CAPSTONE PROJECT - SHINE ROSHAN

TSA Throughput Prediction

2022

YOUR MARKETING COMPANY OR A TOPIC

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

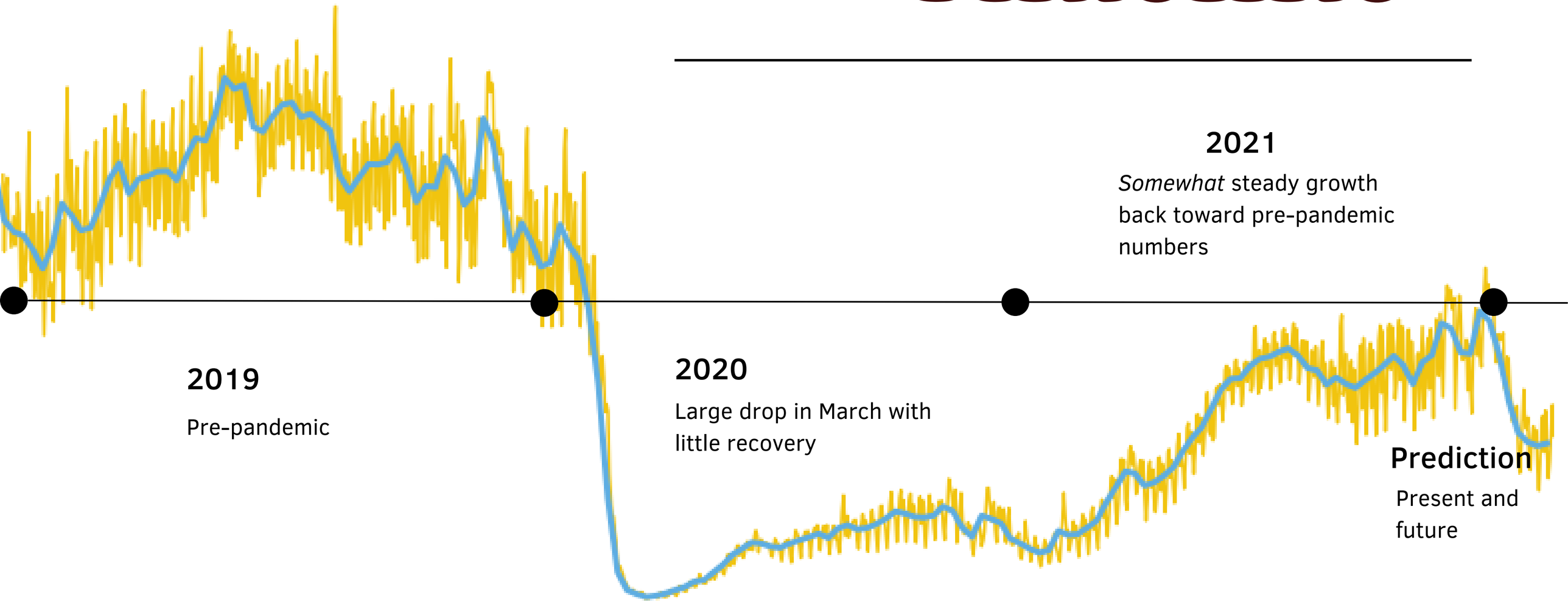
Efficiency

Proper staff and resource allocations

Customer Service

Better customer service will create loyal customers

Timeline



2019

Pre-pandemic

2020

Large drop in March with little recovery

2021

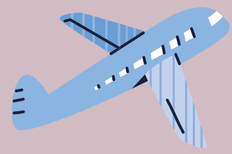
Somewhat steady growth back toward pre-pandemic numbers

Prediction

Present and future

Stakeholders

Hotels nearby



Airlines

Airport businesses

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

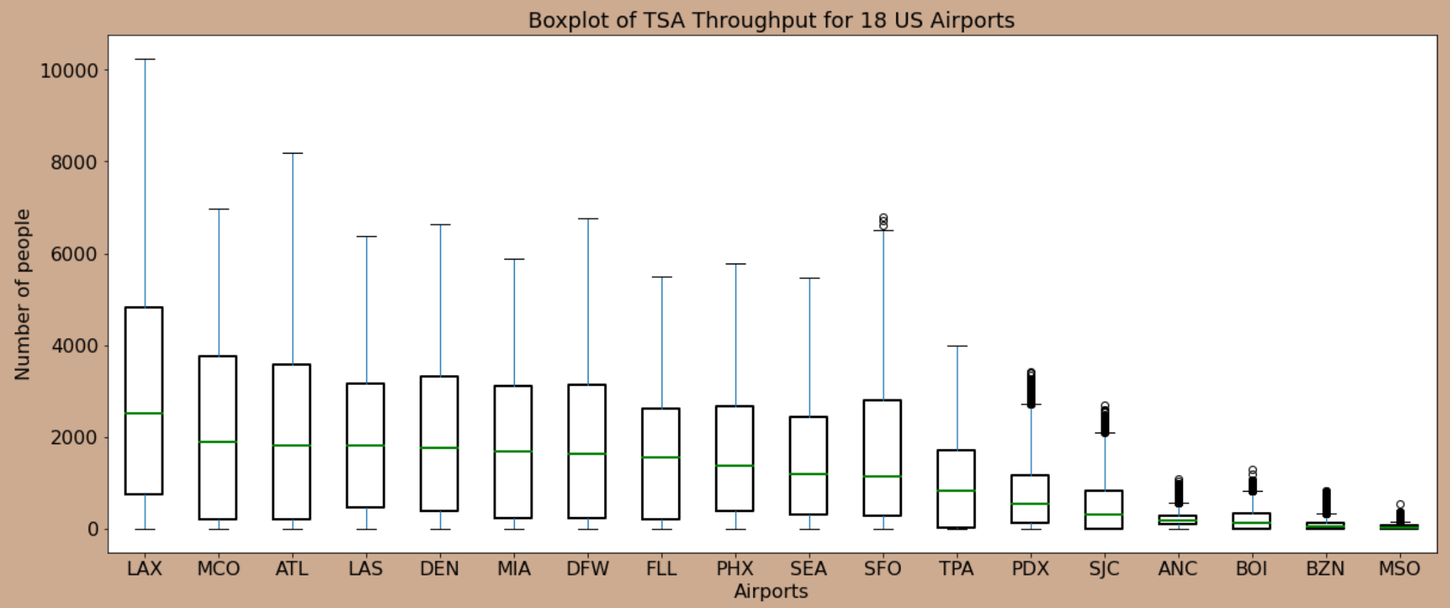
HOURLY DATA

Over 3 years

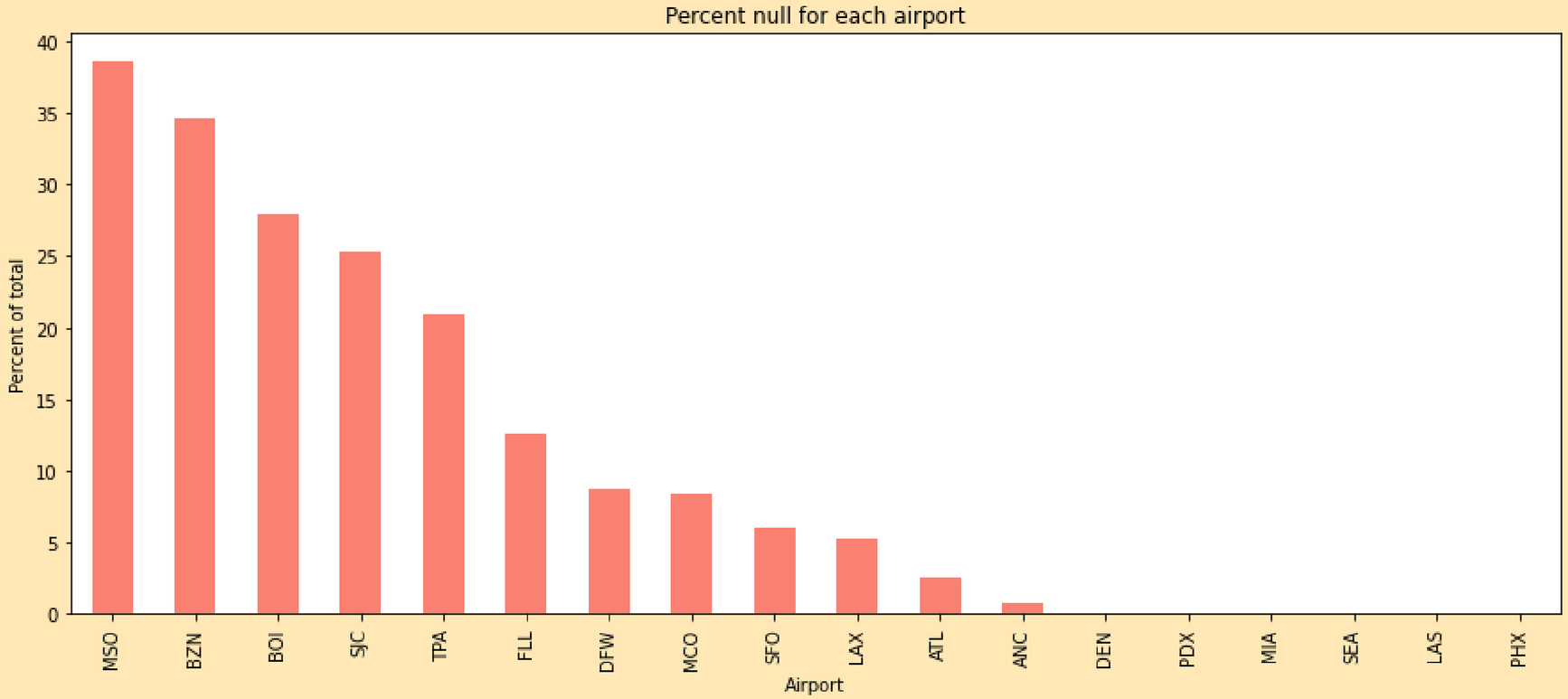
Enough seasons for ARIMA model

18 Columns

AIRPORTS



Null Values



Small Airports

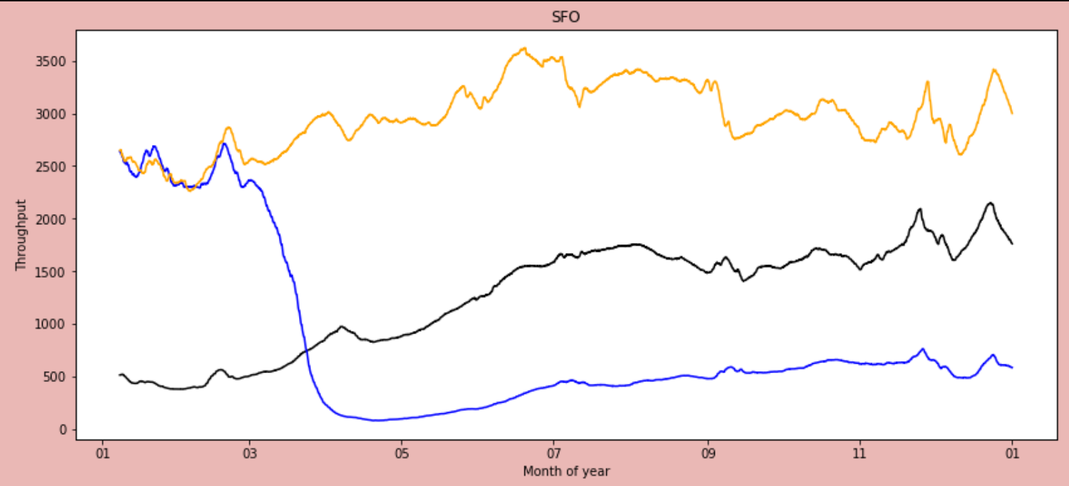
Higher null values

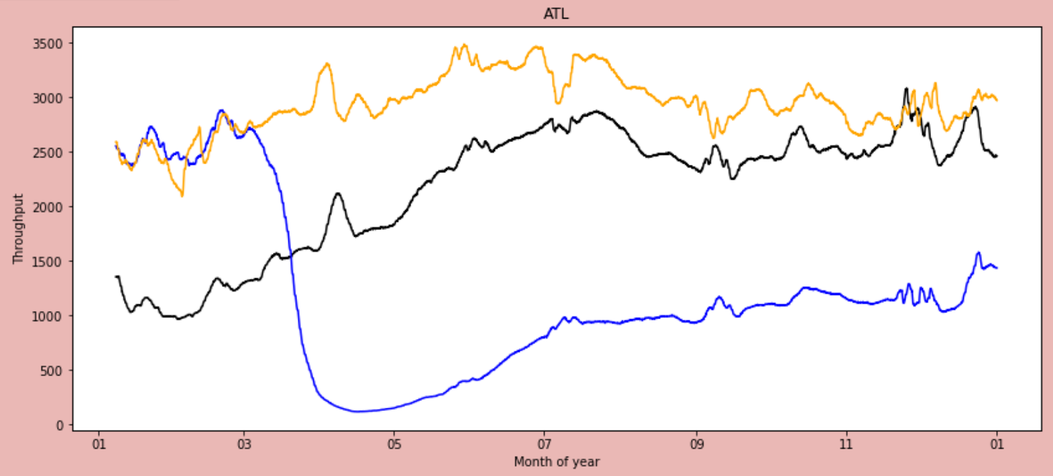
Fewer flights

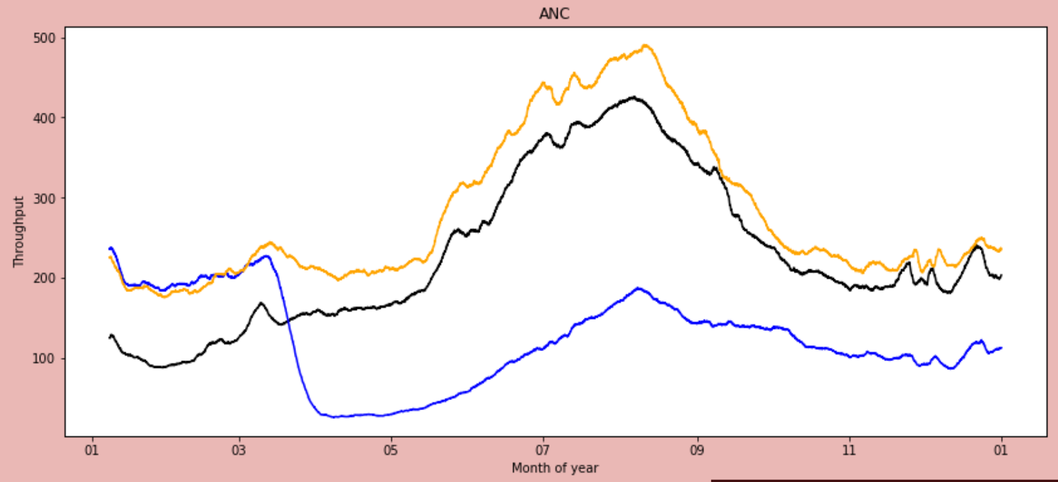
Middle of the night

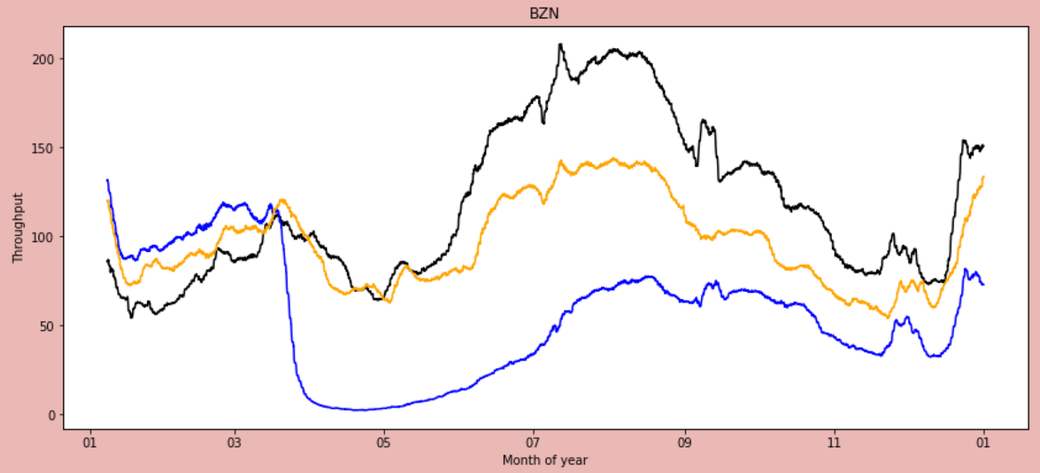
Impute null with ZERO

Yearly Trends







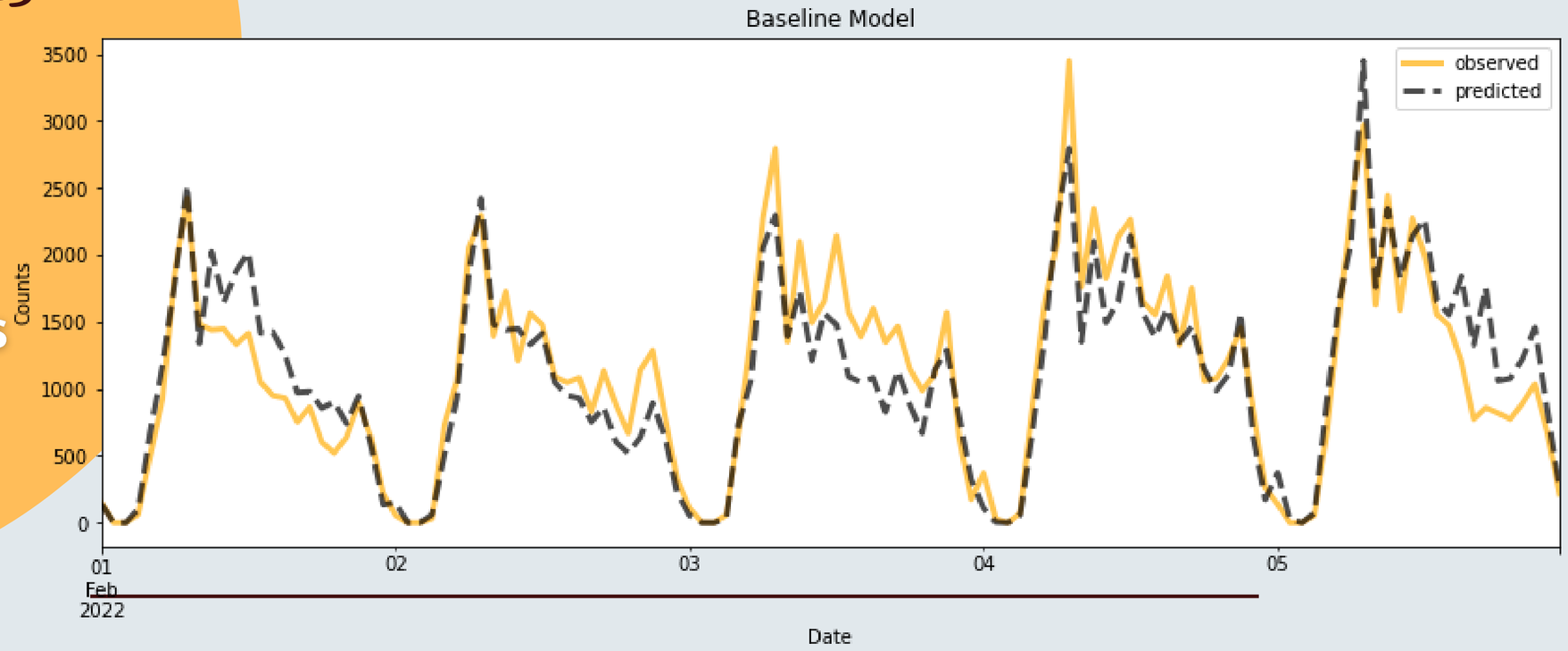


Baseline Model

276 passengers

Average Error

Uses yesterday's values



Model Selection

1. **Seasonality**: D = 1, Perdiod =24
2. **Stationarity**: d = 0, 1 (not stationary)
3. **Autoregressive:** p, P = 0, 1, 2
4. **Moving Average:** q, Q = 0, 1, 2

GRID SEARCH

Training

Train/Test Split

Test - Last 600 observations

Predictive Model

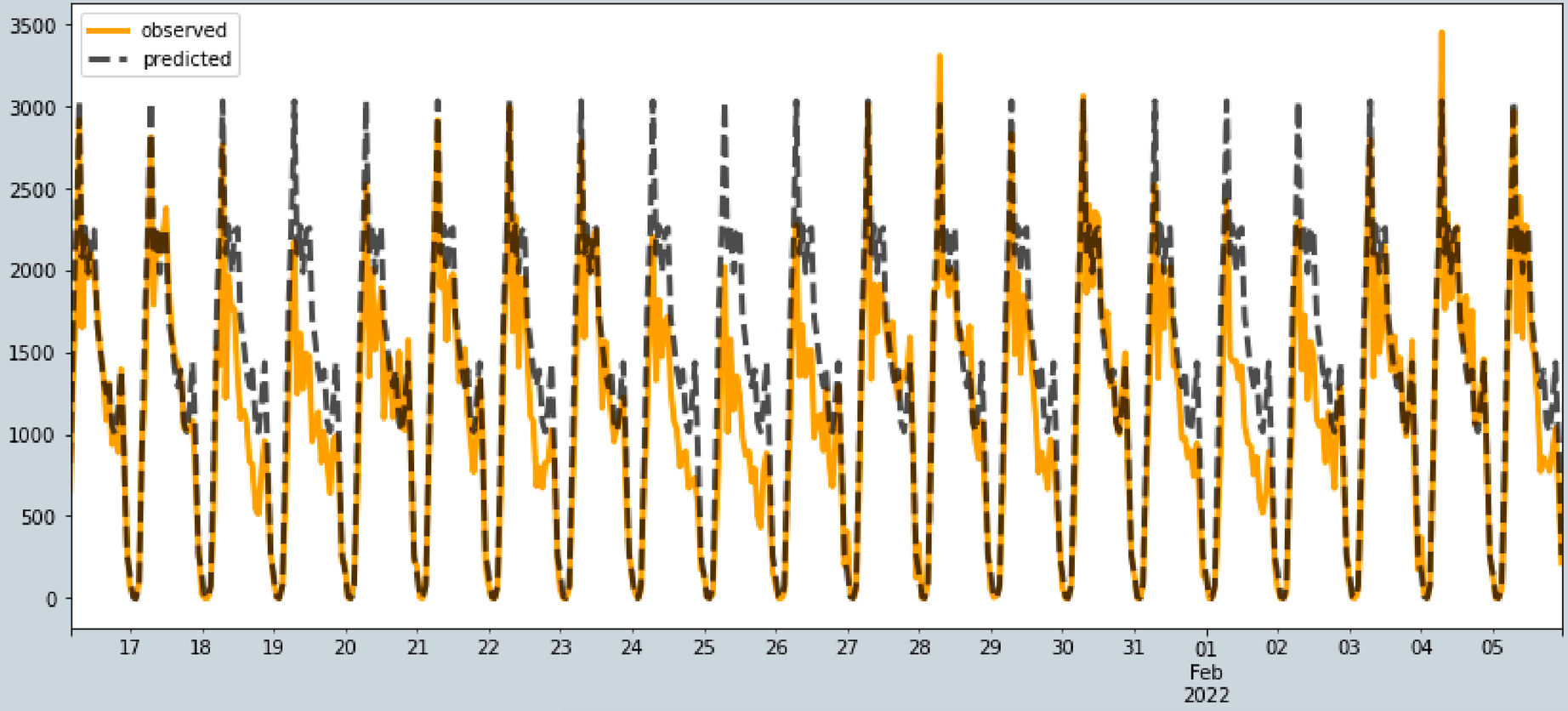
Akaike Information Criteria

Cross-validated AIC

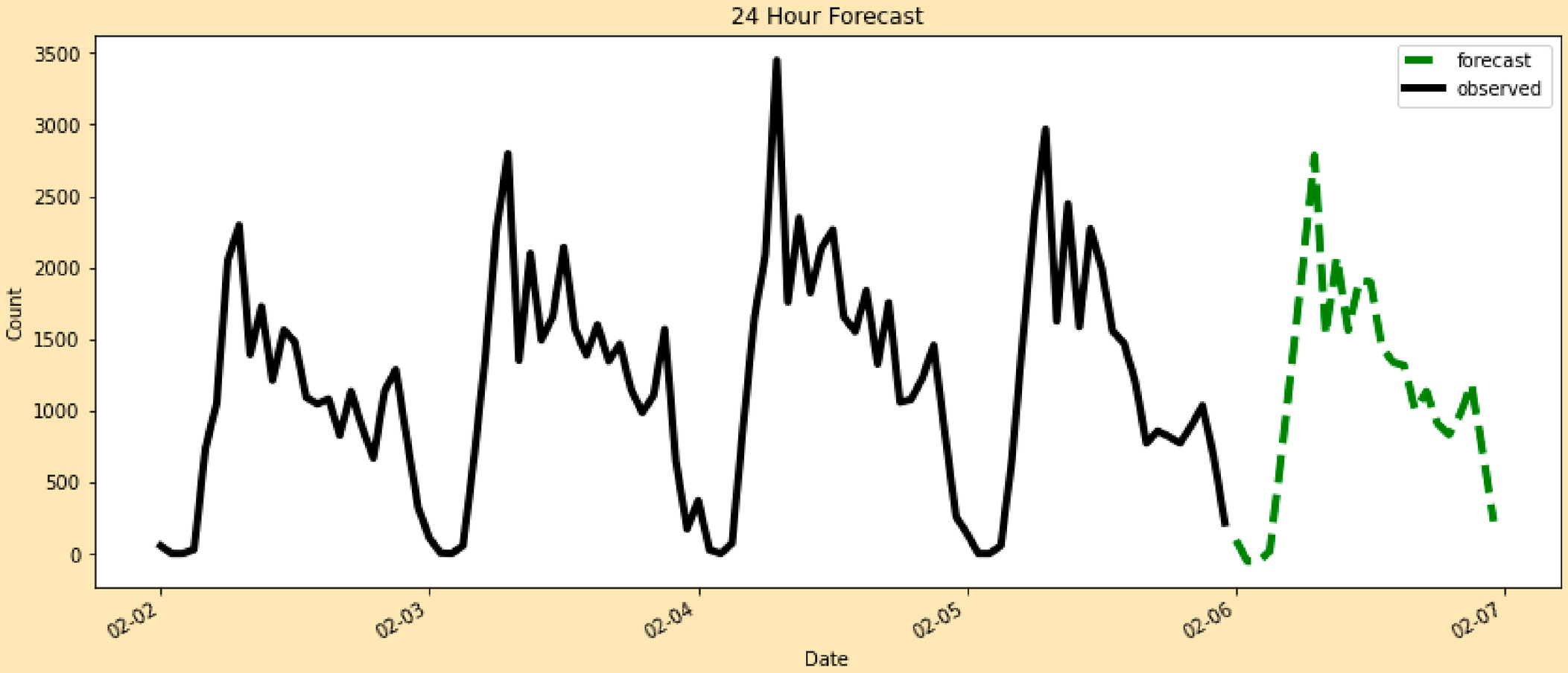
132,854

Prediction

Test MAE - 264 people



Forecast


24 Hour Forecast
forecast
observed
Count
3500
3000
2500
2000
1500
1000
500
0
02-02
02-03
02-04
02-05
02-06
02-07
Date

Next Steps

Automate data collection from source repository as it is updated

Perform time series clustering of all airports