

Project Final Report

Project name: Airline_Delay_and_Cause_Analysis	Participants (Group B): Kanchan Laxman Dombale Chindu Francis
Executive Summary Context: Flight delays and their associated causes are analysed to identify the primary factors contributing to delays and provide actionable recommendations for optimizing flight operations at airlines and airports Objective: This analysis aims to identify the key factors affecting the operational efficiency of airports and airlines and provide recommendations to help improve their performance. Achieved Result: A data-driven visualization of flight delays, cancellations, and diversions, highlighting the impact of various contributing causes.	

1. Business Understanding

Goal:

Analyse the dataset “**Airlines_Delay_Cause.csv**” to identify the primary causes of flight delays and provide actionable recommendations to optimize flight operations for airlines and airports.

Key Business Questions:

- Which are the Best Performing and Worst Performing Airports and Airlines?
- What are the main causes of flight delays and how do they differ depending on the airline, airport and season?
- Which airlines and airports experience the most cancellations and diversions?
- How can airlines and airports optimize their operations to reduce delays and improve punctuality?

2. Data Understanding

This dataset provides detailed information on airline performance in the U.S. from January 2016 to October 2024, including flight arrival times, delays, cancellations, and diversions. It also tracks the causes of delays, such as carrier, weather, national air system issues, security, and late aircraft. It contains a total of 184616 rows and 21 columns.

[Data source](#)

Variables in the Dataset:

year & month: Date spans from 2016 to 2024.

arr_flights: Total arriving flights per record.

carrier: A 2-Character abbreviated code of the airline carrier. One or multiple carriers can belong to one Parent Company.

carrier_name: The full name of the airline carrier.

airport: A 3-character code unique code for Airport.

airport_name: The full name of the airport.

arr_del15: Number of flights delayed by 15+ minutes

carrier_ct: Count of delays that are caused by the carrier itself, including issues like maintenance problems, crew availability, and operational delays.

weather_ct: Count of delays that are caused by weather conditions such as storms, fog, and other weather-related disruptions.

nas_ct: Count of delays that are caused by National Aviation System related issues include air traffic control delays, heavy traffic volume, and other infrastructure-related issues.

security_ct: Count of delays that are caused by evacuation of a terminal or concourse, re-boarding of aircraft because of security breach.

late_aircraft_ct: Count of delays due to previous aircraft arriving late, causing a domino effect on subsequent flights.

arr_cancelled: Number of flights cancelled

arr_diverted: Number of flights diverted

arr_delay: Total Arrival Delay in Minutes

carrier_delay: Carrier-Related Delay in Minutes

weather_delay: Mean: Weather-Related Delay in Minutes

nas_delay (National Aviation System Delay): NAS Delay in Minutes

security_delay: Mean: Security-Related Delay in Minutes

late_aircraft_delay: Late_Aircraft Delay in Minutes

3. Data Preparation

List of all data cleansing measures:

- 1) Elimination of NULL values: All columns were reviewed for Null or Blank entries. Rows with any Null values were removed entirely because the entire row was NULL.
- 2) Elimination of Duplicates: All columns were reviewed for Duplicates. There is No Duplicate data in this Dataset.
- 3) Airline and Parent Company Mapping: This dataset includes multiple airlines operating under a single Parent Company, with each airline mapped to its respective Parent Company name.

For example: Envoy Air, American Airlines Network, PSA Airlines Inc., Air Wisconsin Airlines Corp, Republic Airline to American Airlines Group.

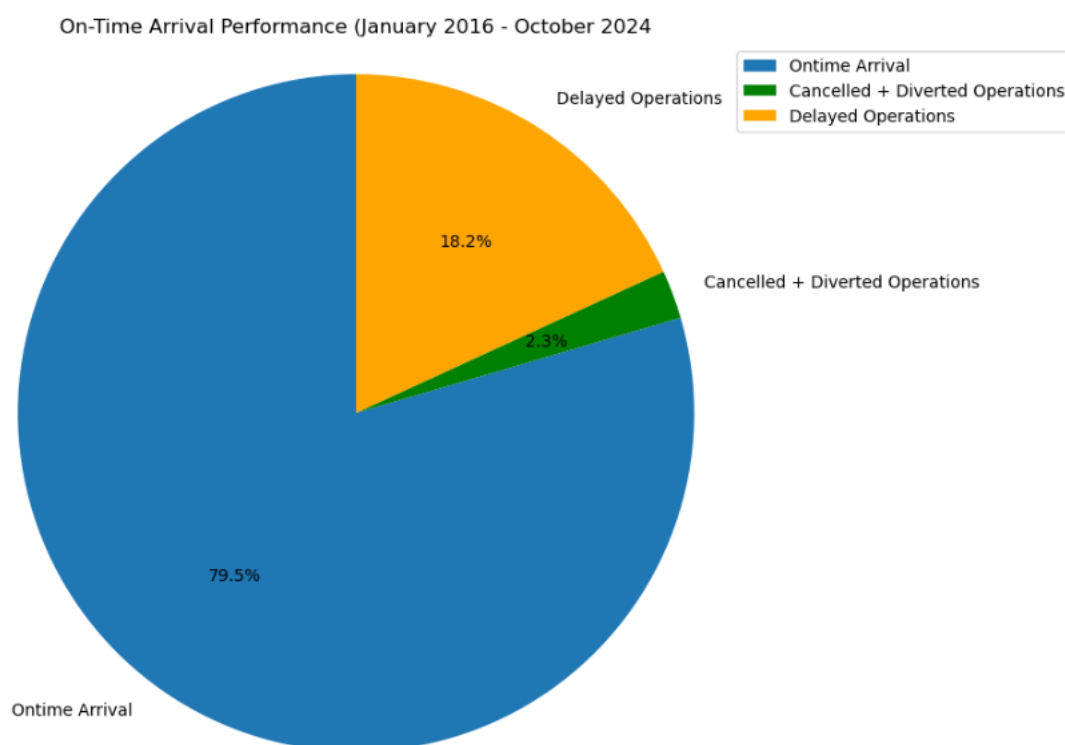
4) Airport-State Mapping: Each airport in the dataset is mapped to its corresponding U.S. state for better geographic analysis.

5) Date column by combining year and month: Created new column “date” from available year, month column combination to do the Analysis over time and Day is considered “1”.

4. Data Modeling

A. Which are the Best Performing and Worst Performing Airports and Airlines?

A.1 What is the overall percentage of flights arriving on time vs. delayed?



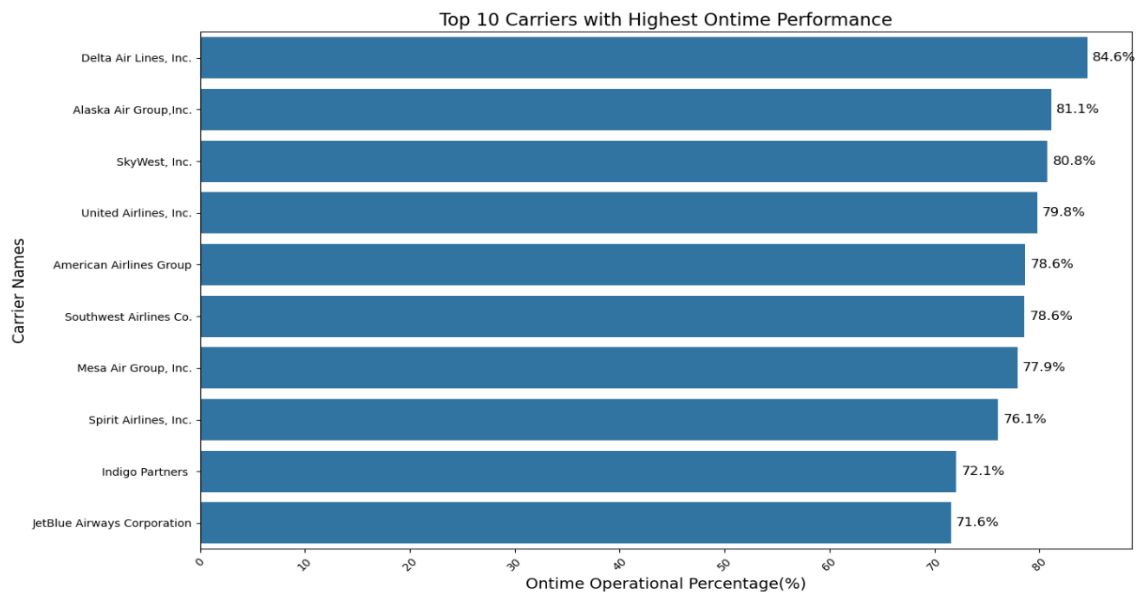
Flight Service Performance Overview

- On-Time Arrivals: 79.5%
- Delayed Flights: 18.15%
- Diverted Flights: 0.24%
- Cancelled Flights: 2.07%

Approximately 80% of flights arrive on time, while about 20% face delays, diversions, or cancellations.

A.2 Which airlines (carriers) have the highest on-time performance and highest Operational Delay?

A.2.1. Top 10 Airlines with Highest Ontime Performance



Higher Percentage = Better On-Time Performance

Delta Air Lines, Inc. leads with 84.58% of its flights arriving on time, making it the most punctual airline on the list. JetBlue Airways Corporation ranks the lowest in the top 10, with 71.58% of its flights arriving on time. Comparison of Airlines

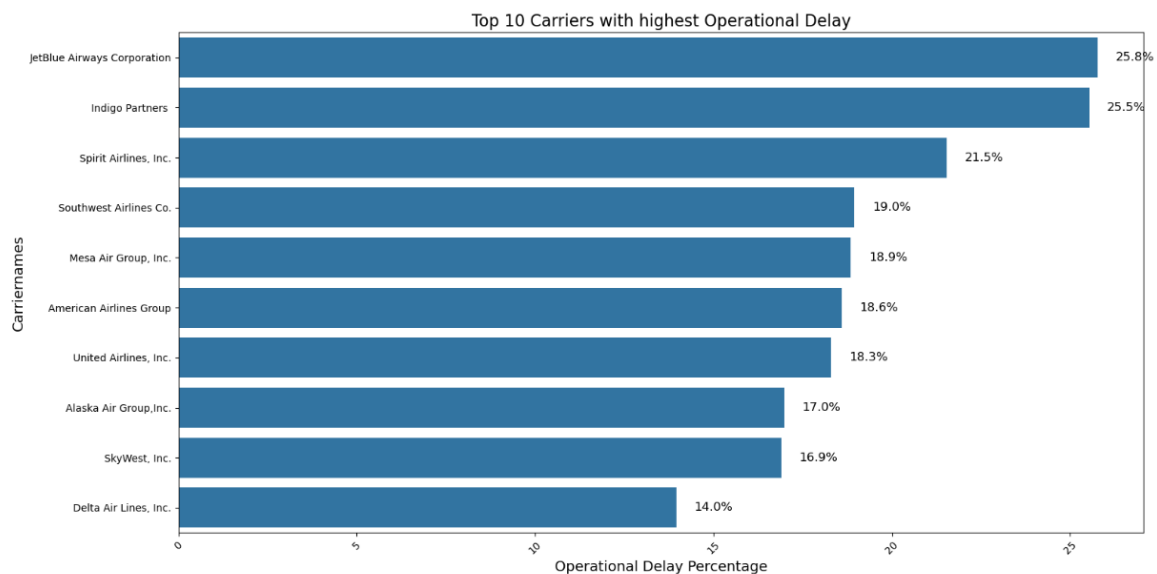
The performance gap between the most and least punctual airlines in the top 10 is around 13%. Airlines such as Alaska Air Group, SkyWest, and United Airlines have percentages above 79%, indicating strong reliability.

Key Takeaways:

- Delta Air Lines is the most punctual airline.
- JetBlue Airways has the lowest on-time performance among the top 10.
- The overall range is between ~71% and ~85%, indicating variability in punctuality.

If on-time performance is a key factor for travel, airlines like Delta, Alaska, and SkyWest may be preferable.

A.2.2 Top 10 Airlines with Highest Operational Delay

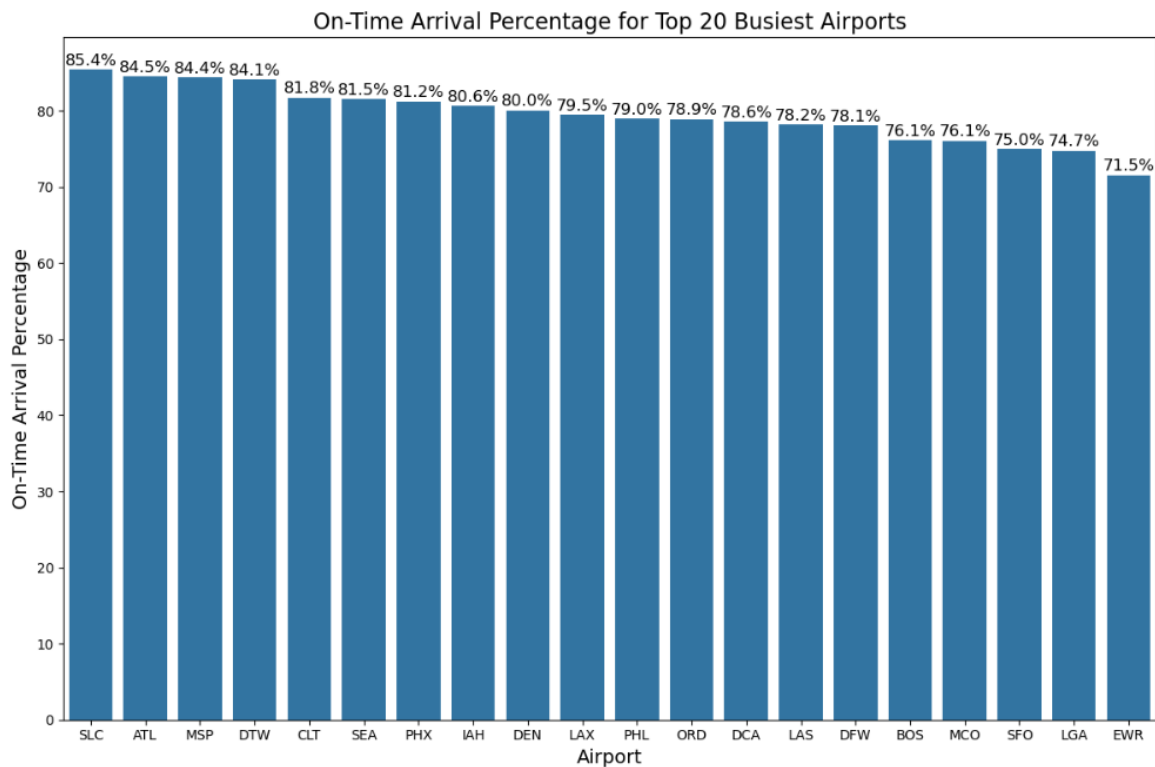


Least Punctual Airlines / The Airline has highest Delays Key Takeaways:

- Delta Air Lines has the lowest delay rate (15.42%), making it the most reliable airline among the top 10.
- JetBlue Airways has the highest delay rate (28.42%), indicating that nearly 1 in 3 flights are delayed.
- Indigo Partners and Spirit Airlines also have higher delay percentages (27.89% and 23.92%), suggesting they may not be the best choice if punctuality is a priority.
- The gap between the most and least delayed airline is approximately 13%, showing noticeable differences in reliability among airlines.
- Most airlines fall within a 15%–25% delay range, which means delays are relatively common even among the top-performing carriers.
- The low-cost and ultra-low-cost carriers (LCCs and ULCCs) often experience higher operational delays.

A.3 Which airports has highest on time Arrival Percentage and Operational delay Percentage?

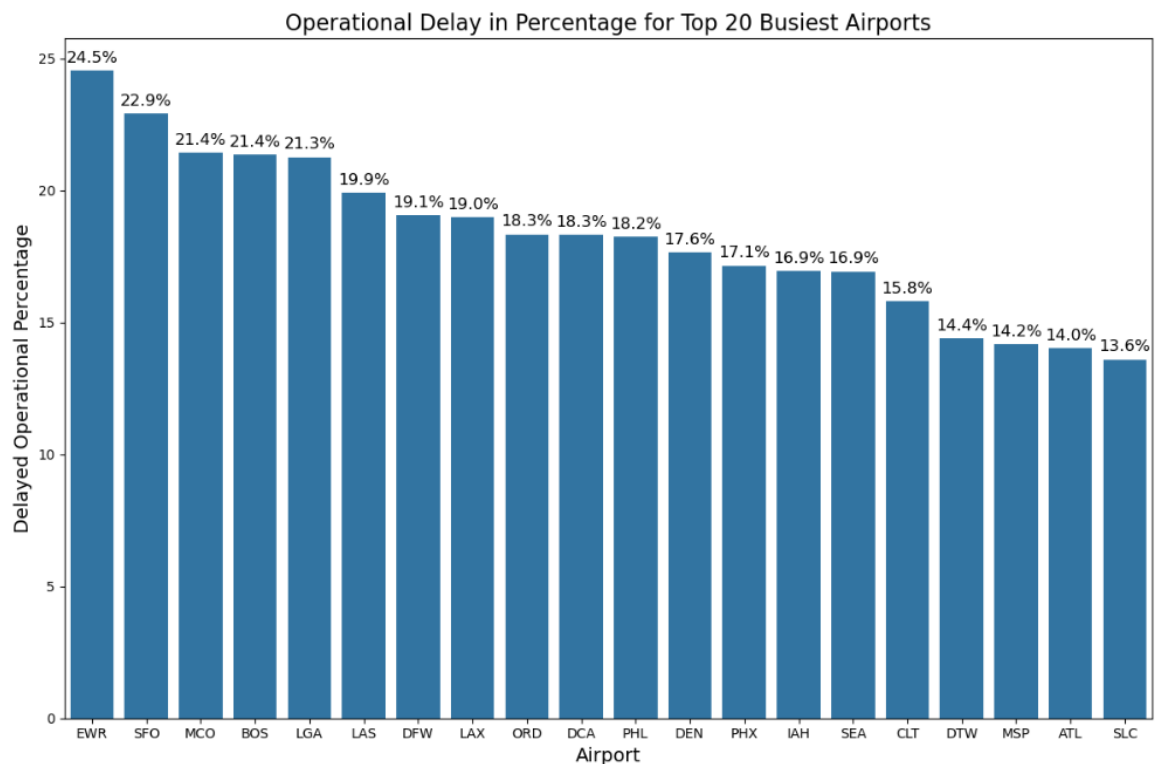
A.3.1 On Time Arrival percentage for Top 20 Busiest Airports



Airports with Highest Operational delay:

- SLC (Salt Lake City International): Highest operational efficiency percentage.
- ATL (Atlanta Intl), MSP (Minneapolis Intl),DTW(Detroit): have on-time performance above 84%.
- Delay Gap: ~14% difference between the most and least delayed airports, highlighting reliability differences.
- Delay Range: Most airports experience delays between 15%–30%, indicating common delays even among top performers.

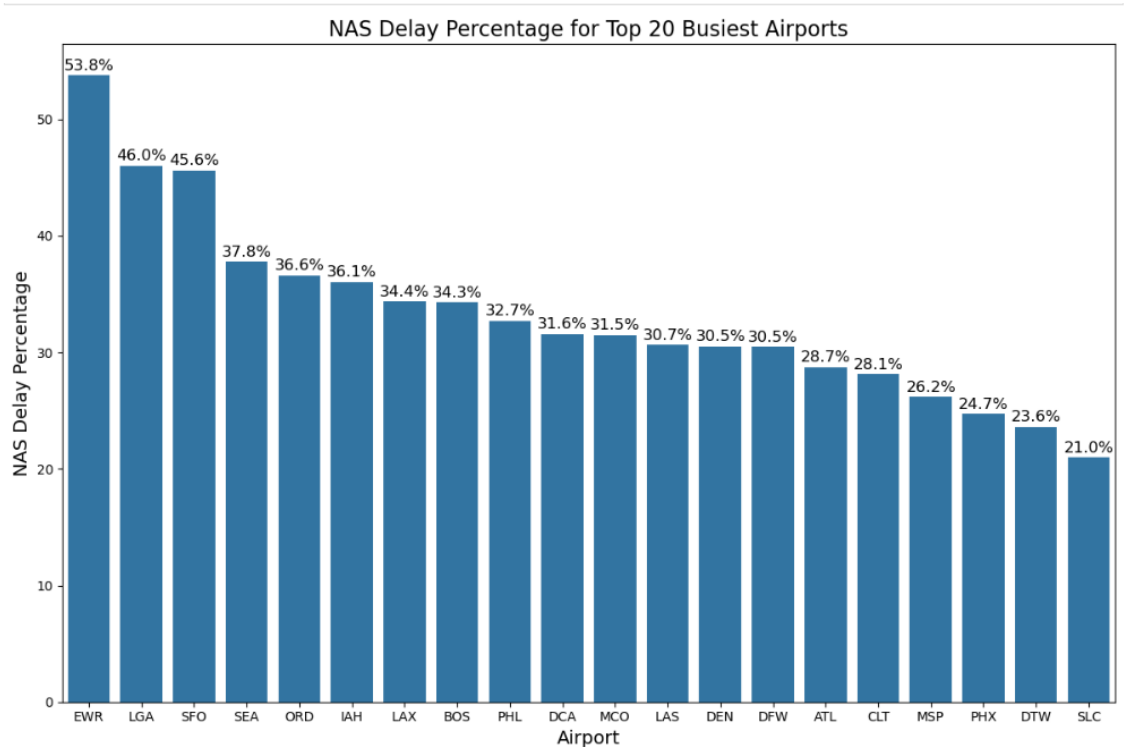
A.3.2 Operational Delay Percentage for Top 20 Busiest Airports



Airports with Highest Operational delay:

- **EWR (Newark, NJ):** Highest delay rate (24.5%), with nearly 1 in 4 flights delayed.
- **SFO and MCO:** Higher delay rates (22.9% and 21.4%)
- **Delay Gap:** ~10% difference between the most and least delayed airlines, highlighting reliability differences.
- **Delay Range:** Most Airports experience delays between 13%–25%, indicating common delays even among top performers.

A.4 Which airports experience the most delays because of Nas (National Aviation System Delays - ATC Congestion, Runway Delays)



Among the top 20 busiest airports of US, EWR experience highest NAS delay percentage followed by LGA and SFO. [1](#)

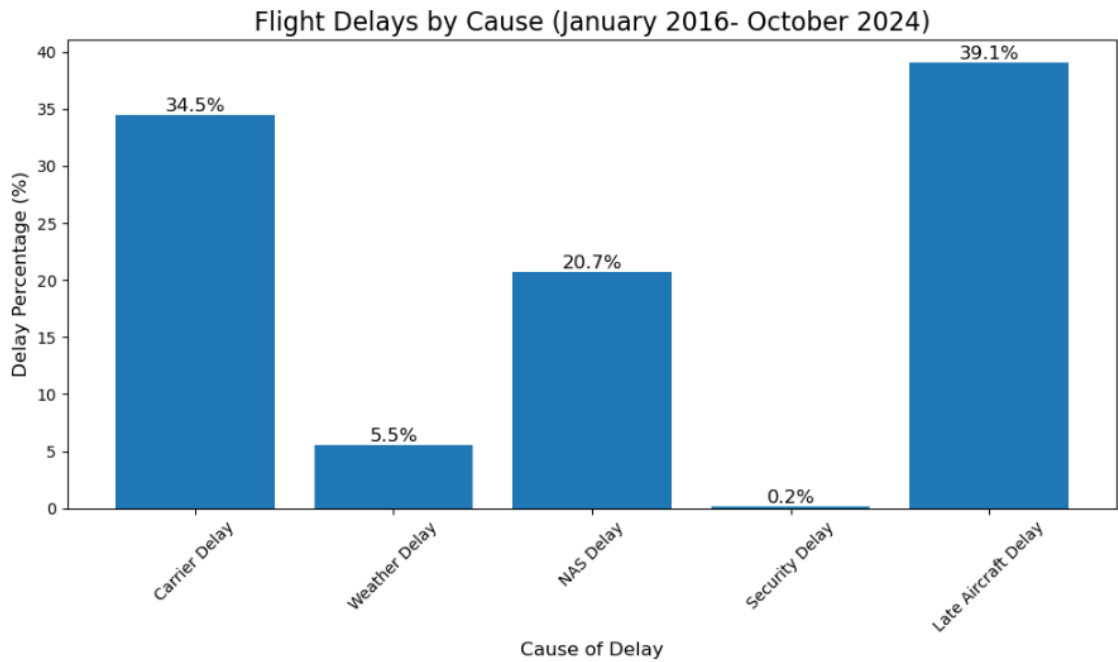
As shown in Chart EWR (Newark Liberty International Airport, located in Newark, New Jersey, USA.) has 53.76 % Delay due to NAS (National Aviation System Delays: ATC, Runway Delays): 3 Runways¶ The combination of high traffic volume, weather impacts, ATC congestion, regional airspace constraints, and airport-specific operational challenges likely contributes to EWR's high NAS delays.

Since NAS delays are not solely under the control of the airport but are part of a broader system, the interdependencies between the airport and the airspace system can lead to significant delays, especially during peak travel times or bad weather conditions.

B. What are the main causes of flight delays and how do they differ they differ depending on the airline, airport and season?

B.1. What are the main causes of flight delays?

B.1.1 Main Causes



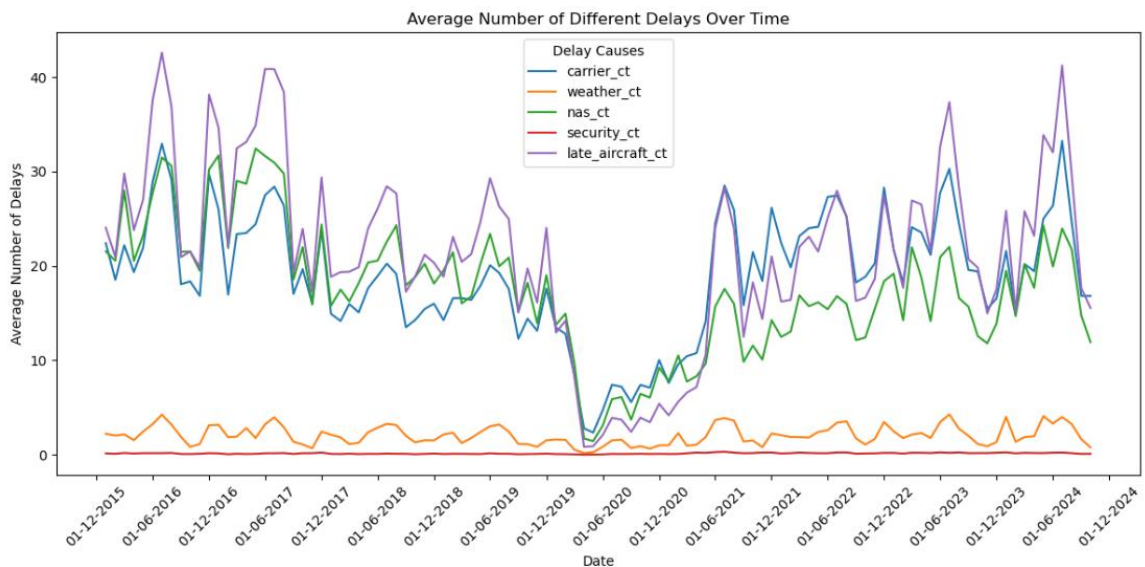
Main Causes of Flight Delays:

- Late Aircraft Delays: Contribute 39.08% of overall delays.
- carrier Delays: Account for 34.46% of overall delays.
- NAS Delays (National Air System): Responsible for 20.75% of delays.
- Weather Delays: Make up 5.54% of overall delays.
- Security Delays: Contribute 0.18% of delays.

Top 3 Delay Causes:

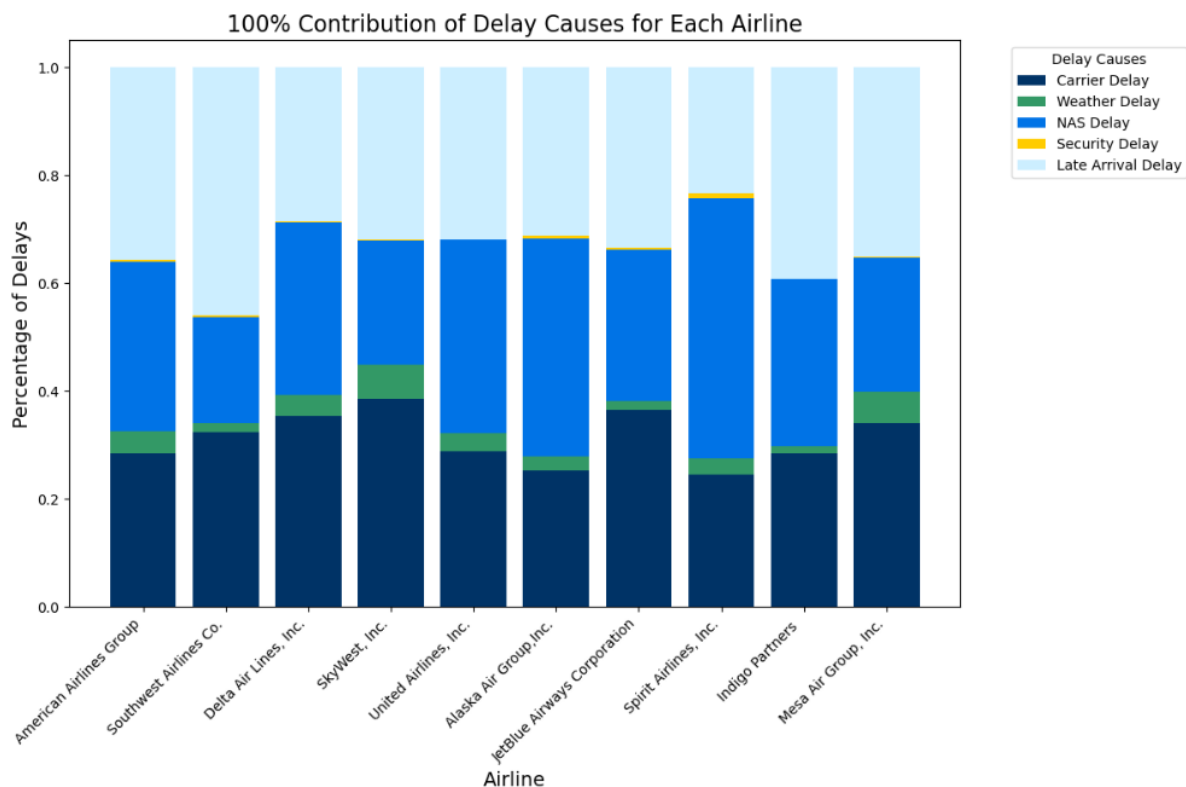
Late Aircraft Delays, carrier Delays, and NAS Delays are the major contributors, accounting for around 95% of all delays.[1](#)

B.1.2 How do delay causes vary with time?



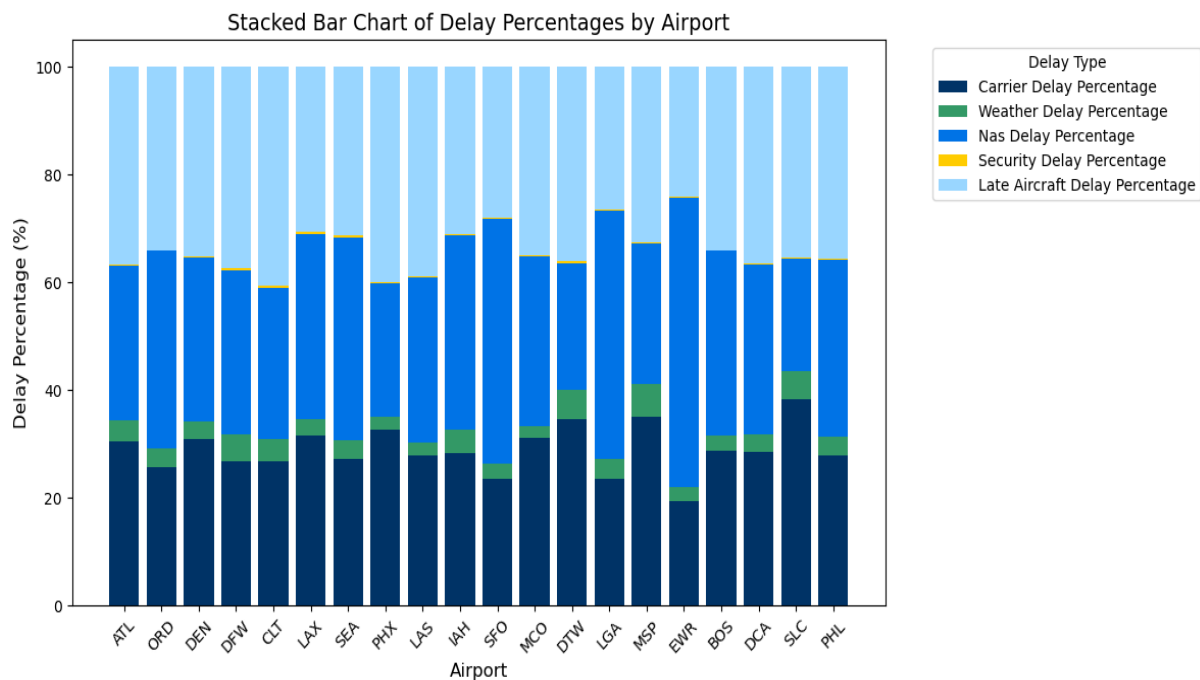
- Late aircraft ct has remained a consistent top contributor for delays from 2016-2020.
- We can see a dip in overall contributors from dec-2019 to June -2020 (covid period)
- From June 2020 - December 2022, Carrier delay has surpassed late career delay and later late aircraft ct again became top contributor. The peak in the count of carrier delays from June 2020 to December 2022 in U.S. traffic is largely linked to the impacts of the COVID-19 pandemic and the subsequent recovery period.
- When airlines began resuming operations after the initial pandemic shutdown, they faced significant operational challenges. Reduced staffing, crew shortages, and reconfigured schedules led to inefficiencies that increased carrier delays. Maintenance schedules of flights were affected during pandemic. Also, there was a rise in demand for flights post lockdown. But the carriers were not sure about the stability of demand (multiple covid waves).
- Reduction in NAS DELAYS in recent years is observed.
- Recent reductions in NAS delays have been achieved through a combination of technological, operational, and collaborative improvements:
- Improved coordination, real-time data integration, and updated procedures have enabled air traffic controllers to manage high traffic volumes more flexibly and responsively. Advances in weather prediction have allowed for proactive adjustments to flight paths, minimizing the impact of adverse conditions.

B.2. How does Flight delays differ depending on the airline?



- late_aircraft delays are Higher at Southwest Airlines (46%) and Indigo Partners (39%).
- Carrier delays are Higher at SkyWest Inc (38%) and JetBlue Airways corporation (36%).
- Nas Delays are Higher at Spirit Airlines (48%) and Alaska Airlines (40%).
- Mesa (6%) and SkyWest Airlines (6%) faces most weather-Related delays.
- Spirit Airlines faces (1%) Security delays.
- The American Airlines Group, Southwest Airlines Co., and Delta Air Lines seem to have a relatively similar distribution of delay causes, with Late Arrival Delay being the dominant factor.

B.3. How does Flight delays differ depending on the airport?



Late Aircraft Delay: are higher at following Airports

- CLT (41%) – Charlotte Douglas International Airport (Domestic + International)
- PHX (40%) – Phoenix Sky Harbor International Airport (Domestic + International)
- LAS (39%) – McCarran International Airport (Domestic + International)

Carrier Delays: are higher at following Airports

- SLC (39%) – Salt Lake City International Airport, Salt Lake City (Domestic + International)
- MSP (35%) – Minneapolis-Saint Paul International Airport (Domestic + International)
- DTW (35%) – Detroit Metropolitan Wayne County Airport (Domestic + International)

National Aviation Delays (NAS): are Higher at following Airports:

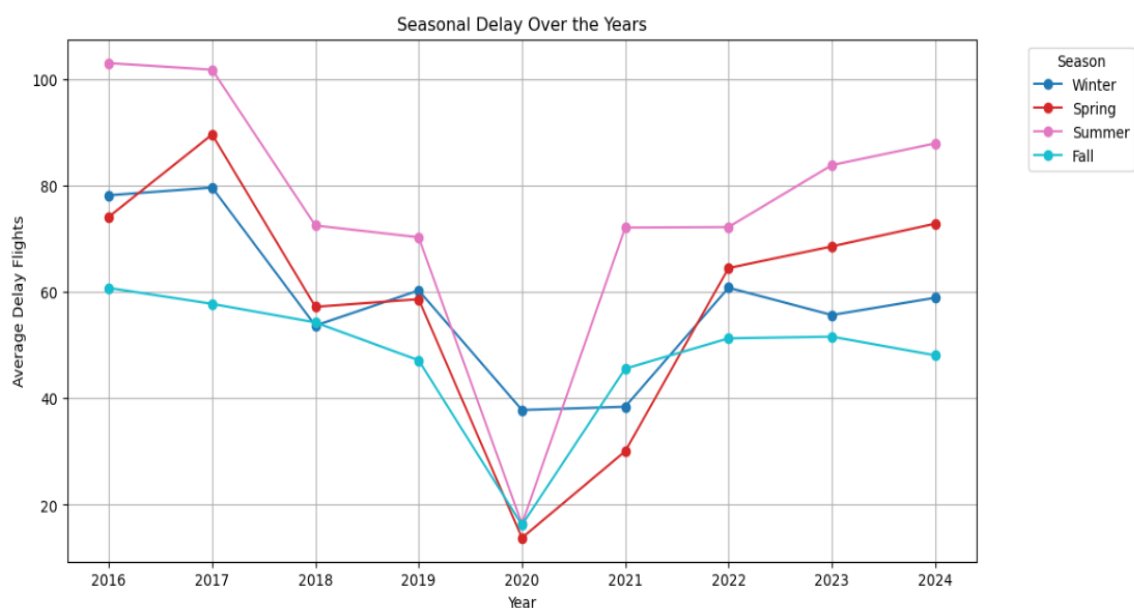
- EWR (53%) – Newark Liberty International Airport, Newark, New Jersey, USA (Domestic + International)
- LGA (46%) – LaGuardia Airport, Queens, New York City, New York, USA (Primarily Domestic, very few international flights)
- SFO (45%) – San Francisco International Airport, San Francisco, California, USA (Domestic + International)

Weather delays: are higher at following Airports:

- DFW: Dallas/Fort Worth International Airport (Dallas/Fort Worth, Texas, USA) (5%)
- IAH: George Bush Intercontinental Airport (Houston, Texas, USA) (4%)

Security Delay: (Very few, But at Some Airports need few Improvements)

B.4. How does Flight delays differ depending on the Season?



Identifying Trends:

- The Summer season shows consistently high delays, peaking around 2019 and then dropping significantly in 2020 before rising again.
- All seasons show a significant drop in average delays in the year 2020, likely due to an external factor affecting all flights (possibly the COVID-19 pandemic).
- Post-2020, delays start increasing again for all seasons, indicating a recovery phase.

Comparing Seasons:

- Summer generally has the highest average delays compared to other seasons.
- Winter and Spring have relatively lower delays, with Fall showing moderate delays.

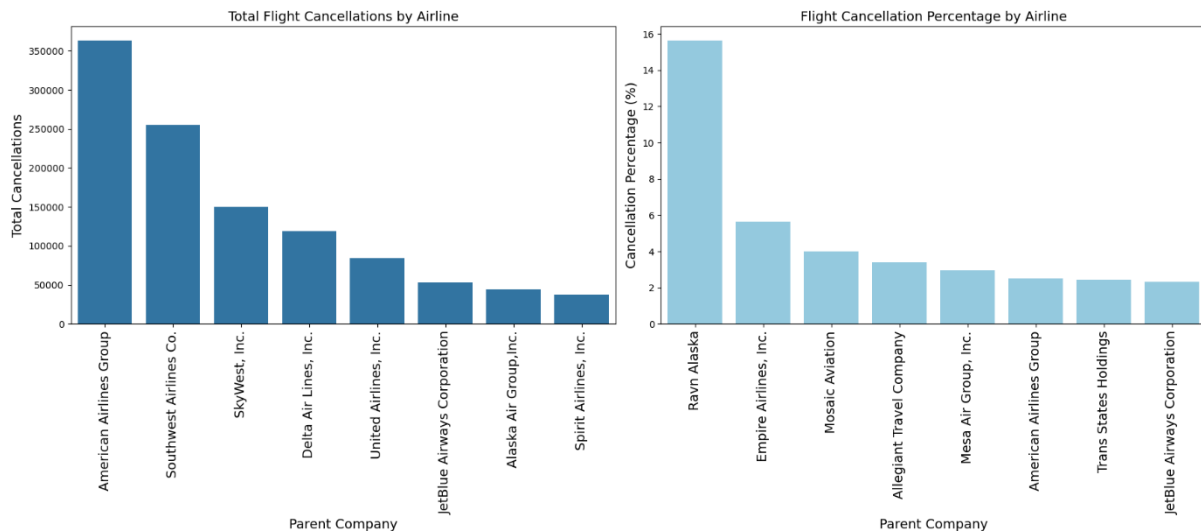
Yearly Analysis:

- Compare the average delays for different seasons in the same year to understand which season experiences more delays.
- Notice the variation in delays from year to year for each season, reflecting changes in operational or external factors.

C. Which airlines and airports experience the most cancellations and diversions and what are the underlying reasons?

C.1. Which airlines experience the most cancellations and diversions?

C.1.1. Which airlines experience the most Cancellations



- **COUNT GRAPH (On Left):**

If focusing on total impact on passengers. It shows Airlines with the highest cancellations. Highlights airlines that cause the most disruptions in total. Larger airlines will naturally have more cancellations, so this can be biased.

Interpretation:

Total Cancellations: American Airlines Group has the highest total number of cancellations, suggesting it experiences significant operational challenges or high traffic volume.

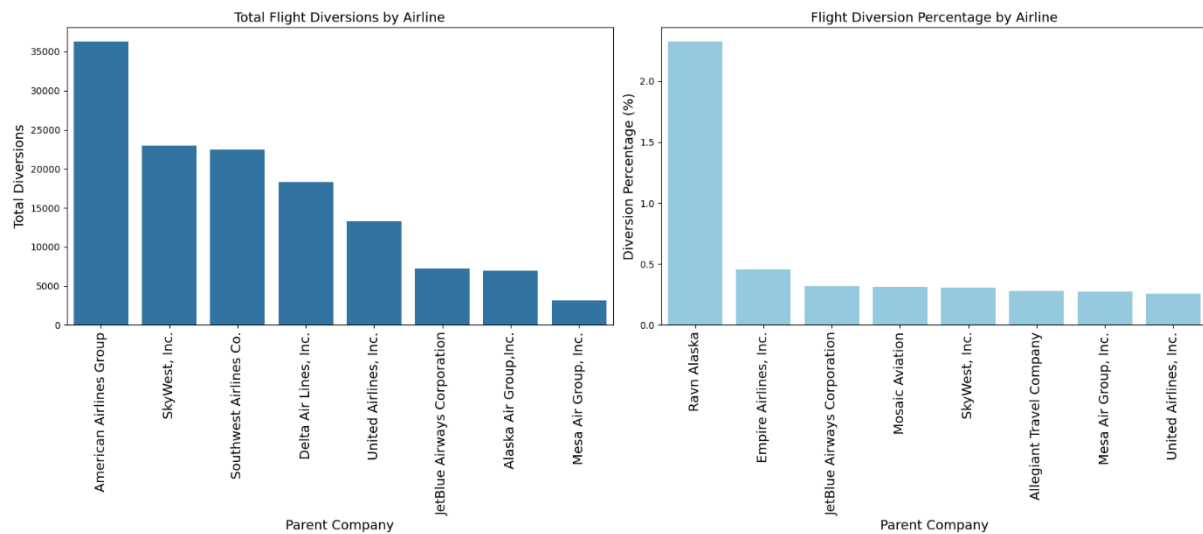
- **Percentage Graph (On Right):**

Used for fair comparisons (since larger airlines operate more flights). Highlights which airlines are less reliable, regardless of size. Better if comparing airline performance (reliability).

Interpretation:

Cancellation Percentage: Ravn Alaska has the highest percentage of cancellations, indicating a higher likelihood of flight disruptions relative to the number of flights operated.

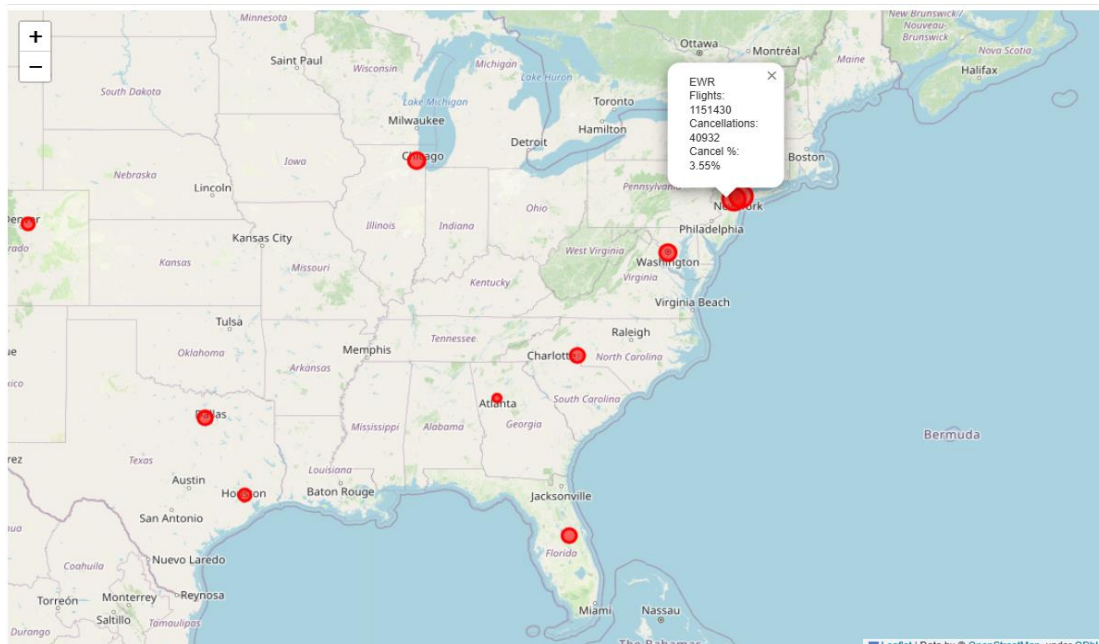
C.1.2. Which airlines experience the most Diversions



- Total Flight Diversion:
 - American Airlines Group leads with the highest number of total flight diversions.
 - SkyWest, Inc. follows closely, followed by Southwest Airlines Co.
 - This indicates that these airlines experience the most diversions overall.
- Diversion Percentage:
 - Ravn Alaska has the highest diversion percentage, followed by Empire Airlines, Inc. and JetBlue Airways Corporation.
 - This suggests that, despite having fewer total diversions, a larger proportion of their flights are diverted compared to larger carriers.

C.2. Which airports experience the most cancellations and diversions?

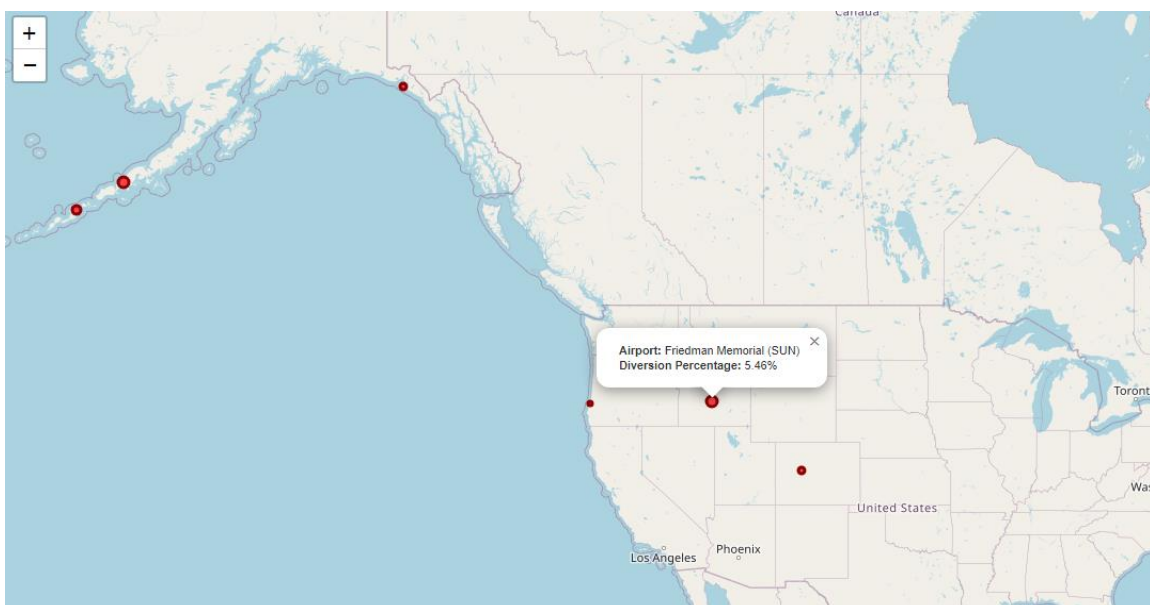
C.2.1 Airports with Maximum cancellations



- EWR (Newark Liberty International Airport) has the highest number of cancellations, with a cancellation percentage of 3.55%.
- LGA (LaGuardia Airport) follows closely with a cancellation percentage of 3.53%.
- DCA (Ronald Reagan Washington National Airport) ranks third with a cancellation percentage of 2.78%.

These airports have a higher proportion of flights cancelled, indicating that they face more operational challenges in terms of cancellations compared to other airports. While EWR and LGA show the highest cancellation rates, DCA, ORD (O'Hare International Airport), and DFW (Dallas/Fort Worth International Airport) also have notable cancellation percentages.

C.2.2 Airports with Maximum diversions



- SUN (Friedman Memorial Airport) – 5.46% Diversion Rate:

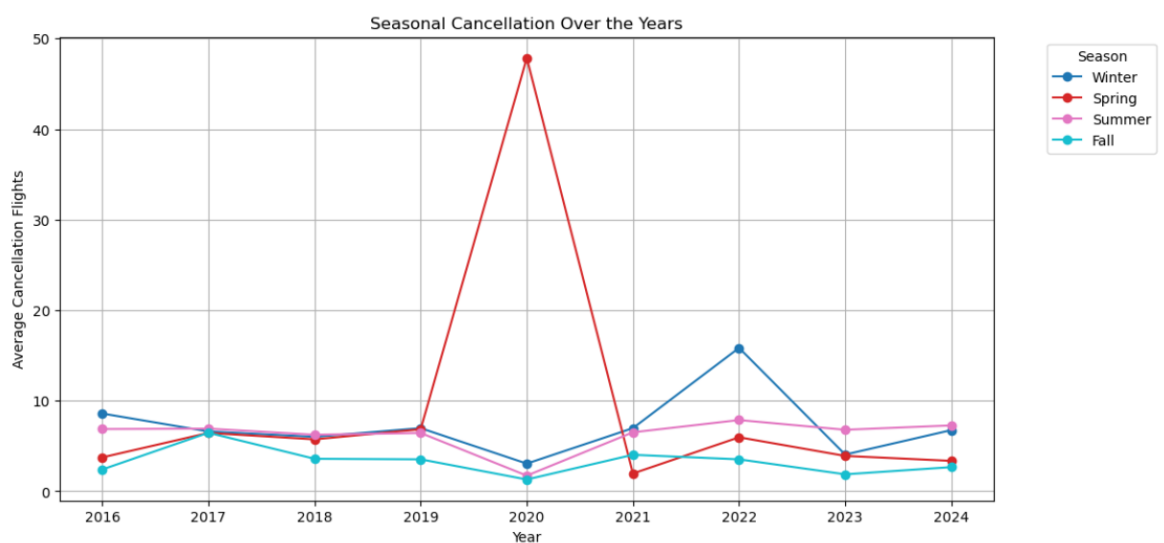
Reason: This airport is in a mountainous region, which can lead to weather challenges like poor visibility and strong winds, often causing flights to be diverted.

- CDB (Cold Bay Airport) – 4.96% Diversion Rate:

Reason: Cold Bay is in a remote part of Alaska, where harsh weather conditions, such as fog and snow, frequently impact flight operations, leading to more diversions.

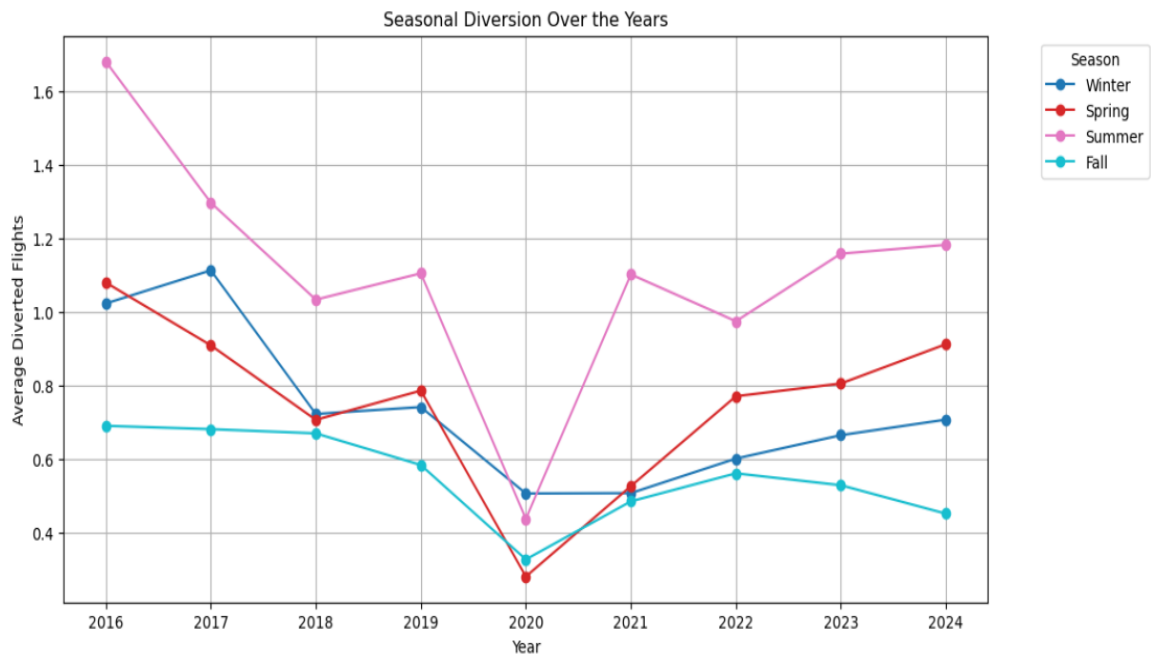
C.3. What are the seasonal trends in flight cancellations and diversions?

C.3.1. Seasonal cancellations over the years



- **COVID-19 Disruption in 2020:** The sharp spike in cancellations during the summer of 2020 is directly linked to the global impact of the COVID-19 pandemic. This period saw massive disruptions to air travel, causing a significant rise in cancellations, particularly in the summer season.
- **Winter of 2022 Surge:** The increase in cancellations during the winter of 2022 highlights the impact of external factors such as adverse weather conditions or ongoing operational challenges, possibly aggravated by the effects of the pandemic on airline operations.
- **Stabilization Post-2020:** After the dramatic spike in 2020, flight cancellations returned to more typical levels with occasional fluctuations, especially in summer and winter. The years 2023 and 2024 show a stable pattern of cancellations, reflecting a return to normalcy in air travel, albeit with some seasonal variability.

C.3.2. Seasonal diversions over the years



- **Impact of 2020:** The year 2020 shows a drastic dip in flight diversions for all seasons, especially summer, which peaks in 2020 before returning to normal levels afterward. This likely reflects the COVID-19 pandemic's significant disruption of air travel, leading to fewer flights overall.
- **Summer vs. Other Seasons:** Summer consistently has the highest average diverted flights over the years, with a notable peak in 2020. This suggests that summer experiences more disruptions on average than the other seasons, possibly due to factors like weather conditions and higher air traffic.
- **Gradual Increase Post-2020:** After 2020, there is a steady increase in diverted flights in all seasons, particularly in spring and fall, indicating that air traffic and diversions began to rise again as travel resumed post-pandemic.

5. Fazit (Evaluation/Deployment)

Our Aim is to identify the primary causes of flight delays and provide actionable recommendations to optimize flight operations for airlines and airports.

Main Causes: Following are the Main Causes for the Delays

1. Late Aircraft Delay
2. Carrier Delay
3. NAS (National Aviation System) Delay

Suggestions to Airlines:

Late Aircraft Delays (Southwest Airlines - 46% & Indigo Partners - 39%):

1. Improved Scheduling: Use data analytics and machine learning to predict delays and adjust schedules accordingly. Adding buffer time for high-traffic routes and implementing real-time adjustments can minimize the impact of late arrivals and operational disruptions.
2. Efficient Turnaround: Automate ground operations, train specialized turnaround teams, and invest in modern equipment to speed up turnaround processes. Streamlining gate management and using real-time data can further improve efficiency.

Carrier Delays (SkyWest Inc - 38% & JetBlue Airways - 36%):

1. Predictive Maintenance: Equip aircraft with IoT sensors and use predictive analytics to forecast potential issues, enabling proactive maintenance during off-peak times to avoid unplanned delays.
2. Resource Allocation: Use workforce management software to optimize ground staff scheduling and track equipment utilization. This ensures resources are available when needed, reducing turnaround delays.
3. Real-time Monitoring: Implement centralized dashboards and real-time flight monitoring tools to track operational conditions and adjust promptly to minimize delays.

NAS Delays (Spirit Airlines - 48% & Alaska Airlines - 40%):

1. Coordination with ATC: Share real-time data with ATC to adjust flight paths and departure times, improving communication and reducing delays due to airspace congestion.
2. Infrastructure Investment: Modernize runways, taxiways, and gate management systems to reduce congestion and improve the efficiency of take-offs, landings, and turnarounds.
3. Advanced Traffic Management: Implement Performance-Based Navigation (PBN) for more efficient flight paths and better air traffic flow, reducing delays and improving fuel efficiency.

Suggestions to Airports:

Late Aircraft Delays: Higher at the Following Airports:

- CLT (40%) – Charlotte Douglas International Airport
 - PHX (40%) – Phoenix Sky Harbor International Airport
 - LAS (39%) – McCarran International Airport
1. Quick Turnaround Processes: Streamline refueling, cleaning, and maintenance processes to minimize aircraft downtime and improve turnaround efficiency.
 2. Gate Availability: Enhance gate management to reduce delays in boarding and disembarking, ensuring smooth transitions between flights.

Carrier Delays: Higher at the Following Airports:

- ATL (31%) – Hartsfield-Jackson Atlanta International Airport
 - DEN (31%) – Denver International Airport
 - PHX (33%) – Phoenix Sky Harbor International Airport
 - LAX (32%) – Los Angeles International Airport
1. Crew Management: Improve crew scheduling systems to ensure timely crew availability and reduce delays.

2. Operational Efficiency: Optimize processes such as boarding, baggage handling, and fuelling to minimize delays.
3. Coordination with Airlines: Enhance communication with airlines to promptly address and resolve carrier-specific delays.

National Aviation System (NAS) Delays: Higher at the Following Airports:

- ORD (37%) – O'Hare International Airport
 - SEA (38%) – Seattle-Tacoma International Airport
 - IAH (36%) – George Bush Intercontinental Airport
1. Infrastructure Investments: Modernize and expand airport facilities to accommodate increased traffic and reduce congestion.
 2. Traffic Flow Management: Implement advanced systems to optimize aircraft movements both on the ground and in the air, improving overall traffic flow efficiency