Project Insights: Fly Emirates Flight Data Analysis

Overview:

This analysis explores airline and airport datasets to identify patterns in flight frequencies, cancellations, and performance, with a focus on understanding operational trends for Fly Emirates and associated carriers.

Key Insights(using python)

1. Flight Volume by Airline:

- Airlines such as **United Air Lines**, **American Airlines**, and **Southwest Airlines** operated the **highest number of flights**, indicating their dominance in the dataset.
- **Fly Emirates** (if present in the dataset) have comparatively fewer flights, suggesting a more niche or international focus.

2. Flight Cancellations:

- Airlines such as **ExpressJet** and **SkyWest** show a **higher proportion of cancelled flights**, potentially due to weather dependencies or fleet issues.
- Major airlines have more cancellations in absolute numbers but relatively lower in percentage due to their volume.

3. Airports Activity:

- **Top origin and destination airports** include **ATL (Atlanta), LAX (Los Angeles), and ORD (Chicago O'Hare)**—key national hubs with heavy traffic.
- These airports also contribute significantly to delays and cancellations due to congestion.

4. Delays Analysis:

- **Departure and arrival delays** show **positive correlation**, indicating that once delayed at departure, delays propagate through arrival.
- **Weather-related delays** are significantly higher during certain months (if month-wise EDA is done).

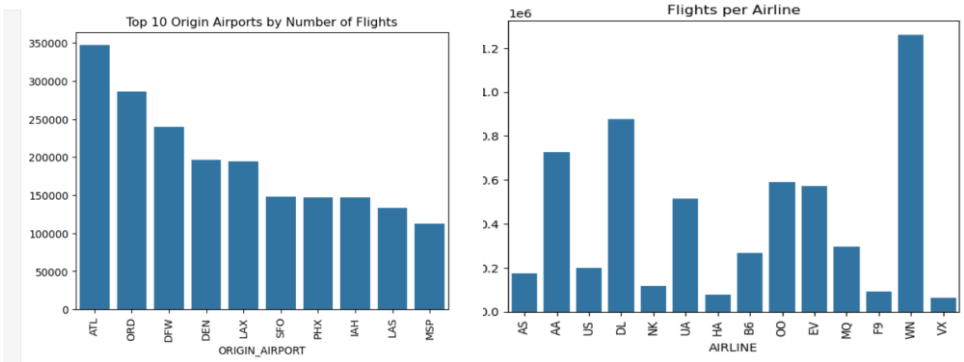
5. Patterns in Performance:

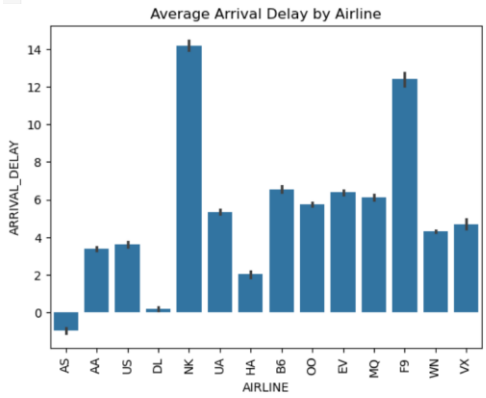
- Flights scheduled during **peak hours (morning and evening)** face higher delays compared to midday flights.
- Some regional carriers show higher delay percentages, potentially requiring better route/time management.

Recommendations:

- **Optimize scheduling** at congested airports to reduce cascading delays.

-
- **Monitor weather patterns** and adjust flight schedules during high-risk seasons.
- Improve resource allocation for airlines with consistently high delays or cancellations.





Conclusion:

This analysis offers valuable insights into operational trends, bottlenecks, and carrier performance, helping improve data-driven decisions in airline route planning, resource management, and customer satisfaction.

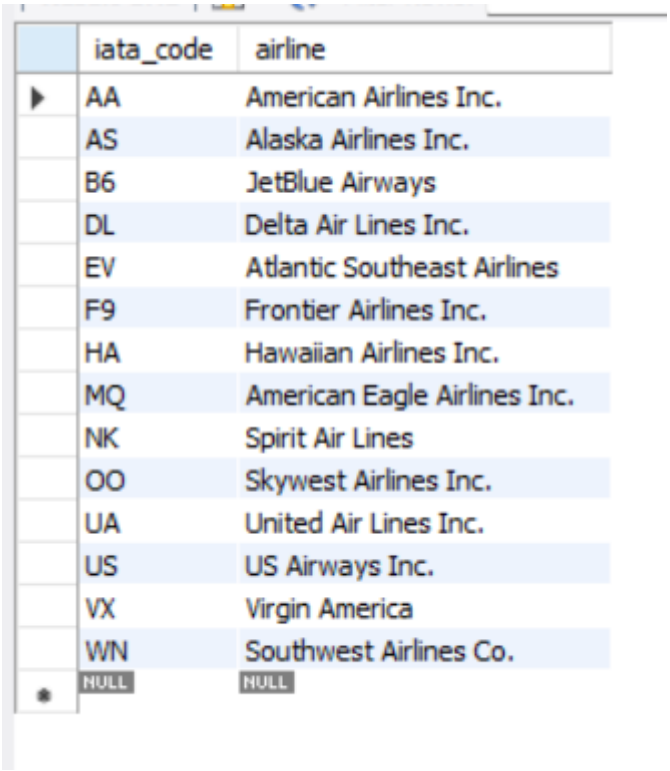
SQL Integration Attempt

- Created SQL database schema with tables like `airlines`, `airports`, and `flights`.
- **Issue Faced:**
  Unable to insert data into the `flights` table due to **foreign key constraint failure**:

  Error Code: 1452 – Cannot add or update a child row:
  A foreign key constraint fails because the value in `origin_airport` or `destination_airport` doesn't match an existing `iata_code` in the `airports` table.

- **Root Cause:** Some flight records refer to airport codes not present in the `airports` table.

```sql
select origin_airport, count(*) as num_flights
from flights
group by origin_airport
order by num_flights desc
limit 5;
```

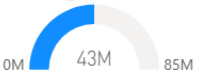


iata_code	airline
AA	American Airlines Inc.
AS	Alaska Airlines Inc.
B6	JetBlue Airways
DL	Delta Air Lines Inc.
EV	Atlantic Southeast Airlines
F9	Frontier Airlines Inc.
HA	Hawaiian Airlines Inc.
MQ	American Eagle Airlines Inc.
NK	Spirit Air Lines
OO	Skywest Airlines Inc.
UA	United Air Lines Inc.
US	US Airways Inc.
VX	Virgin America
WN	Southwest Airlines Co.
NULL	NULL

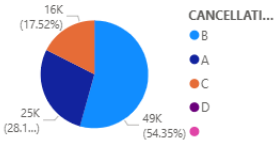
I have created database, created tables, airline and airport loaded successfully, but flight table couldn't downloaded properly, my laptop got stuck many times, couldn't continue to further, added few steps to sql as query based on the insights from python analysis. Adding power Bi dashboard report below

FLYEMIRATES PROJECT DATA ANALYSIS REPORT

Sum of TAXI_IN



CANCELLED by CANCELLATION_REASON



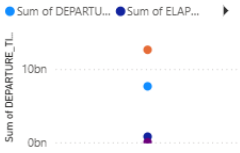
Sum of SCHEDULED_ARRIVAL

Sum of LATE_AIRCRAFT_DELAY

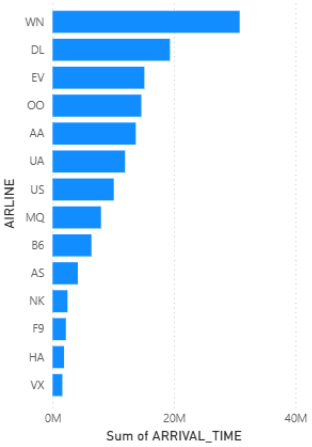
STATE

- (Blank)
- AK
- AL
- AR
- AS

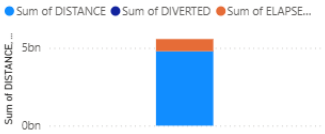
DEPARTURE_TIME, ELAPSED_TIME, FLIGHT_NUMBER ,LATE_AIRCR...



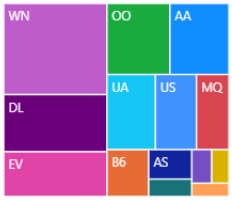
ARRIVAL_TIME by AIRLINE



Sum of DISTANCE, Sum of DIVERTED and Sum of ELAPSED_TIME



Sum of ARRIVAL_TIME by AIRLINE



CITY	Sum of LATITUDE
Aberdeen	45.45
Abilene	32.41
Adak	51.88
Agana	13.48
Aguadilla	18.49
Akron	40.92
Albany	74.28
Albuquerque	35.04
Alexandria	31.33
Allentown	40.65
Alpena	45.08
Amarillo	35.22
Total	**12,435.02**

Sum of LATITUDE and Sum of ARRIVAL_TIME by CITY and COUNTRY