

# Exploring Insights From Synthetic Airline Data Analysis With Qlik

---

Prepared by  
Pratik Nikhade

# Index

S.No	Title	Page No
1	<b>Introduction</b>	2
2	<b>Problem Understanding</b>	4
3	<b>Data Collection</b>	8
4	<b>Data Preparation</b>	10
5	<b>Data Visualisations</b>	12
6	<b>Dashboard</b>	14
7	<b>Report: Overview of Airline Passenger Data Analysis</b>	17
8	<b>Performance Testing</b>	19
9	<b>Conclusion</b>	20
10	<b>References</b>	21

## **1. Introduction**

### **1.1 Overview**

The project titled "Exploring Insights from Synthetic Airline Data Analysis with Qlik" will be empowering the potential benefits of Qlik Tool for preparing various interactive dashboards a top business intelligence and data visualisation platform. This data includes flight schedules, passenger demographics, ticket sales, and performance measures, and it is intended to mimic real-world airline operations. The goal of the project is to find hidden patterns, trends, and correlations in this artificial dataset by utilising Qlik's powerful analytical tools. Enhancing decision-making procedures for airlines, airports, and other aviation industry stakeholders is made possible by these insights.

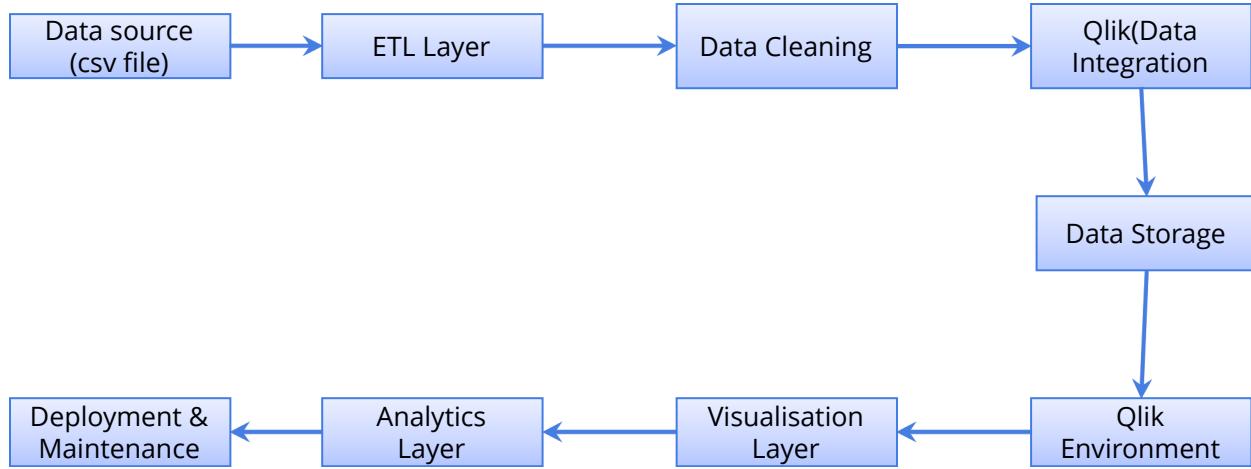
The primary objective is to enhance decision-making processes for airlines, airports, and other stakeholders in the aviation industry. Through detailed visualizations and data-driven insights, the project seeks to optimize revenue strategies, improve operational efficiency, and elevate the customer experience. These insights can lead to more informed strategic decisions, better resource allocation, and personalized marketing efforts, ultimately driving higher profitability and customer satisfaction. The project exemplifies how synthetic data and sophisticated analytics can be harnessed to simulate real-world scenarios and provide actionable intelligence in the highly competitive aviation sector.

### **1.2 Purpose of the Project**

The purpose of the project "Exploring Insights from Synthetic Airline Data Analysis with Qlik" is to utilize the capabilities of Qlik, a leading business intelligence and data visualization platform, to analyze synthetic airline data. This analysis aims to uncover hidden patterns, trends, and correlations within the dataset to enhance decision-making processes for stakeholders in the aviation industry, including airlines, airports, and other related entities.

1. Enhanced Revenue Strategies
2. Improved Operational Efficiency
3. Elevated Customer Experience
4. Informed Strategic Decision-Making
5. Better Resource Allocation
6. Targeted Marketing Efforts
7. Predictive and What-If Analysis
8. Real-Time Data Updates and Interactive Dashboards

### 1.3 Technical Architecture



1. Data Sources and Collection : Generate and gather synthetic data representing various aspects of airline operations.
2. Data Integration and ETL (Extract, Transform, Load): Extract, transform, and load data from multiple sources into a consistent, centralized repository.
3. Data Storage and Management: Store processed data in a scalable and efficient centralized data repository.
4. Qlik Environment Setup: Configure Qlik Sense or QlikView to connect to the data repository and import data.
5. Data Visualization and Dashboard Development: Design and develop interactive dashboards to visualize key metrics and trends.
6. Analytics and Insights: Utilize Qlik's advanced analytics capabilities to uncover patterns and forecast trends.
7. Deployment and Maintenance: Deploy Qlik applications and dashboards, ensuring ongoing monitoring and maintenance.

## **2. Problem Understanding**

### **2.1 Business Problem**

In a fiercely competitive market, the airline sector must maximise revenue while improving customer happiness and operational efficiency. These difficulties involve controlling flight itineraries, comprehending the demographics of passengers, forecasting demand, and enhancing performance indicators. To make well-informed decisions, efficiently manage resources, and customise marketing campaigns, airlines and airports require data-driven insights. However, it is challenging to find hidden patterns and connections using conventional analysis approaches due to the complexity and volume of data gathered from several sources.

Addressing the Business Problems:

The project "Exploring Insights from Synthetic Airline Data Analysis with Qlik" aims to tackle these business problems by leveraging Qlik's powerful analytical tools and interactive dashboards. By analyzing synthetic data that mimics real-world airline operations, the project seeks to provide actionable insights that can help airlines and airports:

- Develop data-driven revenue strategies.
- Optimize flight schedules and resource allocation.
- Enhance customer experiences through personalized services.
- Accurately forecast demand and plan capacities.
- Make informed strategic decisions based on comprehensive data analysis.
- Implement targeted marketing efforts to increase customer engagement and loyalty.

Ultimately, these insights will empower stakeholders in the aviation industry to improve operational efficiency, increase profitability, and elevate customer satisfaction in a highly competitive market.

## 2.2 Business Requirements

The business requirements for "Exploring Insights from Synthetic Airline Data Analysis with Qlik" encompass the need for realistic synthetic datasets representing various facets of airline operations, including flight schedules, passenger demographics, ticket sales, and performance measures. Ensuring data consistency and quality is crucial for dependable analysis, which calls for the deployment of a scalable and effective data storage system. Aggregating various data sources into a centralised repository and preparing the data for Qlik analysis need thorough data integration and transformation processes. Delivering user-friendly interactive dashboards with real-time data updates and a variety of visualisations to emphasise important metrics and trends is a requirement of the project. Actionable insights for revenue optimisation, improvements in customer experience, operational efficiency, precise demand forecasts, help for strategic decision-making, resource allocation optimisation, and targeted marketing insights are among the functional requirements. Strong technological and security controls, along with thorough user training and feedback systems, are also essential for the effective implementation and use of the analytical solution by aviation industry stakeholders.

### 1. Data Requirements

**Synthetic Data Generation:** Develop realistic synthetic datasets representing flight schedules, passenger demographics, ticket sales, and performance measures.

**Data Quality:** Ensure high-quality, clean data with accurate, consistent, and complete records.

**Data Storage:** Use a scalable and efficient data storage solution to manage large volumes of synthetic data.

### 2. Analytical Requirements

**Data Integration:** Integrate data from multiple sources into a centralized repository using ETL processes.

**Data Transformation:** Transform data into a consistent format suitable for analysis in Qlik.

**Advanced Analytics:** Utilize Qlik's advanced analytics capabilities to uncover hidden patterns, trends, and correlations.

### 3. Visualization and Dashboard Requirements

**Interactive Dashboards:** Develop user-friendly, interactive dashboards in Qlik to visualize key metrics and trends.

Real-Time Data Updates: Ensure dashboards provide real-time data updates for accurate and timely decision-making.

#### 4. Functional Requirements

Revenue Optimization: Provide insights to develop dynamic pricing strategies and maximize ticket sales.

Operational Efficiency: Identify inefficiencies and recommend improvements in flight scheduling, crew management, and resource allocation.

Customer Experience: Analyze passenger data to enhance personalized services and improve customer satisfaction.

Demand Forecasting: Use predictive analytics to accurately forecast travel demand and plan capacities accordingly.

#### 5. Technical Requirements

Qlik Environment Setup: Configure and deploy Qlik Sense or QlikView on a secure server or cloud environment.

Data Connectivity: Establish secure and reliable connections between Qlik and the centralized data repository.

Scalability: Ensure the architecture can scale to handle growing data volumes and analytical demands.

Performance Monitoring: Implement tools to monitor the performance of Qlik applications and dashboards.

#### 6. User Requirements

Stakeholder Identification: Identify all key stakeholders, including airline management, operations teams, marketing departments, and airport authorities.

User Training: Provide training for stakeholders to effectively use Qlik dashboards and interpret insights.

User Feedback: Implement mechanisms for stakeholders to provide feedback on the dashboards and analytical tools to continuously improve the solution.

## 2.3 Literature Survey

The project "Exploring Insights from Synthetic Airline Data Analysis with Qlik" is underpinned by a thorough literature survey across various domains critical to its success. Drawing from a plethora of research papers, the project taps into a rich vein of knowledge encompassing synthetic data generation, data quality and preparation methodologies, business intelligence and data visualization strategies, advanced analytics and predictive modeling techniques, operational efficiency and revenue management insights, customer experience enhancement approaches, strategic decision-making frameworks, data security and governance best practices, and user adoption and training strategies. For instance, studies such as Generating Realistic Data for Smart Cities: A Case Study of Urban Mobility Analytics by G. de Felice et al. and Predictive Analytics in Airline Industry: A Comprehensive Review by N. Agarwal et al. provide valuable insights into synthetic data generation and predictive analytics applications in the airline industry, respectively. Additionally, papers like Data Preprocessing Techniques for Data Mining: An Overview by S. Singh et al. and Business Intelligence and Analytics: From Big Data to Big Impact by M. Cordeiro et al. shed light on data preprocessing methodologies and the transformative potential of business intelligence tools. This synthesis of research literature serves as the bedrock for the project's methodology, ensuring alignment with industry best practices and cutting-edge research trends, and positioning it for success in exploring actionable insights from synthetic airline data with Qlik.

Navuluri Madhavilatha et al. authored a paper titled "Airline Data Analysis," which delves into the challenges of contemporary data analysis and the importance of effective analytics in commercial aviation decision-making processes.

### **3. Data Collection**

The dataset was downloaded from Kaggle which is website that provides datasets to work with. (<https://www.kaggle.com/datasets/iamsouravbanerjee/airline-dataset/data>). This dataset includes a variety of parameters related to international airline operations.

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 Flight Status.

All of these columns combined offer deep insights into flight statuses, crew information, flight routes, travel information, and passenger demographics. This information can be used by academics and professionals in the field to examine passenger behaviour patterns, improve traveller experiences, assess pilot performance, and improve overall flight operations.

- 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)

Once the dataset is ready it was loaded into Qlik Sense Cloud based platform with the following steps-

1. Sign in to Qlik Sense Cloud: Visit the Qlik Sense Cloud website (<https://www.qlik.com/us/trial/qlik-sense-cloud>), and sign in to your Qlik Sense Cloud account. If you don't have an account, you can sign up for free.
2. Access the Qlik Sense Cloud Hub: Once signed in, you'll be directed to the Qlik Sense Cloud Hub. This is your main dashboard for managing apps and data.
3. Create a New App: Click on the "Create new" button to create a new app.
4. Add Data to the App: In the newly created app, click on the "Add data" button. You'll be prompted to choose a data source. Qlik Sense Cloud allows you to load data from various sources such as files, web files, databases, and cloud storage.
5. Select Data Source: Choose the appropriate option based on where your dataset is located. For example: If your dataset is stored locally on your computer, select "File" and upload the file.
6. Configure Data Load Settings: Depending on the data source, you may need to configure additional settings such as file format, delimiter, data structure, etc.
7. Preview and Load Data: Once the data source is configured, Qlik Sense Cloud will preview the data. Review the preview to ensure it looks correct. Click on the "Load data" button to load the dataset into your Qlik Sense Cloud app.
8. Save and Publish the App: Once you're satisfied with the data loading and modeling, save the app. Optionally, you can publish the app to share it with other users in your Qlik Sense Cloud environment.

Dataset_Final_1 Dataset_Final_1.csv							Fields: 18	More
Passenger ID	First Name	Last Name	Gender	Age	Nationality	Airport Name	Airport Cou...	Actions
165488	Dionis	Joist	Female	4	Philippines	Bremen Airport	DE	More
824967	Federica	Peters	Female	16	Portugal	Watson Lake Airport	CA	More
0a1ws9	Lammond	Sargood	Male	59	Serbia	Karluk Lake Seaplane Base	US	More
0A1y0a	Arel	Beswick	Male	83	United States	Holy Cross Airport	US	More
0A20Yl	Jack	Mitrikhin	Male	89	Turkey	Rottnest Island Airport	AU	More
0a5x5P	Perry	Pretsell	Male	64	Nigeria	Vallenar Airport	CL	More

Fig.1 Overview of data after loading in Qlik Sense

#### **4. Data Preparation**

The data preparation phase is a crucial step in the data analysis process where raw data is transformed, cleaned, and structured to make it suitable for analysis. This phase involves several key steps:

1. Data Exploration: Explore the collected data to understand its structure, format, and quality. Identify any missing values, outliers, duplicates, or inconsistencies in the data. The dataset did not contain any missing values or outliers.
2. Data Cleaning: Handle missing data: Impute missing values using methods like mean, median, or interpolation, or remove records with missing values if appropriate. Remove duplicates: Identify and remove duplicate records to ensure data integrity. Standardize data formats: Ensure consistency in data formats across different variables (e.g., date formats, units of measurement). Handle outliers: Identify and deal with outliers that may skew the analysis results.

The dataset was in CSV format. Once the dataset was downloaded it was opened in excel and different parameters were studied regarding. The date format was changed and was replaced with suitable format in Excel as the date formats was different.

3. Data Transformation: Feature engineering: Create new features or variables from existing ones to capture additional information or improve model performance.
  - Data aggregation: Aggregate data at different levels (e.g., daily, weekly, monthly) to simplify analysis and reduce complexity. New columns were created from Flight status to On Time Flights, Cancelled, Delayed Flights for further analysis.

In the Add Field option Calculated Field was used to create new Columns by giving expressions.

4. Data Formatting: Format the data according to the requirements of Qlik tool and ensured consistency in data types, variable names, and data structures to facilitate analysis and interpretation.
5. Data Validation: Validated the cleaned and transformed data to ensure it accurately represents the real-world phenomena or processes being studied.

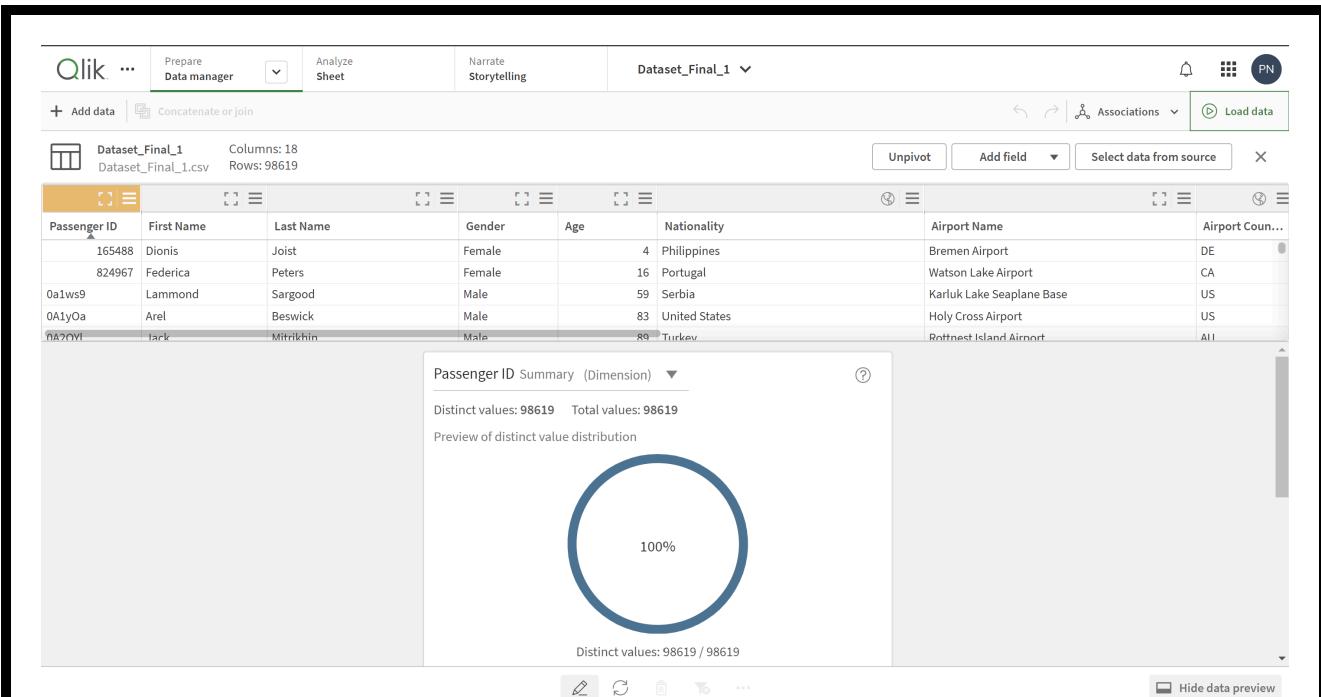


Fig. 2 Represents the data in Data Manager mode for pre-processing steps

The total columns in data is 18.

The total records/rows are 98619.

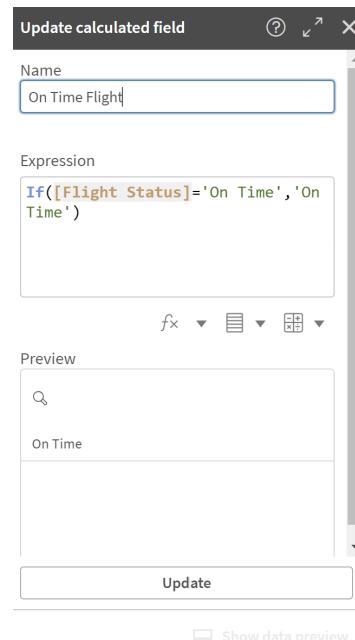


Fig.3 Represents Calculated Field

## **4. Data Visualization**

In Qlik Sense, creating data visualizations is a easy process as it is all GUI based. Here are the following steps to perform Data Visualization.

1. Data Loading: Before we can create visualizations, we need to load the data into Qlik Sense. This involves uploading or connecting to your data source and configuring the data load script to import the necessary fields and tables.
2. Open the App: Once the data is loaded, open the Qlik Sense app where we want to create the visualization.
3. Enter Edit Mode: Click on the "Edit" button at the top-right corner of the screen to enter edit mode.
4. Add a Sheet: In the left sidebar, click on the "Sheets" tab, then click on the "+" icon to add a new sheet.
5. Add a Visualization: On the new sheet, click on the "Add new visualization" button (the plus icon) to start creating visualization.
6. Choose Visualization Type: Qlik Sense offers a variety of visualization types, including bar charts, line charts, pie charts, tables, maps, and more. Choose the appropriate visualization type for your data.
7. Select Data Fields: Drag and drop the data fields you want to visualize onto the visualization canvas. Qlik Sense will automatically generate a basic visualization based on your selection.
8. Customize the Visualization: Use the properties panel on the right side of the screen to customize the appearance and behavior of your visualization. You can adjust settings such as colors, labels, axes, sorting, and filtering to tailor the visualization to your preferences.
9. Interact with the Visualization: Once your visualization is created, you can interact with it in various ways. Hover over data points to view details. Click on data points to filter other visualizations on the sheet. Use selections in other visualizations to dynamically update the data displayed in your visualization.
10. Save and Publish: After creating and customizing your visualization, click on the "Save" button to save your changes.

The screenshot shows the Qlik Sense Analyze Sheet interface. The left sidebar contains 'Fields' (Search, Show by table, All tables), 'Master items', and a 'Properties' panel with sections for 'Data' (First KPI (Measures: On Time Flights), Second KPI (Measures: + Add, Drop field to use or for a suggestion)), 'Visualization' (#1 KPI, #1), 'Filters' (+ Add, Drop field to filter visualization), and 'Presentation' (Labels, Tooltip). The main area displays a dashboard titled 'Dashboard 3' with the following components:

- FLIGHT OPERATIONS** section: A table with 'Airport Name' and three KPIs: 'On Time Flights 32.85k', 'Cancelled Flights 32.94k', and 'Delayed Flights 32.83k'.
- Passengers vs Flight Status** section: A bar chart showing passenger counts from 0 to 35k.
- Details of Passengers** section: A table showing passenger details (e.g., First Name, Last Name, Gender, Age, Nationality, Airport) with a total of 98619 passengers.

At the bottom, it shows 'Source table Dataset\_Final\_1' and '1 - 100 of 98619'.

Fig.4 Here in Analyse Sheet various visualisations can be created and analysed

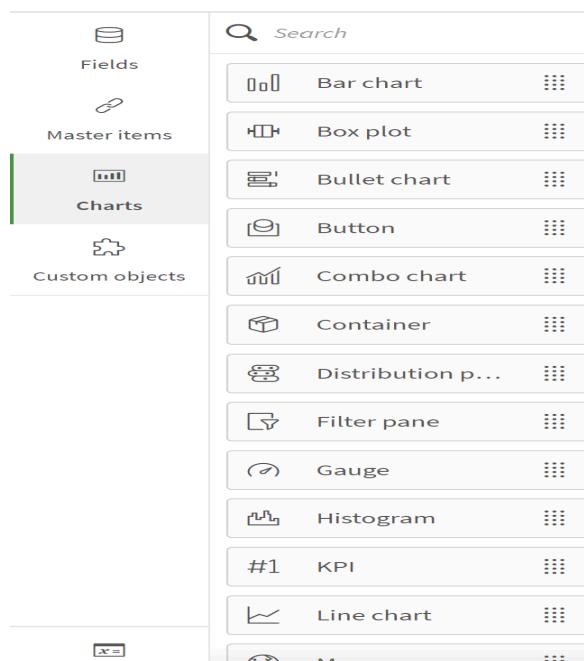


Fig.5 Various Charts can be selected from the following options

## 6. Dashboard

Interactive dashboards can be designed in Qlik sense with various options like Filter Pane. Which helps user to easily navigate .In Qlik Sense, dashboard creation entails combining many sheets and visualisations into a single interface for data analysis and exploration. Users may easily mix different visualisations, including tables, bar charts, and line graphs, with text boxes, photos, and filter panes within Qlik Sense's user-friendly interface to create dynamic and educational dashboards. With the help of these dashboards, users may make well-informed decisions by gaining insights from data through the visual representation of important metrics, trends, and relationships.

The style, appearance, and interactivity of Qlik Sense dashboards can be customised to meet a variety of data analysis requirements, providing flexibility and scalability. Moreover, users can leverage Qlik Sense's powerful features, such as selections and navigation buttons, to enhance usability and facilitate seamless navigation within the dashboard.

Whether it's monitoring KPIs, tracking business performance, or conducting ad-hoc analysis, Qlik Sense dashboards provide a comprehensive solution for leveraging data-driven insights to drive organizational success.

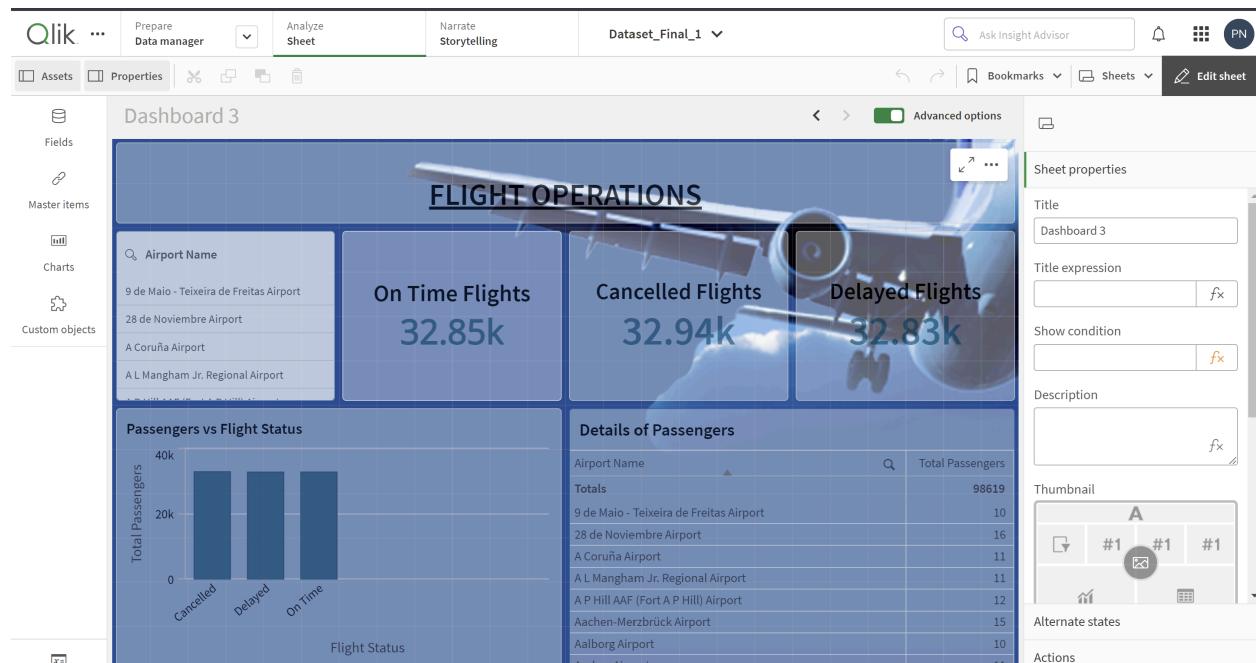


Fig. 6 The advanced option is enabled for making changes in the Dashboard

Once all the changes are made disable advanced mode option and click on edit sheet.

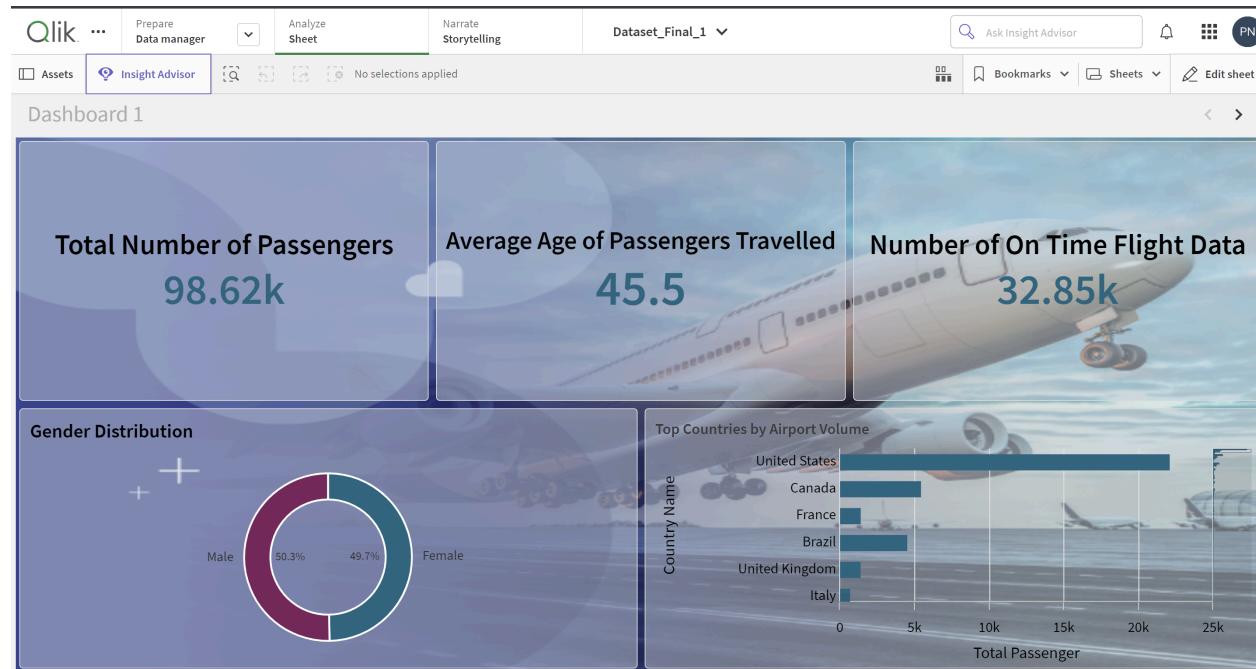


Fig.7 Dashboard 1

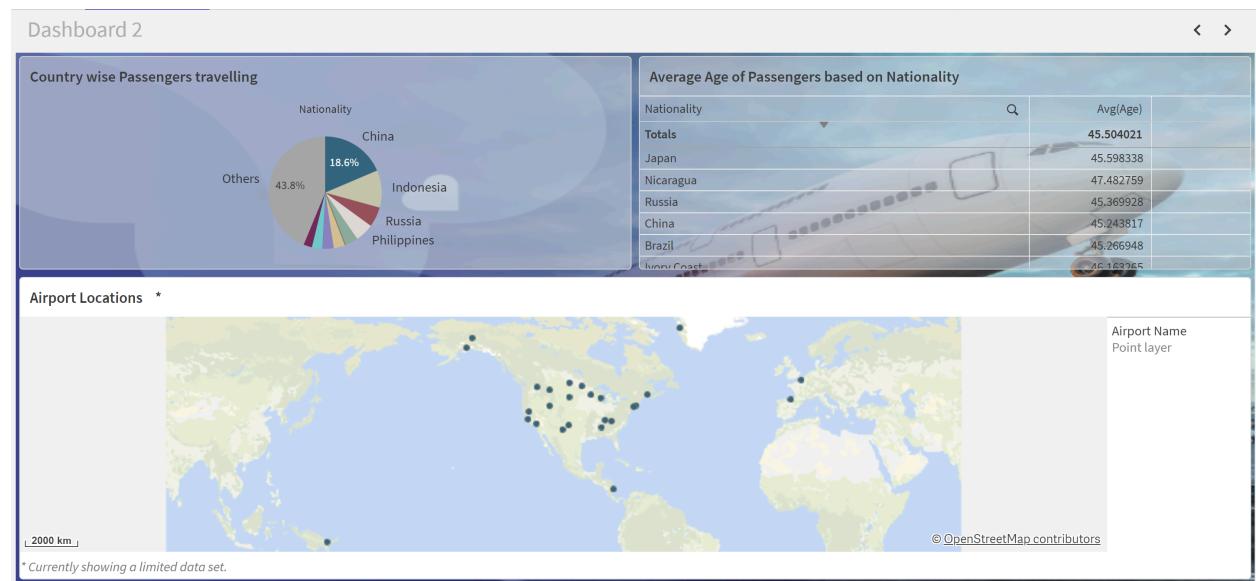
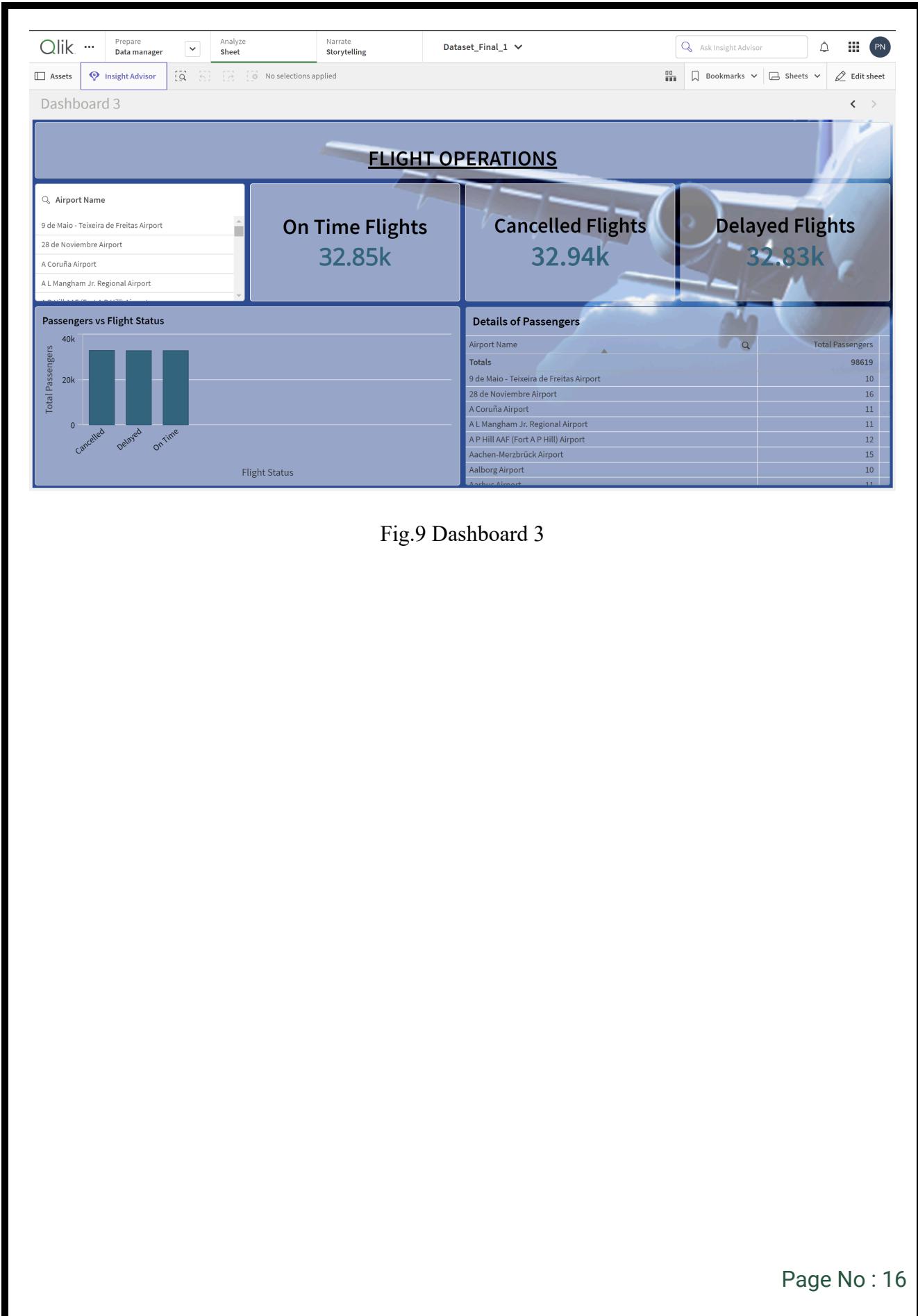


Fig.8 Dashboard 2



## **7. Report: Overview of Airline Passenger Data Analysis**

The report offers a comprehensive analysis of airline passenger data, encompassing key metrics vital for understanding travel patterns, passenger demographics, and flight operations. With approximately 98,000 passengers in the dataset, the analysis provides a detailed overview of travel activity and passenger profiles. Insights into the average age of passengers, on-time flight performance, gender distribution, and top countries and origins of passengers inform strategic decision-making for service customization, marketing strategies, and operational planning. Additionally, the report highlights the significance of analyzing flight operations data for enhancing efficiency and passenger satisfaction. By leveraging these insights, airlines and airports can optimize resources, tailor services to diverse passenger needs, and ultimately improve the overall travel experience.

1. Total Number of Passengers: The dataset comprises approximately 98,000 passengers, offering a comprehensive view of travel activity and passenger data available for analysis. This metric establishes the dataset's scope and scale, crucial for understanding the magnitude of travel patterns.
2. Average Age of Passengers: Understanding the average age of passengers provides insights into the primary demographic, guiding service customization and marketing strategies. Analysis over time can unveil demographic shifts, aiding strategic planning and resource allocation.
3. On-time Flights: The number of on-time flights reflects the reliability and punctuality of flight operations, assessing the efficiency of airlines and airports in adhering to scheduled departure and arrival times.
4. Gender Distribution: Examining the gender distribution among travelers offers insights into demographic balance, informing marketing and service strategies tailored to diverse passenger profiles.
5. Top Countries with Airports: Analysis reveals that the United States hosts the highest number of airports, followed by Canada, France, Brazil, the United Kingdom, and Italy, indicating key geographical hubs for air travel.
6. Top Passenger Origins: China, Indonesia, and Russia emerge as significant sources of passenger traffic, highlighting key travel hubs and peak travel times for resource allocation and operational planning.

7. Average Age of Passengers by Nationality: Delving into age demographics based on nationality allows for tailored services and marketing strategies catering to the preferences of travelers from different countries.
8. Flight Operations Analysis: Analyzing flight data, including punctuality, cancellations, and delays, is pivotal for enhancing operational efficiency, passenger satisfaction, and overall airline performance.
9. Passenger Details at Specific Airports: Detailed information on passenger counts at particular airports offers insights into travel patterns and facilitates strategic decision-making for airport management.

## **8. Performance Testing**

Performance testing in Qlik Sense involves assessing the system's ability to handle large volumes of data and user interactions while maintaining optimal performance levels. This testing evaluates the responsiveness and scalability of the Qlik Sense platform under various scenarios, such as loading large datasets, executing complex calculations, rendering visualizations, and supporting concurrent user access.

One crucial aspect of performance testing in Qlik Sense is measuring the amount of data rendered by the platform within acceptable response times. This metric helps determine the platform's capacity to efficiently process and display data-intensive visualizations to users. The amount of data rendered can vary depending on factors such as the complexity of the data model, the number of dimensions and measures in visualizations, and the level of aggregation applied.

Performance testing typically involves simulating real-world usage scenarios and load conditions to identify potential bottlenecks and areas for optimization. By measuring the amount of data rendered and analyzing response times under different usage scenarios, organizations can ensure that their Qlik Sense deployments can effectively support the needs of users and scale to accommodate growing data volumes.

The amount of data rendered in Qlik was around 98k. The response time was also quick in Qlik. There were instances when Calculated Field was being created it took some time to load but it was negligible.

Data Filters were used to ease of user so that the user can select the data of particular choice and the results would be displayed according to the desired users input. Data filters in analytics platforms like Qlik Sense enable users to interactively select and manipulate data based on their preferences, focusing on relevant subsets for analysis. They provide flexibility by allowing users to specify criteria such as time periods, regions, or product categories, tailoring analyses to specific needs. Real-time feedback facilitates dynamic adjustments, enhancing exploration and insight discovery. Filters simplify navigation through complex datasets, reducing cognitive load and promoting ease of use. Collaboration is facilitated through the ability to save and share filter selections, promoting data-driven decision-making across organizations.

## **9. Conclusion**

In the realm of synthetic airline data analysis, Qlik Sense emerges as an indispensable ally, offering a suite of powerful analytical tools and intuitive visualization features. As outlined, the project's primary objective revolves around enhancing decision-making processes for airlines, airports, and other stakeholders in the aviation industry. Qlik Sense plays a pivotal role in achieving this goal by providing a platform to explore insights derived from synthetic datasets mimicking real-world airline operations.

With its robust analytical capabilities, Qlik Sense enables stakeholders to delve deep into the synthetic data, uncovering hidden patterns, trends, and correlations. The platform's ability to handle large volumes of data facilitates comprehensive analysis, providing stakeholders with a holistic view of various aspects of airline operations, including flight schedules, passenger demographics, ticket sales, and performance measures.

Moreover, Qlik Sense's intuitive visualization features allow stakeholders to translate complex data into interactive dashboards and visualizations. These dashboards offer real-time insights, empowering decision-makers to make informed choices that drive business outcomes. Whether optimizing revenue strategies, improving operational efficiency, or enhancing the customer experience, Qlik Sense provides the tools necessary to derive actionable intelligence from synthetic airline data.

Furthermore, Qlik Sense's versatility extends to its ability to simulate real-world scenarios, thanks to its integration with synthetic data. By generating synthetic datasets that closely resemble actual airline operations, stakeholders can simulate various scenarios and evaluate the potential impact of different strategies and decisions.

In essence, Qlik Sense exemplifies how synthetic data and sophisticated analytics can be leveraged to simulate real-world scenarios and provide actionable intelligence in the highly competitive aviation sector. Through detailed visualizations and data-driven insights, the platform empowers stakeholders to optimize revenue strategies, improve operational efficiency, and elevate the customer experience, ultimately driving higher profitability and customer satisfaction in the aviation industry.

## **10. References**

1. <https://www.kaggle.com/datasets/iamsouravbanerjee/airline-dataset/data>
2. <https://ijsrcseit.com/home/issue/view/article.php?id=CSEIT19514>
3. <https://www.novypro.com/project/indias-airline-data-analysis>