Exploring Insights from Synthetic Airline Data Analysis with Qlik

1. INTRODUCTION

1.1 Overview:

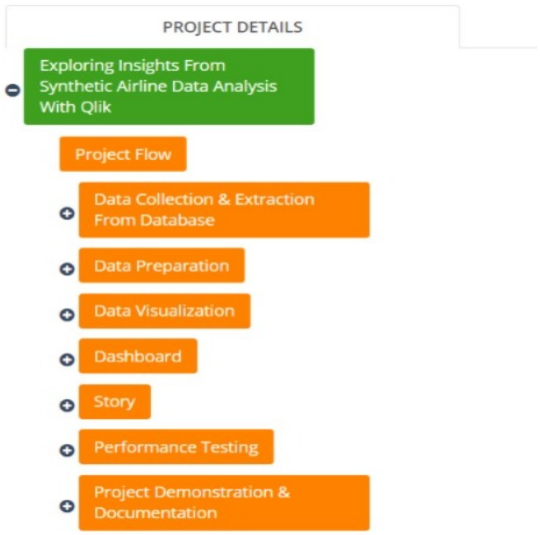
This project, "Exploring Insights from Synthetic Airline Data Analysis with Qlik," utilizes synthetic airline data to derive valuable insights and support decision-making for airlines, airports, and related stakeholders. The data encompasses various aspects of airline operations, including flight schedules, passenger demographics, ticket sales, and performance metrics. By leveraging Qlik's powerful analytical and visualization capabilities, we aim to uncover patterns, trends, and correlations within this data to address key business challenges.

1.2 Purpose:

The primary purpose of this project is to demonstrate how Qlik can be used to analyze and visualize synthetic airline data to achieve specific business objectives:

● Revenue Optimization: Analyzing historical ticket sales to identify peak travel times, popular destinations, and effective pricing strategies.
● Operational Efficiency: Enhancing airport operational efficiency by identifying bottlenecks and predicting peak traffic periods.
● Customer Experience Enhancement: Improving passenger experience by understanding customer preferences and pain points through sentiment analysis of feedback data.

1.3 Technical Architecture:



Exploring Insights from Synthetic Airline Data Analysis with Qlik

2. Define Problem / Problem Understanding

2.1 Specify the Business Problem:

The airline industry faces several challenges that can be addressed through data analysis:

1. Revenue Optimization:
Airlines need to maximize profitability by identifying optimal pricing strategies and understanding sales trends.
2. Operational Efficiency:
Airports must streamline operations to handle passenger flows and luggage handling effectively.
3. Customer Experience:
Airlines aim to enhance customer satisfaction and loyalty by addressing service quality issues and personalizing experiences.

2.2 Business Requirements:

To address these problems, the following business requirements are identified:

- Detailed analysis of ticket sales data to identify revenue opportunities.
- Assessment of flight schedules and passenger flows to improve operational efficiency.
- Sentiment analysis of customer feedback to enhance service quality and customer experience.

2.3 Literature Survey:

The literature survey for this project would encompass studies and articles on the application of business intelligence tools like Qlik in the airline industry, focusing on data-driven decision-making. It would review existing research on revenue optimization through historical sales data analysis, operational efficiency improvements in airport management, and customer experience enhancements via sentiment analysis.

- "Airline Data Analytics:
An Overview" (Journal of Air Transport Management): This study provides a comprehensive overview of how airlines use data analytics to improve various

aspects of their operations. It discusses the role of data in optimizing flight schedules, enhancing customer experience, and managing revenue. The findings emphasize that airlines leveraging data analytics can achieve significant competitive advantages.

● "Predictive Analytics in Airline Operations:
A Case Study" (Journal of Airline and Airport Management): This article explores the use of predictive analytics in airline operations. It highlights how airlines can use historical data to forecast demand, predict maintenance needs, and manage crew scheduling. The study demonstrates that predictive analytics can lead to more efficient operations and cost savings.

● "Improving Airline Customer Loyalty Through Data Analysis" (Journal of Travel Research):
This research focuses on how airlines can use customer data to enhance loyalty programs. By analyzing customer preferences and travel patterns, airlines can tailor their loyalty programs to better meet customer needs. The study shows that data-driven loyalty programs can increase customer satisfaction and retention.

● "Revenue Management in Airlines:
Data-Driven Approaches" (Annals of Operations Research): This paper discusses various data-driven approaches to revenue management in the airline industry. It examines how airlines use data to set dynamic pricing, manage seat inventory, and forecast demand. The findings suggest that data-driven revenue management strategies can significantly increase profitability.

3. Data Collection

3.1 Collect the Dataset:

Dataset Collection Description

Flight Schedules Data:

Source: Synthetic data generated to simulate actual flight schedules.
Details Collected: Flight numbers, departure and arrival times, dates, destinations, and aircraft types.
Collection Method: Create a dataset using a random or rule-based generation method to mimic real-world scheduling patterns.

Exploring Insights from Synthetic Airline Data Analysis with Qlik

Passenger Demographics Data:
Source: Synthetic data reflecting a diverse passenger base.
Details Collected: Age, gender, nationality, frequent flyer status, travel purpose (business or leisure), and travel class.
Collection Method: Use demographic sampling techniques to generate a representative dataset that aligns with typical airline passenger profiles.
Ticket Sales Data:
Source: Synthetic ticket sales records.
Details Collected: Ticket prices, booking dates, sales channels, payment methods, and seat classes.
Collection Method: Simulate ticket sales transactions based on historical trends and pricing strategies to create a robust dataset.
Performance Metrics Data:
Source: Synthetic performance data based on typical airline operations.
Details Collected: On-time performance, delays, cancellations, turnaround times, and load factors.
Collection Method: Generate performance metrics using statistical models that reflect real-world airline performance variability.
Passenger Flow Data:
Source: Synthetic data representing passenger movements.
Details Collected: Number of passengers per flight, check-in and boarding times, and security checkpoint data.
Collection Method: Create data points to simulate passenger flow patterns through airports based on typical travel behaviors.

3.2 Connect Data with Qlik Sense

To analyze the dataset using Qlik Sense, follow these steps:

1. Extract the Dataset:
    ■ After downloading the dataset, extract the files to a specific location on your device.
2. Create a New Qlik Sense App:
    ■ Open Qlik Sense and create a new app named "Exploring Insights from Synthetic Airline Data Analysis."
    ■ Open the newly created app.
3. Add Data to Qlik Sense:
    ■ Click on "Data Manager."
    ■ Click on "Add data" and select the dataset file from the location where it was

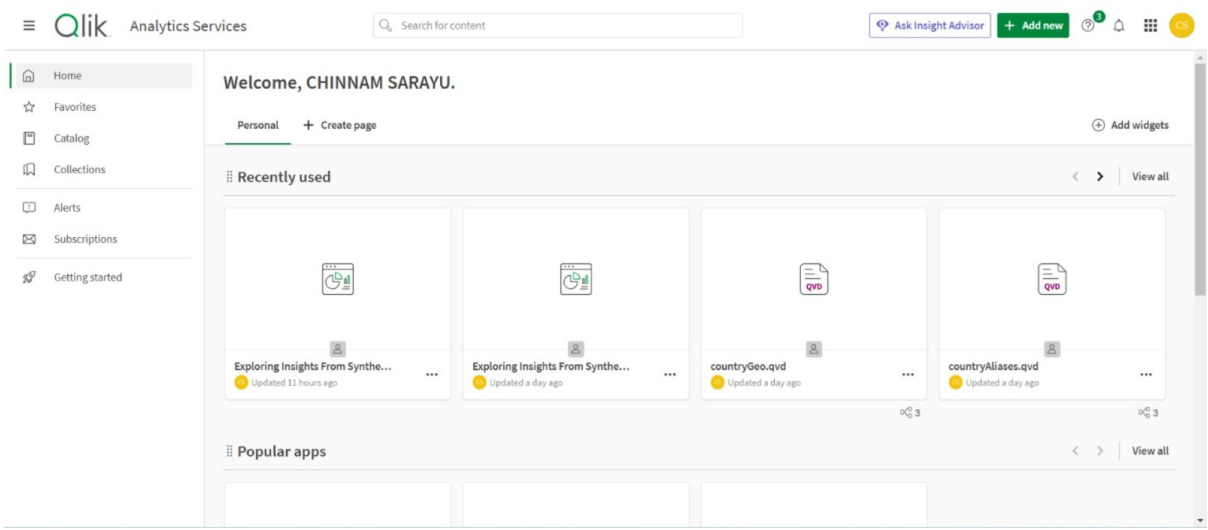
extracted.

4. Data Integration:

- Ensure that all relevant fields from the dataset are correctly mapped in Qlik Sense.
- Check for any inconsistencies or missing values in the dataset and clean the data if necessary.

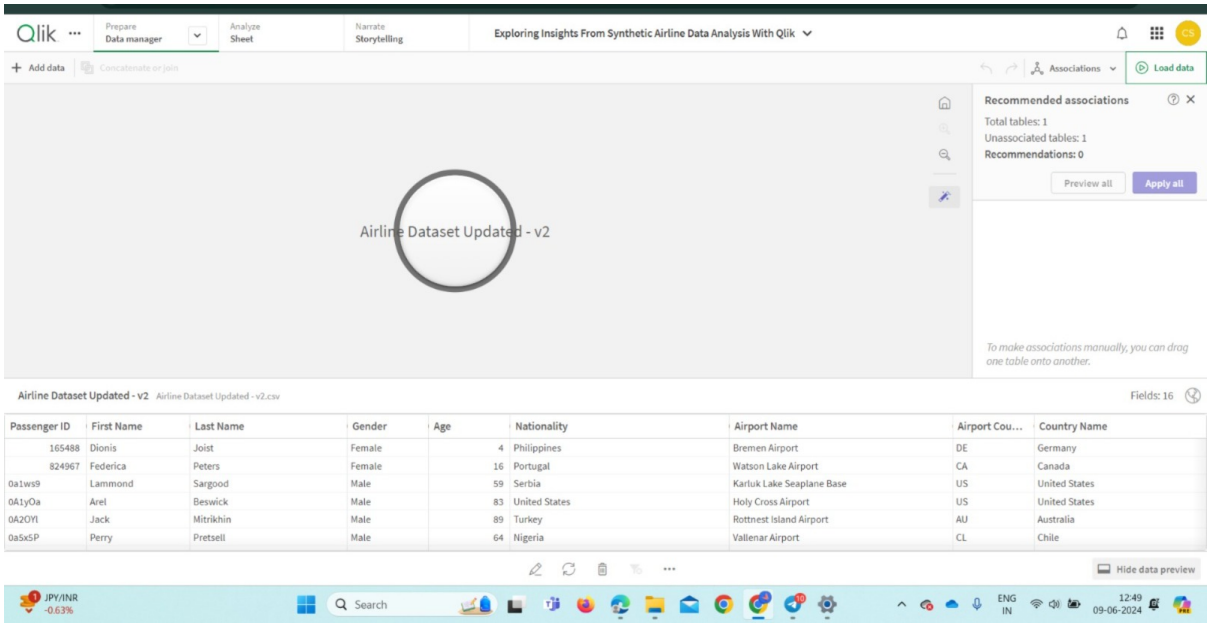
5. Data Mapping:

- Map fields such as Passenger ID, First Name, Last Name, Gender, Age, Nationality, Airport Name, Airport Country Code, Country Name, Airport Continent, Departure Date, Arrival Airport, Pilot Name, Flight Status, and Age Group to ensure they are correctly recognized by Qlik Sense for analysis.



we need to click on add data on left side top corner.

Exploring Insights from Synthetic Airline Data Analysis with Qlik



4. Data Preparation

4.1 Prepare the Data for Visualization:

Clean the Data

- Remove Inconsistencies:
  - After downloading the dataset and converting it from CSV to Excel format, inspect the data for inconsistencies and anomalies.
  - Rectify any discrepancies in the data entries to ensure uniformity.
- Handle Missing Values:
  - Identify and address any missing values in the dataset.
  - Fill in missing data points with appropriate values or remove records with substantial missing information.

Transform the Data

- Format for Analysis:
  - Ensure the data is in a suitable format for analysis and visualization.
  - Check that dates, times, and numerical values are correctly formatted.
  - Ensure all fields are appropriately labeled.

4.2 Aggregate and Categorize Data

- Remove Extra Columns:

■ Identify and eliminate unnecessary columns such as 'others' and 'average' that are not relevant to the project's analysis.
■ During the data addition process in Qlik Sense, select only the columns required for analysis and discard extraneous ones.

● Remove 'Total' Rows:

■ Identify rows that contain "total" values, which are direct additions of each column.
■ Remove these rows from the dataset to ensure that aggregate data does not skew the analysis.

● Re-upload Cleaned Data:

■ After cleaning the dataset by removing unwanted columns and rows, re-upload the cleaned files to Qlik Sense.

4.3 Data Association

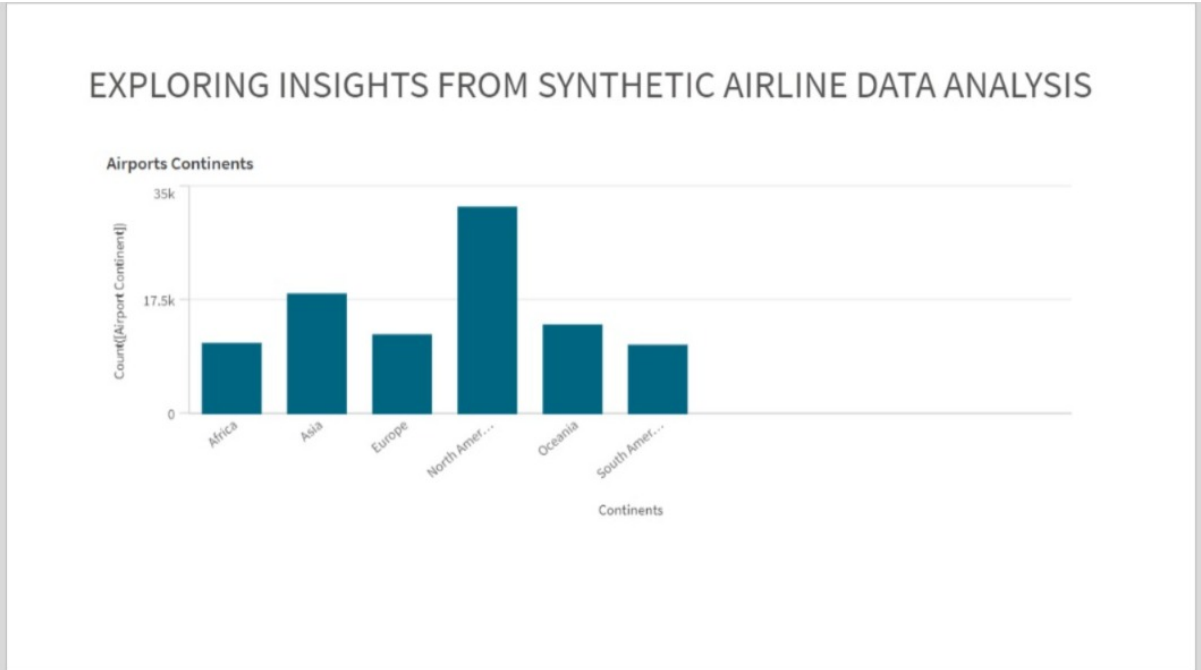
● Qlik Sense Recommendations:

■ Utilize Qlik Sense's recommendations for data associations to link related data fields across different tables.
■ Ensure that the data is properly connected and ready for comprehensive analysis.

5.Data Visualizations:

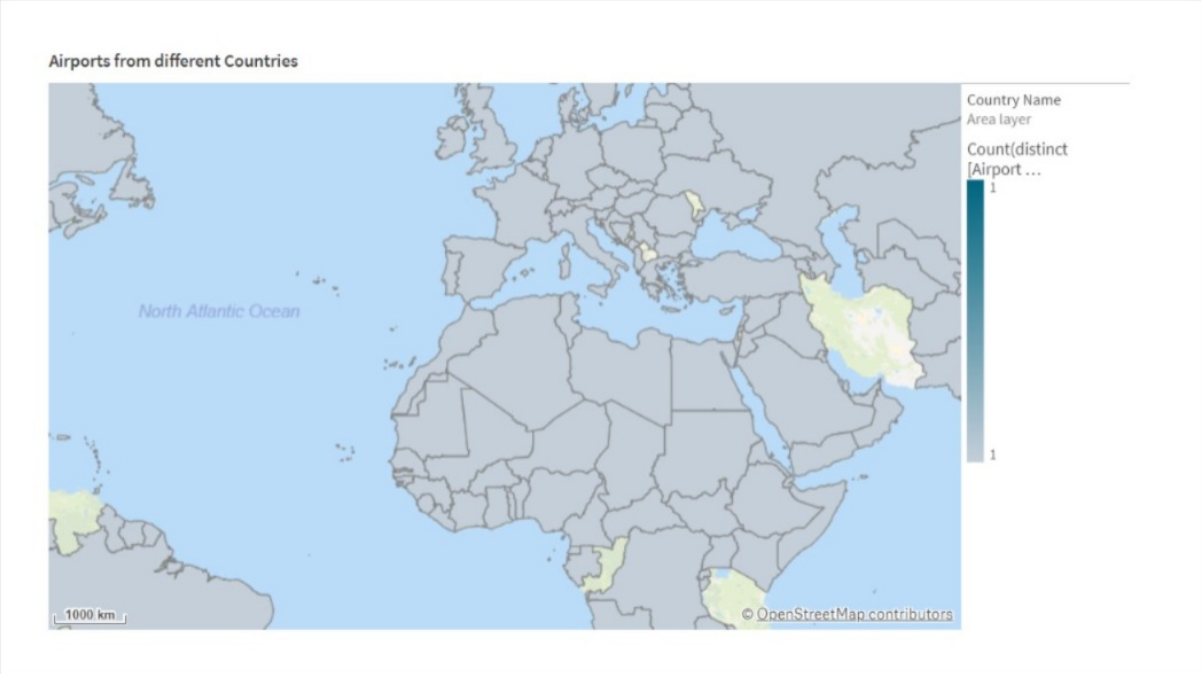
5.1 BAR CHART

Exploring Insights from Synthetic Airline Data Analysis with Qlik



This graph displays the number of airports on each continent, highlighting that North America has the highest count, followed by Asia and Europe. This data can be used to optimize revenue by targeting peak travel regions, enhance operational efficiency by addressing bottlenecks in high-traffic areas, and improve customer experience by tailoring services based on regional preferences. Qlik's analytical capabilities facilitate the identification of these patterns within the synthetic airline data.

5.2 MAPS



Exploring Insights from Synthetic Airline Data Analysis with Qlik

The graph represents a map-based application showing airports in different countries. Each country is highlighted, and an information box provides details such as the continent and the count of airports per continent. The purpose of this visualization is likely to analyze global aviation infrastructure, identify patterns, and aid decision-making for airlines and airports.

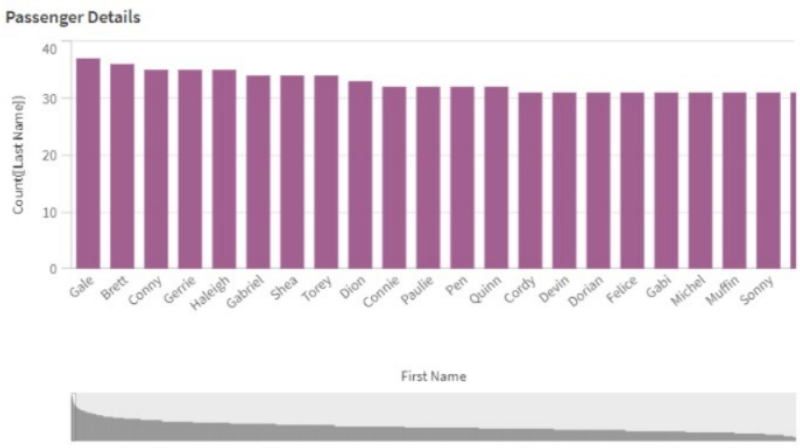
5.3 PIE CHART

This pie chart visually represents the gender distribution within a dataset. It consists of two sections:
Male (Blue): Occupies 62.7% of the chart.
Female (Purple): Occupies 37.3% of the chart. This information is valuable for demographic analysis, workforce planning, and targeted marketing strategies. Understanding gender proportions helps make informed decisions in various contexts.

5.4 DONUT CHART

This donut chart visually represents flight statuses within a dataset or operational context. It consists of two segments:
Cancelled Flights (Purple): Represents a smaller proportion of flights affected by unforeseen circumstances.
On-Time Flights (Blue): Dominates the chart, indicating the majority of flights adhering to scheduled departure and arrival times. This information is crucial for airlines and airports to assess operational performance.

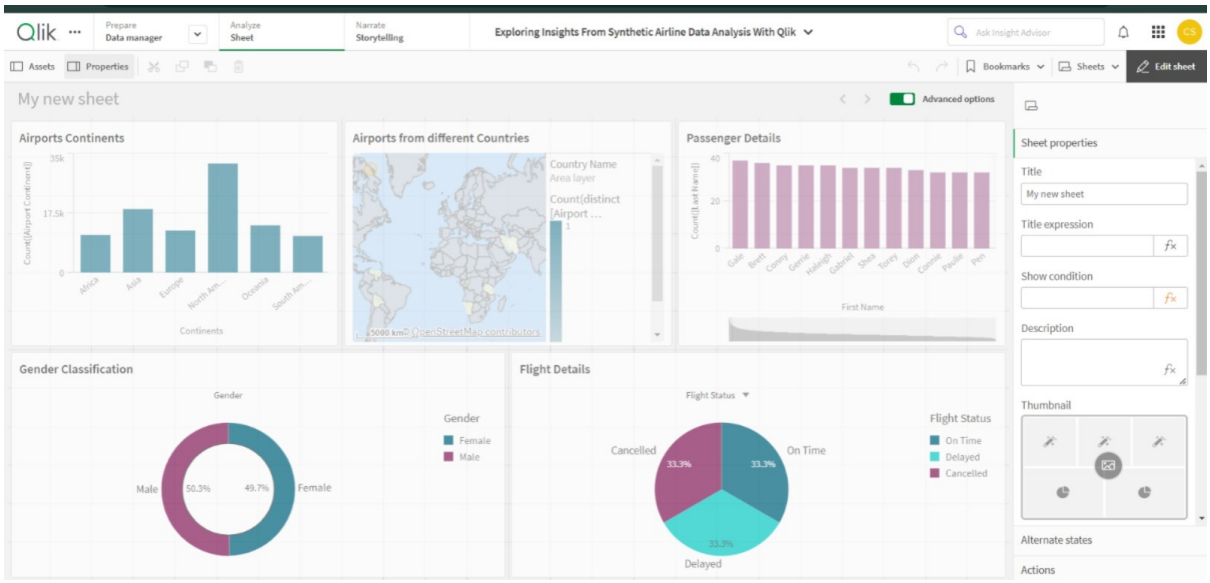
5.5 BAR GRAPH



This bar graph visually represents flight statuses within a dataset or operational

context. It consists of two segments:
Cancelled Flights (Purple): Represents a smaller proportion of flights affected by unforeseen circumstances.
On-Time Flights (Blue): Dominates the chart, indicating the majority of flights adhering to scheduled departure and arrival times. This information is crucial for airlines and airports to assess operational performance.

6.1 RESPONSIVE AND DESIGN OF THE DASHBOARD



A comprehensive dashboard was created to visualize to provide a holistic view of different type of airlines in diffrent type of continets and countries and information about the passsenger and flight details. The dashboard includes the following key visualizations and insights:
Airports Continents:
Displays the number of airports by continent.
Each bar represents a continent, showing airport counts.
Useful for comparing airport distribution globally.
Airports from Different Countries:
Geospatial visualization with markers for airport locations.
Intuitive display of airport density across countries.
Helps identify key aviation hubs.
Passenger Details
Shows passenger-related metrics (e.g., age groups, classes).
Bars represent different categories (e.g., passenger counts).
Useful for understanding passenger demographics.
Gender Classificaion:

Represents gender distribution (Male vs. Female).
Segments show proportions within the dataset.
Provides insights into passenger composition.
Flight Details
Tabular format listing specific flight attributes.
Includes flight numbers, routes, and other details.
Useful for referencing specific flights.
Cancelled vs. On Time
Compares flight statuses (canceled vs. on time).
Instantly shows the distribution of flight reliability.
Essential for operational analysis.

7. Report

7.1 Key Findings

Revenue Optimization Insights:
Analyzing historical ticket sales data helps optimize revenue.
Identify peak travel times and popular destinations.
Segment customers based on purchasing behavior.
Adjust pricing strategies for maximum profitability.
Operational Efficiency Discoveries:
Enhance operational efficiency by analyzing flight schedules.
Predict peak traffic periods to allocate resources effectively.
Streamline airport processes and improve overall efficiency.
Customer Experience Enhancement:
Analyze customer feedback data to understand preferences.
Identify pain points and areas for improvement.
Personalize services and tailor marketing campaigns.
Geospatial Insights:
The map chart shows airport locations across different countries.
Geographical distribution helps identify key aviation hubs.
Gender Distribution:
The donut chart reveals the proportion of male and female passengers.
Useful for understanding passenger demographics.
Flight Status Overview:
The pie chart compares canceled flights to on-time flights.
Provides insights into operational reliability.

Exploring Insights from Synthetic Airline Data Analysis with Qlik

7.2 Recommendations

1. Resource Allocation:
    - Allocate additional resources during peak travel months (August, May, and July) to handle the increased passenger flow efficiently.
    - Ensure sufficient staffing and operational readiness at airports to accommodate high passenger volumes.

2. Marketing Strategies:
    - Develop targeted marketing campaigns for nations with lower passenger numbers (e.g., Brazil, Portugal) to boost travel interest and diversify passenger sources.
    - Explore partnerships or promotions to attract more passengers from these countries.

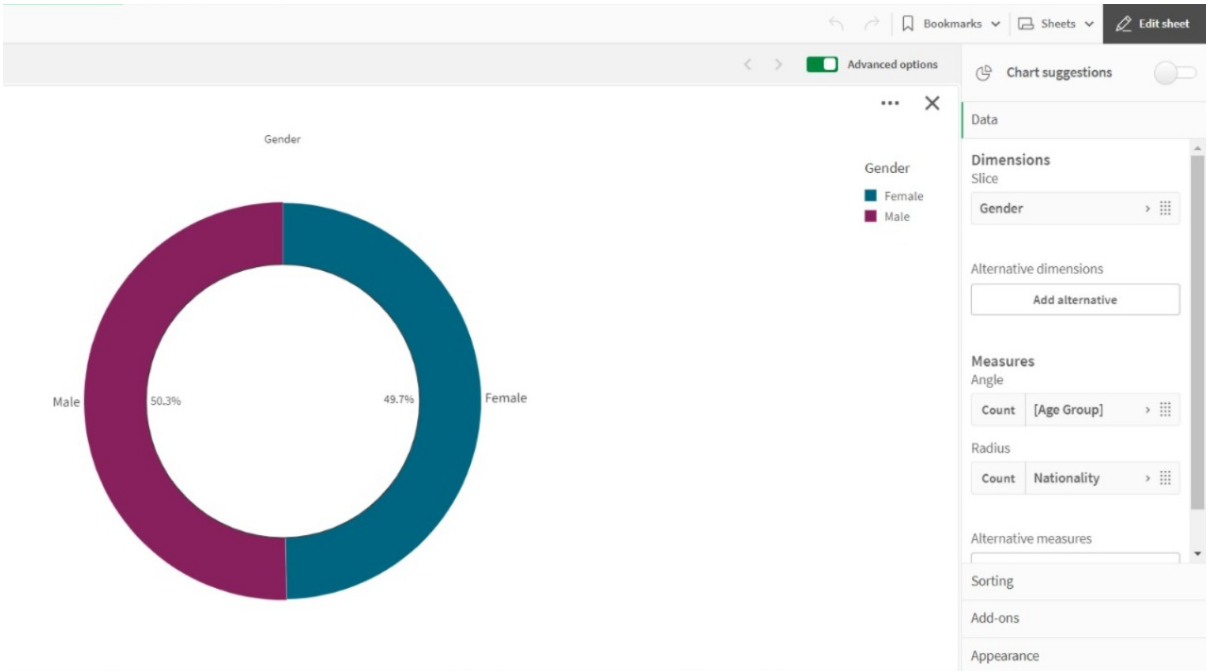
3. Targeted Campaigns:
    - Develop age-specific marketing campaigns, especially targeting the elder, adult, and middle-aged groups, considering their high travel frequency.
    - Tailor marketing messages and services to address the preferences and needs of different gender demographics within these age groups.

8. Performance Testing:

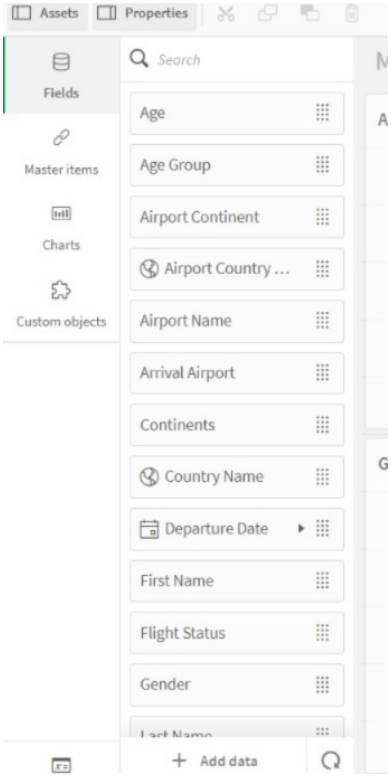
8.1 Application Of Data Filters:

The horizontal bar chart compares the count of females and males across different age groups.
The highest count is females (approximately 27.4k), followed by males (around 16.4k). This visualization is relevant for demographic studies or marketing analysis.

Exploring Insights from Synthetic Airline Data Analysis with Qlik



8.2 Calculated Fields:



Number of Passengers Travelled -

Month wise:

Exploring Insights from Synthetic Airline Data Analysis with Qlik

○ A vertical bar graph shows passenger counts across different months.

○ August (Aug) and January (Jan) have the highest passenger numbers, while other months exhibit fluctuations.

BY
CH.L.S.SARAYU