Travel & Hospitality

Sky Analytics: Navigating the Complexities of Airline and Airport Operations – Part I

Batch: AlmaXPro53

Group: GP05

Team members:

Aparna Praturi

Ayush

Bharat Shewale

Kondle Rajshekar

Kumari Shivangi

Vantakula Gautham Naidu

Background:

SkyNet Analysis Inc. is a leading consultancy firm specializing in aviation analytics. With the aviation industry's rapid expansion and the increasing complexity of global air travel, SkyNet plays a critical role in providing data-driven insights to airlines, airports, and regulatory bodies. The company has access to extensive datasets that cover a wide range of information, including flight schedules, delays, airline operations, and airport traffic details. These datasets offer a unique opportunity to explore and understand the multifaceted nature of the aviation industry, from operational efficiency and customer satisfaction to logistical challenges and environmental impact.

Objective:

The primary objective of this case study, titled "Sky Analytics: Navigating the Complexities of Airline and Airport Operations," is to deeply analyze and interpret the extensive datasets encompassing flights, airlines, and airports - namely "flights.csv", "airlines.csv", and "airports.csv". The analysis aims to uncover critical insights into flight operations, delay patterns, airline efficiency, and airport traffic dynamics. By exploring these datasets, the study seeks to identify key factors influencing operational efficiency, understand the intricacies of flight scheduling and delays, and evaluate the performance metrics of airlines and airports. The ultimate goal is to provide strategic recommendations to enhance operational effectiveness, improve customer experiences in air travel, and contribute to the overall advancement of the aviation industry's standards and practices.

Data Sources:

1. Flights Dataset

This dataset contains detailed flight information, including timings, delays, and other flight-specific data.

- YEAR, MONTH, DAY, DAY_OF_WEEK: Date and day information for the flight.
- AIRLINE: Airline identifier.
- FLIGHT_NUMBER: Flight number.
- TAIL_NUMBER: Aircraft tail number.
- ORIGIN_AIRPORT, DESTINATION_AIRPORT: Airport codes for origin and destination.
- SCHEDULED_DEPARTURE, DEPARTURE_TIME: Scheduled and actual departure times.
- DEPARTURE_DELAY: Delay in departure (minutes).

- TAXI_OUT: The time duration between departure from the gate and wheels off.
- WHEELS_OFF, WHEELS_ON: Time when wheels were off/on the ground.
- SCHEDULED_TIME: Scheduled duration of the flight.
- ELAPSED_TIME: Actual time taken for the flight.
- AIR_TIME: Time in the air.
- DISTANCE: Distance covered by the flight.
- TAXI_IN: The time duration from wheels on to arrival at the gate.
- SCHEDULED_ARRIVAL, ARRIVAL_TIME: Scheduled and actual arrival times.
- ARRIVAL_DELAY: Delay in arrival (minutes).
- DIVERTED, CANCELLED: Indicators for diverted or cancelled flights.
- CANCELLATION_REASON: Reason for cancellation (if any).
- AIR_SYSTEM_DELAY, SECURITY_DELAY, AIRLINE_DELAY, LATE_AIRCRAFT_DELAY, WEATHER_DELAY: Different types of delays (minutes).

2. Airlines Dataset

This dataset provides information about various airlines.

- IATA_CODE: Unique airline code.
- AIRLINE: Full name of the airline.

3. Airports Dataset

This dataset contains information about various airports.

- IATA_CODE: Unique airport code.
- AIRPORT: Full name of the airport.
- CITY: City where the airport is located.
- STATE: State where the airport is located.
- COUNTRY: Country where the airport is located.
- LATITUDE, LONGITUDE: Geographic coordinates of the airport.

1. Identify and address missing data in the dataset. Are there any patterns in the missing data that can be noted?

FLIGHTS TABLE

COLUMN NAME	NUMBER OF MISSING VALUES
YEAR	0
MONTH	0
DAY	0
DAY_OF_WEEK	0
AIRLINE	0
FLIGHT_NUMBER	0
TAIL_NUMBER	457
ORIGIN_AIRPORT	0
DESTINATION_AIRPORT	0
SCHEDULED_DEPARTURE	0
DEPARTURE_TIME	2182
DEPARTURE_DELAY	2182
TAXI_OUT	2215
WHEELS_OFF	2215
SCHEDULED_TIME	0
ELAPSED_TIME	2399
AIR_TIME	2399
DISTANCE	0
WHEELS_ON	2282
TAXI_IN	2282
SCHEDULED_ARRIVAL	0
ARRIVAL_TIME	2282
ARRIVAL_DELAY	2399
DIVERTED	0
CANCELLED	0
CANCELLATION_REASON	67766
AIR_SYSTEM_DELAY	55584
SECURITY_DELAY	55584
AIRLINE_DELAY	55584
LATE_AIRCRAFT_DELAY	55584
WEATHER_DELAY	55584

Observations:

- The missing values are scattered across rows without a specific time-based, airline-based, or airport-based trend.

-

- There is no consistent grouping of missing data in particular columns or for specific airlines or dates.
- The missing data appears to be random rather than systematic, meaning there is no visible pattern that would suggest it's linked to certain conditions.

AIRPORTS TABLE

COLUMN NAME	NUMBER OF MISSING VALUES
IATA_CODE	0
AIRPORT	0
CITY	0
STATE	0
COUNTRY	0
LATITUDE	3
LONGITUDE	3

Observations:

- The missing values are only present in the LATITUDE and LONGITUDE columns.
- These missing entries are very few — only **3 rows**.
- The missing data does not appear to be associated with any specific airport or region.
- There is no time-related, location-related, or other visible pattern in the missing values.

AIRLINES TABLE

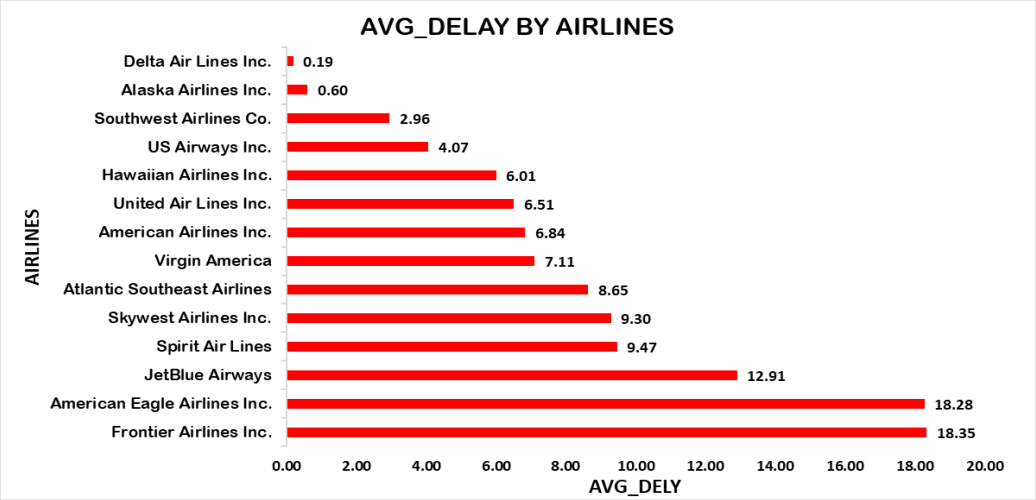
COLUMN NAME	NUMBER OF MISSING VALUES
IATA_CODE	0
AIRLINE	0

Observations:

- All rows in this dataset are complete.

2. Determine the average flight delay per airline. What are the top three airlines with the highest average delays?

AIRLINES	AVG_DELAY
Frontier Airlines Inc.	18.35
American Eagle Airlines Inc.	18.28
JetBlue Airways	12.91
Spirit Air Lines	9.47
Skywest Airlines Inc.	9.30
Atlantic Southeast Airlines	8.65
Virgin America	7.11
American Airlines Inc.	6.84
United Air Lines Inc.	6.51
Hawaiian Airlines Inc.	6.01
US Airways Inc.	4.07
Southwest Airlines Co.	2.96
Alaska Airlines Inc.	0.60
Delta Air Lines Inc.	0.19



Insights:

- **Frontier, American Eagle, and JetBlue** are the three airlines with the highest average delays.
- Frontier and American Eagle show particularly poor on-time performance (18+ minutes avg delay), while JetBlue is affected mainly by hub congestion.
- This highlights how both operational strategy and airport base location play a role in airline punctuality.

3. Identify the top 5 busiest airports based on the number of incoming and outgoing flights.

IATA_CODE	AIRPORT	CITY	STATE	COUNTRY	total_traffic	flights_incoming	flights_leaving
ATL	Hartsfield-Jackson Atlanta International Airport	Atlanta	GA	USA	8684	4303	4381
DFW	Dallas/Fort Worth International Airport	Dallas-Fort Worth	TX	USA	6739	3381	3358
ORD	Chicago O'Hare International Airport	Chicago	IL	USA	6438	3183	3255
LAX	Los Angeles International Airport	Los Angeles	CA	USA	5081	2506	2575
DEN	Denver International Airport	Denver	CO	USA	5029	2501	2528

Insights:

- ATL (Atlanta) dominates with 8,684 total flights - nearly 29% more than the second busiest airport (DFW).
- All major airports maintain balanced traffic - incoming and outgoing flights are nearly equal, indicating efficient hub operations.

4. Analyze the flight cancellations: Which airline has the highest cancellation rate, and what are the most common reasons for cancellations?

AIRLINE	COUNT of FLIGHT_NUMBER	SUM of CANCELLED	Cancellation rate
MQ	4478	413	9.222867351
B6	3282	220	6.703229738
EV	7383	324	4.388459976
US	4896	184	3.758169935
OO	7197	220	3.056829234
WN	14856	374	2.517501346
UA	5915	142	2.400676247
AA	6499	144	2.215725496
VX	641	13	2.028081123
F9	987	17	1.722391084

DL	9621	149	1.548695562
NK	1264	14	1.107594937
AS	2032	17	0.8366141732
HA	949	3	0.3161222339

Insights:

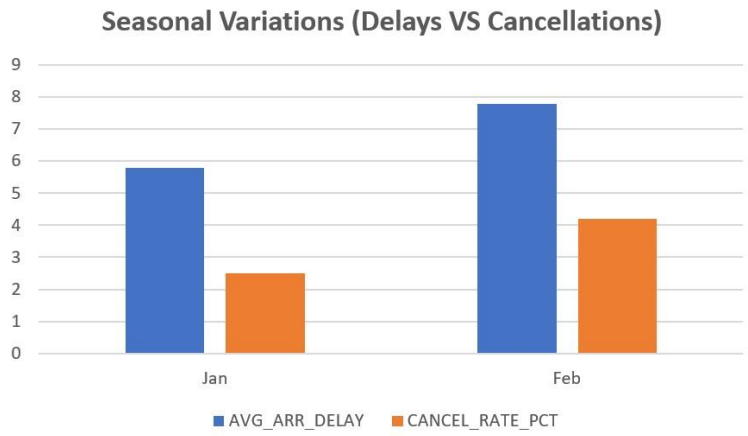
- MQ (Envoy Air) has the *worst reliability* in terms of cancellations — nearly 1 out of every 10 flights cancelled.
- B6 (JetBlue) and EV (ExpressJet) also show significant cancellation risks.
- Hawaiian Airlines (HA) is extremely reliable with cancellations being very rare.

CANCELLATION_REASON	SUM of CANCELLED
B	1453
A	430
C	351

Insights:

- B (Weather) → 1,453 cancellations (by far the most common reason).
- A (Carrier/Airline-related) → 430 cancellations.
- C (NAS – National Airspace System / ATC, congestion, airport ops) → 351 cancellations.

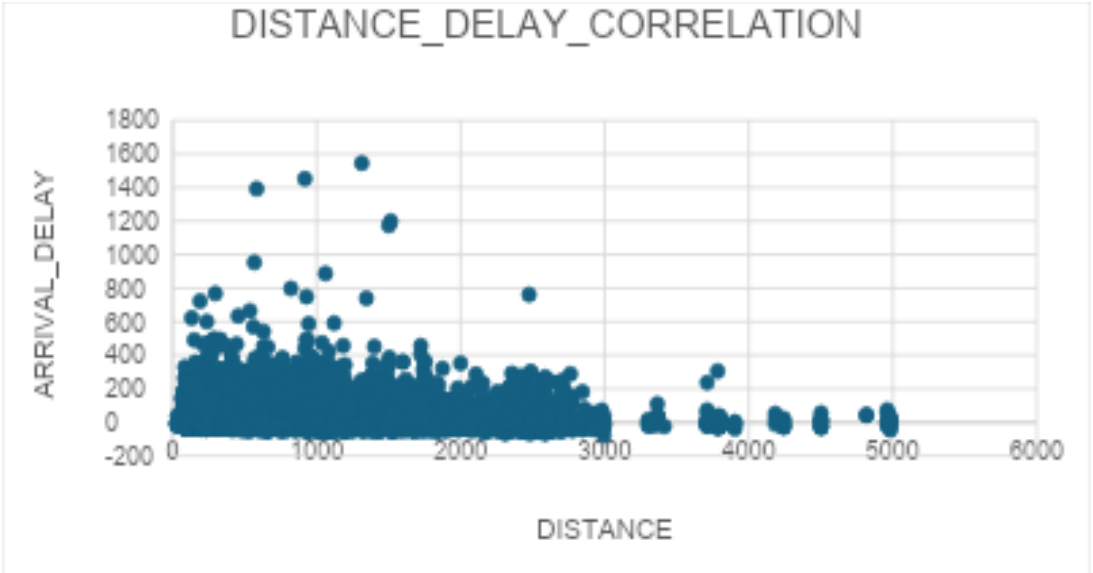
5. Examine if there are seasonal patterns in flight operations. Are certain months more prone to delays or cancellations?



Insights:

- Months covered: 1, 2.
- Average arrival delays range from 5.8 min to 7.8 min.
- Cancellation rates span 2.50% to 4.19%.

6. Investigate if there's a correlation between the distance of the flight and the length of delays. Use scatter plots for visualization.



Insights:

- Correlation value between Distance and Delays = -0.0325.
- Most flights are short-haul (under 2000 miles) sense because domestic U.S. flights are mostly short to medium range.
- Arrival delays are spread across all distances
- Delays range from 0 to 1600+ minutes, but those extreme outliers are rare.
- Most delays are under 200 minutes, regardless of distance.
- A few short-to-medium flights (~500–1500 miles) have very high delays (600–1600 min). These are likely weather disruptions or operational issues, not distance-related.
- The correlation between flight distance and arrival delays is extremely weak (r = –0.03). This indicates that distance is **not** a major factor influencing delays. Instead, external factors such as airport congestion, airline operations, and weather disruptions are the primary contributors to delays.

7. Calculate the on-time performance (percentage of flights that are not delayed) for each airline. Rank them based on this metric.

IATA_CODE	AIRLINE	percentage_flights_delayed
DL	Delta Air Lines Inc.	30.393
AS	Alaska Airlines Inc.	32.171
WN	Southwest Airlines Co.	38.493
VX	Virgin America	39.331
US	US Airways Inc.	40.004
AA	American Airlines Inc.	40.214
UA	United Air Lines Inc.	40.99
EV	Atlantic Southeast Airlines	43.234
OO	Skywest Airlines Inc.	43.307
NK	Spirit Air Lines	45.236
B6	JetBlue Airways	47.597
HA	Hawaiian Airlines Inc.	49.26
MQ	American Eagle Airlines Inc.	53.478
F9	Frontier Airlines Inc.	53.822

Insights:

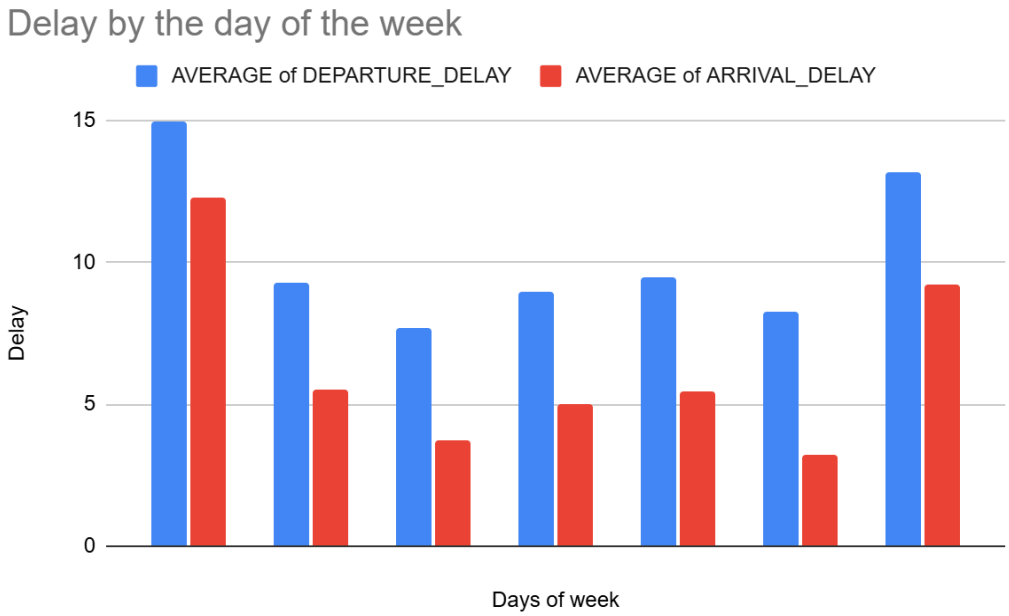
- Delta leads in punctuality with only 30.4% flights delayed - significantly better than all other airlines.

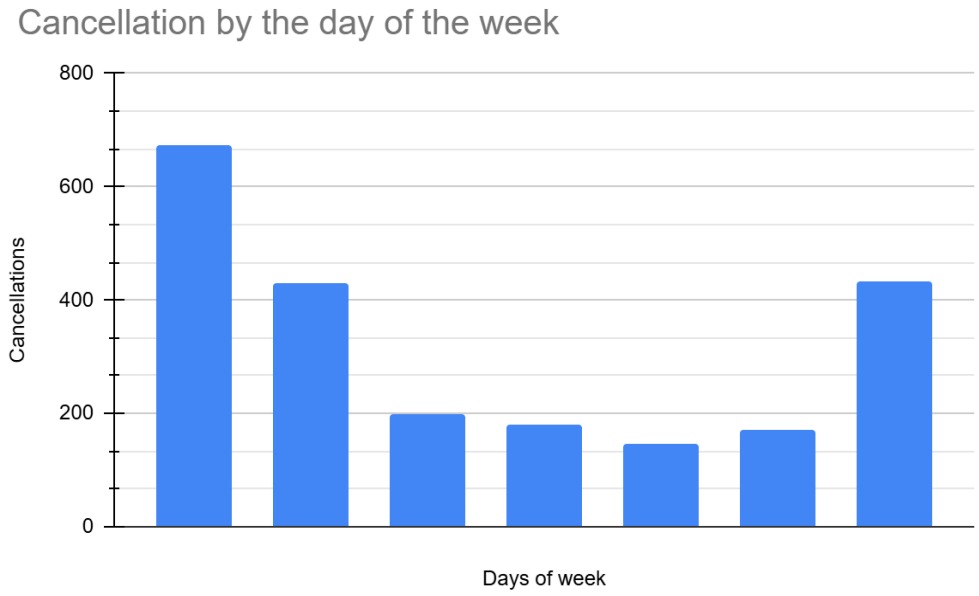
- Performance varies dramatically across carriers - ranging from Delta's 30.4% to Frontier's 53.8% delay rates.
- A wide performance gap exists - nearly a 23-percentage point difference between the best performer (Delta) and the worst (Frontier).

8. Assess how flight operations (delays, cancellations) vary by the day of the week.

DAY_OF_WEEK	SUM of CANCELLED	AVERAGE of DEPARTURE_DELAY	AVERAGE of ARRIVAL_DELAY
1	674	14.96629452	12.28394535
2	431	9.265462204	5.496277364
3	200	7.711330658	3.756293851
4	179	8.934619128	4.974716111

5	146	9.486448557	5.445959019
6	170	8.262711864	3.238984674
7	434	13.19937151	9.190710191





Insights:

1. Cancellations by Day

- Monday (1) has the highest cancellations (674).
- Wednesday (3) and Thursday (4) have the fewest cancellations (200 & 179).

- Sunday (7) also has very high cancellations (434).

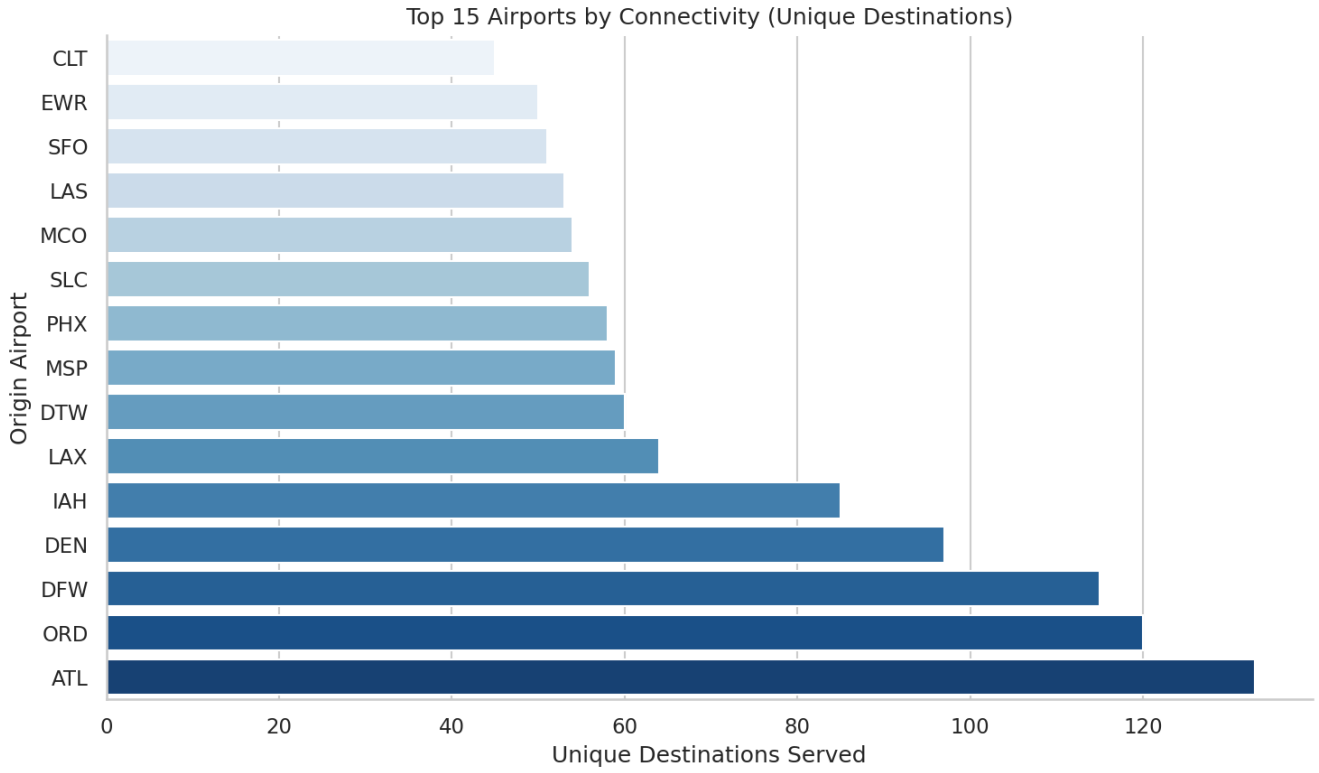
2. Departure Delays by day

- Longest average departure delays: Monday (~15 min), Sunday (~13 min)
- Shortest departure delays: Wednesday (~7.7 min) and Saturday (~8.2 min).

3. Arrival Delays by day

- Worst average arrival delays: Monday (12.3 min), Sunday (9.2 min)
    - Best (shortest) arrival delays: Saturday (3.2 min) and Wednesday (3.8 min).

9. Which airports serve as the most significant hubs in terms of connectivity (most destinations served)?



Top 5 Hubs by Connectivity

Airport (Origin)	Unique Destinations
ATL	133
ORD	120
DFW	115
DEN	97
IAH	85

Insights:

- Top hub: ATL with 133 unique destinations.
- High connectivity hubs are pivotal for network reliability; small disruptions can cascade widely.

10. Compare the scheduled flight duration versus the actual flight duration. Which airlines have the most and least deviation?

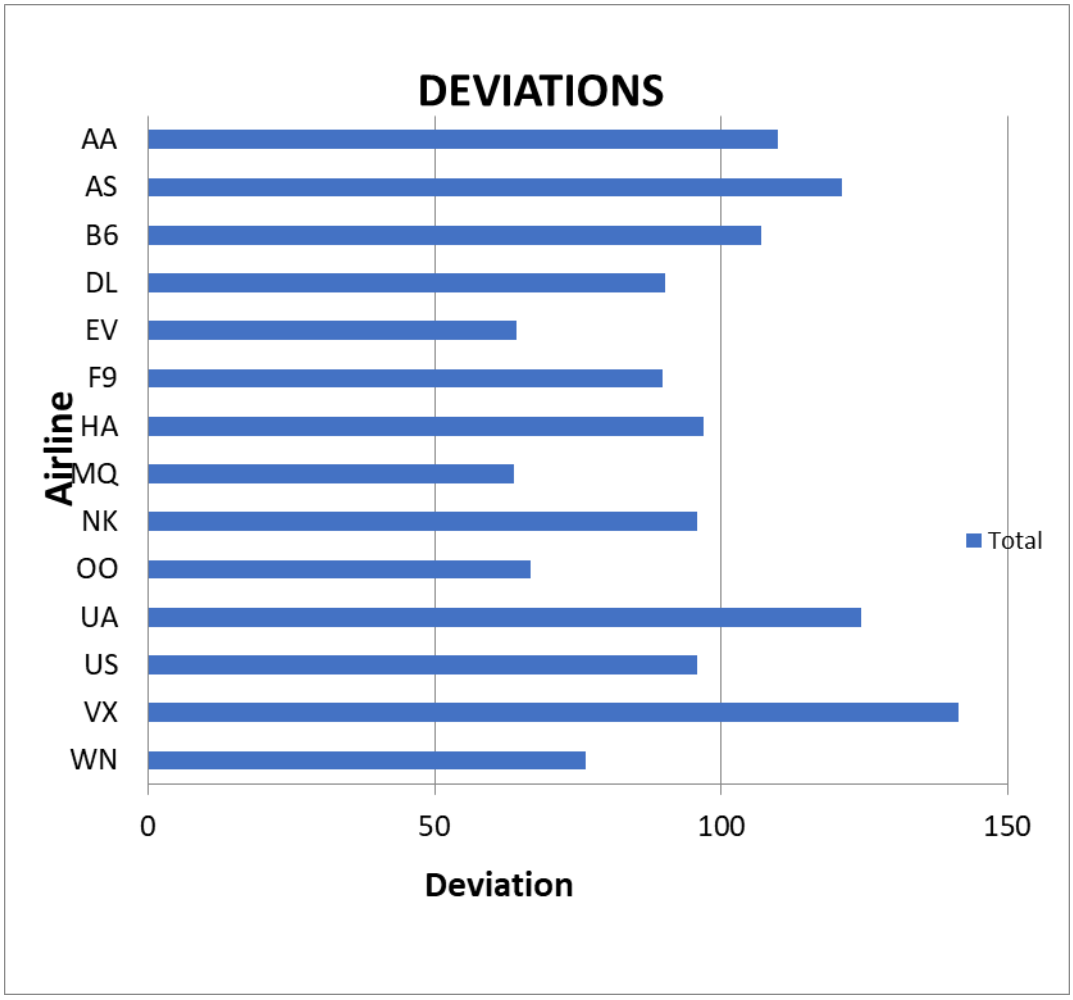
Deviation=ABS(SCHEDULED_TIME - ELAPSED_TIME)

Most Deviation

Airlines	Average of DURATION_DEVIATION
VX	141.5907643
UA	124.3874891
B6	106.9529258
AS	121.061255
AA	110.0211323

Least Deviation

Airlines	Average of DURATION_DEVIATION
EV	64.25216527
F9	89.85743802
MQ	63.91662556
OO	66.80641911
WN	76.47923013



Insights:

- **Virgin America (VX)** stands out with the **highest average deviation** at over 141 minutes, meaning flights for this airline are significantly delayed or inconsistent in duration.
- **United Airlines (UA)** and **Alaska Airlines (AS)** also face major deviations, possibly due to operational challenges, route planning, or weather-related issues.
- **JetBlue (B6)** and **American Airlines (AA)** have deviations exceeding 100 minutes, indicating they frequently experience delays.
- **American Eagle (MQ)** and **Atlantic Southeast Airlines (EV)** have the **lowest deviations**, suggesting they are the most punctual or have better operational efficiency.
- **Southwest Airlines (WN)** is also relatively punctual, making it a reliable choice for travelers.
- Airlines like **Skywest (OO)** and **Frontier (F9)** also perform better in sticking to schedules compared to other airlines.

11. Based on the tail numbers, determine which airline has the highest number of flights per aircraft, indicating fleet utilisation.

IATA_CODE	AIRLINE	flights_per_aircraft
WN	Southwest Airlines Co.	22.25
HA	Hawaiian Airlines Inc.	20.57
MQ	American Eagle Airlines Inc.	20.43
OO	Skywest Airlines Inc.	19.71
EV	Atlantic Southeast Airlines	18.92
NK	Spirit Air Lines	18.66
F9	Frontier Airlines Inc.	17.96
B6	JetBlue Airways	15.01
AS	Alaska Airlines Inc.	14.82
DL	Delta Air Lines Inc.	14.27
US	US Airways Inc.	13.9
VX	Virgin America	11.85
AA	American Airlines Inc.	11.19
UA	United Air Lines Inc.	9.97

Insights:

- **Southwest maximises aircraft utilisation** with 22.25 flights per aircraft - significantly higher than all competitors .
- **Clear utilisation tiers exist** - top performers (WN, HA, MQ) exceed 20 flights per aircraft, while the bottom tier (VX, AA, UA) stays below 12 .
- **Wide efficiency gap exists** - Southwest operates 2.2x more flights per aircraft than United (22.25 vs 9.97).

12. Using latitude and longitude data, analyze the geographical distribution of airports. Which states or regions have the highest concentration of airports?

STATE	COUNTS of AIRPORT
TX	24
CA	22
AK	19
FL	17
MI	15
NY	14
CO	10
MN	8
MT	8
NC	8

Insights:

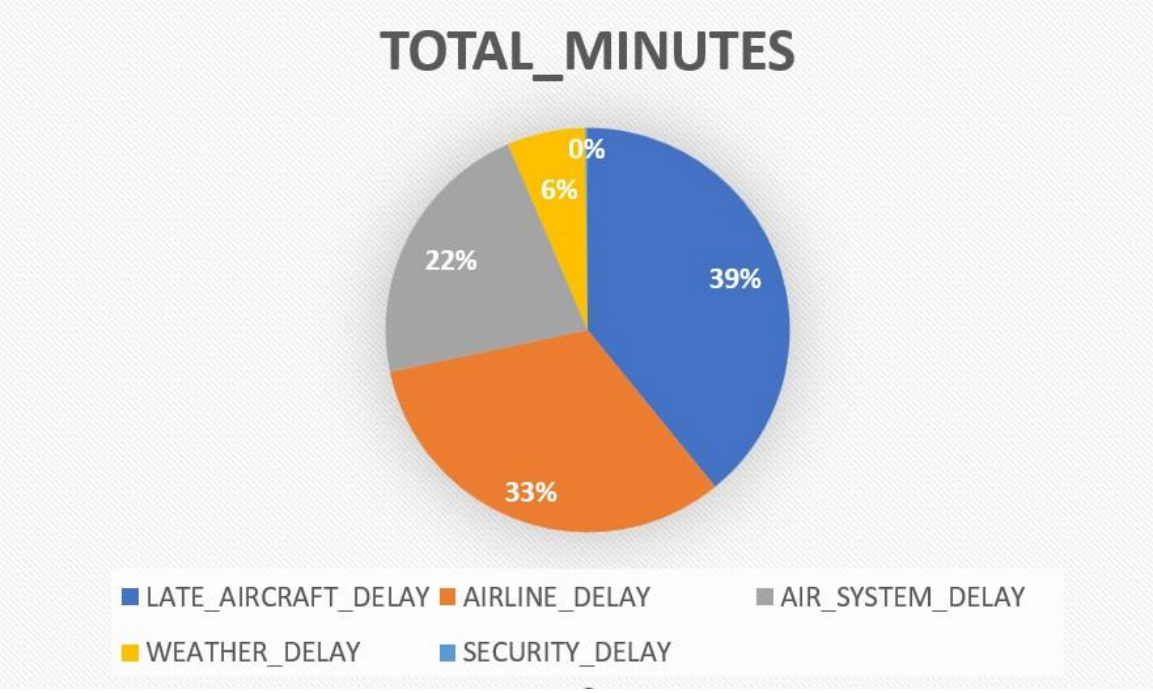
- The highest airport concentrations are found in Texas, California, and Alaska, followed by Florida and New York.
- Texas has the highest airports.

13. For flights that are delayed, break down the delay types (airline, weather, security, etc.) and analyze their proportions.

Delay Type	Percent of Delay Minutes
LATE_AIRCRAFT_DELAY	39.13%
AIRLINE_DELAY	32.55%
AIR_SYSTEM_DELAY	21.88%
WEATHER_DELAY	6.24%
SECURITY_DELAY	0.19%

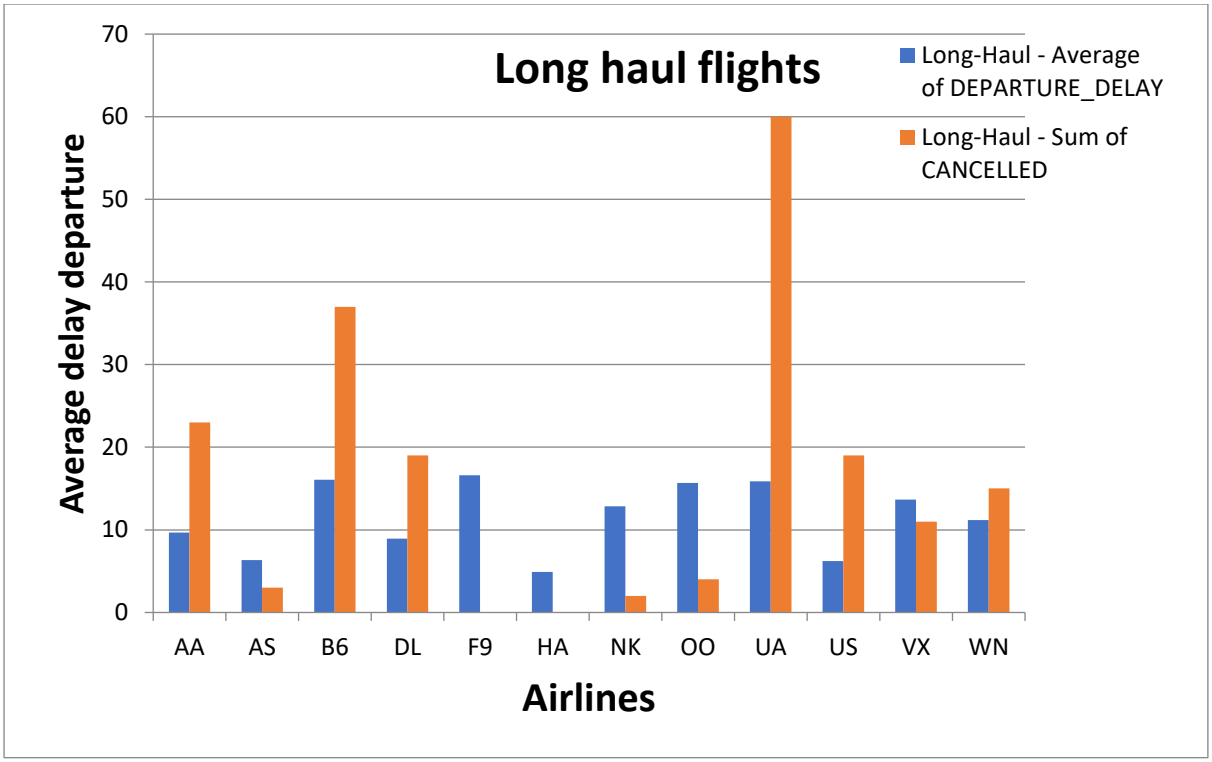
Insights:

- Top contributor: Late Aircraft Delay (39.13% of total delay minutes).
- Operational focus areas: late aircraft turnarounds, airline-controlled processes, and NAS constraints.



14. Compare the operational metrics (delays, cancellations) between long-haul and short-haul flights for different airlines.

	Long-Haul		Short-Haul		Total Average of DEPARTURE_DELAY	Total Sum of CANCELLED
Row Labels	Average of DEPARTURE_DELAY	Sum of CANCELLED	Average of DEPARTURE_DELAY	Sum of CANCELLED		
AA	9.677220756	23	10.80425206	121	10.60270525	144
AS	6.339823009	3	3.135172414	14	4.033746898	17
B6	16.06890459	37	14.97960816	183	15.18063254	220
DL	8.927083333	19	7.445704208	130	7.685877137	149
EV			9.967303609	324	9.967303609	324
F9	16.60204082	0	17.06651376	17	17.01958763	17
HA	4.888888889	0	2.058746736	3	2.597251586	3
MQ			16.58153242	413	16.58153242	413
NK	12.86131387	2	11.13656783	12	11.3256	14
OO	15.67924528	4	10.72110118	216	10.75868974	220
UA	15.88060508	60	13.32322975	82	14.14263459	142
US	6.223557692	19	5.935101725	165	5.986002121	184
VX	13.67680608	11	11.03287671	2	12.14012739	13
WN	11.17635844	15	8.97210029	359	9.131677019	374
Grand Total	**11.17527519**	**193**	**10.05311414**	**2041**	**10.18990533**	**2234**



Insights:

- **Overall Performance:** The Grand Total row shows that the average departure delay across all flights is about 10.19 minutes, and a total of 2,234 flights were canceled. Short-haul flights have a slightly lower average delay (10.05 minutes) compared to long-haul flights (11.18 minutes).

- **Most Canceled Flights:** The airline with the highest number of canceled flights is WN (374 total), followed closely by MQ (413 total), EV (324 total), and B6 (220 total).

- **Long-Haul Delays**: F9 and B6 have the highest average long-haul departure delays, at 16.60 minutes and 16.07 minutes, respectively.

- **Short-Haul Delays**: MQ has the highest average short-haul delay (16.58 minutes), followed by F9 (17.07 minutes) and B6 (14.98 minutes).

- **Best Overall Performance:** AS and HA appear to be the most reliable airlines. AS has a low total average delay (4.03 minutes) and only 17 cancellations. HA has an even lower total average delay (2.59 minutes) and only 3 cancellations.

15. Use pivot tables to summarize key operational metrics (like average delay, number of flights, cancellations) by airline and airport.

Summary metrics by airlines:

AIRLINE	NO_OF_FLIGHTS	CANCELED_FLIGHTS	AVG_DELAY
Southwest Airlines Co.	14856	374	2.955275547
Delta Air Lines Inc.	9621	149	0.194479695
Atlantic Southeast Airlines	7383	324	8.650575039
Skywest Airlines Inc.	7197	220	9.295048935
American Airlines Inc.	6499	144	6.836776534
United Air Lines Inc.	5915	142	6.514509123
US Airways Inc.	4896	184	4.067205444
American Eagle Airlines Inc.	4478	413	18.28021707
JetBlue Airways	3282	220	12.91042824
Alaska Airlines Inc.	2032	17	0.599601594
Spirit Air Lines	1264	14	9.473178543
Frontier Airlines Inc.	987	17	18.34504132
Hawaiian Airlines Inc.	949	3	6.012684989
Virgin America	641	13	7.111464968

Insights:

- **Busiest Airlines Dominate the Market**: Southwest (14,856 flights), Delta (9,621), Atlantic Southeast (7,383), SkyWest (7,197), American (6,499) are the top operators. These carriers handle the majority of domestic US flights, which naturally leads to more cancellations in absolute numbers, even if their cancellation rate is lower.
- Regional Carriers Struggle More.
- American Eagle → highest cancellations (9.2%) and one of the worst delays (18 min).
- SkyWest, Atlantic Southeast → moderate flight volume but consistently poor average delays (8–9 min). Reason: small aircraft + weather exposure + limited backup capacity.
- **Top On-Time Performers (Reliability Leaders):** Delta (0.19 min avg delay), Alaska (0.6 min), Southwest (3 min). Despite their scale, they manage very low delays and cancellations, making them most reliable for passengers.

Summary metrics by airports:

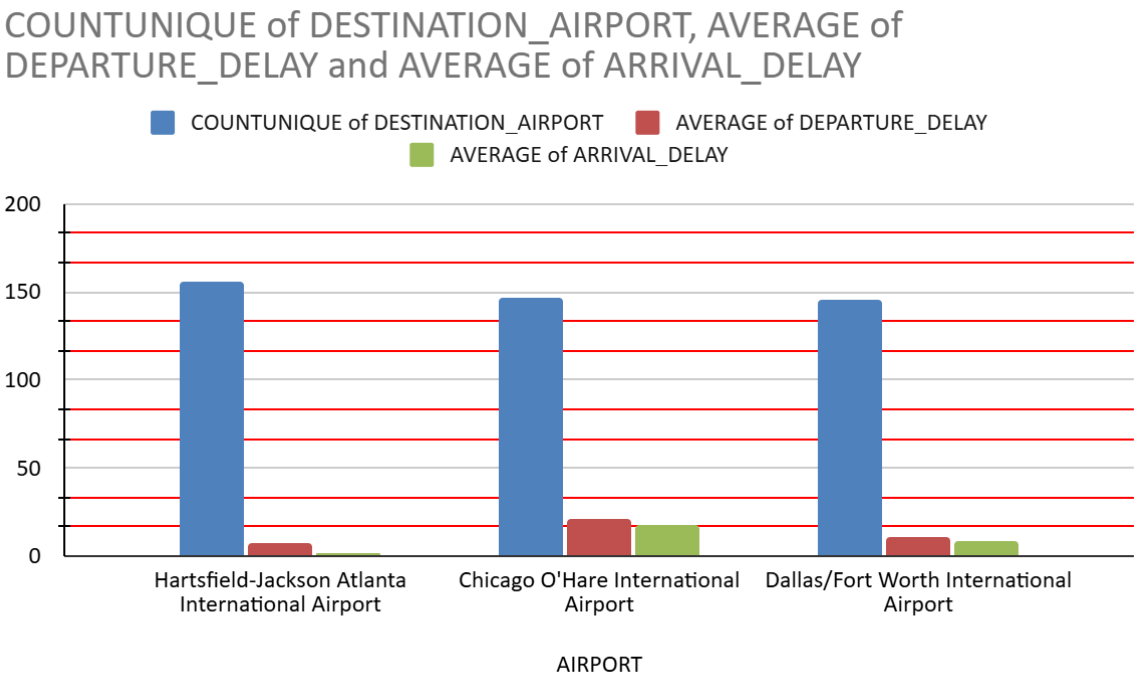
AIRPORTS	NO_OF_FLIGHTS	CACELLED_FLIGHTS	AVG_DELAY
Hartsfield-Jackson Atlanta International Airport	10128823	50	1.84
Chicago O'Hare International Airport	9680853	225	18.14
Dallas/Fort Worth International Airport	7190711	80	8.61
Denver International Airport	6479173	40	9.39
LaGuardia Airport (Marine Air Terminal)	3117892	181	13.32
Gen. Edward Lawrence Logan International Airport	1771793	148	14.29
Newark Liberty International Airport	2937898	99	6.10
Columbia Regional Airport	50839	1	55.43
Del Norte County Airport (Jack McNamara Field)	63979	1	52.82
Central Nebraska Regional Airport	37823	0	45.67

Insights:

- **Mega-Hubs Handle the Majority of Flights:**Atlanta (10M+), Chicago O'Hare (9.6M), Dallas-Fort Worth (7.1M), Denver (6.4M) are the busiest airports in the U.S. These airports serve as major domestic & international hubs, explaining their massive flight volumes.
- **Cancellations Concentrated in Large Hubs:** Chicago O'Hare has 225 cancellations – the highest in absolute numbers. However, cancellation rates are still very low relative to total flights (e.g., 225 out of 9.6M). This shows that scale drives absolute cancellations, not necessarily poor operations.
- **Delay Patterns Differ by Airport Size:**
  - Large hubs: Chicago O'Hare (18.1 min) and LaGuardia (13.3 min) suffer significant delays, likely due to air traffic congestion and slot restrictions. Atlanta (1.8 min) and Dallas (8.6 min) perform much better despite handling more flights.
  - Smaller regional airports: Columbia (55.4 min), Del Norte (52.8 min), Central Nebraska (45.7 min) have extremely high delays. Reason: weather sensitivity, fewer backup flights, limited infrastructure.

16. Combine the "flights.csv" dataset with the "airports.csv" based on airport codes. Use the merged dataset to identify the top 3 airports in terms of flight connectivity (number of unique destinations served) and analyze their average delay times (both departure and arrival).

AIRPORT	COUNTUNIQUE of DESTINATION_AIRPORT	AVERAGE of DEPARTURE_DELAY	AVERAGE of ARRIVAL_DELAY
Hartsfield-Jackson Atlanta International Airport	156	7.632587859	1.842971429
Chicago O'Hare International Airport	147	20.94885145	18.1419335
Dallas/Fort Worth International Airport	145	10.69354359	8.610762332



Insights:

1. Connectivity (COUNTUNIQUE of DESTINATION_AIRPORT)
    - ATL (Atlanta) serves the most destinations (156) → this makes sense because it's the world's busiest hub.
    - ORD (Chicago O'Hare) is second (147 destinations).
    - DFW (Dallas/Fort Worth) follows closely with 145 destinations.

    These are the top 3 most connected airports in the dataset.

2. Departure Delays (Average Minutes Late at Takeoff)
    - ATL: 7.6 mins → relatively efficient for such a huge hub.

- ORD: 21 mins → much higher; likely due to congestion + Chicago weather.
- DFW: 10.7 mins → moderate.

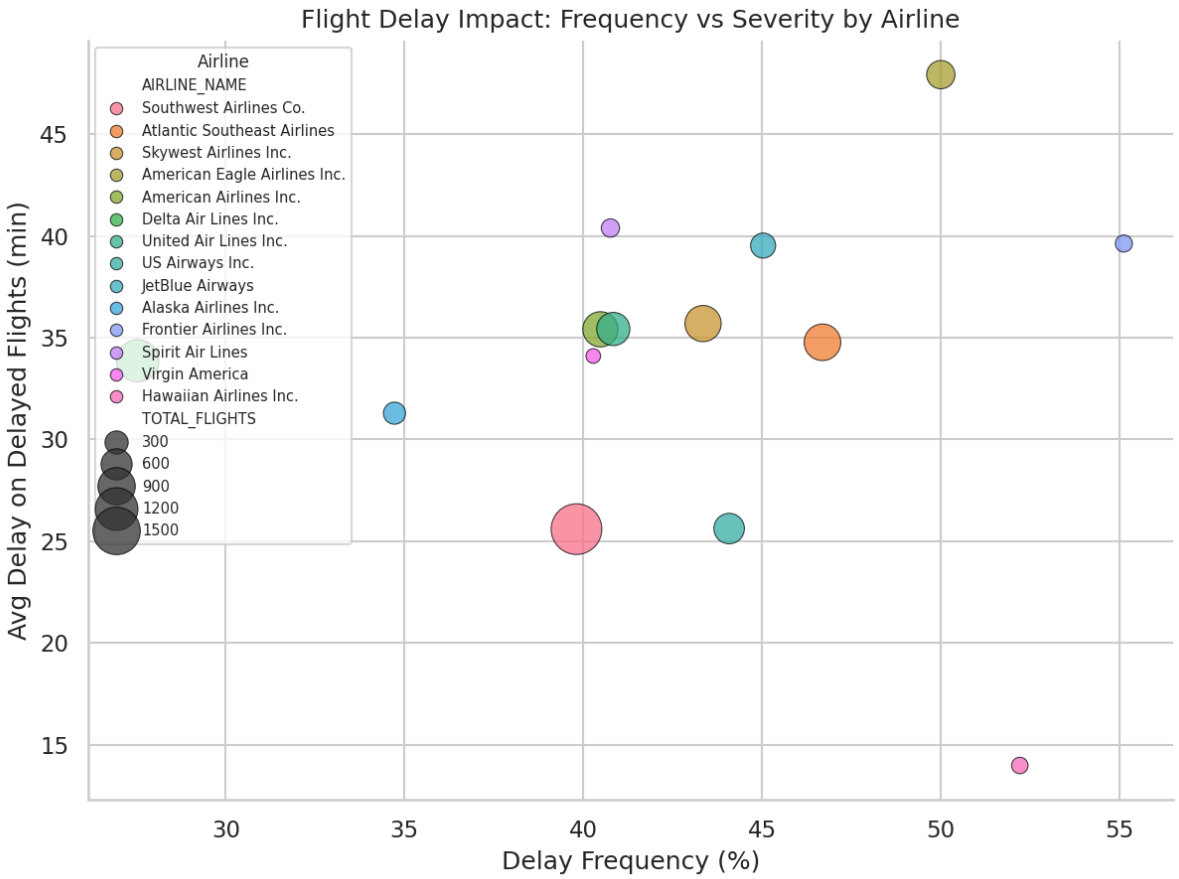
3. Arrival Delays (Average Minutes Late on Landing)
   - ATL: 1.8 mins → almost on time arrivals.
   - ORD: 18.1 mins → very high delays, consistent with departure issues.
     DFW: 8.6 mins → some delays but less severe.

17. Calculate the cumulative impact of delays for each airline. Consider both the frequency of delays and the average delay time. How do these factors combine to affect overall airline performance

- Impact Score = Number of delayed flights × average delay for delayed flights.
- Higher frequency and/or longer delays raise the overall impact.

Top 5 Airlines by Delay Impact

Airline	Total Flights	Delay Frequency (%)	Avg Delay on Delayed (min)	Impact Score
Southwest Airlines Co.	1,723	39.8%	25.6	17,559
Atlantic Southeast Airlines	861	46.7%	34.8	13,980
Skywest Airlines Inc.	842	43.4%	35.7	13,029
American Eagle Airlines Inc.	484	50.0%	47.9	11,598
American Airlines Inc.	793	40.5%	35.4	11,368



Insights:

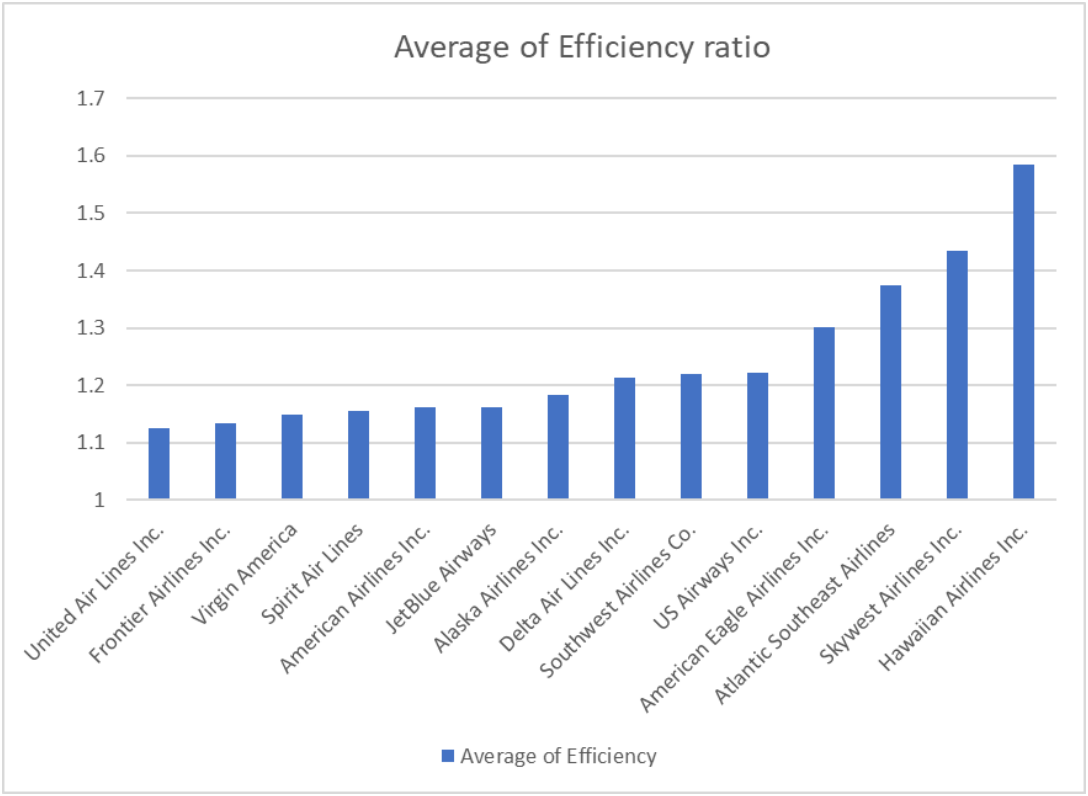
- Highest overall delay impact: Southwest Airlines Co. (Impact Score 17,559, Freq 39.8%, Avg Delay 25.6m); Atlantic Southeast Airlines (Impact Score 13,980, Freq 46.7%, Avg Delay 34.8m); Skywest Airlines Inc. (Impact Score 13,029, Freq 43.4%, Avg Delay 35.7m).

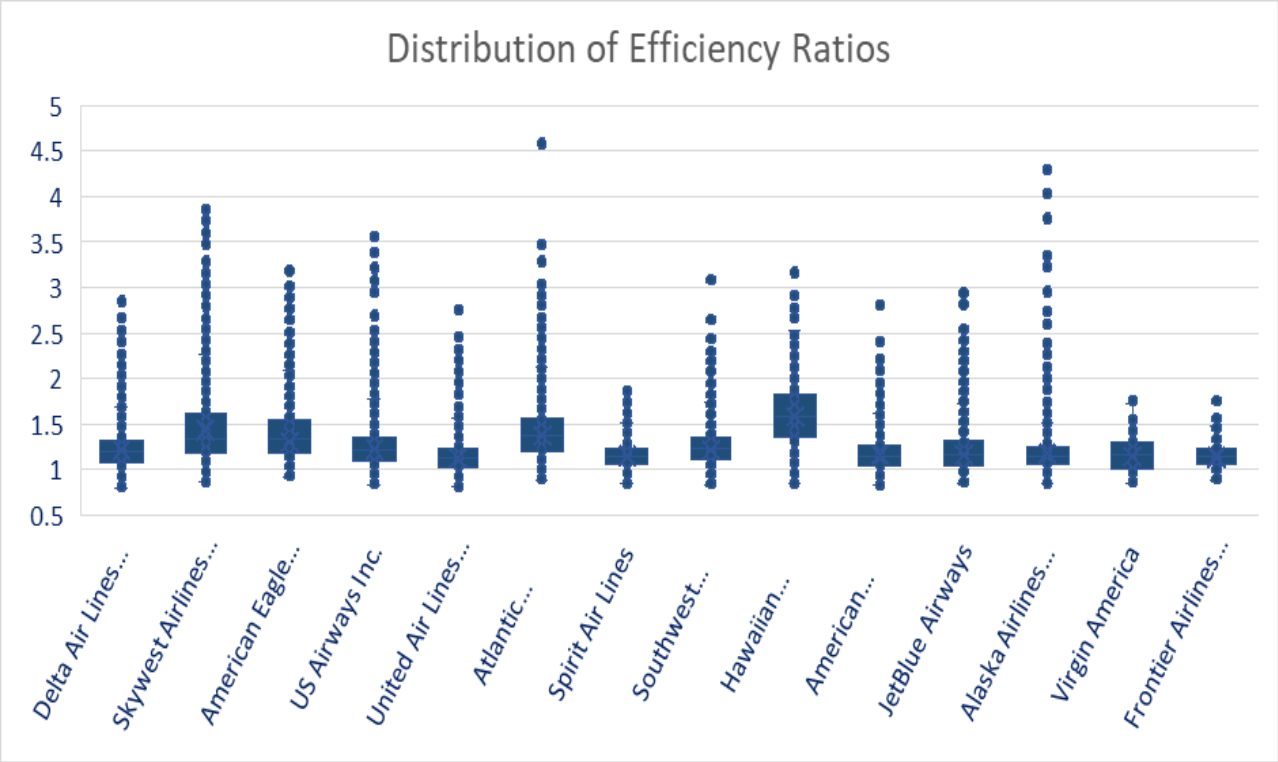
18. Using the distance data from "flights.csv" and geographical coordinates from "airports.csv", calculate the efficiency of various flight paths. Determine if there are significant differences in the efficiency of flights (measured as a ratio of actual flight time to the shortest possible time based on distance) for different airlines.

- Efficiency Ratio=(500Distance×60)Actual Air Time
- Assume 500 mph as standard cruising speed.

- Meaning:
  - =1 → Flight exactly matches expected time.
  - <1 → More efficient (faster than expected).
  - >1 → Less efficient (slower than expected).

Airline Name	Airline code	Optimal Flight Path Efficiency Ratio
United Air Lines Inc.	UA	1.125229794
Frontier Airlines Inc.	F9	1.133065193
Virgin America	VX	1.148814984
Spirit Air Lines	NK	1.15522082
American Airlines Inc.	AA	1.161015172
JetBlue Airways	B6	1.162985012
Alaska Airlines Inc.	AS	1.183093125
Delta Air Lines Inc.	DL	1.213681361
Southwest Airlines Co.	WN	1.218988812
US Airways Inc.	US	1.222848683
American Eagle Airlines Inc.	MQ	1.30188268
Atlantic Southeast Airlines	EV	1.374419441
Skywest Airlines Inc.	OO	1.433905333
Hawaiian Airlines Inc.	HA	1.584951788





Insights:

- Most airlines cluster near **1.1–1.2** (slightly slower than expected).
- **Hawaiian Airlines** has the highest ratio (~1.6) → long-haul Pacific routes slower relative to baseline.
- **SkyWest & Atlantic Southeast** also show higher ratios and wider spread → less predictable efficiency.
- **Delta & Frontier**: lower ratios and tighter distribution → more consistent, efficient operations.

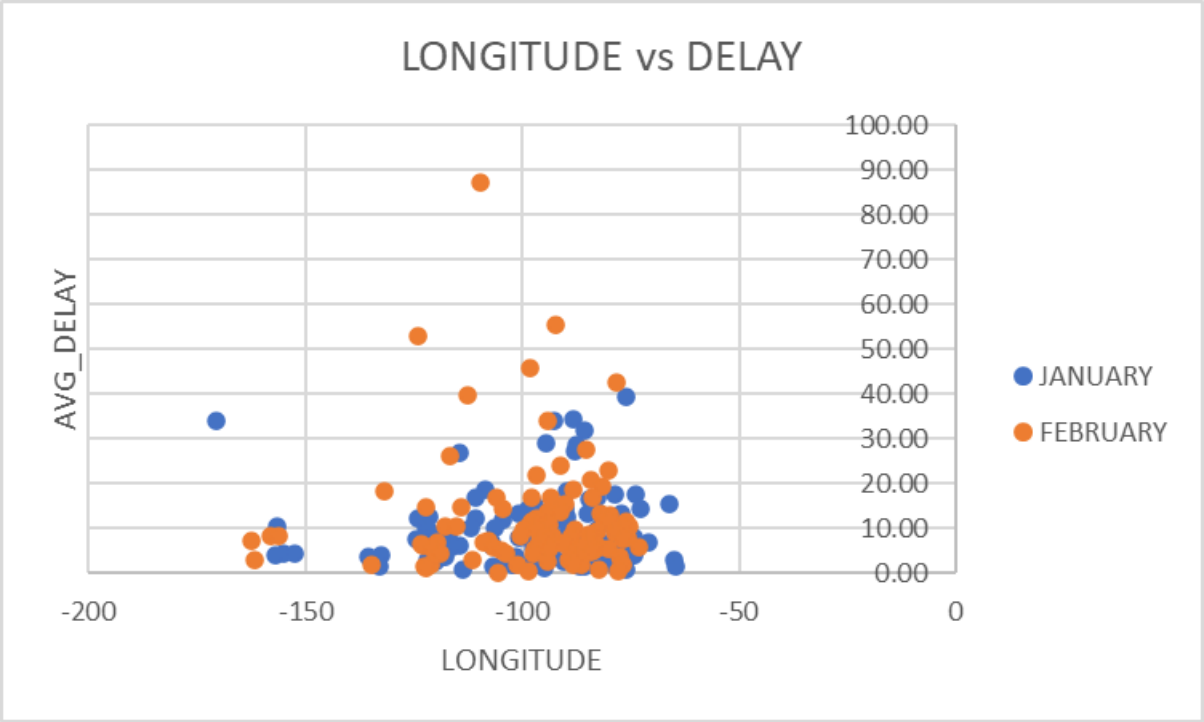
19: Investigate if there's a correlation between airport performance (in terms of delays) and environmental factors like location (latitude and longitude from "airports.csv") and time of year (seasonal weather conditions). Are certain airports more prone to delays due to their geographical location and the associated weather patterns?

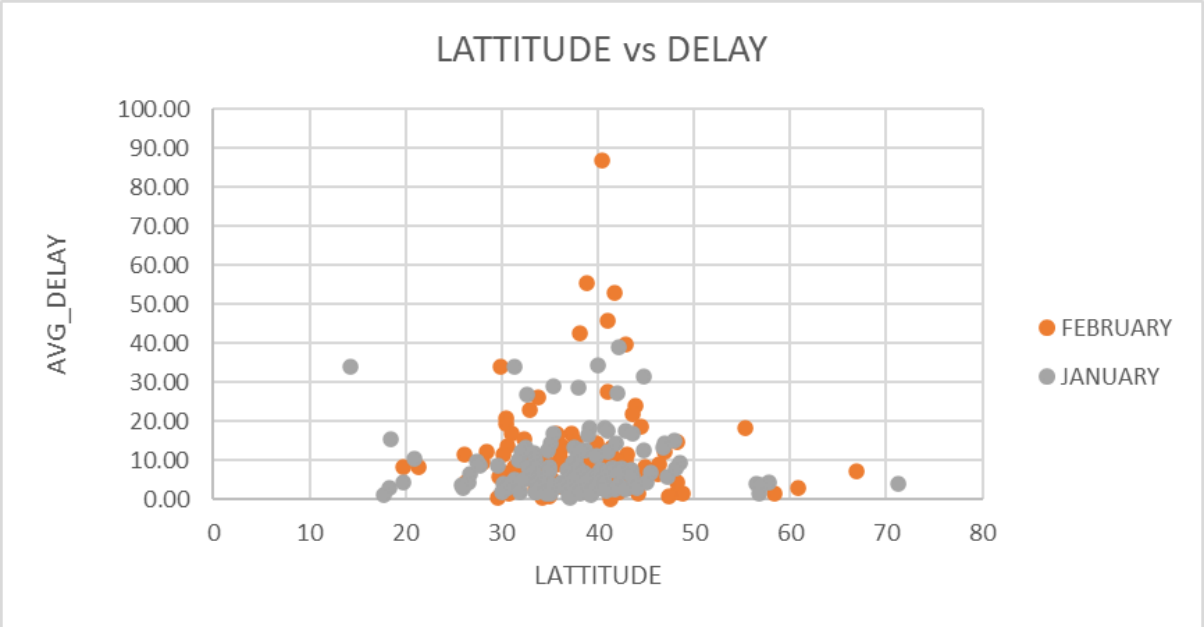
Top 5 Best Airports (Lowest Average Delay – Jan & Feb)

Rank	Airport	State/Region	Avg Delay (min)	Reason (Winter Factors)
1	Laramie Regional Airport	St George	0.6	Mild coastal climate
2	Wilmington International Airport	Norfolk	0.8	warm winters
3	Dallas Love Field (DAL)	Texas (South)	4.3	Mild winters, good ops
4	San Antonio International Airport	Kansas city	1.2	Warmer climate
5	Miami Intl (MIA)	Florida	5.8	Tropical, no snow

Top 5 Worst Airports (Highest Average Delay – Jan & Feb)

Rank	Airport	State/Region	Avg Delay (min)	Reason (Winter Factors)
1	Valdez (VDZ)	Vernal	87.0	Severe winter weather + remote location
2	Columbia Regional (COU)	columbia (Midwest)	55.4	Snow & ice in central U.S.
3	Del Norte County Airport	Crescent City (CEC)	52.8	Coastal fog + storms
4	Central Nebraska Regional	Grand island	45.7	Blizzards, winter storms
5	Charlottesville-Albemarle (CHO)	charlottesville	42.7	Winter storms affecting East





Insights:

1. **Delay vs Latitude:**
   - From this plot, higher delays cluster around ~35°–45° latitude → this is the U.S. "snow belt" (Midwest/Northeast).
   - Very high latitudes (Alaska) do not always show extreme delays because these airports handle fewer flights, but when weather hits, it can be disruptive.

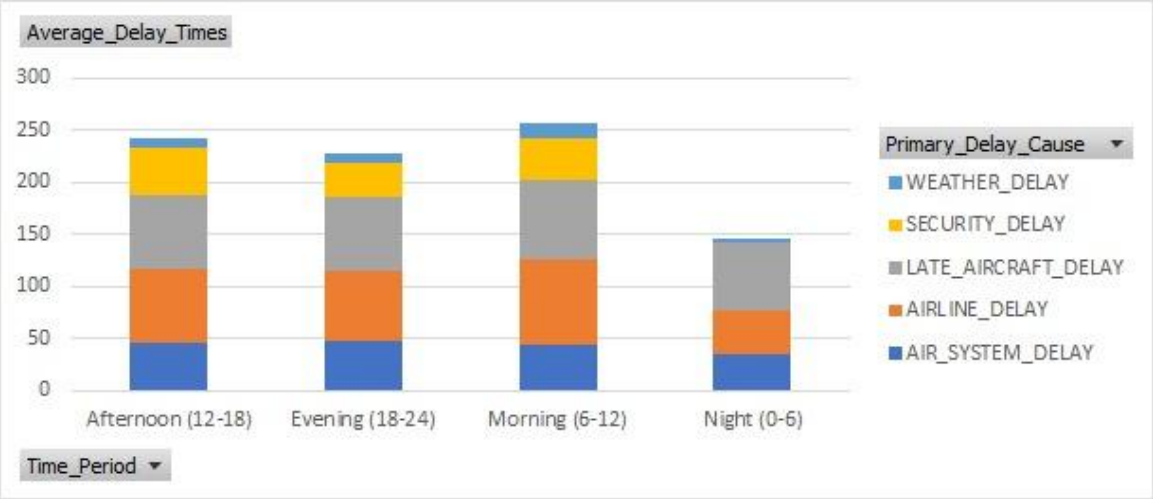
2. **Delay vs Longitude:**
    - Longitude mainly reflects east-west geography.
    - High-delay airports appear across the central/eastern longitudes (–70 to –110), which align with busy hubs in the Midwest and Northeast.

3. In Jan–Feb, smaller Midwest and Northeast airports such as Valdez, Columbia, and Dubuque show the highest delays due to snowstorms. In contrast, southern airports like Miami, Houston, and San Diego perform much better with minimal winter disruption.

20. Use nested functions to analyse the primary cause of delays for each airline. Determine if the predominant cause of delay (such as airline delay, weather delay, or security delay) varies by airline and time of day.

The heatmap provides a clear comparison of delay reasons across time periods, with colour intensity showing severity. It makes it easy to spot peak delay causes and times at a glance.

Time Period **vs** Delay Reason				
Average_Delay_Times	Time Periods			
Reasons For Delay	Afternoon (12-18)	Evening (18-24)	Morning (6-12)	Night (0-6)
AIR_SYSTEM_DELAY	45	47	43	34
AIRLINE_DELAY	72	67	83	43
LATE_AIRCRAFT_DELAY	70	71	76	66
SECURITY_DELAY	47	34	40	
WEATHER_DELAY	9	9	14	2

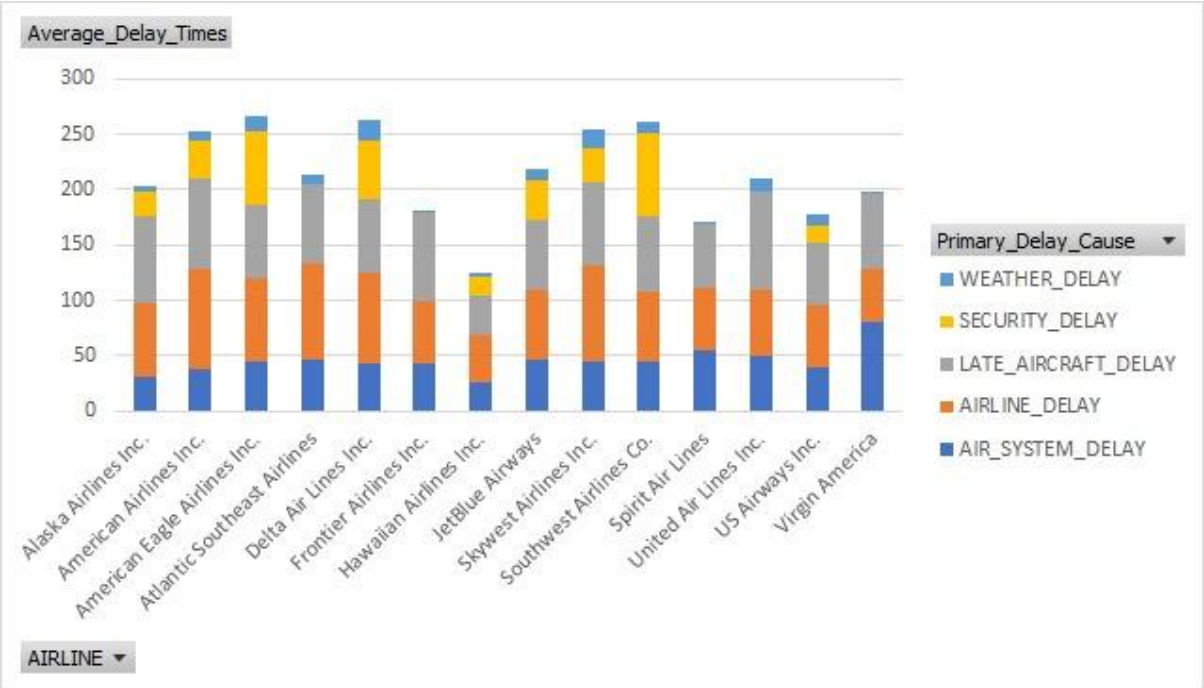


Insights:

- Airline and Late Aircraft delays dominate, peaking in the morning (6–12) at 75–83 minutes, far higher than other causes.
- Weather delays remain minimal (<15 minutes), showing negligible impact compared to operational issues.
- Air System and Security delays stay moderate and stable, averaging 30–50 minutes without sharp fluctuations.

Airlines vs Reason For Delay HeatMap Analysis					
Average_Delay_Times	Reason For Dela				
Airlines	AIR_SYSTEM_DELA	AIRLINE_DELAY	LATE_AIRCRAFT_	SECURITY_[	WEATHER_DELAY
Alaska Airlines Inc.	31	66	79	23	4
American Airlines Inc.	38	90	82	35	8
American Eagle Airlines Inc.	44	76	66	67	14
Atlantic Southeast Airlines	47	86	72		9
Delta Air Lines Inc.	43	81	67	53	20
Frontier Airlines Inc.	42	57	81		1
Hawaiian Airlines Inc.	25	44	36	17	2
JetBlue Airways	46	63	64	36	9
Skywest Airlines Inc.	45	86	75	31	17
Southwest Airlines Co.	44	64	68	75	10
Spirit Air Lines	54	57	58		1
United Air Lines Inc.	50	60	89		12
US Airways Inc.	39	57	57	15	9
Virgin America	81	47	69		2

- The heatmap highlights airline-wise delay reasons, with colour intensity reflecting severity. It helps quickly identify airlines facing the highest delays and the major contributors.



Insights:

- American Airlines and American Eagle face the most severe delays (80–90 mins) mainly due to airline and late aircraft issues, showing significant operational inefficiencies.
- Virgin America shows extreme system delays (81 mins), far higher than peers, indicating infrastructure or scheduling bottlenecks.
- Weather delays remain negligible (<20 mins) across all airlines, confirming that operational causes outweigh environmental factors.

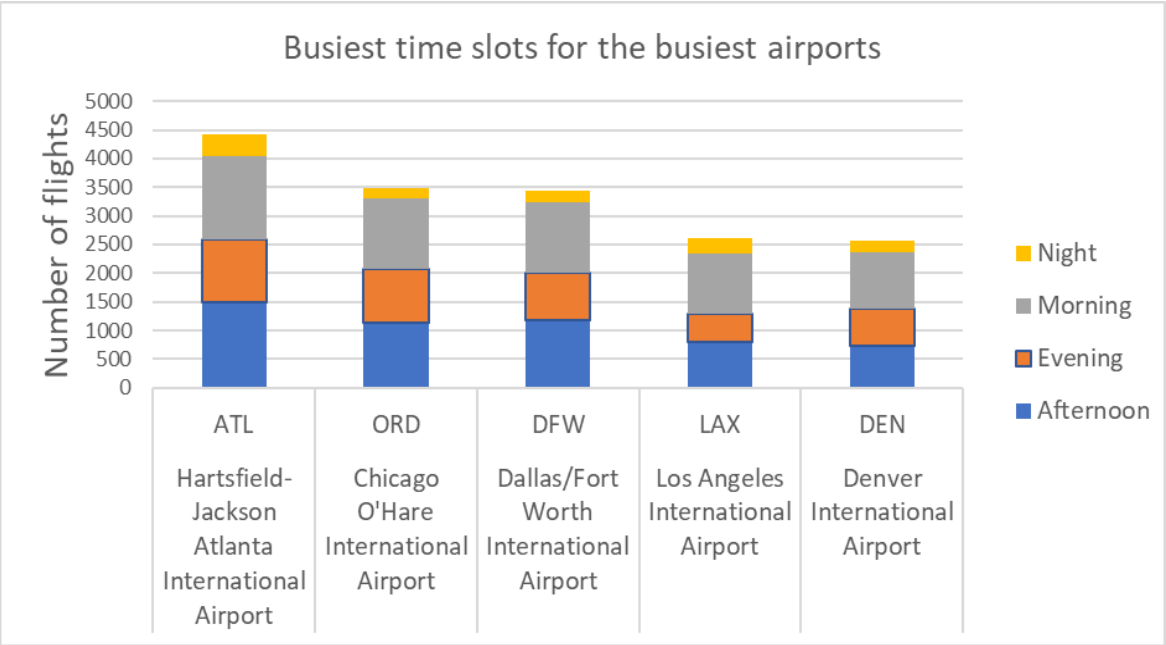
21. Analyze the flight frequencies to determine the peak operating hours for major airports. (Use a combination of Excel functions to categorize flights into different time slots (e.g., morning, afternoon, evening, night) and calculate the number of flights in each slot for the top 5 busiest airports.)

Top 5 busiest airports:

Airport Name	AIrport code	Count of ORIGIN_AIRPORT
Hartsfield-Jackson Atlanta International Airport	ATL	4431
Chicago O'Hare International Airport	ORD	3480
Dallas/Fort Worth International Airport	DFW	3438
Los Angeles International Airport	LAX	2618
Denver International Airport	DEN	2568
	Grand Total	**16535**

Distribution of flights in various timeslots for top 5 busiest airports:

Airport Name	Airport code	Afternoon	Evening	Morning	Night	Grand Total
Hartsfield-Jackson Atlanta International Airport	ATL	1488	1095	1471	377	4431
Chicago O'Hare International Airport	ORD	1141	934	1223	182	3480
Dallas/Fort Worth International Airport	DFW	1178	834	1221	205	3438
Los Angeles International Airport	LAX	796	500	1050	272	2618
Denver International Airport	DEN	729	653	989	197	2568
	Grand Total	5332	4016	5954	1233	16535



Insights:

- **Overall Busiest Airport:** Hartsfield-Jackson Atlanta International Airport (ATL) appears to be the busiest overall, with the highest total number of flights across all time slots.
- **Busiest Time Slot:** The afternoon is consistently the busiest time of day for all five airports shown. The blue "Afternoon" bar is the largest component of each airport's total flights.
- **Time Slot Comparisons:** The evening and morning time slots have a similar number of flights at each airport, though there is some variation. The night time slot is the least busy period for all airports by a significant margin.
- **Airport Specifics:**
    - **ATL (Atlanta):** Has a notably high volume of flights in all time slots, particularly in the afternoon and evening.
    - **ORD (Chicago) and DFW (Dallas/Fort Worth):** These two airports have a very similar total number of flights and a similar distribution across the time slots, with afternoon being the busiest.
    - **LAX (Los Angeles) and DEN (Denver):** These two airports are less busy than the top three, with LAX having slightly more flights than DEN. Their flight distribution across the time slots is also quite similar.