

Project Report

Exploring insights from Airline data analysis with Qlik

Submitted by: Agrima Sharma

1. Introduction:

1.1 Overview:

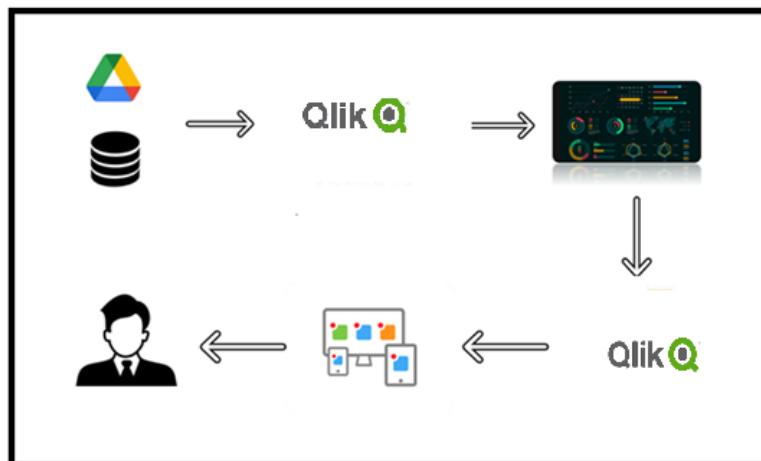
The project "Exploring Insights from Synthetic Airline Data Analysis with Qlik" focuses on using synthetic airline data to generate valuable insights through Qlik, a business intelligence and data visualization tool.

This project employs synthetic airline data to replicate various aspects of airline operations, such as flight schedules, passenger demographics, ticket sales, and performance metrics.

1.2 Purpose:

The goal is to utilize Qlik's analytical tools to reveal patterns, trends, and correlations in the data, thereby supporting decision-making processes for airlines, airports, and related stakeholders.

1.3 Technical Architecture:



- **Collect the dataset:** Gather all relevant data required for analysis from various sources.
- **Connect Data with Qlik Sense:** Integrate the collected data into Qlik Sense for analysis.
- **Prepare the Data for Visualization:** Clean, format, and organize the data to be suitable for visual representation.
- **Visualizations:** Create charts, graphs, and other visual elements to represent the data insights.
- **Responsive and Design of Dashboard:** Design a user-friendly and adaptable dashboard layout.
- **Story Creation:** Develop a narrative using the visualizations to convey the insights and findings effectively.

2. Define Problem / Problem Understanding:

2.1 Specify the business problem:

2.1.1 Revenue Optimization: The airline needs to improve revenue management using historical ticket sales data to understand travel patterns, customer segments, and pricing strategies, leveraging Qlik for visualizations.

2.1.2 Operational Efficiency: The airport authority needs to improve operational efficiency by addressing bottlenecks in flight schedules, passenger flows, and luggage handling using Qlik.

2.1.3 Customer Experience Enhancement: Airlines need to improve passenger experience by understanding customer preferences and pain points, using Qlik integrated with sentiment analysis.

2.2 Business requirements:

2.2.1 Revenue Optimization

1. Data Integration: Import historical ticket sales data into Qlik.
2. Visualization Tools: Create dashboards for revenue trends, peak travel times, and popular destinations.
3. Customer Segmentation: Segment customers based on purchasing behavior.
4. Pricing Strategy Analysis: Assess current pricing strategies.
5. Predictive Analytics: Forecast future trends and adjust pricing dynamically.
6. Reporting: Generate performance and improvement reports.

2.2.2 Operational Efficiency

1. Data Integration: Consolidate flight schedules, passenger flow, and luggage handling data into Qlik.
2. Bottleneck Identification: Create visualizations to identify operational bottlenecks.
3. Peak Period Prediction: Forecast peak traffic periods using predictive models.
4. Resource Allocation: Develop dashboards for effective resource allocation.
5. Efficiency Metrics: Track key performance indicators (KPIs).
6. Optimization Recommendations: Generate insights to streamline processes.

2.2.3 Customer Experience Enhancement

1. Data Integration: Integrate customer feedback and sentiment analysis data into Qlik.
2. Customer Insights: Create visualizations of customer preferences, satisfaction levels, and pain points.
3. Personalization: Identify opportunities for personalized services.
4. Marketing Optimization: Develop tailored marketing campaigns based on customer insights.

5. **Feedback Loop:** Implement continuous feedback collection and analysis.
6. **Loyalty Metrics:** Track customer loyalty and satisfaction metrics.
7. **Improvement Areas:** Provide actionable insights to enhance service quality.

2.3 Literature Survey:

2.3.1 Introduction:

The project uses synthetic airline data with Qlik for deriving insights into flight schedules, passenger demographics, ticket sales, and performance metrics to aid decision-making for airlines and airports.

2.3.2 Revenue Optimization:

1. **Data-Driven Revenue Management:** Research highlights the importance of historical sales data and predictive analytics in optimizing pricing strategies and identifying peak travel periods. Utilizing Qlik for visualizing trends and segmenting customers supports enhanced profitability (Boyd, 2007; Talluri & van Ryzin, 2004).
2. **Customer Segmentation and Pricing Strategies:** Studies emphasize customer segmentation in tailoring pricing strategies. Qlik's capabilities facilitate effective segmentation and dynamic pricing adjustments (Wirtz et al., 2003; Chen & Wu, 2009).

2.3.3 Operational Efficiency:

1. **Enhancing Airport Operations:** Research on operational efficiency addresses bottlenecks in passenger flows and luggage handling. Qlik can identify inefficiencies and predict peak traffic periods, enhancing resource allocation (de Neufville & Odoni, 2003; Ashford et al., 2013).
2. **Predictive Analytics for Resource Allocation:** Predictive models can significantly improve airport operations. Qlik's analytical tools provide actionable insights for better resource management (Lee et al., 2010; Yeh & Kuo, 2003).

2.3.4 Customer Experience Enhancement:

1. **Understanding Customer Preferences:** Enhancing customer experience requires understanding preferences and pain points. Sentiment analysis with Qlik can identify areas for improvement and personalize services (Fodness & Murray, 2007; Parasuraman et al., 2005).
2. **Sentiment Analysis and Personalization:** Sentiment analysis captures customer feedback effectively. Integrating this with Qlik helps tailor marketing campaigns and services, fostering loyalty and satisfaction (Feldman, 2013; Liu, 2012).

2.3.5 Conclusion:

The project aligns with existing research on revenue management, operational efficiency, and customer experience enhancement, leveraging Qlik's capabilities to derive actionable insights from synthetic airline data, benefiting airlines and airports in optimizing operations and enhancing customer satisfaction.

2.3.6 References:

1. Boyd, E. A. (2007). The Future of Pricing.
2. Talluri, K. T., & van Ryzin, G. J. (2004). The Theory and Practice of Revenue Management.
3. Wirtz, J., Mattila, A. S., & Lwin, M. O. (2007). Journal of Service Research.
4. Chen, K., & Wu, H. (2009). Sustainable social relationship between buyers and retailers.
5. de Neufville, R., & Odoni, A. (2003). Airport Systems.
6. Ashford, N., Mumayiz, S., & Wright, P. H. (2013). Airport Engineering.
7. Lee, H. L., Padmanabhan, V., & Whang, S. (2010). Management Science.
8. Yeh, C. H., & Kuo, Y. (2003). Transportation Research Part E.
9. Fodness, D., & Murray, B. (2007). Journal of Services Marketing.
10. Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (2005). Journal of Marketing.
11. Feldman, R. (2013). Communications of the ACM.
12. Liu, B. (2012). Synthesis Lectures on Human Language Technologies.

3. Data Collection:

3.1 Collect the dataset:

A dataset is collected from the following 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)

3.2 Connect Data with Qlik Sense:

Integrate dataset into Qlik Sense by establishing connections to data sources, which can include databases, spreadsheets, and cloud services. Use Qlik Sense's data load editor to import and structure the data efficiently. Ensure that the data connections are secure and configured correctly to enable seamless data integration and subsequent analysis.

4. Data Preparation:

4.1 Prepare the Data for Visualization:

Preparing the data for visualization entails several steps: cleaning the data to eliminate irrelevant or missing entries, transforming the data into a format conducive to visualization, exploring the data to uncover patterns and trends, filtering the data to concentrate on specific subsets, readying the data for visualization software, and verifying its accuracy and completeness. This comprehensive process ensures that the data is clear, understandable, and prepared for creating visualizations that provide insights into performance and efficiency. Given that the data has already been cleaned, we can proceed to the visualization stage.

5. Data Visualizations:

5.1 Visualizations:

Data visualization involves the creation of graphical representations of data to assist individuals in comprehending and exploring the information. The aim of data visualization is to render complex data sets more accessible, intuitive, and easier to interpret. Through the use of visual elements like charts, graphs, and maps, data visualizations enable people to quickly identify patterns, trends, and outliers within the data.

5.2 Activities:

5.2.1 Total Number of Passengers

Total Number of Passengers

157.7k

5.2.2 Number of Passengers effected by Cancelled Flights

Number of passengers affected by Cancelled flights

52.66k

5.2.3 Number of Passengers effected by Delay of Flights

Number of passengers affected by Delay of flights

52.53k

5.2.4 Number of Male Passengers

Number of Male Passengers

79.29k

5.2.5 Number of Female Passengers

Number of Female Pasengers

78.44k

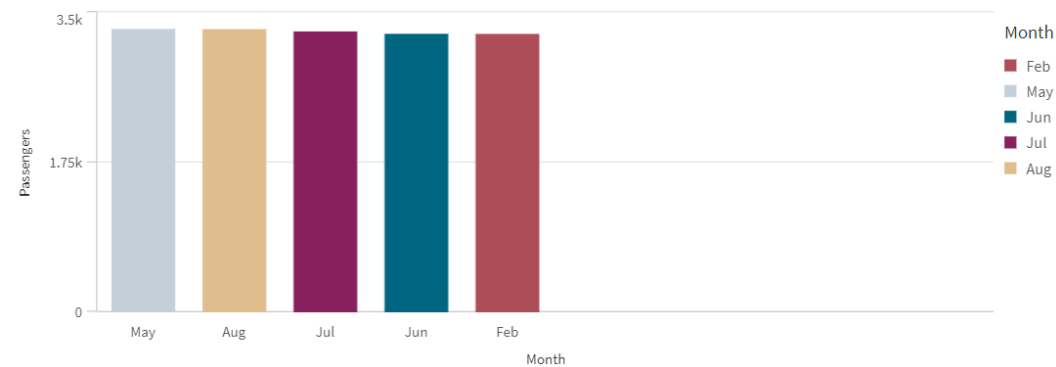
5.2.6 Number of Flights-On-Time

Number of Flights - On Time

52.54k

5.2.7 Top 5 Months where Passengers traveled the most.

Top 5- Number of Passengers travelled - Month Wise



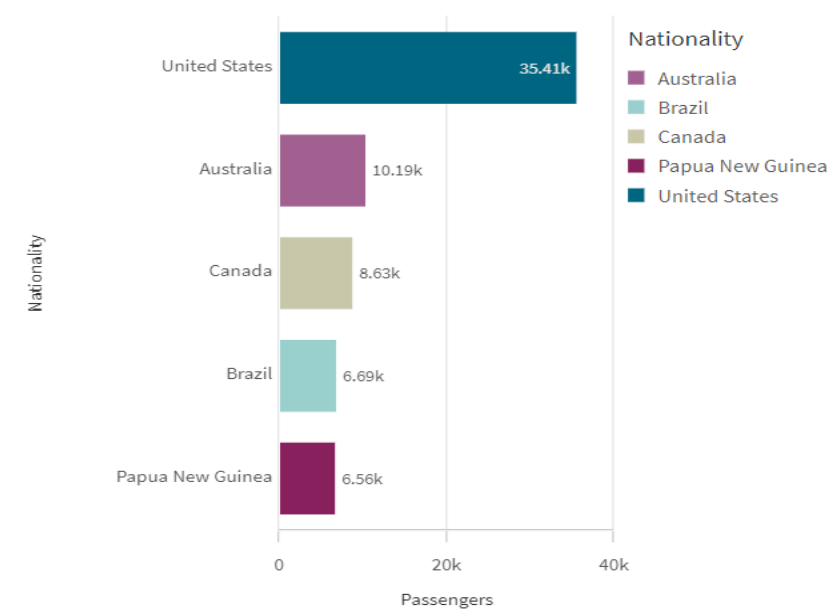
5.2.8 Top 3 Months - Flight status Wise (Delayed/Cancelled/On-Time)

Top 3 Months - Flight Status Wise



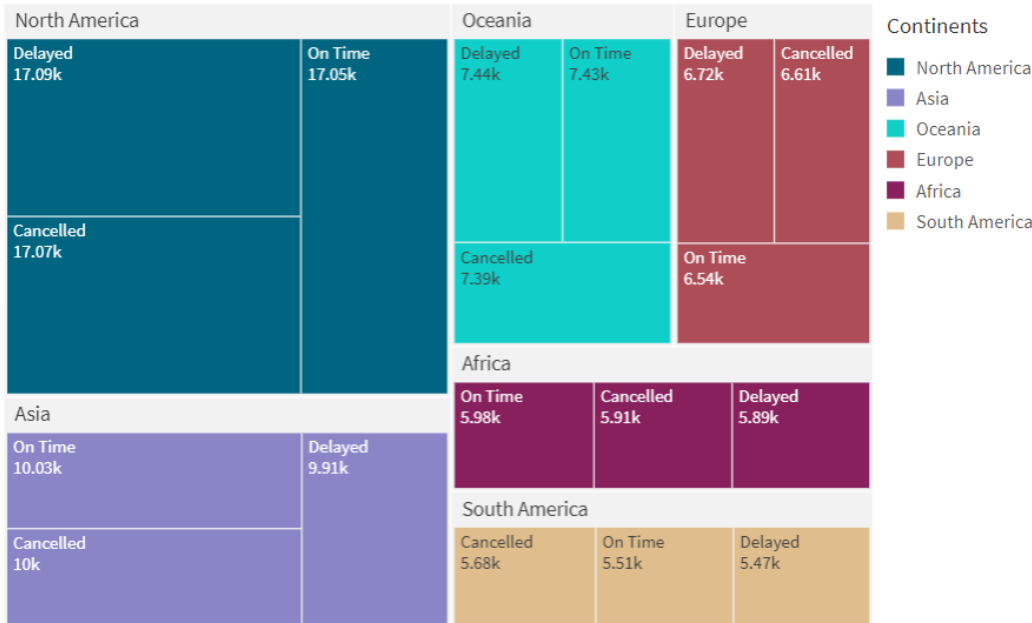
5.2.9 Number of Passengers Nationality Wise

Number of Passengers - Nation Wise



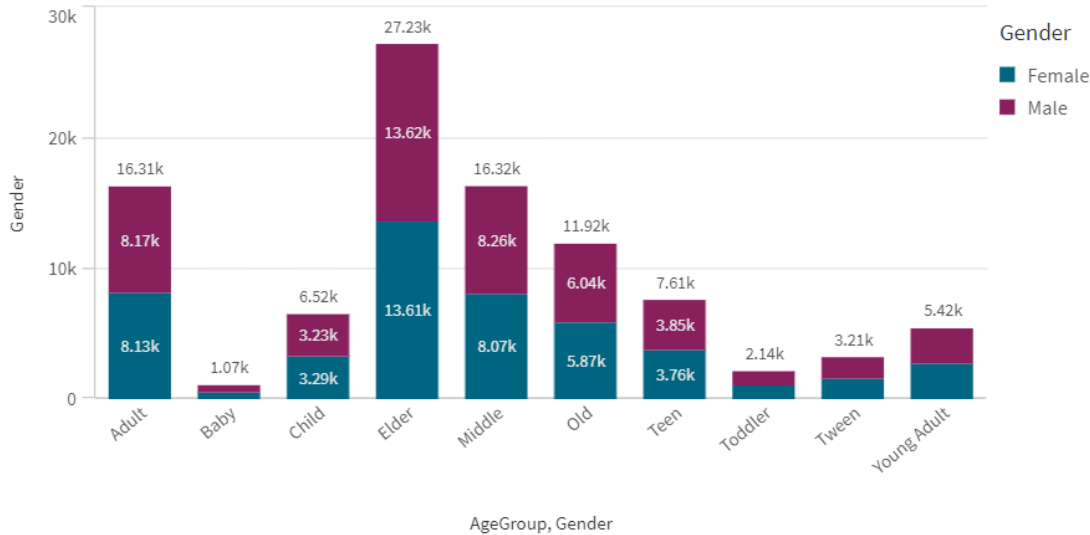
5.2.10 Continent wise – Flight Status

Continents Wise - Flight Status



5.2.11 Age Group – Gender Wise

Age Group - Gender Wise



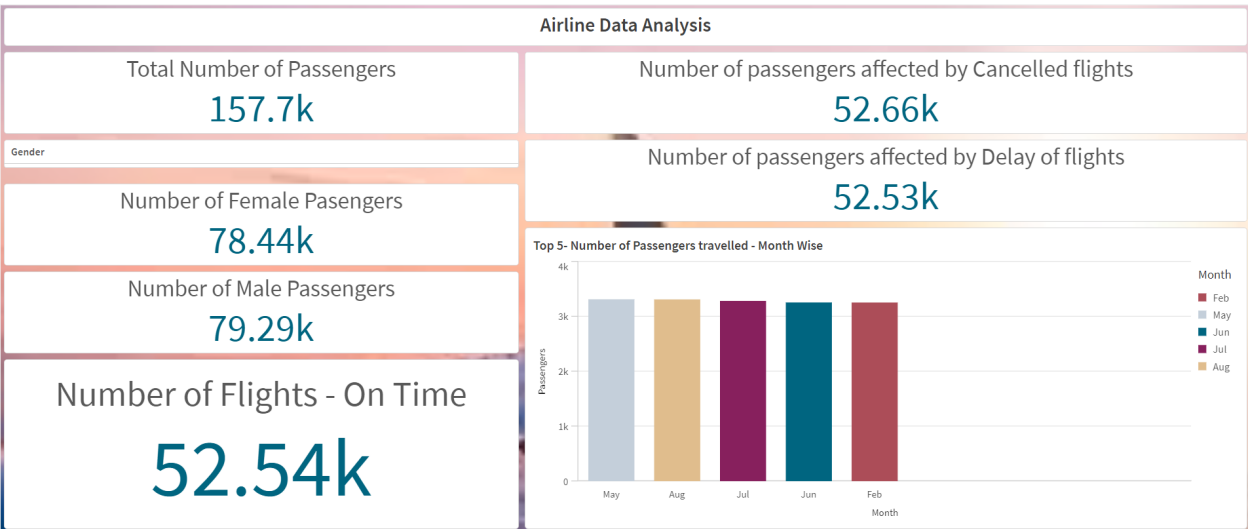
5.2.12 A filter pane consisting of Gender(M/F)



6. Dashboard:

6.1 Responsive and Design of Dashboard:

6.1.1 Dashboard 1



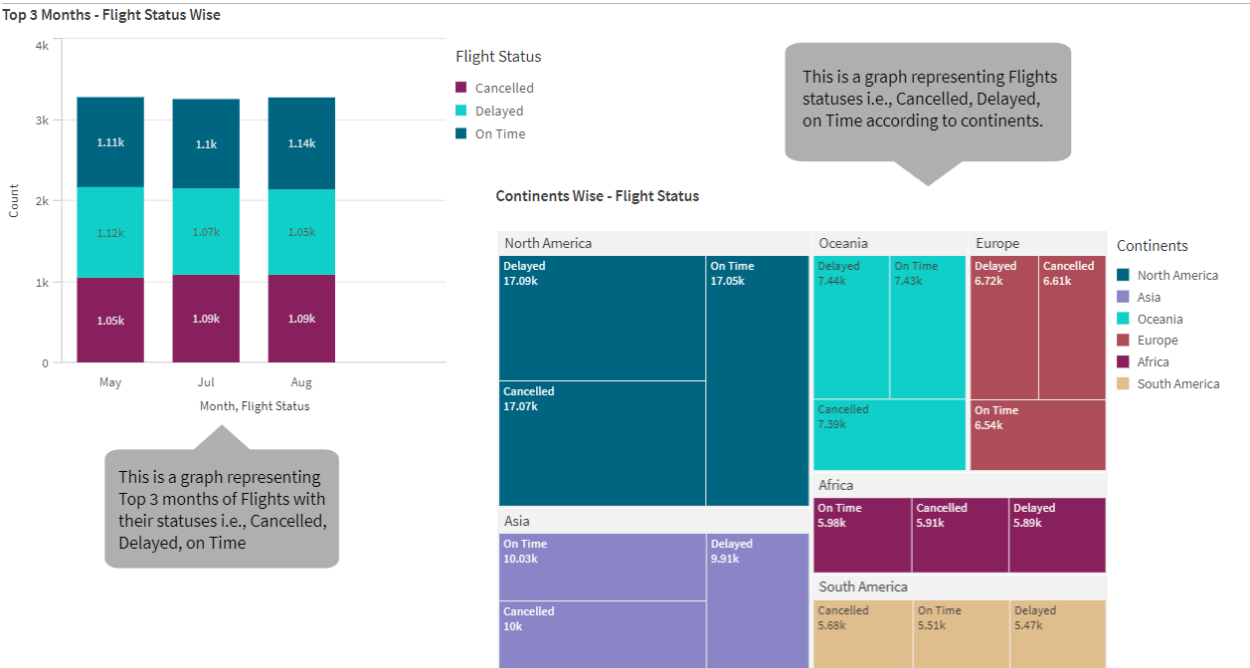
6.1.2 Dashboard 2



7. Report:

7.1 Report Creation:

Slide 1



Slide 2



Slide 3

Total Number of Passengers

157.7k

Number of Female Pasengers

78.44k

Number of Male Passengers

79.29k

Number of passengers affected by Delay of flights

52.53k

Number of passengers affected by Cancelled flights

52.66k

The number of male and female passengers travelled:

Male - [Click Here](#)

Female - [Click Here](#)

Number of Flights - On Time

52.54k

This slide shows a number of values to indicate total no. of passengers and no. of female and male passengers. It also gives data of the no. of flights on time and no. of passengers affected by the delaying or cancellation of flights. The slide also has 2 links that takes us to the sheet to demonstrate the logistics using filter on gender.

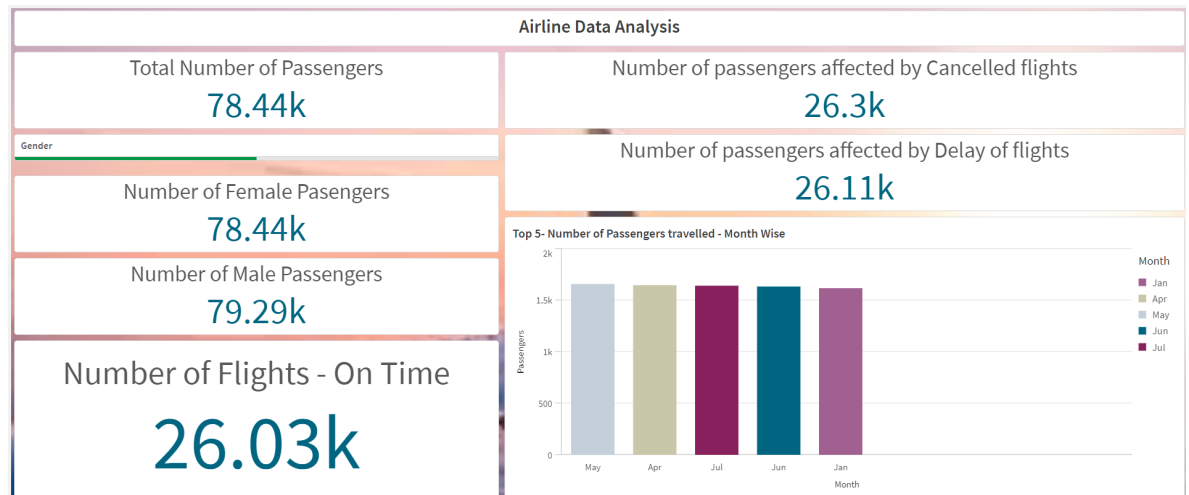
8. Performance Training:

8.1 Amount of Data Rendered:

```
[Airline Dataset]:  
NOCONCATENATE LOAD *,  
if(Age >=0 AND Age <= 1, 'Baby',  
if(Age >= 1 AND Age <= 3, 'Toddler',  
if(Age >= 4 AND Age <= 9, 'Child',  
if(Age >=10 AND Age <= 12, 'Tween',  
if(Age >=13 AND Age <= 19, 'Teen',  
if(Age >=20 AND Age <=24, 'Young Adult',  
if (Age >=25 AND Age <=39, 'Adult',  
if (Age >=40 AND Age <= 54, 'Middle',  
if(Age >=55 AND Age <= 79, 'Elder',  
if(Age >=80, 'Old'))))))) AS AgeGroup,  
Date#([Departure Date], 'MM/DD/YYYY') as [Departure_Date],  
Year([Departure Date]) AS Year,  
Month([Departure Date]) as Month  
RESIDENT [Airline Dataset]  
Where NOT ([Arrival Airport] = '0' OR [Arrival Airport] = '-');
```

8.2 Utilization of Data Filters:

Utilisation of Gender Data Filter with gender set as Female.



Utilisation of Gender Data Filter with gender set as Male.

