Exploring Insights From Synthetic Airline Data Analysis With Qlik

1. INTRODUCTION

1.1 OVERVIEW:

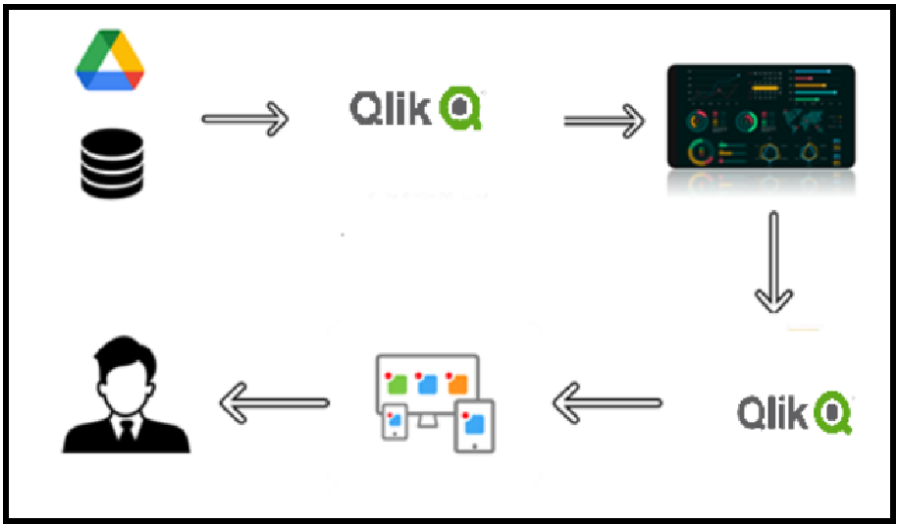
This document provides a detailed guide on "Exploring insights from synthetic airline data using Qlik Sense". The project involves analyzing various aspects of airline operations, performance, and customer satisfaction through interactive visualizations.

1.2 PURPOSE:

The goal of this project is to create a comprehensive dashboard in Qlik Sense that provides insights into the airline industry's performance using synthetic data. In this project, the synthetic airline data simulates various aspects of airline operations, including flight schedules, passenger demographics, ticket sales, and performance metrics. The objective is to leverage Qlik's analytical capabilities to uncover patterns, trends, and correlations within this data, aiding in decision-making processes for airlines, airports, and related stakeholders.

1.3 TECHNICAL ARCHITECHTURE: A diagram showing the major components and their interactions.Components are

- Data Sources
- Data Ingestion and Storage
- Qlik Sense Environment
- User Interface



2. PROBLEM UNDERSTANDING

2.1 SPECIFY THE BUSINESS PROBLEM:

The key business problems are

- **Flight Performance Optimization**:Frequent flight delays and cancellations negatively impact customer satisfaction and airline reputation.
- **Passenger Analysis:**Lack of detailed understanding of passenger travel trendsand preferences.
- **Revenue Management**:Inefficient revenue management leading to potential revenue losses.
- **Operational Efficiency:**Inefficiencies in flight operations, such as prolonged turnaround times and resource allocation issues.
- **Customer Satisfaction and Experience** : Insufficient insights into factors affecting customer satisfaction and loyalty.

2.2 BUSINESS REQUIREMENTS:

- Data Integration and Preparation
- Performance Analysis
- Operational Efficiency
- Customer Experience
- Interactive Dashboards and Reports
- Predictive Analysis

By focusing on these requirements, the project aims to provide comprehensive insights into airline operation, performance, and customer experience.

2.3 LITERATURE SURVEY:

The use of synthetic airline data for analysis offers a versatile and powerful tool for exploring various aspects of airline operations, passenger behavior, and financial performance. Continued advancements in data generation and analysis techniques will further enhance the ability to derive actionable insights from synthetic data, benefiting both researchers and industry practitioners.

3. DATA COLLECTION

3.1 COLLECT THE DATA:

Source Data: Obtain a synthetic airline dataset. Ensure it includes relevant fields such as Date, Airline, PassengerCount, FlightDelay,Destination, etc.
Format: The dataset can be in formats like CSV, Excel, or a database table.

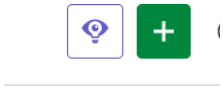
Structure of the Dataset

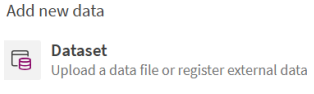
Passenger ID	First Name	Last Name	Gender	Age	Nationality	Airport Name	Airport Country Code	Country Name	Airport Continent	Continents	Departure Date	Arrival Airport	Pilot Name	Flight Status
fs4OZI	Allan	Prime	Male	49	Philippines	Mid-Carolina Regional Airport	US	United States	NAM	North America	6/23/2022	SRW	Ulrick Tutchings	Cancelled
urqtZB	Conrad	Vaun	Male	15	China	Alcides Fernández Airport	CO	Colombia	SAM	South America	02-02-2022	ACD	Giulietta Harler	On Time
Ym0iup	Carmela	Bridal	Female	36	Australia	Cataratas Del Iguazú International Airport	AR	Argentina	SAM	South America	2/14/2022	IGR	Pennie Rizzotto	Cancelled
uTCmlG	Welbie	Shorrock	Male	41	Cameroon	Vatulele Airport	FJ	Fiji	OC	Oceania	9/26/2022	VTF	Griffin Cowey	On Time
HVmTqS	Waldon	Deverale	Male	80	Vietnam	Coen Airport	AU	Australia	OC	Oceania	05-09-2022	CUQ	Oralie Reisenberg	Delayed

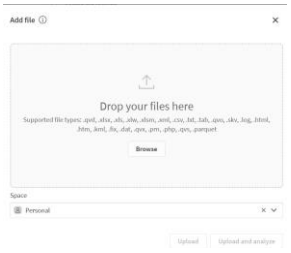
The dataset prominently incorporates fields such as Passenger ID, First Name, Last Name, Gender, Age, Nationality, Airport Name, Airport Country Code, Country Name, Airport Continent, Continents, Departure Date, Arrival Airport, Pilot Name, and FlightStatus.

3.2 CONNECT DATA WITH QLIK SENSE:

After collecting the dataset ,Open Qlik Sense and create a new app.Click + symbol ont he top right corner.Click "add dataset".Click "Upload data file".click on "Browse" and add the required dataseet from your device.Then click on "upload andanalyze".Then the new app will be created to create your visualisations.





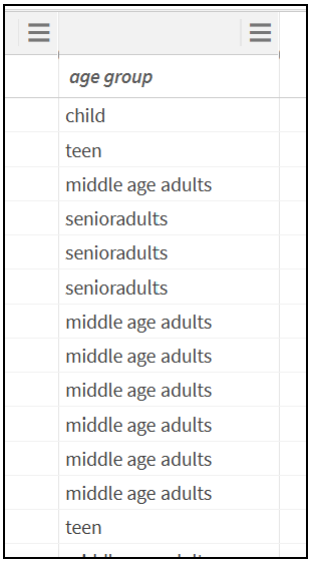


4. DATA PREPARATION

4.1 PREPARING 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 performanceand efficiency.

In the app,go to data manager and then select data load editor and embed the code required to create new columns in the dataset (or) go to the edit option in data manager and select "Add field" and add "calculated field" and select the name for the coloumn and give the expression on which you want to calculate the data of the column. From the dataset,I created "age group" column with the expression : if(Age<4,'toodler',
if(Age<12,'child',
if(Age<19,'teen',
if(Age<30,'adults',
if(Age<=59,'middle age adults',
if(Age>60,'senioradults'))))))
This created a new column in the dataset.



age group
child
teen
middle age adults
senioradults
senioradults
senioradults
middle age adults
middle age adults
middle age adults
middle age adults
middle age adults
middle age adults
teen

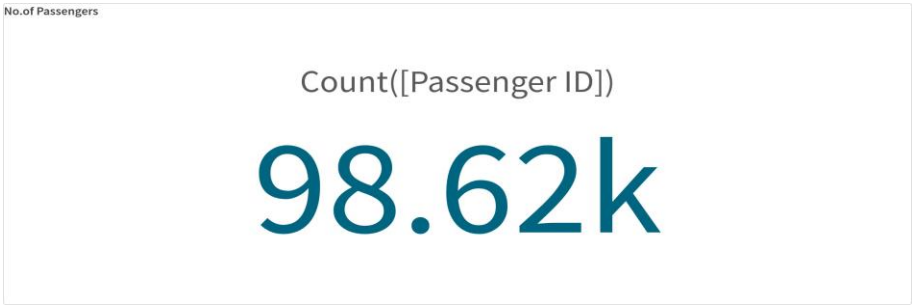
4.1 Creation of age group column for data preparation

5. DATA VISUALISATIONS

5.1 VISUALISATIONS:

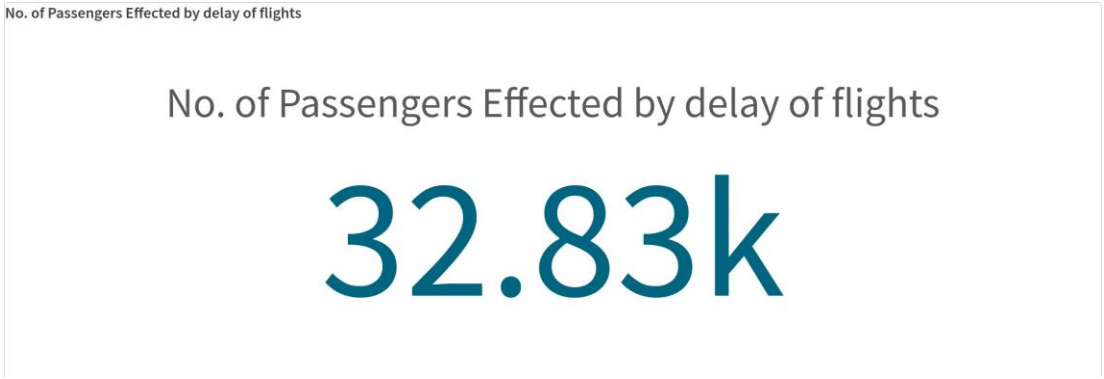
Data visualization refers to the representation of data in graphical formats such as charts, graphs, dashboards, or infographics. This makes it easier for you to see trends, recognize relationships, and uncover data-driven insights from large, complex data sets. These insights can increase efficiency, revenue, and profits for your organization.

ACTIVITY 1: TOTAL NO.OF PASSENGERS USING KPI



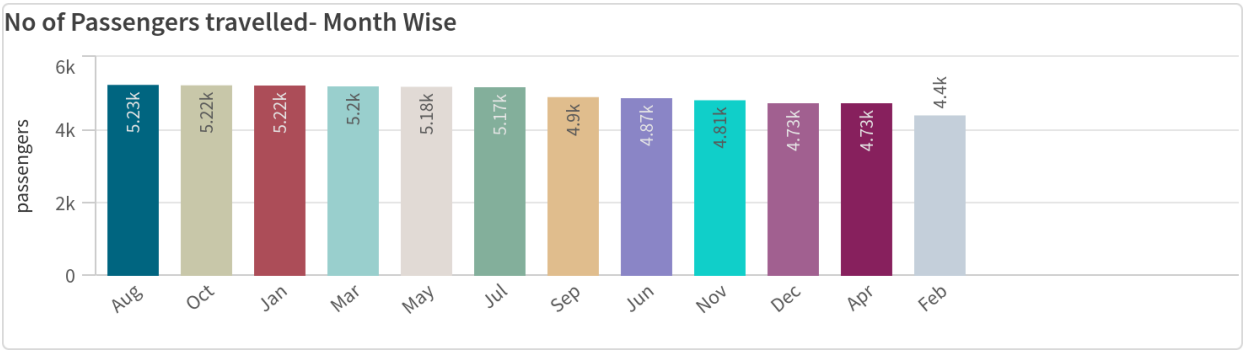
ACTIVITY 2: NUMBER OF PASSENGERS EFFECTED BY CANCELLED FLIGHTSUSING KPI

ACTIVITY 3: NUMBER OF PASSENGERS EFFECTED BY DELAY OF FLIGHTUSING KPI



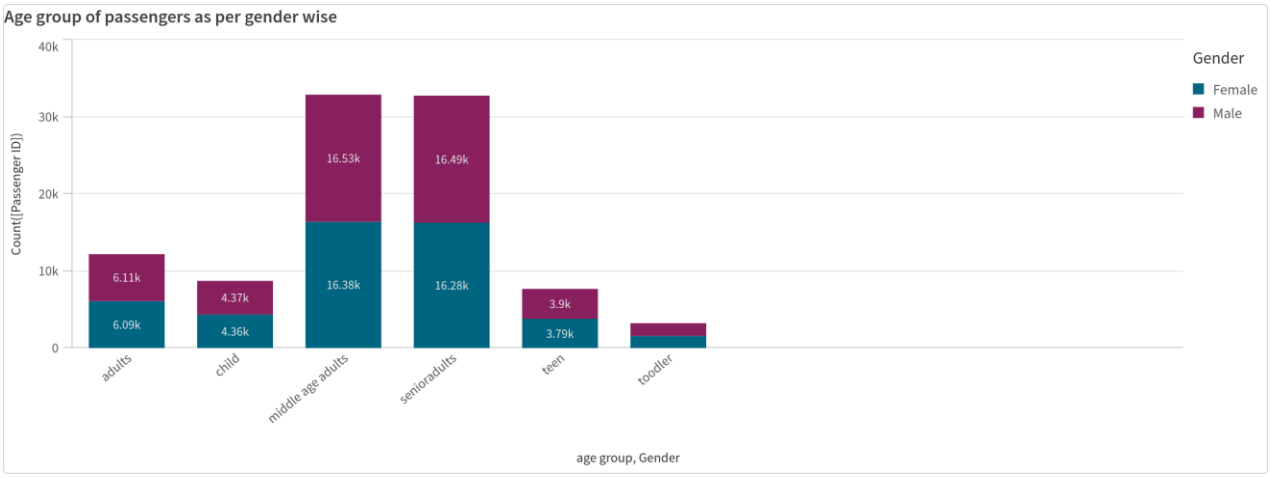
ACTIVITY 4: NUMBER OF FLIGHTS ON TIME USING KPI

ACTIVITY 5: NUMBER OF PASSENGERS TRAVELLED MONTH WISE USING BARCHART

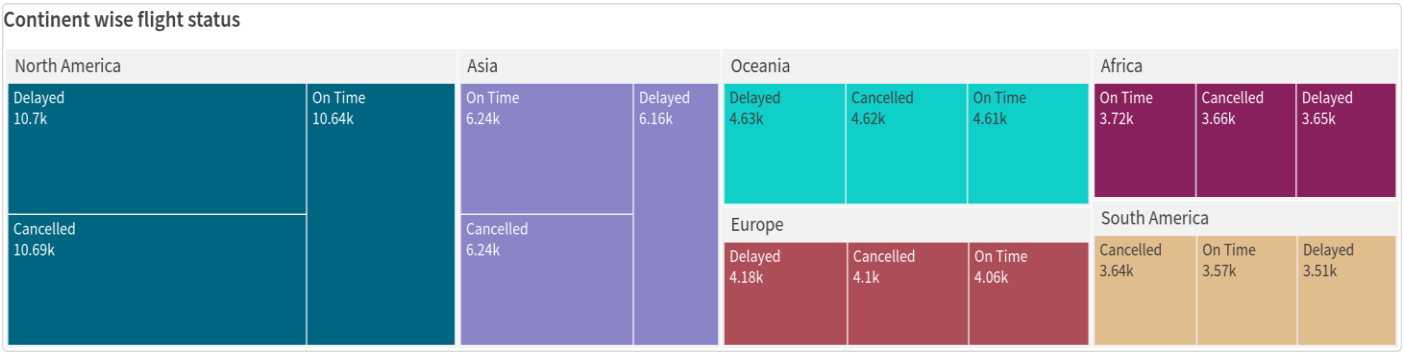


ACTVITY 6: TOP 3 MONTHS FLIGHT STATUS WISE USING STACK BAR CHART

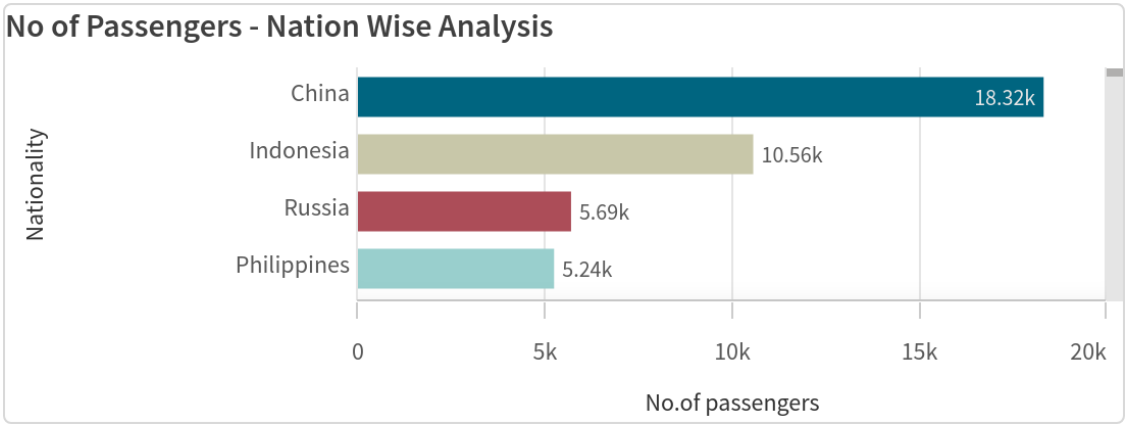
ACTIVITY 7: AGE GROUP OF PASSENGERS AS PER GENDER WISE USING BARCHART STACK



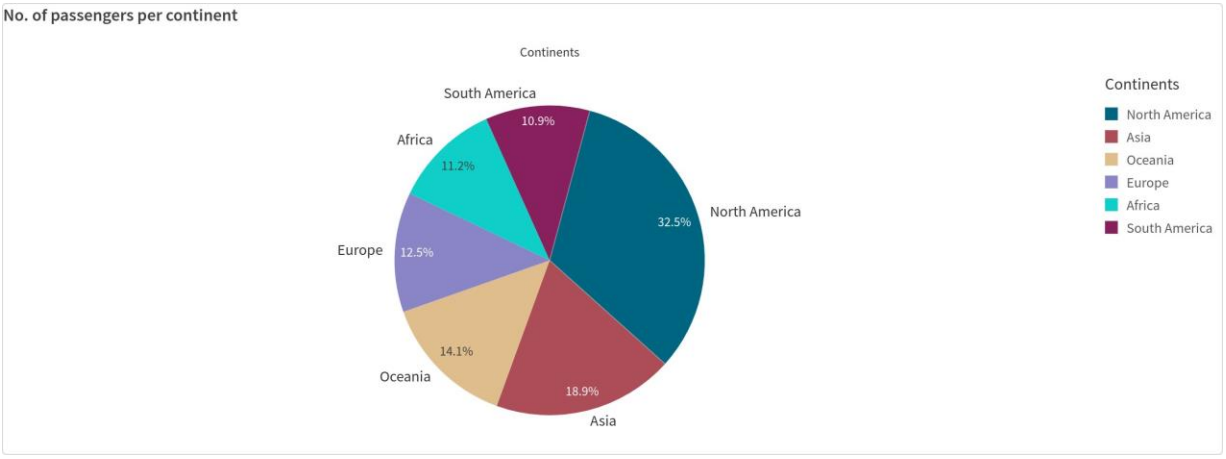
ACTIVITY 8: CONTINENT WISE FLIGHT STATUS USING TREE MAP



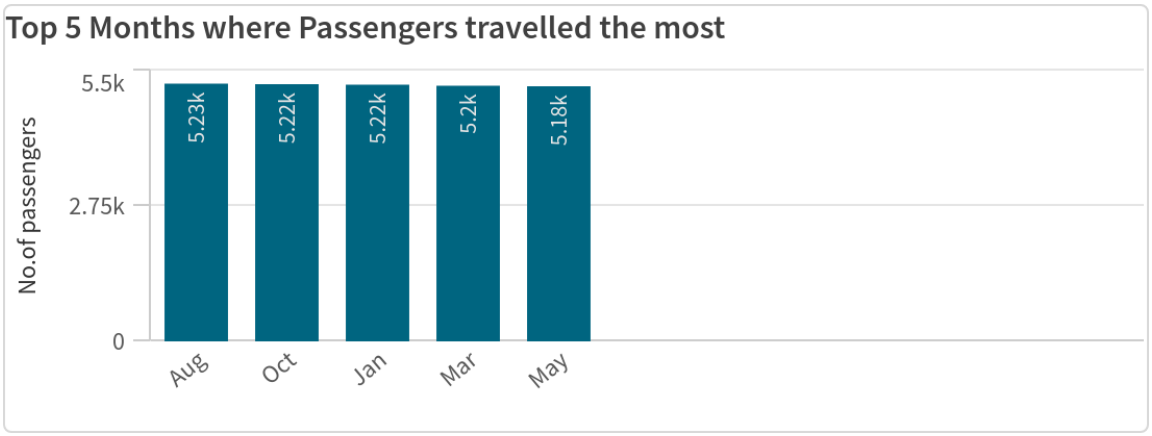
ACTIVITY 9: NUMBER OF PASSENGERS _ NATION WISE ANALYSIS USINGHORIZANTAL BAR CHART



ACTIVITY 10: NUMBER OF PASSENGERS PER CONTINENT USING PIE CHART

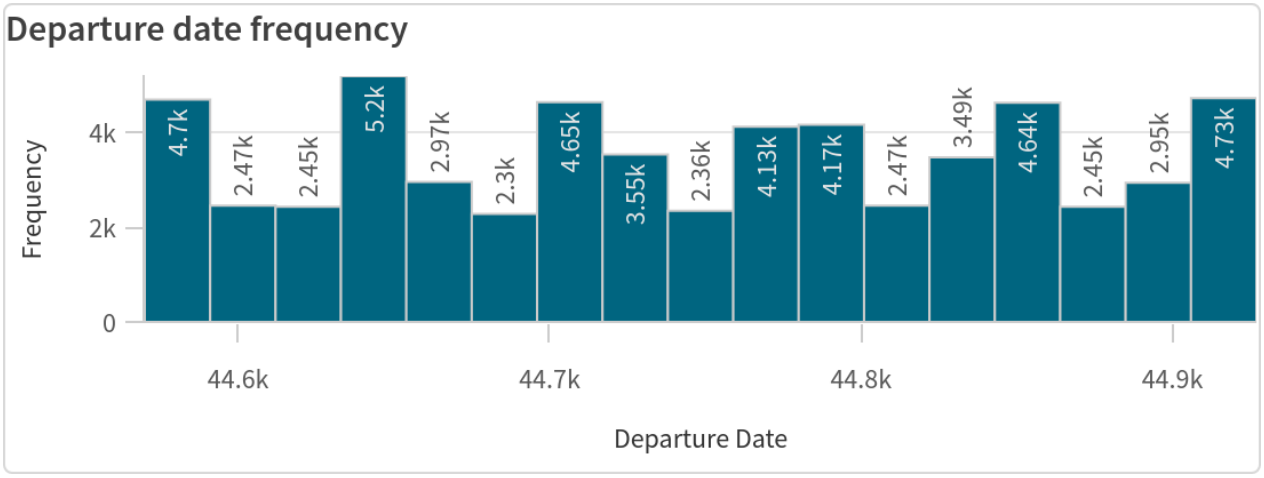


ACTIVITY 11: TOP 5 MONTHS WHERE PASSENGERS TRAVELLED THE MOSTUSING BAR CHART

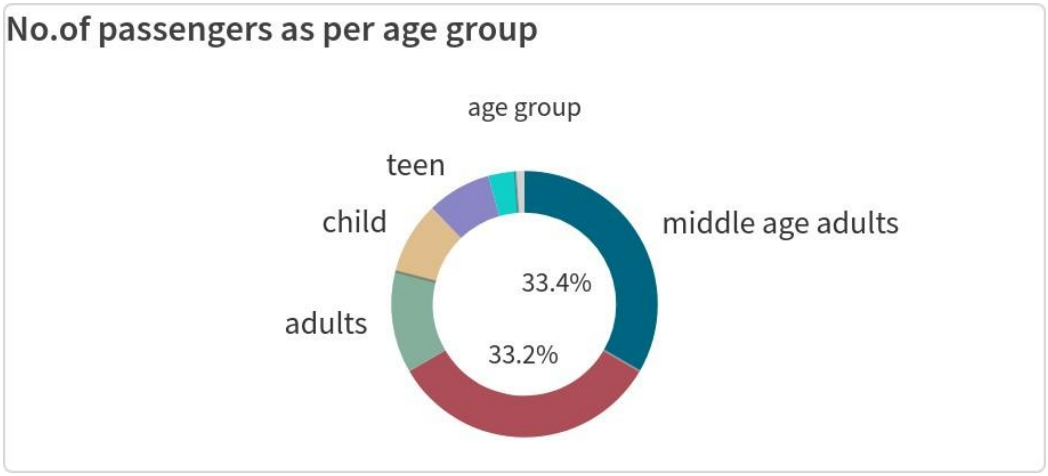


ACTIVITY 12: AIRPORT CONTINENTS IN MAP

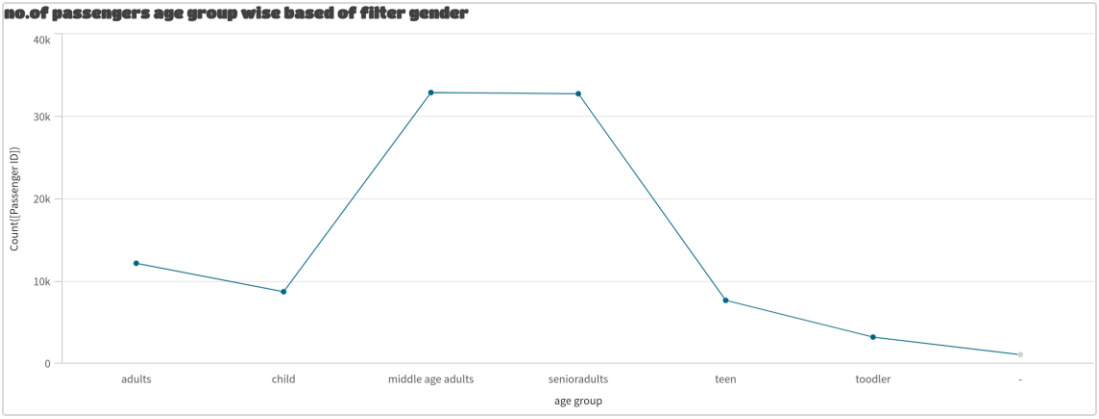
ACTIVITY 13: DEPARTURE DATE FREQUENCY USING HISTOGRAM



ACTIVITY 14: NUMBER OF PASSENGERS AS PER AGE GROUP USING PIECHART

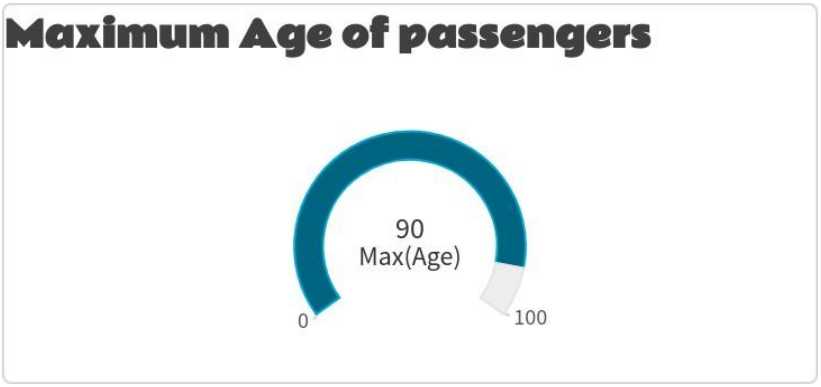


ACTIVITY 15: NO.OF PASSENGERS AS PER GENDER USING LINE CHART



ACTIVITY 16: PERCENTAGE OS PASSENGERS PER CONTINENT-FLIGHT STATUSWISE USING TREE MAP

ACTIVITY 17: MAXIMUM AGE AMONG PASSENGERS USING GAUGE RADIALCHART



ACTIVITY 18: MINIMUM AGE AMONG PASSANGERS USING GAUGE BAR CHART

ACTIVITY 19: TOTAL NUMBER OF COUNTIRES AND COUNTRY HAVINGMAXIMUM PASSENGERS USING NL INSIGHTS

No.of countries and Country having max passengers

Calculated measure (KPI)

- The count of passengers is 98.62k.

Ranking

- The count of passengers is 98619.

Breakdown (geospatial)

- Count of passengers is 98619 across 235 Airport Country Code.
- Top passengers is 22104 where Airport Country Code is US.

ACTIVITY 20: MONTHS HAVING MAXIMUM AND MINIMUM PASSENGERS USINGNL INSIGHTS

Months of max and min passangers

Calculated measure (KPI)

- The count of passangers is 98.62k.

Trend over time

- The count of passangers is 98619.
- The maximum is 38961 where Month is Aug.
- The minimum is 4398 where Month is Feb.

Mutual information

- The mutual dependence between Passenger ID and Month is 99.65%.

ACTIVITY 21: TOTAL NUMBER OF NATIONALITIES AND NATIONALITY HAVINGMAXIMUM PASSENGERS USING NL INSIGHTS

Total no.of nationalities

Calculated measure (KPI)

- The count of passengers is 98.62k.

Ranking

- The count of passengers is 98619.

Breakdown (geospatial)

- Count of passengers is 98619 across 240 Nationality.
- Top passengers is 18317 where Nationality is China.

Mutual information

- The mutual dependence between Passenger ID and Nationality is 99.97%.

INFORMATION ABOUT THE VISUALISATIONS I USED:

- **KPI:** In a KPI visualization, you can have one or two measures and no dimensions. With two measures, the second value automatically becomes a complementary value and is shown with a smaller font size.

- **BARCHART:** Grouping and stacking bars makes it easy to visualize grouped data. Thebar chart is also useful when you want to compare values side by side.

- **TREE MAP:** Treemaps display hierarchical data by using nested rectangles, thatis, smaller rectangles within a larger rectangle.

- **PIE CHARTS**:The pie chart displays the relation between values as well as therelation of a single value to the total.

- **HISTROGRAM:**The histogram is suitable for visualizing distribution of numericaldata over a continuous interval, or a certain time period.

- **MAP CHART**:Maps have many ways to present your data. You can add multiplelayers to your map to display different types of information on the same map.

- **GAUGE**:The gauge is designed to show a single measure value and visualize howto interpret that value.

- **NL INSIGHTS**:The NL insights visualization displays natural language insightsand information about selected dimensions and measures.

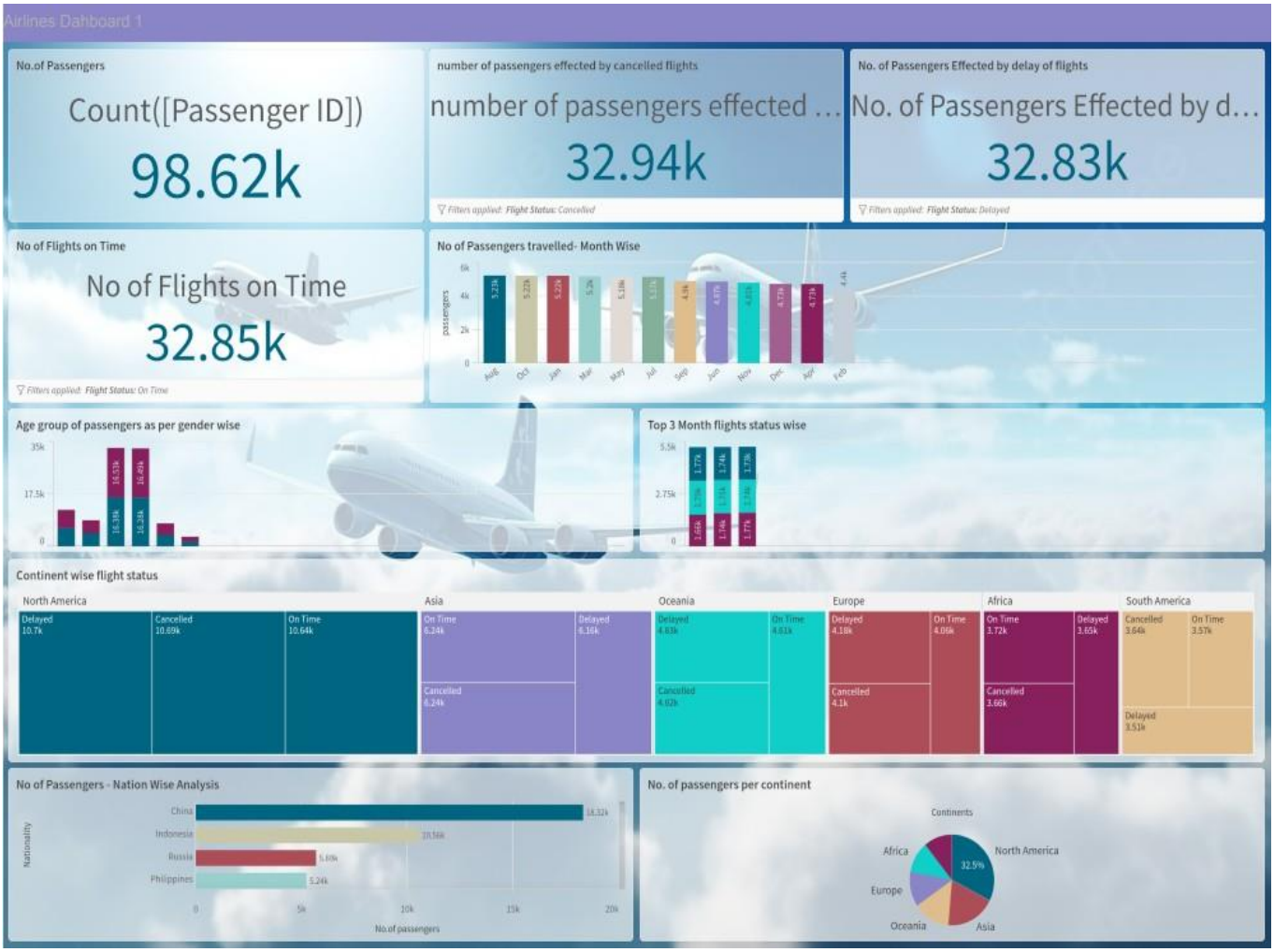
- **LINE CHART:**The line chart is used to show trends over time. The dimension isalways on the x-axis, and the measures are always on the y-axis

6. DASHBOARD

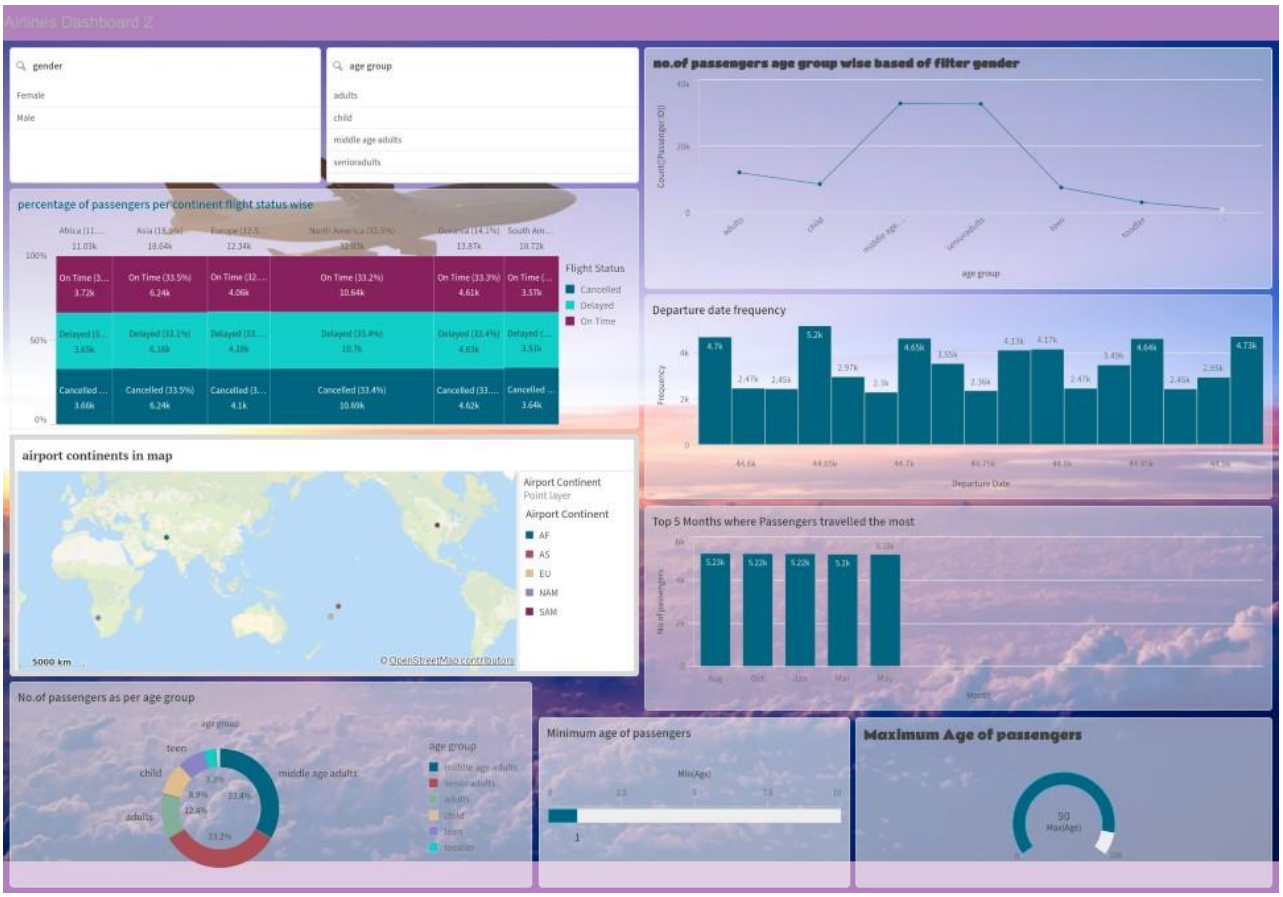
A dashboard is a graphical user interface (GUI) that displays information and data in an organized, easy-to-read format. Dashboards are often used to provide real-time monitoring and analysis of data and are typically designed for a specific purpose or use case. Dashboards can be used in a variety of settings, such as business, finance, manufacturing, healthcare, and many other industries. They can be used to track key performance indicators (KPIs), monitor performance metrics, and display data in the form of charts, graphs, and tables.

6.1 RESPONSIVE AND DESIGN OF DASHBOARD:

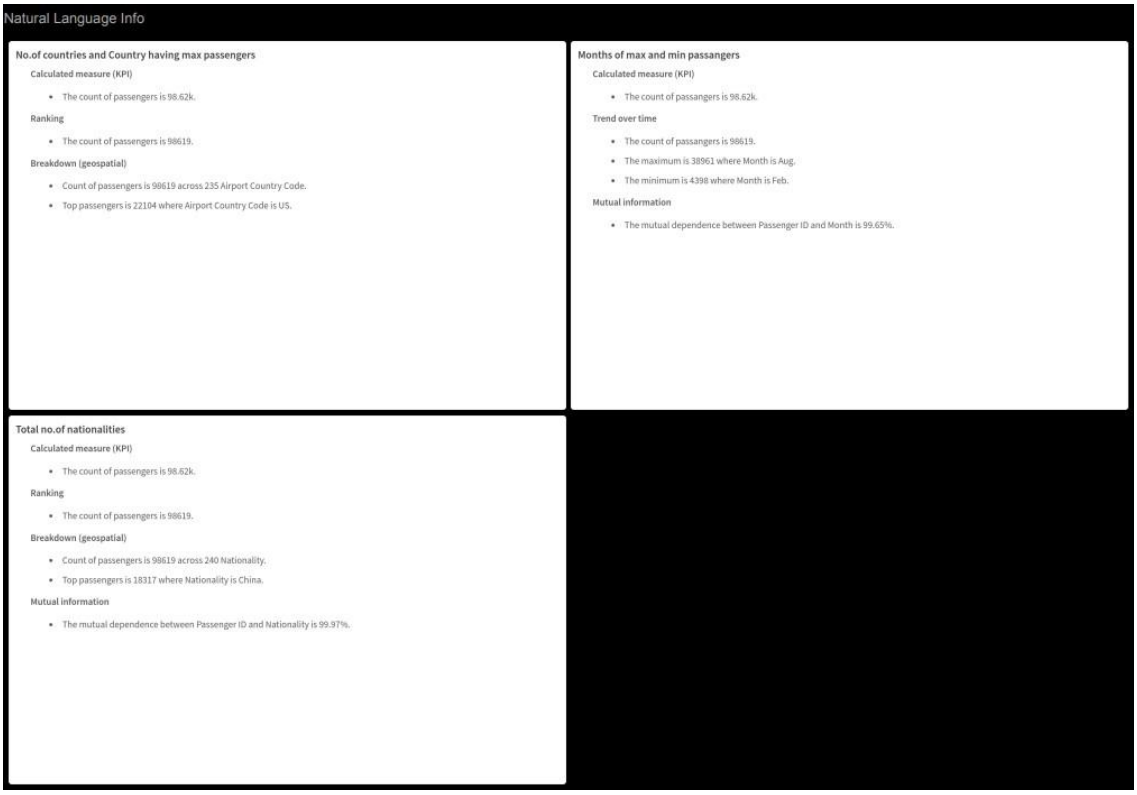
DASHBOARD 1:



DASHBOARD 2:



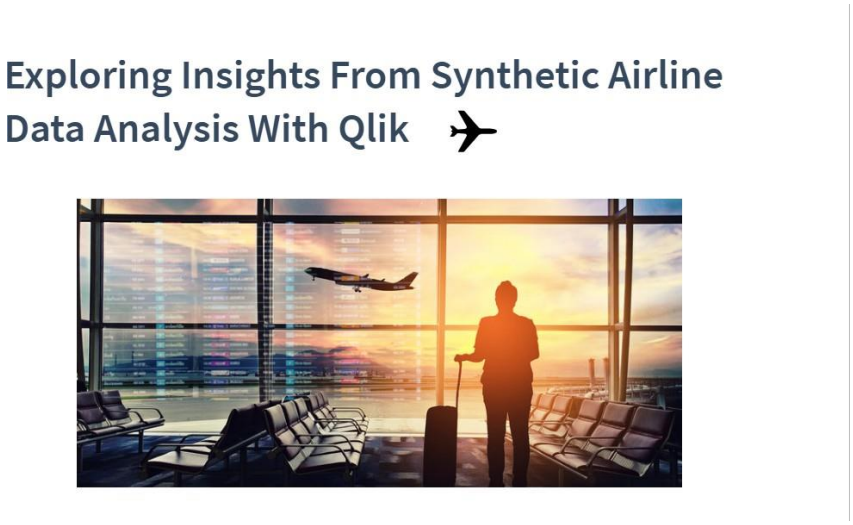
DASHBOARD 3:

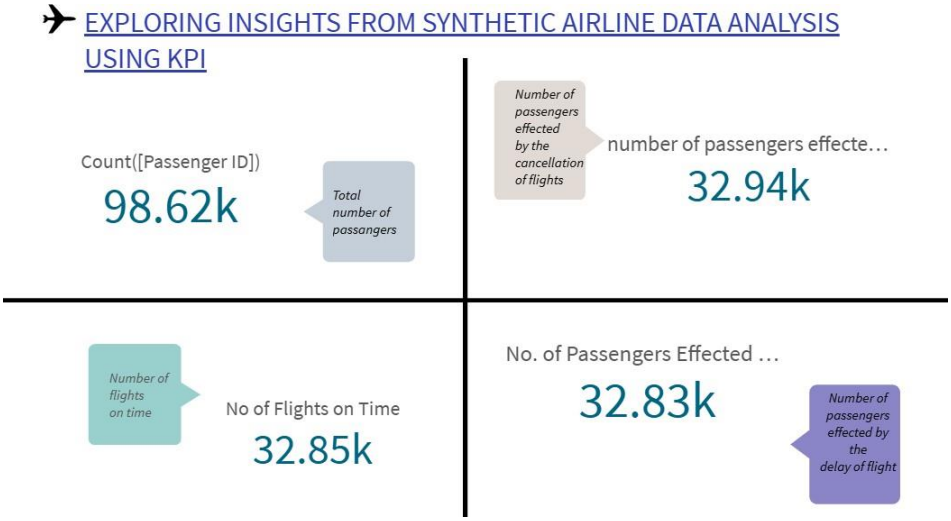


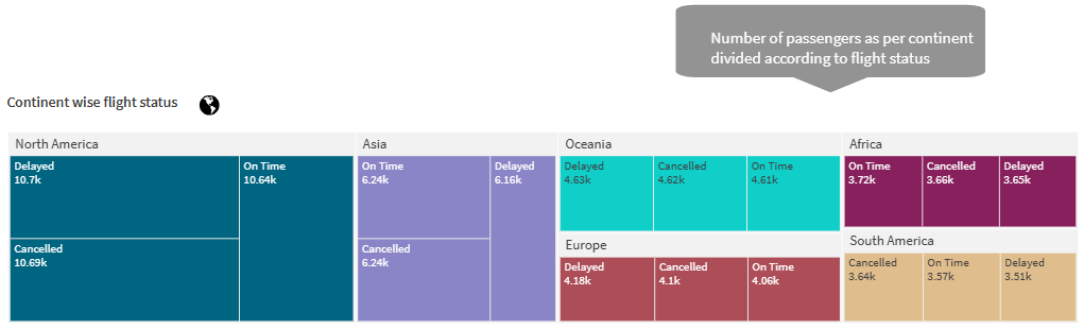
7. STORY

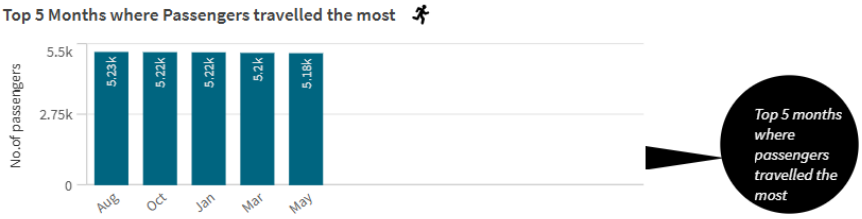
7.1 STORY CREATION:

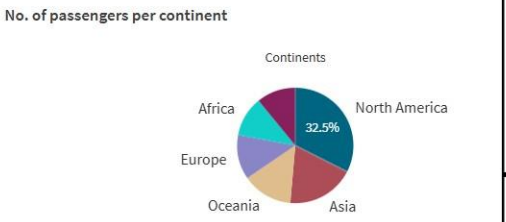
A data story is a way of presenting data and analysis in a narrative format, with the goal of making the information more engaging and easier to understand. A data story typically includes a clear introduction that sets the stage and explains the context for the data, a body that presents the data and analysis in a logical and systematic way, and a conclusion that summarizes the key findings and highlights their implications. Data stories can be told using a variety of mediums, such as reports, presentations, interactive visualizations, and videos.

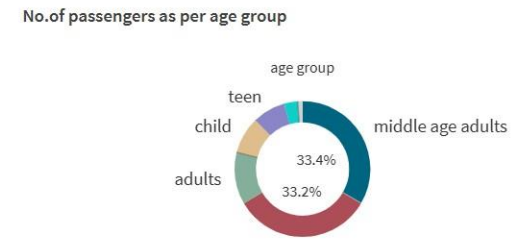


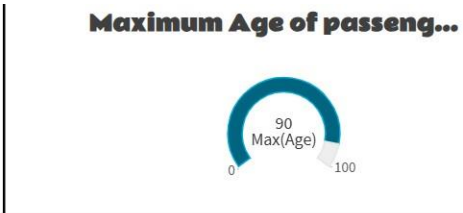


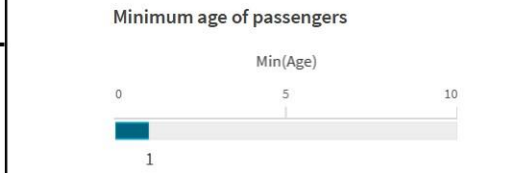








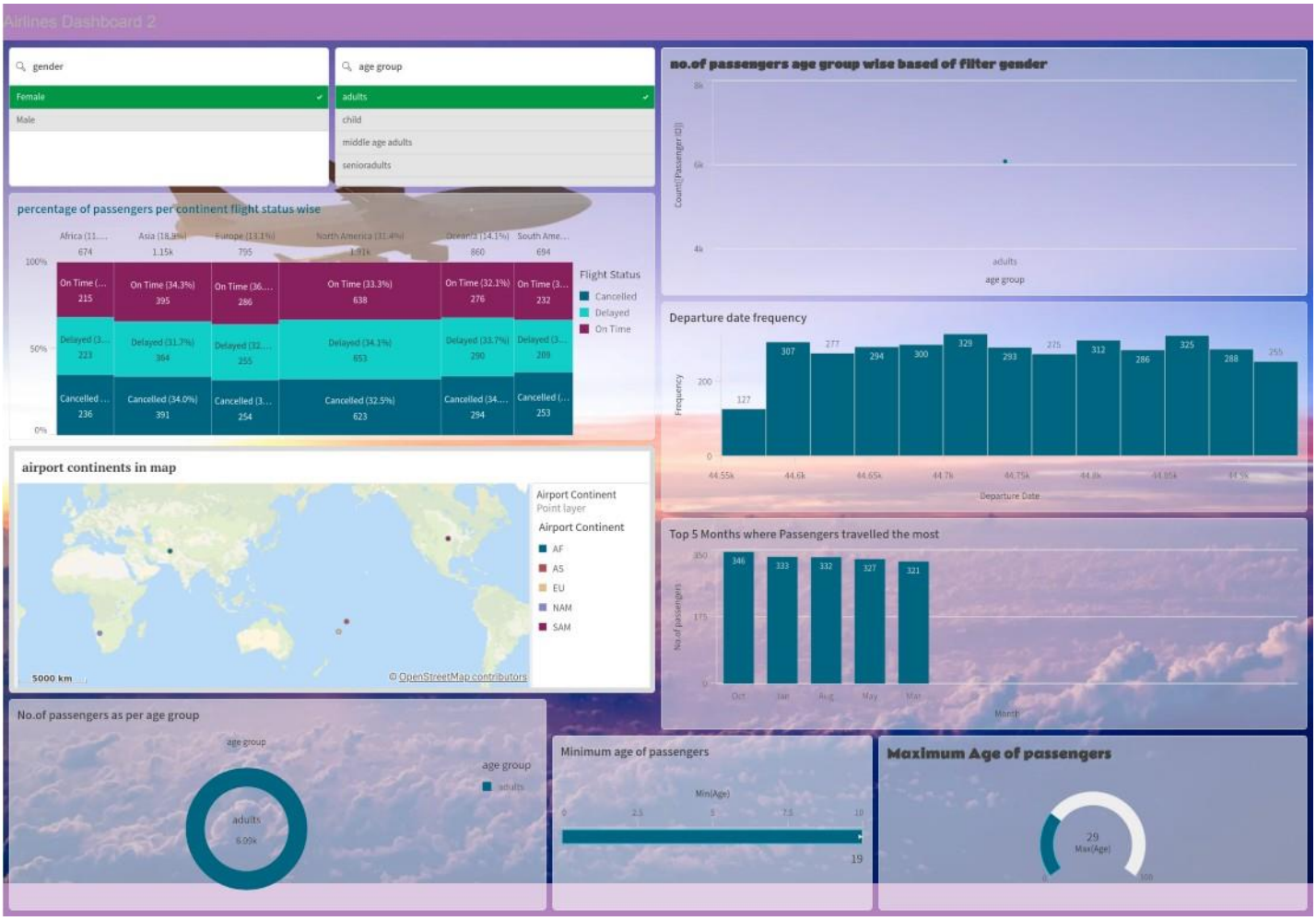




The number of male or female passengers
female- click here
male-click here

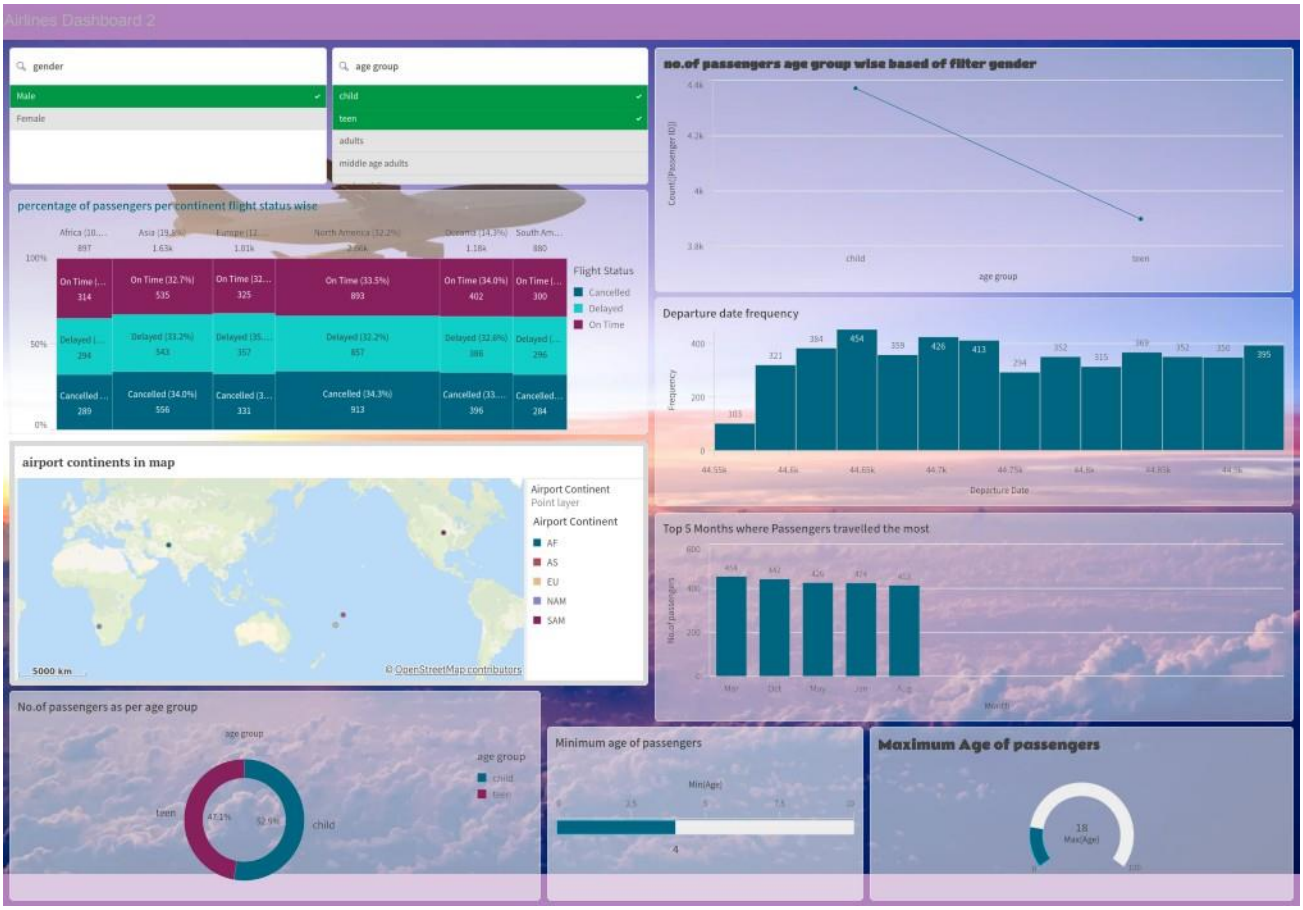
8. PERFORMANCE TESTING

8.1 UTILISATION OF DATA FILTERS



-In this dashboard, I used 2 filters on the basis of gender and age group

-According to female gender and age group adult, the visualisations give the above data.



-The dashboard having filters gender and age group.

-The gender is "male" and taken two age groups "child" and "teen" gives the following information in the view as visualizations.