Descriptive Analysis on US Flight Delays on the Winter Season (2018-19)

Author: Gerard Espejo Bas

Abstract

This paper aims to capture the relation between air traffic congestion, and air traffic delay, in airport operations. Also, identify patterns of behavior in air traffic delay. Using the data of "Airline On-Time Performance Data" of months: December 2018, January 2019 and February 2019; it is found that the concentration of air traffic in major airports doesn't contribute to the increase of flight delays. Actually, the major contributors to the increase of air traffic delay are the airports with less traffic. Also, the weather has an important effect on the geographical disposition of the flight delays.

1. Introduction

Commercial aviation is already the most global of businesses. Official Airline Guide data shows that the global network connects all countries in the world: 3,190 cities and 3,250 airports. Connectivity between these airports, which has doubled since the early nineteen-eighties, continues to grow, and more of the world's people from both emerging and mature economies use aviation as part of their lives.

- Long term growth potential for our industry is confirmed:
  - Traffic forecast to double in the next 15 years.
  - 4.4% average traffic growth over the next 20 years.
- Demand for 37,400 passenger and freight aircraft over the next 20 years:
  - ~36,560 passenger aircraft and 830 new build freighters.
- Asia-Pacific will account for 42% of deliveries, with airlines in North America and Europe together 35% of the passenger aircraft deliveries.[1]

Motivation. Following Airbus global forecast, it is clear that air traffic is going to be more congested, which will suppose an increasing amount of Flight Delays. It is crucial for the health of air traffic, and the aviation industry to understand the behavior of this matter, to be able to prevent it or reduced it.

Flight delays have been one of the important problems in airport management and flight scheduling, blurring the efficiency of air system operations and the choice of passengers. A significantly high number of passengers have suffered in choosing a reliable flight or airport.

Although some airports and/or airlines have put efforts in airport/airline management to reduce the possible delays, flight delays become unavoidable in some airports. In reality, multiple factors that impact flight delay are in many cases independent. A part of the indicators are related to the departure and arrival delay such as aircraft type, flight schedule, and flight departure/arrival. Others may relate to external triggers, like weather and airport capacity. [2]

This study aims to analyze the delay behavior in US airports, in the winter season (critical season due to the augmented chances of weather delays), in order to provide a wider understanding of flight delays behavior in airport operations. This study doesn't study delay factors.

Initial Hypothesis:

- The highest airports per volume flights, have the highest volume of delayed flights.
- The highest airports per volume flights, have the highest average delay.
- Airports located in the north area of the country, have higher average delays than southern airports.
- Most flown routes are directly related to the highest airports per volume of flights.
- Most flown routes have the highest volume of delayed flights.
- Highest delayed routes are directly related to the highest delayed airports.
- The highest carriers per volume flights, are the ones with highest flight delays.
- Highest delayed carriers' destinations are the highest delayed airports.

2. Methodology

Flight delay propagation is the correlation between arrival and departure delay, that is, late arrival of one flight causes late departure for the next flight on the itinerary of the same aircraft, we use the concept of propagated delay. [3]

Correlation Matrix. On that note, an initial exploratory has been conducted, analyzing the correlation and distributions between both delays. Figure 1.

[1] (2018). Global Market Forecast, de AIRBUS, URL: https://www.airbus.com/aircraft/market/global-market-forecast.html

[2] Weiwei Wu ,Cheng-Lung Wu, Tao Feng ,Haoyu Zhang ,and Shuping Qiu. (2017). Comparative Analysis on Propagation Effects of Flight Delays: A Case Study of China Airlines. mayo 20, 2018, de Hindawi, URL: https://www.hindawi.com/journals/jat/2018/5236798/

[3] Weiwei Wu ,Cheng-Lung Wu, Tao Feng ,Haoyu Zhang ,and Shuping Qiu. (2017). Comparative Analysis on Propagation Effects of Flight Delays: A Case Study of China Airlines. mayo 20, 2018, de Hindawi, URL: https://www.hindawi.com/journals/jat/2018/5236798/

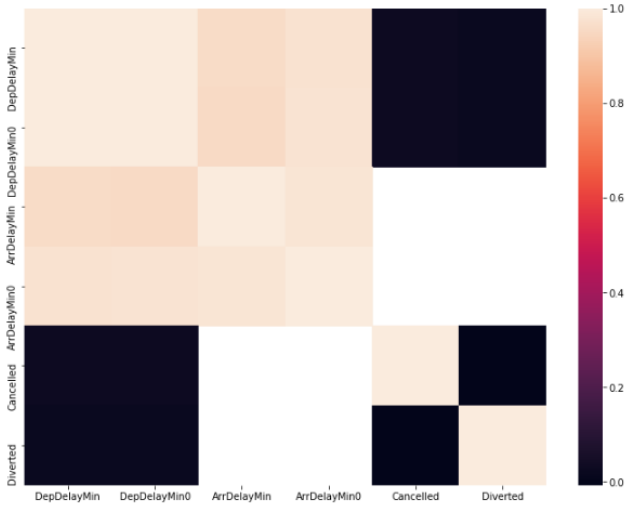

Figure 1: Correlation Matrix

The propagated delay concept is shown as a high value of correlation between Arrival and Departures delays: 0.959005. That indicates that there is high probability that if a flight has a delayed arrival the next flight departure will be delayed. These margin of 0.41 could be attributable to the recovered on-flight delay time.

Distribution. The joint distribution shows a very similar pattern is shown on the comparison on the count vs. mean graphics. Also, a normal distribution is shown for both types of delays, whereas different. Figure 4a.

For the data analysis of this paper, we will focus on the arrival delays. Although the state-of-the-art papers resolve the correlation between delays with the Copula function, having observed the key parameters of each field: count, mean, std, min, max and percentiles; it is decided to only take the arrival delays as a representation of the flight delay.

3. Data and Analysis

Flight data was sourced from the Bureau of Transportation Statistics, of the United States Department of Transportation. The dataset utilized is the "Reporting Carrier On-Time Performance" for months: December 2018, January 2019 and February 2019. The data includes scheduled and actual departure and arrival times, canceled and diverted flights, taxi-out and taxi-in times, causes of delay and cancellation, air time, and non-stop distance, etc.

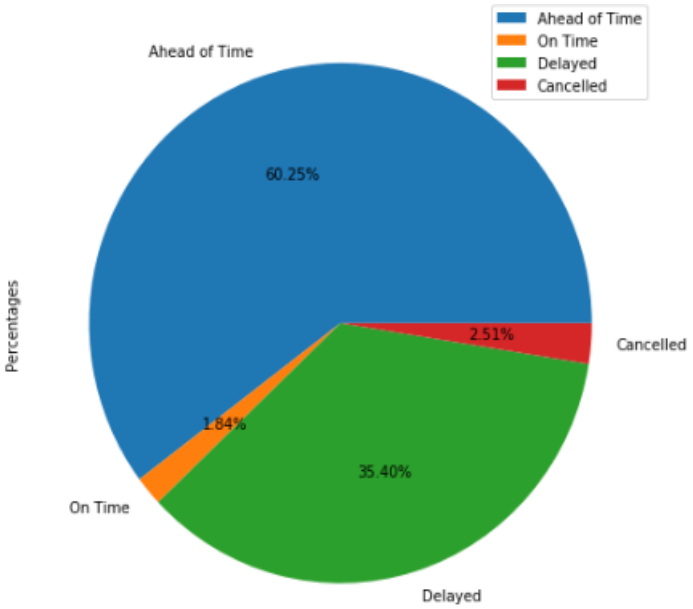

Figure 2: Flight Categoritzaton

3.1 Exploring Air Traffic of Winter Season (2018-2019)

Figure 2 shows that during the winter season of 2018-2019 the 60.25% of flights arrived ahead of time, adding the one-time ones, we have total of 62.09% of the total flights that didn't represent inconvenience to the passenger (delay). But we can also observe that the 35.40% of the flight had some type of delay, or even got cancelled (2.51%).

It is really interesting to see that in amount of percentage the 'Ahead of Time' and 'Delayed' flights, represent the majority of the flights. Whereas, the 'On-Time' flights only represent 1.84% of the total. From this we can rescue that the aviation sector isn't exact on time arrival estimations. Moreover, through this data we can recognize a pattern that airlines use to reduce their rate of delay, divulging higher flight times than real flight times.

The distribution of delayed flights between the three months of the winter season is very similar. December is the month with the most delayed flights (34.31%), Februrary (33.79%), whereas January is the lowest (31.90%).

3.2 Air Traffic Congestion Study

Figure 3. In order to examine the North American air traffic disposition, the study has been divided in three major layers: State, City and Airport. The top 20 states by number of flights indicates that most of the air traffic is located on the center and eastern part of the country. The States that appear on the top 20, that are not included in this area, such as: California or Texas, concentrate the air traffic in one major (or two) major Hub airports. The top 10 airports by volume of flights, don't follow a geographical pattern, whereas between the top 10 and top 20, the majority of the airports belong to the pattern shown on the State and City layers; indicating that the top 10 airports represent the major Hubs in each region.

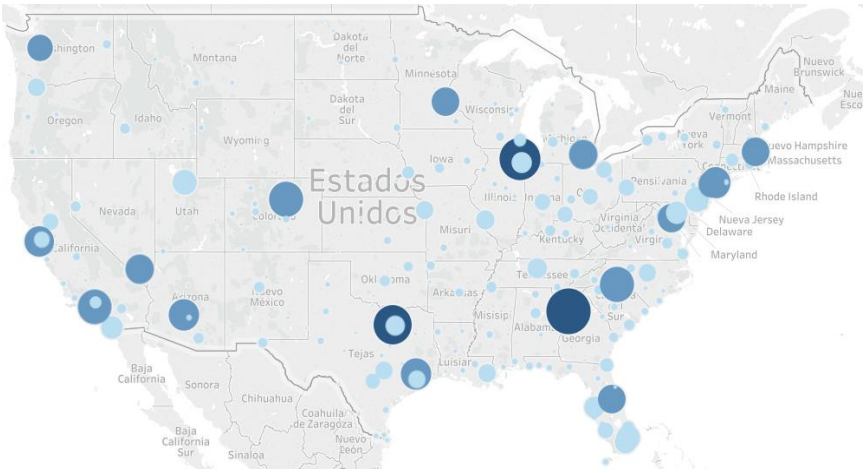

Figure 3: Air Traffic Congestion Map

3.3 Air Traffic Delay Study

Delay Distribution. The joint distribution shows the Departure and Arrivals Delays, distributions. Focusing only on the delayed flights (values above 0), the distribution is very similar. Is on the negative values (ahead and on-time flights) that the distributions change, representing the 62.09% of all flights. Both distributions show similar values on their 75% percentile: Departure=7.00 and Arrival=9.00, and their maximum delay values, Dep=2672 and Arrival=2649. The choice of representing the delay with only arrival delays seems valid. Figure 4b shows the Arrival Delay Distribution, underlining the delayed flights.

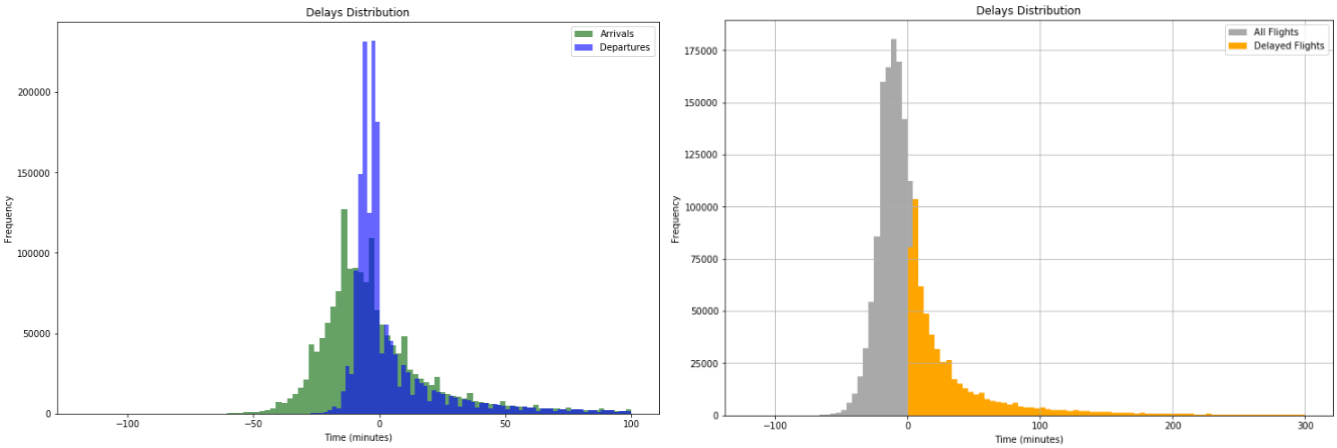

Figure 4: a.Joint distribution (Departure and Arrivals Delays) / b.Flights Delayed Underlined

3.3.1 Airports

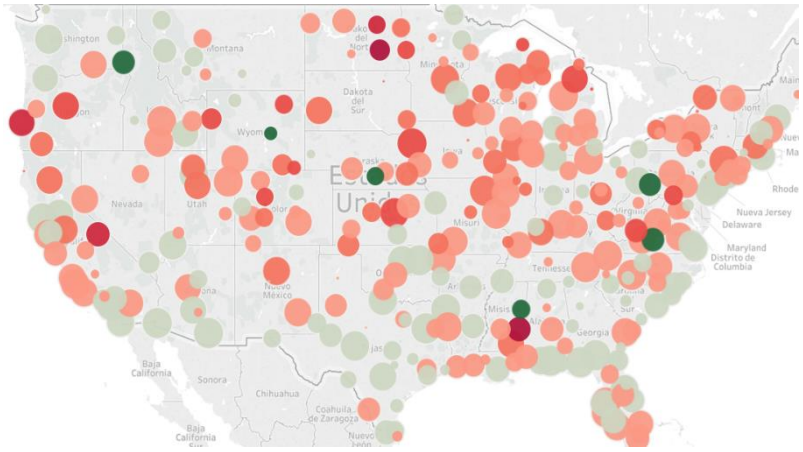

Figure 5: Geographical Delay Disposition by airports

Disposition of Flights Delays.

Figure 5. As it was concluded in the congestion study, the air traffic is mainly concentrated in the center and eastern part of the country. In the western side of the country, the delays are really concentrated in few airports, but very intense. Whereas, looking at the north-eastern part of the map, the concentration of delays seems to be more separated with values of delay often over the mean. Comparing north and south, the difference in average of delay by airports is high.

It can also be rescued from the map, that the highest airports per volume flights (Hub's), are not the highest in average delay. Let's check the following case.

***Case of study**: Comparison of the highest delayed airports delay distribution, against the highest airports by volume of flights.*

In order to examine the characteristics of the delays of each group, the main focus of the comparison is in the range of the following parameters: number of delayed flights (count), average delay (mean), standard deviation (std), percentiles and min/max values.

First thing we can observe is that the highest airports per volume flights (Hub's), are not the highest in average delay. Although, the are by amount of flights delayed.

Figure 6 and 7. Both of the groups show similar distributions but looking at the box plot of the highest delayed airports we can see that the standard deviation values, ranging from (ACV: 40.615654 to LWB: 132.264129) are much higher than the highest airports by volume of flights, which range from (CLT: 8.484677 to BOS: 30.670567). What that means is that the flights delayed in the highest airports by volume of flights, will tend to be close to the mean, having low rate of dispersion. Which in reality transforms to delays that won't normally be exceeding the mean by a lot, and if they do it will be in rare occasions because the number of flights is so high. Whereas looking at the dispersion in the highest delayed airports, is much higher. Also, the difference between the 75th percentile and the 95th is much higher, indicating higher maximum delay values.

Having analyzed this case it is clear that in major Hub airports (highest airports by volume of flights), airport operations are more optimized than in airports with less air traffic.

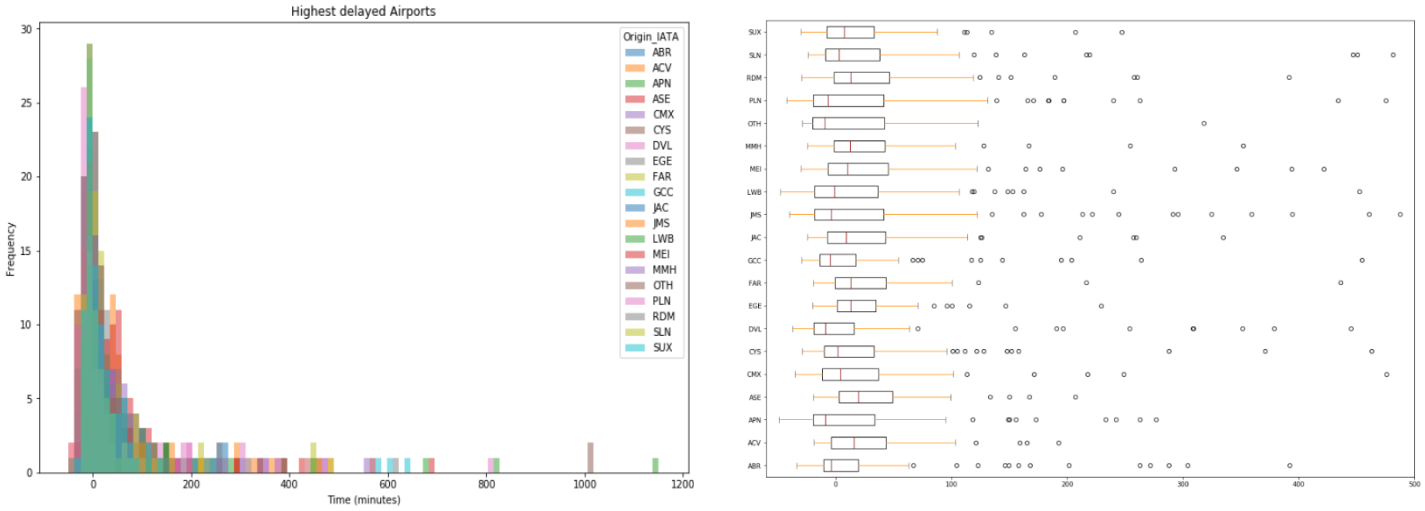

Figure 6: Highest delayed airports: a.Histogram b.Boxplot

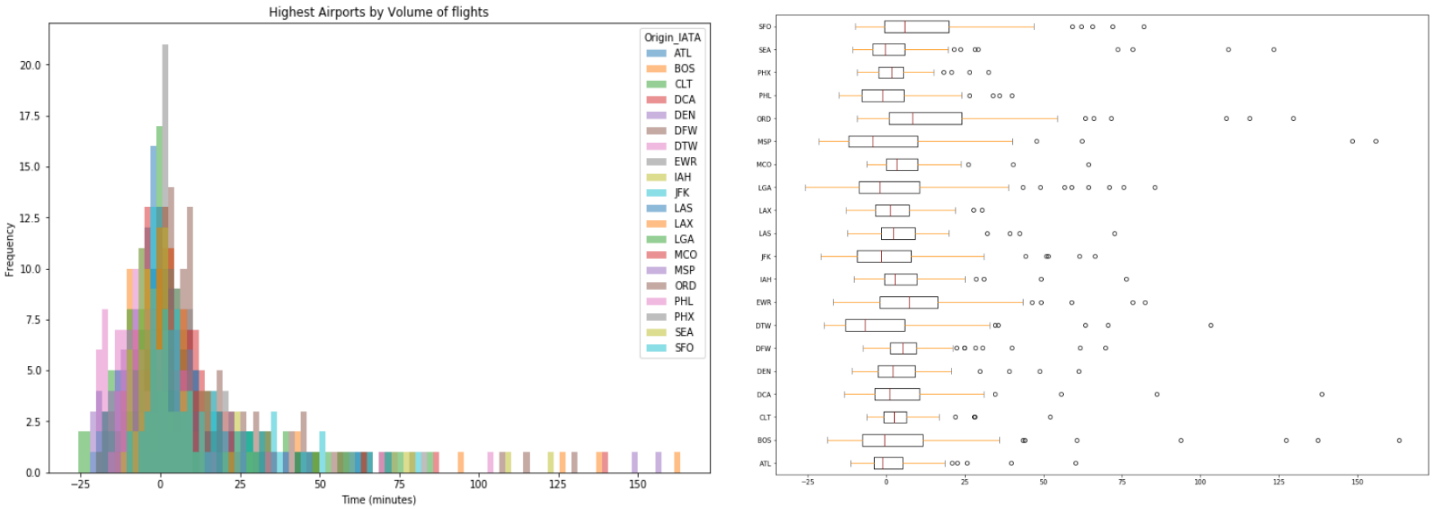

Figure 7: Highest Airports by volume of flights: a.Histogram b.Boxplot

3.3.2 Routes

Origin_IATA	Dest_IATA	FlightNum	
SFO	JFK	16	179
EWR	MCO	1527	165
BOS	BWI	1027	164
ATL	MCO	1153	139
LAX	JFK	24	136
PHL	BOS	1776	134
CMH	ORD	3476	131
LAX	OAK	949	128
EWR	PHX	230	111
JFK	SFO	15	104

Figure 8: Most flown routes

Figure 8 show the ranking of most operated routes in the US, which confirms the direct relation between routes and airports. The airports shown on the chart are part of the highest airports by volume of flights, indicating the concentration of Air Traffic in major Hub airports.

Moreover, in further analysis comparing routes ORD (Chicago) appears to be the major airport with more delayed routes, which is normal considering it is the airport with more delayed flights as shown on earlier airport analysis. But, also it is the most common airport in the most delayed routes by average delay, affecting the idea that highest delayed routes are directly related to the highest delayed airports, by average delay.

3.3.3 Carriers

Origin_IATA	OriginCityName	Carrier	
ATL	Atlanta, GA	DL	55276
DFW	Dallas/Fort Worth, TX	AA	34692
CLT	Charlotte, NC	AA	23530
ORD	Chicago, IL	UA	16973
		AA	15523
DEN	Denver, CO	UA	15046
IAH	Houston, TX	UA	14845
ORD	Chicago, IL	OO	14561
MSP	Minneapolis, MN	DL	14316
PHX	Phoenix, AZ	AA	13955
MIA	Miami, FL	AA	13408
EWR	Newark, NJ	UA	13344
SFO	San Francisco, CA	UA	12788
DTW	Detroit, MI	DL	12580
BOS	Boston, MA	B6	12407

Figure 9: Most common destinations for the most delayed airlines

In the dataset, only 17 airlines are considered. Therefore, the analysis cannot assure that the highest carriers per volume flights, are the ones with highest flight delays.

Figure 9 shows the most common destinations for the most delayed airlines by aircraft delay, which are the highest airports by volume of flights, which also indicates the concentration of Air Traffic in major Hub airports.

4. Results and Conclusions

It is found that the concentration of air traffic in major airports doesn't contribute to the increase of flight delays. Actually, the major contributors to the increase of air traffic delay are the airports with less traffic. Moreover, delayed flights represent the 35.40% of the air traffic.

Geographical Disposition. The Center-Eastern area of the country appears to concentrate a high percentage of air traffic. Also, the weather has an important effect on the geographical disposition of the flight delays, because the northern airports accumulate more amount of flights delayed, an higher average delays than the southern airports.

In my opinion, air traffic delays will always exist due to the fact that they can be originated by external factors. But as the aviation industry is growing, Airport Managers need to invest on new infrastructures capable of handling the increase in air traffic, and promote new procedures or tools of air navigation optimization, in order to prevent delays.