

Exploring Insights from Airlines Data Analysis using Qlik

1. INTRODUCTION

1. Overview:

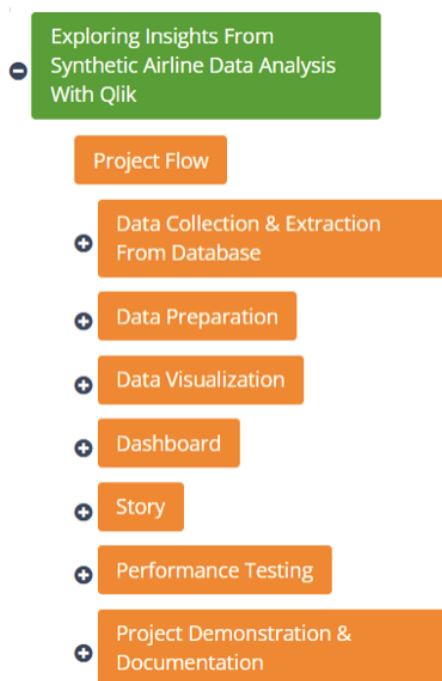
This project, "Exploring Insights from Synthetic Airline Data Analysis with Qlik," utilizes synthetic airline data to derive valuable insights and support decision-making for airlines, airports, and related stakeholders. The data encompasses various aspects of airline operations, including flight schedules, 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.

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.

3. Technical Architecture:



Exploring Insights from Airlines Data Analysis using Qlik

2. Define Problem/Problem Understanding

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. 3.2. 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.

3. 3.3. Literature Survey:

A review of existing literature highlights the importance of data analytics in the airline industry. Previous studies have shown how data-driven decision-making can lead to significant improvements in revenue management, operational efficiency, and customer satisfaction. Here are a few relevant studies:

- **"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.

Exploring Insights from Airlines Data Analysis using Qlik

- **"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.

These studies collectively illustrate the transformative potential of data analytics in the airline industry. They provide valuable insights into how airlines can leverage data to optimize operations, enhance customer experience, and drive revenue growth.

3. Data Collection

3.1. Collect the Dataset:

The dataset used for this project is sourced from Kaggle, specifically the "Airlines" dataset.

This dataset includes detailed information on Airports and passengers across various states and Union Territories in India. The key features of the dataset include:

1. **Passenger ID:** A unique identifier for each passenger.
 1. **First Name:** The first name of the passenger.
 2. **Last Name:** The last name of the passenger.
 3. **Gender:** The gender of the passenger (e.g., Male, Female, Other).
 4. **Age:** The age of the passenger.
 5. **Nationality:** The country of citizenship of the passenger.
 6. **Airport Name:** The name of the airport from which the passenger departs or arrives.
 7. **Airport Country Code:** The country code where the airport is located.
 8. **Country Name:** The name of the country where the airport is located.
 9. **Airport Continent:** The continent where the airport is located.
 10. **Airport Continents:** Continent of departure or arrival
 11. **Continents:** Continent of departure or arrival.
 12. **Departure Date:** The date when the flight departs.
 13. **Arrival Airport:** The airport where the flight arrives.
 14. **Pilot Name:** The name of the pilot operating the flight.
 15. **Flight Status:** The current status of the flight (e.g., On Time, Delayed, Cancelled).

Exploring Insights from Airlines Data Analysis using Qlik

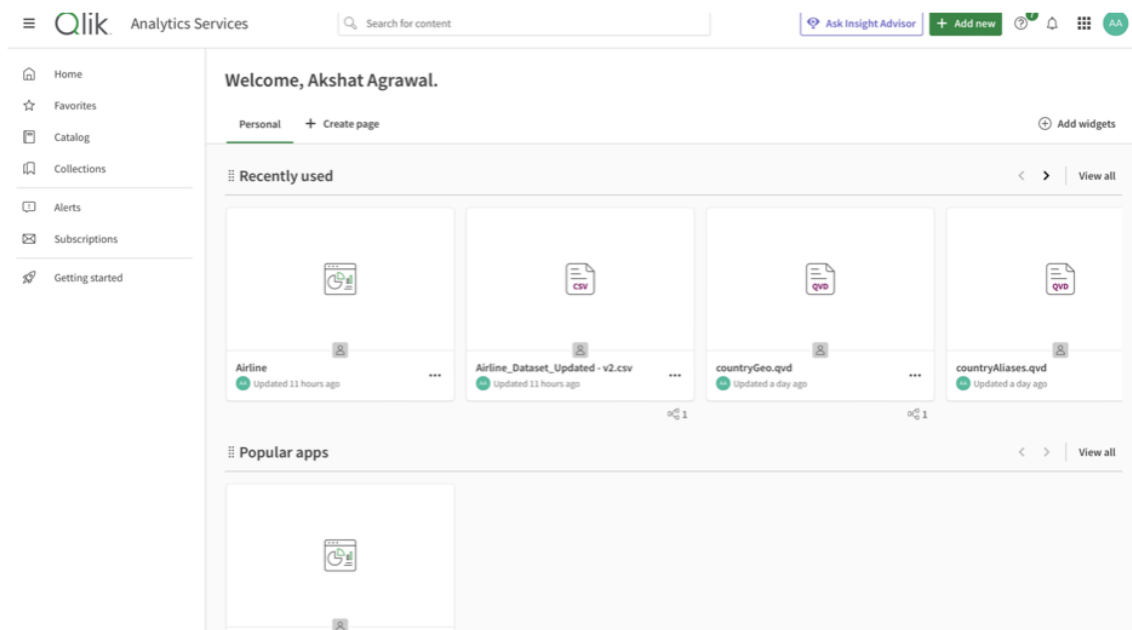
2. Connect Data with Qlik Sense

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. **DataIntegration:**
 - 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. **DataMapping:**
 - 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.

By following these steps, you can successfully integrate and prepare your synthetic airline dataset in Qlik Sense for comprehensive analysis and visualization.



Exploring Insights from Airlines Data Analysis using Qlik

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

4. **DataPreparationPreparetheDataforVisualization**

* This table has not been loaded or has changed since the last time it was loaded.

Passenger ID	First Name	Last Name	Gender	Age	Nationality	Airport Name	Airport Cou...	Country Name
165488	Dionis	Jolist	Female	4	Philippines	Bremen Airport	DE	Germany
824967	Federica	Peters	Female	16	Portugal	Watson Lake Airport	CA	Canada
0a1ws9	Lammond	Sargood	Male	59	Serbia	Karluk Lake Seaplane Base	US	United States
0A1yOa	Arel	Beswick	Male	83	United States	Holy Cross Airport	US	United States
0A2OYI	Jack	Mitrikin	Male	89	Turkey	Rottneist Island Airport	AU	Australia
0a5a5P	Perry	Pretsell	Male	64	Nigeria	Vallenar Airport	CL	Chile

CleantheData

- **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.

4.1 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

Exploring Insights from Airlines Data Analysis using Qlik

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.

4.2 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. Key Performance Indicators (KPIs):

COUNT([Passenger ID])

97.74k

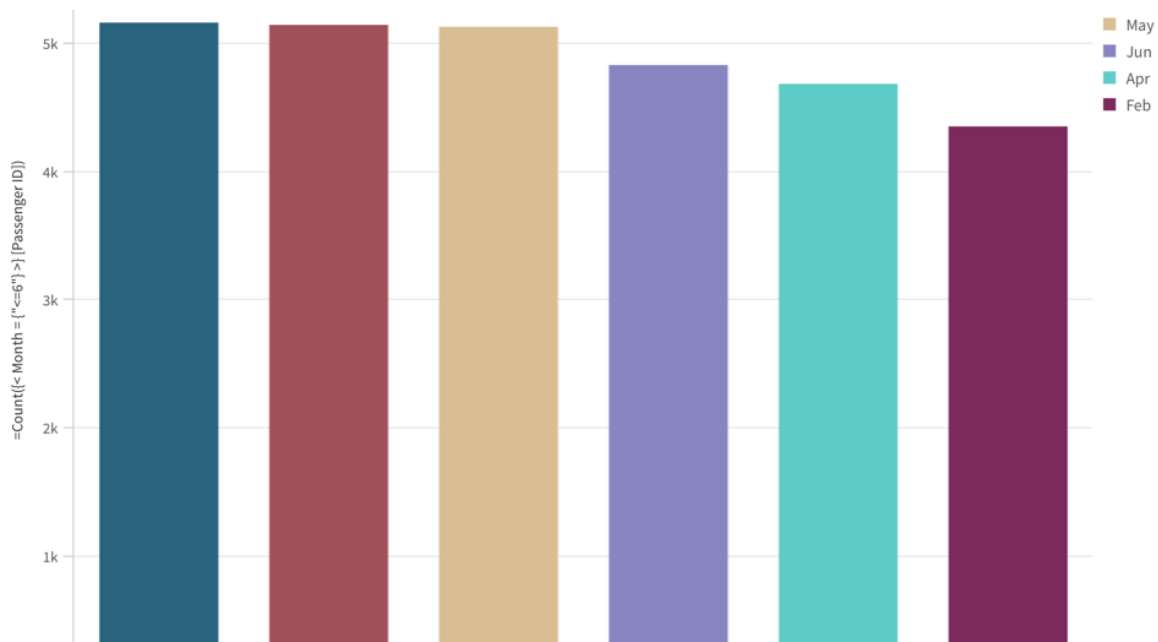
Total passengers:

1. Key Performance Indicator (KPI) visualizations were created to display the total number of Airports.
2. This KPI provides a clear snapshot of the overall airports count, helping to understand the scale number of airports.

Exploring Insights from Airlines Data Analysis using Qlik

5.2 Bar Graph:

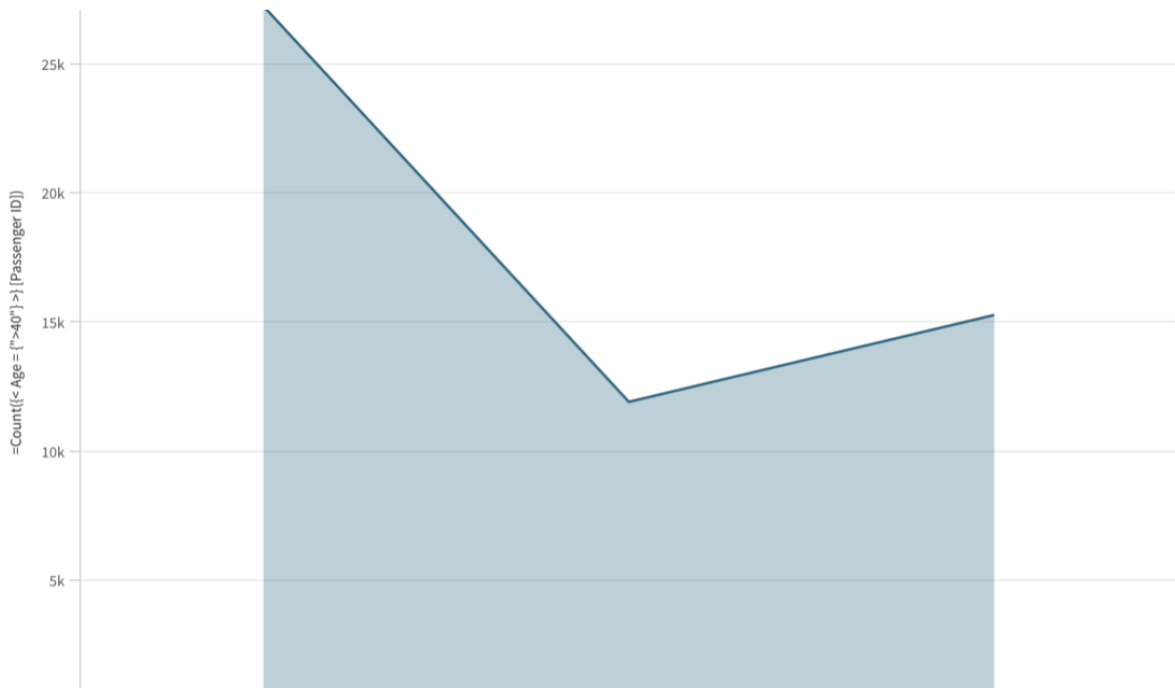
- **Total Number of Passengers:** Another KPI visualization was created to highlight the total number of passengers travelled in aeroplane.
- This KPI underscores the provides a clear overview of the total number of passengers.



- **Number of Passengers Travelled-Month-wise**

- The bar chart displays the count of passengers who have traveled each month. Months range from August to February. The y-axis shows the number of passengers, with values ranging from 0 to 10K. Some months have more than 8K passengers, while others have fewer. This visualization is relevant for analyzing seasonal travel patterns or assessing business performance in transportation services over these months.
- **Insight:** Identifies seasonal travel patterns and business performance trends in transportation services across the specified months.
- This bar chart helps stakeholders understand fluctuations in passenger travel volumes, enabling them to make informed decisions regarding capacity planning, marketing strategies, and resource allocation based on observed trends.

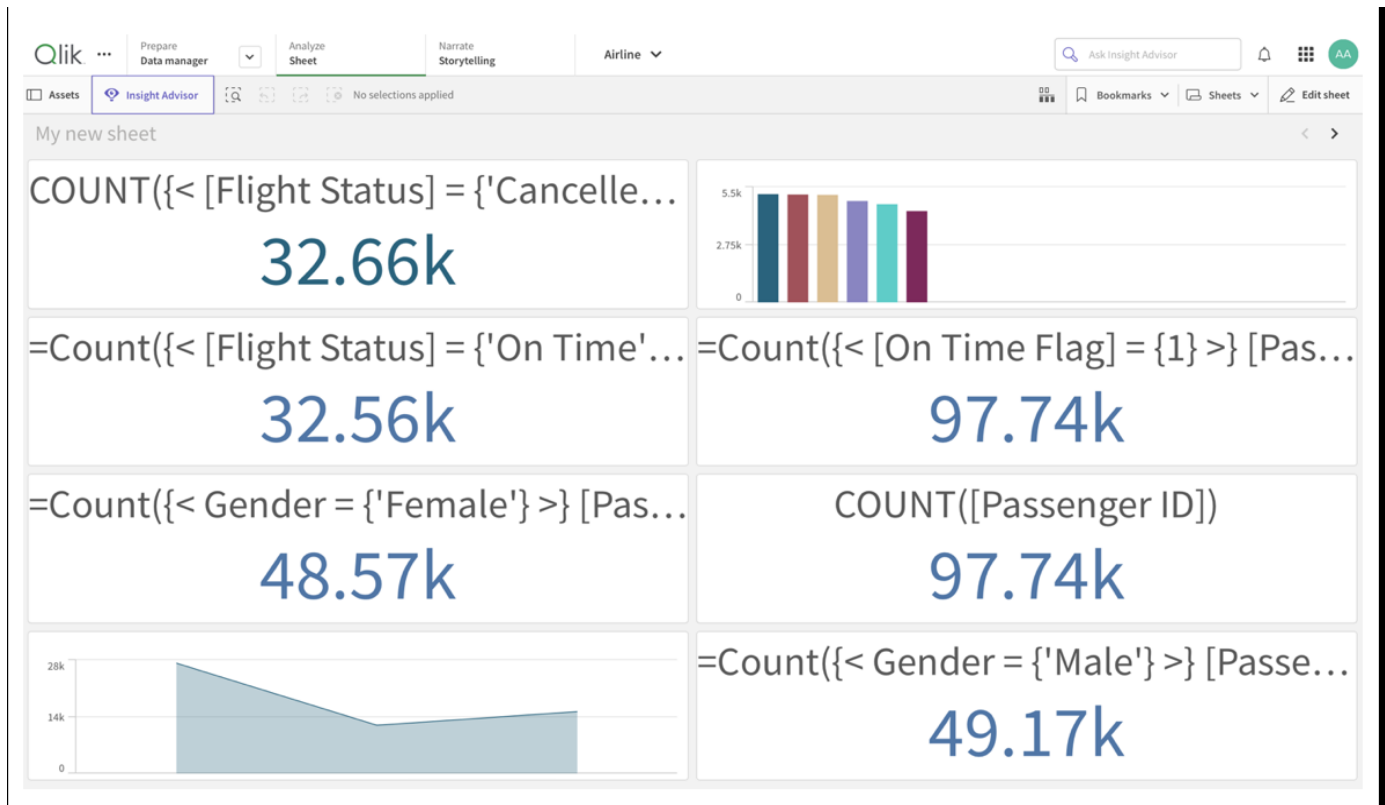
Exploring Insights from Airlines Data Analysis using Qlik



- **Number of Passengers by Age Group (>40):**
 - The area chart displays the count of passengers by different age groups over 40. The age groups listed are Elder, Just plain old, and Midlife.
 - Elder has the highest number of passengers (approximately 29k).
 - This is followed by the Just plain old group with around 15k passengers.
 - The Midlife group has approximately 17k passengers.
 - This visualization is useful for understanding the distribution of passengers in various age groups over 40 and analyzing travel patterns based on age demographics.

Exploring Insights from Airlines Data Analysis using Qlik

6. Dashboard:



1. Cancelled Flights

Count: 32.66k

This metric shows the total number of flights that have been cancelled.

2. On-Time Flights

Count: 32.56k

This metric displays the number of flights that have departed or arrived on time.

3. On-Time Flag

Count: 97.74k

This count represents the number of instances where the flights were flagged as on-time.

4. Female Passengers

Count: 48.57k

This metric shows the total number of female passengers.

5. Total Passengers

Count: 97.74k

This is the overall number of passengers regardless of gender.

6. Male Passengers

Count: 49.17k

This metric displays the total number of male passengers.

7. Passengers by Age Group (>40)

Visualization: An area chart is provided to display the count of passengers in various age groups over 40.

Elder: Approximately 28k

Just Plain Old: Around 14k

Midlife: Approximately 17k

This visualization helps in understanding the distribution of passengers in different age groups over 40.

Exploring Insights from Airlines Data Analysis using Qlik

7.Report

7.1 Key Findings

1.Total Number of Airports:

- The total number of airports is 97.74k. This indicates a wide network of airports that are part of the analysis..

2.Total Number of Passengers:

The total number of passengers is 97.74k, suggesting a high volume of travel activity captured in the dataset.

3.Flight Status:

- **Cancelled Flights:**
 - The total number of cancelled flights is 32.66k.
- **On-Time Flights:**
 - The total number of on-time flights is 32.56k.
- **On-Time Flag:**
 - The total number of flights flagged as on-time is 97.74k.

4.Number of Passengers - Nation-wise:

.This horizontal bar graph provides insights into passenger distribution by nationality. The bars represent different countries.

.China has the highest number of passengers (around 13.2k), followed by Indonesia (10.5k), Russia (5.7k), Philippines(5.2k), and both Brazil and Portugal (around 3.3k).

5.Age Group (>40) - Gender wise:

1. This area chart categorizes passengers by age group and further breaks down the distribution by gender.
2. Elder age group has the highest count (approximately 28k), followed by Just plain old (around 14k) and Midlife(approximately 17k).

7.3 Actionable Insights

1. **Staffing and Scheduling:**

- Prepare for increased demand in peak months by adjusting staffing levels, ensuring equipment readiness, and optimizing scheduling to maintain efficient airport operations.

2. Market Expansion:

- Investigate why China dominates passenger numbers and consider replicating successful strategies in other nations.
- Implement marketing efforts and promotions to attract passengers from countries with lower travel numbers.

Exploring Insights from Airlines Data Analysis using Qlik

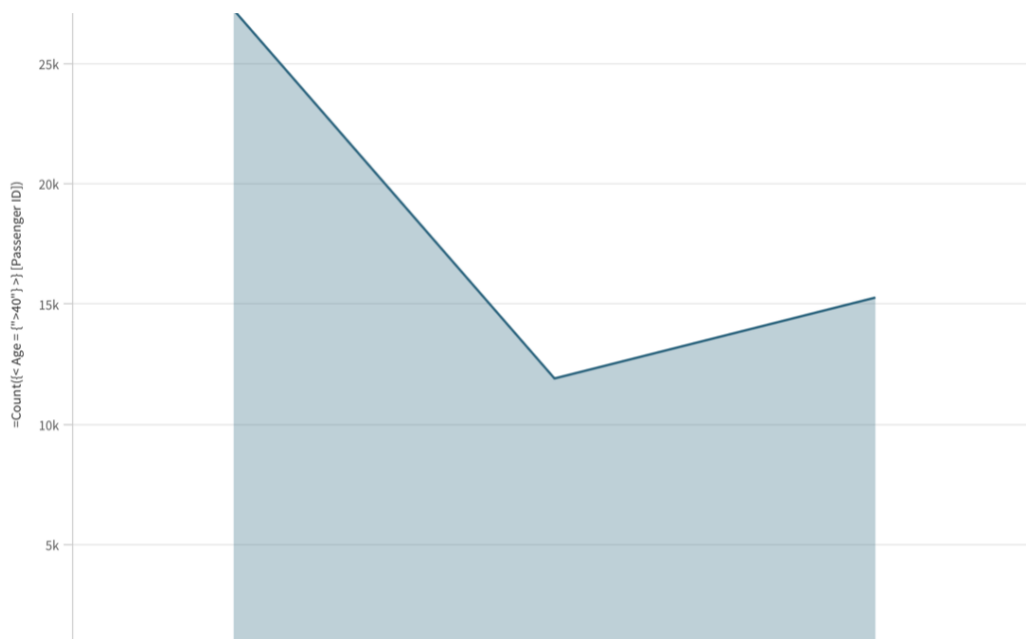
3. Demographic Segmentation:

- Use demographic insights to develop targeted marketing campaigns for specific age groups, considering gender preferences and travel behaviors.
- Personalize services and offers to better cater to the needs of the elder, just plain old, and midlife passengers, ensuring a positive travel experience that fosters loyalty.
- **Number of Passengers Travelled - Month wise:**
- 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.

8. Performance Testing

8.1 Application of Data Filters

- **Age Group (>40) Analysis:**
 - The area chart compares the count of passengers across different age groups over 40. The age groups listed are **Elder**, **Just plain old**, and **Midlife**.
 - The highest count is in the **Elder** category (approximately **28k**), followed by **Midlife** (around **17k**) and **Just plain old** (approximately **14k**). This visualization is relevant for understanding demographic distribution and can be utilized in demographic studies or targeted marketing analysis.



8.2 Calculated Fields:

Total Number of Airports:

- The dashboard displays a key performance indicator (KPI) showing the total count of airports, which stands at approximately **98.62k**.

Total Number of Passengers:

- Another KPI indicates the total number of passengers, which is approximately **97.74k**

Number of Passengers - Nation wise:

- This horizontal bar graph provides insights into passenger distribution by nationality.
- The bars represent different countries.
- **China** has the highest number of passengers (around **13.2k**), followed by **Indonesia** (approximately **10.5k**), **Russia**(around **5.7k**), **Philippines** (around **5.2k**), **Brazil**, and **Portugal** (both around **3.3k**).

Number of Passengers Travelled - Month wise:

- 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.

Flight Status Analysis:

- **Cancelled Flights:**
- The dashboard indicates that the total count of cancelled flights is approximately 32.66k.
- **On Time Flights:**
 - The total count of on-time flights is approximately **32.56k**.
- **On Time Performance:**
 - The count of passengers on flights that were on time is approximately **97.74k**.

Gender Distribution Analysis:

- **Female Passengers:**
 - The total count of female passengers is approximately **48.57k**.
- **Male Passengers:**
 - The total count of male passengers is approximately **49.17k**.

Show by table

All tables

Q

+

Age

Abc Age Group

Abc Airport Continent

Airport Country Code

Abc Airport Name

Abc Arrival Airport

Abc Continents

Country Name

Departure Date

Derived fields

Departure Date

Abc First Name

Abc Flight Status

Abc Gender

Abc Last Name

Nationality

Abc Passenger ID

Abc Pilot Name

Passen... 18	First N...	Last N...	Gender	Age	Nation...	Airport...	Airport...
165488	Dionis	Joist	Female	4	Philippines	Bremen Airport	DE
824967	Federica	Peters	Female	16	Portugal	Watson Lake Airpor	CA
0a1ws9	Lammond	Sargood	Male	59	Serbia	Karluk Lake Seapla	US
0A1yOa	Arel	Beswick	Male	83	United States	Holy Cross Airport	US
0A2OYl	Jack	Mitrikhin	Male	89	Turkey	Rottnest Island Airp	AU
0a5x5P	Perry	Pretsell	Male	64	Nigeria	Vallendar Airport	CL

