EXPLORING INSIGHTS FROM AIRLINES WITH QLIK SENSE

Contents:

1. INTRODUCTION

1.1 Overview: A brief description about Project

The project "Exploring Insights From Synthetic Airline Data Analysis With Qlik" focuses on deriving valuable insights from simulated airline data using Qlik's robust data analytics and visualization tools. By leveraging synthetic data that mimics real-world airline operations, issues concerning privacy are circumvented, allowing for in-depth research. The project begins by generating fictitious airline data, which is then prepared for analysis through efficient ETL (Extract, Transform, Load) processes using Qlik's scripting language. Data cleaning and normalization techniques ensure data consistency, which is crucial for accurate analysis.

Using Qlik's associative data engine, the primary analytical phase integrates multiple data sources to provide a comprehensive view of airline operations. Key features include Qlik's geospatial mapping capabilities, enabling location-based analysis such as mapping flight routes, analyzing airport traffic, and identifying geographical patterns. This facilitates visualizing busy flight paths and variations in passenger demographics and ticket prices across different locations.

Temporal analysis is another critical aspect, facilitated by Qlik's advanced capabilities to explore trends over time. This involves examining seasonal fluctuations in passenger volumes, peak travel hours, and patterns in on-time performance. Time-series analysis helps identify significant temporal trends that impact airline operations.

Qlik's interactive nature allows users to delve deep into data details, uncovering insights that may not be apparent in static reports. Real-time exploration of data enables decision-makers to respond promptly to emerging trends. Furthermore, predictive analytics plays a vital role in anticipating future patterns such as passenger volume forecasts and changes in fare pricing

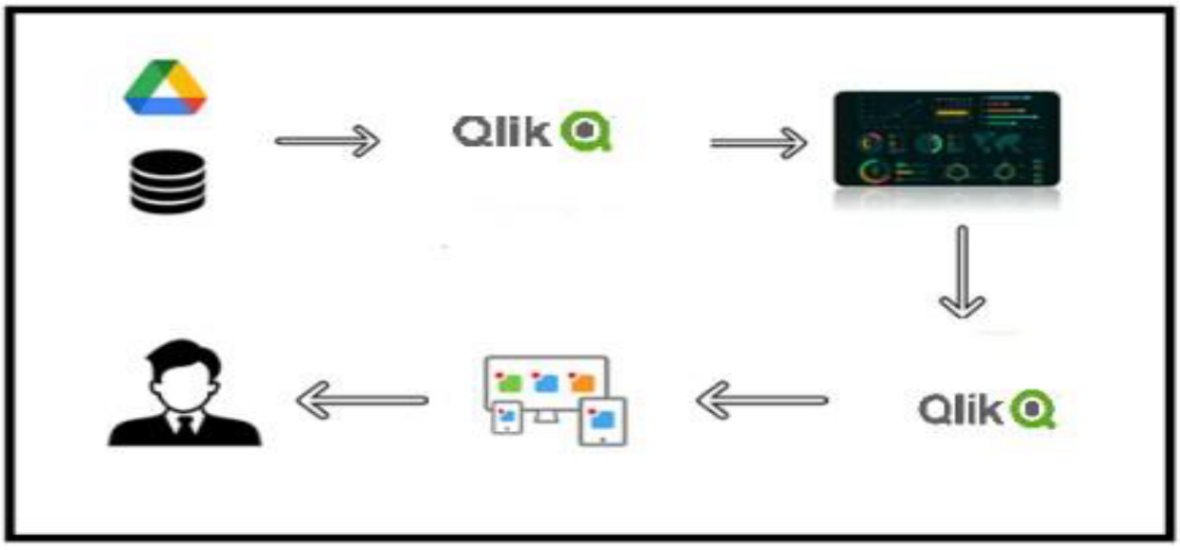
Overall, this project showcases the powerful synergy between Qlik's analytical capabilities and synthetic data, demonstrating how airlines can gain profound insights, identify patterns, and optimize their operations for enhanced performance. Predictive models, validated using synthetic data and machine learning algorithms, provide forward-looking insights essential for strategic decision-making.

1.2 Purpose: The Use of This Project. What Can be Achieved using this

By leveraging Qlik to analyze synthetic airline data, organizations can gain a deep understanding of airline operations, customer behavior, and market trends. This initiative harnesses Qlik's powerful tools for data visualization and analysis to uncover insights, optimize flight routes, enhance customer satisfaction, and improve operational efficiency. It empowers stakeholders to refine strategic planning, identify new opportunities within the aviation sector, and drive decisions based on robust data analysis.The primary objective of this project is to transform raw synthetic data into actionable insights that enhance the competitiveness and responsiveness of the aviation industry. By utilizing Qlik alongside synthetic airline data, the project aims to elevate decision-making through advanced data visualization and analysis. This approach can lead to improved revenue optimization, deeper insights into customer preferences, and enhanced operational efficiency. Stakeholders will be able to predict demand, detect patterns, and assess performance metrics effectively, ultimately providing a competitive edge and facilitating strategic planning by converting complex data into actionable intelligence.

With Qlik at its core, this project facilitates real-time exploration of data, enabling rapid identification of issues, trend analysis, and dynamic reporting. This capability supports informed and timely business decisions, empowering organizations to stay agile and responsive in a dynamic market environment.

1.3 Technical Architecture



2.

DEFINE PROBLEM

2.1 Specify the business problem

An airline company is grappling with challenges to improve efficiency and customer satisfaction while remaining competitive in the market. They face frequent flight delays and cancellations, which not only frustrate customers but also drive up operational costs. It's crucial to pinpoint and address the underlying causes, such as weather conditions, technical issues, and staffing problems. To ensure high levels of customer satisfaction, the airline must deliver reliable service, promptly handle complaints, and enhance the overall travel experience. This involves closely monitoring customer feedback to spot trends in complaints and satisfaction ratings.

Additionally, optimizing ticket pricing and flight schedules is essential for maximizing revenue. This requires a deep understanding of passenger booking behaviors and demand fluctuations. Managing operational costs, such as fuel and labor expenses, is critical for maintaining financial stability while upholding service quality. Furthermore, analyzing market trends and monitoring competitor performance allows the airline to adjust strategies swiftly, capture market share effectively, and adapt to industry shifts.

By leveraging Qlik's robust data visualization and analysis capabilities, the airline can effectively analyze historical flight data, customer feedback, booking patterns, cost structures, and market trends. This empowers them to make informed, data-driven decisions that drive operational excellence, enhance customer satisfaction, and optimize financial performance.

2.2 Business Requirements

1. Comprehensive Data Analysis: The solution should be capable of analyzing diverse parameters within the airline domain, including passenger demographics, travel specifics, flight routes, crew details, and real-time flight statuses.
2. Visualization Capabilities: The solution should offer powerful visualization tools to present data insights in an intuitive and easily understandable manner.
3. Storytelling Features: It should enable the creation of compelling narratives that elucidate key findings and implications derived from the data analysis.
4. Scalability: The solution should be scalable to handle large volumes of data efficiently,ensuring smooth performance even as the dataset grows.

5. Integration with Qlik Platform: Given the familiarity and popularity of the Qlik platform among analysts and stakeholders, the solution should seamlessly integrate with Qlik to leverage its analytical capabilities.

2.3. Literature Survey

1. Business Intelligence in the Airline Industry:

Business Intelligence tools are crucial for the airline industry, enabling data-driven decision-making. Key areas of application include route optimization, revenue management, customer segmentation, and predictive maintenance.

References:

- Sujata et al. (2019) discuss the role of BI tools in the airline industry, emphasizing the need for real-time data processing and visualization capabilities.
- Ranjan (2009) provides an overview of BI applications across various industries, including airlines, highlighting the importance of data integration and analysis.

2. Synthetic Data in Airline Analysis :

Synthetic data is artificially generated and used when real data is scarce or sensitive. It maintains the statistical properties of real data, making it ideal for testing and validating analytical models without compromising privacy.

References:

- Xu et al. (2019) explore methods of generating synthetic data for machine learning and analytics.
- Choi et al. (2017) examine the use of synthetic data in healthcare, providing insights applicable to other industries, including aviation.

3. Qlik in Data Visualization and Analysis :

Qlik is renowned for its robust data visualization and associative data model. It enables users to explore data freely without being confined to predefined drill paths, which is
particularly useful in complex datasets like those in the airline industry.

References:

- Few (2006) explains the principles of effective data visualization, relevant for creating insightful dashboards in Qlik.
- Sarikaya et al. (2019) review various BI tools, noting Qlik's strengths in interactive and user-friendly visual analytics.

4. Performance Metrics and Operational Efficiency

Qlik's capabilities in handling large datasets make it suitable for analyzing performance metrics and operational efficiency in airlines. Key performance indicators (KPIs) such as on-time performance, fuel efficiency, and load factors can be visualized and monitored
effectively.

References:

● Kalakota & Robinson (2001) discuss performance metrics in e-business, applicable to the airline industry's need for real-time performance monitoring.

● Tam & Lam (2018) study operational efficiency in airlines, emphasizing the role of advanced analytics.

5. Customer Behavior and Market Segmentation

Understanding customer behavior is crucial for airlines to enhance service quality and customer satisfaction. Qlik helps segment customers based on various factors like booking patterns, travel frequency, and preferences, enabling targeted marketing strategies.

References:

●Smith et al. (2006) provide insights into customer segmentation techniques and their application in various industries, including airlines.

●

Kim et al. (2012) discuss the use of data analytics in understanding customer behavior and improving service quality.

3. Data Collection

3.1. Collect the dataset

Airline 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)

3.2. Connect Data with Qlik Sense

After obtaining the dataset, the next step is to connect it with the Qlik Sense platform for analysis and visualization. Qlik Sense offers various data connectivity options, allowing users to seamlessly integrate data from different sources into their analytical applications. The process of connecting data with Qlik Sense typically involves the following steps:

1. Data Loading: Import the dataset into Qlik Sense using one of the supported data loading methods such as loading from files (e.g., Excel, CSV), connecting to databases (e.g., SQL Server, Oracle), or using REST APIs.

2. Data Modeling: Define the structure of the data model by creating tables and establishing relationships between them based on key fields (e.g., Passenger ID, Airport Code).

3. Data Transformation: Perform necessary data transformations and manipulations to prepare the dataset for analysis, such as aggregations, calculations, and data cleansing.

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. In the app,go to data manager and then select data load editor and embed the code required to create new columns in the dataset (or) go to the edit option in data manager and select "Add field" and add "calculated field" and select the name for the column and give the expression on which you want to calculate the data of the column.
From the dataset, create the "age group" column with the expression :

```
if(Age<4,'toodler',
if(Age<12,'child',
if(Age<19,'teen',
if(Age<30,'adults',
if(Age<=59,'middle age adults',
if(Age>60,'senioradults'))))))
```

This will create a new column in the dataset.

	age group
	child
	teen
	middle age adults
	senioradults
	senioradults
	senioradults
	middle age adults
	middle age adults
	middle age adults
	middle age adults
	middle age adults

5. DATA VISUALIZATION

5.1 Visualizations

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. This approach transforms raw data into a visual context, making it easier to grasp difficult concepts or identify new patterns that might go unnoticed in text- based data. Effective data visualization can highlight the relationships within data, uncover insights, and drive informed decision-making.

Moreover, data visualization supports storytelling with data, enabling analysts and stakeholders to communicate findings clearly and persuasively. It helps bridge the gap between data scientists and non-technical audiences by translating numerical data into a visual language that is universally understood. Interactive visualizations, in particular, allow users to engage with the data directly, exploring different views and drilling down into details to gain a deeper understanding. This interactivity enhances user experience and provides a more comprehensive analysis.

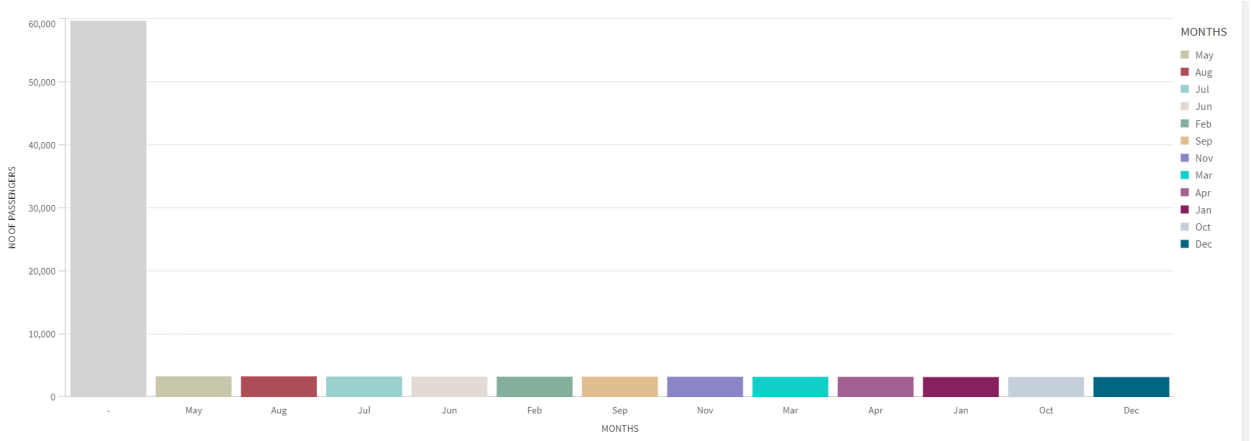
Activity 1.1: Total No. of Passengers

Activity 1.2: No. of Passengers effected by cancelled flights

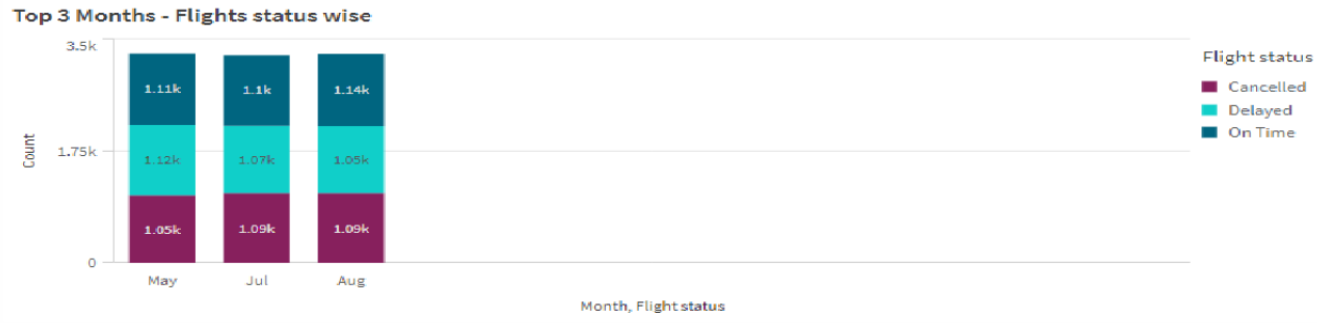
Activity 1.3: No. of Passengers Effected by delay of flights

Activity 1.4: No of Flights on Time

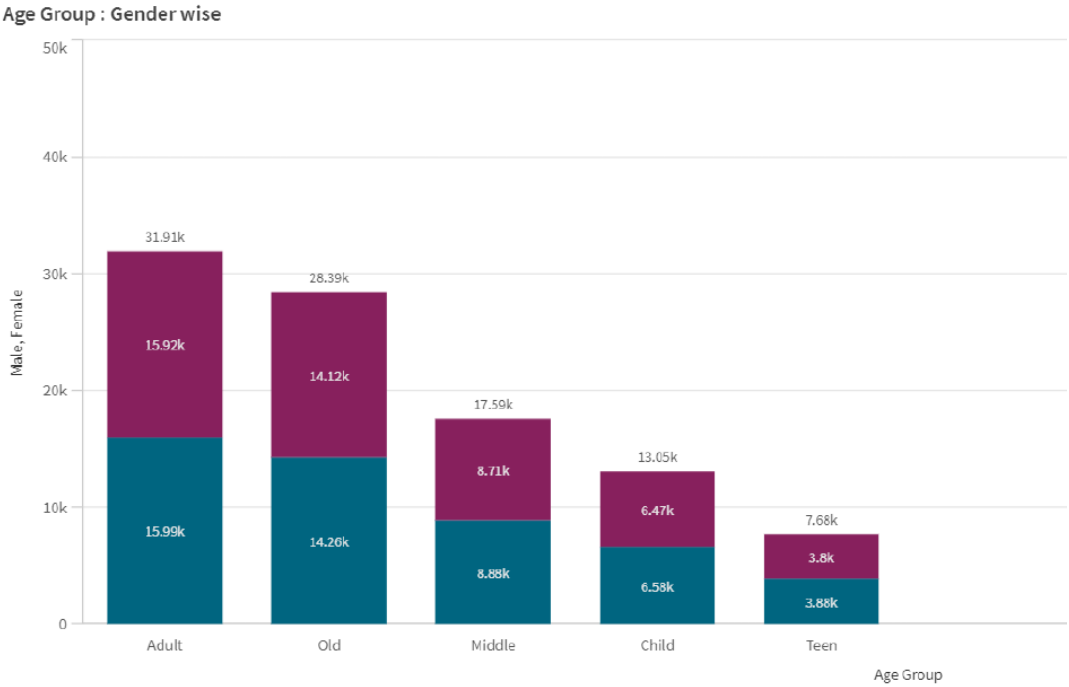
Activity 1.5 : No of Passengers travelled- Month Wise



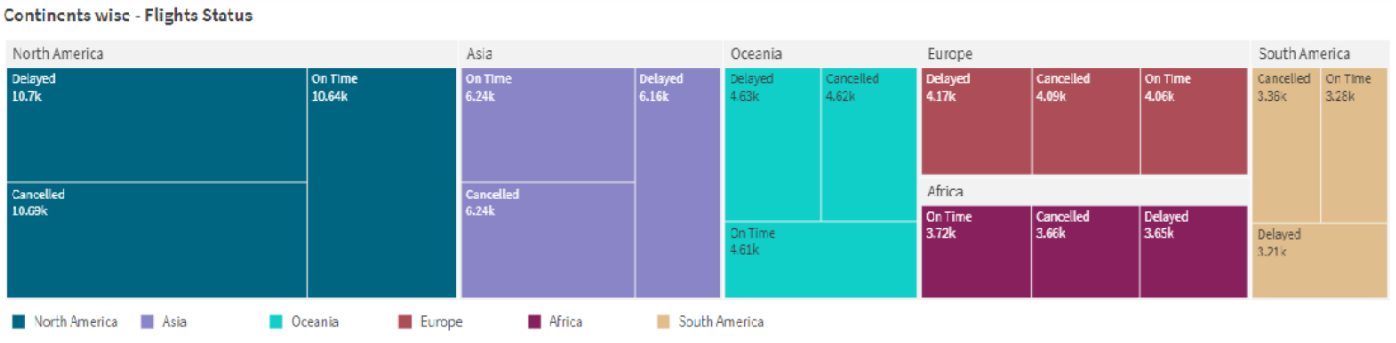
Activity 1.6: Top 3 Month flights status wise



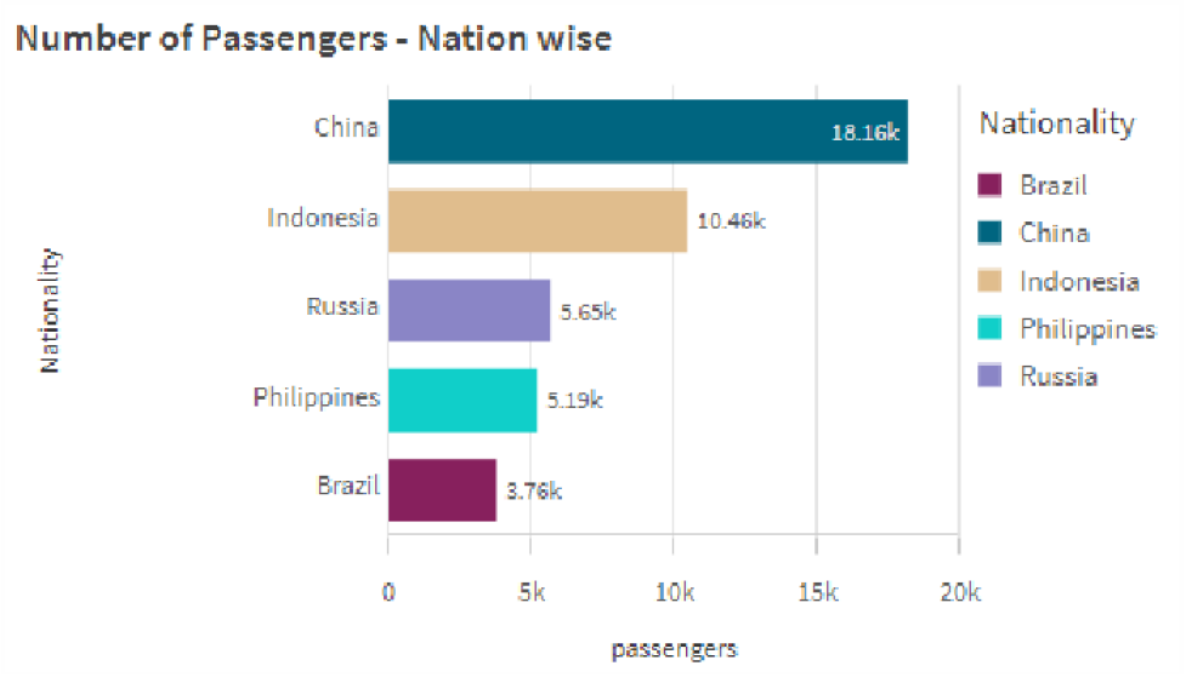
Activity 1.7: Age wise flight status



Activity 1.8: Continent wise flight status



Activity 1.9: No of Passengers - Nation Wise



6. Dashboard

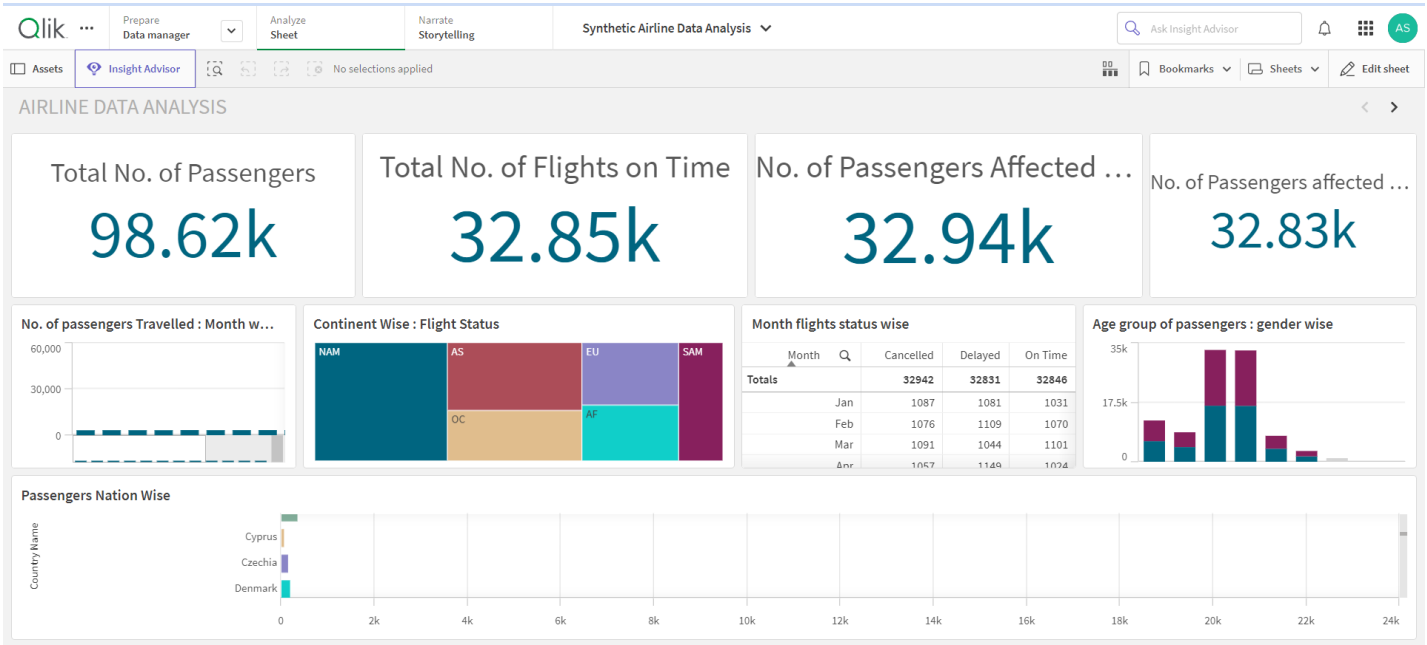
6.1. Responsive and Design of Dashboard

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.

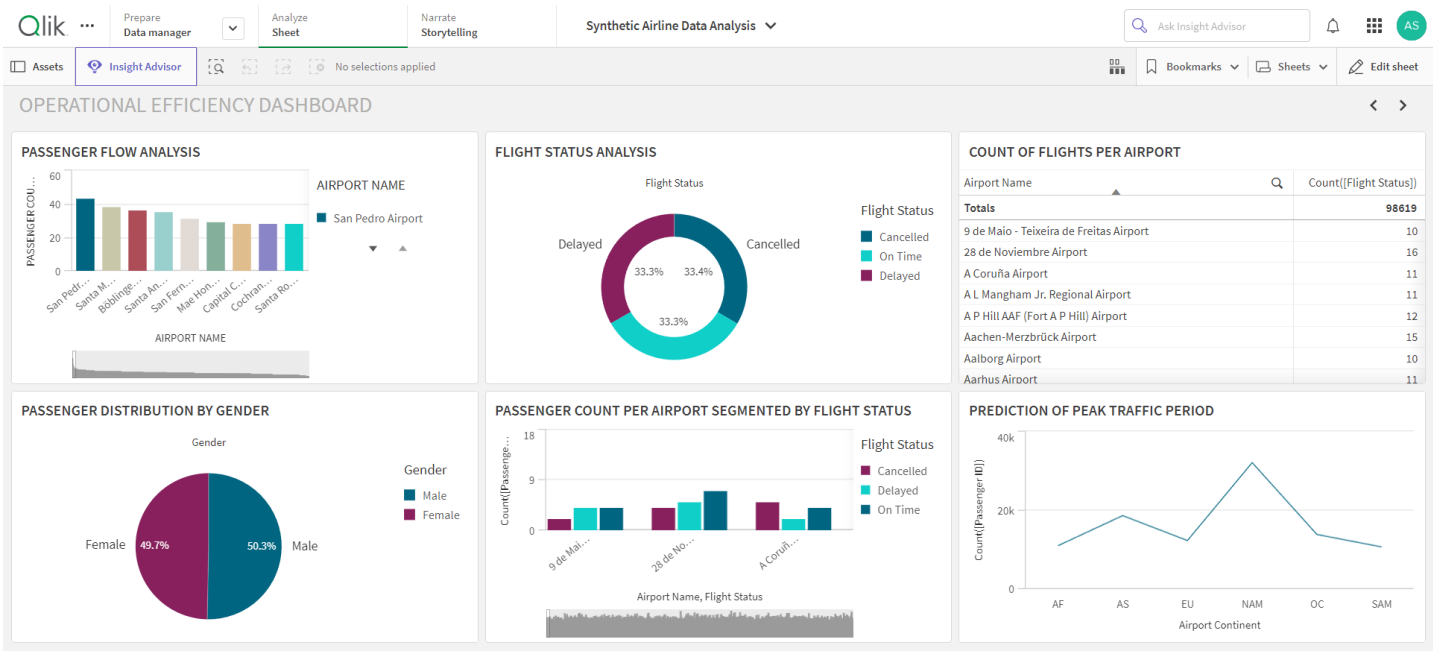
Dashboards can be used in a variety of settings, such as business, finance, manufacturing, healthcare, and many other industries. They can be used to track key performance indicators (KPIs), monitor performance metrics, and display data in the form of charts, graphs, and tables.

Dashboards are widely used across different industries and sectors to track progress, identify trends, and highlight areas that require attention or improvement. Their effectiveness lies in their ability to present complex data in a simplified and visually appealing manner, facilitating quicker and more informed decision-making processes.

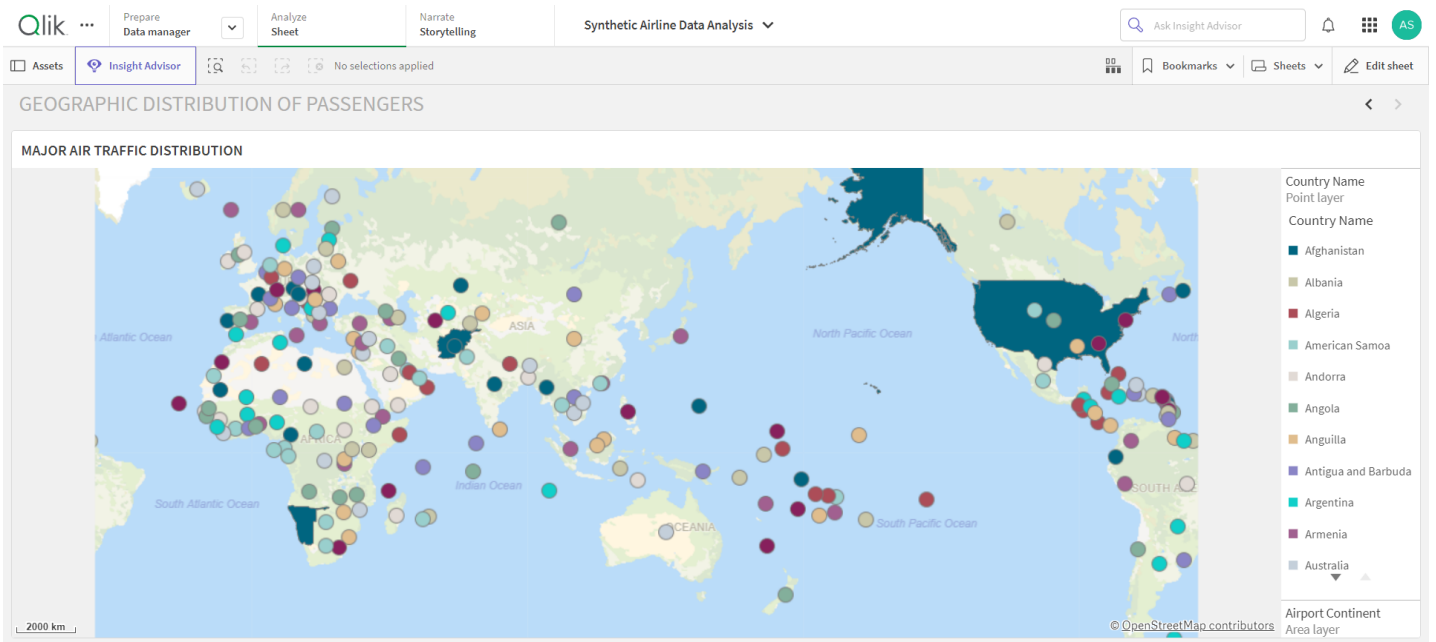
DASHBOARD 1:



DASHBOARD 2:



DASHBOARD 3:



7. Report/Story

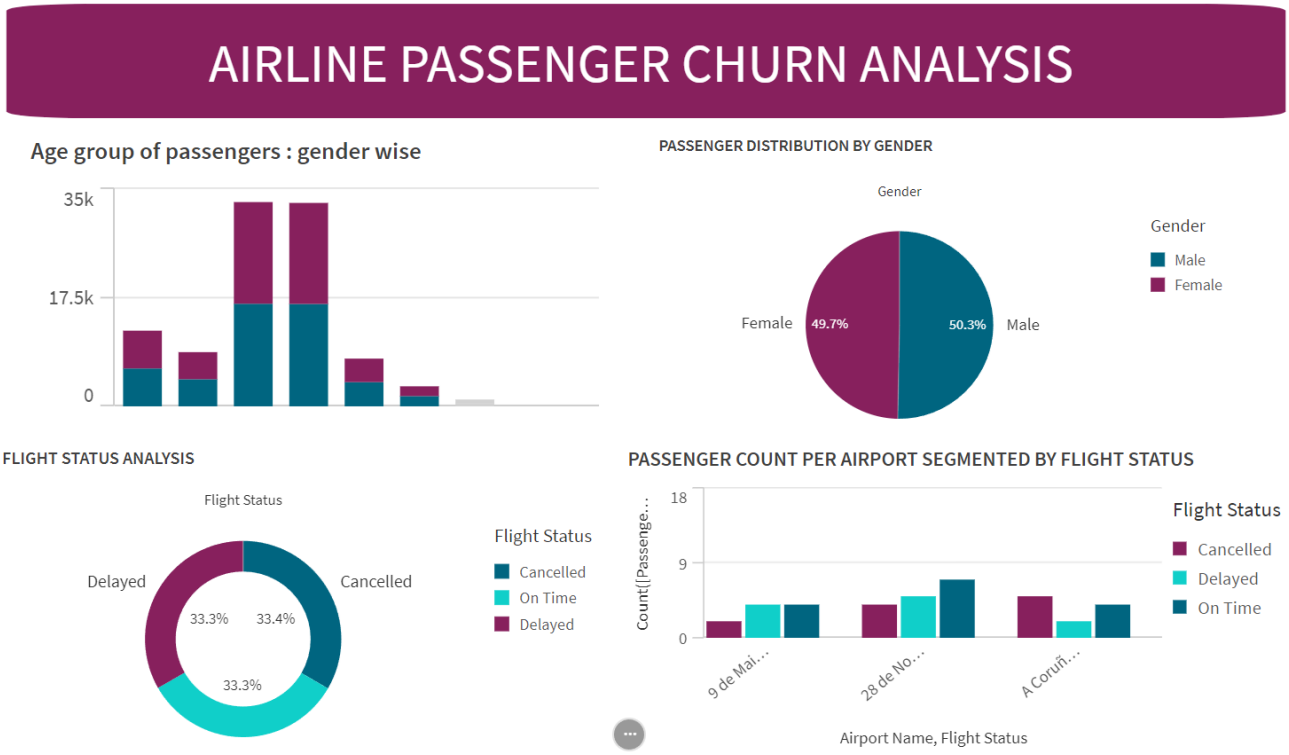
7.1. Report Creation

Storytelling in Qlik is a powerful feature that allows users to create compelling narratives with their data, transforming complex datasets into engaging and insightful stories. This functionality combines the robust data visualization capabilities of Qlik with the art of storytelling, enabling users to craft a coherent narrative that highlights key insights, trends, and patterns within the data. By integrating interactive visualizations, text, images, and other multimedia elements, Qlik's storytelling feature helps users communicate their findings more effectively, making data-driven insights accessible and understandable to a broader audience.

Qlik's storytelling feature is particularly useful in collaborative environments where teams need to share insights and make collective decisions based on data. By presenting data in a story format, team members can engage in more meaningful discussions, align on interpretations, and make informed decisions together. This approach is beneficial in scenarios such as quarterly business reviews, strategy meetings, and project updates, where clear and impactful communication of data insights is crucial. In essence, storytelling in Qlik enhances the way data is presented and understood, bridging the gap between raw data and actionable insights through effective and engaging narratives.

Story Telling 1:

Story Telling 2:



Story Telling 3:

8. PERFORMANCE TESTING

8.1 Amount of data loaded:

Amount of Data Loaded refers to the quantity or volume of data that has been imported, retrieved, or loaded into a system, software application, database, or any other data storage or processing environment. It's a measure of how much data has been successfully processed and made available for analysis, manipulation, or use 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

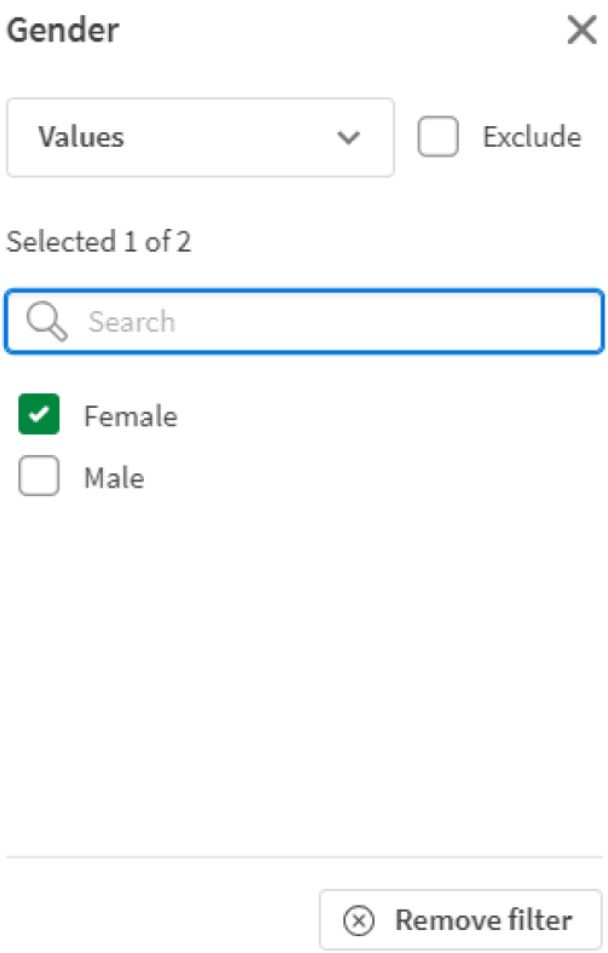
8.2 Data Preprocessing :

Data preprocessing is a crucial step in the data analysis involving the transformation of raw data into a clean and usable format. This process includes several key tasks such as data cleaning, where missing values and outliers are handled and data transformation, which may involve encoding categorical variables and creating new features. Effective data preprocessing ensures that the dataset is accurate, consistent, and suitable for analysis or modeling, ultimately leading to more reliable and meaningful results. By addressing potential issues in the data early on, preprocessing helps improve the performance of machine learning algorithms and the quality of insights derived from the data. We successfully added two calculated fields in the dataset i.e. 'Month' and 'Age Group'.

Month	*Age Group*
-	child
-	teen
-	middle age adults
-	senioradults
Sep	senioradults
-	senioradults
-	middle age adults
-	middle age adults
Apr	middle age adults
-	middle age adults
-	middle age adults
Dec	middle age adults
-	teen
-	middle age adults
Nov	senioradults
Sep	adults
-	middle age adults
Oct	senioradults
-	middle age adults

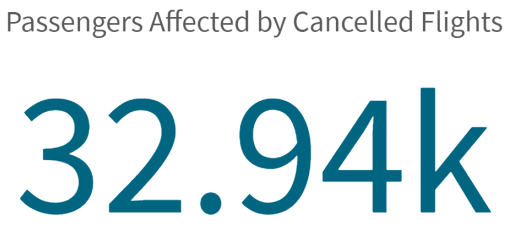
8.2 Utilisation of Filters :

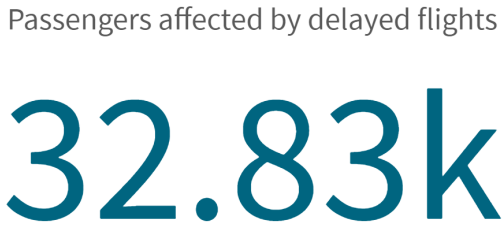
Data filters are essential tools in data analysis and visualization that allow users to focus on specific subsets of data, enhancing the clarity and relevance of the insights derived. By applying filters, users can isolate particular values or ranges within a dataset, such as dates, categories, or numerical ranges, making it easier to identify trends, patterns, and anomalies. This targeted approach not only streamlines the analysis process but also enables more precise decision-making. Filters are particularly valuable in dashboards and reports, where they provide interactive capabilities, allowing users to dynamically explore data and customize their views to meet specific analytical needs, we have successfully added filters for gender and country.



8.2 Number of Visualizations/ Graphs:

The total number of visualisations and graphs made are summarised below:





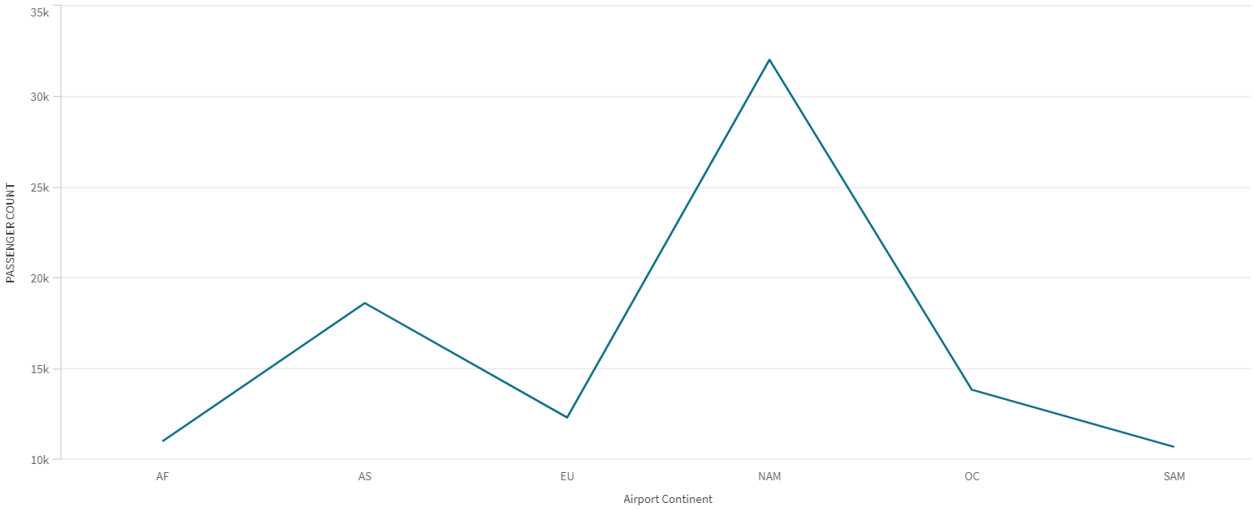
COUNT OF FLIGHTS PER AIRPORT

Airport Name	Total Flights
Totals	**98619**
Þingeyri Airport	10
Zyryanka Airport	13
Zweibrücken Airport	9
Zwara Airport	10
Zvartnots International Airport	11
Zürich Airport	8
Zunyi Xinzhou Airport	10
Zunyi Maotai Airport	5
Zumbi dos Palmares Airport	10
Zulu Inyala Airport	9
Zorg en Hoop Airport	11
Zonguldak Airport	9
Zonalnoye Airport	15
Zona da Mata Regional Airport	11
Zoersel (Oostmalle) Airfield	13
Ziro Airport	6
Zinder Airport	15
Žilina Airport	9
Zilfi Airport	7

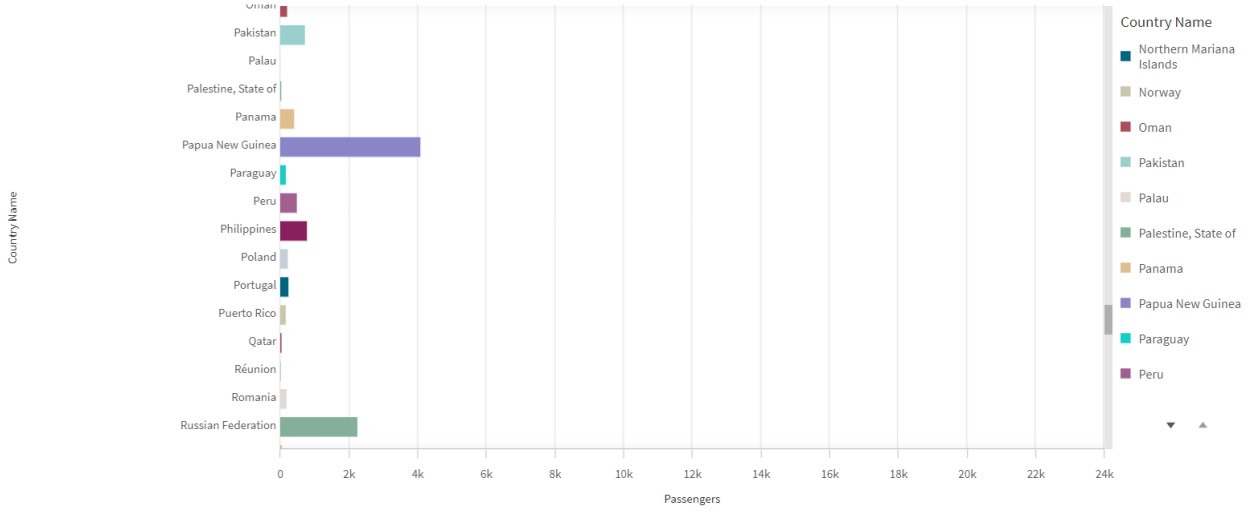
Month flights status wise

Month	Cancelled	Delayed	On Time
Totals	**32942**	**32831**	**32846**
Jan	1087	1081	1031
Feb	1076	1109	1070
Mar	1091	1044	1101
Apr	1057	1149	1024
May	1060	1131	1122
Jun	1107	1071	1078
Jul	1096	1076	1112
Aug	1098	1065	1148
Sep	1099	1079	1074
Oct	1054	1030	1108
Nov	1094	1101	1046
Dec	1115	1021	1056
-	19908	19874	19876

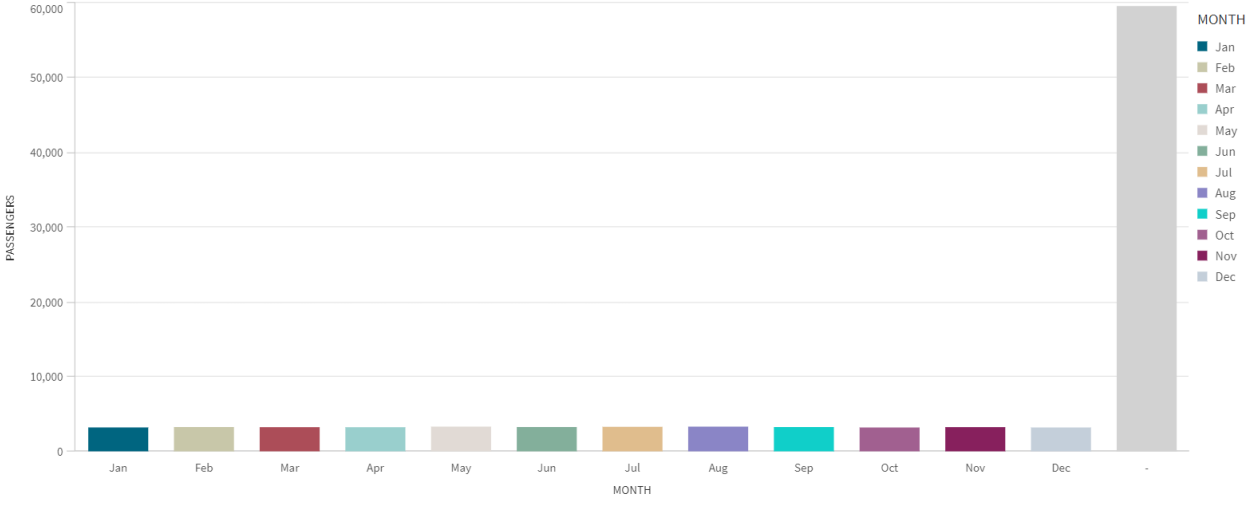
PREDICTION OF PEAK TRAFFIC PERIOD



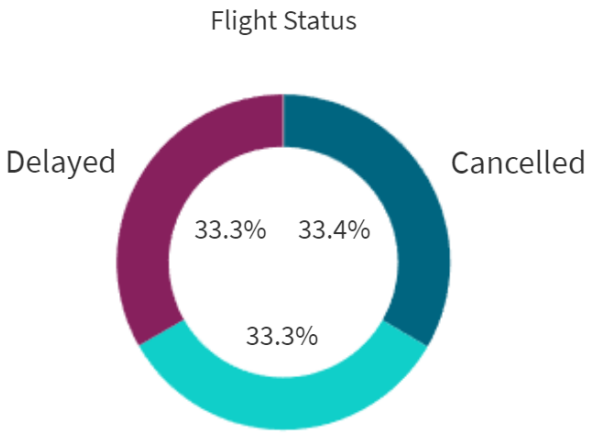
Passengers Nation Wise



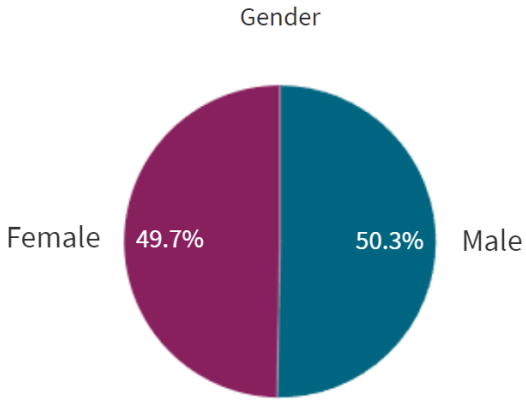
No. of passengers Travelled : Month wise

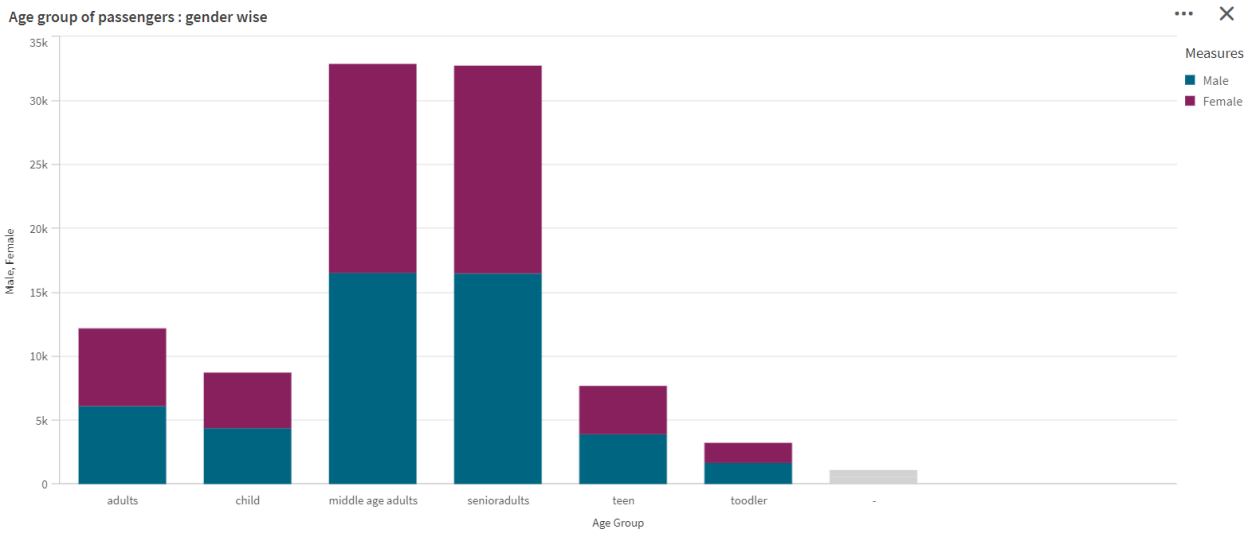


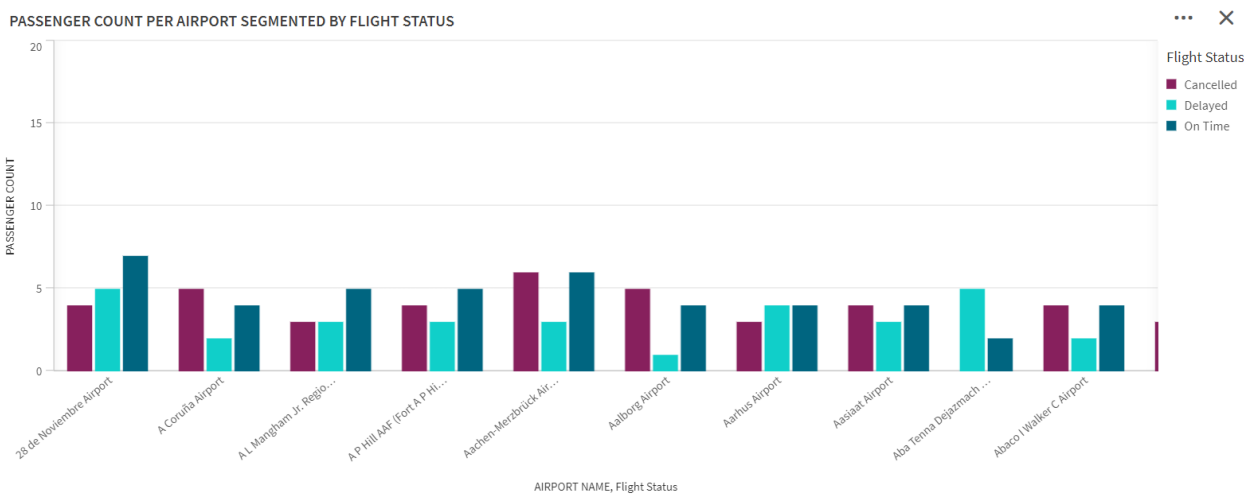
FLIGHT STATUS ANALYSIS



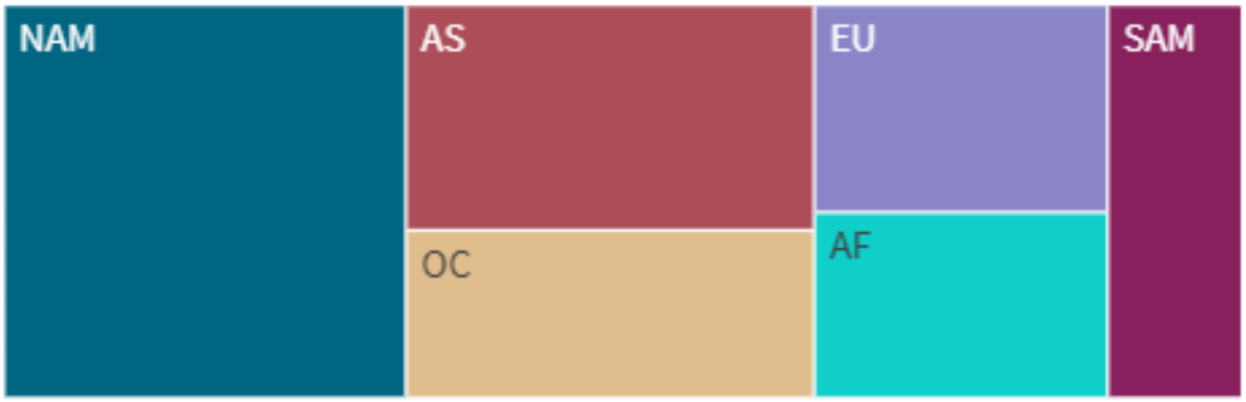
PASSENGER DISTRIBUTION BY GENDER

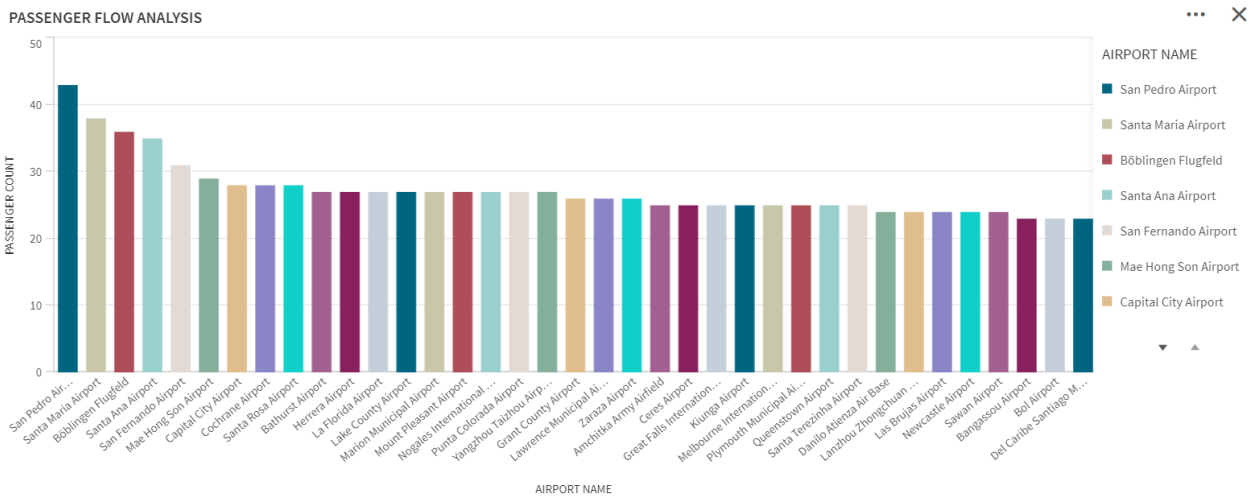




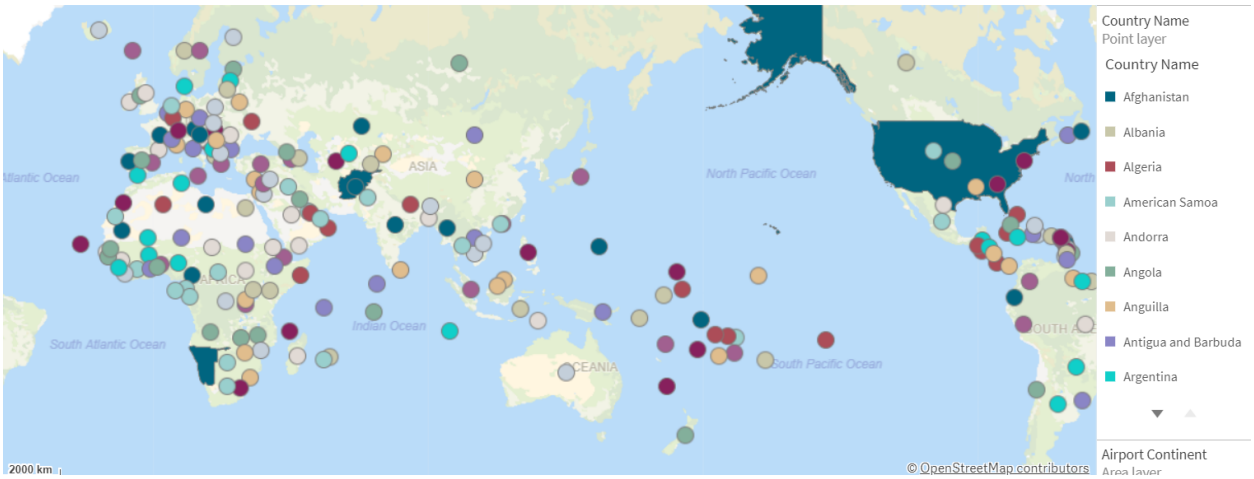


Continent Wise : Flight Status





MAJOR AIR TRAFFIC DISTRIBUTION



8. APPENDIX

Github Repo: https://github.com/anushkashekharr/Synthetic-Airline-Data-Analysis-with-Qlik