

SDTMSA Project



Pavel Shumkovskii

Summary

- Preliminary analysis
- Advanced analysis
- Predictive modelling



Preliminary analysis

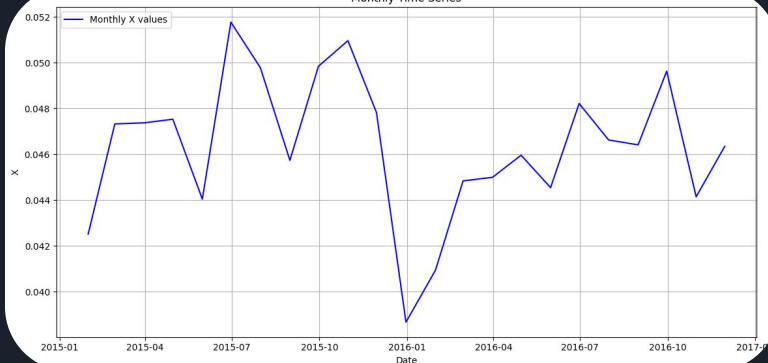


Data

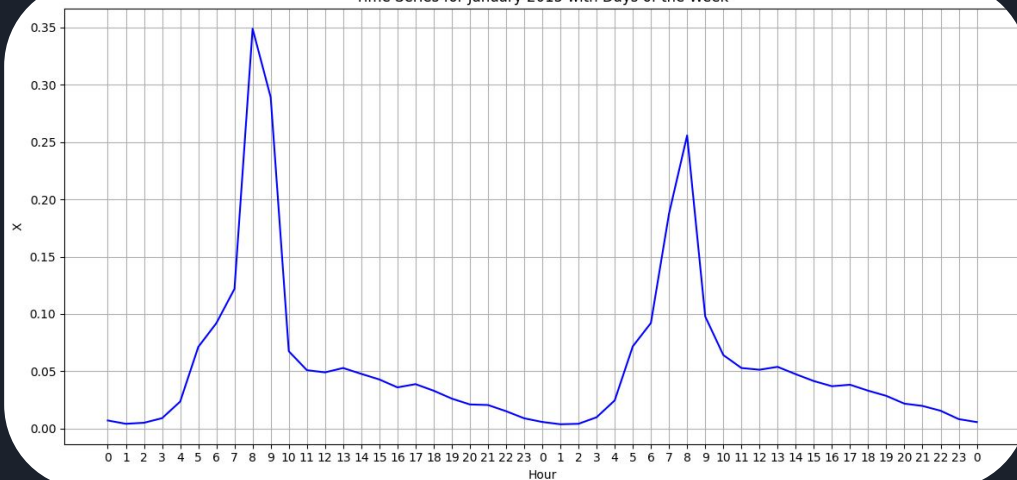


- *Signal-like*
- $[0, 0.45]$
- *Seasonal-cycle patterns*

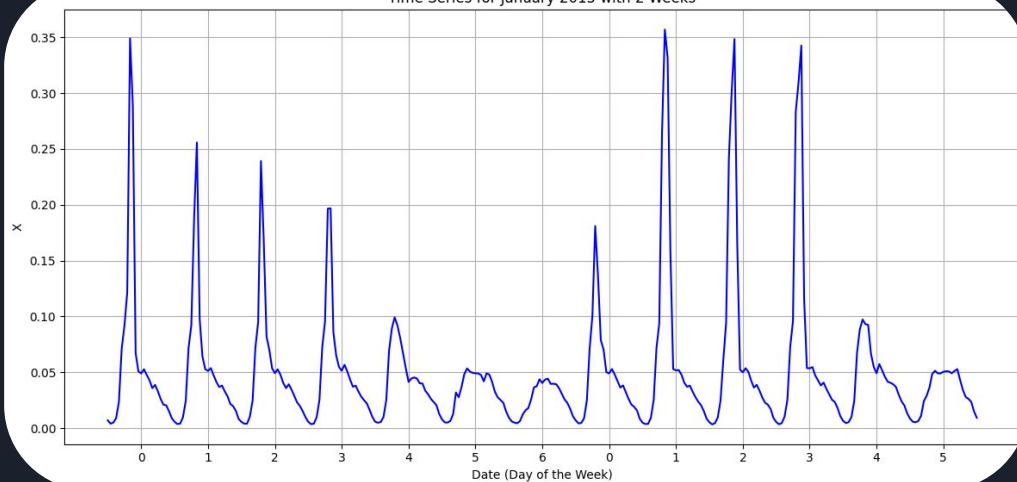
Monthly Time Series



Time Series for January 2015 with Days of the Week



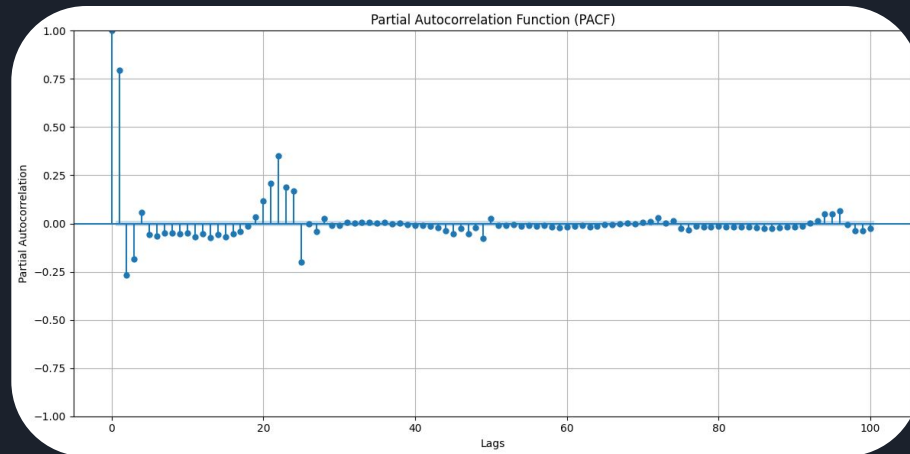
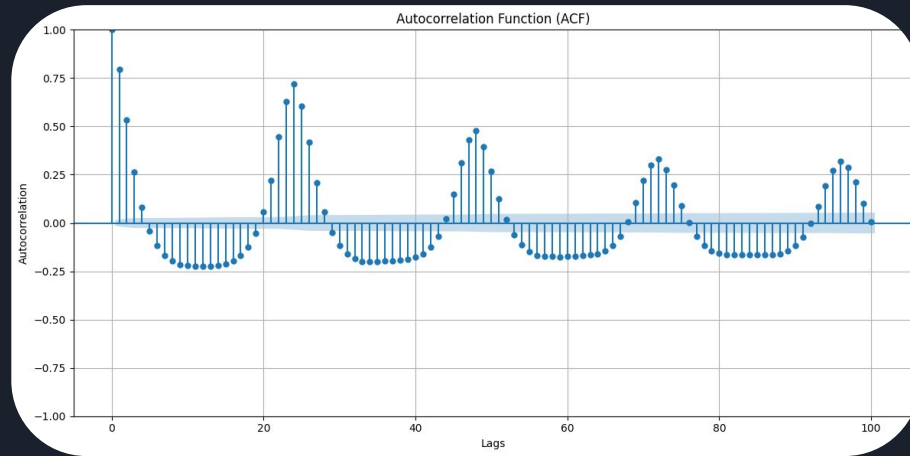
Time Series for January 2015 with 2 Weeks



Stationary tests



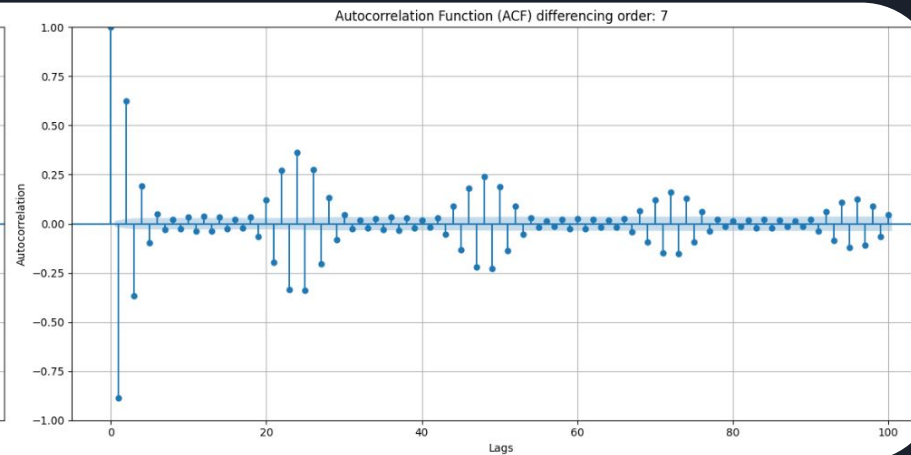
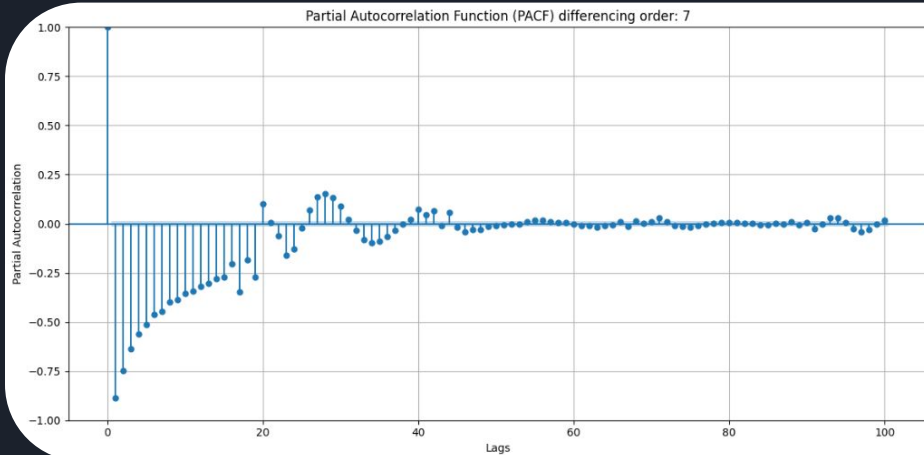
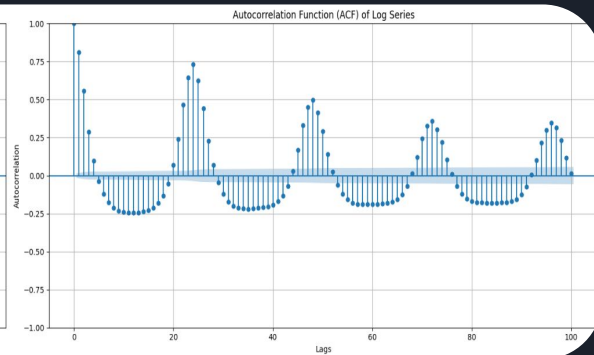
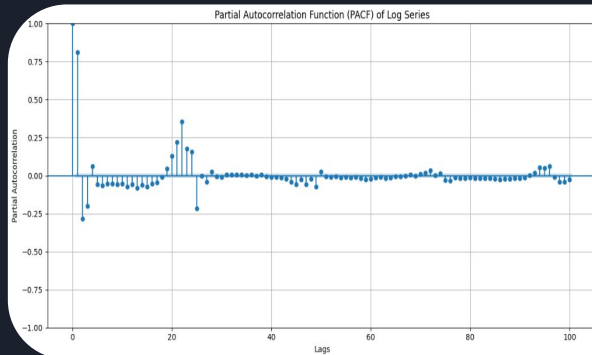
- ADF: 0.00
- KPSS: 0.1



Differencing



- 7-order differencing
- Log differencing



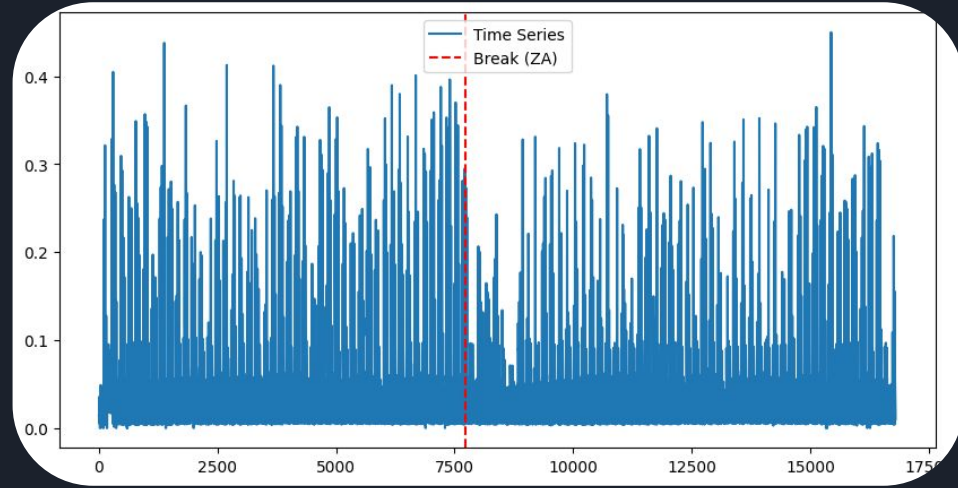
Advanced analysis



Manual approach

Tested approaches

- **Hurst exponent (0.9) + fractional differencing (0.1)**
- **Zivot-Andrews test** -> structural break
- **Fast Fourier Transform (FFT)**
- **Multiple seasonalities**





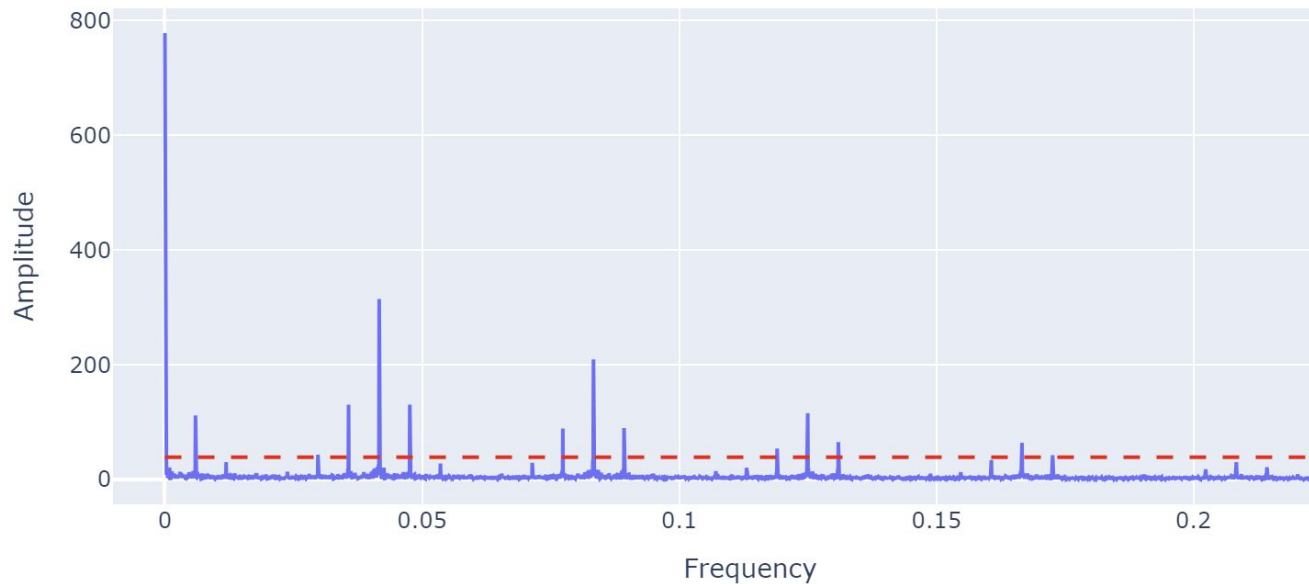
Fast Fourier Transform

Found seasonality



- 168 hours
- 24 hours
- 12 hours
- 8 hours

Fourier Spectrum

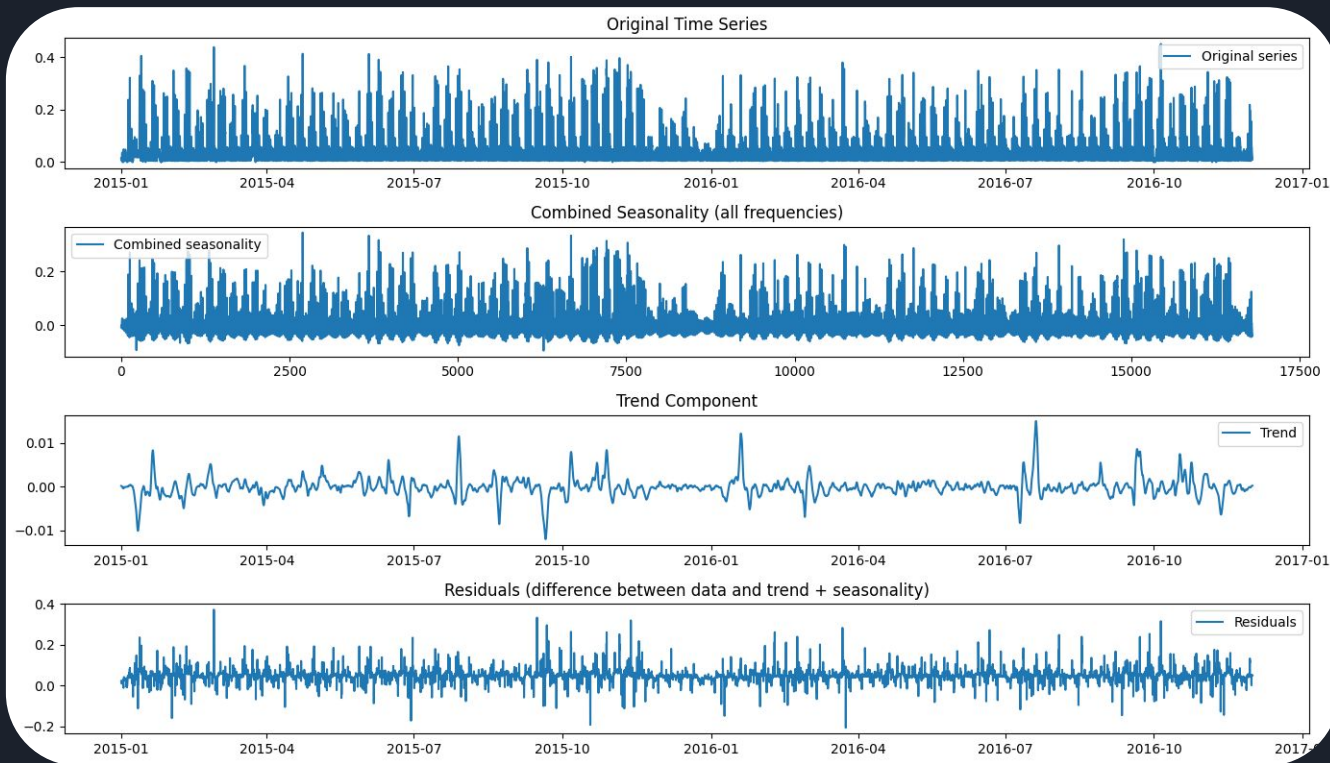


Multi - seasonal decomposition

Found seasonality



- *High errors*
- *Noisy trend*



Predictive modelling

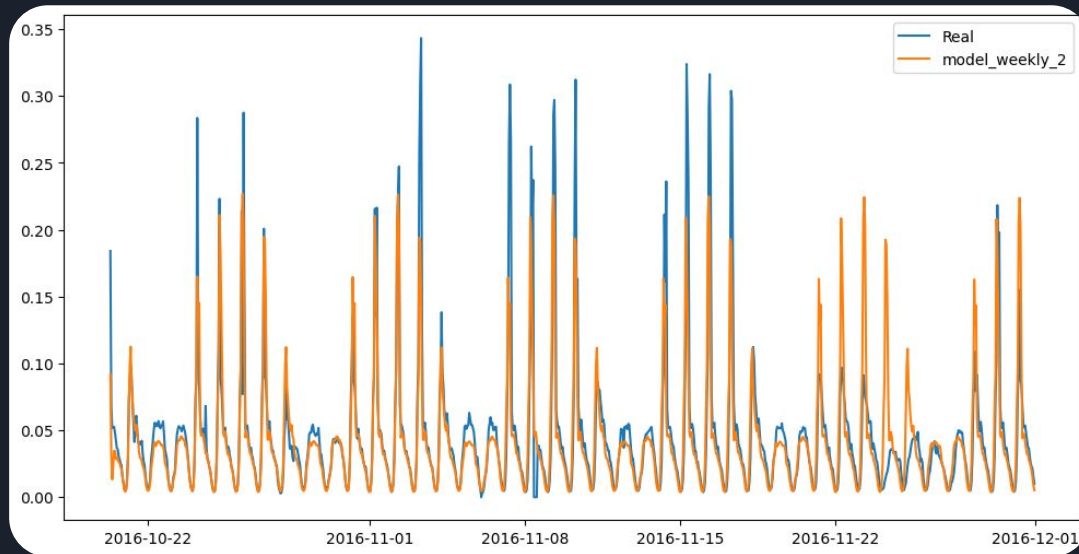




Arima

SARIMAX

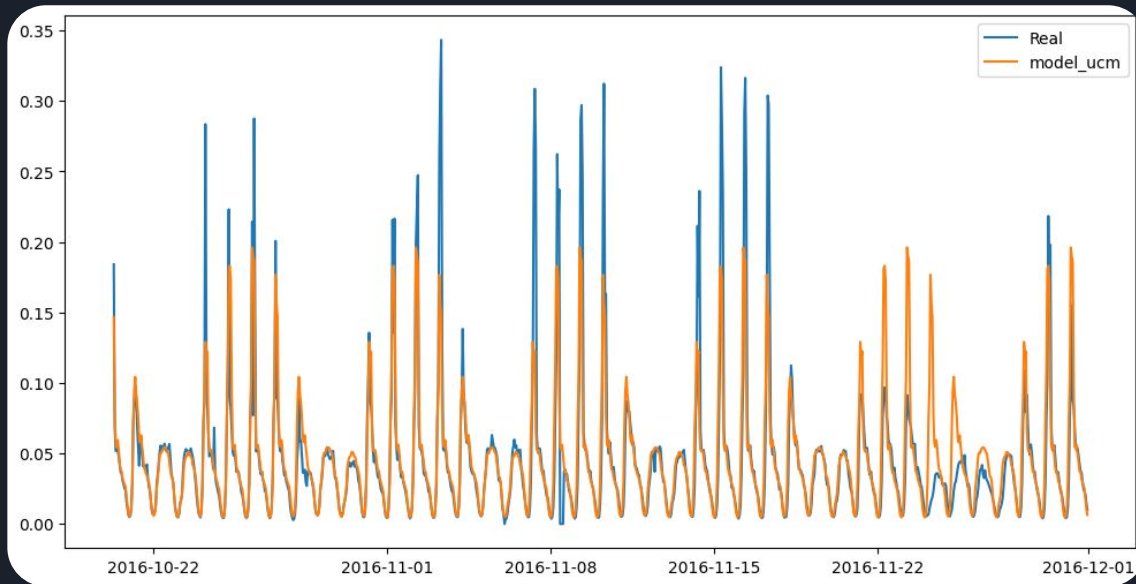
- ARIMA (p,d,q) - **(2, 0, 2)**
- Seasonal order - **(1, 0, 1, 24*7)**
- Train set - last **[-2000:-1000]** values
- Test set - last **[-1000:]** values
- MAE = **0.014**



UCM

UCM

- Local level
- Seasonal component (24*7)
- Irregular component
- Stochastic cycle component
- MAE = **0.012**





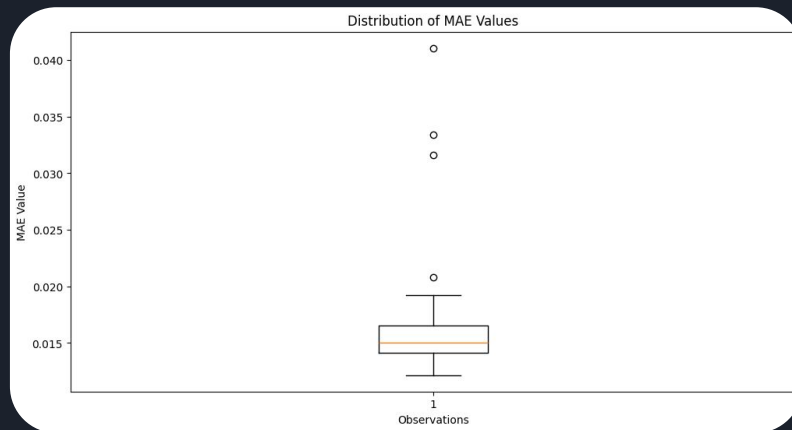
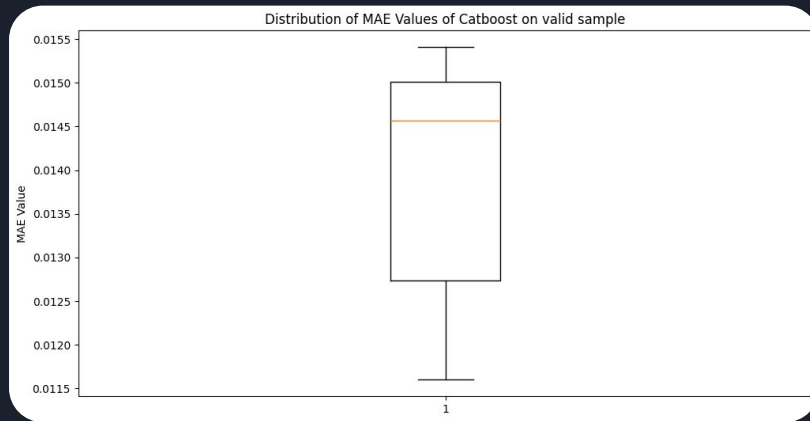
CatBoost

- 1000 lags
- 1-day-ahead predictions
- MAE \approx 0.0145

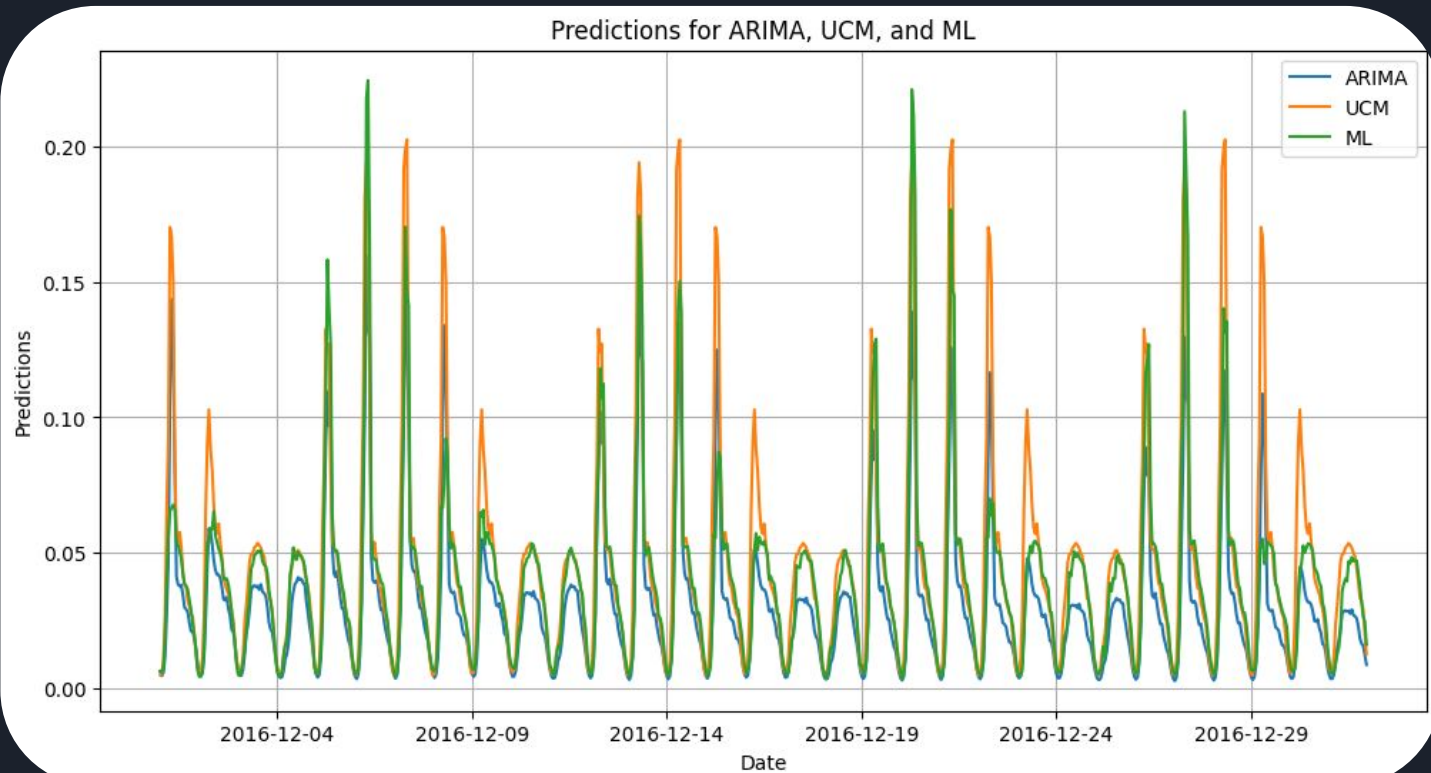
LSTM

- 1000 lags
- 1-shot-predictions
- 4 layers, 128 hidden
- MAE \approx 0.015

Catboost + LSTM



Final predictions





**Thanks
for your attention!**



[repo](#)