

CAPSTONE PROJECT - SHINE ROSHAN

TSA Throughput Prediction

2022



Question



Using historic hourly data starting at the beginning of 2019, can we predict the number of passengers going through airport security for a given hour?

Motivation

Recovery

Predict if and when
the travel industry
may recover fully
from Covid 19



Efficiency

Proper staff and resource
allocations

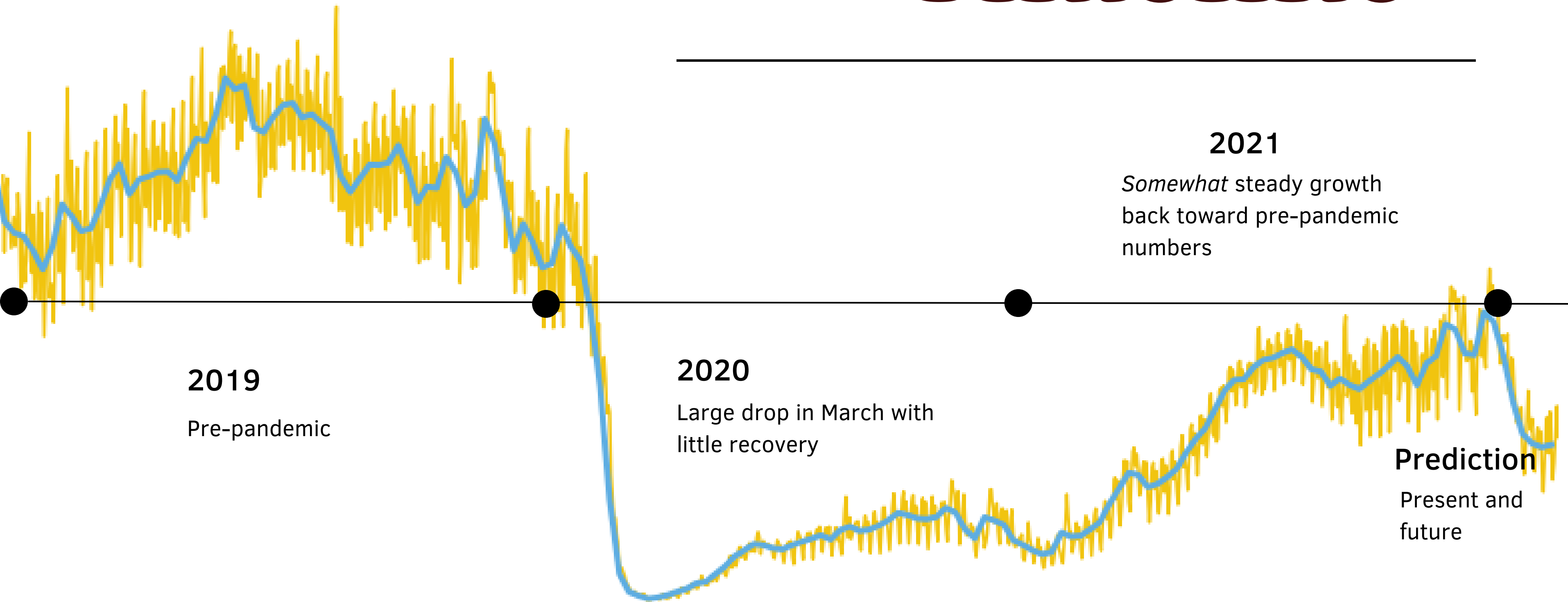


Customer Service

Better customer service
will create loyal
customers



Timeline



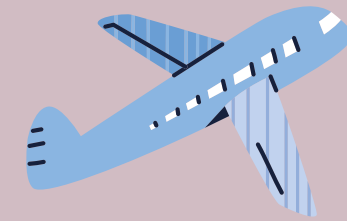
Stakeholders



Airport
businesses



Hotels
nearby



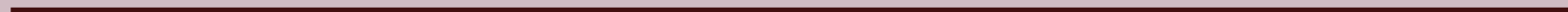
Airlines



Airports



Shuttles
and Taxis



Potential Challenges

**Global
events**



Pandemics



Weather

Data Collection

01

Source - Repository

- Github repository (updated regularly)
- csv files for each airport
- Gate level counts

02

Dataframe - individual time seires

- Aggregate throughput for each airport
 - Column - Single airport
 - Datetime index (time series analysis)
-

Data Structure

HOURLY DATA

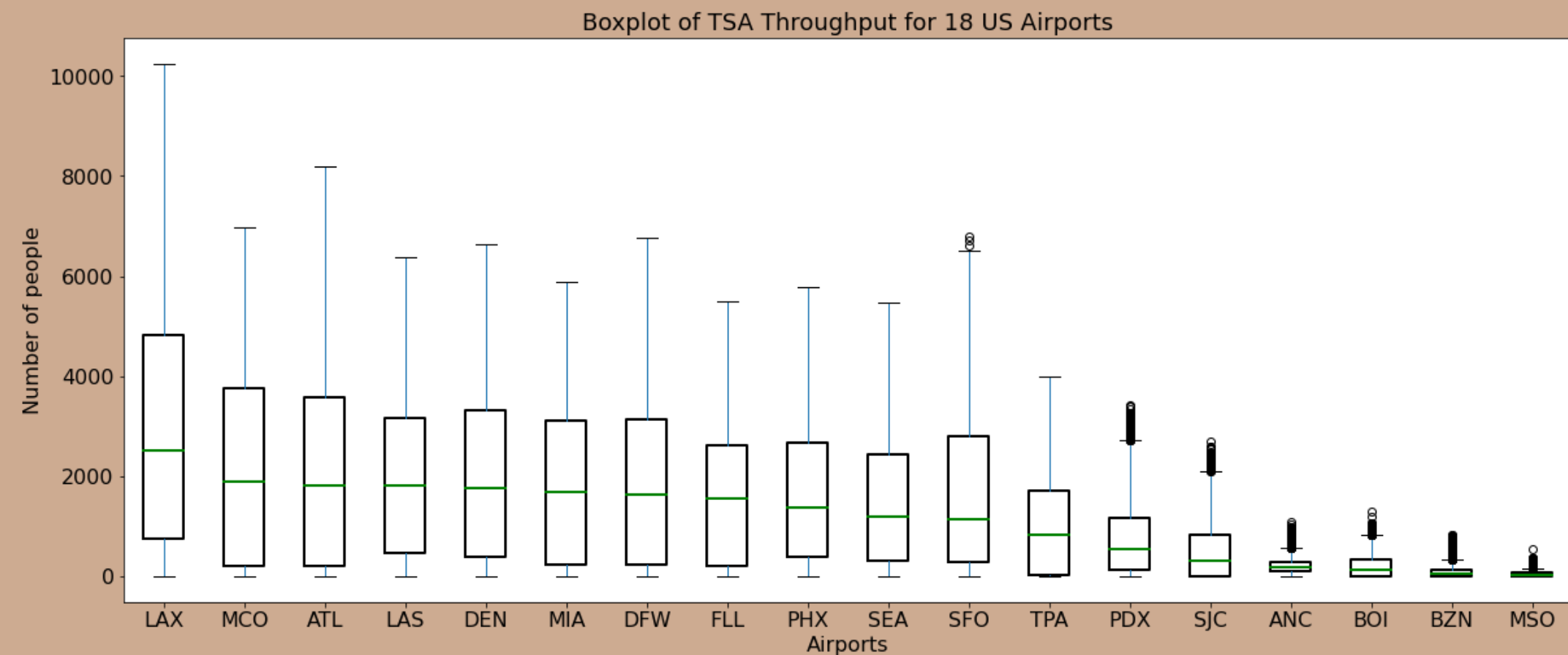
Over 3 years

Start - Dec 31, 2018

End - Feb 5, 2022

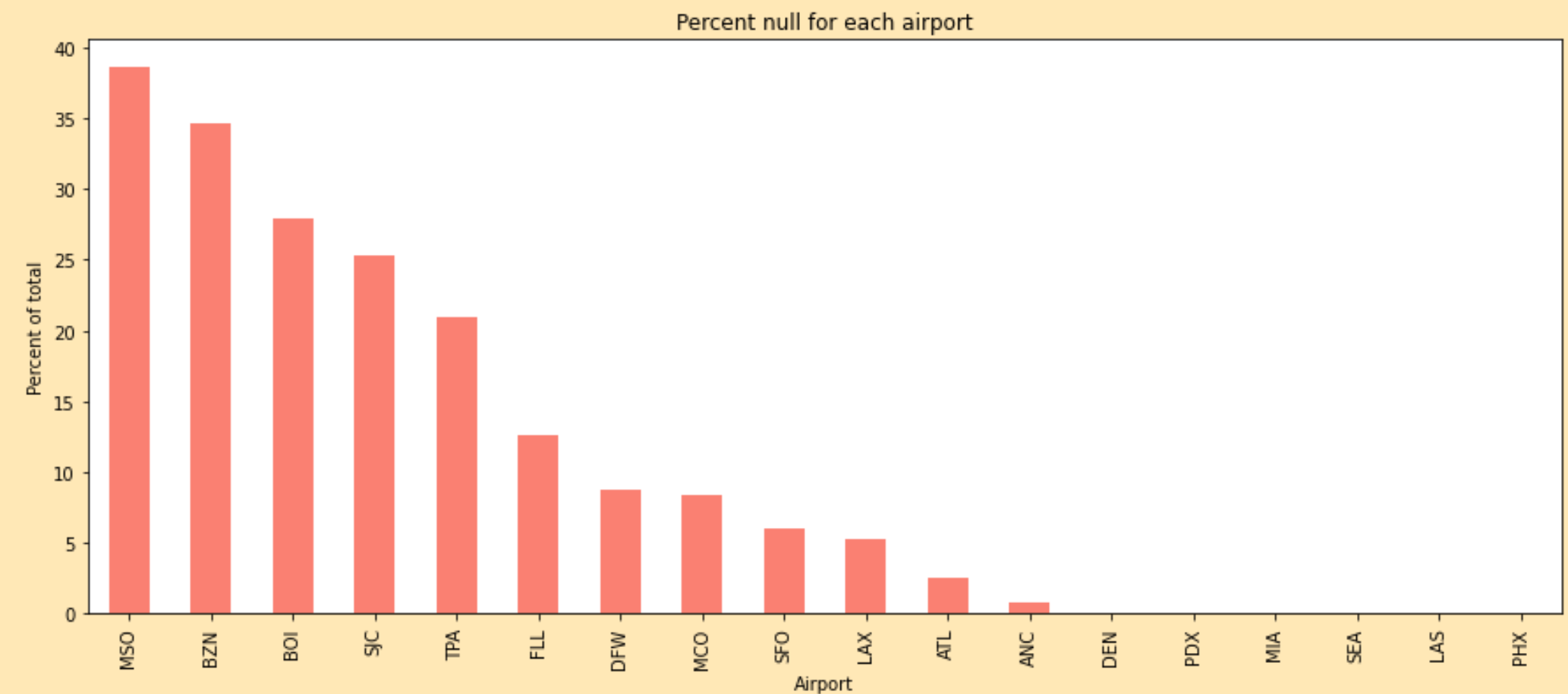
18 Columns

A I R P O R T S



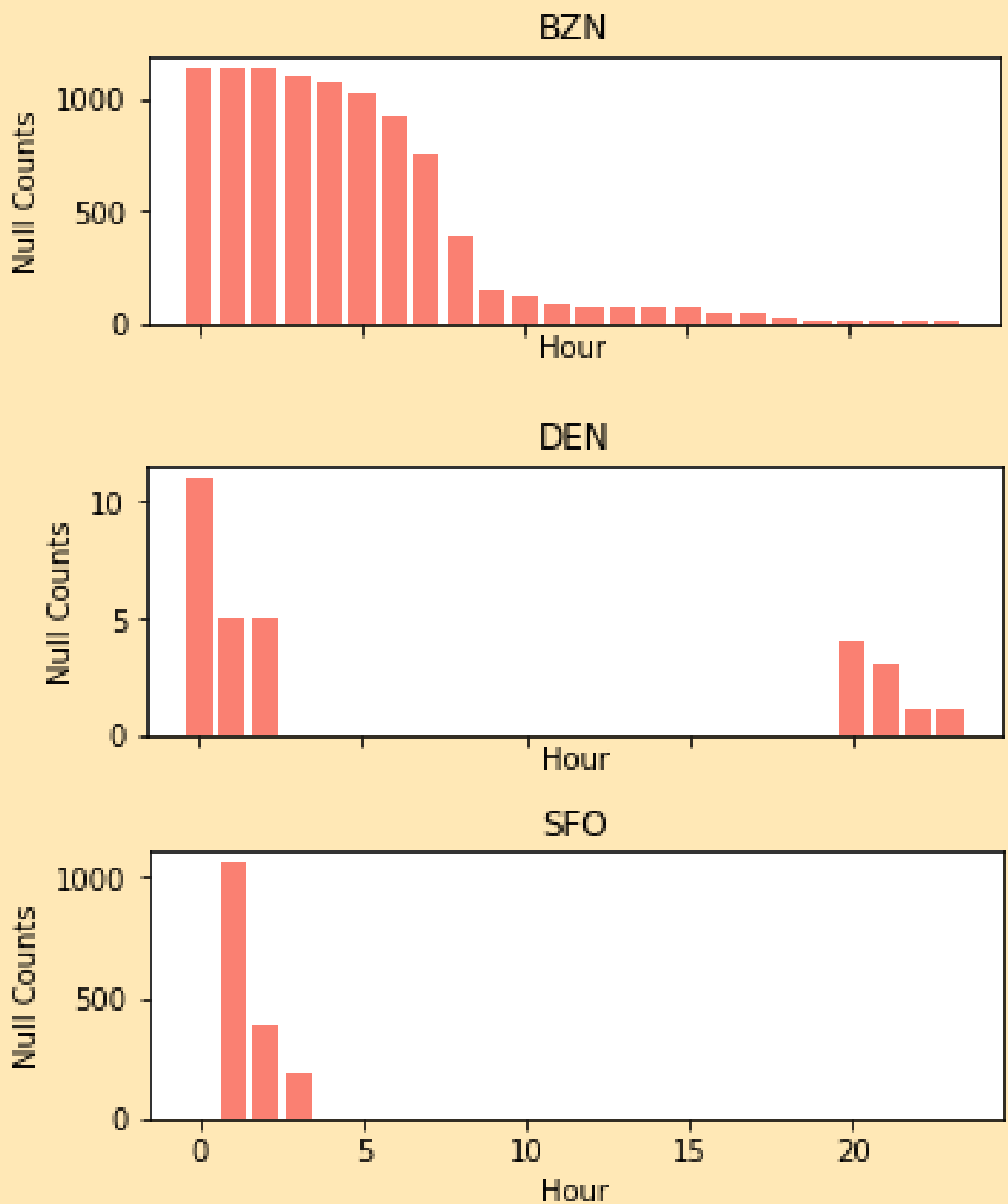
Null Values

Small Airports have
higher null values

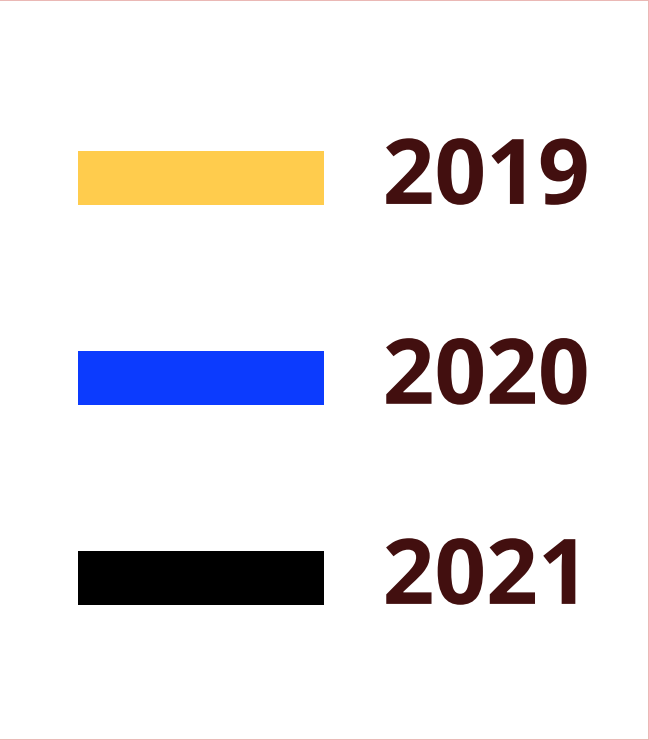
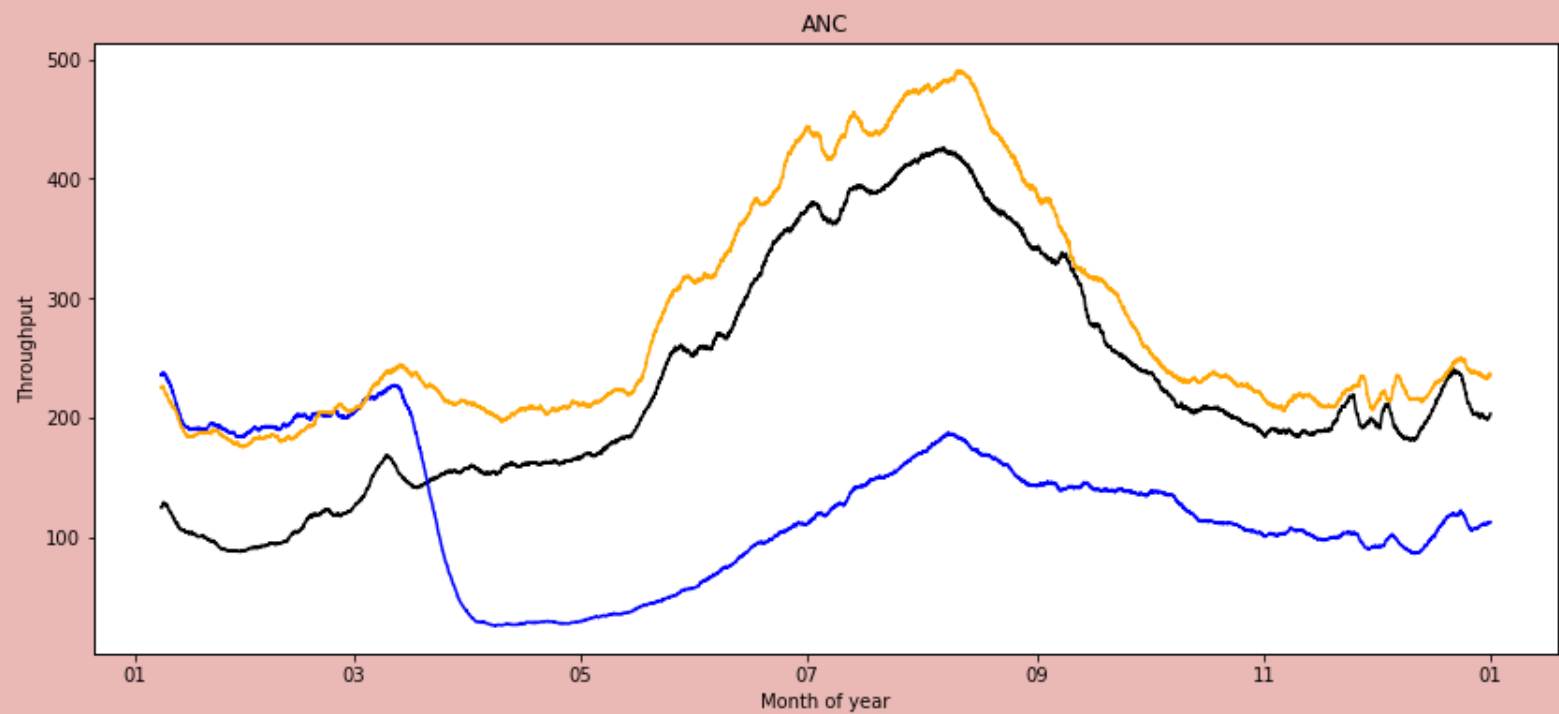
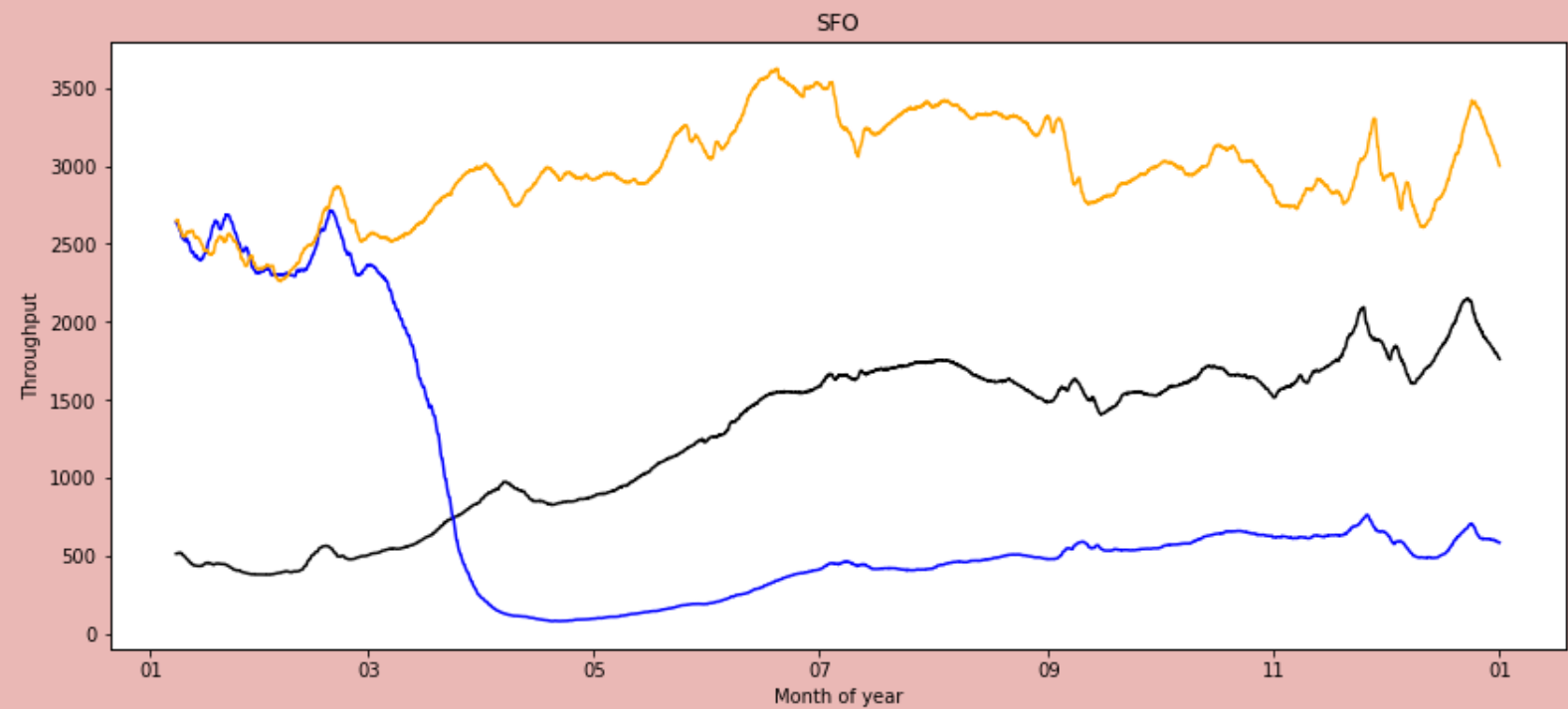


Impute null with ZERO

Fewer Flights in the
Middle of the night



Yearly Trends

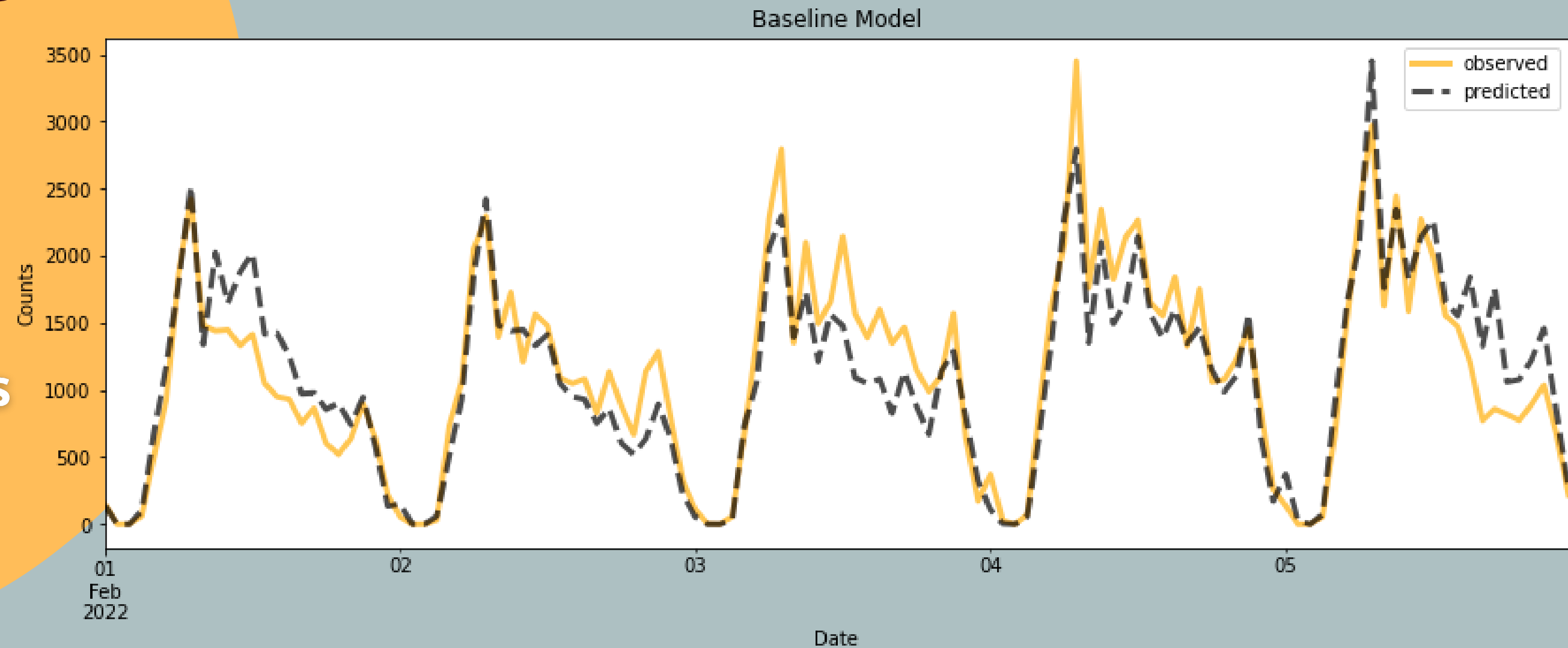


Baseline Model

276 passengers

Average Error

Uses yesterday's values



Seasonal ARIMA Model

- AR - Auto-regressive: q
- MA - Moving Average: p
- I - Differencing: d
- Seasonal parameters: P, D, Q, S

$(p, d, q) \times (P, D, Q, S)$

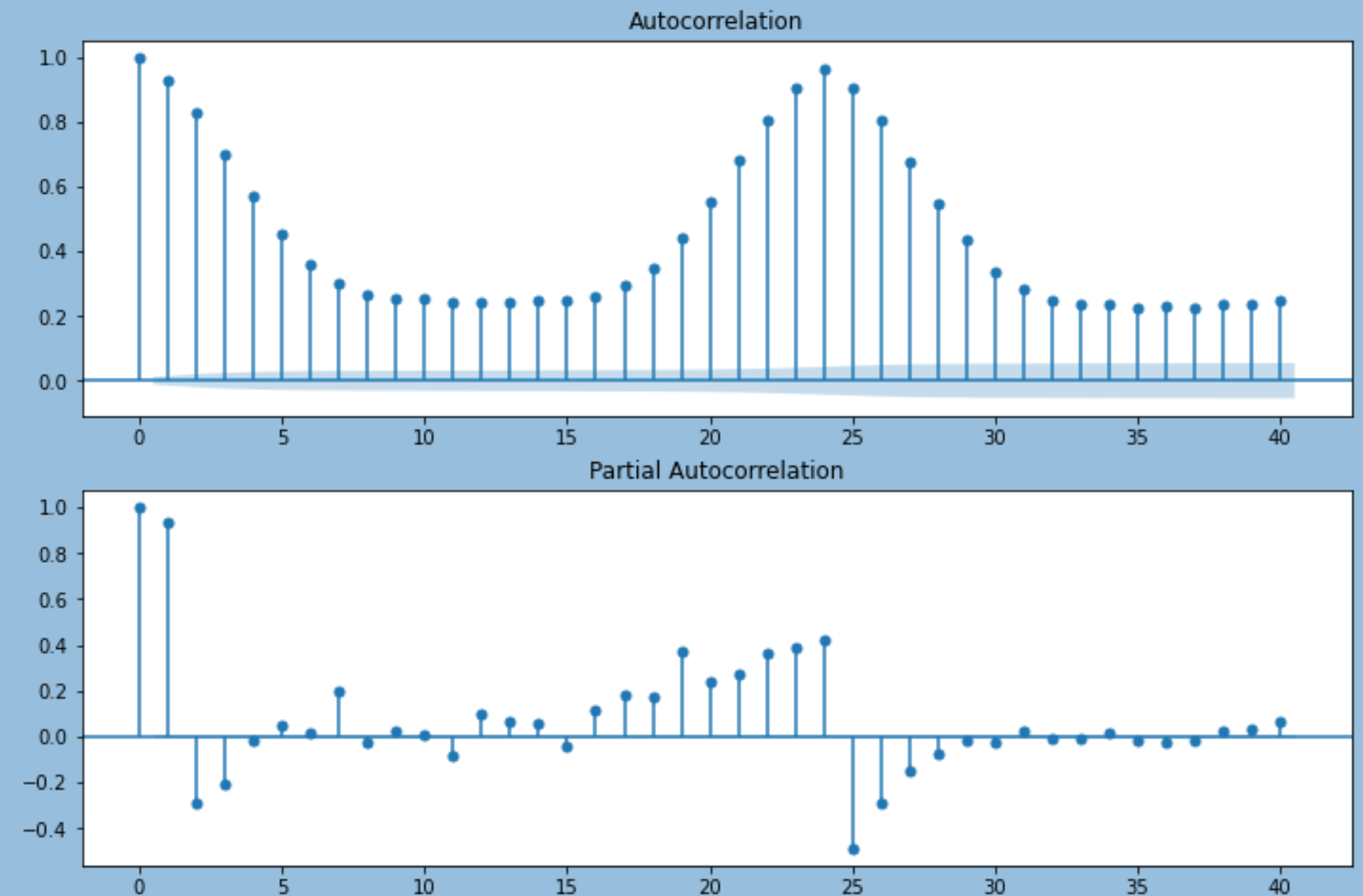
Seasonality/Stationarity

Seasonal:

- Period (S) = 24
- D = 1 (max values)

Non-stationary:

- $d+D \leq 2$
- $d = 0$ or 1



Best Parameters

- 1 Seasonality: $D=1, S=24$
- 2 Stationarity: $d = 0, 1$
- 3 Autoregressive: $p, q = 0, 1, 2$
- 4 Moving Average: $P, Q = 0, 1, 2, 3$

**GRID
SEARCH:
 p, d, q, P, Q**

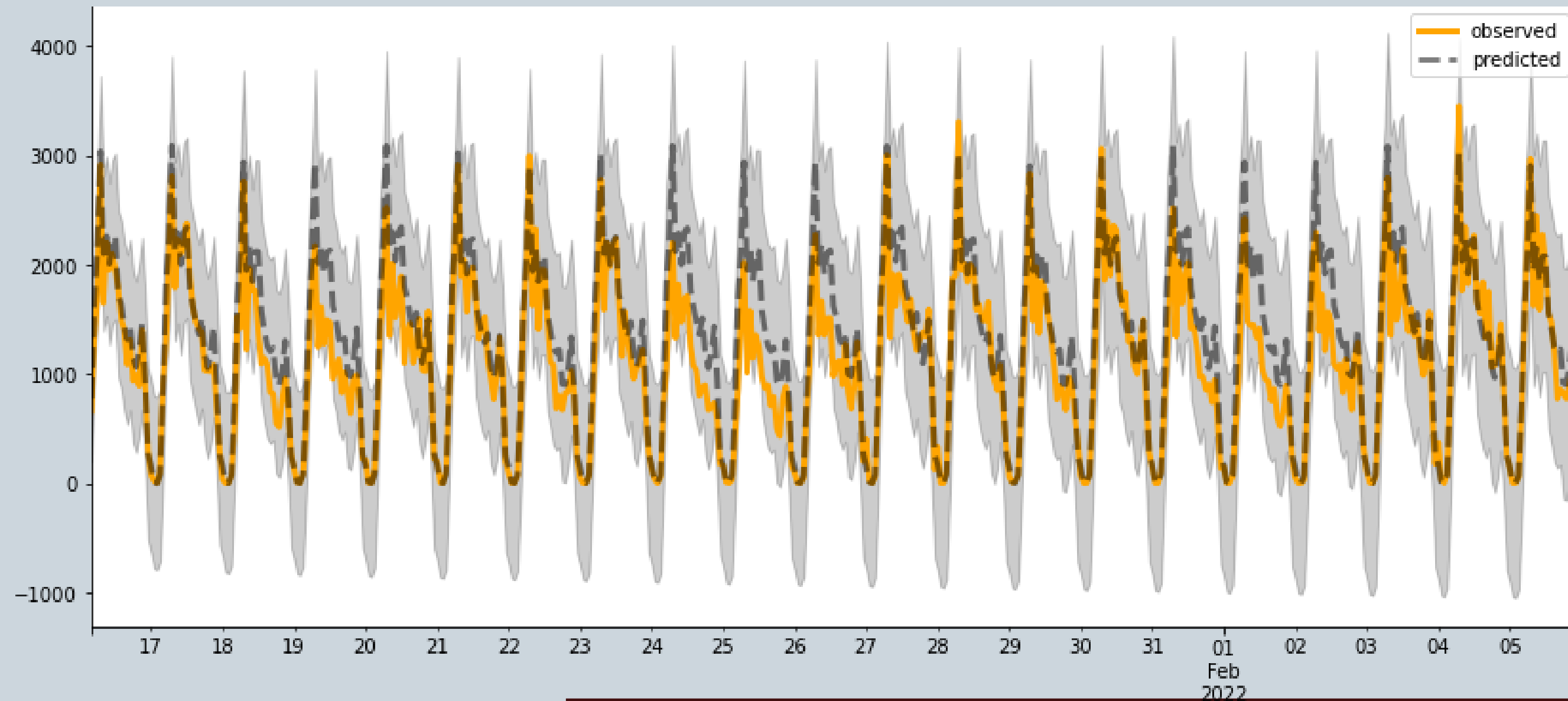
Training

- Train/Test split: test against last 500 values
- Cross-validation for best parameters
- Metric - Akaike Information Criteria (AIC) - Best predictive model

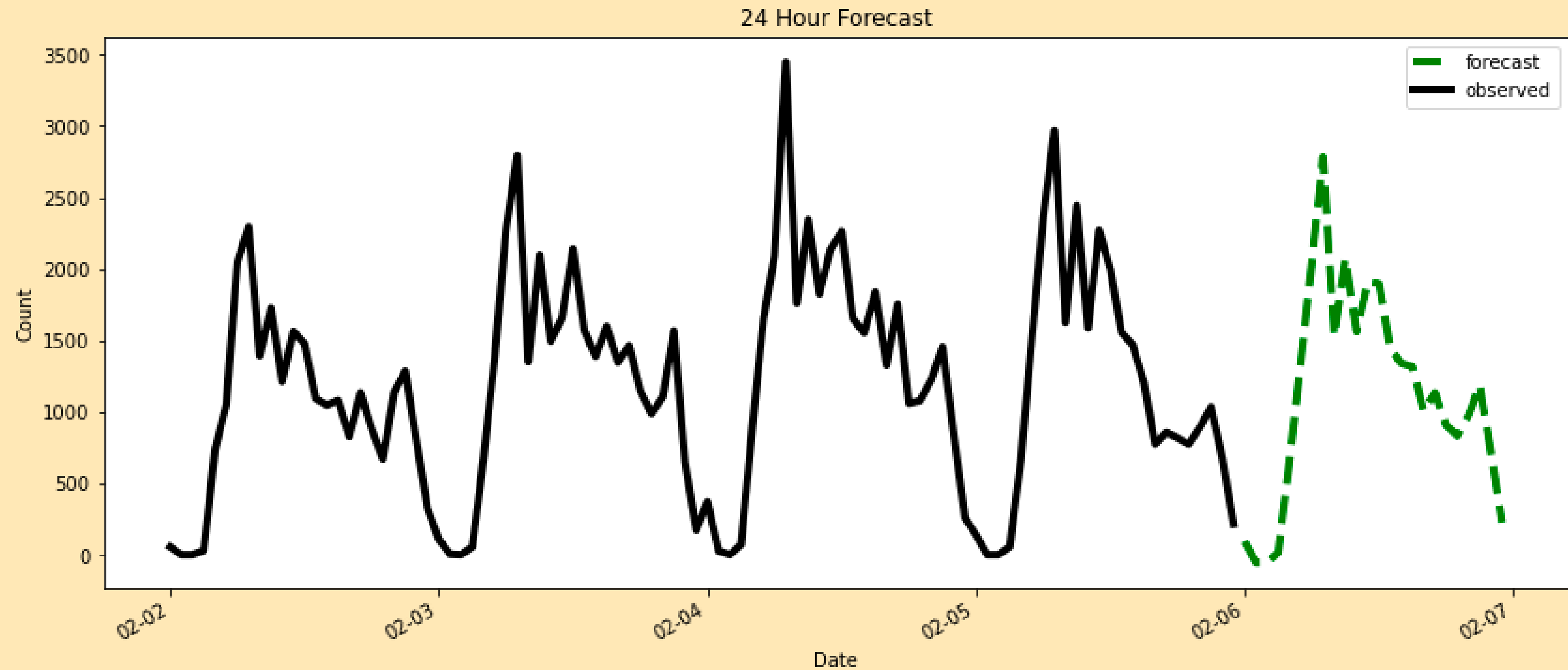
(1, 0, 1) X (3, 1, 3, 24)

Prediction

Test MAE - 227 people



Forecast



Next Steps

Perform time series clustering of all airports

Automate data collection from source repository as it is updated
