Exploring Insights From Synthetic Airline Data Analysis With Qlik

1. Introduction

1.1. Overview

In today's dynamic aviation industry, understanding the intricacies of airline operations is paramount for enhancing efficiency, optimizing services, and ensuring passenger satisfaction. Exploring Insights From Synthetic Airline Data Analysis With Qlik provides a comprehensive examination of diverse parameters within the airline domain. This analysis is facilitated by a rich dataset encompassing various facets of airline operations on a global scale.

The dataset, meticulously curated for comprehensive analysis, contains an array of crucial fields including Passenger ID, First Name, Last Name, Gender, Age, Nationality, Airport Name, Airport Country Code, Country Name, Airport Continent, Continents, Departure Date, Arrival Airport, Pilot Name, and Flight Status. Each of these columns contributes to a holistic understanding of passenger demographics, travel specifics, flight routes, crew details, and real-time flight statuses.

1.2. Purpose

With such a robust dataset at our disposal, researchers and industry experts can delve deep into the realm of aviation analytics, uncovering invaluable insights to drive informed decision-making. Through the power of Qlik's visualization capabilities, we embark on a journey of exploration, uncovering trends, patterns, and correlations that shed light on various aspects of airline operations.

By harnessing the analytical prowess of the Qlik platform, we not only present visually engaging representations of the data but also weave compelling narratives that elucidate key findings and implications. This synthesis of data visualization and storytelling serves as a powerful tool for stakeholders across the aviation spectrum, empowering them to optimize travel experiences, evaluate pilot performance, and enhance overall flight operations.

1.3. Technical Architecture
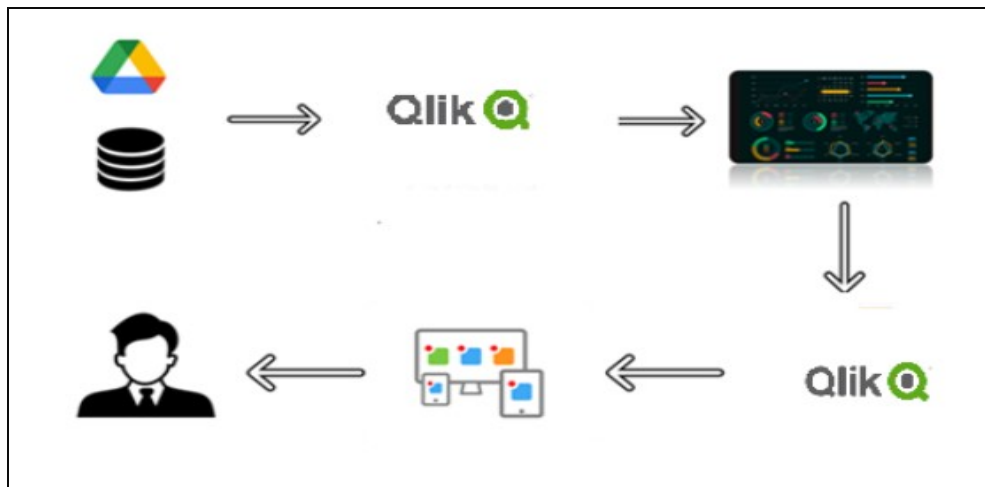
2. Problem Understanding

2.1. Specification of the Business Problem

The aviation industry operates within a complex ecosystem, where numerous factors influence operational efficiency, passenger satisfaction, and safety. However, gaining comprehensive insights into airline operations can be challenging due to the vast amount of data generated across various touchpoints. To address this challenge, there is a need for a robust data analysis solution that can provide actionable insights into passenger demographics, travel details, flight routes, crew information, and flight statuses.

2.2. Business Requirements

1. **Comprehensive Data Analysis:** The solution should be capable of analyzing diverse parameters within the airline domain, including passenger demographics, travel specifics, flight routes, crew details, and real-time flight statuses.
2. **Visualization Capabilities:** The solution should offer powerful visualization tools to present data insights in an intuitive and easily understandable manner.
3. **Storytelling Features:** It should enable the creation of compelling narratives that elucidate key findings and implications derived from the data analysis.
4. **Scalability:** The solution should be scalable to handle large volumes of data efficiently, ensuring smooth performance even as the dataset grows.
5. **Integration with Qlik Platform:** Given the familiarity and popularity of the Qlik platform among analysts and stakeholders, the solution should seamlessly integrate with Qlik to leverage its analytical capabilities.

2.3. Literature Survey

A literature survey reveals the significance of data-driven decision-making in the aviation industry. Various studies have highlighted the importance of leveraging data analytics to optimize airline operations, enhance passenger experiences, and improve overall efficiency. Additionally, research articles, whitepapers, and industry reports provide insights into the challenges faced by airlines and the potential benefits of adopting advanced analytics solutions.

Moreover, existing literature emphasizes the role of data visualization and storytelling in conveying complex insights to stakeholders effectively. Solutions like Qlik have been widely adopted in various industries for their ability to transform raw data into actionable insights through interactive visualizations and storytelling features.

By conducting a literature survey, we can gain a deeper understanding of the current state of data analytics in the aviation industry, identify best practices, and leverage existing knowledge to inform the development of our solution for exploring insights from synthetic airline data analysis with Qlik.

3. Data Collection

3.1. Collect the dataset

Airline data contains all the meta information regarding the columns described in the CSV files. Column Description of the Dataset:
- **Passenger ID -** Unique identifier for each passenger
- **First Name -** First name of the passenger
- **Last Name -** Last name of the passenger
- **Gender -** Gender of the passenger
- **Age -** Age of the passenger
- **Nationality -** Nationality of the passenger
- **Airport Name -** Name of the airport where the passenger boarded
- **Airport Country Code -** Country code of the airport's location
- **Country Name -** Name of the country the airport is located in
- **Airport Continent -** Continent where the airport is situated
- **Continents -** Continents involved in the flight route
- **Departure Date -** Date when the flight departed
- **Arrival Airport -** Destination airport of the flight
- **Pilot Name -** Name of the pilot operating the flight
- **Flight Status -** Current status of the flight (e.g., on-time, delayed, canceled)

3.2. Connect Data with Qlik Sense

After obtaining the dataset, the next step is to connect it with the Qlik Sense platform for analysis and visualization. Qlik Sense offers various data connectivity options, allowing users to seamlessly integrate data from different sources into their analytical applications. The process of connecting data with Qlik Sense typically involves the following steps:

1. **Data Loading:** Import the dataset into Qlik Sense using one of the supported data loading methods such as loading from files (e.g., Excel, CSV), connecting to databases (e.g., SQL Server, Oracle), or using REST APIs.
2. **Data Modeling:** Define the structure of the data model by creating tables and establishing relationships between them based on key fields (e.g., Passenger ID, Airport Code).
3. **Data Transformation:** Perform necessary data transformations and manipulations to prepare the dataset for analysis, such as aggregations, calculations, and data cleansing.
4. **Data Visualization:** Create interactive visualizations and dashboards using Qlik Sense's intuitive drag-and-drop interface, leveraging the power of associative data modeling to explore relationships and uncover insights within the dataset.

4. Data Preparation

4.1. Prepare the Data for Visualization

1. **Clean and Transform Data:** Handle missing values, correct data types, remove duplicates, extract date features, standardize text fields, and create new fields.
2. **Engineer and Enrich Data:** Create features like passenger age group, add geographical info, categorize flight statuses (e.g., on-time, delayed, canceled), and others.
3. **Summarize and Export:** Summarize demographics, airport trends and flight trends, then export the cleaned dataset for use in Qlik for visualizations and analysis.

5. Data Visualizations
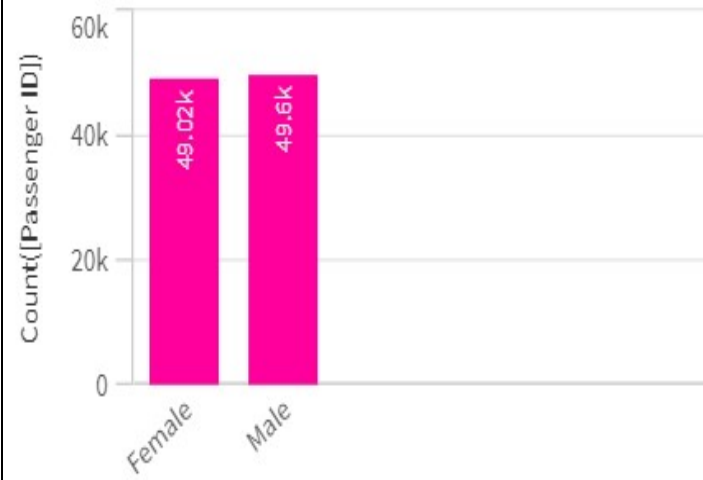
5.1. Visualizations

Total number of visualizations: 27

1. Gender Distribution among Passengers (Visualization: Bar chart)
2. Nationality of Passengers (Visualization: Pie Chart)
3. Total Passenger Count (Visualization: KPI)
4. Age Distribution of Passengers (Visualization: Histogram)
5. Flight Status of Passengers (Visualization: Line Chart)
6. Arrival Airports in the World (Visualization: Map)
7. Passenger Count by Continent (Visualization: TreeMap)
8. Flight Status (Visualization: Filter Pane)
9. Passenger Count by Airport Country Code (Visualization: Horizontal ComboChart)
10. Number of Arrival Airports (Visualization: Gauge)
11. Passenger Arrivals by Country (Visualization: Mekko Chart)
12. Flight Status Distribution (Visualization: Bar Chart)
13. Flight Status of Passengers by Airport Country Code (Visualization: Mekko Chart)
14. Total Flight Status Count (Visualization: KPI)
15. Flight Status Over Time (Visualization: Line Chart)
16. Monthly Flight Status (Visualization: Bar Chart)
17. On-Time Flight Status by Nationality (Visualization: Bar Chart)
18. Arrival airport by Passengers (Visualization: Bar Chart) (Filter: Country Name(India))
19. count of Passengers (Visualization: KPI) (Filter: Country Name(India))
20. Average age of Indian people who are boarding the Flight (Visualization: Guage) (Filter: Country Name(India))
21. Passenger by Arrival airport (Visualization: Donut Chart) (Filter: Country Name(India))
22. Count of Passengers, Age, Gender, Flight Status (Visualization: Waterfall Chart) ((Filter: Country Name(India))
23. The Gender of Indian Passengers (Visualization: Treemap) (Filter: Country Name(India))
24. Number of Flights by Pilot (Visualization: Bar Chart)
25. Flight Status by Pilot (Visualization: Bar Chart)
26. Nationality and Gender Distribution (Visualization: Distribution plot)
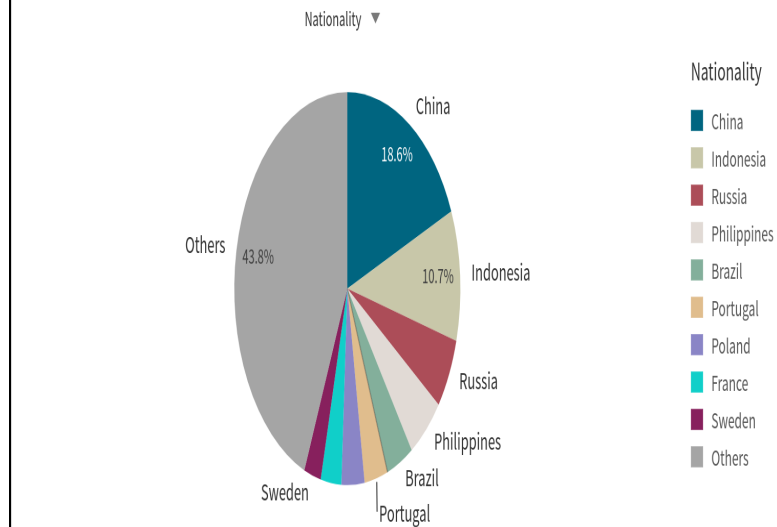27. Age Group and Flight Status Correlation (Visualization: Treemap

Some of the visualizations are:

1. Passenger Demographics

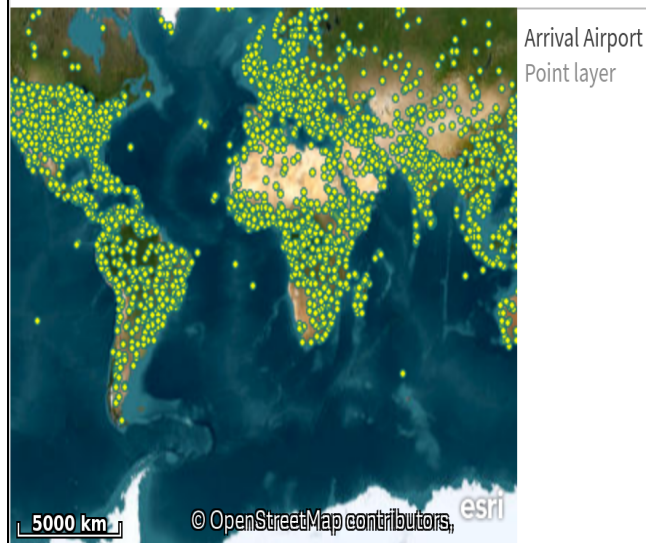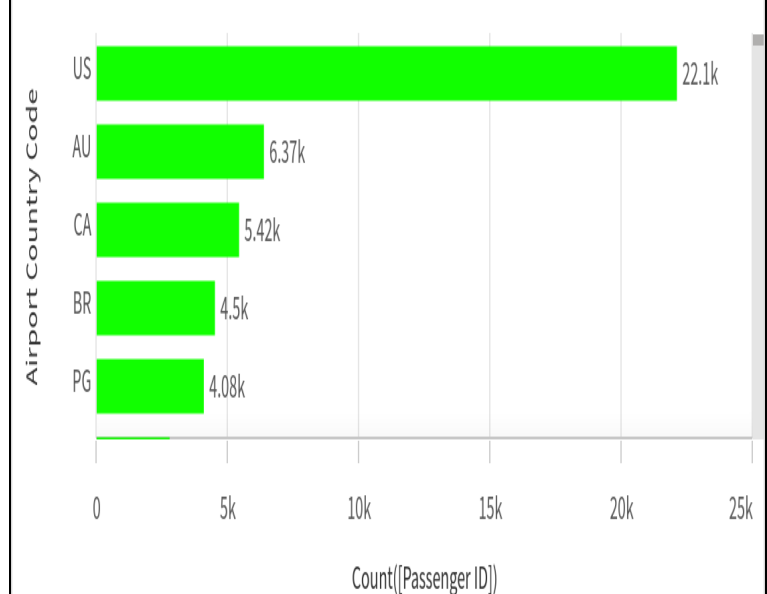## Gender Distribution among the Passengers.

Female: 49.02k
Male: 49.6k

(Count([Passenger ID]))

Nationality of the Passengers.

Nationality ▼

- China 18.6%
- Indonesia 10.7%
- Russia
- Philippines
- Brazil
- Portugal
- Sweden
- Others 43.8%

Nationality
- China
- Indonesia
- Russia
- Philippines
- Brazil
- Portugal
- Poland
- France
- Sweden
- Others

2. Flights and Airports

Arrival Airports in the World. *

Arrival Airport
Point layer

5000 km
© OpenStreetMap contributors,

esri

* Currently showing a limited data set.

Passenger Count by Airport Country Code.

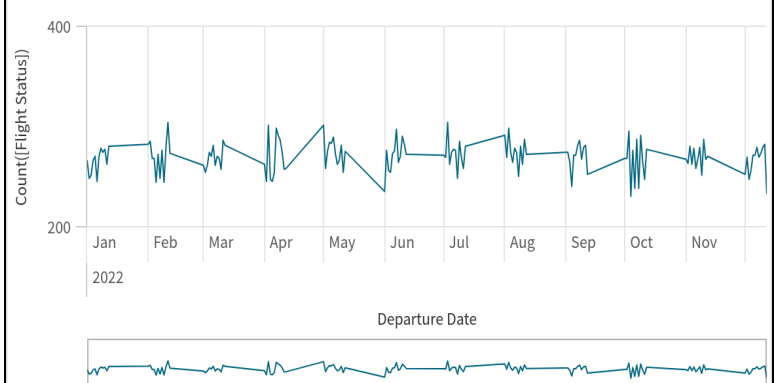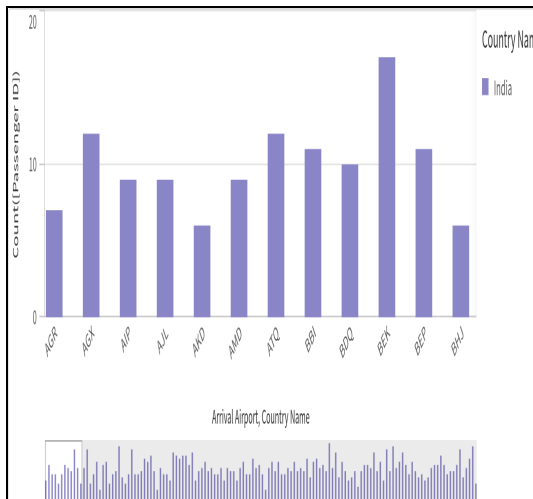| Airport Country Code	Count([Passenger ID])
US	22.1k
AU	6.37k
CA	5.42k
BR	4.5k
PG	4.08k

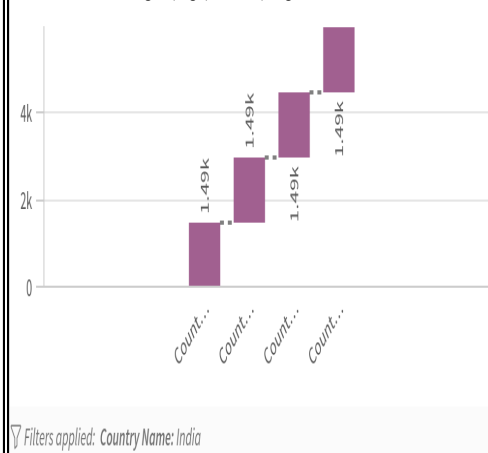## 3. Flight Status Analysis

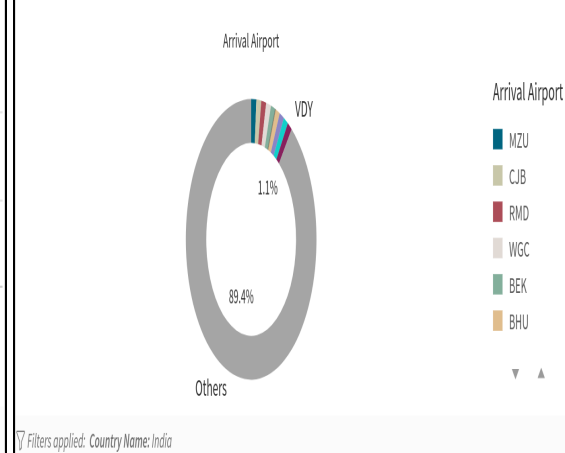**Flight Status Distribution.**
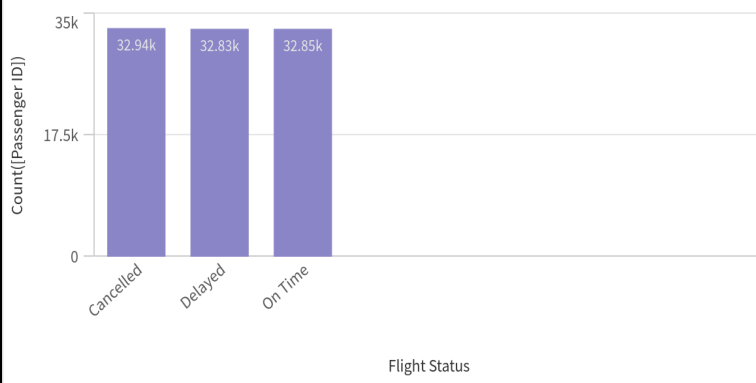
Flight Status Over Time.

4. Visualizations for India

The count of Passengers, Age, Gender, Flight Status.
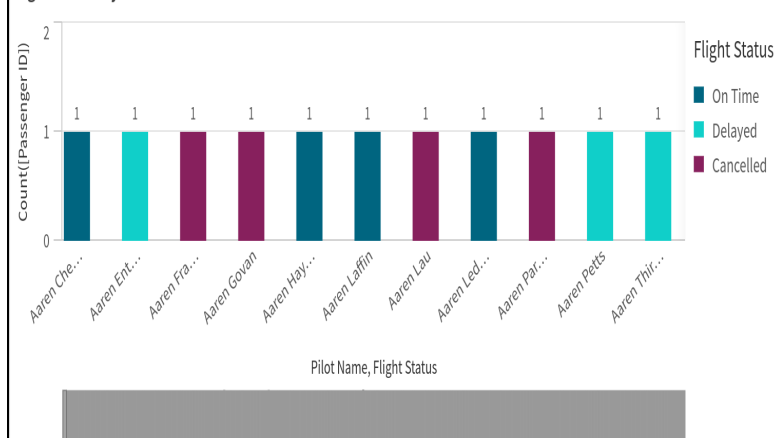

Filters applied: **Country Name:** India

Passengers by Arrival Airport.
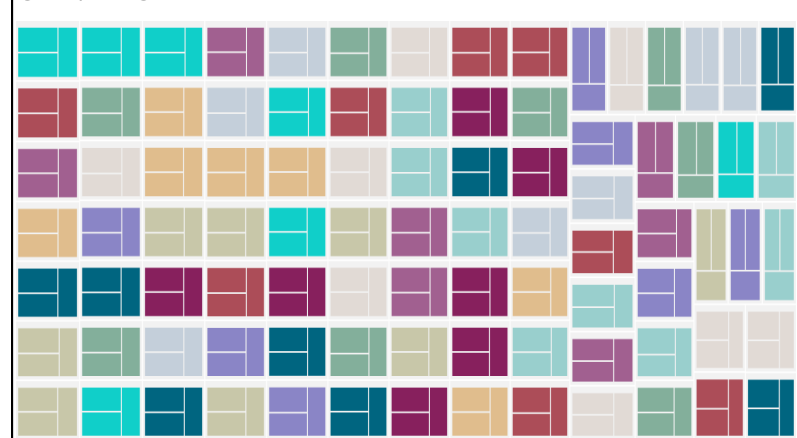

Filters applied: **Country Name:** India

5. Additional Insights
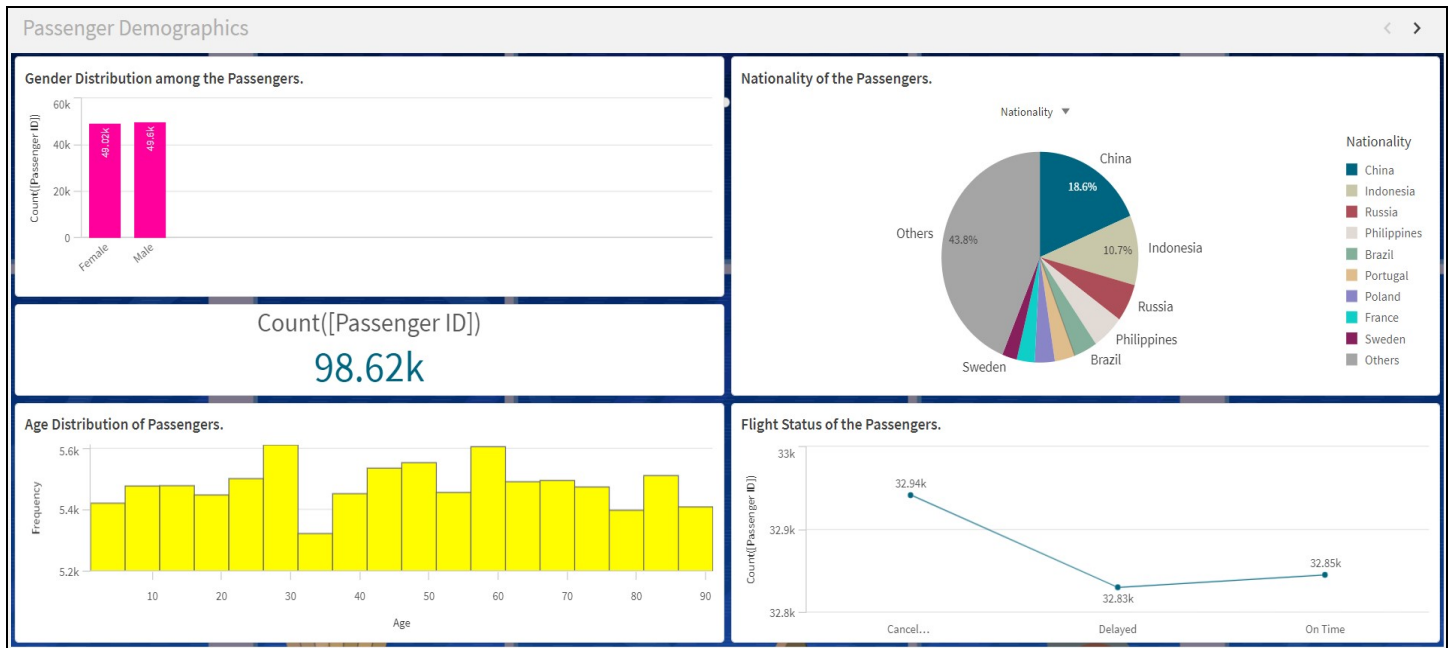
Flight Status by Pilot.

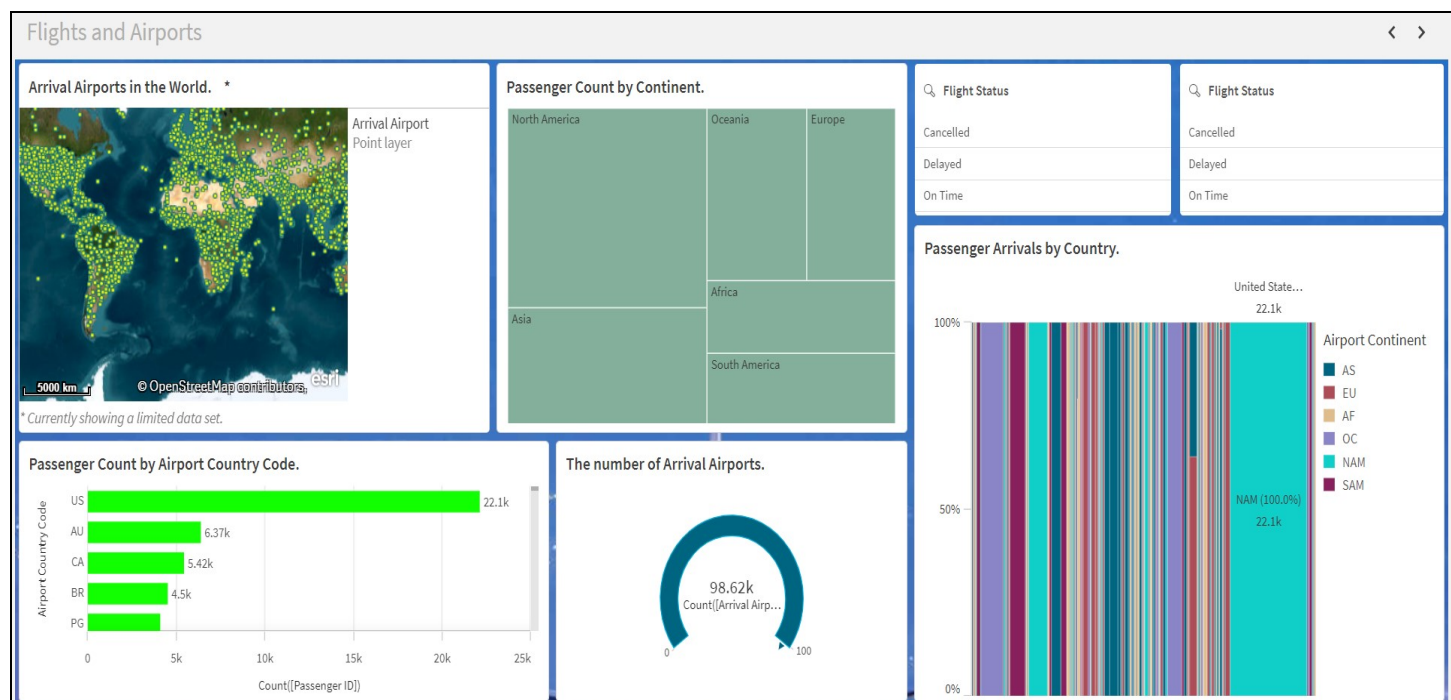**Age Group and Flight Status Correlation.**

6. Dashboard

6.1. Responsive and Design of Dashboard

Dashboard 1:

Dashboard 2:

Dashboard 3:

Flight Status Analysis

Flight Status Distribution.

Flight Status of Passengers by Airport Country Code.

Flight Status Over Time.

Count([Flight Status])
98.62k

Monthly Flight Status.

On-Time Flight Status by Nationality.

Dashboard 4:

Visualizations for India

Count([Passenger ID])
1.49k

Average Age of Indian People Who are Boarding the Flight.

45.13
Avg(Age)

Passengers by Arrival Airport.

The count of Passengers, Age, Gender, Flight Status.

The Gender of the Passengers.

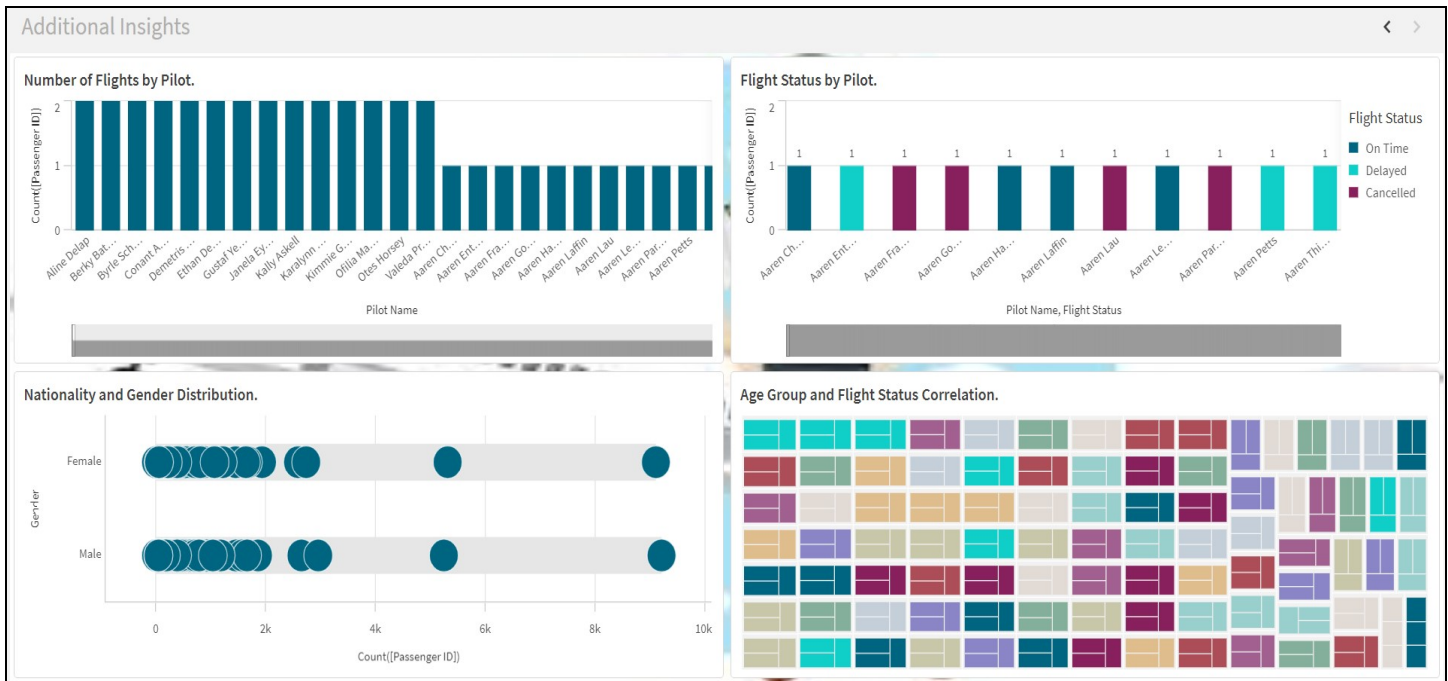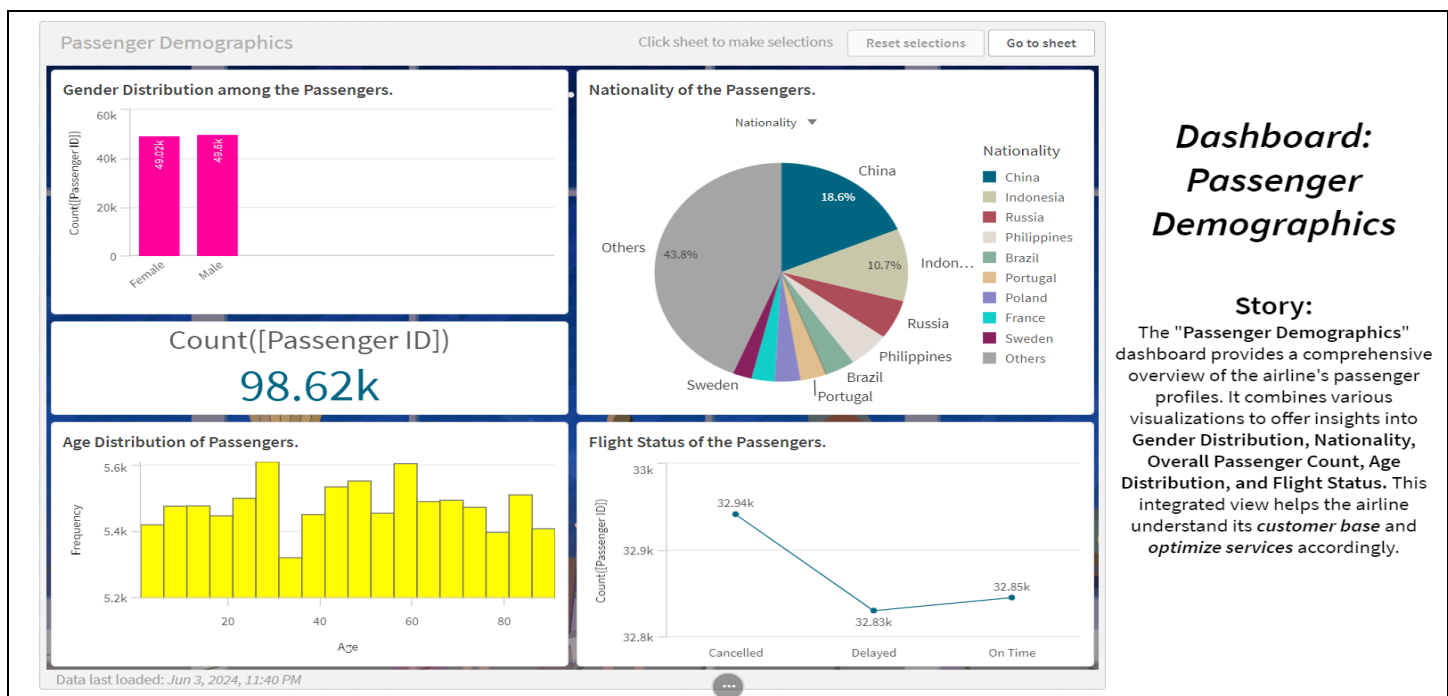**Dashboard 5:**

7. Report/Story

7.1. Report Creation

Story Telling 1:

1. Gender Distribution Among Passengers (*Visualization: Bar Chart*)

X-Axis: Gender (Male, Female)
Y-Axis: Count of Passenger IDs
Story: This bar chart shows the number of male and female passengers, revealing the gender composition. Here, there are 49.02k male and 49.6k female passengers, males make up 49.71% and females 50.29%. This insight can inform the airline's marketing and service strategies.

2. Nationality of Passengers (*Visualization: Pie Chart*)

Slices: Nationality
Size: Count of First Names
Story: The pie chart illustrates the distribution of passengers by nationality, highlighting the most common nationalities. For instance, here Chinese passengers account for 18.6%, Indonesians 10.7%, and Russians 5.8%, this helps the airline understand where most of its passengers come from and potentially cater to the preferences of these groups.

3. Total Passenger Count (*Visualization: KPI*)
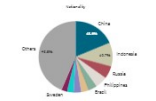
Count([Passenger ID])
98.62k

Metric: Count of Passenger IDs
Story: The KPI displays the total number of passengers, providing a quick snapshot of the airline's customer base size. Here, it shows "Total Passengers: 89.62k," which immediately conveys the operations' scale and helps assess growth or decline over time.

4. Age Distribution of Passengers (*Visualization: Histogram*)
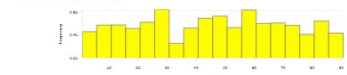
X-Axis: Age
Y-Axis: Frequency (Count of Passenger IDs)
Story: The histogram presents the age distribution of passengers, highlighting which age groups are most prevalent. For instance, the majority of passengers are aged 26-31, the airline can tailor its services to suit the needs and preferences of this age group.

5. Flight Status of Passengers (*Visualization: Line Chart*)

X-Axis: Flight Status (e.g., On Time, Delayed, Cancelled)
Y-Axis: Count of Passenger IDs
Story: The line chart tracks the number of passengers by flight status over time, showing trends and peaks. For example, there is a peak in the number of canceled flights, and the airline can investigate and address potential causes for these cancellations to improve service punctuality.
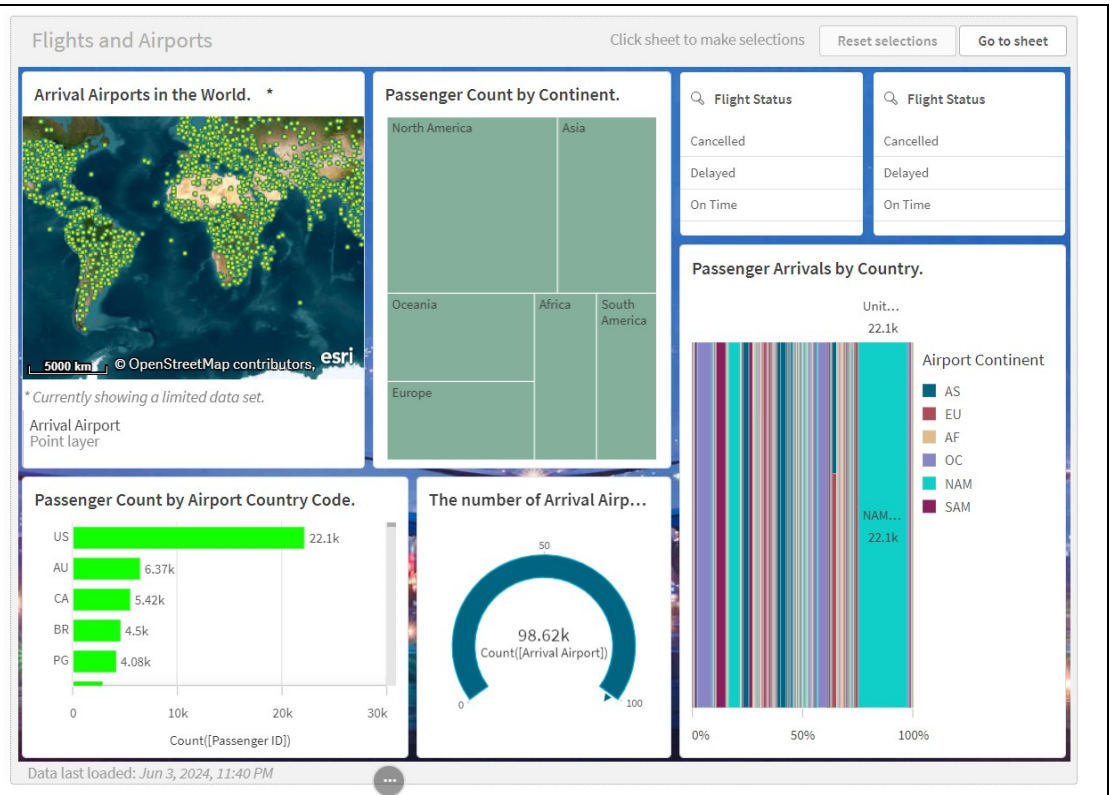
Story Telling 2:

Dashboard: Flights and Airports

Story:
The "Flights and Airports" dashboard provides a detailed analysis of **Arrival Airports, Passenger Distribution by Continent and Country, Flight Status, and Passenger Counts.** This narrative helps stakeholders understand *flight patterns, passenger demographics, and operational efficiency* across different regions.

Flights and Airports

Click sheet to make selections | Reset selections | Go to sheet

Arrival Airports in the World. *
* Currently showing a limited data set.
Arrival Airport Point layer
5000 km | © OpenStreetMap contributors, esri

Passenger Count by Continent.
North America | Asia
Oceania | Africa | South America
Europe

Flight Status
Cancelled
Delayed
On Time

Flight Status
Cancelled
Delayed
On Time

Passenger Arrivals by Country.
Unit...
22.1k

Airport Continent
- AS
- EU
- AF
- OC
- NAM
- SAM

NAM...
22.1k

Passenger Count by Airport Country Code.
US — 22.1k
AU — 6.37k
CA — 5.42k
BR — 4.5k
PG — 4.08k
0 | 10k | 20k | 30k
Count([Passenger ID])

The number of Arrival Airp...
50
98.62k
Count([Arrival Airport])
0 | 100

0% | 50% | 100%

1. Arrival Airports in the World (*Visualization: Map*)

Data Points: Arrival Airports
Story: This map visualization shows the locations of arrival airports worldwide. By visualizing arrival airports on a global map, we can see the geographic spread and identify key hubs. For instance, major cities with high passenger traffic are highlighted, indicating important destinations for the airline.

2. Passenger Count by Continent (*Visualization: TreeMap*)

Dimensions: Continents
Measure: Count of Passenger IDs
Story: The treemap illustrates the distribution of passengers by continent. This visualization shows which continents have the highest passenger counts. For example, here the largest section represents North America which indicates that North America has the highest number of passengers, guiding regional marketing and operational strategies.

3. Flight Status (*Visualization: Filter Pane*) ✔

Dimensions: Flight Status (e.g., On Time, Delayed, Cancelled)
Story: The filter pane allows users to filter the dashboard by different flight statuses. Users can dynamically explore data based on the current status of flights, providing a real-time understanding of operational conditions. For example, filtering to show only "Delayed" flights can help identify patterns and areas needing improvement.

4. Passenger Count by Airport Country Code (*Visualization: Horizontal ComboChart*)

X-Axis: Count of Passenger IDs
Y-Axis: Airport Country Code
Story: The combo chart shows the number of passengers for each airport country code. This visualization highlights which countries have the highest passenger counts. For example, "US" has the highest bar which indicates that the United States have the most passengers, helping to prioritize resources and marketing efforts in that country.

5. Number of Arrival Airports (*Visualization: Gauge*)

Measure: Count of Arrival Airports
Story: The gauge shows the total number of arrival airports served by the airline. This provides a quick overview of the airline's reach. For instance, a gauge showing 150 arrival airports indicates the airline's extensive network and helps in assessing network expansion goals.

6. Passenger Arrivals by Country (*Visualization: Mekko Chart*)

Dimensions: Country Name, **Cells** - Airport Continent
Measure: Count of Passenger IDs
Story: The Mekko chart illustrates passenger arrivals by country and continent.Insight: This chart provides a detailed view of passenger distribution, showing both the volume of passengers and their regional distribution. For example, if the chart shows a large block for "USA" in "North America", it highlights the significant passenger traffic to the United States, guiding strategic decisions.

Story Telling 3:

Dashboard: Flight Status Analysis

Story:

The **"Flight Status Analysis"** dashboard provides a detailed examination of the **flight statuses across different dimensions** such as distribution by status, airport country code, and nationality. It also tracks the **flight statuses over time and by month.** This analysis helps stakeholders understand the *performance and reliability of flights, identifying patterns and areas for improvement.*
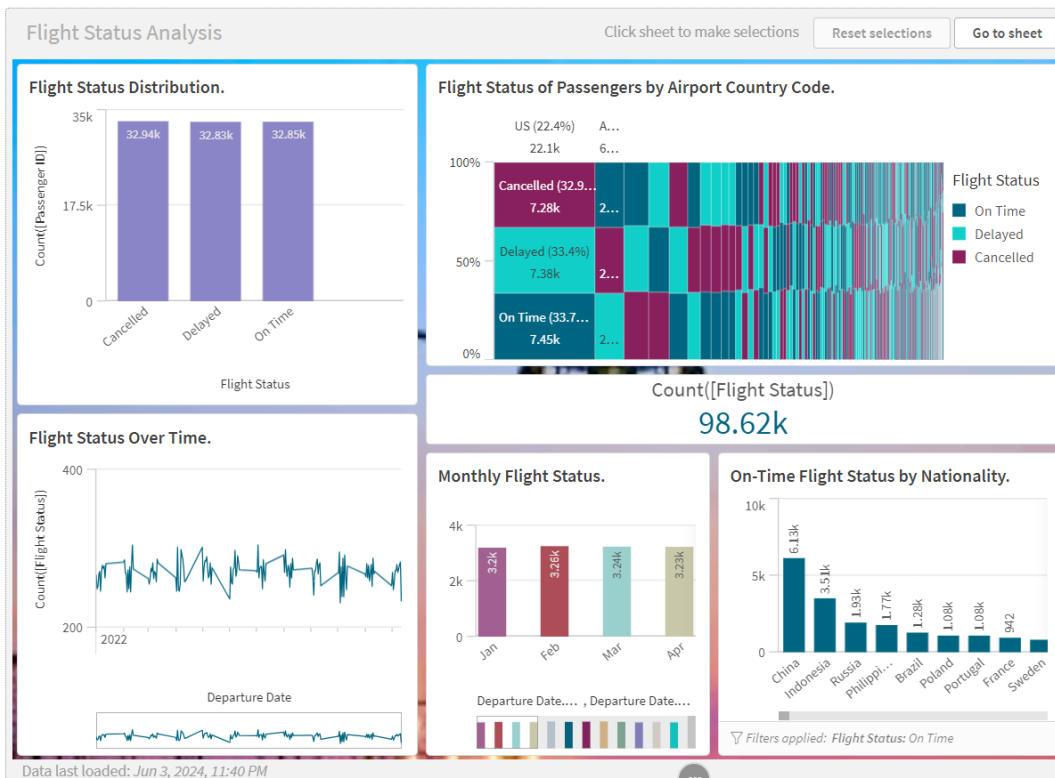
1. Flight Status Distribution (*Visualization: Bar Chart*)

X-Axis: Flight Status (e.g., On Time, Delayed, Cancelled)
Y-Axis: Count of Passenger IDs
Story: This bar chart displays the distribution of flight statuses, highlighting the *count of passengers affected by each status.* Here the *"Cancelled"* bar is significantly *higher* than *"Delayed"* and *"On Time"* which indicates that the majority of flights aren't punctual, indicating the airline's to *increase their reliability.*

2. Flight Status of Passengers by Airport Country Code (*Visualization: Mekko Chart*)

Dimensions: Airport Country Code, Flight Status
Measure: Count of Passenger IDs
Story: The Mekko chart illustrates the distribution of flight statuses across different airport country codes. This visualization *shows which countries experience more delays or cancellations.* For instance, flights from the *US* have a *higher proportion of delays* which indicates *potential operational issues* in that region, *prompting further investigation.*

3. Total Flight Status Count (*Visualization: KPI*)

Measure: Count of Flight Statuses
Story: The KPI displays the *total number of recorded flight statuses.* For example, showing *"Total Flight Status Records: 98.62k"* provides a quick snapshot of the *dataset size*, indicating the *volume of flight data being analyzed.*
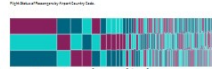
4. Flight Status Over Time (*Visualization: Line Chart*)

X-Axis: Departure Date
Y-Axis: Count of Flight Statuses
Story: This line chart tracks the *number of flights by status over time, showing trends and patterns.* If the chart reveals an *increasing* trend in *delayed flights* during *certain months*, it can indicate *seasonal challenges or operational inefficiencies* that need addressing.

5. Monthly Flight Status (*Visualization: Bar Chart*)

X-Axis: Month of Departure Date
Y-Axis: Count of Flight Statuses
Story: The bar chart shows the *monthly distribution of flight statuses.* Here, *April* shows a *higher* count of *"Delayed"* statuses, it could be due to *weather conditions* affecting flight schedules, highlighting the *need for better contingency planning* during these months.

6. On-Time Flight Status by Nationality (*Visualization: Bar Chart*)

X-Axis: Nationality
Y-Axis: Count of Flight Statuses
Filter: Flight Status (On Time)
Story: This bar chart displays the *count of on-time flights for passengers of different nationalities.* Here, *"Chinese"* passengers have the *highest* count of *on-time flights* which indicates *reliable service* for this demographic, informing *marketing strategies* and *customer service enhancements* for other nationalities.
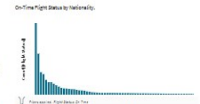
8. Performance Testing

8.1. Amount of Data Rendered

Amount of Data Rendered refers to the volume of data that is processed and displayed within a visualization or dashboard in Qlik Sense. This is crucial for performance optimization, as large datasets can slow down the rendering and responsiveness of visualizations.

Airline Dataset Updated - v2

Passenger ID

First Name

Last Name

Gender

Age

Nationality

Airport Name

Airport Country Code

Country Name

Airport Continent

Continents

Departure Date

Arrival Airport

Pilot Name

Flight Status

Airline Dataset Updated - v2.@6_GeoInfo

Airline Dataset Updated - v2.@9_GeoInfo

Airline Dataset Updated - v2.Airport Country Code_GeoInfo

8.2. Utilization of Data Filters

For the airline dataset analysis in Qlik Sense, I have implemented a filter on the Country Name field to focus exclusively on passengers from India. The dashboard 4 created includes several key visualizations to provide comprehensive insights. Firstly, it displays the count of Passenger IDs by Arrival Airport, offering a clear view of the most frequented destinations for Indian travelers. Additionally, the dashboard includes a metric for the total count of Passenger IDs, reflecting the overall volume of Indian passengers in the dataset.

Another crucial visualization presents the average age of Indian passengers boarding flights, providing demographic insights into the age distribution. Moreover, the dashboard features a breakdown of passengers by Arrival Airport, further segmented by count, age, gender, and flight status. This enables detailed analysis of passenger demographics and

flight details. The gender of passengers is also highlighted, showcasing the distribution between male and female travelers.

Throughout all visualizations, the filter for Country Name (India) is prominently applied, ensuring that the data presented is specific to Indian passengers and allowing for targeted analysis and insights.