

# Airline Passenger Travel & Flight Performance Analysis

## 1. Project Overview

The Airline Passenger Travel & Flight Performance Analysis project focuses on analyzing airline passenger behavior and flight performance metrics using real-world aviation data. The goal of this project is to extract actionable insights related to passenger demographics, travel trends, flight punctuality, cancellations, and geographical patterns.

This project demonstrates an **end-to-end data analytics workflow** using:

- Python for data cleaning
- SQL for querying and analysis
- Power BI for interactive dashboarding and visualization

## 2. Business Objectives

The key business questions addressed in this project are:

1. What is the total number of passenger trips recorded?
2. What is the distribution of passengers by gender?
3. Which age group accounts for the highest number of travellers?
4. Which departure airport is used most frequently?
5. Identify the country with the maximum number of passenger trips?
6. Which continent has the highest number of departing flights?
7. During which month do the most departures occur?
8. What is the overall distribution of flight statuses (on-time, delayed, cancelled)?
9. How many trips were made by each age group?
10. What percentage of flights were cancelled?

### **3. Dataset Description**

- **Dataset Name:** Airline Dataset
- **Records:** ~98,000+ passenger trips
- **Data Type:** Structured CSV data

#### **Key Columns:**

- Passenger Name
- Age & Age Category
- Gender
- Nationality
- Departure & Arrival Airport
- Country & Continent
- Departure Date & Month
- Flight Status (On-Time, Delayed, Cancelled)
- Flight ID

### **4. Data Cleaning & Preprocessing (Python)**

**Tool Used:** Python (Pandas, NumPy)

#### **Cleaning Steps Performed:**

- Removed duplicate passenger records
- Handled missing values in key columns
- Standardized column names
- Created derived columns:
  - age\_category
  - departure\_month
  - departure\_year

- Removed redundant columns (e.g., duplicate continent fields)
- Ensured correct data types for dates and numerical fields

Output of this step was a **clean and analysis-ready dataset** used for SQL and Power BI.

### **Data Integration from Python to MySQL:**

The cleaned airline dataset was integrated from Python into a MySQL database using a database connector. Record counts and sample queries were used to validate successful data transfer, enabling efficient SQL analysis and seamless integration with Power BI for visualization.

## **5. SQL Analysis & Results**

**Tool Used:** MySQL

SQL was used to answer analytical business questions efficiently.

### **Key SQL Insights:**

- Q1. Total Number of Passenger Trips Recorded
- Q2. Passenger Distribution by Gender
- Q3. Age Group with the Highest Number of Travelers
- Q4. Most Frequently Used Departure Airport
- Q5. Country with the Highest Number of Passenger Trips
- Q6. Continent with Highest Departing Flights
- Q7. Peak Travel Month Based on Departures
- Q8. Overall Distribution of Flight Statuses
- Q9. Passenger Trips by Age Group
- Q10. Percentage of Flights Cancelled

## Results:

### Q1. Total Passenger Trips

#### SQL Query

```
-- Q1. What is the total number of passenger trips recorded?
```

```
SELECT COUNT(passenger_full_name) AS Total_Count FROM airline_data;
```

#### Result

	Total_Count
▶	98610

### Q2. Distribution of Passengers by Gender

#### SQL Query

```
-- Q2. What is the distribution of passengers by gender?
```

- ```
SELECT gender, COUNT(*) AS total_values FROM airline_data GROUP BY gender;
```

#### Result

|   | gender | total_values |
|---|--------|--------------|
| ▶ | Female | 49021        |
|   | Male   | 49598        |

### Q3. Highest Traveling Age Group

#### SQL Query

```
-- Q3. Which age group accounts for the highest number of travelers?
```

```
SELECT age_category, COUNT(*) AS highest_travelled_age_category FROM airline_data GROUP BY age_category  
ORDER BY highest_travelled_age_category DESC LIMIT 1;  
/
```

## Result

|   | age_category   | highest_travelled_age_category |
|---|----------------|--------------------------------|
| ▶ | Senior_citizen | 43831                          |

## Q4. Most Frequently Used Departure Airport

### SQL Query

```
-- Q4. Which departure airport is used most frequently?  
SELECT airport_name, COUNT(*) AS highest_departed_airport FROM airline_data GROUP BY airport_name  
ORDER BY highest_departed_airport DESC LIMIT 1;
```

## Result

|   | airport_name      | highest_departed_airport |
|---|-------------------|--------------------------|
| ▶ | San Pedro Airport | 43                       |

## Q5. Top Passenger Nationality

### SQL Query

```
-- Q5. Identify the country with the maximum number of passenger trips.  
SELECT nationality, COUNT(*) AS highest_passengers FROM airline_data GROUP BY nationality ORDER BY highest_passengers DESC LIMIT 1;
```

## Result

|   | nationality | highest_passengers |
|---|-------------|--------------------|
| ▶ | China       | 18317              |

## Q6. Continent with Highest Departing Flights

### SQL Query

```
-- Q6. Which continent has the highest number of departing flights?  
SELECT continents, COUNT(*) AS highest_total_flights FROM airline_data GROUP BY continents ORDER BY highest_total_flights DESC LIMIT 1;
```

### Result

|   | continents    | highest_total_flights |
|---|---------------|-----------------------|
| ▶ | North America | 32033                 |

## Q7. Peak Travel Month

### SQL Query

```
-- Q7. During which month do the most departures occur?
```

```
SELECT departure_month, COUNT(*) AS highest_travel_month FROM airline_data GROUP BY departure_month  
ORDER BY highest_travel_month DESC LIMIT 1;
```

### Result

|   | departure_month | highest_travel_month |
|---|-----------------|----------------------|
| ▶ | 8               | 8544                 |

## Q8. Flight Status Distribution

### SQL Query

```
-- Q8. What is the overall distribution of flight statuses (on-time, delayed, cancelled)?
```

```
SELECT flight_status, COUNT(*) AS total_values FROM airline_data GROUP BY flight_status;
```

## Result

|   | flight_status | total_values |
|---|---------------|--------------|
| ▶ | On Time       | 32846        |
|   | Delayed       | 32831        |
|   | Cancelled     | 32942        |

## Q9. Passenger Trips by Age Group

### SQL Query

```
-- Q9. How many trips were made by each age group?
```

```
SELECT age_category, COUNT(*) AS total_values FROM airline_data GROUP BY age_category;
```

## Result

|   | age_category   | total_values |
|---|----------------|--------------|
| ▶ | Senior_citizen | 43831        |
|   | Adult          | 18605        |
|   | Middle_aged    | 16537        |
|   | Teen           | 6599         |
|   | Child          | 7627         |
|   | Baby           | 5420         |

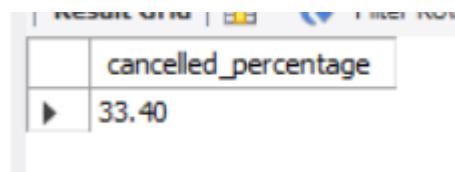
## Q10. Percentage of Flights Cancelled

### SQL Query

```
-- Q10. What percentage of flights were cancelled?
```

```
SELECT ROUND((SUM(CASE WHEN flight_status = 'Cancelled' THEN 1 ELSE 0 END) * 100.0)/ COUNT(*),2) AS cancelled_percentage FROM airline_data;
```

## Result



|   |                      |
|---|----------------------|
|   | cancelled_percentage |
| ▶ | 33.40                |

## 6. Power BI Dashboard

**Tool Used:** Power BI

An interactive dashboard was created to visually present insights.

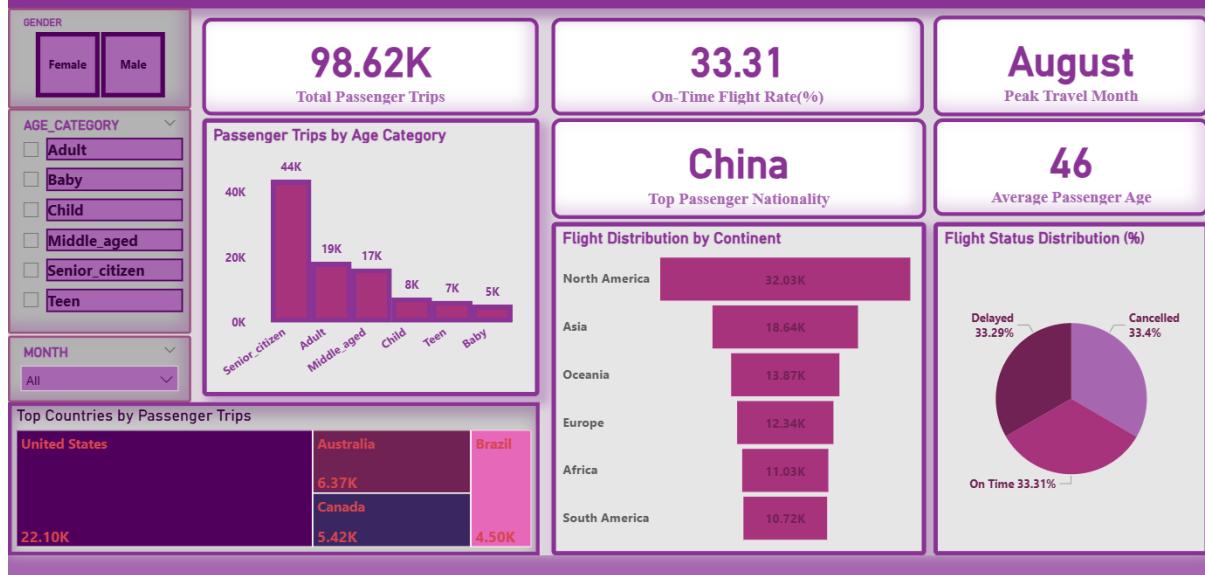
### KPI Cards:

- **Total Passenger Trips:** 98.62K
- **On-Time Flight Rate:** 33.31%
- **Peak Travel Month:** August
- **Top Passenger Nationality:** China
- **Average Passenger Age:** 46

### Visualizations Included:

- Passenger Trips by Age Category (Bar Chart)
- Flight Distribution by Continent (Funnel Chart)
- Top Countries by Passenger Trips (Treemap)
- Flight Status Distribution (Pie Chart)
- Interactive slicers for:
  - Gender
  - Age Category
  - Month

# Airline Passenger Travel & Flight Performance Dashboard



## 7. Key Insights & Findings

- **Senior citizens** represent the highest number of passenger trips.
- **August** is the peak travel month, indicating seasonal demand.
- **China** contributes the highest number of passengers.
- **North America** has the highest number of departing flights.
- Flight status distribution is nearly even across:
  - On-Time (~33%)
  - Delayed (~33%)
  - Cancelled (~33%)
- Average passenger age suggests a mature traveller demographic.

## **8. Business Impact**

This analysis can help airlines:

- Optimize staffing during peak months
- Improve punctuality strategies
- Target high-volume passenger demographics
- Plan route expansions based on continent and country demand
- Enhance customer experience by reducing delays and cancellations

## **9. Tools & Technologies Used**

| <b>Category</b> | <b>Tools</b> |
|-----------------|--------------|
|-----------------|--------------|

Data Cleaning Python (Pandas,Numpy)

Database MySQL

Visualization Power BI

## **10. Conclusion**

This project successfully demonstrates a **real-world data analytics pipeline**, covering:

- Raw data processing
- SQL-based analysis
- Insightful visual storytelling using Power BI