Exploring Insights from Synthetic Airline Data Analysis with Qlik

1. Introduction

1.1 Overview

The "Exploring Insights from Synthetic Airline Data Analysis with Qlik" project aims to utilize synthetic airline data to derive valuable insights using Qlik, a business intelligence and data visualization tool. The project involves analyzing various aspects of airline operations, including flight schedules, passenger demographics, ticket sales, and performance metrics, to uncover patterns, trends, and correlations that can aid in decision-making processes for airlines, airports, and related stakeholders.

1.2 Purpose

The purpose of this project is to demonstrate the capabilities of Qlik in analyzing synthetic airline data to:

- Optimize revenue by identifying peak travel times, popular destinations, and pricing strategies.
- Enhance operational efficiency by identifying bottlenecks in airport operations and predicting peak traffic periods.
- Improve customer experience by understanding customer preferences, satisfaction levels, and pain points

The purpose of this project is to explore insights from synthetic airline data using Qlik, a powerful data visualization and analytics platform. This document provides a comprehensive overview of the project, including data preparation, visualization creation, and insights derived from the analysis.

1.3 Technical Architecture

1. **Data Source:** Synthetic airline data simulating various aspects of airline operations, including flight schedules, passenger demographics, ticket sales, and performance metrics.

2. **Data Integration:** Qlik's data integration capabilities will be used to integrate the synthetic airline data with other relevant data sources, such as customer feedback data.

3. **Data Analysis:** Qlik's analytical capabilities will be used to analyze the integrated data, including data visualization, filtering, and drill-down capabilities.

4. **Data Visualization:** Qlik's data visualization capabilities will be used to create

interactive dashboards and reports to present insights and trends to stakeholders.

5. **Scenario-Based Analysis:** The project will involve scenario-based analysis, where Qlik will be used to analyze specific business scenarios, such as revenue optimization, operational efficiency, and customer experience enhancement.

6. **Qlik Platform:** The project will utilize Qlik's business intelligence and data visualization platform, including Qlik Sense, QlikView, or Qlik Cloud.
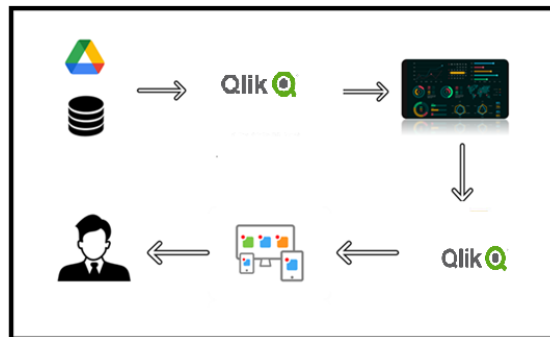
2. Define Problem / Problem Understanding

2.1 Business Problem

i. **Revenue Optimization:** Airlines struggle to optimize their revenue due to lack of insights into historical ticket sales data, peak travel times, popular destinations, and pricing strategies.

ii. **Operational Efficiency:** Airports face challenges in enhancing operational efficiency due to inadequate analysis of flight schedules, passenger flows, and luggage handling processes, leading to bottlenecks and inefficiencies.

iii. **Customer Experience Enhancement:** Airlines find it difficult to enhance the passenger experience due to limited understanding of customer preferences, satisfaction levels, and pain points, resulting in poor customer loyalty and satisfaction.

2.2 Business Requirements

- Analyze historical ticket sales data to identify trends and patterns.
- Segment customers based on purchasing behavior to develop targeted pricing strategies.
- Visualize revenue trends over time to inform business decisions.
- Identify bottlenecks in airport operations and predict peak traffic periods.
- Allocate resources effectively to streamline processes and improve overall efficiency.

2

2.3 Literature Survey

The literature survey highlights the importance of business intelligence and data analytics in the airline industry. Studies have shown that the use of data analytics can improve revenue optimization, operational efficiency, and customer experience enhancement.

    i.   A study by Microsoft found that 56% of small business owners believe their businesses could gain efficiency through AI solutions, including analyzing customer data and trends.

    ii.   Another study by Optimum Business emphasized the role of AI in driving improved outcomes for small businesses, including automating repetitive tasks and enhancing customer service.

    iii.   Research by Sprout Social demonstrated the effectiveness of AI-powered social media analytics in improving social media performance and optimizing content strategies.

These studies underscore the potential of Qlik's analytical capabilities in uncovering patterns, trends, and correlations within synthetic airline data, ultimately aiding in decision-making processes for airlines, airports, and related stakeholders.

3. Data Collection

3.1 Collect the dataset

Data collection is the process of systematically gathering and measuring information on variables of interest to answer research questions, test hypotheses, evaluate outcomes, and generate insights. For this project, data was collected to explore and analyze passenger demographics and flight information from an airline dataset.

The main objectives of data collection for this project are:

- To gather comprehensive information on passengers and flights to analyze and understand passenger demographics.

- To identify patterns and trends in passenger behavior and flight operations.

- To generate insights that can help improve airline services, customer satisfaction, and operational efficiency.

The dataset for this project was obtained from Kaggle, a popular platform for data science and machine learning competitions. The dataset contains detailed information about airline passengers, flights, and related attributes.

The data collection process involved the following steps:

1. Identify Data Requirements: Determine the key variables and attributes needed for the analysis.

2. Data Acquisition: Download the dataset from Kaggle.

3. Data Understanding: Review the dataset to understand the structure, types of data, and the information provided in each column.

4. Data Extraction: Extract relevant data fields for further analysis and visualization.

The dataset includes the following columns, each providing specific information about passengers and flights:

- 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 where the airport is located.

- 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

Connecting and preparing data in Qlik Sense is a crucial step in ensuring that the dataset is ready for analysis and visualization. This section outlines the process of loading the dataset into Qlik Sense and verifying its integrity.

Steps to Connect Data with Qlik Sense:

    i.   Load Data

        1. Open Qlik Sense and navigate to the Data Load Editor.

        2. Click on "Add data" and select the CSV file containing the dataset.

        3. Preview the data to ensure it is correctly loaded and that all columns are present.

    ii.   Load Data using Script

        1. Use Qlik's scripting language to load the data.

        2. The script written will load the dataset into Qlik Sense for further analysis

    iii.   Data Verification

        1. Verify Data: Check the data to ensure it is correctly loaded.

        2. Data Integrity and Consistency: Ensure that the dataset is consistent and that there are no integrity issues such as duplicate records or mismatched data types.

4. 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, 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.

Steps for Data Preparation-

    i.   Data Cleansing:-

        1. Handle Missing Values: Identify and manage any missing values in the dataset. This may involve imputing missing values with appropriate substitutes or removing records with missing data.

        2. Standardize Data Formats: Ensure that all data fields are in consistent formats, particularly date fields. This may involve converting date strings to date objects or standardizing numeric formats.

    ii.   Data Transformation:-

        1. Create Derived Fields: Transform raw data into meaningful categories or groups. For example, categorize ages into age groups (Child, Teen, Adult, Middle Aged,

Senior) using nested IF statements.

2. Standardize Column Names: Rename columns to ensure they are clear and consistent, facilitating easier reference and analysis.

iii.   Data Exploration:-

Identify Patterns and Trends: Explore the data to uncover patterns and trends. Use Qlik Sense's data profiling features to understand the distribution and relationships within the data.

iv.   Data Filtering:-

1. Focus on Specific Subsets: Filter the data to focus on specific subsets relevant to your analysis. For instance, you may want to analyze data for a particular age group or gender.

v.   Data Loading Script:-

Use Qlik's scripting language to load and transform the data as needed.

vi.   Data Verification:-

1. Verify Transformations: Check the data to ensure that all transformations have been applied correctly. This includes verifying that the AgeGroup field correctly categorizes the age of each passenger.

2. Ensure Accuracy and Completeness: Verify that the dataset is accurate and complete before proceeding to visualization.
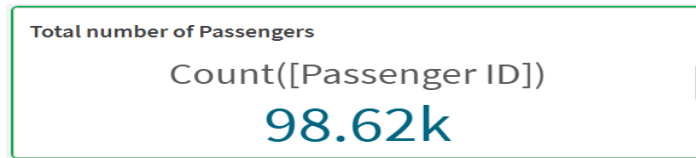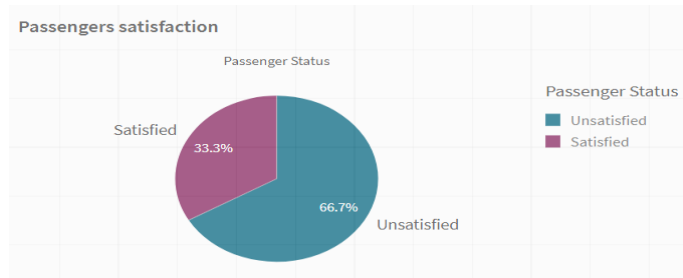

5. Data Visualizations

5.1 Visualizations

Data visualization is the process of converting raw data into graphical representations to make the information easily understandable and to gain insights into performance and efficiency. With Qlik Sense, any one can create a variety of visualizations that help in exploring and presenting data trends, patterns, and relationships. The goal of data visualization is to make complex data sets more accessible, intuitive, and easier to interpret. This section outlines the key visualizations created for the project, the steps to create them, and the insights derived from these visualizations.
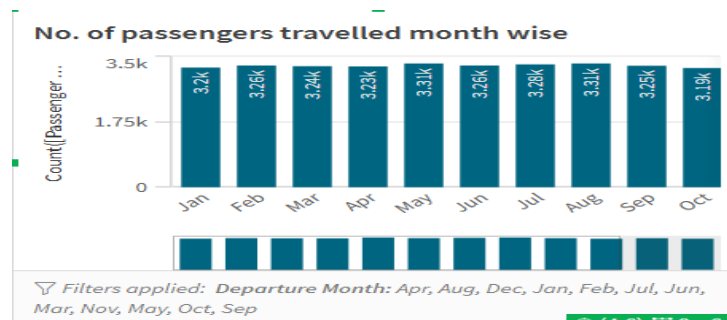
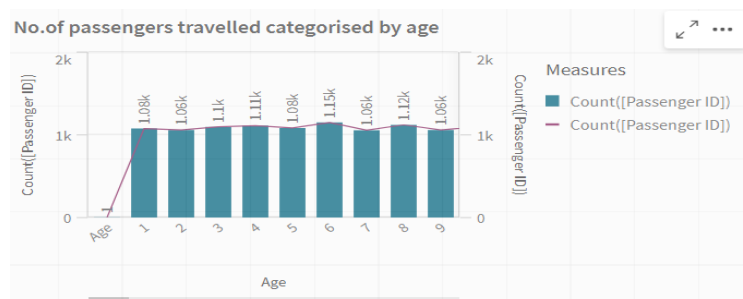**Airline Data Analysis**

   a.   **Total number of Passengers (KPI)**

Total number of Passengers

Count([Passenger ID])

98.62k

b. **Passenger Satisfaction (Pie Chart)**

Passengers satisfaction

Passenger Status

Satisfied

33.3%

66.7%

Unsatisfied

Passenger Status
- Unsatisfied
- Satisfied

c. **Number of Passengers travelled per month (Vertical Bar Chart)**

No. of passengers travelled month wise

Count([Passenger ...

3.5k

1.75k

0

Jan 3.2k | Feb 3.26k | Mar 3.24k | Apr 3.23k | May 3.31k | Jun 3.26k | Jul 3.28k | Aug 3.31k | Sep 3.25k | Oct 3.19k

Filters applied: Departure Month: Apr, Aug, Dec, Jan, Feb, Jul, Jun, Mar, Nov, May, Oct, Sep

d. **Number of Passengers travelled categorised by age (Vertical Combo Chart)**

No.of passengers travelled categorised by age

2k

Count([Passenger ID])

1k

0

Age | 1 1.08k | 2 1.06k | 3 1.1k | 4 1.11k | 5 1.08k | 6 1.15k | 7 1.06k | 8 1.12k | 9 1.06k

Age

Measures
- Count([Passenger ID])
- Count([Passenger ID])

e. **Number of flights were onTime per month (Vertical Bar Chart)**

No. of Flights were on Time month wise

Count([Flight Stat...

1.2k

600

0

Aug 1.15k | May 1.12k | Jul 1.11k | Oct 1.11k | Mar 1.1k | Jun 1.08k | Sep 1.07k | Feb 1.07k | Dec 1.06k | Nov 1.05k | Jan 1.03k | Apr 1.02k

Filters applied: Departure Month: Apr, Aug, Dec, Feb, Jul, Jan, Jun, May, Mar, Nov, Sep, Oct Flight Status: On Time
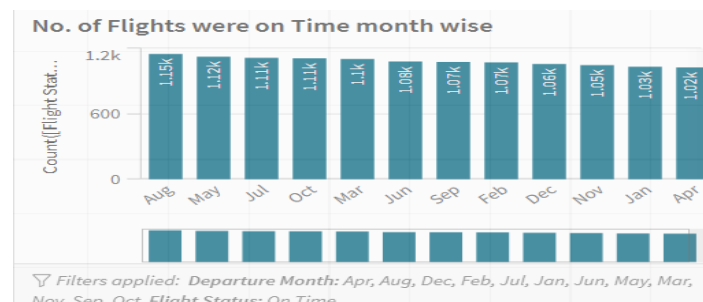
f. Number of passengers travelled from each airport (Vertical Combo Chart)

No. of passengers travel from each airport
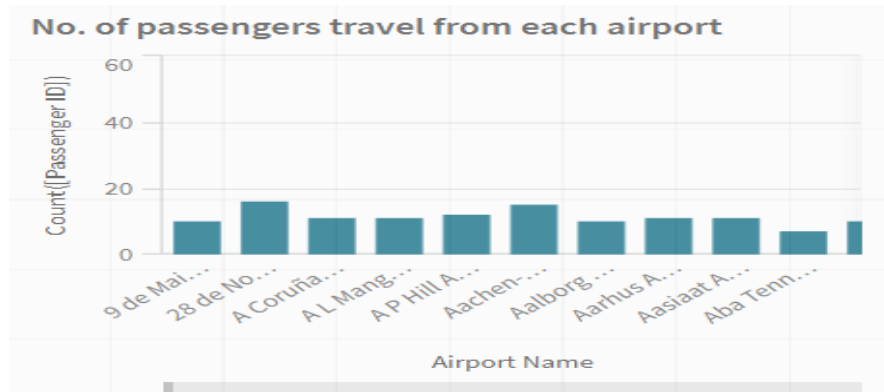
g. Donut Chart to show measure of Flight Status (Donut Chart)

Donut Chart that shows measure of onTi...

Delayed   Cancelled

33.3%   33.4%

33.3%

Filters applied: Flight Status: Delayed, Cancelled, On Time
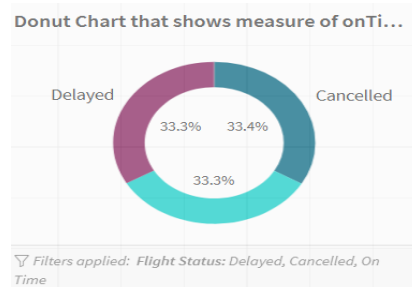
h. Number of airports per each country (Table)

No. of airports per each country

Country Name		countDistinct Airport Name
Totals		**9063**
Country Name		1
United States		2018
Canada		493
France		133
Brazil		411
United Kingdom		126
Italy		64

i. Number of pilots at each airport (Horizontal Bar Chart)

Number of Pilots by Airport Name

Airport Name	countDistinct Pilot Name
San Pedro Airport	43
Santa Maria Airport	38
Böblingen Flugfeld	36
Santa Ana Airport	35
San Fernando Airport	31
Mae Hong Son Airport	29

j. Total No. of Pilots (KPI)

Total no. of Pilots

Count([Pilot Name])

98.62k
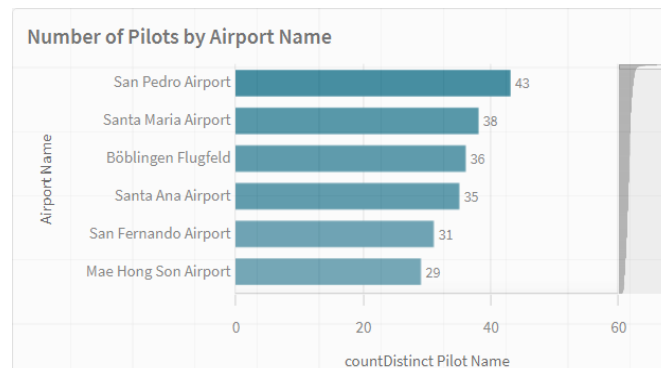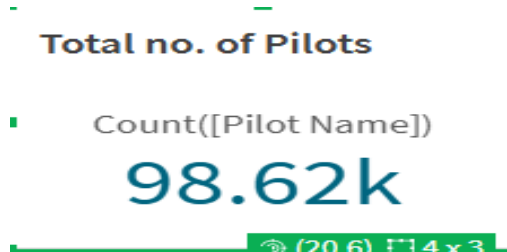
k. Average Age of Passengers (Gauge)

Average age of passengers

50

45.5
Avg(Age)

0   100
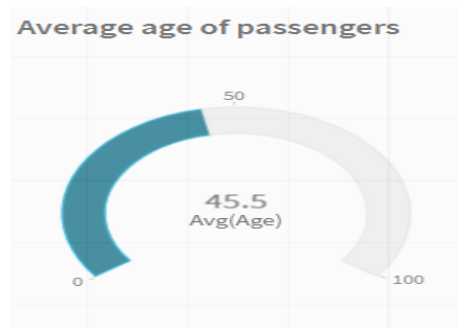
l. Number of Passengers travelling from each country (Horizontal Bar Chart)

No. of Passengers travelling from each country

Country	countDistinct Passenger Name
United States	22.1k
Australia	6.37k
Canada	5.42k
Brazil	4.5k
Papua New Guinea	4.08k
China	2.78k

Country Name

0   5k   10k   15k   20k   25k

countDistinct Passenger Name
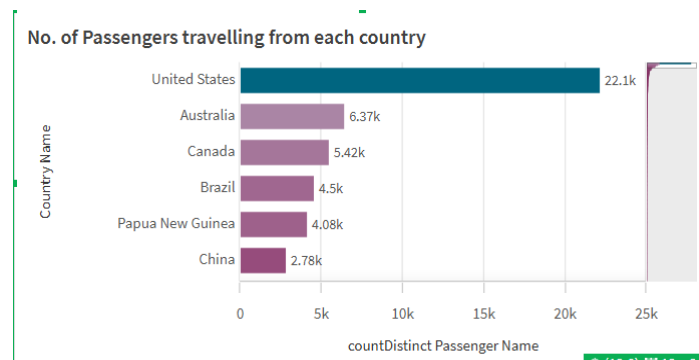
m. Passengers and their Gender (Table)

Passengers and their gender

Passenger Name	Gender
AarenAyre	Female
AarenBoyse	Female
AarenGreenalf	Female
AarenHornig	Female
AarenKohnen	Female
AarenLattka	Female
AarenThrussell	Female
AarenTours	Female
AarenWalkingshaw	Female

n. Age Group of Passengers Gender wise (Vertical Stacked Bar Chart)
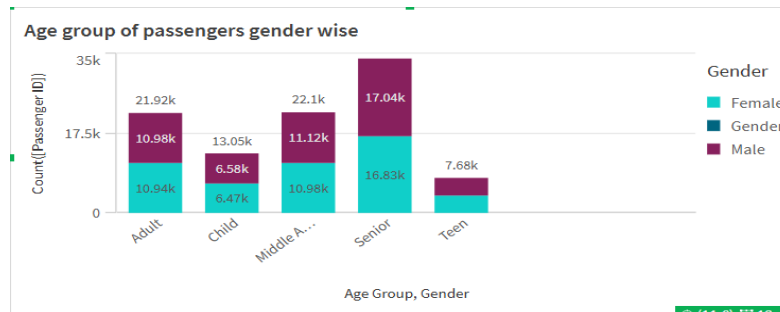
o. Continent wise Flight Status (Tree Map)
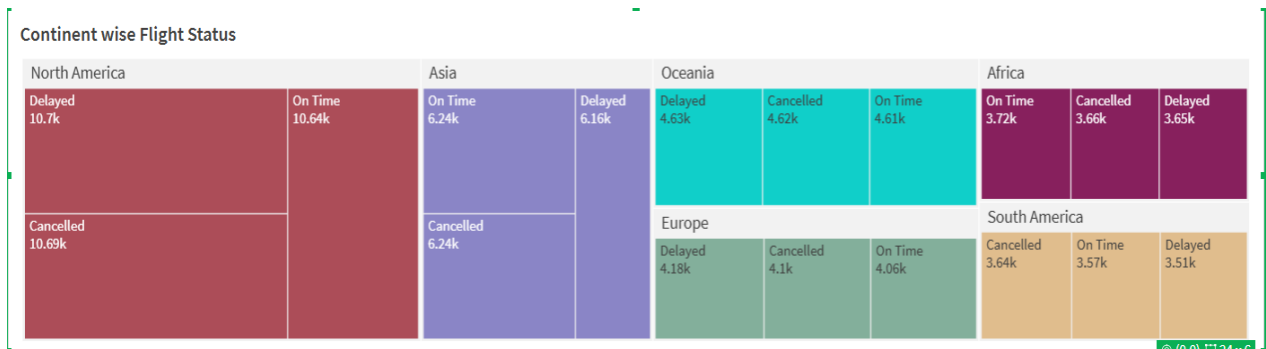
6. Dashboard

6.1 Responsive and Design of Dashboard

A dashboard is a collection of visualizations that provides a comprehensive overview of key metrics and insights at a glance. In this project, the dashboard created in Qlik Sense integrates various visualizations to help users understand passenger demographics, booking trends, and other critical aspects of airline operations. The goal is to create a user-friendly and informative dashboard that effectively communicates key insights from the data.

In this project, I have created following 4 dashboards:

    i.   Passenger Data

   ii.   Flight Data

  iii.   Filter data on basis of Gender

  iv.   Continents Data

Dashboard 1:
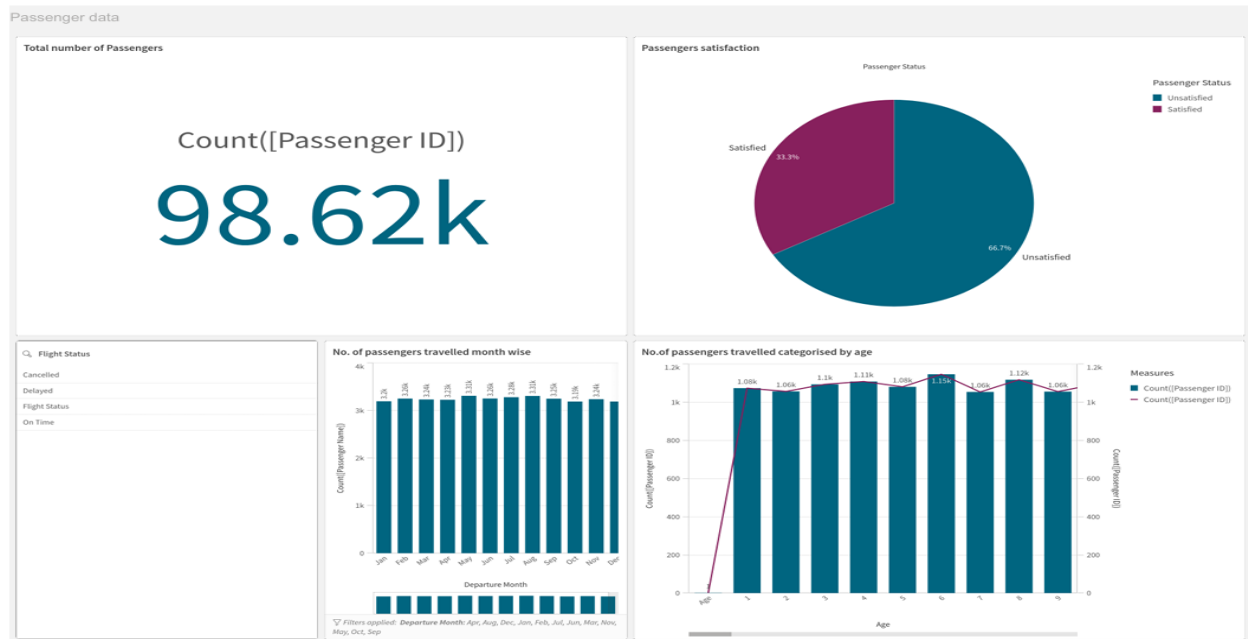
Dashboard 2:
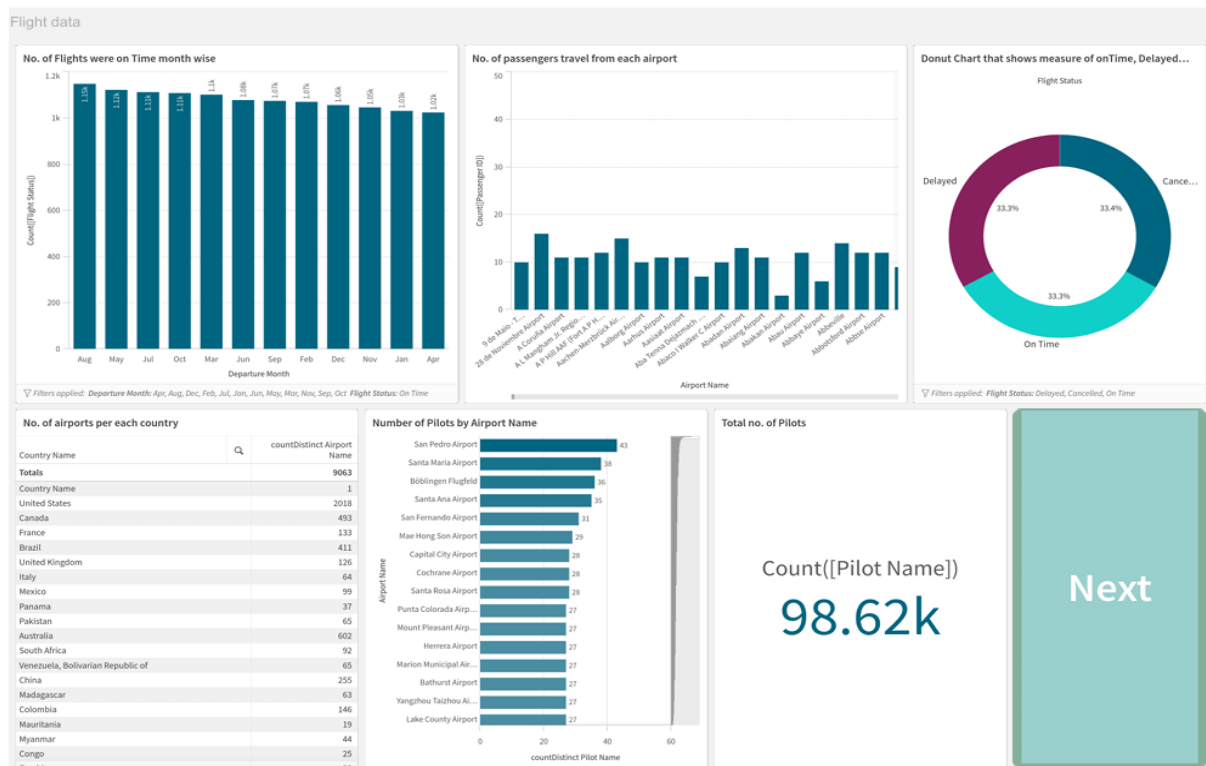
Dashboard 3:
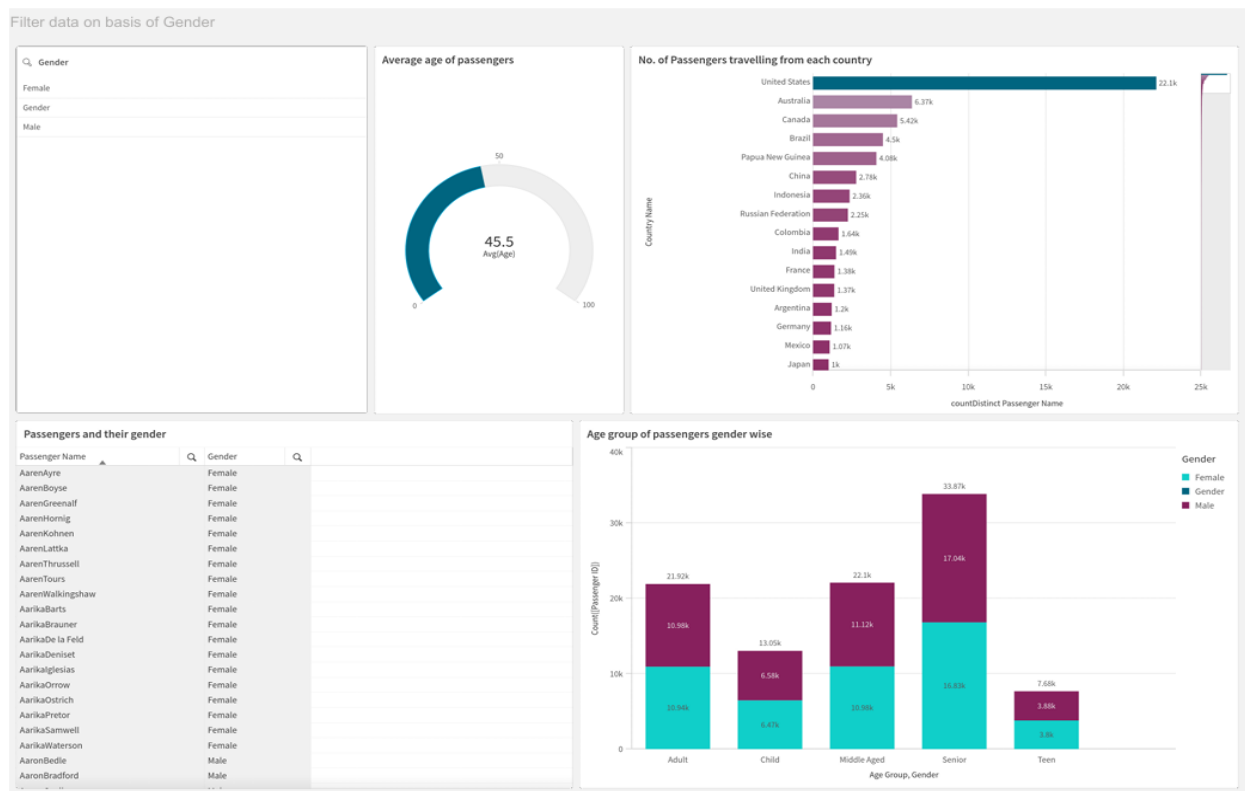
Dashboard 4:
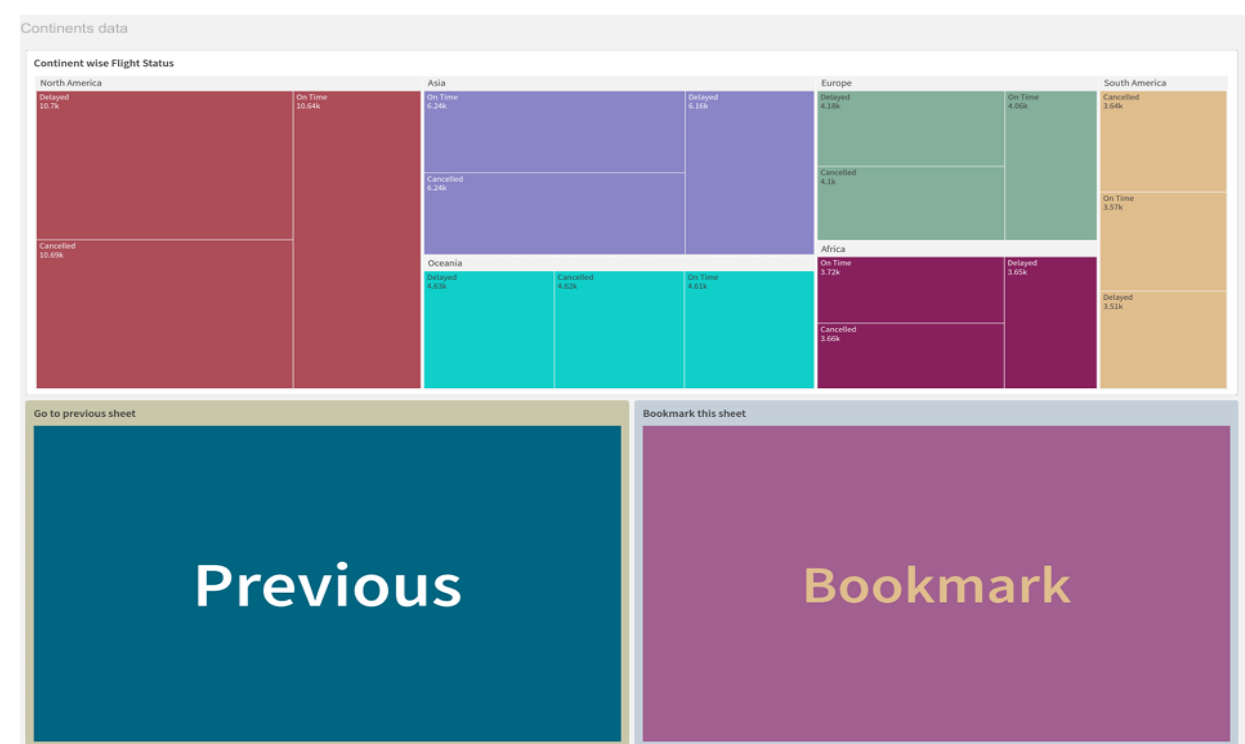
7. Report

7.1 Report Creation

The storytelling feature in Qlik Sense enables user to create a compelling narrative that guides user's audience through the data, highlighting key insights and providing context. By combining visualizations, text, and interactive elements, user can create a powerful data story that enhances understanding and supports data-driven decision-making.
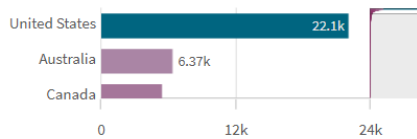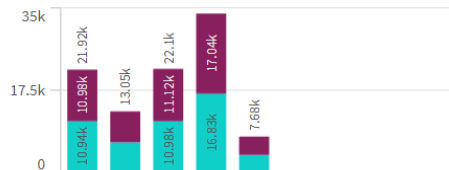
Following is Airline Data Analysis story:
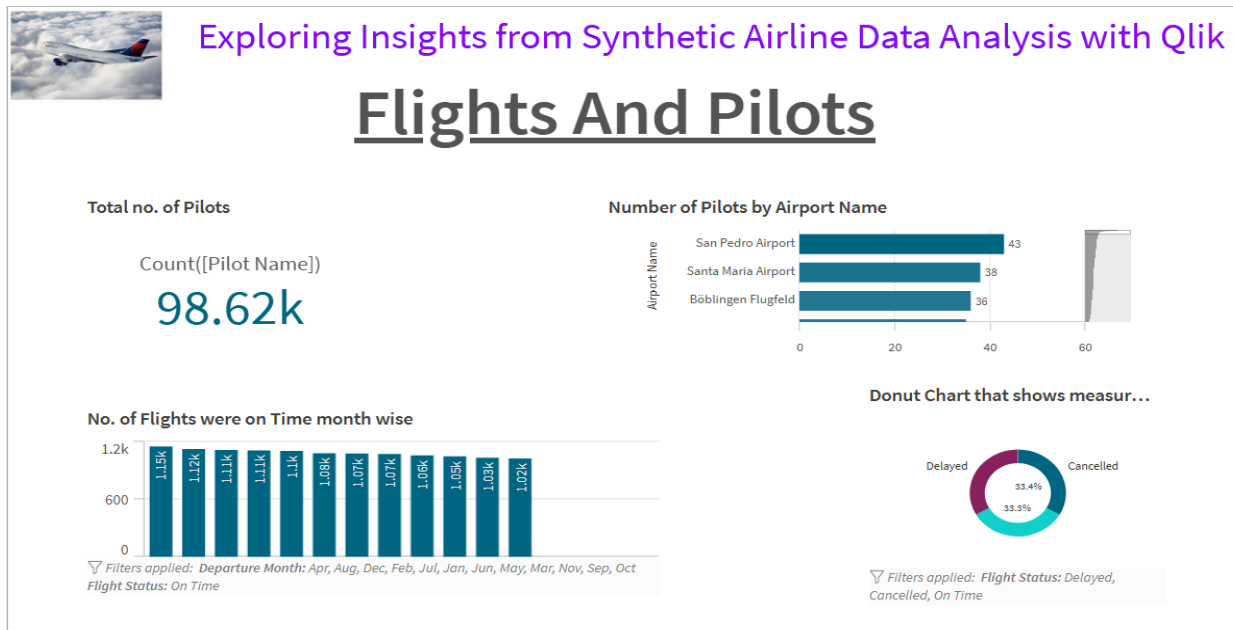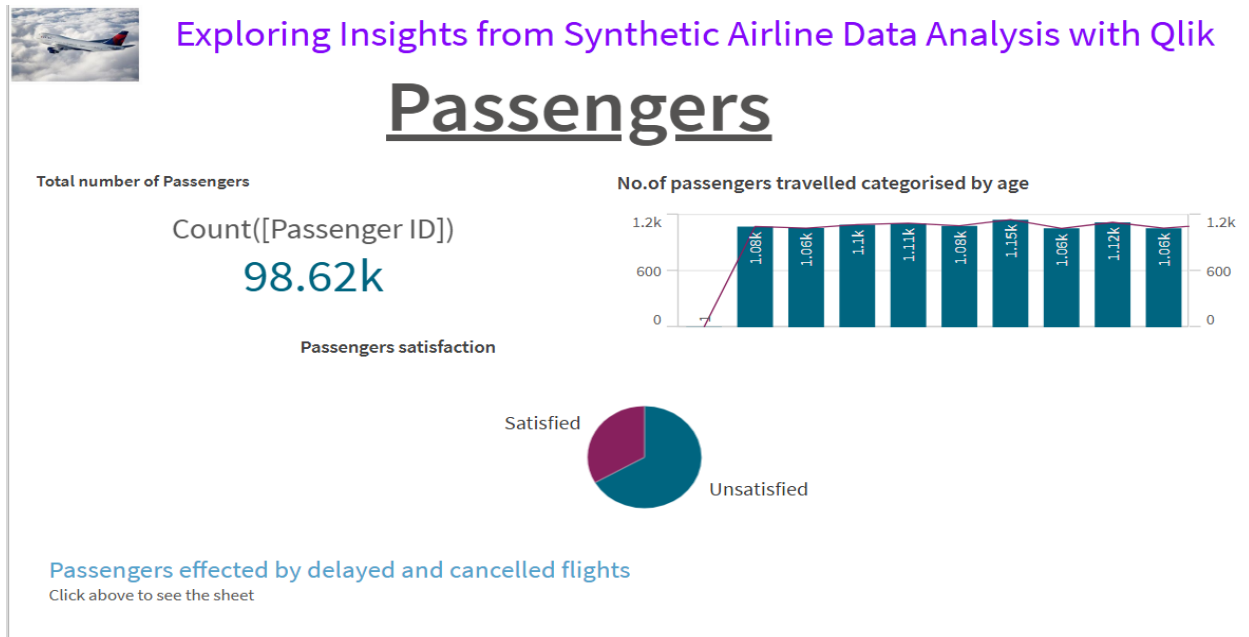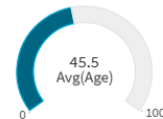
Gender based data

No. of Passengers travelling from each country

Age group of passengers gender wise

Average age of passe...

Filtered data as gender Male

click above to see the sheet

Filtered data as gender Female
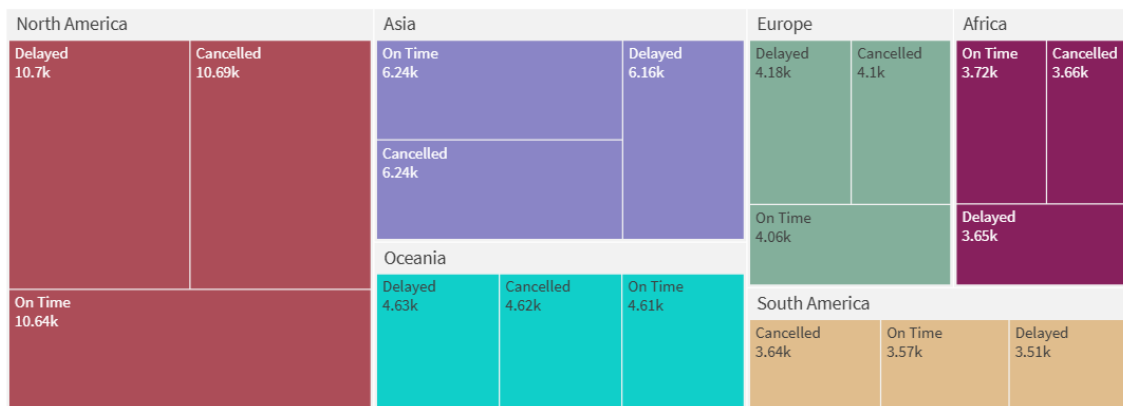
click above to see the sheet

Continents

Continent wise Flight Status

North America		Asia		Europe		Africa	
Delayed 10.7k	Cancelled 10.69k	On Time 6.24k	Delayed 6.16k	Delayed 4.18k	Cancelled 4.1k	On Time 3.72k	Cancelled 3.66k
		Cancelled 6.24k		On Time 4.06k		Delayed 3.65k	
On Time 10.64k		Oceania		South America			
		Delayed 4.63k	Cancelled 4.62k	On Time 4.61k	Cancelled 3.64k	On Time 3.57k	Delayed 3.51k

8. Performance Testing

8.1 Amount of data rendered

Understanding the volume and structure of the data is crucial for performance testing, ensuring efficient data handling, and validating the data processing steps.

The dataset used for this analysis includes various details about airline passengers, flights, and related attributes. The data was obtained from a synthetic dataset provided by Kaggle, comprising multiple fields relevant to our analysis.
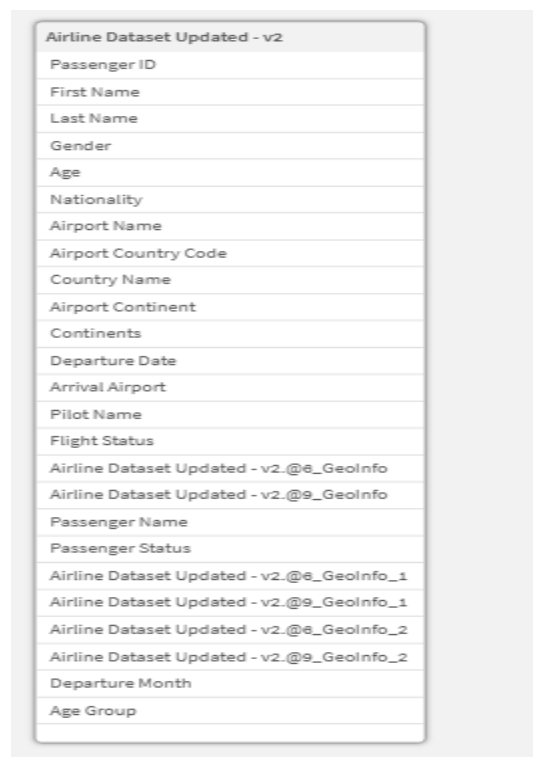
The data was loaded into Qlik Sense using the Data Load Editor. The following summary provides an overview of the amount of data loaded into table within the Qlik Sense application:

No. of Tables: 1

Table Name: Airline Dataset

Total records: 98620

Fields:

| 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 |
| Passenger Name |
| Passenger Status |
| Airline Dataset Updated - v2.@6_GeoInfo_1 |
| Airline Dataset Updated - v2.@9_GeoInfo_1 |
| Airline Dataset Updated - v2.@6_GeoInfo_2 |
| Airline Dataset Updated - v2.@9_GeoInfo_2 |
| Departure Month |
| Age Group |

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.

In this project, after analysis of Airline data the filters are as follows:
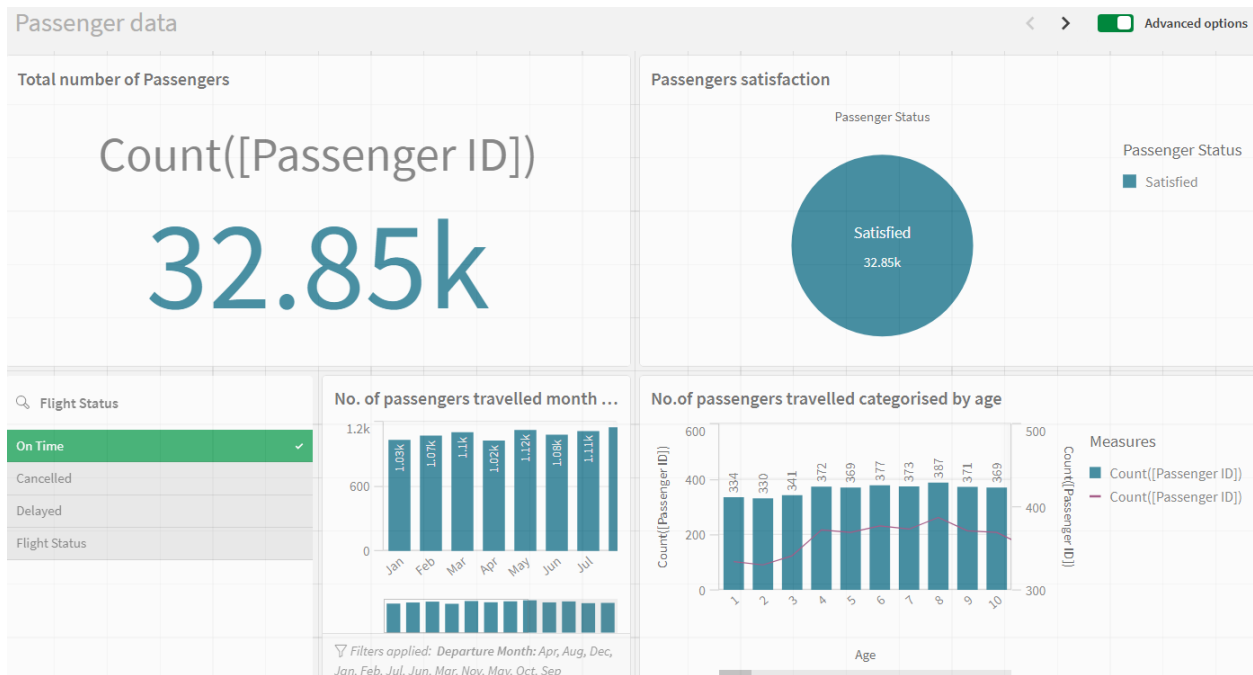
Filtered data based on Gender - Male
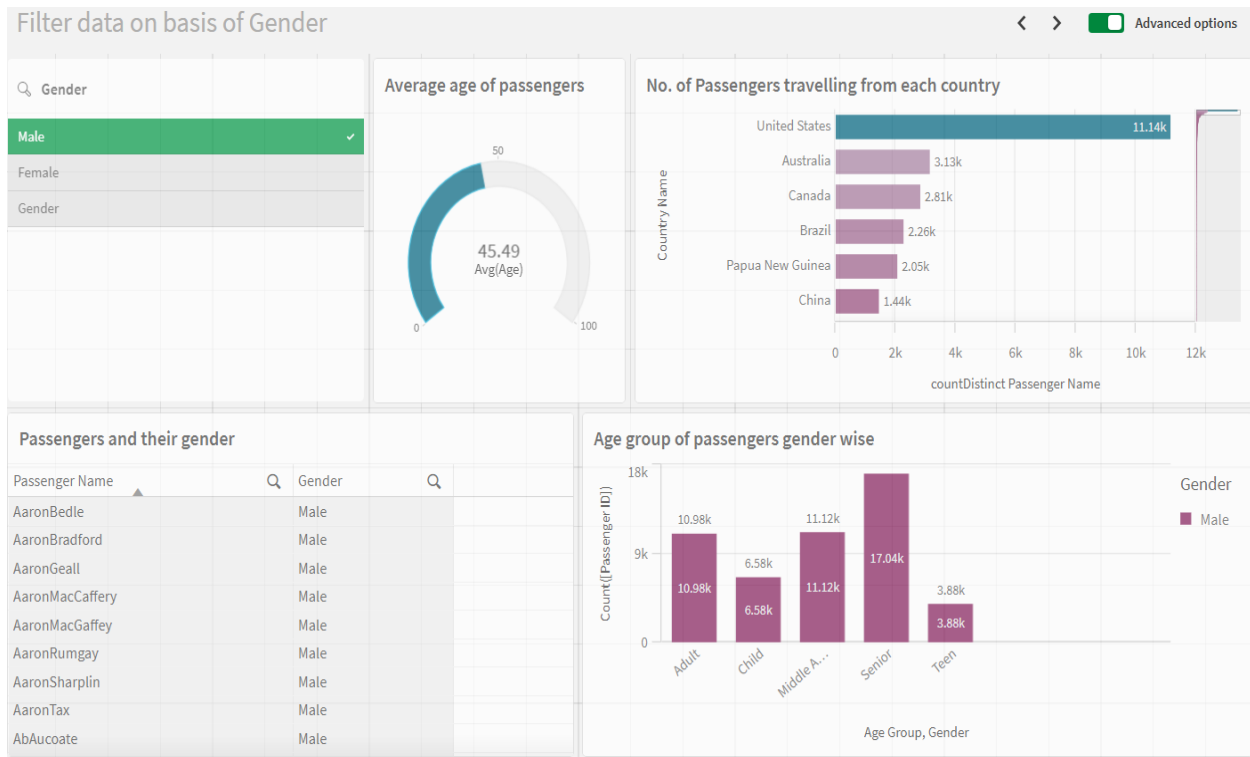
Filtered data based on Gender - Female
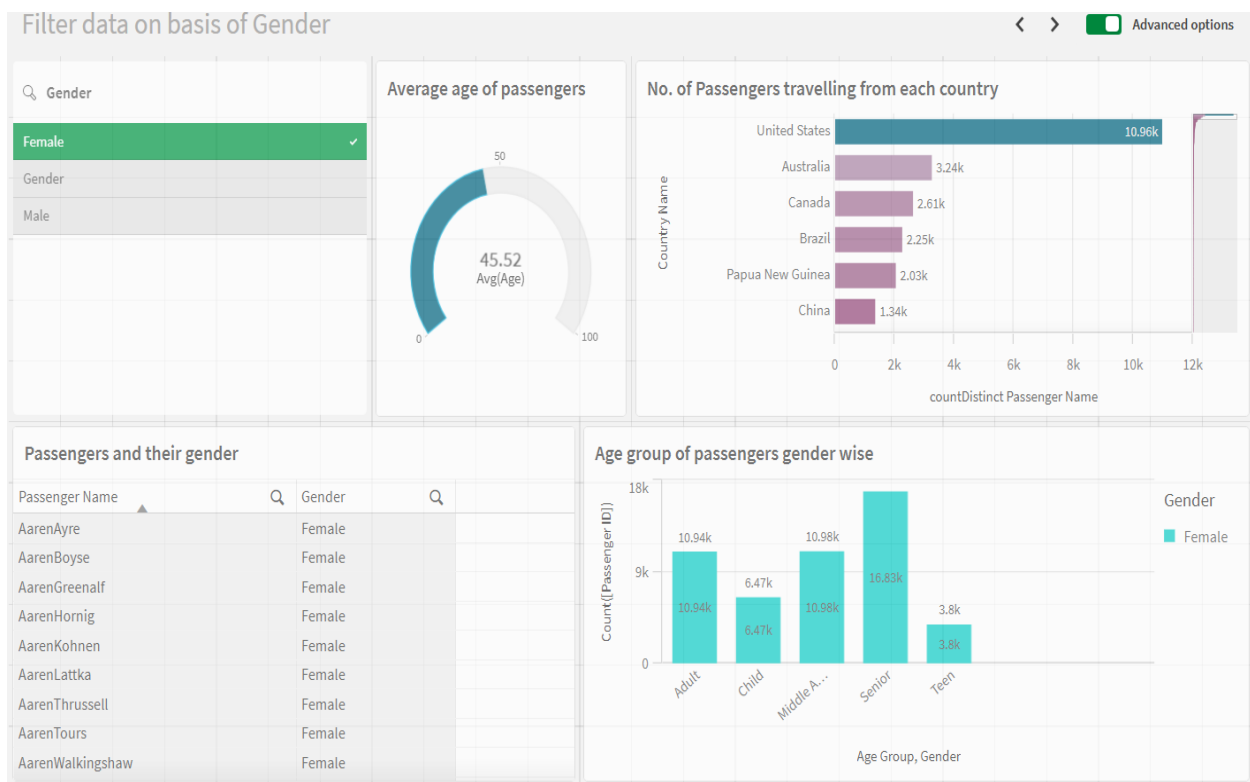
Filtered data based on flight status - onTime

Project documentation by,

Sajid Shaik

Email- shaiksajid.sk3@gmail.com