

# **BUSINESS ANALYTICS**

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

### **1. INTRODUCTION**

#### **1.1 OVERVIEW: A BRIEF DESCRIPTION ABOUT YOUR PROJECT:**

The goal of the comprehensive project "Exploring Insights from Synthetic Airline Data Analysis with Qlik" is to use synthetic airline data analysis capabilities of Qlik, a top business intelligence and data visualization platform. This data, which includes a wide range of components like flight schedules, passenger demographics, ticket sales, and performance measures, is painstakingly created to mimic actual airline operations.

This project's main goal is to use Qlik's sophisticated analytical tools to find important correlations, new trends, and complex patterns in the synthetic data. By doing this, the research hopes to offer useful insights that help improve the way different stakeholders, such as airlines, airports, and industry partners, make decisions. Stakeholders will have a better grasp of customer behavior, market dynamics, and operational efficiencies thanks to this in-depth analysis. In the end, the project aims to support industry players' strategic planning initiatives, airline operational optimization, and passenger experience enhancement.

#### **1.2 PURPOSE : THE USE OF THIS PROJECT. WHAT CAN BE ACHIEVED USING THIS:**

##### **Scenario 1: Revenue Optimization**

An airline is examining past ticket sales data in-depth in an effort to improve its revenue optimization tactics. Qlik's sophisticated data visualization and analytical tools allow the airline to pinpoint popular travel dates, high-demand locations, and profitable pricing schemes. The airline may segment clients based on their purchase patterns, visualize revenue trends across a range of time periods, and modify pricing models to optimize profitability thanks to this thorough study. By using data-driven decision-making, the airline is able to use dynamic pricing strategies, improve market responsiveness, and stimulate revenue development.

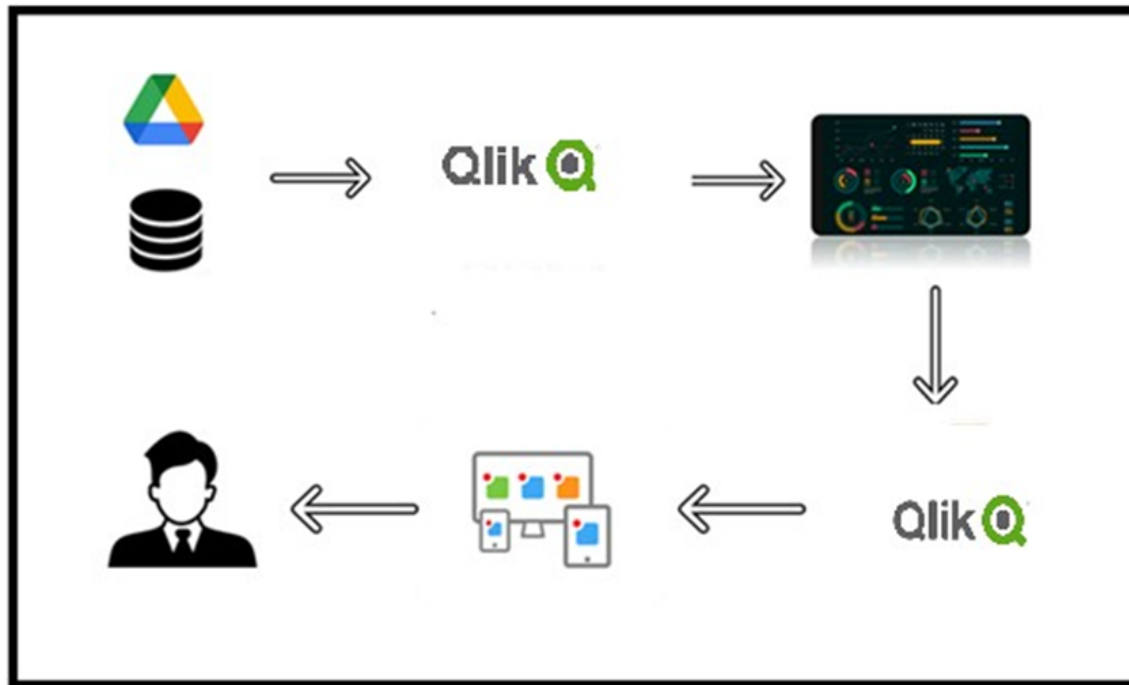
##### **Scenario 2: Operational Efficiency**

An airport authority looks at detailed information on flight schedules, passenger flows, and baggage handling procedures in order to increase its operational efficiency. The authority may identify operational bottlenecks, forecast high traffic periods, and optimize resource allocation by integrating Qlik with synthetic airline data. The airport can optimize its operations, including gate assignments and baggage handling, by utilizing Qlik's robust data integration and visualization technologies. This leads to improved passenger experiences and more effective utilization of airport resources. By taking a proactive stance in handling airport operations, turnaround times are shortened, delays are decreased, and overall efficiency is increased.

### Scenario 3: Customer Experience Enhancement

Airlines are putting more and more effort into studying their customers' preferences, levels of happiness, and pain spots in order to improve the passenger experience. Airlines can find important areas for improvement and consumer sentiment trends by using Qlik to do sentiment analysis on customer feedback data. Airlines are better able to customize services, target marketing efforts, and cater to individual customer needs thanks to this data-driven strategy. Airlines can cultivate increased customer loyalty and satisfaction by improving the quality of their services and taking proactive measures to resolve consumer issues. This will ultimately lead to a more competitive position in the market. Airlines are able to guarantee a more pleasurable and customized journey experience for their patrons by utilizing these insights.

#### 1.3 TECHNICAL ARCHITECTURE:



## 2. DEFINE PROBLEM / PROBLEM UNDERSTANDING:

### 2.1 SPECIFY THE BUSINESS PROBLEM:

Enhancing customer experience, optimizing income, and improving operational efficiency are major problems for the airline and airport authorities. Due to inadequate insights into past ticket sales, travel trends, and customer purchasing behaviors, the airline is unable to maximize revenue, which results in less-than-ideal pricing tactics. At the same time, the airport authority struggles with ineffective management of aircraft schedules, passenger flows, and luggage handling, which leads to inefficient resource allocation and operational bottlenecks.

Moreover, airlines are unable to fully comprehend the preferences and satisfaction levels of their customers, which hinders their capacity to efficiently handle customer complaints and customize services. These stakeholders hope to accomplish sentiment analysis on consumer comments, streamline airport operations, and evaluate historical ticket sales data by utilizing Qlik's sophisticated data visualization and analytical capabilities. By using an integrated strategy, pricing strategies can be optimized, operational efficiency can be increased, and customer satisfaction and loyalty will all rise.

## **2.2 BUSINESS REQUIREMENTS:**

### **Data Integration and Management**

**Centralized Data Repository:** Integrate and consolidate data from various sources, ensuring accuracy and consistency.

### **Analytical Capabilities**

**Revenue Analysis:** Tools to visualize revenue trends, analyze peak travel periods, and segment customers based on purchasing behavior.

**Operational Efficiency Analysis:** Analyze flight schedules, passenger flows, and luggage handling to identify and address bottlenecks.

**Customer Experience Enhancement:** Perform sentiment analysis on customer feedback to identify trends and areas for improvement.

### **Visualization and Reporting**

**Interactive Dashboards:** Develop user-friendly dashboards to display key performance indicators (KPIs) and insights.

**Customizable Reports:** Generate reports tailored to different stakeholder needs.

### **Predictive Analytics**

**Demand Forecasting:** Implement predictive models to forecast passenger demand and adjust pricing strategies.

**Operational Predictions:** Predict peak traffic periods to optimize resource allocation.

**Comprehensive Training:** Provide training programs for staff on using Qlik for data analysis and reporting.

By focusing on these core requirements, the airline can effectively leverage to optimize revenue, improve operational efficiency, improve customer experience.

### **Literature Survey :**

#### **1) Data Integration and Management**

Title: "Data Integration in Business Intelligence: Concepts and Practices" Authors: Rick Sherman

Summary: This work explores best practices and methodologies for integrating disparate data sources into a cohesive BI platform. It emphasizes the importance of data quality, consistency, and the technical challenges involved in consolidating various data types.

Relevance: Provides foundational knowledge on integrating and managing diverse data sources, crucial for the centralized data repository requirement.

#### Analytical Capabilities

2) Title: "Business Analytics: Data Analysis & Decision-Making" Authors: S. Christian Albright, Wayne L. Winston

Summary: This book covers comprehensive analytical techniques used in business settings, including revenue analysis, customer segmentation, and operational efficiency analysis.

Relevance: Offers methodologies for analyzing revenue trends, customer behaviors, and operational processes, aligning with the analytical capabilities needed for revenue optimization and operational efficiency.

3) Title: "Sentiment Analysis and Opinion Mining" Authors: Bing Liu

Summary: This book provides an in-depth look into sentiment analysis techniques and their applications in understanding customer feedback and improving service quality.

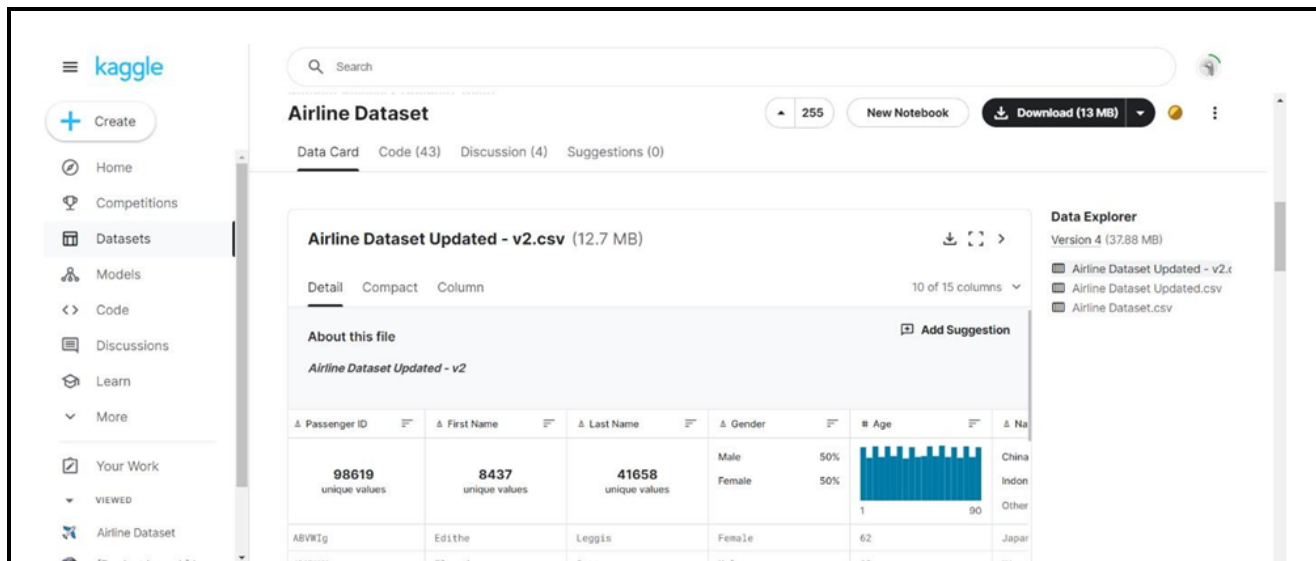
Relevance: Directly supports the requirement for enhancing customer experience through sentiment analysis of feedback data.

### **3. DATA COLLECTION:**

#### **3.1 COLLECT THE DATASET:**

The process of collecting data is comprehensive and entails the methodical gathering and measurement of information on a range of relevant characteristics. This method is intended to produce insights, test hypotheses, assess results, and respond to research questions. The first step is to pinpoint the important variables and create a strategy for reliably and consistently collecting this data. Data is gathered using a variety of techniques, depending on the requirements of the research, including surveys, interviews, observations, and experiments. To guarantee compliance with prescribed guidelines, the gathered information undergoes periodic assessments for precision and dependability. The ultimate goal of this procedure is to provide a solid dataset that will support research goals and offer insightful information for analysis and decision-making. Here, we obtained the data in tabular format from the Kaggle website.

The link is [here](#).



### **3.2 CONNECT DATA WITH QLIK SENSE:**

**STEP 1:** Access your Qlik cloud account by logging in.

**STEP 2:** Choose New Analytics App from the list of apps to create by clicking the add new button located on the right side of the Qlik Cloud platform.

**STEP 3:** After loading the data, upload the dataset from your local computer by selecting the Upload Data option.

Await the loading of the data.

### **4. DATA PREPARATION:**

In order to ensure that the data is ready for analysis, there are numerous crucial procedures involved in data preparation for visualization. To ensure accuracy, the data is first cleansed to eliminate any unnecessary or missing information. Subsequently, the data is formatted appropriately for use with visualization tools, which facilitates the creation of insightful visual representations.

After then, the data is examined to find trends and patterns that shed light on its structure. By focusing the analysis on particular subsets, filtering the data might improve knowledge in particular domains.

The data is then ready for visualization tools, making sure that it is accurate, complete, and formatted correctly. This guarantees that the data may be easily understood and prepared for the production of insightful infographics.

Now that the data cleaning procedure is over, we can go to the visualization phase with ease, using the prepared data to produce lucid, perceptive visuals that aid in decision-making and enhance comprehension of the implications of the data.

Data preprocessing and cleansing are the steps.

1)Through data cleaning, we identify the information that is necessary solely by eliminating unnecessary, incomplete, and missing data.

2) During the data preprocessing step, we add any extra information that is needed or eliminate any useless data. Preprocessing the data at this point would make it ready for improved visuals.

Data Preprocessing Code:

```
[Airline Dataset Updated - v2]:
Load;
[Airline Dataset Updated - v2]:
NOCONCATENATE LOAD
if(Age 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 >=55 AND Age <=79, 'Elder',
if(Age > 48 AND Age < 54, 'Middle',
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 Updated - v2]
WHERE NOT ([Arrival Airport] ='0' OR [Arrival Airport]= '-');
```

## 5. DATA VISUALISATIONS:

The technique of creating graphical representations of data to aid in understanding and effective information exploration is known as data visualization. The main goal is to convert complicated data sets into formats that are easier to understand, navigate, and comprehend. Data visualizations let users rapidly spot patterns, trends, and outliers in the data by including visual components like charts, graphs, and maps. This method provides facts in an understandable and visually appealing way to help in decision-making while also improving the capacity to comprehend complex data relationships.

STEP-1 : Go to Sheet in the analyze tab on the top menu and click edit sheet and choose advanced options in the right corner at top.

STEP-2: Drag the required chart on to the sheet area and add required measures to create a visualization

The visualizations created are :

1) Total no of passengers : Take a Kpi chart and enter passengerid count as the formula and label as total no of passengers



2)Gender : Add a filterpane and select the measure as gender and add it to the sheet

| Gender |
|--------|
| Female |
| Male   |

3) No of passengers effected by the cancelled flights : Drag a Kpi chart on to the sheet and add the required measure as passenger id and select count and in the fx tab write the query as =COUNT({< [FlightStatus] = {'Cancelled'} >} [Passenger ID]) and label as ='Number of ' & GetFieldSelections([Gender]) & ' Passengers effected by Canceled flights'

Male:

Number of Male Passengers effected by Cancelled flights  
16.35k

Female:

Number of Female Passengers effected by Cancelled flights  
16.31k

4) No of passengers effected by delayed flights : Drag a Kpi chart on to the sheet and add the required measure as passenger id and select count and in the fx tab write the query as =COUNT({< [FlightStatus] = {'Delayed'} >} [Passenger ID]) and label as ='Number of ' & GetFieldSelections([Gender]) & ' Passengers effected by Cancelled flights'

Number of Male Passengers effected by Delayed flights  
16.39k

Number of Female Passengers effected by Delayed flights  
16.13k

5) No of Flights on time : Drag a Kpi chart on to the sheet and add the required measure as passenger id and select count and in the fx tab write the query as =COUNT({< [Flight Status] = {'On Time ' } >} [Passenger ID]) and label as ='Number of flights on time '

No of flights-On time

32.56k

6) No of female passengers : Drag a Kpi chart on to the sheet and add the required measures as passenger id and select count and in the fx tab write the query as =If(Only(Gender) = 'Female',COUNT({<[Gender]='Female'}>} DISTINCT [Passenger ID]),0) and label as 'Number of femalepassengers '

No of female Passengers

48.57k

7) No of Male passengers : Drag a Kpi chart on to the sheet and add the required measures as passenger id and select count and in the fx tab write the query as =If(Only(Gender) = 'Female',COUNT({<[Gender]='Male'}>} DISTINCT[Passenger ID]),0) and label as 'Number of Male passengers '

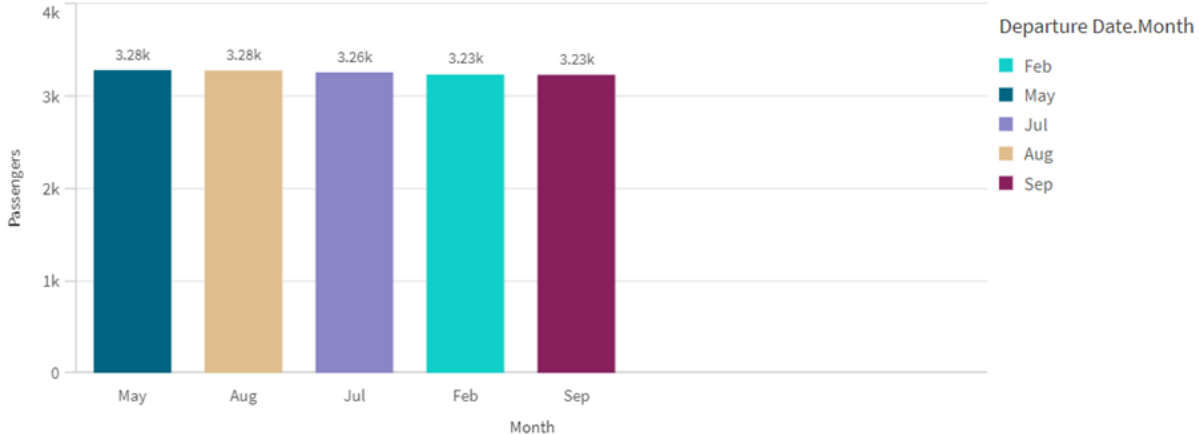
No of Male Passengers

49.17k

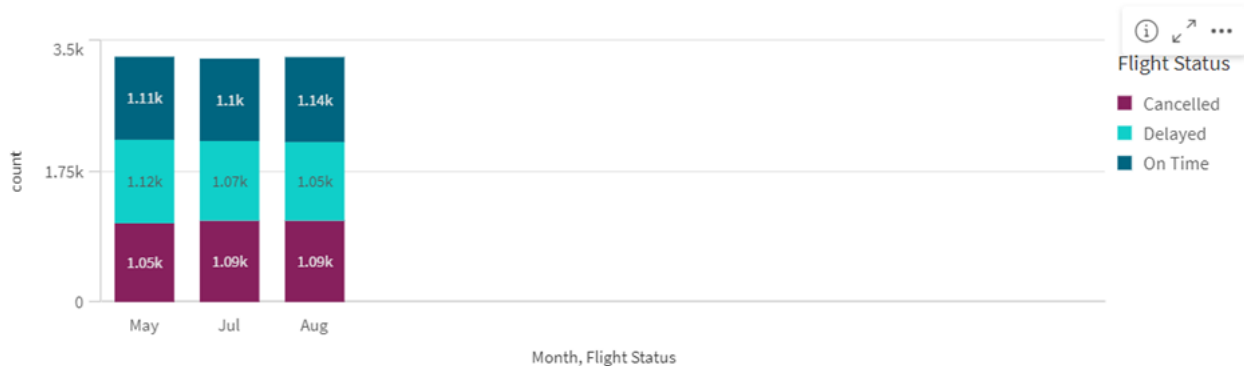
8) Top 5 months in which passengers traveled the most : Drag a bar chart on to the workstation and add passengerid count as bars and Departure date.Month as dimension and in limitation section choose fixed number and enter 5 and In appearance choose color by dimension and turn on the labels .



Top 5 Months - No of Passangers Travelled

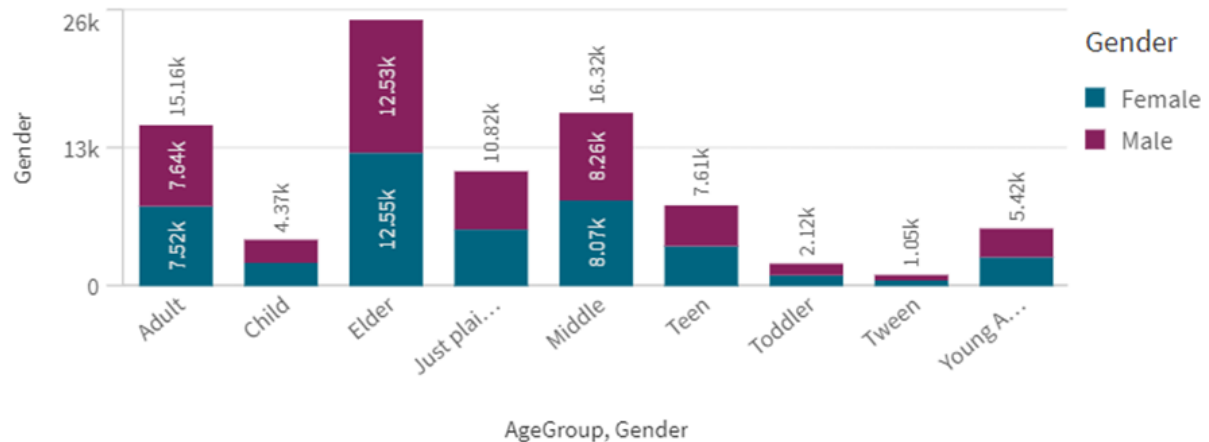


9) Top 3 months of people effected by flight status against the passengers count : Drag a bar chart on to the workstation and add passenger id count as bars and Departure date.Month as dimensionand in limitation section choose fixed number and enter 3. Now drag Flight status from fields on to the chart an click add flight status this will generate a grouped chart .Now go to appearanceand choose a stacked chart and turn on labels and segment labels on

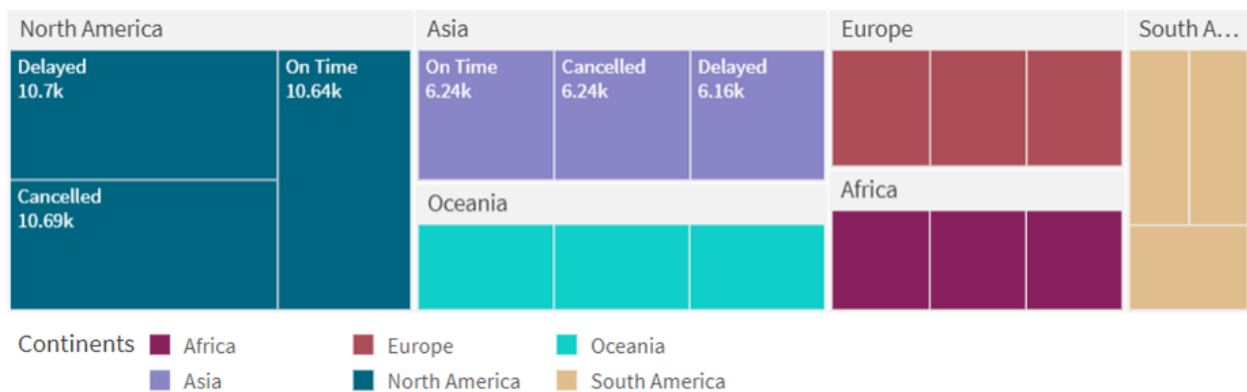


10) Age Group - Gender Wise : Drag a bar chart on to the workstation and add gender as bars and Agegroup as dimension. Now drag Gender from fields on to the chart and click add gender. This will generate a grouped chart.Now go to appearance and choose a stacked chart and turn on labels and segment labelson.

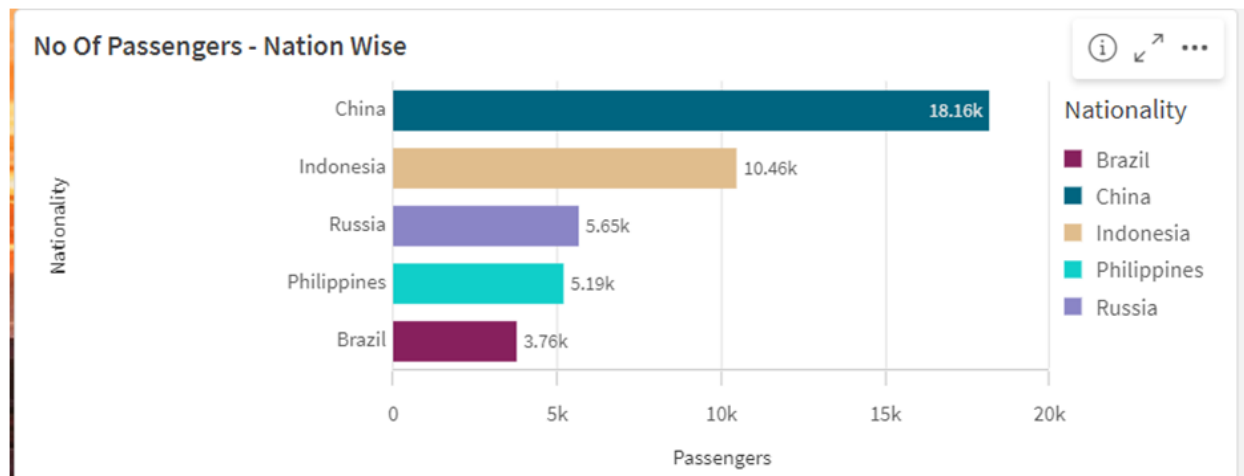
Age Group - Gender Wise



11) Continents wise - Flight Status : Drag a Tree map charton to the workstation and addContinents as dimension. Now drag flight status from fields on to the chart and click add flight status. This will generate the chart .Now go to appearance and choose color as by dimensionand turn on labels and segment labels on.



12) No of passengers -Nation wise : Drag abar chart on to the workstation and addpassengerid count as bars and Nationality as dimension and in the limitation section choose fixed number and enter 5 and In appearance choose color by dimension and turn on the labels andalso choose orientation to be horizontal.

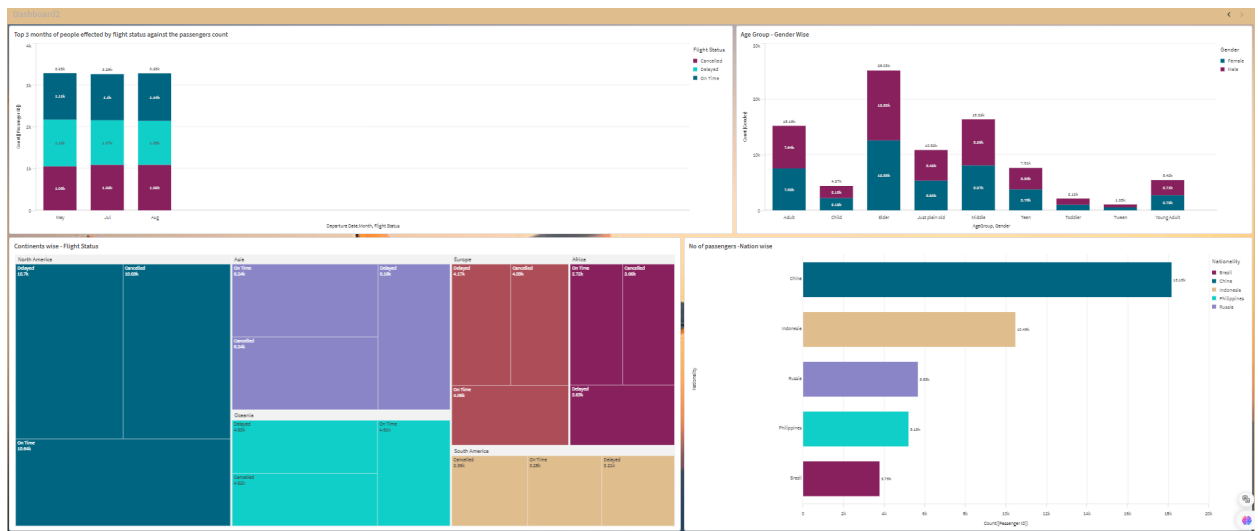
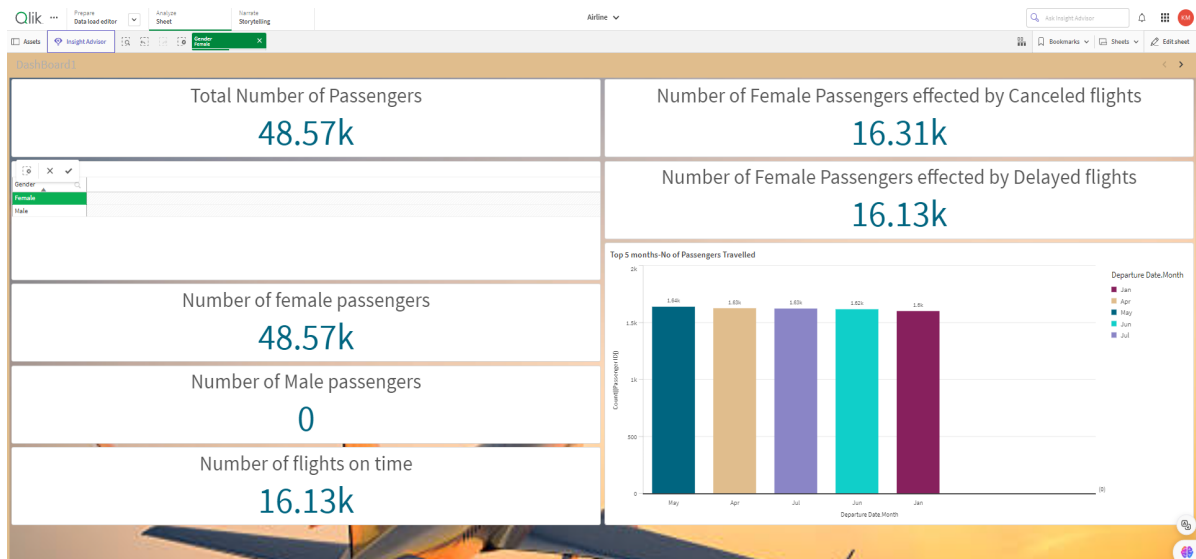


## **6. DASHBOARDS:**

A graphical user interface (GUI) called a dashboard is made to provide facts and information in an organized, readable manner. Dashboards are customized for certain goals or use cases and are frequently used for real-time monitoring and data analysis. They are extensively used in many different contexts, such as business, finance, manufacturing, healthcare, and a host of other industries. Dashboards facilitate fast and well-informed decision-making by allowing users to manage key performance indicators (KPIs), monitor performance measures, and visually show data through charts, graphs, and tables.

**STEP-1 :** Now we need to assemble all the visualizations we made into a place add a background image to these dashboards :

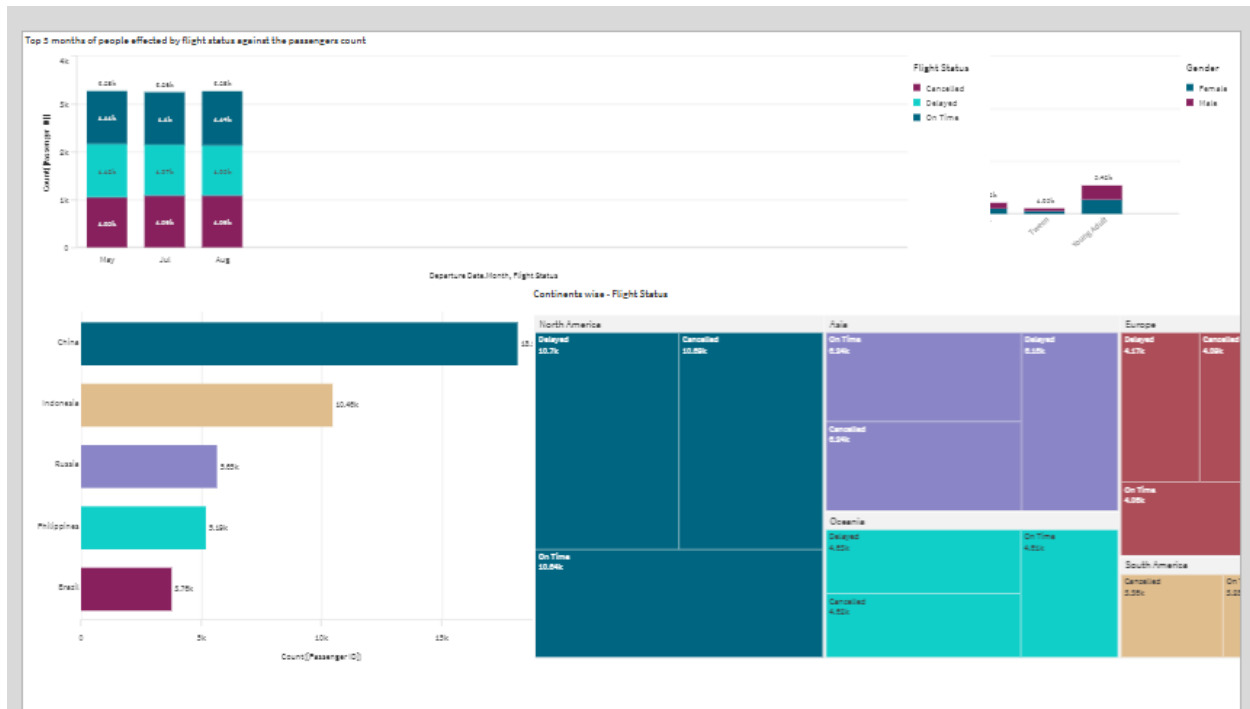
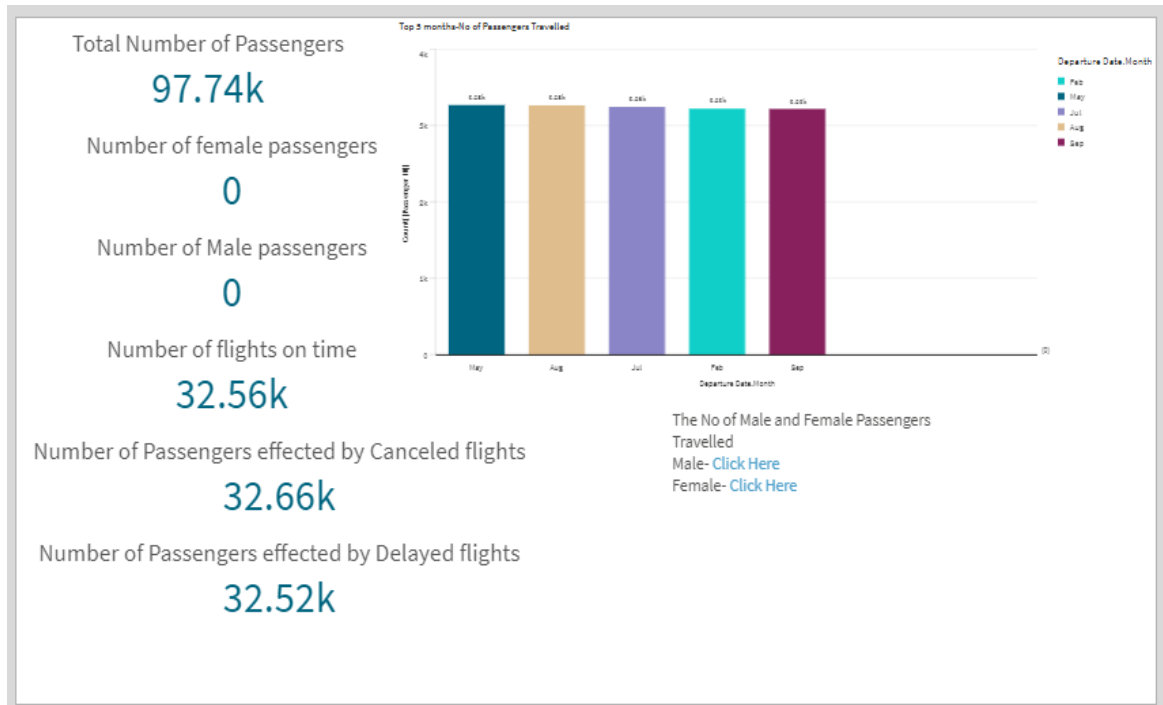




In this way, we can make dashboards

## 7. REPORT

For report we use storytelling and take a snapshot of every visualization we made and make those on story telling slides. We arrange them, describe them using shapes and form a slide



## 8. PERFORMANCE TESTING :

### 8.1 AMOUNT OF DATA RENDERED :

The "Amount of Data Loaded" refers to the volume or quantity of data that has been imported, retrieved, or loaded into a system, software application, database, or any other

datastorage or processing environment. This measure indicates the extent of data that has been successfully processed and is now accessible for analysis, manipulation, or use within the system. It encompasses all types of data that have been transferred from various sources into the target environment, highlighting the system's capacity to handle and make data available for subsequent operations and decision-making processes.

## 8.2 UTILIZATION OF DATA FILTERS:

"Utilization of Filters" refersto the strategic application or use of filtering mechanisms withina system, software application, or data processing pipeline to extract selectively, manipulate, or analyze data based on specified criteria or conditions. This process involves setting predefined parameters that determine which data points arerelevant and should be includedin the analysis, while excluding those that do not meet the established criteria.

Filters play a critical role in managing large datasets by narrowing down the scope of data, thereby enhancing the efficiency and effectiveness of data processing and analysis. By focusing only on the relevant information that meets the predefined criteria, filters help in reducing data noise, improving the clarity of insights, and enabling more precise decision- making. The application of filters can be dynamic, allowing users to adjust criteria based onevolving needs and ensuring that the data remains relevant and actionable throughout the analysis process. This targeted approach to data management is essential in various fields, including business intelligence, data science, and analytics, where the ability to quickly and accurately isolatepertinent information can significantly impactoutcomes and performance.

