



# US Civil Aviation Analysis in 2023

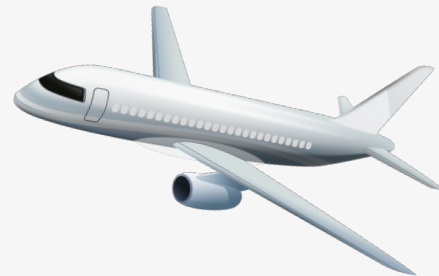
Thủy Triều, Đại Việt - BI57 - 3/2 - Nhóm 2

# TỔNG QUAN ĐỀ TÀI

<b>Chủ đề</b>	US Airline 2023 - US Flight Delay Time
<b>Nội dung phân tích</b>	Dựa vào các chỉ số như chuyến bay, khí tượng, hãng bay, etc... Tìm hiểu nguyên nhân về việc Delay và chuyển hướng chuyến bay
<b>Mục đích phân tích</b>	Phân tích để tìm ra các đặc điểm ảnh hưởng có thể tác động đến việc delay các chuyến bay
<b>Công cụ sử dụng</b>	Python & Power BI

Dataset : <https://www.kaggle.com/datasets/bordanova/2023-us-civil-flights-delay-meteo-and-aircraft/data>

# NỘI DUNG



## 01 TỔNG QUAN VỀ BỘ DỮ LIỆU

Giới thiệu về  
dữ liệu

## 02 TIỀN XỬ LÝ DỮ LIỆU

EDA dữ liệu, xử lý và  
modeling data

## 03 PHÂN TÍCH TỔNG QUAN & KHÁM PHÁ

Tổng quan và phân tích  
thăm dò

## 04 PHÂN TÍCH CÁC YẾU TỐ ẢNH HƯỞNG

Đi chi tiết vào các  
yếu tố liên quan

## 05 KẾT LUẬN & ĐỀ XUẤT

Đưa ra kết luận và các  
đề xuất

# 01

## TỔNG QUAN VỀ BỘ DỮ LIỆU



## US 2023 Civil Flights

US\_Flights\_2023

- Flight\_info
- Categories & delay time
- Result

6.7M  
rows

Cancelled\_Diverted\_2023

- Flight\_info
- Categories & delay time
- Result

104k  
rows

airports\_geolocation

- IATA\_CODE
- Airport\_name
- Country
- State
- City
- Latitude
- Longitude

364  
rows

weather\_by\_airport

- Time
- Airport\_id
- tagv (Avg Temperature (°C))
- tmin (Min Temperature (°C))
- tmax (Max Temperature (°C))
- snow (Snow Depth)
- wdir (Wind Direction)
- wspd (Wind Speed - km/h)
- prcp (Precipitation - mm)
- pres (Sea-Level Air Pressure - hPa)

133K  
rows



```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 6743404 entries, 0 to 6743403  
Data columns (total 23 columns):
```

```
#   Column                Dtype
```

0	FlightDate	object
1	Airline	object
2	Tail_Number	object
3	Dep_Airport	object
4	Dep_CityName	object
5	Arr_Airport	object
6	Arr_CityName	object
7	Flight_Duration	int64
8	Manufacturer	object
9	Model	object
10	Aircraft_age	int64
11	Day_Of_Week	int64
12	DepTime_label	object
13	Distance_type	object
14	Dep_Delay	int64
15	Arr_Delay	int64
16	Delay_Carrier	int64
17	Delay_Weather	int64
18	Delay_NAS	int64
19	Delay_Security	int64
20	Delay_LastAircraft	int64
21	Dep_Delay_Type	object
22	Arr_Delay_Type	object

```
dtypes: int64(10), object(13)
```

Flight\_info

Category &  
Delay\_time

Results



US\_FIGHT\_2023 TABLES

RangeIndex: 104488 entries, 0 to 104487

Data columns (total 22 columns):

#	Column	Non-Null Count		Dtype
0	FlightDate	104488	non-null	object
1	Airline	104488	non-null	object
2	Tail_Number	104488	non-null	object
3	Dep_Airport	104488	non-null	object
4	Dep_CityName	104488	non-null	object
5	Arr_Airport	104488	non-null	object
6	Arr_CityName	104488	non-null	object
7	Flight_Duration	104488	non-null	float64
8	Day_Of_Week	104488	non-null	int64
9	DepTime_label	104488	non-null	object
10	Distance_type	104488	non-null	object
11	Dep_Delay	104488	non-null	float64
12	Arr_Delay	104488	non-null	float64
13	Delay_Carrier	104488	non-null	float64
14	Delay_Weather	104488	non-null	float64
15	Delay_NAS	104488	non-null	float64
16	Delay_Security	104488	non-null	float64
17	Delay_LastAircraft	104488	non-null	float64
18	Dep_Delay_Type	104488	non-null	object
19	Arr_Delay_Type	104488	non-null	object
20	Cancelled	104488	non-null	float64
21	Diverted	104488	non-null	float64

dtypes: float64(10), int64(1), object(11)



Cancelled & diverted  
Flight\_info

Cancelled\_Diverted 2023 Tables

Category &  
Delay\_time

Results



# 02

## TIỀN XỬ LÝ DỮ LIỆU



—Vấn đề

“Các bảng dữ liệu chưa thể  
liên kết được với nhau vì  
không có key”



# Tạo Primary Key & clean data cho bảng US\_Flight\_2023

```
[ ] #Tạo PK cho bảng Flights
    flights_df['ids'] = flights_df['FlightDate'] + flights_df['Dep_Airport']
                        + flights_df['Arr_Airport'] + flights_df['Tail_Number'] + flights_df['DepTime_label']
```

```
[ ] # Remove duplicates in-place Flights
    flights_df.drop_duplicates(subset='ids', keep='first', inplace=True)
```

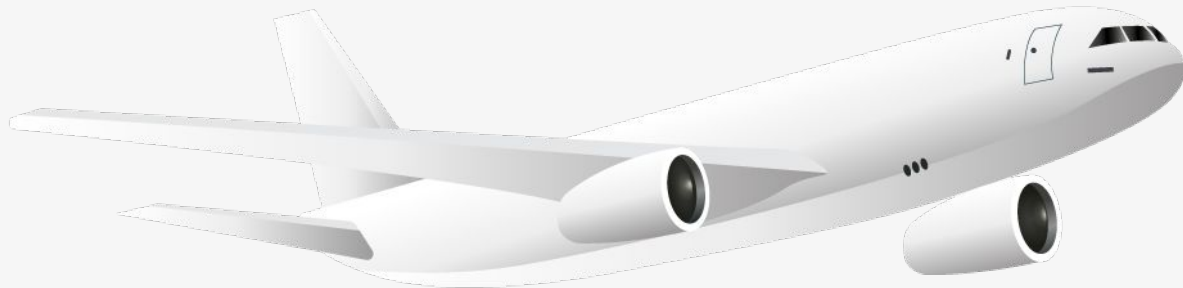
	FlightDate	Day_Of_Week	Airline	Tail_Number	Dep_Airport	Dep_CityName	DepTime_label	Dep_Delay	Dep_Delay_Tag	Dep_Delay_Type	...	Distance_type	Delay_Carrier	Delay_Weather	Delay
71414	2023-01-01	7	American Airlines Inc.	N819AW	CLT	Charlotte, NC	Evening	-7	0	Low <5min	...	Short Haul >1500Mi	0	0	
71099	2023-01-01	7	American Airlines Inc.	N819AW	CLT	Charlotte, NC	Evening	-4	0	Low <5min	...	Short Haul >1500Mi	0	0	
426762	2023-01-01	7	Skywest Airlines Inc.	N751SK	DFW	Dallas/Fort Worth, TX	Evening	-11	0	Low <5min	...	Short Haul >1500Mi	0	0	
422732	2023-01-01	7	Skywest Airlines Inc.	N751SK	DFW	Dallas/Fort Worth, TX	Evening	-6	0	Low <5min	...	Short Haul >1500Mi	0	0	
445519	2023-01-01	7	Skywest Airlines Inc.	N608UX	EGE	Eagle, CO	Afternoon	-7	0	Low <5min	...	Short Haul >1500Mi	0	0	

Dữ liệu trùng lặp

# Tạo Primary Key & clean data cho bảng Cancelled\_Diverted\_2023

```
[ ]  
cancelled_diverted_df['ids'] = cancelled_diverted_df['FlightDate'] + cancelled_diverted_df['Dep_Airport']  
    + cancelled_diverted_df['Arr_Airport'] + cancelled_diverted_df['Tail_Number']  
    + cancelled_diverted_df['DepTime_label']
```

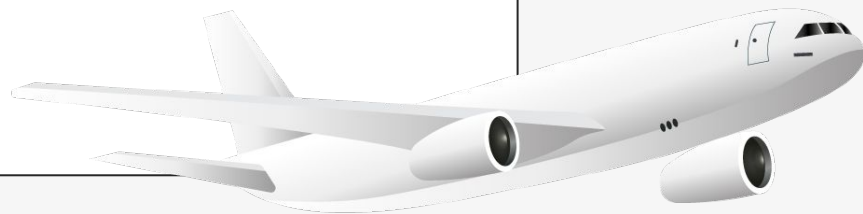
```
[ ] # Remove duplicates in-place  
cancelled_diverted_df.drop_duplicates(subset='ids', keep='first', inplace=True)
```



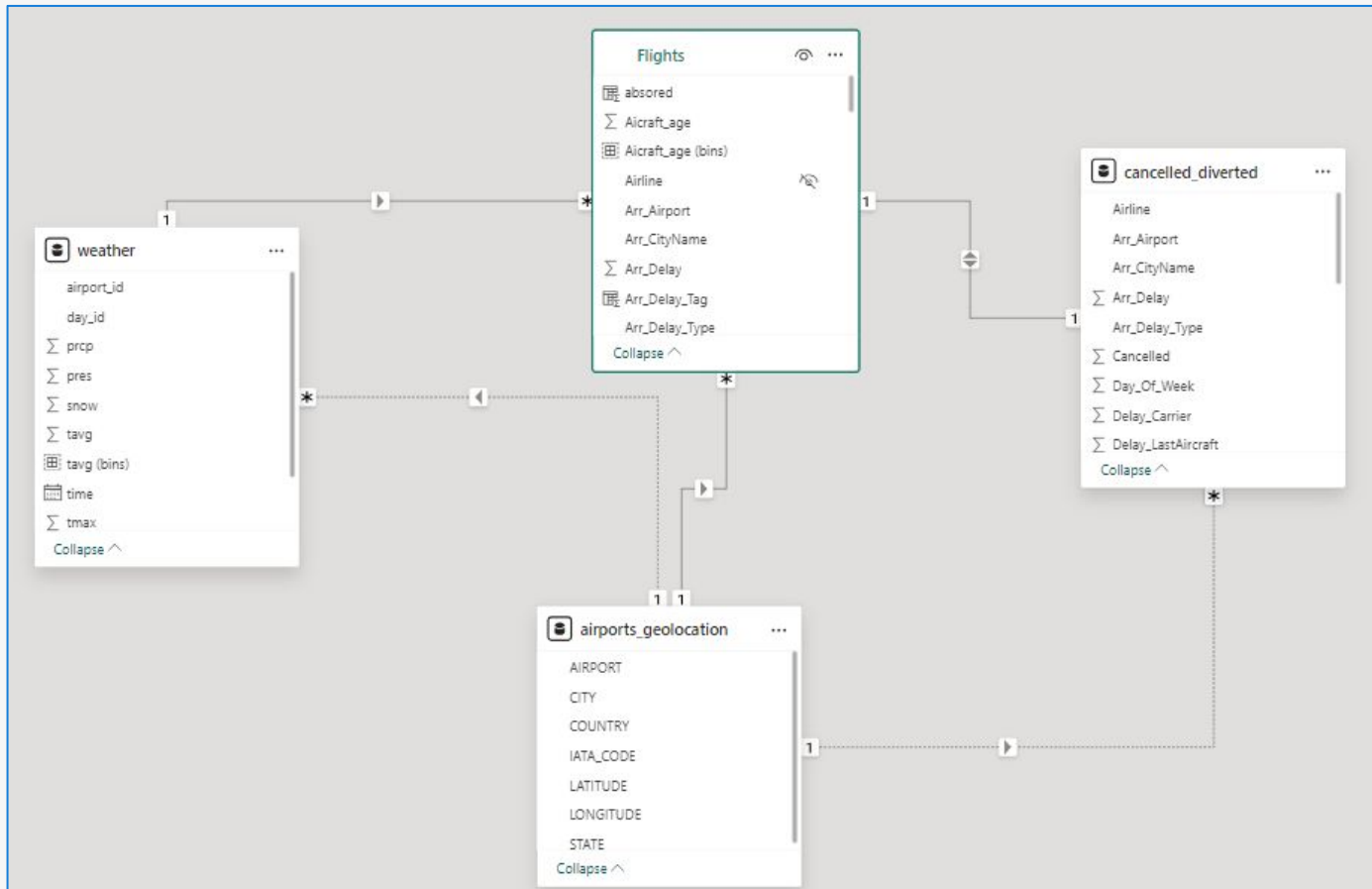
# Tạo Primary Key & clean data cho bảng Weather

```
[ ] weather_df['day_id'] = weather_df['time'] + weather_df['airport_id']  
# Find duplicate rows based on 'ids'  
duplicates = weather_df.duplicated(subset='day_id', keep=False)  
  
# Display the duplicate rows  
weather_df[duplicates].count()
```

```
⇒ time          0  
   tavg          0  
   tmin          0  
   tmax          0  
   prcp          0  
   snow          0  
   wdir          0  
   wspd          0  
   pres          0  
   airport_id    0  
   day_id        0  
   dtype: int64
```



# Tạo relationship - Thiết lập modeling data



**Table Flights & Cancelled diverted là bảng Fact**

**Table Weather & Airports Geolocation là các bảng Dim mô tả cho 2 table Fact**



# 03

## PHÂN TÍCH TỔNG QUAN VÀ KHÁM PHÁ

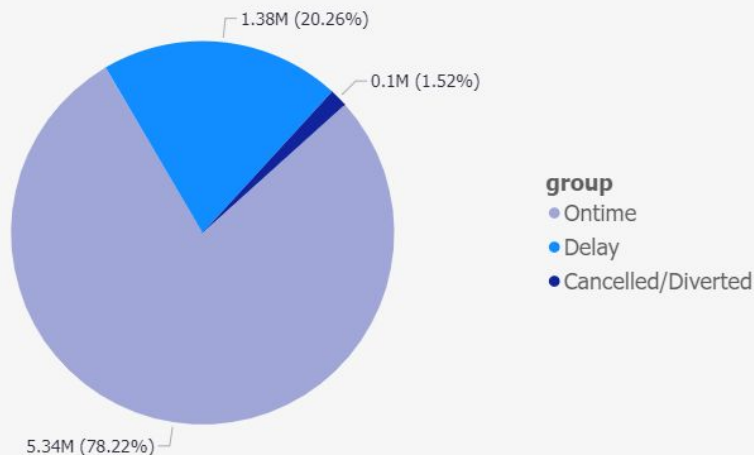
# Thông tin về chuyến bay

Total Flights 2023

6.82M



Ontime Ratio of Total Flights



## Total Flights

Tổng xấp xỉ 6,82 triệu chuyến bay

- Hủy và chuyển hướng bay khoảng 100 ngàn chuyến.
- Delay khoảng 1.38 triệu chuyến.
- Đúng giờ 5.34 triệu chuyến.

## Rate

- Tỷ lệ chuyến bay bị delay ~ 20%
- Trung bình thời gian delay ~ 69 phút





# Danh mục các trường hợp Delay



## Delay By NAS

Nguyên nhân chậm trễ nằm trong kiểm soát của hệ thống hàng không quốc gia, bao gồm: Hoạt động của sân bay, lưu lượng giao thông đông đúc và kiểm soát không lưu.



## Delay By Carrier

Nguyên nhân nằm trong tầm kiểm soát của hãng hàng không (ví dụ: các vấn đề về bảo trì hoặc phi hành đoàn, vệ sinh máy bay, xếp hành lý, tiếp nhiên liệu, v.v.).



## Delay By last aircraft

Chuyến bay trước với cùng máy bay đã đến muộn, khiến chuyến bay hiện tại khởi hành muộn



## Delay By Security

Nguyên nhân do sơ tán khỏi nhà ga hoặc phòng chờ, lên máy bay lại do vi phạm an ninh, thiết bị soi chiếu không hoạt động và/hoặc xếp hàng dài quá 29 phút tại các khu vực soi chiếu.



## Delay By Weather

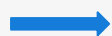
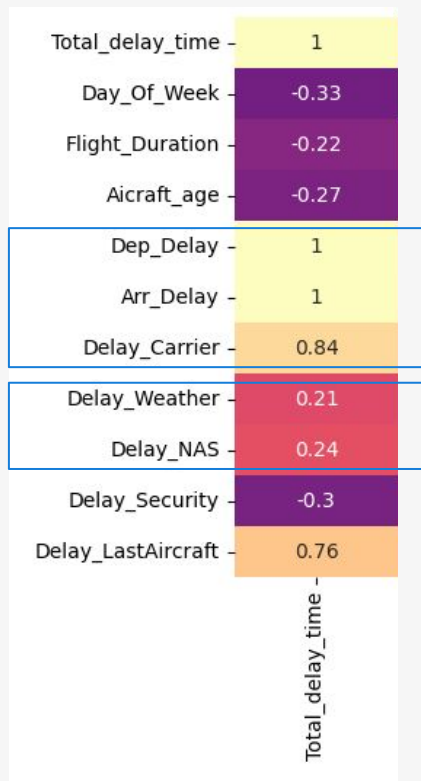
Nguyên nhân do các điều kiện thời tiết khắc nghiệt hoặc nguy hiểm được dự báo hoặc biểu hiện tại điểm khởi hành, trên đường hoặc tại điểm đến

# Tính chỉ số tổng thời gian delay

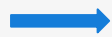


```
1 Total_delay_time = Normal_Flights[Delay_Carrier]+Normal_Flights[Delay_LastAircraft]+Normal_Flights[Delay_NAS]+Normal_Flights  
[Delay_Security]+Normal_Flights[Delay_Weather]
```

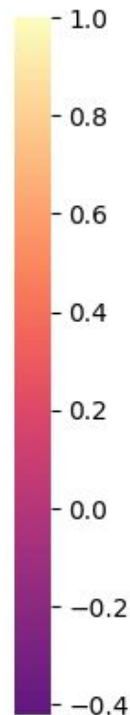
# Tìm hiểu sự tương quan giữa các yếu tố



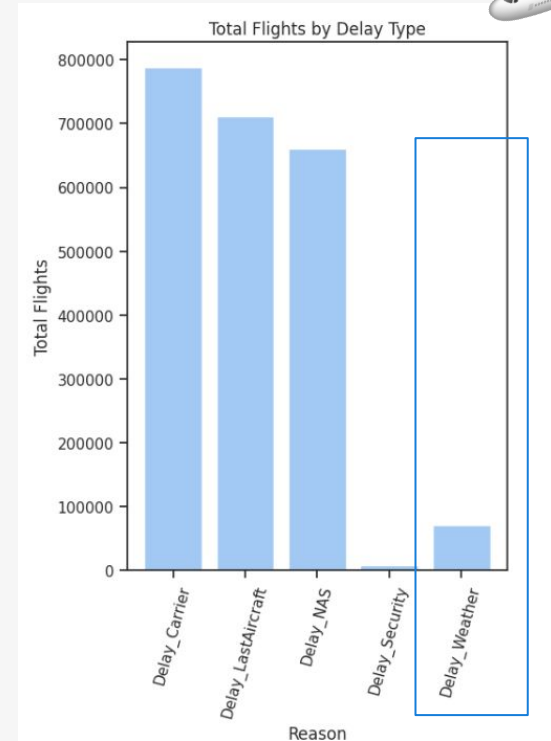
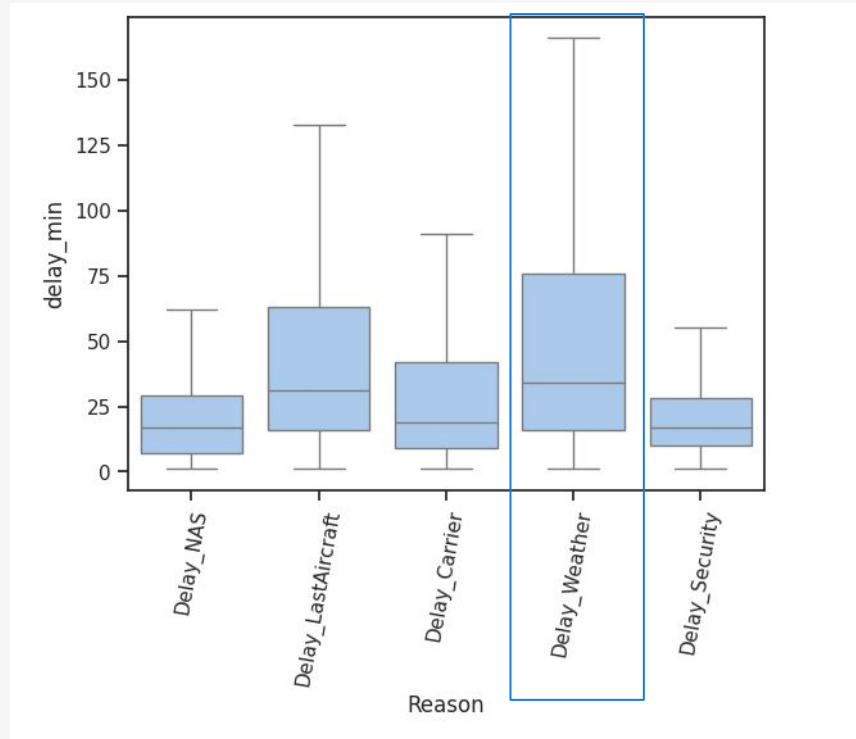
Tác động cùng chiều, **mạnh** đến thời gian delay



Tác động cùng chiều, **yếu** đến thời gian delay



# Phân tích thăm dò về các mục delay



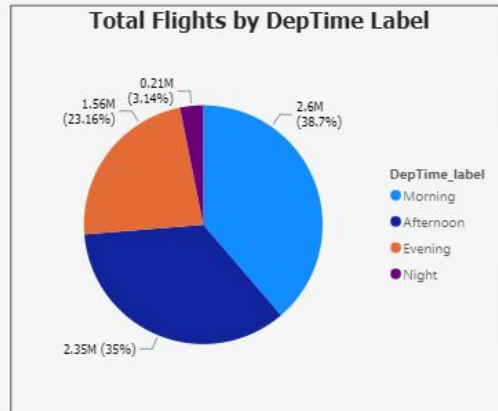
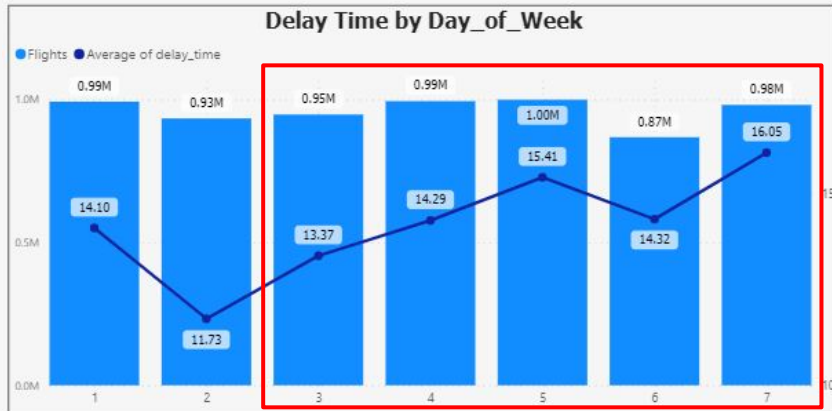
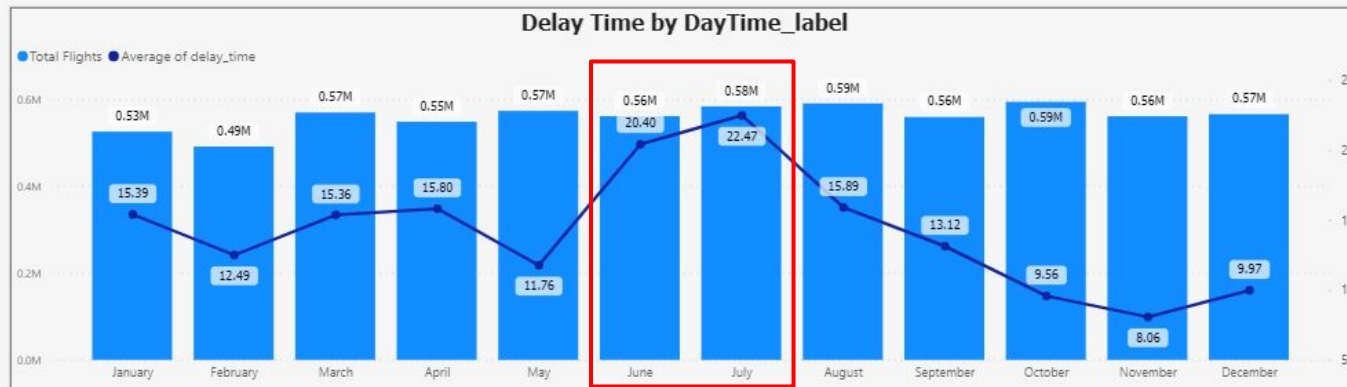
Số lượng chuyến bay delay do thời tiết **khá ít**, nhưng thời gian delay khá dài



# 04

## PHÂN TÍCH CÁC YẾU TỐ ẢNH HƯỞNG

# Delay time by season - Yếu tố thời gian



Tỷ lệ số chuyến bay theo tháng và ngày tương đối đồng đều nhau, chỉ khác biệt ở các buổi trong ngày.

- Về tháng, khi đi vào các tháng cao điểm du lịch như tháng 6 và 7, lượng chuyến bay có thể sẽ quá tải dẫn đến tình trạng delay đột biến.
- Về ngày trong tuần, delay xuất hiện tăng tiến từ thứ 4 đến cuối tuần.
- Số lượng chuyến bay tập trung nhiều vào buổi sáng và trưa

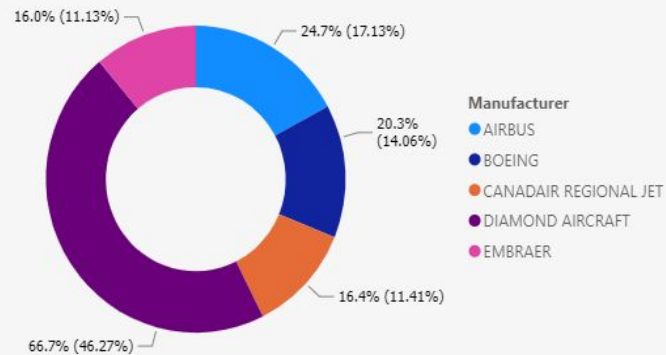
=> Các đoạn cao điểm như nghỉ lễ, mùa du lịch hoặc các thời điểm cuối tuần sẽ làm tăng cao chỉ số delay.

# Delay time by Aircraft - Yếu tố hãng máy bay

Delay Frequency by Aircraft age



Delay rate of flights according to each manufacturer



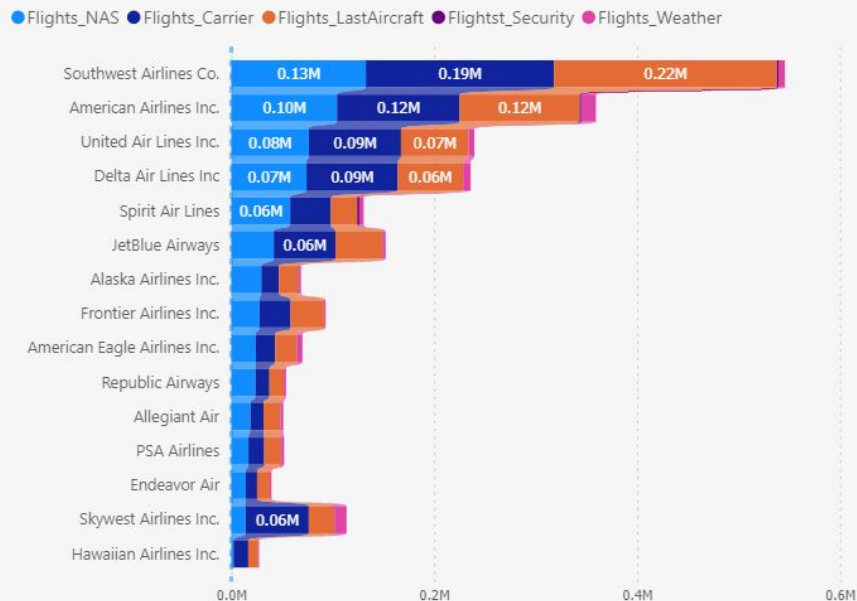
- Máy bay có tuổi đời càng lớn thì tỷ lệ delay chuyến bay càng cao.
- Tỷ lệ delay của hãng DIAMOND AIRCRAFT là cao nhất. Nguyên nhân do lượt bay của hãng máy bay này quá thấp với 3 lượt bay và trễ đến 2 lượt nên số liệu bị cao bất thường.

Ontime / Late Flights by Manufacturer

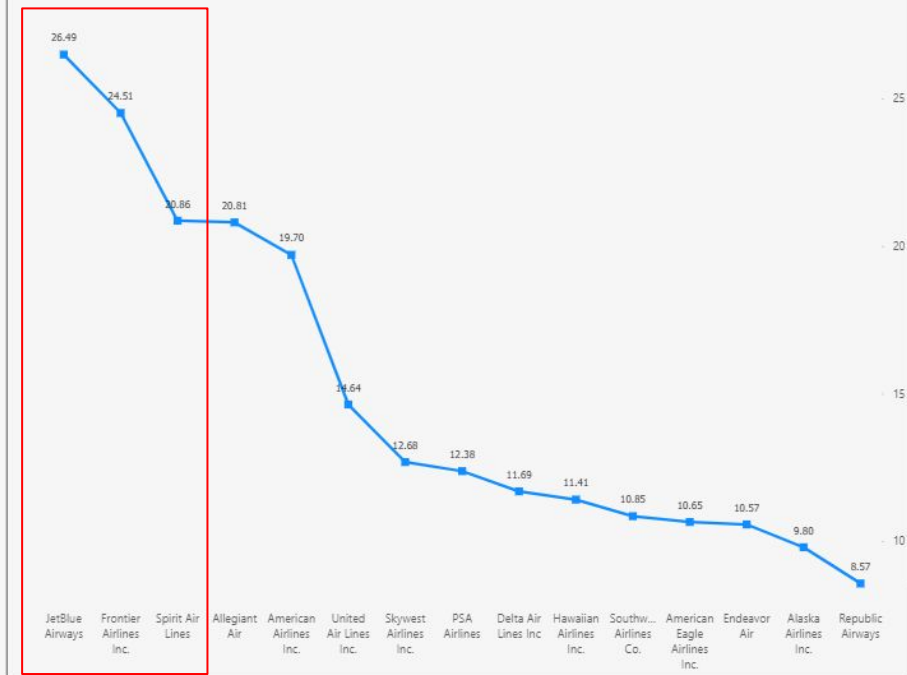


# Delay time by Airline

Delay Categories by Airlines



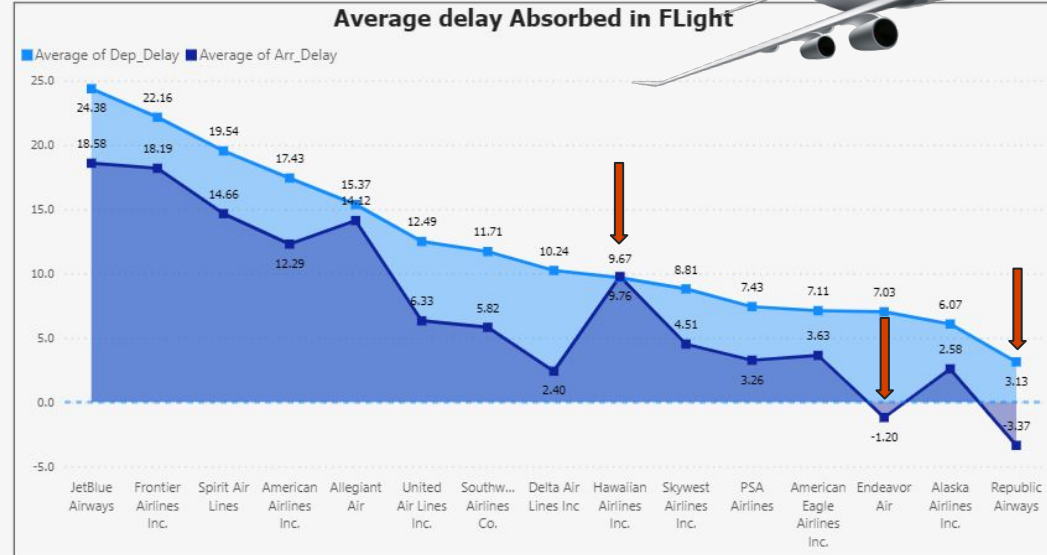
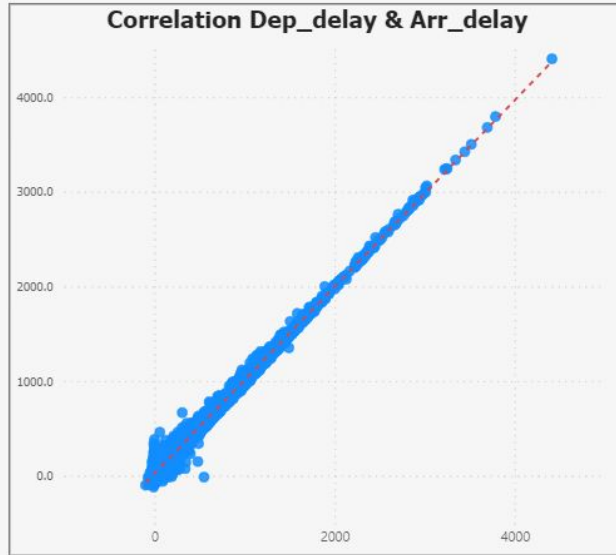
Ranking Airline by delay time



- Hầu hết các chuyến bay trễ của các hãng hàng không đều chịu ảnh hưởng **bởi nguyên nhân NAS, Carrier & Last aircraft.**
- Top 3 hãng hàng không tập trung nhiều chuyến bay nhất là **Southwest, American và United Airlines**
- Top 3 các hãng hàng không có **thời gian delay** cao nhất : **JetBlue, Frontier, Spirit**



# Airline plans



Thời gian trễ khi khởi hành có mối tương quan thuận chiều với thời gian trễ khi hạ cánh

Các hãng hàng không thường tận dụng khoảng thời gian bay, bay với tốc độ nhanh hơn để giảm thiểu thời gian trễ khi hạ cánh

# Delay time by weather

Month	Average of Total_delay_time	Temperature (°C)	Wind Speed (km/h)	Wind Direction	Snow Depth	Precipitation (mm)	Pressure (hPa)
January	15.39	29.70	55.20	360.00	1780	1,045.80	107.20
February	12.49	29.30	56.20	360.00	910	1,051.00	282.70
March	15.36	29.20	62.30	360.00	940	1,040.70	105.70
April	15.80	33.10	45.00	360.00	810	1,049.20	571.50
May	11.76	34.20	66.60	360.00	250	1,035.20	316.70
June	20.40	36.40	41.70	360.00	0	1,027.40	131.80
July	22.47	42.20	37.40	359.00	0	1,033.50	176.50
August	15.89	40.30	41.00	360.00	0	1,031.70	177.80
September	13.12	38.50	48.20	360.00	0	1,030.50	204.50
October	9.56	33.00	55.40	360.00	250	1,039.50	169.90
November	8.06	30.00	78.30	360.00	584	1,040.90	191.30
December	9.97	29.60	59.40	360.00	660	1,041.80	225.00
<b>Total</b>	<b>14.21</b>	<b>42.20</b>	<b>78.30</b>	<b>360.00</b>	<b>1780</b>	<b>1,051.00</b>	<b>571.50</b>



Thời gian delay cao nhất năm 2023 thuộc tháng 6,7 , tương quan với điều kiện nhiệt độ trung bình trong 2 tháng này cũ tăng cao

Độ sâu tuyết tập trung vào các tháng mùa đông t12, t1

→ Các điều kiện thời tiết năm 2023 không có gì bất thường

# Delay time by weather



Top 5 cities - Delay flights by weather

STATE	City	Delay Flights by weather	°C	Avg wdir	Avg wspd	Avg snow	Avg prcp	Avg pres
TX	Dallas/Fort Worth, TX	6,178	21.46	167.06	16.29	0.01	1.88	1,014.72
IL	Chicago, IL	4,336	12.70	173.92	14.78	2.06	2.40	1,015.69
FL	Orlando, FL	4,087	23.67	160.82	11.92	0.00	3.07	1,017.12
GA	Atlanta, GA	3,238	17.64	205.78	7.50	0.00	3.16	1,017.11
CO	Denver, CO	3,124	10.79	183.89	15.41	9.44	1.28	1,013.94
<b>Total</b>		<b>20,963</b>	<b>16.51</b>	<b>180.63</b>	<b>13.15</b>	<b>2.42</b>	<b>2.33</b>	<b>1,015.64</b>

Bottom 5 cities - Delay flights by weather

STATE	City	Delay Flights by weather	°C	Avg wdir	Avg wspd	Avg snow	Avg prcp	Avg pres
AK	Adak Island, AK	1	5.63	217.25	22.98	0.00	3.43	1,009.69
AK	Dillingham, AK	1	10.73	173.21	12.44	0.00	3.62	1,010.60
AK	Kotzebue, AK	1	-6.74	155.19	16.30	0.00	0.95	1,009.65
GU	Guam, TT	1	27.96	98.87	16.29	0.00	8.07	1,010.22
MN	St. Cloud, MN	1	-0.80	204.26	12.90	191.48	1.26	1,017.19
NC	New Bern/Morehead/Beaufort, NC	1	18.18	163.53	11.72	0.00	4.38	1,016.93
WA	Walla Walla, WA	1	13.88	179.03	11.06	0.00	1.26	1,016.65
WY	Cody, WY	1	7.08	172.06	12.38	0.00	0.95	1,015.89
<b>Total</b>		<b>8</b>	<b>12.65</b>	<b>146.24</b>	<b>15.03</b>	<b>10.76</b>	<b>4.05</b>	<b>1,011.92</b>

Không có sự khác biệt đáng kể về điều kiện khí hậu ở các thành phố thường xuyên bị delay do thời tiết và các thành phố ít bị delay bị delay vì lý do thời tiết

# 05

## KẾT LUẬN & ĐỀ XUẤT





## *Các yếu tố ảnh hưởng*

Các khoảng thời gian có số lượng chuyến bay cao sẽ ảnh hưởng đến thời gian delay càng cao mùa

3 Airlines không có thời gian chờ dài nhất : **JetBlue, Frontier, Spirit**

3 Airline có số lượng chuyến bay nhiều nhất: **Southwest, Delta Air Lines và American Airlines**

Máy bay có tuổi đời càng lớn thì tần suất delay chuyến bay càng cao

Nguyên nhân do thời tiết ảnh hưởng nhiều đến việc delay, các hãng hàng không cần phân tích thêm các điều kiện thời tiết để sắp xếp lịch bay phù hợp

6.72M

Total Flights

5963

Airlines

350

Airports

335

Cities

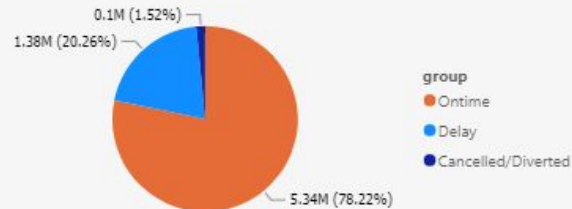
5

Manufacturer

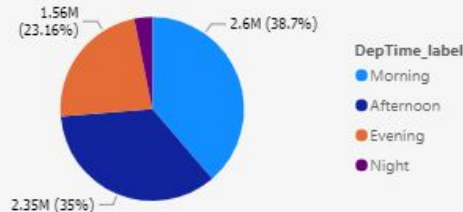
20

Model

Ontime Ratio of Total Flights



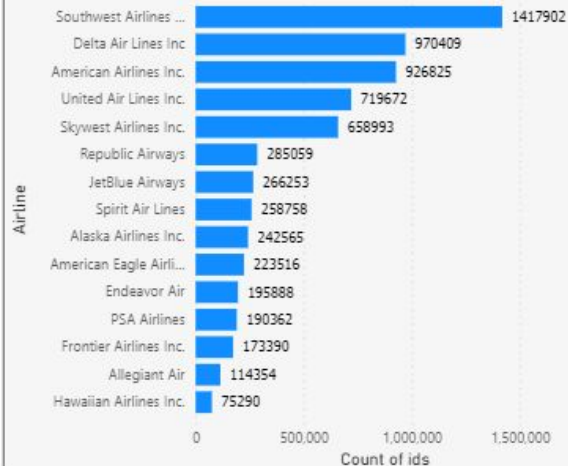
Total Flights by DepTime Label



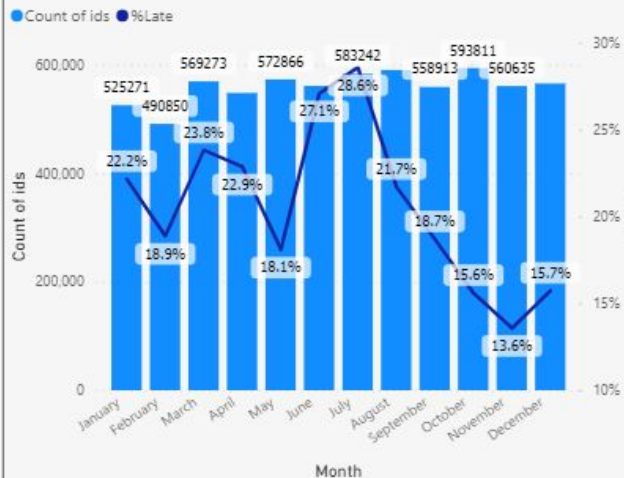
Ontime / Late Flights by Manufacturer



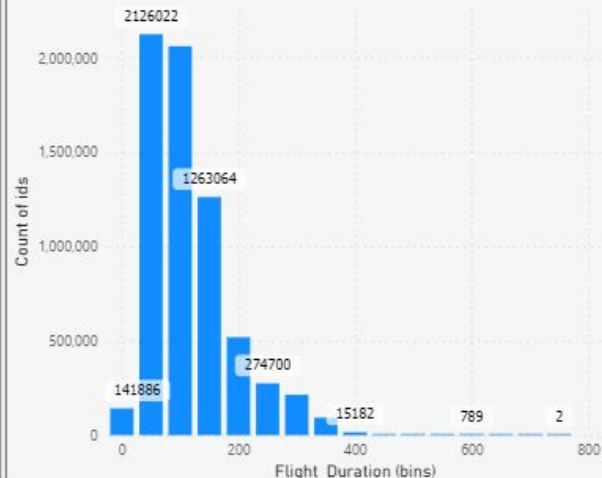
Count of ids by Airline



Count of ids and %Late by Month



Distance



# Thank you !

Credits: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, infographics & images by **Freepik**

