REPORT ON U.S FLIGHT ANALYSIS

The US Airline Performance Analysis project aims to evaluate the operational efficiency and punctuality of major US airlines using a comprehensive dataset of one million flights. The primary objectives are to identify patterns in flight delays, cancellations, and route distribution, and to provide actionable insights for improving airline performance and passenger satisfaction.

The project leverages Power BI to visualize key metrics, including total flights, average delays, airline-wise performance, seasonal trends, and geographical coverage. Through a combination of line charts, bar charts, pie charts, and maps, the analysis highlights high-delay periods, major contributing airlines, and key flight routes.

By integrating quantitative metrics with intuitive visualizations, this project enables stakeholders to monitor performance, pinpoint operational bottlenecks, and implement strategies to enhance on-time arrivals and overall efficiency in the US airline industry.

This project involves cleaning and analyzing a large U.S. flight dataset to identify flight performance patterns, particularly focusing on on-time arrival rates of airlines. The data includes detailed flight information such as schedules, delays, and cancellations. The analysis was conducted using MySQL for data preparation and Power BI for visualization.

Database and Table Creation

A new database named **flight_data** was created to organize and manage all flight-related information.A structured table called **flights** was designed with appropriate data types to store flight attributes like year, month, day, airline, flight number, origin and destination airports, scheduled and actual departure/arrival times, and delay information.

Duplicate Removal

Duplicate records were identified based on key columns such as:

- Year, Month, Day
- Airline
- Flight Number
- Origin and Destination Airports

Duplicates were then removed to ensure data accuracy and consistency. Each record was assigned a unique **ID** to simplify record management.

Handling Missing Values

The dataset was examined for missing or blank values in critical fields such as **AIRLINE**, **FLIGHT_NUMBER**, **DEPARTURE_DELAY**, and **ARRIVAL_DELAY**.
Records with missing airline or flight number information were deleted, as these were essential for reliable analysis.

This step ensured that only valid and complete data remained in the dataset.

Removing Unused or Empty Columns

Columns that contained only null values or were not relevant to the analysis (such as **SECURITY_DELAY**) were dropped to simplify the dataset and improve performance.

Creating a Flight Date Field

A new derived column, **flight_date**, was created by combining the year, month, and day columns.
This made it easier to analyze time-based trends and visualize data chronologically in Power BI.

Data Type Consistency

- Delay and distance columns were converted to **integer** data types.

Time Format Standardization

- SCHEDULED_DEPARTURE values were converted to a readable 4-digit format using LPAD().

Removing Canceled or Diverted Flights

- Rows where CANCELLED = 1 or DIVERTED = 1 were deleted.
  This ensures only completed flights are analyzed.

Airline Code Standardization

- Joined with an airline reference table to replace IATA codes with full airline names.

US Airline Performance Analysis using Power BI

The report provides an overview of US airline operations, highlighting flight volumes, delays, and route coverage based on one million recorded flights. The analysis identifies key performance patterns and areas for operational improvement.

1. Key Metrics

- **Total Flights:** 1,000,000
- **Average Departure Delay:** 54.95 minutes
- **Average Arrival Delay:** 58.91 minutes

These metrics indicate that, on average, flights are delayed by nearly an hour, suggesting operational and scheduling challenges that affect on-time performance.

2. Monthly Delay Trends

Analysis of average arrival delays by month reveals a **seasonal pattern**:

- **Peak delays:** June, coinciding with high summer travel demand.
- **Lowest delays:** September–October, reflecting off-peak periods.

This information can guide staffing adjustments, flight scheduling, and customer communication during high-delay periods.

3. Airline Performance

- **Top carriers by flight volume:** Southwest (WN), American Airlines (AA), Delta (DL).
- **Mid-sized carriers:** Atlantic Southeast Airlines (EV) handles ~109,000 flights.
- **Smaller carriers:** NK, US, B6 have comparatively fewer flights.

While large carriers contribute the most to cumulative delays due to flight volume, smaller carriers may show better punctuality. Targeted analysis of high-traffic airlines is recommended for delay reduction initiatives.

4. Geographical Coverage

The origin and destination map highlights:

- Dense flight activity across North America.
- Sparse international connections to Europe, Asia, South America, Africa, and Australia.

This visualization helps identify high-traffic routes, potential bottlenecks, and opportunities for route optimization.

5. Flight Status Distribution

- **On-time or minor delays:** ~53.5%
- **Significant delays:** Remaining flights
- **Cancellations:** 0.21%

The low cancellation rate reflects strong operational continuity, while delays remain the primary challenge affecting customer satisfaction.

Recommendations

1. **Focus on High-Volume Airlines:** Prioritize delay mitigation for WN, AA, and DL.
2. **Seasonal Planning:** Increase staffing and improve scheduling during peak summer months.
3. **Route Optimization:** Analyze high-delay routes for targeted operational improvements.