Airline

BOARDING PASS

- FLIGHT
DS10
- GATE
B10
- SEAT
1A

Capstone: Forecasting of Air Passenger Traffic

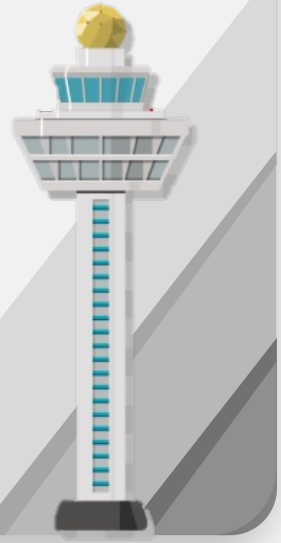
Presented by: Sherlyn Tan

Table of Contents

Airline

BOARDING PASS

01
Overview



The Transformation of Changi Airport

1981

Commencement of airport operations

1984

Commissioning of second runway

1990

Commencement of operations at T2

The Transformation of Changi Airport

2006

Commencement of operations at budget terminal

2007

Commencement of operations at T3

2012

Closure of budget terminal

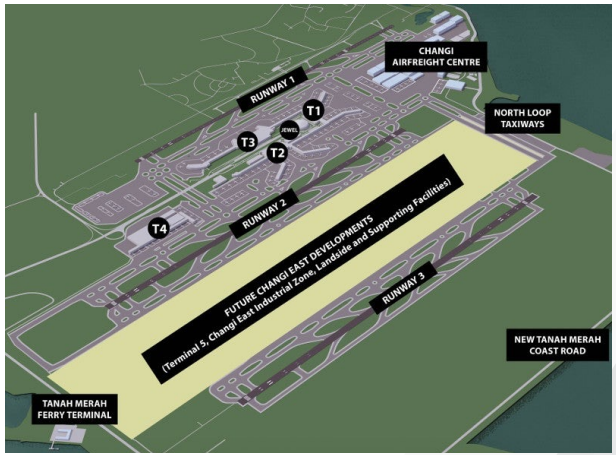
The Transformation of Changi Airport

2017
Commencement of operations at T4

2020
Commissioning of third runway

2030
Planned completion of T5

Problem Statement



Retrieved from: https://www.airport-technology.com/projects/terminal-5-changi-international-airport/

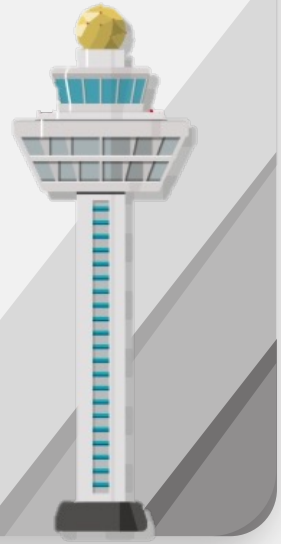
As part of CAG's airport operations team, I would like to utilise **regression** and **time series forecasting** models to forecast passenger traffic to assess sufficiency of the **passenger handling capacity** of Changi Airport with the addition of Terminal 5.

Airline

BOARDING PASS

02

Data Acquisition & Cleaning



Data Acquisition

Data Source	Description
Air Passenger Departures By Region / Country Of Disembarkation	• Monthly air passenger departures by region / country of disembarkation between January 1961 to April 2023
Air Passenger Arrivals By Region / Country Of Embarkation	• Monthly air passenger arrivals by region / country of embarkation between January 1961 to April 2023
Civil Aircraft Arrivals And Departures, Passengers, Air Cargo Tonnage, Direct And Transhipment Tonnage And Mail	• Monthly air passenger transit between January 1961 to April 2023

Data Cleaning

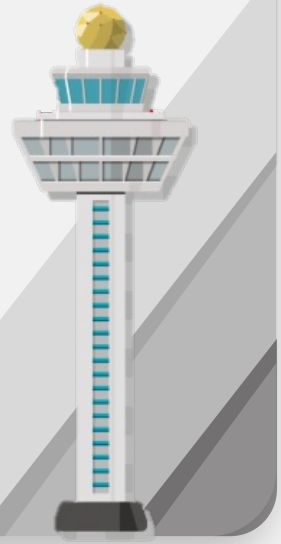
Dataset	Cleaning Steps
Monthly air passenger departures	• Formatting (i.e., transposition, resetting of index, improve readability of columns, reorder rows) • Check for missing values • Convert data types to int or datetime • Drop observations prior to January 1989 (since figures from January 1989 onwards are for Changi Airport alone)
Monthly air passenger arrivals	
Monthly air passenger transit	

Airline

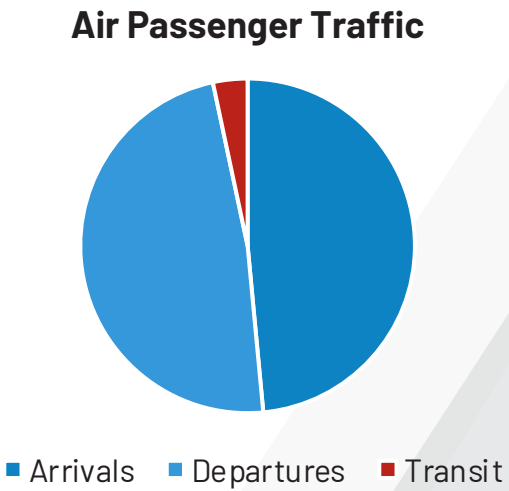
BOARDING PASS

03

Exploratory Data Analysis

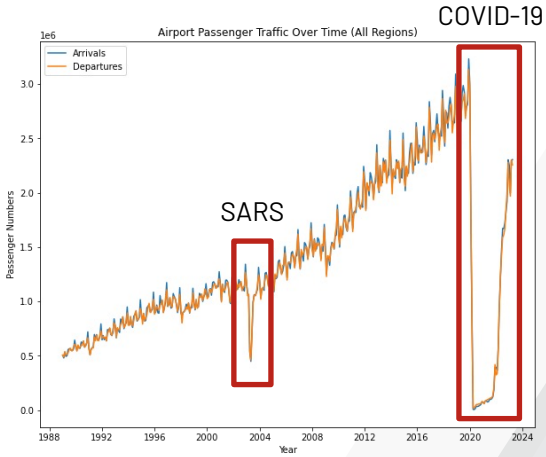


Overview of Datasets

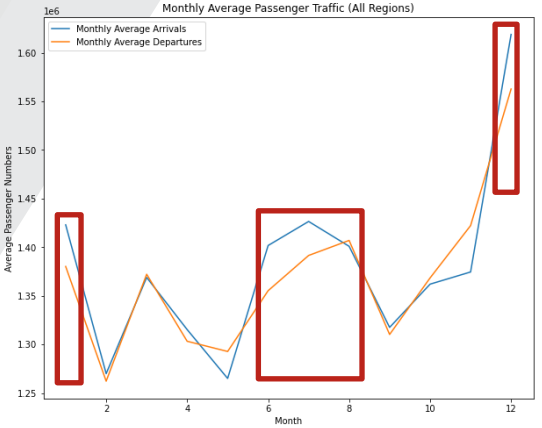


- Changi Airport handles mostly passenger arrivals and departures
- Proportion of arrivals and departures are roughly equal

Overview of Datasets (Arrivals & Departures)

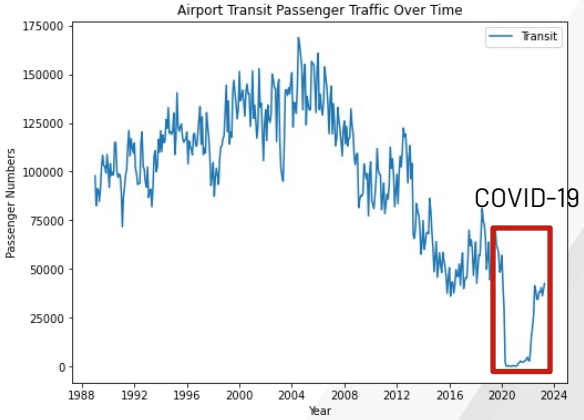


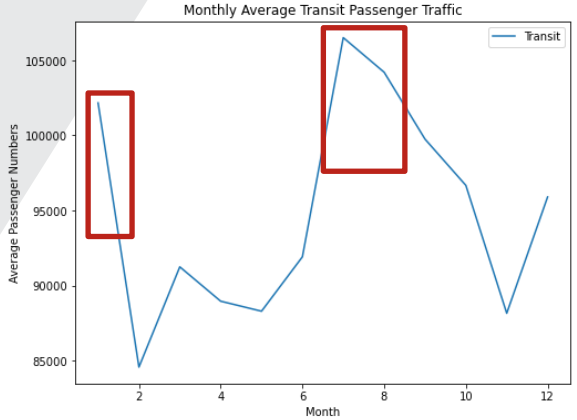
- General increasing trend
- Fluctuation pattern for both arrivals and departures mirrors each other quite closely



- General increasing trend
- Peaks in (1) early of the year, (2) mid-year and (3) year-end

Overview of Datasets (Transit)





- Increasing trend in earlier decades and peaks at around 2005 to 2006
- Decreasing trend beyond the peak

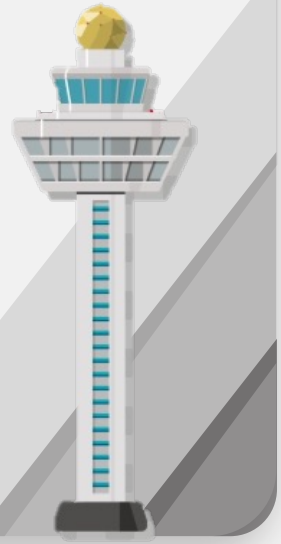
- Differs from the trends seen in arrival and departure traffic
- Peaks in early of of the year and in mid-year

Airline

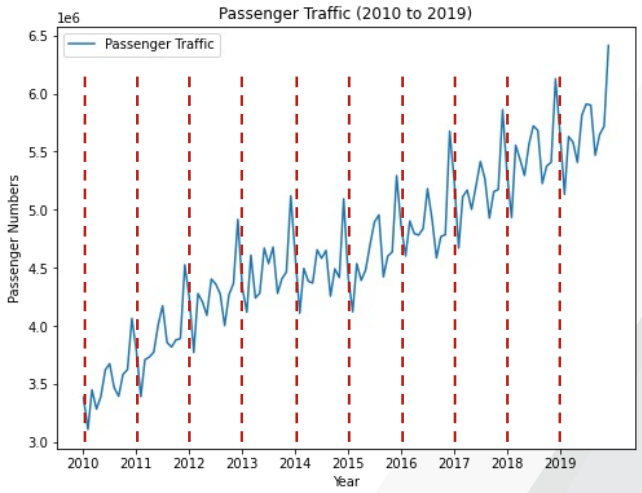
BOARDING PASS

04

Preprocessing & Modeling

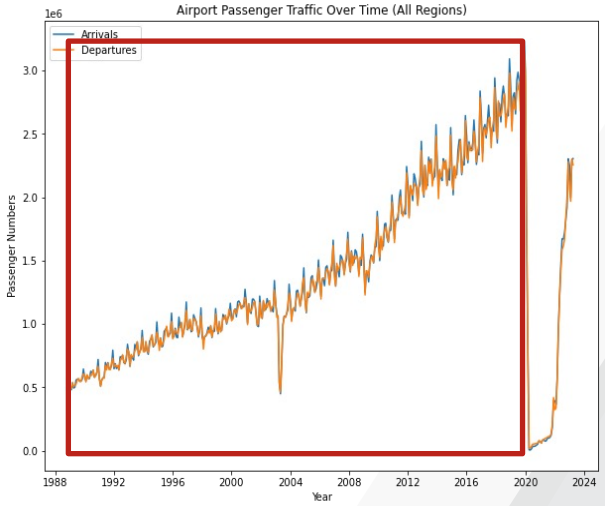


Preprocessing



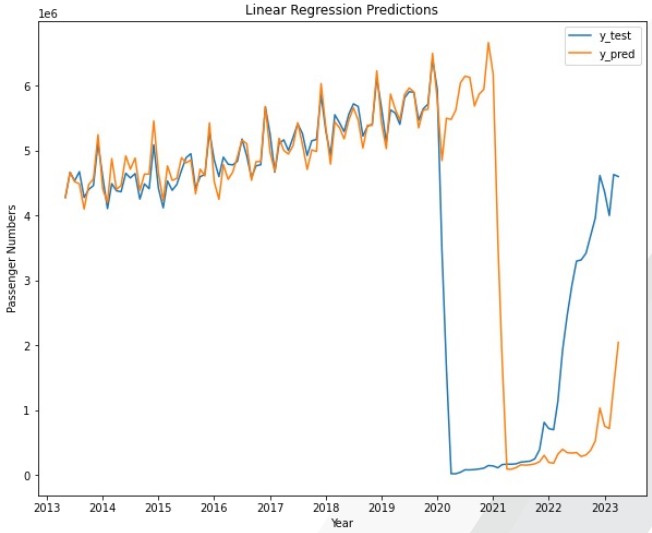
- Check for stationarity
  - Data is not stationary and thus, an additional column on differenced data is added
- Examine patterns in data to determine suitable lag interval
  - Each year experiences a spike in mid year before going down and spiking again at the end of the year and thus, an additional column for lag interval of 12 months is added

Modeling

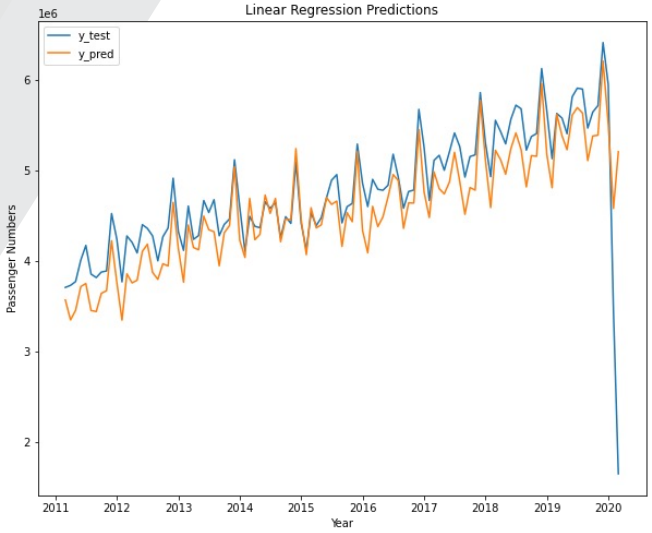


- Part 1: Before truncating dataset
  - All models performed consistently badly on the test set
  - Data in the train set are relatively stable in nature while the data in the test set contain drastic drop due to the halt in air travel during the pandemic
- Part 2: After truncating dataset
  - Rerun of all models using a truncated dataset to only retain data before the halt in air travel in 2020
  - Facilitates objective assessment of the models' performances

Modeling (Linear Regression)



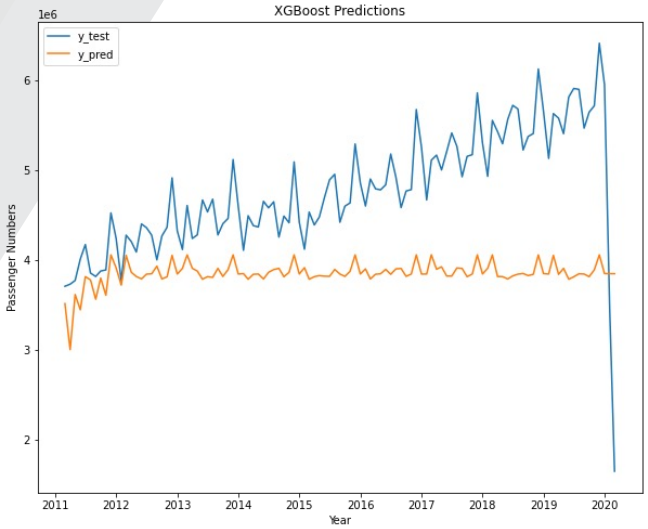
Before truncating dataset



After truncating dataset

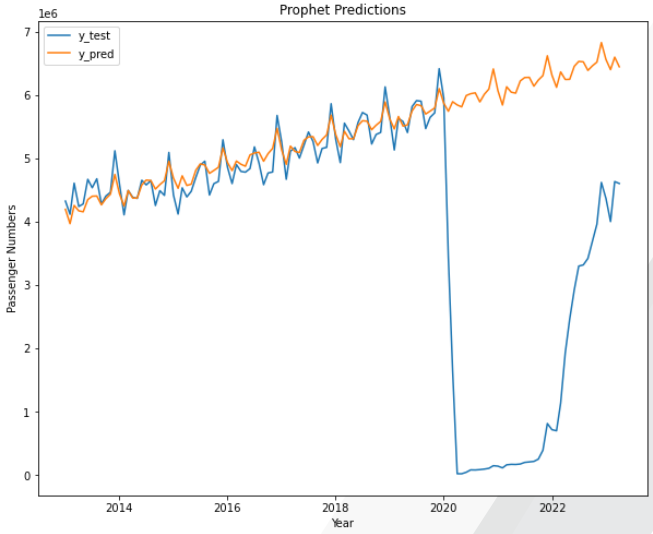
Modeling (XGBoost)

Before truncating dataset

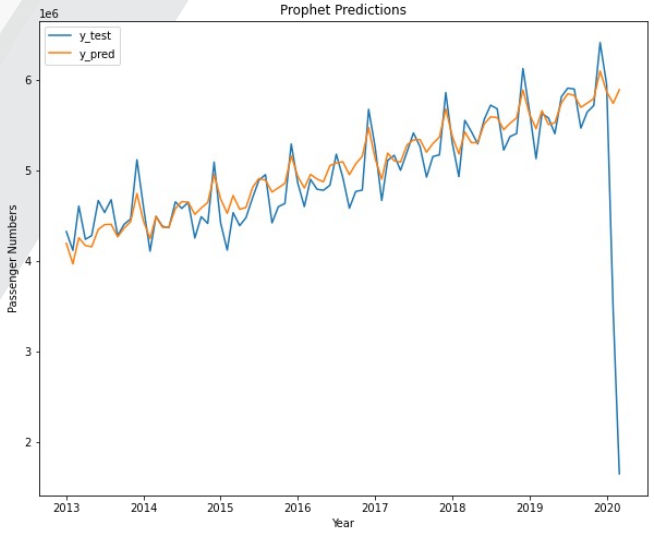


After truncating dataset

Modeling (Prophet)



Before truncating dataset



After truncating dataset

Outcome

- MAPE test values (rounded to 3 significant figures)

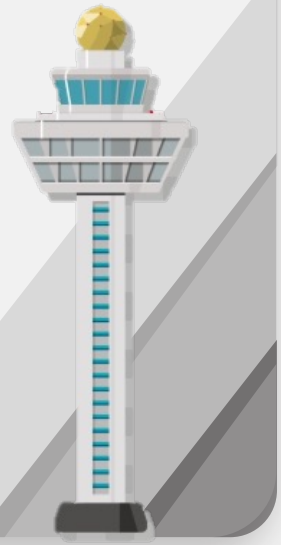
	Linear Regression	XGBoost	Prophet
MAPE (Test, before truncating)	8.59	7.36	10.9
MAPE (Test, after truncating)	0.0731	0.198	0.0668

Airline

BOARDING PASS

05

Conclusion & Future Steps



Conclusion

- While there are significant improvements in the performance of all models on the test set but the performance of the XGBoost model remains lacklustre as compared to Linear Regression and Prophet
  - Decision tree-based model may not be suitable in forecasting time series data
- The Prophet model was eventually selected as the final model to be used for the forecasting of air passenger traffic in view of its test MAPE score (of about 6.68%) being the lowest

Recommendation

- Air passenger traffic is expected to return to pre-pandemic level by 2024
  - Assign early 2020 (pre-pandemic period) as January 2024 and forecast into the future to arrive at the period of (pseudo) June 2030
- Using the Prophet model, the forecasted air passenger traffic in (pseudo) June 2030 would reach about 6.9 million which translates to about 82.5 million passengers passing through Changi Airport annually
  - Well within the airport's maximum handling capacity of 135 million with the addition of T5
  - By further stretching out the forecasting timeline, the annual forecasted traffic would reach about 133 million by (pseudo) August 2049

Future Steps

- Exploration of additional models to forecast traffic by regions to further assist in planning the flight distribution across the terminals
- Source for alternative sources of data which records the air passenger traffic observations daily to increase the number of observations and investigate its impact on the models' performances
- For Linear Regression model, additional features could be added to indicate local / regional holiday periods or socioeconomic indicators
- Further research and exploration of ConvLSTM2D model in view of the relatively low MAPE values of about 3% to 9% despite including observations affected by the COVID-19 pandemic (as published in the International Journal of Aviation, Aeronautics, and Aerospace).

Airline

BOARDING PASS

- **FLIGHT**
  DS010
- **GATE**
  B10
- **SEAT**
  1A