2015 Flight Delays: An Empirical Study

Jesseca Johnson
Udacity Data Foundations Nanodegree
October 2018

The following insights were found using 2015 Flight Delays and Cancellations data collected by the US Department of Transportation.

Insight One: How many? When?

https://public.tableau.com/profile/jesseca.johnson#!/vizhome/JJohnson_2015FlightDelays/InsightOne

How do we define a delayed flight?

What should we count as a delayed flight? If it lands one minute late, should that count as a delay? That seems a little harsh. For our purposes, let's use the standard that Paul Blackwood uses in his master's thesis and say that a flight will be considered delayed if it arrives 15 or more minutes past its scheduled arrival time.

What proportion of flights experienced a delay in 2015?

I determined that about 79% of flights were on time or early and 19% experienced a delay. (For the remaining 2%, we cannot tell from this data.)

How does the number of delays change with time?

Day of the Week

The largest number of delays occur on Day 4 of the week and the smallest number occur on Day 6. Since this is American data, Day 1 should be Sunday, meaning Day 4 would be Wednesday. However, in other parts of the world, the week actually starts on Monday, so we can't be totally sure about this.

Month of the Year

According to our data, the greatest number of delays occurs in June and the smallest number in September.

One limitation of both of these is that we are looking at raw numbers rather than a proportion of all flights for that month, so it may be, for example, that there are more delayed flights in June simply because there are more flights overall in June. It's the same thing for the day of the week.

Insight Two: Where? Who?

https://public.tableau.com/profile/jesseca.johnson#!/vizhome/JJohnson_2015FlightDelays/InsightTwo

One way to explore this data is to simply look at where delays are happening and which airlines have the most delays. Here we have the top 10 origin airports associated with delays, the top 10 destination airports associated with delays, and the top five airlines associated with delays according to the 2015 data.

Top 10 Origin Airports Associated with Delays

1. ORD (Chicago)
2. ATL (Atlanta)
3. DFW (Dallas-Ft. Worth)
4. DEN (Denver)
5. LAX (Los Angeles)
6. IAH (Houston)
7. SFO (San Francisco)
8. PHX (Phoenix)
9. LAS (Las Vegas)
10. LGA (New York City)

Top 10 Destination Airports Associated with Delays

1. ORD (Chicago)
2. ATL (Atlanta)
3. DFW (Dallas-Ft. Worth)
4. LAX (Los Angeles)
5. DEN (Denver)
6. SFO (San Francisco)
7. IAH (Houston)
8. PHX (Phoenix)
9. LAS (Las Vegas)
10. MCO (Orlando)

Top Five Airlines Associated with Delays

1. Southwest
2. American Airlines
3. Delta
4. Atlantic Southeast
5. Skywest

I did not use varying colors for this data as this would have been a distraction for the reader.

For the Origins board, there is a Destinations filter and vice versa so that you can play with the different origins and destinations to see who had the most delays.

Insight Three: Why?

https://public.tableau.com/profile/jesseca.johnson#!/vizhome/JJohnson_2015FlightDelays/InsightThree

If our goal is to reduce flight delays, it's important to look at the reasons. Our data offers five columns related to the delay reason:

1. Air system delay
2. Security Delay
3. Airline Delay
4. Late Aircraft Delay
5. Weather Delay

The data in these columns appears to be the number of minutes that each given reason caused a delay, and a delayed flight can have multiple delay reasons. The sum of these five columns comprises the ARRIVAL_DELAY column for flights that have a positive arrival delay.

It appears that "Late Aircraft Delay" is the most common reason with over 19,000 minutes recorded. Air System Delays and Airline Delays were the second and third most common reasons with just over 15,000 minutes of recorded delay time for each. However, a major limitation of this data is that we don't know more details about these causes. For example, what is causing late aircraft delays? What is an air system delay, exactly? And what caused airline delays? We simply aren't given much information to work with here.

There is a month filter attached to this worksheet so that you can see how the reasons change with the months.

Resources

1. Image credit: Pexels Free Stock Photos
2. Blackwood, Paul. *Understanding Flight Delays at U.S. Airports in 2010, Using Chicago O'Hare International Airport as a Case Study.* Western Michigan University, April 2012. Accessed at https://scholarworks.wmich.edu/cgi/viewcontent.cgi?article=1033&context=masters_theses on October 11, 2018.