

Introduction

1.1 Overview

The project "Exploring Insights from Synthetic Airline Data Analysis with Qlik" focuses on utilizing Qlik, a prominent business intelligence (BI) and data visualization platform, to analyze synthetic airline data. This dataset mirrors real-world airline operations, including flight schedules, passenger demographics, ticket sales, and performance metrics. By employing Qlik's sophisticated analytical tools, the project aims to uncover significant patterns, trends, and relationships within the data, thereby providing actionable insights that can inform and enhance decision-making for airlines, airports, and other related stakeholders.

1.2 Purpose

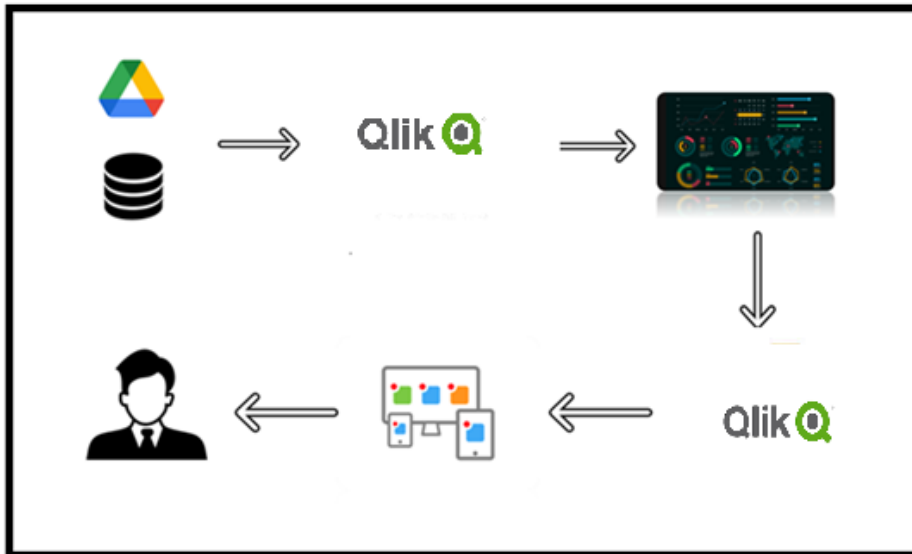
The primary purpose of this project is to utilize Qlik's advanced data visualization and analytics features to dissect and interpret synthetic airline data. The synthetic nature of the data allows for comprehensive analysis without compromising real-world privacy concerns, making it an ideal candidate for testing and demonstrating Qlik's capabilities. This analysis can serve multiple purposes:

- **Operational Efficiency:** Identifying inefficiencies and potential areas for improvement in airline operations.
- **Customer Insights:** Understanding passenger demographics and preferences to enhance customer satisfaction and tailor services.
- **Sales and Revenue Analysis:** Analyzing ticket sales patterns to optimize pricing strategies and maximize revenue.
- **Performance Metrics:** Monitoring key performance indicators (KPIs) to ensure adherence to industry standards and improve overall performance.

From this project, the following key outcomes are anticipated:

- **Trend Analysis:** Identifying trends in flight schedules, passenger numbers, and ticket sales over time.
- **Correlation Insights:** Uncovering correlations between various data points, such as the relationship between ticket prices and sales volumes or the impact of flight delays on customer satisfaction.
- **Data-Driven Decisions:** Equipping airline management with actionable insights to make informed decisions that enhance operational efficiency, customer experience, and profitability.
- **Visualization of Complex Data:** Creating intuitive and interactive visual representations of complex data sets to facilitate easy understanding and analysis.

1.3 Technical Architecture.



Problem Understanding

2.1 Specify the business problem

The airline industry operates in a highly competitive and dynamic environment, where efficiency, customer satisfaction, and profitability are critical for success. Airlines need to optimize their operations continually, understand customer preferences, and enhance their revenue management strategies to stay ahead. However, gaining these insights from complex and vast datasets can be challenging.

Key Business Problems:

- **Operational Efficiency:** Identifying and mitigating inefficiencies in flight operations, scheduling, and resource allocation.
- **Customer Satisfaction:** Understanding passenger demographics and preferences to improve service offerings and enhance the customer experience.
- **Revenue Optimization:** Analyzing ticket sales data to develop effective pricing strategies and maximize revenue.
- **Performance Monitoring:** Keeping track of key performance indicators (KPIs) to ensure standards are met and performance is continuously improved.

2.2 Business Requirements

To address these business problems, the project requires a comprehensive set of analytical capabilities provided by Qlik.

The following are the specific business requirements for the project:

- **Data Integration:** The ability to import and integrate various sources of synthetic airline data into Qlik.
- **Interactive Dashboards:** Development of user-friendly and interactive dashboards that visualize key aspects of the data.
- **Trend and Pattern Analysis:** Tools to identify trends and patterns in the data over time.
- **Correlation and Causation Analysis:** Capabilities to uncover relationships between different data variables, such as ticket prices and sales volumes.
- **Custom Reporting:** The ability to generate customized reports summarizing the analysis and findings.
- **Real-time Insights:** Providing real-time insights and updates as new data is integrated.

2.3 Literature Survey

To ensure a robust approach to the project, a review of existing literature and prior research in the fields of airline data analytics and business intelligence tools was conducted. The following key points summarize the findings from the literature survey:

- **Importance of Data Analytics in Airlines :** Studies highlight the critical role of data analytics in enhancing operational efficiency, customer satisfaction, and revenue management within the airline industry. Data-driven decision-making has been shown to provide a competitive edge by enabling more accurate forecasting and strategic planning.
- **Use of BI Tools in Airline Industry :** Business intelligence tools like Qlik are widely used in the airline industry for their ability to handle large datasets and provide intuitive visualizations. Research indicates that BI tools can significantly reduce the time and effort required to derive insights from complex data.
- **Synthetic Data in Analytics :** Synthetic data is increasingly used for analysis in situations where real-world data is sensitive or unavailable. Studies suggest that synthetic data, while fictional, can closely mimic real-world scenarios, making it valuable for testing and demonstrating analytical capabilities.
- **Challenges and Best Practices :** Common challenges in airline data analytics include data integration from disparate sources, ensuring data quality, and managing data privacy. Best practices identified include the use of automated data cleaning processes, maintaining data integrity, and ensuring that analytical tools are user-friendly and accessible to various stakeholders.

By reviewing existing literature, the project benefits from a foundation of established knowledge and best practices, ensuring that the analysis conducted using Qlik is both relevant and effective in addressing the identified business problems.

Data Collection

3.1 Collect the dataset

Data collection is the process of gathering and measuring information on variables of interest, in an established systematic fashion that enables one to answer stated research questions, test hypotheses, evaluate outcomes and generate insights from the data.

The dataset - Airline Dataset-Navigating the Skies: Exploring Insights from Synthetic Airline Data has been downloaded from Kaggle

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

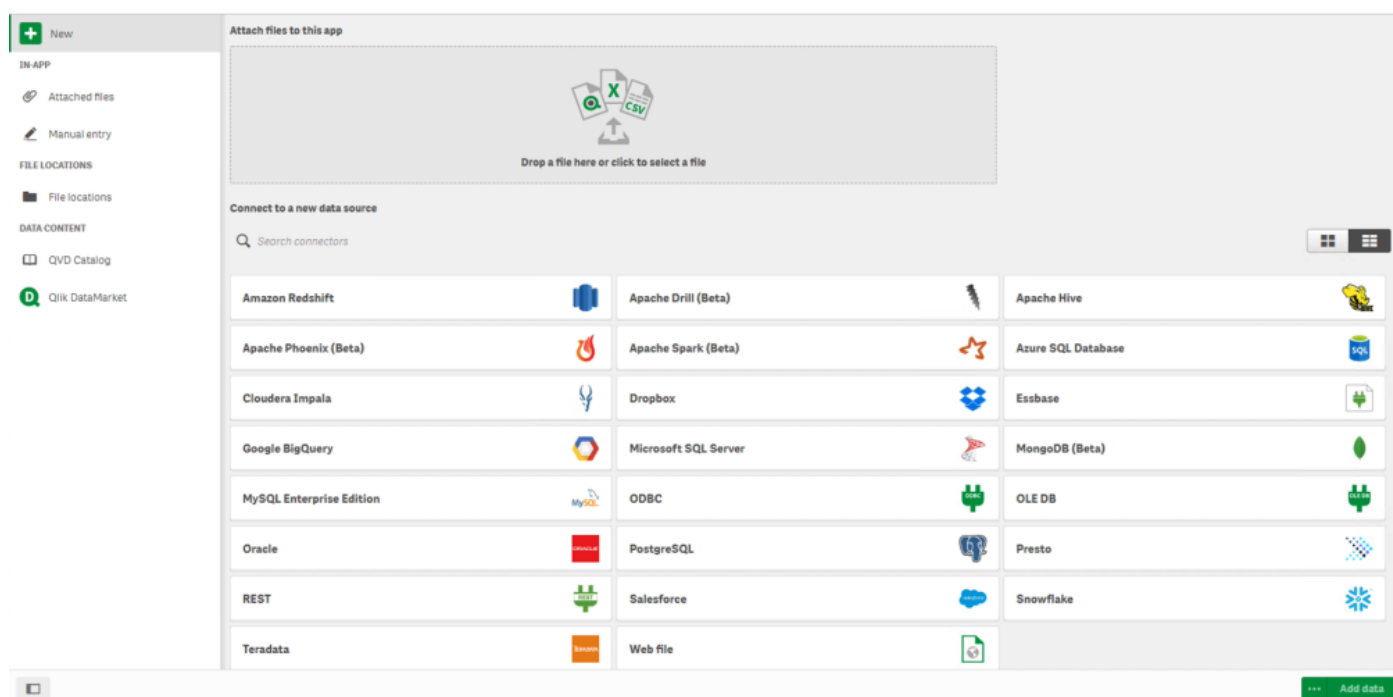
- Qlik Sense lets you to connect to your data, wherever it is stored, with a wide range of Qlik Connectors and other data connection types.
- When you create a data connection it is saved, so you can quickly select and load data from the data sources that you commonly use. Connect to databases, social media data, local files, remote files, and web files.

To select data from a data source, you can either create a new data connection or use a saved data connection. You can create data connections and access your saved connections from:

- **Add data** in the data manager. Add new data to your app quickly and get assistance creating associations.
- **Data connections** in the data load editor. Select data from a new or existing data connection, or use the script editor to load data from a data connection. You can also edit existing data connections.

Adding data to the app

- You can add data to your app quickly. Open the **Data manager** and then click **New**. You can also click **Add data**. You are also prompted to add data when you create a new app. When you are editing a sheet, you can also click **Add** in the **Fields** panel to add data.



Data Preparation

4.1 Prepare the data for visualization

Preparing the data for visualization involves cleaning the data to remove irrelevant or missing data, transforming the data into a format that can be easily visualized, exploring the data to identify patterns and trends, filtering the data to focus on specific subsets of data, preparing the data for visualization software, and ensuring the data is accurate and complete. This process helps to make the data easily understandable and ready for creating visualizations to gain insights into performance and efficiency. Since the data is already cleaned, we can move to visualization.

There are two steps in data preparation :

1. Data Loading:

Steps for Data Loading:

- Data Source Identification:** Identify and gather the necessary data files, which may include CSV files, Excel spreadsheets, or databases.
- Connecting to Data Sources:** Use Qlik's data connection features to link to the data

sources. This might involve setting up connections to local files or cloud storage services.

- Data Import: Import the data into Qlik's environment. This can be done through Qlik's data load editor, which allows for scripting and automation of data import processes.
- Data Validation: Ensure that the data is correctly imported by checking for completeness and consistency. Validate that all necessary fields and records are present and accurately represented.

2.Data Cleaning and Pre-processing:

Once the data is loaded into Qlik, it needs to be preprocessed to ensure it is clean, consistent, and ready for analysis.

In Qlik, the preprocessing steps can be performed using a combination of its data load editor and data manager features. The data load editor allows for scripting and automation of data preprocessing tasks, while the data manager provides a more visual and interactive interface for data preparation.

Add this code in the data load editor to create a field called AgeGroup to divide the passengers based on their age:

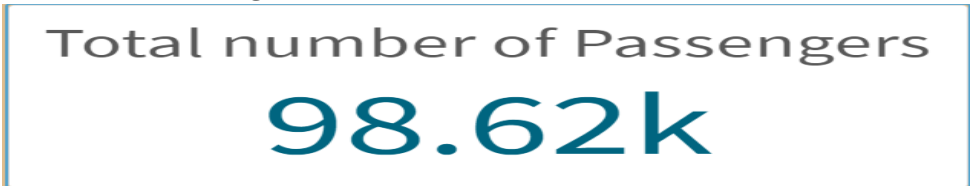
```
[Airline_Dataset]:  
Load *;  
[Airline_Dataset]:  
NoConcatenate Load *,  
If(Age >= 0 and Age <=1 , 'Baby',  
    if(Age >=1 and Age<=3, 'Toddler',  
    if(Age >=4 and Age<=9, 'Child',  
    if(Age >=10 and Age<=12, 'Tween',  
    if(Age >=13 and Age<=19, 'Teen',  
    if(Age >=20 and Age<=24, 'Young Adult',  
    if(Age >=25 and Age<=39, 'Adult',  
    if(Age >=40 and Age<=54, 'Middle',  
    if(Age >=55 and Age<=79, 'Elder',  
    if(Age >=80 , 'Old'))))))) as AgeGroup
```

By meticulously loading and preprocessing the data, we ensure a solid foundation for subsequent analysis, enabling accurate and insightful exploration of the synthetic airline data using Qlik.

Data Visualizations

5.1 Synthetic Airlines Data Analysis Visualisations:

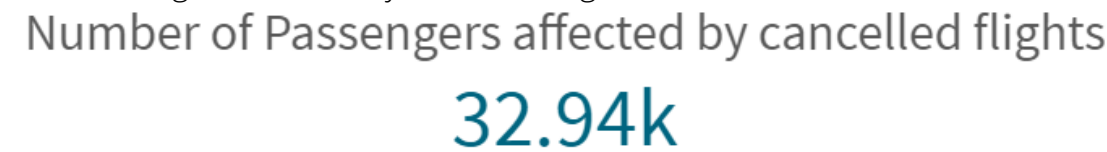
- Total Number of Passengers:



- Number of Flights on time:



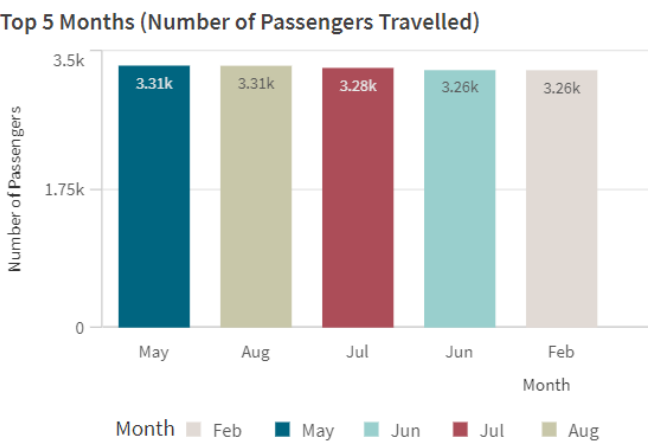
- Number of Passengers affected by Cancelled flights:



- Number of Passengers affected by Delayed flights:

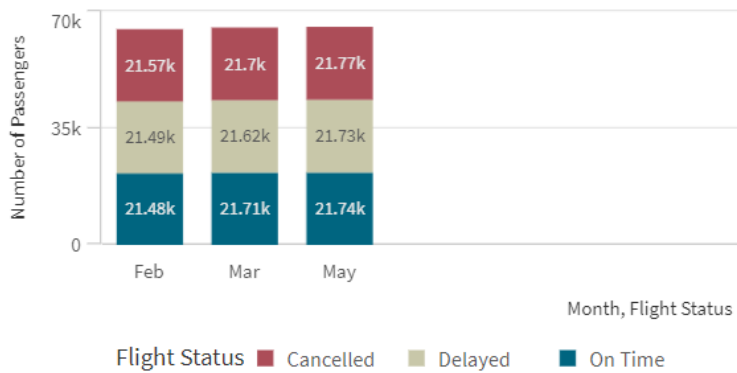


- Top 5 months with highest number of passengers:

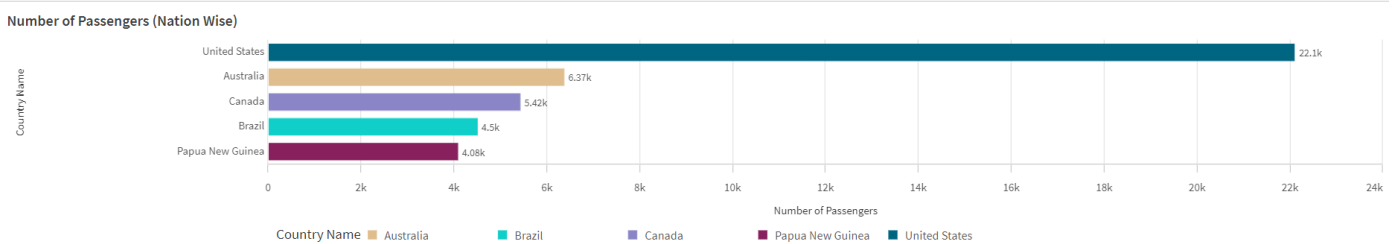


- Top 3 Months - Flight status Wise (Delayed/Cancelled/On-Time):

Top 3 Months (Flight Status Wise)

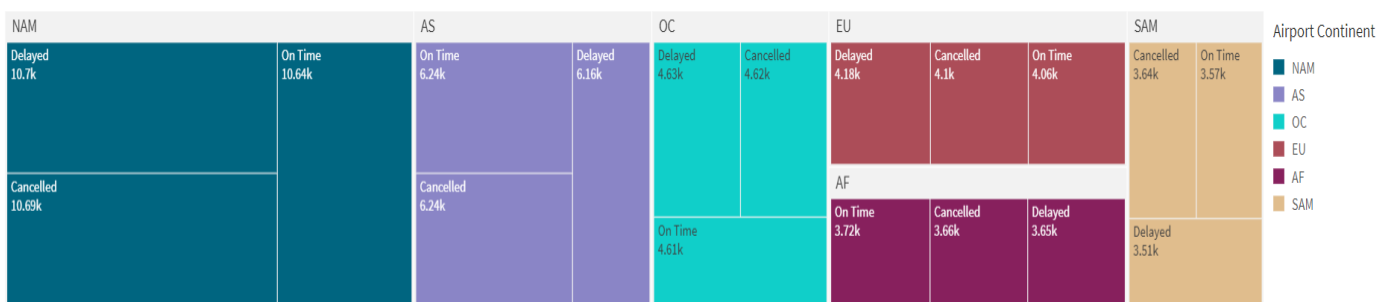


- Number of Passengers Nationality Wise:



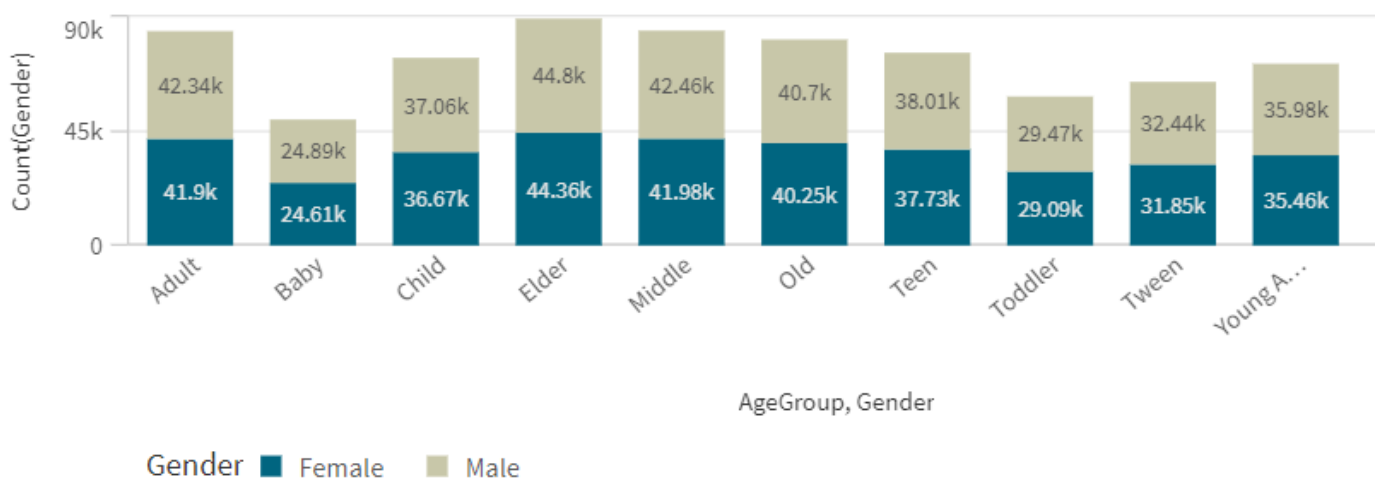
- Continent wise – Flight Status:

Flight Status (Continent Wise)



- Age Group – Gender Wise

Age Group (Gender Wise)



Dashboard

6.1 Responsive and Design of Dashboard

Overview

This project includes the design and implementation of two interactive dashboards in Qlik Sense. Each dashboard is tailored to provide specific insights into synthetic airline data, covering passenger demographics, flight statuses, and operational metrics. The design focuses on delivering clear, actionable insights through a combination of visualizations and interactive elements.

Dashboard 1: Passenger and Flight Insights

Visualizations:

- Total Number of Passengers
Type: KPI Tile
Description: Displays the total number of passengers across all flights.
- Number of Passengers Affected by Cancelled Flights
Type: KPI Tile
Description: Shows the total number of passengers who experienced flight cancellations.
- Number of Passengers Affected by Delayed Flights
Type: KPI Tile
Description: Shows the total number of passengers who experienced flight delays.
- Number of Male Passengers
Type: KPI Tile
Description: Displays the total number of male passengers.
- Top 5 Months Where Passengers Traveled the Most
Type: Bar Chart
Description: Highlights the top 5 months with the highest passenger traffic.
- Number of Female Passengers
Type: KPI Tile
Description: Displays the total number of female passengers.

- Number of Flights On-Time

Type: KPI Tile

Description: Shows the total number of flights that were on time.

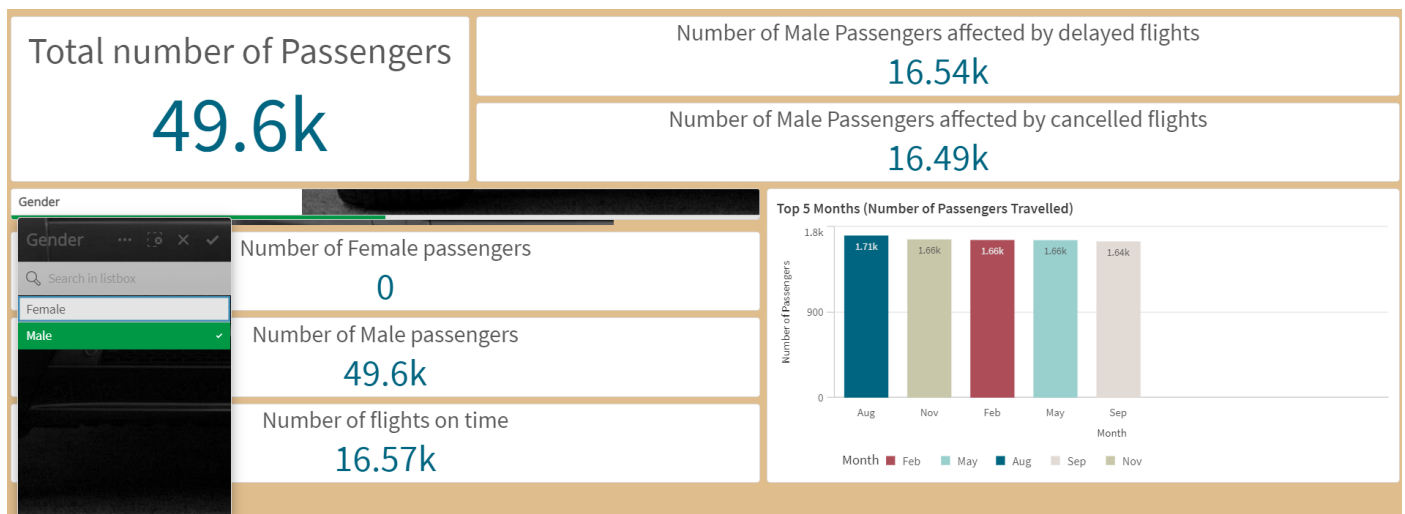
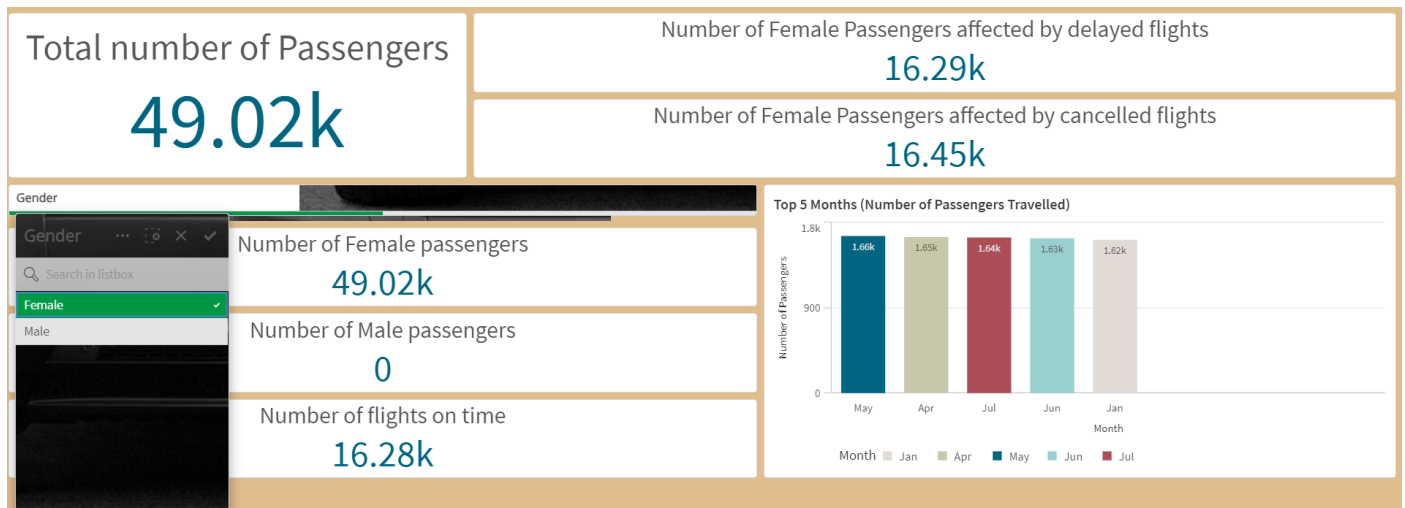
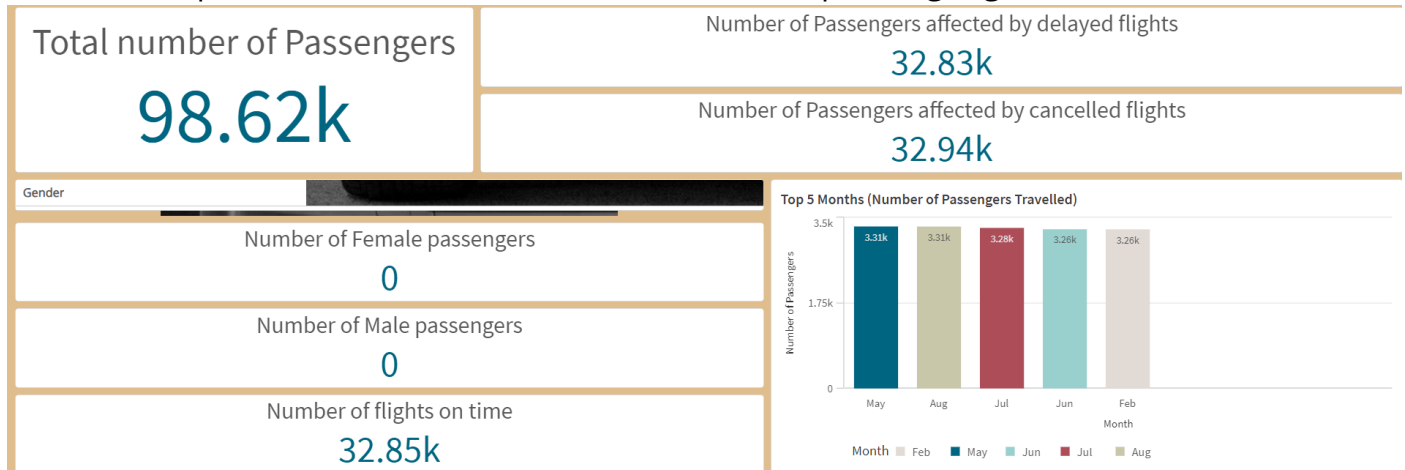
Filters:

- Gender Filter

Type: Filter Pane

Options: Male (M) and Female (F)

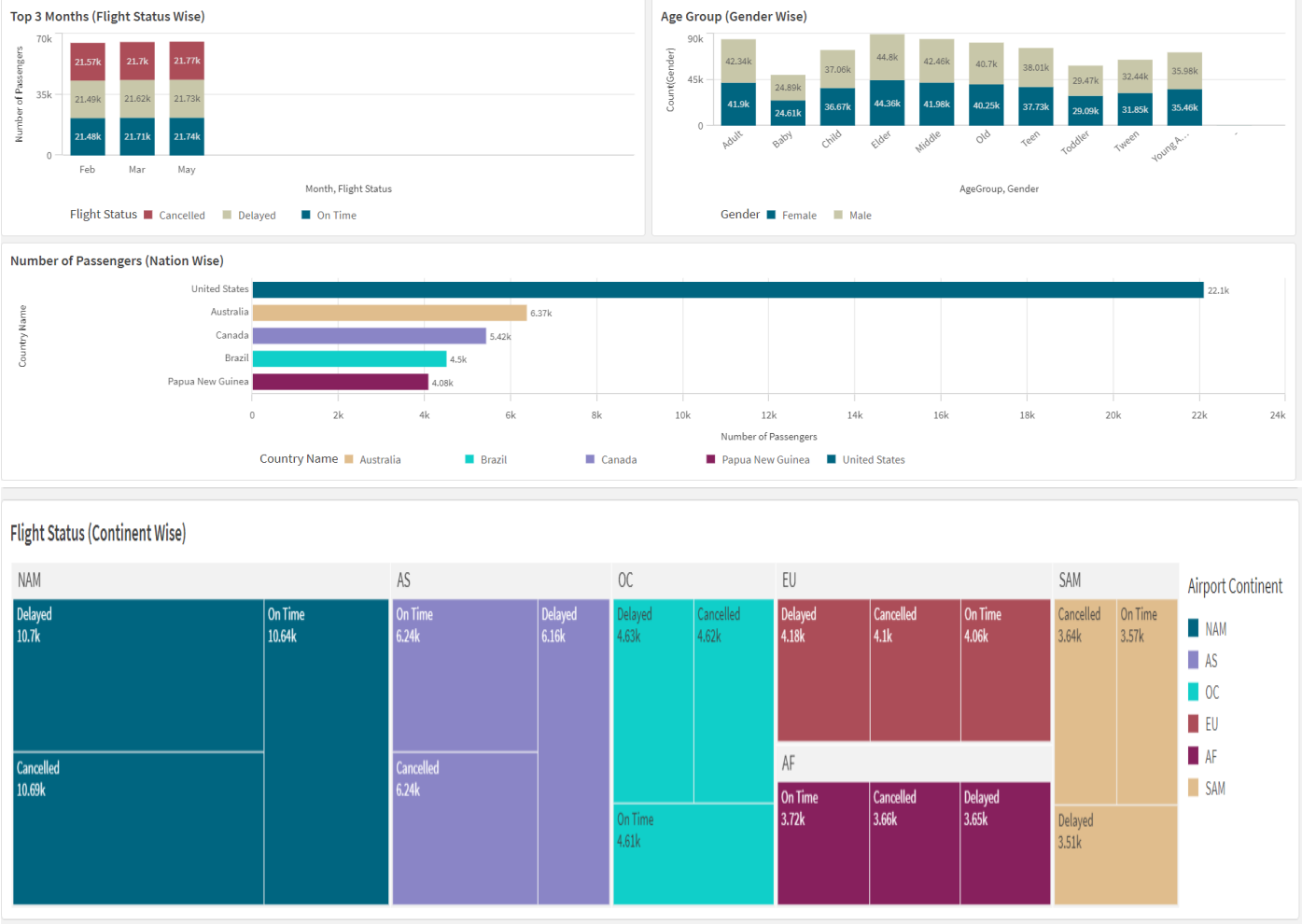
Description: Allows users to filter data based on passenger gender.



Dashboard 2: Detailed Flight and Passenger Analysis

Visualizations:

- Top 3 Months - Flight Status Wise (Delayed/Cancelled/On-Time)
Type: Stacked Bar Chart
Description: Displays the top 3 months with the highest number of flights, categorized by flight status (delayed, cancelled, on-time).
- Number of Passengers Nationality Wise
Type: Pie Chart
Description: Shows the distribution of passengers based on nationality.
- Continent Wise – Flight Status
Type: Map Chart
Description: Visualizes flight status (on-time, delayed, cancelled) by continent.
- Age Group – Gender Wise
Type: Stacked Bar Chart
Description: Displays the distribution of passengers by age group and gender.

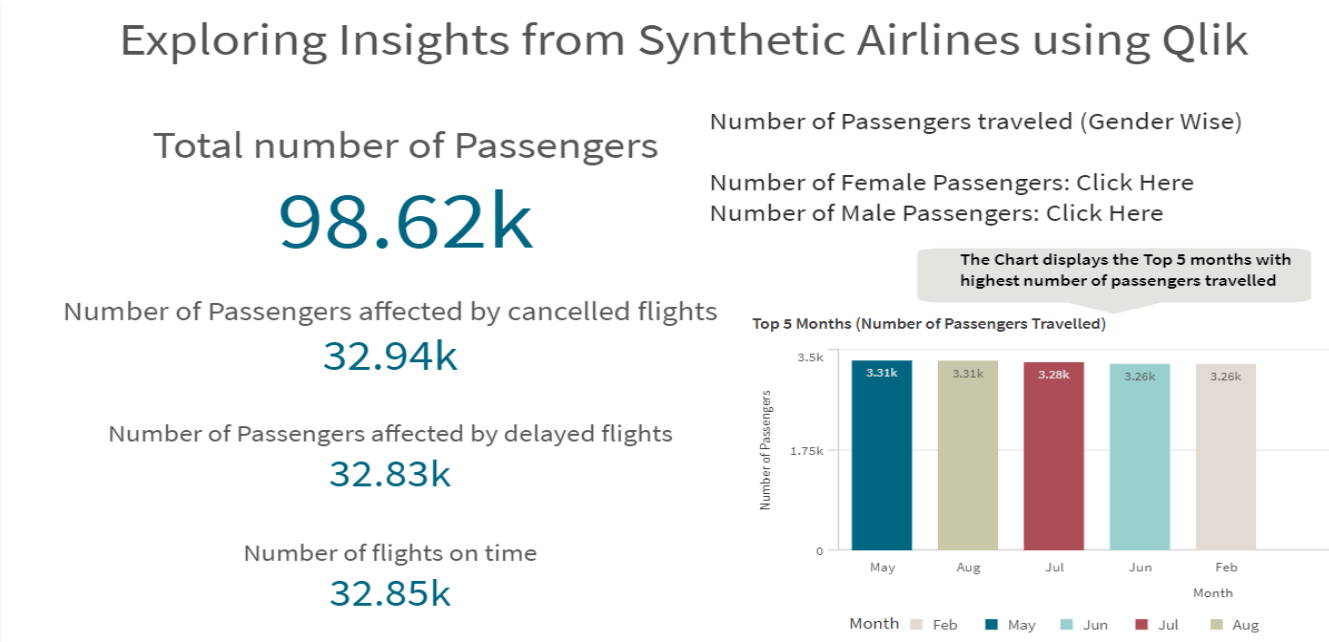
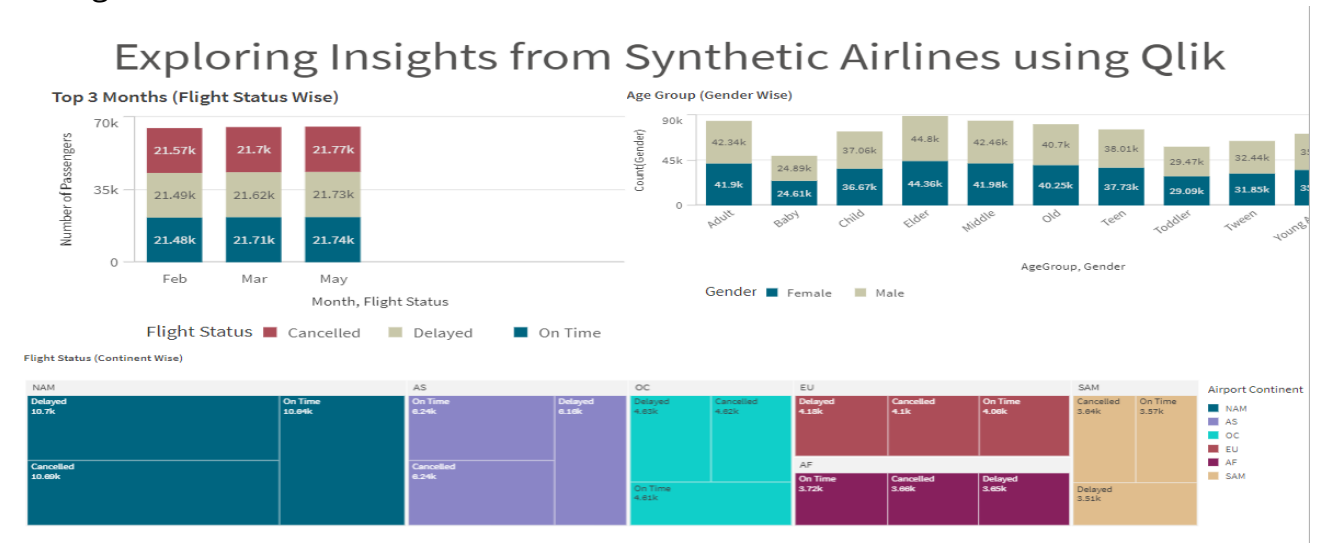


Report

7.1 Report Creation

Overview


Storytelling with data involves presenting analytical findings in a compelling and narrative format, making it easier for stakeholders to understand and act on the insights. In this project, Qlik Sense is used to create a data-driven story that guides users through the key findings from the synthetic airline data analysis. This narrative approach helps highlight important trends, patterns, and correlations, and provides context to support decision-making.



Performance Testing

8.1 Amount of Data loaded

The amount of data loaded into Qlik Sense for this project is critical to ensure comprehensive analysis while maintaining performance efficiency. The synthetic airline data includes various datasets covering different aspects of airline operations.

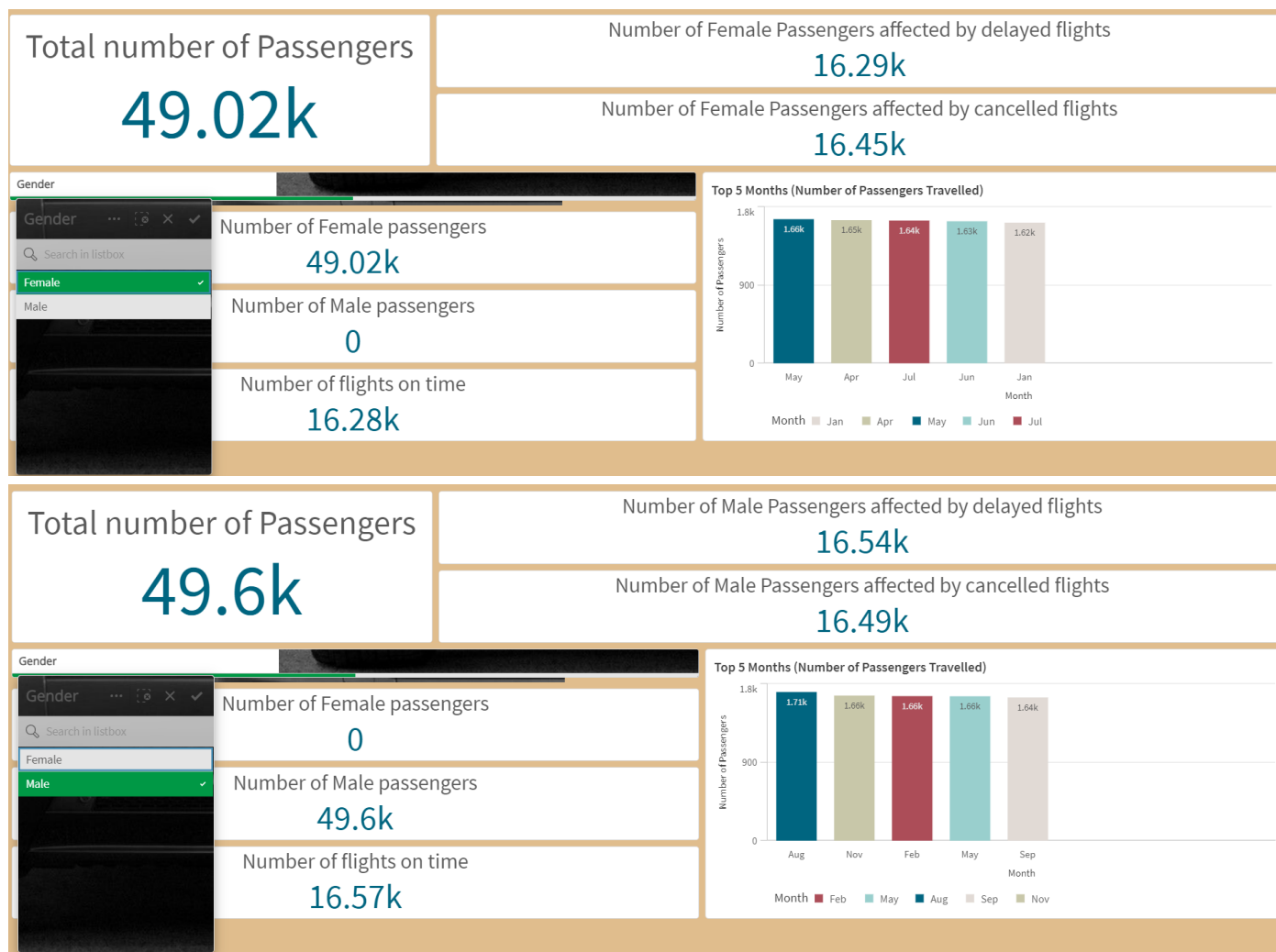
Airline_Dataset_
\$Syn 1 
AgeGroup
Departure_Date
Year
Month
Airline_Dataset-1.Passenger ID
Airline_Dataset-1.First Name
Airline_Dataset-1.Last Name
Airline_Dataset-1.Gender
Airline_Dataset-1.Age
Airline_Dataset-1.Nationality
Airline_Dataset-1.Airport Name
Airline_Dataset-1.Airport Country Code
Airline_Dataset-1.Country Name
Airline_Dataset-1.Airport Continent
Airline_Dataset-1.Continents
Airline_Dataset-1.Departure Date
Airline_Dataset-1.Arrival Airport
Airline_Dataset-1.Pilot Name
Airline_Dataset-1.Flight Status
Airline_Dataset-1.Airline_Dataset-1.Nationality_GeoInfo
Airline_Dataset-1.Airline_Dataset-1.Airport Country Code_GeoInfo
Airline_Dataset-1.Airline_Dataset-1.Country Name_GeoInfo

8.2 Utilization of Filters

Filters are a crucial component of interactive data analysis in Qlik Sense, enabling users to refine their view of the data and focus on specific subsets of interest. In this project, various filters have been implemented to enhance the exploration and analysis of the synthetic airline data.

Gender Filter-

- Description: Allows users to filter the data based on passenger gender.
- Options: Male (M), Female (F).
- Usage: Applied to visualizations to analyze gender-specific trends and insights. For example, users can examine the number of male vs. female passengers or filter flight status data by gender.



8.3 Data Pre-processing - Qlik Sense Script

This is the code to add new fields like AgeGroup , Month and Year.

```
[Airline_Dataset]:
Load *;
[Airline_Dataset]:
NoConcatenate Load *,

If(Age >= 0 and Age <=1 , 'Baby',
    if(Age >=1 and Age<=3, 'Toddler',
    if(Age >=4 and Age<=9, 'Child',
    if(Age >=10 and Age<=12, 'Tween',
    if(Age >=13 and Age<=19, 'Teen',
    if(Age >=20 and Age<=24, 'Young Adult',
    if(Age >=25 and Age<=39, 'Adult',
    if(Age >=40 and Age<=54, 'Middle',
    if(Age >=55 and Age<=79, 'Elder',
    if(Age >=80 , 'Old')))))))) as AgeGroup,
Date#([Departure Date], 'MM/DD/YYYY') AS [Departure_Date],
Year([Departure Date]) as Year,
Month([Departure Date]) as Month
Resident [Airline_Dataset]
Where not ([Arrival Airport] = '0' or [Arrival Airport] = '-');
```

8.4 Number of Visualizations/Graphs

- Total Number of Passengers
Type: KPI Tile
Description: Displays the total number of passengers across all flights.
- Number of Passengers Affected by Cancelled Flights
Type: KPI Tile
Description: Shows the total number of passengers who experienced flight cancellations.
- Number of Passengers Affected by Delayed Flights
Type: KPI Tile
Description: Shows the total number of passengers who experienced flight delays.
- Number of Male Passengers

Type: KPI Tile

Description: Displays the total number of male passengers.

- Top 5 Months Where Passengers Traveled the Most

Type: Bar Chart

Description: Highlights the top 5 months with the highest passenger traffic.

- Number of Female Passengers

Type: KPI Tile

Description: Displays the total number of female passengers.

- Number of Flights On-Time

Type: KPI Tile

Description: Shows the total number of flights that were on time.

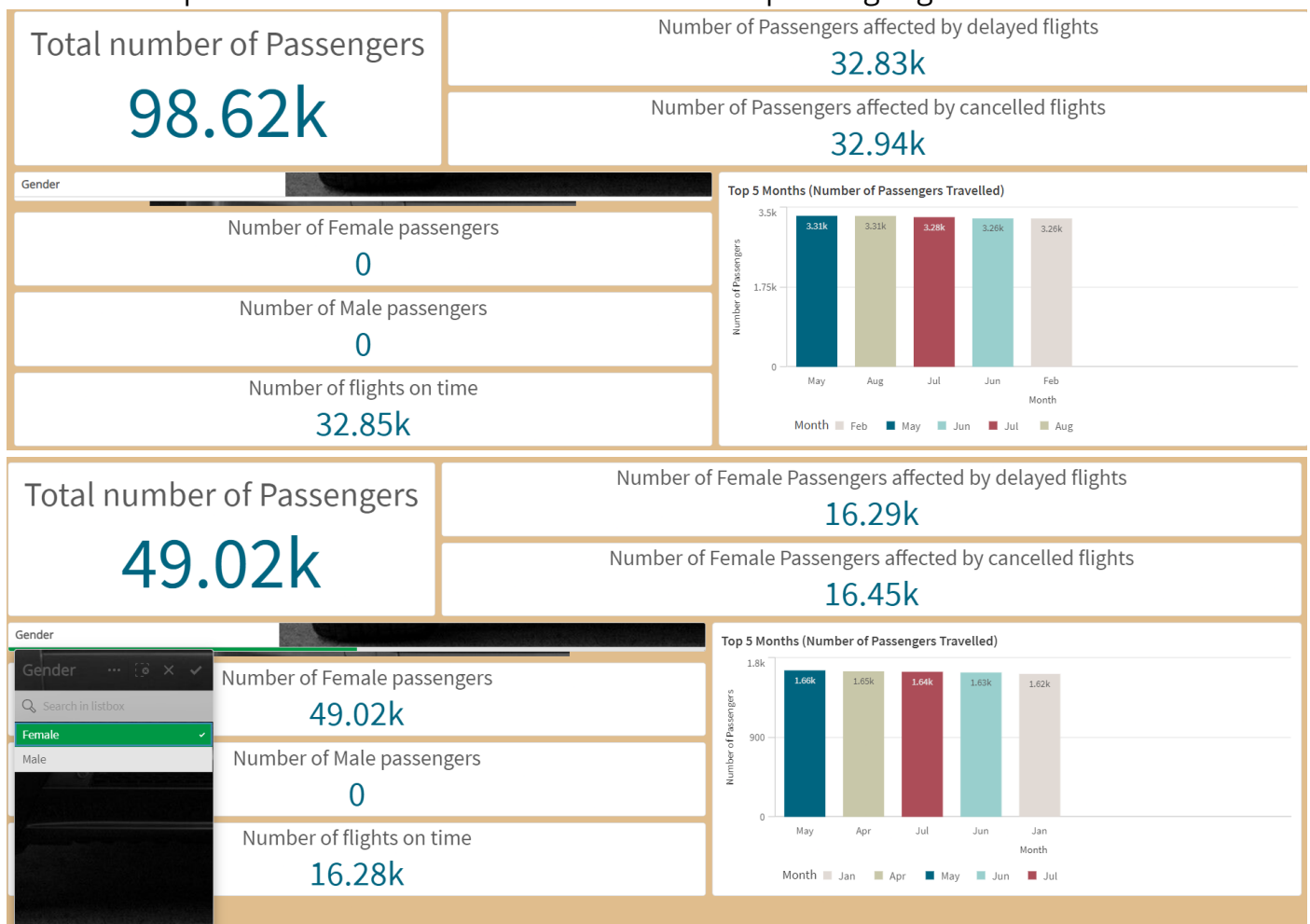
Filters:

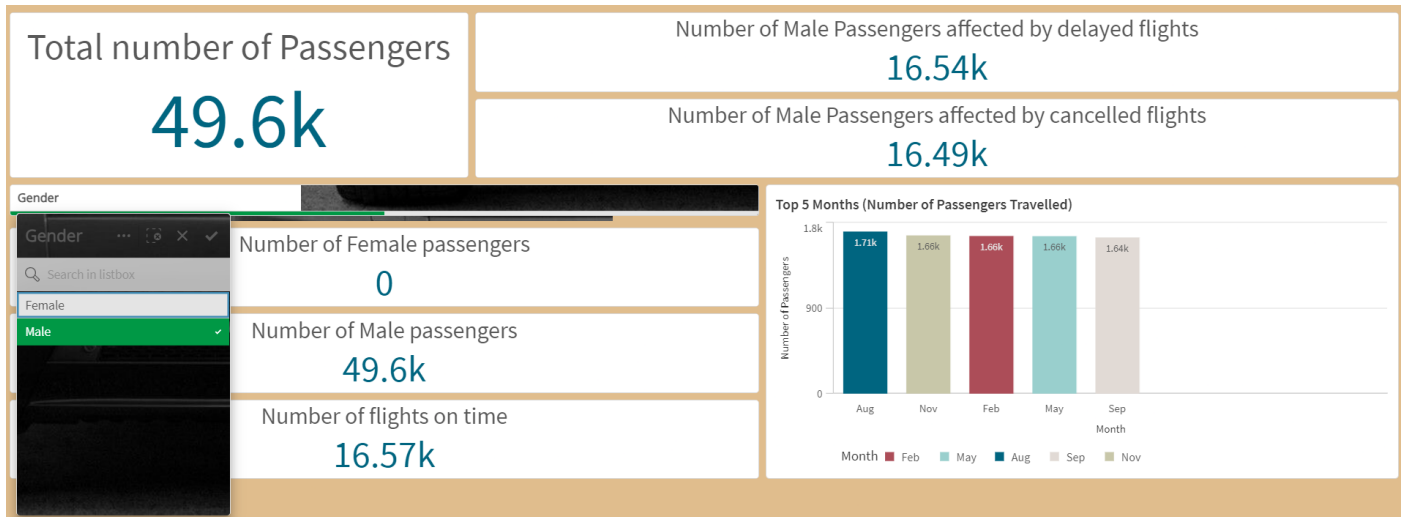
- Gender Filter

Type: Filter Pane

Options: Male (M) and Female (F)

Description: Allows users to filter data based on passenger gender.

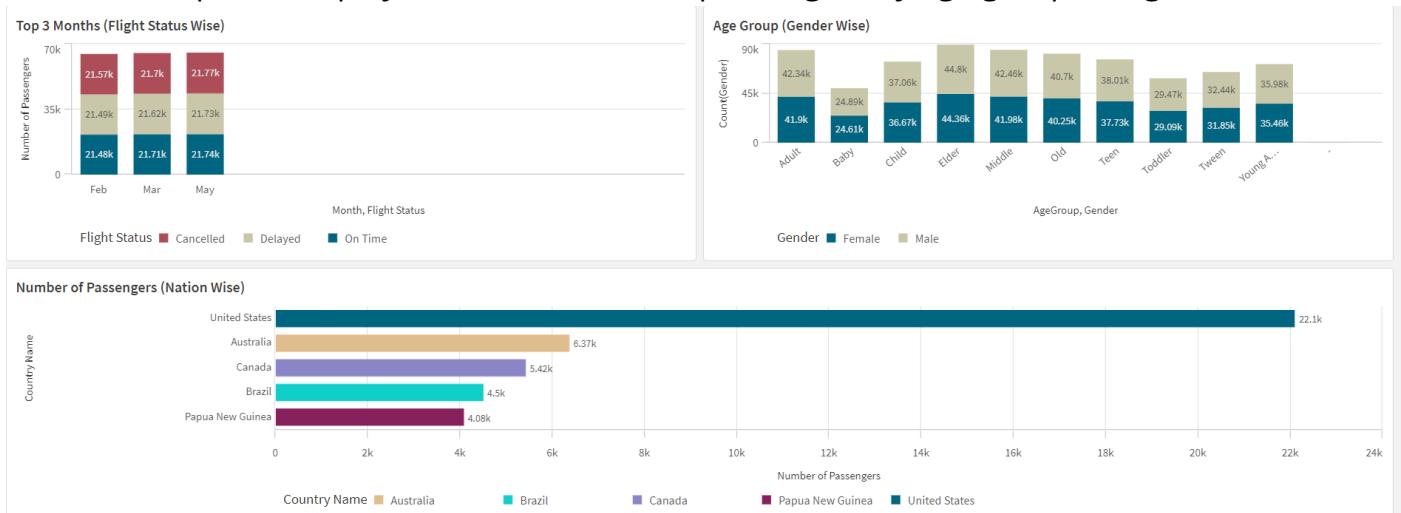




Dashboard 2: Detailed Flight and Passenger Analysis

Visualizations:

- Top 3 Months - Flight Status Wise (Delayed/Cancelled/On-Time)**
 Type: Stacked Bar Chart
 Description: Displays the top 3 months with the highest number of flights, categorized by flight status (delayed, cancelled, on-time).
- Number of Passengers Nationality Wise**
 Type: Pie Chart
 Description: Shows the distribution of passengers based on nationality.
- Continent Wise – Flight Status**
 Type: Map Chart
 Description: Visualizes flight status (on-time, delayed, cancelled) by continent.
- Age Group – Gender Wise**
 Type: Stacked Bar Chart
 Description: Displays the distribution of passengers by age group and gender.



Flight Status (Continent Wise)

