The Impact of Covid-19 on Air Traffic:

Spatiotemporal/Time Series Forecasting and Benchmarking

Bo Yan, Yuan Hu, & Adelle Driker
(Group 6)

Team Introductions

Professors Rose Yu & Ilkay Altintas de Callafon

- Advisors/Mentors

Bo Yan

- Record Keeper
- Software/ML/DL Engineer

Yuan Hu

- Budget Manager
- Data Engineer/Solution Architect

Adelle Driker

- Project Coordinator/Manager
- Data/Business Analyst

Recap - Problem Definition

Many global industries have been affected by the COVID 19 pandemic, the airline industry being one of the most heavily hit

- E.g. London's Heathrow Airport reported a 97% decrease in passenger numbers between May 2019 and May 2020

Creates uncertainty for both passengers and airline companies, especially due to the multiple waves of virus mutations

- How should airlines plan future flights? When should passengers schedule their travels?

In other words, given a country's COVID situation, how should an airline/passengers plan ahead?

Recap - Data

OpenSky Flight Data (Jan 2019 - Present)

- As-is, new files released on a monthly basis, multiple entries with missing data
- CallSign*, Number, ICAO24, Reg, TypeCode, Origin, Dest, First/Last Seen DT, Lat/Long/Alt of Origin & Dest

Johns Hopkins COVID19 Data (Jan 2020 - Present)

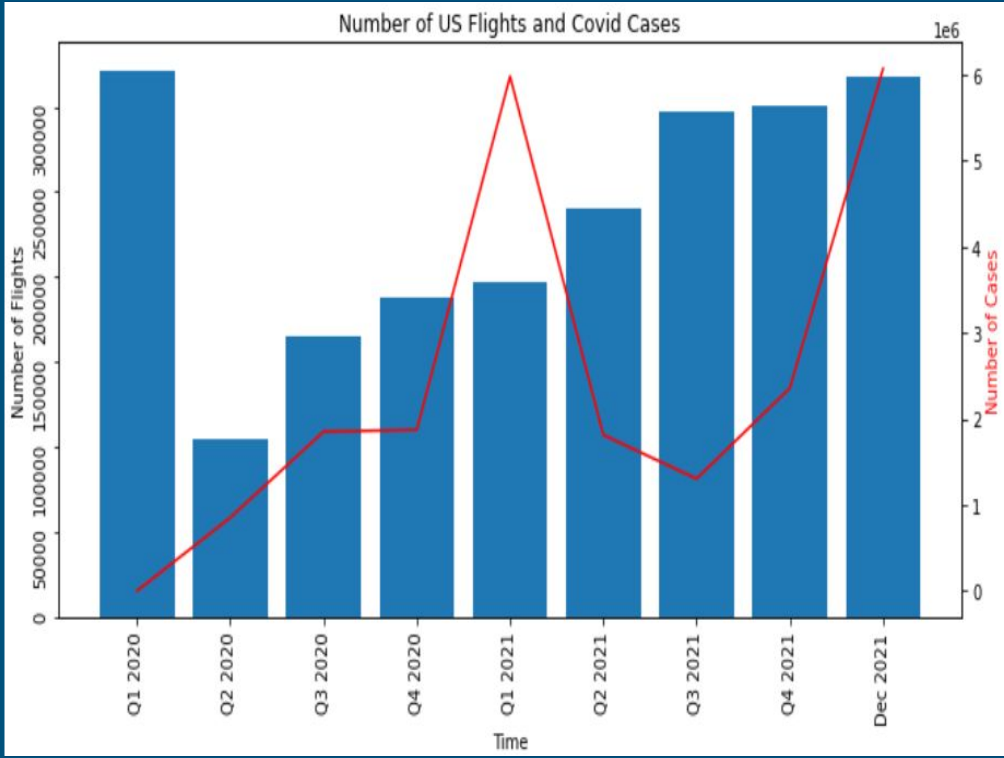
- Updates for historical inaccuracies, new files released daily
- Province/State, Country/Region**, Lat, Long, Dates

Airline Code and Country Mapping

- Sourced from IATA and ICAO, mostly complete
- Airline Name, IATA Designator, 3-Digit Code, ICAO Designator*, Country**
- Will be used to link together the OpenSky Flight and COVID19 datasets

Main Insights - Combining Flight and Covid Data

- Drastic dip in Flights in Q2 2020 signifying beginning of pandemic
- Large Covid surge in Q1 2021
- First decline in early 2021 as vaccines are introduced
- Another increase in cases in mid-late 2021 indicates Delta, then Omicron waves
- While Covid data fluctuates, Flight data shows a steady recovery

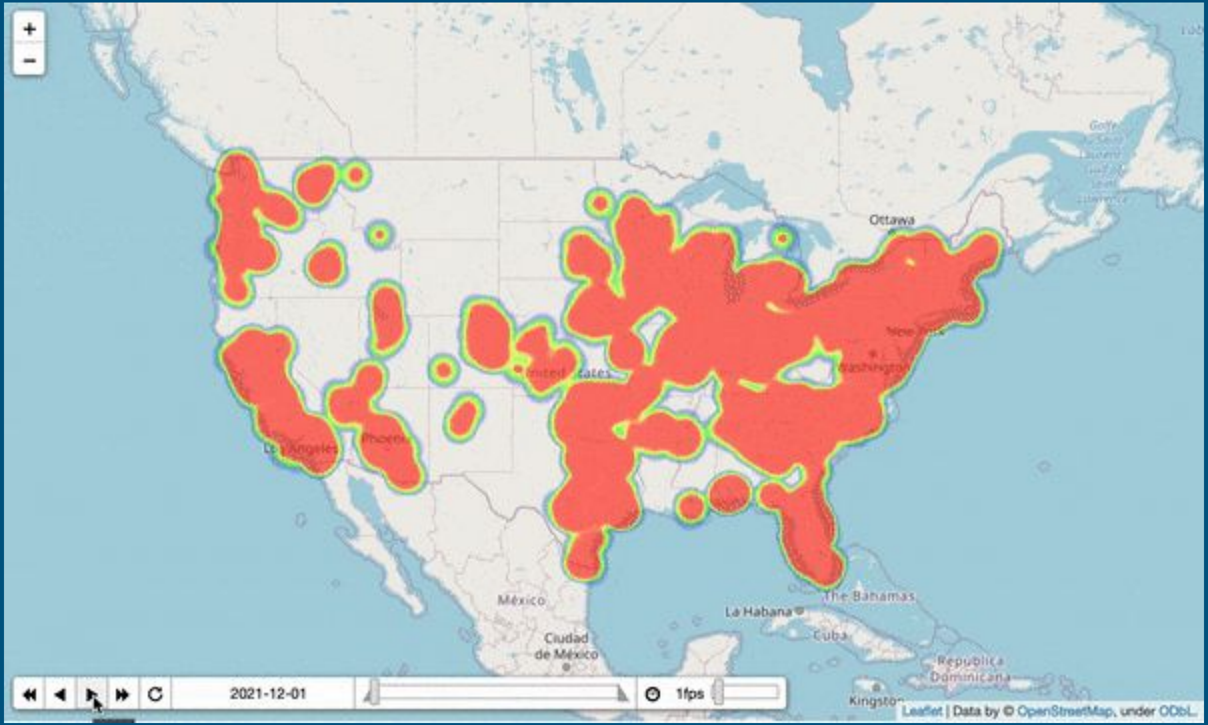


Main Insights - Monitoring Historical Departing Flights US

Scenario:

- # Departing Flights @US Airports

  @Dec 2021



Main Insights - Visualizing Global Flight Trajectories

Scenario:

- Node-Link network:

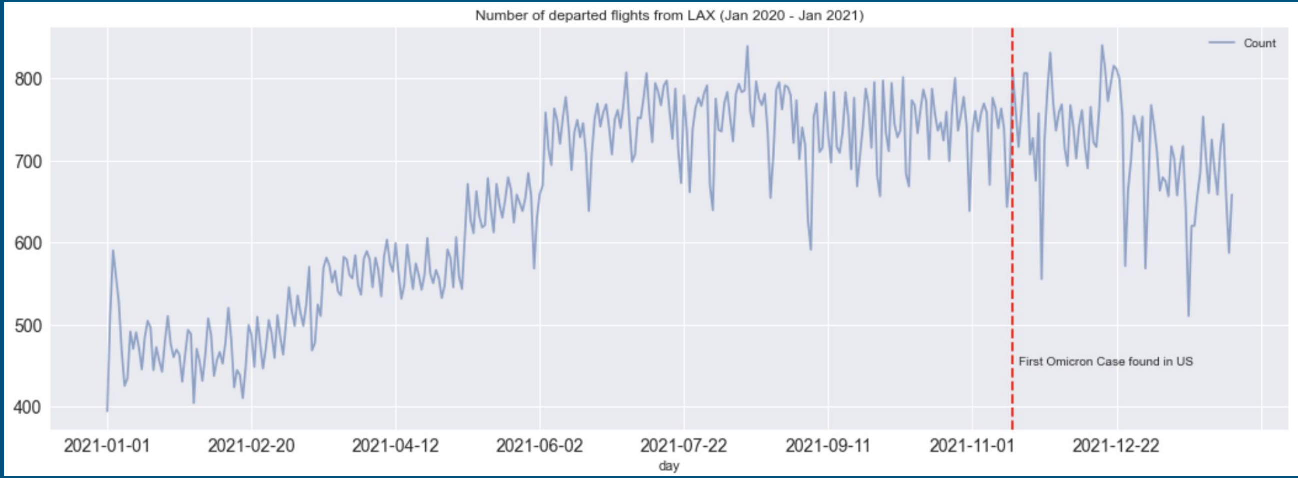
  Global flight trajectory
  @Dec 26, 2021

Accuracy/Significance of Results

Scenario:

- # Departing Flights @LAX(Los Angeles)

  Jan 2021 - Jan 2022



Steps:

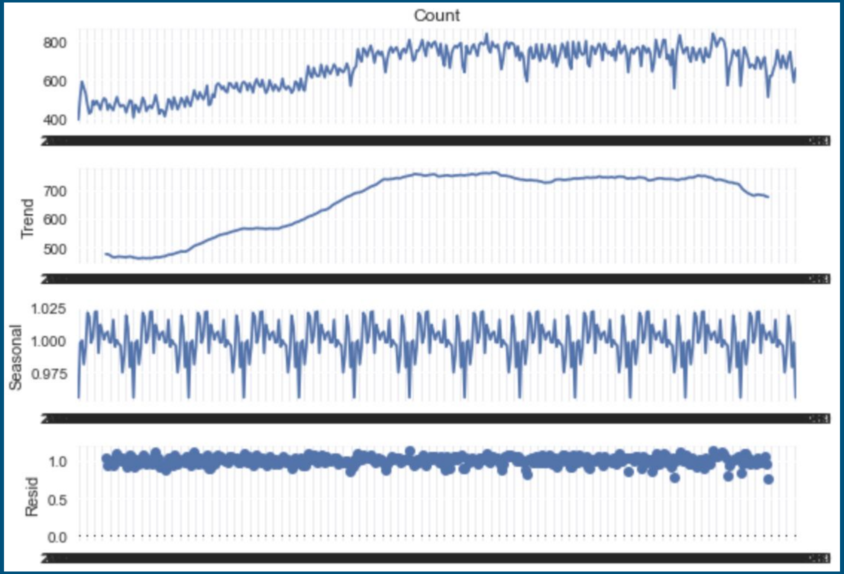
- Time-series decomposition
- Residuals' normality check for t-test assumption
- Hypothesis Testing: whether the outbreak of Omicron affected #flights @LAX

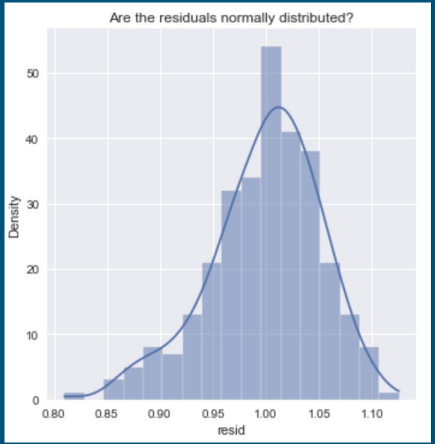
Hypothesis Testing

- Null: Outbreak of Omicron doesn't affect the number of departing flights from LAX.
- H1: Outbreak of Omicron affect the number of departing flights from LAX.

```
Ttest_indResult(statistic=0.0355954955174674, pvalue=0.971682513102905)
```

- Failed to reject null hypothesis
- Conclusion: Omicron doesn’t significantly affect the number of departing flights @LAX





Preliminary Modeling Results

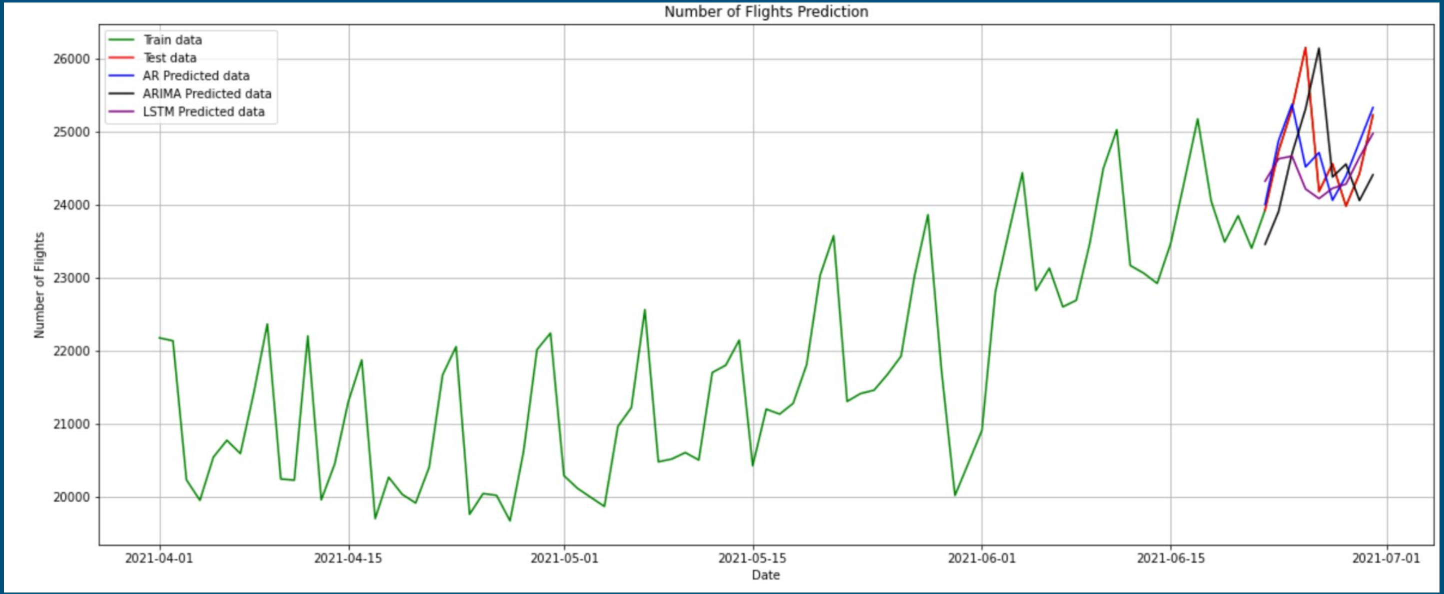
- Sample:

  1 Quarter
- Data:

  Flight Datasets
- Baseline Methods:

  AR and ARIMA
- Deep Learning Methods:

  LSTM



Main Insights

- Sample: 1 Quarter
- The trend shows the growing number of flights with time.
- The seasonality information shows a biweekly cycle.
- The residuals shows periods of high variability in around 7 days of the series.

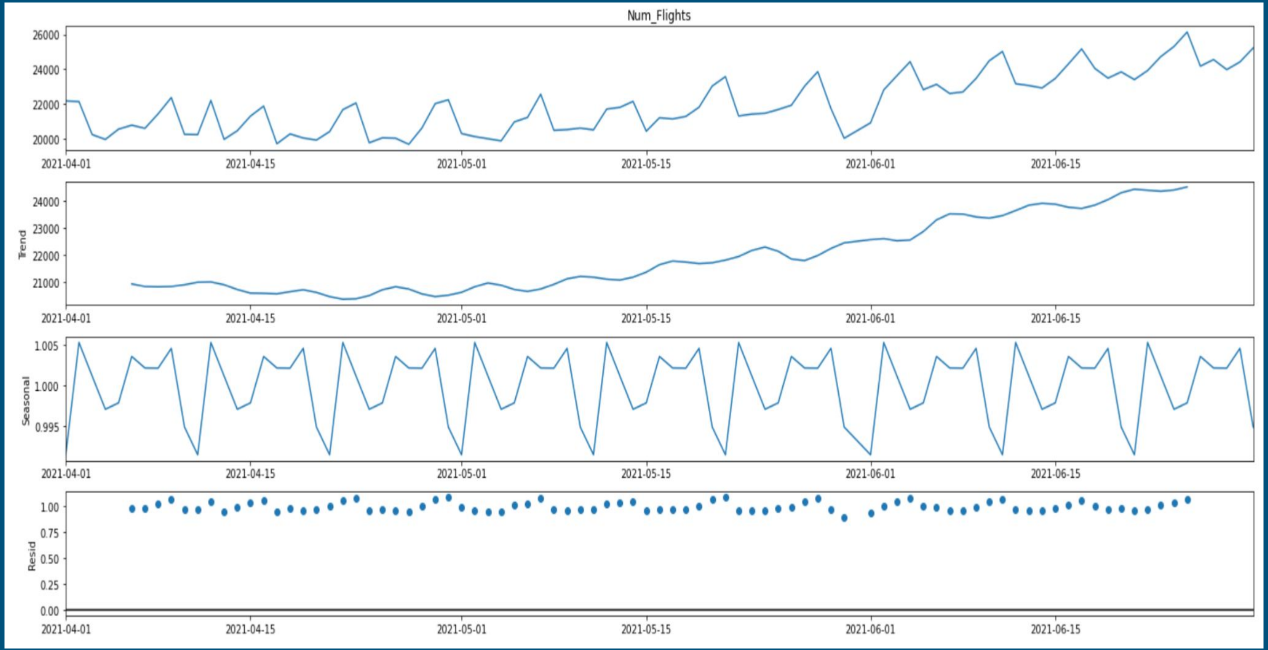


Fig. Decomposed Dataset

Main Insights

- Sample: 1 Quarter
- Autocorrelation: Gradual decrease
- Partial Autocorrelation: Sharp cut-off
- Guides us to find the optimal parameters for our models

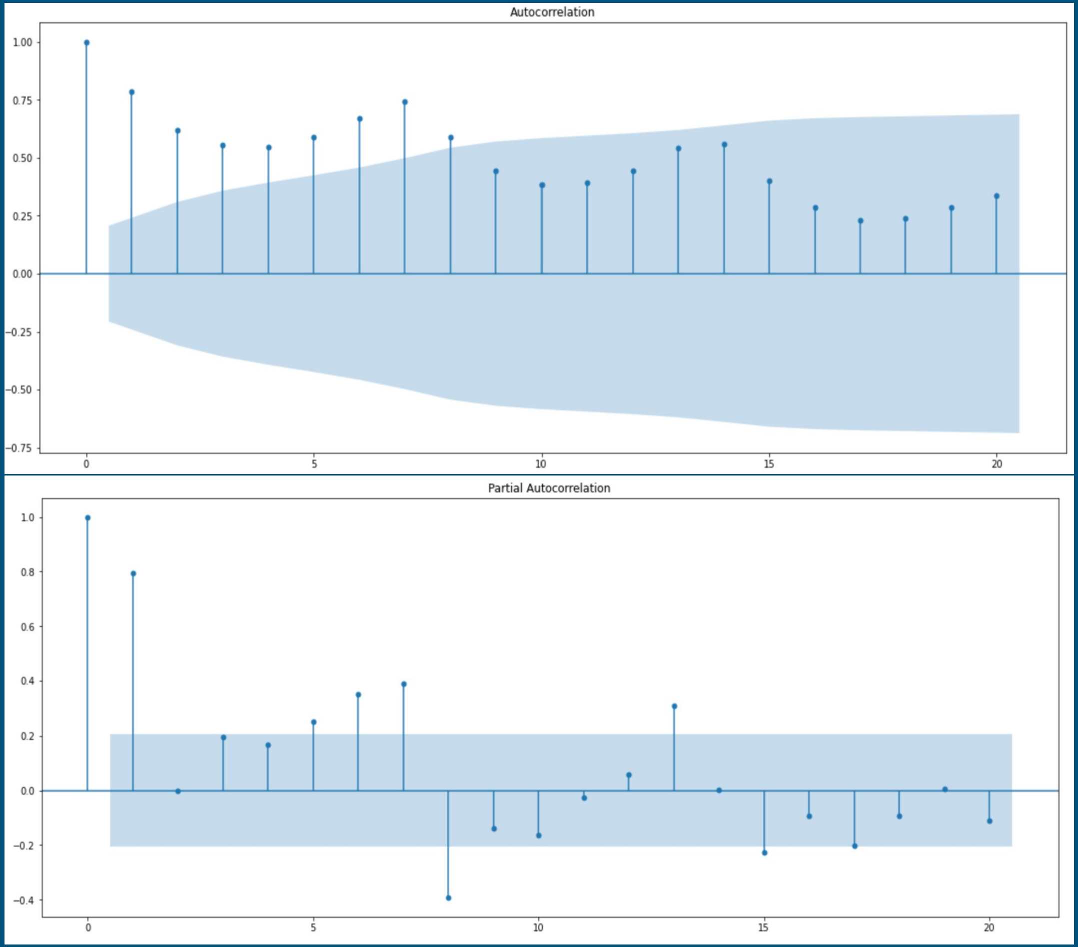


Fig. Autocorrelation and Partial Autocorrelation

Main Insights

- Sample: 1 Quarter
- Using AR model and ARIMA model as the baseline models can reflect the trend information
- Trend component grows the flight number month by month
- Seasonal component has a cycle less than 2 weeks
- The variance in the data keeps on increasing with time

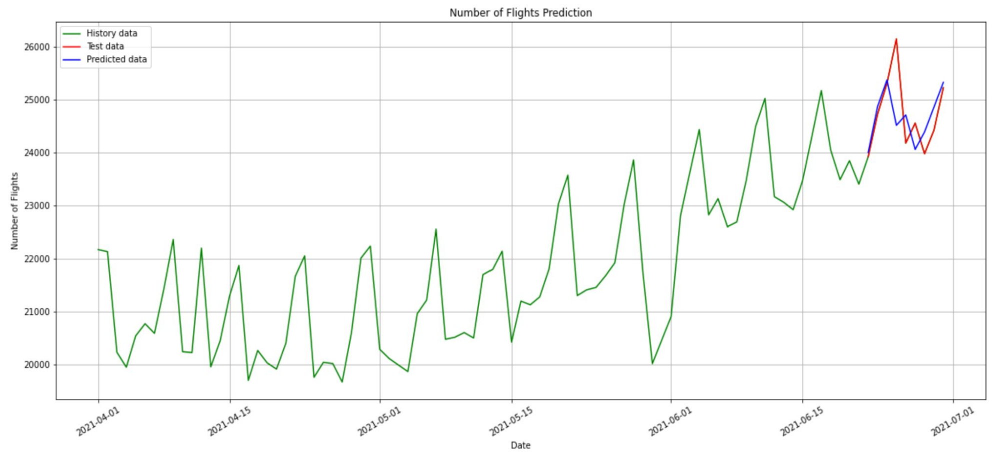


Fig. AR model prediction results

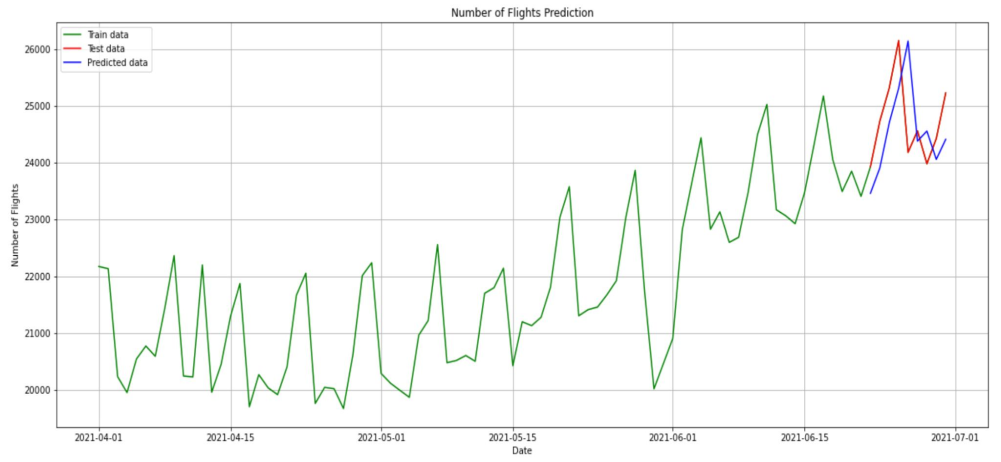


Fig. ARIMA model prediction results

Main Insights

- Sample: 1 Quarter
- Using the LSTM model to train the data can reflect the trend information
- Need further performance tuning to improve performance

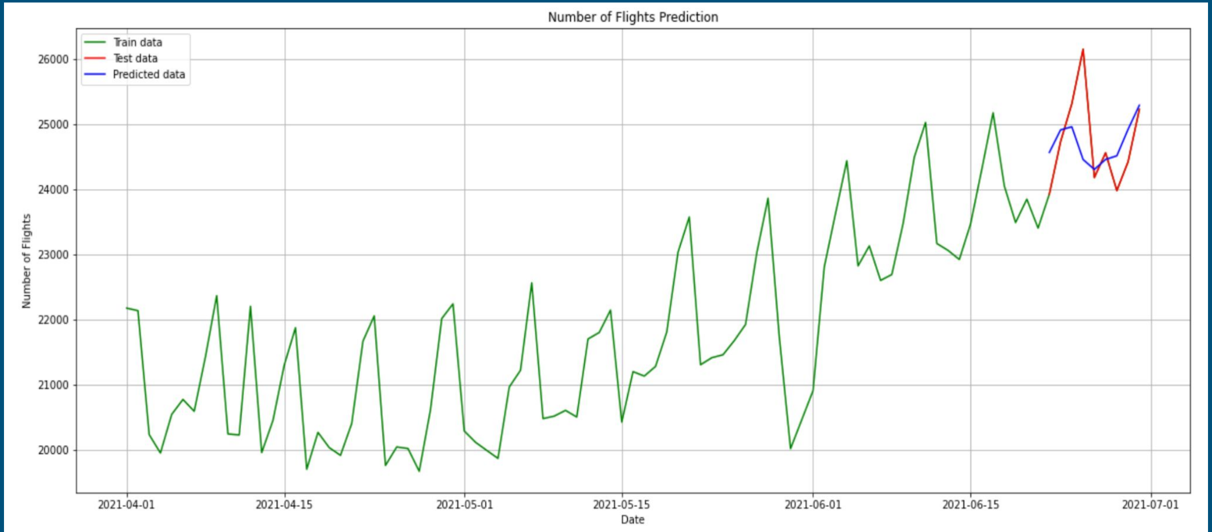


Fig. LSTM model prediction results

Recommendations

Given the preliminary insights from EDA and Modeling, we recommend that the following actions are taken:

- Test the model on a larger portion of the dataset
  - Get a better idea of performance
- Conduct further hypothesis testing on the impact of the Delta variant, since the effects on the numbers of flights are not very apparent with Omicron
- Create a set of network graphs to compare conditions on the same date but different years
- Update the Heat Map to illustrate effects of Covid throughout the entire pandemic

Effect on Data Pipeline & Architecture

- Ingest all the data across the entire covid pandemic period
- Some data processing for EDA requires additional steps
- Structure of data need be further finalized to fit deep learning modules such as TorchTS and other specified dataLoader class objects

Next Steps - Modeling

- More feature engineering
  - Used features: Timestamp and Number of Flights features
  - Plan to add Covid Number and Location related features
  - Plan to add seasonality information
- Training on large datasets
  - Used 3 months of data (total 25 months of data)
  - Plan to run our models against the whole dataset
- Performance tuning
  - LSTM prediction results aren’t as good as the baseline models results
  - Plan to tune performance further to improve it
- Other deep learning methods
  - Seq2Seq and DCRNN, to achieve better performance
- Benchmarks
  - Create benchmarks against different models

Thank You

Q&A