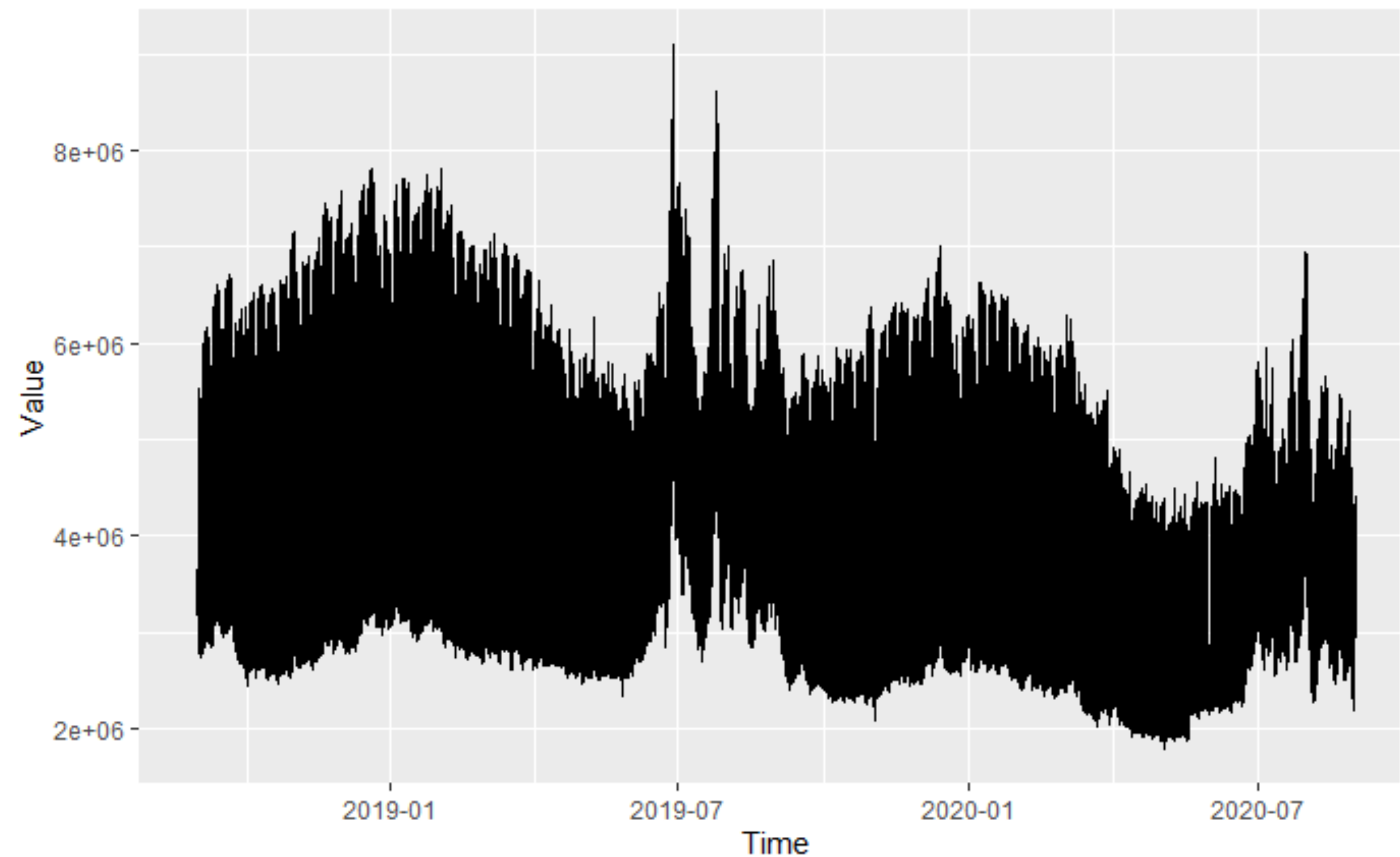


# **Streaming Data Management and Time Series Analysis**

**Corso di laurea magistrale in Data Science  
A.A. 2020/2021**

# INTRODUZIONE



Implementare almeno un modello:

- ARIMA
- UCM
- Machine Learning

per prevedere i dati di settembre e ottobre.

# PREPROCESSING

- I dati non presentano particolari criticità

- Manca l'ora 03:00:00 nei giorni 2019-03-31 e 2020-03-29 (cambio dell'ora)

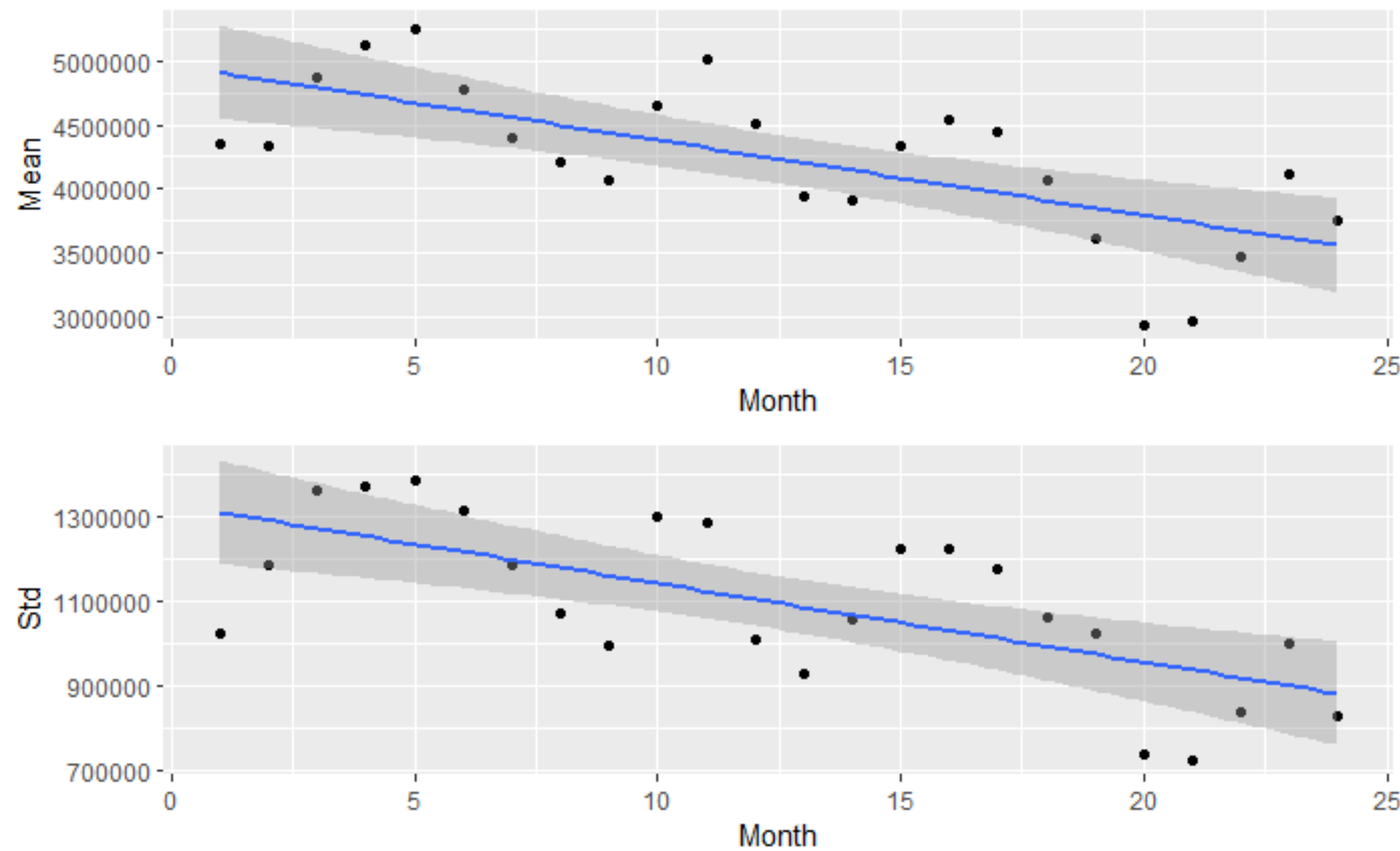


Il valore è stato aggiunto ed è pari a quello dell'ora precedente 02:00:00

- Non vi sono duplicati o valori NaN

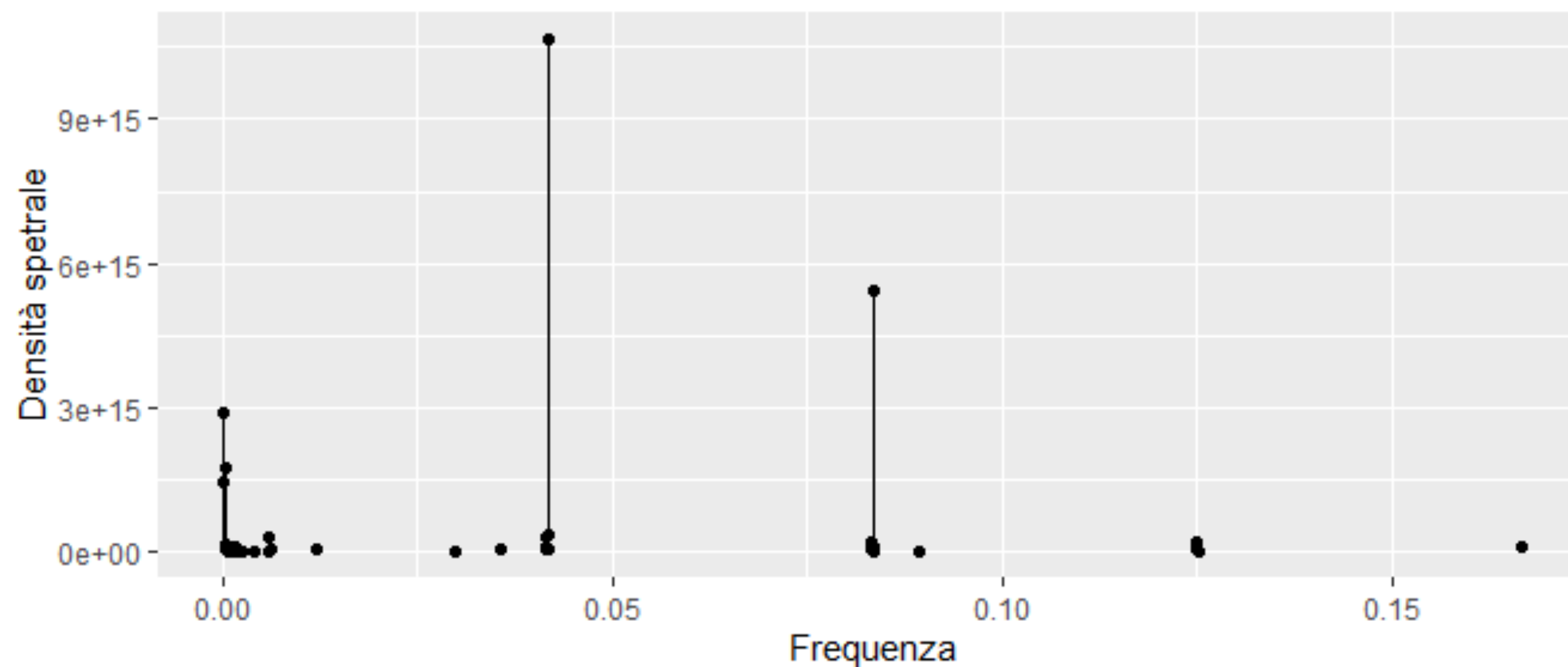
# ESPLORAZIONE

## ANDAMENTO GLOBALE



La serie si caratterizza di un trend decrescente e la varianza dei dati diminuisce con il tempo.

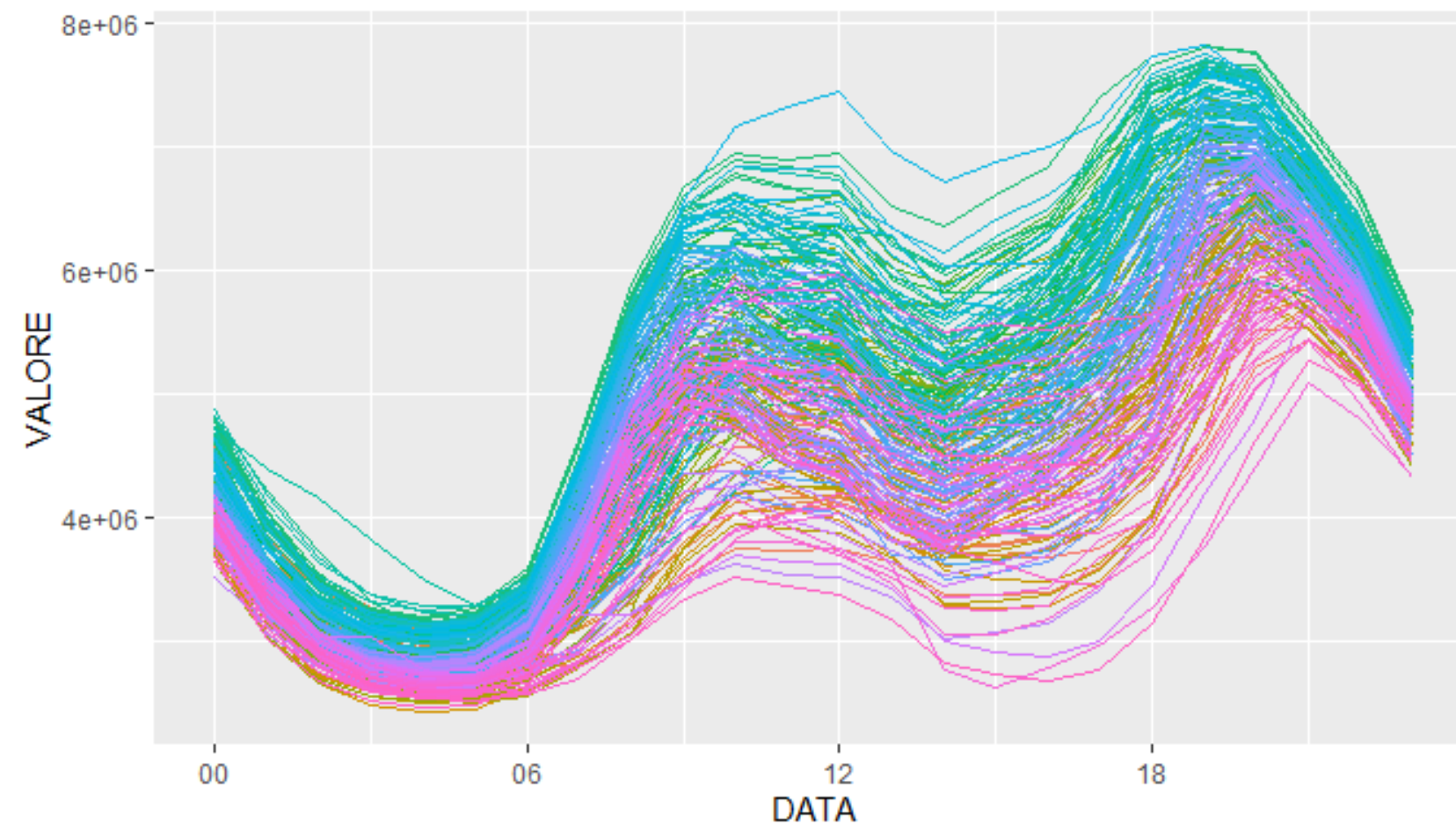
# ESPLORAZIONE STAGIONALITÀ



Il *periodigramma* definisce le frequenze più rilevanti nei dati, associate quindi a:

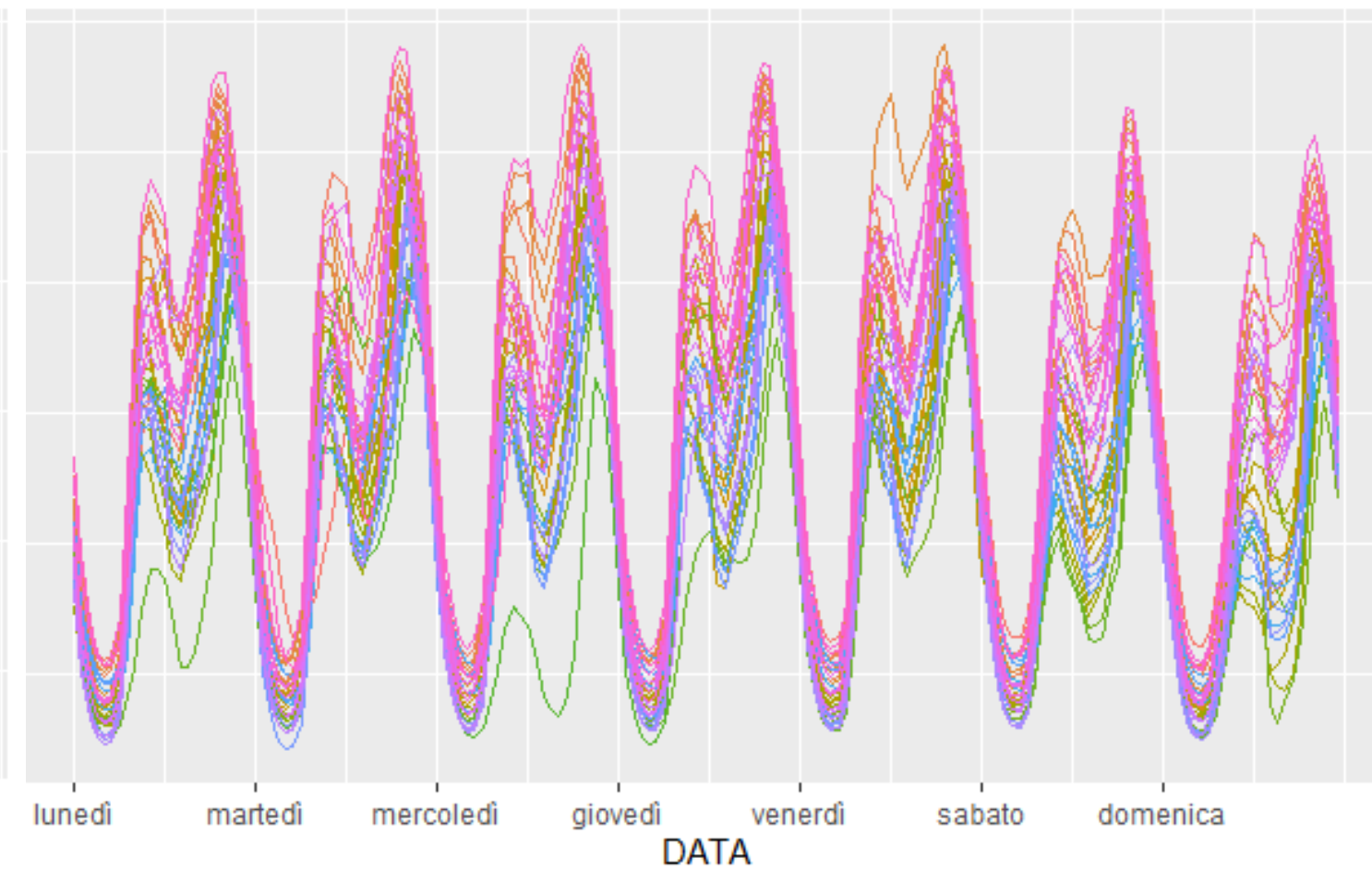
- Stagionalità giornaliera
- Stagionalità annuale.
- Stagionalità settimanale.

# ESPLORAZIONE STAGIONALITÀ



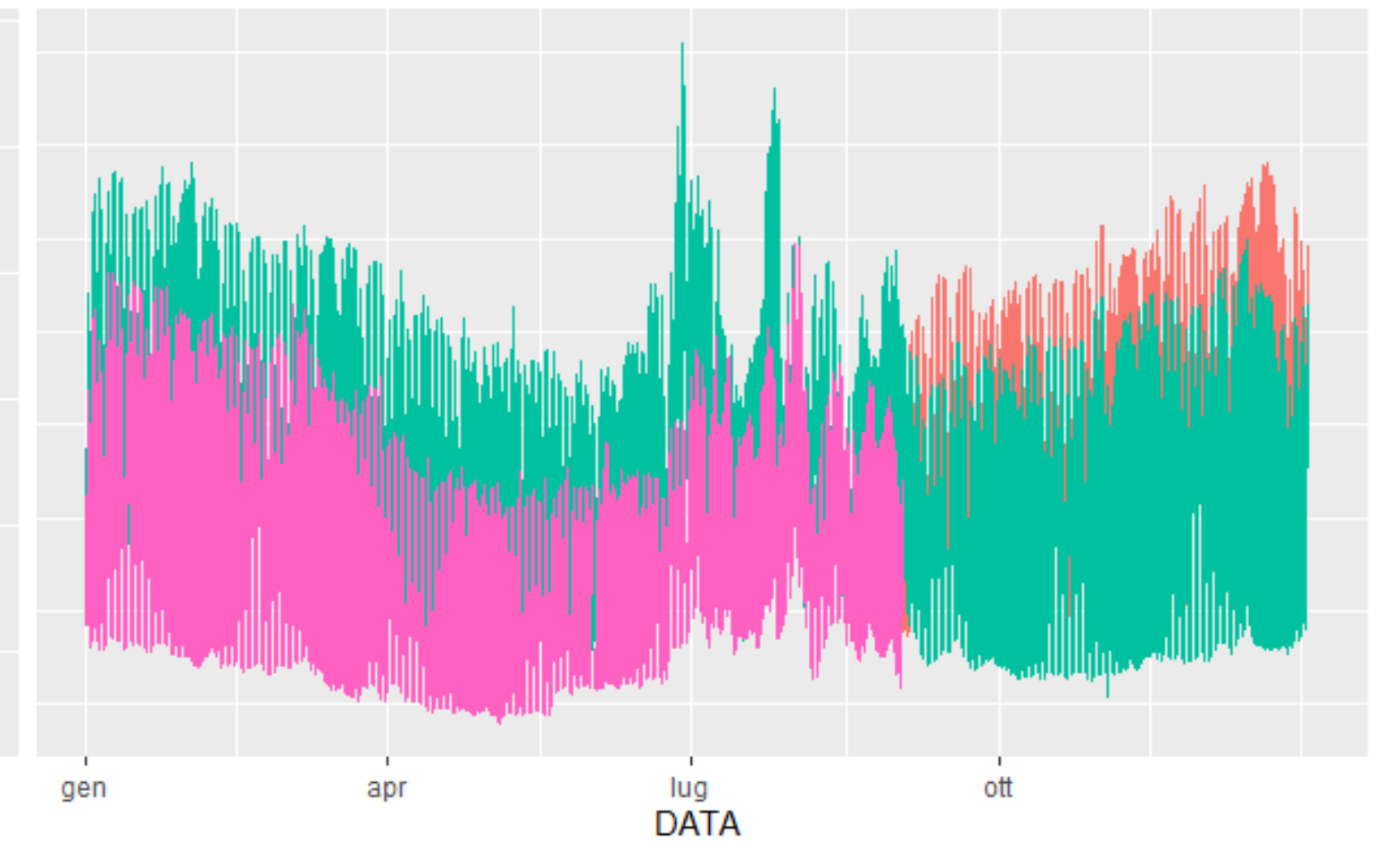
↓

Profilo  
giornaliero



↓

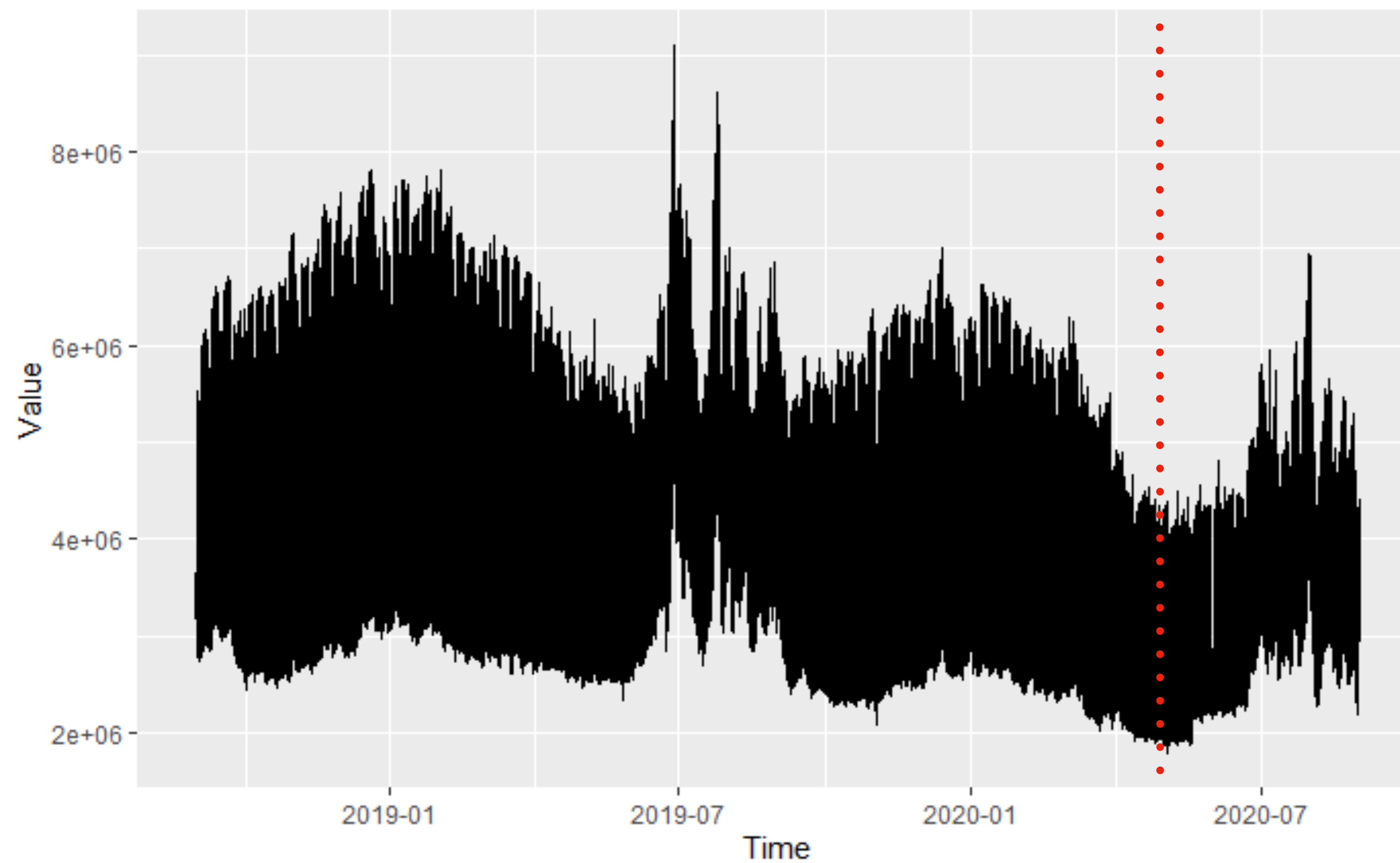
Profilo  
settimanale



↓

Profilo  
annuale

# TRAIN E VALIDATION SET



Train:

fino a 2020-04-01 23:00:00

~80% del dataset

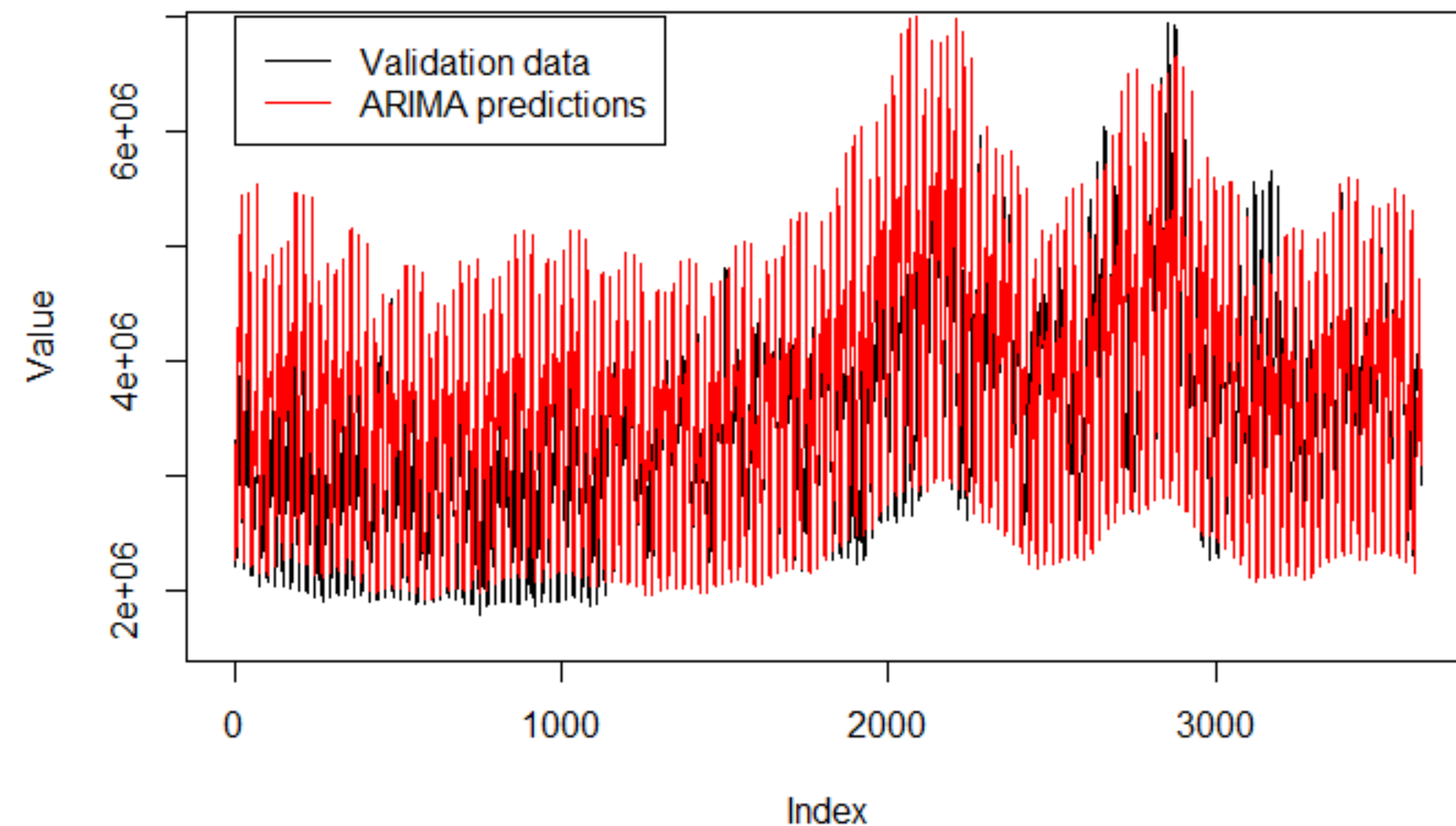
Validation:

la parte rimanente

~20% del dataset



# ARIMA

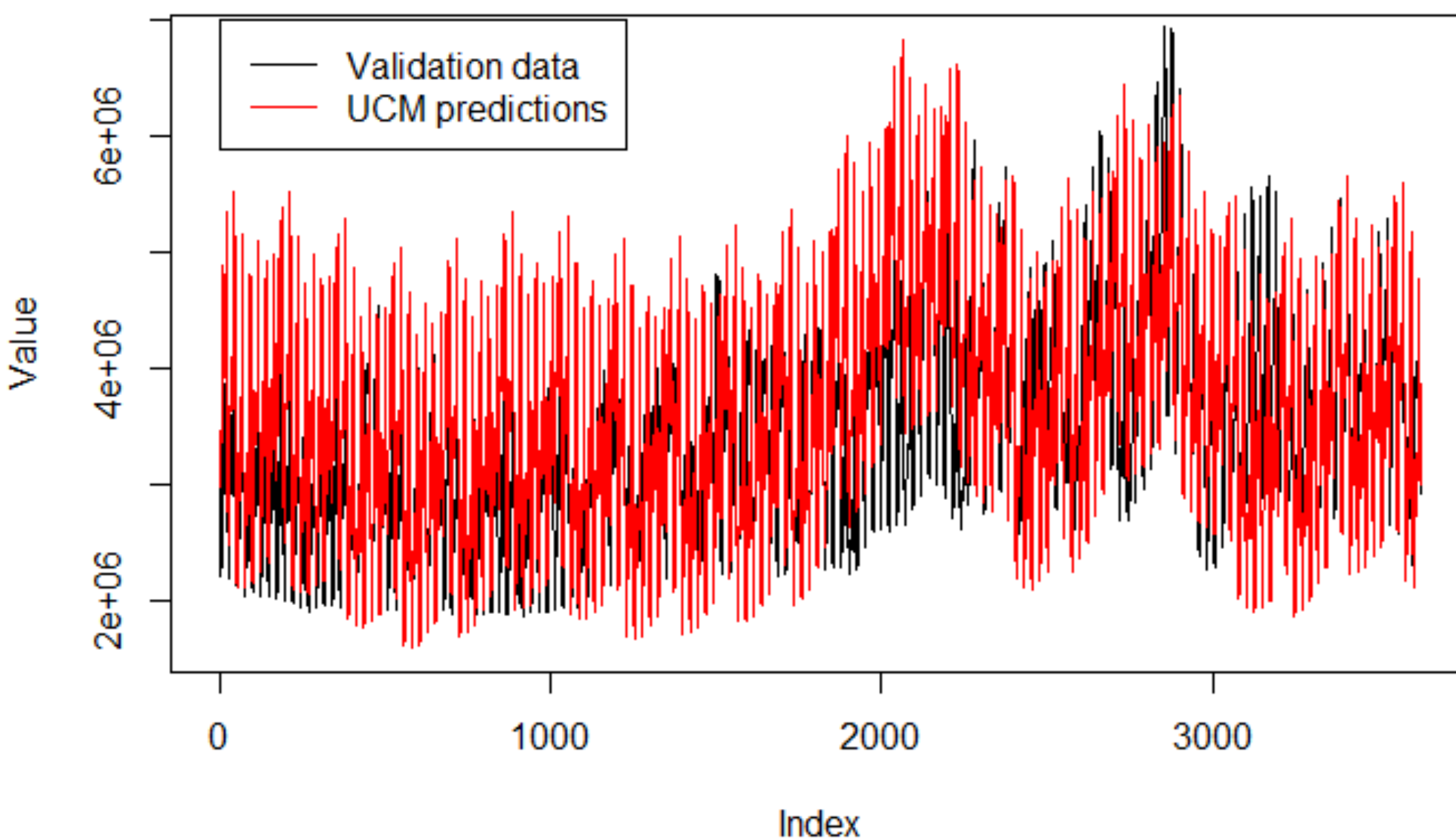


Gli iperparametri migliori:

- $p = 3$ ;  $d = 0$ ;  $q = 2$ ;
- **Stagionalità settimanale**  $24 \times 7$  stocastica  
 $P = 1$ ;  $D = 1$ ;  $Q = 1$ ;
- **Regressori esterni**: stagionalità annuale (16 armoniche) e festività
- Trasformazione **logaritmica**



# UCM



Il miglior modelli si compone di:

- **Local linear trend**
- Stagionalità **giornaliera** (dummy)
- Stagionalità **settimanale** (8 armoniche)
- Stagionalità **annuale** (16 armoniche)

I valori iniziali sono stati modificati opportunamente per ottenere questo risultato.

# RETI NEURALI

Sono state implementate due tipi di architetture per la previsione dei dati attraverso un approccio supervised:

- LSTM
- GRU

con le **stesse** caratteristiche, e.g. numero di neuroni di output, learning rate, tasso di dropout, **dimensione delle finestre di input** (2 mesi, 2 settimane etc.).

# RETI NEURALI

Utilizzando Google Colab:

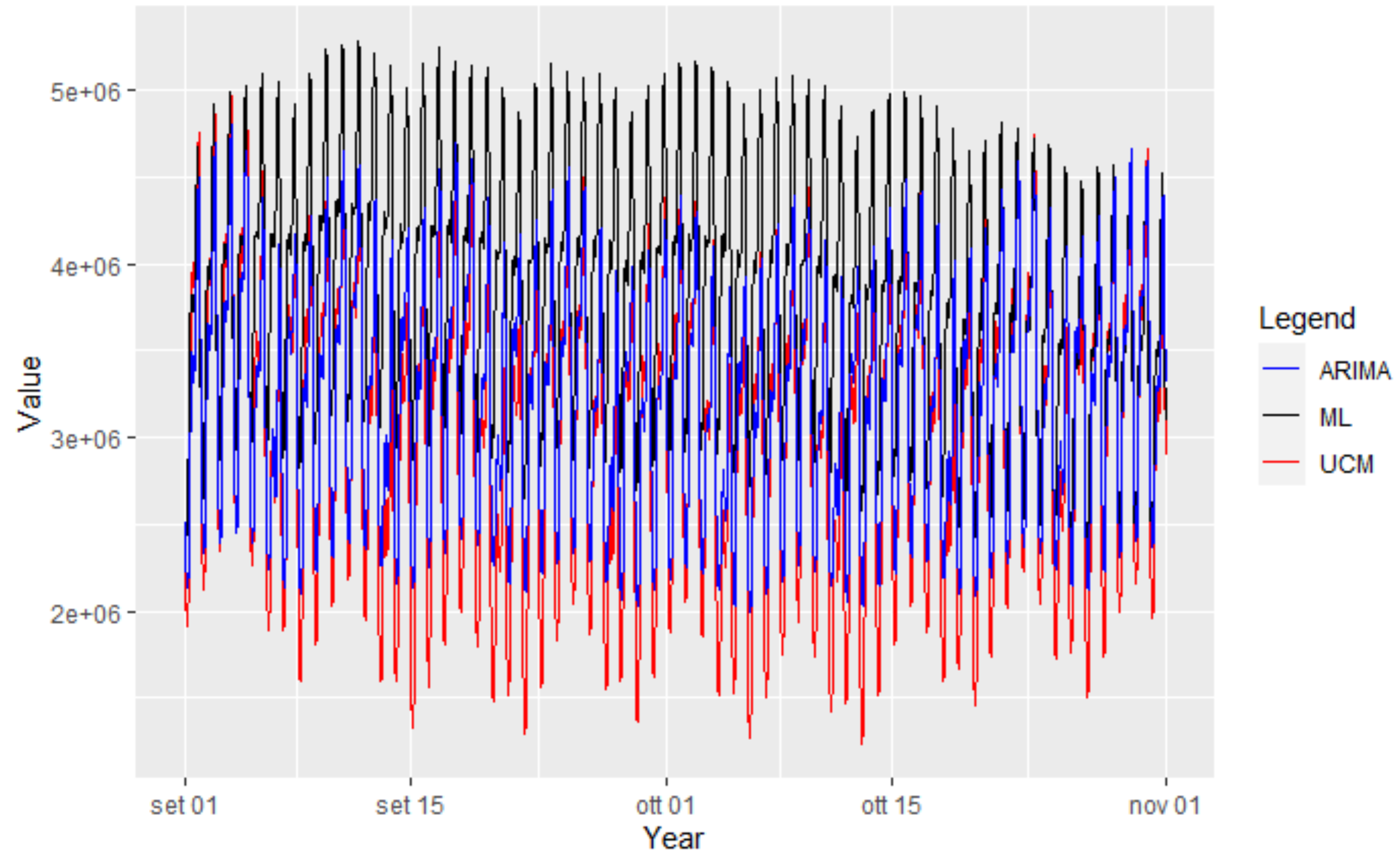
- è possibile sfruttare la tecnologia cuDNN attraverso l'uso della GPU
- non è stato possibile implementare modelli complessi (probabile underfitting) a causa delle limitazioni della RAM.

# MODELLO FINALE

Il modello migliore in termini di training e validation error è l'ARIMA.

MODELLO	MAE (Training)	MAE (Validation)
ARIMA	59699	480520
UCM	107765	513543
LSTM	4496664	3604482

# PREVISIONI



# CONCLUSIONI

- L'ARIMA e gli UCM si sono rivelati dei buoni modelli predittivi.
- Si potrebbero usare altre tecniche come Prophet, NeuralProphet o KNN

GRAZIE PER  
L'ATTENZIONE

