Impact of COVID-19 on the Aviation Industry

Vedanta Trivedi[a, *] and V. Pattabiraman[b]

[a]*B.Tech(CSE) Student, Vellore Institute of Technology, Chennai, Tamil Nadu, India*

[b]*Vellore Institute of Technology, Chennai, Tamil Nadu, India*

*****e-mail: vedantatrivedi.trivedi@gmail.com*

Abstract—Covid-19 almost shut down the operations for the transport sector as a whole, while the aviation industry took the biggest hit. With the restriction on the International and Domestic flights, the airline industry is struggling to survive, with mounting debt and all the major airlines in the world reporting huge losses in the Q1 and Q2 of this fiscal year. The pandemic is also expected to decrease the number of passengers that travel on airlines in 20-21 to about 25% to the number that travelled in 19-20, a number that has scared off any possible ways of getting help from the capital markets for these companies. This study analyzes The Impact Of Covid-19 On The Aviation Industry in detail, comparing data from 2019 - 2020 to previous years and comment on it's recovery.

Keywords: aviation industry, time series analysis, data exploration, facebook prophet, long short term memory

1. INTRODUCTION

The airline industry has always been one of the most competitive industries in the world. With high amounts of capital tied up in fixed costs, and high swings in the oil prices around the world, the airline industry has always struggled to put profits on the table, even though they are critical to the success of an economy.

The airline industry had to effectively stop all operations when the restrictions to air travel were announced in April 2020. With effectively no income and high operational expenses, the entire industry faced huge losses, with the companies forced to downsize and layoff their employees to stay afloat. While the impact of the pandemic on the industry is well known, the extent of the impact is still poorly understood. Using statistical analysis and time series forecasting, this study aims at understanding the financial implications of the pandemic on the industry, the trends and forecasts its recovery.

The factors analysed are:

1. Passenger Volumes
2. Air Traffic Volumes
3. Airfare
4. Revenue
5. Stock Prices

The problem statement can therefore be reduced to answer 5 important questions:

1. How has Covid-19 affected the volume of passengers between 2019 and 2020?
2. How has Covid-19 affected the number of domestic flights between 2019 and 2020?
3. What was the impact on airline fares?
4. What was the financial impact on the airline industry in terms of revenue?
5. What are the forecasted stock prices of the top airlines?

2. RELATED WORKS

In this section, related studies in time series modelling and the airlines industry are analyzed. In [1], annotated collections of over 35,000 real-world and model-generated time series, and over 9000 time-series analysis algorithms are analysed. The study employed a new approach to comparing across diverse scientific data and methods that allows them to organize time-series datasets automatically according to their properties, retrieve alternatives to particular analysis methods developed in other scientific disciplines and automate the selection of useful methods for time-series classification and regression tasks.

To understand the pattern in stock price forecasting and various models used, in [2], the researchers are using four types of deep learning architectures i.e Multilayer Perceptron (MLP), Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN) for predicting the stock price of a company based on the historical prices available. The network was able to predict for NYSE even though it was trained with NSE data. This was possible because both the stock markets share some common inner dynamics, which showed that our model could work on different types of stock market data.

However, there are serious challenges associated with producing reliable and high quality forecasts — especially when there are a variety of time series and analysts with expertise in time series modeling are relatively rare. To address these challenges, [3] describe a practical approach to forecasting “at scale” that combines configurable models with analyst-in-the-loop performance analysis. They propose a modular regression model with interpretable parameters that can be intuitively adjusted by analysts with domain knowledge about the time series.

Statistical models play a huge role in understanding the impact of an event on the industry. In [4], researchers attempt to analyze the impact of lockdown and covid crisis on airlines in India as of 2020 using the Altman Z-score model. The study also suggests the possible way-out for mitigating the expected losses. The results show that the sustainability of airlines warrants turnaround changes in their revenue strategies and operating models. Focus on minimizing losses rather than profit maximization possibly can help the airlines to combat the current situation.

3. METHODOLOGY

Data from official sources is collected and analysed using Python programming, with the libraries Pandas and Numpy used to perform percentage change analysis on the data. For the stock prices, time series analysis is performed, comparing the following models.

3.1 Facebook Prophet

The most commonly used models for forecasting predictions are the autoregressive models. Briefly, the autoregressive model specifies that the output variable depends linearly on its own previous values and on a stochastic term (an imperfectly predictable term).

Recently, in an attempt to develop a model that could capture seasonality in time-series data, Facebook developed the famous Prophet model that is publicly available for everyone.

Prophet is able to capture daily, weekly and yearly seasonality along with holiday effects, by implementing additive regression models.

The mathematical equation behind the Prophet model is defined as:

$$y(t) = g(t) + s(t) + h(t) + e(t)$$

$g(t)$: piecewise linear or logistic growth curve for modeling non-periodic changes in time series

$s(t)$: periodic changes (e.g. weekly/yearly seasonality) h(t): effects of holidays (user provided) with irregular schedules

e(t): error term accounts for any unusual changes not accommodated by the model

Prophet provides us with two models(however, newer models can be written or extended according to specific requirements). One is the logistic growth model and the other one is a piecewise linear model. By default, Prophet uses a piecewise linear model, but it can be changed by specifying the model. Choosing a model is delicate as it is dependent on a variety of factors such as company size, growth rate, business model etc., If the data to be forecasted, has saturating and non-linear data(grows non-linearly and after reaching the saturation point, shows little to no growth or shrink and only exhibits some seasonal changes), then logistic growth model is the best option. Nevertheless, if the data shows linear properties and had growth or shrink trends in the past then, a piecewise linear model is a better choice.

The logistic growth model is fit using the following statistical equation,

$$g(t) = \frac{C}{1 + e^{-k(t-m)}} \quad (1)$$

where,

C is the carry capacity

k is the growth rate

m is an offset parameter

Piecewise linear model is fit using the following statistical equations,

$$y = \begin{cases} \beta_0 + \beta_1 x & x \leq c \\ \beta_0 - \beta_2 c + (\beta_1 + \beta_2) x & x > c \end{cases} \quad (2)$$

where c is the trend change point(it defines the change in the trend). B1 is a trend parameter and can be tuned as per requirement for forecasting

3.2 Long Short Term Memory(LSTM) Model

Long short-term memory (LSTM) is an artificial recurrent neural network (RNN) architecture used in the field of deep learning. Unlike standard feedforward neural networks, LSTM has feedback connections.

This characteristic is extremely useful when we deal with Time-Series or Sequential Data. When using an LSTM model we are free and able to decide what information will be stored and what discarded.

LSTM has an internal state variable, which is passed from one cell to another and modified by Operation Gates.

1. Forget Gate: This gate decides what information should be thrown away or kept. Information from the previous hidden state and information from the current input is passed through the sigmoid function. Values come out between 0 and 1. The closer to 0 means to forget, and the closer to 1 means to keep.
2. Input Gate:To update the cell state, we have the input gate. First, we pass the previous hidden state and current input into a sigmoid function. That decides which values will be updated by transforming the values to be between 0 and 1. 0 means not important, and 1 means important. You also pass the hidden state and current input into the tanh function to squish values between -1 and 1 to help regulate the network. Then you multiply the tanh output with the sigmoid output. The sigmoid output will decide which information is important to keep from the tanh output.
3. Output Gate: This gate controls how much of the internal state is passed to the output and it works in a similar way to the other gates.

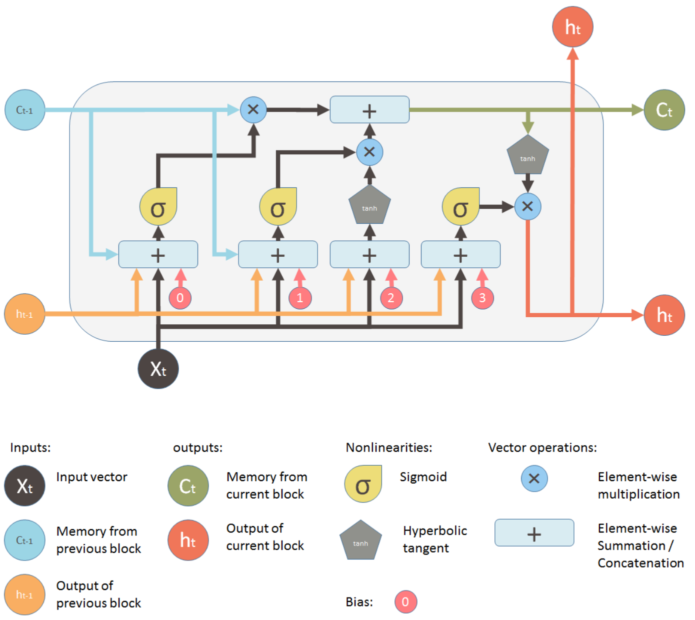


Fig. 1. Flow diagram of Long Short Term Memory Model

4. DATASETS

There were a total of 5 datasets which were used for this study. All these are used to analyze the impact of the pandemic on the US Airline Industry. They are :

a. Passenger Volume : TSA Coronavirus checkpoint: https://www.tsa.gov/coronavirus/passenger-throughput

Table 1. Dataset description for Passenger Volume

Columns	Description
Total Traveler Throughput one year ago	Number of Travelers on the same day a year ago
Total Traveler Throughout	Number of Travelers on that day
Date	The date of the flight

b. Airports : Data World - Airports, Airlines, and Routes: https://data.world/tylerudite/airports-airlines-and-routes/workspace/file?filename=airlines.csv

Table 2. Dataset description for Airports

Columns	Description
Name	Airport Name
IDs	Airport ID
City	City of the airport
Country	Country of the Airport
IATA	IATA Code of the Airport
Latitude	Latitude of the Airport
Longitude	Longitude of the Airport
Altitude	Altitude of the Airport

c. Airfare: Bureau of Transportation Statistics: https://www.transtats.bts.gov/AverageFare/

Table 3. Dataset description for Airfare

Columns	Description
Year	Year of the data
Domestic	Domestic Airlines(USA)
Latin_America	Latin American Airline
Atlantic	Atlantic Airlines
Pacific	Pacific Airlines
International	International Airlines
Total	Total Airfare

d. Financial Impact: Bureau of Transportation Statistics: https://www.transtats.bts.gov/Data_Elements_Financial.aspx?Data=6

Table 4. Dataset description for Financial Dataset

Columns	Description
Year	Year of the data
Quarter	Quarter of the year(1,2,3 or 4, or total)
Domestic	Domestic Airlines(USA)
Latin_America	Latin American Airline
Atlantic	Atlantic Airlines
Pacific	Pacific Airlines
International	International Airlines
Total	Total Airfare

e. Stock Prices(American Airlines) : Yahoo Finance: https://finance.yahoo.com/quote/AAL/history/

Table 5. Dataset description for Stock Prices

Columns	Description
Date	Date of the price
Open	Opening Price
High	Highest Price of the day
Low	Lowest Price of the day
Close	Closing Price
Adj. Close	Adjusted Close
Volume	Total Volume for the day

5. IMPLEMENTATION

To understand how covid-19 has affected the volume of passengers, we calculate the percentage change in the number of passengers before and after the 2 busiest days for airlines in the year - Christmas and independence day. We then plot the passenger volumes monthwise for both years side by side to understand the drop in numbers.

We use a similar, percentage change mechanism to analyze the changes in the number of flights, revenue and air fares.

For the time series model, we forecasted the closing price of the American Airline stock, in order to predict the recovery of the airline industry as a whole.

5.1 *Implementation of the LSTM Model*

1. A multi-layer LSTM recurrent neural network was built to predict the last value of a sequence of values.
2. The data was then normalised using the MinMax scaler.
3. The LSTM was built with 50 neurons and 4 hidden layers. 1 neuron was assigned in the output layer for predicting the normalized stock price. MSE loss function and the Adam stochastic gradient descent optimizer was used for optimization of the parameter values.

```
Model: "sequential_3"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
lstm_8 (LSTM)                (None, 50, 50)            10400
_________________________________________________________________
dropout_8 (Dropout)          (None, 50, 50)            0
_________________________________________________________________
lstm_9 (LSTM)                (None, 50, 50)            20200
_________________________________________________________________
dropout_9 (Dropout)          (None, 50, 50)            0
_________________________________________________________________
lstm_10 (LSTM)               (None, 50, 50)            20200
_________________________________________________________________
dropout_10 (Dropout)         (None, 50, 50)            0
_________________________________________________________________
lstm_11 (LSTM)               (None, 50)                20200
_________________________________________________________________
dropout_11 (Dropout)         (None, 50)                0
_________________________________________________________________
dense_2 (Dense)              (None, 1)                 51
=================================================================
Total params: 71,051
Trainable params: 71,051
Non-trainable params: 0
_________________________________________________________________
```

Fig. 2. Structure of Long Short Term Memory Model

5.2 *Implementation of the Facebook Prophet Model*

For this model, data was not split into training and test sets but instead we will use all the data to fit the model and then ask the model to predict future values i.e. the stock price in 2021.

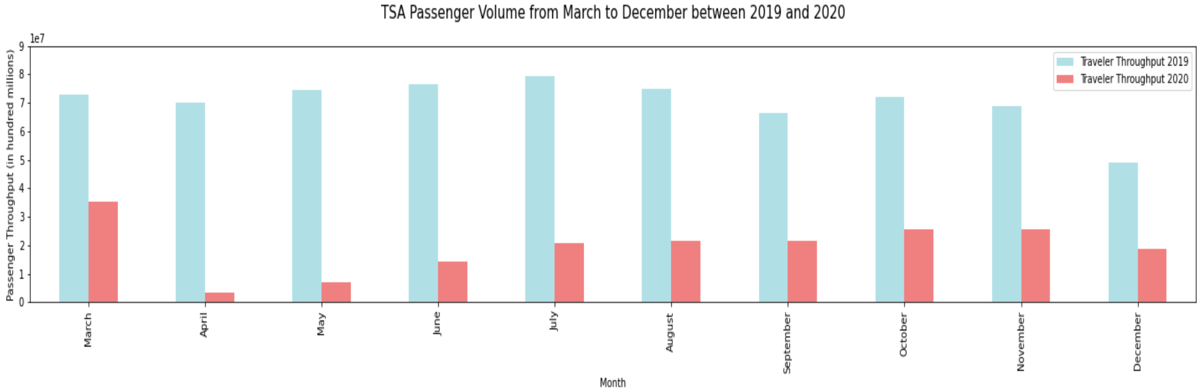
We then proceed to plot the trend, weekly, seasonally, yearly and daily components in order to gain a deeper understanding of the forecasted values.

6. RESULT AND DISCUSSIONS

6.1 *Passenger Volumes*

a. The largest deficit in passengers is seen in the months of April and May.

Fig. 3. Passenger volumes of 2019 and 2020



b. July 4th (Independence Day) volume was more heavily affected than the volume for December 23rd (before Christmas)

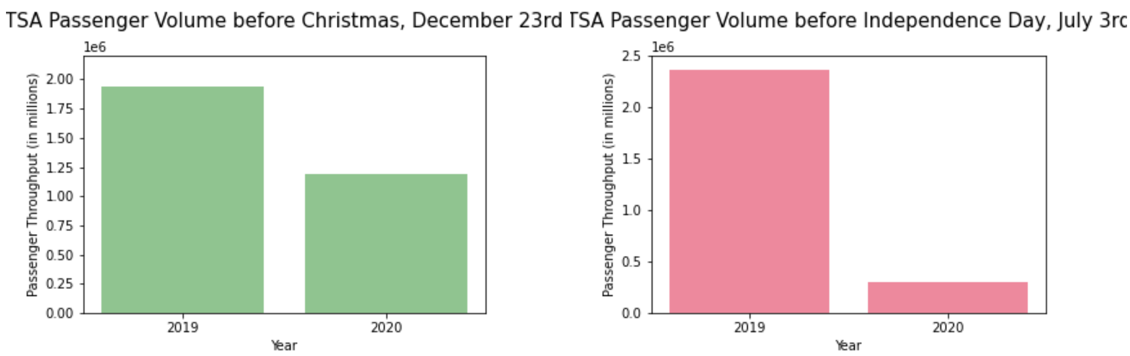


Fig. 4. Passengers volume comparison between Christmas and Independence Day

c. The percent change was most drastic in April, steadily decreases after July

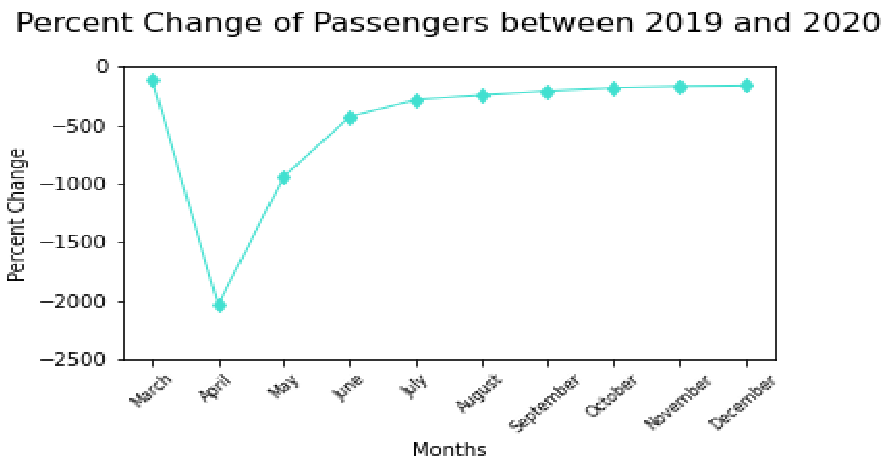


Fig. 5. Percentage Change of Passengers between 2019 and 2020

6.2 *Flights*

a. The total number of flights taken in 2019 is 9771281.
b. The total number of flights taken in 2020 is 4577349.
c. United Airlines had the highest percentage change in the number of flights between 2019 and 2020. (-65.98%).
d. United Parcel Service (UPS) had the lowest percentage change in flights between 2019 and 2020. (-16.90%)
e. FedEx and UPS were the only two mail services that were a part of the top companies with a high number of domestic flights.

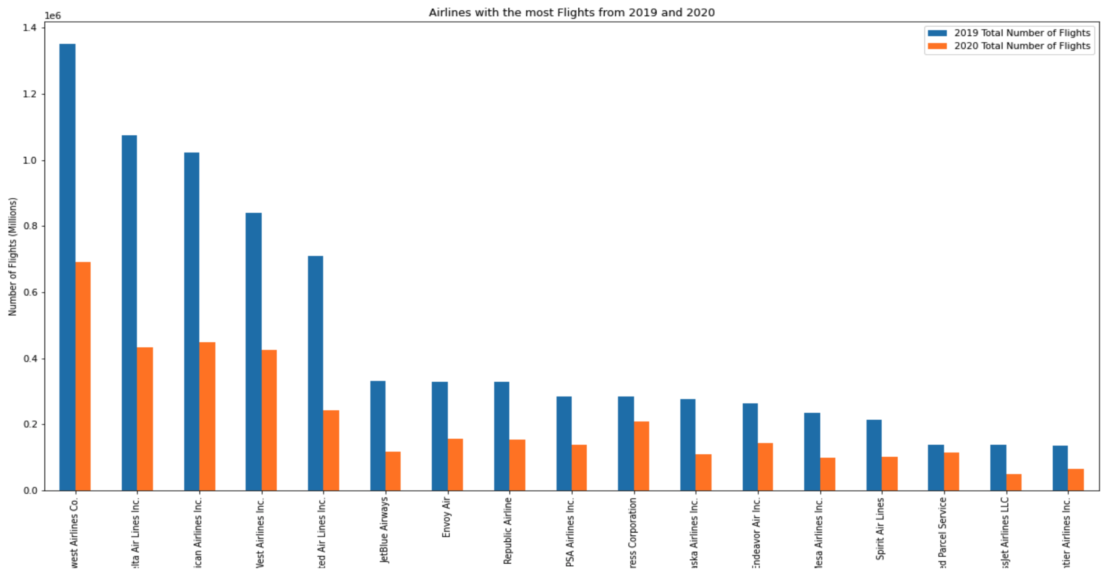


Fig. 6. Airlines with the most flights from 2019 and 2020

f. All top airline companies experience over 45% negative percentage change in the total number of domestic flights between 2019 and 2020 with the exception of UPS.
g. Southwest, Delta and American Airlines were the only companies to have over 1,000,000 domestic

flights in 2019.

h. In 2020, all top airlines had less than 1,000,000 domestic flights.

i. The change in the number of flights from top airlines from 2019 to 2020 is -53.0%. In other words, in 2020, the number of flights from around the world declined by 50%.

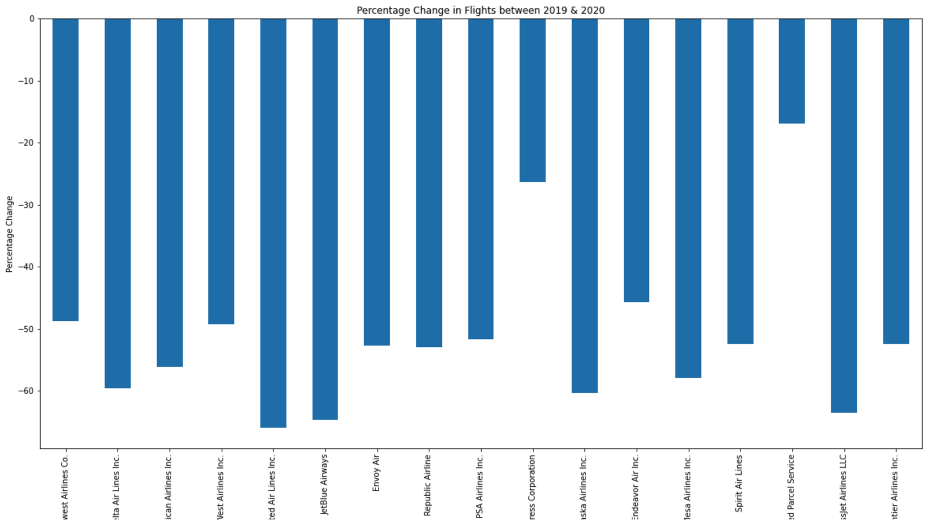


Fig. 7. Airlines with the most flights from 2019 and 2020

6.3 Airfare

a. From the graphs, we can see the strong effect that COVID-19 had on the price of airline tickets in the US

b. Every single airport saw a decrease of 15% or more

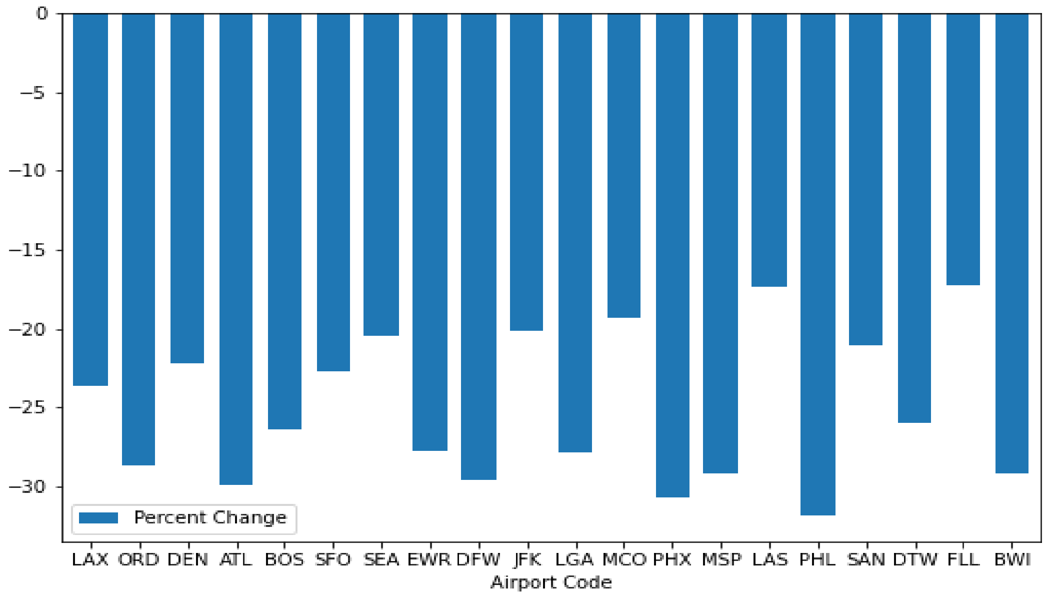


Fig. 8. Percentage change in Airline Ticket Prices in Airports

c. Airports surrounding major city hubs (ORD, ATL, DFW, PHL) saw the greatest decreases in airfare (> -25%)
d. In 2019, 17/20 airports had an average ticket price greater than $300. In 2020, only 3/20 airports had an average ticket price of $300

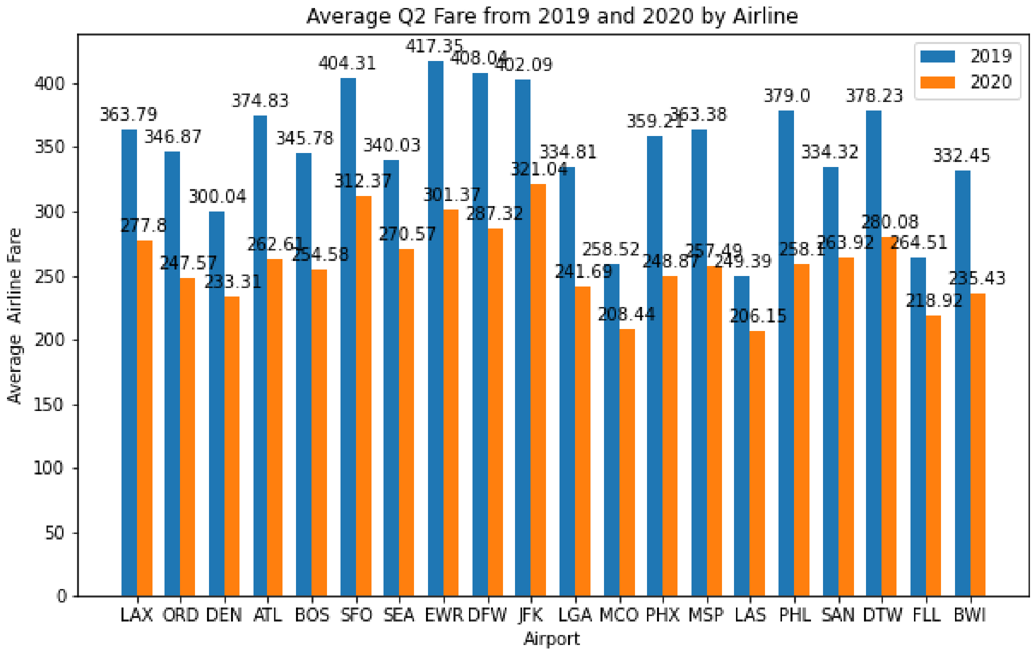


Fig. 9. Average Q2 Fare from 2019 and 2020, airline wise

6.4 Financials

a. Annual total revenue for domestic flights in the U.S. took a major plunge in 2020. The prior years show some consistency in the 10-20 million range. Travel restrictions and stay at home orders had a major impact on domestic flights beginning in March 2020.

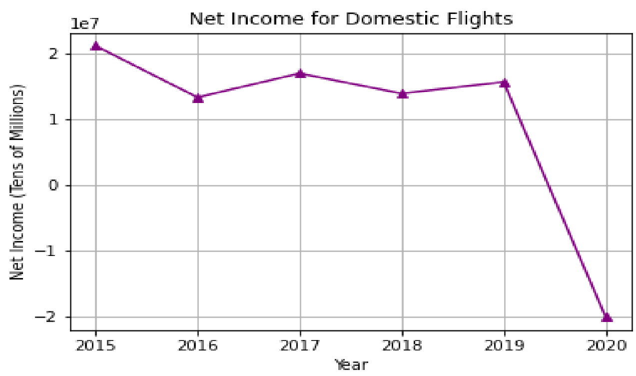


Fig. 10. Year Wise Net Income from Domestic Flights

b. Annual total revenue for international flights for U.S. carriers found substantial gains in 2020. Although the gains were high relative to prior years, domestic flights typically generate the highest revenue for U.S. carriers. The gains here were not enough to offset losses.
c. The increase in international flights reveals the increase in demand to leave the country during the pandemic.

d. An additional analysis was completed to display the average annual net income for U.S. carriers across all regions. The data frame supports our claim that drastic losses are directly related to the pandemic.

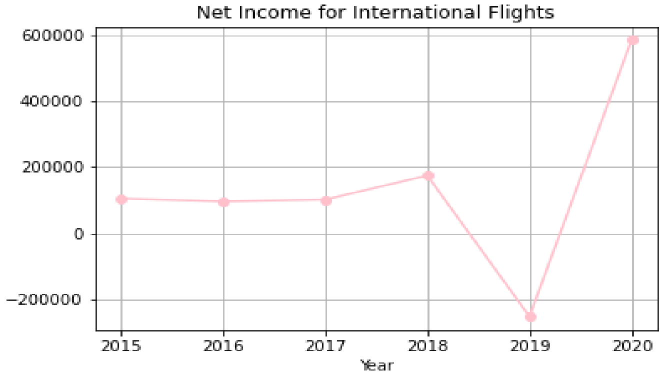


Fig. 11. Year Wise Net Income from International Flights

6.5 *LSTM Model Forecasted Values*

Our model was able to accurately follow most of the unexpected jumps/drops however, for the most recent date stamps, we can see that the model expected (predicted) lower values compared to the real values of the stock price.
Hence, this model was discarded for the other choice.

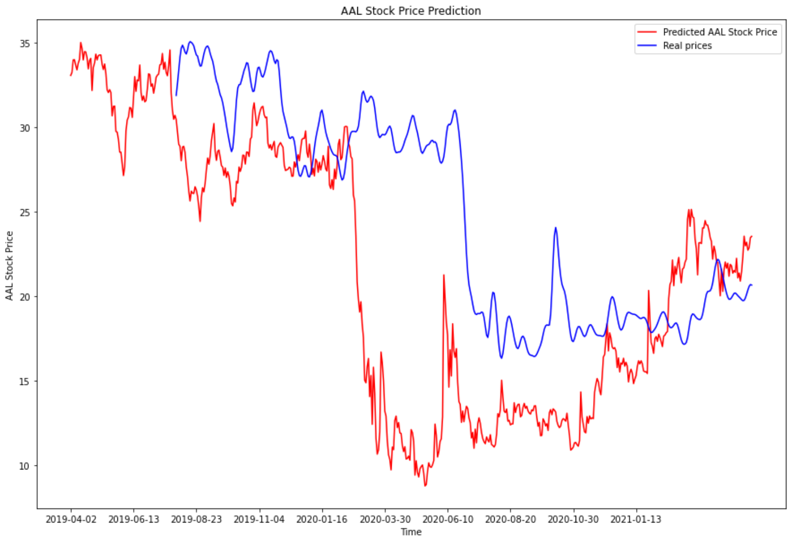


Fig. 12. Results from the LSTM Model, Huge drop in March 2020 due to lockdown restrictions

6.6 *Facebook Prophet Model*

a. The model used all the data for the training (black dots) and predicted the future stock price from Jan 2016 till May 2021. Blue shadow is the confidence interval.
b. From this forecast, we can observe another upward trend in the airline stock prices by the end of the year, but is still far away from the 2019 values.

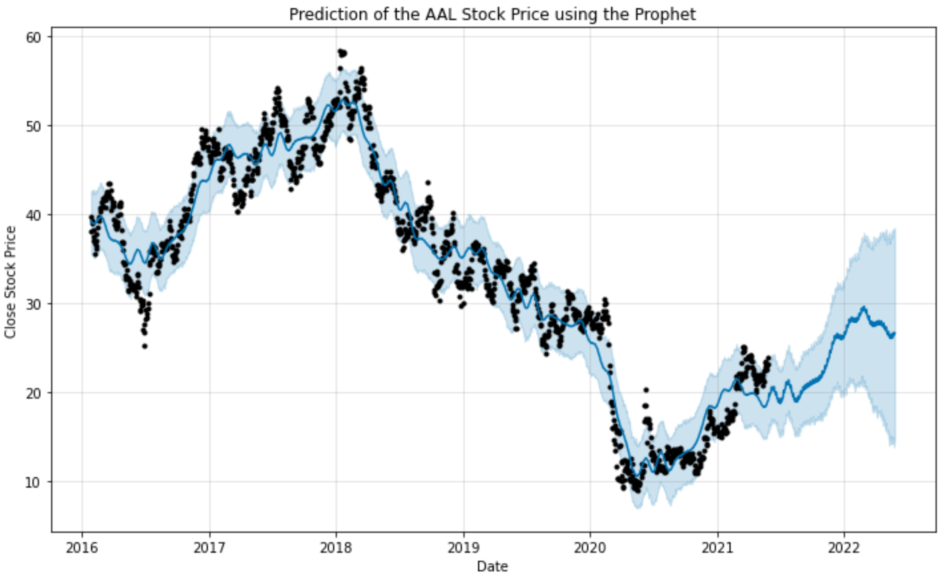


Fig. 13. Forecasted Results of the Prophet Model

7. CONCLUSION

a. The airline industry's recovery is highly dependent on the government restrictions being relaxed, as the passenger volumes are the lowest around this period causing the overall financials to go down as well.

b. International travel has been more profitable during covid-19, and there is an increased demand to leave the country during the pandemic.

c. Due to their low demand, airfares have gone down significantly, with airports around popular cities having the largest decrease in ticket prices.

d. The airline industry is not expected to recover before the start of 2022, unless there is a drastic change in the vaccination drive which facilitates easier travel and relaxation of travel restrictions.

8. REFERENCES

1. Fulcher Ben D., Little Max A. and Jones Nick S. 2013Highly comparative time-series analysis: the empirical structure of time series and their methodsJ. R. Soc. Interface.102013004820130048 - http://doi.org/10.1098/rsif.2013.0048

2. Hiransha M, Gopalakrishnan E.A., Vijay Krishna Menon, Soman K.P., NSE Stock Market Prediction Using Deep-Learning Models, Procedia Computer Science, Volume 132, 2018, Pages 1351-1362, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2018.05.050.

3. Taylor SJ, Letham B (2017a) Forecasting at scale. Am Stat. https://doi.org/10.1080/00031305.2017.1380080
4. Agrawal, A. Sustainability of airlines in India with Covid-19: Challenges ahead and possible way-outs. J Revenue Pricing Manag (2020). https://doi.org/10.1057/s41272-020-00257-z

5. Facebook Prophet Model Documentation - https://github.com/facebook/prophet
6. Keras Documentation - https://keras.io