DATA VISUALIZATION & NARRATIVES

FINAL REPORT

(TEAM: FLIGHT DELAY)

Table of Contents

Table of Figures

1. Introduction

Nowadays, traveling and shipping have a different definition compared with the last decades. Many countries' economies depend on aircraft in terms of tourism, shipping goods, or other flights connected to any country. However, this report will show some reasons for the delayed or cancelled flights and the possible reasons for that by presenting some visuals and finding some narratives that can lead to finding new insights. In addition, the data that will be used in the report is related to the US Department of Transport "DoT," and the report might show a covered angle for the strategy evaluators in US DoT as it can illustrate some patterns of the reasons of delay or cancel flights.

2. Data source & Data Preparation

The dataset was downloaded from the Kaggle website in a CSV format and stored in our local computer (the link is provided in Appendix). The dataset was loaded as a pandas data-frame to explore the quality of the data.

- The data was examined for outliers, Null values, and duplicates.
- Data type for all the variables was inspected.
- All the values were in a standardised format.
- Data does not contain any sensitive or confidential information.

a. Data description

- The data has 5,819,079 rows and 31 columns in total.
- Number of unique values for each column in the data-frame is shown in below figure
- size of the data is 592.4 MB.
- data contains 26 numerical variables and 5 object type variables.
- data has records only for the year 2015.

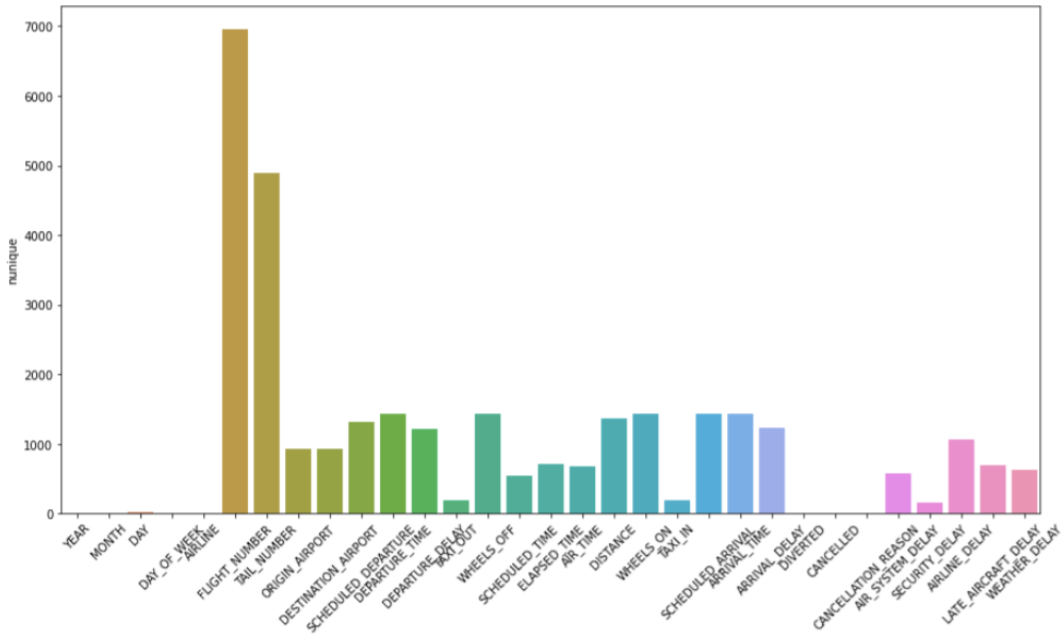


Figure 1. Number of Unique Values per Column

b. Data Dictionary

The airports dataset (airports.csv) contains the list of all US airports, including their location. The attributes of these datasets are:

Column Name	Data Type	Null Values	Description
IATA_CODE	Factor	0	A three-letter geocode that uniquely identifies each airport, for example, JFK for John F. Kennedy International Airport
AIRPORT	String	0	The official name of an airport
CITY	Factor	0	The city name where the airport is located
STATE	Factor	0	The two-letter code of the state where the airport is located, for instance, NY for New York
COUNTRY	Factor	0	Filled with 'USA' as the dataset only comprises of US airports
LATITUDE	Decimal	0	The coordinates of the airport in a decimal form
LONGITUDE	Decimal	0	

Table 1

The airlines dataset (airlines.csv) is the list of US airlines that serve domestic routes, that contains:

Table 1. Airline dataset data dictionary

Column Name	Data Type	Null Values	Description
IATA_CODE	Factor	0	A two-letter code that uniquely identifies each airline, for example, UA for United Airlines
AIRLINE	String	0	The name of an airline

Table 2

Flight dataset (flight.csv) is the list of flight trips of delayed or cancelled flights in the USA in 2015.

Table 3. Flight dataset data dictionary

Column Name	Data Type	Null Values	Description
YEAR	Factor	0	Year of the flight trip
MONTH	Factor	0	Month of the flight trip
DAY	Factor	0	Day of the flight trip
DAY_OF_WEEK	Factor	0	Day of week of the flight trip
AIRLINE	Factor	0	The IATA Code of an airline
FLIGHT_NUMBER	Factor	0	The unique code of a flight route

Column Name	Data Type	Null Values	Description
TAIL_NUMBER	Factor	14,721	The unique code of an airplane
ORIGIN_AIRPORT	Factor	0	The IATA code of the starting airport of the flight trip
DESTINATION_AIRPORT	Factor	0	The IATA code of the destination airport of the flight trip
SCHEDULED_DEPARTURE	HR:MM	0	Scheduled departure time of the flight trip
DEPARTURE_TIME	HR:MM	86,153	The actual departure time of the flight
DEPARTURE_DELAY	Integer	86,153	Total delay on departure in minutes
TAXI_OUT	Integer	89,047	The time duration elapsed between departure from the origin airport gate and wheels off
WHEELS_OFF	HR:MM	89,047	The time point that the aircraft's wheels leave the ground
SCHEDULED_TIME	Integer	6	Planned time amount needed for the flight trip
ELAPSED_TIME	Integer	105,071	AIR_TIME+TAXI_IN+TAXI_OUT
AIR_TIME	Integer	105,071	The time duration between wheels_off and wheels_on in minutes
DISTANCE	Integer	0	Distance between the origin and destination airport in miles
WHEELS_ON	HR:MM	92,513	The time point that the aircraft's wheels touch on the ground
TAXI_IN	Integer	92,513	The time duration elapsed between wheels-on and gate arrival at the destination airport
SCHEDULED_ARRIVAL	HR:MM	0	Planned arrival time
ARRIVAL_TIME	HR:MM	92,513	WHEELS_ON+TAXI_IN
ARRIVAL_DELAY	Integer	10,5071	ARRIVAL_TIME-SCHEDULED_ARRIVAL
DIVERTED	Factor	0	Aircraft landed on airport that out of schedule
CANCELLED	Boolean	0	Flight Cancelled Status (1 = cancelled)
CANCELLATION_REASON	Factor	5,729,195	Reason for cancellation of flight: A - Airline/Carrier; B - Weather; C - National Air System (NAS)/Air System; D - Security
AIR_SYSTEM_DELAY	Integer	4,755,640	Delay in minutes caused by National Aviation System (NAS) such as non-extreme weather condition, heavy traffic volume, etc
SECURITY_DELAY	Integer	4,755,640	Delay time in minutes caused by the security reasons, such as terminal evacuation, security breach, long lines more than 29 minutes at screening areas

Column Name	Data Type	Null Values	Description
AIRLINE_DELAY	Integer	4,755,640	Delay time in minutes due to any circumstances under airline company control, such as maintenance, fuelling
LATE_AIRCRAFT_DELAY	Integer	4,755,640	Delay time in minutes caused by late arrival of the same aircraft from the previous airport
WEATHER_DELAY	Integer	4,755,640	Delay time in minutes that caused by an extreme weather conditions (actual or forecasted) that delays or prevents the operation of a flight, for instance: blizzard, tornado, hurricane.

Table 3

c. Data cleansing

The dataset did not require much data cleaning. All the values in the dataset were in an acceptable range. There were no outliers present in the dataset and it did not contain any duplicate values. However, there were a few variables which had several NULL values. Despite these records containing a NULL value, these rows won't be deleted as the NULL value indicates that the flight was or wasn't delayed, cancelled, or diverted. For instance, the AIR_SYSTEM_DELAY column is NULL if the flight arrived on time, the TAIL_NUMBER column is NULL if the flight was cancelled and there are no aircraft assigned to the flight number yet.

3. Research questions

a. Who Are the Stakeholders?

Stakeholders are the strategy evaluators in the U.S. Department of Transport, and C-suit executives. They are the ones who approve policies and strategies to make sure the flyers / customers are happy with the services.

b. The Big Idea

Stakeholders wish to understand the trend of delayed and cancelled flights and the causes of such incidents in 2015. This analysis will help the stakeholders to explore if the changes in the policies at the start of 2015 have had any positive impact and to decide if any further changes need to be made in flight routes / times based on the various parameters.

c. What is the Issue?

The U.S. DoT understands that the most delays and cancellations occur due to weather but also want to understand other variables causing delay / cancellations along with some additional statistics to back their claims.

d. How will Data Visualisation Address this Issue?

Visualisations would allow strategy planners to understand the dynamics of flight delays and cancellation with respect to the various parameters, some of them are listed below-

1. Which airport has the most delays and how they are spread over the various delay causes?
2. What are the causes of delay?
3. What is flight cancellation statistics?
4. What are the crucial factors causing the arrival delay?
5. How season impacts the dynamics?

4. Literature review

Air travel plays a pivotal role in the global economy by connecting people and businesses around the world. Since it provides regional, national, and multinational connectivity, it can create economic opportunities such as tourism and local and international trades. In the end, it leads to job creation and drives economic growth. Air Transport Action Group (ATAG) estimates that the aviation industry generated USD 2.7 trillion and supported 65.5 million jobs globally in 2016 (ATAG, 2018). ATAG also predicts that aviation will create 98 million jobs and contributes 5.7 trillion to the global economy in 2036 (ATAG, 2018).

The tourism sector relies on aviation industries, where air transport can provide an affordable and faster means for the tourists to reach the destinations. Statista recorded a 1.49 billion international tourist arrival in 2019 which more than half travelled to the destination using air transport (Statista, 2021). Several countries are highly dependent on tourism to boost their economies, such as the Maldives, where tourism contributes ~39% to their GDP, Macao with 28%, and Aruba with 28% (World Bank, 2021). The disruption in the aviation industries will impact these countries' tourism industries and, in the end, their economy.

Flights delay and cancellations are common issues that affect various entities such as the passengers, the airlines, and the airports. It increases passengers' travel time, disrupts their schedule, and the inconvenience may lead to passengers' stress. US Department of Transportation publishes a weekly report regarding air travel consumer complaint, where most of the complaints from the airline passengers are flight cancellation and flight delay(U.S. Department of Transportation, 2020). Research found that the flight delay can raise the passenger's dissatisfaction's level and can lead to their decision to use another airline (Mazzeo, 2003).

On the other hand, it also costs the airlines an additional expense, such as passengers compensation, extra fees for holding the plane on the tarmac. Study shows that flight delays cost reaches USD 30 billion in the US alone in 2007 where USD 8.7 billion is attributed to the airline cost (Ball et al., 2010).

5. Results & Visualisations

This section is divided into two:

Part a: comprises of the visuals for the strategy makers to deep dive into the data who take decisions about changes in the transport plans and work with Air Traffic Controllers to make sure the delays and cancellations are minimised.

Part b: is for the higher management and C-suit management for a high-level view of the air transport situation.

Part 'a' answers question 1 and 2, whereas part 'b' answers question 3 and 4. Both the sections also cover more visuals for additional analysis.

a. Exploratory Dashboard for Strategy-makers

The Flights Distribution

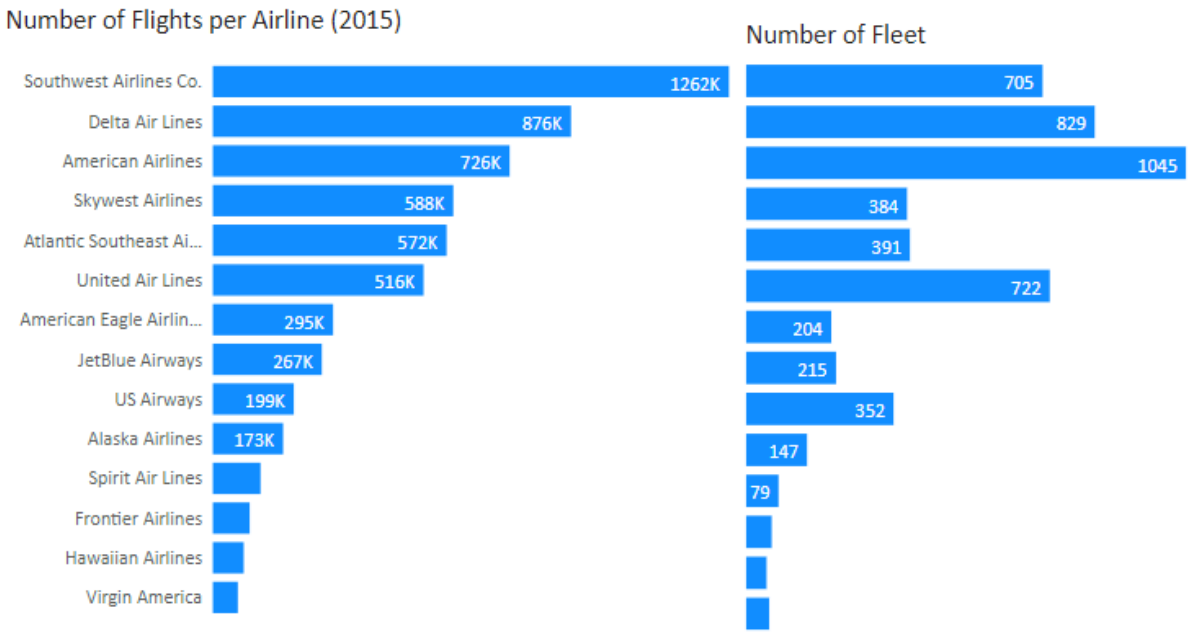


Figure 2.Number of Flight and Fleet per Airline (2015)

The plots above illustrate that Southwest Airlines, Delta Airlines, and American Airlines are the largest airlines in the US based on domestic flights. These three companies alone control ~50% of domestic flights in the US. However, despite serving the most extensive domestic route, Southwest airlines only rank 4th in the number of aircraft operated, behind American Airlines, Delta Airlines, and United Airlines.

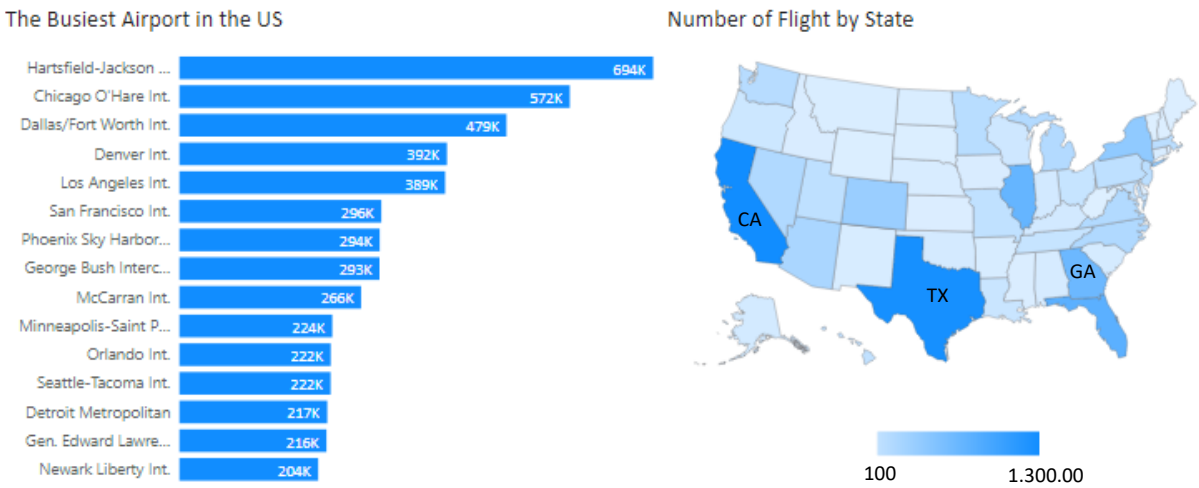


Figure 3. Number of Flight Served per Airport and State

On the other hand, Hartsville-Jackson in Atlanta, Georgia (GA), was the busiest airport where it served almost 700.000 incoming and outgoing domestic flights per year. Apart from being Delta Airlines' Hub, Atlanta is a strategically located city where more than 80% of the US population can reach Atlanta only by two hours flight (Atlanta International Airport, 2021). On the state-wise, Texas (TX) and California (CA) are the states where the airports served the domestic flights the most. Each state has two busiest airports on the list; Texas has Dallas/Fort Worth and George Bush Intercontinental (Houston), while Los Angeles and San Francisco Airport are in California.

Flight On-Time Performance (OTP)

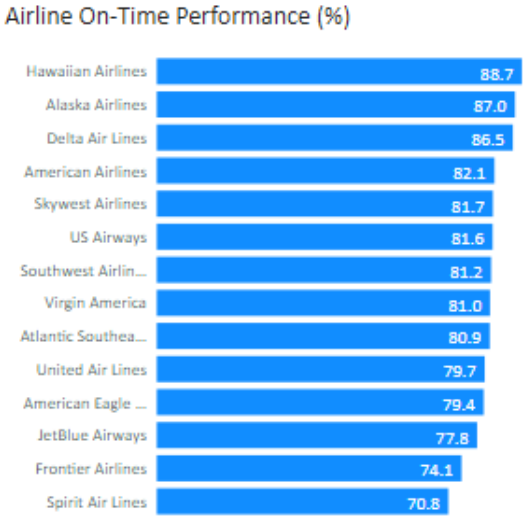


Figure 4. Airline On-Time Performance on Domestic Flights in 2015

The On-Time Performance (OTP) measures the ability of an airplane to arrive on time. The chart above demonstrates that Hawaiian Airlines, Alaska Airlines, and Delta Airlines are the best airlines based on OTP, where more than 85% of their flights arrived on time in the destination airport.

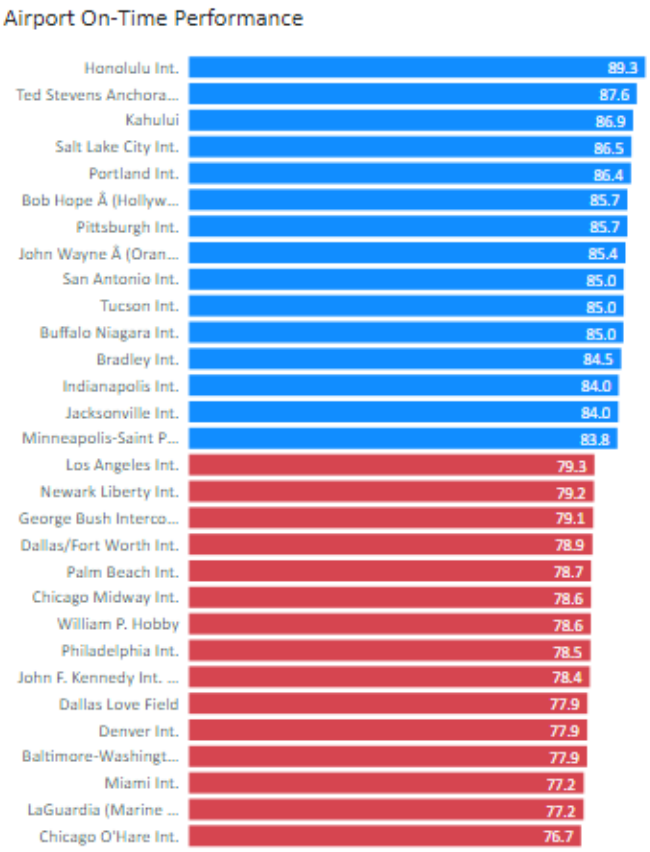


Figure 5. Airline On-Time Performance on the Domestic Flights in 2015

The OTP is also applied to the airports where it is determined by the on-time departures of an airplane. The figures show that, for the airport with more than 15.000 departures per year, Honolulu International Airport has the best OTP. In contrast, several busiest airports in the US, such as Los Angeles Airport, George Bush Intercontinental in Houston, Texas, and Dallas/Fort Worth in Dallas, Texas, are the worst airports based on their OTP.

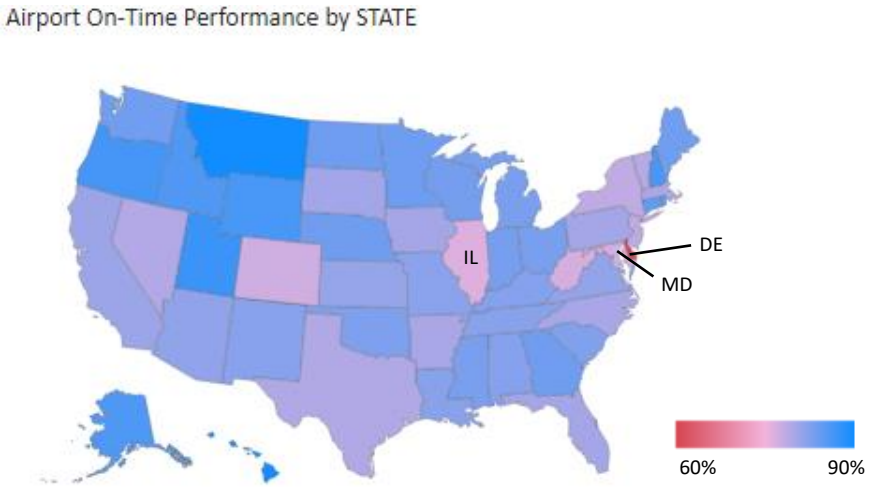


Figure 6. State-wise Airport On-Time Performance

Illinois (IL) and Maryland (MD) are the states with the lowest OTP, 77% and 78%, respectively. The Chicago O'Hare Airport, the worst-performing airport in the US, is in Illinois, while Maryland has Baltimore-Washington Airport in the bottom 4.

From Figure 6, Delaware's (DE) OTP is 63%. However, the only airport in the state, the Wilmington Airport, only serves 100 flight from January to June 2015 before Frontier Airlines cease its operation in the airport in 2015 (Goss, 2013).

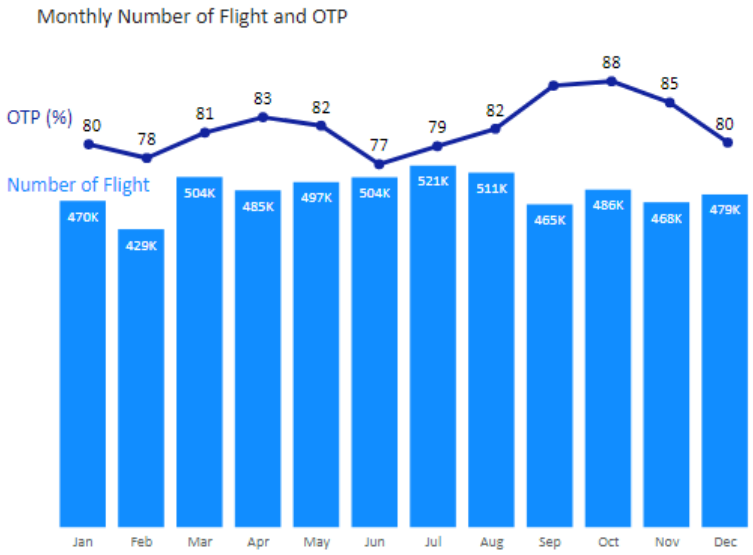


Figure 7. Domestic Flight Overall On-Time Performance

The plot above shows that the number of domestic flights from June to August is the highest compared to the other. The summer season, which happens from June, drives the airline companies to add their flight route to meet passengers' increasing demand to travel during that period. However, the domestic flights OTP reached the lowest in June and February 2015, 77% and 78%, respectively. It is interesting to explore why February's OTP is low despite having the lowest number of flights.

Flight cancellation statistics with respect to weekday

It looks like the maximum number of cancellations happened on a Monday which would impact the travellers significantly since it is the first working day of the week.

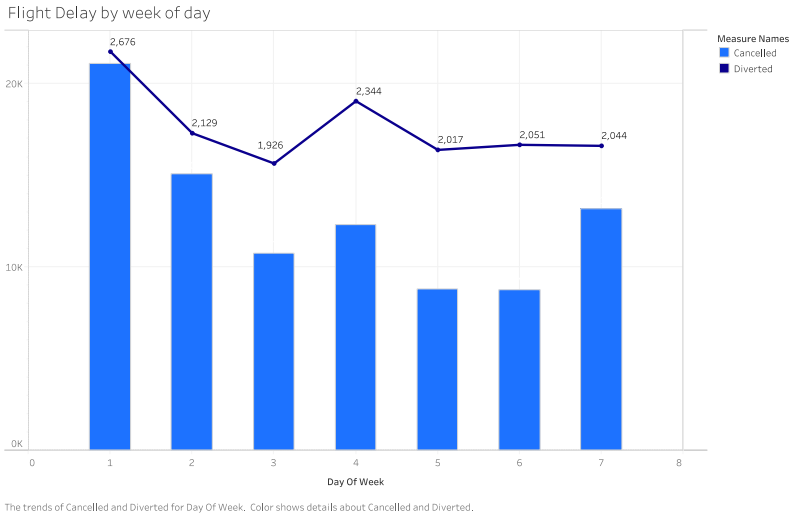


Figure 8. Flight cancellation statistics with respect to weekday

Flight Delay Causes

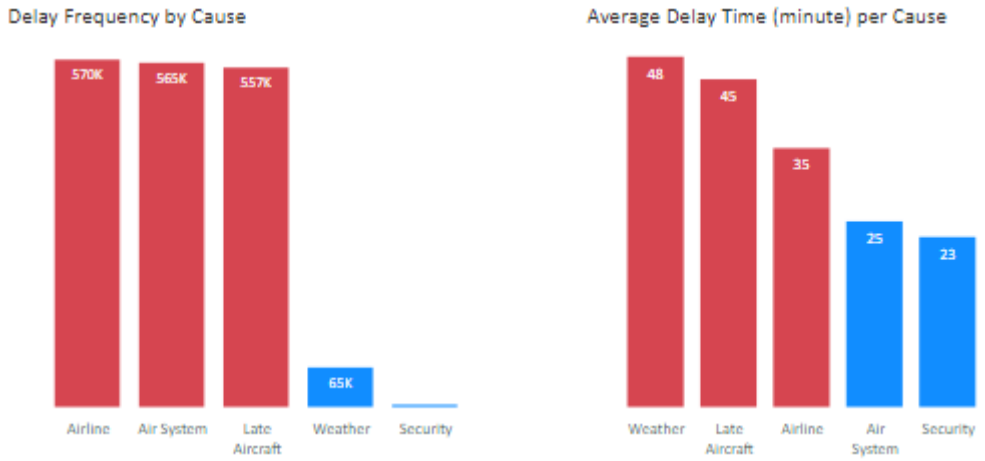


Figure 9. The number of delays and average delay time by cause

The plot above illustrates that airline (internal air carrier issue), air system (national aviation system issue), and late aircraft are the most frequent delay causes, leading to more than 500.000 delays in 2015. It should be noted that one or more factors can cause delays for a single flight. For instance, flight AA1563 from Tulsa Airport (TUL) to Dallas/Fort Worth Airport (DFW) was late for 113 minutes caused by multiple factors as shown below.

Flight Number	Air System Delay	Airline Delay	Late Aircraft Delay	Weather Delay	Total Delay
AA 1563	32	16	3	62	113

Table 4. Flight delay records

Even though extreme weather rarely causes flight delays, the weather causes a longer delay time compared to the other causes, where on average, it pushes the flights to take off 48 minutes late.

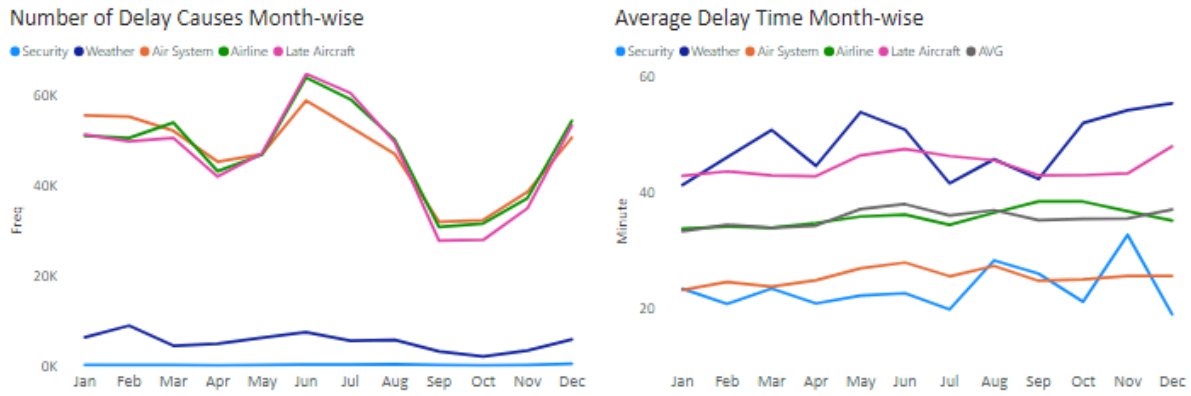


Figure 10. Monthly Delay Frequency and Average Delay Time (minutes)

June has the highest number of delays and the average delay time. These findings align with Figure 7, where the domestic flight's OTP reaches its lowest in the same month. Besides that, there is an upward trend of delay frequency from October to December leading to a significant drop in domestic OTP.

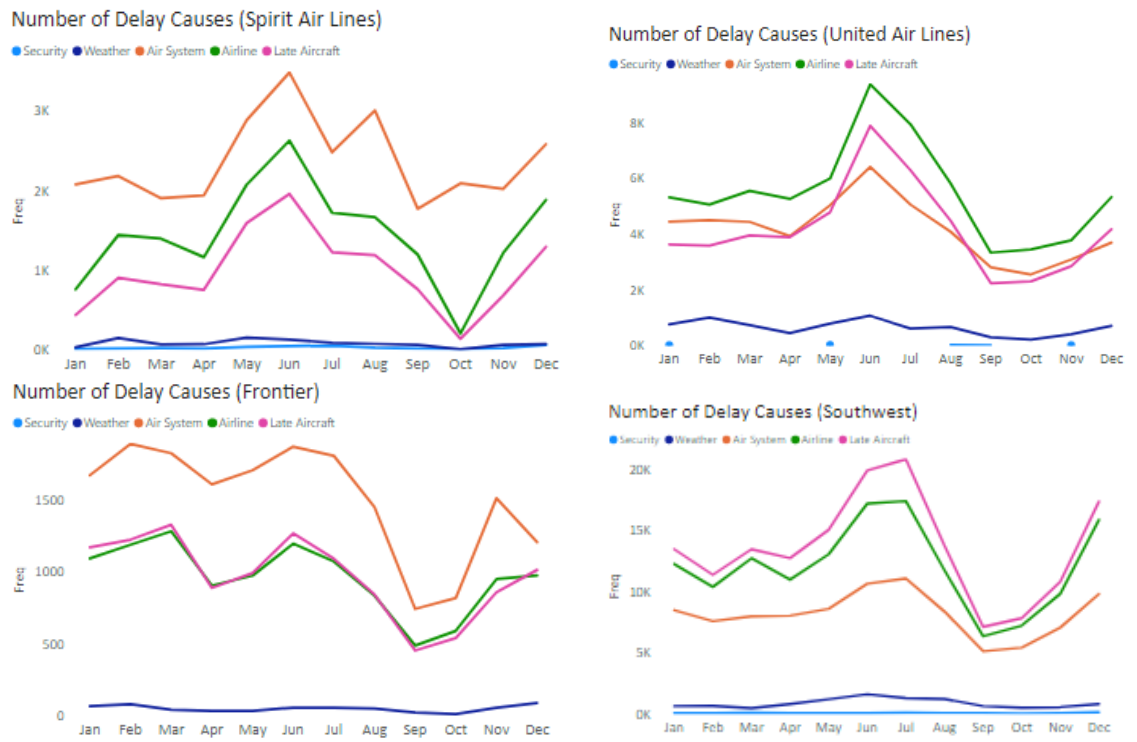


Figure 11. Delay Causes by Airline

Looking into the individual airlines, especially Spirit and Frontier (low-cost carrier and has the worst OTP), United (a full-service airline with the lowest OTP), and Southwest (the busiest airline), we found that each air carriers have different major delay causes. Air system (National Aviation System) is the leading delay cause for Spirit and Frontier, while internal United issues and Southwest late aircraft arrival are causing more delays in the respective airline than the other factors.

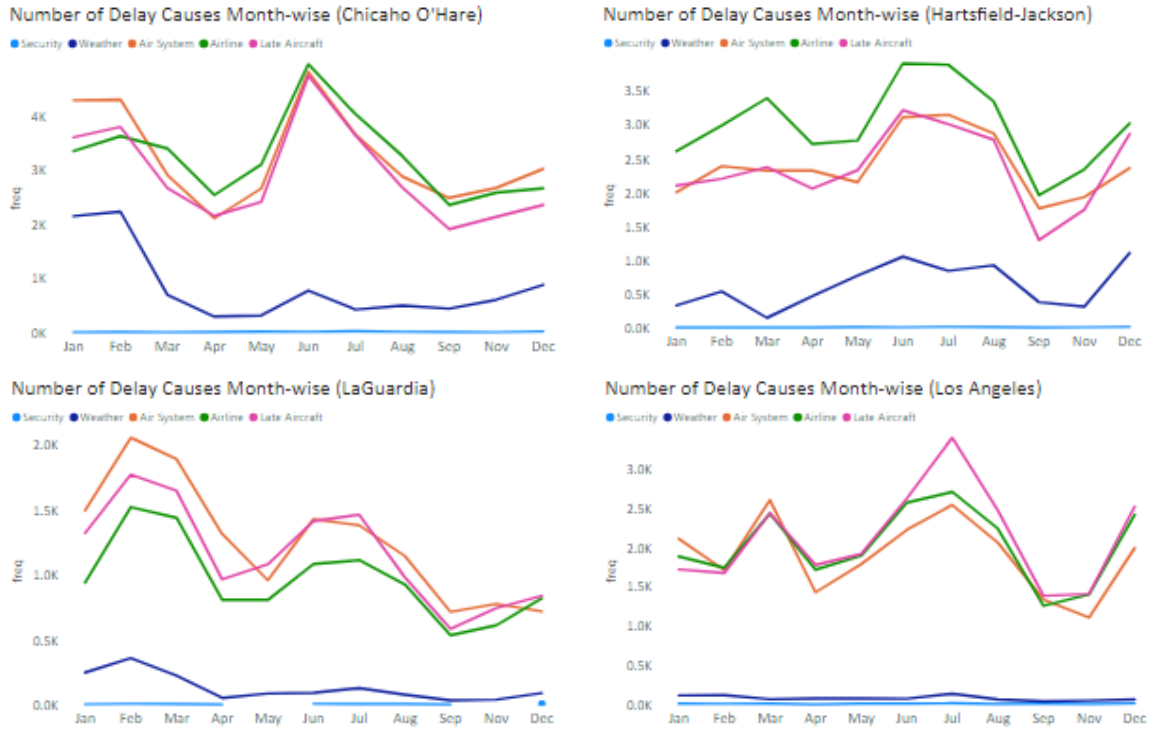


Figure 12. Delay Causes by Airport

Each airport also has different main delay causes. For instance, most of the delays in Chicago and Hartsfield airport were due to circumstances under the airline control, while the air system is the major delay cause in LaGuardia airport (New York). Interestingly, the peak of delay frequency varies among the airports; for example, in Chicago, delays are more frequent in February and June, while LaGuardia happens in February and March.

Flight Cancelled Causes

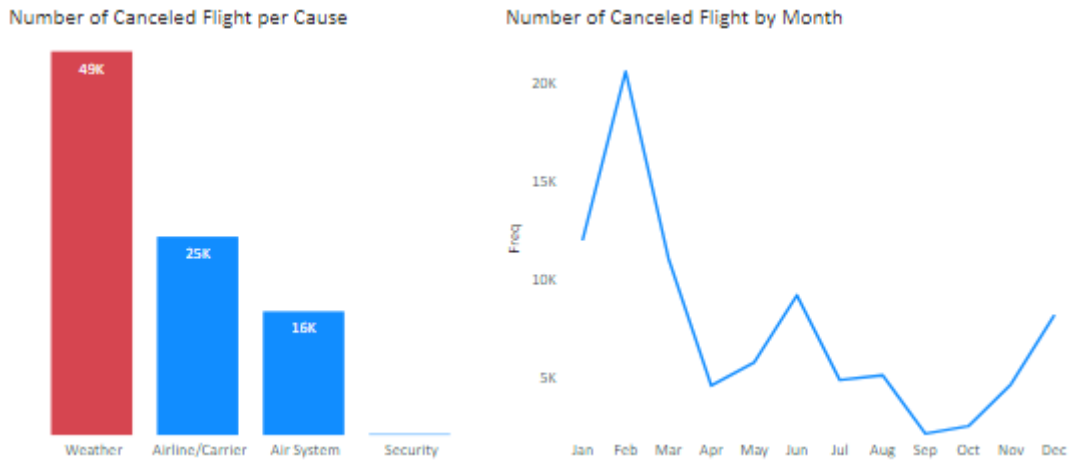


Figure 13. Flight Cancelation Frequency by Cause and Month

Extreme weather is the main factor in flight cancelations, where it accounts for more than 50% of all cancelled flights. On the other hand, the number of cancelled flights in February is significantly higher compared to the other months.

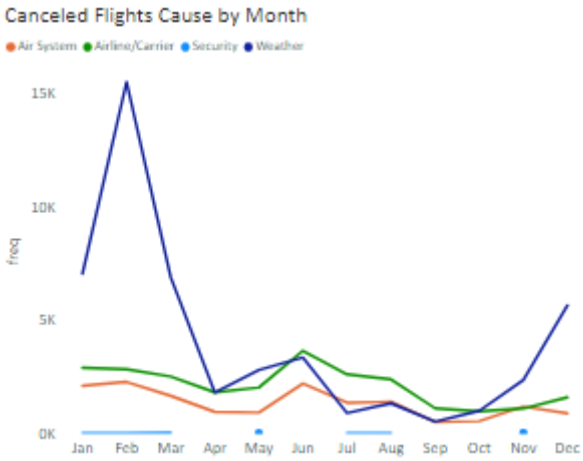


Figure 14. Monthly-wise Cancelled Flight Causes Frequency

It was the extreme weather that leads to severe flight cancelations in February 2015. As it was reported, there is a series of winter storms hitting across the US – especially in the northeast part of

the country – from the end of January to early March (Fritz, 2015; Taylor, 2015). There are around 28.000 flights cancelled due to extreme weather conditions from January to March alone.

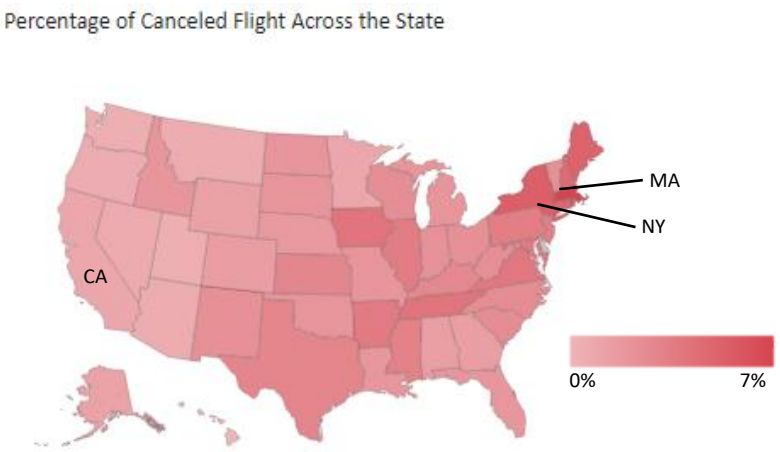


Figure 15. Percentage of Flight Cancelled related to extreme blizzard in January – March 2015

The plot above shows that the blizzard impacted the airports in the northeast the most, such as Massachusetts (MA) and New York (NY). From Figure 12, there is a sudden jump in the number of flight delays caused by the extreme weather in the LaGuardia Airport located in NY, which high likely caused by the North American blizzard. On the other side, it doesn't significantly impact the western region, such as California, where the number of cancelled flights is minimal, and there is no significant jump in flight delay in the Los Angeles airport, as can be seen in Figure 12.

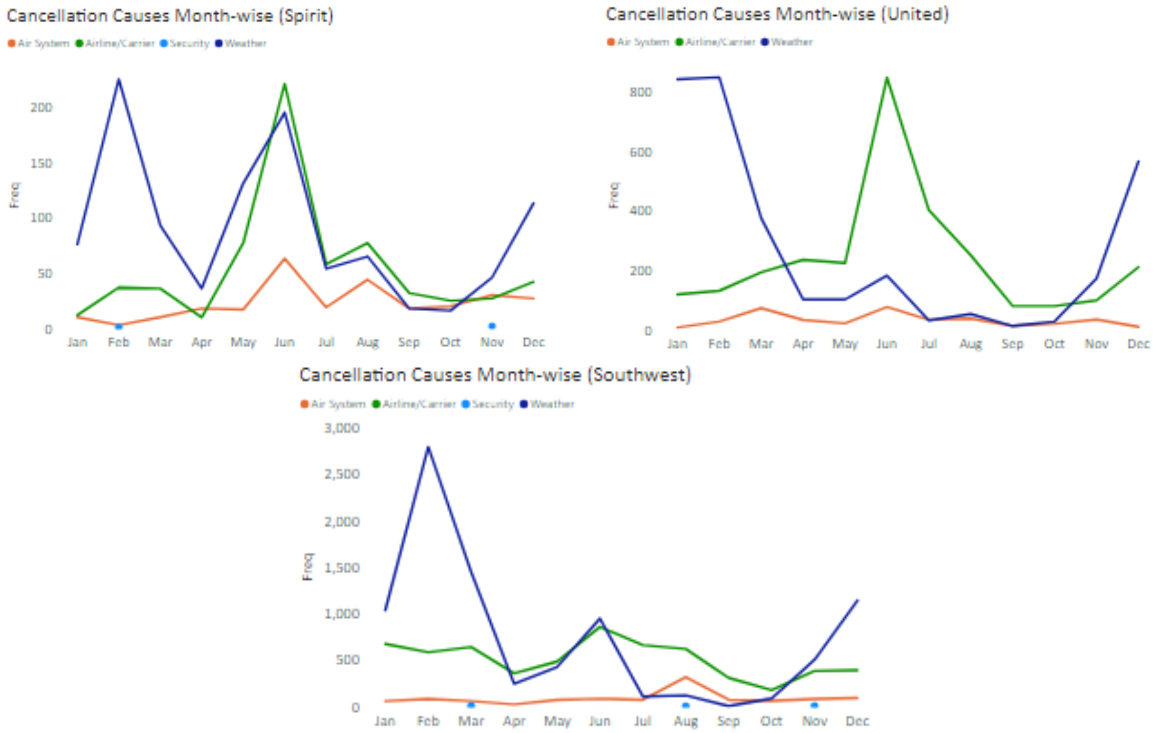


Figure 16.Monthly Flight Delay Causes per Airline

Besides the extreme weather-related cancelations around February 2015, there is a significant increase in flights cancelled in June, especially in United and Spirit Airlines. Phys reported that in June 2015, United experienced a computer glitch two times in just over a month (Koenig & Mayerowitz, 2015) that causes United to cancel more than 800 flights in June alone. This glitch is not their first-

time encounter for United, where the same article reported that they have struggled with the same outage since 2012 after migrating to a new system following the merge with Continental Airlines. From Figure 12, Chicago O'Hare airport, the largest hub for United Airlines, experiences a high flight delay that high likely triggered by United's computer outage.

The visuals are published on UTS PowerBI server – Link to Visuals

b. Dashboards for C-Suit

Higher management may not always be interested in all the details about flight delays and cancellations but would always like to have an overview. Without giving out a lot of details, below dashboards give them a snapshot of top contributors and the trend for cancelled and delayed flights.

The dashboards are published on UTS PowerBI server – Link to Dashboard

All the dashboards are interactive and give details when cursor is hovered over the graph elements. If any value on the graph is clicked, the other graphs get filtered based on the selection.

There are various filters available on all the dashboards as shown in below image-

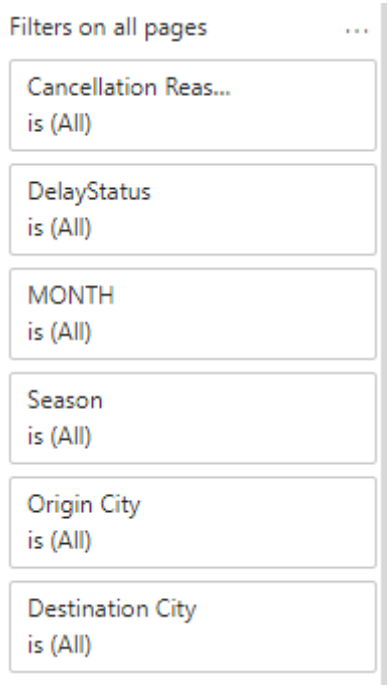


Figure 17: Page level filters on Dashboards

Dashboard 1 – Overall Flights Analysis

Below dashboard image shows overall trend in terms of the number of flights and top contributors in Airline, Airports and Flight numbers. This dashboard also gives trend of scheduled flights over the period of 2015 along with Day of week analysis.

Summer has the most flights scheduled, and monthly trend shows that the most flights were scheduled between March and August. It also shows that there were significantly lower number of flights scheduled in February.

Day of week analysis shows that the most flights are scheduled during weekdays and Saturdays have lowest number of flights on average.

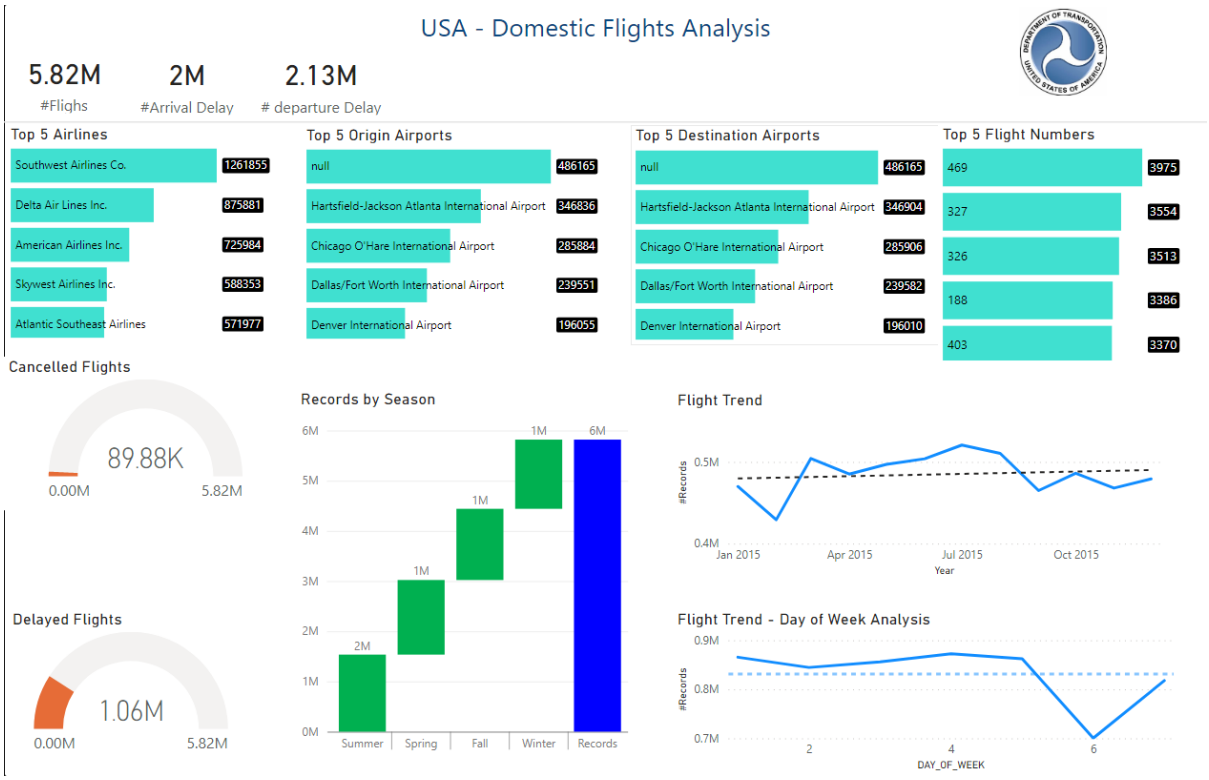


Figure 18: USA Scheduled flights Analysis.

Below image shows an example of the filtered Dashboard 1 – with filter on the Origin City as "New York". In this filter it is very clear to see the difference in Flight trend. Day of week analysis seems close to the overall flight schedules. Unlike overall flight schedule, New York has most flights scheduled in Spring.

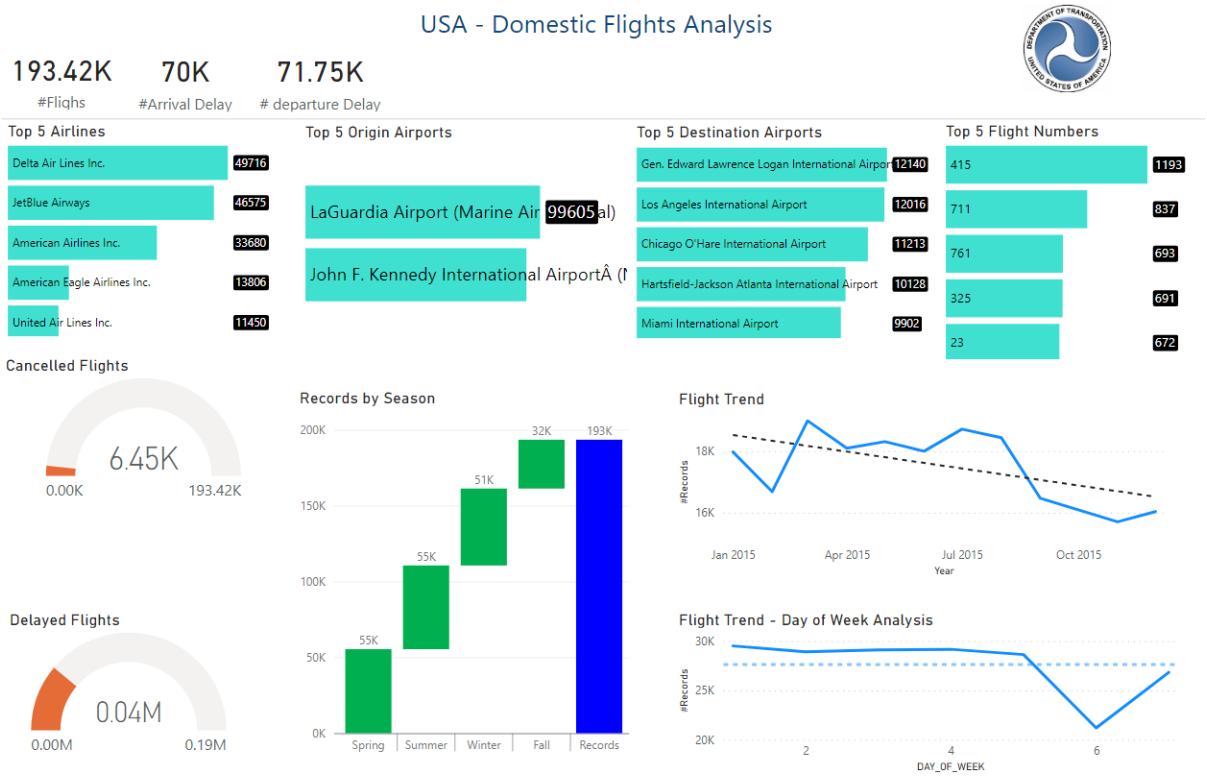


Figure 19 USA Scheduled flights Analysis – Filter on Origin City "New York".

Dashboard 2 - Analysis of Cancelled flights (Answers Research Q3)

Below image of cancelled flights dashboard shows the total number of flights cancelled, number of airlines for which the flights were cancelled, number of origin airports and number of destination airports along with below graphs-

- Top 5 airlines for which the flights were cancelled.
- Top 5 origin airports from where the flights were cancelled.
- Top 5 destination airports where the cancelled flights were supposed to reach.
- Top 5 flight numbers
- Distribution of flight cancellation based on the cancellation reason.
- Distribution of flight cancellation based on season.
- Trend of flight cancellation– hierarchy
- Day of week analysis

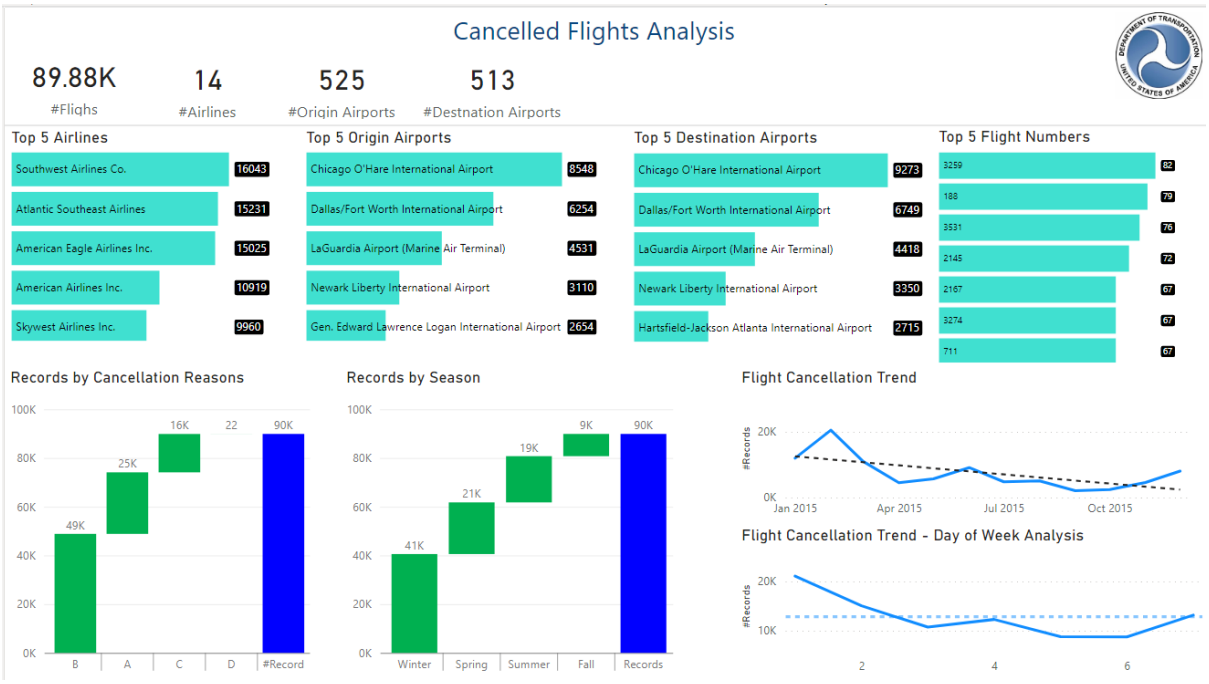


Figure 20 Analysis for cancelled Flights

As per the above dashboard, the most flights are cancelled on Monday, but the flight cancellation number is below average during other weekdays unlike the number of flights scheduled. Additionally, February Month shows significantly lower number of flights scheduled as compared to January and March but in cancellation statistics, February has higher number of flights cancelled as compared to January and March. Winter has the most flights cancelled, though this ranks the last in number of flights scheduled.

Below image of Dashboard 2, shows an example of filtered cancelled flights dashboard. The topmost airline in cancelled flights is selected and all the visuals on this dashboard have got filtered based on the selection.

It is easy to understand different dynamics with respect to the specific airline chosen. It is important to note that the dynamics would highly depend on the flights scheduled for relevant airline. Below image is an example / representation of a within dashboard filter to demonstrate its usefulness.

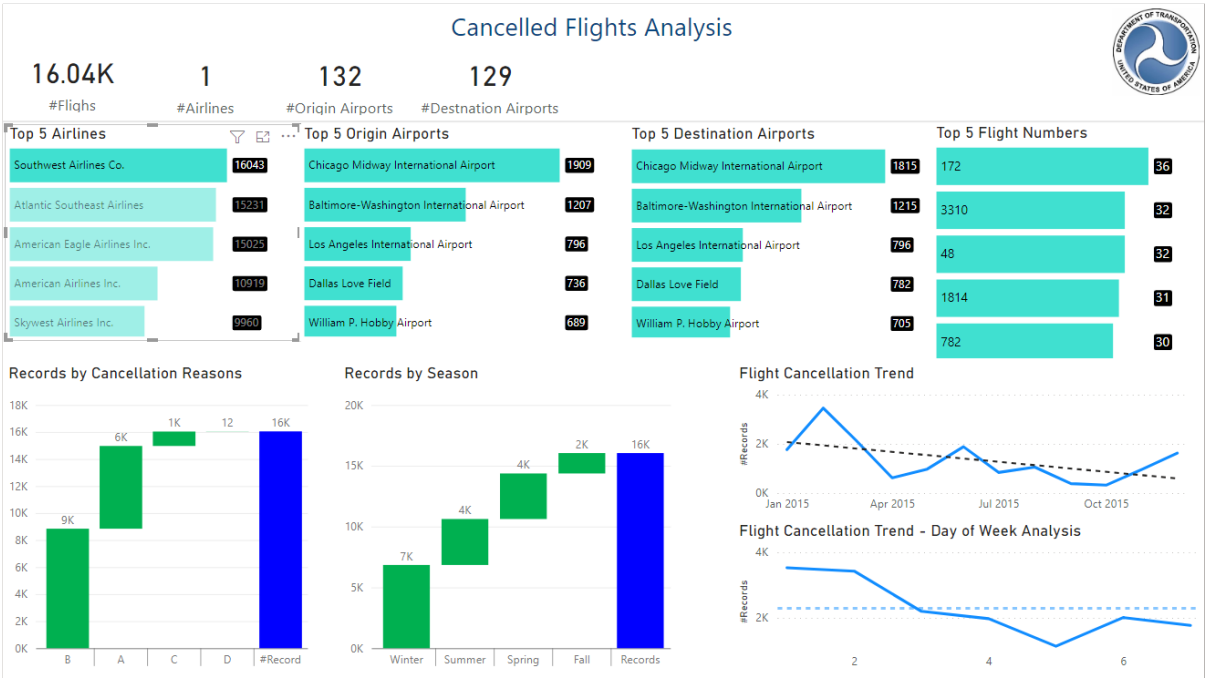


Figure 21 Cancelled flights - Example of filter based on selection of a top airline in first graph.

Dashboard 3 - Analysis of Delayed flights (Answers Research Questions 1,2, 5)

Below image of delayed flights dashboard shows the total number of flights delayed, number of airlines for which the flights were delays, number of origin airports and number of destination airports, number of flights by cause of delay, average delay in minutes by cause of delay, along with below graphs-

- Top 5 airlines for which the flights were delayed.
- Top 5 origin airports from where the flights were delayed.
- Top 5 destination airports for delayed flights.
- Top 5 flight numbers.
- Distribution of flight delay based on season.
- Trend of flight delays – hierarchy

Day of week analysis

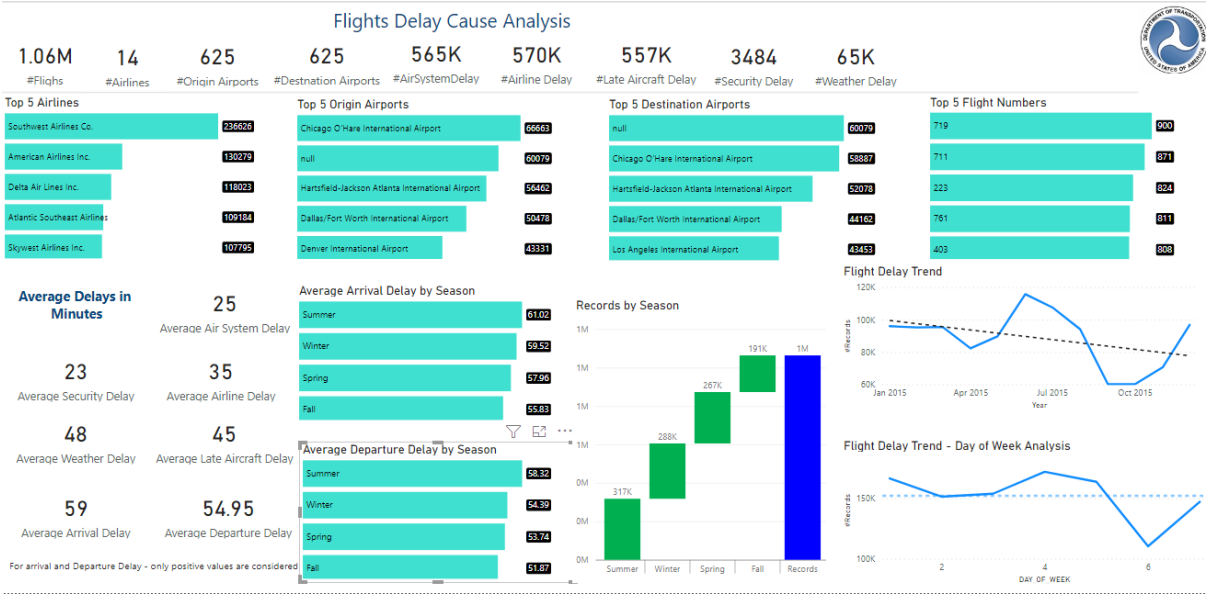


Figure 22 Analysis for Delayed Flights

Above dashboard shows that the summer season and June month has highest flight delays. Fall has the lowest average delay for arrival as well as departure. Out of all the causes of arrival / departure delay – Late aircraft delay has highest average delay and security delay has the lowest.

As per the examples in earlier two dashboards, this dashboard can also be filtered on the page level filters or by selection of a value on the dashboard.

Below images of filtered dashboards show detailed filtered on top Origin airport and top destination airport (Answer to research question 1):

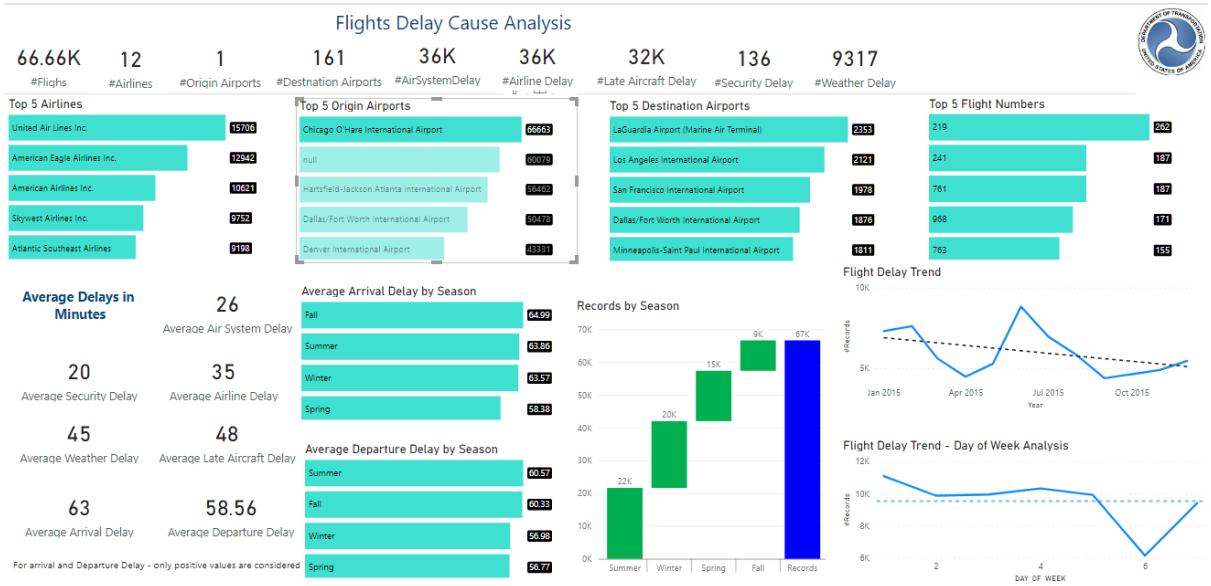


Figure 23 Analysis of Delayed Flights - Filtered on Top Origin Airport

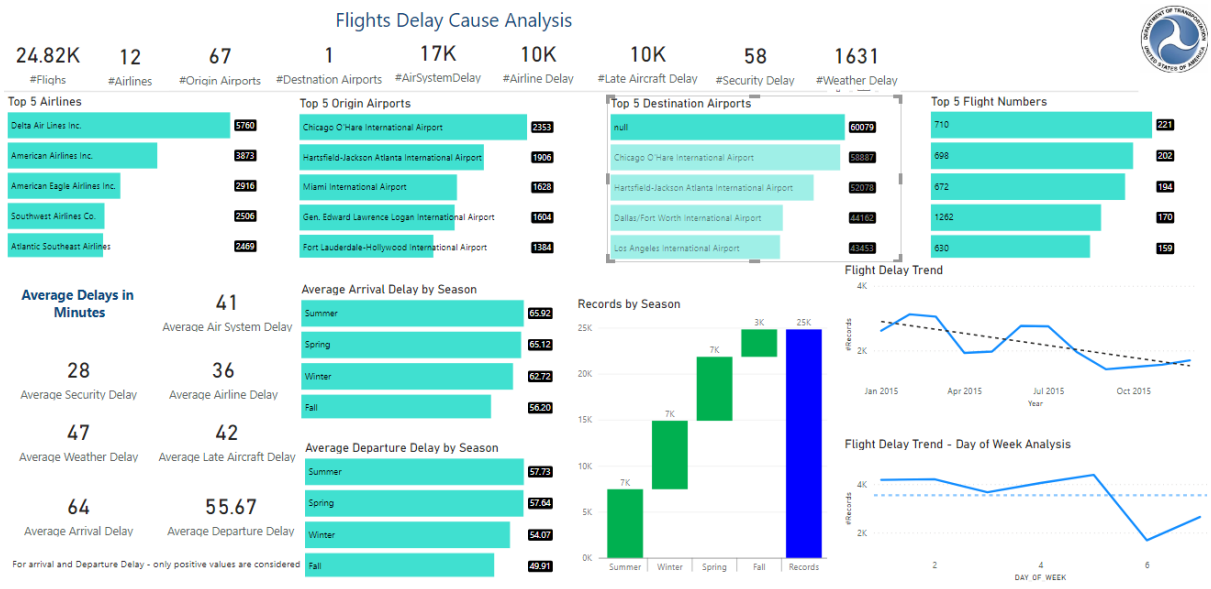


Figure 24 Analysis of Delayed Flights - Filtered on Top Destination Airport

Dashboard 4 – Factors contributing to Arrival Delay. (Answer to Research Question 4)

Below dashboard shows how the arrival delay is correlated with departure delay and various delay causes. Arrival delay and departure delay are highly correlated. Weather delay, Airline delay causes have positive correlation with arrival delay. Page level filters provide more in depth analysis on the season wise impact.

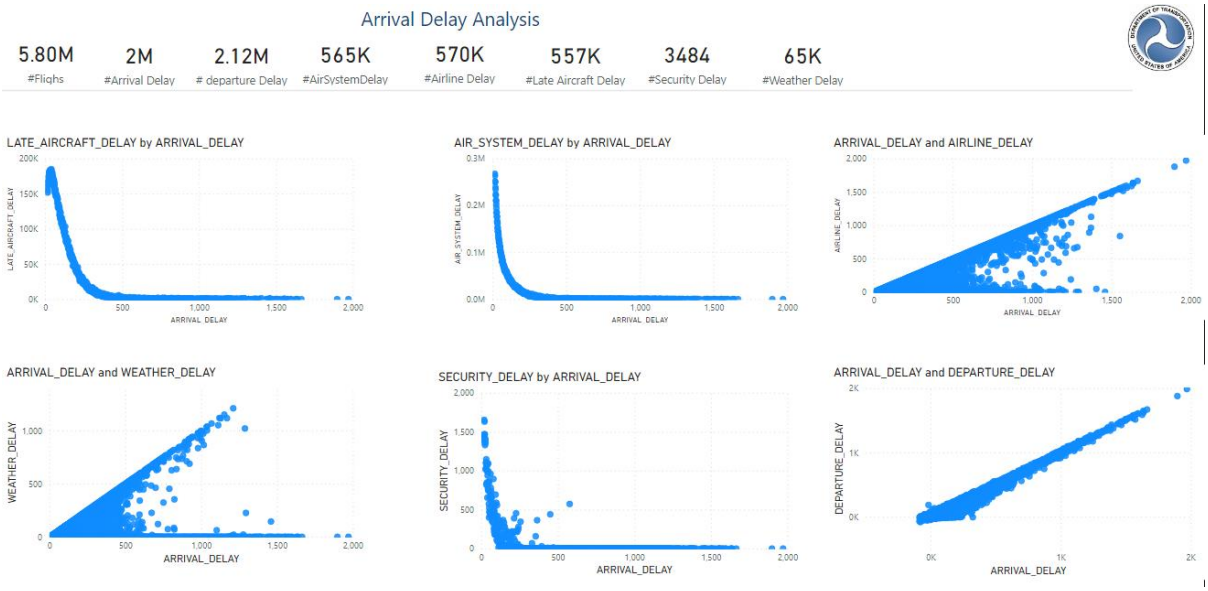


Figure 25 Analysis of Delayed Flights - Filtered on Top Destination Airport

6. Discussion

The analysis performed above helped us get a better understanding of the factors affecting flight delays. We also explored what percentage of the airlines control the greatest number of domestic Flights in the US, which airports were the busiest and what is the traffic of incoming and Outgoing flights per year. We also focused on which airlines arrived on time and what time of the year had the lowest On Time Performance. More research is being conducted to explore the reasons for these delays.

Flight delay is a multifaceted issue (multiple factors can cause the flight delay), despite airline, air system, and late aircraft the main factor in causing the flight delay and weather causing the longest delay time. For instance, one aircraft was delayed because of the weather/airline, but since the aviation system is a network and the late aircraft has a ripple effect, it affects the subsequent flight to also be delayed. We also came across a few instances where the cause of delay was not due to just one factor, but due to multiple factors that added up and resulted in a total delay of 113 minutes. In such cases, it can be challenging for the airlines to perform on time. However, with detailed analysis of the causes for aircraft delay, the airlines can have a better understanding on which of these factors can cause the maximum delay and focus on fixing that issue before moving on to other causes.

While considering the performance of airlines based on months, it was discovered that June had the highest number of delays and the average delay time. An assumption is that the delay could be due to June being the month of peak winter in the United States. Another interesting fact is most delays in June is with domestic flights. A reason for which could be the weather

being almost similar in a few states, hence if the weather is bad in the origin, it can be bad in the destination region as well.

One of the analysis shows that the low-cost airlines have the maximum delays. Low-cost airlines might not have the best engine performance to function effectively during bad weather conditions, at times they might not have the luxury to replace it with another flight to avoid flight delays. In a few other cases, low-cost airlines might not be able to pay their crew on time which could also affect the performance of these airlines.

Throughout the analysis we have tried to explore the different causes for flight delays. Based on this analysis, a group of dashboards was created which provides a quick reference for higher management to find all the necessary information which would help them understand the performance of the airlines and take precautionary measures. Each dashboard represents a specific point of interest and provides complete information about the causes for delays and cancellations.

7. Recommendations

This report has provided an exploration of the delayed and cancelled flights and illustrated the reasons behind that. After analysing the data, we find that some actions can help the strategy evaluators to reduce the chance of the deletion or cancellation of the flights:

- The internal airline is one of the significant issues in flight delay/cancelled. Thus, the authority may need to push the industry to provide a more reliable system to avoid any delay or even cancellation.
- Providing a simulation flight system that can predict the flights that can be delayed or cancelled can increase the satisfaction of the customers as they can benefit from this system for their plans.
- Further analysis on Time of Day will give deeper insights on strategy planning. Though the times would revolve around the demand by passengers, it would be wise to revisit the frequently cancelled or delayed flight schedules.
- More data needs to be added to perform additional time series analysis for better strategy planning.

8. Reflection

Team Dynamics:

The team worked well, had no conflicts, and used divide and conquer technique to complete the tasks.

At the start we came up with 4 datasets and one of them was chosen by voting. We all started exploring it and came up with some basic visuals which are shared in the report. That helped in formulating our research question.

We came together on zoom calls and WhatsApp group for brainstorming what we can do / need to do and then distributed the work. These discussions made sure that everyone was on the same page and to come up with better ideas for visualisation and data stories. We had created an online document where we all could work collaboratively at the same time and could review each other work.Overall,

our communication and collaboration plan helped us to do better and keep healthy relationship within the team.

Different perspectives during discussions helped to expand thinking.

Team Tasks:

- Introduction was led by Khalid.
- Data dictionary was written by Irfan (Airline and Airport) and Siddhi (Flights).
- Data Description and Data Cleaning was written by Priyanka.
- Research question and its' subparts were written by Siddhi.
- Literature Review was written by Irfan.
- Results and visualisations: Part a developed and written by Irfan, visual 11 by Priyanka and Part b developed and written by Siddhi.
- Discussion, Recommendation was written by Khalid and Priyanka.

Learnings for DVN:

We chose a dataset which is more relevant for prediction and machine learning. We could have gone for some dataset specific for data visualisation to get beautiful visuals. But this dataset helped us to learn DVN better by allowing us to take some extra efforts to share a better story.

Personal Learnings:

Over the period, this assignment helped me to correlate various visuals and experiment PowerBI as a visualisation tool. I could create dashboards as shared in part b of the section 'Results and Visualisations' based on the learnings from DVN subject and a recent PowerBI course that I completed on LinkedIn learning.

If I were developing this dashboard before DVN subject, I would have probably added borders to each graph within the dashboard, but the lectures helped me to understand various ways of removing clutter. I have tried to remove duplicate data - have made use of data value labels or axis value labels and not both for showing the details. I probably would not have understood that using both is a clutter without DVN subject. I understand that there is still some clutter in the dashboards - including in the labels and titles, but in the limited time, I have tried to do my best.

I feel the dashboards that I have created look clean, consistent and I have also tried to cross check them with an aviation SME to see if they are useful. His feedback helped to improve it a little further.

What could have been done differently:

- We could have explored some more datasets which were specific for data visualisation practice before finalising the flights dataset.
- We could also have chosen the correct colour theme for consistent visualisations for part a and part b - we missed on this but otherwise tried to be consistent
- For progress report - the team confirmed in the WhatsApp group that everyone had finished their part so based on that I formatted the report, downloaded it, and submitted the assignment well before time. Later after a while everyone started correcting some errors in

their sections of the report. For final submission, we tried to handle it better with clear communication.

What would I do differently (Specific to the time frame between progress report and final report) -

- I feel Irfan and I completed most visuals by 22nd May, most of the narrative part for discussion and references was left for the other two group members. Next time, I would prefer to have more frequent status calls to divide work evenly between visuals and narratives to give everyone chance to learn everything.
- About the part b dashboards that I created - Limited time did not allow me to make changes to all the graphs, I would spend a little more time in improving titles and other visual effects to declutter it further. I would also add some notes to explain certain visuals (Need to figure out a way to add meaningful notes when the graphs are filterable and may have different information and insights based on filters). I would add more details in tooltips for better cross comparisons of the visuals.
- More detailed group interactions with SME at the beginning to understand the real problems would help in improving it further and to make it more meaningful.
- I would like to change the Recommendations section a bit, but I did not want to be too critical in the work contribution by others, but I would like to improve points 1 and 2 in Recommendations.

9. References

ATAG. (2018). *Aviation Benefits Beyond Borders*. Retrieved from https://aviationbenefits.org/media/166344/abbb18_full-report_web.pdf

Atlanta International Airport. (2021). Atlanta International Airport (ATL) Fact Sheet. Retrieved from https://www.atl.com/about-atl/atl-factsheet/

Ball, M., Barnhart, C., Dresner, M., Hansen, M., Neels, K., Odoni, A., . . . Zou, B. (2010). Total Delay Impact. Study A Comprehensive Assessment of the Costs and Impacts of Flight Delay in the United States. *Nextor*. Retrieved from https://rosap.ntl.bts.gov/view/dot/6234

Fritz, A. (2015). Boston clinches snowiest season on record amid winter of superlatives. Retrieved from https://www.washingtonpost.com/news/capital-weather-gang/wp/2015/03/15/boston-clinches-snowiest-season-on-record-amid-winter-of-superlatives/

Goss, S. (2013). Retrieved from https://www.usatoday.com/story/todayinthesky/2015/09/21/airline-less-delaware-airport-hopes-ad-campaign-boosts-image/72553788/

Koenig, D., & Mayerowitz, S. (2015). United Airlines hit by 2nd massive grounding in 2 months. Retrieved from https://phys.org/news/2015-07-glitch-grounds-flights-hour.html

Mazzeo, M. J. (2003). Competition and Service Quality in the U.S. Airline Industry. *Review of Industrial Organization, 22*, 275-296. Retrieved from https://www.kellogg.northwestern.edu/faculty/mazzeo/htm/airline_ontime.pdf

Statista. (2021). *Air Transportation - Statistics & Facts*. Retrieved from: https://www.statista.com/topics/1707/air-transportation/

Taylor, A. (2015). What Record-Breaking Snow Really Looks Like. Retrieved from https://www.theatlantic.com/photo/2015/02/winter-storms-bury-freeze-northeastern-us/385566/

U.S. Department of Transportation. (2020). *Air Travel Consumer Report*. Retrieved from https://www.transportation.gov/sites/dot.gov/files/2020-01/january-2020-atcr-1-16-2020.pdf

World Bank. (2021). *Travel and Tourism direct contribution to GDP*. Retrieved from: https://tcdata360.worldbank.org/indicators/tot.direct.gdp

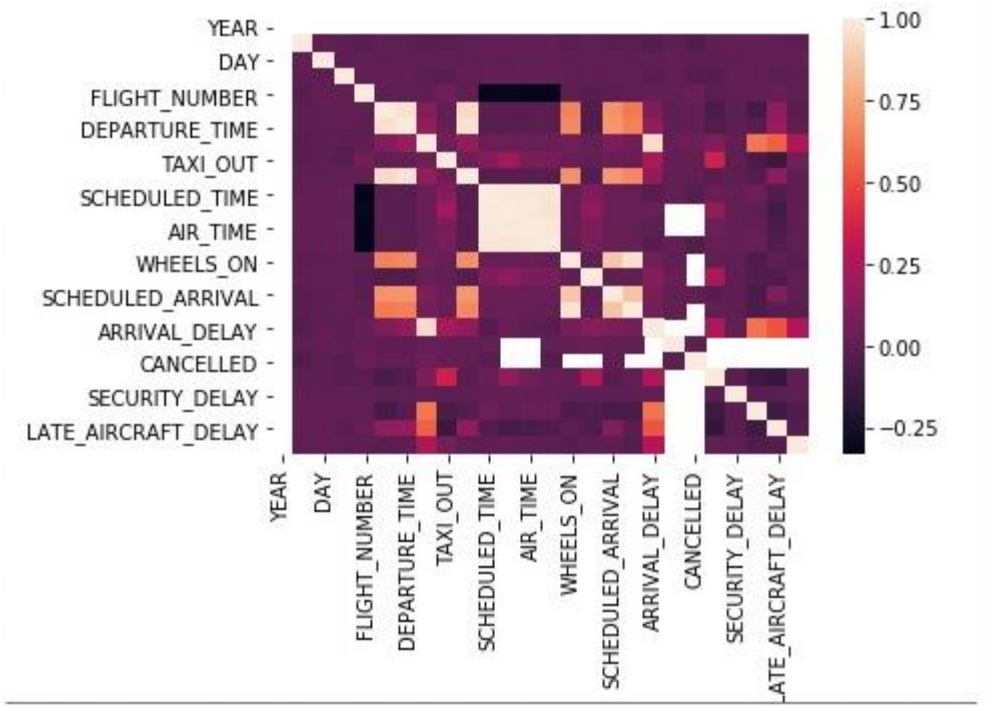
10. Appendix

Dataset

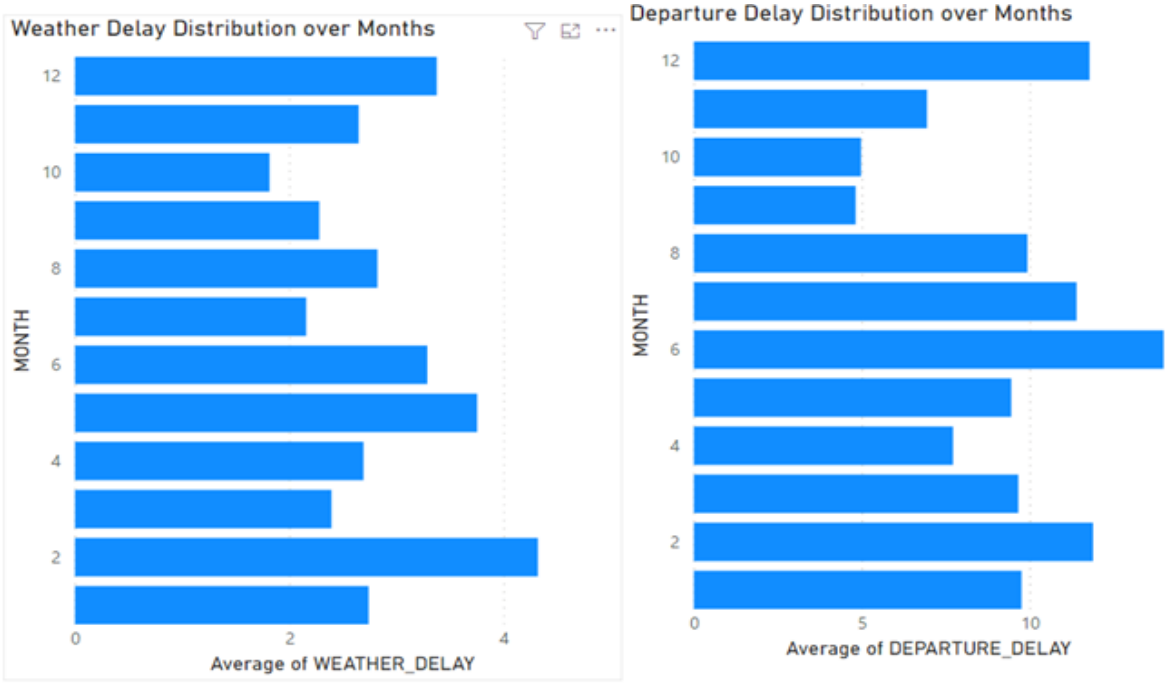
https://www.kaggle.com/usdot/flight-delays?select=flights.csv

Other Visuals

Correlation between the numeric variables

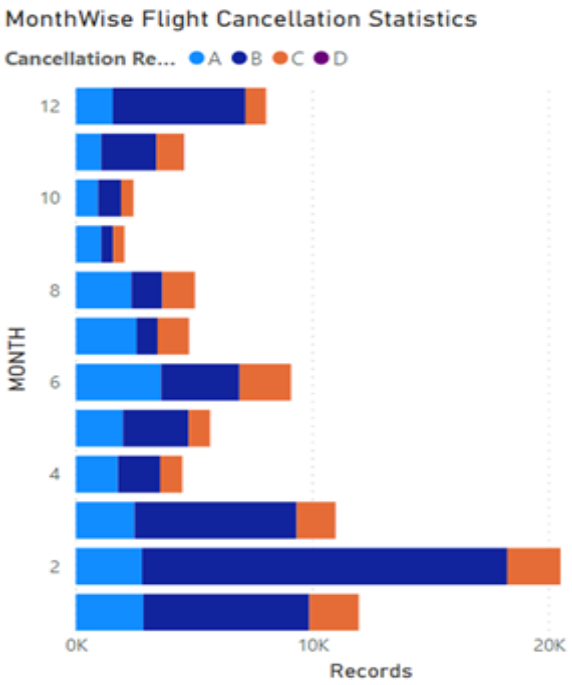


Weather and Departure delay Distribution over Months



The above distribution shows that the average delay caused due to weather is highest in February whereas departure delay is highest during June and December. The delay could be related to the holiday season but needs to be explored further with more granular details.

Flight cancellation statistics over the year



Most flights are cancelled in the month of February in 2015 - due to reason A: Aircraft or Carrier Cancellation. This needs to be analysed further as the ratio between flights cancelled and the planned flights in that month, and with the details if any specific Airline / aircraft is related to this high number of cancellations.