

Exploring Insights From Synthetic Airline Data Analysis With Qlik

1. Introduction

1.1. Project Overview:

The goal of the project "Exploring Insights from Synthetic Airline Data Analysis with Qlik" is to extract insightful information from synthetic airline data by leveraging the robust analytical capabilities that Qlik offers. This dataset, which includes a wide range of variables like flight statuses (e.g., on-time, delayed, cancelled), passenger demographics (e.g., age, gender), and geographic distribution (e.g., routes, destinations), is meticulously designed to replicate the complexities of actual airline operations. We intend to delve deeply into these various facets of airline operations using Qlik's capabilities in order to find hidden patterns, recognise new trends, and find connections that could influence important choices made by the airline sector.

1.2. Project Purpose:

This project's two main goals are as follows:

Encouraging Stakeholders in the Airline Industry: The objective of this project is to give participants valuable insight obtained through an analysis of synthetic airline data. These data have the potential to improve pricing tactics, increase consumer satisfaction, increase operational efficiency, and optimise flight schedules.

Showcasing Qlik's Capabilities: The project also demonstrates Qlik's capacity to manage complex datasets and derive valuable insights. The study effectively illustrates the utility of Qlik as a potent instrument for data-driven decision-making in a variety of industries by analysing synthetic airline data.

1.3. Technical Architecture

The technical architecture of the project includes an extensive workflow that begins with data preparation and gathering and ends with analysis and visualisation.

Data collection: Kaggle is the source of synthetic airline data, which guarantees a thorough depiction of airline operations.

Data Preparation: In the Qlik environment, data is cleaned, processed, and made ready for analysis.

Data analysis: To find patterns and insights in the data, Qlik's analytical tools are used to explore and analyse the data.

Visualisation and Dashboards: Qlik Sense is used to generate interactive dashboards that display insights in an understandable and user-friendly way.

Performance testing: To guarantee a flawless user experience, dashboards are evaluated for efficiency and responsiveness.

Project Documentation: A comprehensive record of all project procedures, conclusions, and suggestions is kept.

2. Define Problems / Problem Understanding

2.1. Specify the Business Problem

The airline industry faces multiple challenges that impact operational efficiency, customer satisfaction, and revenue optimization. Key problems include:

Revenue Management: To optimise revenue, airlines must identify popular routes and peak travel times, as well as optimise ticket pricing and sales techniques.

Operational Efficiency: To handle growing passenger counts, cut down on delays, and efficiently manage resources, airports and airlines need to streamline their operations.

Customer Experience: Maintaining a positive passenger experience is essential to business. In order to improve overall happiness and offer customised services, airlines must comprehend the demographics and interests of their passengers.

2.2. Business Requirements:

In order to tackle the highlighted business issues, the project needs to fulfil the subsequent requirements:

Capabilities for Data Analysis:

Revenue Optimisation: Examine data on ticket sales to find revenue patterns, periods of high travel demand, and efficient pricing schemes.

Operational Insights: To find bottlenecks and boost operational effectiveness, look at aircraft schedules, passenger flows, and flight statuses.

Customer Demographics: Examine traveller demographics to learn about their preferences and adjust services as necessary.

Dashboards that interact:

Passenger Analytics Dashboard: See nationality distribution, monthly trends, and passenger demographics visually to learn more about how customers behave.

Flight Status Insights Dashboard: Show flight status information, such as delayed, cancelled, and on-time flights, to spot trends and improve operational choices.

Dashboard for Geographical and Demographic Insights: Map the distribution of passengers geographically and examine passenger demographics regionally to spot market trends and opportunities.

3. Data Collection

Data collection or data gathering is the process of gathering and measuring information on targeted variables in an established system, which then enables one to answer relevant questions and evaluate outcomes. Data collection is a research component in all study fields, including physical and social science, humanities and business.

3.1. Source of Dataset

The dataset used in this project was sourced from Kaggle, a well-known platform for high-quality datasets. The synthetic airline dataset provides a detailed representation of airline operations, including passenger information, flight details, and performance metrics.

[Click here](#) to download the dataset.

Dataset Fields

The synthetic airline dataset includes the following fields:

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, cancelled).

This comprehensive dataset provides the necessary information to perform detailed analysis and derive valuable insights into airline operations, passenger demographics, and flight performance.

3.2. Connect Data with Qlik

After downloading the dataset, the following steps were undertaken to connect the data for analysis in Qlik Sense:

Creating a New Analytic App: With Qlik Cloud open, I made a new analytical application called "Airline Data Analysis App." The data analysis process is managed and organised using this app as a workspace.

Uploading the Dataset: I made the choice to upload data in the Qlik Cloud setting. Selecting "Upload the dataset from files and other resources," I uploaded the Kaggle dataset that I had previously downloaded. The newly built app now has access to the successfully uploaded dataset.

4. Data Preparation

4.1. Prepare the Data for Visualisations:

The following actions have been carried out to get the data ready for analysis and visualisation after the dataset was uploaded to Qlik Sense:

Changing Field Names:

Initially I made changes to the selections so that "No Field Names" became "Embedded Field Names." This guaranteed that the fields in the dataset had accurate labels and were prepared for further processing.

Data Filtration:

I used the filter Arrival Airport > 0 to deal with missing values in the Arrival Airport field. In order to preserve the accuracy and relevance of the data for further analysis, this filter eliminated any rows with null values in the Arrival Airport field. Eliminating these null values was essential since they can cause errors in the analysis and affect the conclusions drawn from the data.



Figure 1: Filter Applied to Remove Null Values in Arrival Airport Field

By filtering out incomplete data, I ensured that all remaining entries had valid and actionable information, thereby improving the quality of the analysis and the reliability of the insights generated from the dataset.

Calculated Field Creation:

Month: I took the month out of the Departure Date column and put it into a calculated field called Month. Time-based data analysis is made easier by this field.

Age Category: I made the Age Category calculated field. Based on their age, passengers are categorised into four age groups using this field: child, teen, adult, or senior. The following is the classification:

Child: aged 0–12

Teen: Ages 13 to 19

Adult: 20–64 years old

Senior: more than 65 years old

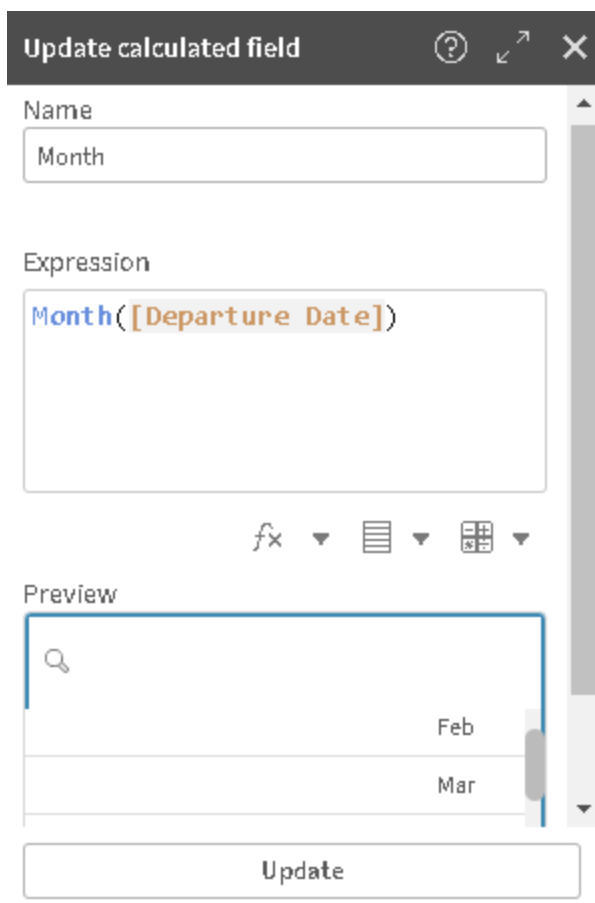


Figure 2: Expression for Calculated Field - Month

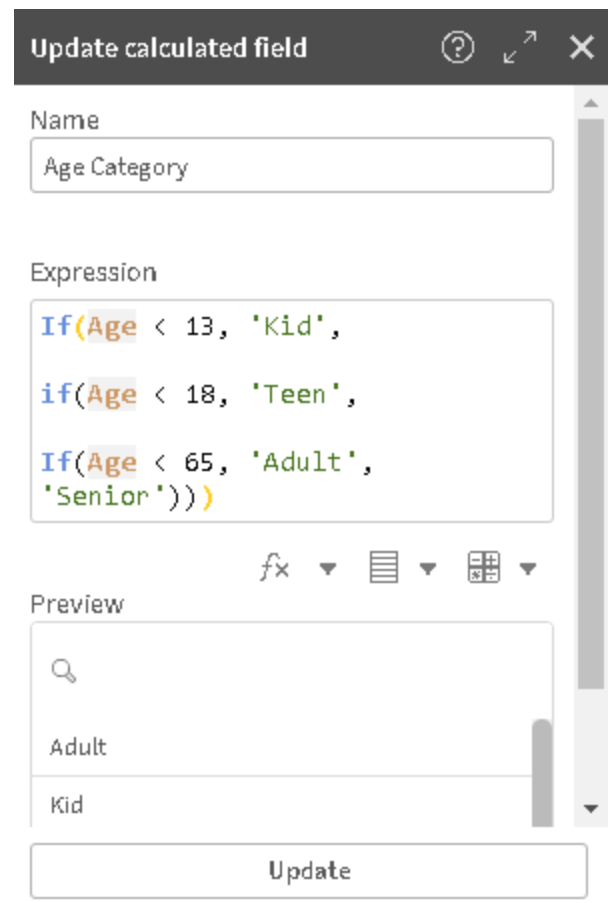


Figure 3: Expression for Calculated Field - Age Category

These data preparation procedures cleaned, organised, and improved the dataset with new computed fields, preparing it for further Qlik Sense analysis and visualisation. More detailed insights into passenger demographics and travel patterns were made possible by the calculated fields, while the applied filter guaranteed data quality.

5. Data Visualisation

Data visualization is the graphical representation of information and data. Data visualisation tools offer an easy-to-use means of seeing and comprehending trends, outliers, and patterns in data through the use of visual elements such as charts, graphs, and maps. Using the most suitable graphical methods to convey information to consumers in an understandable and effective manner is the primary objective.

Effective data visualisations can uncover insights that may be hidden in raw data, making complex data easier to interpret. They are essential to the processes of data analysis and decision-making because they help people swiftly understand complex ideas or spot novel trends.

5.1. Visualizations:

- **KPI Chart:**

- KPI Chart displaying Passenger Count: This KPI Chart gives an instantaneous picture of the total number of passengers.

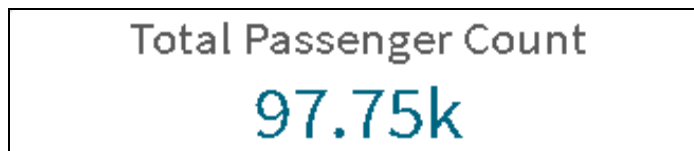


Figure 4: Total Passenger Count KPI Chart

- KPI Chart displaying Average Age: This KPI Chart provides information about the passengers' demographic makeup by exhibiting the average age of the travellers.

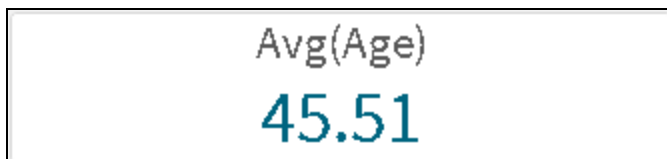


Figure 5: Average Age of Passengers KPI Chart

- **Scatter Plot:**

- Scatter Plot showing Passenger Distribution by Age and Nationality: With nationality as the dimension, average age on the x-axis, and passenger count on the y-axis, this scatter plot illustrates the distribution of passengers by age and nationality.

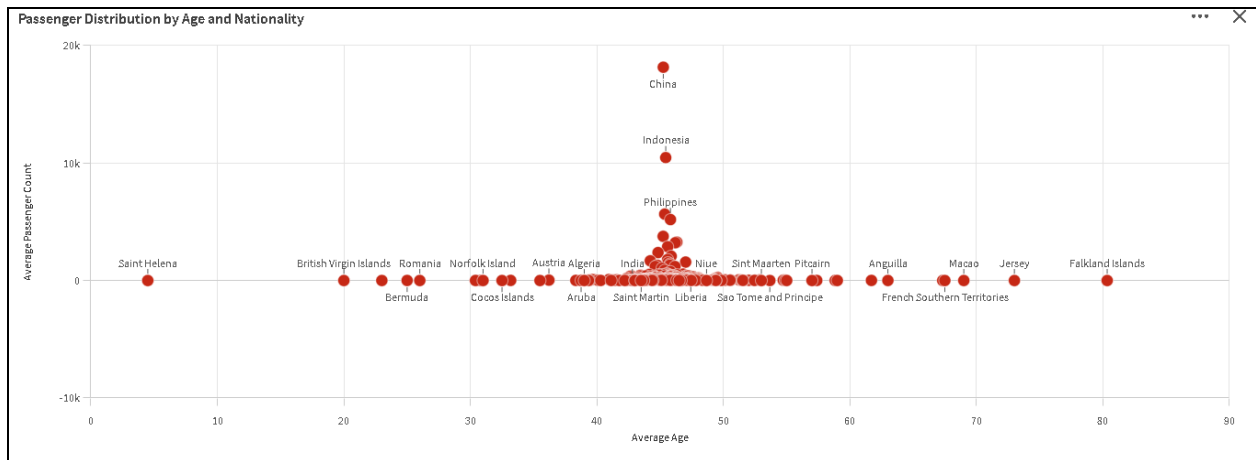


Figure 6: Passenger Distribution by Age and Nationality Scatter Plot

- **Line Chart:**

- Line Chart showing Monthly Passenger Count: The monthly passenger count is shown in this line chart, which also offers insights into seasonal fluctuations and trends in passenger volume.

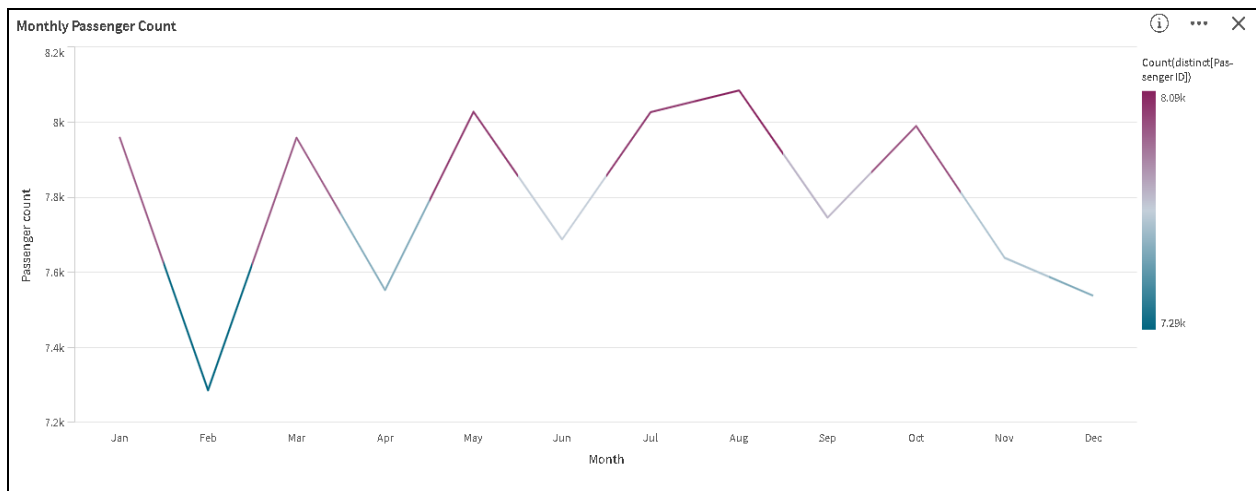


Figure 7: Monthly Passenger Count Line Chart

- **Stacked Bar Chart:**

- Stacked Bar Chart showing Monthly Distribution of Passengers by Age Group: Every bar in this stacked bar chart represents a separate age group (adult, teen, kid, senior), and it shows the monthly distribution of passengers by age group.

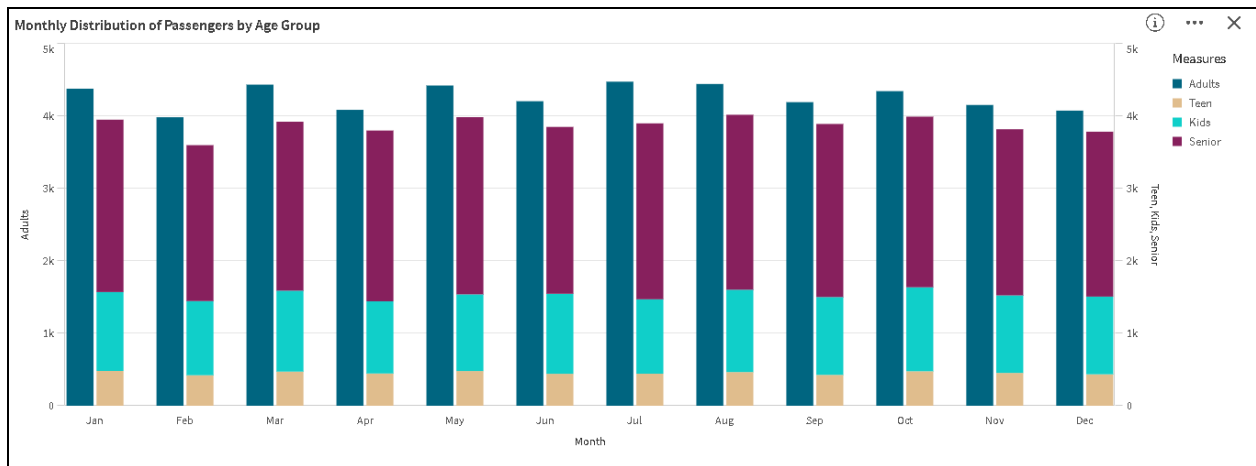


Figure 8: Monthly Distribution of Passengers by Age Group Stacked Bar Chart

- **Donut Chart:**

- Donut Chart showing Passenger Gender Distribution: This donut chart provides information on the gender makeup of the passengers by displaying the distribution of passengers by gender.

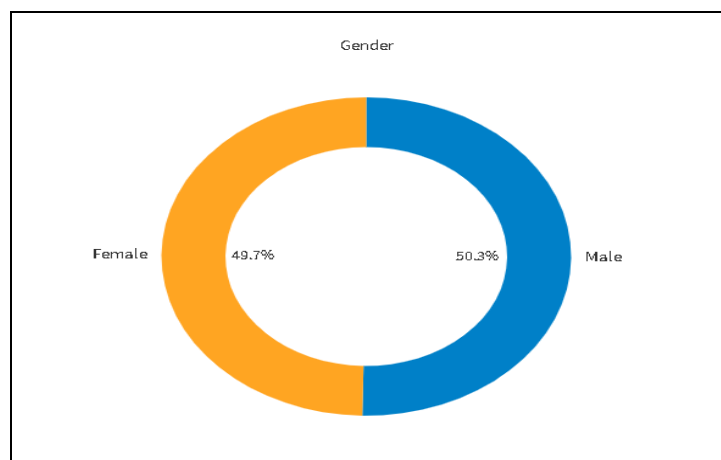


Figure 9: Passenger Gender Distribution Donut Chart

- **Pivot Table:**

- Offering a thorough analysis of geographic data, this Pivot Table displays the quantity of arrival airports and passenger counts by nation.

Country Name	Number of Airports	Number of Passengers
Afghanistan	32	369
Albania	1	12
Algeria	43	449
American Samoa	4	45
FTI	1	14
OFU	1	9
PPG	1	10
TAV	1	12
Andorra	1	7
Angola	40	445
Anguilla	1	14
Antigua and Barbuda	2	20
Argentina	108	1195
Armenia	2	21
Aruba	1	9
Australia	602	6131
Austria	7	91

Figure 10: Arrival Airports and Passenger Counts by Country Pivot Table

- **Bar Chart:**

- Bar Chart showing Passenger Distribution by Age Category: To better comprehend the demographic distribution of passengers, this bar chart shows the distribution of passengers by age group.

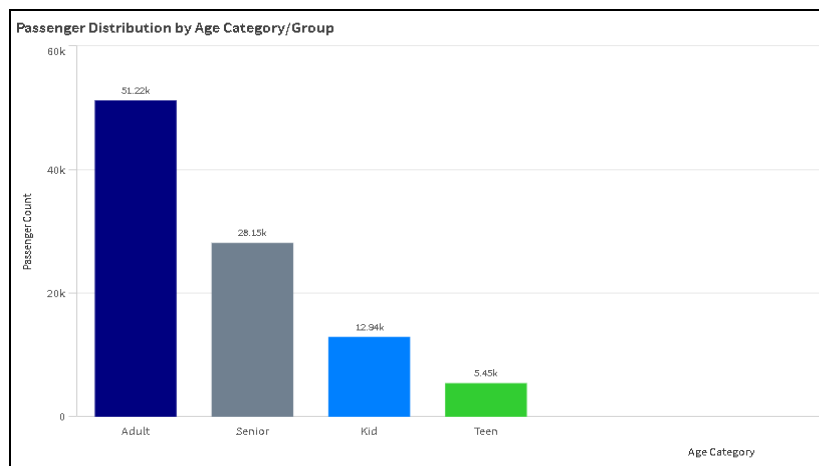


Figure 11: Passenger Distribution by Age Category Bar Chart

- **Map:**

- Map showing Global Passenger Distribution by Nationality: This map illustrates how passengers are dispersed throughout various nations and offers a worldwide perspective on passenger distribution by nationality.

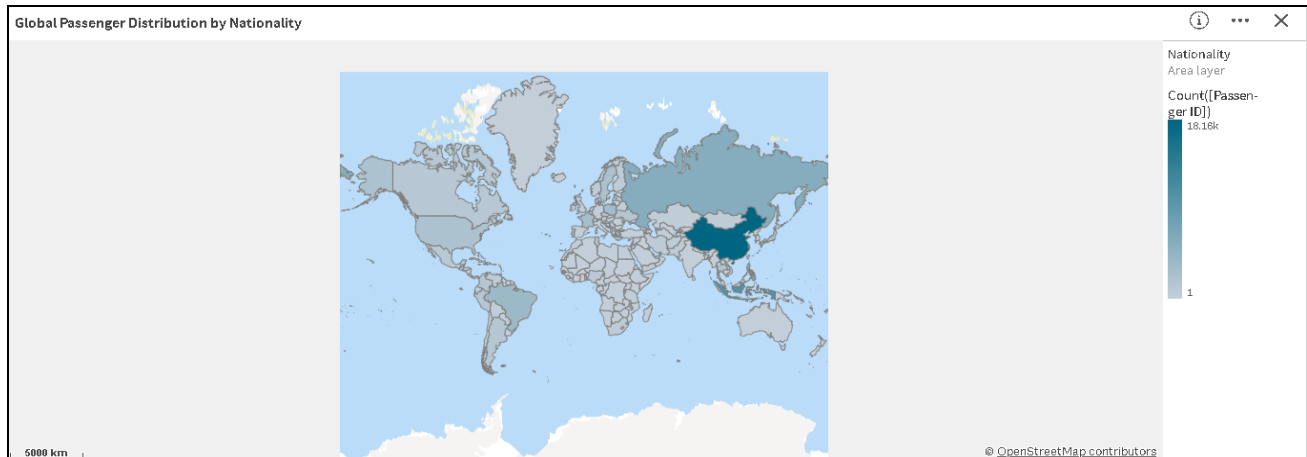


Figure 12: Global Passenger Distribution by Nationality Map

These visualizations collectively provide a comprehensive analysis of the synthetic airline data, enabling stakeholders to explore various aspects of airline operations, passenger demographics, and flight statuses interactively.

6. Dashboards

Dashboards are effective tools for data analytics because they offer a unified perspective of many data visualisations, which simplifies the understanding and analysis of complex datasets. They give customers the ability to engage with and study data in a consistent and user-friendly way by combining different graphs, charts, and key performance indicators (KPIs) into a single interface. Dashboards give important information in an easily readable format, enabling real-time data monitoring and speedy decision-making. They are intended to draw attention to significant trends, patterns, and anomalies, offering insights that motivate calculated business decisions.

Dashboards built with Qlik Sense provide a thorough analysis of synthetic airline data in the context of this project. They address a range of topics, including flight status updates, passenger demographics, and airline operations. Every dashboard is customised to meet certain analytical requirements, such as comprehending traveller characteristics, gauging operational effectiveness, and analysing geographic distributions. These dashboards convert raw data into useful insights by utilising Qlik's sophisticated visualisation capabilities, assisting stakeholders in decision-making and streamlining airline operations.

Dashboard 1: Passenger Analytics

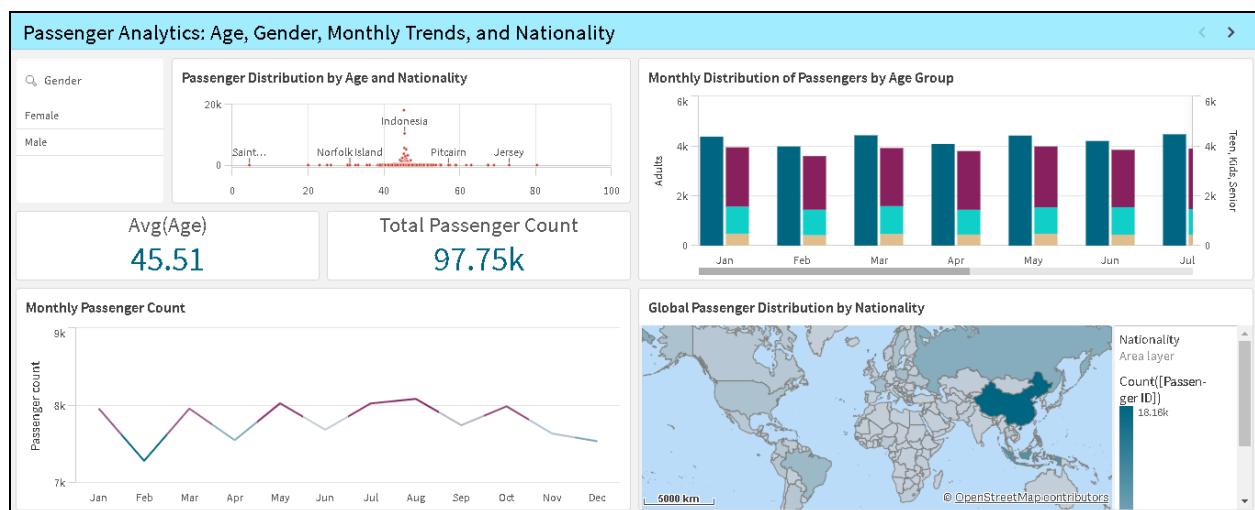


Figure 13: Passenger Analytics Dashboard

The purpose of this dashboard is to offer a thorough examination of travel patterns and passenger demographics. It has visualisations and key performance indicators (KPIs) that provide information about the age, gender, nationality, and monthly travel habits of the passengers. Airlines can better understand passenger characteristics and customise their services to cater to various demographic groups by analysing these variables. Additionally, the dashboard aids in spotting patterns and trends in the way passengers travel over time.

Dashboard 2: Overview of Flight Status

The examination of flight statuses and how they affect the traveller experience is the main emphasis of this dashboard. It contains graphics showing how flight statuses (such on-time, delayed, and cancelled) are distributed among various months and continents. It also sheds light on the distribution of genders among passengers impacted by different flight statuses. In order to improve the entire passenger experience, this dashboard assists airport authorities and airlines in understanding operating efficiency and identifying areas for improvement.

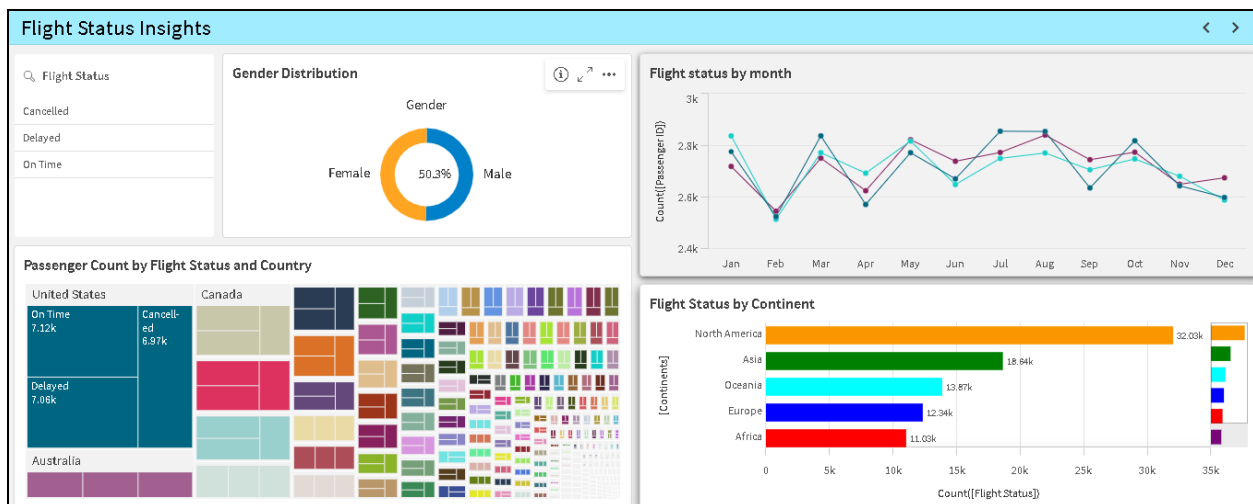


Figure 14: Flight Status Overview Dashboard

Dashboard 3: Passenger Demographic and Geographical Data



Figure 15: Passenger Demographic and Geographical Data Dashboard

An extensive examination of the passenger population's demographics and geographic dispersion is possible with this dashboard. It has KPIs and visualisations that display the total number of nations, airports of arrival, and average age of travellers by nation. The distribution of passengers by age group and continent is also displayed on the dashboard. In order to plan their services and operations appropriately and to understand the demographics of their passengers, airports and airlines need to know this information.

7. Story

In Qlik Sense, a story is a narrative presentation that employs data visualisations to effectively and clearly communicate insights and discoveries. Qlik stories are dynamic and interactive, in contrast to static reports, enabling users to explore the data via guided analytics. By using a storytelling technique, complicated data insights are communicated more effectively and are comprehensible to a broader audience.

7.1. Story in Qlik: Exploring Insights from Synthetic Airline Data Analysis

"Exploring Insights from Synthetic Airline Data Analysis," the story created for this project, integrates a number of dashboard visualisations into a seamless narrative that focuses on important discoveries and useful insights. The narrative is set up to lead interested parties through the investigation, progressing from a broad summary of passenger demographics to in-depth analyses of flight statuses and geographic distributions.

7.2. My Stories

Story 1: Passenger Demographics and Trends:

This story provides a comprehensive analysis of passenger demographics and trends, highlighting important data such as monthly travel patterns, age groups, gender distribution, and overall passenger numbers. The story provides insights into how passengers are dispersed across different demographics and how their travel behaviours vary over time, offering important information for understanding passenger dynamics using a variety of visualisations, including KPI charts, line charts, and maps.

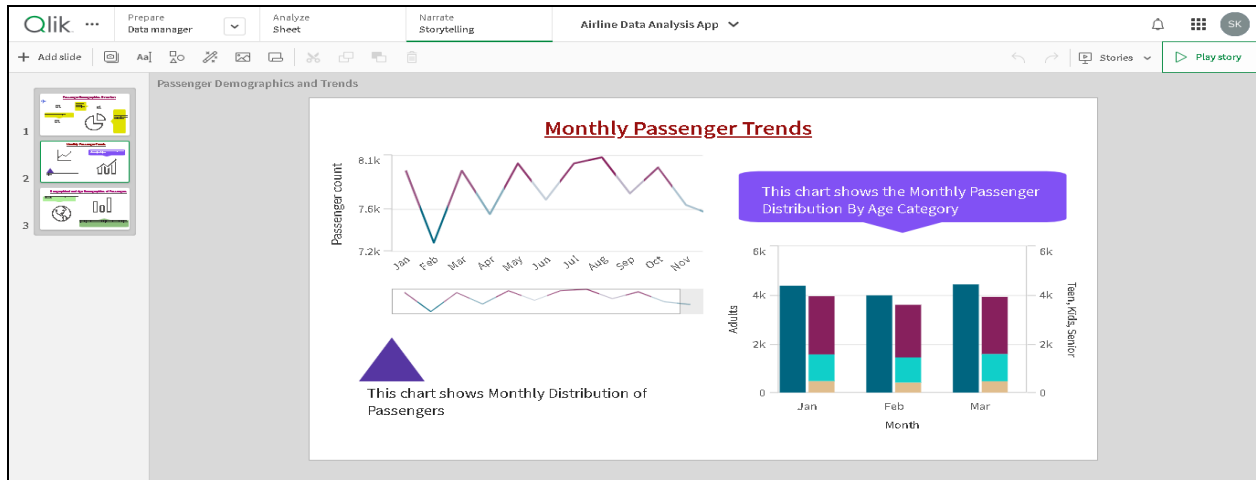


Figure 16: Passenger Demographics and Trends Story

Story 2: Comprehensive Aviation Analysis in Asia

This story analyses the number of passengers, airline infrastructure, and flight status developments with an emphasis on the airline industry in Asia. The story illustrates the number of airports and nations in Asia, the monthly distribution of passengers, and the effects of flight statuses like on-time, delayed, and cancelled using KPI charts, pivot tables, and line charts. With important insights for stakeholders, this report offers a thorough look at the operational and demographic elements of aviation in the Asian region.

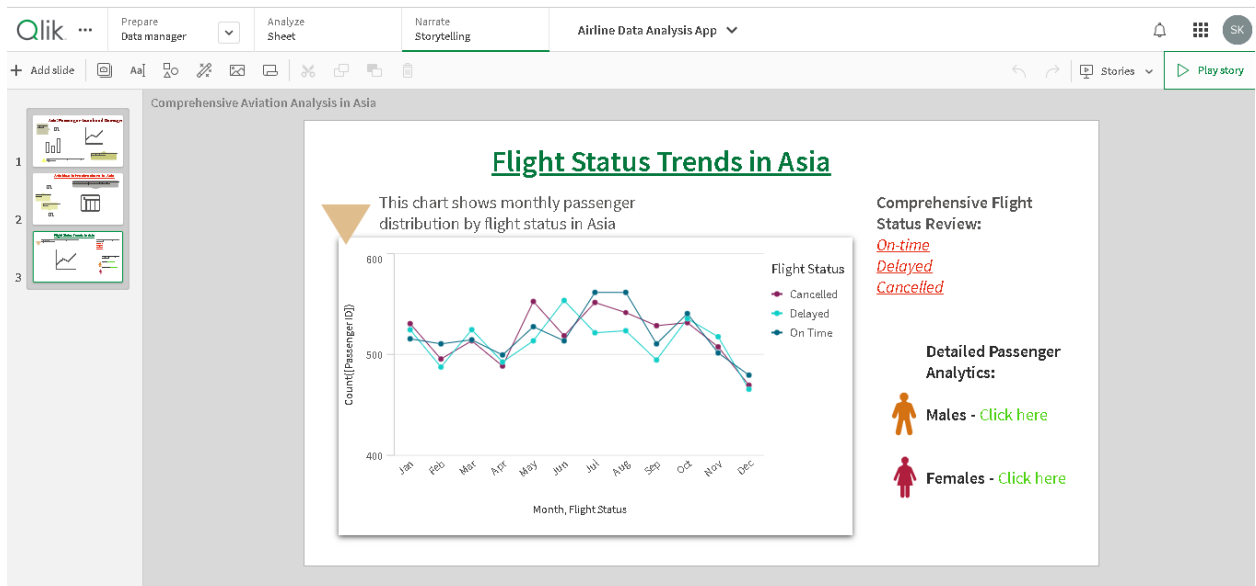


Figure 17: Comprehensive Aviation Analysis in Asia Story

8. Performance Testing

Any data analytics project must include performance testing to make sure the dashboards and visualisations are quick to respond, effective, and able to handle the amount of data being processed. In order to verify that the "Airline Data Analysis App" built in Qlik Sense could handle huge datasets and preserve interaction, its performance was put through a rigorous testing process.

8.1. Amount of Data Rendered

The thousands of entries in the artificial airline dataset utilised for this study include specific details like passenger demographics, aircraft schedules, and performance measures. A thorough load test was performed to evaluate the application's performance with high data volumes. To assess the application's effectiveness in rendering data, the entire dataset was used for testing. We kept an eye on several visualisations, such as scatter plots, line charts, KPI charts, and maps, to make sure they loaded in a reasonable amount of time. To improve data load times, optimisation strategies including data aggregation and indexing were used. Reducing the volume of data processed in real-time also involved the use of calculated fields and filters. The outcomes demonstrated that, even while processing sizable sections of the data, the application presented the data effectively and without noticeable lag. The visualisations also maintained their interactive and responsive qualities.

8.2. Utilization of Filters

During data analysis, filters are crucial for condensing datasets and concentrating on particular sections. Several filters were applied to the dashboards in this project, including filters for gender, continents, and flight status.

Gender Filter:

By dividing the data according to the gender of the passengers, the gender filter enables a thorough examination of trends and patterns that are unique to one gender over the other. The performance and responsiveness of the visualisations were evaluated once this filter was applied.

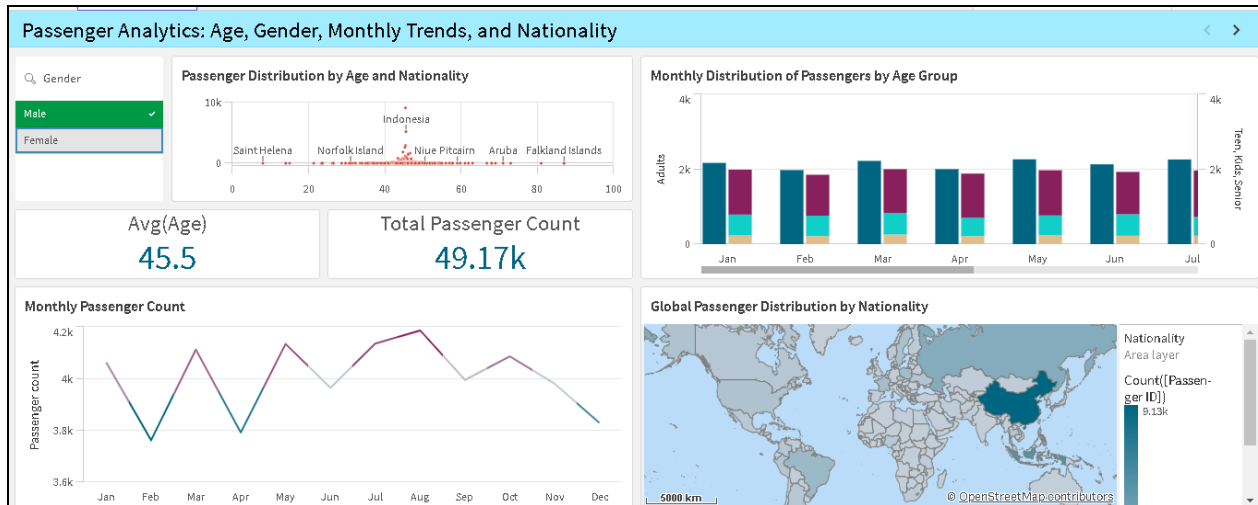


Figure 18: Dashboard - Gender Filter Pane

Flight Status Filter:

This filter aids in data analysis by taking into account the present flight status (e.g., on-time, delayed, cancelled). The effectiveness of this filter in rapidly updating visualisations was tested so that users may examine the relationship between different flight statuses and different data.

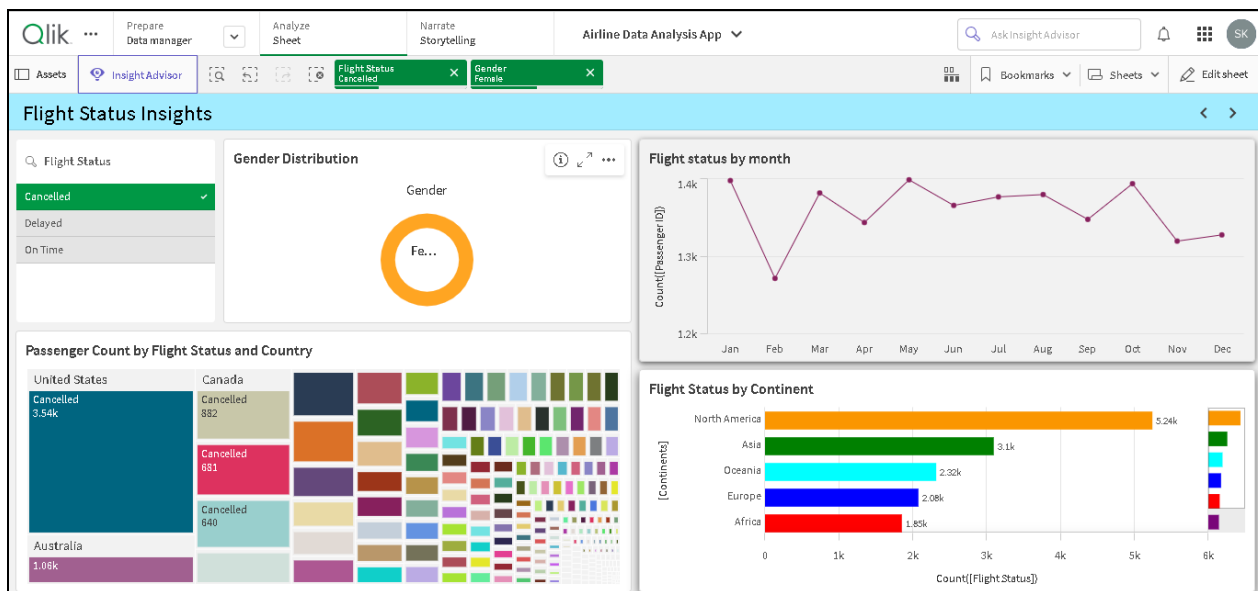


Figure 19: Dashboard - Flight Status Filter Pane

Continent Filter:

By breaking down passenger statistics and flight operations by continent, users can concentrate on particular geographic areas. The effectiveness of this filter was assessed by using it on geographic visualisations and monitoring the interaction and load times.

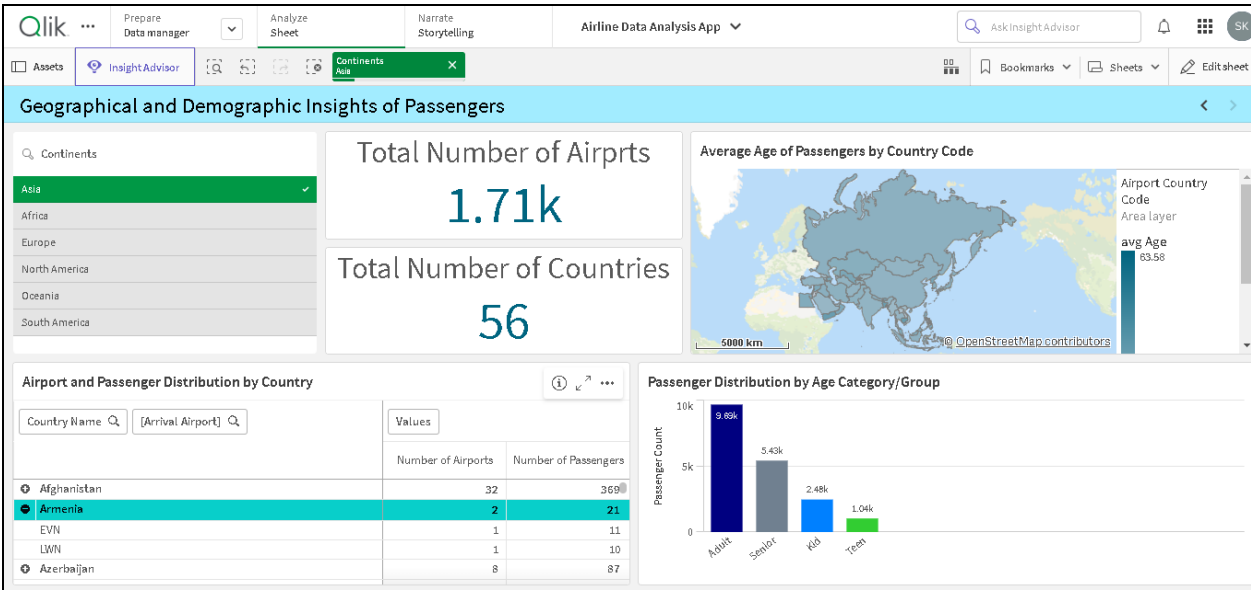


Figure 20: Dashboard - Continent Filter Pane

By using these filters on a range of visualisations, the effectiveness of the filters was evaluated in terms of how they affected responsiveness and load times. To determine each filter's efficiency, both solo and combined filters were examined. The filters improved the data and updated the visualisations fast, especially the ones that affected significant dataset chunks like month and arrival airport. The outcomes demonstrated how well the filters worked, enabling users to easily modify the data and see nearly instantaneous modifications to visualisations.

9. Conclusion

The "Exploring Insights from Synthetic Airline Data Analysis with Qlik" project demonstrates how powerful data analytics can be in the aviation sector and highlights Qlik's extensive capabilities for turning complicated datasets into insights that can be put to use. This study has found important patterns, trends, and relationships by carefully examining synthetic airline data. These findings have the potential to have a big influence on the decision-making processes inside the airline industry.

Stakeholders are able to obtain vital information on passenger demographics, flight statuses, operational efficiencies, and revenue optimisation initiatives by means of thorough data visualisation and dashboarding. These insights enable airport authorities and airlines to make well-informed decisions to meet the changing demands of the sector, as well as to optimise operational efficiency, increase customer satisfaction, and improve pricing techniques.

The project also demonstrates Qlik's efficacy as an analytics platform by showing off its capacity to manage huge datasets, carry out intricate analyses, and provide user-friendly, interactive visualisations. Organisations may leverage data-driven decision-making to promote innovation, improve performance, and maintain competitiveness in the ever-changing aviation industry by utilising Qlik's analytical tools and features.

The research shows the transformational potential of data analytics in influencing the future of the aviation sector in addition to offering insightful information about the synthetic airline data. In the quickly changing aviation industry, stakeholders may overcome obstacles, seize opportunities, and promote sustainable growth and success by adopting data-driven strategies and utilising cutting-edge analytics tools like Qlik.