# HOW MUCH ELECTRIC POWER WILL WE NEED IN THE NEXT DAY?

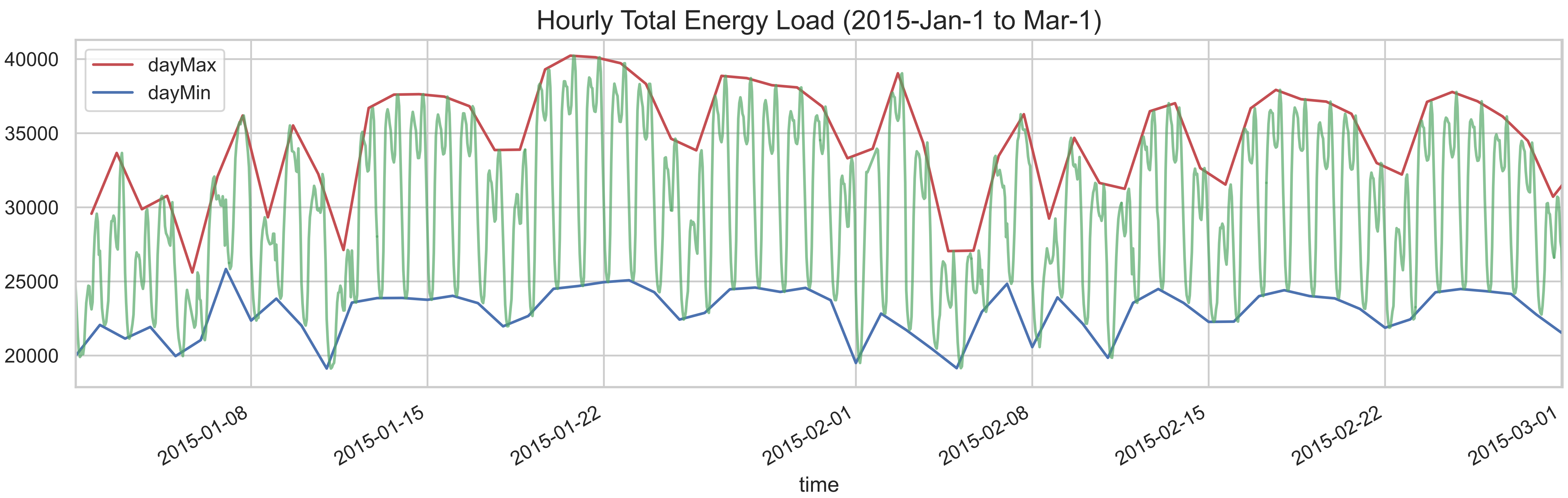
Forecasting Energy Load using Seasonal ARIMA

Mendy Hsu

# Problem Statements

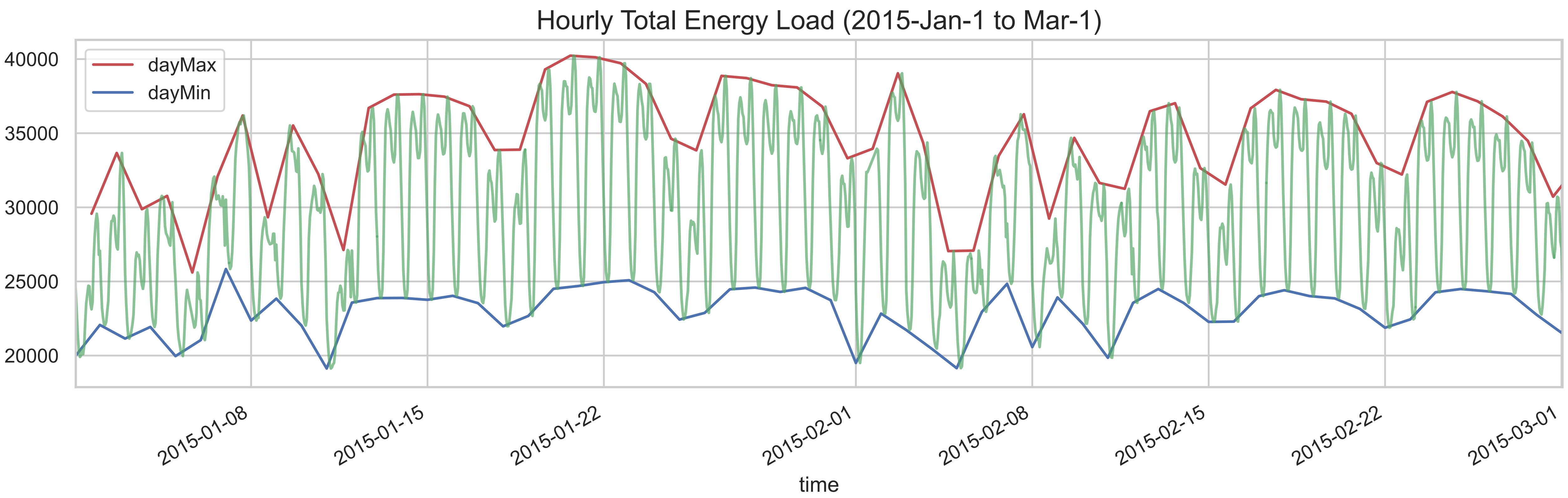
- Energy suppliers/companies need to ensure the electricity supply always meets their customers' demands (reliability of supply) and prevents the supply shortage.

- **Load forecasting** helps an **electric utility** to make important decisions, such as purchasing and generating electric power, and electrical infrastructure development.



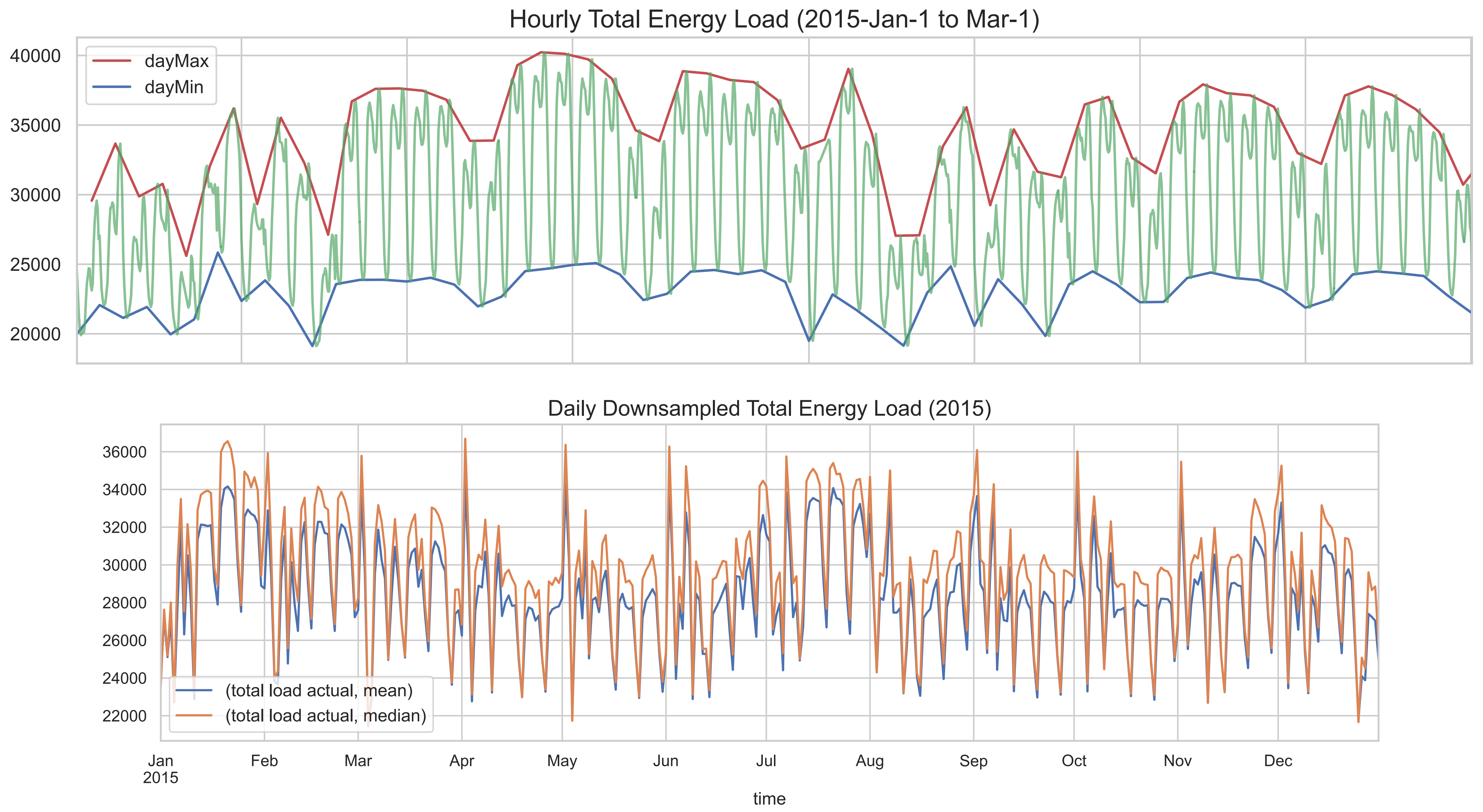

# Data Information

- **Period and location**: four years (2015 - 2018) recording in **Spain**.
- **Number of fields**: 29 (electrical consumption, generation and pricing, etc.)
- **Number of rows**: **35064** hourly records.

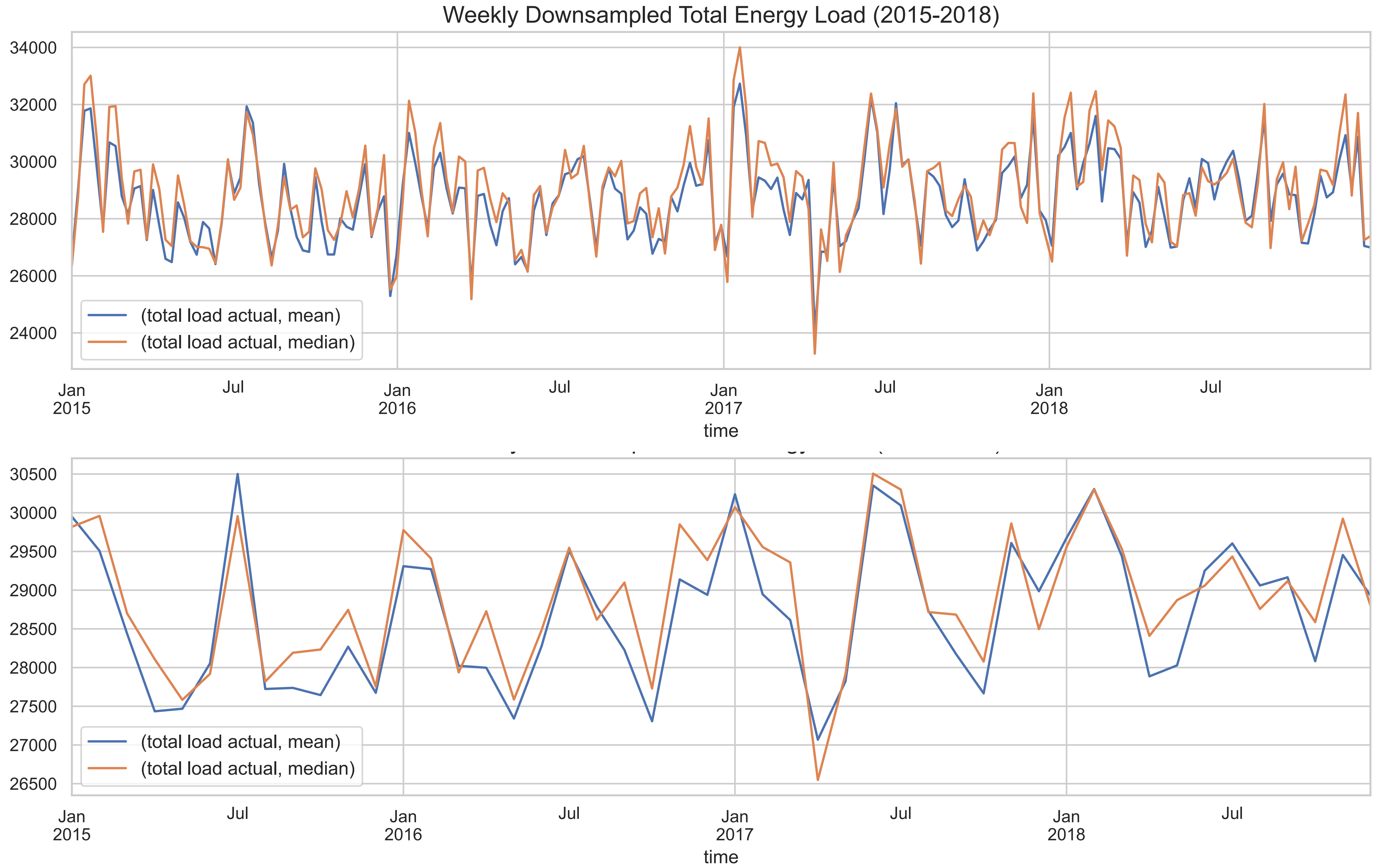



- Daily pattern (green line): two maximums and one minimum.

- Weekly pattern (red and blue envelope).



Hourly Total Energy Load (2015-Jan-1 to Mar-1)
dayMax
dayMin
40000
35000
30000
25000
20000
Daily Downsampled Total Energy Load (2015)
36000
34000
32000
30000
28000
26000
24000
22000
(total load actual, mean)
(total load actual, median)
Jan
2015
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
time



Weekly Downsampled Total Energy Load (2015-2018)
34000
32000
30000
28000
26000
24000
(total load actual, mean)
(total load actual, median)
Jan
2015
Jul
Jan
2016
Jul
Jan
2017
Jul
Jan
2018
Jul
time
30500
30000
29500
29000
28500
28000
27500
27000
26500
(total load actual, mean)
(total load actual, median)
Jan
2015
Jul
Jan
2016
Jul
Jan
2017
Jul
Jan
2018
Jul
time

# Problem Formulation

- **Goals**: Based on 2015 – 2017 records, predict the *total load* hourly in 2018.

- **Issue**: With **Seasonal ARIMA** model, it takes much longer time to train hourly points over 3 years.

  → Option 1 : train the model for sort-term period (ex: train data of 2 to 6 months; predict the next month).

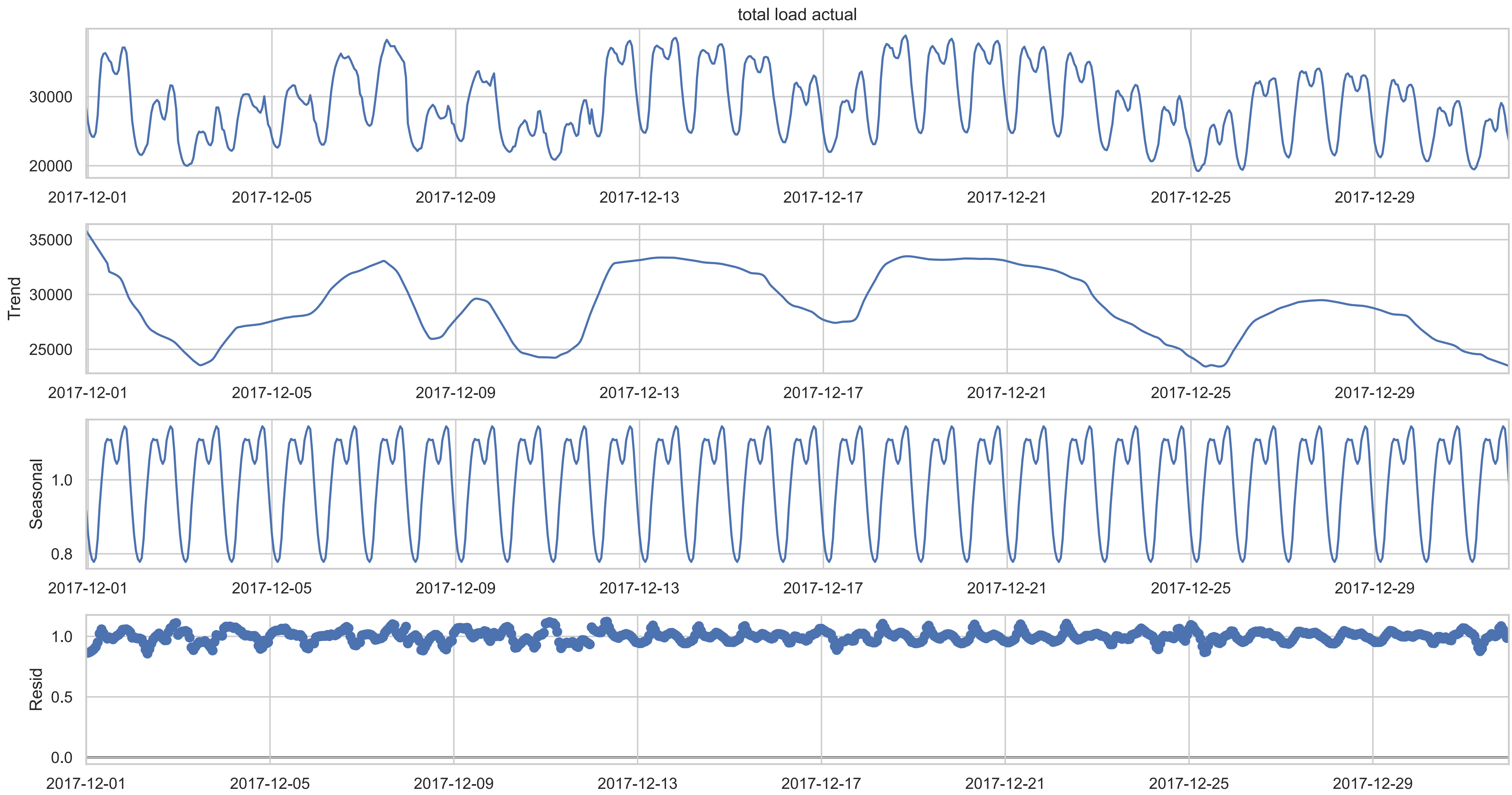
  → Option 2 : train the model of daily load (downsampled the time series).
  This reduced the number of points by 1/24.

# Problem Formulation

- **Alternative Goals**:
  - Predicting daily **minimum** and **maximum** *total load*.
    → **Seasonal ARIMA (this is our focus).**
  - The typical pattern of *hourly total load* each day.
    → **Time series decomposition**

- **Data splits:**
  - **Training**: 2015-1-1 to 2017-12-31
  - **Validation**: 2018-1-1 to 2018-12-31

# Seasonal Decomposition ($y_t = T_t \times S_t \times R_t$)



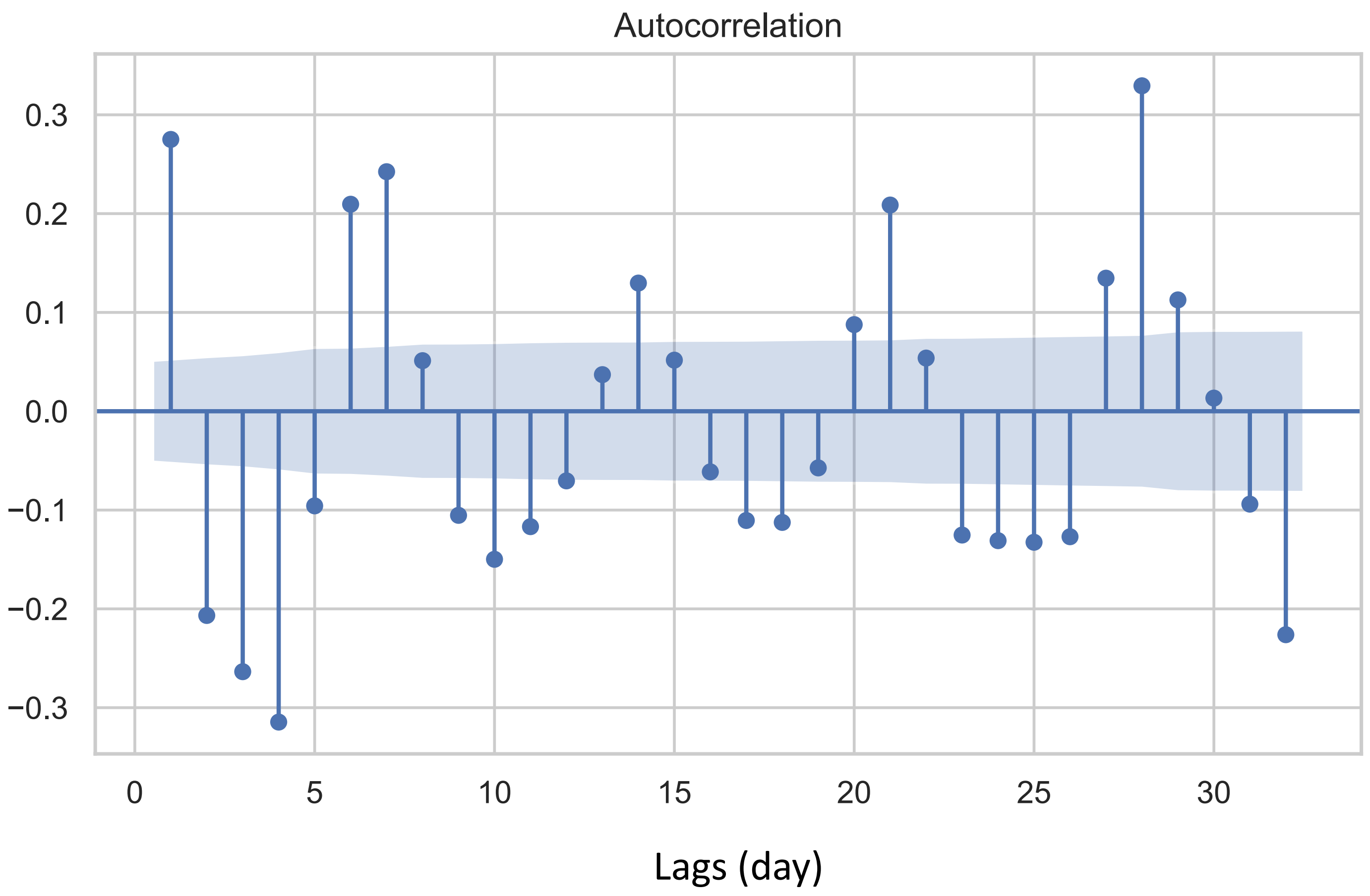
# Seasonal ARIMA (SARIMAX)

- ARIMA - **A**uto-**R**egressive **I**ntegrated **M**oving **A**verage
- X – e**X**ogenous factors

- Hyperparameters:
  - Non-seasonal orders: (p, d, q)
  - Seasonal orders: (P, D, Q, s)

- P, p – autoregressive
- D, d – differencing order
- Q, q – moving average
- s – number of steps for a season

# Hyperparameter Tuning (Seasonal steps s = 7)



# Hyperparameter Tuning (non-seasonal p, d, q)

- Difference  (d) - **Stationarity**

- AR(p)-MA(q)
  - **A**uto**C**orrelation **F**unction (ACF)
  - **P**artial **A**uto**C**orrelation **F**unction (PACF)

| | **AR (p)** | **MA(q)** | **ARMA(p, q)** |
|---|---|---|---|
| ACF | Tail off | Cuts off after lag q | Tail off |
| PACF | Cuts off after lag p | Tail off | Tail off |

Ref: DataCamp

# Hyperparameter Tuning (seasonal P, D, Q)

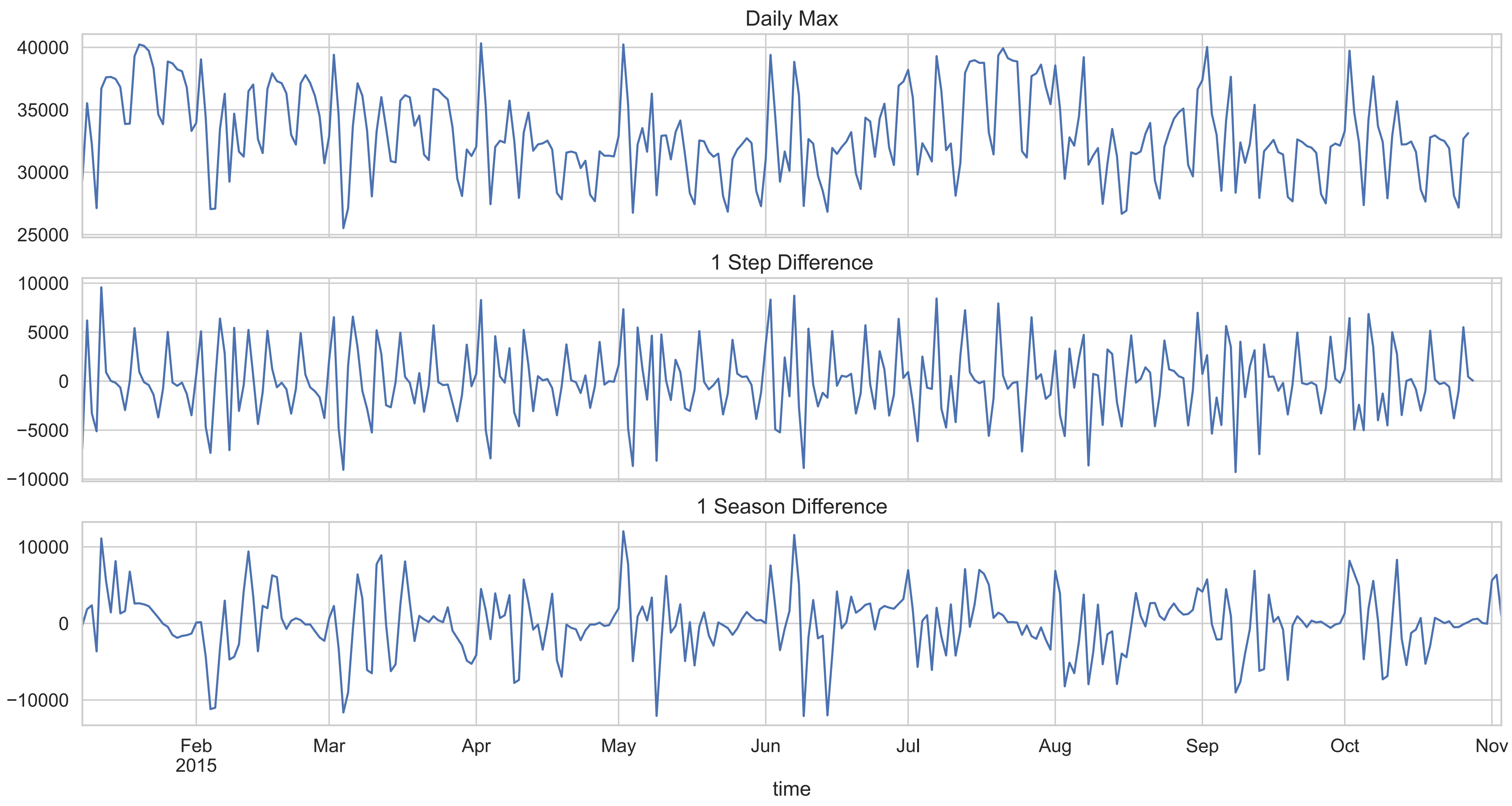
- Difference (D*s) - **Stationarity**

- Seasonal AR(P)-MA(Q)
  - **A**uto**C**orrelation **F**unction (ACF)
  - **P**artial **A**uto**C**orrelation **F**unction (PACF)

| | Seasonal AR (P) | Seasonal MA(Q) | SARMA(P, Q) |
|---|---|---|---|
| ACF | Tail off | Cuts off after lag (Q*s) | Tail off |
| PACF | Cuts off after lag (P*s) | Tail off | Tail off |

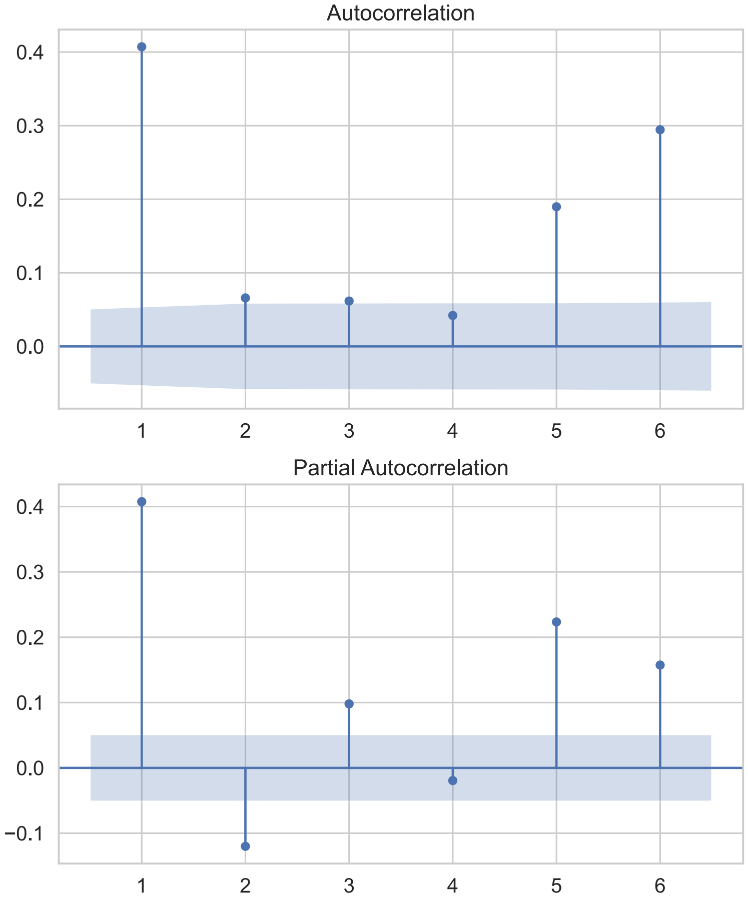
Ref: DataCamp

# Differencing

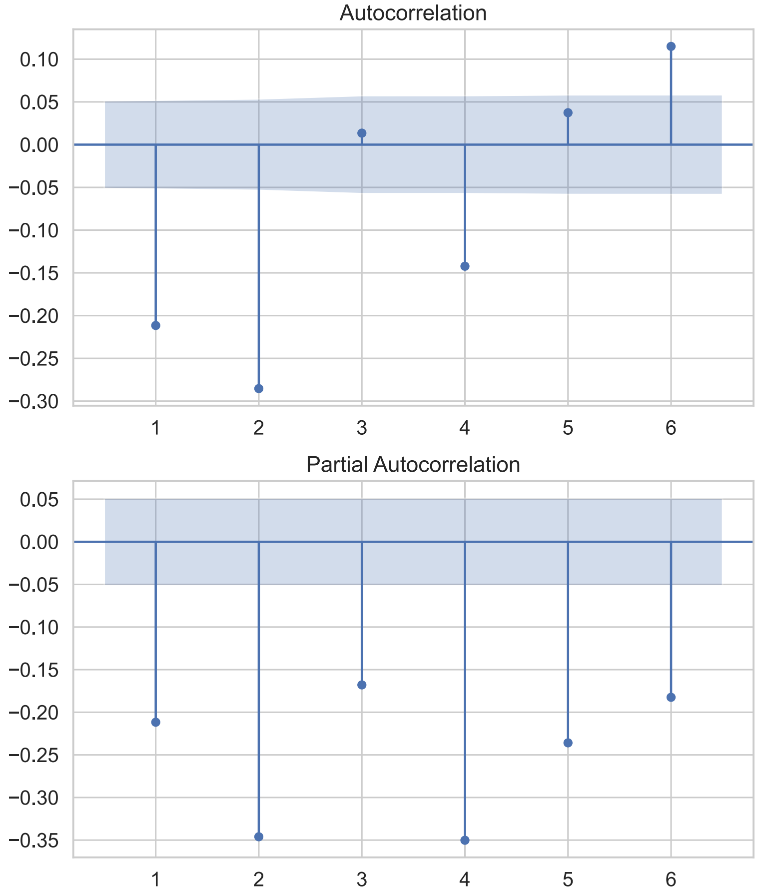


# ACF & PACF for (p, q)

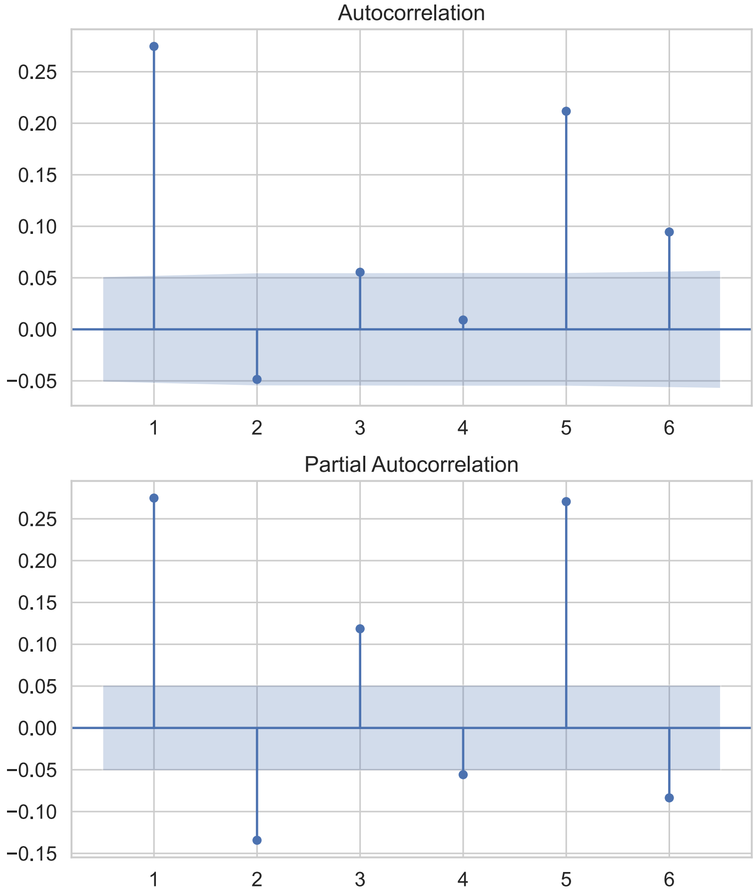
## d, D = 0




## d = 1, D =0

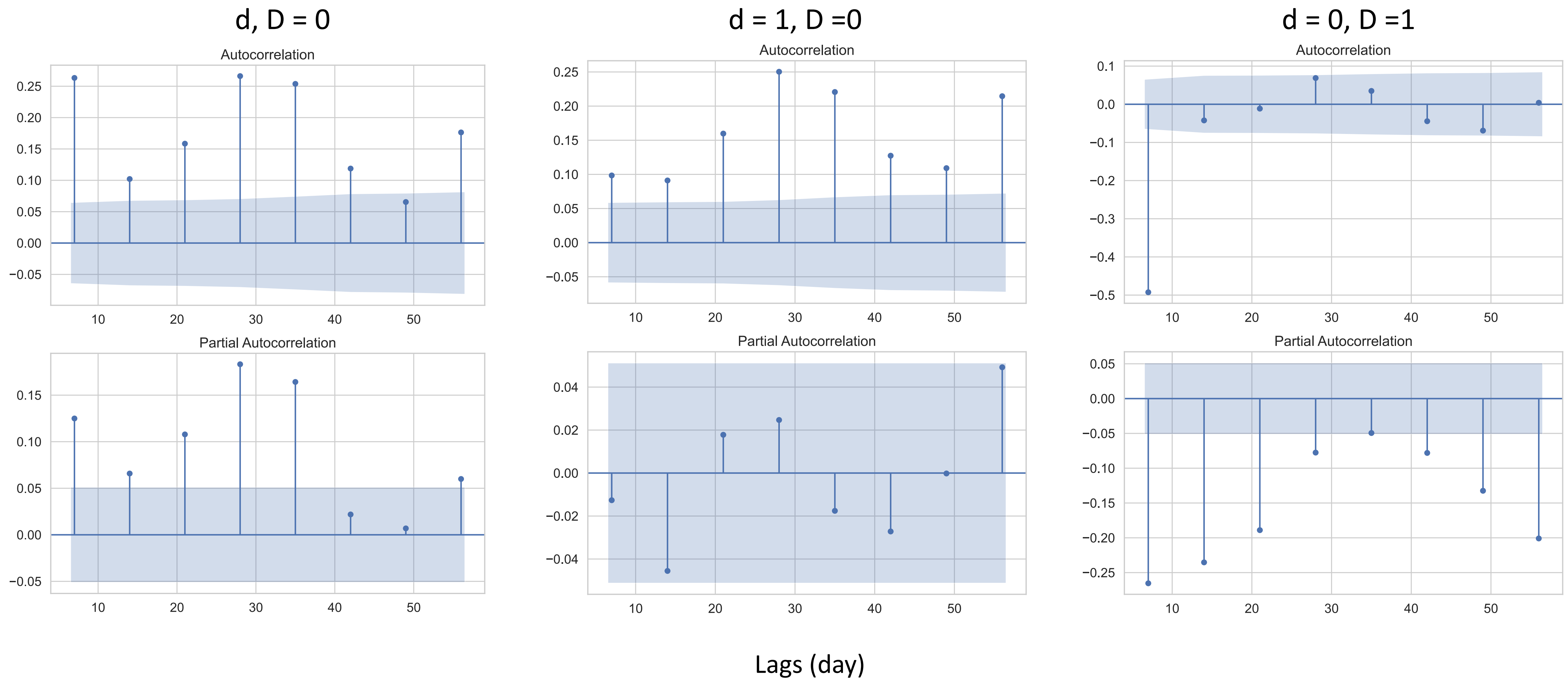



## d = 0, D =1




Lags (day)

# ACF & PACF for (P, Q)



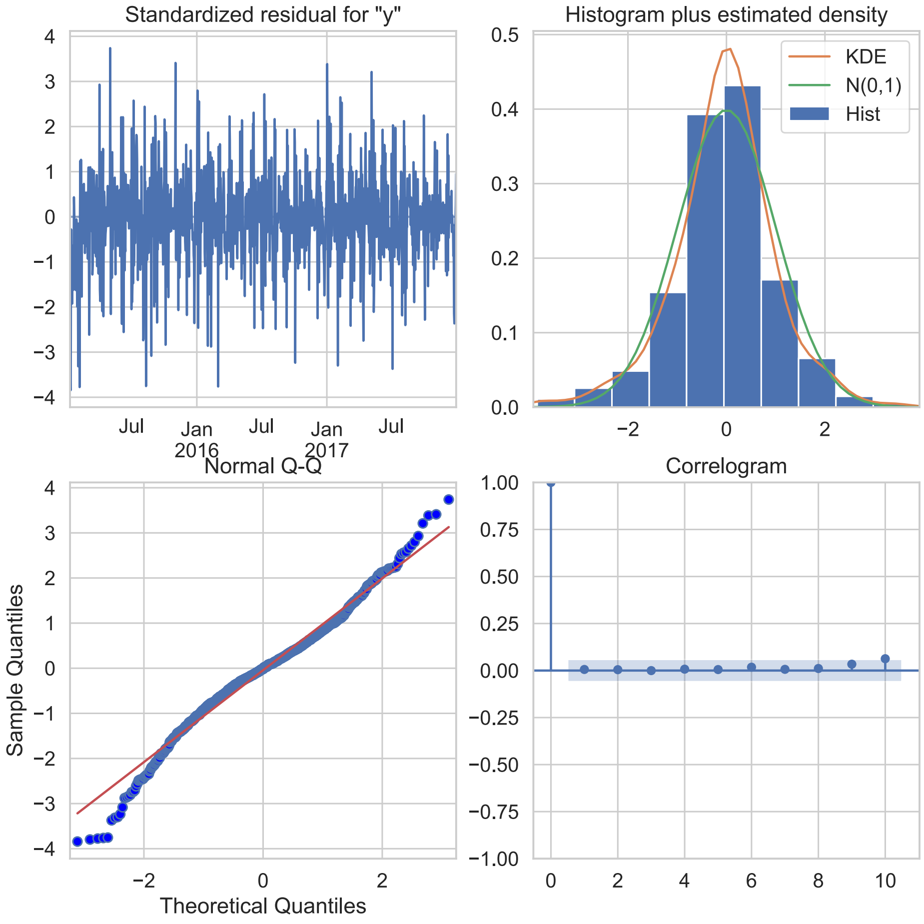
# Hyperparameter Tuning with Auto-ARIMA (Day Min)

SARIMAX Results

| | | | |
|---|---|---|---|
| Dep. Variable: | y | No. Observations: | 1096 |
| Model: | SARIMAX(5, 1, 0)x(5, 1, 0, 7) | Log Likelihood | -9432.661 |
| Date: | Mon, 22 Nov 2021 | AIC | 18887.322 |
| Time: | 16:51:43 | BIC | 18942.235 |
| Sample: | 01-01-2015 | HQIC | 18908.107 |
| | - 12-31-2017 | | |
| Covariance Type: | opg | | |

| | coef | std err | z | P>\|z\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| ar.L1 | -0.5150 | 0.022 | -23.035 | 0.000 | -0.559 | -0.471 |
| ar.L2 | -0.5459 | 0.024 | -23.123 | 0.000 | -0.592 | -0.500 |
| ar.L3 | -0.3549 | 0.026 | -13.898 | 0.000 | -0.405 | -0.305 |
| ar.L4 | -0.3766 | 0.028 | -13.409 | 0.000 | -0.432 | -0.322 |
| ar.L5 | -0.1603 | 0.029 | -5.512 | 0.000 | -0.217 | -0.103 |
| ar.S.L7 | -0.8577 | 0.029 | -29.419 | 0.000 | -0.915 | -0.801 |
| ar.S.L14 | -0.7938 | 0.041 | -19.220 | 0.000 | -0.875 | -0.713 |
| ar.S.L21 | -0.6203 | 0.042 | -14.847 | 0.000 | -0.702 | -0.538 |
| ar.S.L28 | -0.3728 | 0.040 | -9.329 | 0.000 | -0.451 | -0.294 |
| ar.S.L35 | -0.1141 | 0.036 | -3.137 | 0.002 | -0.185 | -0.043 |
| sigma2 | 1.9e+06 | 6.52e+04 | 29.154 | 0.000 | 1.77e+06 | 2.03e+06 |

| | | | |
|---|---|---|---|
| Ljung-Box (L1) (Q): | 0.04 | Jarque-Bera (JB): | 102.38 |
| Prob(Q): | 0.85 | Prob(JB): | 0.00 |
| Heteroskedasticity (H): | 0.69 | Skew: | -0.26 |
| Prob(H) (two-sided): | 0.00 | Kurtosis: | 4.41 |



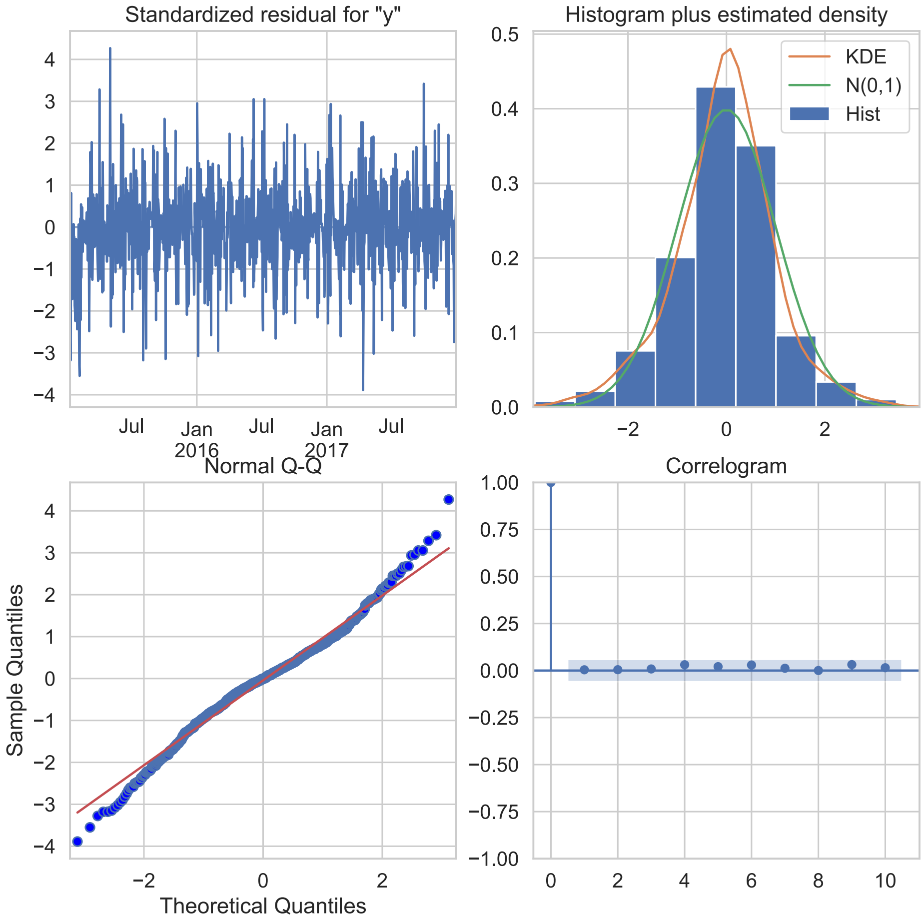
# Hyperparameter Tuning with Auto-ARIMA (Day Max)

SARIMAX Results

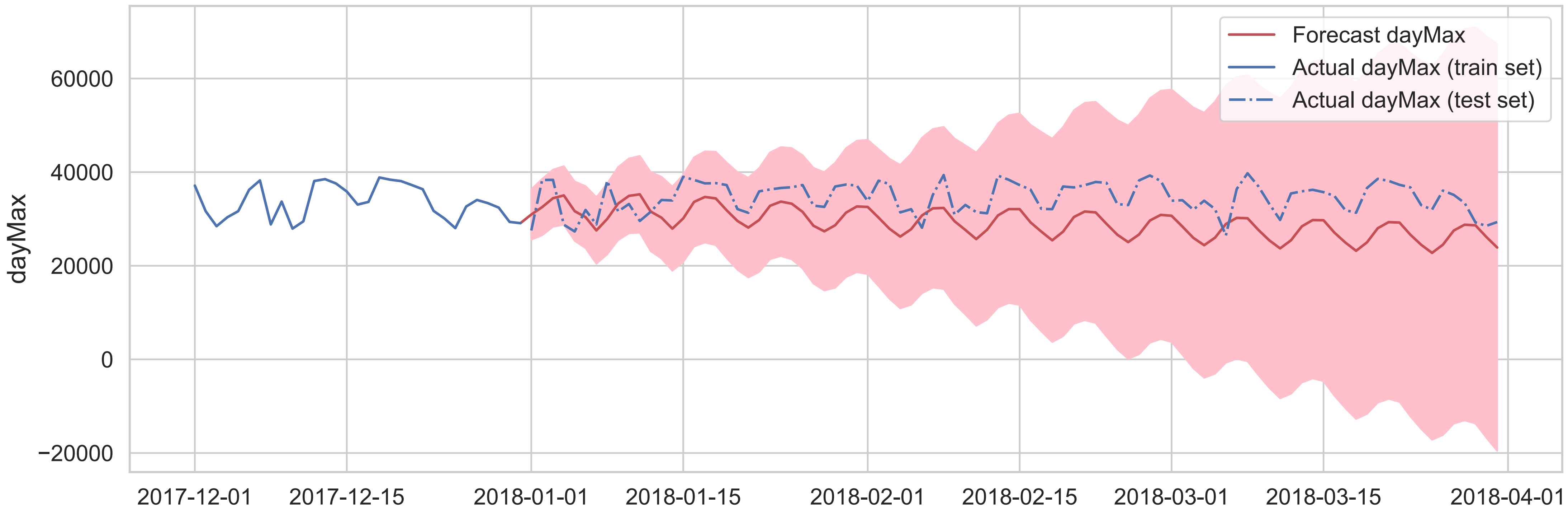
| | | | |
|---|---|---|---|
| Dep. Variable: | y | No. Observations: | 1096 |
| Model: | SARIMAX(5, 1, 0)x(5, 1, 0, 7) | Log Likelihood | -10176.442 |
| Date: | Mon, 22 Nov 2021 | AIC | 20374.884 |
| Time: | 16:48:09 | BIC | 20429.797 |
| Sample: | 01-01-2015 | HQIC | 20395.669 |
| | - 12-31-2017 | | |
| Covariance Type: | opg | | |

| | coef | std err | z | P>\|z\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| ar.L1 | -0.5002 | 0.023 | -21.951 | 0.000 | -0.545 | -0.456 |
| ar.L2 | -0.5636 | 0.025 | -22.965 | 0.000 | -0.612 | -0.516 |
| ar.L3 | -0.3773 | 0.028 | -13.716 | 0.000 | -0.431 | -0.323 |
| ar.L4 | -0.3881 | 0.028 | -13.824 | 0.000 | -0.443 | -0.333 |
| ar.L5 | -0.1339 | 0.028 | -4.751 | 0.000 | -0.189 | -0.079 |
| ar.S.L7 | -0.8839 | 0.032 | -28.042 | 0.000 | -0.946 | -0.822 |
| ar.S.L14 | -0.8352 | 0.043 | -19.420 | 0.000 | -0.919 | -0.751 |
| ar.S.L21 | -0.6653 | 0.045 | -14.926 | 0.000 | -0.753 | -0.578 |
| ar.S.L28 | -0.3723 | 0.042 | -8.948 | 0.000 | -0.454 | -0.291 |
| ar.S.L35 | -0.1109 | 0.037 | -2.973 | 0.003 | -0.184 | -0.038 |
| sigma2 | 7.559e+06 | 2.62e+05 | 28.819 | 0.000 | 7.04e+06 | 8.07e+06 |

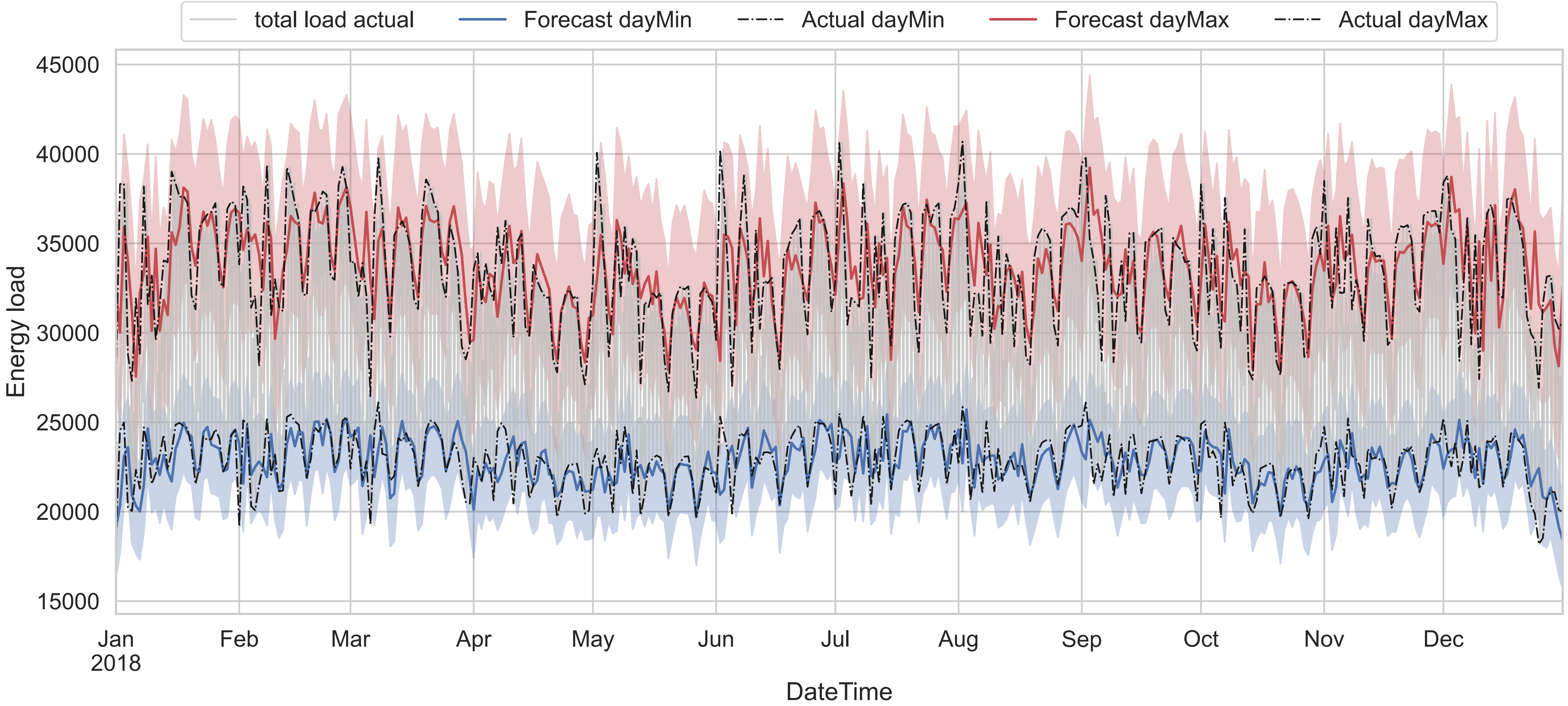
| | | | |
|---|---|---|---|
| Ljung-Box (L1) (Q): | 0.02 | Jarque-Bera (JB): | 68.02 |
| Prob(Q): | 0.90 | Prob(JB): | 0.00 |
| Heteroskedasticity (H): | 0.83 | Skew: | -0.12 |
| Prob(H) (two-sided): | 0.07 | Kurtosis: | 4.20 |



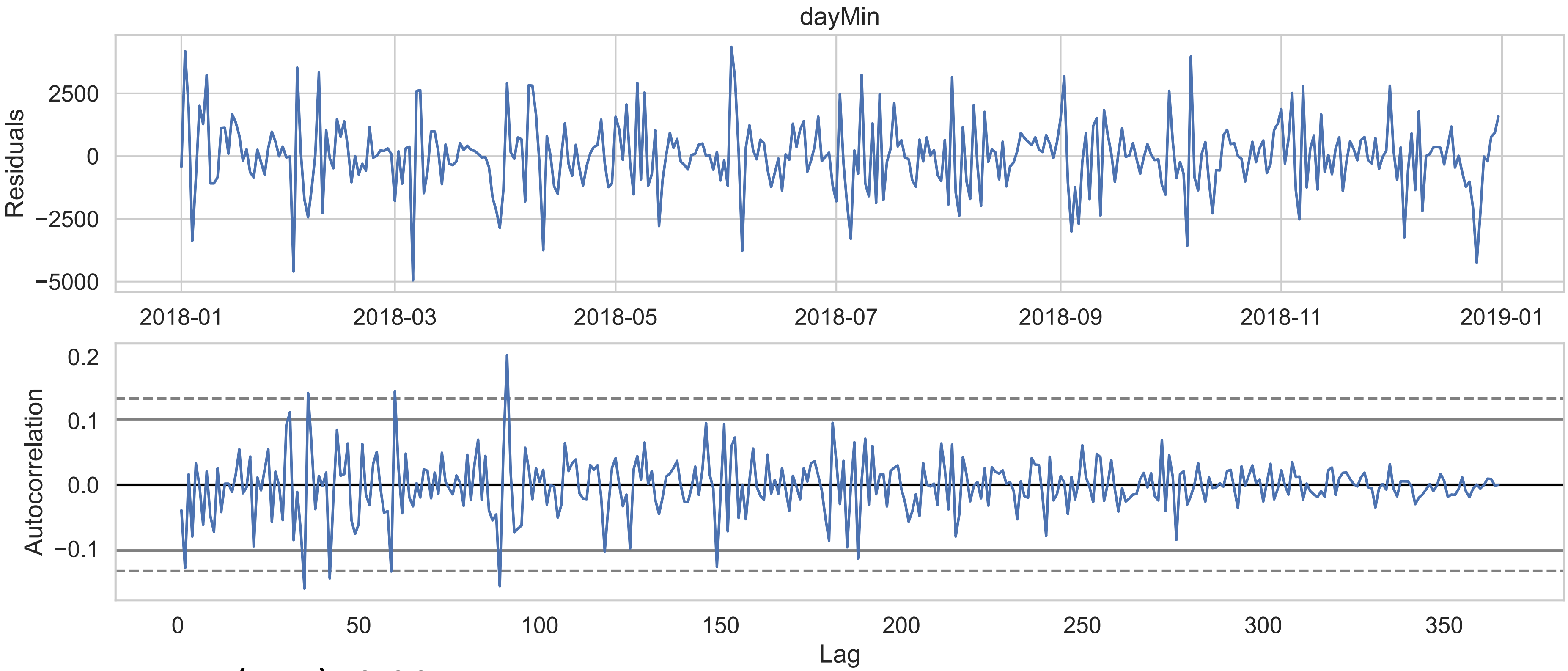
# Forecasting (without updating the new point)

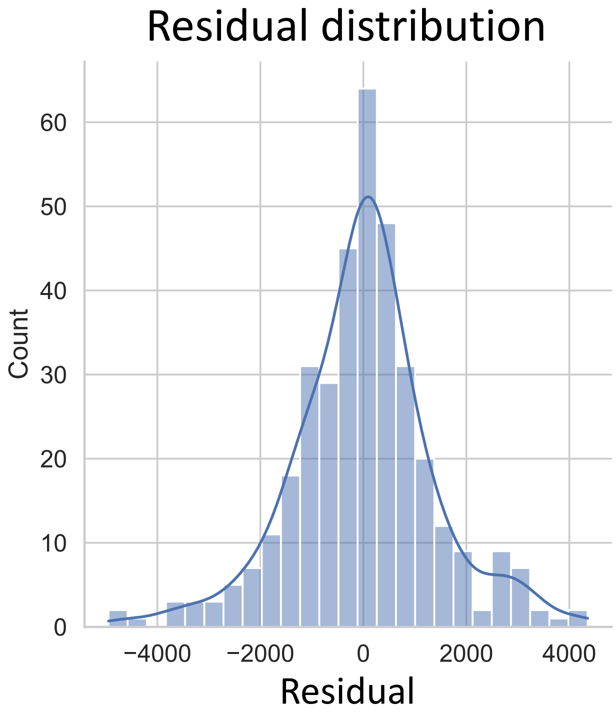


# Forecasting (updating the new point daily)



# Accuracy




R-square (test): 0.227

MSE (test): 1884187.588

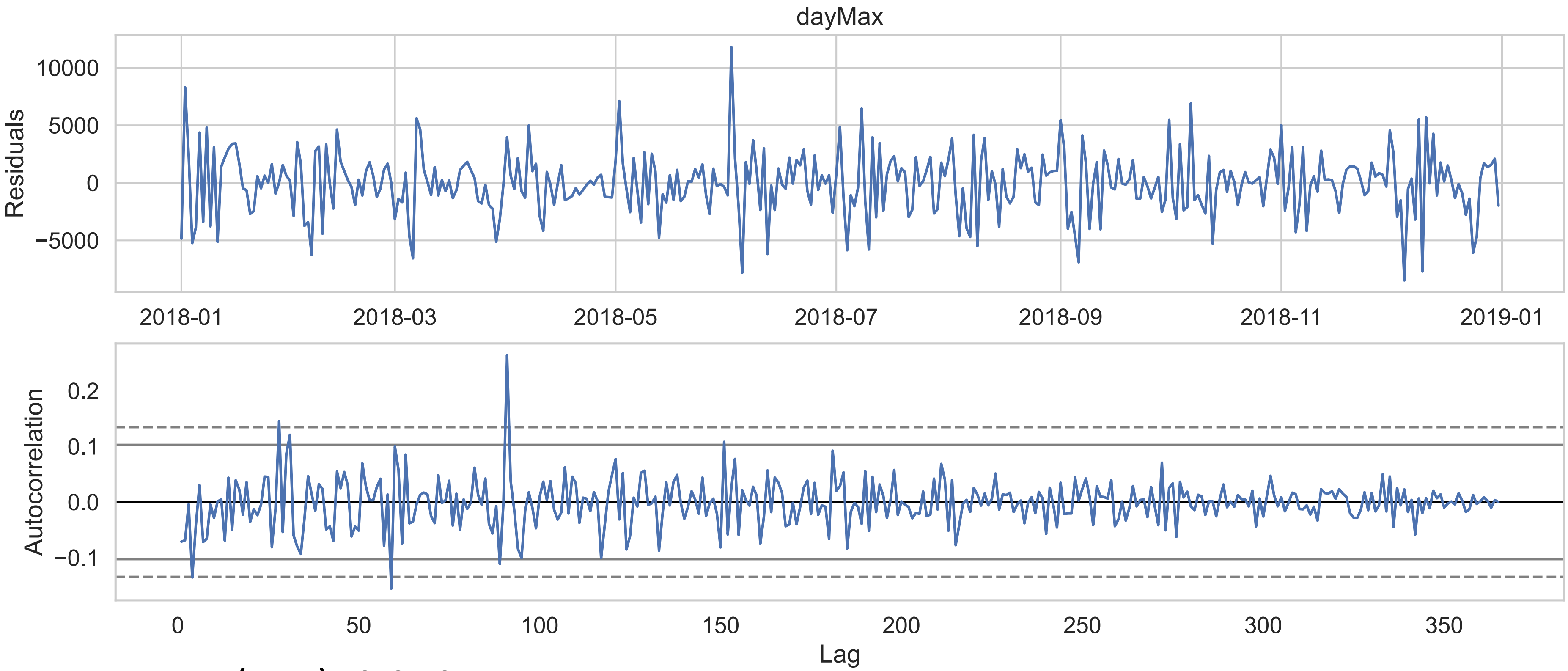
MAE (test): 985.713

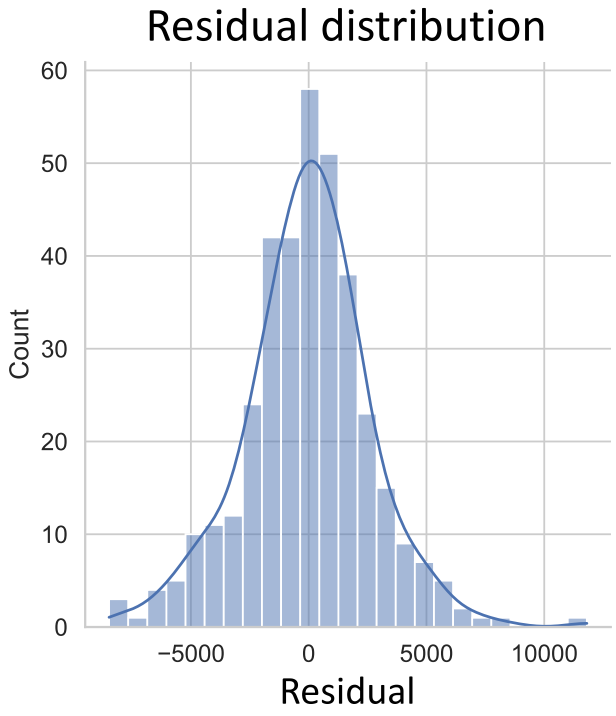
MAPE (test): 1.997

Corr(actual y, prediction) = 0.547, p-value = 0.000

$$\text{MAPE}(y, \hat{y}) = \frac{1}{n_{\text{samples}}} \sum_{i=0}^{n_{\text{samples}}-1} \frac{|y_i - \hat{y}_i|}{max(\epsilon, |y_i|)}$$

# Accuracy




R-square (test): 0.310

MSE (test): 7084440.930

MAE (test): 1987.432

MAPE (test): 2.547

Corr (actual y, prediction) = 0.577, p-value = 0.000

# Conclusion & Future Work

- SARIMAX model on forecasting min/max energy load.
- Used Auto ARIMA for hyperparameter tunning.
- Accuracy: MAPE = 2 (dayMin) and 2.55 (dayMax).

- Future work:
  - Combining the predicted daily Min/Max and hourly load pattern of a day (decomposition).
  - Other models:
    LSTM (Long-Short Term Memory);
    NNETAR (Neural NETwork AutoRegression).