



---

Sense™

**Project Title:**

**Exploring Insights from  
Synthetic Airline Data  
Analysis With Qlik**

**Name:B Srikanth**

**College:Vidya Jyothi Institute of Technology**

## PROBLEM STATEMENT:

# Exploring Insights from Synthetic Airline Data Analysis with Qlik



## INTRODUCTION

- Airline data holds immense importance as it offers insights into the functioning and efficiency of the aviation industry. Aims to identify patterns in delays, cancellations, and
- on-time performance to develop strategies that optimize flight schedules, minimize disruptions, and enhance the overall travel experience for passengers.
- The objective is to use Qlik's capabilities to identify patterns, trends, and correlations within the data, aiding decision-making for airlines, airports, and related stakeholders.

## PROJECT FLOW

### DEFINING PROBLEM/PROBLEM UNDERSTANDING

This project provides insights for optimizing flight routes, enhancing on-time performance, and improving overall operational efficiency within the aviation industry.

#### Social Impact:

- By analyzing passenger demographics and preferences, airlines can tailor services to better meet customer needs, enhancing overall satisfaction and loyalty.

#### Business Impact:

Airlines can optimize flight routes, schedules, and resources based on data insights, reducing costs and improving overall efficiency.

### DATA COLLECTION & EXTRACTION:

Data collection is the process of gathering and measuring information from various sources to get a comprehensive and accurate view of a specific domain or subject.

#### Steps for Data Collection

##### 1. Accessing Kaggle

- Visit the Kaggle website and navigate to the provided dataset link in the project documentation.
- Create a Kaggle account or sign in if you already have one.

##### 2. Downloading the Dataset

- Download the dataset by clicking on the download button, typically provided in a compressed format such as ZIP.
- Save the downloaded file to your PC.

##### 3. Extracting the Dataset

- Locate the downloaded compressed file and extract its contents.
- Store the extracted files in a designated project directory for easy access.

### Data Collection And Extraction-[Link](#)

### DATA PREPARATION:

#### Data Loading:

Steps for Loading Data into Qlik Cloud:

##### 1. Setting Up Qlik Cloud Environment

- **Access Qlik Cloud:** Log in to your Qlik Cloud account using your credentials.
- **Create a New App:** Navigate to the Qlik Cloud hub and click on "Create new app".

Provide a name for your app and click "Create".

## 2. Uploading the Dataset

- **Open the App:** Click on the newly created app to open it.
- **Upload Data:** Click on "Add data" from the main screen of the app. Choose "My files" to upload data from your local machine.
- **Select the File:** Browse to the location of your extracted dataset (e.g., CSV or Excel file) on your PC, select the file, and click "Open".
- **Confirm Upload:** Once the file is uploaded, it will appear in the atmanager. Confirm that the data has been uploaded correctly by previewing it.

## Data Cleaning and Pre-Processing :

Data cleaning and preprocessing are crucial steps to ensure the dataset is accurate, consistent, and ready for analysis. Here's a detailed explanation on how to clean and preprocess the airline dataset in Qlik Cloud.

### Data Cleaning Using Set Nulls Function

- **Navigate to Data Manager:**
  - Open the Data Manager from the Qlik Cloud interface.
- **Set Nulls:**
  - Select the dataset and locate fields with null values.
  - Click on the "Set Nulls" option.
  - Specify how you want to handle null values. You can choose to replace null values with a default value, remove rows with null values.
  - Apply the changes and review the dataset to ensure null values are handled correctly.

### Correct Data Types

- **Ensure Correct Data Types:**
  - Verify that all fields have the correct data types. Adjust data types as necessary within the Data Manager (e.g., dates should be in date format, numeric fields in number format).

### Remove Duplicates:

- Identify and remove duplicate records using the "Remove duplicates" option in the data preparation interface.

### Data Transformation

- **Create New Fields:**
  - Create new fields based on existing data. For example, extract the year, month, or

- day from a date field or calculate flight durations.
- Use the "Add calculated field" option in the Data Manager to create these new fields.
- **Rename Fields:**
  - Rename fields to be more descriptive and meaningful for analysis directly in the Data Manager.
- **Combine Fields:**
  - Combine multiple fields into a single field if necessary (e.g., combine first and last names into a full name field).

### **Data Cleaning And Prep-Processing-[Link](#)**

## **Data Visualization:**

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

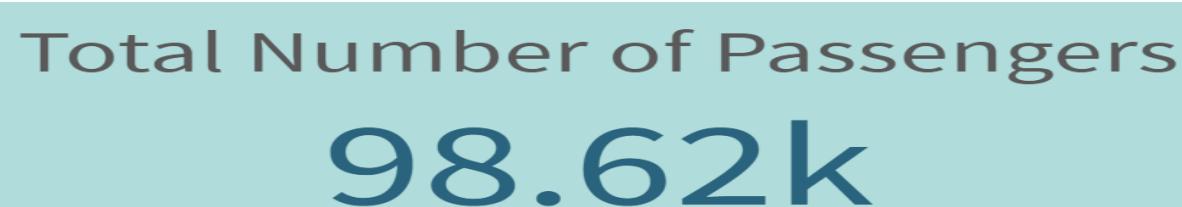
Steps for Data Visualization:

1. **Selecting Data:** Choose the dataset and fields you want to visualize.
2. **Drag-and-Drop:** Drag the fields from the data pane onto the canvas.
3. **Choose Visualization Type:** Select the appropriate chart type from the chart menu.
4. **Customize:** Customize the visualization by adjusting settings such as colors, labels, and axis scales.
5. **Interact:** Interact with the visualization by selecting data points, filtering, and drilling down for deeper insights.
6. **Create Dashboards:** Combine multiple visualizations into dashboards to tell a coherent story or present key insights.

### **Examples:**

#### **KPI(Key Performance Indicator):**

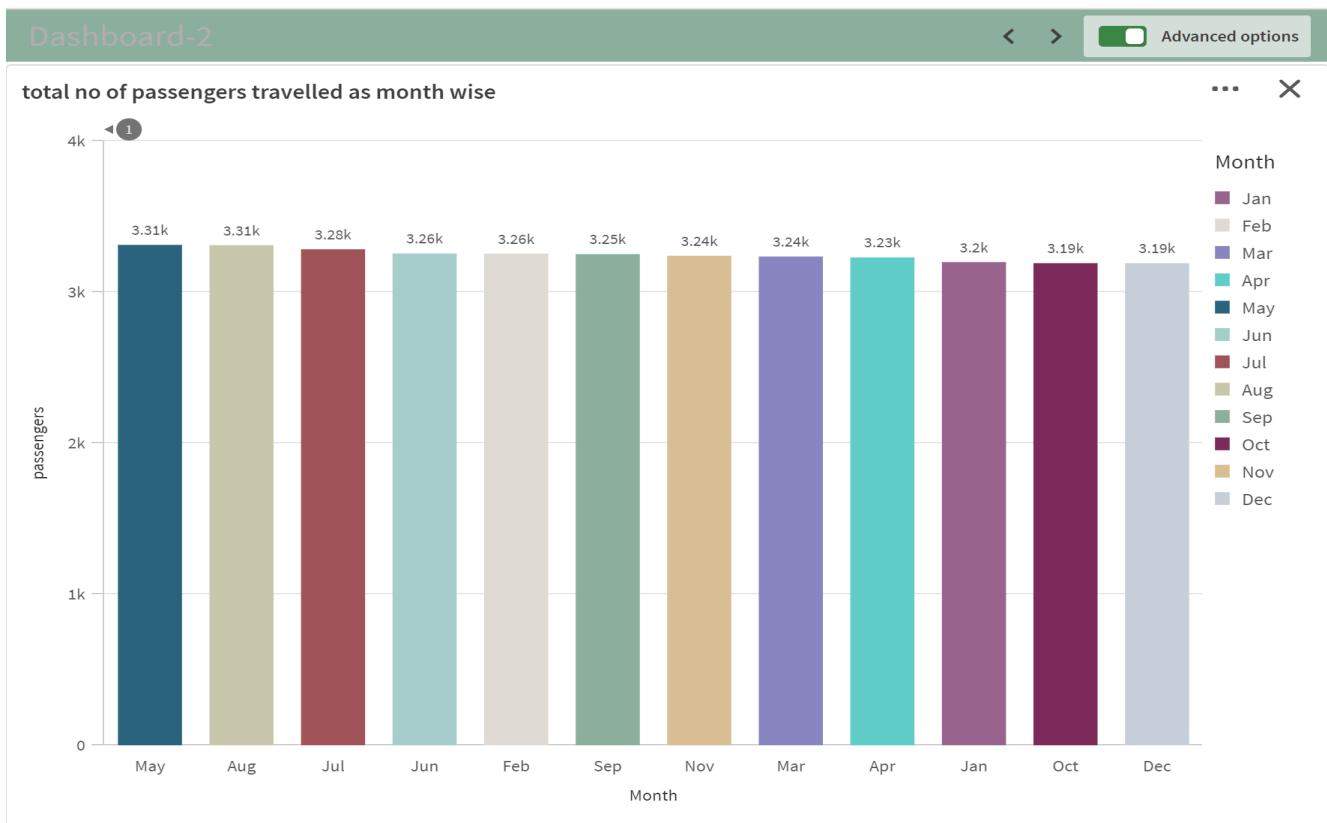
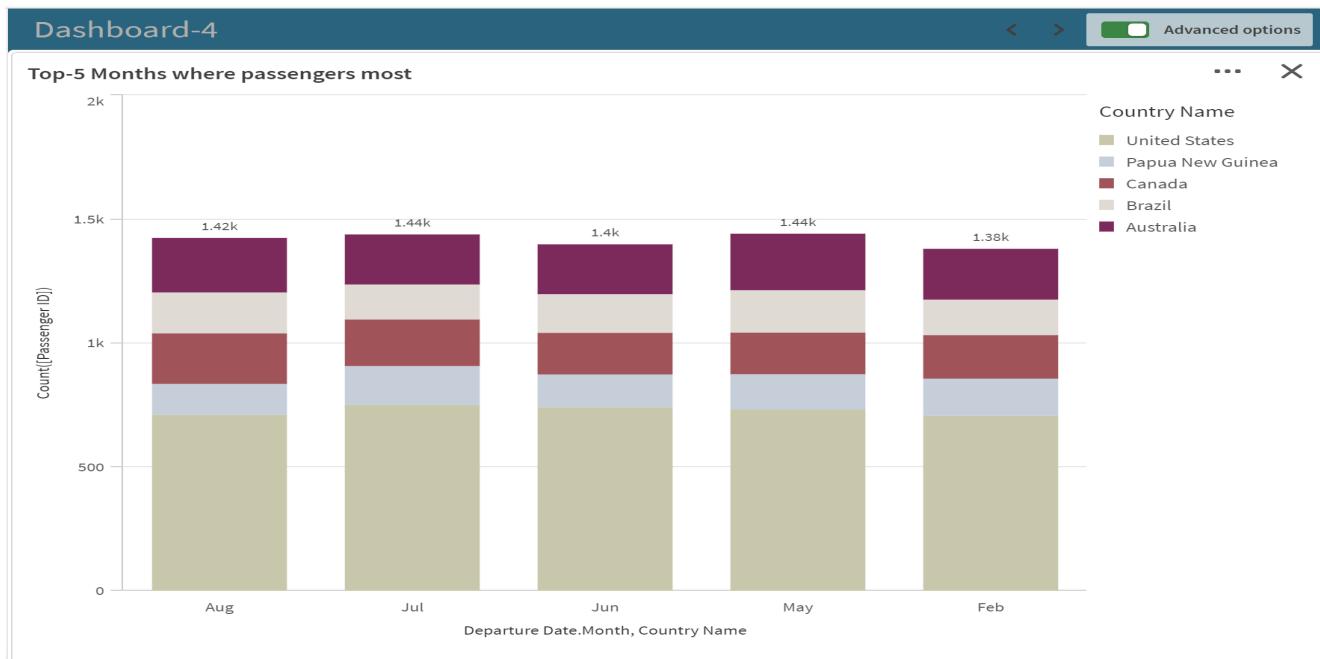
Key Performance Indicator (KPI) is a quantifiable metric used to evaluate the performance of an organization, team, process, or individual against predefined goals or objectives.



## Bar Charts:

- Bar charts visually represent data using rectangular bars, where the length or height of each bar corresponds to the value of a variable.
- They are effective for comparing values across different categories or groups, allowing viewers to quickly identify patterns and trends in the data.

## Examples:



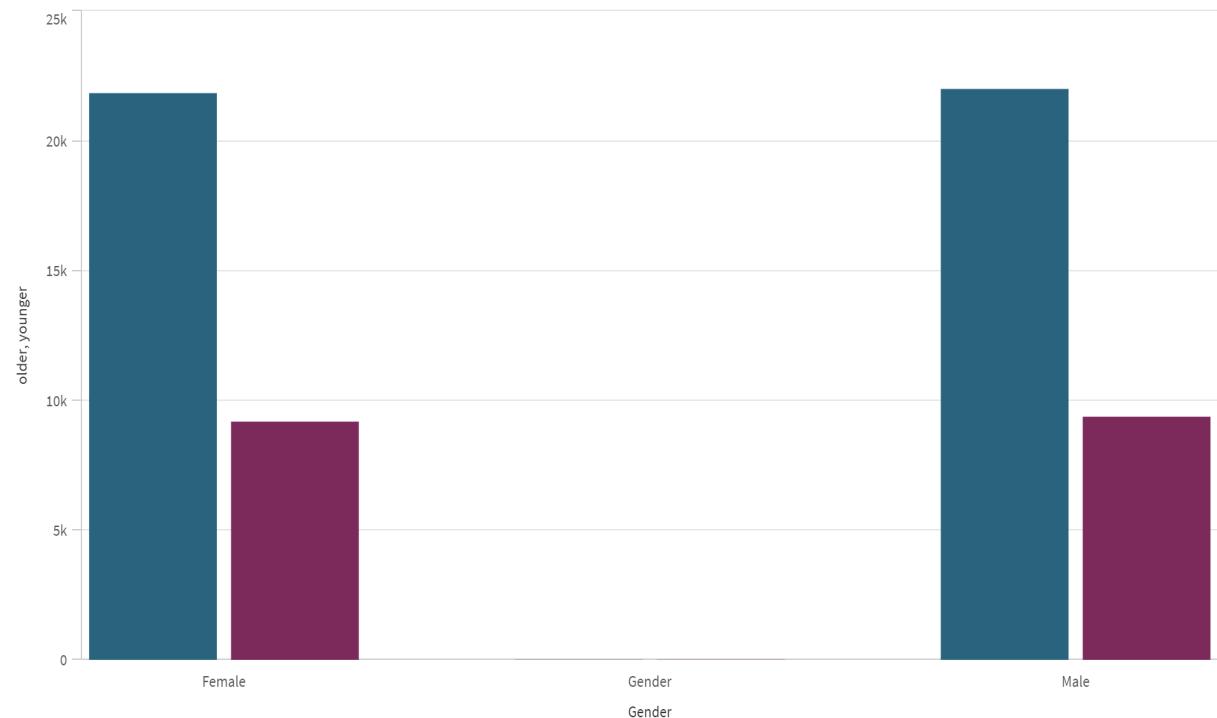
## Dashboard-3

< >

Advanced options

... X

### gender wise passengers



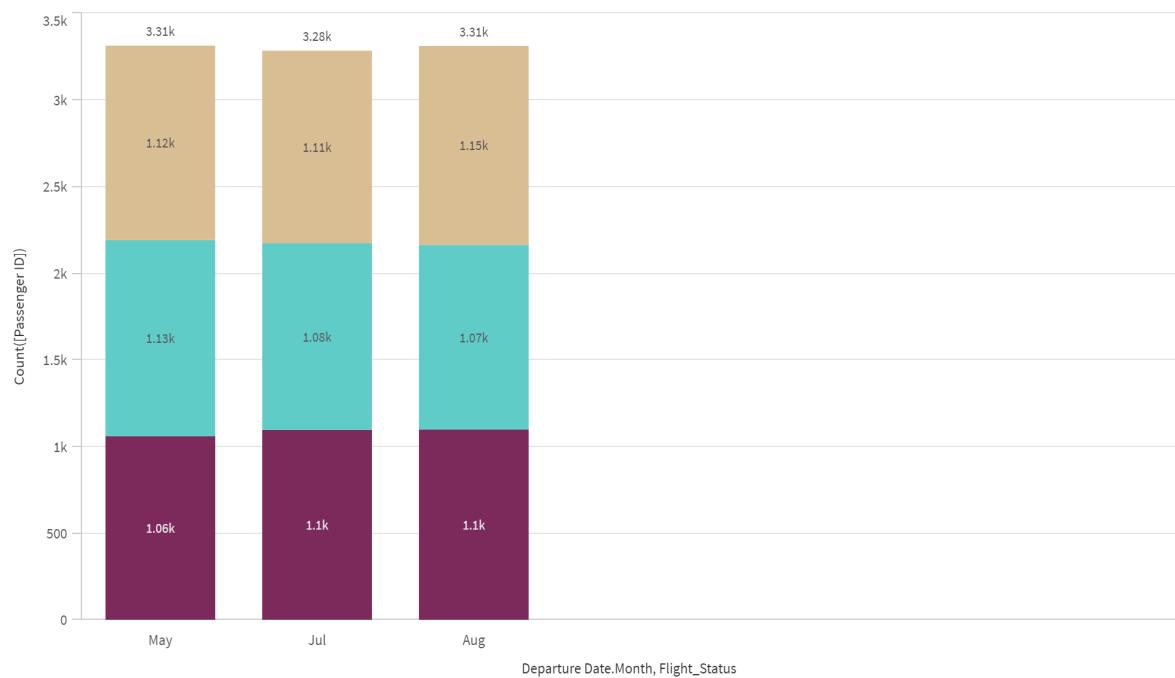
## Dashboard-4

< >

Advanced options

... X

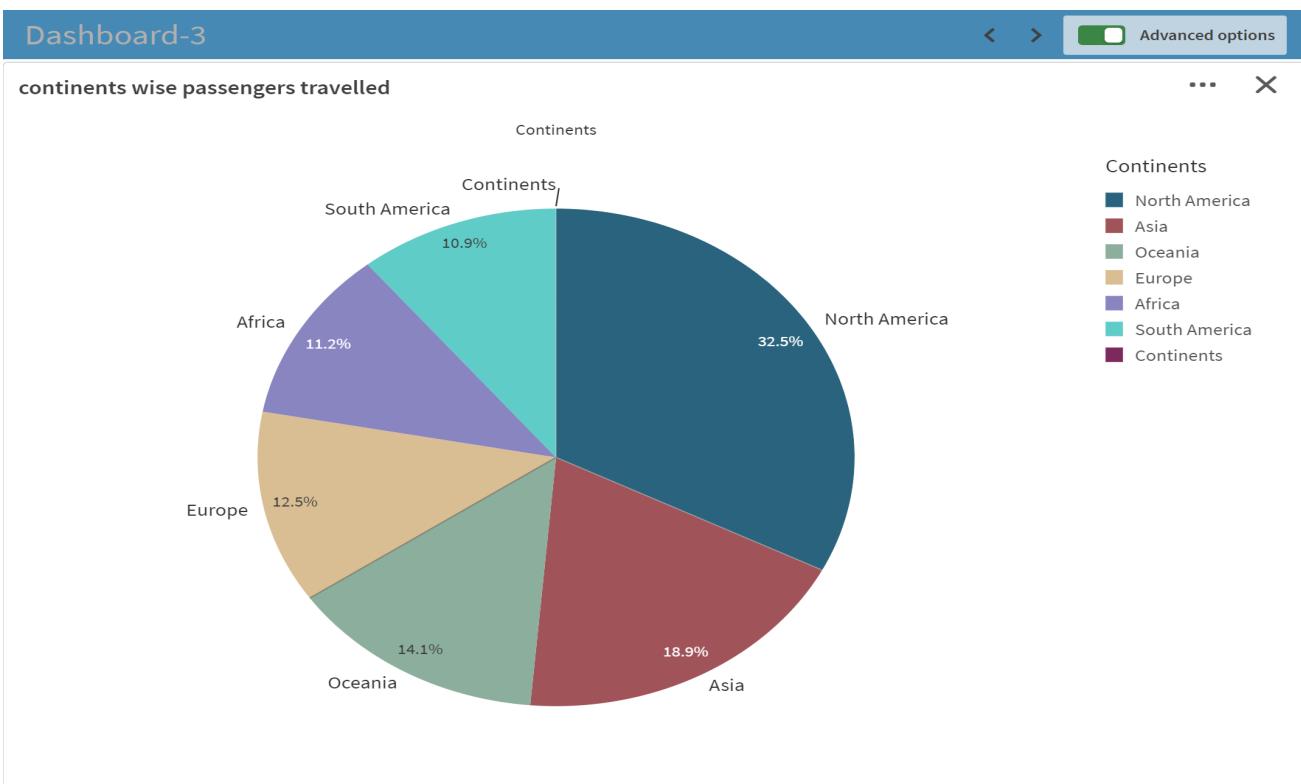
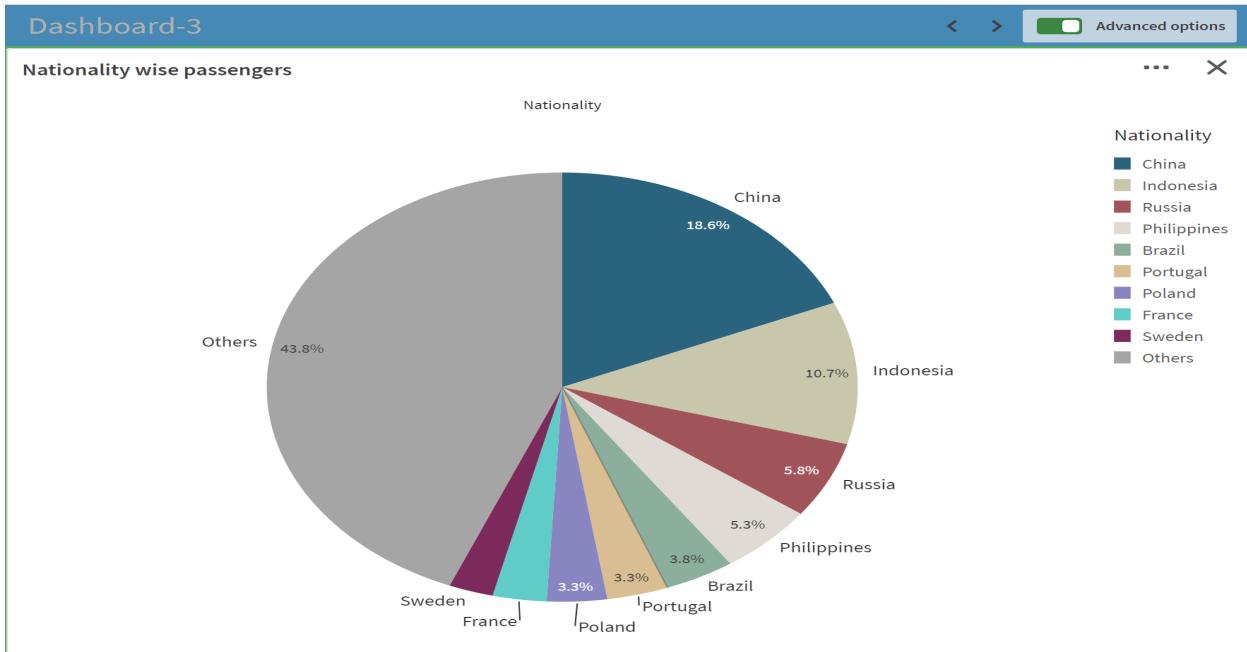
### Top-3 months Flight status



## Pie Charts:

Pie charts are circular statistical graphics that use slices to represent proportional parts of a whole data set, where each slice's size is proportional to the quantity it represents. They provide a visual representation of the composition or distribution of a categorical data set.

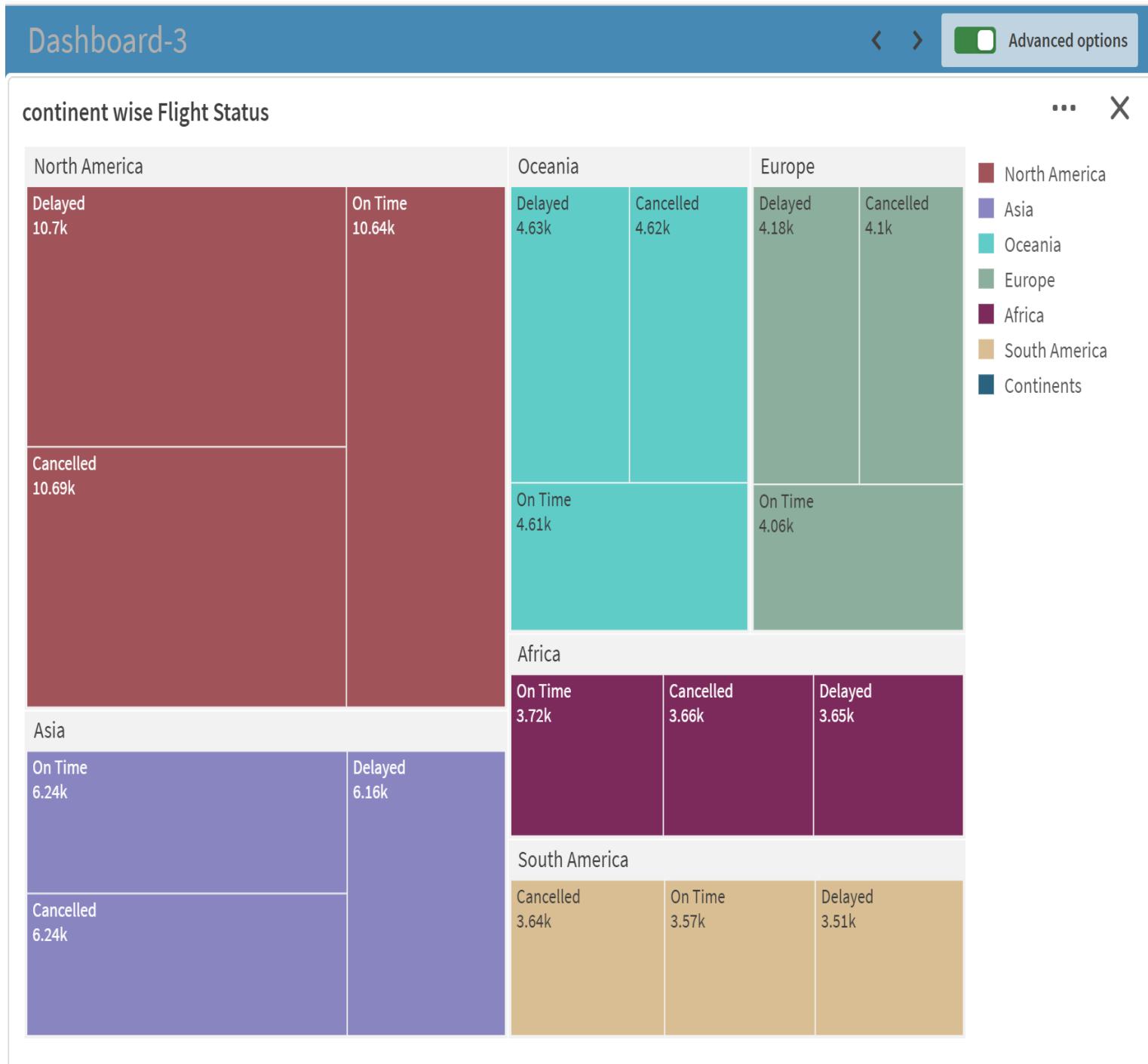
### Examples:



## TreeMap:

A treemap is a data visualization technique that displays hierarchical data using nested rectangles. Each branch of the hierarchy is represented by a rectangle, which is then subdivided into smaller rectangles representing sub-branches. The size of each rectangle is proportional to a specific data dimension, such as value or volume, making it easy to compare proportions within the hierarchy.

### Examples:



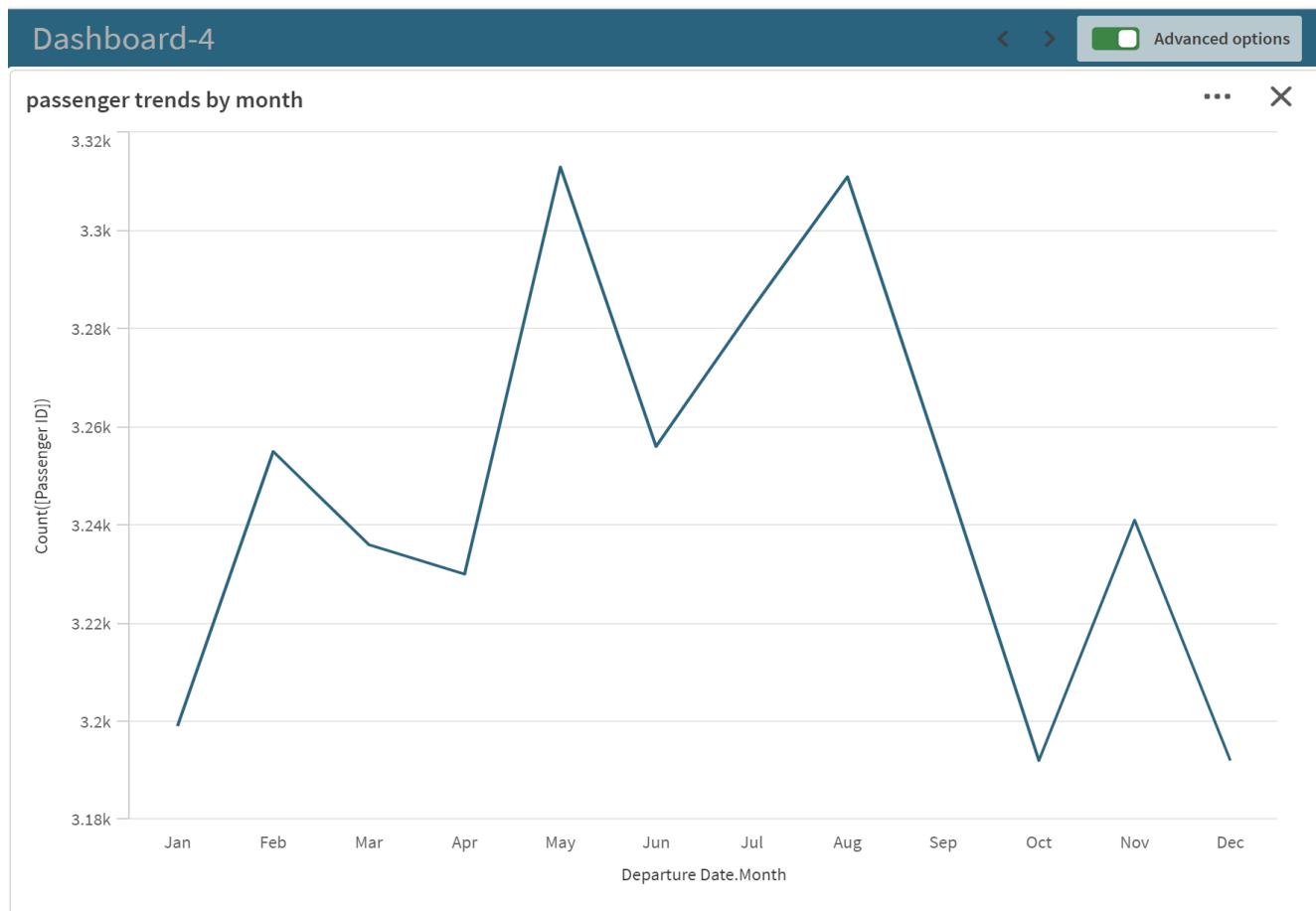
## Line Charts:

Line charts are graphical representations used to display data points connected by straight lines, illustrating trends or changes over time. They are particularly effective for showing the relationship between two variables and how one variable changes in response to another.

### Key Features of Line Charts:

- Trends Over Time:** Line charts are ideal for visualizing trends, patterns, and fluctuations in data over continuous intervals, such as time (e.g., days, months, years).
- Comparative Analysis:** Multiple lines can be plotted on the same chart to compare different datasets or variables, making it easy to observe correlations and differences.

### Examples:



## Table:

- Table offers a clear and organized way to display information. Each row represents a data point or record, and each column represents a specific variable or attribute.
- This format allows for easy comparison between different data points and makes it suitable for presenting precise numerical information.

The screenshot shows a dashboard titled "flight status list according to country". The table has four columns: "Country Name", "Count(Delayed)", "Count(Cancelled)", and "Count(OnTime)". The data includes a total row and numerous country entries. The table is styled with a green header bar and a light gray background for the data rows. A search icon is located in the top right of the table area.

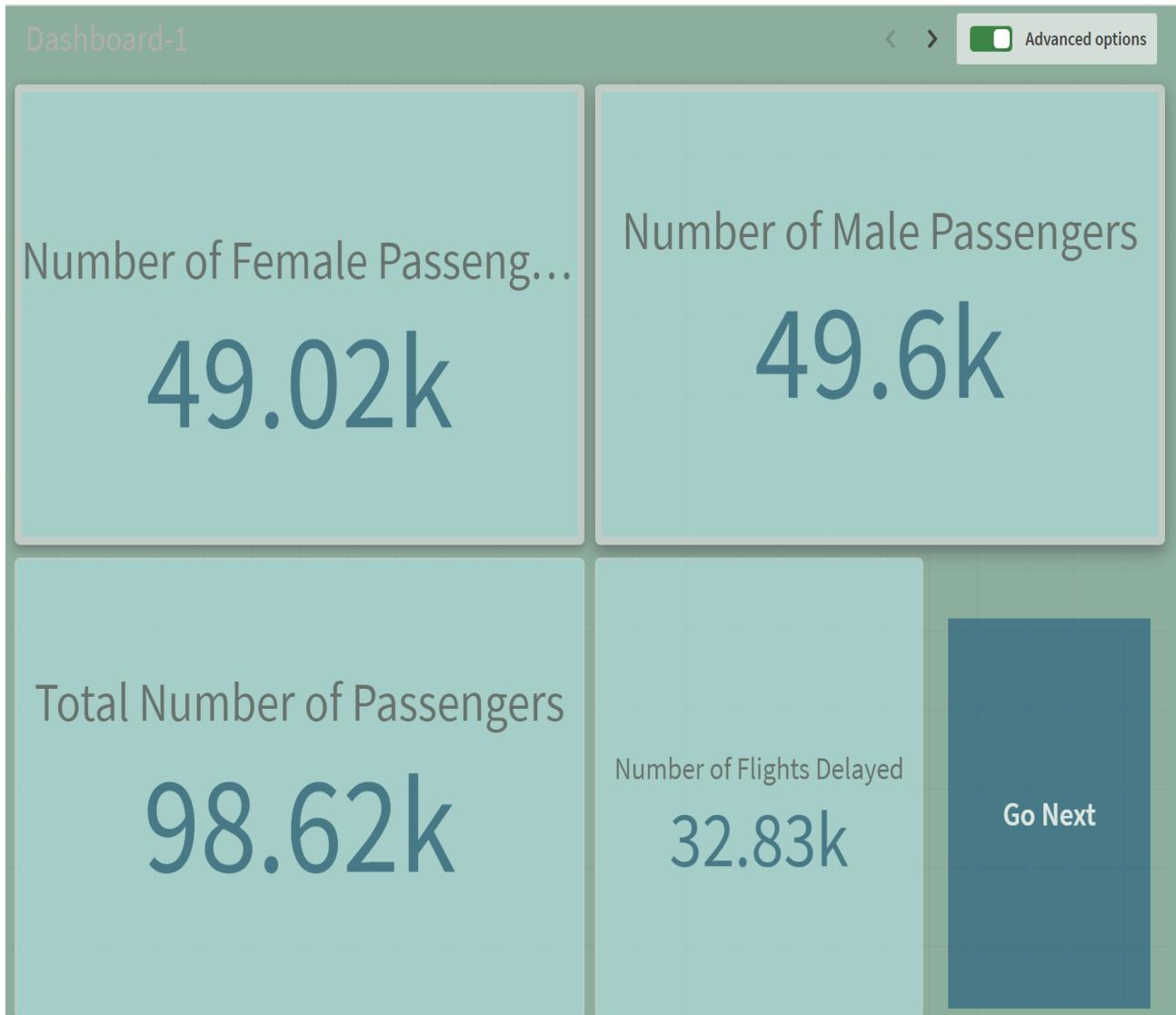
Country Name	Count(Delayed)	Count(Cancelled)	Count(OnTime)
Totals	32831	32942	32846
Afghanistan	121	137	114
Albania	5	5	2
Algeria	151	151	147
American Samoa	10	21	14
Andorra	1	4	2
Angola	154	142	149
Anguilla	4	7	3
Antigua and Barbuda	7	7	6
Argentina	427	394	382
Armenia	6	7	8
Aruba	2	4	3
Australia	2142	2131	2097
Austria	27	37	27
Azerbaijan	25	30	32
Bahamas	134	134	134
Bahrain	3	5	4
Bangladesh	55	68	57
Barbados	3	7	3
Belarus	26	20	21
Belgium	20	29	23
Belize	68	78	72
Benin	11	16	24

[Data Visualization-Link](#)

## Dashboards:

Dashboards are powerful tools used to present data visualizations and insights in a user-friendly format. They essentially act as the final layer, summarizing key findings and metrics from your data analysis.

### Dashboard-1



## Dashboard-2



## Dashboard-3



## Dashboard-4:



## Dashboard-5:



## Story:

In Qlik, the concept of "story" refers to a specific feature designed for data storytelling. It's not about creating stories within the software itself, but rather using Qlik to present your data analysis findings in a narrative format.



### Airline Data Analysis Story



Total Number of Passengers  
**98.62k**

Number of Female Passengers  
**49.02k**

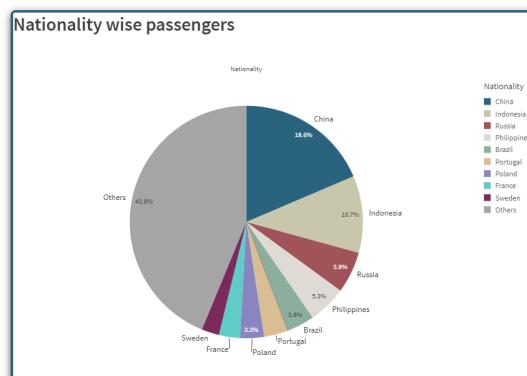
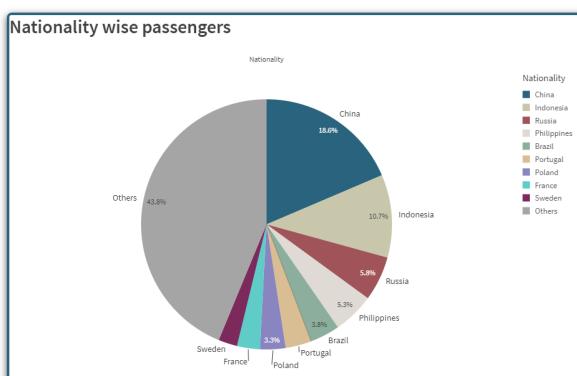
Number of Male Passengers  
**49.6k**

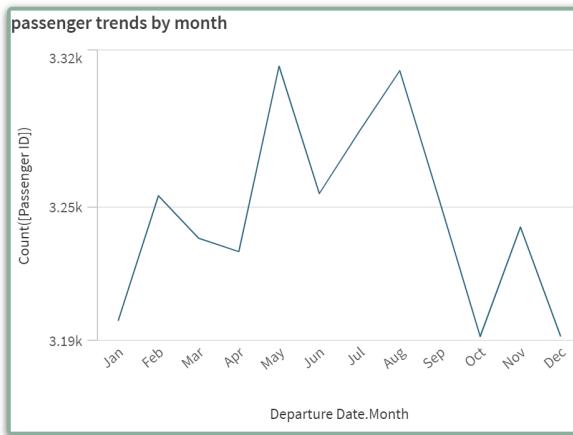
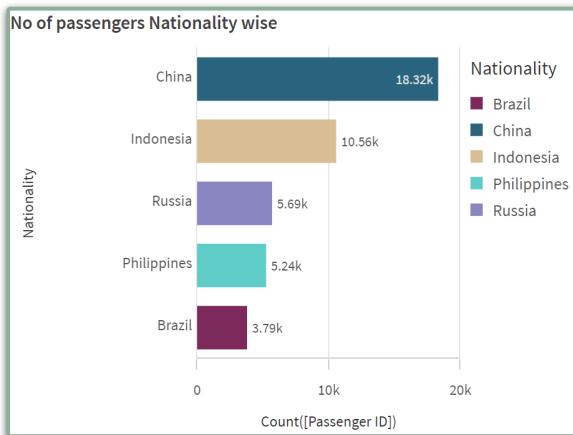
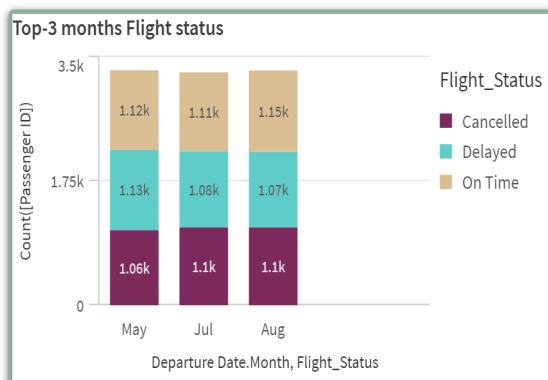
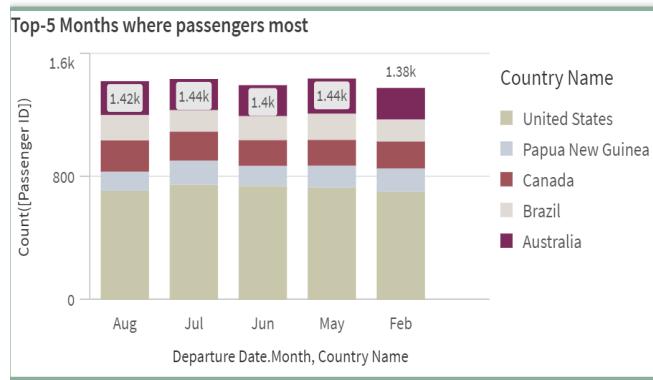
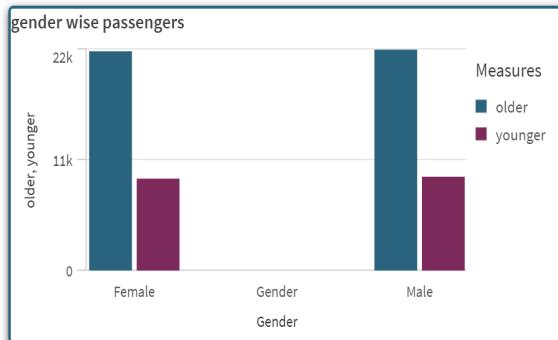
Number of Flights Delayed  
**32.83k**

No of flights Cancelled  
**32.94k**

No of flights on time  
**32.85k**

### pie charts





## Treemap:

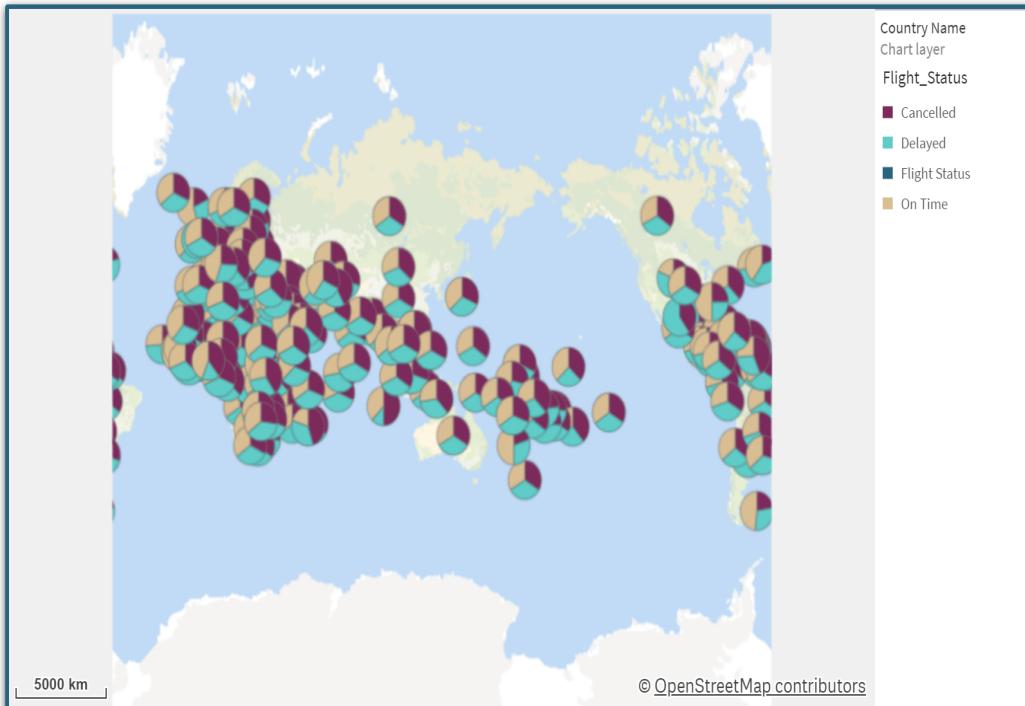


## Table

flight status list according to country

Country Name	Q	Count(Delayed)	Count(Cancelled)	Count(ontime)
Totals		32831	32942	32846
Afghanistan		121	137	114
Albania		5	5	2
Algeria		151	151	147
American Samoa		10	21	14
Andorra		1	4	2
Angola		154	142	149
Anguilla		4	7	3
Antigua and Barbuda		7	7	6
Argentina		427	394	382
Armenia		6	7	8
Aruba		2	4	3
Australia		2142	2131	2097
Austria		27	37	27
Azerbaijan		25	30	32
Bahamas		134	134	134
Bahrain		3	5	4

Map:



**Bookmarks:**

young passengers-[click Here](#)

No Of Old Passengers-[click Here](#)

No of Delayed Flights-[click Here](#)

country wise passengers list-[click Here](#)