

**A PROJECT REPORT**

**ON**

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

**(Batch May - June 2024)**

**SUBMITTED**

**BY**

**DODDI GOPINADH**

(Skill Wallet ID: SWUID20240006338)

**IN**

**Qlik Platform**



## CONTENTS

### **1. Introduction**

- 1.1. Overview: A brief description of the project
- 1.2. Purpose: The use of this project. What can be achieved by doing this
- 1.3. Technical Architecture

### **2. Define Problem/ Problem Understanding**

- 2.1. Specify Business Problem
- 2.2. Business Requirements
- 2.3. Literature Survey
- 2.4. Social/Business Impact

### **3. Data Collection**

- 3.1. Collect the data set
- 3.2. Connect the Data with Qlik Sense

### **4. Data Preparation**

- 4.1. Prepare the data for Visualization

### **5. Data Visualization**

- 5.1. Visualization

### **6. Dashboard**

- 6.1. Responsive and Design of the Dashboard
- 6.2. Design of story

### **7. Report**

- 7.1. Report Creation

### **8. Performance Testing**

- 8.1. Amount of Data Rendered
- 8.2. Utilization of Data Filters

# 1. INTRODUCTION

## 1.1 Overview:

The project titled "Exploring Insights from Synthetic Airline Data Analysis with Qlik" aims to utilize synthetic airline data to extract valuable insights through Qlik, a business intelligence and data visualization tool. The dataset simulates various aspects of airline operations, such as flight schedules, passenger demographics, ticket sales, and performance metrics. The objective is to leverage Qlik's analytical capabilities to identify patterns, trends, and correlations within the data, aiding decision-making processes for airlines, airports, and other related stakeholders.

## 1.2 Purpose

The purpose of this project is to leverage the analytical capabilities of Qlik to address key areas in airline operations, leading to significant business improvements. This is achieved through four main objectives:

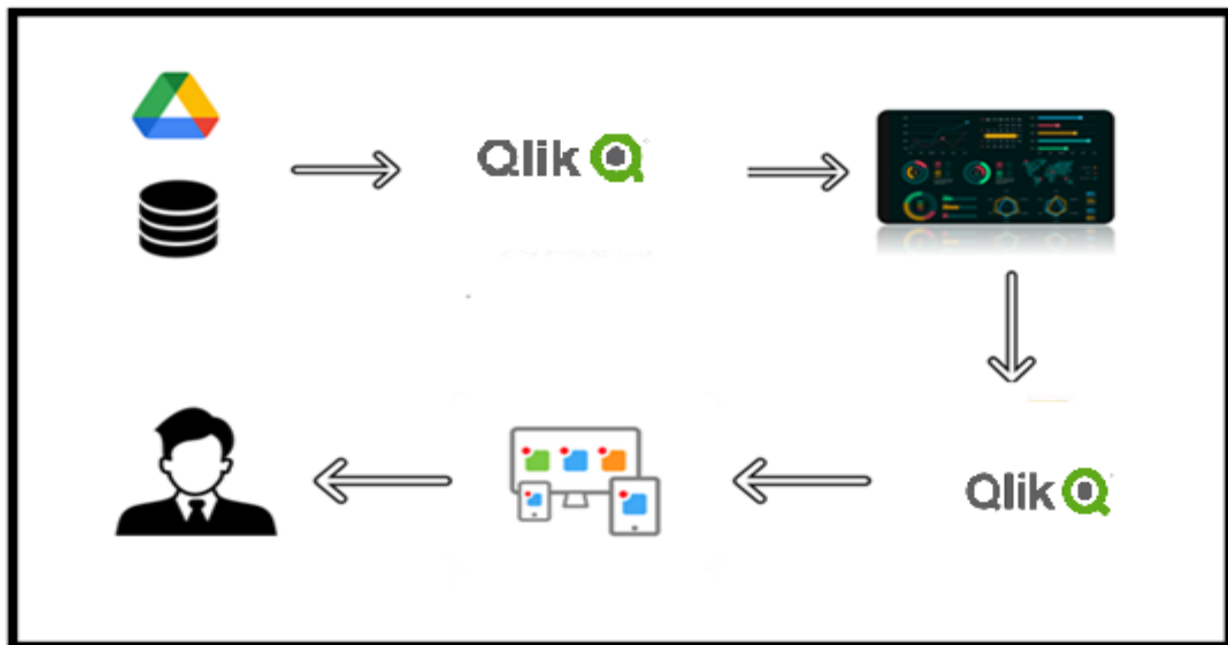
- ✓ **Revenue Optimization:** The project aims to help airlines maximize their revenue by analyzing historical ticket sales data. By identifying peak travel times and popular destinations, airlines can adjust their pricing strategies to capture the most value. Visualization of revenue trends over time and segmentation of customers based on their purchasing behavior allows for more precise pricing adjustments. This data-driven approach ensures that airlines can set optimal prices to boost profitability while meeting market demand efficiently.
- ✓ **Operational Efficiency:** Another primary objective is to enhance the operational efficiency of airports. This involves analyzing flight schedules, passenger flows, and luggage handling processes to identify any bottlenecks. By predicting peak traffic periods and understanding the flow of passengers and luggage, airports can allocate resources more effectively. The insights gained from this analysis can lead to streamlined operations, reduced delays, and improved overall efficiency. This not only helps in managing the current operations better but also in planning for future demands more accurately.
- ✓ **Customer Experience Enhancement:** The project focuses on improving the passenger experience by gaining a deeper understanding of customer preferences, satisfaction levels, and pain points. Through sentiment analysis of customer feedback data, airlines can identify specific areas where improvements are needed. Personalized services and targeted marketing campaigns can be developed based on these insights, leading to enhanced customer satisfaction and loyalty. By addressing the specific needs and

concerns of passengers, airlines can foster a more positive travel experience, which is crucial in a highly competitive industry.

- ✓ **Informed Decision-Making:** The ultimate goal of the project is to support informed decision-making across various facets of airline operations. By creating comprehensive and interactive visualizations in Qlik, stakeholders can explore data patterns, trends, and correlations that might not be evident otherwise. This empowers airline and airport management to make data-driven decisions that enhance strategic planning, operational management, and customer relations. The ability to visualize complex data in an accessible format ensures that decisions are based on accurate and comprehensive insights, leading to better outcomes for the business.

### 1.3 Technical Architecture

The technical architecture for this project includes the following components:



- ✓ **Data Source:** Synthetic airline data, which includes flight schedules, passenger demographics, ticket sales, and performance metrics.
- ✓ **Data Storage:** Data is stored in a format compatible with Qlik, ensuring easy access and processing.
- ✓ **Data Processing:** The data undergoes cleaning, transformation, and preparation to ensure it is accurate, complete, and ready for visualization.

- ✓ **Visualization Tool:** Qlik, used for creating interactive dashboards and visualizations to explore patterns, trends, and correlations in the data.
- ✓ **Analysis and Reporting:** Insights derived from Qlik visualizations aid in decision-making processes for revenue optimization, operational efficiency, and customer experience enhancement.

## 2. DEFINE PROBLEM/ PROBLEM UNDERSTANDING

### 2.1 Specify Business Problem

The primary business problems addressed by this project include:

- ✓ Identifying peak travel times and popular destinations to optimize ticket sales and pricing strategies.
- ✓ Enhancing operational efficiency at airports by streamlining flight schedules, passenger flows, and luggage handling processes.
- ✓ Improving customer satisfaction and loyalty by understanding and addressing customer preferences and pain points.

### 2.2 Business Requirements

The business requirements for this project are:

- ✓ **Data Integration:** Integrate synthetic airline data with Qlik for seamless analysis and visualization.
- ✓ **Revenue Analysis:** Analyze historical ticket sales data to identify revenue trends and optimize pricing strategies.
- ✓ **Operational Analysis:** Analyze flight schedules, passenger flows, and luggage handling processes to identify and address operational bottlenecks.
- ✓ **Customer Analysis:** Perform sentiment analysis on customer feedback data to understand and enhance the passenger experience.
- ✓ **Visualization:** Create interactive dashboards and visualizations to facilitate data exploration and decision-making.

### 2.3 Literature Survey

The literature survey involves reviewing existing research and publications on:

- ✓ **Data Analytics in Aviation:** Studies on how data analytics is used in the aviation

industry to optimize operations and enhance customer experience.

- ✓ **Business Intelligence Tools:** Research on the capabilities and applications of business intelligence tools like Qlik in various industries.
- ✓ **Revenue Management:** Literature on pricing strategies and revenue management techniques in the airline industry.
- ✓ **Operational Efficiency:** Studies on airport operations and efficiency improvement methods.
- ✓ **Customer Experience:** Research on factors influencing customer satisfaction and loyalty in the airline industry.

## 2.4 Social/Business Impact

### ▶ **Social Impact:**

By analyzing passenger demographics and preferences, airlines can tailor services to better meet customer needs, enhancing overall satisfaction and loyalty.

Improved operational efficiency can lead to reduced wait times and smoother airport experiences for passengers.

### ▶ **Business Impact:**

Airlines can optimize flight routes, schedules, and resource allocation based on data insights, leading to cost reduction and improved overall efficiency.

Enhanced customer experience can result in increased customer loyalty and positive word-of-mouth, driving more business.

Data-driven decision-making can lead to better strategic planning and competitive advantage in the aviation industry.

## 3. DATA COLLECTION

### 3.1 Collect the data set

Data collection is the process of gathering and measuring information on variables of interest, in an established systematic fashion that enables one to answer stated research questions, test hypotheses, evaluate outcomes and generate insights from the data.

📁 **DATASET Download link:** <https://www.kaggle.com/datasets/iamsouravbanerjee/airline-dataset/data>

## Understand the Data:

Data contains all the meta information regarding the columns described in the CSV files

Column Description of the Dataset:

- 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)

## 3.2 Connect the Data with Qlik Sense

Connecting the Data with Qlik Sense is the first step of the project.

- After downloading the data set from the Kaggle website we need to extract the data from the zip file.
- we have to login to the Qlik Sense Account
- Navigate to your Qlik Sense hub and create a new app or open an existing app where you want to add the stacked bar chart.
- Click on Load Data to load the Data in the Qlik Sense and use this data for the next process.

## 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, preparing the data for visualization software, and ensuring that the data is accurate and complete. This process helps to make data easily understandable and ready for creating visualizations to gain insights.

Relationships among different tables have been established in order to make it easy for visualizing.

Script is for adding new columns to the dataset:

```
[Airline_Dataset]:  
Load *;  
  
// Remove rows with '0' and '-' from ArrivalAirport column  
[Airline_Dataset]:  
NOCONCATENATE LOAD *,  
if(Age >= 0 AND Age <= 1, 'Baby',  
  if(Age >= 1 AND Age <= 3, 'Toddler',  
    if(Age >= 4 AND Age <= 9, 'Child',  
      if(Age >= 10 AND Age <= 12, 'Tween',  
        if(Age >= 13 AND Age <= 19, 'Teen',  
          if(Age >= 20 AND Age <= 24, 'Young Adult',  
            if(Age >= 25 AND Age <= 39, 'Adult',  
              if(Age >= 40 AND Age <= 54, 'Middle',  
                if(Age >= 55 AND Age <= 79, 'Elder',  
                  if(Age >= 80, 'Just plain old')))))))) AS AgeGroup,  
Date#([Departure Date], 'MM/DD/YYYY') as [Departure_Date],  
Year([Departure Date]) AS Year,  
Month([Departure Date]) as Month  
RESIDENT [Airline_Dataset]  
WHERE NOT ([Arrival Airport] = '0' OR [Arrival Airport] = '-');
```



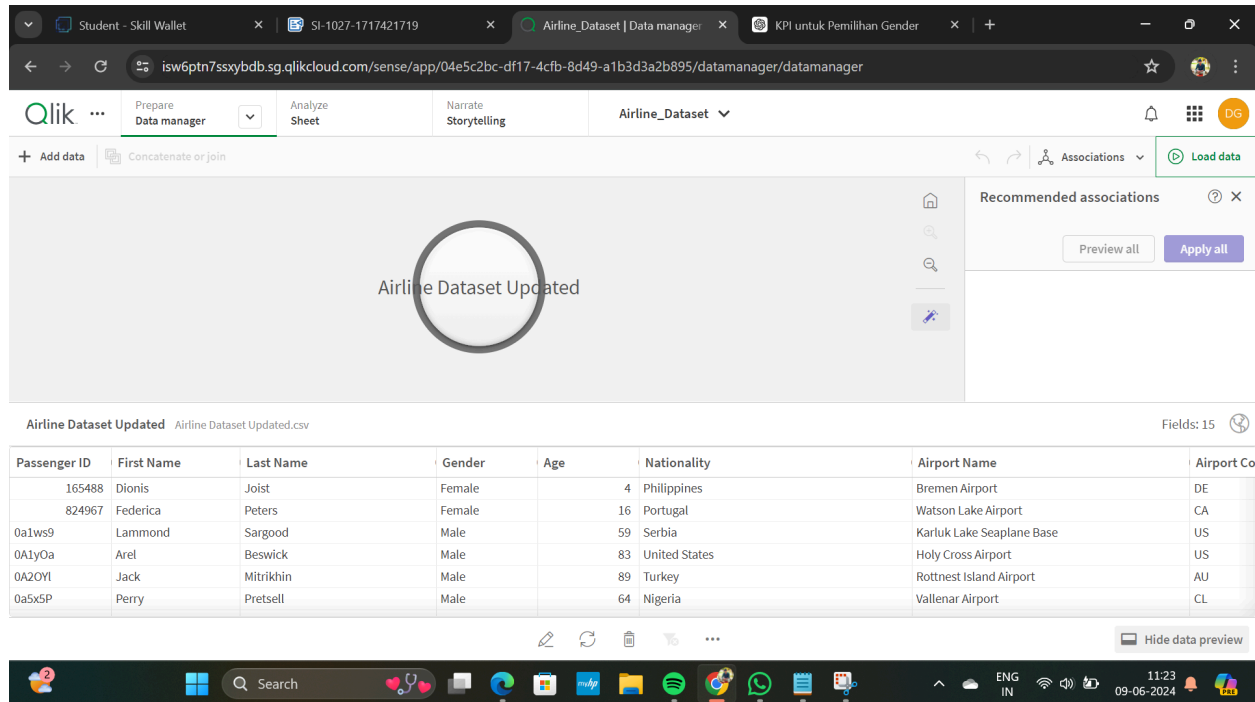


Fig: 4.1 Data Preparation

## 5. DATA VISUALIZATION

### 5.1 Visualization

Data visualization is the process of creating graphical representations of data to help people understand and explore the information. The goal of data visualization is to make complex data sets more accessible, intuitive, and easier to interpret. By using visual elements such as charts, graphs, and maps, data visualizations can help people quickly identify patterns, trends, and outliers in the data.

#### Some of the types of Visualizations

**Bar Charts:** Compare performance across different categories.

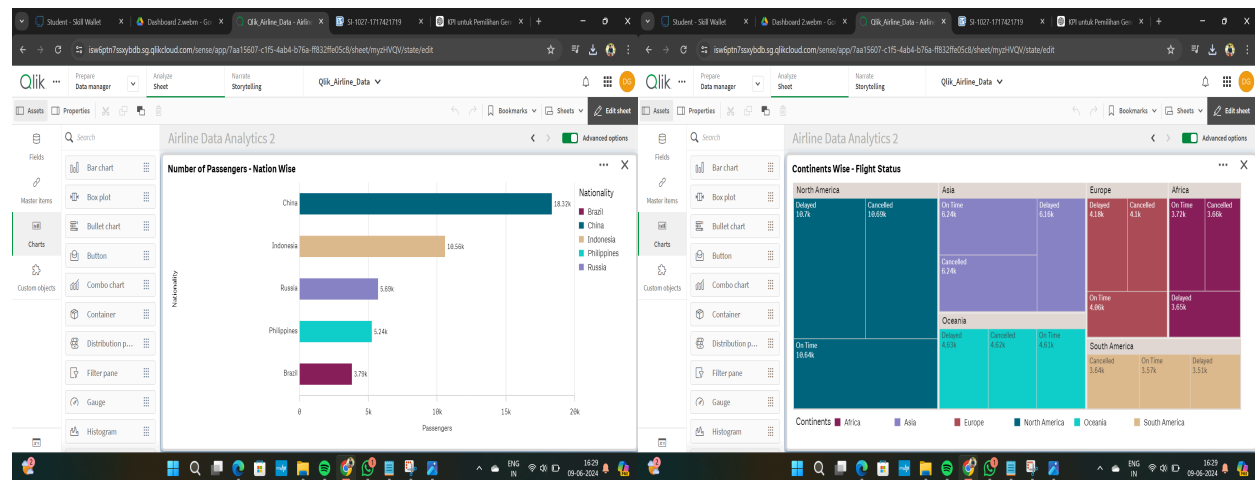
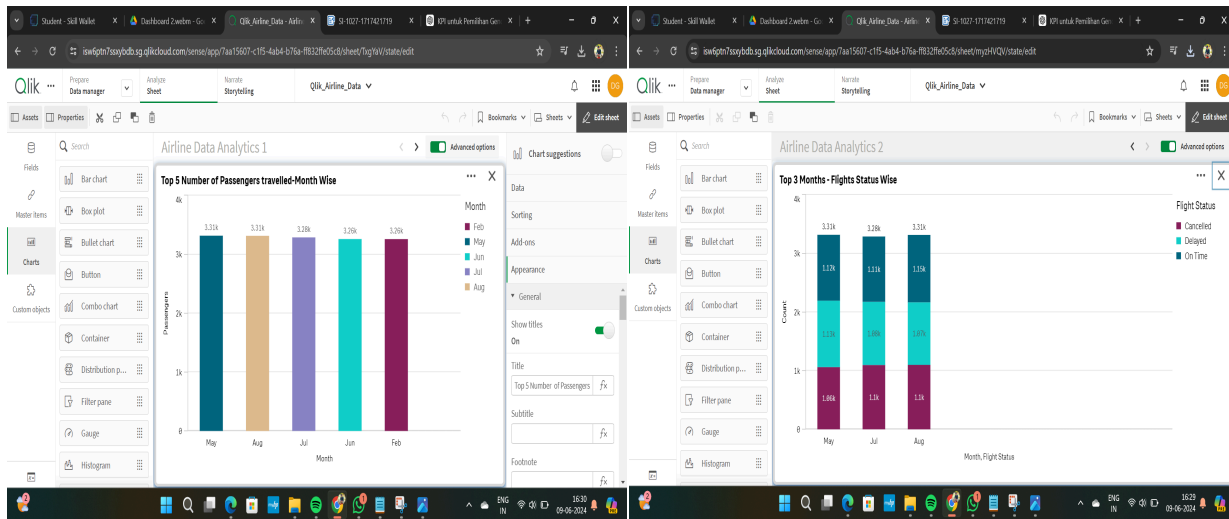
**Line Charts:** Track changes over time.

**Heat Maps:** Show data intensity and identify high-activity areas.

**Scatter Plots:** Visualize relationships between two variables.

**Pie Charts:** Display composition or breakdown of a whole.

**Maps:** Illustrate geographic data distributions



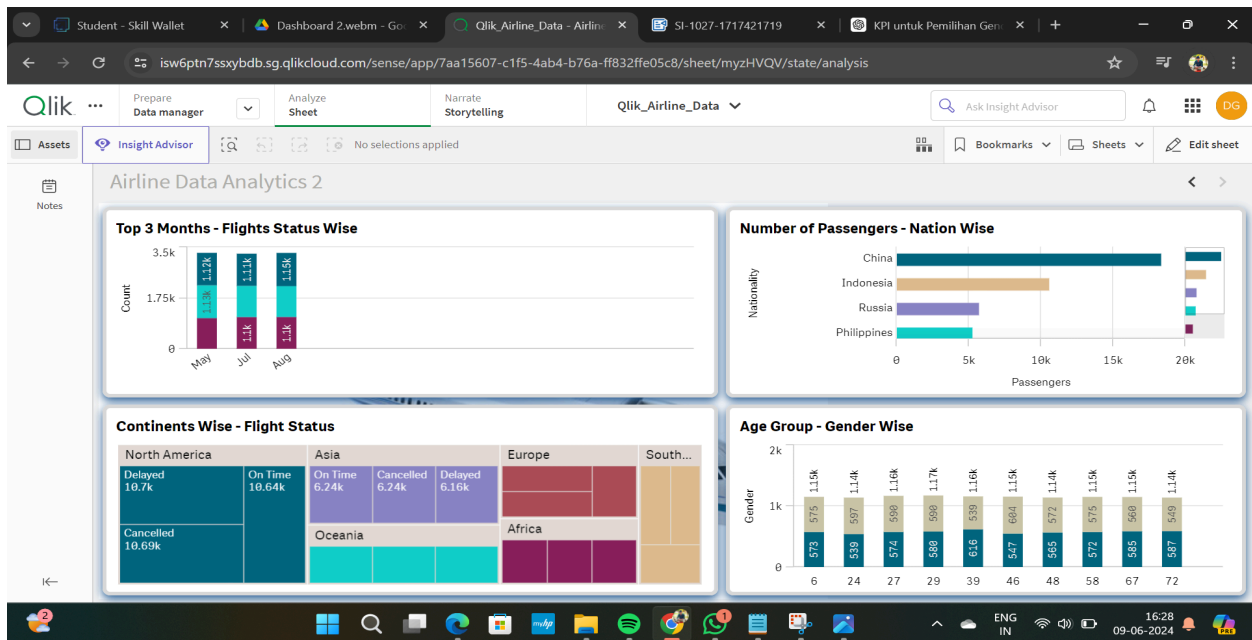
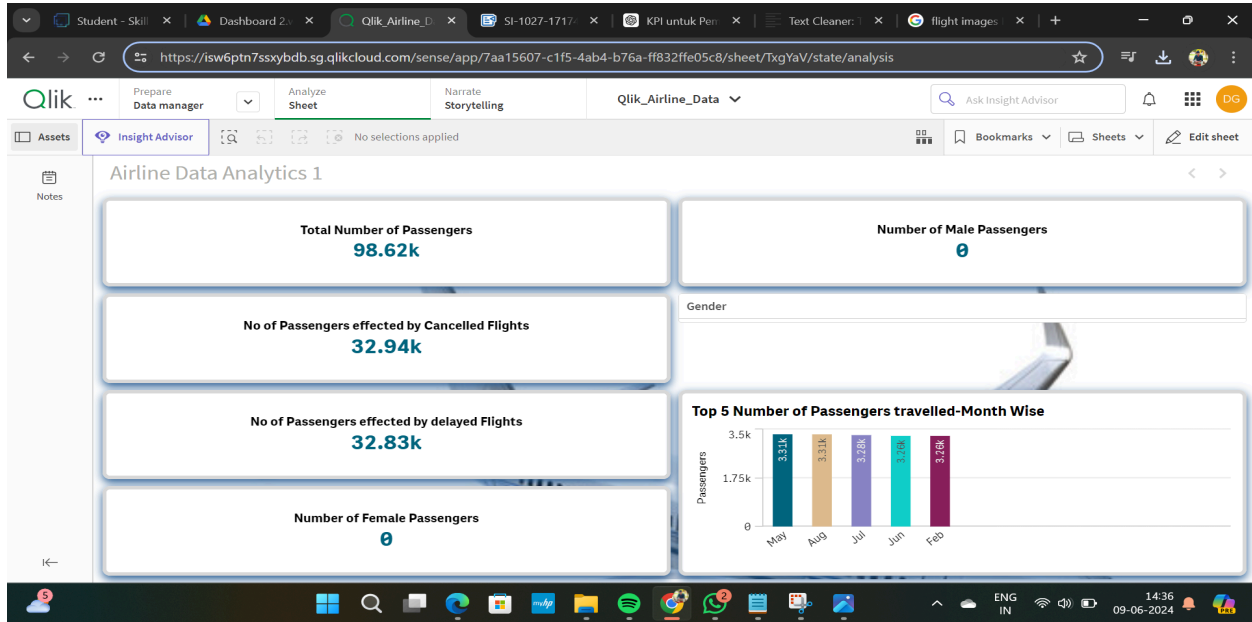
**Fig 5.1 Sample Visualizations**

## 6. DASHBOARD

### 6.1 Responsive and Design of the Dashboard

A dashboard is a graphical user interface (GUI) that displays information and data in an organized, easy-to-read format. Dashboards are often used to provide real-time monitoring and analysis of data and are typically designed for a specific purpose or use case. Dashboards can be

used in a variety of settings, such as business, finance, manufacturing, healthcare, and many other industries. They can be used to track key performance indicators (KPIs), monitor performance metrics, and display data in the form of charts, graphs, and tables.



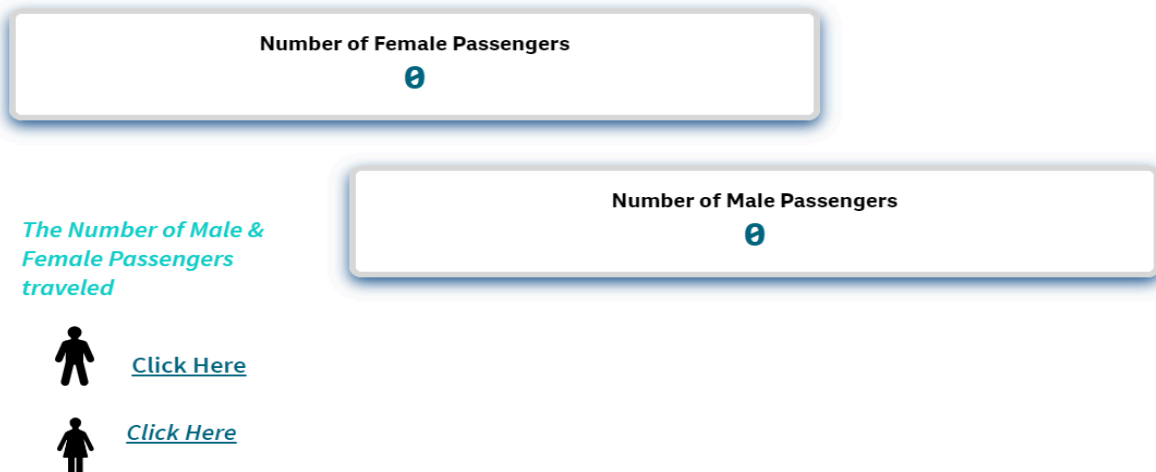
## 6.2 Design of story

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

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



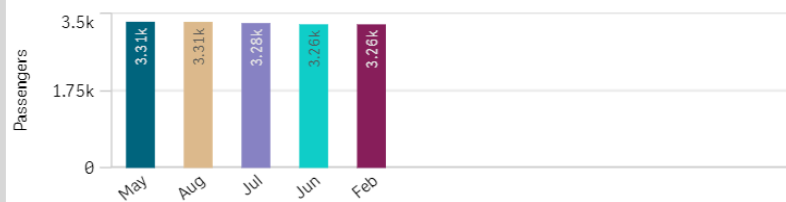
### *Exploring Insights From Synthetic Airline Data Analysis With Qlik*



## Exploring Insights From Synthetic Airline Data Analysis With Qlik

*This Chart represents the Top 5 Months in which the Passengers traveled the most across the Continents given in Dataset*

**Top 5 Number of Passengers travelled-Month Wise**



## Exploring Insights From Synthetic Airline Data Analysis With Qlik

**Top 3 Months - Flights Status Wise**



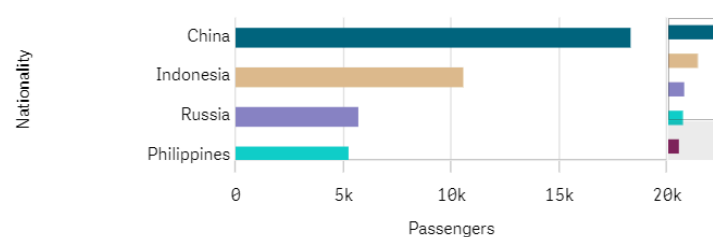
*This Chart represents the Top 3 Months Flight Status on [Cancelled/ Delayed/ On Time] given in the Dataset.*

*Which helps in the control and monitor of Operational Efficiency.*

*This Chart represents the Number of Passengers traveled across the Nationwide like [ China/ Indonesia/ Russia etc.,]*

*This Chart helps in monitor and analyze the Revenue Optimization and Economy of a Nation.*

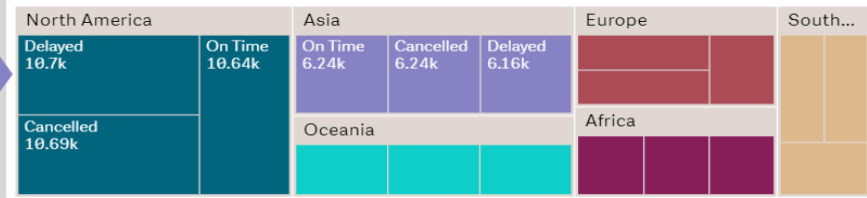
**Number of Passengers - Nation Wise**



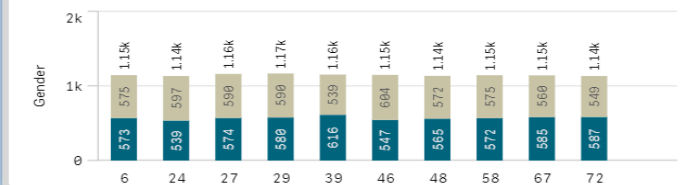
## Exploring Insights From Synthetic Airline Data Analysis With Qlik

This TreeMap represents the Continent wise Flight Status of the Flight which is [ Cancelled/ Delayed/ OnTime]  
Which helps in the control and monitor of Operational Efficiency.

Continents Wise - Flight Status



Age Group - Gender Wise



This Stacked BarChart represents the Gender wise count of Passengers traveling plotted against the Age of the Passenger given in the Dataset.

## 7. REPORT

### 7.1 Report Creation

Designing a report in Power BI involves several key steps. First, connect to data sources and create visualizations like charts and graphs. Customize their appearance and interactivity, then organize them logically on the canvas. Format elements for consistency and clarity, considering the audience's needs to ensure effective communication of insights. Optionally, create dashboards for a summarized view. Finally, iterate based on feedback to continually improve the report's design and usefulness.

## 8. PERFORMANCE TESTING

### 8.1 Amount of Data Rendered

"Amount of Data Loaded" refers to the quantity or volume of data that has been imported, retrieved, or loaded into a system, software application, database, or any other dataset or age or processing environment. It's a measure of how much data has been successfully processed and made available for analysis, manipulation, or use within the system.

Airline_Dataset_
AgeGroup
Departure_Date
Year
Month
Airline_Dataset_Passenger ID
Airline_Dataset_First Name
Airline_Dataset_Last Name
Airline_Dataset_Gender
Airline_Dataset_Age
Airline_Dataset_Nationality
Airline_Dataset_Airport Name
Airline_Dataset_Airport Country Code
Airline_Dataset_Country Name
Airline_Dataset_Airport Continent
Airline_Dataset_Continents
Airline_Dataset_Departure Date
Airline_Dataset_Arrival Airport
Airline_Dataset_Pilot Name
Airline_Dataset_Flight Status
Airline_Dataset_Airline_Dataset_Nationality_GeoInfo
Airline_Dataset_Airline_Dataset_Airport Country Code_GeoInfo
Airline_Dataset_Airline_Dataset_Country Name_GeoInfo

## 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.