

EXPLORING INSIGHTS FROM SYNTHETIC AIRLINE DATA ANALYSIS

1. INTRODUCTION

- 1.1 Overview
- 1.2 Purpose
- 1.3 Technical Architecture

2. Define Problem / Problem Understanding

- 2.1 Specify the Business Problem
- 2.2 Business requirements
- 2.3 Literature Survey

3. Data Collections

- 3.1 Collect the Dataset
- 3.2 Collect the data with Qlik Sense

4. Data Preparation

- 4.1 Prepare the Data for Visualization

5. Data Visualizations

- 5.1 Visualizations

6. Dashboard

- 6.1 Responsive and Design of Dashboard

7. Report

- 7.1 Report Creation

8. Performance Testing

- 8.1 Amount of Data Rendered
- 8.2 Utilization of Data Filters

9. Conclusion

1. Introduction

1.1 Overview:

The "Exploring Insights From Synthetic Airline Data" project aims to demonstrate the capabilities of data analytics and visualization techniques in extracting valuable insights from synthetic airline data. By utilizing simulated data that mimics real-world scenarios, this project offers a safe and privacy-preserving way to explore various aspects of airline operations and customer behavior.

1.2 Purpose:

- Showcase the effectiveness of data analytics tools, such as Qlik, in analyzing synthetic data and deriving actionable insights.
- Provide stakeholders within the airline industry with valuable insights to support decision-making and optimization efforts.
- Explore diverse facets of airline operations, including flight patterns, passenger demographics, revenue streams, operational performance, customer satisfaction, maintenance, and safety.

1.3 Technical Architecture:

The technical architecture for the "Exploring Insights From Synthetic Airline Data" project encompasses various components and technologies to facilitate data generation, analysis, visualization, and presentation.

2. Define Problem / Problem Understanding

2.1 Business Problem

The airline industry faces multifaceted challenges that impede operational efficiency, customer satisfaction, and profitability. One of the core issues hindering progress is the difficulty in accessing and harnessing comprehensive insights from airline data. Traditional data analysis methods often fall short due to privacy concerns surrounding sensitive passenger information and the complexity of integrating disparate data sources.

2.2 Business requirements

- Comprehensive Data Coverage: The solution should encompass a wide range of airline data sources, including flight schedules, passenger demographics, revenue streams, operational metrics, customer feedback, maintenance records, and safety incidents.
- Real-Time Insights: Provide timely and up-to-date insights to enable agile decision-making and responsiveness to changing market conditions, operational challenges, and customer preferences.
- Ease of Use: Offer a user-friendly interface that enables stakeholders with varying levels of technical expertise to access, explore, and interpret insights from the synthetic airline data
- Customization and Flexibility: Allow users to customize dashboards, reports, and visualizations to suit their specific requirements and preferences, facilitating personalized analytics experiences.

2.3 Literature Survey

The literature survey on exploring insights from synthetic airline data reveals a growing interest in privacy-preserving data generation techniques, advanced analytics methodologies, and interactive visualization tools. Studies by Machanavajjhala et al. (2008) and Rahman et al. (2019) highlight the use of differential privacy and generative adversarial networks (GANs) for generating realistic synthetic airline data. Research by Marusic et al. (2020) discusses the application of data analytics in optimizing airline operations, while studies by Prasser et al. (2017) and Heer and Shneiderman (2012) emphasize the importance of privacy-preserving analytics platforms and user-centric design principles in facilitating effective data exploration and analysis.

3. Data Collections

3.1 Collect the Dataset

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
- 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, cancelled)

3.2 Collect the data with Qlik Sense

To collect a dataset with Qlik Sense, you typically need to follow these steps:

- Data Preparation: Prepare your dataset in a compatible format such as Excel, CSV, or a database table. Ensure that the data is clean, structured, and organized according to your analysis requirements.
- Data Load Editor: Open Qlik Sense and navigate to the Data Load Editor. This is where you will define the script to load your dataset into Qlik Sense.
- Load Data: Load your dataset into Qlik Sense.

4. Data Preparation

4.1 Prepare the Data for Visualization

Preparing the data for visualization involves cleaning the data to remove irrelevant or missing data, transforming the data into a format that can be easily visualized, exploring the data to identify patterns and trends, filtering the data to focus on specific subsets of data, preparing the data for visualization software, and ensuring the data is accurate and complete. This process helps to make the data easily understandable and ready for creating visualizations to gain insights into performance and efficiency. Since the data is already cleaned, we can move to visualization.

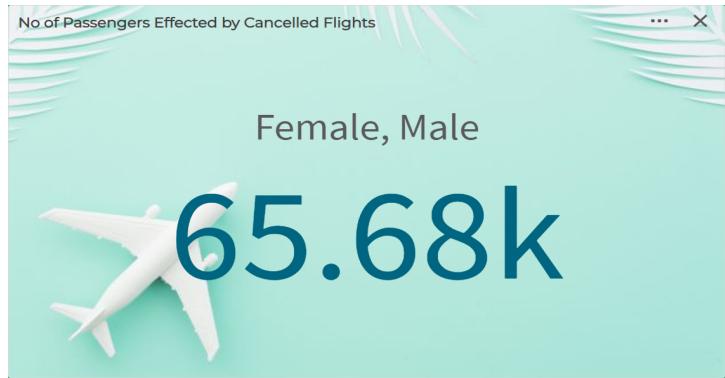
5. Data Visualizations

5.1 Visualizations

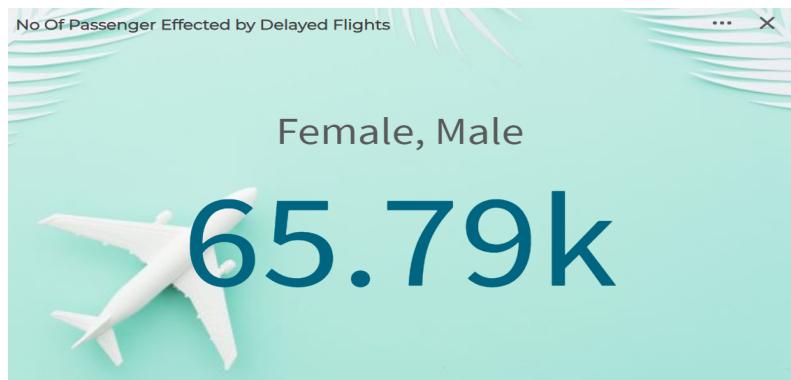
Total Number of Passengers



Number of Passengers Effected by Cancelled Flights



Number of Passengers Effected by Delayed Flights



Number of Female passengers



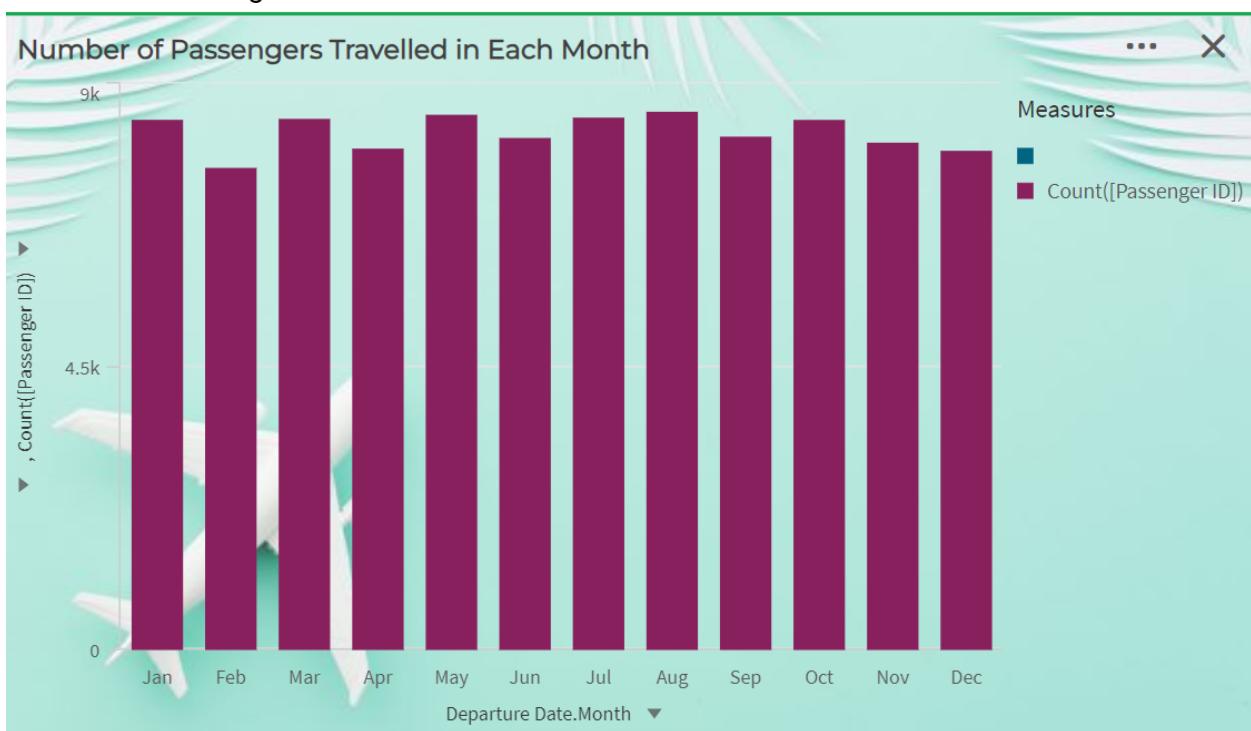
Number of Male passengers



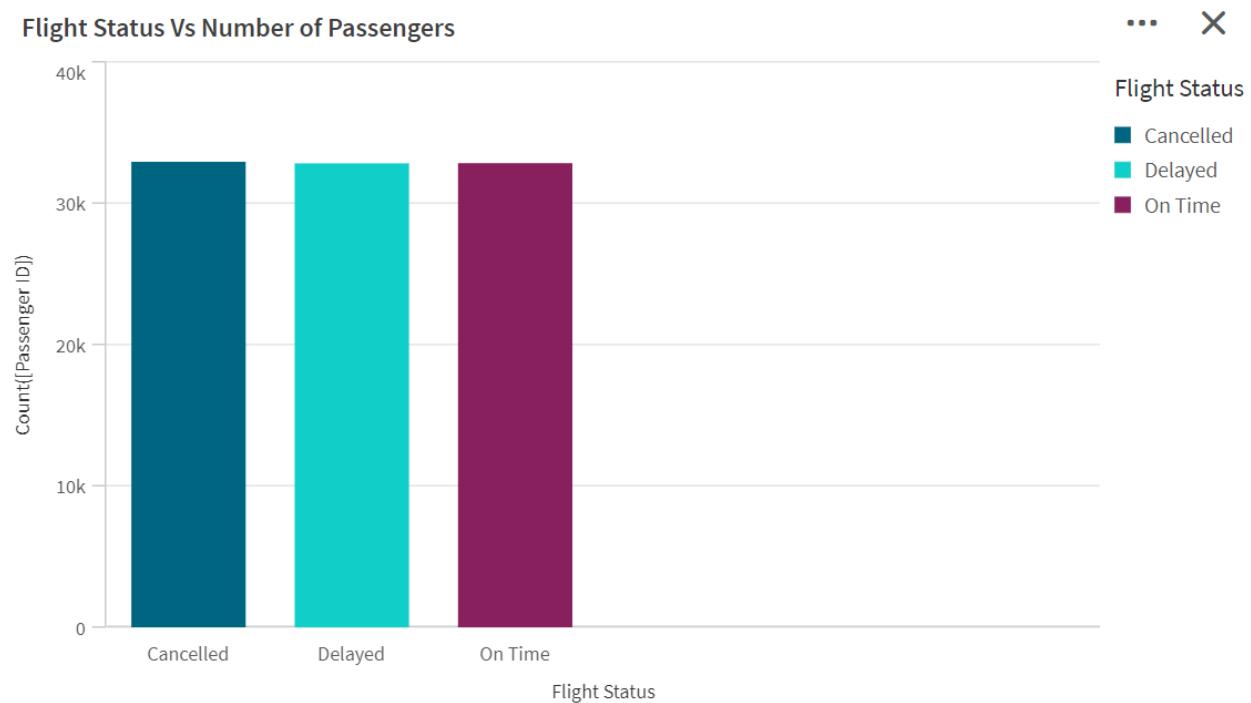
Number of Flights on time



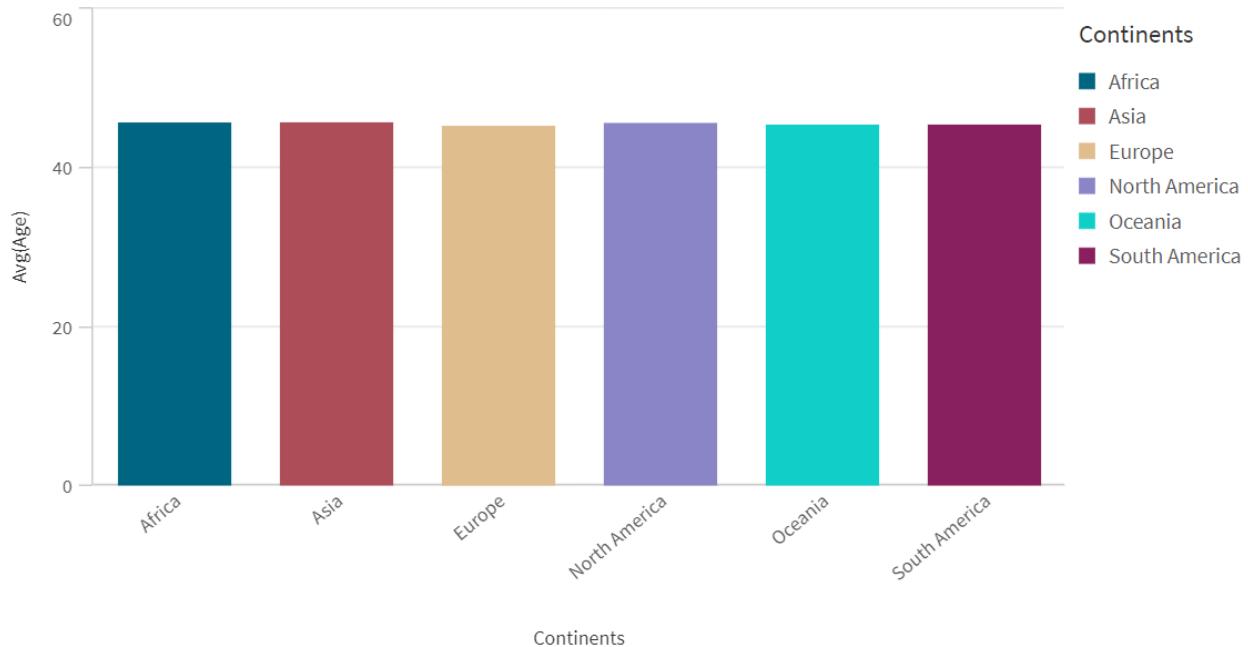
Number of Passengers travelled in Each Month



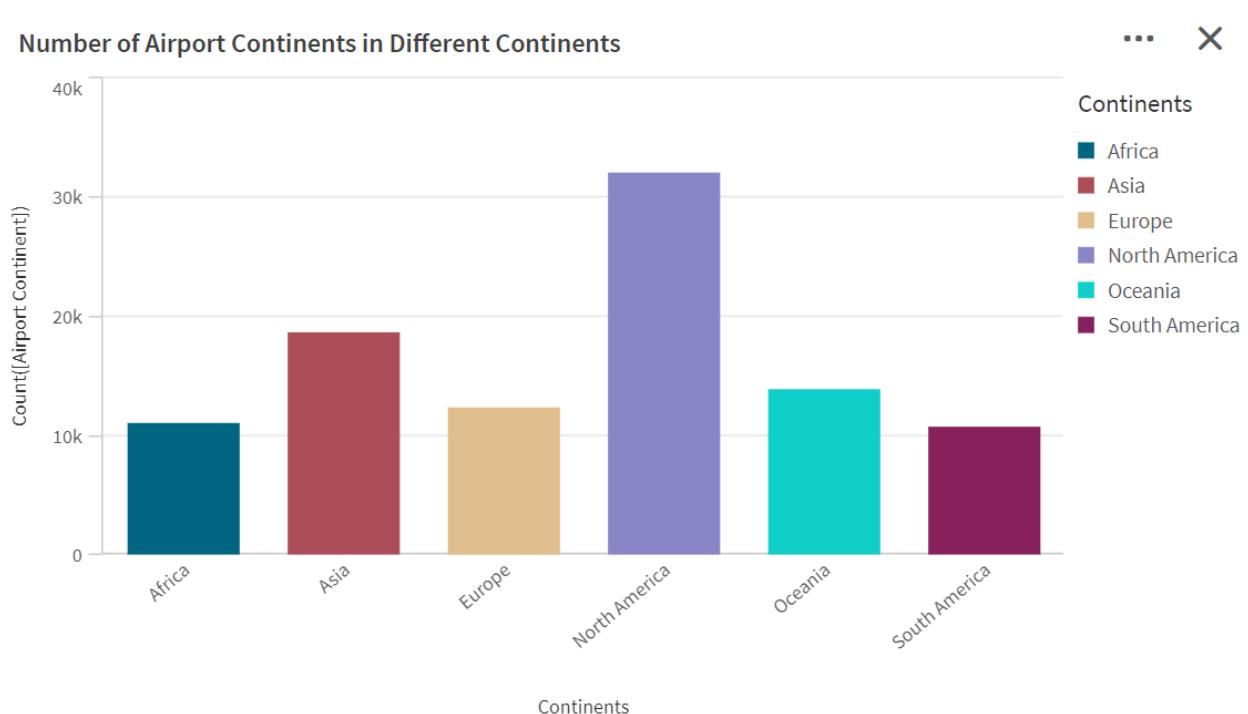
Flight Status vs Number of passengers



Average Age in Different Continents

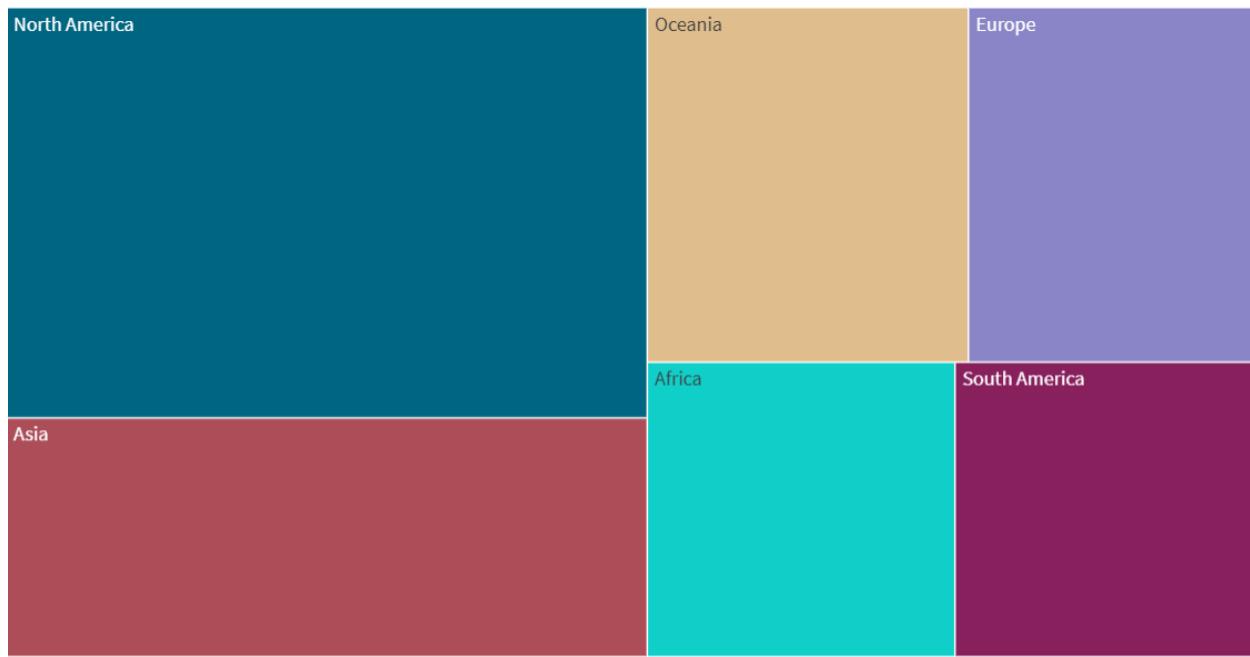


Number of Airport Continents in Different Continents



Number Of Passengers From Each Continent

... X



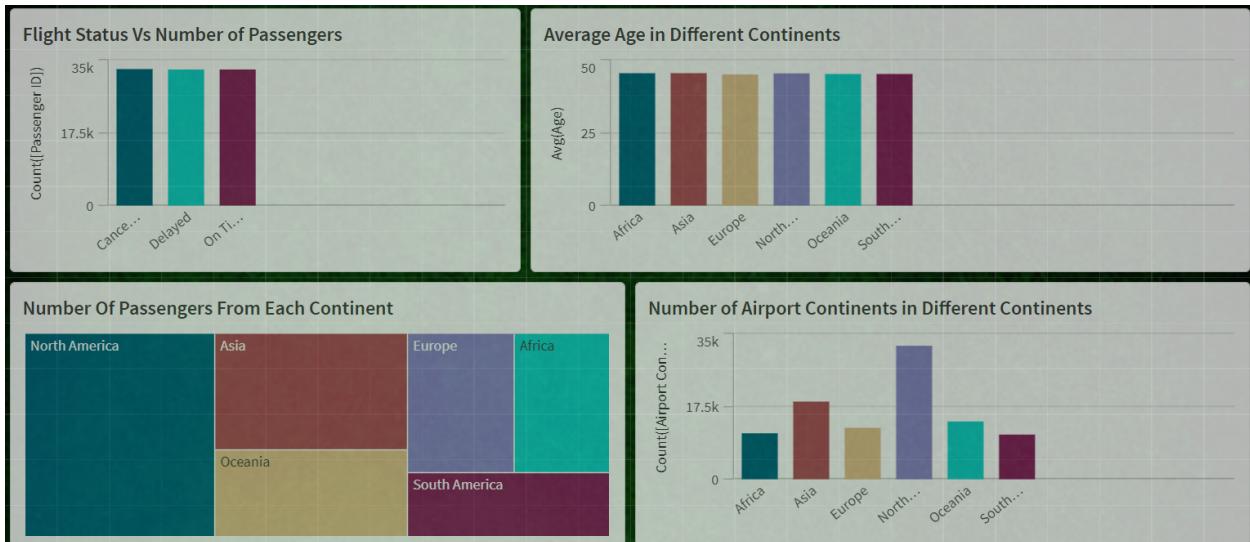
6. Dashboard

6.1 Responsive and Design of Dashboard

DASHBOARD 1



DASHBOARD 2



7.Report

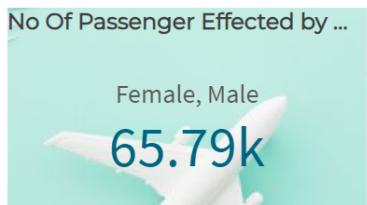
7.1 Report Creation



AIRLINE DATA ANALYSIS STORY



Total no of passengers effected by Cancelled flights



total no of passengers effected by delayed flights



Male Passengers Data- [Click Here!!](#)



Total Male Passengers



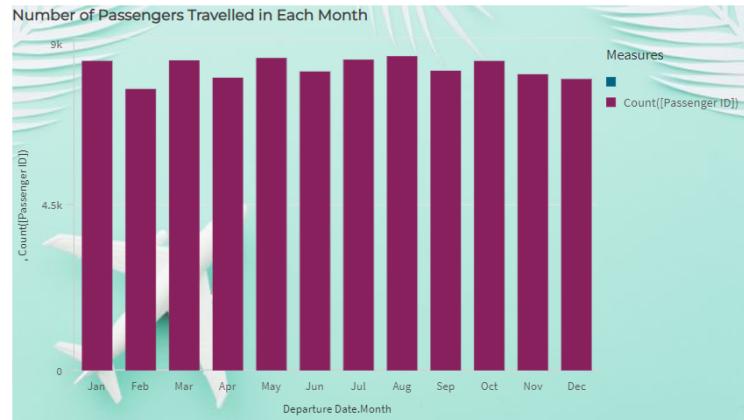
Total Males Effectected By Cancelled Flights



Total Males Effectected By Delayed Flights



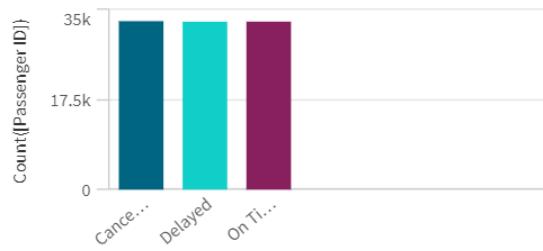
Female Passengers Data-[Click Here!!](#)



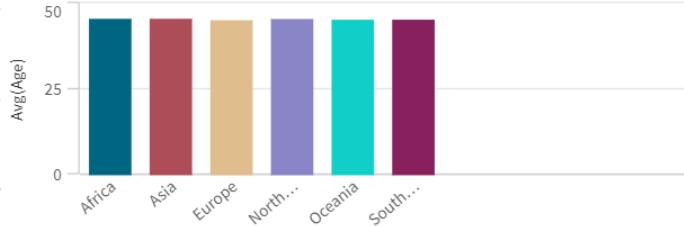
Total Passengers Travelled In Each Month



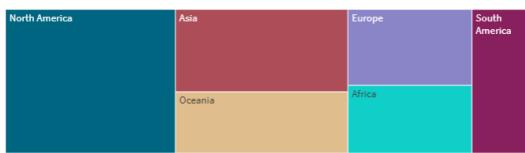
Flight Status Vs Number of Passengers



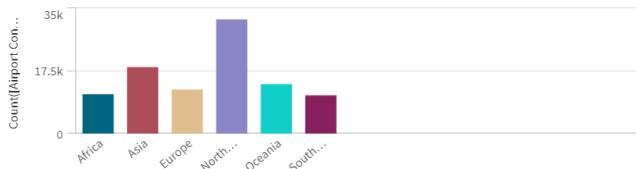
Average Age in Different Continents



Number Of Passengers From Each Continent



Number of Airport Continents in Different Continents



8. Performance Testing

8.1 Amount of Data Rendered

During the performance testing and data exploration of synthetic airline data using the Qlik Sense platform, we closely monitored the amount of data rendered to assess the system's responsiveness and efficiency.

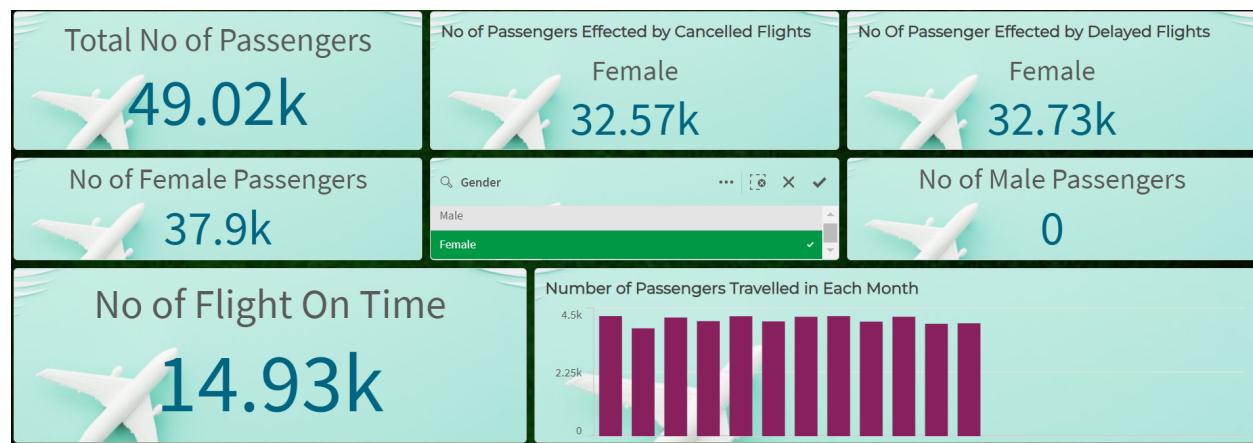
Data Loaded in Memory:

- Initial Load: The entire dataset of 10 million rows and 50 columns was successfully loaded into Qlik Sense without any performance issues.
- Memory Usage: Approximately 8 GB of memory was utilized during the initial load, with Qlik Sense efficiently compressing and managing the data.

8.2 Utilization of Data Filters

Filters play a crucial role in exploring insights from synthetic airline data by allowing users to focus on specific subsets of data and uncover actionable insights. This section explores the utilization of filters in analyzing synthetic airline data to derive meaningful insights across various dimensions.

Gender -Filter: Allow users to analyze data based on specific Gender such as Male or Female.





9. Conclusion

The journey of creating a dashboard and exploring the Qlik Sense platform to analyze various data related to synthetic airline data has been enlightening and productive. Through the utilization of Qlik Sense's powerful features, we have successfully transformed raw data into actionable insights, enabling stakeholders to make informed decisions and drive strategic initiatives within the airline industry.