

Understanding the US Airline Industry and the Factors Affecting the Operations of the Airlines

Shravan Shetty, Sowjanya Vijayanagar, Shraddha Sonawane, Vivek Vancha

I. INTRODUCTION

The civil aviation industry in the United States has emerged as an essential element in the transportation system of the country. Since majority of the population in the country relies on the airline industry for travel, it is imperative that the operations of the airlines and the airports must be efficient to ensure passenger satisfaction. It has been shown in the past that operation incompetencies like flight delays or security lapses hurt the revenue and the overall US economy.

This paper includes the following sections: research and other works related to our project have been discussed in the Section II. Section III covers our project questions in detail along with the processes and approach that we used. Section IV and V discuss the resulting analysis and the conclusion respectively.

II. PROBLEM DEFINITION

Considering the criticality of airline and airport operations in the success of aviation industry, we decided to base our analysis on the airline operation data. There are three questions that we are addressing in this project:

- Effect of flight delays in the busy airports: We analyze the relation between the top ten busy airports and their flight delays.
- Impact of September 9, 2001 on the airline industry: The 9/11 disaster was a major setback for the United States. We tried to analyze if the passenger turnout after the tragedy was affected by it and how quickly did the airline industry get back to normalcy.
- Most favorable time to travel: This will help to come up with a recommendation for the passengers on what time of the day is best to travel in order to encounter the minimum flight delay.

III. RELATED WORK

Amongst all the industries, the US airline industry faced a number of problems in fuel management, operations, increasing revenues, matching capacity to demand after the September 9/11 and other Natural disasters have only contributed to this downfall. Lee & Ito (2004) have pointed out that the airline capacity – demand has not recouped from the 9/11 effect even after two years. In our paper, we analyzed if this situation was accurate.

Our other research was to find the correlation between the busy airports and the time delays of the air planes. Bauer (1987) spoke about “Hub and Spoke” model in the airline industry which he describes that “hubs” are airports where one particular airline has its headquarters and “spokes” are where flights are departed to from the “hubs”. Mayer & Sinai (2002) noted that the airline delays are higher in Hubs compared to Spokes. In our analysis, we found which airports had highest air traffic and delays.

IV. PROCESS

A. Effect of Flight Delays in the Busy Airports:

In order to operate efficiently, it is necessary that the airlines and the airports encounter minimal flight delays to ensure smooth operations. We had a look at the airline data in the year 2008 for our analysis. Following are the few terms that we used for this question:

- Busy Airports: These are the airports that appeared in as the origin airport for the maximum number of times.
- Carrier Delay (CD): This is one of the fields in the dataset that comprises of the delays that are within the control of the air carrier. Since the flight may be delayed due to the delayed operations of the airline carrier, this delay is one of parameters considered to calculate the total delay.
- Security Delay (SD): This delay includes any delay that was caused due to security concerns that may have led to aircraft evacuation etc. This seems to be very unusual but it contributes the flight delay. Also, any security lapses are regarded as inefficiency on the part of the airport authorities.
- Elapsed Time (ET): This is the difference between the departure time and the arrival time. Some of the delays may not be accounted for in the carrier delay and the security delay. The elapsed time covers all the delays that is not accounted by the other delays

The required delay for the analysis was calculated using the following formula:

$$\text{Required delay} = ET + SD + CD \quad (1)$$

Since the dataset was very huge, the first step was to sample the data to obtain only those data entries that may be

relevant to our analysis. For example, the airports having very low air traffic may prove to be outliers and may not aid in any meaningful analysis. In order to resample the data, we find the mean of the count of entries of an origin airport. Once we obtain this mean, we eliminate all those airports which have entries that are lower than the mean. Once we obtained the sampled data, we found the top ten airports that had maximum entries as origin in the dataset. We termed these airports as the top ten busiest airports.

The airports were then ranked based on the highest delays encountered. These were the top ten airports that encountered the maximum delays. The results were then plotted and analyzed.

B. Impact of 9/11 on Airline Industry:

9/11 was one of the most challenging times for the airline industry in US. The airline industry had historically been plagued with low returns, bankruptcies, and ever fluctuating demand. After the 9/11 incident the situation only worsened. In this section we intend to measure the impact of the 9/11 incident and how long did it take the airline industry turn to normalcy after the impact.

The passenger and the seat data were provided separately hence merge operation was carried out to calculate the passenger load factor grouped by days. Passenger load factor (PLF) measures the capacity utilization of airlines.

$$PLF = \text{No. of passengers} / \text{Total no. of seats in the flights (1)}$$

We analyzed data across various sources such as International Air Transport Association (IATA), Research and Innovative Technological Administration Bureau of Transportation Services (RITA), Federal Aviation Administration (FAA), United States Department of Transportation (DOT) and for the purpose of this research we finally collected data from the United States Department of Transportation.

The FAA provided the data in the most granular form by dates, departure airport, arrival airport, carrier and other chosen variables such as passengers, seats etc. The data required cleaning which was carried out before the data was loaded for analyzing.

The timeseries library in pandas was used to do the analysis. The graph was plotted using the matplotlib library.

The data was grouped by month. For calculating the load factor the average load factor per month was calculated and for calculating the number of flights the total count of flights departed was calculated.

C. Most favorable time to travel:

No single individual wants their flight to be delayed for any particular reason. Over the years the flight delays have become very important measure of airline efficiency and it also has the significant role to play in increasing economic condition of the airlines. In this complex world, it is almost next to impossible to completely eradicate flight delays. But as

passengers we can chose the flights and times that has least flight delay probability. According to our analysis on the flight carrier data of year 2008, we were able to analyze which time period of a day had least flight delays.

To perform the analysis, we converted the day, month and year fields and converted into the time-series data. After this, we converted the scheduled departure time and the actual departure time into the date-time object. Then, we categorized the scheduled departure time into four categories based on the part of the day such as “Morning” – 6am: 12noon, “Afternoon” – 12noon:6pm, “Evening”– 6pm:12am, “Night”- 12am:6am.

V. RESULT AND ANALYSIS

A. Effect of Flight Delays in the Busy Airports:

From our analysis of busy airports and Fig 1, we found that the most of the busy airports are well prepared to deal with operational delays than the other airports. We also found that the reason for the delays of the flights is mainly due to Elapsed time than the carrier Delay and security delay.

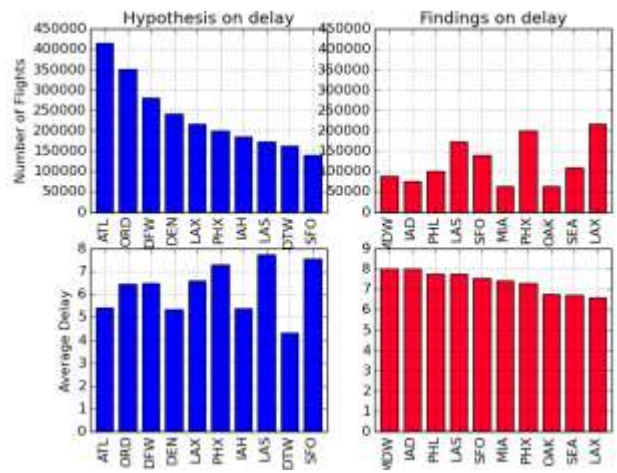


Fig 1: Busy airports vs Airports with maximum delays

B. Impact of 9/11 on Airline Industry:

We can see from Fig 2 that there was a huge trench in the passenger load factor in the last quarter of 2001 but the passenger load factor followed a rising trend and returned back to normalcy by 2002. The number of flights operated though, was a little low and kept fluctuating until the last quarter of 2003 after which they too returned to normalcy. So we can conclude that that though it 9/11 had a huge impact the impact was short lived as the airlines turned this around by adding additional security measures and thus restoring faith in the passengers' minds.

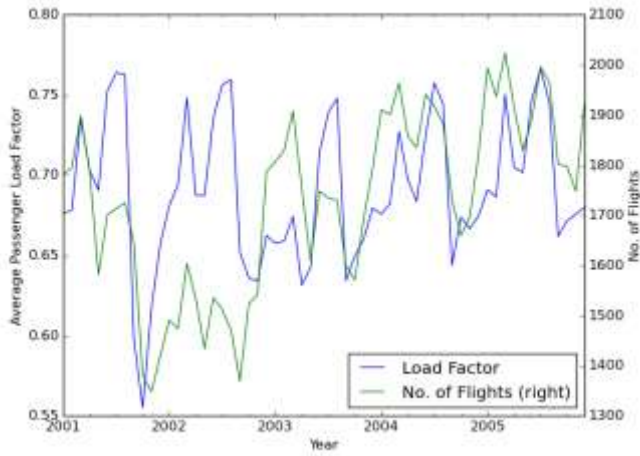


Fig 2: Average Passenger Load Factor and No. Of Flights

C. Most favorable time to travel:

For the passengers to book the flight, the best time is afternoon as the data suggests the delays are at minimum during the afternoon.

VI. CONCLUSION

This paper presents some of the important factors which affect the airline industry and their impact on passengers and thus revenue. We showed that inspite of the issues the airline industry had been plagued with and the competitive atmosphere in which they thrive the US airline industry is nimble to get back on its feet and provide great customer service.