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 in-depth initiative designed to harness the power of Qlik, a leading business intelligence and data visualization platform, to analyze synthetic airline data. This data is meticulously crafted to emulate real-world airline operations, encompassing a wide array of elements such as flight schedules, passenger demographics, ticket sales, and performance metrics.

The primary objective of this project is to leverage Qlik's advanced analytical capabilities to uncover intricate patterns, emerging trends, and significant correlations within the synthetic data. By doing so, the analysis aims to provide actionable insights that can enhance decision-making processes for various stakeholders, including airlines, airports, and industry partners.

Through this sophisticated analysis, stakeholders will be equipped with a deeper understanding of operational efficiencies, market dynamics, and customer behavior. The project ultimately seeks to contribute to the optimization of airline operations, the enhancement of passenger experience, and the strategic planning efforts of industry participants.

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

Scenario 1: Revenue Optimization

An airline seeks to enhance its revenue optimization strategies by deeply analyzing historical ticket sales data. Leveraging Qlik's advanced data visualization and analytical capabilities, the airline can identify peak travel times, popular routes, and effective pricing tactics. This detailed analysis enables the airline to track revenue trends over different periods, segment customers by purchasing behavior, and refine pricing models to maximize profits. With these insights, the airline can implement dynamic pricing strategies, respond more effectively to market demands, and drive revenue growth through informed decision-making.

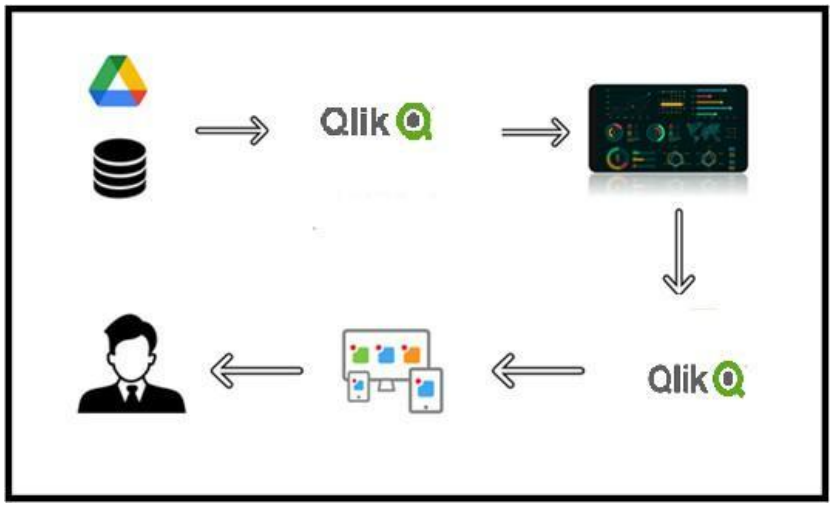
Scenario 2: Operational Efficiency

An airport authority aims to boost operational efficiency by examining extensive data on flight schedules, passenger flows, and luggage handling processes. By integrating Qlik with synthetic airline data, the authority can detect operational bottlenecks, predict peak traffic times, and optimize resource allocation. Qlik's powerful data integration and visualization tools allow the airport to streamline operations, from gate assignments to baggage handling, ensuring smoother passenger experiences and better utilization of airport resources. This proactive management approach results in faster turnaround times, fewer delays, and enhanced overall efficiency.

Scenario 3: Customer Experience Enhancement

Airlines are increasingly focused on improving the passenger experience by gaining a deeper understanding of customer preferences, satisfaction levels, and pain points. Using Qlik for sentiment analysis on customer feedback data, airlines can identify key improvement areas and trends in customer sentiment. This data-driven approach allows airlines to personalize services, tailor marketing campaigns, and address specific customer needs more effectively. By proactively enhancing service quality and resolving customer concerns, airlines can increase customer loyalty and satisfaction, ultimately achieving a stronger competitive edge. These insights help ensure 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 significant challenges in revenue optimization, operational efficiency, and customer experience enhancement. The airline struggles with maximizing revenue due to a lack of insights into historical ticket sales, travel patterns, and customer purchasing behaviors, leading to suboptimal pricing strategies. Simultaneously, the airport authority deals with inefficiencies in managing flight schedules, passenger flows, and luggage handling, causing operational bottlenecks and poor resource allocation. Additionally, airlines lack a comprehensive understanding of customer preferences and satisfaction, limiting 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 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.

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.

2.3 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.

2. Analytical Capabilities

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.

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. Visualization and Reporting
Title: "The Visual Display of Quantitative Information"
Authors: Edward R. Tufte
Summary: Tufte's work is a seminal guide on the principles of data visualization and the

effective presentation of data insights through graphical means.
Relevance: Essential for developing interactive dashboards and customizable reports in Qlik, ensuring data is presented clearly and effectively to stakeholders.

4. Predictive Analytics

Title: "Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die" Authors: Eric Siegel
Summary: This book explains the fundamentals of predictive analytics and how predictive models can forecast future events and behaviors in various business contexts.
Relevance: Supports the implementation of demand forecasting and operational predictions, crucial for dynamic pricing and resource optimization.

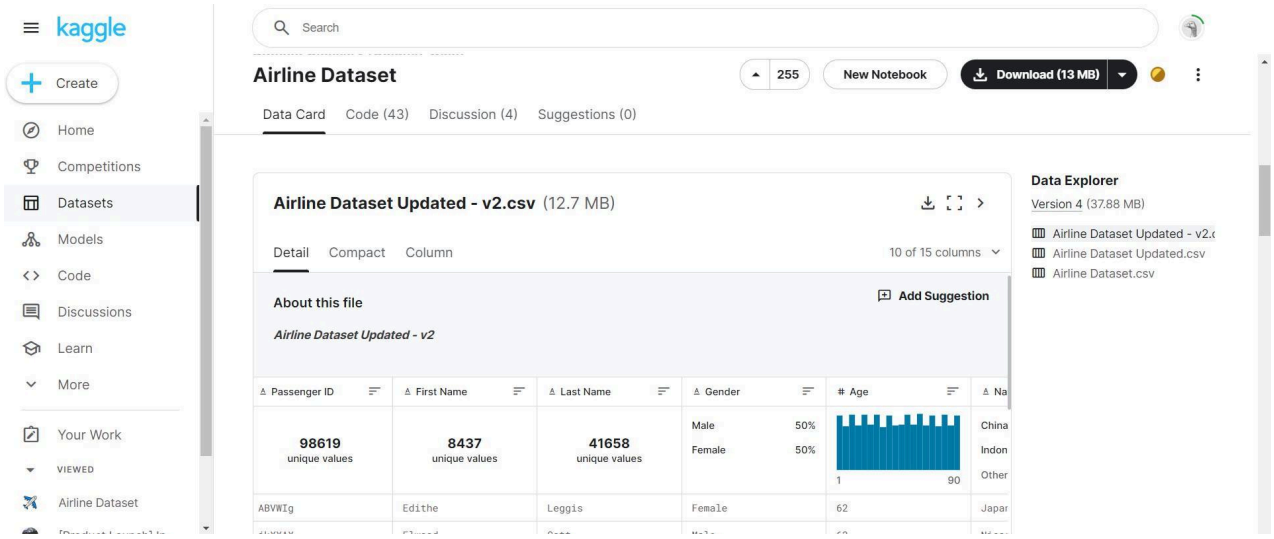
Summary :

The literature survey covers key areas critical to the successful implementation of Qlik for airline and airport operations, including data integration, analytical techniques, data visualization, predictive analytics, user training, and performance monitoring. Each source provides theoretical and practical insights that inform the project's requirements and objectives.

3. Data Collection :

3.1 Collect the dataset :

Data collection is a detailed process that involves systematically gathering and measuring information on various variables of interest. This approach is designed to answer research questions, test hypotheses, evaluate outcomes, and generate insights. It begins with identifying key variables and developing a plan to gather this information accurately and consistently. Various methods, such as surveys, interviews, observations, and experiments, are employed to collect data based on the research needs. Ensuring adherence to established standards, the collected data is regularly reviewed for accuracy and reliability. Ultimately, this process aims to create a robust dataset to support research objectives and provide meaningful insights for decision-making and further analysis.
here we downloaded the data as a tables format from the website Kaggle.
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 essential steps to ensure the data is ready for analysis. Initially, the data is cleaned to remove irrelevant or missing information, guaranteeing its accuracy. Next, the data is transformed into a suitable format for visualization tools, making it easier to create meaningful visual representations.
Following this, the data is explored to identify patterns and trends, providing valuable insights into its structure. Filtering the data allows for a more focused analysis on specific subsets, enhancing understanding in particular areas.
Subsequently, the data is prepared for visualization software, ensuring correct formatting, completeness, and accuracy. This ensures the data is easily interpretable and ready for creating visualizations that effectively communicate insights.
With the data cleaning process completed, we can now seamlessly move on 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.
Steps are Data cleaning and Data Preprocessing :
In data cleaning we remove incomplete missing and unwanted irrelevant data, thereby establishing the data needed only
In data Preprocessing we add any required additional data or remove data that is not useful. In this stage the data preprocessing would result in preparing the data for better visualizations.
Data Preprocessing Code :

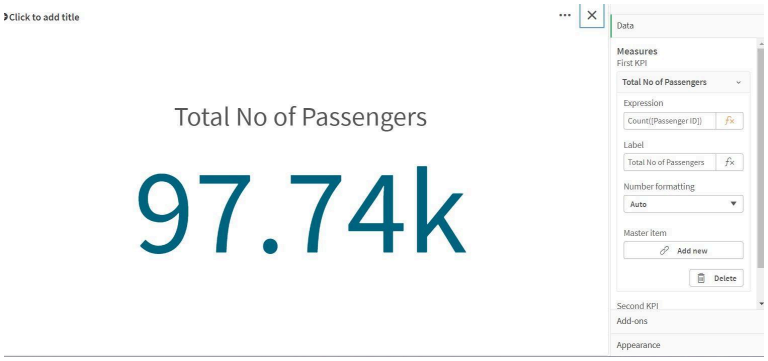
```
[Airline Dataset Updated - v2]:
Load;
[Airline Dataset Updated - v2]:
NOCONCATENATE LOAD
if(Age AND Age <=1, 'Baby',
if(Age >= 1 AND Age <=3, 'Toddler',
if(Age >=4 AND Age <=9, 'Child',
if(Age >= 10 AND Age <=12, 'Tween',
if(Age > 13 AND Age < 19, 'Teen',
if(Age >=20 AND Age <=24, 'Young Adult',
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]= '-');
```

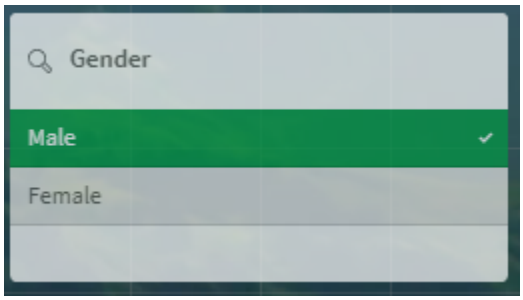
5.Data Visualizations :

Data visualization is the process of crafting graphical representations of data to help individuals comprehend and explore the information effectively. The primary objective is to transform complex data sets into more accessible, intuitive, and interpretable formats. By utilizing visual elements such as charts, graphs, and maps, data visualizations enable users to identify patterns quickly, trends, and outliers within the data. This approach not only enhances the ability to grasp intricate data relationships but also aids in making informed decisions by presenting data in a clear and visually engaging manner.
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 passnegerid count as the formula and label as total no of passengers .



2) 

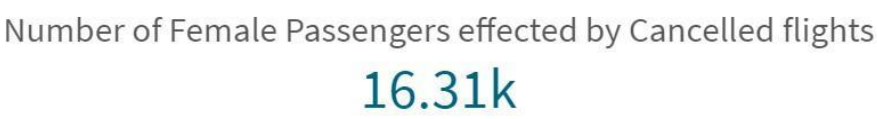
Gender : Add a filter pane and select the measure as gender and add it to the sheet

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

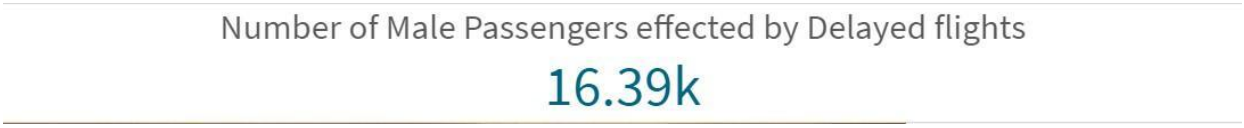
Female :



4) No of passengers affected 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 affected by Canceled flights'

Male :

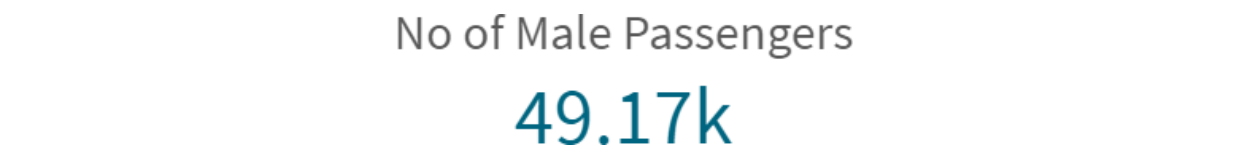


Female :

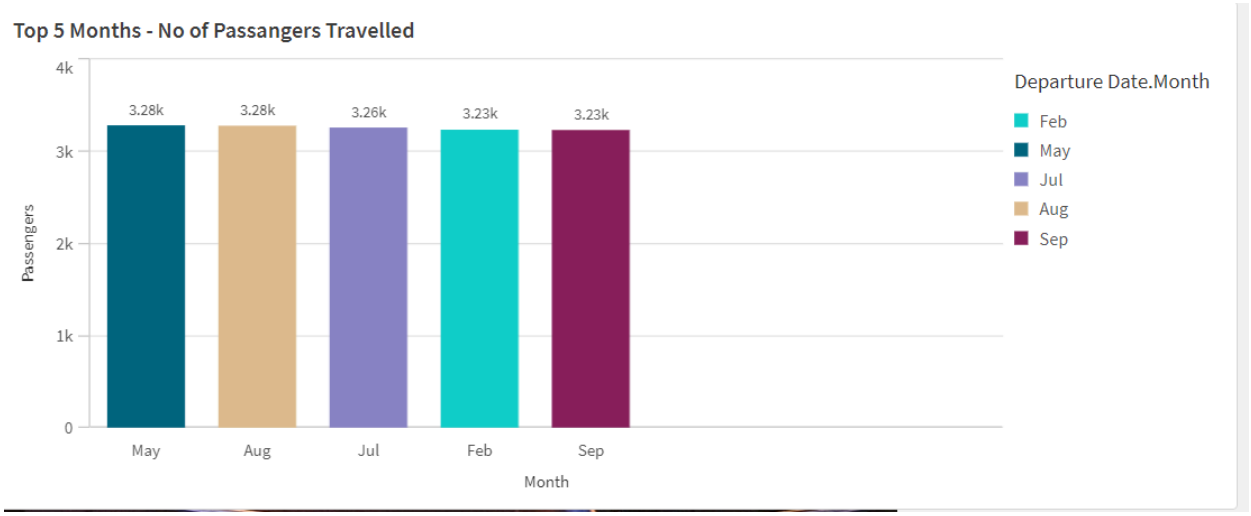
5) No of Flights on time : Drag a Kpi chart on to the sheet and add the required measures as passenger id and select count and in the fx tab write the query as
=COUNT({< [Flight Status] = {'On Time '} >} [Passenger ID]) and label as ='Number of flights on time '.

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 '.

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 '.

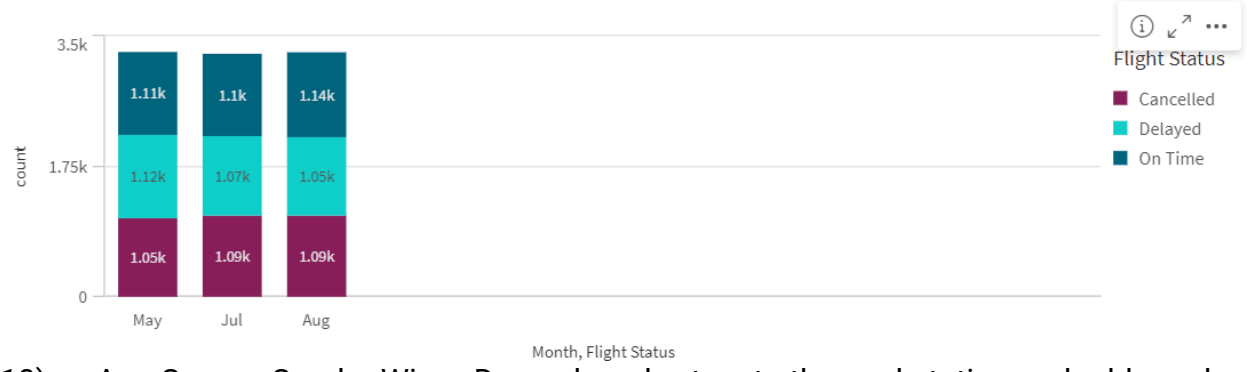


8) Top 5 months in which passengers traveled the most : Drag a bar chart on to the workstation and add passenger id count as bars and Departure date.Month as dimension and in limitation section choose fixed number and enter 5 and In appearance choose color by dimension and turn on the labels .

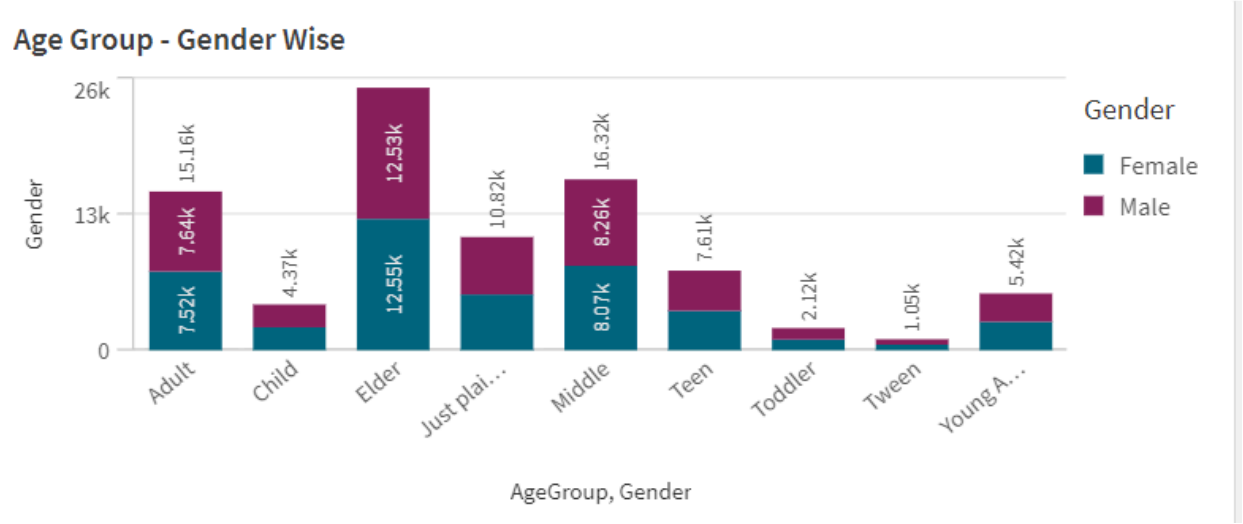


9) Top 3 months of people effected by flight status against the passengers count : Drag a bar chart on to the workstation and add passenger id count as bars and Departure date.Month as dimension and in limitation section choose fixed number and enter 3.

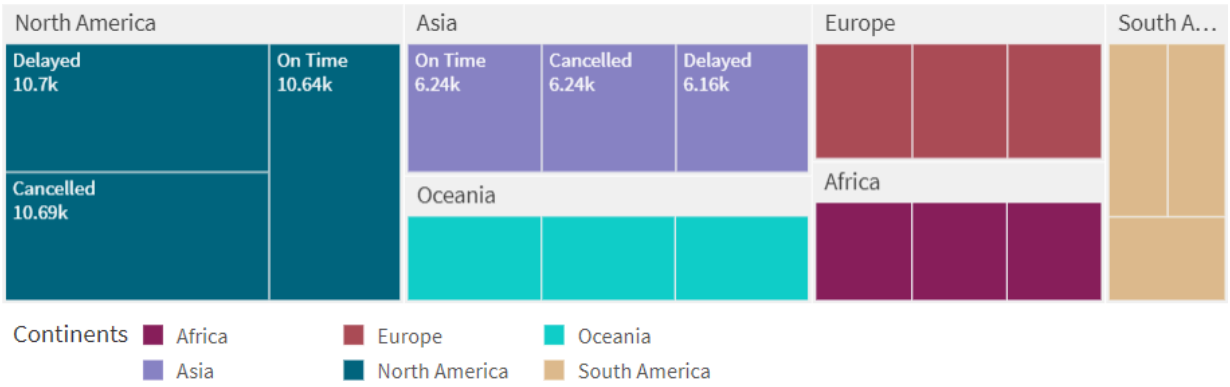
Now drag Flight status from fields on to the chart an click add flight status this will generate a grouped chart .Now go to appearance and choose a stacked chart and turn on labels and segment labels on.



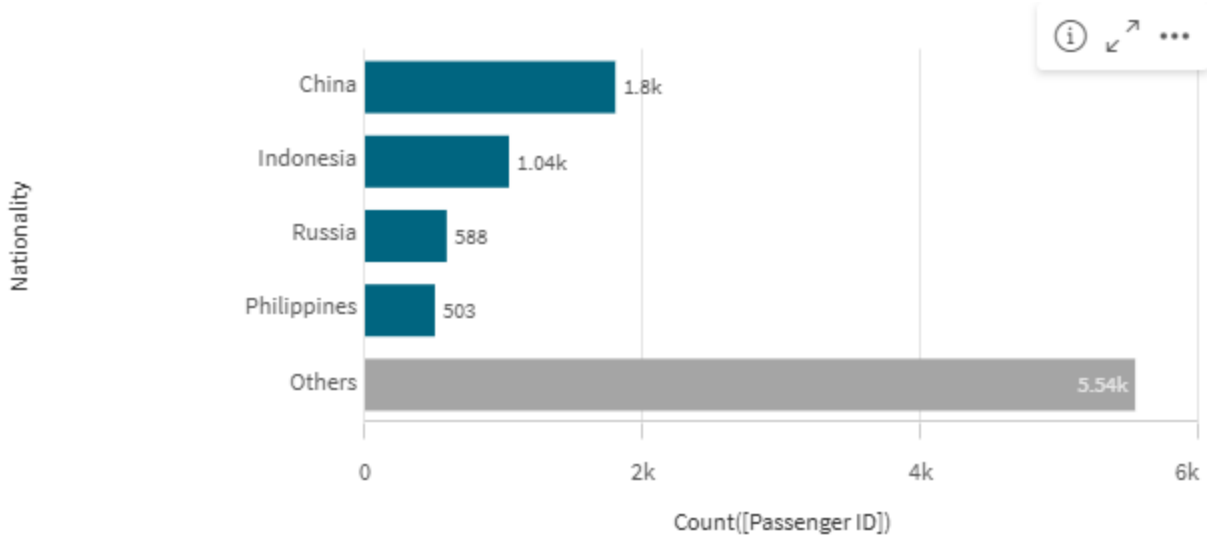
10) Age Group - Gender Wise : Drag a bar chart on to the workstation and add gender as bars and Agegroup as dimension. Now drag Gender from fields on to the chart and click add gender. This will generate a grouped chart.Now go to appearance and choose a stacked chart and turn on labels and segment labels 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 fixed number and enter 5 and In appearance choose color by dimension and turn on the labels and also choose orientation to be horizontal.

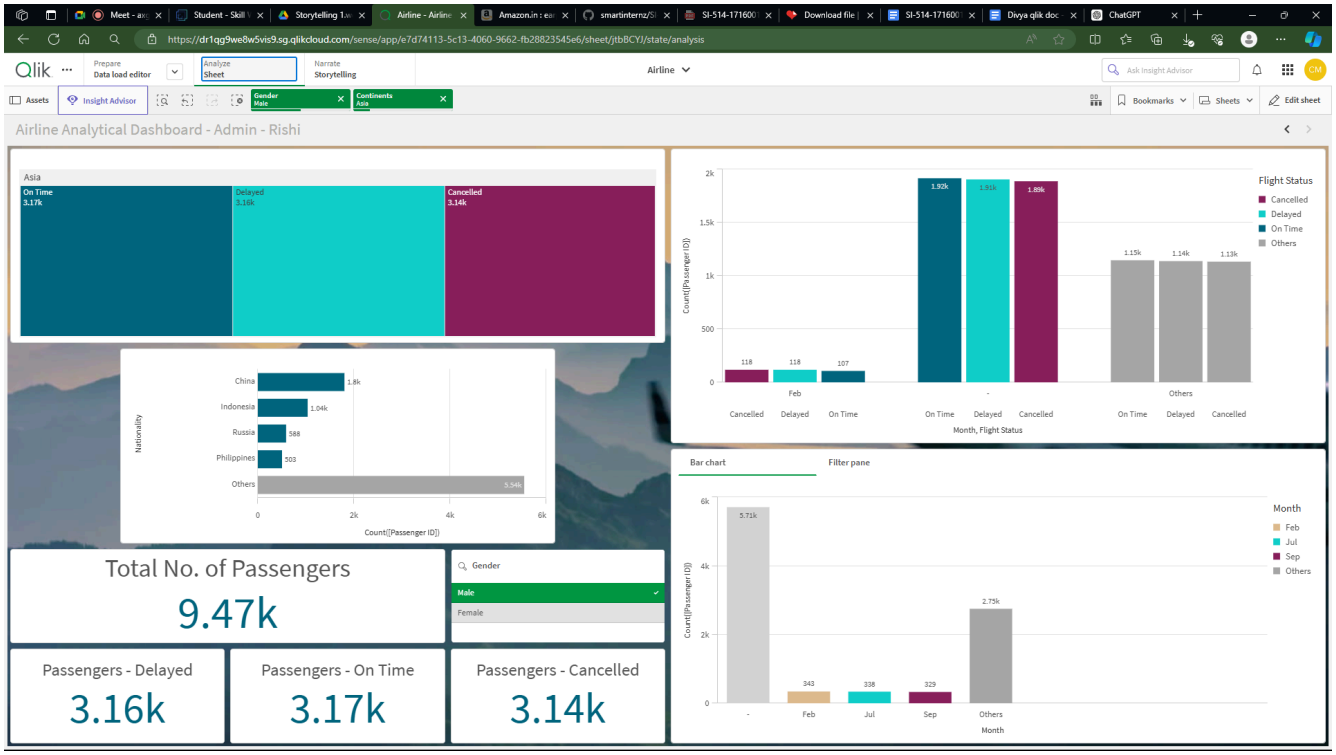


these are the visualizations.

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 :

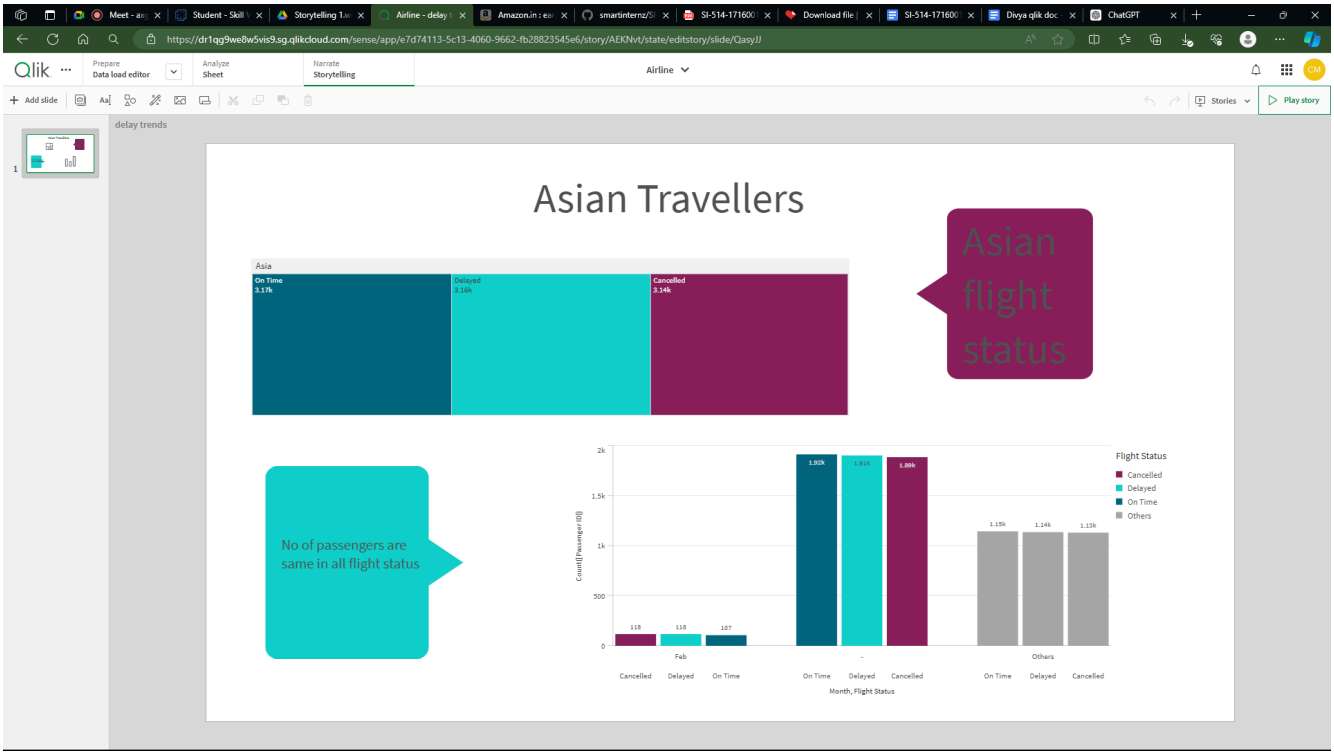
Dashboard - 1

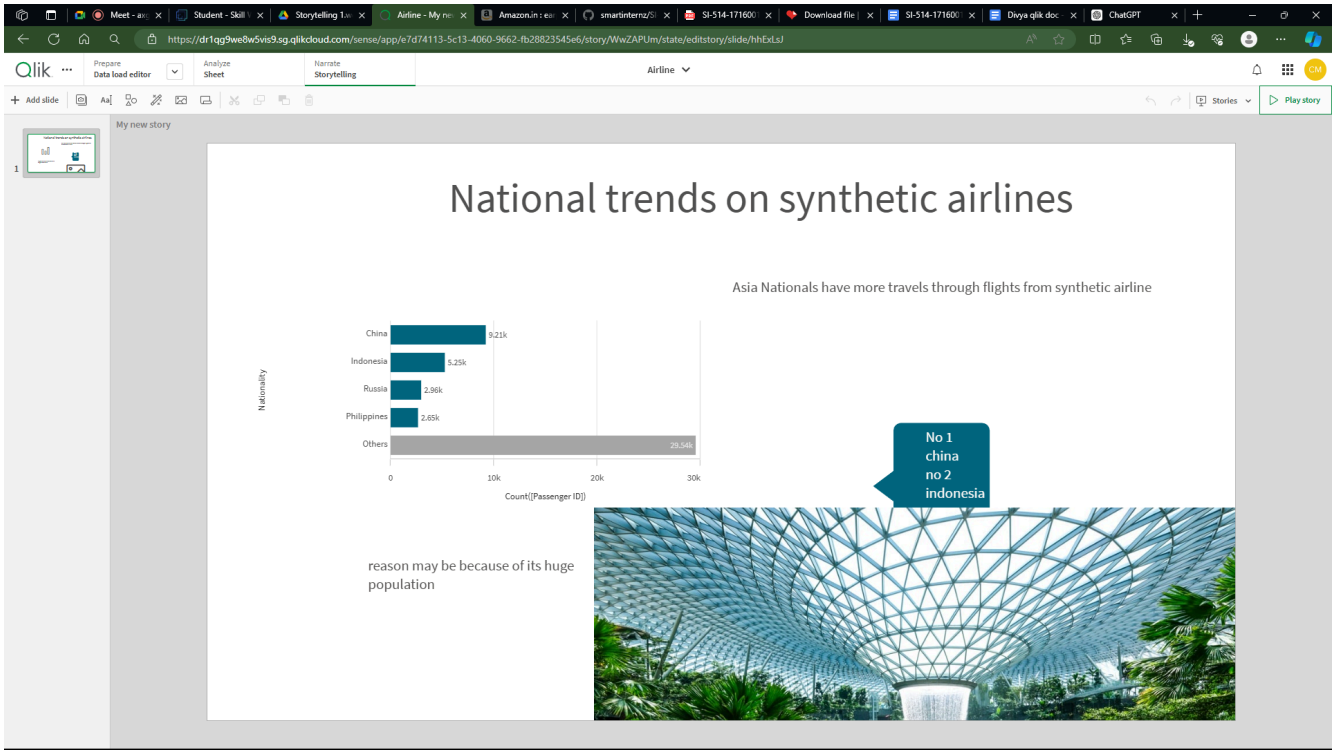


In this way we can make dashboards.

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.