Project Report On Synthetic Airline Data Analysis With Qlik

1. INTRODUCTION

1.1 Overview of the Project

The project "Exploring Insights from Synthetic Airline Data Analysis with Qlik" focuses on leveraging synthetic airline data to derive meaningful insights using Qlik, a leading business intelligence and data visualization tool. The synthetic data simulates key aspects of airline operations, including flight schedules, passenger demographics, ticket sales, and performance metrics. The objective is to utilize Qlik's powerful analytical and visualization capabilities to uncover patterns, trends, and correlations within this data. This can significantly aid decision-making processes for airlines, airports, and related stakeholders by providing actionable insights into various operational and customer-centric aspects of airline management.

1.2 Purpose

The primary purpose of this project is to demonstrate how synthetic airline data can be analyzed using Qlik to support decision-making and strategic planning in the aviation industry. By exploring different scenarios, the project aims to showcase the potential benefits of using Qlik for:

- Revenue Optimization: Analyzing ticket sales data to identify peak travel times, popular destinations, and effective pricing strategies.
- Operational Efficiency: Enhancing airport operational efficiency by identifying bottlenecks and optimizing resource allocation.
- Customer Experience Enhancement: Improving passenger experience by understanding customer preferences and feedback, leading to personalized services and targeted marketing.

Using this project, stakeholders can achieve:

- Increased Revenue: By identifying and capitalizing on high-demand travel periods

and popular destinations, and by refining pricing strategies based on customer purchasing behavior.
- Improved Operational Efficiency: Through the identification and mitigation of operational bottlenecks, optimal resource allocation, and enhanced process management.
- Enhanced Customer Satisfaction: By analyzing customer feedback to identify areas for service improvement, personalizing passenger services, and designing targeted marketing campaigns.

1.3 Technical Architecture

1. Data Sources
    - Synthetic Airline Data: Simulated data sets representing various aspects of airline operations.

2. Data Processing and Integration
    - Data Extraction: Importing synthetic data into Qlik.
    - Data Transformation: Cleaning and organizing data for analysis, including merging different data sets where necessary.

3. Qlik Analytical Tools
    - Qlik Sense: Tools for data visualization and dashboard creation.

4.Visualization and Reporting
    - Dashboards: Interactive dashboards for real-time data visualization.
    - Reports: Customizable reports for detailed analysis and insights.

2. PROBLEM DEFINATION

2.1 Business Problem

The aviation industry faces multiple challenges, including fluctuating demand, operational inefficiencies, and varying levels of customer satisfaction. Airlines and airports need robust data analysis tools to optimize revenue, improve operational efficiency, and enhance the passenger experience.

2.2 Business Requirements

1. Revenue Optimization
    - Historical Data Analysis: Evaluate past ticket sales to identify patterns.
    - Peak and Off-Peak Analysis: Determine peak travel periods and adjust pricing.
    - Customer Segmentation: Classify customers based on purchasing behavior for targeted marketing.

2. Operational Efficiency
    - Flight Schedule Analysis: Optimize flight schedules to minimize delays and maximize utilization.
    - Passenger Flow Analysis: Understand passenger movements to improve resource allocation.
    - Process Optimization: Identify and resolve operational bottlenecks in areas such as luggage handling.

3. Customer Experience Enhancement
    - Sentiment Analysis: Analyze customer feedback to gauge satisfaction levels.
    - Service Personalization: Use customer preferences to tailor services.
    - Marketing Campaigns: Design targeted campaigns based on customer demographics and behavior.

2.3 Literature Survey

- Studies on revenue management in airlines have emphasized the importance of dynamic pricing and demand forecasting (Talluri & Van Ryzin, 2004).
- Research on operational efficiency has highlighted the role of data analytics in optimizing airport operations and reducing delays (Cook & Tanner, 2011).
- Customer experience research has shown that personalized services and effective handling of customer feedback significantly enhance passenger satisfaction (Gursoy et al., 2014).
- The application of business intelligence tools like Qlik in the aviation industry has been explored, demonstrating significant improvements in decision-making and operational efficiency (Watson & Wixom, 2007).
- The integration of synthetic data for simulation and testing purposes has been

validated as an effective method for preparing and enhancing real-world data analysis (Biem et al., 2010).

3. DATA COLLECTION

3.1 Data Collection

Data collection is a fundamental process that involves gathering and measuring information on variables of interest in a structured and systematic manner. This process enables researchers and analysts to answer specific research questions, test hypotheses, evaluate outcomes, and generate meaningful insights from the data. In the context of this project, we are utilizing a synthetic airline dataset, which simulates various aspects of airline operations. The dataset includes comprehensive meta-information about the columns described in the CSV files, providing a robust foundation for analysis using Qlik.

The dataset used in this project contains detailed information related to passengers, flights, and airports. Each record in the dataset represents a unique passenger's flight details. Below is a detailed description of each column included in the dataset:

- **Passenger ID:** A unique identifier assigned to each passenger, ensuring the distinctness of each record.
- **First Name:** The first name of the passenger.
- **Last Name:** The last name of the passenger.
- **Gender:** The gender of the passenger, categorized as male, female, or other.
- **Age:** The age of the passenger.
- **Nationality:** The nationality of the passenger, indicating their country of origin.
- **Airport Name:** The name of the airport where the passenger boarded the flight.
- **Airport Country Code:** The country code where the airport is located, providing a standardized reference.
- **Country Name:** The name of the country where the airport is situated.
- **Airport Continent:** The continent where the airport is located, facilitating regional analysis.
- **Continents:** The continents involved in the passenger’s flight route, useful for understanding the geographical span of the flights.
- **Departure Date:** The date when the flight departed, essential for temporal analysis.
- **Arrival Airport:** The destination airport of the flight.

- **Pilot Name:** The name of the pilot operating the flight.
- **Flight Status:** The current status of the flight, which can be on-time, delayed, or canceled.

3.2 Connecting the Data with Qlik Sense

Connecting data with Qlik Sense involves integrating various data sources into the Qlik Sense environment to enable comprehensive analysis and visualization. This process includes:
1.Data Importation
2.Data Transformation
4.Data Modeling
3.Data Loading

4. DATA PREPARATION

4.1 Data Loading:

Qlik Sense provides a user-friendly interface that simplifies the data loading process through a drag-and-drop method. This approach allows users to quickly import data without needing to write complex scripts.

4.2 Data Cleaning and Preprocessing:

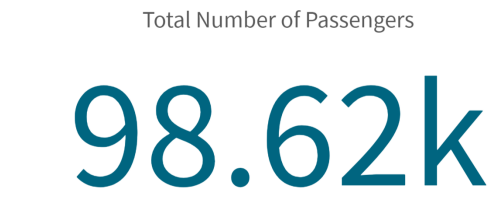
This step involves ensuring that the dataset is free of irrelevant or missing data and that all entries are accurate and complete. Also transforming the data into a suitable format for visualization by renaming fields, adjusting data types, applying necessary filters, and establishing relationships between tables to integrate the dataset seamlessly.
In context to this project, the field age group was used to create a new field named Age_group which could be used to analyse the number of passengers based on age group.
The code shown below is used in Data load editor to create a new field as part of preprocessing:

```
if([Age] >= 0 and [Age] <= 1, 'Baby',
   if([Age] >= 2 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-Aged',
                        if([Age]>=55 and [Age]<=79,'Elder','Just plain old'))))))))) AS [Age Category]
```

5. DATA VISUALISATION

5.1 Total Number of Passengers:

The first step is understanding the overall volume of passengers. Visualizing the total number of passengers over a specific period can help airlines gauge demand and plan accordingly.



5.2 Number of Passengers affected by Cancelled flights:

Identifying the number of passengers affected by delayed and canceled flights helps in assessing the customer impact and developing strategies to mitigate disruptions.I

5.3 Number of Passengers affected by Delayed flights:

Identifying the number of passengers affected by delayed and canceled flights helps in assessing the customer impact and developing strategies to mitigate disruptions.

Number of passengers affected by delayed flights

32.83k

5.4 Number of flights on Time:

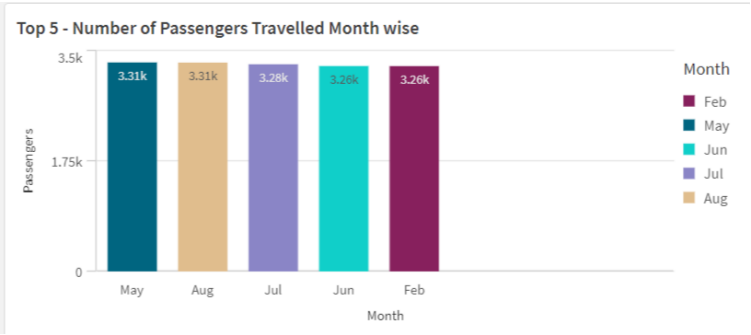
Identifying the number of passengers affected by delayed and canceled flights helps in assessing the customer impact and developing strategies to mitigate disruptions.

Number of Flights on time

32.85k

5.5 Top 5 - Number of Passengers Travelled - Month Wise :

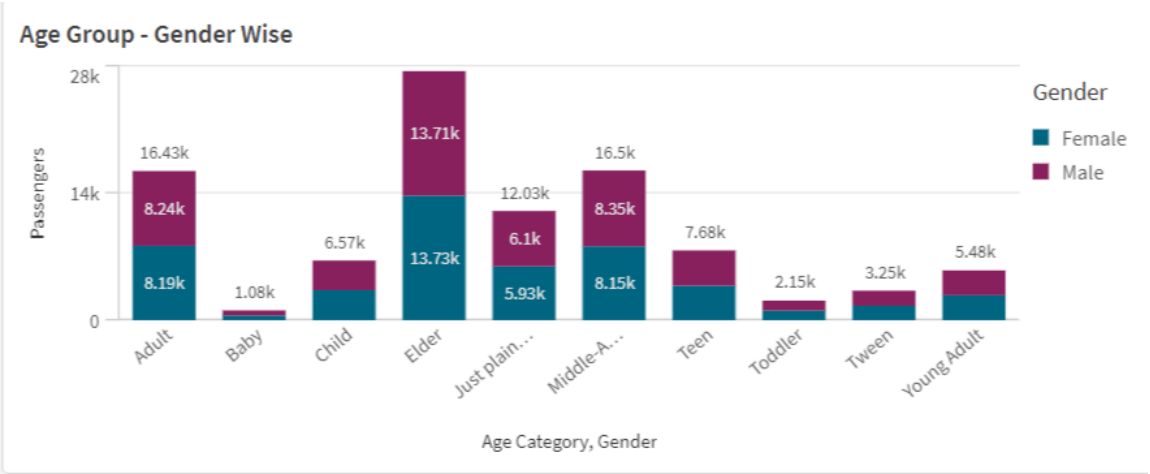
Visualizing passenger numbers on a month-by-month basis can identify peak travel periods and seasonal trends, informing capacity planning and marketing efforts.



5.6 Age Group - Gender Wise:

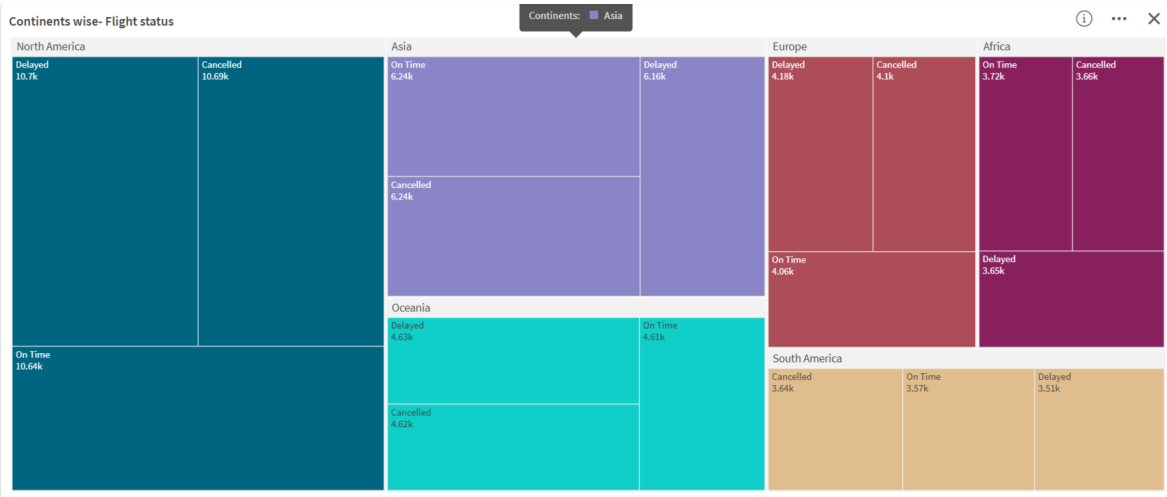
Creating an age group field allows for a detailed breakdown of passengers by age category. This analysis can reveal demographic trends and help in tailoring marketing strategies. Examining the gender distribution across different age groups provides

insights into the passenger profile, enabling more targeted services and campaigns.



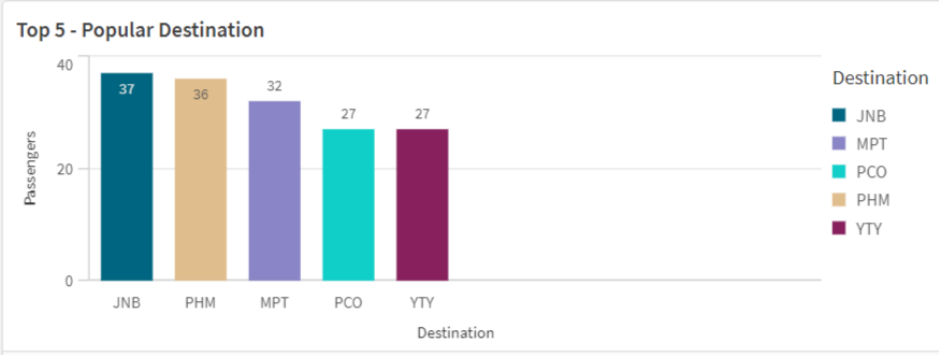
5.7 Continent wise- Flight Status:

Breaking down flight status by continent provides insights into regional performance and helps in understanding geographical challenges.



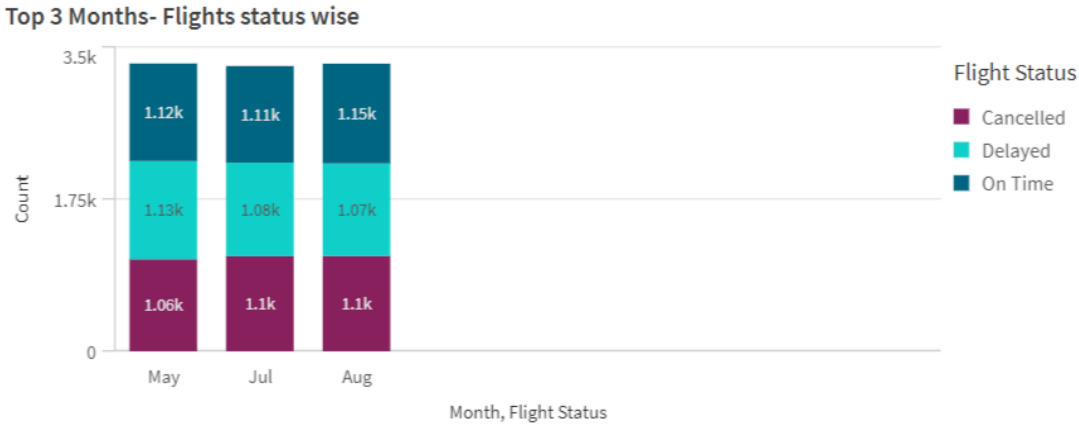
5.8 Top 5- Popular Destination:

Identifying the top 5 destinations based on passenger numbers highlights popular routes, aiding in network planning and route optimization.



5.9 Top 3 Months - Flights status wise:

Analyzing flight status data across different months can uncover seasonal patterns in operational performance, guiding resource allocation and contingency planning.

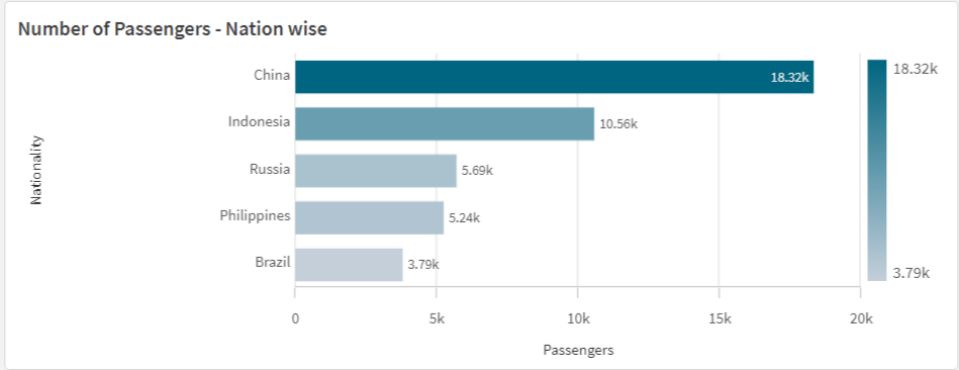


5.10 Top 5 - Flights status - Airpot wise:

Identifying airports with the highest number of on-time, delayed, or canceled flights helps in pinpointing operational bottlenecks and areas for improvement.

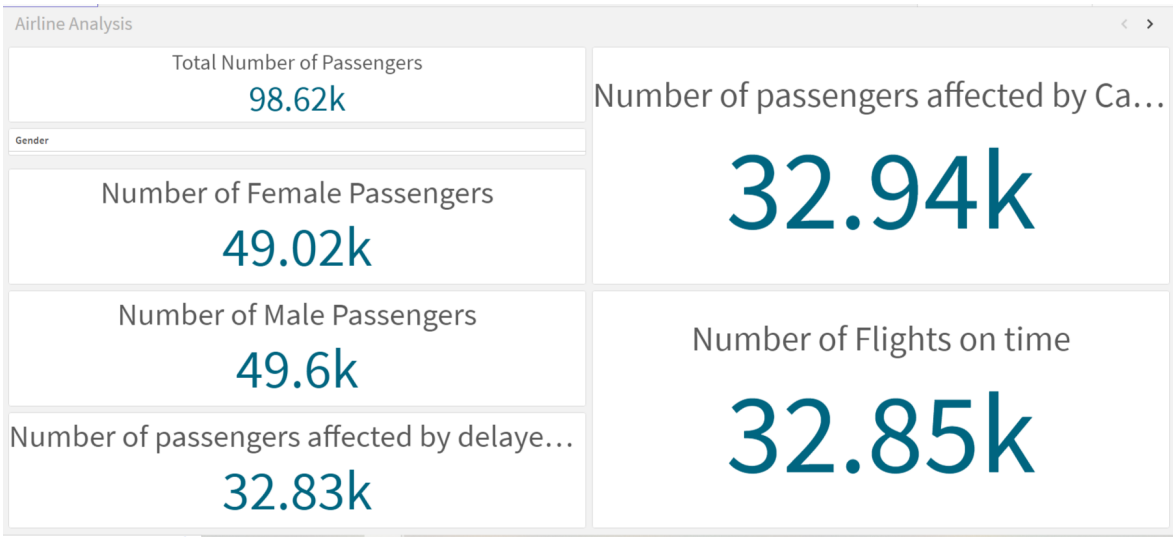
5.11 Number of Passengers - Nation wise:

Analyzing passenger numbers by nationality can reveal key markets and support the development of region-specific strategies.



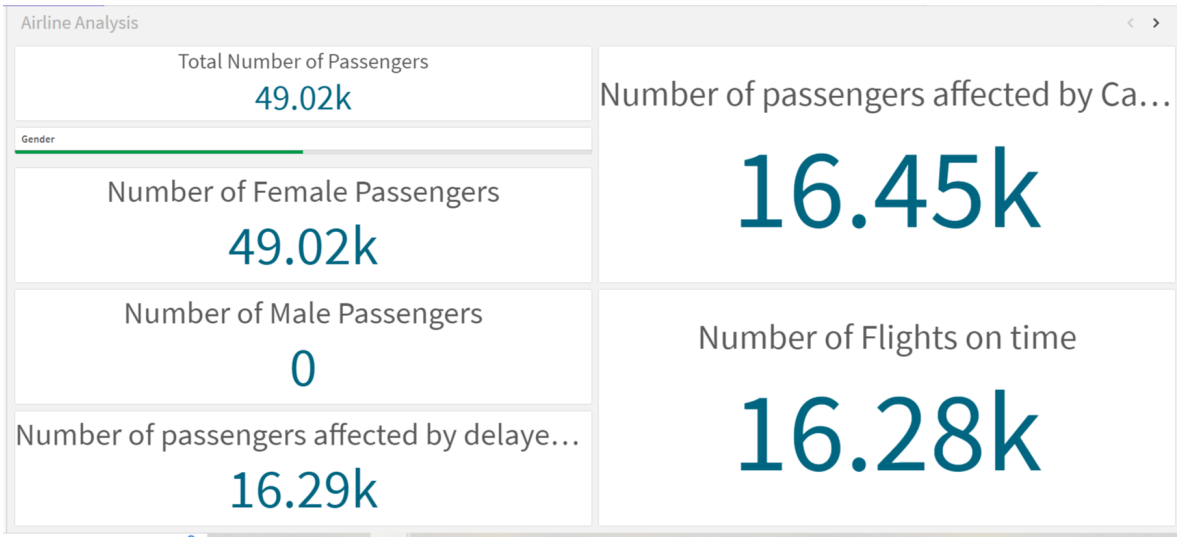
6.DASHBOARD

Airline Analysis Dashboard 1:



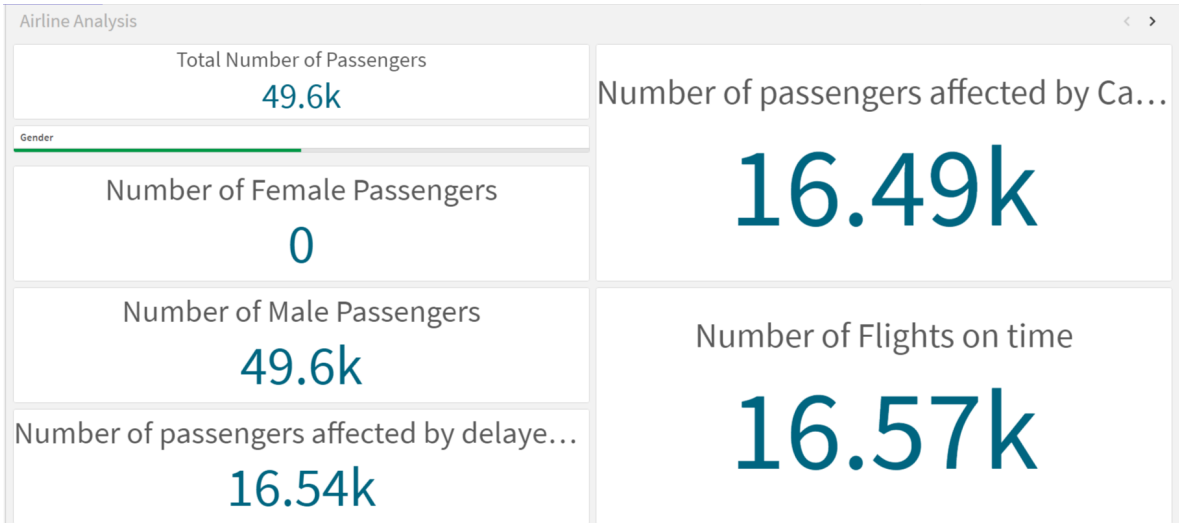
Airline Analysis Dashboard 2(Gender Based Analysis using Filter pane :Female):

The filter pane allows for the analysis of the total number of male and female passengers separately.

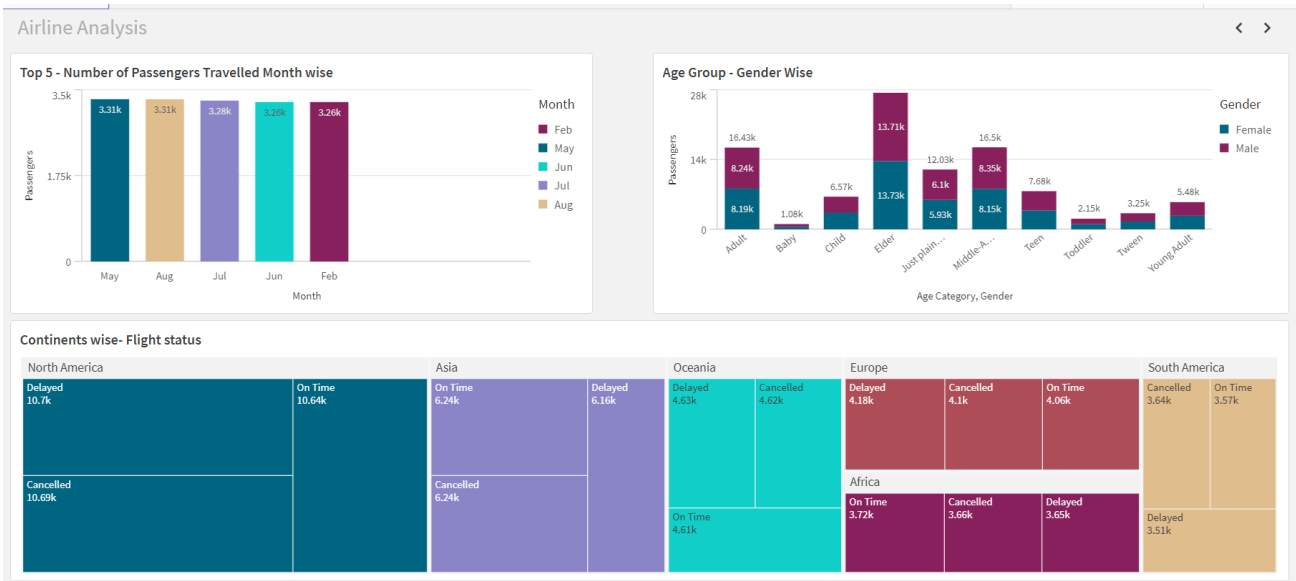


Airline Analysis Dashboard 3 (Gender Based Analysis using Filter pane :Male):

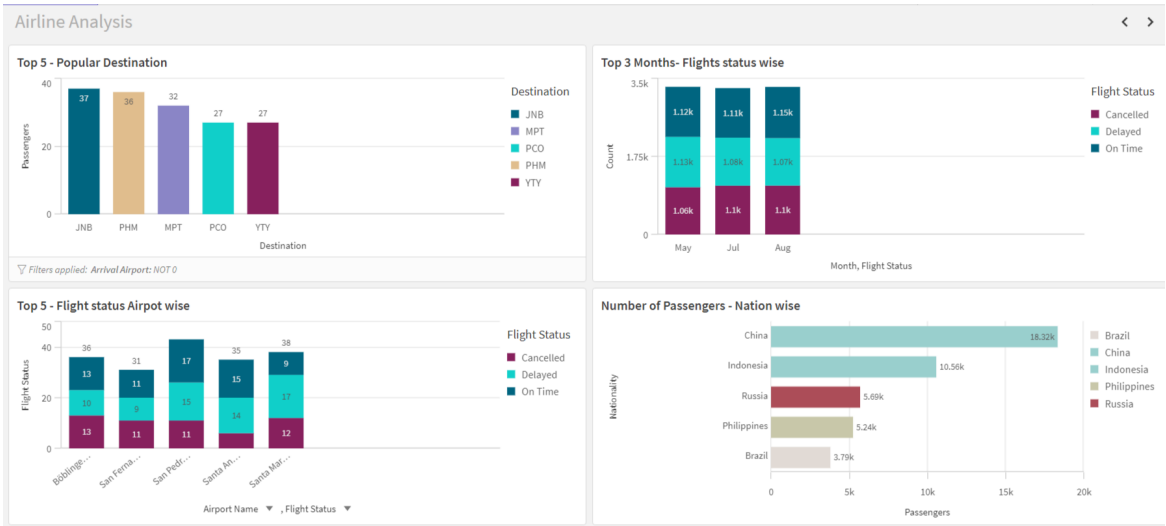
The filter pane allows for the analysis of the total number of male and female passengers separately.



Airline Analysis Dashboard 4:



Airline Analysis Dashboard 5:



7. STORY

<u>Observations:</u>

1.Total number of passengers: 98.62k passengers have been recorded, showcasing a substantial dataset for analysis.

2.Cancelled and Delayed FLights:

- 32.94k passengers were affected by cancelled flights.

- 32.83k passengers experienced delays.
- The number of flights on time was 32.85k, which is comparable to number of delated and cancelled flights, indicating a significant number of operational issues.

3. Peak Travel Periods: The top 5 months for passenger travel (Feb, Jun, Jul, Aug) each had around 3.26k to 3.3k passengers, showing consistent high travel volumes during these periods.

4.Flight Status by Continent:

- North America showed the highest number of delays(10.7k) and cancellations(10.69k), but also had a substantial number of on-time flights(10.64k).
- Europe and Asia had notable delyas and cancellations but were relatively balanced with on-time flights.

5.Passenger Demagraphics: Age groups and gender analysis indicate that the majority of passengers fall within the elder age group, with a higher number of male passengers compared to females across most age categories.

6.Popular Destinations: The top destinations include JNB, PHM<MPT,PCO and YTY, with JNB being popular at 37 passengers.

7.Flight Status By Airport: Analysis of the top 5 airports by flight status shows a variation in the number of cancelled, delayed and on-time flights, indicating potential operational challenges at these airports.

8.Nation-wise Passengers: The highest number of passengers come from China(18.32k), followed by Indonesia(10.56k), Russia(5.69k), Philippines (5.24k), and Brazil(3.79k).

Based on the observations, potential courses of action could include:

1.Improve operational Efficiency:

- Addressing the hight number of cancelled and delayed flights by investigating and resolving common causes. Implementing better scheduling, maintenance routines, and resource allocation can help improve on-time performance.

- Analysing the reason behind operational inefficiencies in specific regions like North America and Europe. Implementing region-specific solutions to enhance overall efficiency.
- Expanding successful practices in regions with high on-time performance to other regions.
- Using insights from passengers flow and flight status data to optimize flight schedules.
- Ensuring that high-demand routes have sufficient capacity and that resources are allocated efficiently to reduce delays and cancellaions.

2.Revenue Optimization:

- Using the peak travel period data to adjust pricing strategies, offering promotions during off-peak time to balance the load and maximize revenue. Dynamic pricing models could also be beneficial.
- Utilizing data on popular destinations and peak travel periods to create targeted marketing campaigs.
- Highlighting destinations that are trending and offer special deals to attract more passengers.

3.Enhance Customer Experience:

- Focusing on the passenger demographics to tailor services and marketing strategies. Providing personalized services for frequent travelers, different age groups, and high-demand nationalities can improve satisfaction and loyalty.
- Improving communication and compensation strategies for passengers affected by cancellations and delays to maintain trust and satisfaction.
- Continuonusly collecting and analysing customer feedback to identify pain points and areas for improvement.
- Implementing Changes based on this feedback to enhance the overall travel experience.

By implementing these actions, airlines can improve operational efficiency, enhance customer satisfaction, and optimize revenue, ultimately leading to a more competitive and successful operation.