Project on

Flight Delay

Agenda

1. Introduction of Project
2. Meta Data
3. Data Pre-Processing
4. Analysis and Interpretation
5. Conclusion

Introduction

Reasons for domestic flight delays according to the US Federal Aviation Administration:

- Passenger-related delays
- Airline-related delays
- Airport-related delays
- Weather-related delays
- Other delays

Metadata

- Year
- Month
- DayofMonth
- DayOfWeek
- DepTime
- CRSDepTime
- ArrTime
- CRSArrTime
- UniqueCarrier
- FlightNum
- ActualElapsedTime
- CRSElapsedTime
- AirTime
- ArrDelay
- DepDelay
- Origin
- Dest
- Distance
- TaxiIn
- TaxiOut
- Cancelled
- CancellationCode
- Diverted
- CarrierDelay
- WeatherDelay
- NASDelay
- SecurityDelay
- LateAircraftDelay

Data Pre- Processing

- - Find duplicate values
- - Deal with duplicate values
- - Find null values
- - Deal with null values
- - Data Formatting (Datetime)

Identifying and removing null values

Step-1

	Null Percentage
LateAircraftDelay	35.59%
SecurityDelay	35.59%
NASDelay	35.59%
WeatherDelay	35.59%
CarrierDelay	35.59%
ActualElapsedTime	0.43%
AirTime	0.43%
ArrDelay	0.43%
TaxiIn	0.37%
ArrTime	0.37%
TaxiOut	0.02%
CRSElapsedTime	0.01%
CRSArrTime	0.0%
Month	0.0%
DayofMonth	0.0%
DayOfWeek	0.0%
DepTime	0.0%
Diverted	0.0%
CancellationCode	0.0%
Cancelled	0.0%
Distance	0.0%
UniqueCarrier	0.0%
Dest	0.0%
Origin	0.0%
DepDelay	0.0%
Year	0.0%
CRSDepTime	0.0%
TailNum	0.0%
FlightNum	0.0%
Unnamed: 0	0.0%

Step-2

```
# Fill null values in 'CarrierDelay', 'WeatherDelay', 'NASDelay', 'SecurityDelay', 'LateAircraftDelay' with 0
df['CarrierDelay'].fillna(0, inplace=True)
df['WeatherDelay'].fillna(0, inplace=True)
df['NASDelay'].fillna(0, inplace=True)
df['SecurityDelay'].fillna(0, inplace=True)
df['LateAircraftDelay'].fillna(0, inplace=True)
```

Step-3

```
cancelled_diverted_df = df[(df['Diverted'] == 1) | (df['ActualElapsedTime'].isna())]

cancelled_diverted_df.drop("ArrTime", 1, inplace=True) #1 for column
cancelled_diverted_df.drop("ActualElapsedTime", 1, inplace=True) #1 for column
cancelled_diverted_df.drop("CRSElapsedTime", 1, inplace=True) #1 for column
cancelled_diverted_df.drop("AirTime", 1, inplace=True) #1 for column
cancelled_diverted_df.drop("ArrDelay", 1, inplace=True) #1 for column
cancelled_diverted_df.drop("TaxiIn", 1, inplace=True) #1 for column
cancelled_diverted_df.drop("TaxiOut", 1, inplace=True) #1 for column
cancelled_diverted_df.drop("CarrierDelay", 1, inplace=True) #1 for column
cancelled_diverted_df.drop("WeatherDelay", 1, inplace=True) #1 for column
cancelled_diverted_df.drop("NASDelay", 1, inplace=True) #1 for column
cancelled_diverted_df.drop("SecurityDelay", 1, inplace=True) #1 for column
cancelled_diverted_df.drop("LateAircraftDelay", 1, inplace=True) #1 for column

# Print the new DataFrame
cancelled_diverted_df
```

Processing data for further analysis

Step-4

```
# Create Date from 3 columns:

df["DepDate"] = pd.to_datetime(df.Year*10000 +
                               df.Month*100 +
                               df.DayofMonth, format="%Y%m%d")
```

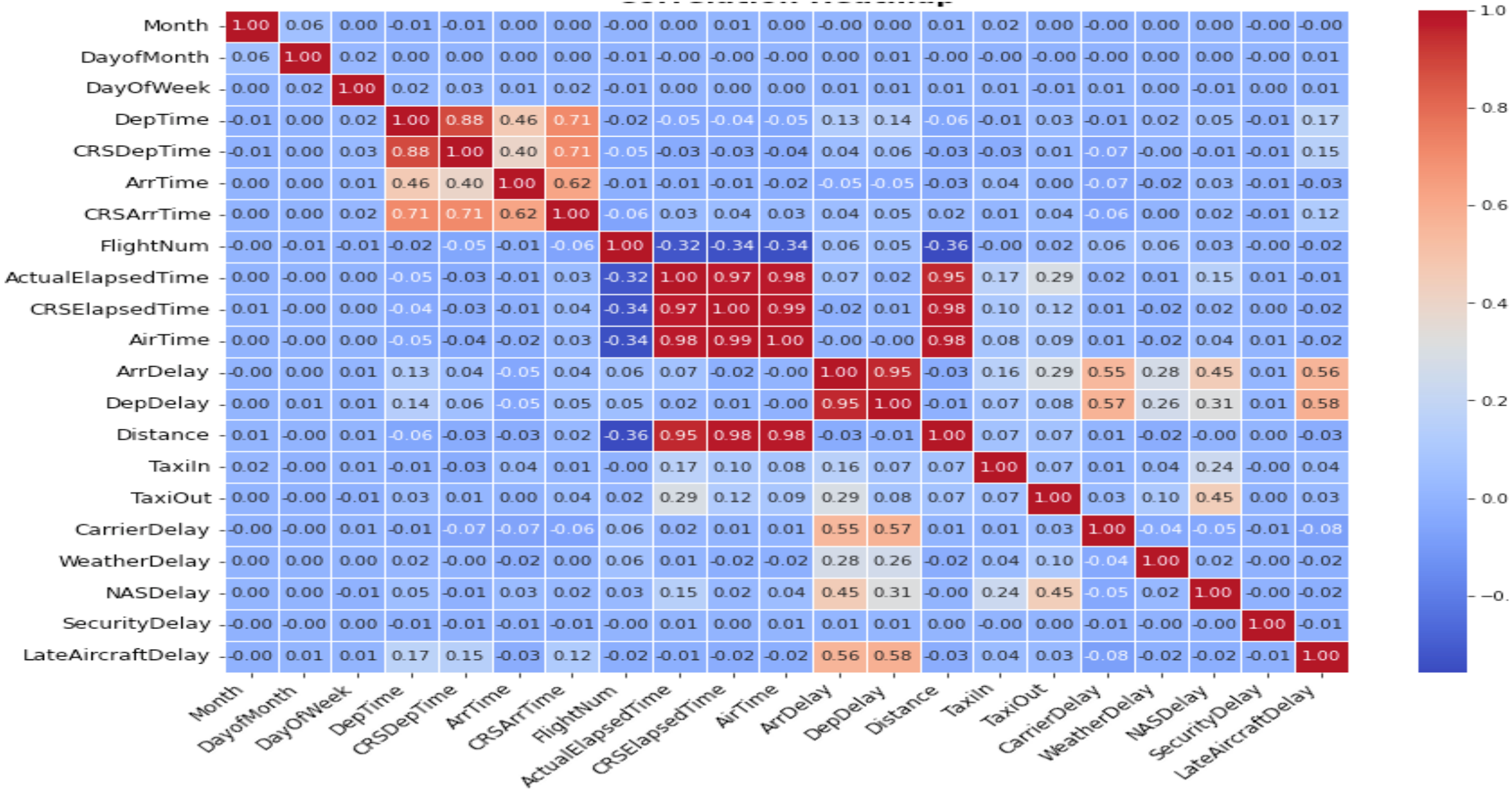
Step -5

```
# Create temporary concate table for analysis:

concatenated_df = pd.concat([df, cancelled_diverted_df])

for dataset in concatenated_df:
    concatenated_df.loc[concatenated_df["ArrDelay"] <= 15, "Status"] = 0
    concatenated_df.loc[concatenated_df["ArrDelay"] >= 15, "Status"] = 1
    concatenated_df.loc[concatenated_df["ArrDelay"] >= 60, "Status"] = 2
    concatenated_df.loc[concatenated_df["Diverted"] == 1, "Status"]  = 3
    concatenated_df.loc[concatenated_df["Cancelled"] == 1, "Status"] = 4
```

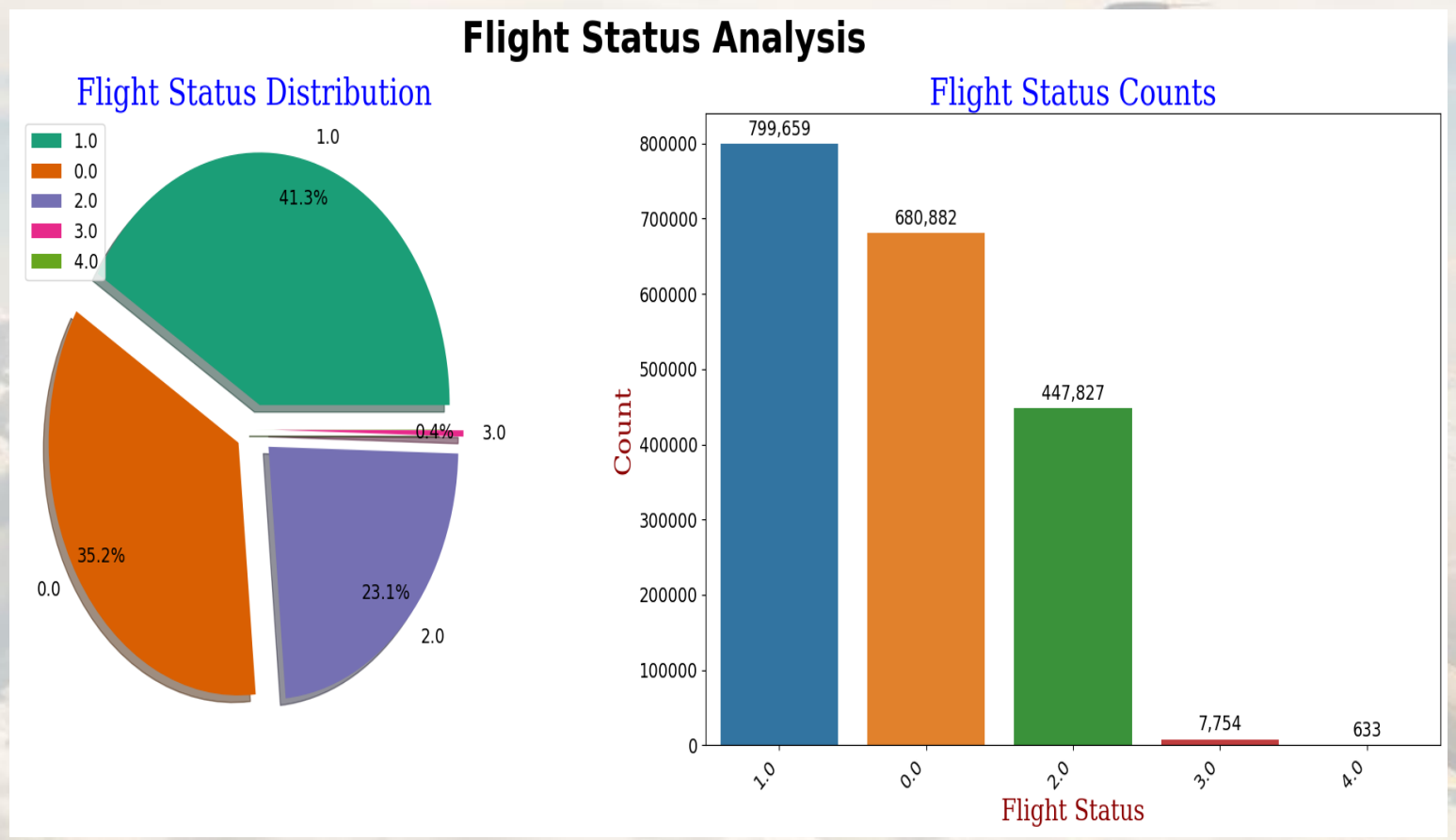
Heat map


Month 1.00 0.06 0.00 -0.01 -0.01 0.00 0.00 -0.00 0.00 0.01 0.00 -0.00 0.00 0.01 0.02 0.00 -0.00 0.00 0.00 -0.00 -0.00
DayofMonth 0.06 1.00 0.02 0.00 0.00 0.00 0.00 -0.01 -0.00 -0.00 -0.00 0.00 0.01 -0.00 -0.00 -0.00 -0.00 0.00 0.00 -0.00 0.01
DayOfWeek 0.00 0.02 1.00 0.02 0.03 0.01 0.02 -0.01 0.00 0.00 0.00 0.01 0.01 0.01 0.01 -0.01 0.01 0.00 -0.01 0.00 0.01
DepTime -0.01 0.00 0.02 1.00 0.88 0.46 0.71 -0.02 -0.05 -0.04 -0.05 0.13 0.14 -0.06 -0.01 0.03 -0.01 0.02 0.05 -0.01 0.17
CRSDepTime -0.01 0.00 0.03 0.88 1.00 0.40 0.71 -0.05 -0.03 -0.03 -0.04 0.04 0.06 -0.03 -0.03 0.01 -0.07 -0.00 -0.01 -0.01 0.15
ArrTime 0.00 0.00 0.01 0.46 0.40 1.00 0.62 -0.01 -0.01 -0.01 -0.02 -0.05 -0.05 -0.03 0.04 0.00 -0.07 -0.02 0.03 -0.01 -0.03
CRSArrTime 0.00 0.00 0.02 0.71 0.71 0.62 1.00 -0.06 0.03 0.04 0.03 0.04 0.05 0.02 0.01 0.04 -0.06 0.00 0.02 -0.01 0.12
FlightNum -0.00 -0.01 -0.01 -0.02 -0.05 -0.01 -0.06 1.00 -0.32 -0.34 -0.34 0.06 0.05 -0.36 -0.00 0.02 0.06 0.06 0.03 -0.00 -0.02
ActualElapsedTime 0.00 -0.00 0.00 -0.05 -0.03 -0.01 0.03 -0.32 1.00 0.97 0.98 0.07 0.02 0.95 0.17 0.29 0.02 0.01 0.15 0.01 -0.01
CRSElapsedTime 0.01 -0.00 0.00 -0.04 -0.03 -0.01 0.04 -0.34 0.97 1.00 0.99 -0.02 0.01 0.98 0.10 0.12 0.01 -0.02 0.02 0.00 -0.02
AirTime 0.00 -0.00 0.00 -0.05 -0.04 -0.02 0.03 -0.34 0.98 0.99 1.00 -0.00 -0.00 0.98 0.08 0.09 0.01 -0.02 0.04 0.01 -0.02
ArrDelay -0.00 0.00 0.01 0.13 0.04 -0.05 0.04 0.06 0.07 -0.02 -0.00 1.00 0.95 -0.03 0.16 0.29 0.55 0.28 0.45 0.01 0.56
DepDelay 0.00 0.01 0.01 0.14 0.06 -0.05 0.05 0.05 0.02 0.01 -0.00 0.95 1.00 -0.01 0.07 0.08 0.57 0.26 0.31 0.01 0.58
Distance 0.01 -0.00 0.01 -0.06 -0.03 -0.03 0.02 -0.36 0.95 0.98 0.98 -0.03 -0.01 1.00 0.07 0.07 0.01 -0.02 -0.00 0.00 -0.03
TaxiIn 0.02 -0.00 0.01 -0.01 -0.03 0.04 0.01 -0.00 0.17 0.10 0.08 0.16 0.07 0.07 1.00 0.07 0.01 0.04 0.24 -0.00 0.04
TaxiOut 0.00 -0.00 -0.01 0.03 0.01 0.00 0.04 0.02 0.29 0.12 0.09 0.29 0.08 0.07 0.07 1.00 0.03 0.10 0.45 0.00 0.03
CarrierDelay -0.00 -0.00 0.01 -0.01 -0.07 -0.07 -0.06 0.06 0.02 0.01 0.01 0.55 0.57 0.01 0.01 0.03 1.00 -0.04 -0.05 -0.01 -0.08
WeatherDelay 0.00 0.00 0.00 0.02 -0.00 -0.02 0.00 0.06 0.01 -0.02 -0.02 0.28 0.26 -0.02 0.04 0.10 -0.04 1.00 0.02 -0.00 -0.02
NASDelay 0.00 0.00 -0.01 0.05 -0.01 0.03 0.02 0.03 0.15 0.02 0.04 0.45 0.31 -0.00 0.24 0.45 -0.05 0.02 1.00 -0.00 -0.02
SecurityDelay -0.00 -0.00 0.00 -0.01 -0.01 -0.01 -0.01 -0.00 0.01 0.00 0.01 0.01 0.01 0.00 -0.00 0.00 -0.01 -0.00 -0.00 1.00 -0.01
LateAircraftDelay -0.00 0.01 0.01 0.17 0.15 -0.03 0.12 -0.02 -0.01 -0.02 -0.02 0.56 0.58 -0.03 0.04 0.03 -0.08 -0.02 -0.02 -0.01 1.00
Month
DayofMonth
DayOfWeek
DepTime
CRSDepTime
ArrTime
CRSArrTime
FlightNum
ActualElapsedTime
CRSElapsedTime
AirTime
ArrDelay
DepDelay
Distance
TaxiIn
TaxiOut
CarrierDelay
WeatherDelay
NASDelay
SecurityDelay
LateAircraftDelay
1.0
0.8
0.6
0.4
0.2
0.0
−0.

1. Flight Status Distribution and their Counts

Interpretation:

- Most flights are on-time (35.2%), followed by delays (41.3%) and cancellations (23.1%).
- On-time flights are increasing, while delayed and cancelled flights are steady.
- Data for one month, one airport.

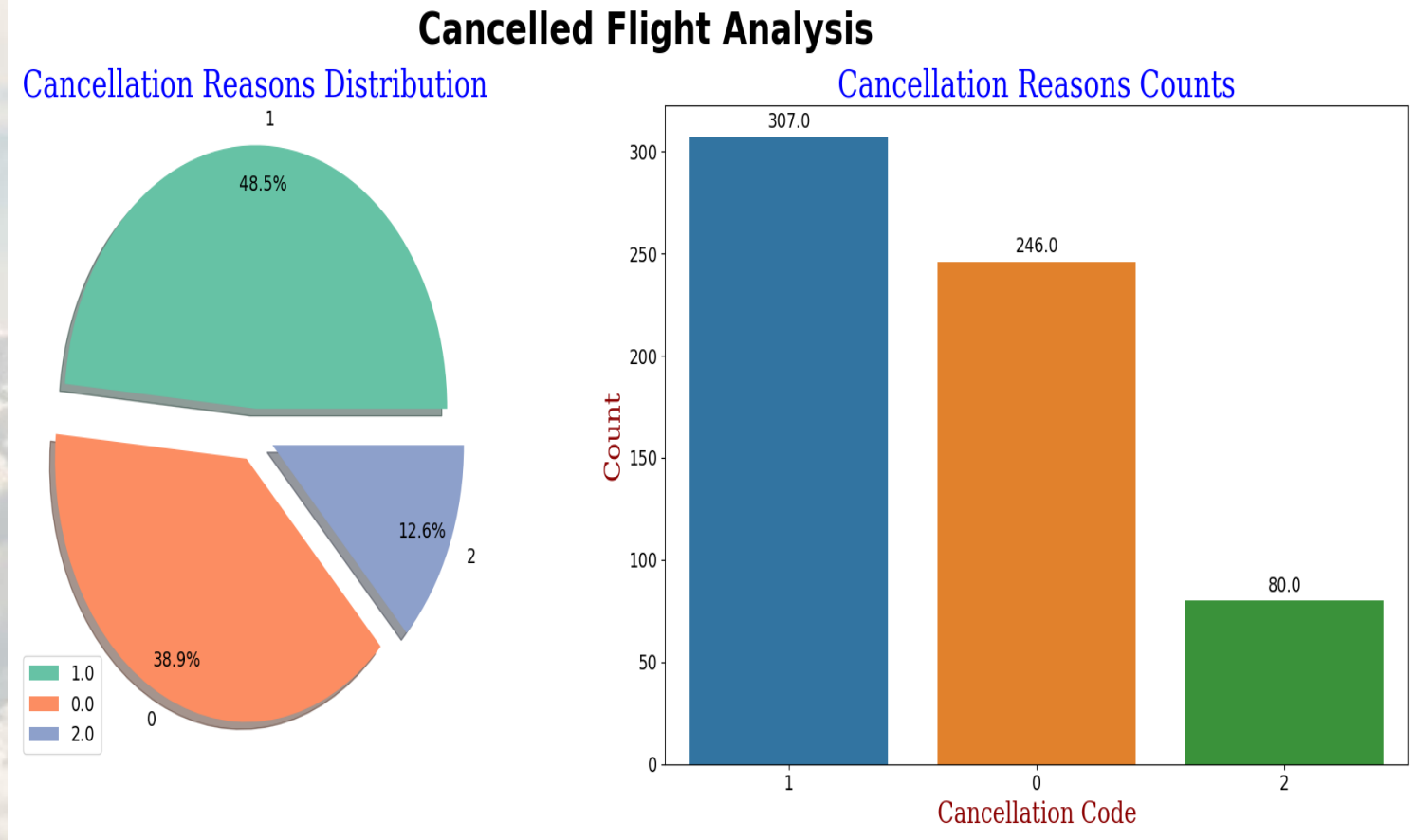


Flight was on time (0), Slightly delayed (1), Highly delayed (2), Diverted (3), or Cancelled (4)

2. Reasons For cancelled Flights

Interpretation:

- Weather is the top reason for cancellations (48.5%).
- Cancellations vary daily.
- Data from one airport, one month.

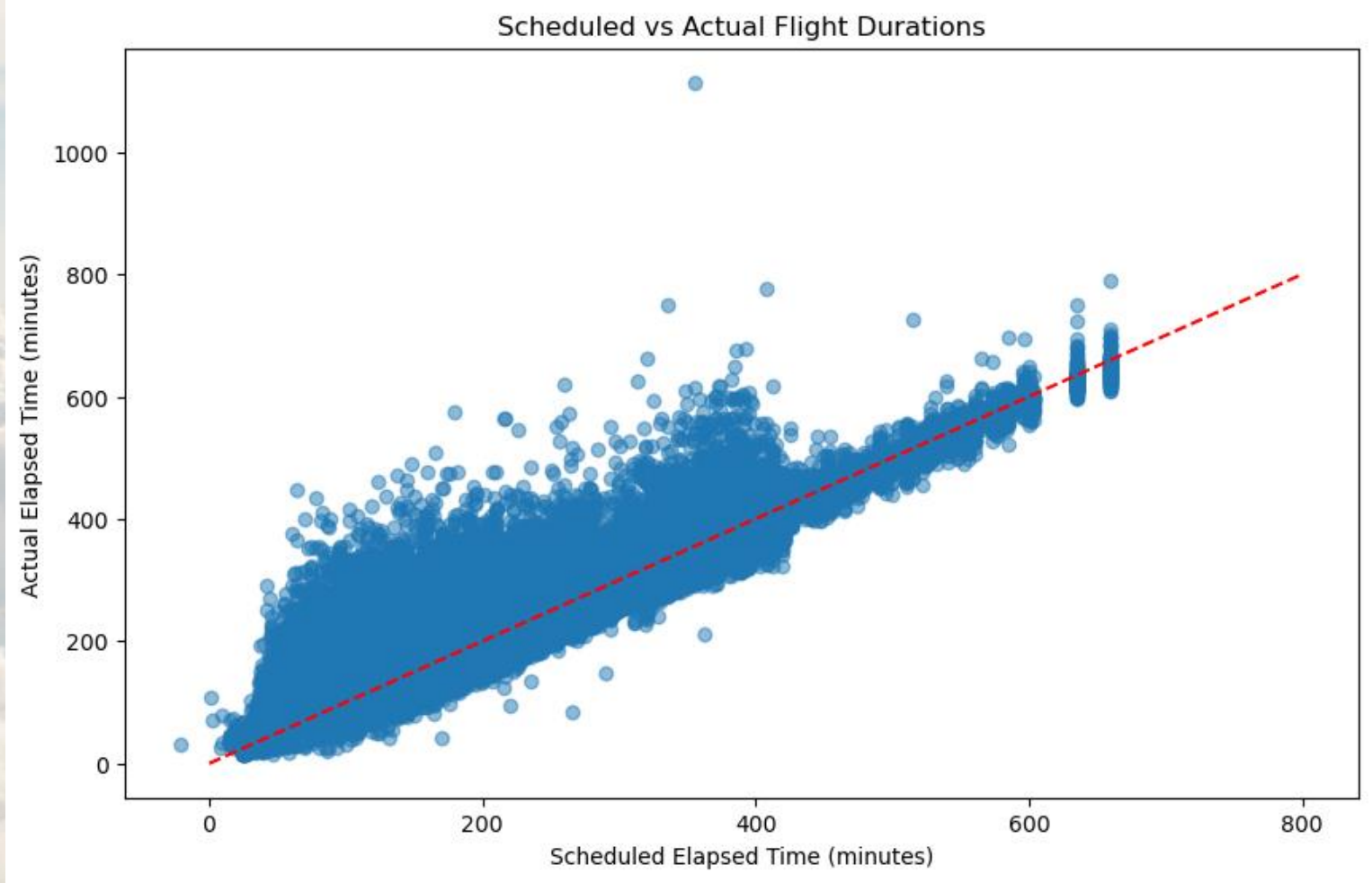


0 = carrier,
1 = weather,
2 = NAS

3. Flight Actual Durations vs Scheduled Duration (Excluding Cancelled or Diverted flight)

Interpretation:

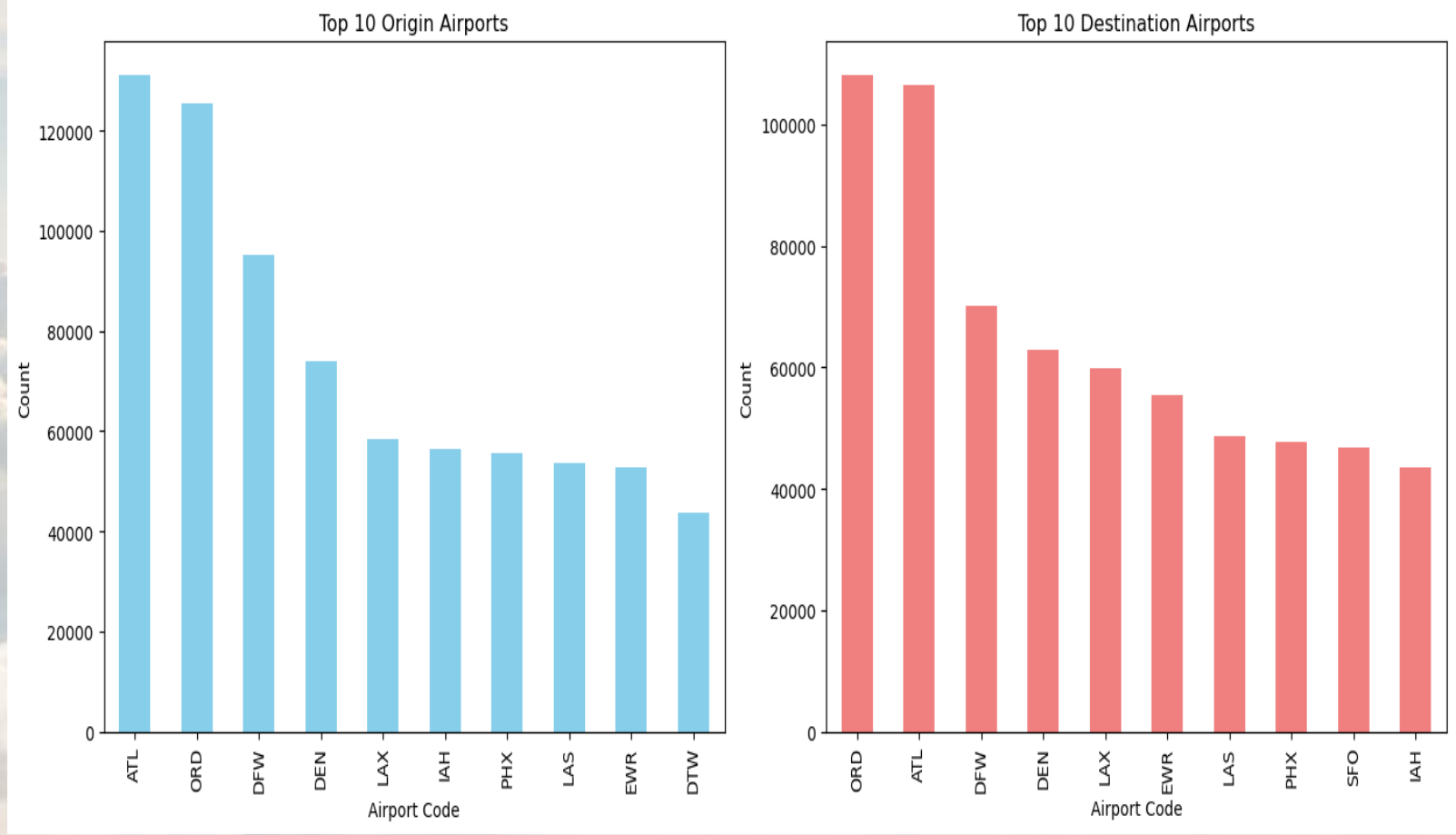
- Most flights (upward trend) take longer than planned.
- Big difference in actual flight times (spread-out data).
- Reasons: weather, air traffic, technical issues, etc.



4.Explore Origin and Destination Analysis

Interpretation:

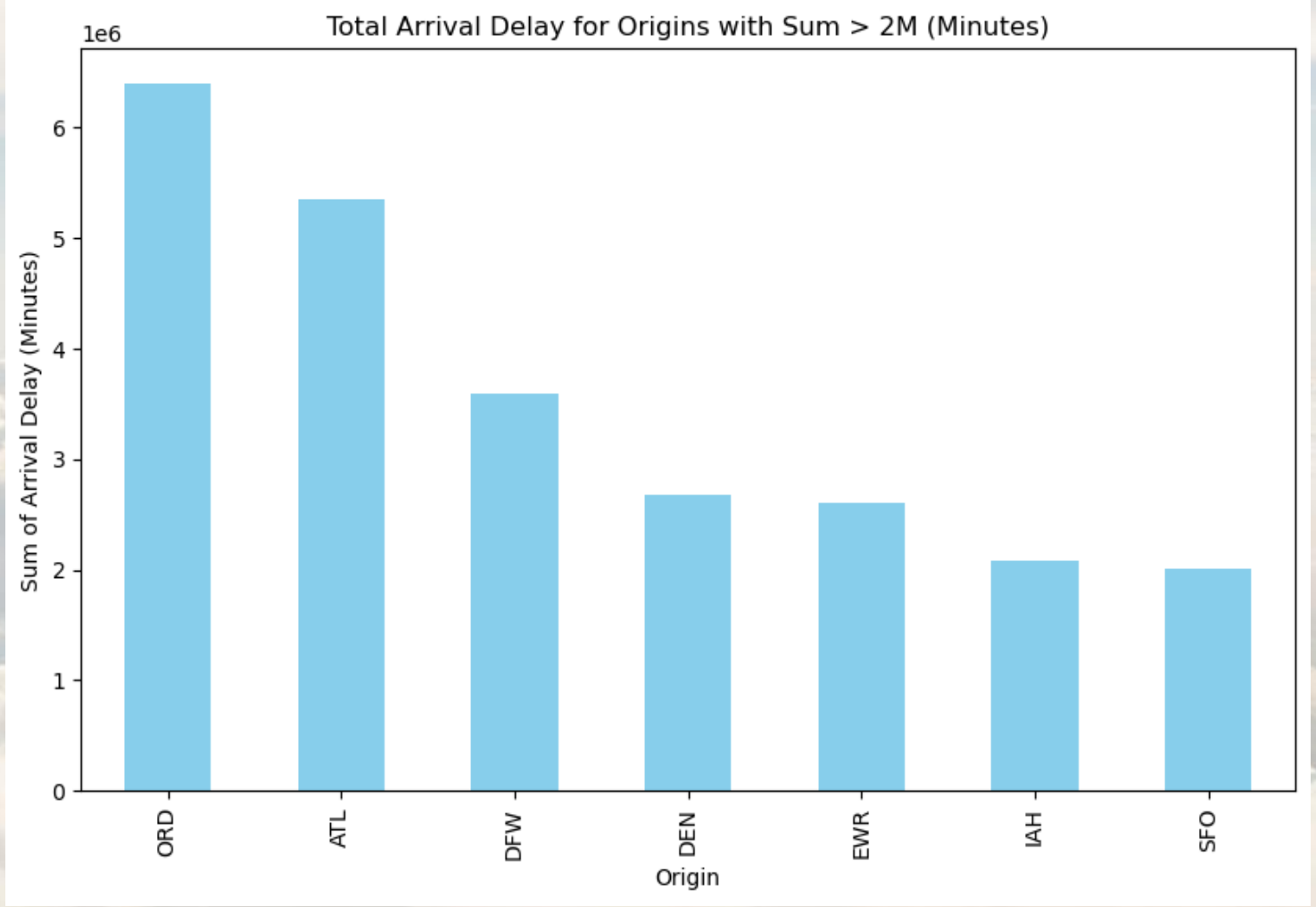
- Busiest origins: ATL, DFW, DEN
- Busiest destinations: MCO, ATL, LAX
- Numbers show annual passenger traffic at each airport.



5.Arrival Delay for Origin Airports

Interpretation:

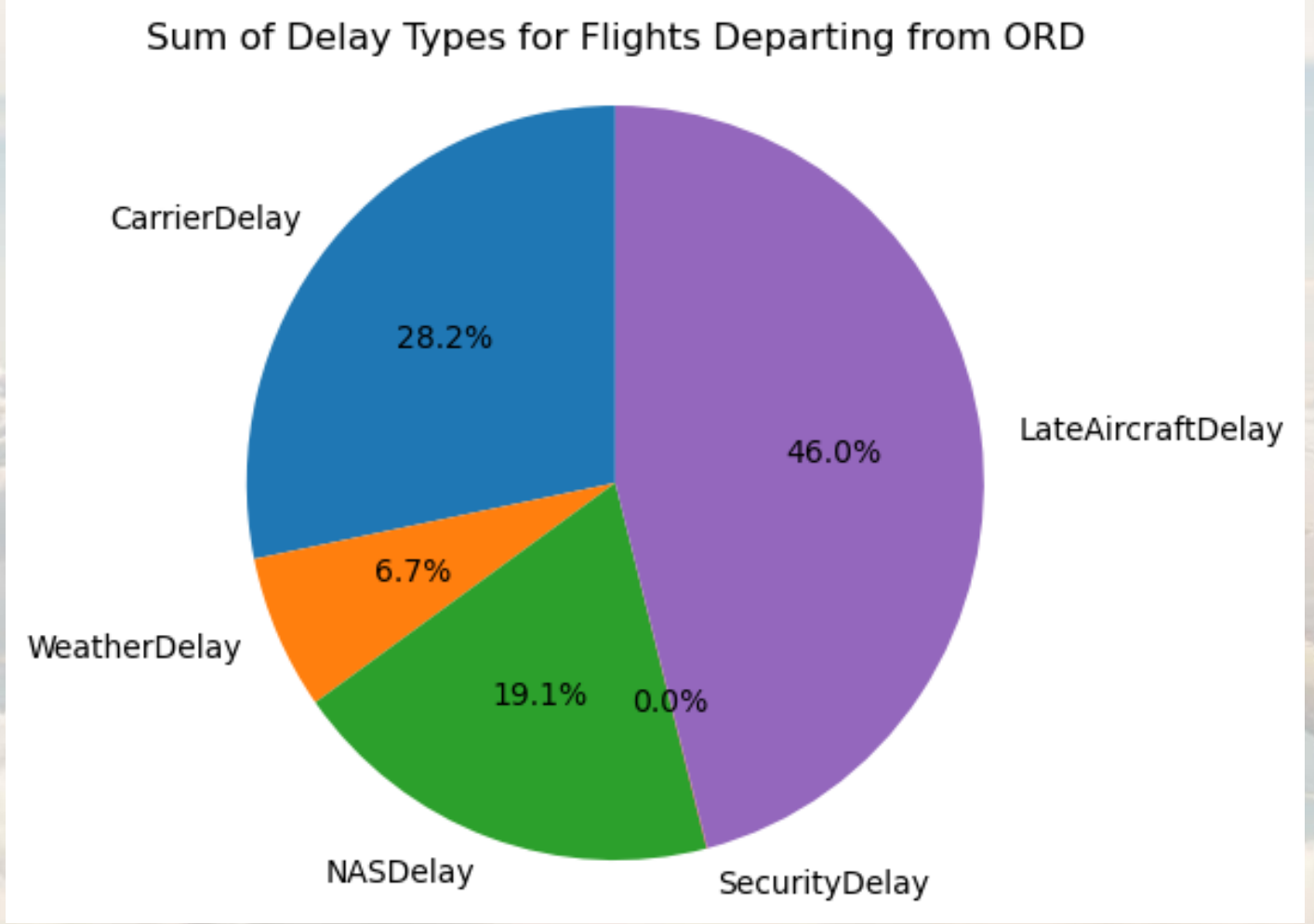
- Find out Max Arrival delay at which origin.
- ORD having high delays.
- It might be due to different delays.



6.Different Delays for ORD (Chicago)

Interpretation:

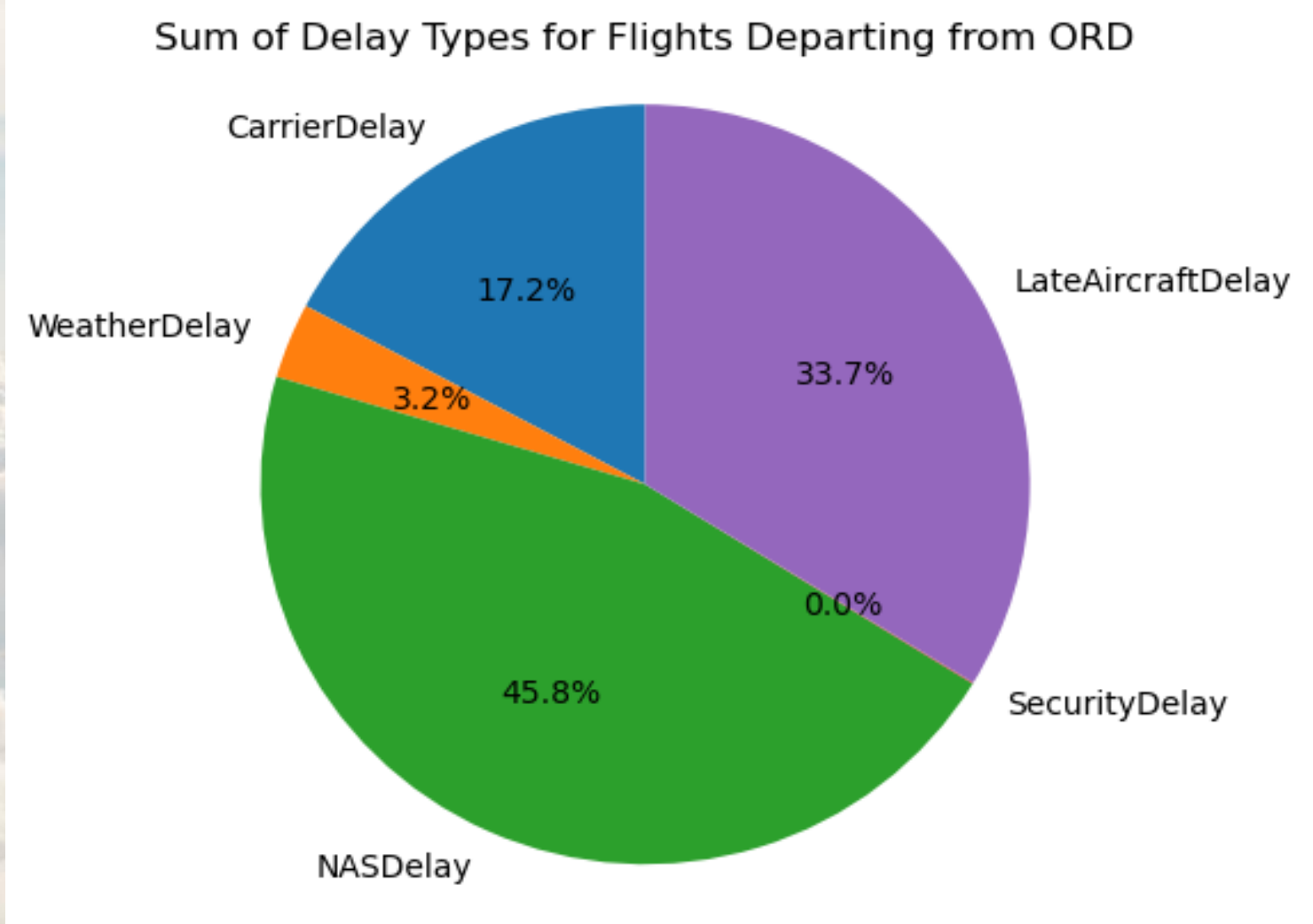
- Find out different delays at ORD origin.
- Max delay is due to Late aircraft.
- At origin we can take care of carrier delay and NAS delay.



7.From where does the Late Aircraft Delay occur at ORD (chicago)

Interpretation:

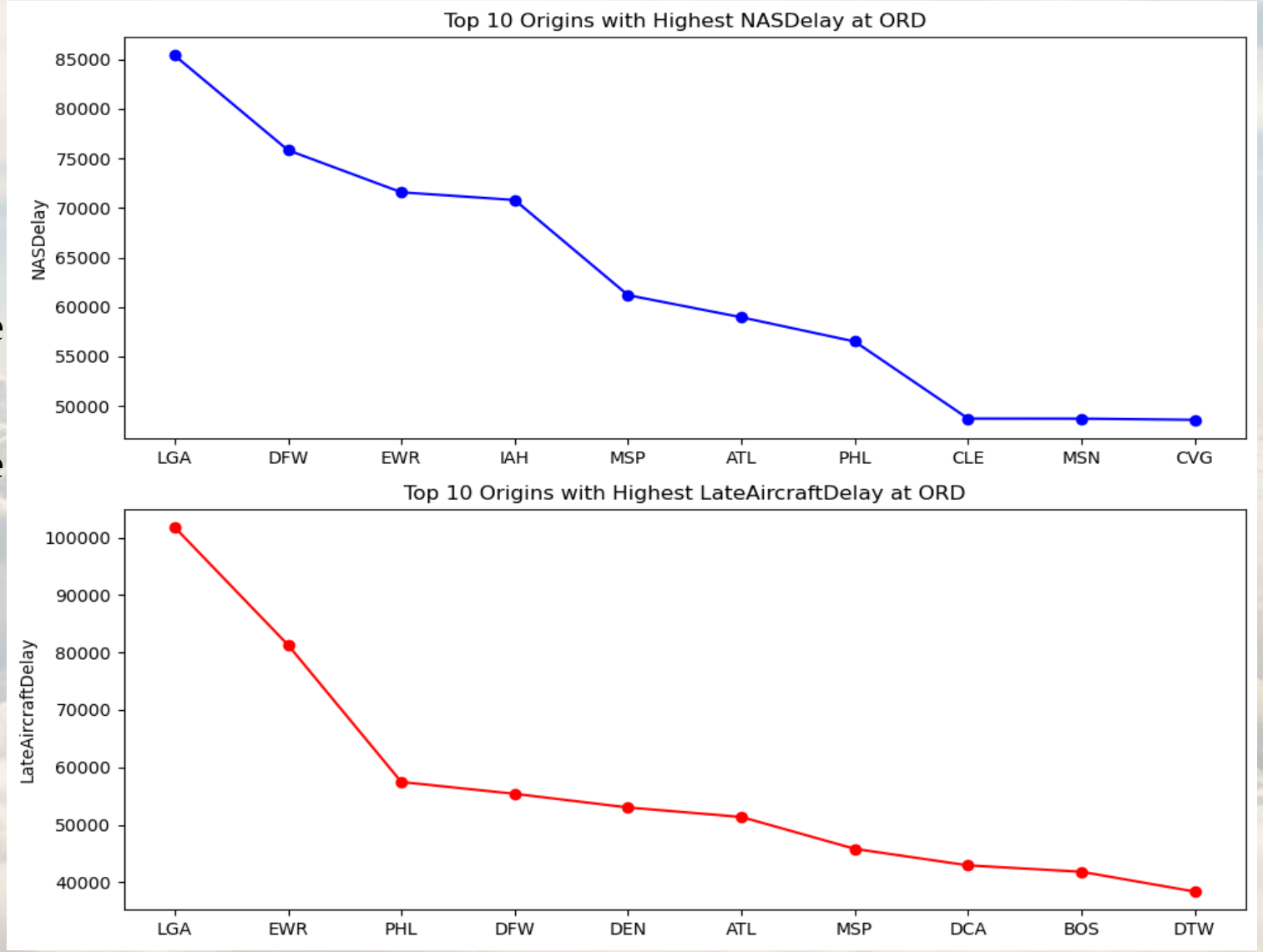
- We want to find out where does aircraft delays at previous airport.
- Find out that max delay was due to NAS Delay (45.8%)
- Every airport which destined for ORD should take care of NAS Delay.
- Other delays are Late Aircraft and Carrier delay.



8.Which airports causing Aircraft delays at the ORD

Interpretation :

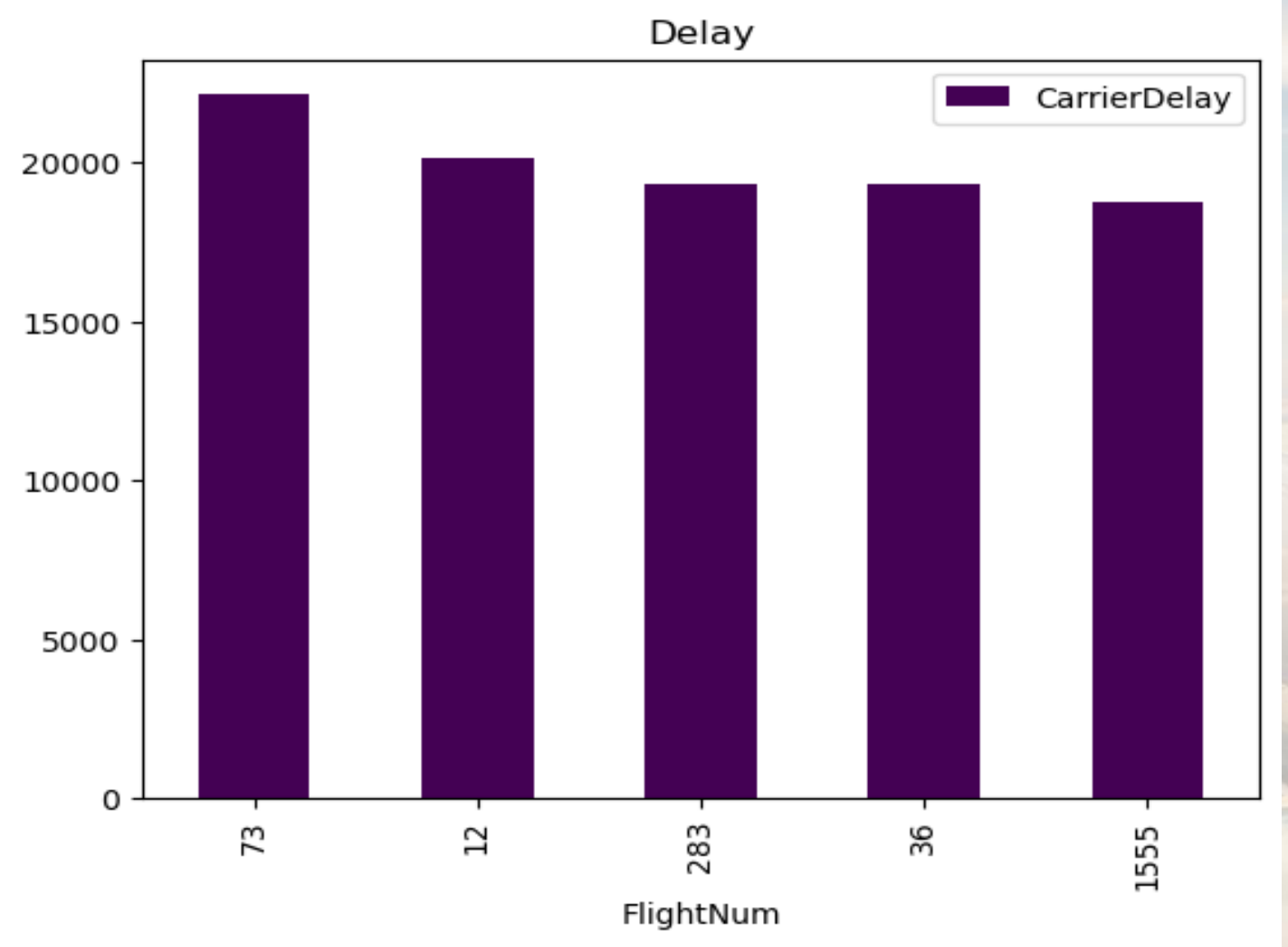
- Which different airport causing Late aircraft delay.
- Find out by NAS Delay and Late aircraft delay.
- Found out these Top 10 delays.



9.Which Aircraft got highest Carrier Delay

Interpretation :

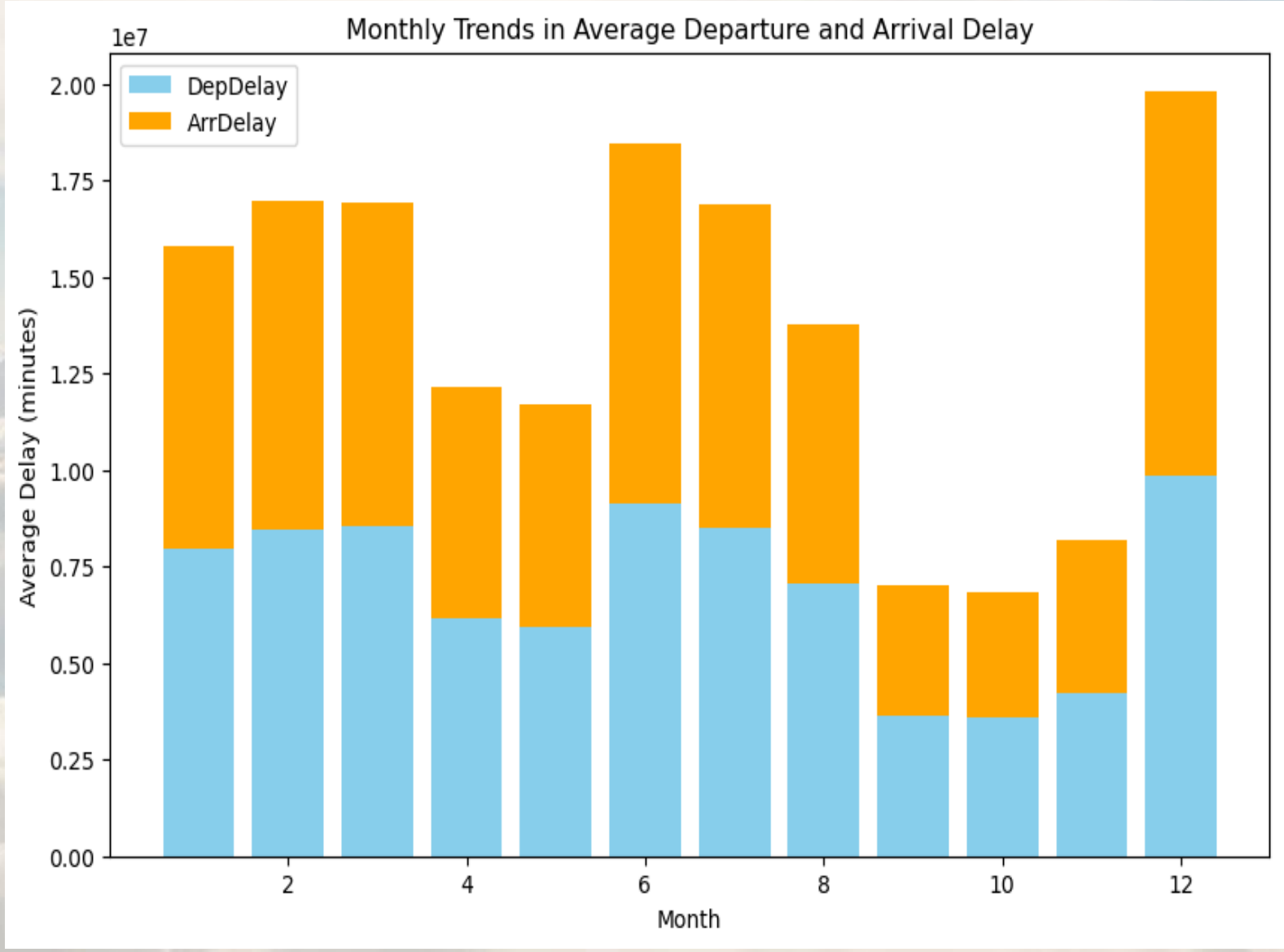
- Which are the aircraft having high carrier delay.
- It may be due to poor maintenance of aircrafts.
- Chart shows top 5 aircraft delays.



10. Delay Analysis (Monthly)

Interpretation :

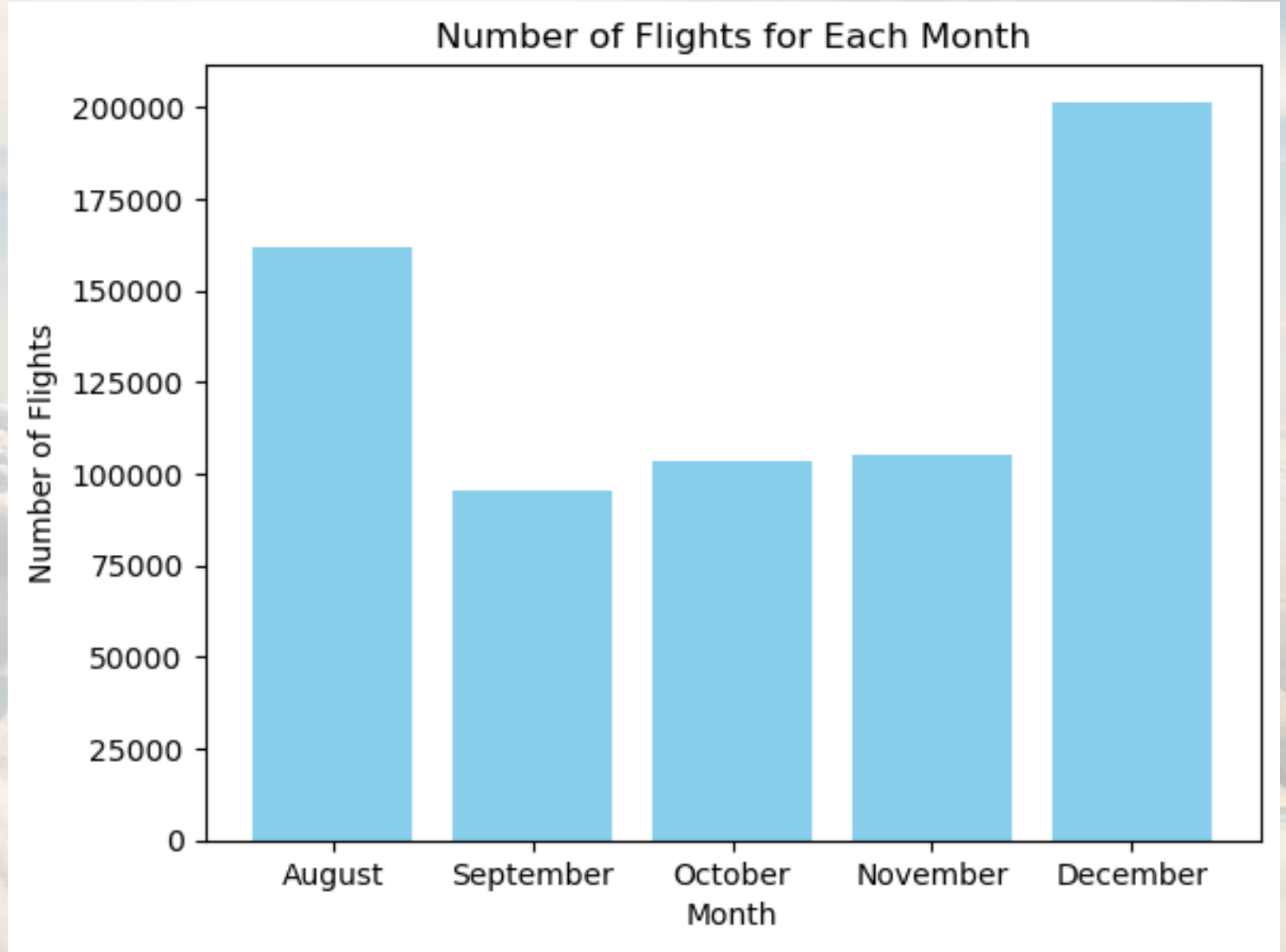
- It is seen that decrease in delay for the month September, October and November.
- It might be due to less people travel during that period.
- It might also be due to high prices of flight during that period.



11.What is the cause that Month 9, 10, 11 got less Delays?

Interpretation :

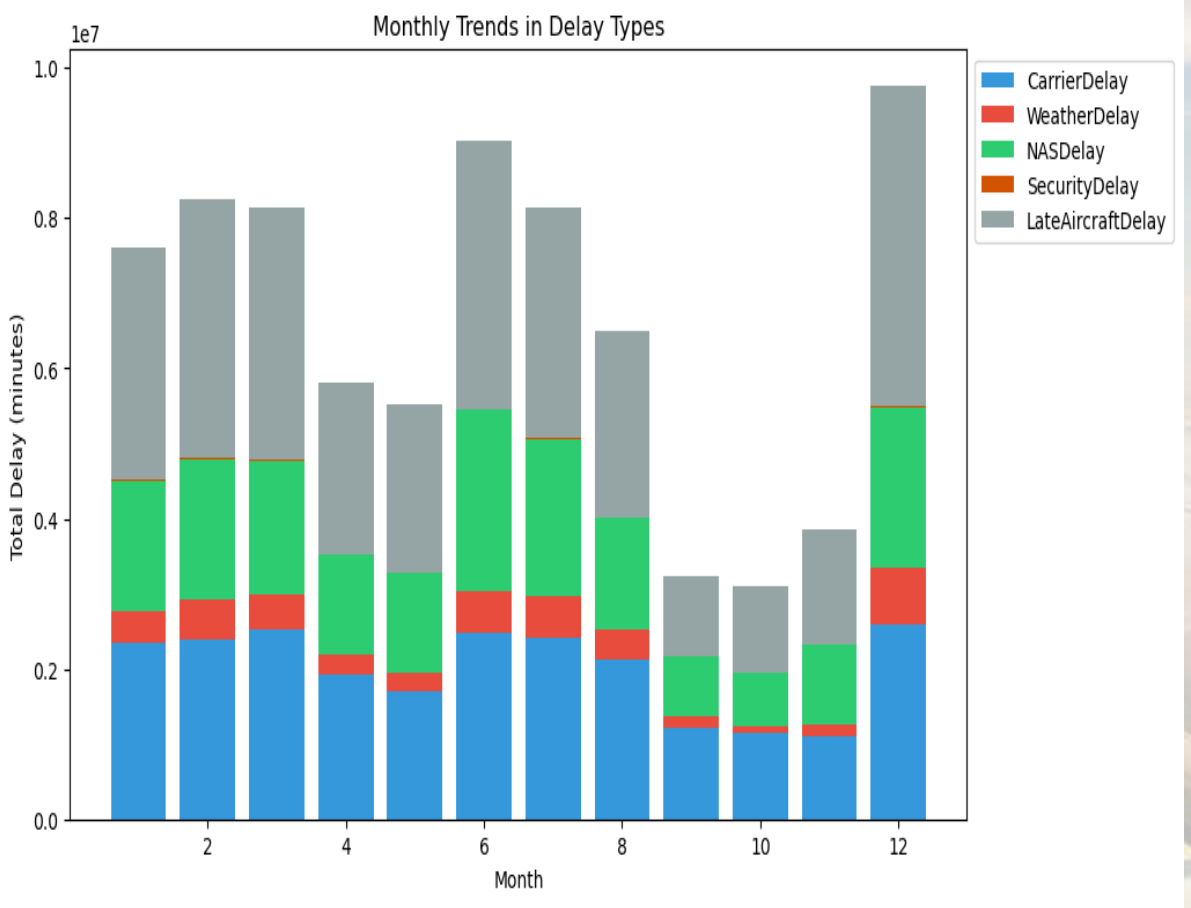
- The graph depicts a consistent rise in the number of flights from August to December, with a sharper increase noticeable in December.
- This surge could be due to heightened holiday travel demand, airlines adjusting schedules to meet seasonal preferences, and an upswing in end-of-year business travel, all contributing to the significant rise in air travel during these months.



12.What are the different delays (Monthly)

Interpretation :

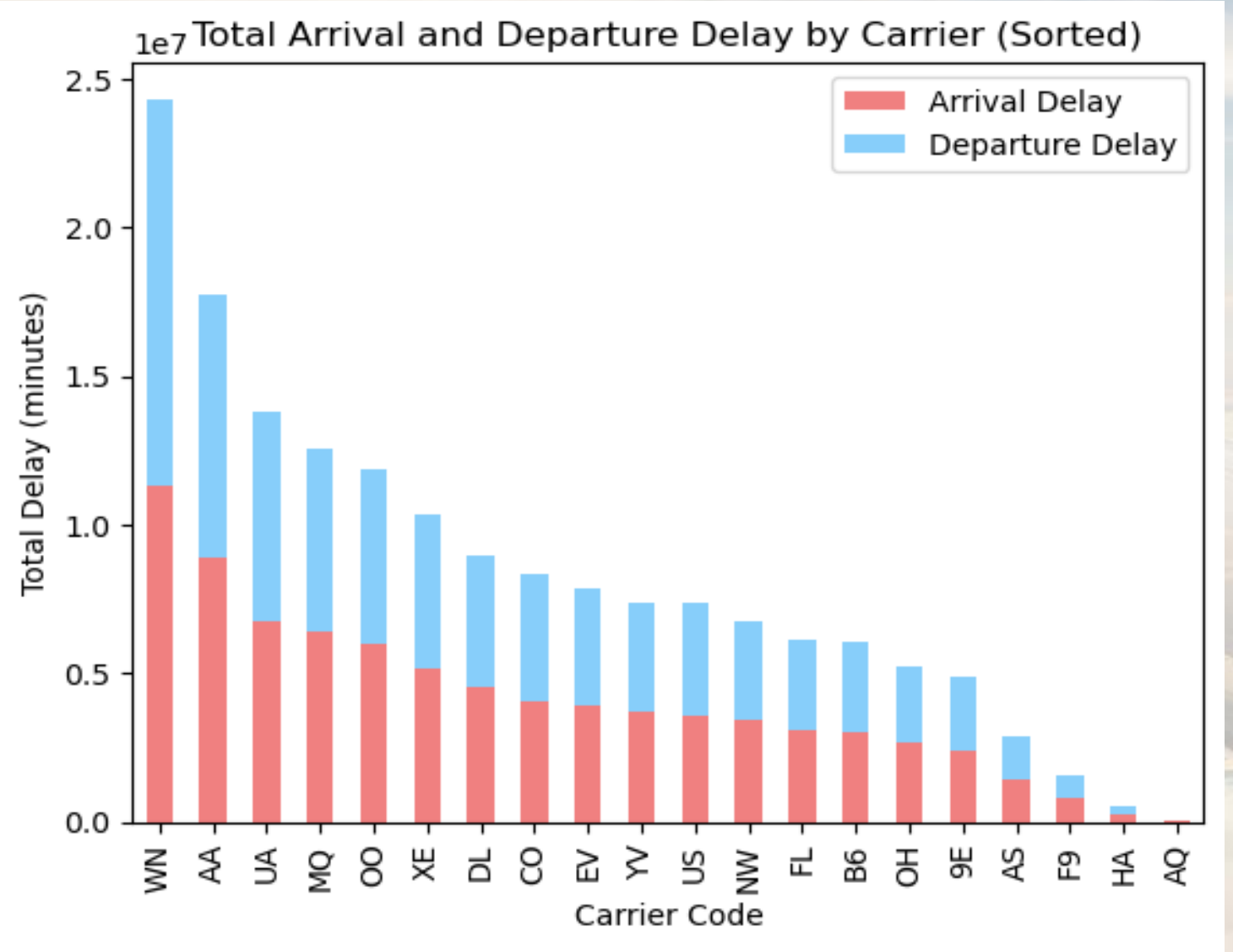
- The data reveals that CarrierDelay consistently dominates as the most significant delay type, indicating that airline-related issues play a major role in overall delays.
- WeatherDelay and NASDelay show a gradual increase from February to December, potentially influenced by changing weather patterns and airspace system issues.
- Security Delay remains relatively stable without notable fluctuations.
- LateAircraftDelay is the least common delay type, consistently showing the lowest values throughout the year, implying it has less impact on overall delays compared to other factors.



13.Explore Arrival Delays with respect to Carrier

Interpretation :

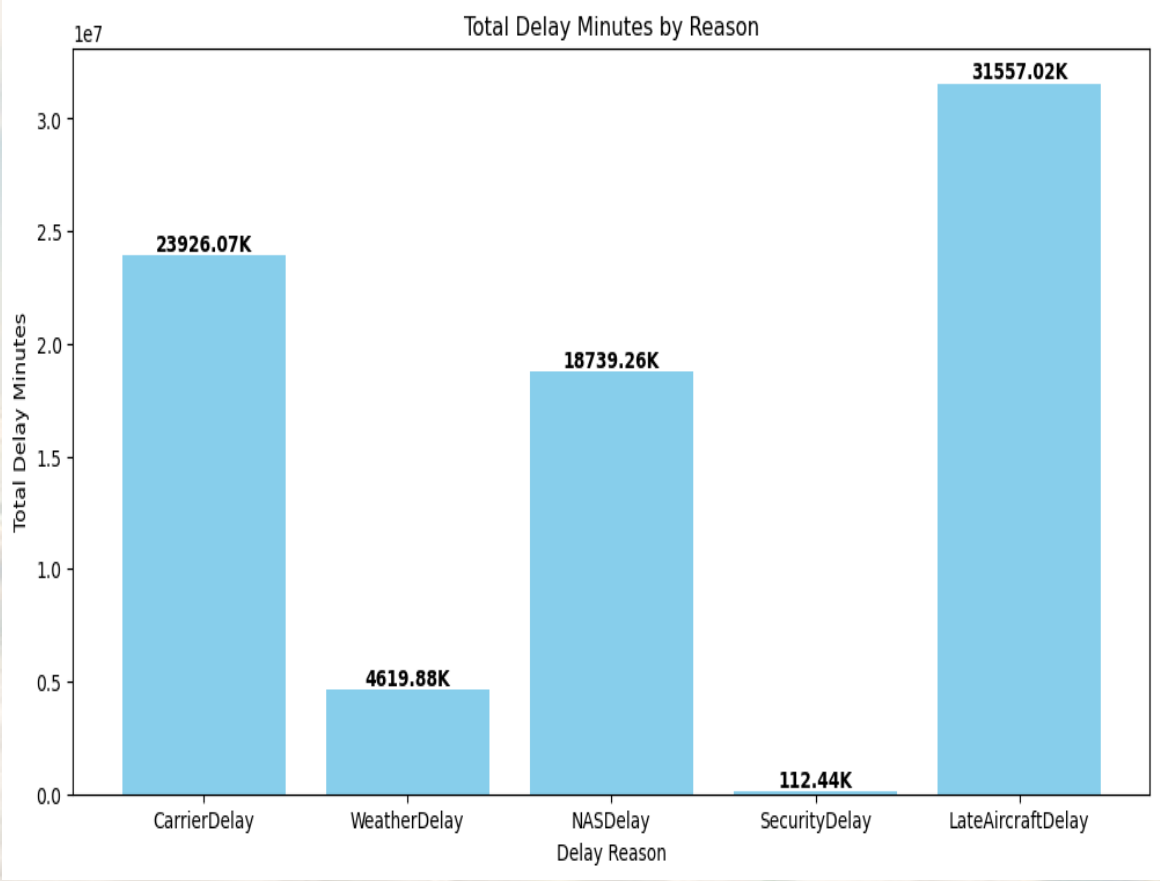
- From this we can say that there is lest amount of delay in Aloha Airlines and highest in Southwest Airlines.



14.Explore the Reasons for Delays

Interpretation :

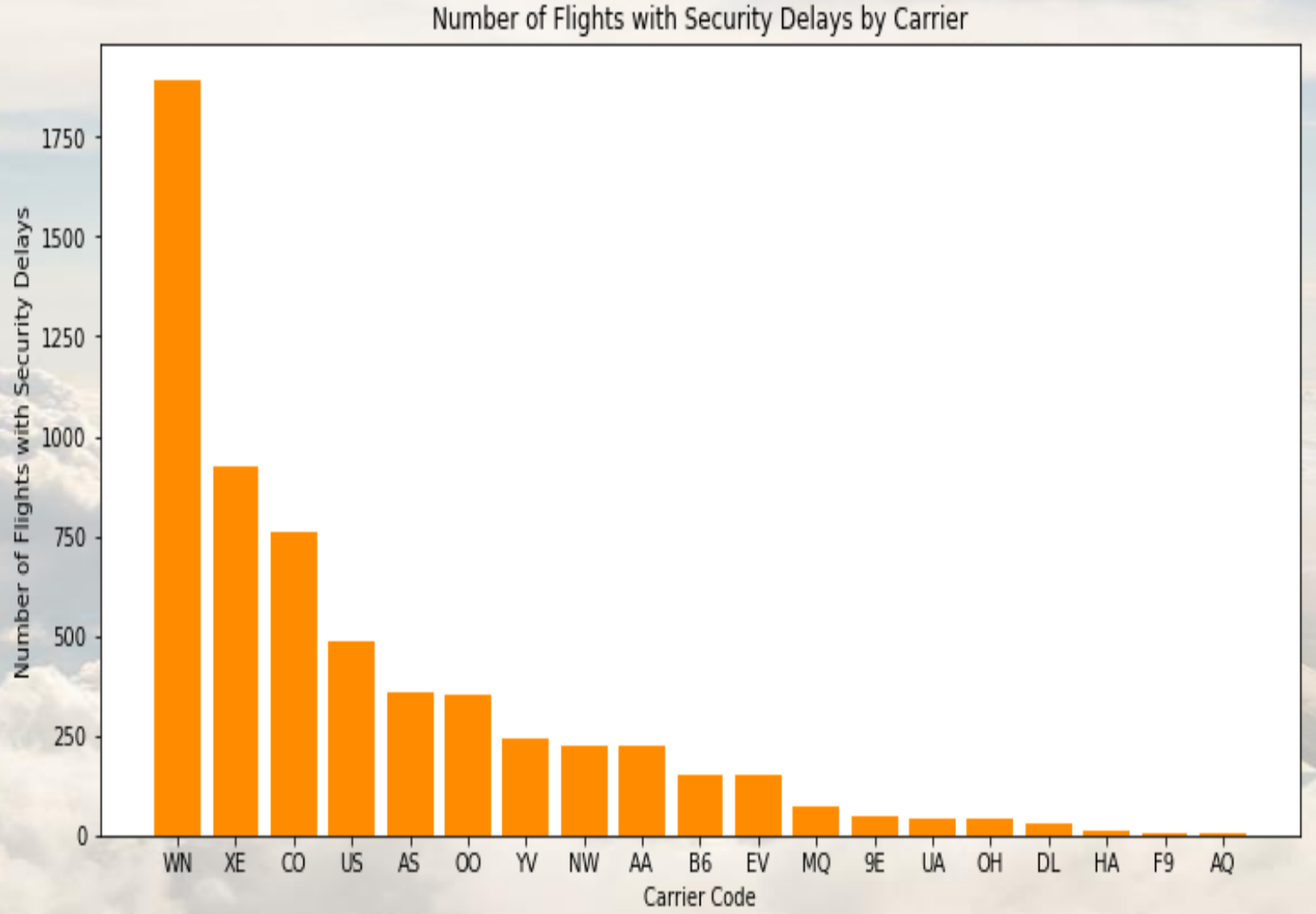
- The data underscores CarrierDelay as the primary cause of delays, surpassing 31 million minutes, prompting a need for airlines to address internal issues for better performance.
- Weather-related delays, totaling over 23 million minutes, emphasize the impact of weather disruptions, necessitating robust contingency plans.
- Additionally, delays attributed to the National Airspace System (NASDelay), exceeding 18 million minutes, highlight the significance of efficient air traffic management and infrastructure.
- SecurityDelay and LateAircraftDelay are less frequent contributors, with 4.6 million and 0.1 million minutes, respectively, signaling their comparatively lower impact on overall delays



15. Investigate Security Delays

Interpretation :

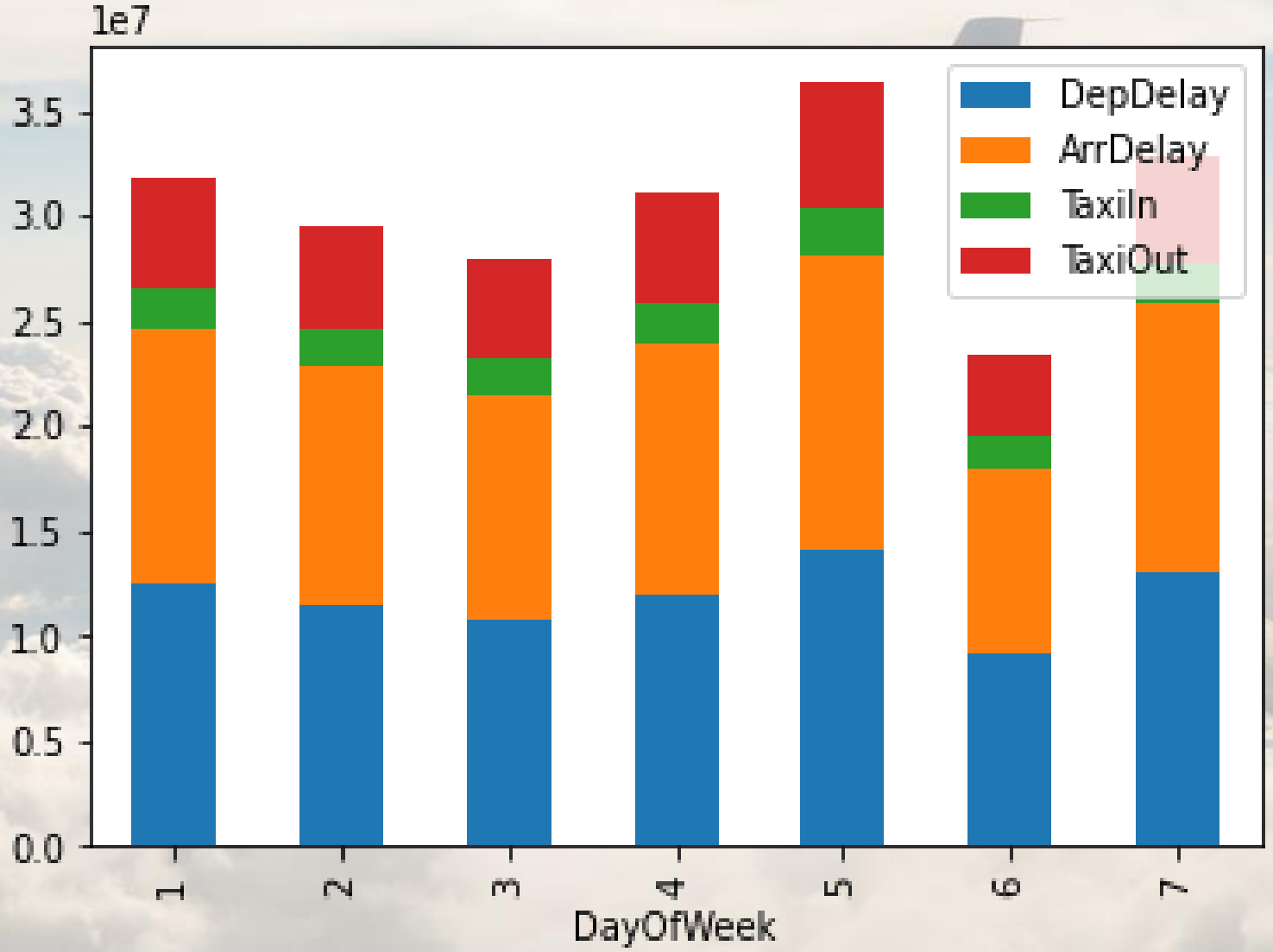
- we can see the highest security delay is happening in the carrier(WN, XE, XE, US, AS).
- There might be a reason that there are fewer baggage counters and technical malfunctions



16. Analysing weekdays delays

Interpretation :

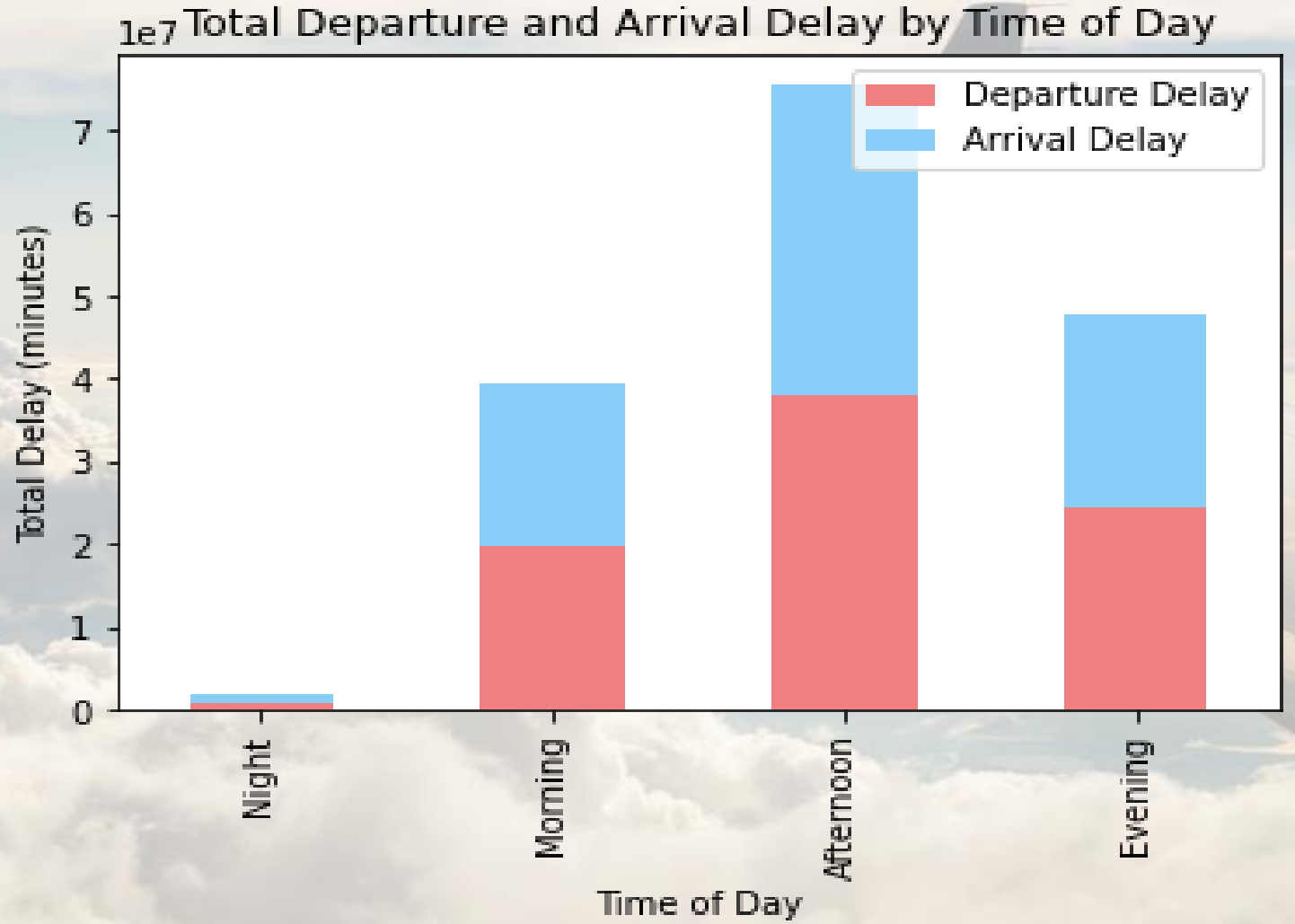
- We observe that Fridays experience the highest delays.
- This could be due to more people travelling home on Fridays and returning on Sundays and also The we can see the trend indicates a decrease in delays on Saturdays.



17. Daily shift wise Analysis

Interpretation :

- We observe that most flights experience delays in the evening, while delays are less frequent at night.
- Adjusting flight timings and optimizing airport infrastructure can help manage crowds. Consider rescheduling flights to nighttime if it is feasible



18.Scatter Animation

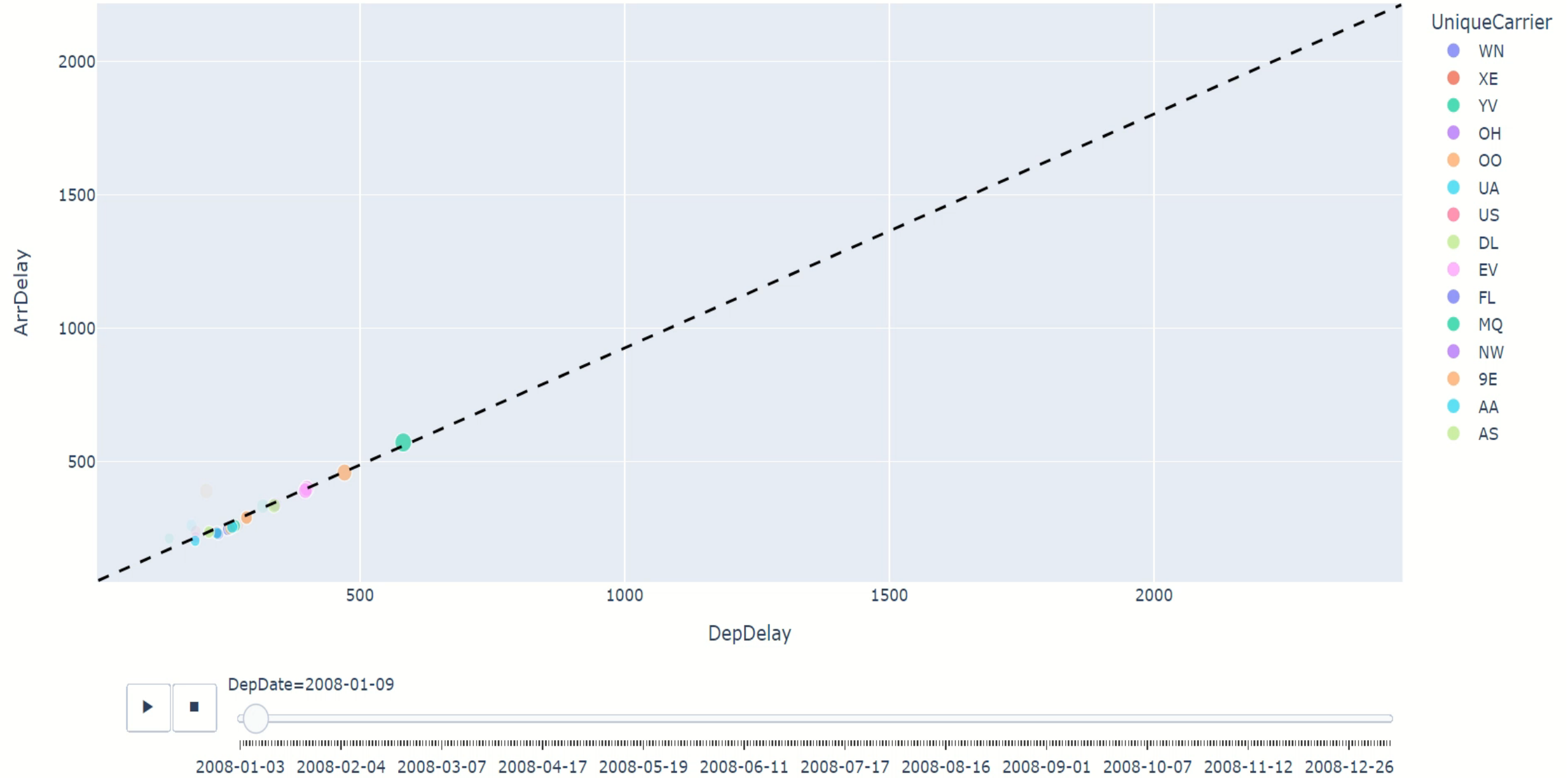
Interpretation:

- we observe the daily delays of unique carriers, specifically OH, NH, MQ, and AA.
- These carriers are suggested for optimization in terms of their timings and maintenance

'WN': 'Southwest Airlines',
'AA': 'American Airlines',
'UA': 'United Airlines',
'DL': 'Delta Airlines',
'CO': 'Continental Airlines',
'EV': 'Atlantic Southeast Airlines',
'FL': 'AirTran Airways',
'B6': 'JetBlue Airways',
'9E': 'Pinnacle Airlines',
'F9': 'Frontier Airlines',

'AQ': 'Aloha Airlines'
'MQ': 'American Eagle Airlines',
'OO': 'Skywest Airlines',
'XE': 'ExpressJet',
'US': 'US Airways',
'NW': 'Northwest Airlines',
'YV': 'Mesa Airlines',
'OH': 'Comair',
'AS': 'Alaska Airlines',
'HA': 'Hawaiian Airlines'

Delay Scatter Animated Chart



Conclusion

- **Understanding Operational load. (by busiest carrier)**
- **Allocating resources (for the busiest airport)**
- **By knowing the Highest airtime of the carrier, we use this to optimize schedules**
- **By busiest month can help in managing the operation of airports**
- **Day-wise help in resource management of the specific day**
- **By understanding weekday delay, we can say weekdays are more delayed when compared to weekends(staffing decisions more on weekdays less on weekends)**
- **By shift delays, we can focus more on improving times(shifting)**
- **By understanding aircraft with the highest delay we can address its operation and maintained**

Data Source

- https://www.kaggle.com/datasets/giovamata/airlinedelaycauses/


Thank You