Airport & Airline Delays

By: Pedro Gomez, Antara Choudhury, Sridevi Madduri and Anna De Jesus

Motivation Behind Project

The traveler’s heart in each of us led us to question what really keeps us from reaching our destinations on time. With this project we hope to shed some light on what really goes on behind the scenes when it comes to flight delays.

Overview

The share of US flight delays caused by airlines, as opposed to weather or air traffic control, has surged to the highest level on record. We are pretty positive that if you've traveled recently in the past years...you've been curious about these DELAYS!

OUR data...

Data exploration and cleanup process

We used Pandas to clean and format our dataset.

The **Bureau** of Transportation Statistics website (data on Airline On-Time Statistics and Delay Causes):

https://www.transtats.bts.gov/OT_Delay/OT_DelayCause1.asp?20=E

The Analysis Process

Jupyter notebooks to analyze our data and Matplotlib to create 6 - 8 visualizations of the data (2 per question).

Research questions to answer:

1. What types of delays were seen in airports for all of the years?

2. Which of those types of delays have the greatest impact?

3. What is the correlation between the types of delays from 2019 to 2022?

   Hypothesis: As the years increase, the percentage of each type of delay decreases.

4. Which airline & airport had the greatest delay?

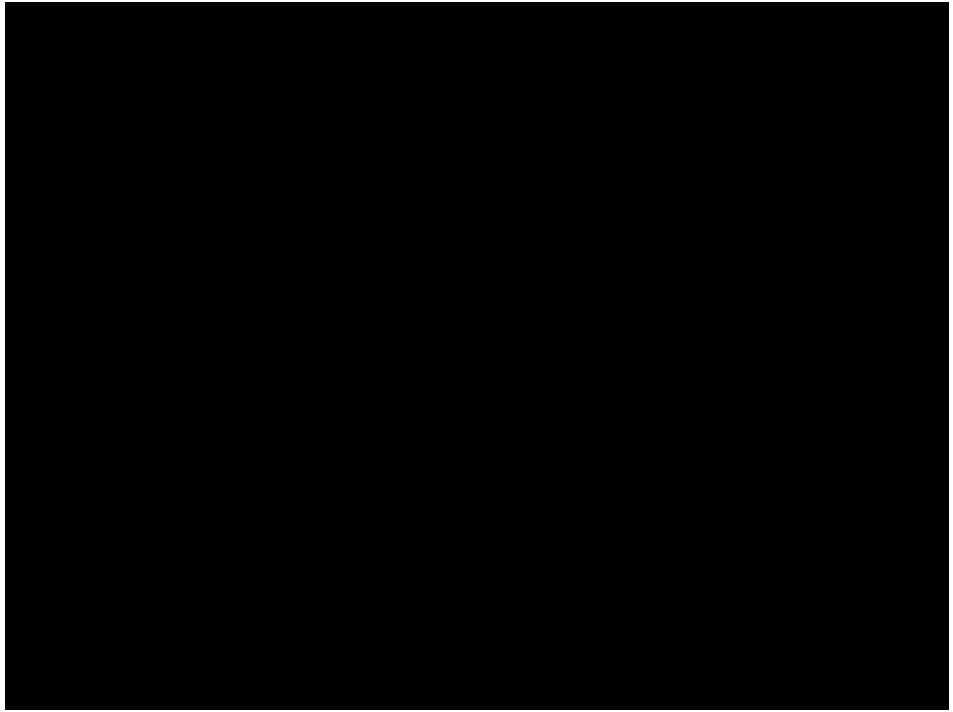
5. Which month had the highest delays for all of the years?

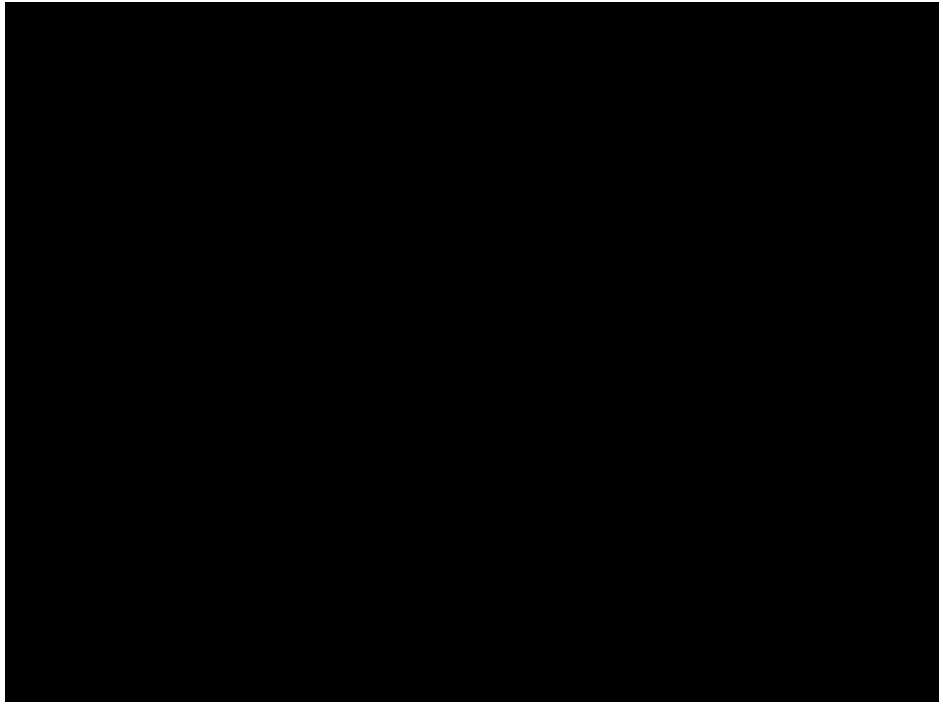
1. What types of delays were seen in airports for all of the years?

Lets review the 5 categories reported for airport delays.

- **Air Carrier:** Any cancellation or delay due to circumstances within the airline's control (e.g. maintenance or crew problems, aircraft cleaning, baggage loading, fueling, etc.).
- **Extreme Weather:** Significant meteorological conditions (actual or forecasted) such as tornado, blizzard or hurricane.
- **Security:** Delays or cancellations caused by evacuation of a terminal or concourse, re-boarding of aircraft because of security breach, inoperative screening equipment and/or long lines in excess of 29 minutes at screening areas.

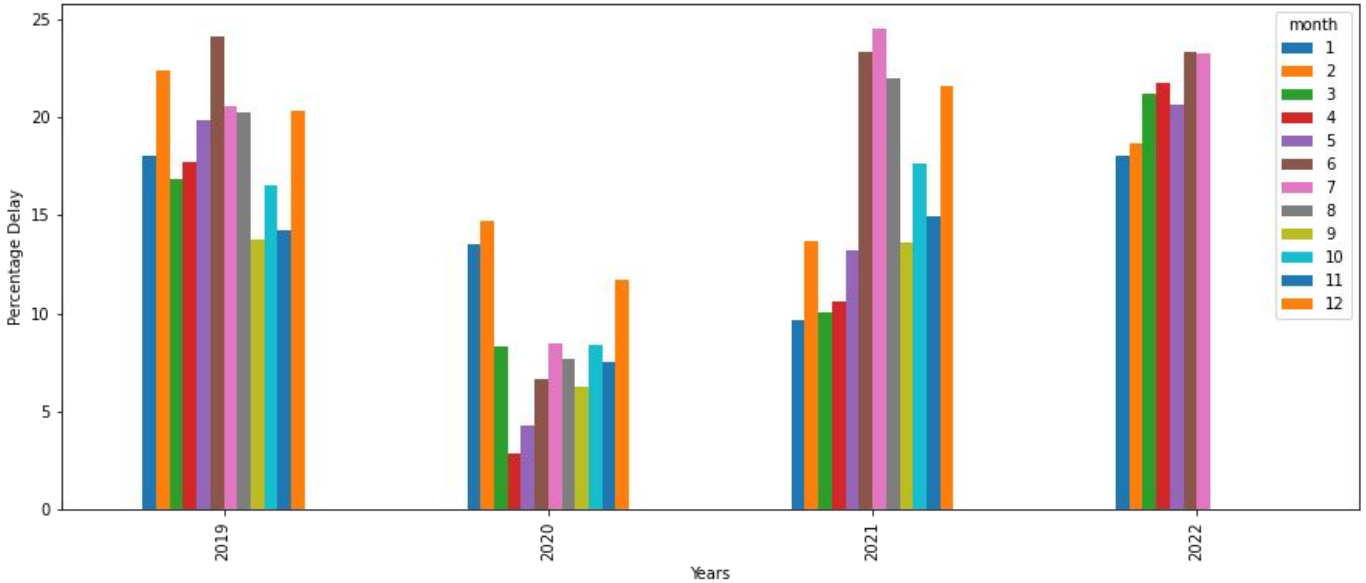
- **National Aviation System (NAS):** Delays and cancellations attributable to the national aviation system that refer to a broad set of conditions, such as non-extreme weather conditions, airport operations, heavy traffic volume, and air traffic control.
- **Late-arriving aircraft:** A previous flight with same aircraft arrived late, causing the present flight to depart late.





1b. Which month had the highest delays for all of the years?

- Group plot below shows percentage delay for each month over the years.
- From the group plot the month of June and July has highest percentage of delay.



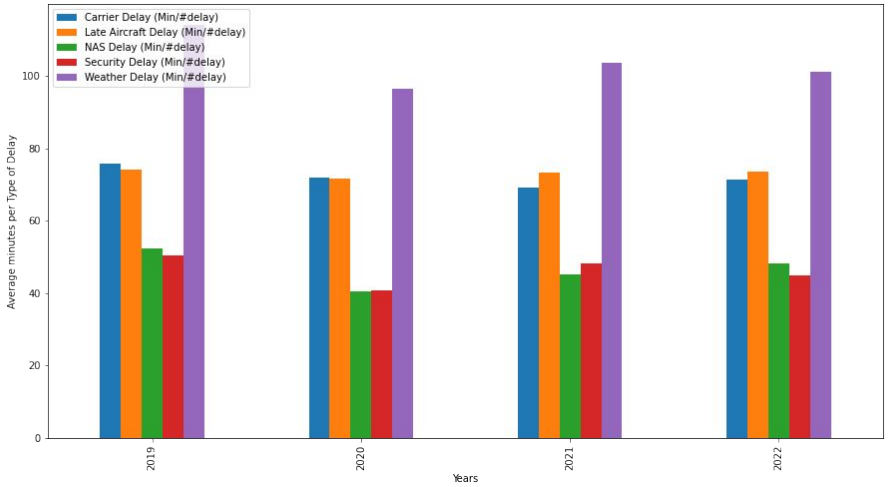
Conclusions

2. Which of those types of delays have the greatest impact?

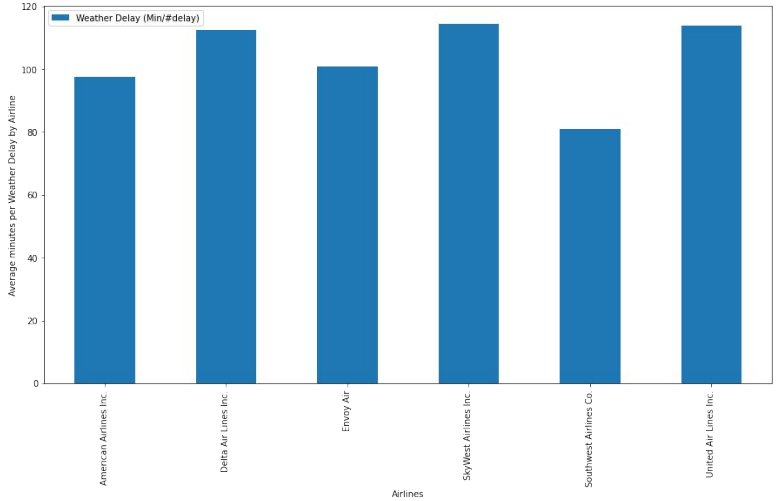
- To measure the impact of the delays we used the minutes registered to calculate and indicator of average minutes per type of delay

- For all four years evaluated, a weather delay took on average 25 to 39 minutes longer than any other type of delay. Being 2019 the year with the highest difference (114 min per weather delay)

The most impactful type of delays are the ones caused by weather:



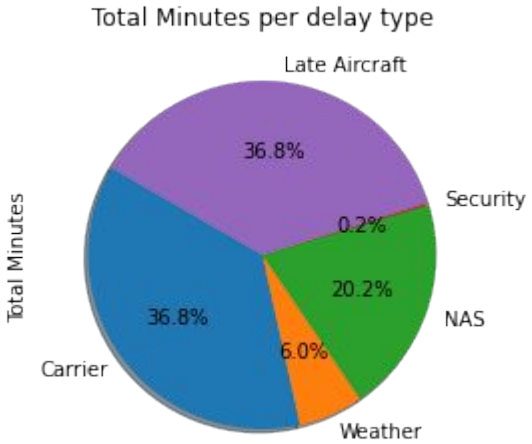
For airlines with more than 10k delayed arrivals registered, the impact of weather delays is represented below:

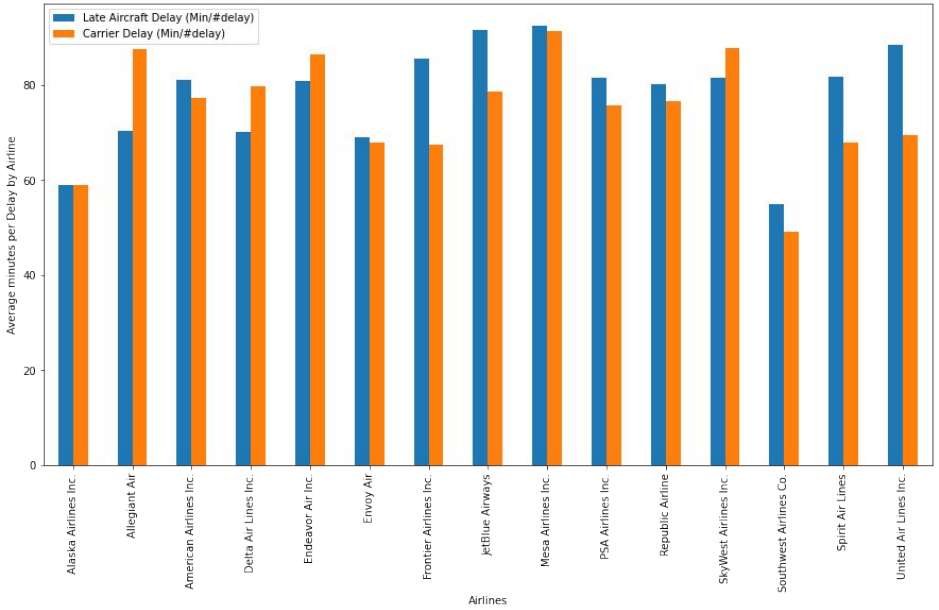


2. Which of those types of delays have the greatest impact?

- Despite weather delays being the most impactful, the total volume of minutes of delays is dominated by “Carrier” and “Late Aircraft” delays, which are under the airlines control.

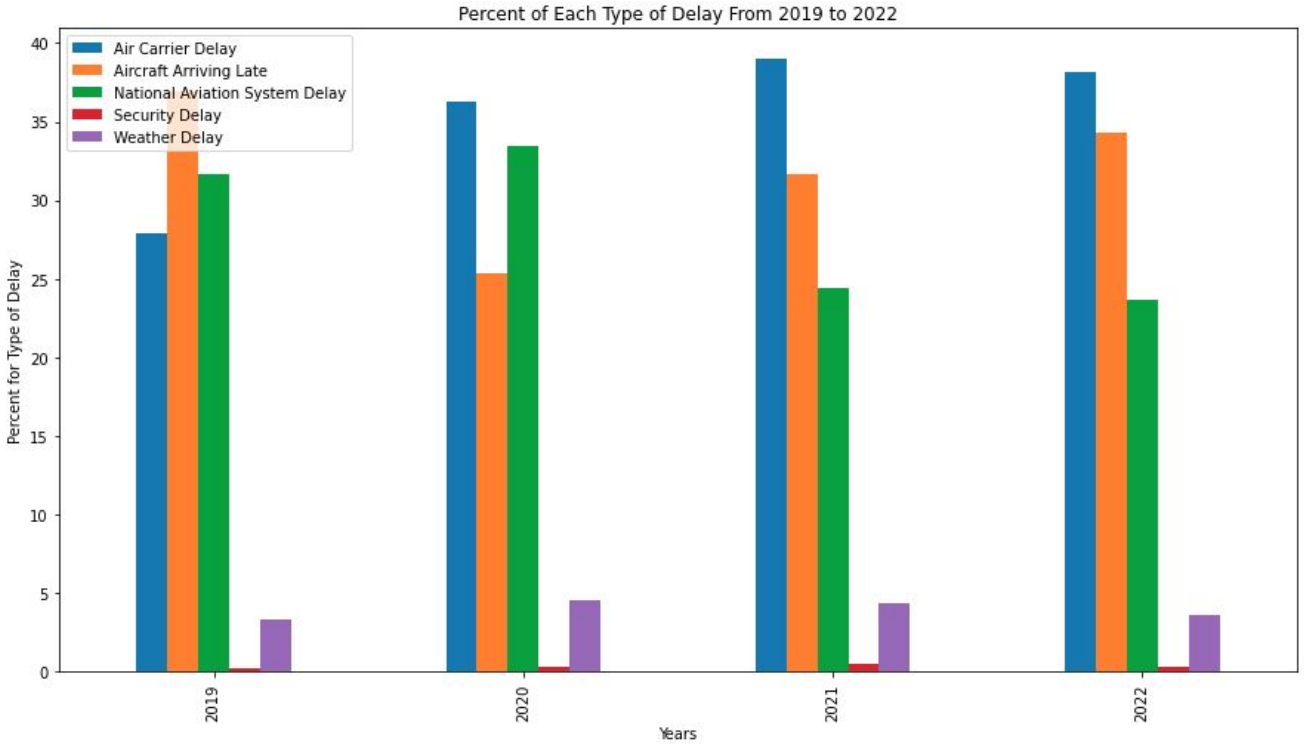
- Using the measure of average minutes per delay we found that the longest delays on average come from **Mesa Airlines** and the shortest on average come from **Southwest**, followed by **Alaska Airlines**





Conclusions

3. What is the relationship between the types of delays from 2019 to 2022?
Hypothesis: As the years increase, the percentage of each type of delay decreases.



Firstly, after viewing the data and visualizations that we have generated, we can **conclude** that there was no relationship between the types of delays from 2019 to 2022.

- Air Carrier Delays: Increased, then decreased

- Aircrafts Arriving Late: Decreased, then increased

- National Aviation System Delays: Increased, then decreased

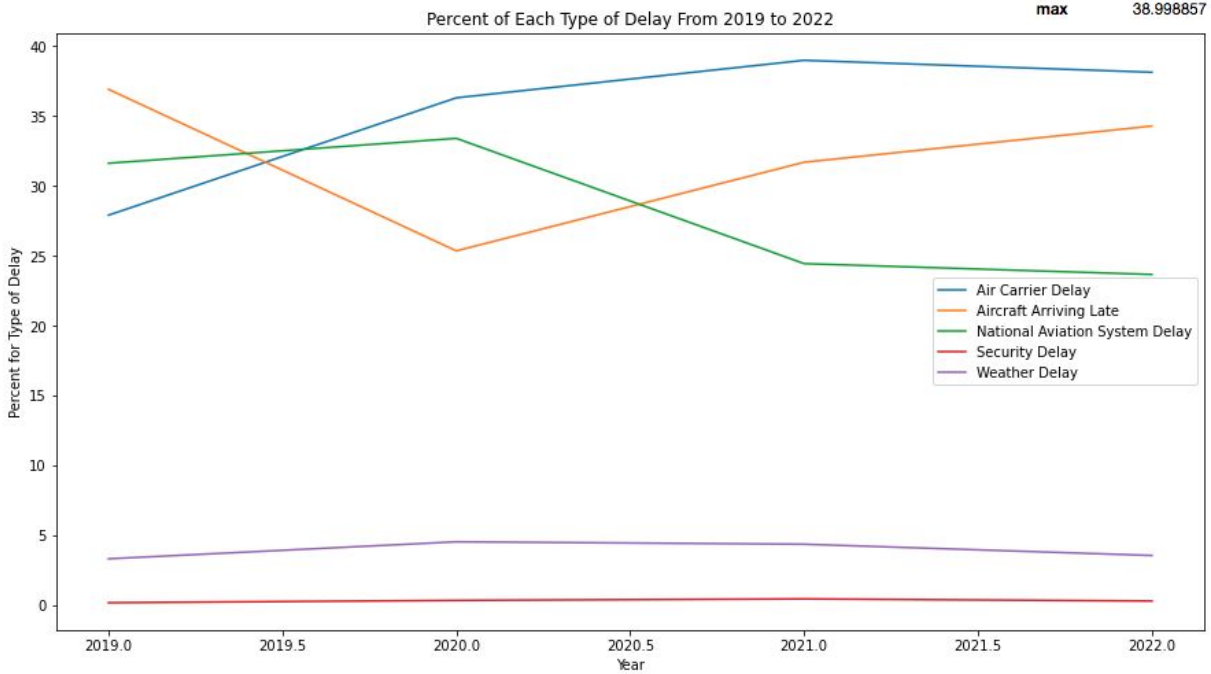
- Security Delays: Increased, then decreased

- Weather Delays: Increased, then decreased

Implications of these findings: as years increase, percent of each type of delay did not decrease; airlines were not able to figure out a way to lower the percent of delays occurring.

3. What is the relationship between the types of delays from 2019 to 2022?

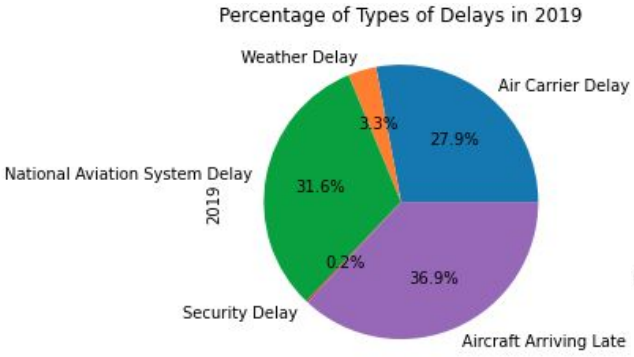
	Air Carrier Delay	Weather Delay	National Aviation System Delay	Security Delay	Aircraft Arriving Late
count	4.000000	4.000000	4.000000	4.000000	4.000000
mean	35.344079	3.957684	28.294559	0.331223	32.072562
std	5.075619	0.595995	4.948595	0.113918	4.953587
min	27.918126	3.332908	23.676731	0.190194	25.364205
25%	34.215896	3.511644	24.258021	0.281567	30.120795
50%	37.229666	3.974496	28.043408	0.334669	33.001114
75%	38.357849	4.420537	32.079946	0.384325	34.952881
max	38.998857	4.548836	33.414692	0.465362	36.923814

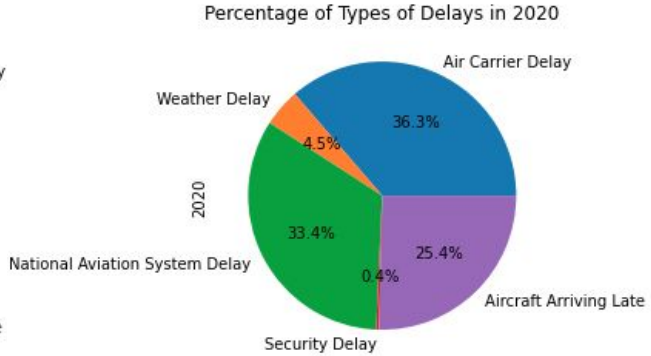


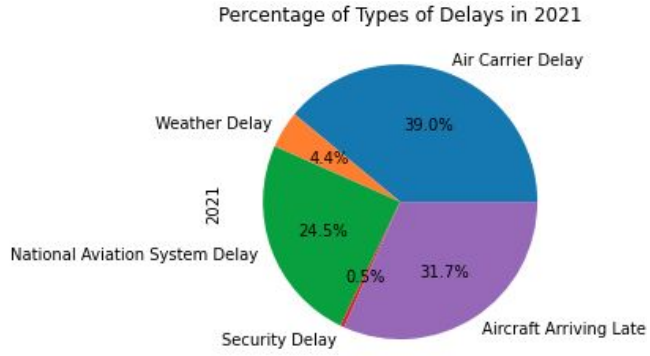
Secondly, after viewing the data and this visualization, we can **conclude** that as the years passed, the percent of Air Carrier Delays, Aircrafts Arriving Late, and National Aviation System Delays remained higher than the percent of Security Delays and Weather Delays.

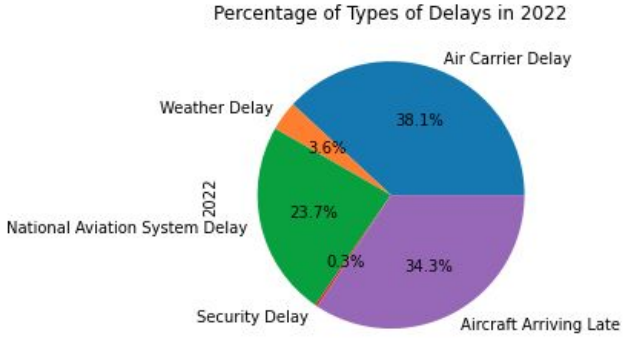
Implications of these findings: as the years passed, more passengers were affected by Air Carrier Delays, Aircrafts Arriving Late and National Aviation System Delays than Security Delays and Weather Delays. The airlines were not able to devise a way to lower the percent of flights delayed due to these three factors.

3. What is the relationship between the types of delays from 2019 to 2022?









Thirdly, we can conclude that as the years passed:

- Total percent of flights delayed due to Security Delays and Weather Delays: did not change a lot

- Flights delayed due to Aircrafts Arriving Late: lower in 2020 than in other years

- Flights delayed due to Air Carrier Delays: lower in 2019 than in later years

- Flights delayed due to National Aviation System Delays: higher in 2019 and 2020 than in 2021 and 2022

Implications of these findings: after 2020, passengers were less delayed due to National Aviation System Delays in flights. Airlines were able to devise a way to lower delays due to this cause. The implications of these findings also indicate that as the years passed, passengers were however more delayed due to Air Carrier Delays. Airlines were not able to come up with a way to lower delays due to this cause.

4. Which airline & airport had the greatest delay?

4.1 Top 3 Worst Airports analysis by Delay Percent

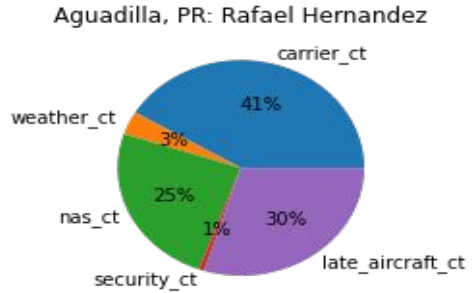
- Top 3 Worst Airports by Delay Percent

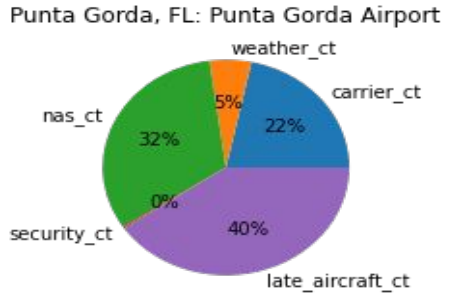
1. Aguadilla, Puerto Rico
2. Punta Gorda, Florida
3. Phoenix, Arizona

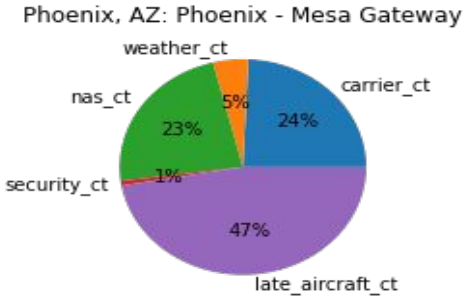
- The pie charts to the right show top 3 worst Airports by Delay Percent breaking down by different categories of delays

- Carrier, Late Aircraft and National Aviation System (NAS) are the main contributors for the delays

- Delay occurrences due to Weather and Security is minimal. However, if the delay due to Weather occurs, delay time by category type analysis shows that Weather delays on average can have more impact.







4. Which airline & airport had the greatest delay?

4.2 Top 3 Worst Airlines Analysis by Delay Percent

- Top 3 Worst Airlines by Delay Percent

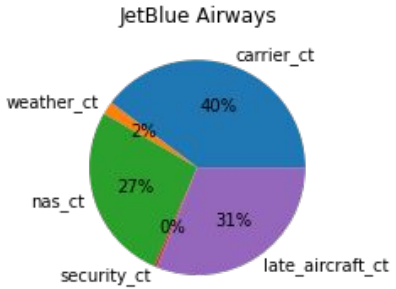
1. JetBlue Airways
2. Allegiant Air
3. Expressjet Airlines

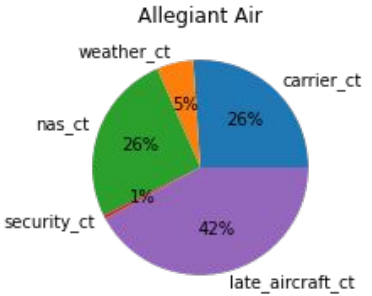
- The pie charts to the right show top 3 worst Airlines by Delay Percent breaking down by different categories of delays.

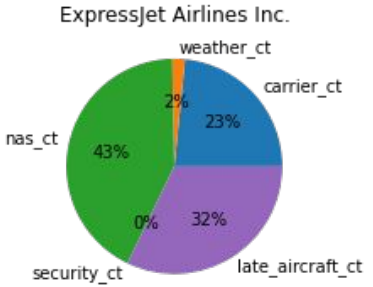
- Carrier, Late Aircraft and National Aviation System (NAS) are the main contributors for the delays.

- Delay occurrences due to Weather and Security is minimal. However, if the delay due to Weather occurs, delay time by category type analysis shows that Weather delays on average can have more impact.

- In our Analysis we see the same trend in the delay types for both Worst Airport & Airlines.

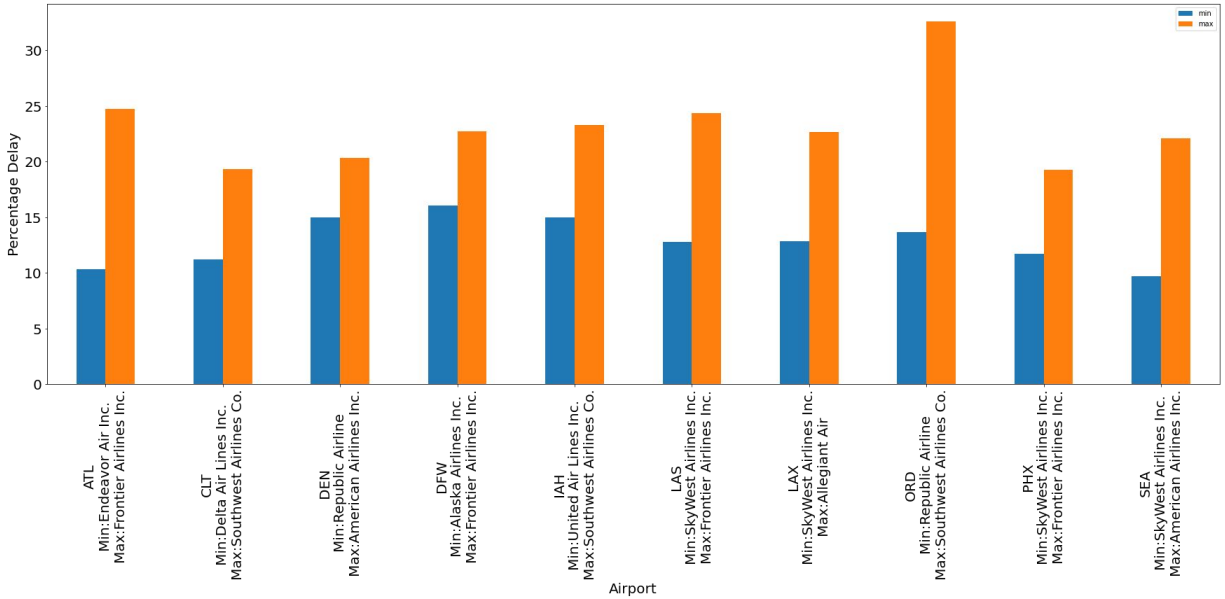






4. Which airline & airport had the greatest delay?

4.3 Worst & Best Airlines Analysis in Top 10 Busy Airports



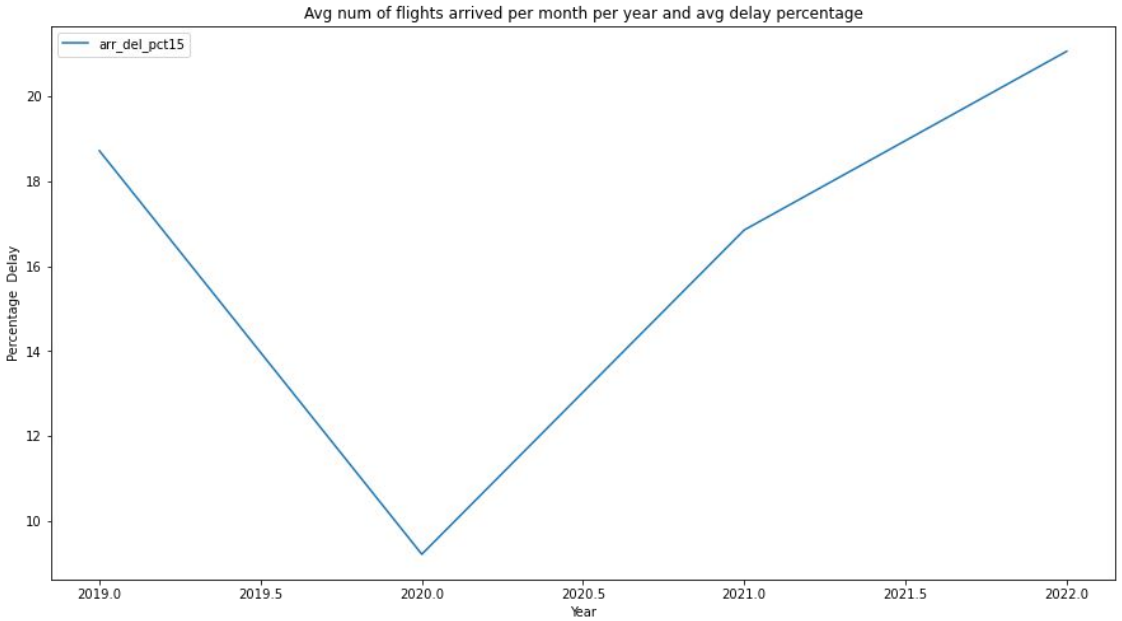
- Top 10 Busy Airports are identified based on the total Arrival Flights

- The group plot to the left shows the worst and best Airlines by Delay Percentage for each airport.

- For example, for the the busiest airport Atlanta (ATL), the worst Airline is Frontier Airlines with 25% delays and the best airline is Endeavor Air Inc with 10% delays.

- Frontier Airlines has max delay percentage in 4 out of 10 busy airports while SkyWest Airlines has the least delay percentage in 4 out of 10 busy airports.

5. Correlation between the average number of flights arrived and percentage delay.



- line plot below shows the correlation between the number of flights arrived and percentage delay.

- From the line plot we can see that as the number of flights arrived in airport increased the delay percentage increases.

-The correlation factor between average number of flights arrived in airport per month and percentage delay is 0.96.

-The correlation factor 0.96 indicates there is strong and positive association between number of flight arrived and percentage delay.


Thank you.