

Report

Flight Cancellation and NAS Study

by Qifang Zheng

Submission to fulfill Udacity Data Analyst Nanodegree

April 2019

I travel often. Having my flight depart on-time is the start of a good trip. I have experienced flight delays and the inconvenience that they have caused, such as itinerary changes, missing the next connection, and etc. Furthermore, flight was canceled occasionally for various reasons. This is the worst situation for a traveler. Flight cancellation not only causes inconvenience but also brings emotional distress. A “Canceled” label on the big TV screen is every traveler’s nightmare. I want to study flight cancellation reason from this project.

I used the Flights dataset from ASA website <http://stat-computing.org/dataexpo/2009/the-data.html>. There are data available from 1987 through 2008. I downloaded the data from 2003 to 2008, the most recent six years, to analyze the flight cancellation and its trend.

First, I downloaded 2008.csv and loaded it to a dataframe. There are four flight cancel reasons, weather, carrier, NAS (National Aviation System) and security. In 2008, there are 2% of the total flights were canceled, within which, 11% cancellation occurred at Chicago O’Hare (ORD) airport. I made a list of the top 10 airports based on how many flights were canceled in 2008. 40% of the cancellation occurred in the 10 airports below.

No. 1. ORD – 15050 flights

No. 2. DFW – 7272 flights

No. 3. ATL – 5830 flights

No. 4. LGA – 5753 flights

No. 5. EWR – 4511 flights

No. 6. BOS – 3655 flights

No. 7. IAH – 3261 flights

No. 8. JFK – 3196 flights

No. 9. LAX – 2838 flights

No. 10. SFO – 2790 flights

Now let’s look at the reasons why the flights were canceled. Particularly, those for O’Hare (ORD). Figure 1, 2 and 3 show the percentage of four flight cancellation reasons. The difference among the three figures is their location. Figure 1 is for all U.S. airports. Figure 2 is for the top 10 airports above. Figure 3 is for O’Hare airport.

Figure 1: Nationwide Cancellation Reason

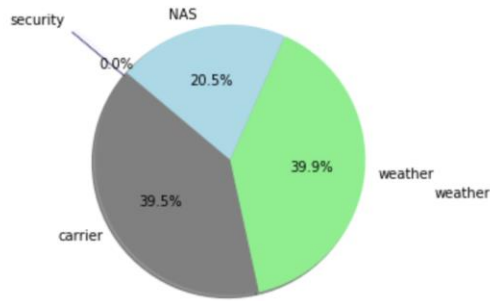


Figure 2: Cancellation Reason for the top 10 airports

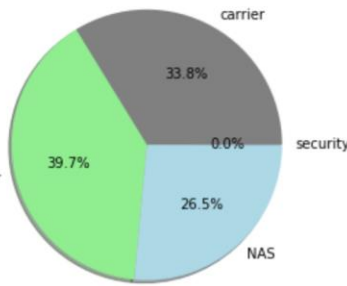
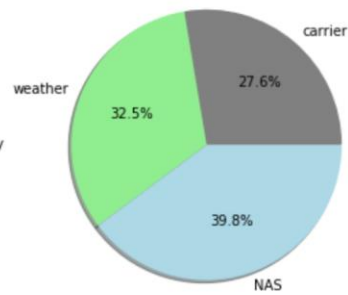


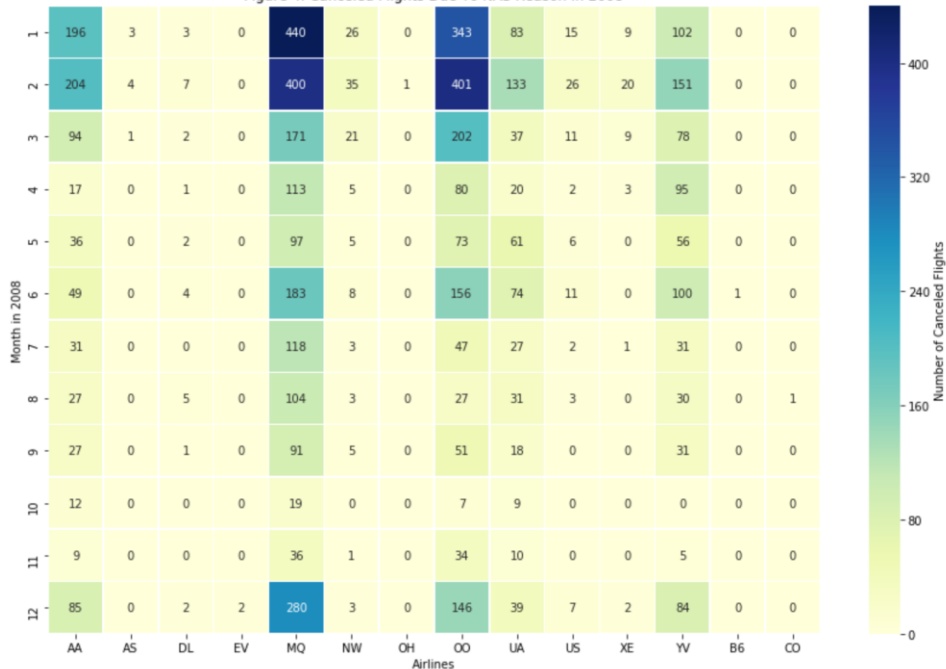
Figure 3: Cancellation Reason for ORD



Weather, carrier and NAS are the three majority reasons for flight cancellation. Only a couple flights were canceled due to security process (it is very unlikely a flight is canceled because passengers are stuck in the security line). Weather is the largest factor leading to cancellation. However, for Chicago O'Hare airport, NAS caused nearly 40% of the total number of cancellations. There were 6,020 flights cancelled due to NAS at O'Hare airport in 2008.

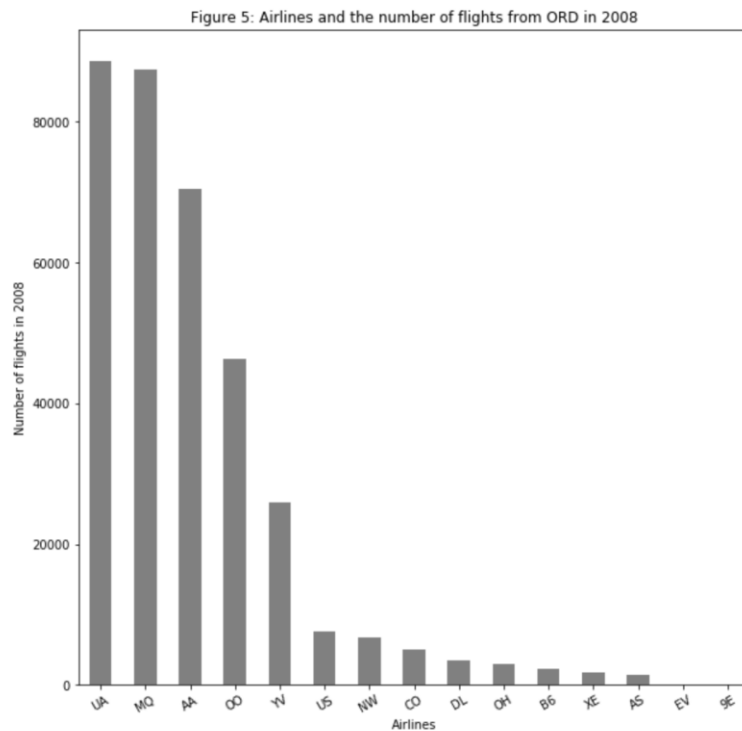
Next, I want to look at which airlines had the most cancellations at O'Hare due to NAS. There are fourteen airlines having outbound flights from O'Hare in 2008. In total, the issue of NAS cancelled 5,991 flights. The result shows MQ (Envoy Air Inc., formerly American Eagle Airlines) and OO (SkyWest) owned 3,619 flights, 60% of the total cancelled flights.

Figure 4: Canceled Flights Due To NAS Reason In 2008



2,602 flights, 43% of the cancellation, were cancelled in January and February due to NAS. This is surprising since I expected December would have the most cancellations because of the heavy holiday traffic. Only 650 flights were cancelled in December.

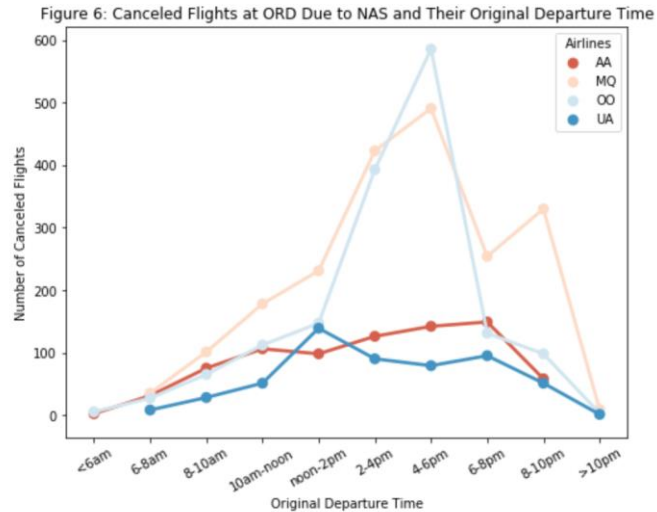
Why did Envoy and SkyWest cancel so many flights? What could be the reasons? Would it be possible they cancelled because they had a lot of flights to depart from O'Hare? What were other airlines departing from O'Hare? Did they depart more or less than Envoy Air and SkyWest? I will explore next.



It turns out Envoy Airline had a lot of flights from O'Hare but it was not the #1. United Airlines (UA) had the most flights from O'Hare in 2008. Envoy, American Airlines (AA) and SkyWest ranked #2, #3 and #4. United and Envoy were tie.

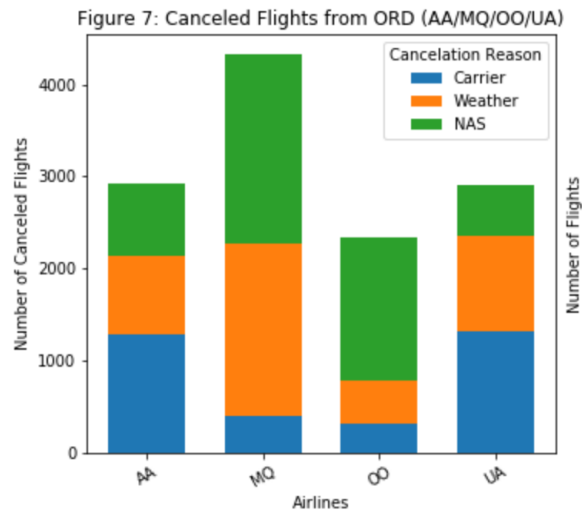
Why did Envoy cancel more flights than United? Maybe the departure time can provide further information. Was there a timeframe in which more flights were canceled? I am going to look at the scheduled departure time.

In 2008, the earliest departure time was 6:00am. The latest departure time was 22:47. Hence, the departure times were divided into ten timeframes, before 6am, 6-8am, 8-10am, 10am-noon, noon-2pm, 2-4pm, 4-6pm, 6-8pm, 8-10pm and after 10pm.



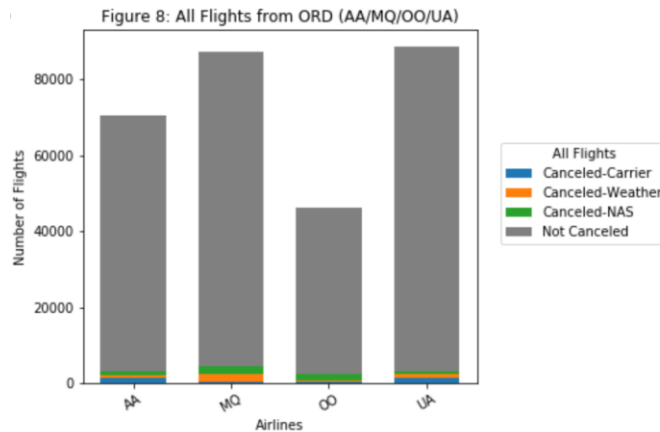
From noon to 6pm, the number of cancelled flights increased for Envoy, SkyWest and American. Possible evidence for this is because there are more flights in the afternoon. The cancellation increased dramatically for Envoy and SkyWest. However, United cancelled less flights in the afternoon. I assume there was an NAS resource competition between United and Envoy/SkyWest.

What about other cancellation reasons? Are there any clues suggesting a correlation?



From Figure 7, Envoy and SkyWest cancelled less flights than United and American due to an airline reason. More Envoy flights were cancelled due to weather. Envoy and SkyWest cancelled a lot of flights due to NAS. Whereas, NAS impacted United and American less. Furthermore, United had the least number of flight cancellations due to NAS.

By bringing the total number of flights into the picture, it is clear to see United had the most flights (the gray portion in Figure 8). Figure 8 shows Figure 7 in scale of total flights, in contrast to only cancelled flights in Figure 7. Though United and Envoy had a similar number of flights, Envoy cancelled more flights due to NAS. Similarly, though American had more flights than SkyWest, SkyWest cancelled more flights due to NAS.



“O’Hare is unusual in that it serves a major hub for more than one of the three U.S. mainline carriers; it is United’s largest hub in both passengers and flights, while it is American’s third-largest hub.”
[Wikipedia]

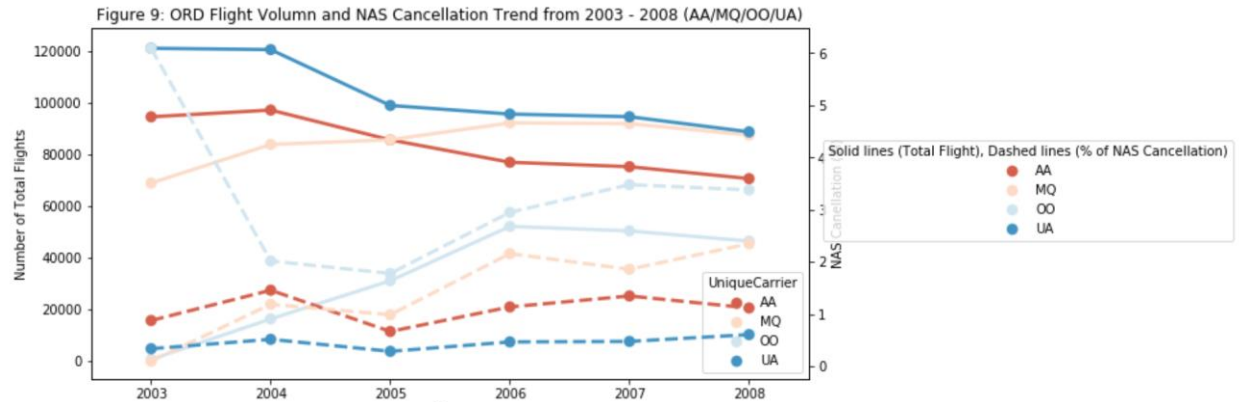
United and United Express take the entirety of Terminal 1 and part of Terminal 2, leaving Terminal 3 to American and other smaller airlines. Suggesting that Envoy and SkyWest do not get the same NAS resource as United and American.

Last, I would like to look at the trend of flight cancellation at O’Hare for American, Envoy, SkyWest and United between 2003 and 2008 to find out the NAS-cancellation situation a few years before.

Both United and American decreased the total number of outbound flights from O’Hare airport since 2003. However, Envoy and SkyWest increased the number flights. SkyWest increased the number flights for nearly 117 times, from 394 flights in 2003 to 46,316 flights in 2008, showing great determination to expand the business using the SkyWest strategy.

Figure 9 shows the total number of flights for the four airlines between 2003 and 2008. It also shows the percentage of NAS-cancellation rate, which did not change much for United and American. The rate increased for Envoy. The rate for SkyWest was the highest in 2003, mainly because there were only around 400 flights in that year. It dropped dramatically in 2004 but has climbed up steadily since 2005.

Overall, Envoy and SkyWest had higher NAS-cancellation rate than United and American. The trend shows the gap is large between the two groups. The fact that the NAS-cancellation rate of United and American had not had much change indicates O’Hare maintains NAS-scheduling priority for United and American, two of the mainline carriers in the U.S., over smaller carriers.



It is concerning to see the NAS-cancellation rate keeps increasing for Envoy and SkyWest. A high number of cancellation rate decreases the reputation for an airline and the management team must be aware of it. Is it possible to move more flights departing from other airports than O'Hare?

Possible solutions for avoiding flight cancellation are for other travelers to avoid booking Envoy or SkyWest flights from O'Hare, especially in January and February. Generally, to prevent flight cancellation, avoid leaving from O'Hare as much as possible since it has the highest flight cancellation.

References:

Wikipedia – www.wikipedia.com

Udacity Data Analyst Nanodegree – www.udacity.com

Pandas online specification – <http://pandas.pydata.org/>

Proof readed by my daughter, Adora Gao

Dataframe Reference

df_2008	all data in 2008
df_canceled	all canceled flights in 2008
df_canceled_origin	all canceled flights in 2008 with the originating airports
df_cancel_reason	all canceled flights in 2008 and the reason for cancellation
df_canceled_ord_nas	canceled flights from Chicago ORD airport due to NAS
df_canceled_ord_2008_month_carrier	canceled flights from ORD due to NAS and the month and airline information
df_heatmap	a temporary df to plot the heatmap of df_canceled_ord_2008_month_carrier
df_2008_ord	all flights from ORD in 2008
df_2008_ord_count	Count flights from ORD by airline
df_canceled_ord_nas_deptime	ORD canceled flights and their original departure time (UA/MQ/AA/OO only)
df_2008_ord_4	UA/MQ/AA/OO flights from ORD
df_ord_cancel_non	df_2008_ord_4 group by NAS cancel, Weather Cancel, Carrier cancel and "Not Canceled".
df_ord_cancel_non_temp	A temp table
df_ord_years	store the data for the final plot from six years within 1987 and 2008