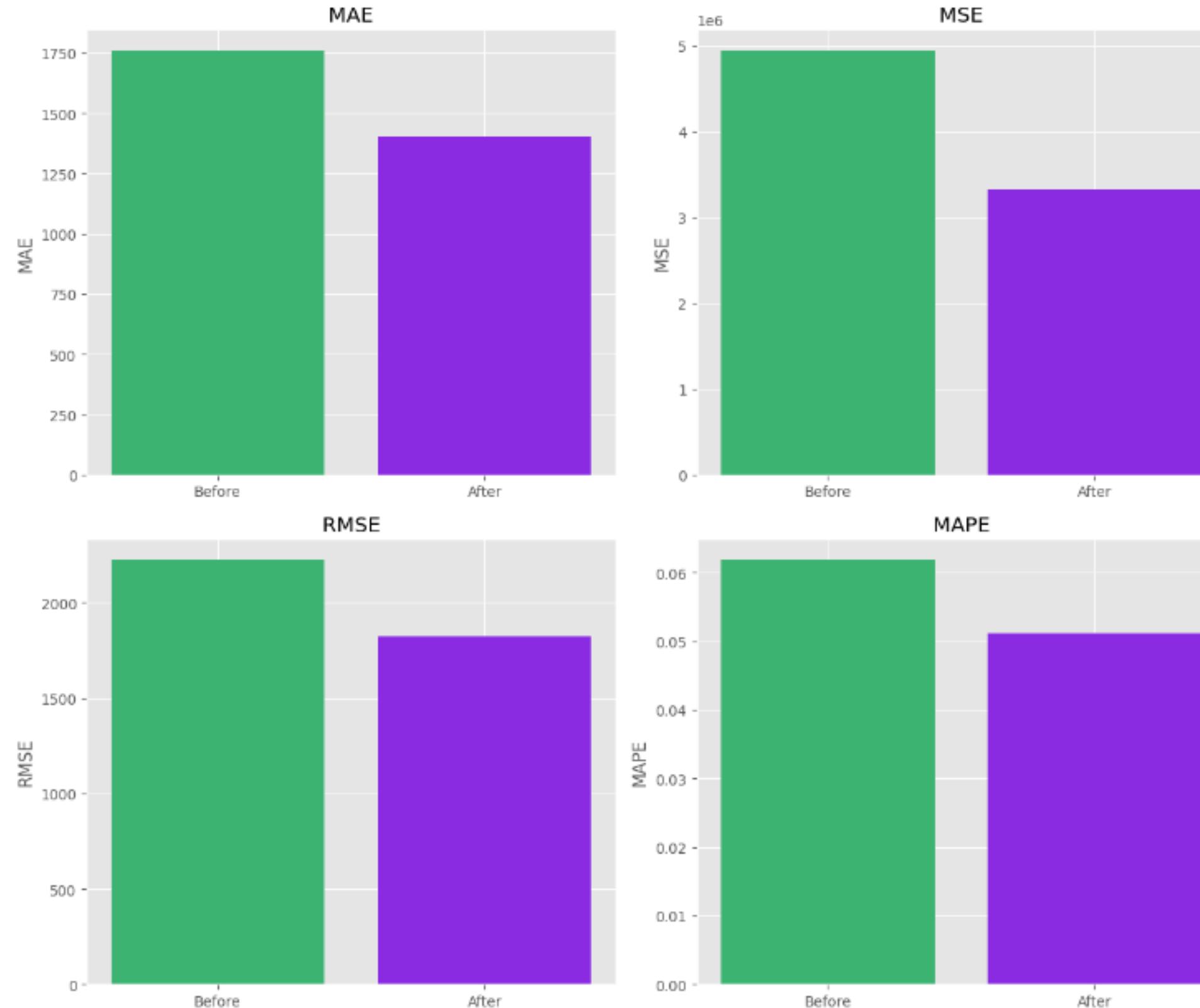
Model	MAE	MSE	RMSE	R2	RMSLE	MAPE
Extra Trees Regressor	1758.9160	4938389.9006	2222.2488	0.8654	0.0759	0.0618



Model	MAE	MSE	RMSE	R2	RMSLE	MAPE
Extra Trees Regressor	1402.4187	3326944.3727	1823.9913	0.9093	0.0653	0.0512

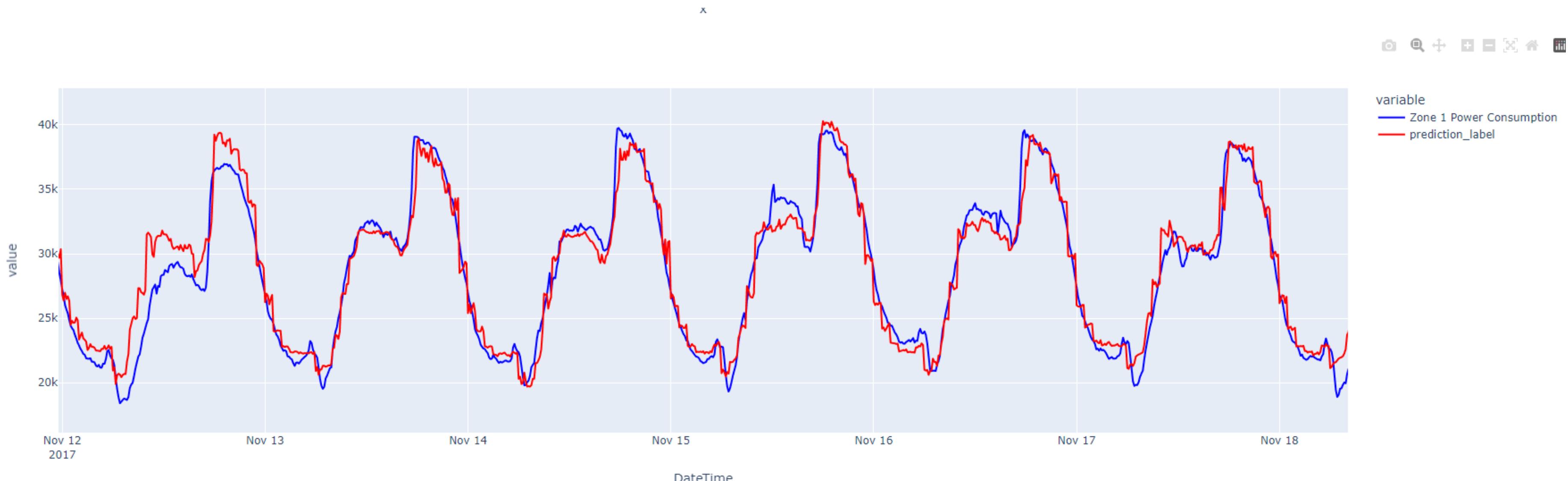
...

Final results after tuning



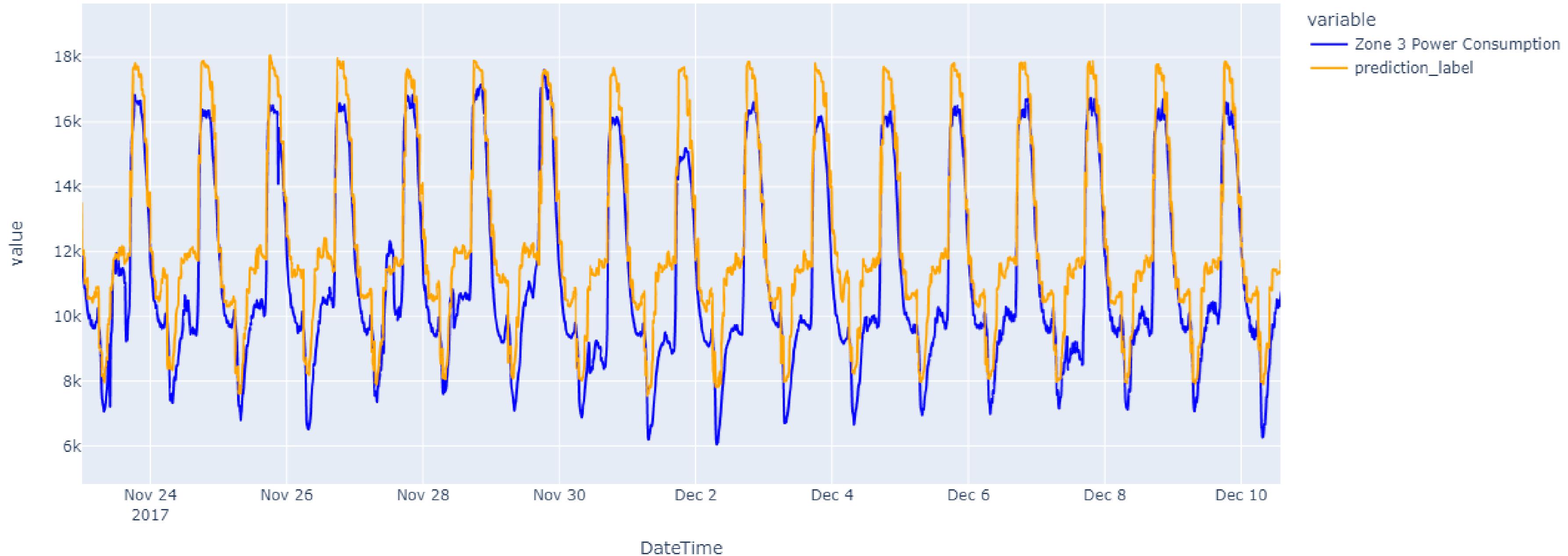
...

Final Results for zone 1



...

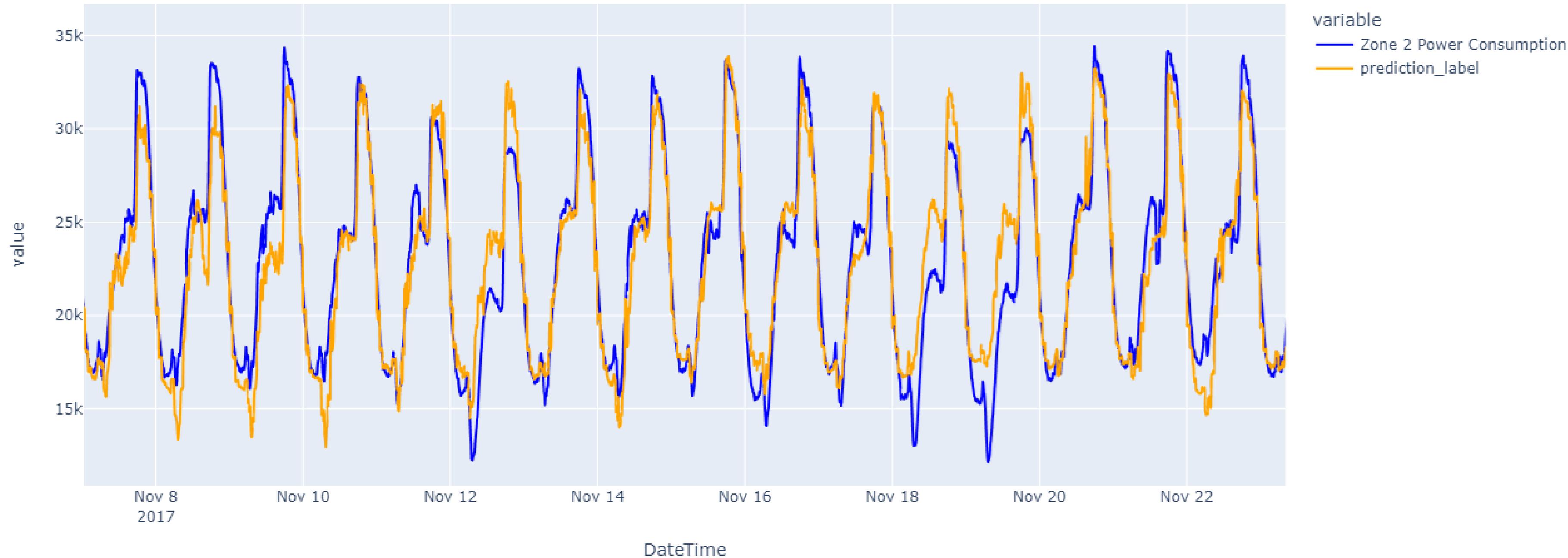
Final Results for zone 2



Model	MAE	MSE	RMSE	R2	RMSLE	MAPE
Extra Trees Regressor	1549.9796	3815324.4134	1953.2855	0.8778	0.0877	0.0681

...

Final Results for zone 3



Model	MAE	MSE	RMSE	R2	RMSLE	MAPE
Extra Trees Regressor	1484.3660	3919511.7616	1979.7757	0.6527	0.1584	0.1249

...
Thank you

Nicola Szwaja
Piotr Droś

CUTIE
PIE