

# ***Synthetic Airline Data Analysis***

## **1. Introduction:**

### 1.1 Overview :

This project is about using Qlik tools to analyze synthetic airline data. The dataset includes information like passenger details, airport names, continent and country names, and dates.

The main challenge is to understand trends in the number of passengers. Reading through tables full of names and locations is hard, but business analytics tools like Qlik make it easier. With Qlik, we can turn the data into clear and simple visualizations.

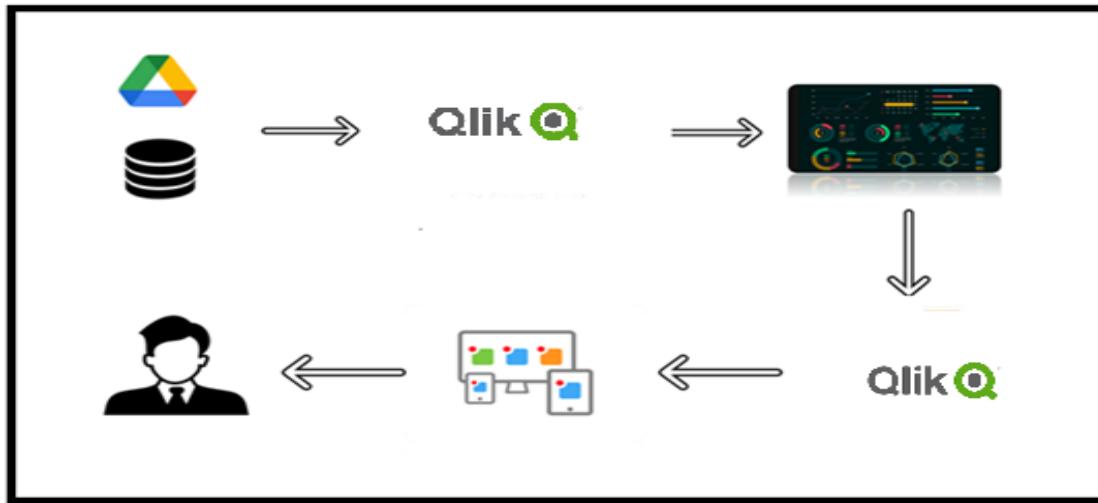
These visualizations help us see patterns and trends quickly, making it much easier to understand the data compared to looking at raw tables. This way, we can make better decisions based on the insights we get from the data.

### 1.2 Purpose:

Predicting trends from past data is crucial in business for several reasons, such as understanding customer interests and maximizing profits. However, analyzing raw data can be challenging. This is where visual analysis comes in handy, as it makes complex data easy to understand for everyone. Visualizations bridge language barriers, allowing anyone to grasp insights without needing to understand the language. By using these visual analyses, businesses can avoid financial risks by not producing items that people don't want, saving costs, and increasing profits. In terms of market trends, we can identify the most needed products and focus on manufacturing those to boost business profits. For example, in this dataset, we can analyze the number of passengers coming from different countries and provide facilities tailored to their needs. We can also make improvements based on gender demographics. Analyzing data on delayed or canceled flights can help improve overall passenger satisfaction.

All these tasks can be efficiently accomplished with proper data visualizations, and Qlik tools are instrumental in making this possible.

### 1.3 Technical Architecture:



## 2. Problem Understanding:

### 2.1 Specifying the Business Problem:

Airlines need to optimize revenue by setting the right ticket prices and identifying busy travel times and popular destinations. Managing schedules, passenger flow, and luggage handling efficiently is crucial for both airlines and airports. Enhancing customer satisfaction by understanding their preferences and addressing pain points is also essential. Additionally, airlines need to understand market demand to avoid overproduction and unnecessary costs. Frequent flight delays and cancellations can significantly impact customer satisfaction and operational efficiency.

Using Qlik tools, airlines and airports can visualize sales data, adjust prices, and identify profitable routes to maximize earnings. Qlik helps spot operational bottlenecks, predict busy periods, and allocate resources effectively to streamline operations. Analyzing customer feedback with Qlik enables personalized services and targeted marketing campaigns, boosting satisfaction and loyalty. Qlik also helps analyze passenger trends to tailor services to different customer segments and visualize delay and cancellation patterns to improve scheduling and reliability. Overall, Qlik transforms complex data into clear visual insights, leading to better decisions, higher profits, and happier customers.

## 2.2 Business Requirements:

To make the most of Qlik tools for analyzing airline data, we need to focus on a few key things. Firstly, we must gather all the necessary data about passengers and flights. Then, we can use Qlik to help us understand things like ticket sales, how efficiently flights are running, and what customers think about their experiences. It's also important to find and fix any problems that slow down operations. By analyzing when people travel most and adjusting prices accordingly, airlines can make more money. Plus, by listening to customer feedback and making changes based on it, airlines can keep passengers happy. Finally, we should always keep improving our tools and methods based on what we learn.

## 2.3 literature survey:

A literature survey involves looking at what others have written about a topic. For our project on analyzing airline data with Qlik tools, we would check studies and articles on how businesses use tools like Qlik in the airline industry. We can also look into different ways data can be shown visually and how airlines predict things like passenger numbers and flight delays. Plus, see how airlines use data to make customers happier and earn more money. It's important to see what's already been done to guide our own research and make sure we're using the best methods.

# 3. Data Collection:

## 3.1 Collecting the Dataset:

We are using a dataset from Kaggle that has synthetic airline data. To get it, you need to sign up on Kaggle, log in, and then click on the dataset link to download it. This data isn't real but mimics what you'd find in the airline industry. It's great for practicing analysis using Qlik tools.

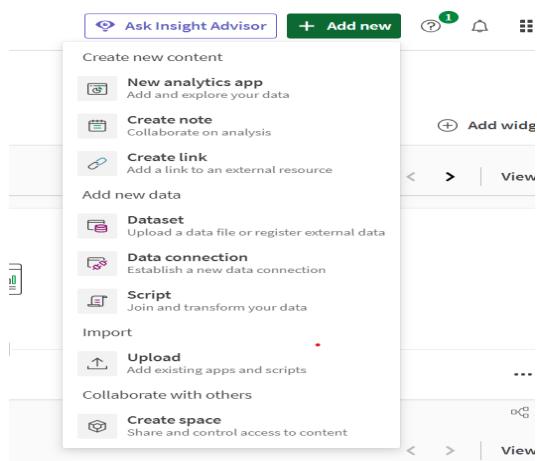
link for dataset: <https://www.kaggle.com/datasets/iamsouravbanerjee/airline-dataset/data>

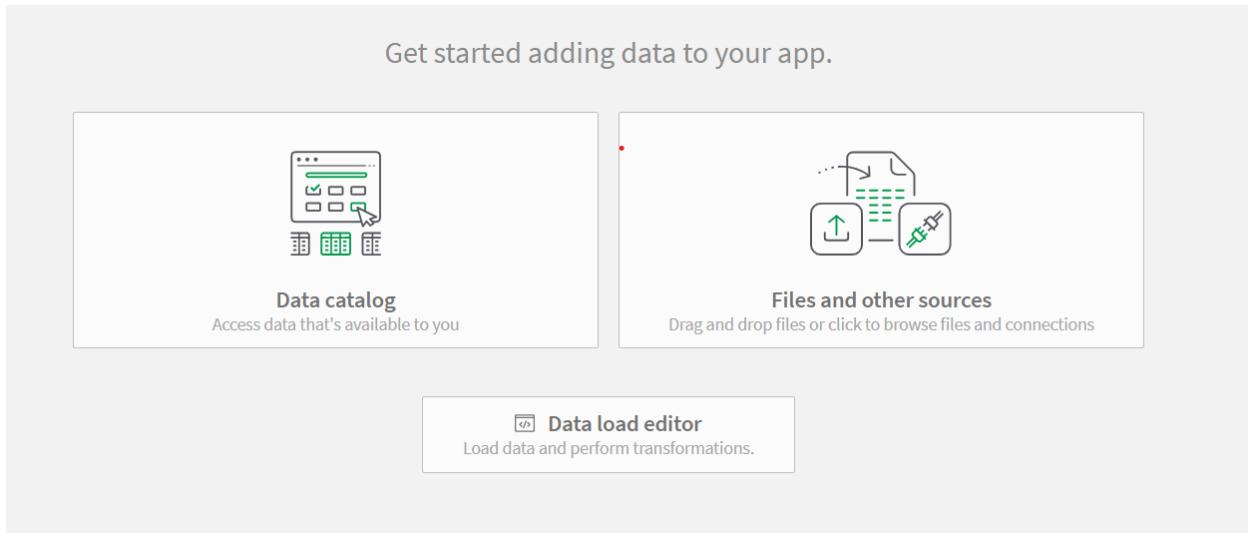
## Available data columns:

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 data to Qlik sense:

To start, log in to the Qlik Cloud website. Then, locate the "Add New" option and click it to create a new app. Once the app is created, you can add data by browsing your files and selecting the dataset you want to use.

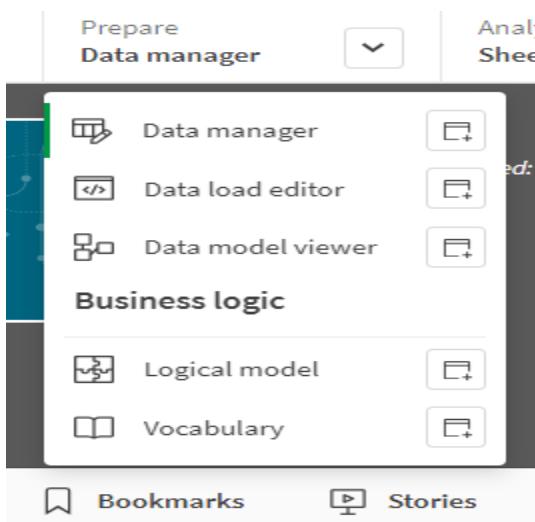




## 4. Data Preparation:

### 4.1 Preparation of the data for visualization:

To get the data ready for visualization, we clean it up by removing anything we don't need and making sure nothing is missing. Then, we put it in a format that makes it easy to show visually. After that, we look at the data to see if there are any patterns or trends we can spot. If we want to focus on specific parts of the data, we can filter it. Once everything is completed, we make sure the data is accurate and complete.



we can add additional script in data load editor.

If there is need of association of tables we can take care of it automatically by using qlik association feature.

Once everything done we click on load data.

## 5. Data Visualizations:

### 5.1 Visualization:

Data visualization is all about turning data into pictures that are easy to understand. Instead of looking at rows of numbers, we can use charts, graphs, and maps to see trends and patterns. This makes it much easier to spot important information and make decisions based on it.

Total Number of Passengers

**98.62k**

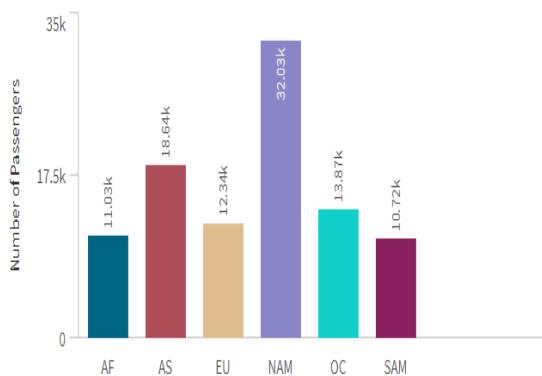
Number of Passengers affected by delayed flights

**32.83k**

Number of Passengers affected by cancelled flights

**32.94k**

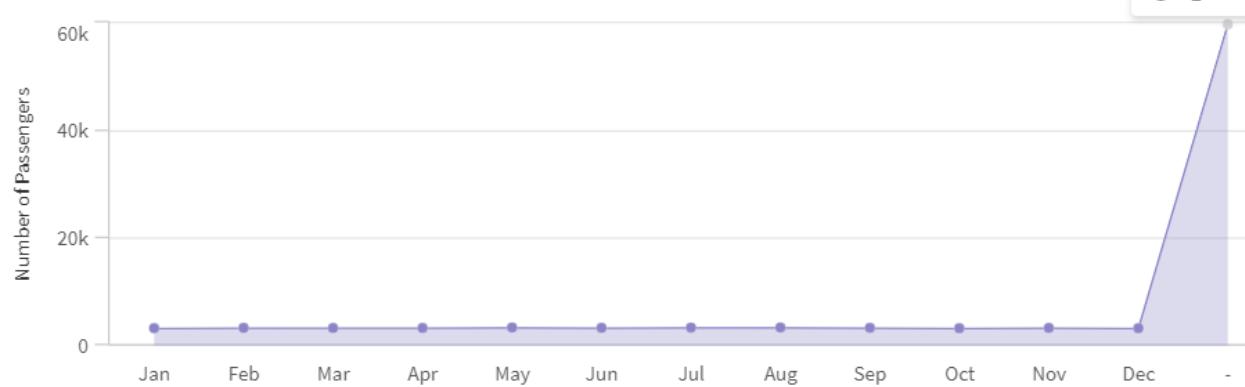
Number of Passengers (Continent Wise)



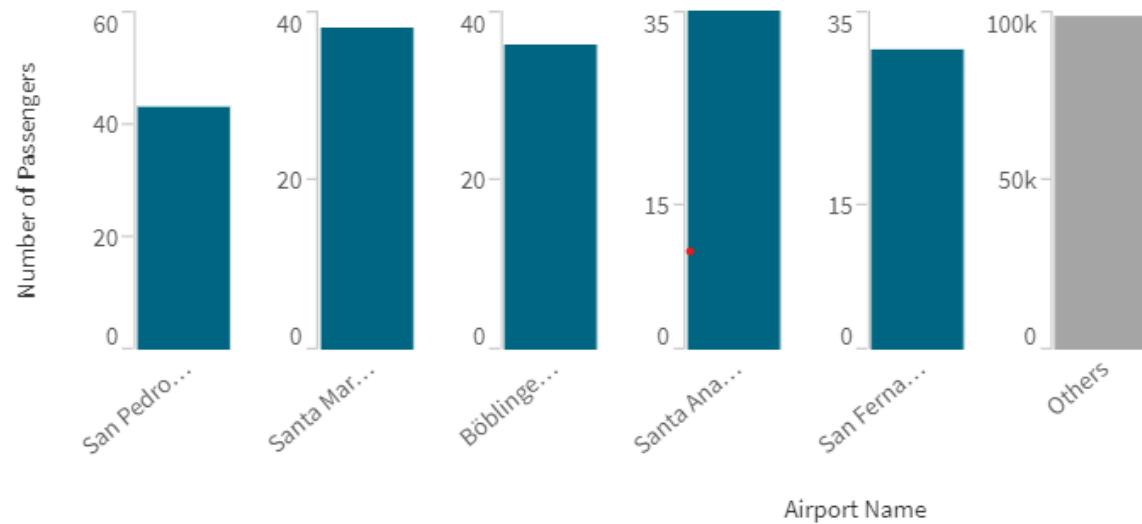
Average Age Group

**45.5**

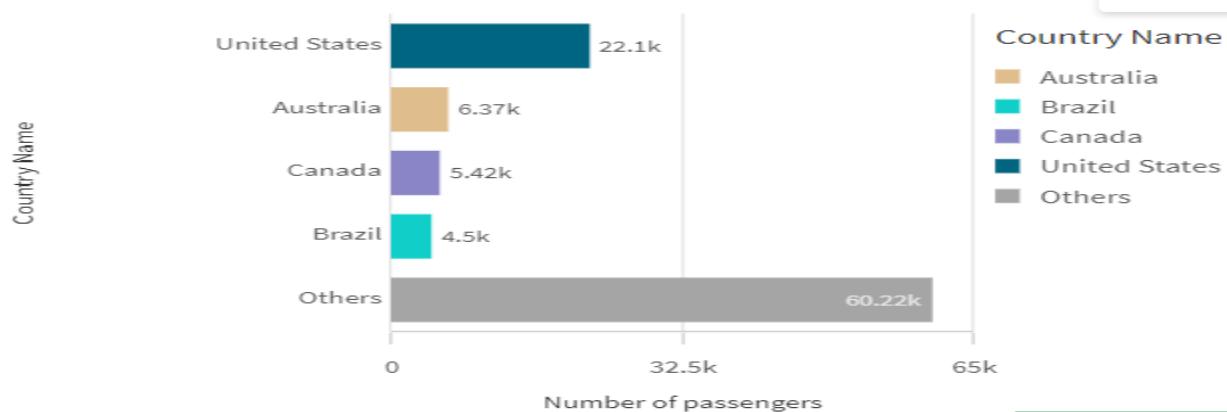
### Number of Passengers Travelled (Month Wise)



### Number of Passengers ( Airport Wise)



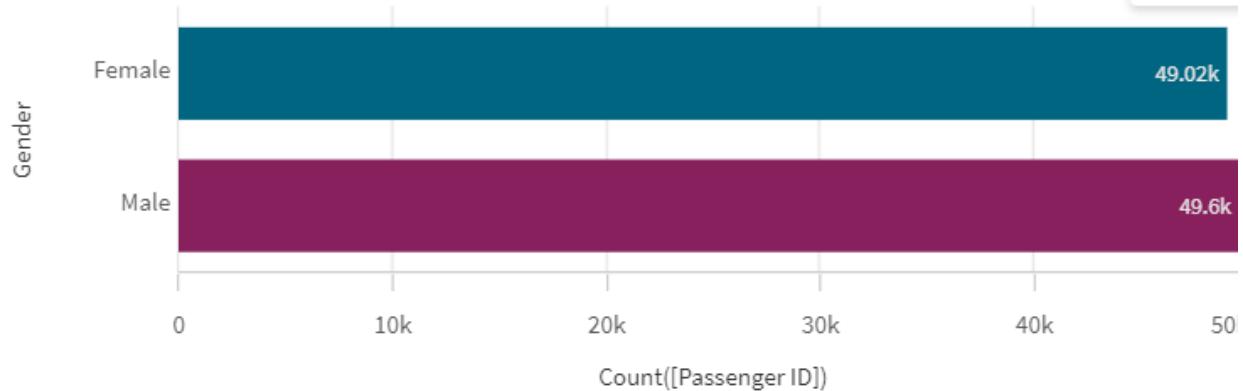
### Number of Passengers(Country Wise)



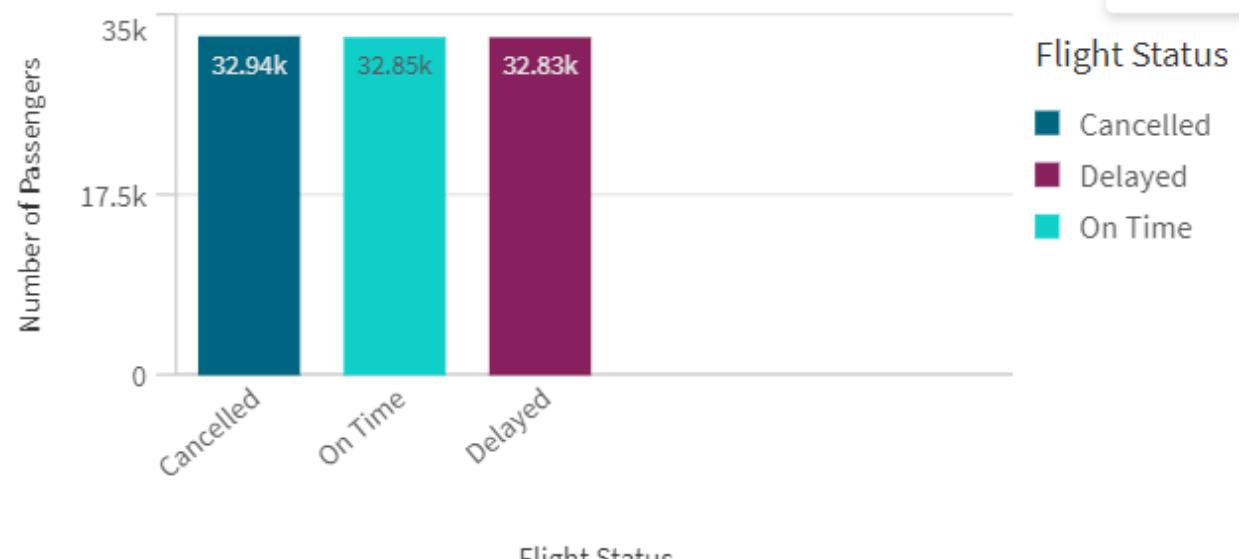
# Number of Airports

9.06k

## Number of Passengers(Gender Wise)



## Number of Passengers (Flight Status Wise)



## Flight Status (Continent Wise)



we can create these charts, fields etc using qlik tools by adding few commands.

we can change the appearance using the available options.

## 6. Dashboard:

Dashboards are like a control center where you can see all your important information at a glance. They're designed to be easy to read and organized so you can quickly understand what's going on.

### 6.1 Responsive and Designing of Dashboard:

we can use the above visualization and create a responsive and understandable dashboard.





## 7. Report:

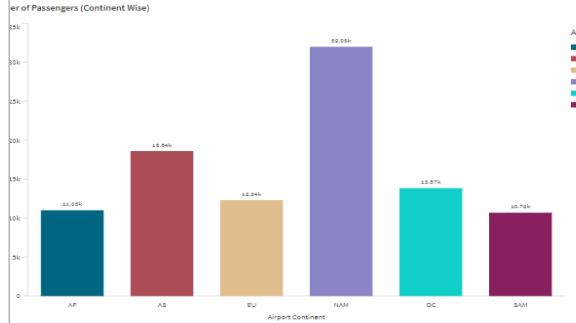
### 7.1 Creation of Report:

A data story is like telling a story with data. It's a way of presenting information in a more engaging and understandable way.

Total Number of Passengers  
**98.62k**

Number of Passengers affected by delayed flights  
**32.83k**

Number of Passengers affected by cancelled flights  
**32.94k**



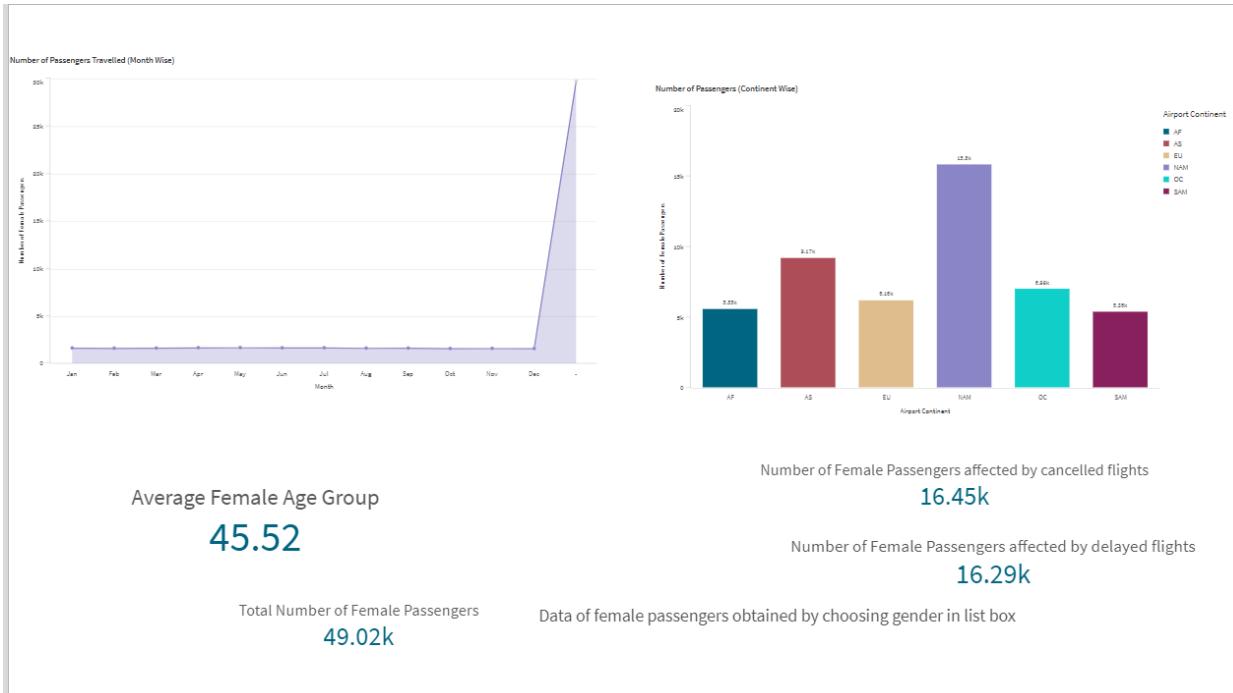
A bar graph representing number of Passengers continent wise

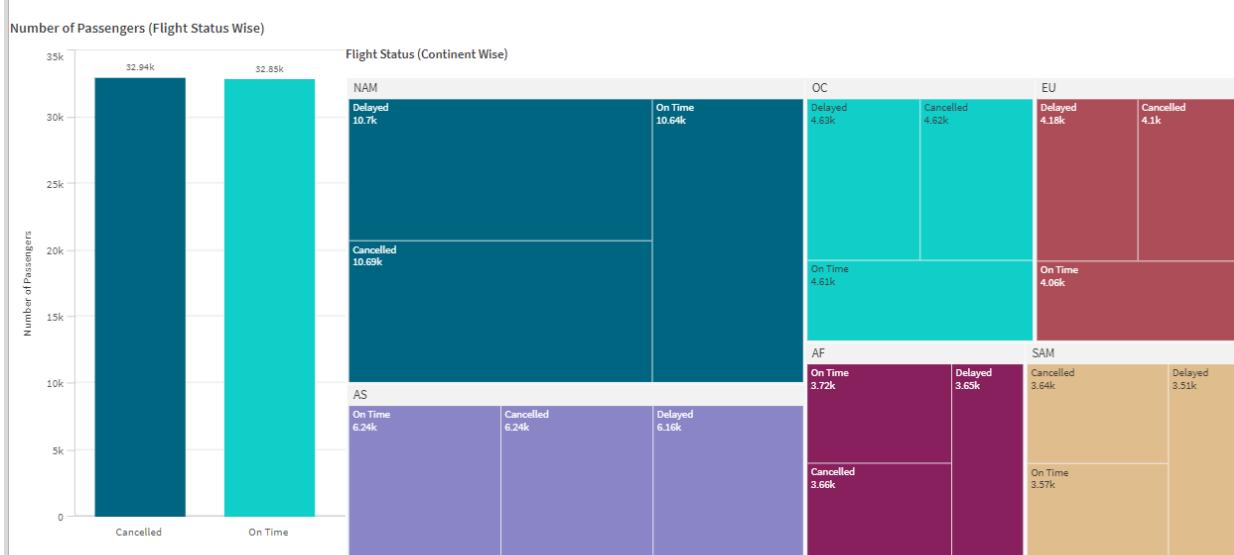
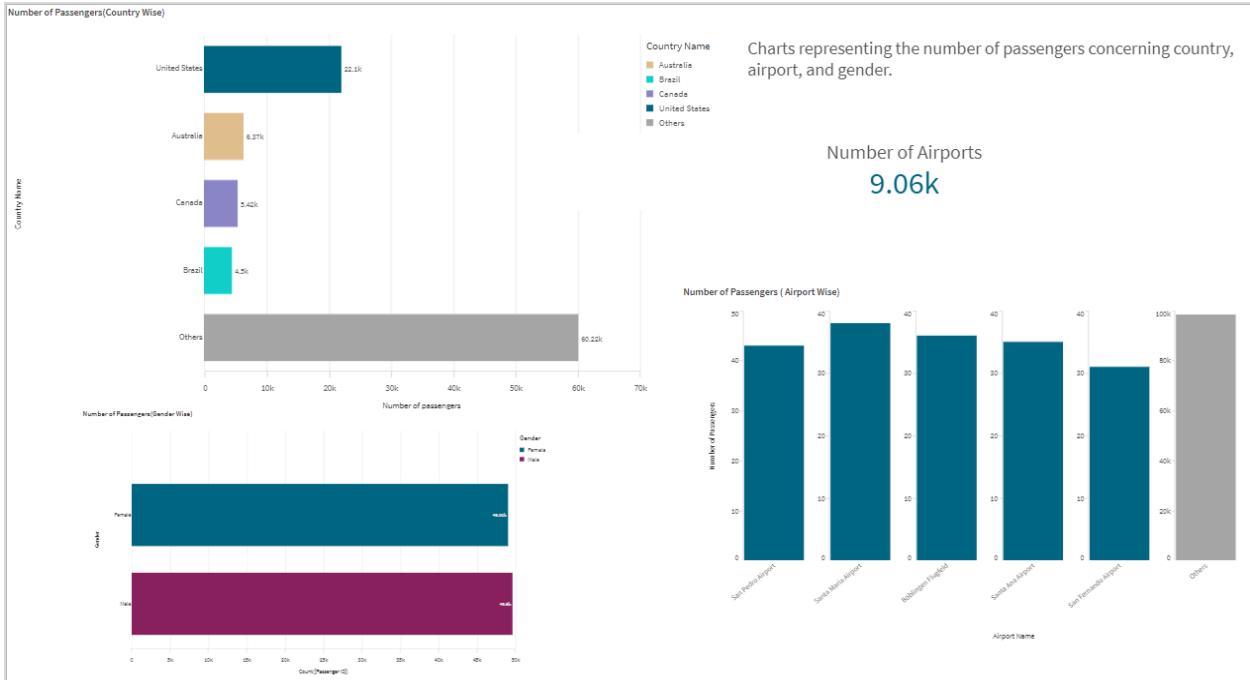
Average Age Group  
**45.5**

Number of Passengers Travelled (Month Wise)

Month	Number of Passengers
Jan	~10k
Feb	~10k
Mar	~10k
Apr	~10k
May	~10k
Jun	~10k
Jul	~10k
Aug	~10k
Sep	~10k
Oct	~45k
Nov	~50k
Dec	~55k

Number of passengers Travelled month wise (line chart)





Charts representing the number of passengers flight status wise and number passengers with the flight status with respect to the continent

## 8. Performance Testing:

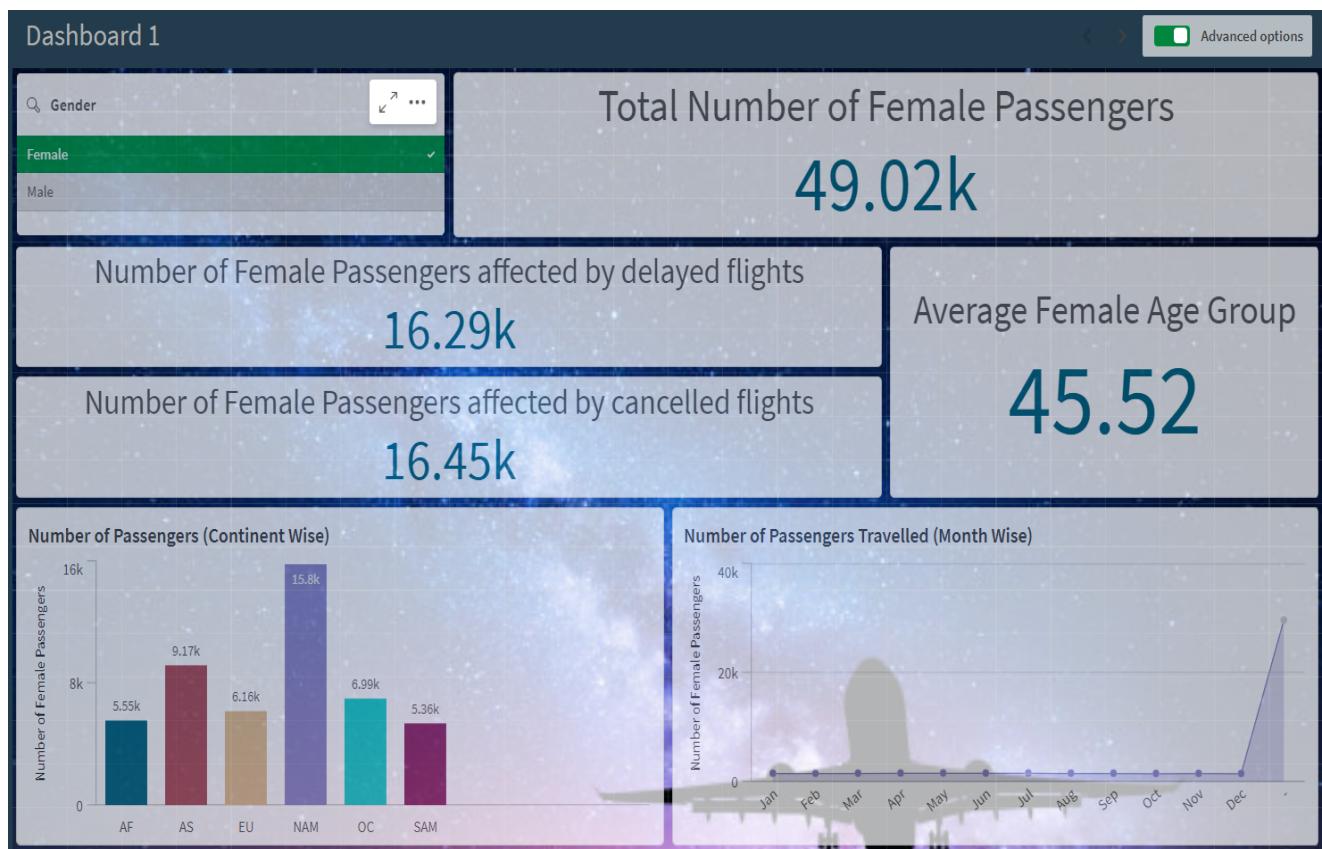
### 8.1 Amount of data rendered:

"Amount of Data Loaded" simply means the quantity or volume of data that has been brought into a system, software, or database for processing or storage. It's a measure of how much data has been successfully imported or retrieved and is now available for analysis or manipulation within the system.

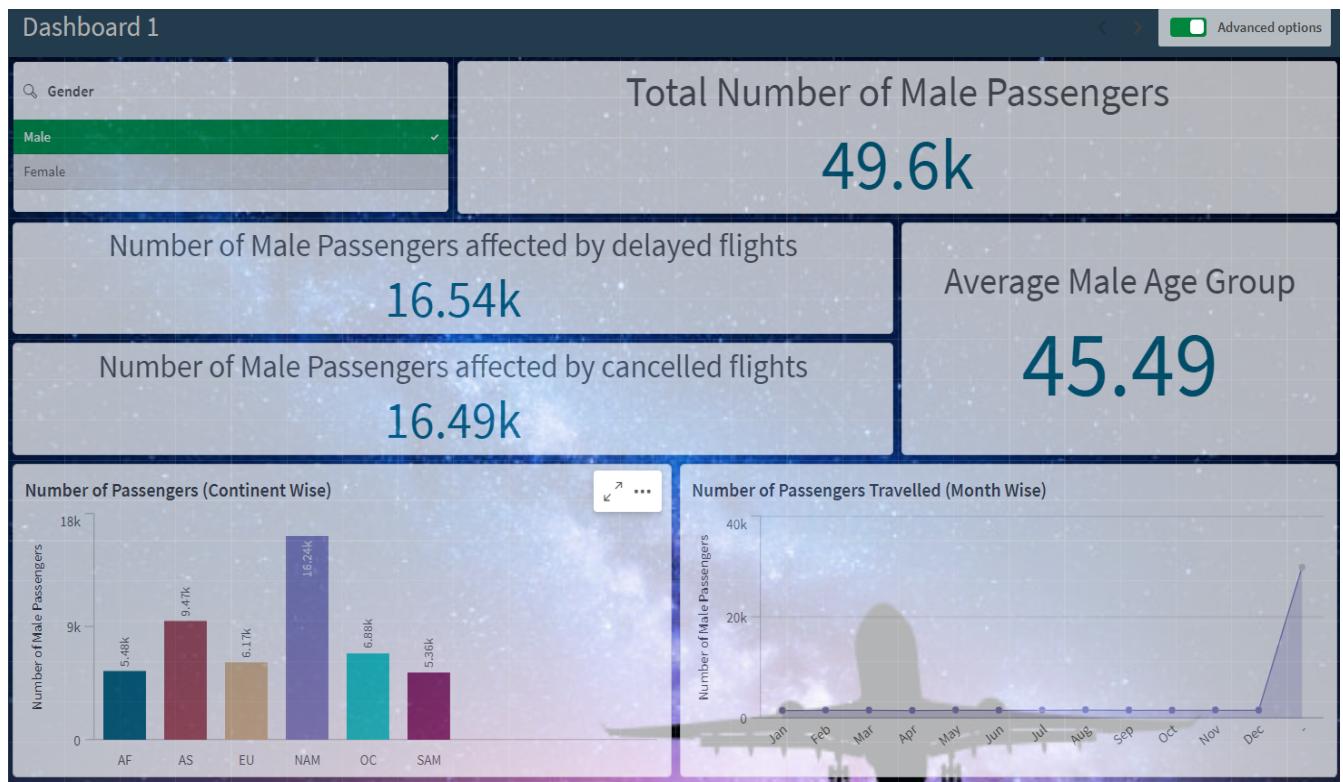
Airline Dataset Updated - v2
Passenger ID
First Name
Last Name
Gender
Age
Nationality
Airport Name
Airport Country Code
Country Name
Airport Continent
Continents
Departure Date
Arrival Airport
Pilot Name
Flight Status
Airline Dataset Updated - v2.Nationality_GeoInfo
Airline Dataset Updated - v2.Airport Country Code_GeoInfo
Airline Dataset Updated - v2.Country Name_GeoInfo

## 8.2 Utilization of data filters:

"Utilization of Filters" refers to using filters within a system or software to selectively extract, manipulate, or analyze data based on specific criteria. Filters help narrow down the data, focusing only on relevant information that meets predefined conditions. They are essential for refining data and ensuring that only the most relevant information is considered during analysis or processing.



Dashboard 1 with filter gender is female



Dashboard 1 with filter gender is male.