Business Analytics

Exploring Insights From Synthetic Airline Data Analysis With Qlik

1.INTRODUCTION

1.1 Overview : A brief description about your project :

The project "Exploring Insights from Synthetic Airline Data Analysis with Qlik" is an extensive initiative aimed at utilizing Qlik, a top-tier business intelligence and data visualization platform, to analyze synthetic airline data. This data is carefully designed to mimic real-world airline operations, including various elements like flight schedules, passenger demographics, ticket sales, and performance metrics.

The main goal of this project is to employ Qlik's advanced analytical features to uncover complex patterns, emerging trends, and significant correlations within the synthetic data. The analysis is intended to provide actionable insights to enhance decision-making processes for airlines, airports, and industry partners.

By conducting this sophisticated analysis, stakeholders will gain a deeper understanding of operational efficiencies, market dynamics, and customer behavior. The project aims to optimize airline operations, improve the passenger experience, and support strategic planning efforts for industry participants.

1.2 Purpose : The use of this project. What can be achieved using this :

Scenario 1: Revenue Optimization

An airline aims to enhance its revenue optimization strategies through a detailed analysis of historical ticket sales data. Utilizing Qlik's advanced data visualization and analytical capabilities, the airline can identify peak travel periods, popular destinations, and effective pricing strategies. This comprehensive analysis allows the airline to visualize revenue trends over different time frames, segment customers based on purchasing behaviors, and adjust pricing models to maximize profitability. These insights enable the airline to implement dynamic pricing strategies, improve market responsiveness, and drive revenue growth through data-informed decision-making.

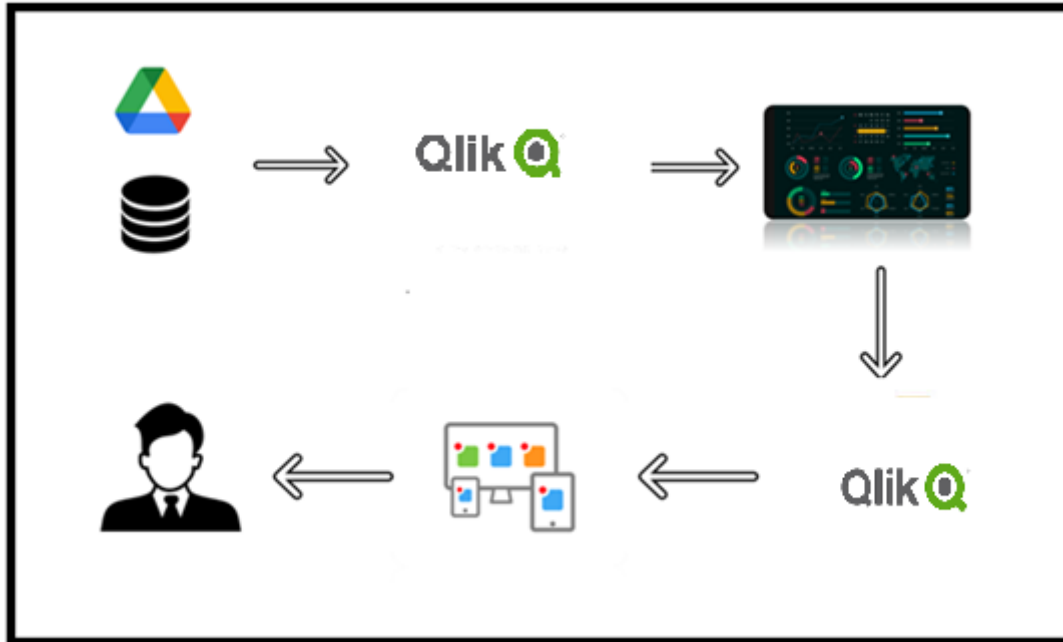Scenario 2: Operational Efficiency

An airport authority is committed to improving operational efficiency by analyzing extensive data on flight schedules, passenger flows, and luggage handling processes. By integrating Qlik with synthetic airline data, the authority can identify operational bottlenecks, predict peak traffic times, and optimize resource allocation. Qlik's robust data integration and visualization tools enable the airport to streamline operations, from gate assignments to baggage handling, ensuring smoother passenger experiences and more efficient use of airport facilities. This proactive approach to managing airport operations leads to improved turnaround times, reduced delays, and enhanced overall efficiency.

Scenario 3: Customer Experience Enhancement

Airlines are increasingly focused on improving the passenger experience by gaining deeper insights into customer preferences, satisfaction levels, and pain points. Using Qlik to perform sentiment analysis on customer feedback data, airlines can identify key areas for improvement and detect trends in customer sentiment. This data-driven approach allows airlines to personalize services, tailor marketing campaigns, and address specific customer needs more effectively. By enhancing service quality and addressing customer concerns

proactively, airlines can foster greater customer loyalty and satisfaction. Ultimately, this leads to a more competitive market position, ensuring a more enjoyable and personalized travel experience for passengers.

1.3 Technical Architecture :

2.Define Problem / Problem Understanding :

2.1 Specify the business Problem :

The airline and airport authority face considerable challenges in revenue optimization, operational efficiency, and enhancing the customer experience. The airline struggles to maximize revenue due to inadequate insights into historical ticket sales, travel patterns, and customer purchasing behaviors, resulting in suboptimal pricing strategies. Simultaneously, the airport authority deals with inefficiencies in managing flight schedules, passenger flows, and luggage handling, leading to operational bottlenecks and poor resource allocation. Moreover, airlines lack a thorough understanding of customer preferences and satisfaction levels, which hinders their ability to personalize services and address customer concerns effectively. By leveraging Qlik's advanced data visualization and analytical capabilities, these stakeholders aim to analyze historical ticket sales data, streamline airport operations, and perform sentiment analysis on customer feedback. This integrated approach will help optimize pricing strategies, enhance operational efficiency, and significantly improve overall customer satisfaction and loyalty.

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.

<u>**Visualization and Reporting**</u>

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

Customizable Reports: Generate reports tailored to different stakeholder needs.

<u>**Predictive Analytics**</u>

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.

2.3 Literature Survey :

1. Data Integration and Management
Title: "Data Warehousing for Business Intelligence"
Authors: Paulraj Ponniah
Summary: This book focuses on integrating disparate data sources to form a unified data warehouse, ensuring data quality and consistency. It provides foundational knowledge essential for establishing a centralized data repository in Qlik, aligning with the project's requirement for data integration and management.

2. Analytical Capabilities
Title: "Data Science for Business"
Authors: Foster Provost, Tom Fawcett
Summary: This book provides an overview of data mining techniques for business analytics, including revenue analysis and customer segmentation. It aligns with the need for advanced analytical capabilities in Qlik, supporting the project's objectives to optimize revenue and improve operational efficiency.

3. Visualization and Reporting
Title: "Storytelling with Data"
Authors: Cole Nussbaumer Knaflic
Summary: This book emphasizes storytelling in data visualization to create compelling visuals. It is crucial for developing effective dashboards and reports in Qlik, ensuring clear communication of data insights to stakeholders.

4. Predictive Analytics
Title: "Data Science and Predictive Analytics"
Authors: Ivo D. Dinov
Summary: This book covers the fundamentals of predictive analytics and data science, with practical applications. It supports predictive analytics for demand forecasting and operational predictions in Qlik, contributing to the project's goal of enhancing revenue optimization and operational efficiency.

3. Data Collection :
3.1 Collect the dataset :

Data collection involves the systematic gathering and measurement of information across various variables of interest. It is geared towards addressing research inquiries, testing hypotheses, evaluating outcomes, and deriving insights. The process commences with the identification of key variables and the formulation of a plan to collect relevant data accurately and consistently. Methods such as surveys, interviews, observations, and experiments are utilized based on research requirements. Maintaining adherence to established standards, the collected data undergoes regular scrutiny to ensure accuracy and reliability. Ultimately, the objective is to compile a robust dataset that underpins research objectives and facilitates informed decision-making and subsequent analysis. In this instance, the data was sourced in tabular format from the Kaggle website.
the link is here.
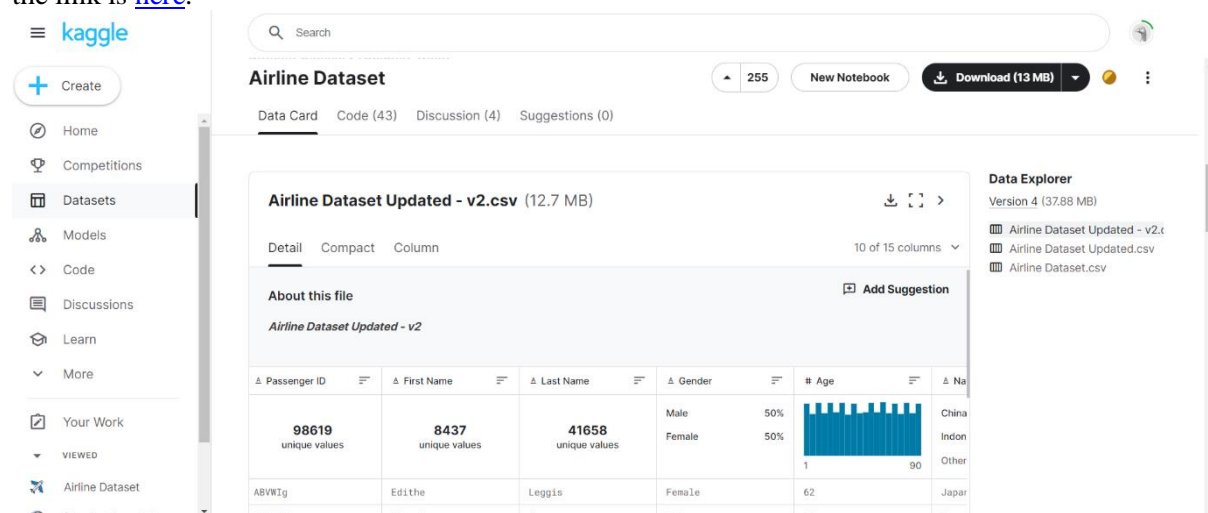
3.2 Connect data with Qlik Sense :
Step-1 : Login into your Qlik cloud account .
Step-2 : Create a new app using add new button on the right side of the qlik cloud platform and select new analytics app .
Step-3 : Now load the data, clicking on upload data option to upload the dataset from your local computer
Wait for the data to be loaded.

4.Data Preparation:

Data preparation for visualization involves several crucial steps to ensure that the data is well-suited for analysis. Initially, the data undergoes a cleaning process to eliminate irrelevant or missing information, thereby ensuring its accuracy. Following this, the data is transformed into a format suitable for visualization tools, facilitating the creation of meaningful visual representations.
Next, the data is explored to identify patterns and trends, offering valuable insights into its structure. Filtering the data enables a more focused analysis on specific subsets, enhancing understanding in particular areas.
Subsequently, the data is prepared for visualization software, ensuring appropriate formatting, completeness, and accuracy. This ensures that the data is easily interpretable and ready for creating visualizations that effectively communicate insights.

With the data cleaning process completed, we can seamlessly transition to the visualization stage, leveraging the prepared data to create clear, insightful visualizations that support decision-making and deepen understanding of the data's implications.

Data cleaning and preprocessing involve the removal of incomplete, missing, and irrelevant data to establish the necessary dataset. Additionally, data preprocessing may include adding required additional data or removing data that is not useful, ultimately preparing the data for enhanced visualizations.

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 Visualizations :

Data visualization involves the creation of graphical representations of data to facilitate effective comprehension and exploration. Its primary aim is to simplify complex data sets into more accessible and intuitive formats. By employing visual elements like charts, graphs, and maps, data visualizations enable users to identify patterns swiftly, trends, and outliers within the data. This not only enhances understanding of intricate data relationships but also supports informed decision-making by presenting data in a visually engaging and clear manner.
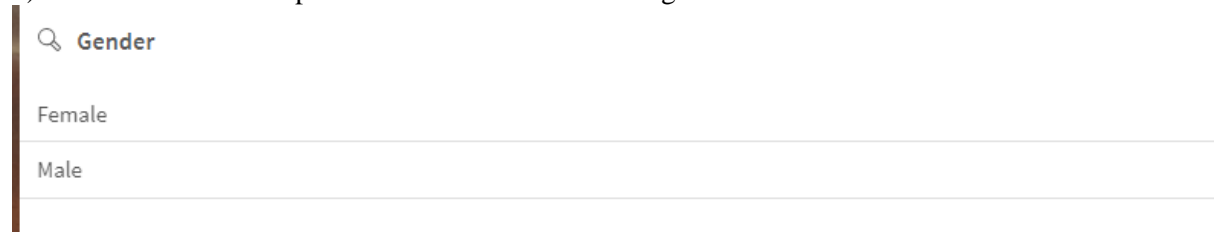
To create a visualization:

1. Navigate to the "Sheet" section in the "Analyze" tab located on the top menu.

2. Click "Edit Sheet" and select "Advanced Options" from the top-right corner.

3. Drag the desired chart onto the sheet area and add the necessary measures to generate the visualization.

The visualizations created are :

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

‹ Click to add title   ··· ×

Total No of Passengers

97.74k

Data

Measures
First KPI

Total No of Passengers   ˅

Expression

Count([Passenger ID])   *fx*

Label

Total No of Passengers   *fx*

Number formatting

Auto   ▼

Master item

⊘   Add new

🗑   Delete

Second KPI
Add-ons
Appearance

2) Gender : Add a filter pane 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 measures as passenger id and select count and in the fx tab write the query as =COUNT({< [Flight Status] = {'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 measures as passenger id and select count and in the fx tab write the query as =COUNT({< [Flight Status] = {'Delayed'} >} [Passenger ID]) and label as ='Number of ' & GetFieldSelections([Gender]) & ' Passengers effected by Cancelled flights'
Male :

Number of Male Passengers effected by Delayed flights
16.39k

Female :

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 measures 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 female passengers '.

## 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 passenger id 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

Month	Passengers
May	3.28k
Aug	3.28k
Jul	3.26k
Feb	3.23k
Sep	3.23k

Departure Date.Month
- Feb
- May
- Jul
- Aug
- Sep
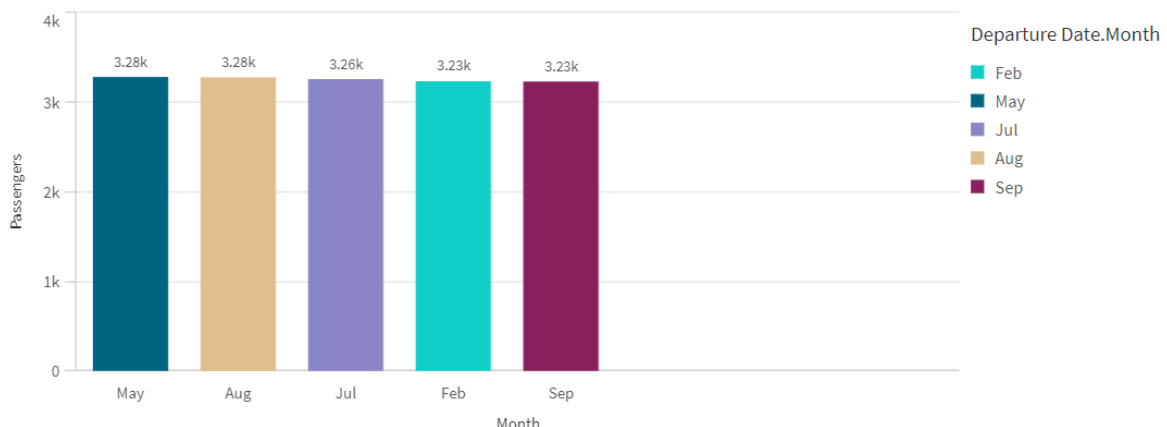
9)Top 3 months of people effected by flight status against the passengers count : Drag a bar chart on to the workstation and add passenger id count as bars and Departure date.Month as dimension and in
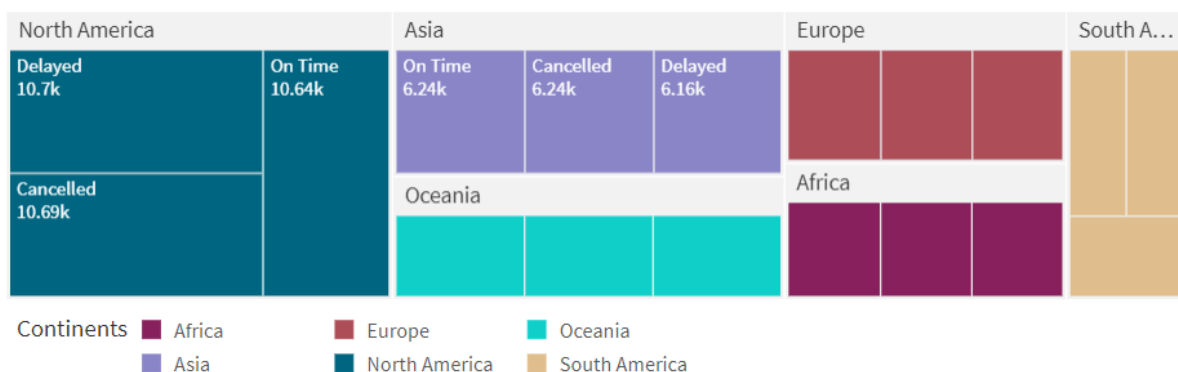
limitation section choose fixed number and enter 3. Now drag Flight status from fields on to the chart an click add flight status this will generate a grouped chart .Now go to appearance and choose a stacked chart and turn on labels and segment labels on.
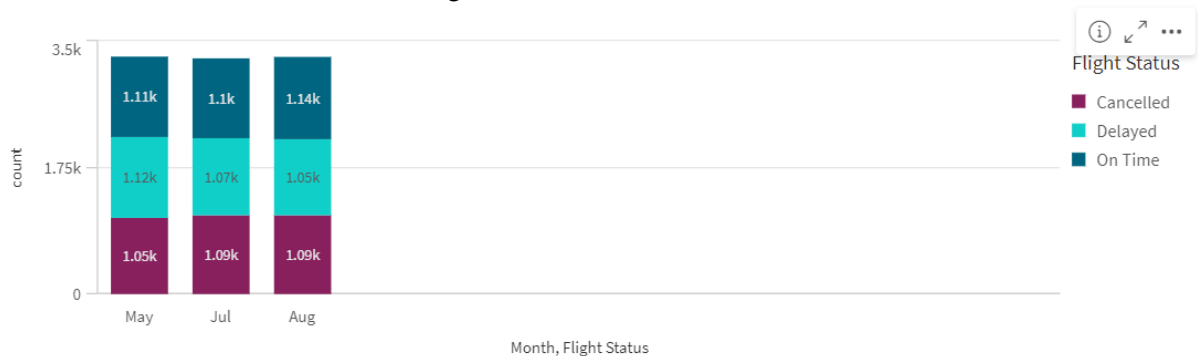
10) Age Group - Gender Wise : Drag a bar chart on to the workstation and add gender as bars and Agegroup as dimension. Now drag Gender from fields on to the chart and click add gender. This will generate a grouped chart.Now go to appearance and choose a stacked chart and turn on labels and segment labels on.
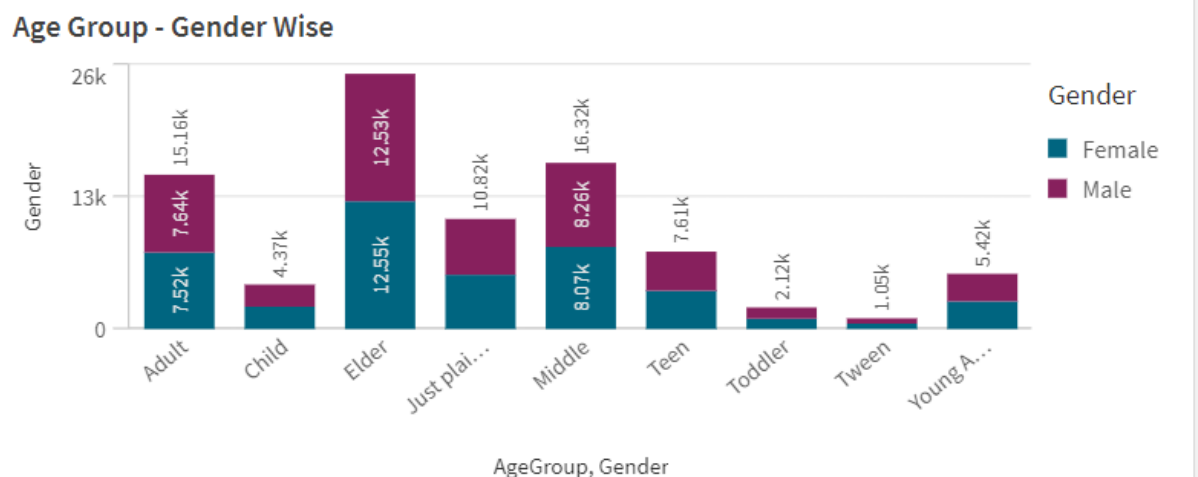

11) Continents wise - Flight Status : Drag a Tree map chart on to the workstation and add Continents as dimension. Now drag flight status from fields on to the chart and click add flight status. This will generate the chart.Now go to appearance and choose color as by dimension and turn on labels and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

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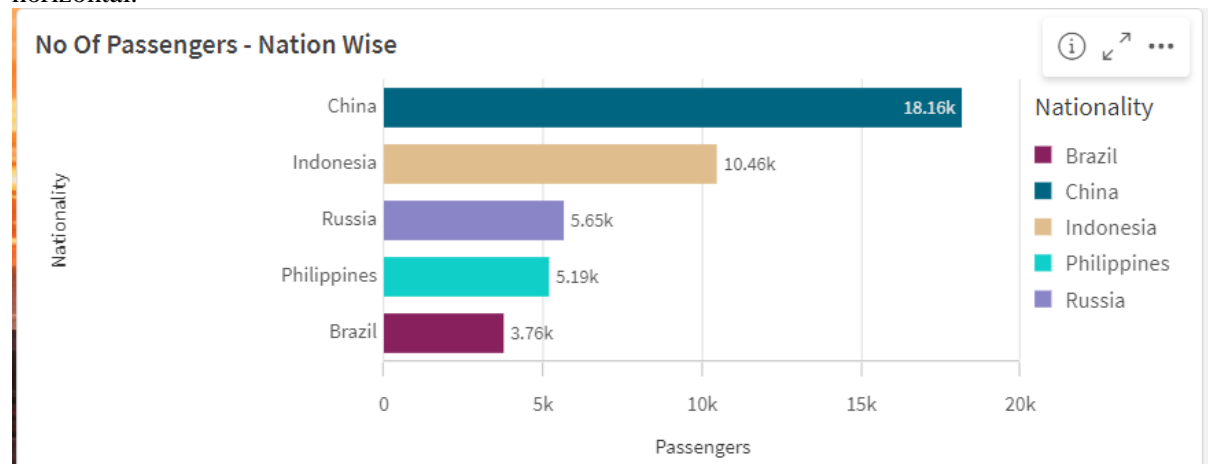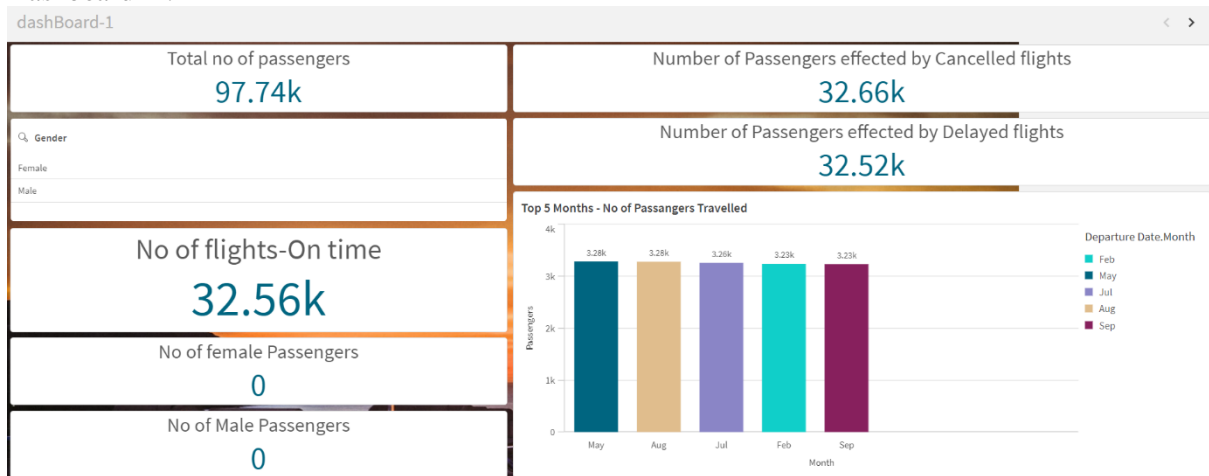

these are the visualizations.

6.Dashboards :

A dashboard is a graphical user interface (GUI) designed to display information and data in a structured and easy-to-read format. Often used for real-time monitoring and data analysis, dashboards are tailored for specific purposes or use cases. They are widely utilized across various settings, including business, finance, manufacturing, healthcare, and numerous other industries. Dashboards enable users to track key performance indicators (KPIs), monitor performance metrics, and present data visually through charts, graphs, and tables, facilitating quick and informed decision-making.
Step-1 : Now we need to assemble all the visualizations we made into a place
add a background image to these dashboards :

Dashboard-1 :

Total no of passengers	Number of Passengers effected by Cancelled flights
97.74k	32.66k

Q Gender

Female

Male

	Number of Passengers effected by Delayed flights
	32.52k

Top 5 Months - No of Passangers Travelled

No of flights-On time

32.56k

No of female Passengers

0

No of Male Passengers

0

Departure Date.Month
- Feb
- May
- Jul
- Aug
- Sep

3.28k 3.28k 3.26k 3.23k 3.23k

May Aug Jul Feb Sep

Month
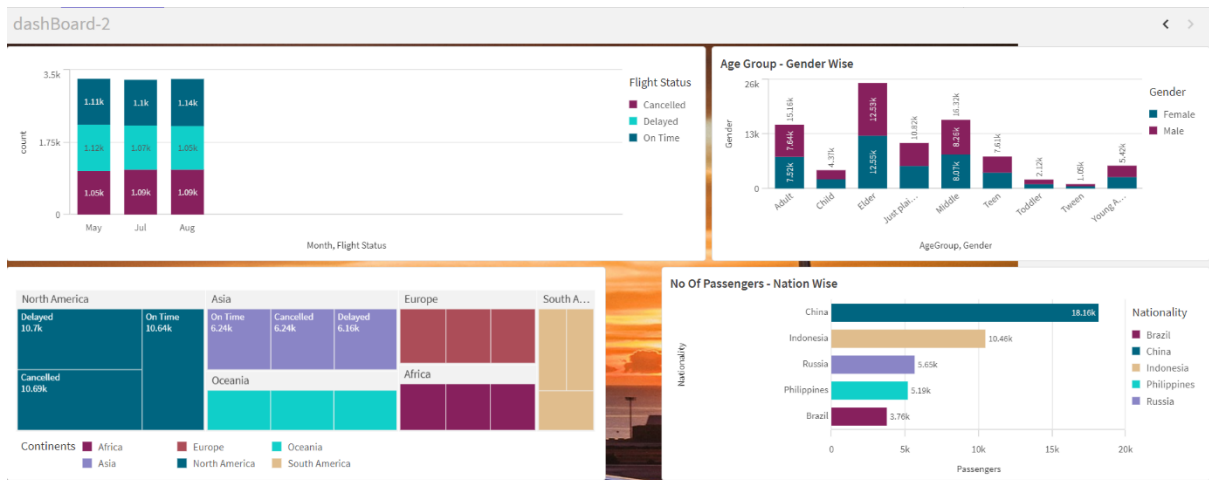
DashBoard-2 :

In this way we can make dashboards.

7.Report :

For report we use story telling 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
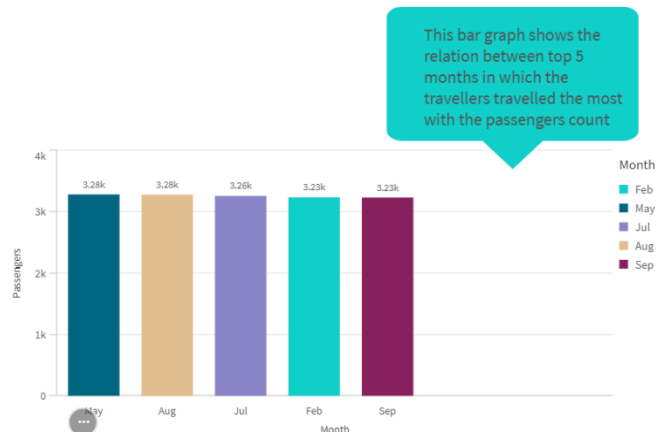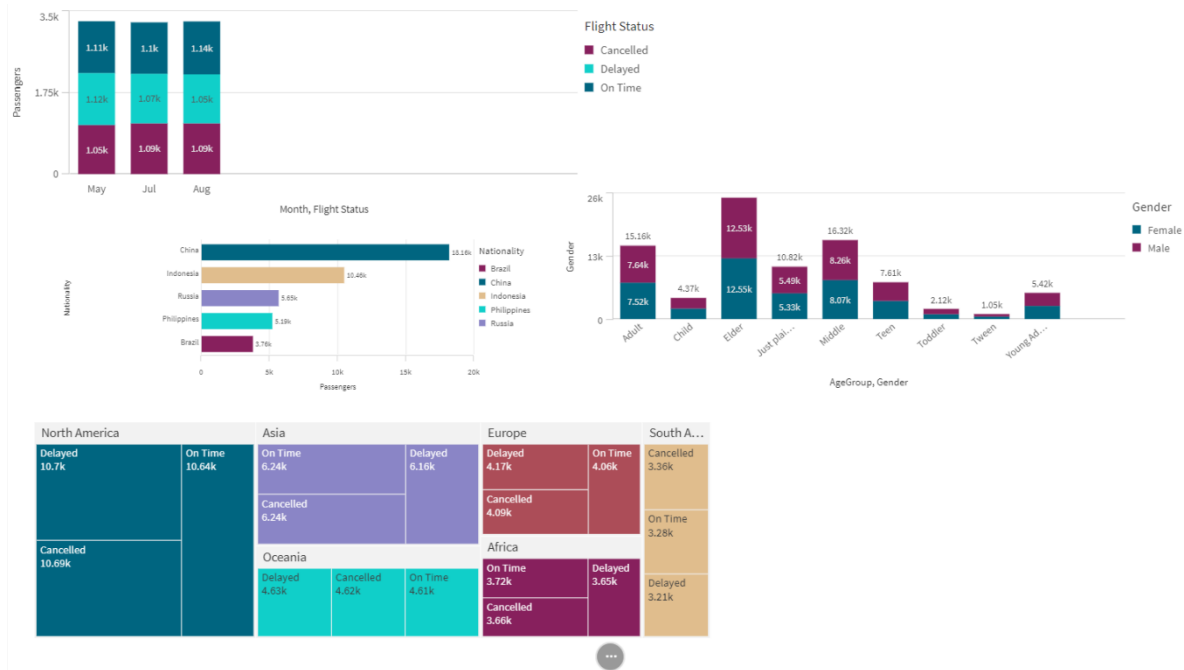
8.Performance Testing :
8.1.Amount of data rendered :
The "Amount of Data Loaded" refers to the volume or quantity of data that has been imported, retrieved, or loaded into a system, software application, database, or any other data storage or processing environment. This measure indicates the extent 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" refers to the strategic application or use of filtering mechanisms within a 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 are relevant and should be included in 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 on evolving 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 isolate pertinent information can significantly impact outcomes and performance.

dashBoard-1

Total no of passengers	Number of Female Passengers effected by Cancelled flights
48.57k	16.31k

Gender

Female ✓
Male

Number of Female Passengers effected by Delayed flights

16.13k

No of flights-On time
16.13k

No of female Passengers
48.57k

No of Male Passengers
0

Top 5 Months - No of Passangers Travelled

Departure Date.Month
- Jan
- Apr
- May
- Jun
- Jul

Assets | Insight Advisor | Gender Male ✕ | Bookmarks ⌄ | Sheets ⌄ | Edit sheet

dashBoard-1

Total no of passengers	Number of Male Passengers effected by Cancelled flights
49.17k	16.35k

Gender

Female
Male ✓

Number of Male Passengers effected by Delayed flights

16.39k

No of flights-On time
16.42k

No of female Passengers
0

No of Male Passengers
49.17k

Top 5 Months - No of Passangers Travelled

Departure Date.Month
- Feb
- May
- Aug
- Sep
- Nov