Time Series Project

Pedestrian Traffic Forecasting

A.Y. 2025/2026
Time Series Project

Presented By
Giovanni Noè

Table of Contents

Exploratory Data Analysis

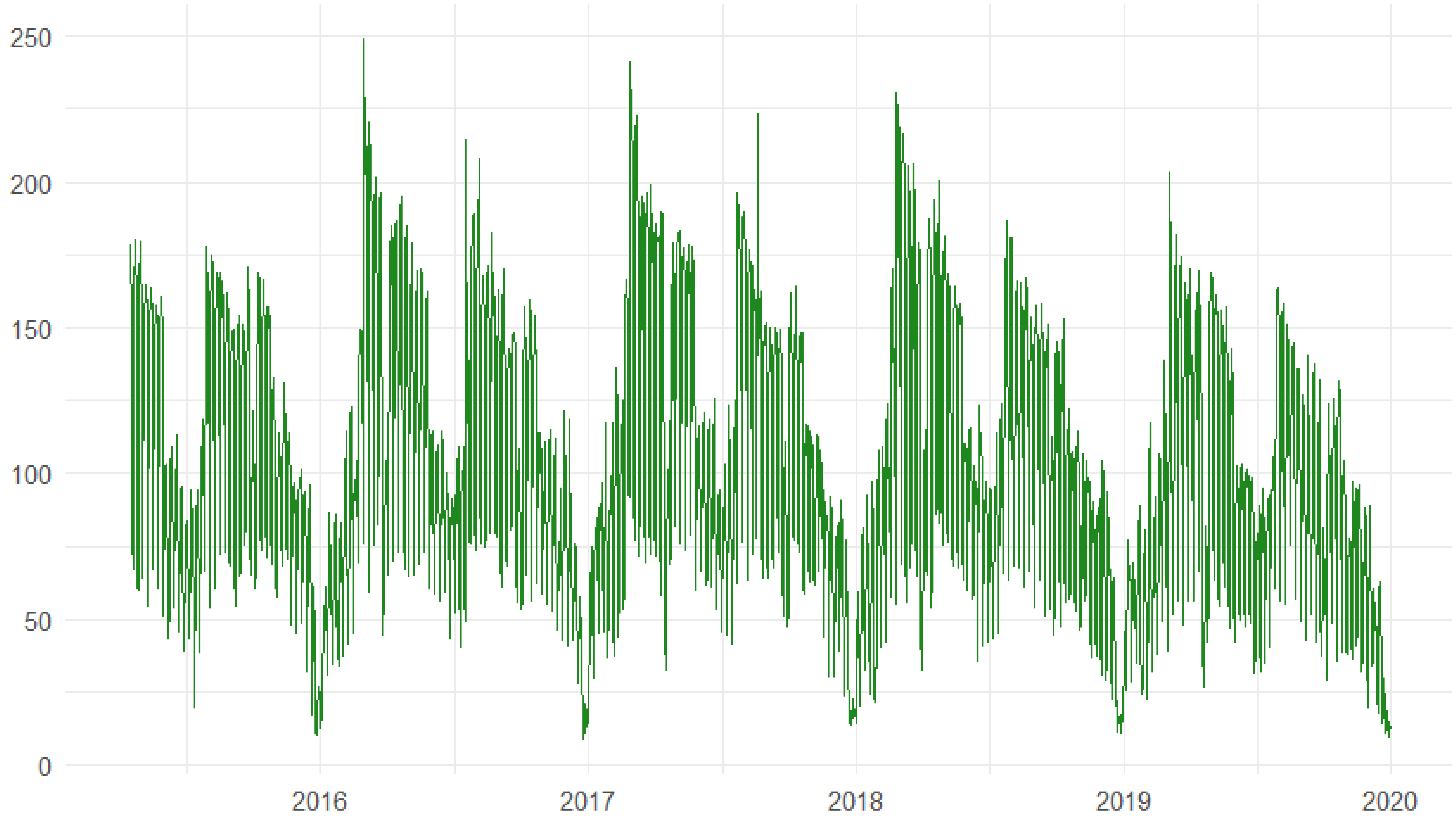
EDA - Preliminary Steps

- Creation of the variables **date** and **hour**
- **Missing values** are only the observations to predict
- **Inferring values** for September 15, 2017
- Checking for **Daylight Saving Time**

EDA - Descriptive Statistics

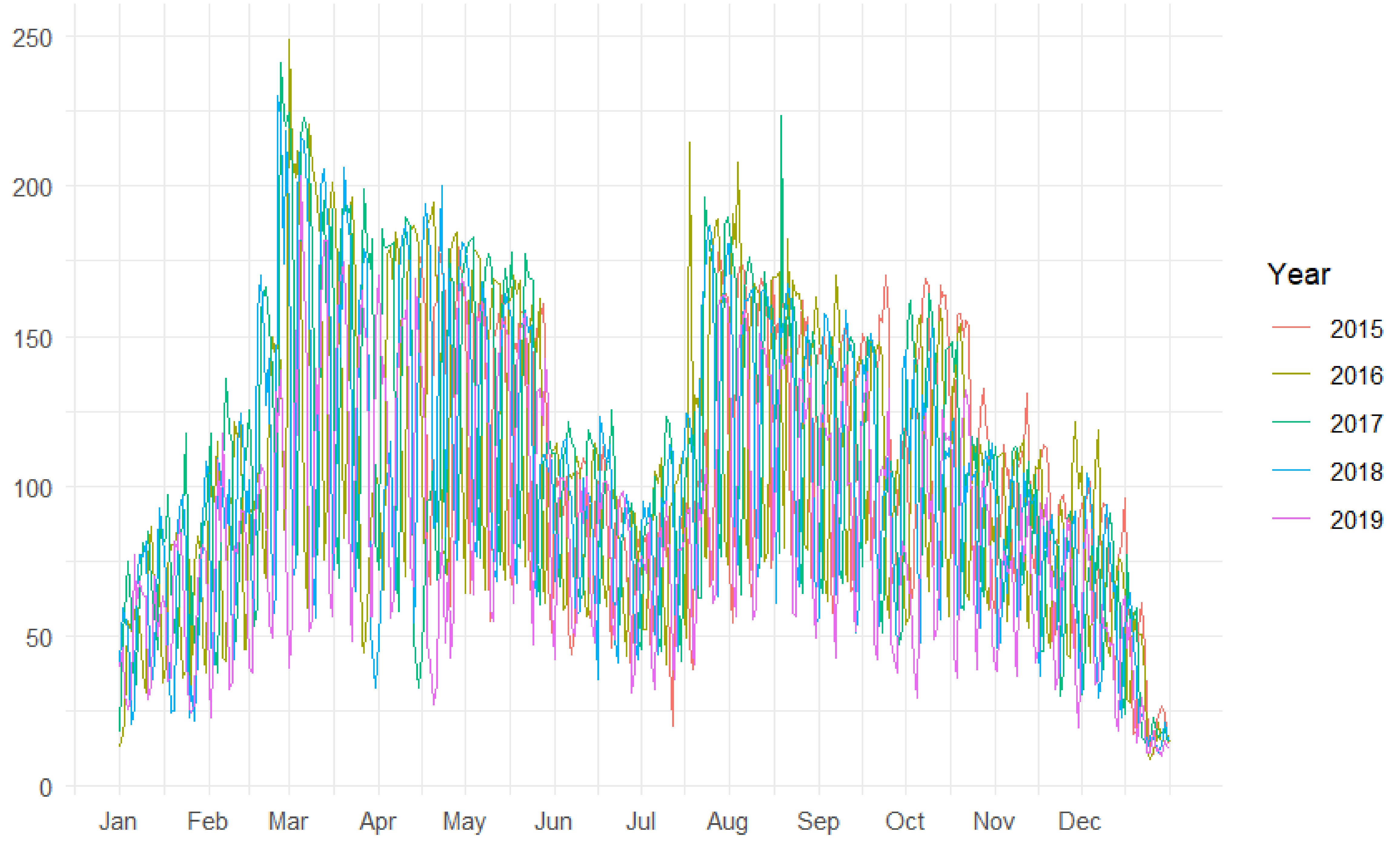
- **Mean** value → 101.5
- **Maximum** value at 11 a.m. → 906
- **Minimum** value between 9 p.m. and 7 a.m. → 0
- **Peak hour** is 4 p.m. → 196.70
- **Lowest average** at 2 a.m. → 8.85

EDA - Trend



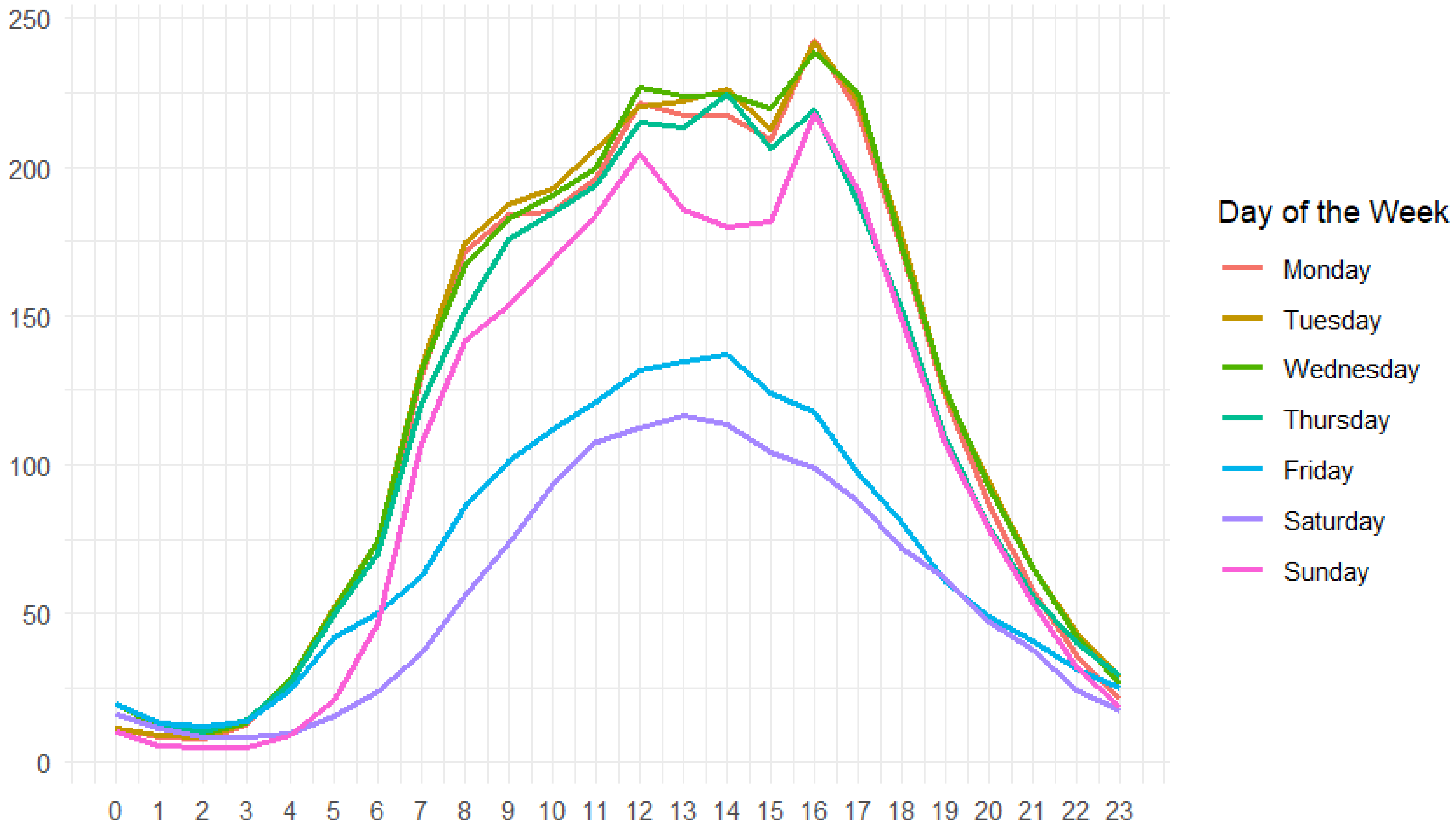
- Slightly **decreasing** trend

EDA - Yearly Seasonality



- Highest decrease in **December-January**

EDA - Daily and Weekly Seasonality



- Evident **daily** and **weekly** seasonalities

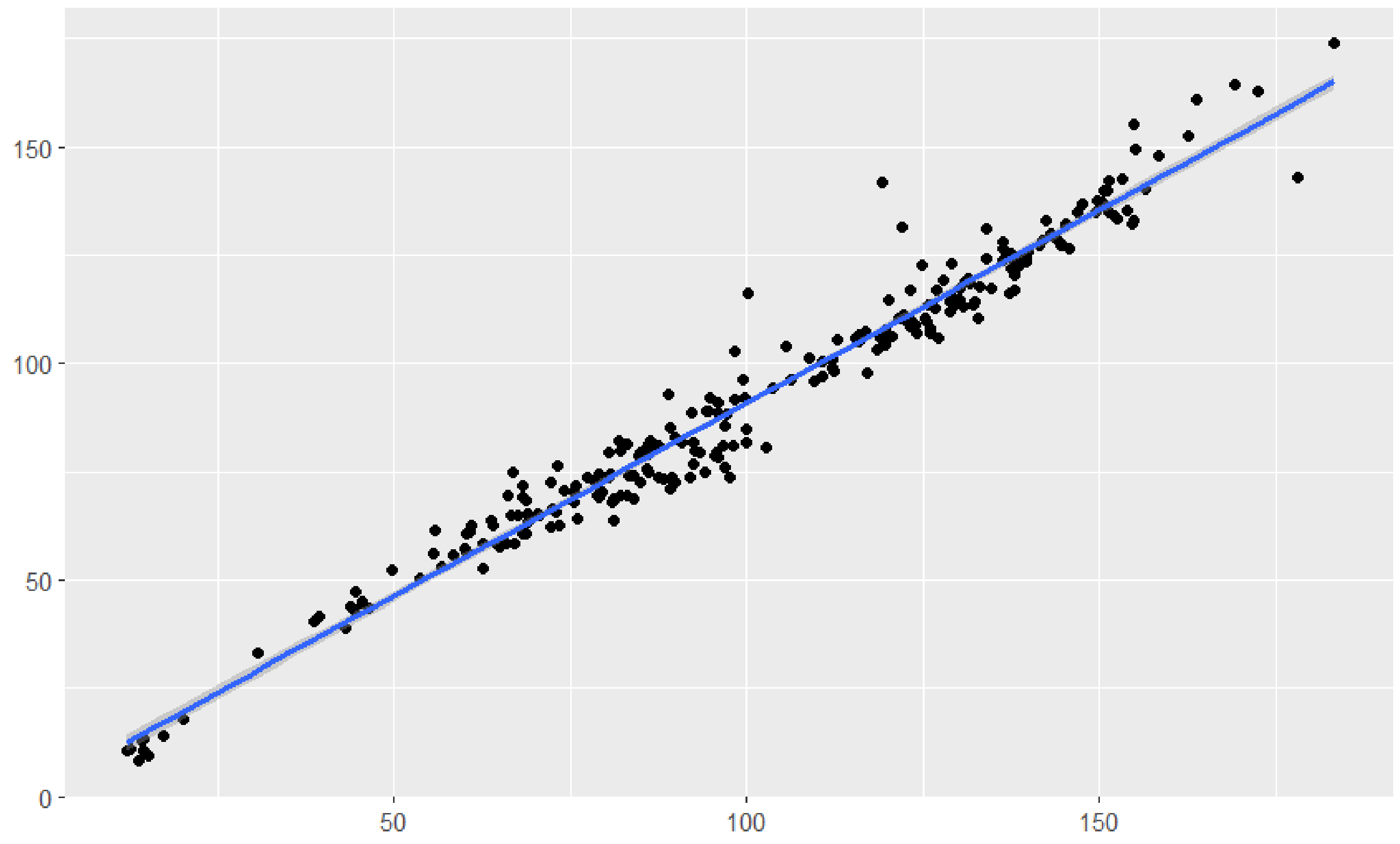
Models

Methodology

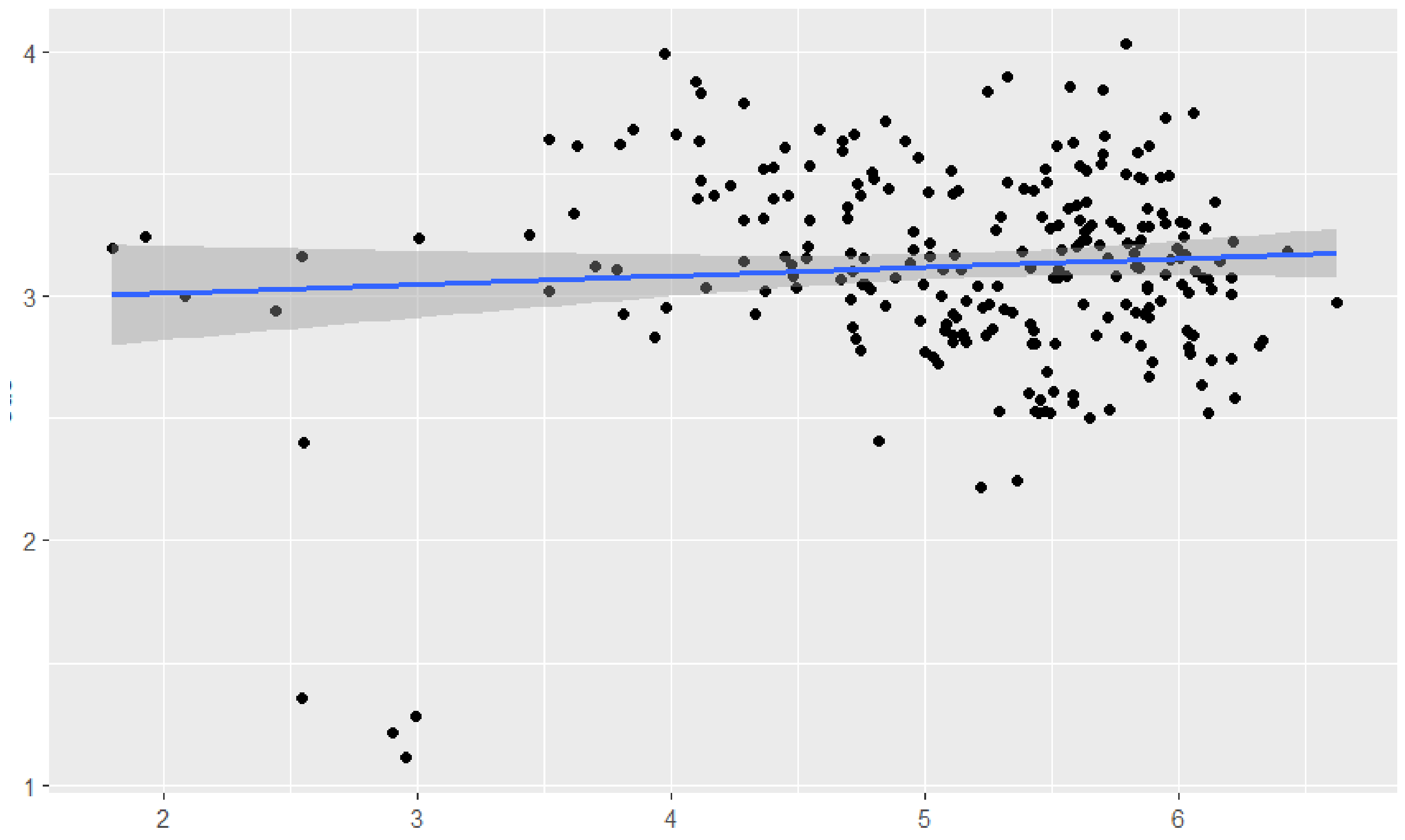
- **Australian Holidays** dummy variable:
  - Australia Day - 01/26
  - ANZAC Day - 04/25
  - Easter
  - Easter Monday
  - Christmas Day - 12/25
  - Boxing Day - 12/26

- Models assessed leaving out **last 30 days of observations**:
  - True values vs forecast **plot**
  - **MAE**

ARIMA Model - 1



Box-Cox Plot of the original TS



Box-Cox transformation (**lambda=0.15**)

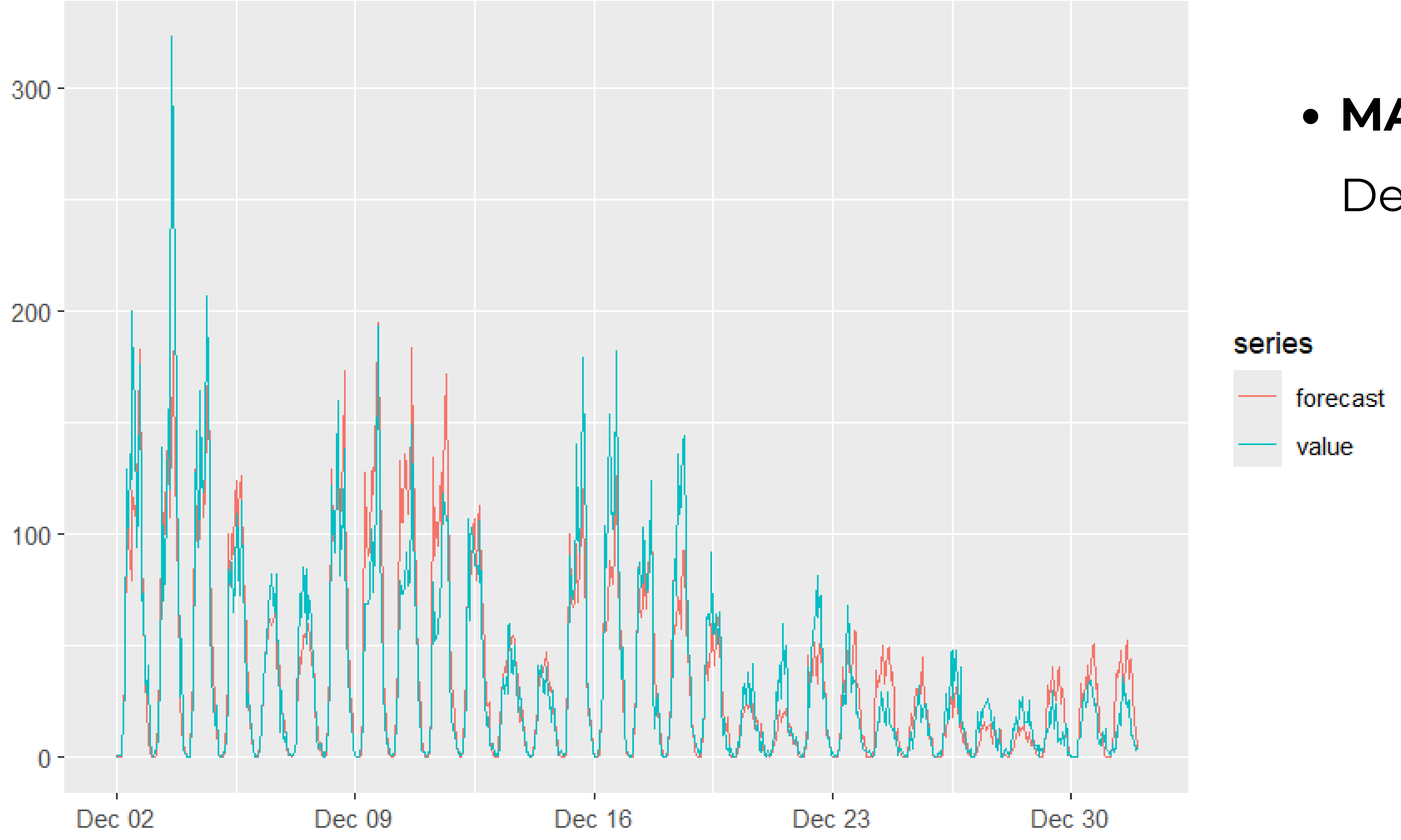
for a constant variance

ARIMA Model - 2

- **15 cosine-sine pairs** to model yearly seasonality
- **One model for each hour** (same parameters)
- **d=1** and **D=1** with period 7
- Recoursive ACF-PACF inspection and **process identification**:
  - **AR(1)** - hardest choice
  - **SMA(1)**
  - **MA(3)**

$$ARIMA(1,1,3)(0,1,1)_7$$

ARIMA Model - 3

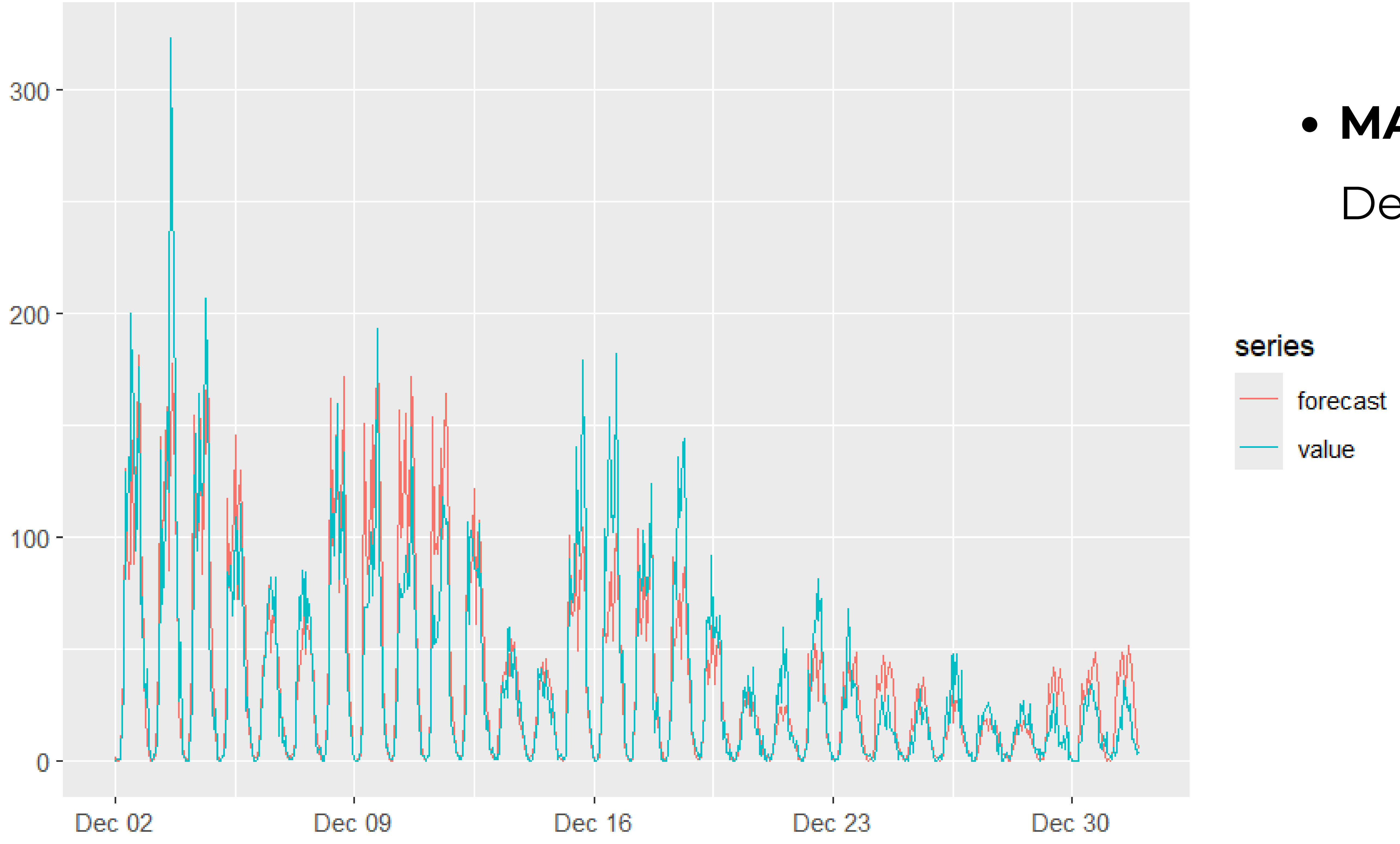


- **MAE = 12.02** on December observations

Unobserved Components Model - 1

- **log(y+1)** transformation
- **One model for each hour** (same parameters)
- Components:
  - **Dummies** - Holidays, Summer vacations, Fridays and Saturdays on holidays
  - First-order **Trend**
  - **Seasonal Dummy** with period 7
  - **Seasonal Trigonometric** with period 365 and 15 harmonics

Unobserved Components Model - 2



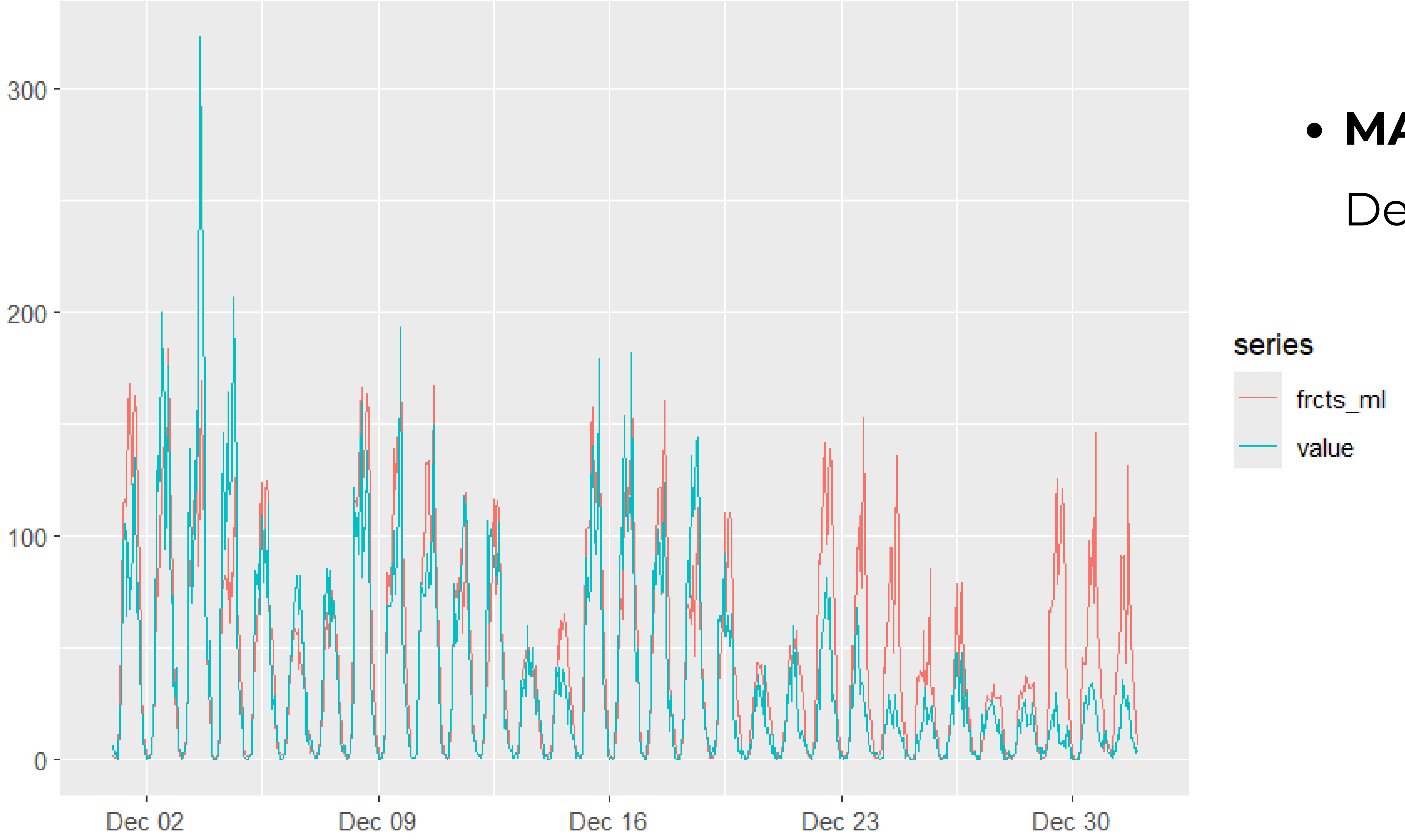
- **MAE = 12.63** on December observations

Gradient Boosting Machine - 1

- **Features:**
  - 24 hours lag
  - 168 hours lag
  - 365 days lag
  - Year day
  - Week day
  - Holiday dummy
  - Summer vacations dummy

- **Lightweight** but strong performances
- 1000 Trees
- **Recoursive Forecasting**

Gradient Boosting Machine - 2



- **MAE = 18.36** on December observations

Conclusions

Conclusions

- **ARIMA** and **UCM** achieved **similar results** on out-of-sample observations
- **Classical** models **outperformed** the **Machine Learning** model
- Discussed **results are partial**

- Final considerations **about predictions:**
  - **ARIMA** → **highest** forecasts (62 mean, and 263 max)
  - **UCM** → **intermediate** predictions (37 mean 195 max)
  - **ML** → **lowest** forecasts (24 mean, 113 max)

Thank You

For Your Attention