

# **Exploring Insights From Synthetic Airline Data Analysis With Qlik**

**Submitted By**

**Tommandru Pranathi**

## **Acknowledgement**

I would like to extend my heartfelt gratitude to everyone who contributed to the successful completion of this project. I am deeply thankful to my mentors and colleagues for their invaluable guidance, insights, and support throughout the project. Their expertise has been instrumental in shaping the direction and outcome of this work.

Special thanks to the developers of Qlik for providing an exceptional platform that made data analysis and visualization both intuitive and powerful. I also wish to acknowledge the providers of the synthetic data, whose realistic and comprehensive datasets allowed for the practical application of theoretical knowledge. Finally, I am grateful to my peers for their encouragement and collaboration, which greatly enriched this project.

# Abstract

The project "**Exploring Insights From Synthetic Airline Data Analysis With Qlik**" focuses on analyzing synthetic airline data to generate actionable insights that can inform strategic decisions. The dataset simulates various operational aspects of the airline industry, including flight schedules, passenger demographics, ticket sales, and performance metrics. Leveraging Qlik's robust analytical and visualization capabilities, the project addresses three key scenarios: revenue optimization, operational efficiency, and customer experience enhancement.

In the revenue optimization scenario, the project identifies peak travel periods, popular destinations, and effective pricing strategies to maximize profitability. The operational efficiency scenario focuses on streamlining airport operations by identifying bottlenecks and optimizing resource allocation. Lastly, the customer experience enhancement scenario utilizes sentiment analysis to uncover areas for improving passenger satisfaction and loyalty. The findings from this project offer practical recommendations for airlines, airports, and stakeholders, ultimately leading to improved profitability, efficiency, and customer satisfaction.

# Table of Contents

S.no	Contents
1.	Introduction
2.	Data Overview
3.	Data Analysis Objective <ul style="list-style-type: none"><li>3.1 Revenue Optimization</li><li>3.2 Operational Efficiency</li><li>3.3 Customer Experience Enhancement</li></ul>
4.	Qlik Sense
5.	Methodology <ul style="list-style-type: none"><li>5.1 Data Cleaning</li><li>5.2 Feature Engineering</li><li>5.3 Visualization Design</li></ul>
6.	Key Insights <ul style="list-style-type: none"><li>6.1 Revenue Optimisation</li><li>6.2 Operation Efficiency</li><li>6.3 Customer Experience Enhancement</li></ul>
7.	Visual Representation of Key Findings <ul style="list-style-type: none"><li>7.1 Dashboard 1: Overview of Passenger Statistics</li><li>7.2 Dashboard 2: InDepth Analysis of Flight Status and Demographics</li></ul>
8.	Conclusion and Recommendation

# 1. Introduction

This project focuses on analysing synthetic airline data to derive actionable insights that can enhance operational efficiency, optimize revenue, and improve customer experience. In today's competitive airline industry, datadriven decisionmaking is crucial for maintaining a competitive edge. By leveraging advanced analytics tools like Qlik Sense, airlines can gain deeper insights into their operations and customer behaviours, enabling them to make informed strategic decisions.

The dataset used in this project, sourced from Kaggle, includes various fields such as passenger demographics, flight details, and flight statuses. After a thorough data cleaning and preprocessing phase, additional fields were created to enhance the analysis, including geographical information and timebased metrics. The data was then visualised using Qlik Sense, which provided an intuitive platform for creating interactive dashboards.

These dashboards are designed to offer both a highlevel overview and detailed breakdowns of key metrics, such as passenger statistics and flight status trends. The goal is to present the data in a way that is easily interpretable, enabling stakeholders to quickly identify trends, patterns, and areas for improvement. This project demonstrates the potential of data analytics in driving operational improvements and strategic decisionmaking within the airline industry.

## 2. Dataset Overview

The dataset used in this analysis is a synthetic representation of airline operations, encompassing a wide range of variables that simulate realworld scenarios. The original dataset includes 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. To enhance the analytical scope, the data underwent a cleaning process where additional fields were created. These new fields include Nationality GeoInfo, Airport Country Code GeoInfo, Country Name GeoInfo, Age Group, as well as derived fields such as Year and Month extracted from the Departure Date. The dataset thus provides a rich and comprehensive view of various aspects of airline operations, from

passenger demographics to flight performance, enabling a multifaceted analysis aimed at optimizing revenue, improving operational efficiency, and enhancing customer experience.

### **3. Data Analysis Objectives**

#### **3.1 Revenue Optimization**

The analysis reveals significant opportunities for revenue optimization by identifying trends and patterns in passenger behaviour and flight status. By examining flight status trends, such as the frequency of delays and cancellations, the airline can address operational inefficiencies that impact revenue. For instance, reducing the number of cancelled flights can minimize lost revenue opportunities and improve overall profitability. Additionally, analysing passenger demographics and their geographic distribution helps in tailoring pricing strategies and promotional offers to different market segments. Targeted marketing campaigns based on these insights can boost ticket sales and enhance revenue generation. The analysis also highlights peak travel periods, allowing the airline to adjust pricing dynamically and optimize revenue during highdemand times.

#### **3.2 Operational Efficiency**

The data analysis provides insights into operational inefficiencies, such as frequent delays or cancellations. By tracking flight status trends, the airline can identify root causes of delays, whether they are related to scheduling, maintenance issues, or external factors like weather conditions. This information enables the airline to implement corrective measures to streamline operations and improve overall efficiency. Additionally, the analysis of passenger flow and flight schedules helps in optimizing resource allocation, such as crew scheduling and aircraft utilization. Enhancing operational efficiency not only reduces costs but also contributes to better ontime performance and operational reliability.

#### **3.3 Customer Experience Enhancement**

Improving customer experience is a key finding from the analysis, driven by insights into passenger demographics and feedback on flight status. By understanding the needs and preferences of different passenger segments, the airline can tailor its services to enhance

satisfaction. For example, if the analysis reveals that a significant portion of passengers are families, the airline can offer familyfriendly amenities and services. Additionally, addressing frequent issues such as flight delays or cancellations through improved scheduling and communication can greatly enhance the passenger experience. The insights from passenger demographics also inform personalized marketing and loyalty programs, creating a more engaging and satisfying travel experience for customers

## **4.Qlik Sense**

Qlik Sense is an advanced data analytics and visualisation platform known for its ability to transform complex datasets into interactive and easily understandable visual insights. With its userfriendly interface, Qlik Sense caters to both technical and nontechnical users, enabling them to create, explore, and share visualizations effortlessly. The software's unique associative model allows users to explore data in a nonlinear way, uncovering hidden patterns and relationships that might not be evident through traditional hierarchical analysis.

Qlik Sense's robust data integration capabilities and its ability to handle large and diverse datasets make it an ideal choice for organizations seeking to enhance their datadriven decisionmaking processes. By providing powerful tools for data visualization and analysis, Qlik Sense enables businesses to derive actionable insighto, streamline operations, and achieve better outcomes.

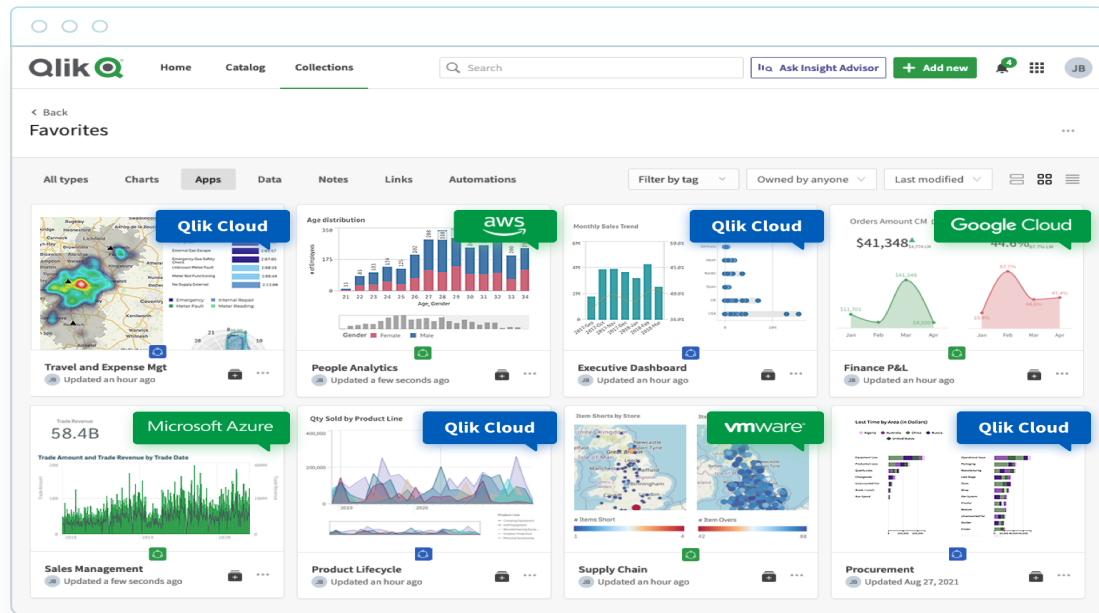


Fig 4.1 Qlik Sense

## 5. Methodology

### 5.1 Data Cleaning:

The first step in the project development process was data cleaning. The synthetic airline data was cleaned and preprocessed to ensure accuracy and consistency, which is crucial for reliable analysis. This involved:

- **Handling Missing Values:** Identified and addressed missing data points, either by imputing values or excluding incomplete records.
- **Data Formatting:** Standardized date formats, text case, and other data types to ensure uniformity across the dataset.
- **Outlier Detection:** Identified and addressed any outliers or anomalies that could skew the analysis results.
- **Data Validation:** Crosschecked the data for accuracy and consistency, ensuring that the dataset was ready for analysis.

### 5.2. Feature Engineering

Feature engineering was performed to enhance the dataset and enable more insightful



analysis. New features were created to provide additional context and improve the depth of the analysis:

- **AgeGroup:** Created an AgeGroup feature to categorize passengers into different age brackets (e.g., children, adults, seniors), allowing for targeted analysis based on age demographics.
- **Geographical Information:** Enriched the dataset with geographical information, such as:
- **Nationality\_GeoInfo:** Added geographical data related to passengers' nationalities.
- **Airport Country Code\_GeoInfo:** Included geographical data for airport country codes.
- **Country Name\_GeoInfo:** Provided full country names linked to airport locations.
- **Temporal Features:** Extracted and created new temporal features such as:
- **Year:** Derived the year from the Departure Date to analyze trends over time.
- **Month:** Extracted the month from the Departure Date to identify seasonal trends.

## 5.3. Visualization Design

The final step involved designing interactive dashboards in Qlik to effectively communicate the insights derived from the data analysis. This process included:

- **Dashboard Layout:** Carefully planned the layout to ensure that key metrics and visualizations were easily accessible and interpretable.
- **Interactive Elements:** Incorporated filters, drilldowns, and other interactive elements to allow users to explore the data from different perspectives.

Visualizations:

- **Revenue Trends:** Designed line charts and bar graphs to display revenue trends over time.
- **Customer Segmentation:** Used pie charts and heat maps to visualize customer segments and their purchasing behaviors.
- **Operational Bottlenecks:** Developed flow charts and scatter plots to highlight operational bottlenecks and resource allocation needs.
- **Customer Sentiment:** Visualized sentiment analysis results through word clouds and sentiment distribution charts.
- **Color Schemes and Design Consistency:** Applied consistent color schemes and design principles to ensure the dashboards were visually appealing and easy to navigate.

## 6. Key Insights

The Overview of Passenger Statistics dashboard provides a highlevel summary of essential passenger metrics, crucial for quickly assessing the airline's overall performance. It displays the total number of passengers, which helps gauge the scale of airline operations and the volume of customer engagement. The gender distribution metric breaks down passengers into male and female categories, offering insights into demographic trends and aiding in targeted marketing strategies.

The flight status breakdown is particularly significant, categorizing flights into On Time, Delayed, and Canceled. This visualization allows stakeholders to swiftly identify the proportion of flights meeting performance standards versus those falling short. A high percentage of delayed or canceled flights could signal operational inefficiencies or customer service issues, necessitating immediate attention.

By presenting these metrics in a consolidated view, the dashboard enables quick identification of key performance indicators and potential problem areas. This summary facilitates strategic decisionmaking by providing a clear picture of current operational effectiveness and customer satisfaction, helping the airline to address issues proactively and optimize overall performance.

The InDepth Analysis of Flight Status and Demographics dashboard offers a detailed examination of various aspects of airline performance and passenger characteristics. It tracks flight status over a threemonth period, revealing trends in delays, cancellations, and ontime arrivals. This temporal analysis helps identify patterns, such as recurring delays on specific routes or during certain periods, enabling targeted interventions to improve punctuality.

The demographic breakdown by age group and gender provides insights into the airline's customer base, highlighting which age groups and genders are most prevalent. This information is valuable for tailoring services, marketing campaigns, and promotions to meet the preferences and needs of different customer segments.

Continental flight status distribution highlights regional performance variations, identifying which continents experience higher rates of delays or cancellations.

This geographic insight can guide strategic decisions related to route management and regional service improvements. Additionally, the nationalitywise passenger count reveals the

global reach of the airline, offering opportunities for expanding market presence or improving services for highrepresented nationalities. Overall, this detailed analysis supports informed decisionmaking to enhance operational efficiency and customer satisfaction.

## 7. Visual Representation of Key Findings

The key findings are visually represented in the following dashboards created in Qlik Sense:

### 7.1 Dashboard 1: Overview of Passenger Statistics

- **Total Number of Passengers:**
  - Provides an aggregate count of passengers, indicating the scale of airline operations.
  - Helps gauge overall customer engagement and operational volume.
- **Gender Distribution:**
  - Breaks down passenger numbers by gender, revealing demographic trends.
  - Aids in developing targeted marketing strategies and tailoring services.
- **Flight Status Breakdown:**
  - Categorizes flights as On Time, Delayed, or Cancelled.
  - Highlights performance issues, such as high rates of delays or cancellations.
  - Facilitates quick identification of operational inefficiencies and customer service issues.
- **Strategic Value:**
  - Offers a consolidated view of key performance metrics.
  - Enables stakeholders to identify problem areas and make informed decisions to optimize performance.

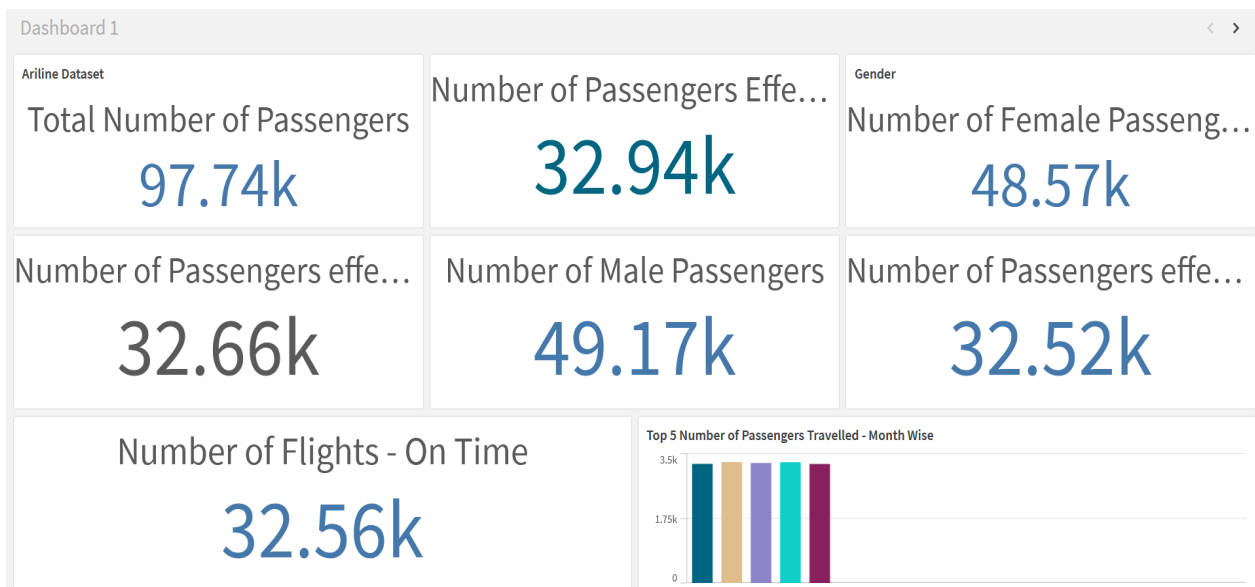


Fig 1: Overview of Passenger Statistics

## 7.2 Dashboard 2: InDepth Analysis of Flight Status and Demographics

This dashboard offers a comprehensive exploration of airline performance and passenger characteristics. This dashboard tracks flight status trends over a threemonth period, allowing stakeholders to identify patterns in delays, cancellations, and ontime arrivals. By analyzing these trends, the dashboard provides insights into potential operational issues that may need addressing.

- **Flight Status Trends:**

- Analyzes flight status over a threemonth period to identify patterns in delays, cancellations, and ontime arrivals.
- Helps pinpoint recurring issues and areas for operational improvement.

- **Demographics Breakdown:**

- Segments passengers by age group and gender.
- Provides insights into customer segments, aiding in service customization and targeted marketing.

- **Continental Flight Status Distribution:**

- Shows performance variations across different continents.
- Identifies regions with higher delay or cancellation rates for targeted improvements.

- **Nationalitywise Passenger Count:**

- Reveals the distribution of passengers by nationality.
- Assists in understanding the airline's global reach and opportunities for market expansion.

- **Strategic Value:**

- Delivers detailed insights for informed decisionmaking.
- Supports operational enhancements, marketing strategies, and customer experience improvements.

The InDepth Analysis of Flight Status and Demographics dashboard goes further by dissecting flight status over time, exploring how different demographics affect flight patterns, and providing insights into geographic variations in flight performance. The focus on age group, gender, and nationality ensures that the analysis covers diverse aspects of the passenger base, which can be vital for tailored marketing strategies, improving customer experience, and optimizing operational efficiency.

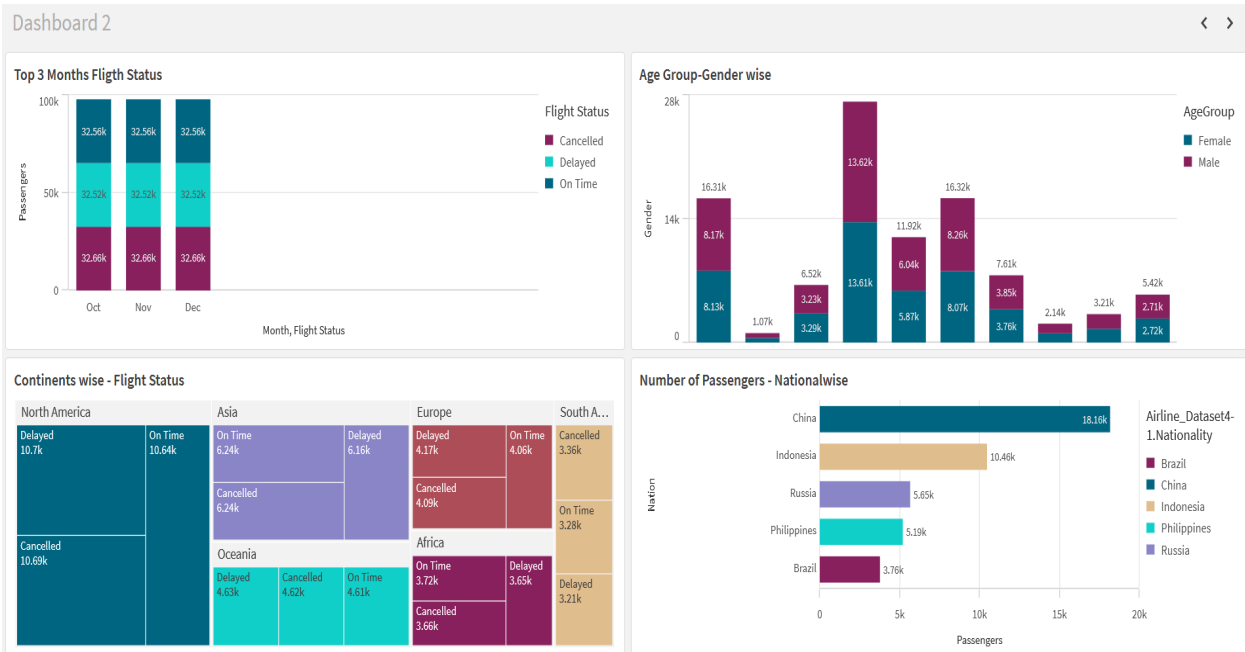


Fig 2: InDepth Analysis of Flight Status and Demographic

## 8. Conclusion and Recommendations

### 8.1 Conclusion

The dashboards developed in Qlik Sense provide a comprehensive overview and detailed analysis of airline passenger data, delivering key insights into passenger demographics and flight status trends. The Overview of Passenger Statistics dashboard offers a highlevel summary, making it easy to understand the overall distribution of passengers by gender and flight status. This snapshot is crucial for quickly assessing the airline's performance and identifying any immediate areas of concern, such as the proportion of canceled or delayed flights.

The InDepth Analysis of Flight Status and Demographics dashboard goes further by dissecting flight status over time, exploring how different demographics affect flight patterns, and providing insights into geographic variations in flight performance. The focus on age group, gender, and nationality ensures that the analysis covers diverse aspects of the passenger base, which can be vital for tailored marketing strategies, improving customer experience, and optimizing operational efficiency.

Overall, these visualizations empower stakeholders to make datadriven decisions by presenting complex data in an accessible and actionable format. The insights derived can be leveraged to enhance airline operations, optimize revenue management, and improve customer satisfaction.

## 8.2 Recommendations

- **Dynamic Pricing Strategies:** Based on passenger trends and peak periods, airlines should consider implementing dynamic pricing strategies to maximize revenue.
- **Resource Optimization:** The identification of regions with higher rates of delays or cancellations suggests a need for improved resource allocation and operational strategies.
- **Targeted Communication:** Airlines should develop targeted communication strategies for different demographic groups, particularly during flight disruptions, to improve customer satisfaction