Exploring Insights from Synthetic Airline Data Analysis with Qlik

---

Define Problem / Problem Understanding :

➤ **Specify the business problem:**

Analyze the demographic and travel patterns of passengers to identify key trends and insights. Specifically, investigate the relationship between passenger demographics (such as age, gender, and nationality) and their travel behaviors (such as departure and arrival airports, flight status, and travel dates). The goal is to uncover patterns that can help improve airline services, optimize flight schedules, and enhance customer satisfaction.

➤ **Business requirements :**

To effectively analyze and address the business problem, the following requirements are established:

1. **Data Acquisition**: Obtain comprehensive data on passenger demographics, travel patterns, flight schedules, and statuses.
2. **Integration with Qlik Sense**: Connect and model the data within Qlik Sense to facilitate detailed analysis and visualization.
3. **Interactive Dashboards**: Develop user-friendly and interactive dashboards that allow stakeholders to explore data and uncover insights.
4. **Advanced Analytics**: Implement data analytics techniques to identify correlations and trends between passenger demographics and travel behaviors.
5. **Performance and Responsiveness**: Ensure that the dashboards and visualizations perform efficiently with large datasets and provide real-time interactivity.

➤ **Literature Survey**

A thorough literature survey is conducted to explore existing research and methodologies relevant to the project. Key areas of focus include:

1. **Demographic Analysis in Airlines**: Studies on how passenger demographics influence travel preferences and behaviors.
2. **Travel Pattern Analysis**: Research on identifying and analyzing travel patterns to

optimize airline operations.
3. **Flight Schedule Optimization**: Best practices and strategies for optimizing flight schedules based on passenger data.
4. **Customer Satisfaction in Airlines**: Insights into improving customer satisfaction through data-driven decision-making.
5. **Qlik Sense in Data Analysis**: Case studies and best practices for utilizing Qlik Sense in business intelligence and data visualization.

Data Collection:

Data collection is the process of gathering and measuring information on variables of interest, in an established systematic fashion that enables one to answer stated research questions, test hypotheses, evaluate outcomes, and generate insights from the data.

➤ **Collect the dataset:**

For this project, the dataset was sourced from Kaggle, a well-known platform for datasets and data science competitions. The selected dataset includes comprehensive information on passenger demographics and travel patterns, making it suitable for analyzing and uncovering key trends and insights. The dataset includes:

1. **Passenger Demographics**: Age, gender, nationality.
2. **Travel Patterns**: Departure and arrival airports, travel dates, flight statuses (on-time, delayed, canceled).
3. **Additional Variables**: Booking lead time, flight duration, and other relevant factors influencing travel behavior.

Dataset: https://www.kaggle.com/datasets/iamsouravbanerjee/airline-dataset/data

➤ **Connect Data with Qlik Sense:**

After collecting the dataset, the next step is to connect it with Qlik Sense for analysis and visualization. This involves several key steps:

1. **Importing the Dataset**: Load the dataset from Kaggle into Qlik Sense. Ensure that all necessary fields and variables are included during the import process.
2. **Ensuring Data Integrity**: Validate the dataset to confirm its accuracy and completeness. Address any issues such as missing or inconsistent data points through cleaning and preprocessing.
3. **Creating Data Connections**: Establish relationships between different data tables

within Qlik Sense. This step is crucial for enabling comprehensive analysis and ensuring that visualizations are based on accurately linked data.

4. **Data Modeling**: Develop a data model that reflects the structure and relationships of the dataset. This model will serve as the foundation for all subsequent analyses and visualizations.

Data Preparation:

Data preparation is a critical step in the data analysis process, involving the cleaning, transformation, and structuring of data to ensure it is ready for analysis and visualization. Proper data preparation enables accurate and meaningful insights

➤ **Prepare the Data for Visualization**

The following steps are undertaken to prepare the dataset for visualization in Qlik Sense:

- **Data Cleaning**:
  - **Remove Duplicates**: Eliminate duplicate records to ensure data integrity.
  - **Handle Missing Values**: Address missing data points through imputation, interpolation, or deletion.
  - **Correct Inconsistencies**: Standardize inconsistent data entries, such as variations in airport codes or date formats.
- **Data Transformation**:
  - **Normalization**: Convert all date formats to a standard format.
  - **Aggregation**: Summarize data to the required level of granularity, such as monthly flight counts.
  - **Calculated Fields**: Create new fields for additional insights, like passenger age or total flight duration.
- **Data Structuring**:
  - **Data Modeling**: Develop a data model in Qlik Sense to represent relationships between tables.
  - **Hierarchy Creation**: Establish hierarchies for drill-down analysis, such as geographic or time dimensions.
  - **Filtering and Sorting**: Predefine filters and sorting options to enhance user experience.

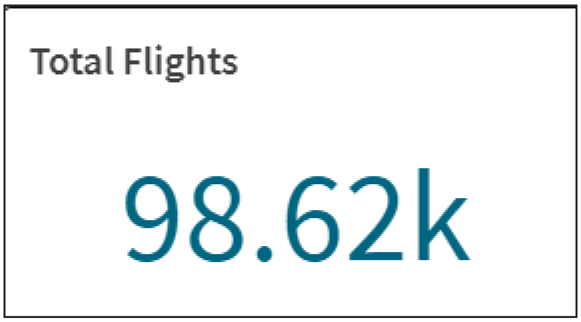
Data Visualizations:

Creating effective data visualizations is a crucial step in analyzing and interpreting the dataset.

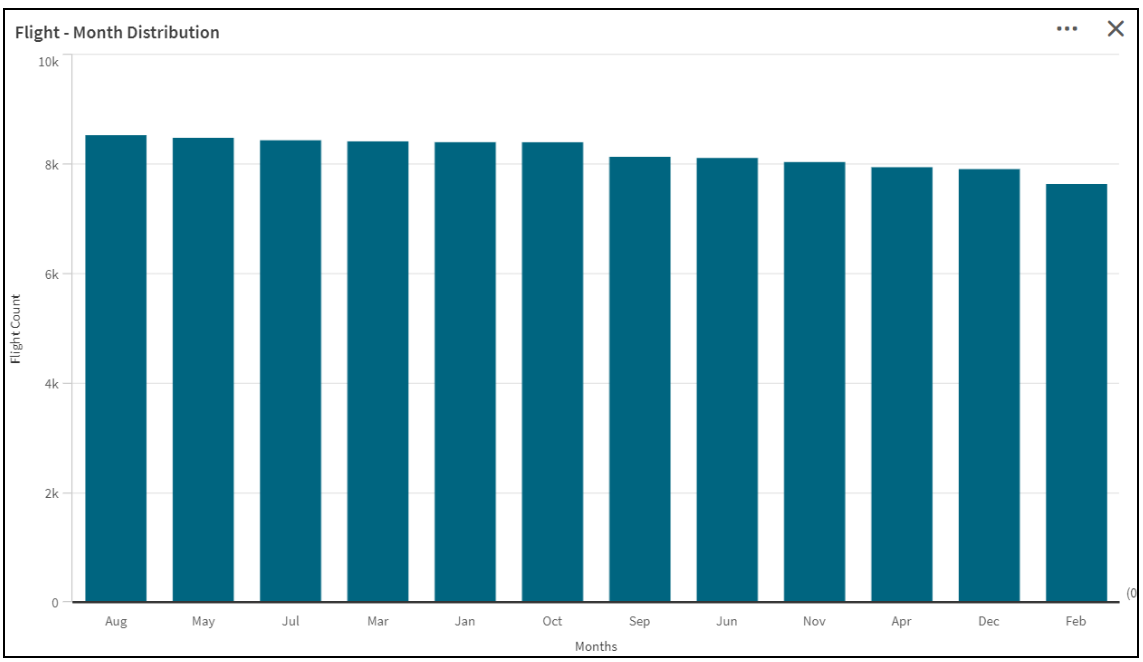
Visualizations help transform raw data into meaningful insights that can be easily understood and acted upon.

- **Visualizations :**

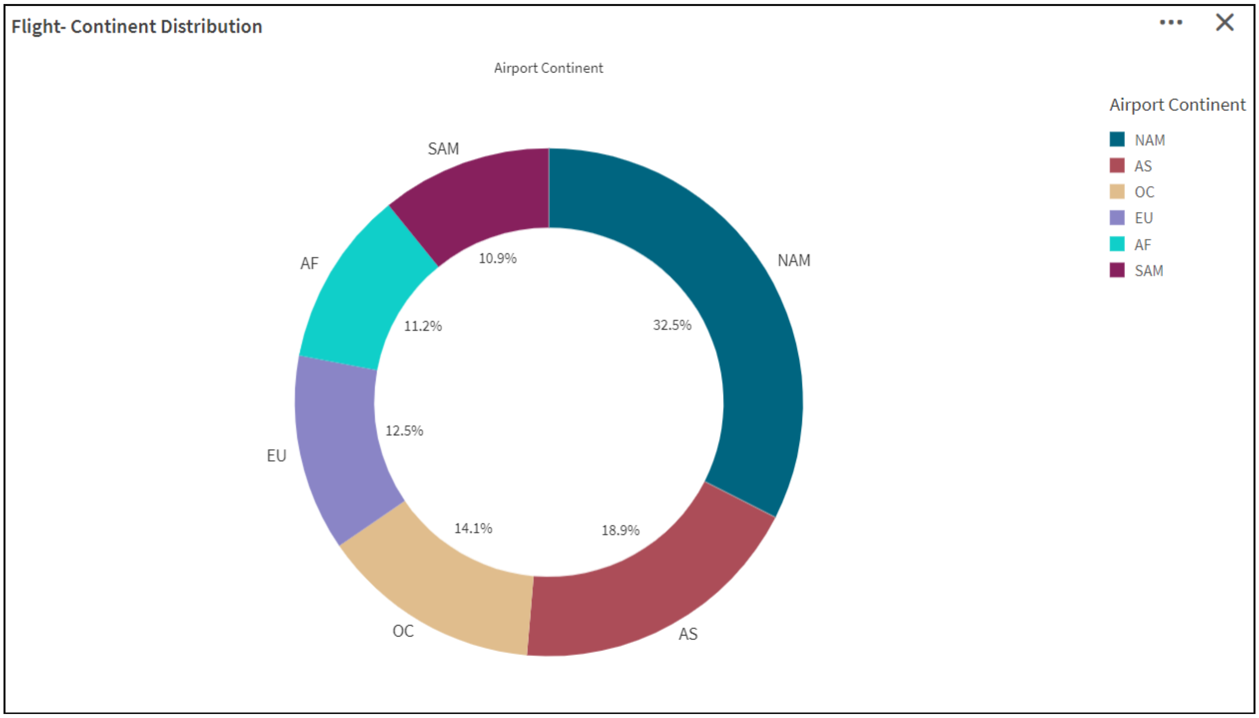
1. Total Flights KPI:



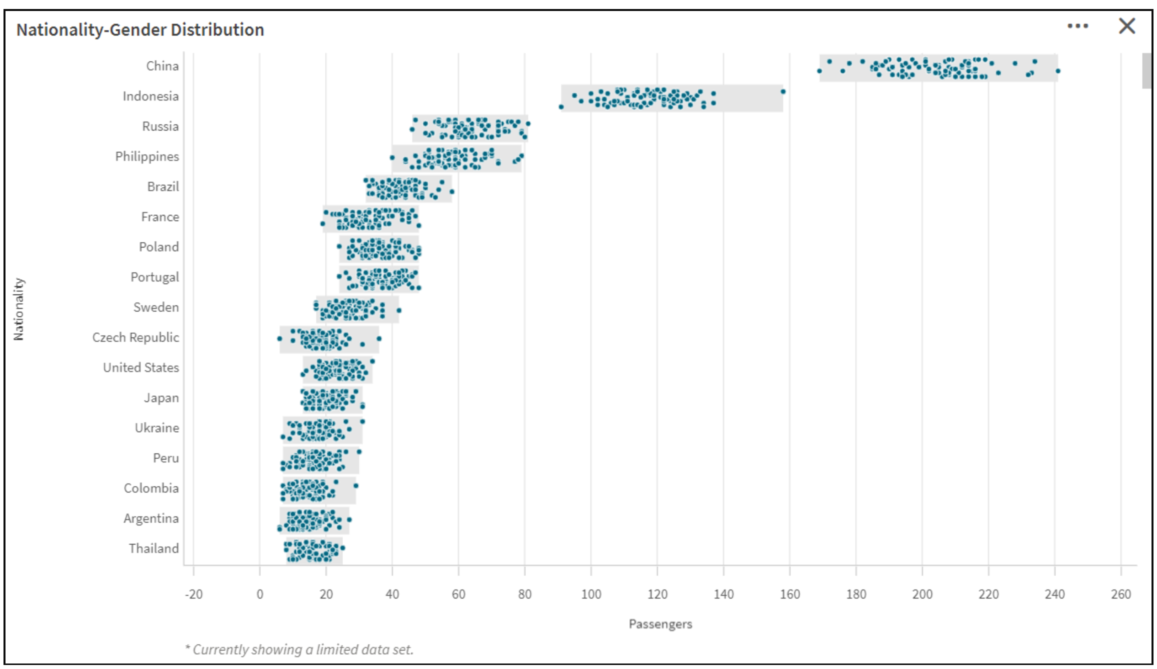
2. Flight Month Distribution Bar Graph:



3. Flight Continent Distribution Pie Chart:



4. Nationality Gender Distribution Plot:



5. Nationality Gener Description NL Insights:

Nationality - Gender Description

Calculated measure (KPI)

- The count of Passengers is 98.62k.

Breakdown (geospatial)

- Count of Passengers is 98.62k across 240 Nationality.
- Top Passengers is 18.32k where Nationality is China.

Breakdown

- Highest Passengers is 241 where Nationality is China and Age is 4.

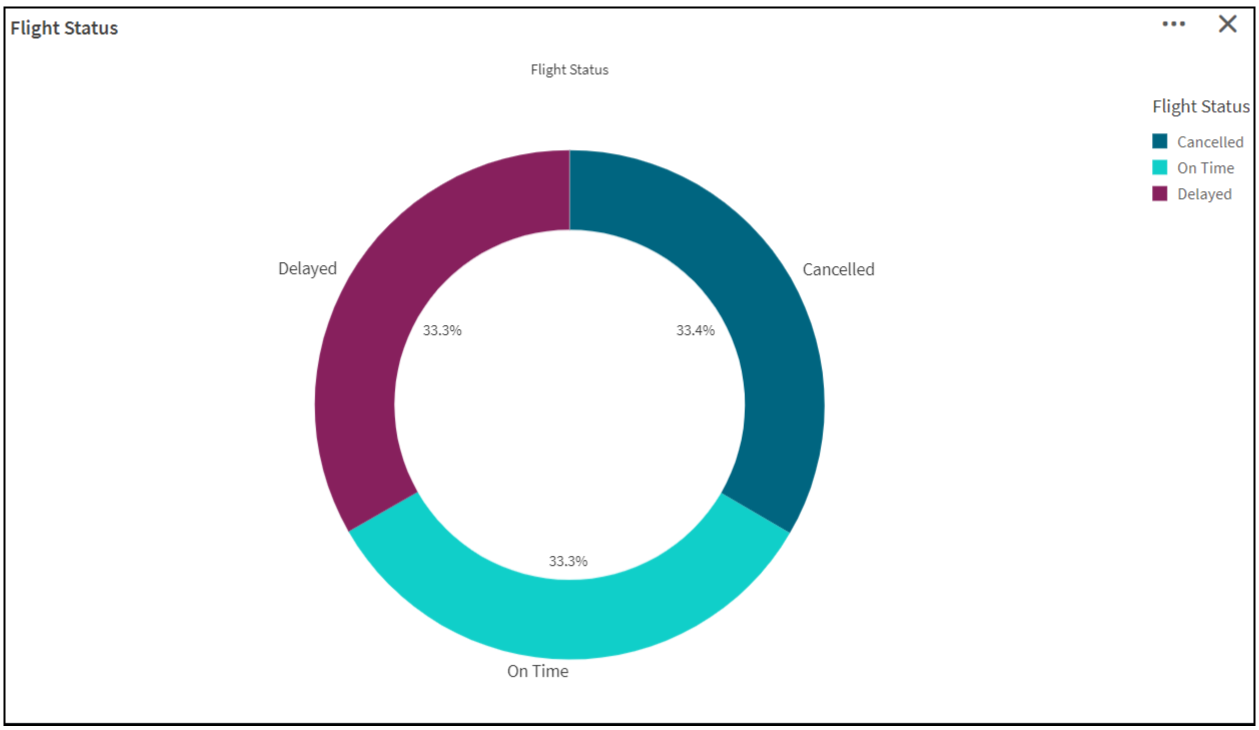
Comparison

- Comparison: Age is 0 and count of Passengers is 98.62k.

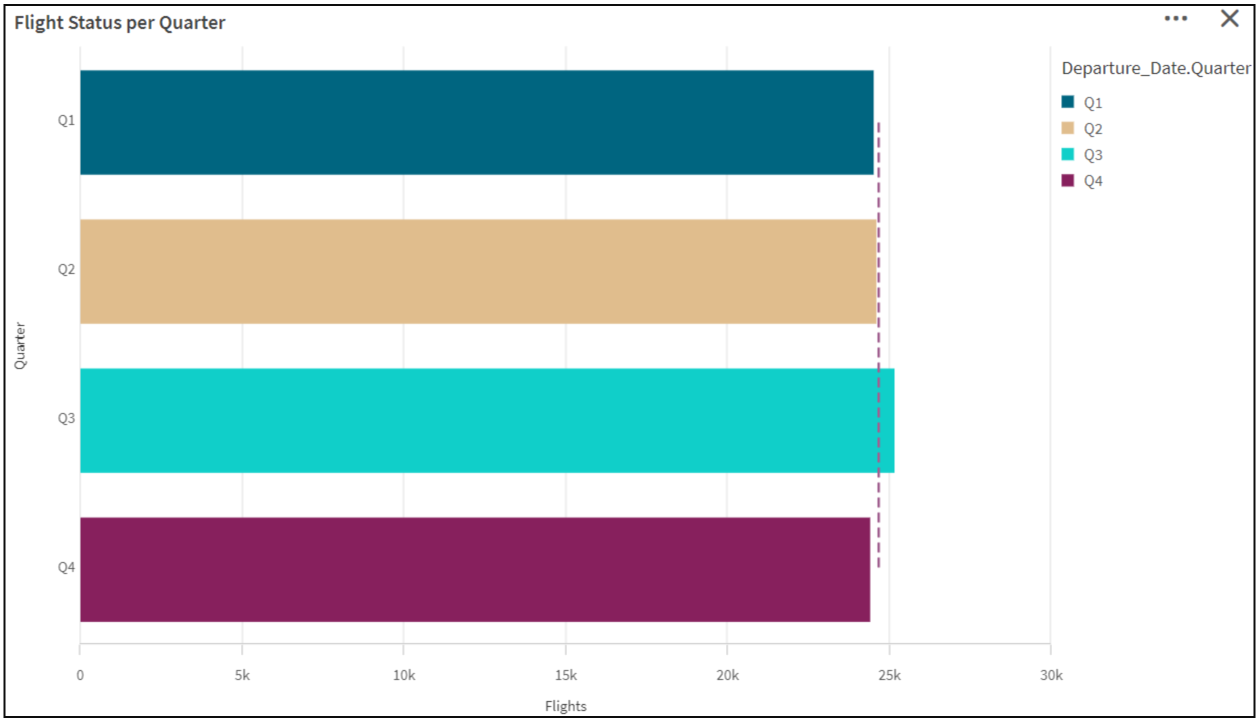
Mutual information

- The mutual dependence between Gender and Nationality is 15.58% while Gender and Age is 8.73%.

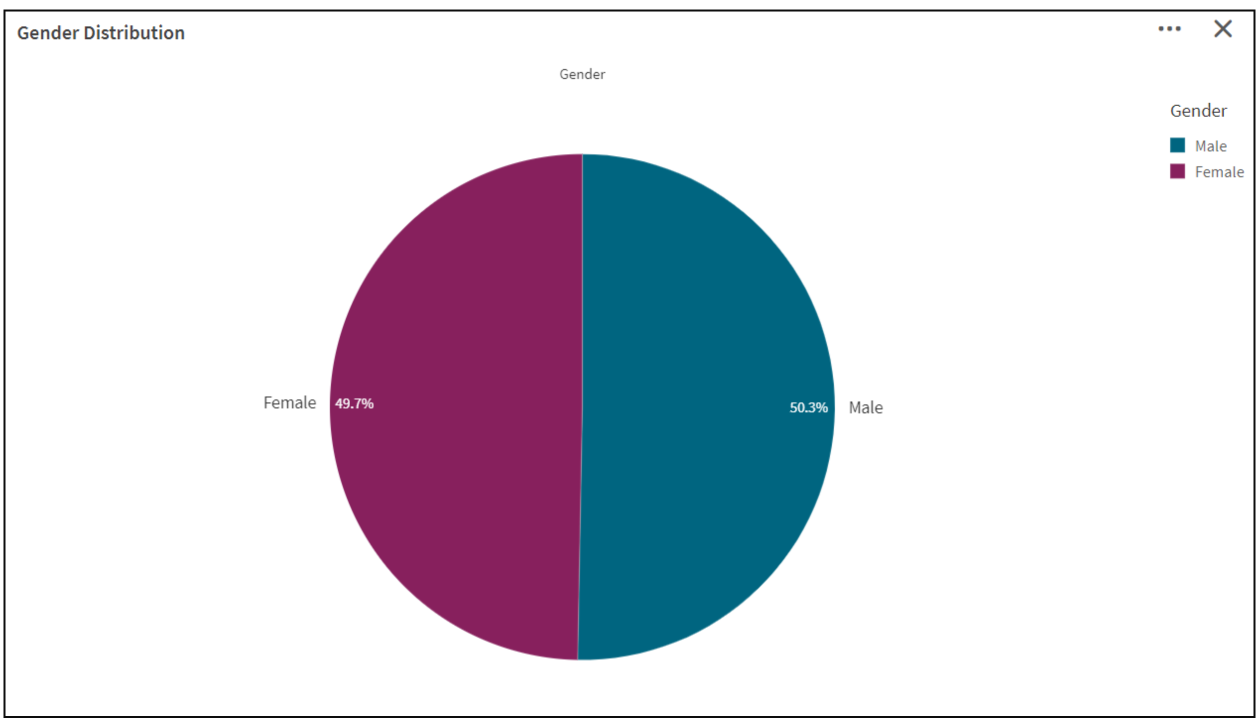
6. Flight Status Pie Chart:



7. Flights per Quarter Bar Graph:



8. Gender Distribution Pie Chart:



9. Passenger Age Slider:

10. Flight Status Filer Pane:

Flight Status

Cancelled

Delayed

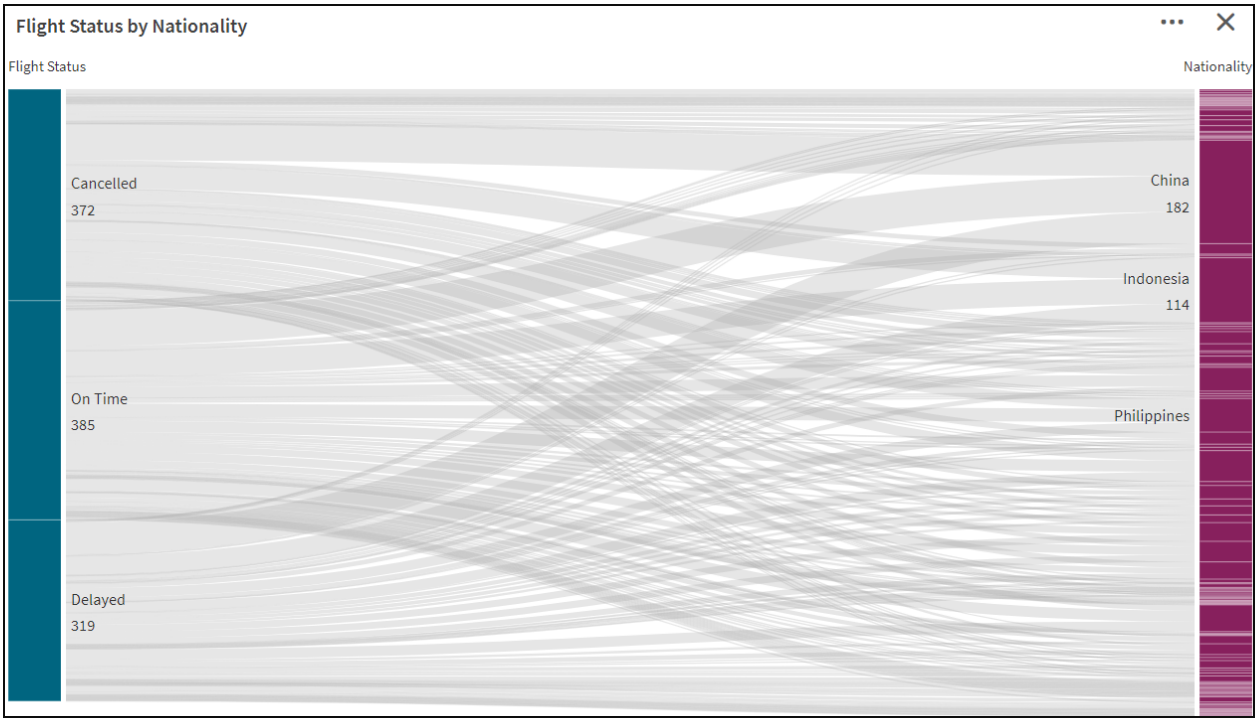
On Time

11. Average Age of Nationalities Table:

Average Age of Nationalities

Nationality	Avg(Age)
Totals	**45.504021**
Afghanistan	45.461929
Aland Islands	46.3
Albania	45.858427
Algeria	39
American Samoa	52.5
Andorra	41.777778
Angola	47.160494
Anguilla	63
Antigua and Barbuda	43.888889
Argentina	44.860856
Armenia	46.136816
Aruba	38.75
Australia	41.022472
Austria	38.486486
Azerbaijan	47.435798
Bahamas	42.529412
Bahrain	42.666667
Bangladesh	43.29918
Barbados	61.7
Belarus	44.431548
Belgium	50.52
Belize	45.210526

12. Flight Status By Nationality Sankey Chart:



13. Flight Status By Nationality NL Insights:

Flight Status by Nationality Description

Calculated measure (KPI)

- The count of Count([Flight Status]) is 1.08k.

Ranking

- The count of Count([Flight Status]) is 1076.

Breakdown (geospatial)

- Count of Count([Flight Status]) is 1076 across 125 Nationality.
- Top Count([Flight Status]) is 182 where Nationality is China.

Mutual information

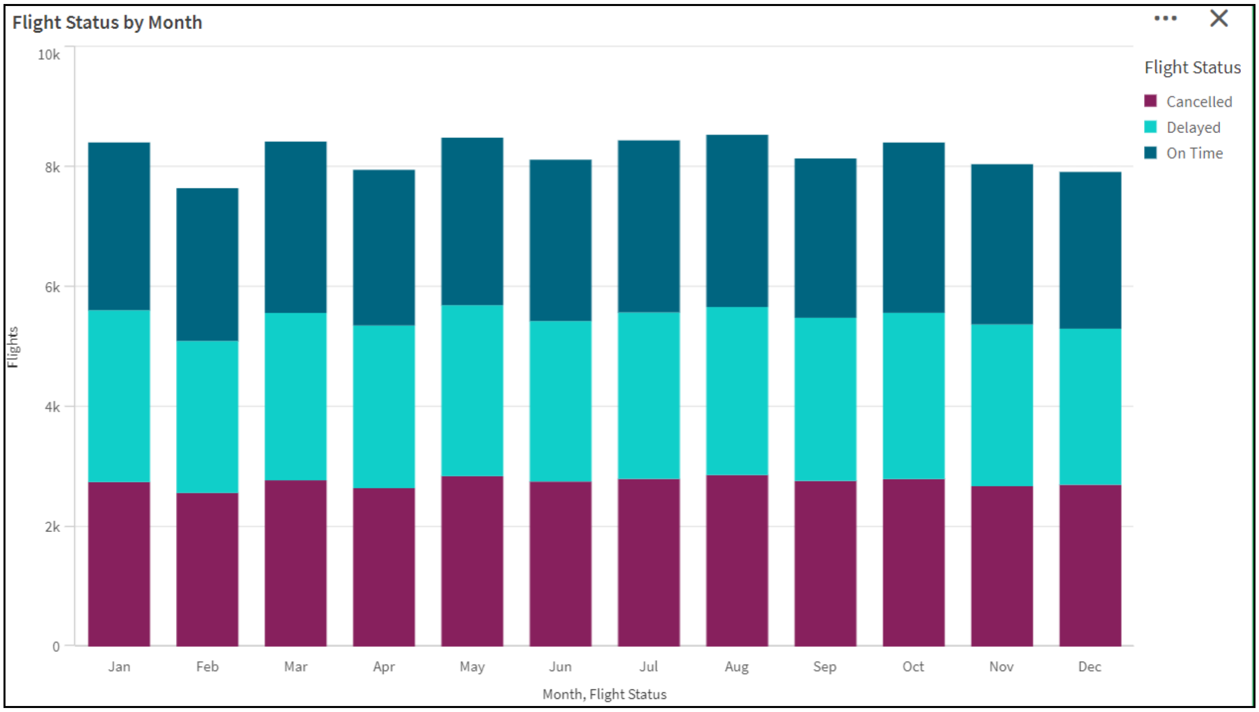
- The mutual dependence between Flight Status and Nationality is 50.44%.

14. Pilot Name Filter Pane:

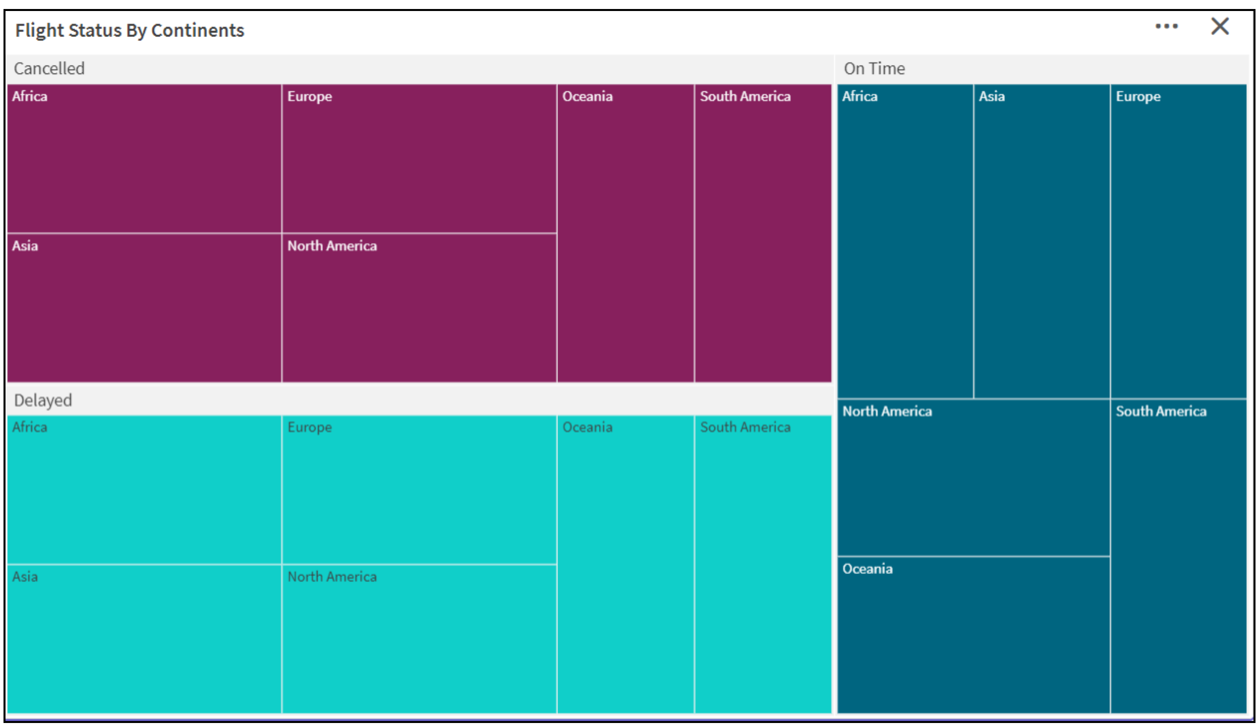
Pilot Name
Edithe Leggis
Elwood Catt
Darby Felgate
Dominica Pyle
Bay Pencost
Lora Durbann
Rand Bram
Perceval Dallosso
Aleda Pigram
Burlie Schustl
Porty Jori
Briant De La Haye
Kalie Scoble
Catriona Beaument
Amberly Handling
Dyna De'Vere - Hunt
Janella Hardaker
Alvin Wenzel

15. Passenger ID Filter Pane:

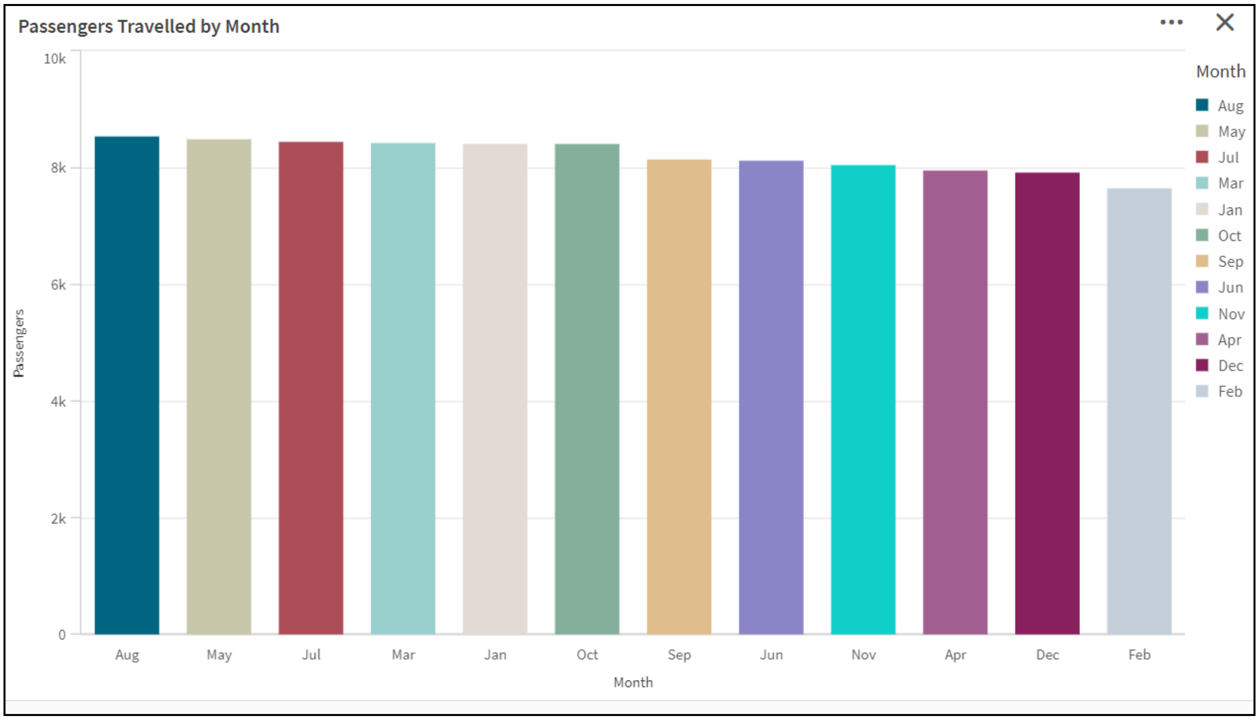
16. Flight Status by Month:



17. Flight Status by Continent:



18. Passengers Travelled By Month:



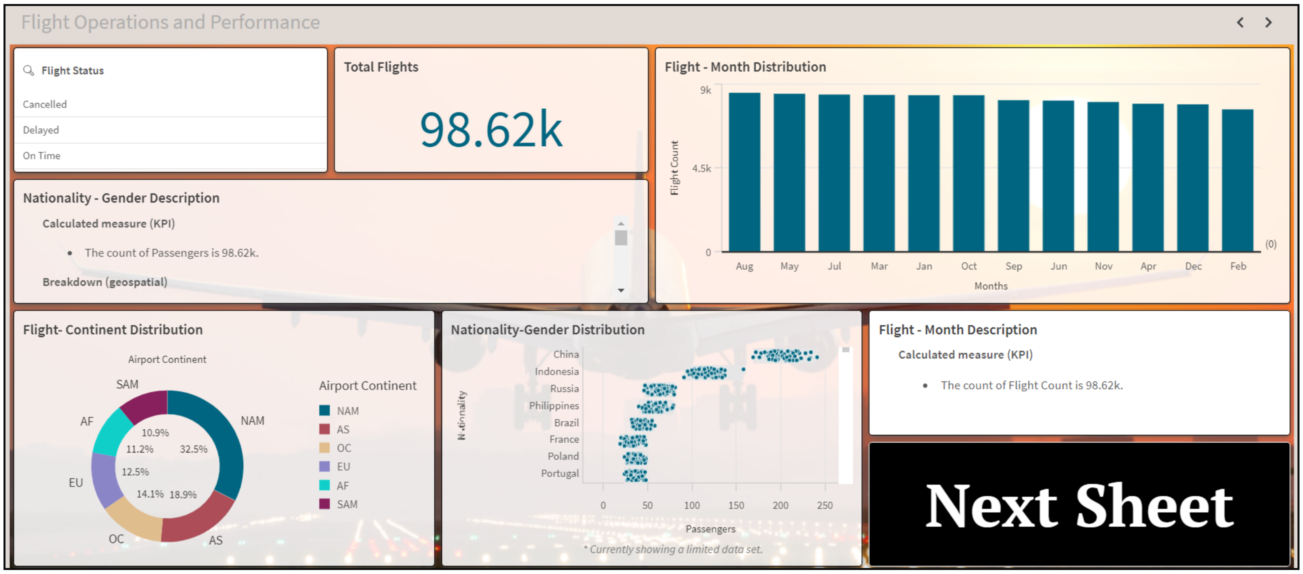
Dashboard

The dashboard is designed to be interactive and user-friendly, enabling stakeholders to explore the data in depth. Key features include:

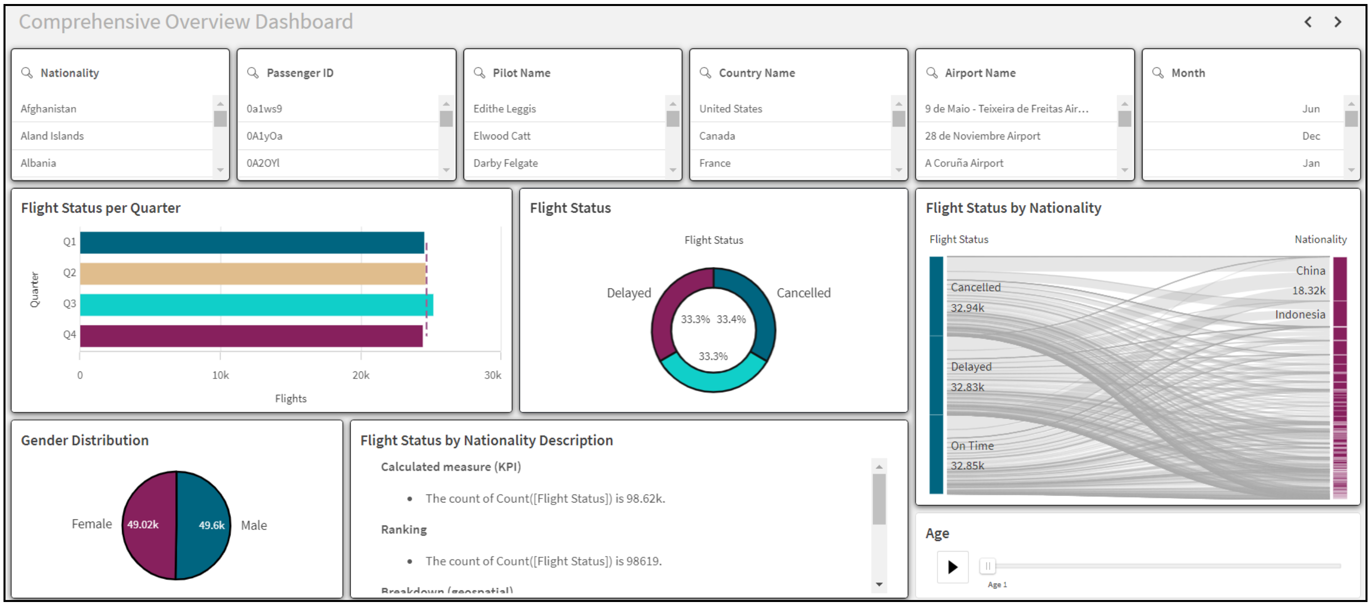
- **Filters and Selections**: Allow users to filter data by various dimensions, such as date range, flight status, and passenger demographics.
- **Responsive Design**: Ensure the dashboard adapts to different devices, providing a seamless experience on desktops, tablets, and smartphones.
- **Dynamic Updates**: Visualizations update in real-time as users interact with filters and selections, providing immediate insights.

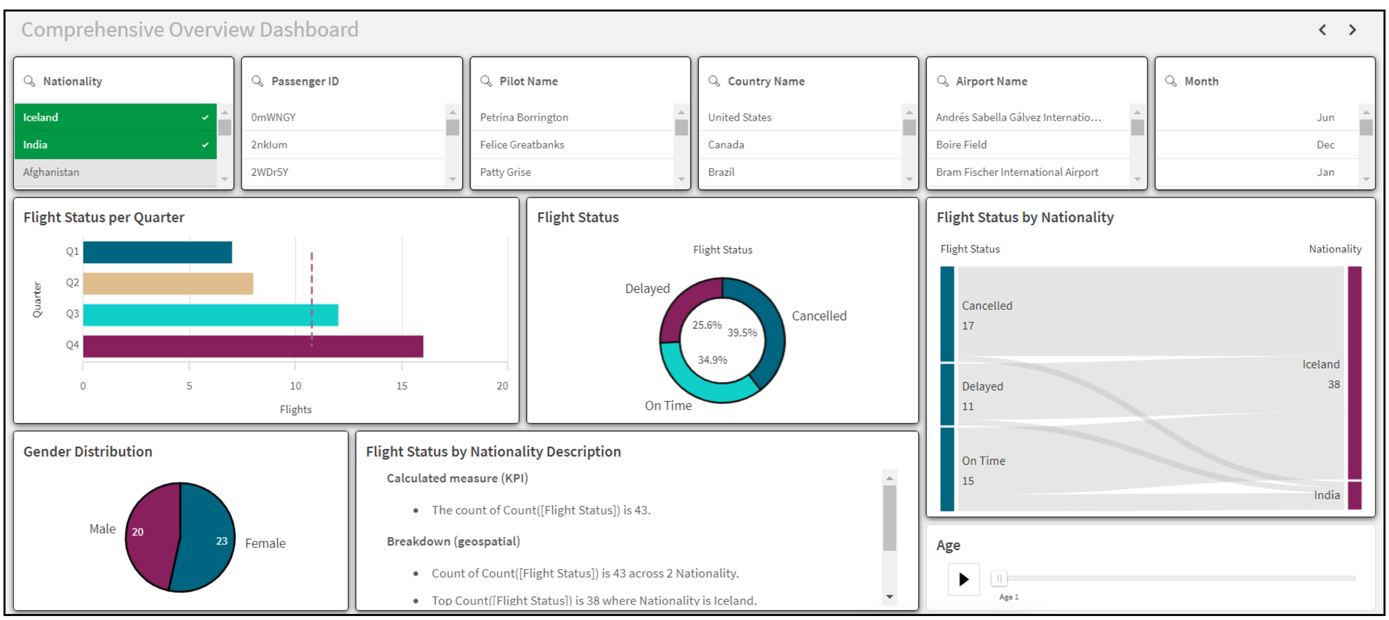
➤ **Responsive and Design of Dashboard:**

Flight Operation and Performance

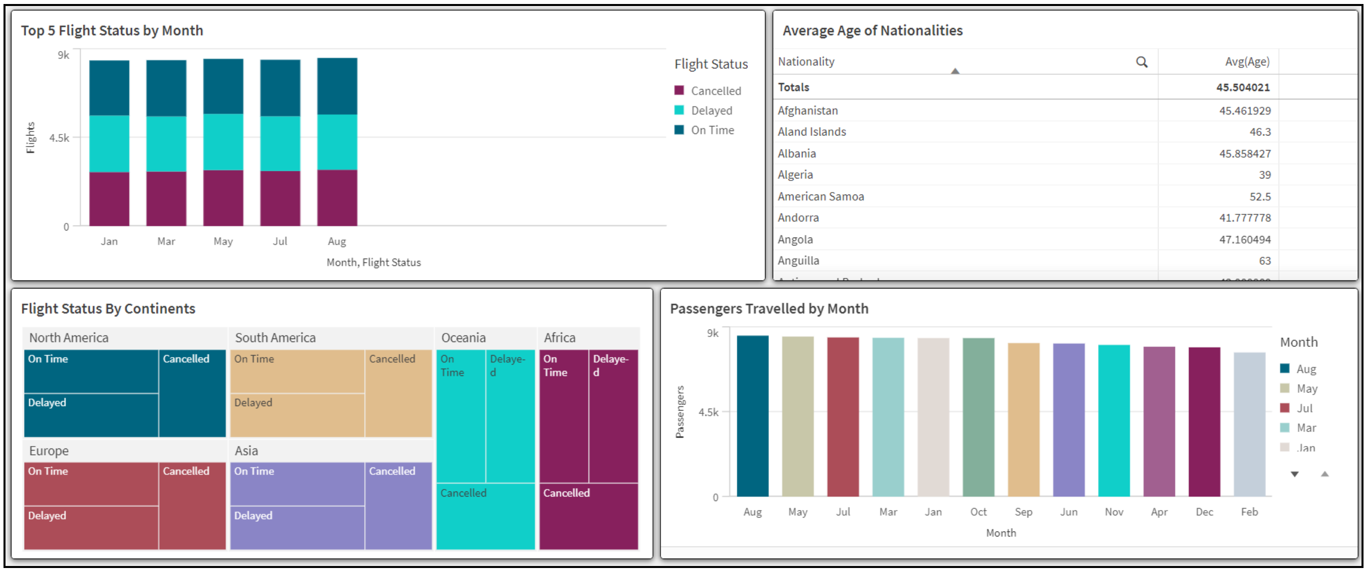


Comprehensive Overview Dashboard





<u>Dashboard -3</u>

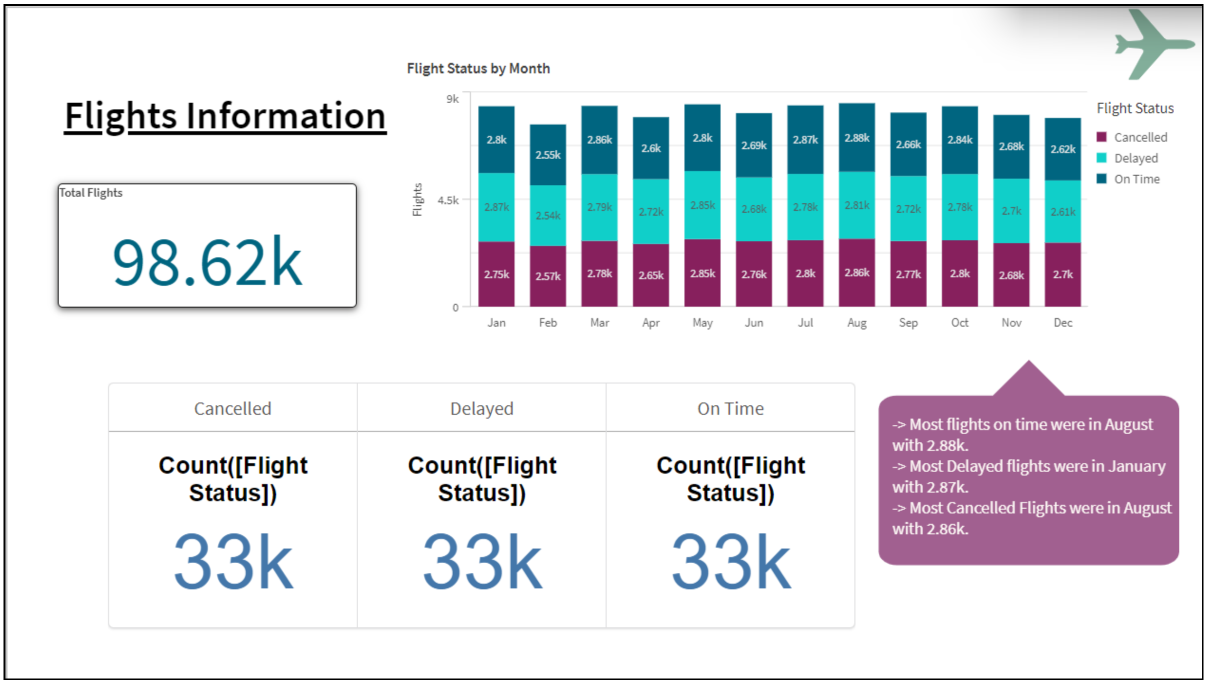

Top 5 Flight Status by Month
9k
4.5k
0
Flights
Jan
Mar
May
Jul
Aug
Month, Flight Status
Flight Status
Cancelled
Delayed
On Time
Average Age of Nationalities
Nationality
Avg(Age)
Totals
45.504021
Afghanistan
45.461929
Aland Islands
46.3
Albania
45.858427
Algeria
39
American Samoa
52.5
Andorra
41.777778
Angola
47.160494
Anguilla
63
Flight Status By Continents
North America
On Time
Cancelled
Delayed
South America
On Time
Cancelled
Delayed
Oceania
On Time
Delaye-d
Cancelled
Africa
On Time
Delaye-d
Cancelled
Europe
On Time
Cancelled
Delayed
Asia
On Time
Cancelled
Delayed
Passengers Travelled by Month
9k
4.5k
0
Passengers
Aug
May
Jul
Mar
Jan
Oct
Sep
Jun
Nov
Apr
Dec
Feb
Month
Month
Aug
May
Jul
Mar
Jan

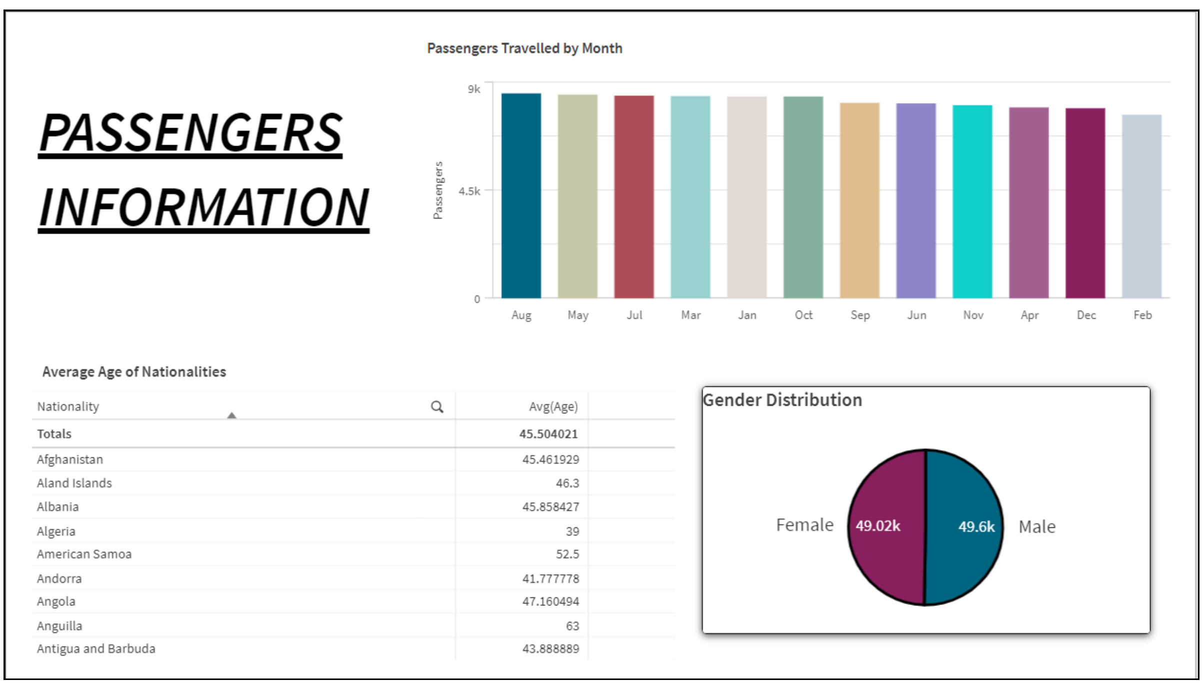

Top 5 Flight Status by Month
2
1
0
Flights
Feb
Mar
May
Jun
Nov
Month, Flight Status
Flight Status
Cancelled
On Time
Delayed
Average Age of Nationalities
Nationality
Avg(Age)
Totals
40.166667
Australia
40.166667
Flight Status By Continents
South America
On Time
Delayed
Cancelled
Passengers Travelled by Month
2
1
0
Passengers
Feb
May
Jun
Nov
Mar
Jul
Sep
Oct
Month
Month
Feb
May
Jun
Nov
Mar

Story:

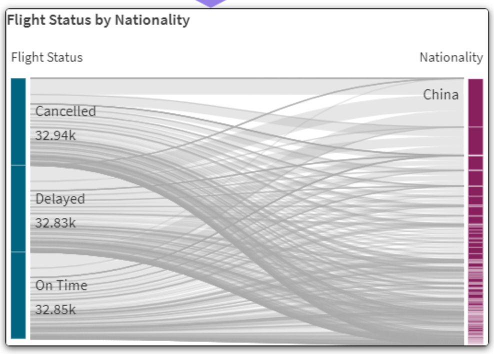
Creating a compelling story from data visualizations is essential to communicate insights effectively. A well-constructed story in Qlik Sense guides the audience through the data, highlighting key findings and enabling informed decision-making.

- **<u>Story Creation</u>**





This chart shows that maximum flights that took has Chinese Nationality



Flight Status By Continents

Division of Flights
(Flight Status) by continents.

Distribution According to Flight Status :-

- Cancelled Flights: Click Here
- Delayed Flights: Click Here
- On-Time Flights: Click Here

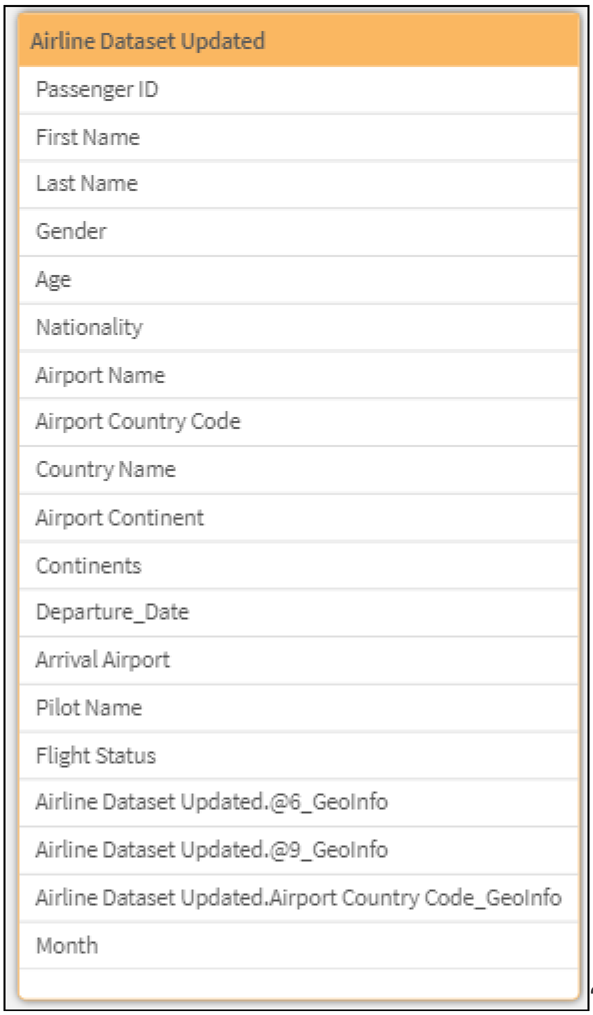
Distribution According to Countries :-

- Distribution for India: Click Here
- Distribution for US : Click Here

<u>Performance Testing:</u>

Performance testing is essential to ensure that the Qlik Sense application can handle large volumes of data efficiently and provide a smooth user experience. This includes testing the amount of data rendered to the database, the utilization of data filters, and the efficiency of data preprocessing scripts

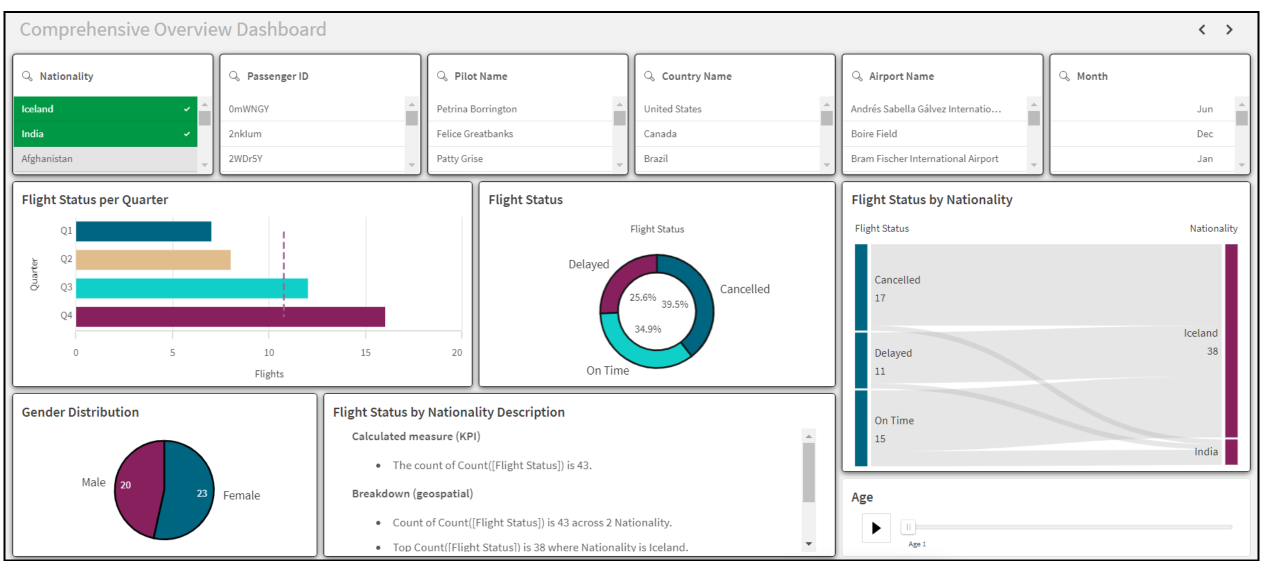
- ➤ **<u>Amount of Data Rendered to DB:</u>**



- **Utilization of Data Filters:**

  Ensuring that data filters in Qlik Sense are effective and responsive is crucial for providing users with a seamless and interactive experience. The following steps outline the process of testing and optimizing the utilization of data filters:

  1. Filter Response Time
  2. Combination Filters
  3. Dynamic Filtering
  4. Optimization Techniques



- **Data Preprocessing – Qlik Sense Script**

```
SET ThousandSep=',';
SET DecimalSep='.';
SET MoneyThousandSep=',';
SET MoneyDecimalSep='.';
SET MoneyFormat='$ ###0.00;-$ ###0.00';
SET TimeFormat='h:mm:ss TT';
SET DateFormat='M/D/YYYY';
SET TimestampFormat='M/D/YYYY h:mm:ss[.fff] TT';
SET FirstWeekDay=6;
SET BrokenWeeks=1;
SET ReferenceDay=0;
SET FirstMonthOfYear=1;
SET CollationLocale='en-US';
SET CreateSearchIndexOnReload=1;
SET MonthNames='Jan;Feb;Mar;Apr;May;Jun;Jul;Aug;Sep;Oct;Nov;Dec';
SET LongMonthNames='January;February;March;April;May;June;July;August;September;October;November;December';
SET DayNames='Mon;Tue;Wed;Thu;Fri;Sat;Sun';
SET LongDayNames='Monday;Tuesday;Wednesday;Thursday;Friday;Saturday;Sunday';
SET NumericalAbbreviation='3:k;6:M;9:G;12:T;15:P;18:E;21:Z;24:Y;-3:m;-6:μ;-9:n;-12:p;-15:f;-18:a;-21:z;-24:y';
```

- **No. of Visualisations and Graph**: 19

Project Demonstration & Documentation

- **Record explanation Video for project end to end solution:**

  https://drive.google.com/file/d/1m5ikAZODkfa94sF_AS0pomLpQOmbm6YN/view?usp=sharing