



SYNTHETIC AIRLINE ANALYSIS/QLIK

K. VYSHNAVI

Table of Contents:

- Introduction
- Understanding the Problem
- Data Collection and Extraction
- Data Preparation
- Data Visualization
- Dashboard
- Story
- Performance Testing

Introduction:

The airline industry encompasses various businesses, called airlines, which offer air transport services for paying customers or business partners. These air transport services are provided for both human travelers and cargo, and are most commonly offered via jets, although some airlines also use helicopters.

In a competitive industry like aviation we need to more competitive by using the strategic and innovation applications..

Qlik insights provides the organization with vast amount of data generated across the world. Qlik provides the Comprehensive view of data. This integration allows the businesses to identify inefficiencies , predict demands , optimize inventory levels ,streamline operations.This helps in decision- making processes.

Purpose or Objective:

The project "Exploring Insights from Synthetic Airline Data Analysis with Qlik" involves utilizing synthetic airline data to derive valuable insights using Qlik, a business intelligence and data visualization tool.

In this project, the synthetic airline data simulates various aspects of airline operations, including flight schedules, passenger demographics, ticket sales, and performance metrics. The objective is to leverage Qlik's analytical capabilities to uncover patterns, trends, and correlations within this data, aiding in decision-making processes for airlines, airports, and related stakeholders

Gains of the Project:

Revenue optimization:

An airline wants to optimize its revenue by analyzing historical ticket sales data, identifying peak travel times, popular destinations, and pricing strategies. Using Qlik, they can visualize revenue trends over time, segment customers based on purchasing behavior, and adjust pricing strategies accordingly to maximize profitability.

Operational Efficiency:

An airport authority aims to enhance operational efficiency by analyzing flight schedules, passenger flows, and luggage handling processes. By integrating Qlik with synthetic airline data, they can identify bottlenecks in airport operations, predict peak traffic periods, and allocate resources effectively to streamline processes and improve overall efficiency

Customer Experience Enhancement

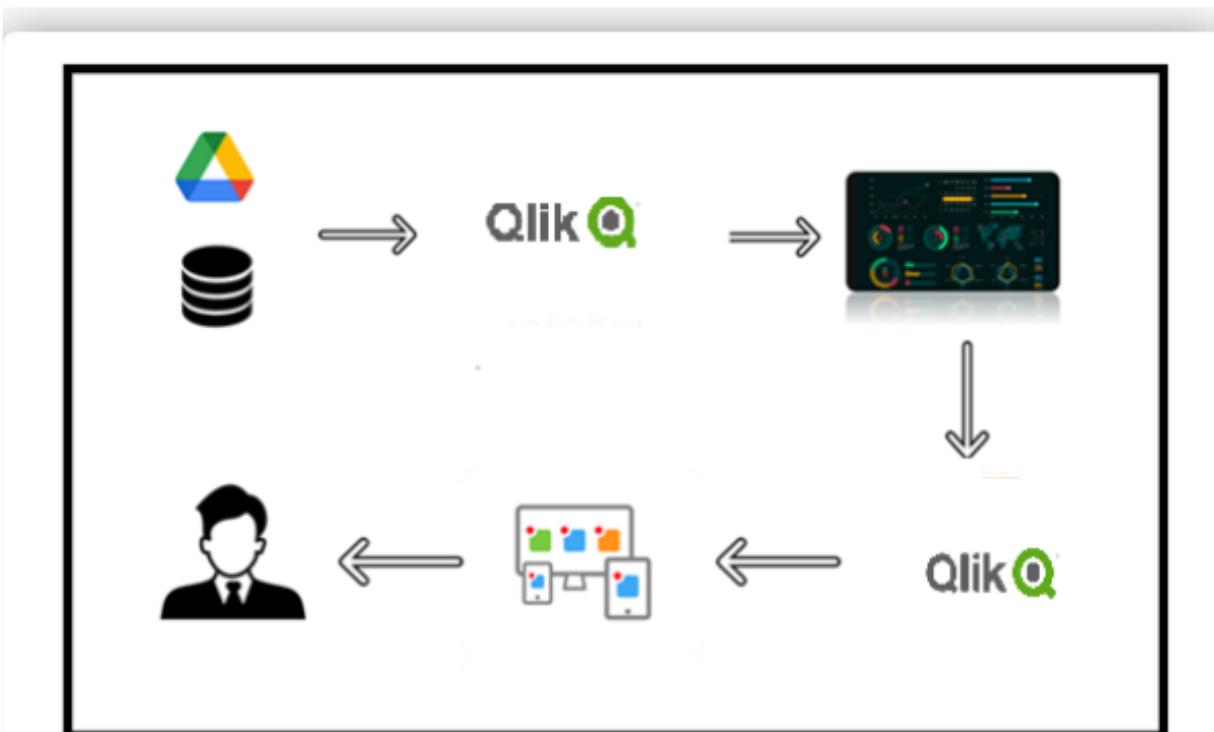
Airlines are keen to enhance the passenger experience by understanding customer

preferences, satisfaction levels, and pain points. Through sentiment analysis on customer feedback data integrated with Qlik, airlines can identify areas for improvement, personalize services, and tailor marketing campaigns to better meet customer needs, ultimately fostering loyalty and satisfaction.

Increasing the customer satisfaction:

By knowing the customer feedback we can help the customer with preferences. The customer preferences plays a crucial role in the revenue and operational management which makes the business run smoothly.

Technical Architectures:



Understanding the Problem:

- **Specifying the problem:**

This project revolves around the synthetic airlines data , where the main operations of this data collection to increase the revenue and better management resources, as well as customer satisfaction. Leveraging the advanced analytics , it seeks for better insights and forecasting the strategies.Using the Qlik sense analytics we can understand the data for our necessities.

- **Business requirements:**

As the airline industry handles the massive data , it requires the real time data, without any false distractions of the data. The real time and primary data provides the complete architecture about the airlines industry so that we can elaborate the requirements for the necessities . The industry should able to adopt the trends and patterns.

To identify such patterns we use the Qlik analytics capabilities to examine the historical data and optimize the revenue and customer satisfaction. It provides the user-friendly and dynamic dashboards , giving the viewers clear insights.

- **Literature Survey:**

A literature survey refers to the students analyse critically and concisely earlier research and literature related to a particular research problem , and utilise them for their own research.

To be precise , it is the secondary data as it is used before , it is used for the reference to know the historical patterns and identify if it is occurred before.

Literature survey helps the students in understanding the significance of new research and its connections to earlier world.

Literature survey helps the students in understanding the significance of new research and its connections to earlier work.

From the start, the airline industry has remarkably connected countries all over the world through rapid long-distance transportation, helping people overcome geographic barriers.

The airline industry produces vast amounts of data, capturing a diverse set of information about their operations, including data related to passengers, freight, flights, and much more. Analyzing air travel data can advance the understanding of airline market dynamics, allowing companies to provide customized, efficient, and safe transportation services.

Revenue Management (RM) was first introduced into the airline industry in the 1970s, which promoted the use of differentiated pricing—charging passengers different fares for multiple booking classes . Traditional RM derives marketing and pricing strategies based on economic theory and is popularly applied and developed in many industries to design the appropriate products that satisfy the customers' interests.

This is survey taken from the dl.acm.org form , this analysis is based on revenue management of the airlines industry which is mainly focused on the RM , operational cost etc..In this survey they used the machine learning for finding the insights of the survey, but here we use Qlik Sense for insight and better decision making.

Data Collection and Extraction:

Data collection refers the collecting the data based on the requirements .

Download the data set:

The data has been downloaded from the kaggle :

<https://www.kaggle.com/datasets/iamsouravbanerjee/airline-dataset>

Understand 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)

Connect the data with Qlik:

>>Open the Qlik cloud
 >> Login
 >>Create new analytics app
 >>Create the app with personal space
 >>Click on create
 >>Open the app
 >>Load the dataset
 >>Cleaning
 >>Data preparation
 >>Following visualizations.

Data Preparation and Extraction:

Data preparation refers to the process of cleaning, transforming, and organizing raw data into a structured format that is suitable for analysis or processing. This step is crucial for ensuring that the data is accurate, consistent, and usable.

This process involves few steps to ensure the data is effective and efficient for the visualizations. The process begins with cleaning the dataset and null values. This transformation is followed by exploring the data to identify the patterns and trends.

These main steps like cleaning, transform collectively help to make the data easily understandable and ready for creating the visualizations that provide insights into performance and efficiency.

Extraction:

Extraction refers to collecting the data from different sources. It is raw data collected from various sources. The data collected should be in way that it is required for further analysis.

Data Visualizations:

Data visualization is the process of creating graphical representations of data to help people understand and explore the information. The primitive goal of 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.

This approach helps to transform the raw data into meaningful insights, making the tool available for decision making and optimization.

Different types of visualizations:

There are several numbers of charts which can be used for data visualizations. Common types of visualizations used to analyse the performance and efficiency of banks include bar charts, line charts ..etc., These maps or charts can be analysed to determine the performance efficiency of the data and track the changes for better understanding the data and make data driven decisions.

The following are the visualizations based on the dataset.

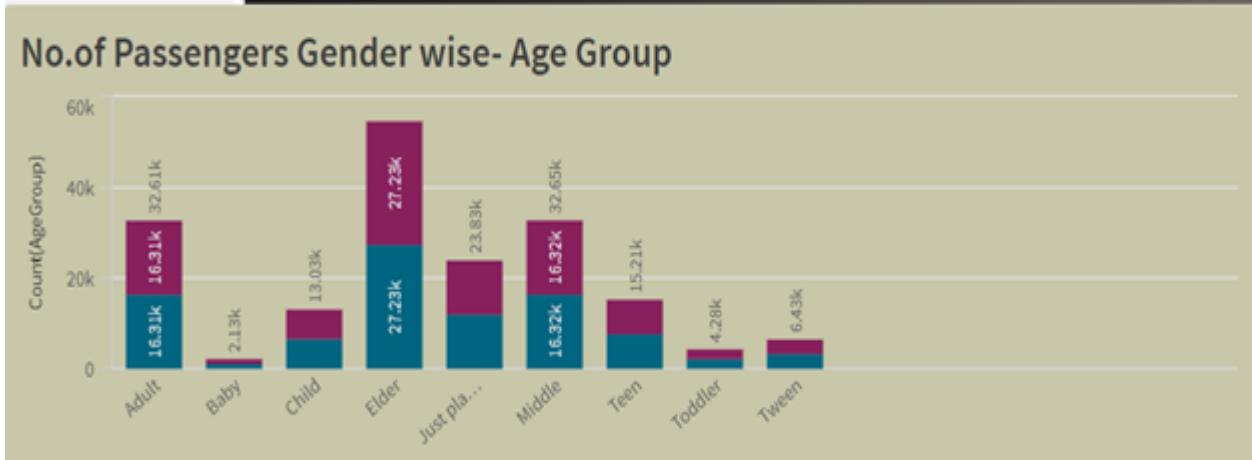
- Total no. of Passengers:



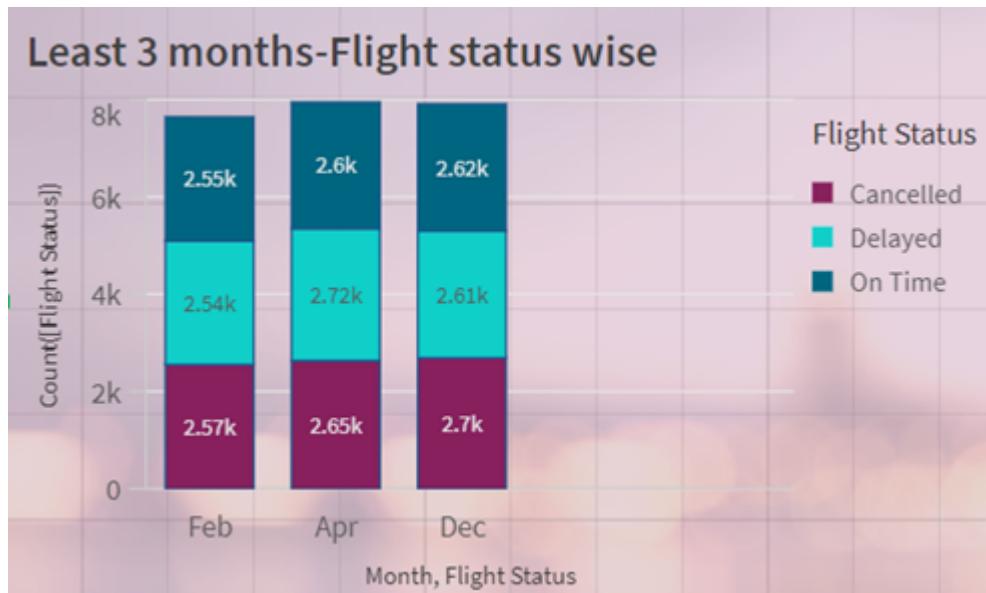
- No. of Flights cancelled per month:



- Passengers as per age group:



- Least 3 months- Flight wise:



- Top 5 months – Flight Cancelled:



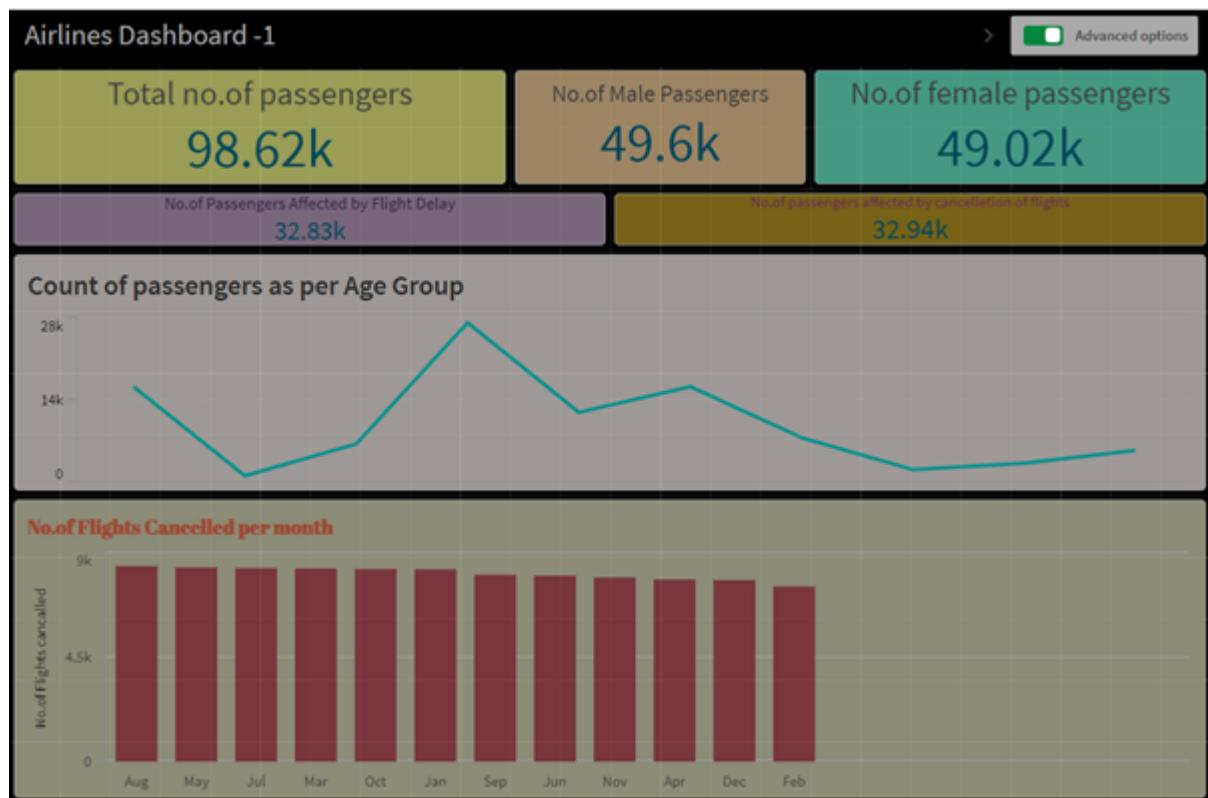
- Count of passengers travelling to different countries:

Passengers travelling to different countries		
Country Name	▲	Count([Passenger ID])
Totals		98819
Afghanistan		372
Albania		12
Algeria		449
American Samoa		45
Andorra		7
Angola		445
Anguilla		14

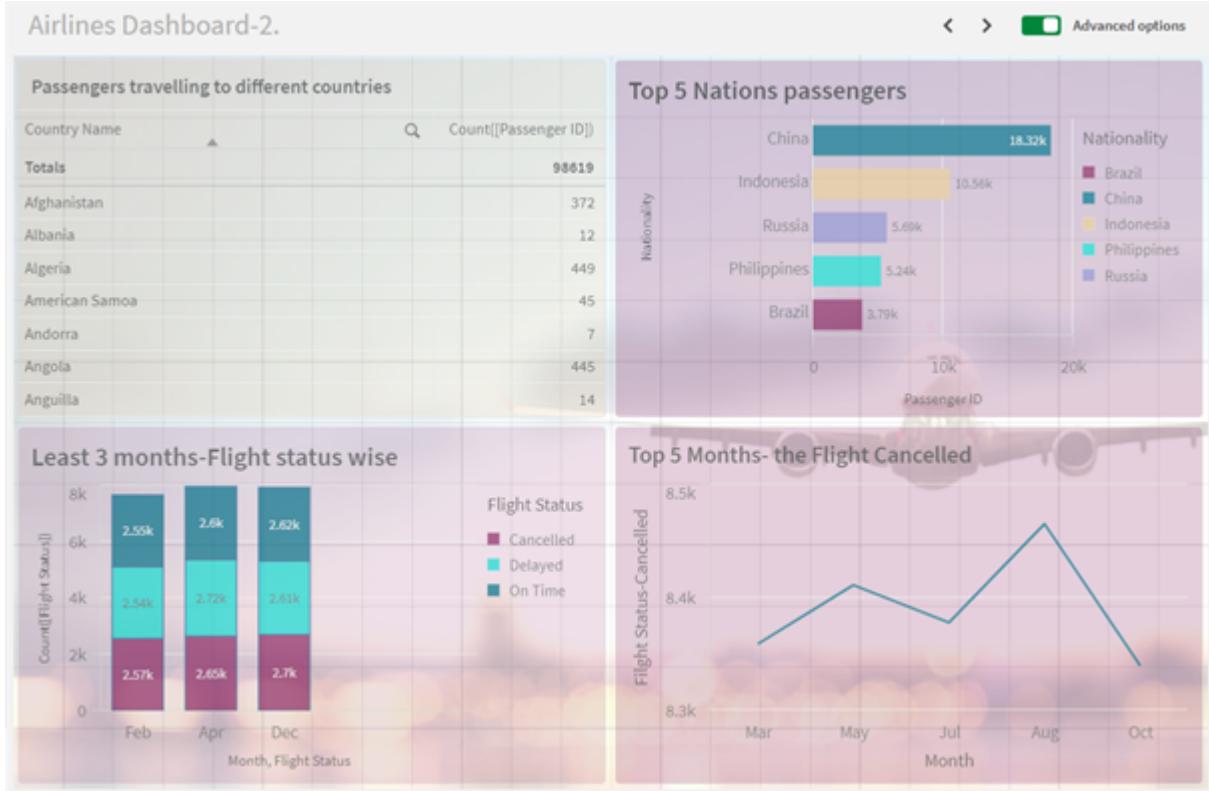
Dashboard:

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.

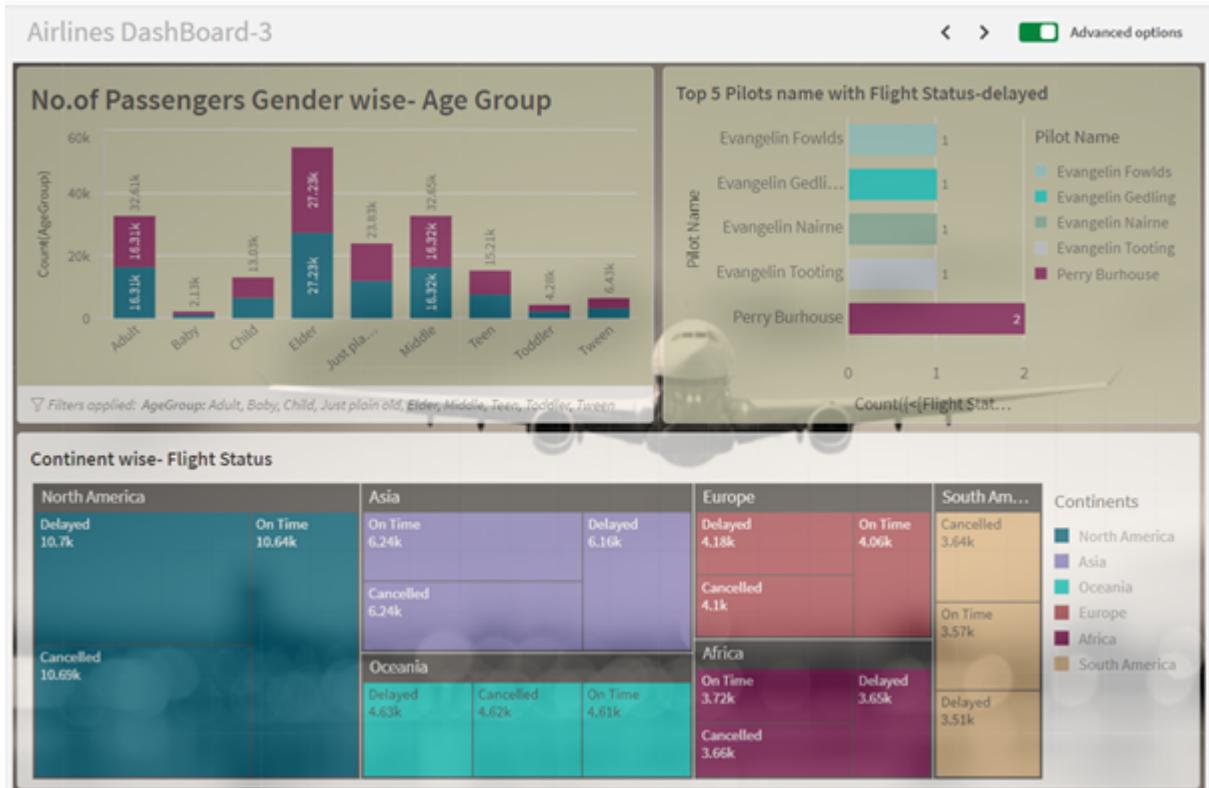
Dashboard :1



Dashboard:2



Dashboard :3



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.

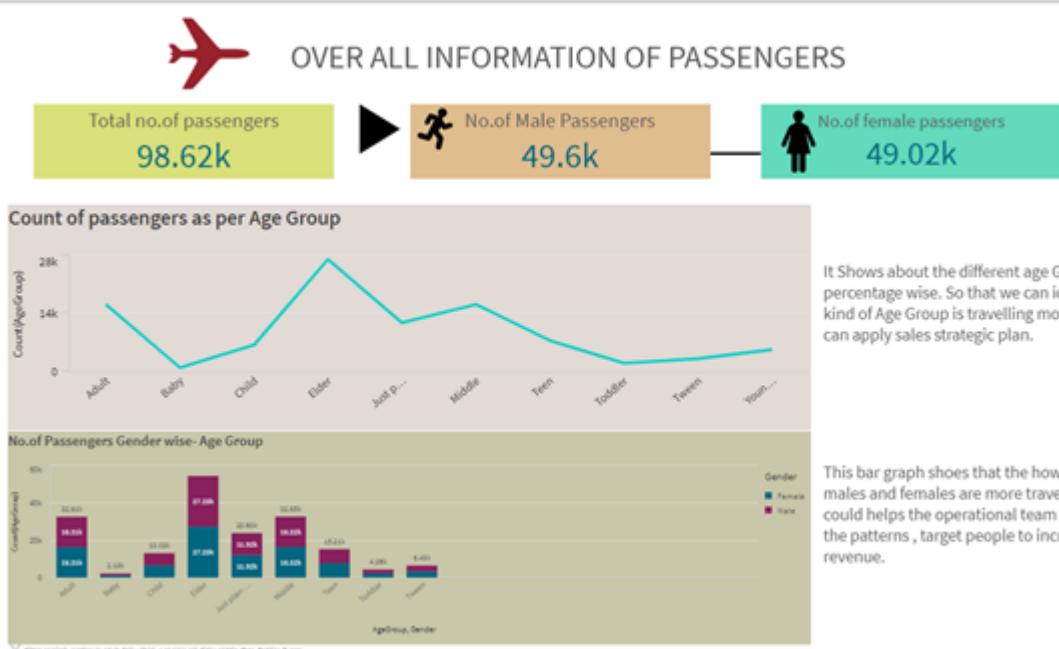
Building a Story:

The purpose of a story is to gather insight and build narratives from your data. You structure the story to convince your audience, by taking snapshots of your visualizations and put them in slides. To make your slides stand out you can apply visual effects, add text and shapes, and can embed sheets to make your story interactive. A story is connected to its app, so you can return to the live data anytime, to discover new and hidden stories.

Organising the visualizations logically will ensure the coherent flow of information. Dashboards will help to make the story building which provides the summarized view to key insights.

Throughout the process the main aim of the story making will ensure to reach the viewers and communicates the data more clearly and concisely.

Story-1:



Story- 2:



Flight Status Affects on the Passengers

No.of passengers affected by cancellation of flights

32.94k

No.of Passengers Affected by Flight Delay

32.83k

No.of Flights Cancelled per month X



From this Graph, we can understand how many flights is been cancelled in every individual month. We can the predict for next year which month the flight can be cancelled and we can identify the patterns and develop the plan regarding the certain circumstances.

Story -3:



Flight Status based on the months

Top 5 Months- the Flight Cancelled



In this graph, we can identify top 5 months which are commonly had the flight cancellation and estimating the top month where there are highest number of cancellations and can improve the operational services and decrease the issue.

In this graph we can understand least 3 months having the issues with flight status , as it is showing there are mostly similar , having the same count. So that it could based on the climatic conditions , engineer working or etc.

Least 3 months-Flight status wise



Story 4:



Continents wise- Passengers and Flight Status



Continent wise- Flight Status



Performance testing :

Amount of Data Loaded

"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 data storage 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
Departure Date-Departure_Date ↗
Passenger ID
First Name
Last Name
Gender
Age
Nationality
Airport Name
Airport Country Code
Country Name
Airport Continent
Continents
Arrival Airport
Pilot Name
Flight Status
Airline Dataset.Nationality_GeoInfo
Airline Dataset.Airport Country Code_GeoInfo
Airline Dataset.Country Name_GeoInfo
Airline Dataset.Nationality_GeoInfo_1
Airline Dataset.Airport Country Code_GeoInfo_1
Airline Dataset.Country Name_GeoInfo_1

Data Pre-Processing – Qlik Sense Script:

```
[Airline Dataset]:
Load* ;
[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] = '-');
```

Here we added the column using the script , we can also add the column using the calculate field. This added field will help in analysing the key insights.

Utilization of 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.

Here we can use the filter in the case of male and female criteria.

No.of Visualizations /graph:

- Total no.of passengers
- No.of male passengers
- No.of female passengers
- No.of filghts Cancelled per month
- No.of passengers affected by cancelled of flights
- No.of passengers affected by flights delayed
- Count of passengers travelling to different countries

- Top 5 nations passengers
- Top 5 months – filght cancelled
- Least 3 month – flight status wise
- No.of passengers Gender wise- Age Group
- Continent wise- Flight status