Project Title: Synthetic Airline Data Analysis with Qlik

## Project Flow

**Project Overview:**

The Synthetic Airline Data Analysis project aims to analyze simulated airline data using the powerful data visualization and analytics capabilities of Qlik. The project involves exploring various aspects of airline operations, passenger demographics, flight statuses, and geographical patterns to derive actionable insights for improved decision-making within the aviation industry.

**Project Objectives:**

1. Explore and analyze synthetic airline data to understand passenger behavior, flight performance, and operational trends.
2. Identify key factors influencing flight delays, cancellations, and passenger satisfaction.
3. Create interactive dashboards and visualizations using Qlik to present insights in a user-friendly manner.
4. Perform demographic analysis to understand the age and gender distribution of passengers.
5. Analyze seasonal trends and geographical variations in flight demand and performance.
6. Provide recommendations for optimizing airline operations, improving passenger experience, and enhancing overall efficiency.

**Project Activities:**

1. Data Acquisition: Obtain synthetic airline data from simulated sources representing passenger information, flight details, and operational metrics.
2. Data Preparation: Clean, transform, and integrate the data to ensure its suitability for analysis within the Qlik platform.
3. Dashboard Development: Design and develop interactive dashboards using Qlik Sense to visualize key performance indicators (KPIs), trends, and insights.
4. Exploratory Data Analysis (EDA): Explore the data to identify patterns, correlations, and anomalies that can provide valuable insights for decision-making.
5. Advanced Analytics: Apply advanced analytical techniques such as predictive modeling, clustering, and time-series analysis to extract deeper insights from the data.
6. Stakeholder Engagement: Present findings and insights to stakeholders, including airline executives, operations managers, and marketing teams, to facilitate data-driven decision-making.
7. Documentation and Reporting: Document the analysis process, methodologies, and findings in a comprehensive report for future reference and knowledge sharing.

**Expected Deliverables:**

1. Interactive Qlik Sense dashboards showcasing key insights and trends in airline data.
2. Comprehensive analysis report detailing findings, recommendations, and actionable insights.
3. Presentation slides summarizing the project objectives, methodology, and key findings for stakeholder engagement.

### Data Collection & Extraction From Database

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.

link to download the dataset: [Link](https://www.kaggle.com/datasets/iamsouravbanerjee/airline-dataset/data)

Data contains all the meta information regarding the columns described in the CSV files  
Column Description of 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)

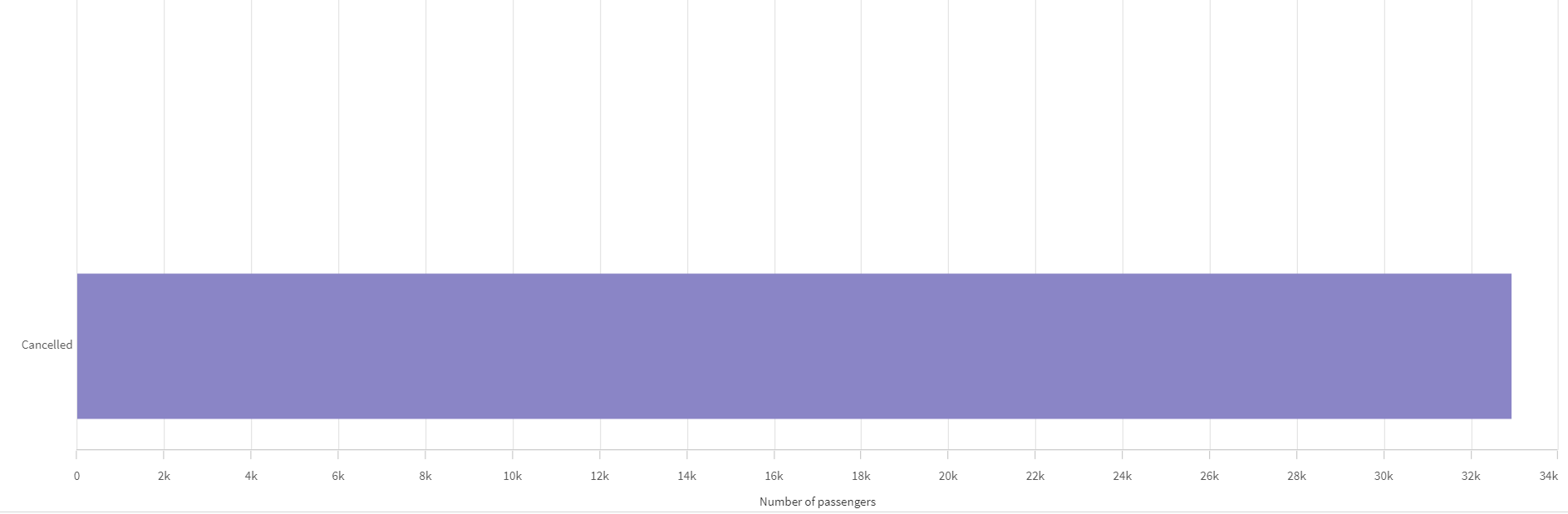
Data preparation involves transforming raw data into a format suitable for analysis.

1. **Data Collection**: Gather data from various sources such as databases, APIs, files, or sensors.
2. **Data Cleaning**: Identify and handle missing values, outliers, duplicates, and inconsistencies. This may involve techniques like imputation, filtering, and deduplication.
3. **Data Transformation**: Convert data into a usable format. This includes standardizing units, encoding categorical variables, and normalizing numerical data.
4. **Feature Engineering**: Create new features or modify existing ones to improve model performance. This can involve techniques like one-hot encoding, binning, and feature scaling.
5. **Data Integration**: Combine data from multiple sources if necessary, ensuring consistency and compatibility.
6. **Data Reduction**: Reduce the dimensional of the data while preserving its essential characteristics. Techniques like principal component analysis (PCA) and feature selection can be used.
7. **Data Splitting**: Divide the data into training, validation, and testing sets for model development and evaluation.
8. **Data Validation**: Check for data integrity and correctness. Perform sanity checks and validate against domain knowledge.
9. **Documentation**: Document the data preparation steps taken, including any assumptions or transformations made. This helps ensure reproducibility and transparency.
10. **Iterate**: Data preparation is often an iterative process. As you analyze the data and build models, you may discover the need for further cleaning or transformation.

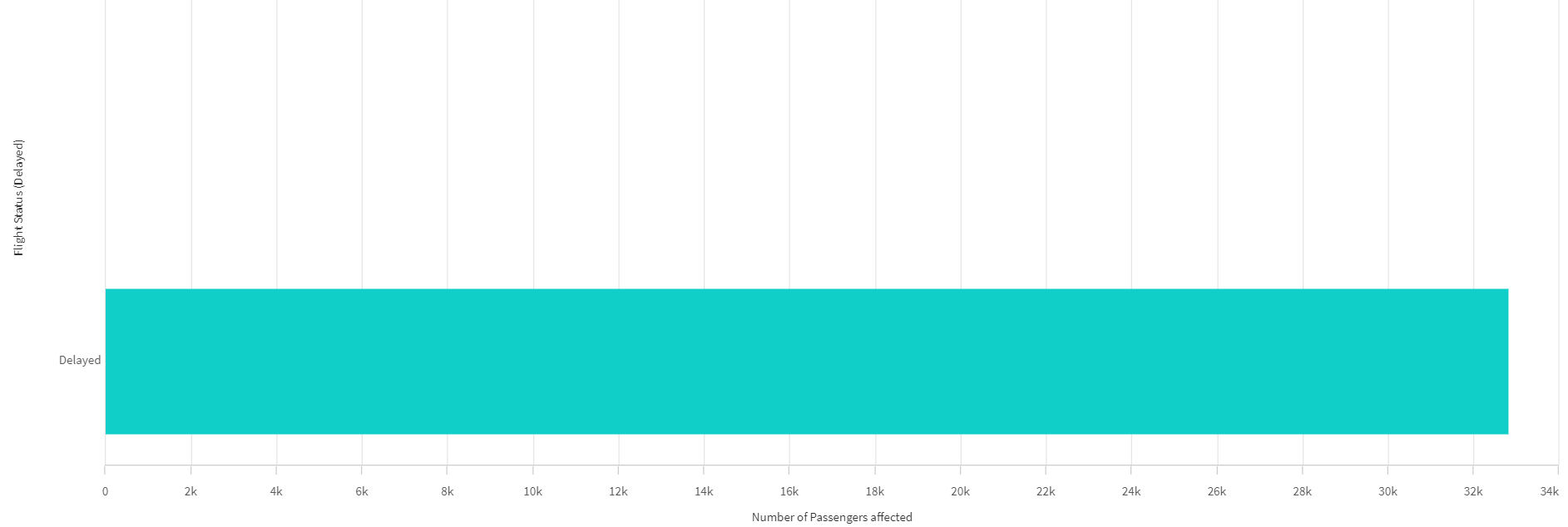
### Data Visualization

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

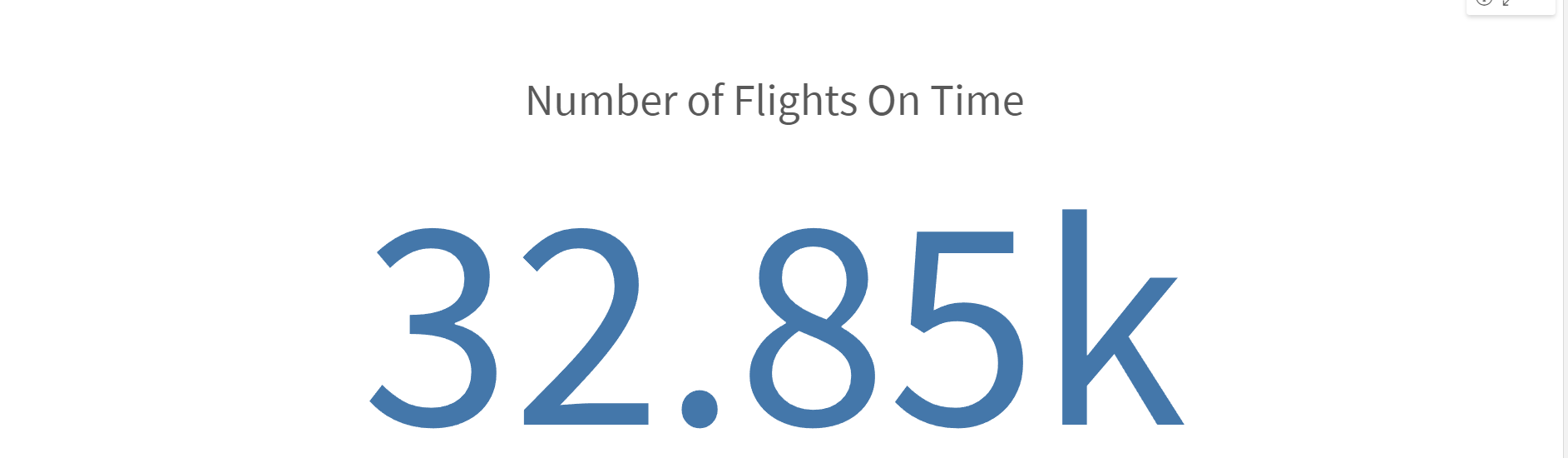
**Activity 1.1: Total No. of Passengers**

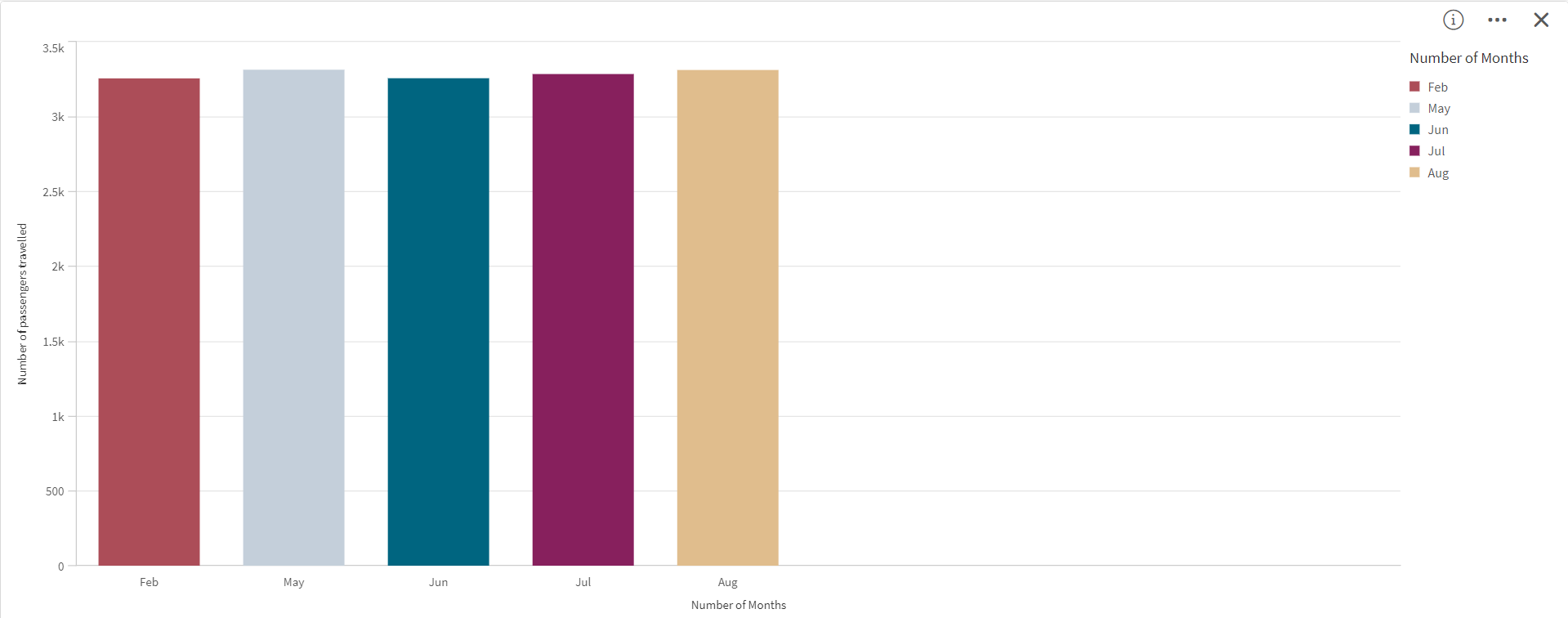
**Activity 1.2: Number of Passengers affected by cancelled flights**

**Activity 1.3: Number of Passengers affected by flight delays**

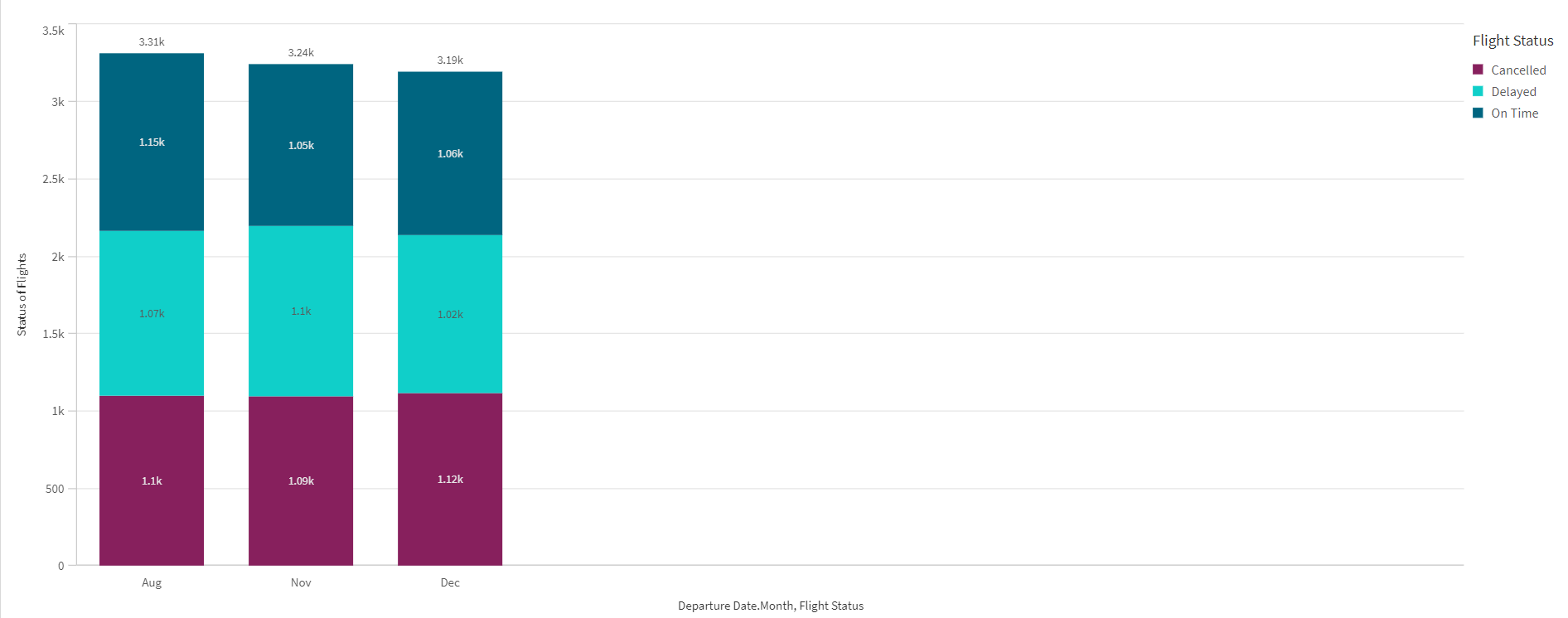


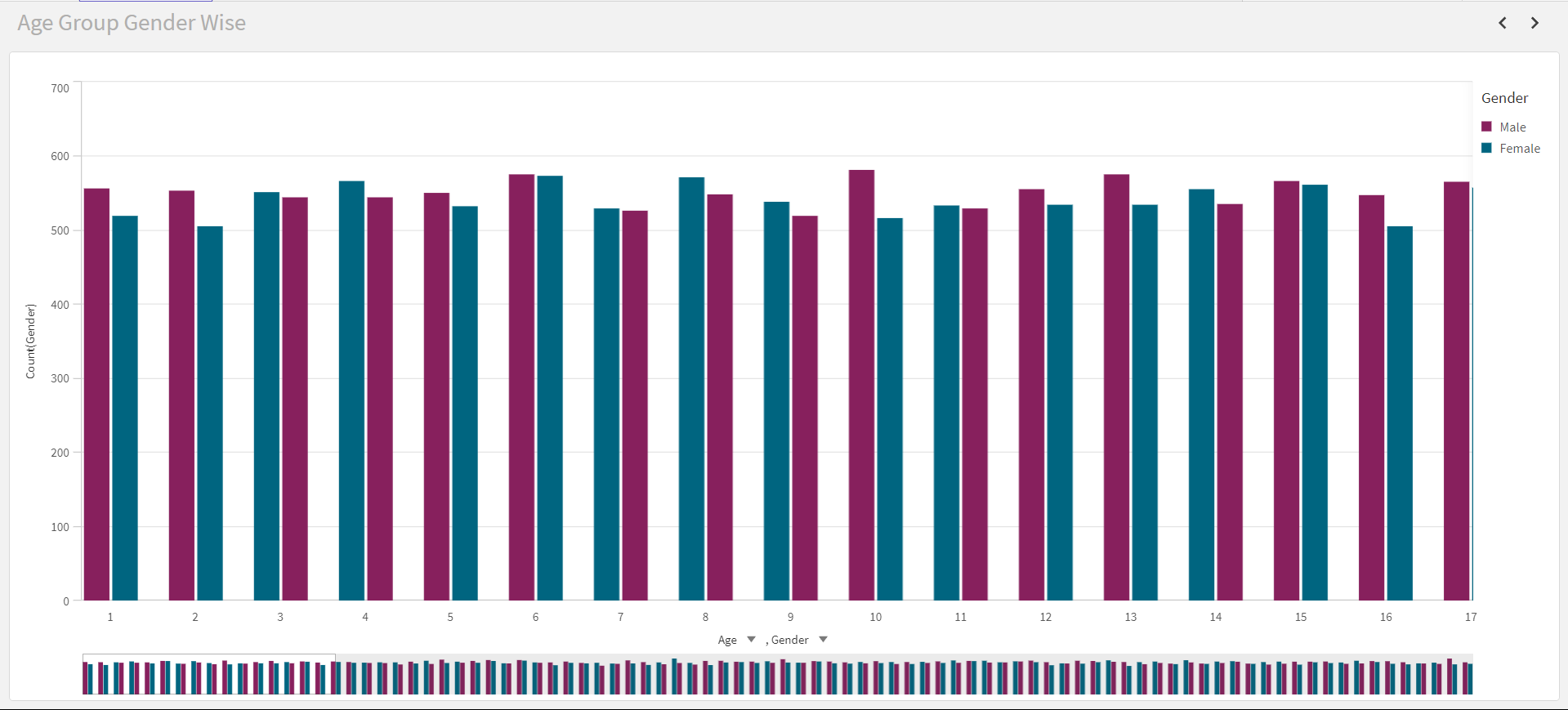
**Activity 1.4: Number of Flights on Time**

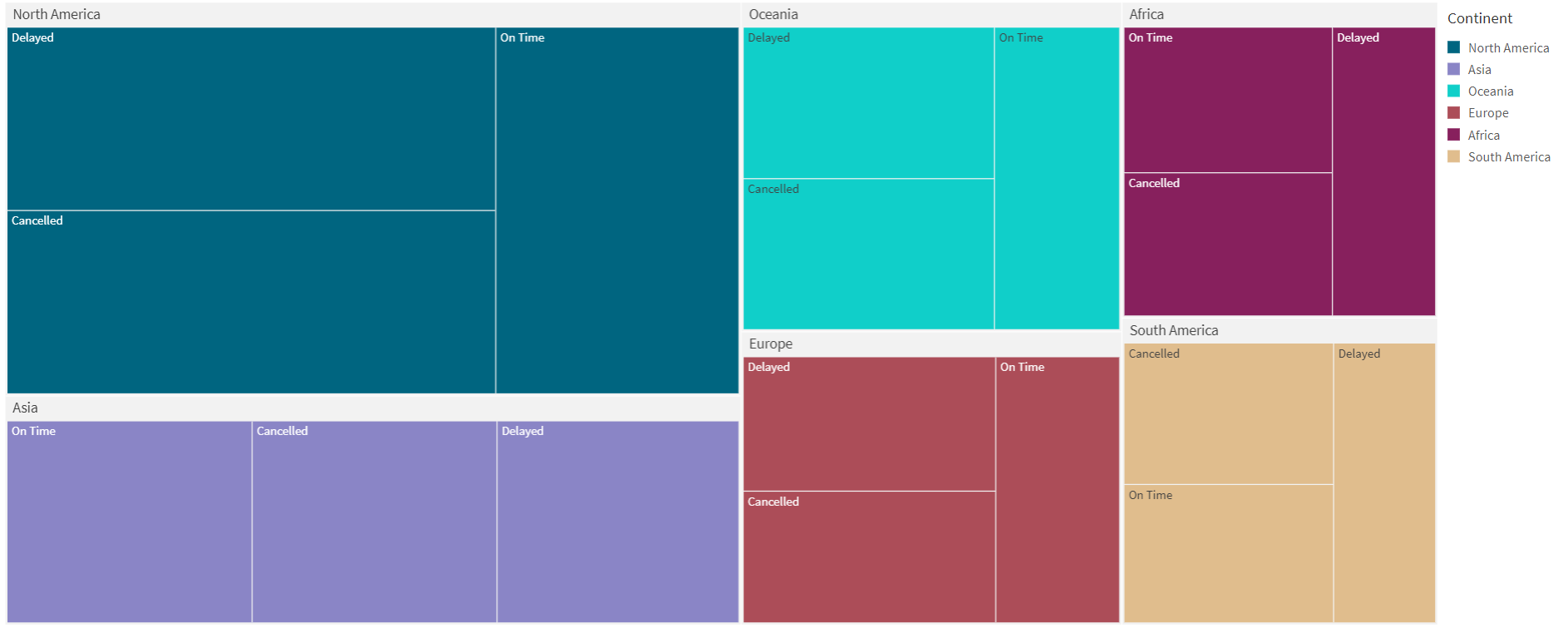


**Activity 1.5: Number of Passengers traveled - Month Wise**

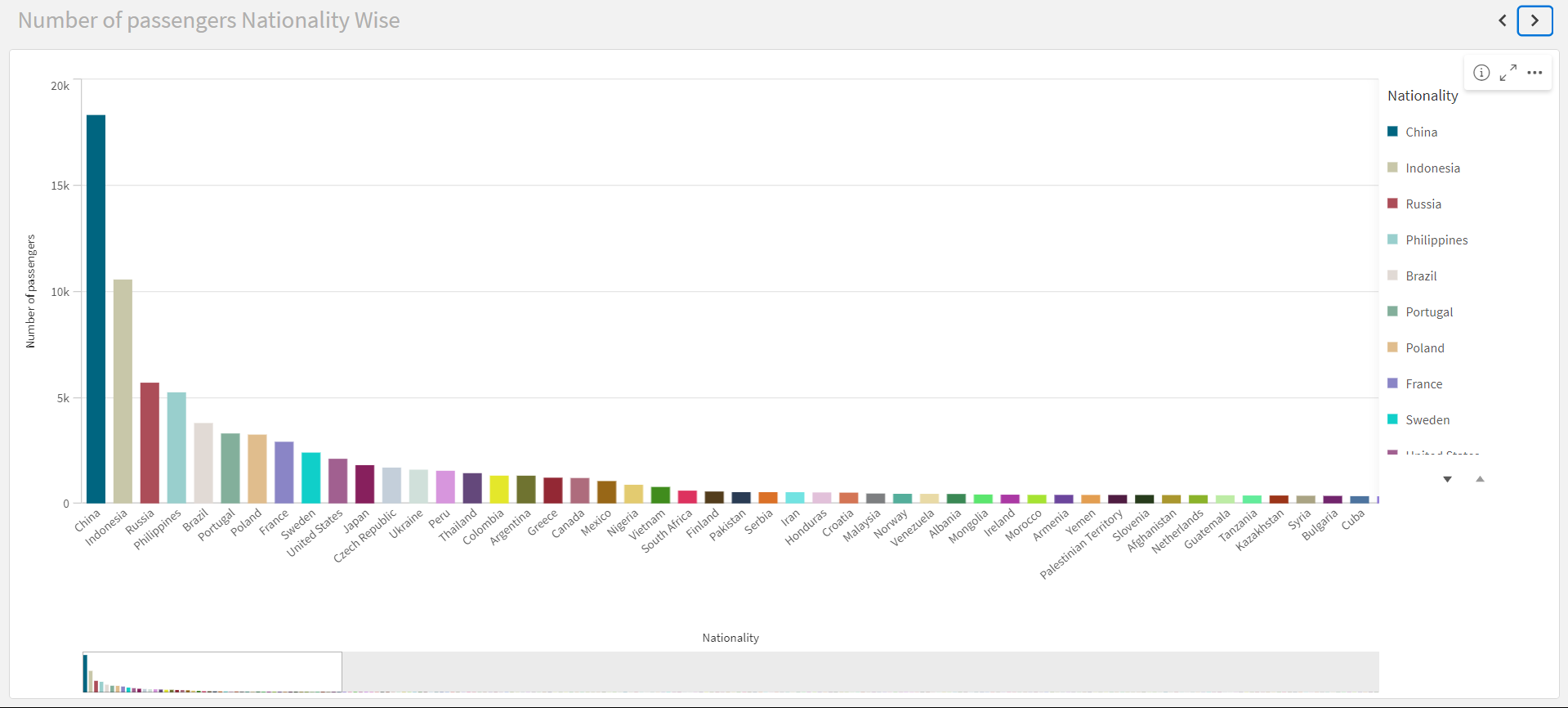
**Activity 1.6: Top 3 Month flights status wise**



**Activity 1.7: Age group of passenge as per gender wise**

**Activity 1.8: Continent-wise flight status**

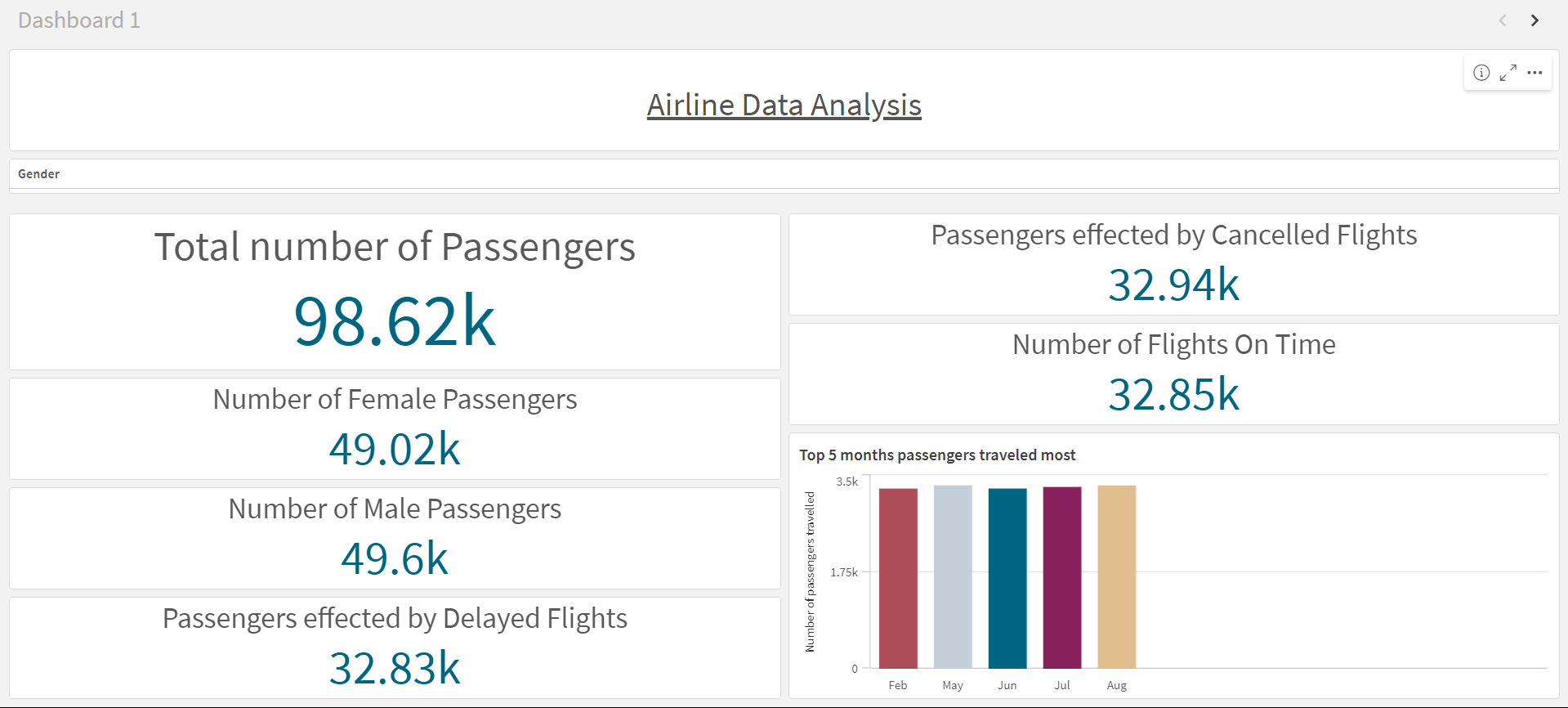
**Activity 1.9: Number of Passengers - Nation Wise Analysis**

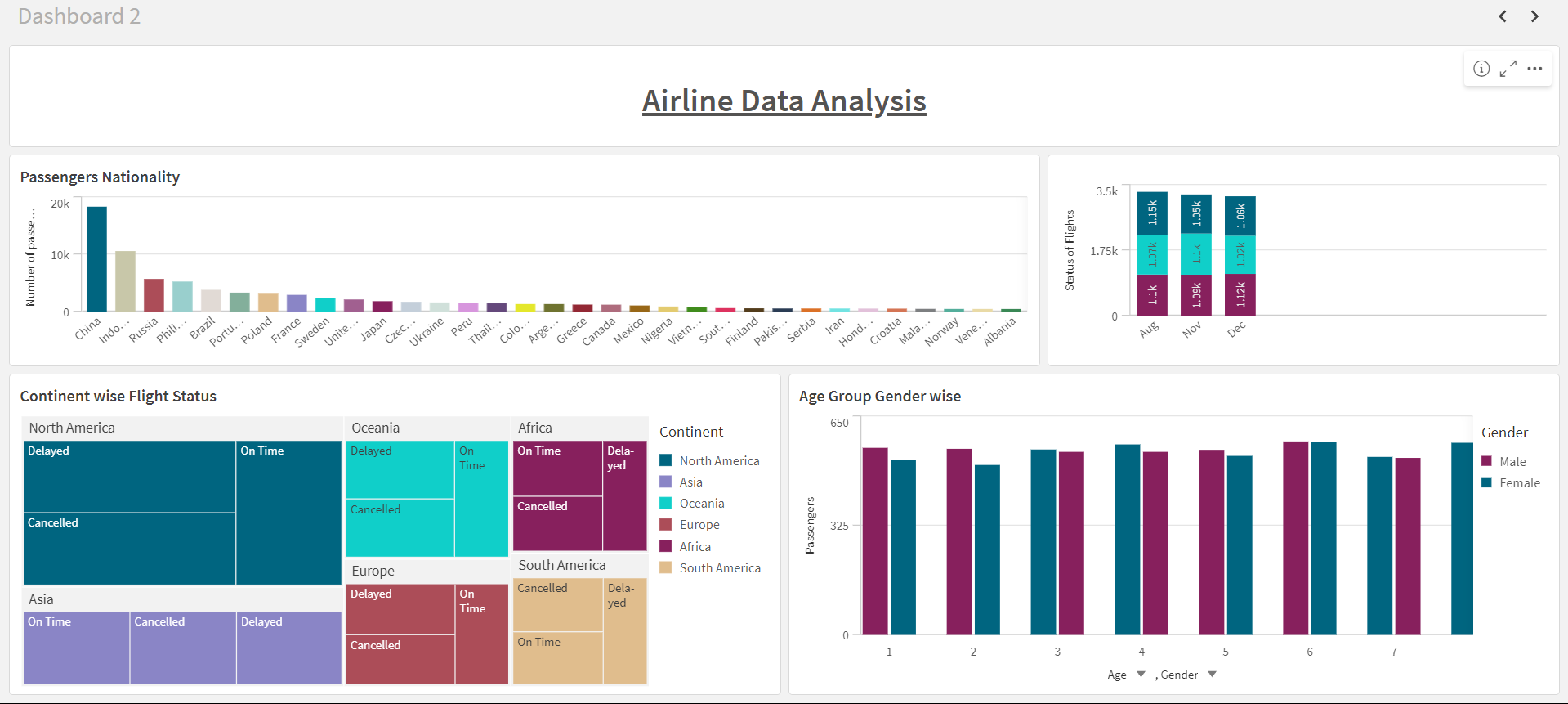


### Responsive And Design Of Dashboard

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.

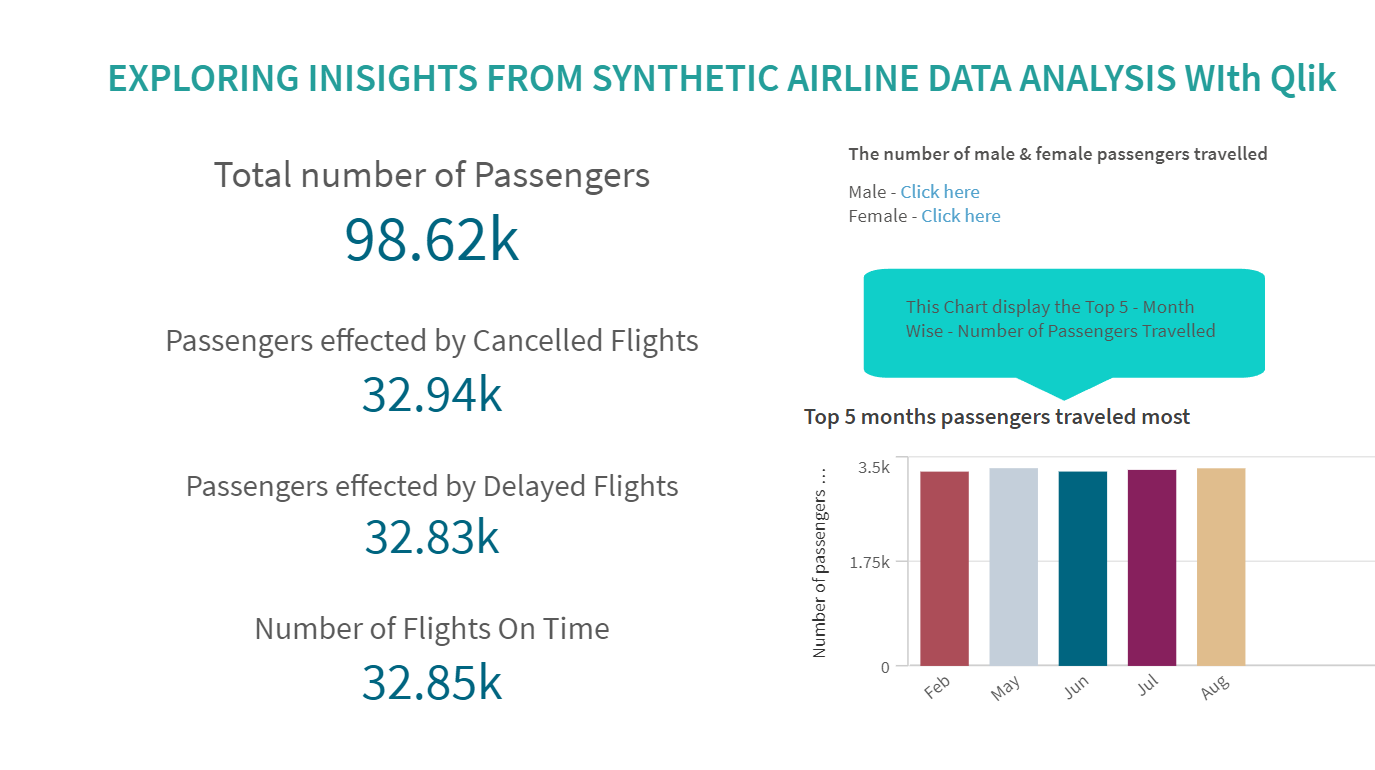
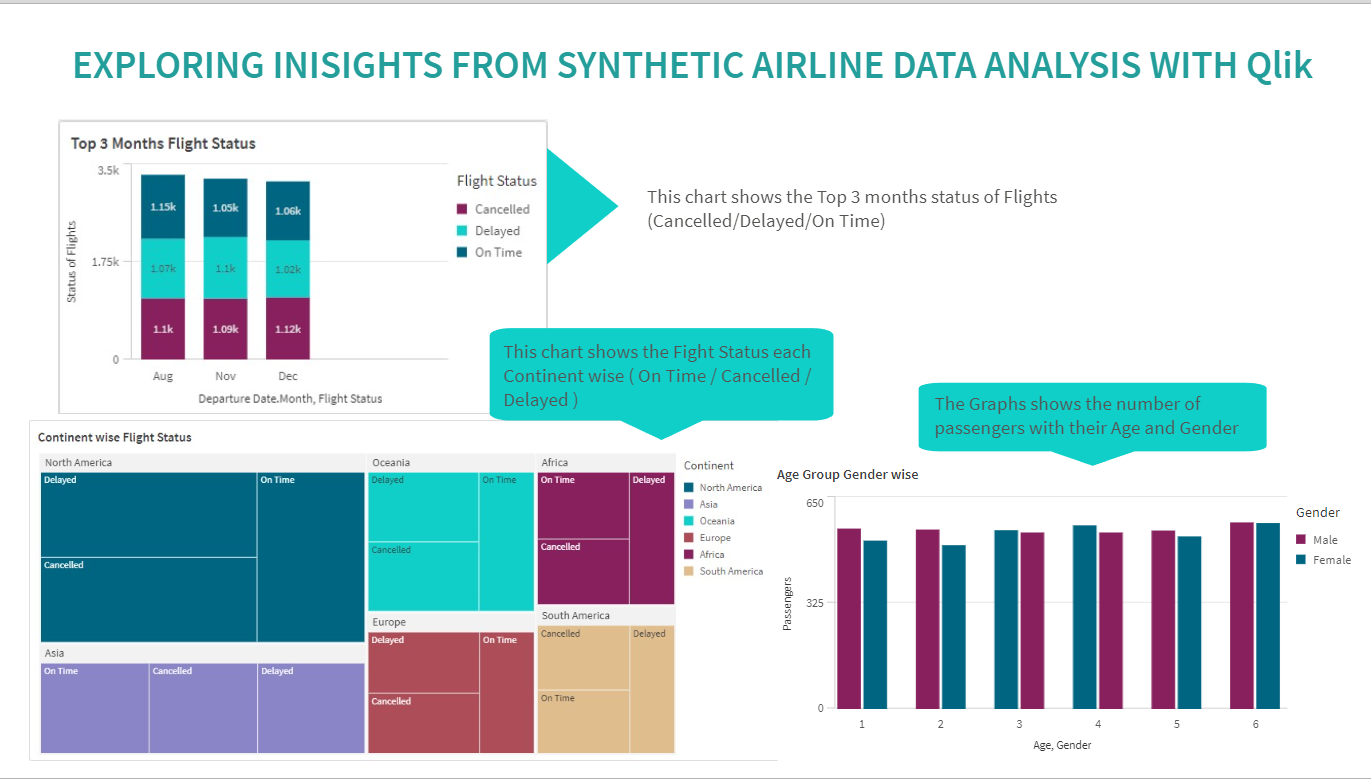
Dashboard 1:



Dashboard 2Story

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.

### Design Of Story



**Project Outcome:**

The Synthetic Airline Data Analysis project aims to provide valuable insights into airline operations, passenger behavior, and flight performance using Qlik's powerful analytics platform. By leveraging data-driven insights, airlines can optimize their operations, improve customer satisfaction, and enhance overall performance in a competitive industry landscape.

In conclusion, the Synthetic Airline Data Analysis project demonstrates the power of data analytics in transforming raw data into actionable insights that drive business growth and innovation within the aviation sector.

DATASET LINK -

<https://gtxq0jcw9nc9onj.sg.qlikcloud.com/sense/app/9ccfd40d-543d-47e6-b2af-0940a64cc0ab>

PRE-PROCESSED LINK-<https://gtxq0jcw9nc9onj.sg.qlikcloud.com/dataset/6654509769d4f4bef471fba7>

PROJECT DEMONSTRATION -

<https://drive.google.com/file/d/167pdwAUsYr-3FWhqPPfDUBQhNg4lY8dB/view?usp=sharing>