UC San Diego
JACOBS SCHOOL OF ENGINEERING

# The Impact of Covid-19 on Air Traffic: Spatiotemporal Forecasting with Deep Learning and Benchmarks

**Advisor:** Professor Rose Yu
**Team:** Adelle Driker, Bo Yan, Yuan Hu

## Introduction

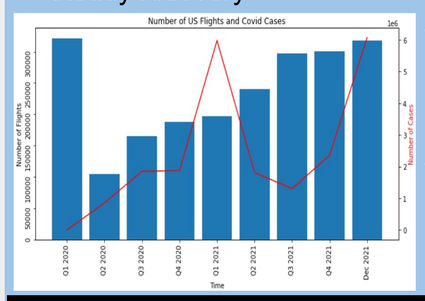
Many global industries have been affected by the COVID 19 pandemic, the airline industry being one of the most heavily hit. As a result, flight trends around the world have drastically deviated from their normal patterns. This presents an interesting scenario to explore, especially with the transience and abnormality of effects of the Covid data. The phenomenon also created uncertainty for both passengers and airline companies, especially due to the multiple waves of virus mutations prompting the following questions: 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? To solve this problem, we present a solution in the form of an air traffic forecast using baseline models: AR and ARIMA, and deep learning models: LSTM(Long Short Term Memory), LSTM-mis-regression, LSTM-quantile-regression, and other models that capture the temporal complexity of such data. These algorithms are developed for time series analysis that are available to the open-source community, and can better serve airlines and passengers.

## Data Sourcing and Storage

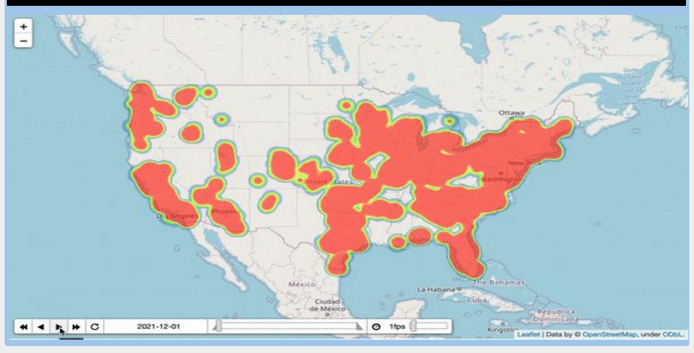
- **Flight Dataset**: Sourced online from OpenSky Flight network (Jan 2019 - Present).
- **Covid-19 Dataset**: Sourced online from Johns Hopkins (Jan 2020 - Present).
- **Airline Code and Country Mapping Dataset**: Sourced online from IATA and ICAO.
- **Data Storage**:
  - Amazon S3: Raw data
  - InfluxDB: Processed data and modeling results

## EDA and Findings

- Drastic dip in flights in Q2 2020 signifying the 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



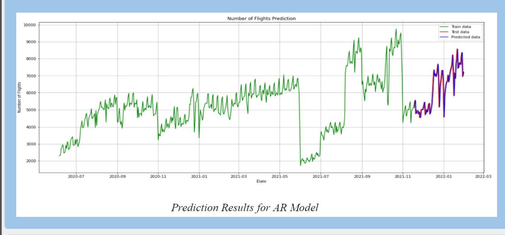

- The Node-Link network below shows the global flight trajectory
- The Geospatial map shows the US Historical Departing Flights



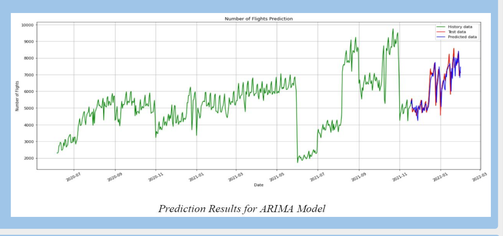

## Modeling Preparation

**Train test split**: Air Traffic data was ingested from 01-22-2020 to 01-31-2022, when the COVID-19 outbroke. The data was subdivided into train (90%) and test (10%). Various methods of infusing COVID-19 and Air Traffic datasets were also explored.

## Baseline Models Results

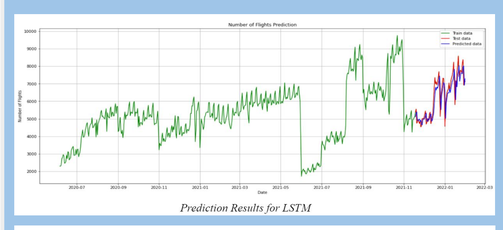



*Prediction Results for AR Model*

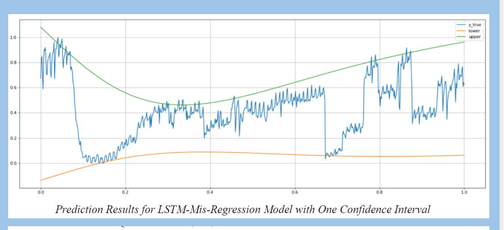



*Prediction Results for ARIMA Model*

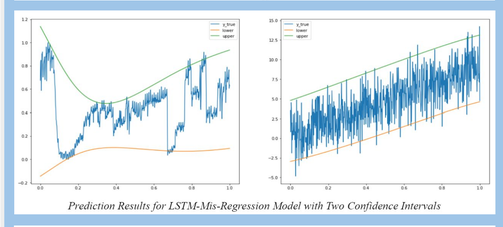
## Deep Learning Modeling Results



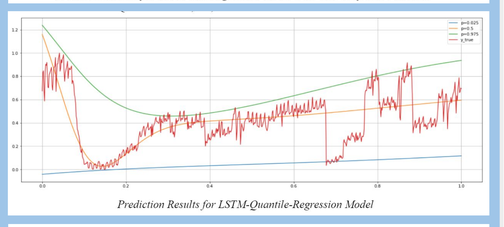

*Prediction Results for LSTM*



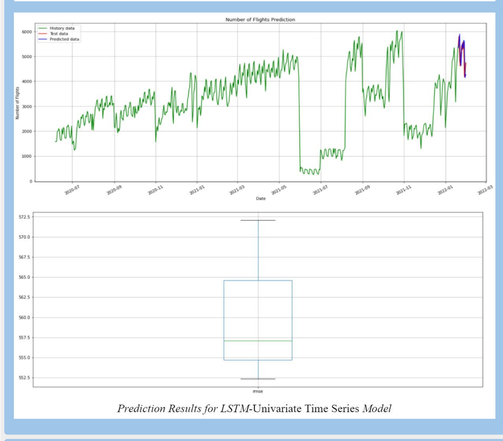

*Prediction Results for LSTM-Mis-Regression Model with One Confidence Interval*



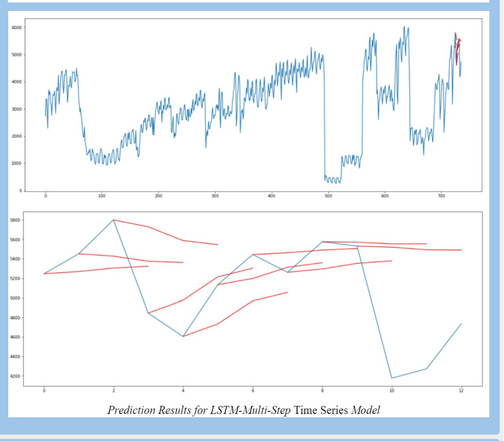

*Prediction Results for LSTM-Mis-Regression Model with Two Confidence Intervals*




*Prediction Results for LSTM-Quantile-Regression Model*

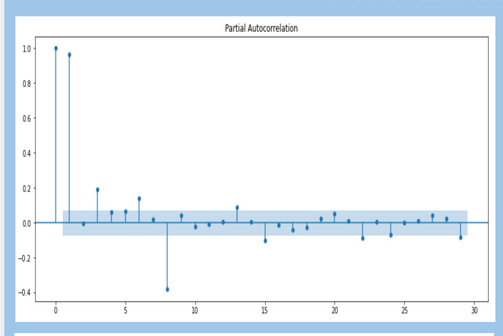


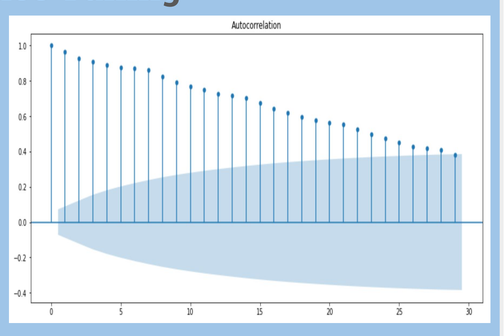

*Prediction Results for LSTM-Univariate Time Series Model*

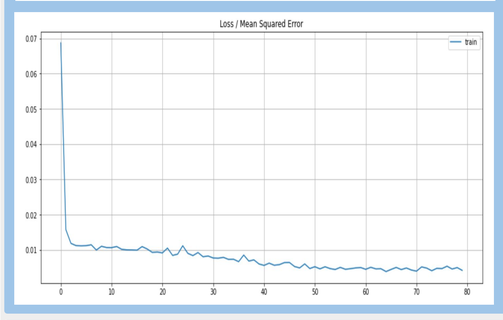


*Prediction Results for LSTM-Multi-Step Time Series Model*

## Performance Tuning



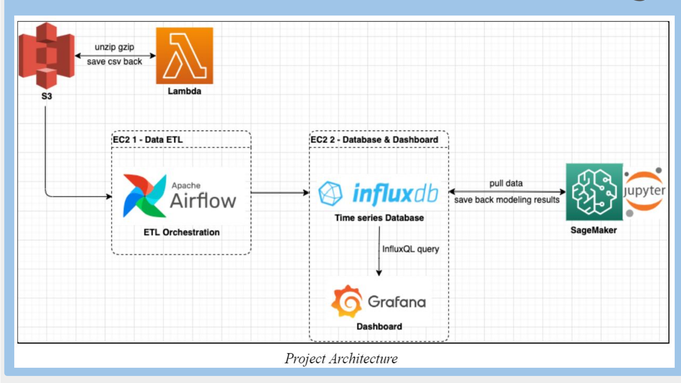

**Baseline models:**
- Autocorrelation
- Partial Autocorrelation

**Deep learning models:**
- Loss function

## Insights

- Deep learning models performed quite well and provided forecast results with high accuracy.
- By comparing deep learning modeling results against classical baseline modeling results, we confirmed that the deep learning models performed as well as and better than the baseline models.
- The success was quantitatively measured by taking the $R^2$, RMSE, and MAE values.

## Architecture Design



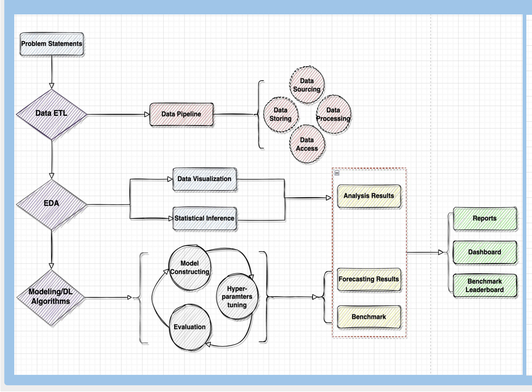

*Project Architecture*

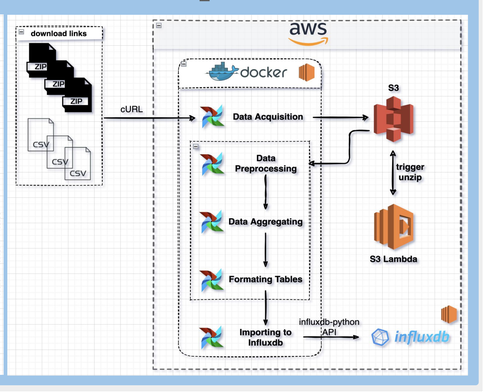
**AWS S3:**
- Store raw datasets

**Lambda function:**
- Target S3 bucket

**EC2 instances:**
- Launch Docker airflow server
- Launch InfluxDB server
- Launch Grafana

## Product Workflow and Data Pipeline

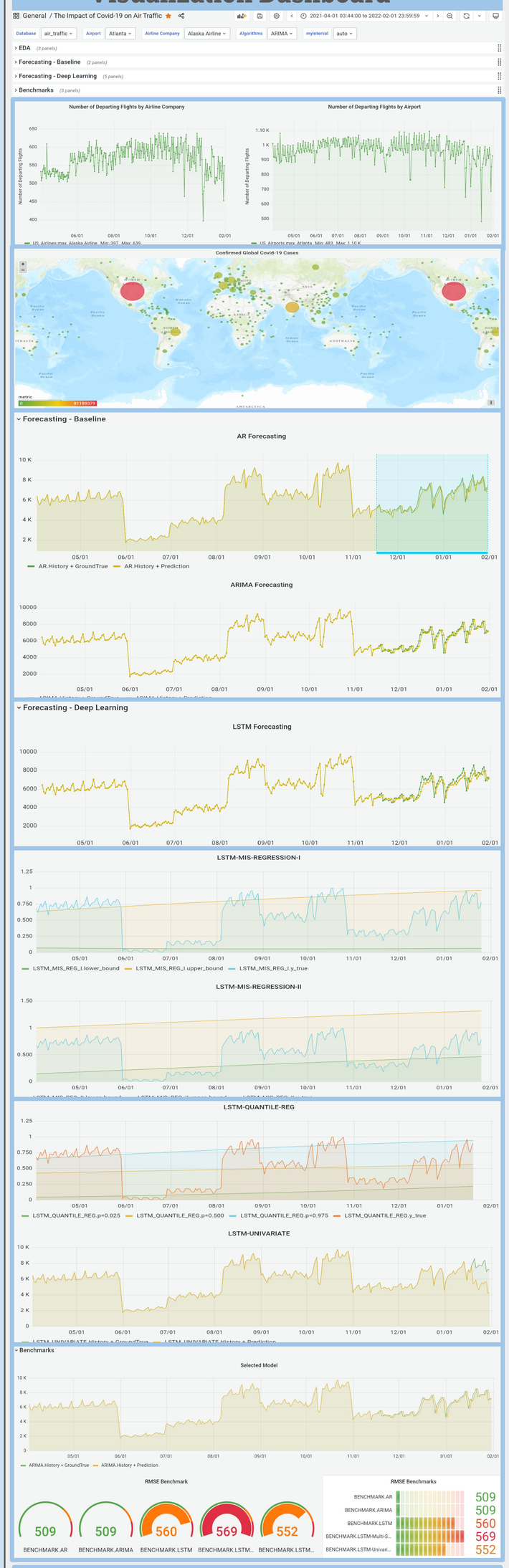



## Scalability

- **Data Scalability**: Model performance evaluated by incorporating different subsamples of the whole datasets, 10%, 20%, and 100% to analyze data scalability.
- **Model Scalability**: Model scalability is tested by assessing the compute time for different model network architecture.
- **Compute Scalability**:Compute scalability was achieved by wrapping deep learning models around PyTorch Lightning, AWS S3 and Celery executor to elastic scalability.

## Visualization Dashboard




## Conclusions

The deep learning models performed as well as and better than the baseline models. The results from our experiments are mainly geared towards analyst and management teams at airline companies so that they accurately and confidently predict future demand, optimize future prices, and adjust their fleets of planes as needed.