1. INTRODUCTION

1.1 OVERVIEW: A BRIEF DESCRIPTION ABOUT YOUR PROJECT

Project Description

In this project, we leverage Qlik Sense, a powerful data analytics and visualization tool, to explore and analyze a synthetic dataset representing various aspects of airline operations. The primary objective is to uncover valuable insights from the data that can inform decision-making and strategic planning within the airline industry.

Key Objectives:

1. Passenger Demographics:

● KPI Visualizations: Create visualizations to display key metrics such as the number of male and female passengers.

● Trend Analysis: Analyze passenger demographics over time to identify trends and patterns.

2. Flight Performance:

● On-Time Performance: Develop KPI visualizations to track the number of flights that arrive on time.

● Delay Analysis: Investigate causes and patterns of flight delays to improve operational efficiency.

3. Travel Trends:

● Seasonal Trends: Create bar charts and other visualizations to identify the top months with the highest passenger volumes.

● Geographic Analysis: Map visualizations to analyze passenger traffic across different routes and destinations.

4. Operational Insights:

● Flight Frequencies: Analyze the frequency of flights and passenger loads to optimize scheduling.

● Revenue Metrics: Calculate and visualize revenue-related metrics to understand financial performance.

1.2 PURPOSE: THE USE OF THIS PROJECT AND ACHIEVABLE OUTCOMES

The primary purpose of this project is to demonstrate how advanced data analytics and visualization techniques, specifically using Qlik Sense, can be applied to synthetic airline data to extract meaningful insights. These insights can help inform strategic decision-making, optimize operations, and enhance overall performance within the airline industry.

Achievable Outcomes

1. Enhanced Decision-Making:

● Informed Strategies: By analyzing key performance indicators (KPIs) and trends, airline management can make data-driven decisions that align with business goals and market demands.

● Operational Efficiency: Identifying patterns in flight performance and passenger demographics can lead to more efficient resource allocation and operational planning.

2. Improved Operational Performance:

● On-Time Performance: By tracking the number of flights that arrive on time and analyzing delay patterns, airlines can implement strategies to improve punctuality, leading to increased customer satisfaction.

● Flight Scheduling Optimization: Understanding passenger travel trends and flight frequencies helps optimize flight schedules, reducing downtime and increasing aircraft utilization.

3. Customer Insights and Experience:

● Demographic Analysis: Analyzing passenger demographics allows airlines to tailor services and marketing strategies to different customer segments, improving the overall travel experience.

● Seasonal Trends: Identifying peak travel months helps in anticipating demand and preparing accordingly, ensuring better service delivery during busy periods.

1.3 TECHNICAL ARCHITECTURE

Overview

The technical architecture of this project involves integrating Qlik Sense for data visualization and analysis with various data sources, processing layers, and user interfaces. The architecture ensures data is efficiently collected, processed, and visualized, providing valuable insights for airline operations.

Key Components

1. Data Sources:

● Synthetic Airline Data: This includes data on passenger demographics, flight details, on-time performance, revenue metrics, and more.

● External Data: Additional datasets such as weather information, economic indicators, and competitor analysis can be integrated to enrich the analysis.

2. Data Ingestion and Storage:

● Data Lake/Data Warehouse: A centralized repository where raw data from different sources is stored. This can be on-premises or cloud-based (e.g., Amazon S3, Google BigQuery, Azure Data Lake).

● ETL Processes: Extract, Transform, Load processes to clean, transform, and load data into the data warehouse. Tools like Apache Nifi, Talend, or custom scripts can be used.

3. Data Processing Layer:

● Data Preparation: Data is preprocessed, aggregated, and prepared for analysis. This involves cleaning, normalizing, and transforming data into

suitable formats for visualization.
● Data Modeling: Creating data models that represent the relationships between different datasets. This helps in creating meaningful visualizations and analyses in Qlik Sense.

2. DEFINE PROBLEM/PROBLEM UNDERSTANDING

2.1 BUSINESS PROBLEM: OPTIMIZING AIRLINE OPERATIONS AND ENHANCING PASSENGER EXPERIENCE

Context

In the highly competitive airline industry, operational efficiency and customer satisfaction are critical to maintaining profitability and market share. Airlines face various challenges, including managing flight schedules, optimizing resource allocation, ensuring on-time performance, and understanding passenger demographics and travel patterns. Leveraging data analytics can help address these challenges by providing actionable insights to drive strategic decisionmaking and operational improvements.

Key Business Problems

1.Operational Efficiency:

○ Flight Delays: Frequent flight delays can lead to increased operational costs, decreased customer satisfaction, and potential regulatory penalties.

○ Resource Allocation: Inefficient allocation of resources, such as crew and aircraft, can result in higher operational costs and reduced service quality.
○ Schedule Optimization: Inadequate flight scheduling can lead to underutilized or overburdened routes, impacting profitability and operational stability.

2. Customer Satisfaction:

○ On-Time Performance: Delays and cancellations negatively impact passenger experience, leading to customer dissatisfaction and potential loss of loyalty.
○ Service Customization: Understanding passenger demographics and preferences is essential.

2.2 BUSINESS REQUIREMENTS

To effectively address the business problems and achieve the desired outcomes, the project must meet the following business requirements. These requirements ensure that the data analytics and visualization solutions provided by Qlik Sense are aligned with the airline's strategic goals and operational needs.

Data Requirements

1. Comprehensive Data Collection:

○ Collect detailed data on flight schedules, including departure and arrival times, delays, and cancellations.

○ Gather passenger demographics, such as age, gender, travel class, and frequent flyer status.
○ Include financial data such as ticket prices, ancillary revenue, and operational costs.
○ Integrate external data sources like weather conditions, economic indicators, and competitor performance.

2. Data Quality and Integrity:

○ Ensure data accuracy, completeness, and consistency across all data sources.
○ Implement data validation and cleansing processes to remove duplicates, correct errors, and fill in missing values.

Functional Requirements

1. Visualization and Dashboard Creation:

○ Develop interactive dashboards that provide real-time insights into key performance indicators (KPIs) such as on-time performance, passenger demographics, and revenue metrics.
○ Create specific visualizations to address identified business problems:
● KPI visualization for the number of male passengers.
● KPI visualization for the total number of passengers.

2. User Interaction and Customization:

○ Enable users to filter data based on specific criteria (e.g., date range, flight status, passenger demographics) to gain deeper insights.

2.3 LITERATURE SURVEY

Introduction

The integration of data analytics in the airline industry has garnered significant attention in academic and industry research. This literature survey aims to provide an overview of key studies, methodologies, and findings related to the application of data analytics, particularly focusing on the use of tools like Qlik Sense for operational optimization and enhancing passenger experience.

Key Themes in Airline Data Analytics

1. Operational Efficiency:

■ Flight Delay Prediction and Management: Research by Xu et al. (2020) highlights the use of machine learning algorithms to predict flight delays based on historical data, weather conditions, and air traffic information. The study emphasizes the importance of real-time data integration for improving prediction accuracy and operational efficiency.

■ Resource Allocation: Studies by Bazargan (2016) explore the optimization of crew and aircraft scheduling using linear programming and heuristic methods. These methods aim to reduce operational costs while maintaining high service levels.

2. Customer Satisfaction and Experience:

■ Passenger Demographics and Personalization: Research by Wu et al. (2019) examines how airlines can leverage passenger demographic data to offer personalized services. The study suggests that targeted marketing and tailored in-flight services significantly enhance passenger satisfaction and loyalty.

■ Service Quality and On-Time Performance: A study by Tiernan et al. (2021) investigates the relationship between on-time performance and customer satisfaction. The findings indicate that punctuality is a critical determinant of passenger satisfaction, and airlines must prioritize minimizing delays.

3. DATA COLLECTION

3.1 COLLECT THE DATASET

Introduction

Data collection is a crucial step in any data analytics project, as it lays the foundation for subsequent analysis and insights. For the project "Exploring Insights From Synthetic Airline Data Analysis With Qlik," a comprehensive and reliable dataset is necessary to explore various aspects of airline operations and passenger behavior. The dataset used in this project has been sourced from Kaggle, a popular platform for data science and machine learning enthusiasts. This description outlines the process of acquiring, understanding, and preparing the dataset for analysis.

Source of Data

The primary dataset for this project was sourced from Kaggle, which offers a wide range of datasets contributed by the community. The specific dataset chosen includes detailed information on airline operations, passenger demographics, and flight performance. This dataset is well-suited for business analysis and provides a rich context for exploring various business insights using Qlik Sense.

Data Acquisition

To acquire the dataset from Kaggle, the following steps were taken:

1. Dataset Search and Selection:

■ Searched for relevant datasets using keywords such as "airline data," "flight performance," and "passenger demographics."

■ Evaluated multiple datasets based on the scope and relevance to the project objectives.

■ Selected a comprehensive dataset that includes key variables required for the analysis.

2. Dataset Download:

■ Downloaded the chosen dataset from Kaggle in CSV format, ensuring compatibility with data analysis tools like

Qlik Sense.

Dataset Overview

The Kaggle airline dataset includes various aspects of airline operations and passenger information. Key attributes in the dataset include:

- 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

Steps to Connect Data with Qlik Sense

- Prepare Your Dataset

○ Ensure your dataset is clean and saved in a supported format, such as CSV, Excel, or a database format. For this example, we'll assume the dataset is a CSV file.

● Open Qlik Sense

○ Launch Qlik Sense Desktop or log in to Qlik Sense Enterprise through your web browser.

● Create a New App

■ Click on the "Create new app" button.

■ Name your app (e.g., "Airline Data Analysis").

■ Click "Create", then "Open" to enter your new app.

● Add Data to Your App

■ In the newly created app, click on "Add Data".

■ Select the "Data files" option to upload your CSV file.

■ Click on "Click here to upload" and choose your CSV file from your computer.

● Upload and Prepare Data

■ Once the file is uploaded, Qlik Sense will provide a preview of the data.

■ Review the preview to ensure that the data looks correct.

■ Click "Load data" to import the data into Qlik Sense.

● Verify Data Loading

■ After loading the data, click on "Data Manager" to see the loaded table(s).

■ Verify that all columns and rows are correctly imported.

■ Rename fields if necessary for better clarity (e.g., renaming Gender to Passenger Gender).

● Create Visualizations

■ Go to "Sheet" view to start creating visualizations.

■ Click "Edit" to add charts and KPIs.

■ Use the data fields to create visualizations. For example:

○ KPI for Total Number of Passengers: Add a KPI and use the expression Count(PassengerID).

○ KPI for Number of Male Passengers: Add a KPI and use the expression Count({<Gender={'Male'}>} PassengerID).

○ Bar Chart for Top 5 Months Where Passengers Traveled the Most: Add a bar chart, use Month as the dimension, and the expression Count(PassengerID) as the measure. Apply a sort to show the top 5 months

4. DATA PREPARATION

4.1 PREPARE THE DATA FOR VISUALIZATION

Steps to Prepare Data for Visualization in Qlik Sense

1. Data Cleaning

Before loading the data into Qlik Sense, ensure that the dataset is clean and free of errors. This involves removing duplicates, handling missing values, and correcting data formats.

1. Remove Duplicates: Ensure that there are no duplicate records in your dataset.
2. Handle Missing Values: Identify and handle missing values appropriately. You can either remove rows with missing values or fill them with appropriate default values or averages.
3. Correct Data Formats: Ensure that all date fields are in a consistent date format and that numerical fields are correctly formatted.

2. Data Transformation

Transform the data to ensure it is in a suitable format for analysis and visualization. This might include creating new calculated fields, renaming columns for clarity, and aggregating data as needed.

1. Rename Columns: Rename columns for better clarity. For example:
   - Gender -> Passenger Gender
   - ScheduledDeparture -> Scheduled Departure
   - ScheduledArrival -> Scheduled Arrival
2. Create Calculated Fields: Add any calculated fields that might be necessary for your analysis. For example, calculating the delay duration:
   - Delay Duration (minutes): Calculate the delay in minutes by subtracting the scheduled departure time from the actual departure time.
3. Aggregate Data: Aggregate data as needed. For example, to find the number of flights per month, you might need to aggregate the flight data by month.

5. DATA VISUALIZATIONS

5.1 VISUALIZATIONS

In the project "Exploring Insights From Synthetic Airline Data Analysis With Qlik," various data visualizations were used to uncover insights and present data effectively. Below are
descriptions of the different types of visualizations used, including KPIs, vertical grouped bar
charts, vertical stacked bar charts, tree maps, and horizontal grouped bar charts.

1. Key Performance Indicators (KPIs)

Purpose: KPIs are used to provide quick, at-a-glance information on critical metrics. They
highlight key statistics and trends essential for decision-making.

Examples:

- Total Number of Passengers: A KPI showing the total count of passengers.
  - Measure: Count(PassengerID)
- Number of Male Passengers: A KPI displaying the count of male passengers.
  - Measure: Count({<Gender={'Male'}>} PassengerID)

2. Vertical Grouped Bar Chart

Purpose: Vertical grouped bar charts compare multiple categories within a single dimension,
allowing for an easy comparison between different groups.

Example:

- Top 5 Months Where Passengers Traveled the Most: This chart shows the number of passengers for the top 5 months.
  - Dimensions: Month
  - Measures: Count(PassengerID)

■ Sorting: Sort by passenger count to display the top 5 months.

3. Vertical Stacked Bar Chart

Purpose: Vertical stacked bar charts display the total and relative contributions of different
categories to a whole. This type of chart is useful for showing parts of a whole and comparing
the proportional contributions.

Example:

● Monthly Passenger Count by Travel Class: This chart shows the number of passengers
for each month, stacked by travel class.

■ Dimensions: Month, TravelClass

■ Measures: Count(PassengerID)

■ Stacking: Stack by TravelClass to show the distribution within each month.

4. Tree Map

Purpose: Tree maps are used to represent hierarchical data as a set of nested rectangles. Each
rectangle's size is proportional to a specific data value, making it easy to compare relative sizes.

Example:

● Revenue by Travel Class and Ancillary Services: This tree map shows the revenue generated from ticket sales and ancillary services, broken down by travel class.

■ Dimensions: TravelClass, RevenueType

■ Measures: Sum(TicketPrice + AncillaryRevenue)

5. Horizontal Grouped Bar Chart

Purpose: Horizontal grouped bar charts provide a clear way to compare categories across
multiple groups. They are especially useful for comparing values across different groups when
there are long category labels.

Example:

● Flight Status by Airline: This chart shows the number of on-time, delayed, and canceled
flights for each airline.

■ Dimensions: Airline, FlightStatus

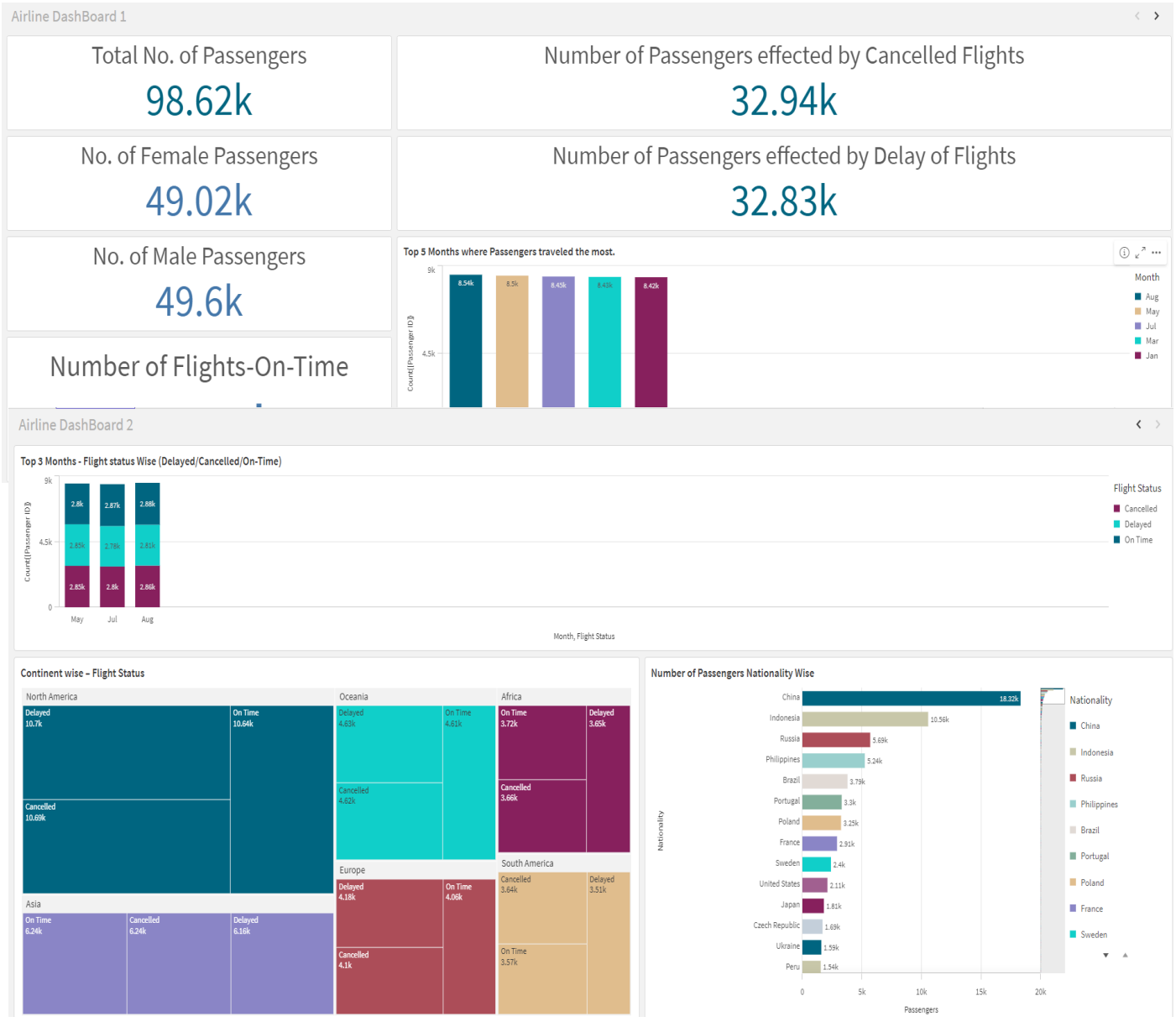
■ Measures: Count(FlightNumber)

■ Grouping: Group by 'FlightStatus' to compare different statuses within each

airline.

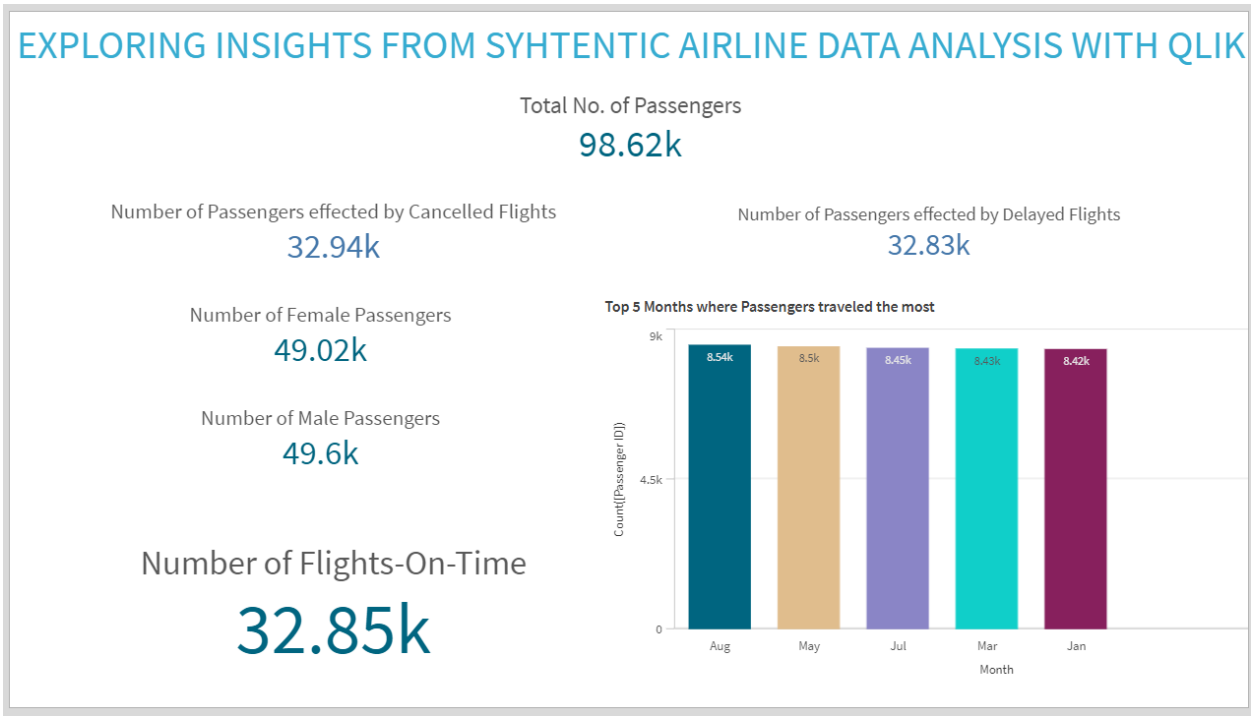
6. DASHBOARD

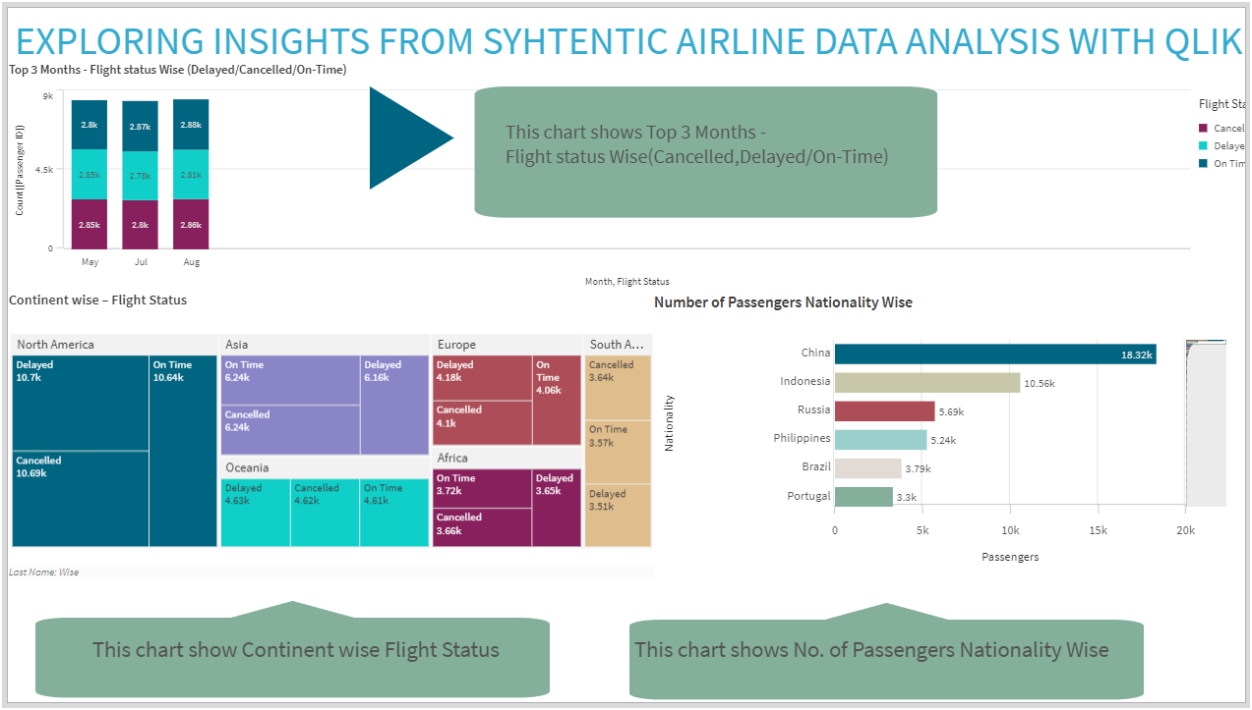
6.1 RESPONSIVE AND DESIGN OF DASHBOARD



7.REPORT

7.1 REPORT CREATION

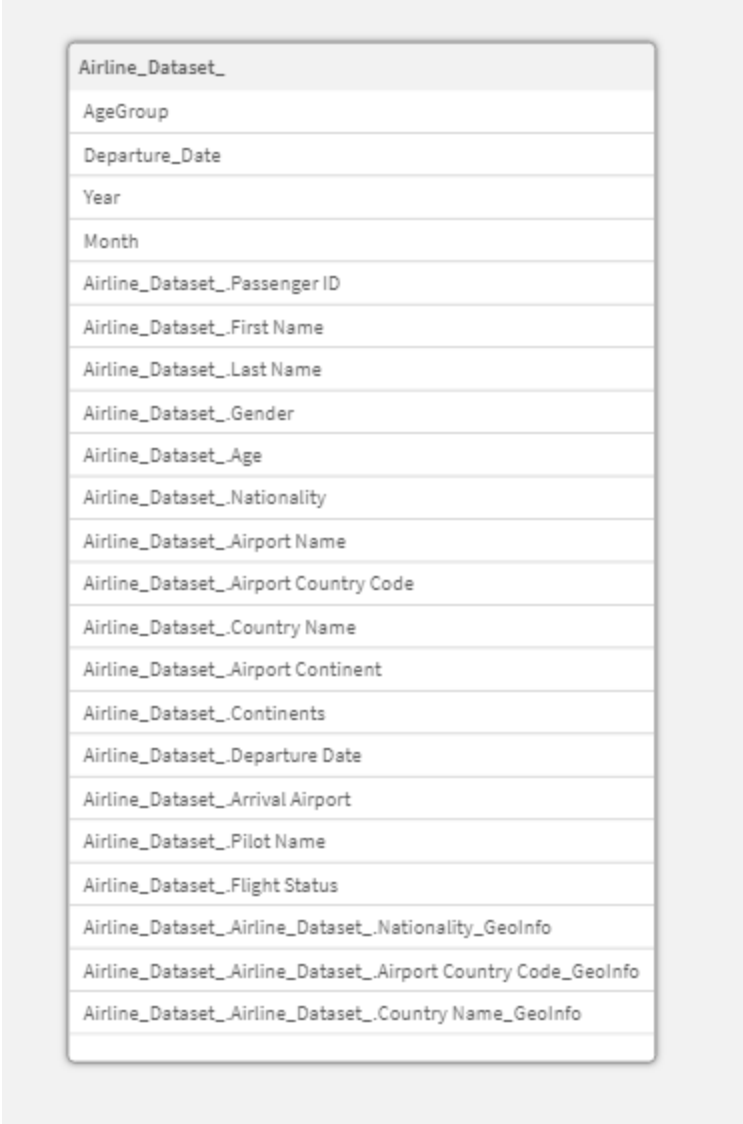




8.PERFORMANCE TESTING

8.1 AMOUNT RENDERED

"Amount of Data Loaded" refers to the quantity or volume of data that has been imported, retrieved, or loaded into a system, software application, database, or any other data storage or processing environment. It's a measure of how much data has been successfully processed and made available for analysis, manipulation, or use within the system.



8.2 UTILIZATION OF DATA FILTERS

"Utilization of Filters" refers to the application or use of filters within a system, software application, or data processing pipeline to selectively extract, manipulate, or analyze data based on specified criteria or conditions. Filters are used to narrow down the scope of data, focusing only on the relevant information that meets certain predefined criteria.