Project Report

Exploring Insights From Synthetic Airline Data Analysis With Qlik

-Suhani Thakur

1) Introduction:

1.1 Overview:

The project "Exploring Insights from Synthetic Airline Data Analysis with Qlik" leverages synthetic airline data to derive meaningful insights using Qlik, a business intelligence and data visualization tool. This data simulates key aspects of airline operations, such as flight schedules, passenger demographics, ticket sales, and performance metrics. The goal is to use Qlik's analytical capabilities to uncover patterns, trends, and correlations, supporting decision-making for airlines, airports, and stakeholders.

1.2 Purpose:

1. Revenue Optimization:

- To identify peak travel times and popular destinations.
- To analyze historical ticket sales data for effective pricing strategies.
- To segment customers based on purchasing behavior.
- To maximize profitability through data-driven pricing adjustments.

2. Operational Efficiency:

- To analyze flight schedules, passenger flows, and luggage handling processes.
- To identify bottlenecks and inefficiencies in airport operations.
- To predict peak traffic periods and allocate resources effectively.

- To streamline processes and enhance overall airport operational efficiency.

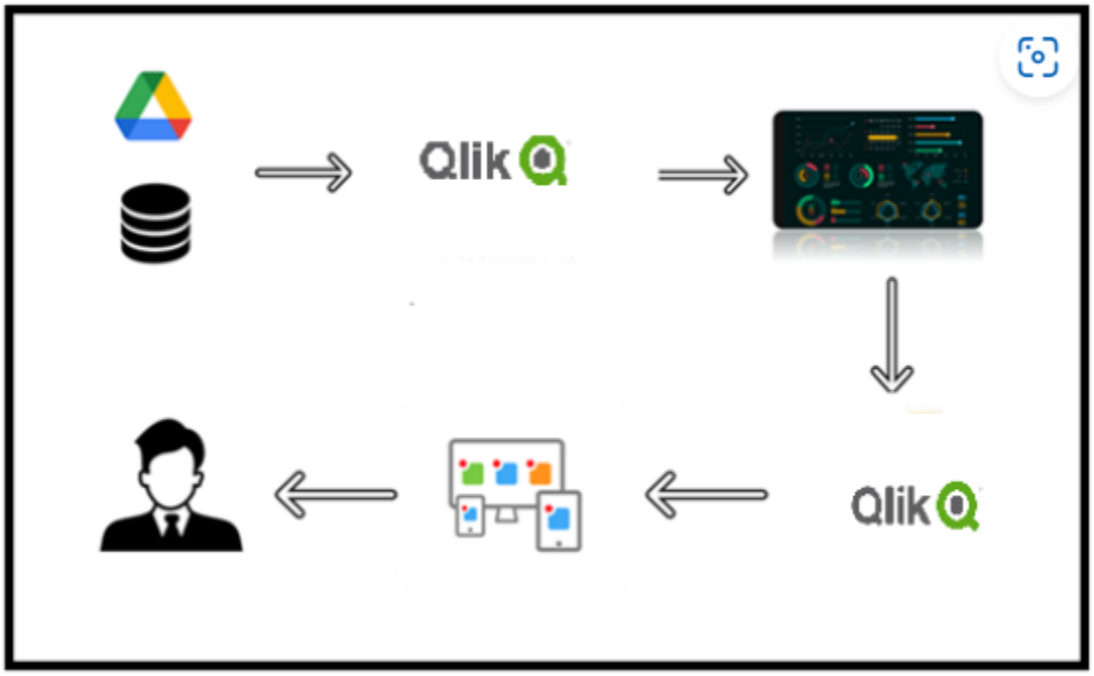
3. Customer Experience Enhancement:

- To understand customer preferences, satisfaction levels, and pain points.
- To perform sentiment analysis on customer feedback data.
- To identify areas for service improvement.
- To personalize services and tailor marketing campaigns.
- To foster customer loyalty and increase satisfaction.

Overall, the project aims to leverage Qlik's powerful data visualization and analytical capabilities to transform synthetic airline data into actionable insights, ultimately aiding in better decision-making for airlines and related stakeholders.

1.3 Technical Architecture:

4o



The technical architecture of the "Exploring Insights from Synthetic Airline Data Analysis with Qlik" project involves several key steps:

1. **Data Acquisition**:
    - **Dataset Download**: The synthetic airline data was downloaded and prepared for analysis.
2. **Data Analysis and Visualization**:
    - **Qlik Cloud Utilization**: The dataset was uploaded to Qlik Cloud, where visualizations and dashboards were created. Qlik's powerful analytical tools were used to uncover patterns, trends, and insights from the data.
3. **Solution Development and Presentation**:
    - **Story Creation**: Using Qlik Cloud, a comprehensive story was developed to explain the findings and proposed solutions.
    - **Consumer Engagement**: The final story and visualizations are designed to be easily interpreted and utilized by stakeholders for informed decision-making.

This architecture ensures a seamless workflow from data acquisition to insightful storytelling and decision support.

2) Problem Understanding :

2.1 Business problem:

The business problem for airlines centers around two main issues:

1. **Operational Inefficiencies Leading to Passenger Inconvenience**:
    - **Delayed and Cancelled Flights**: Frequent delays and cancellations cause significant inconvenience to passengers, leading to dissatisfaction and potential loss of customers.
2. **Underutilization of Data for Resource Optimization and Customer Satisfaction**:
    - **Resource Management**: Airlines are not fully leveraging important statistics and data insights to optimize their resources, such as flight schedules, crew allocations, and maintenance planning.
    - **Customer Satisfaction**: Lack of data-driven strategies prevents airlines from effectively addressing customer preferences and pain points, leading to missed opportunities for enhancing the passenger experience and fostering loyalty.

By addressing these problems through data analysis and visualization with Qlik, airlines can improve operational efficiency, reduce passenger inconvenience, and enhance overall customer satisfaction.

2.2 Business Requirements:

1. **Revenue Optimization**:
    - **Historical Data Analysis**: Ability to analyze past ticket sales data to identify peak travel times, popular destinations, and successful pricing strategies.
    - **Customer Segmentation**: Tools to segment customers based on purchasing behavior and demographics to tailor marketing and pricing strategies.
    - **Trend Visualization**: Visualization of revenue trends over time to inform pricing adjustments and promotional campaigns.
2. **Operational Efficiency**:
    - **Flight Schedule Analysis**: Tools to analyze and optimize flight

schedules to reduce delays and cancellations.
    - **Passenger Flow Management**: Insights into passenger flows to identify and mitigate bottlenecks in boarding, security, and luggage handling processes.
    - **Resource Allocation**: Data-driven predictions of peak traffic periods to effectively allocate staff and resources.
3. **Customer Experience Enhancement**:
    - **Customer Feedback Analysis**: Tools to perform sentiment analysis on customer feedback to identify pain points and areas for improvement.
    - **Service Personalization**: Ability to understand and cater to individual customer preferences for a more personalized travel experience.
    - **Marketing Campaign Optimization**: Data-driven insights to tailor marketing campaigns to better meet customer needs and preferences, fostering loyalty and satisfaction.
4. **Integration and Accessibility**:
    - **Unified Dashboard**: A centralized dashboard that integrates various data sources, providing a comprehensive view of operations and performance.
    - **Interactive Visualizations**: Easy-to-use, interactive visualizations that allow stakeholders to explore data and insights intuitively.
    - **Scalability**: Solutions that can scale with the growth of data and evolving business needs.

These requirements aim to leverage Qlik's capabilities to transform synthetic airline data into actionable insights, driving improved operational efficiency, revenue growth, and enhanced customer experiences.

2.3 Literature Survey:

1. Business Intelligence in Airlines
Business intelligence (BI) is crucial for airlines to optimize operations, enhance customer experiences, and drive revenue growth. BI tools like Qlik transform data into actionable insights through advanced visualization and analytics.
2. Revenue Optimization
Effective revenue management relies on analyzing historical sales data to identify trends and optimize pricing strategies. Studies highlight the benefits of data-driven

customer segmentation and dynamic pricing to maximize profitability.

3. Operational Efficiency

Operational efficiency is vital for minimizing delays and improving passenger satisfaction. Research shows that integrating flight schedules, passenger flows, and luggage handling data helps identify and address operational bottlenecks.

4. Customer Experience Improvement

Understanding customer preferences and feedback is essential for enhancing service quality. Sentiment analysis helps identify pain points and personalize services, leading to improved passenger satisfaction and loyalty.

5. Synthetic Data for Insights

Synthetic data allows airlines to simulate real-world scenarios without compromising sensitive information. It provides a scalable way to test BI tools and strategies, enabling safe exploration of revenue optimization, operational efficiency, and customer experience enhancement.

6. Qlik's Capabilities

Qlik is recognized for its robust BI features, offering interactive visualizations and powerful analytics. It integrates various data sources into comprehensive dashboards, facilitating better decision-making through deeper data exploration.

7. Case Studies

Case studies demonstrate the successful application of BI tools in the aviation industry. Examples include optimizing flight schedules and improving on-time performance, as well as data-driven marketing strategies that boost customer retention and revenue.

8. Conclusion

Leveraging BI tools like Qlik with synthetic airline data offers significant potential for optimizing revenue, enhancing operational efficiency, and improving customer experiences. The literature underscores the importance of data-driven decision-making in achieving competitive advantages in the aviation industry.

3) Data Collection:

3.1 Collect the Dataset:

The dataset used in this project was downloaded from a reliable source that provides synthetic airline data. This dataset is designed to mimic real-world airline operations, ensuring the inclusion of various aspects necessary for comprehensive analysis.

The synthetic airline data includes detailed information on:

- **Flight Schedules**: Data on departure and arrival times, routes, and flight frequencies.
- **Passenger Demographics**: Information about passenger age, gender, travel class, and loyalty status.
- **Ticket Sales**: Historical ticket sales data, including booking dates, prices, and sales channels.
- **Performance Metrics**: Operational metrics such as on-time performance, delays, cancellations, and baggage handling statistics.

3.2 Connect Data with Qlik Cloud:

To seamlessly integrate the dataset with Qlik Cloud, the following steps were executed:

1. **Preparation**: Prior to integration, the dataset underwent thorough cleaning and formatting procedures to ensure data quality and consistency. This involved removing duplicates, addressing missing values, and standardizing data formats.
2. **Accessing Qlik Cloud**: Access to the Qlik Cloud platform was obtained by logging into the account or creating a new one if necessary. A dedicated workspace was established within Qlik Cloud to host the project.
3. **Uploading the Dataset**: Using the intuitive interface of Qlik Cloud's 'Data Manager' tool, the dataset file was uploaded from the local system. Qlik Cloud automatically detected the structure of the data and suggested initial field mappings.
4. **Integration and Data Modeling**: Following the upload, the dataset was

integrated into Qlik Cloud's environment. This involved confirming and refining field mappings, establishing connections between related data tables, and structuring the data model for optimal analysis.

5. **Verification and Validation**: To ensure data integrity and correctness, the dataset was thoroughly verified within Qlik Cloud. Data previews were examined, and initial queries were run to validate the accuracy of the imported data.
6. **Visualization and Dashboard Creation**: Leveraging Qlik Cloud's powerful visualization tools, interactive dashboards were created to visualize insights derived from the dataset. Various chart types, graphs, and visualization elements were utilized to effectively communicate key findings.
7. **Continuous Updates and Maintenance**: Scheduled data refreshes were set up within Qlik Cloud to ensure that the dataset remained up-to-date. Additionally, the data connection was monitored regularly to address any potential issues and maintain seamless integration.

By following these steps, the dataset was seamlessly connected to Qlik Cloud, enabling the project team to leverage the platform's robust BI capabilities for in-depth analysis and visualization of synthetic airline data.

4) Data Preparation:

4.1 Prepare the data for visualisation:

Data preprocessing plays a crucial role in preparing the dataset for effective visualization. The following steps were undertaken as part of the preprocessing phase:

1. **Handling Null Values**:
    - Null values were identified and handled appropriately to ensure data integrity and accuracy.
    - Missing values were either imputed using suitable techniques or removed from the dataset, depending on the impact on analysis and visualization.

2. **Removing Irrelevant Columns**:
    - Columns that did not contribute to the analysis or visualization goals were identified and dropped from the dataset.
    - Removing irrelevant columns helped streamline the dataset and focus on the most relevant information for visualization.

3. **Data Transformation**:
    - Data transformation techniques, such as normalization or standardization, were applied to ensure consistency and comparability across different variables.
    - Transformation also included converting data types to appropriate formats for visualization tools, ensuring accurate representation of the data.

4. **Handling Outliers**:
    - Outliers were identified and addressed to prevent skewing the visualization results.
    - Techniques such as winsorization or trimming were used to handle outliers without compromising the overall integrity of the dataset.

5. **Feature Engineering**:
    - Feature engineering techniques were employed to extract additional

meaningful information from the dataset.
- This involved creating new features or combining existing ones to capture complex relationships and patterns within the data.

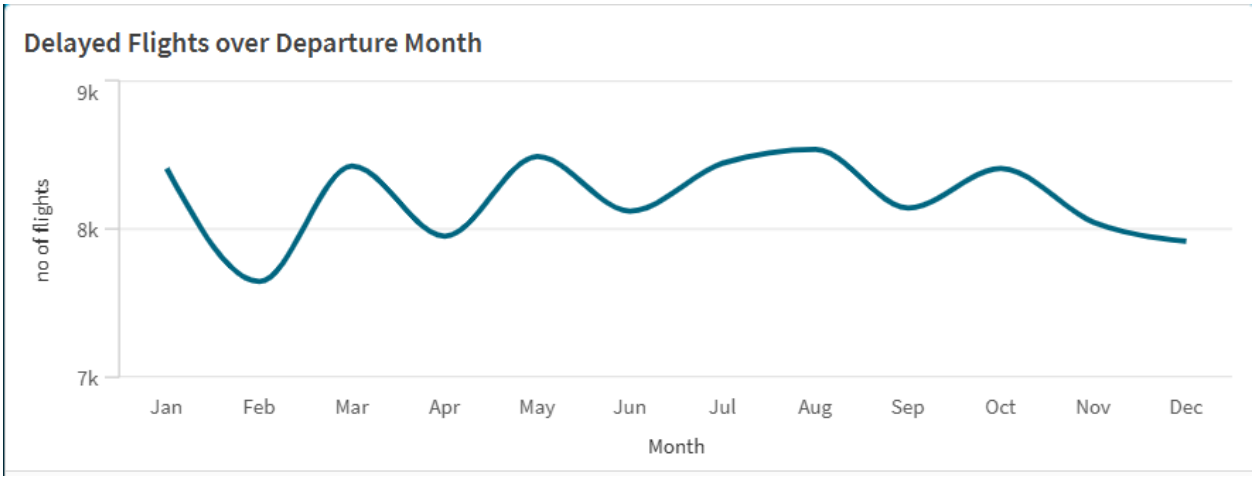
6. **Data Aggregation**:
    - Aggregating data at different levels, such as aggregating daily data into monthly or yearly summaries, was performed to facilitate higher-level analysis and visualization.
    - Aggregation helped in uncovering long-term trends and patterns that may not be apparent at a granular level.

By executing these preprocessing steps, the dataset was refined and optimized for visualization, ensuring that the resulting visualizations accurately represented the underlying insights and supported effective decision-making processes.

5) Data Visualisations:

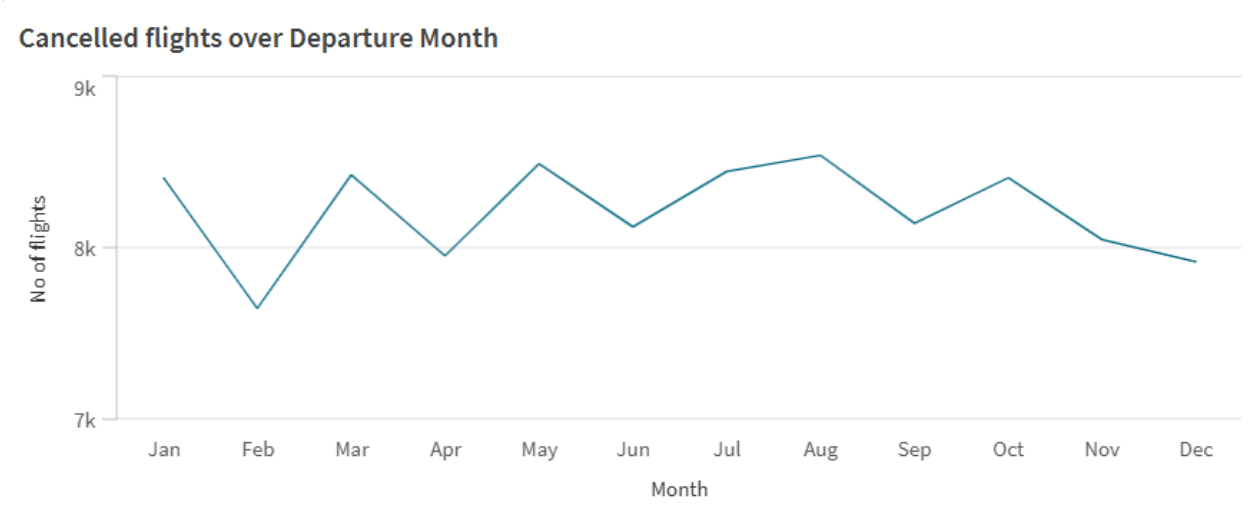
5.1 Visualisations:

- **Delayed Flights Over Departure Month**



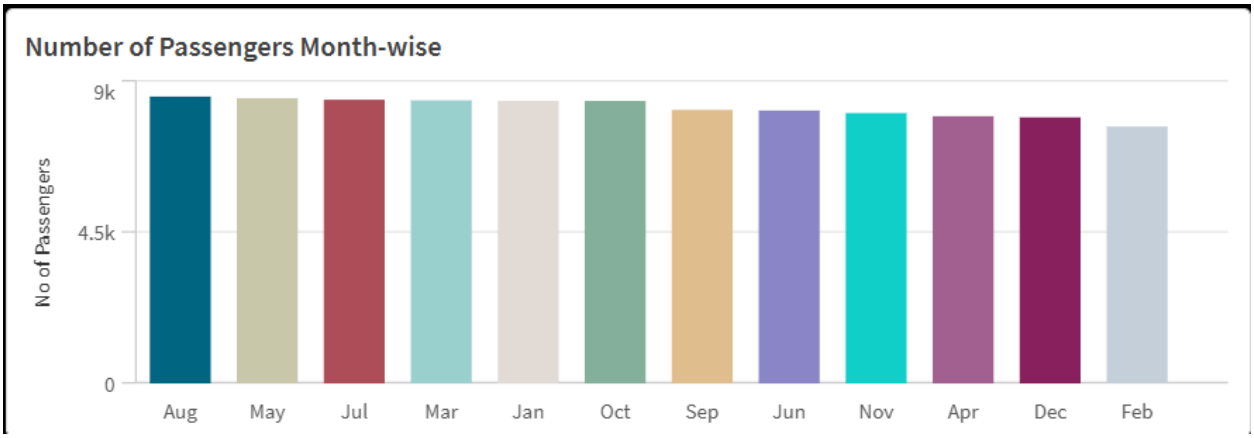
The line chart displays the trend of delayed flights over departure months, with the number of flights (y-axis) plotted against the months (x-axis). Each point on the line represents the average or total number of delayed flights for a specific month. This visualization allows for easy identification of seasonal variations or trends in flight delays throughout the year. Additionally, it provides insights into potential patterns or correlations between flight delays and specific months, enabling airlines to implement targeted strategies to mitigate delays during peak periods.

- **Cancelled Flights over departure month**:



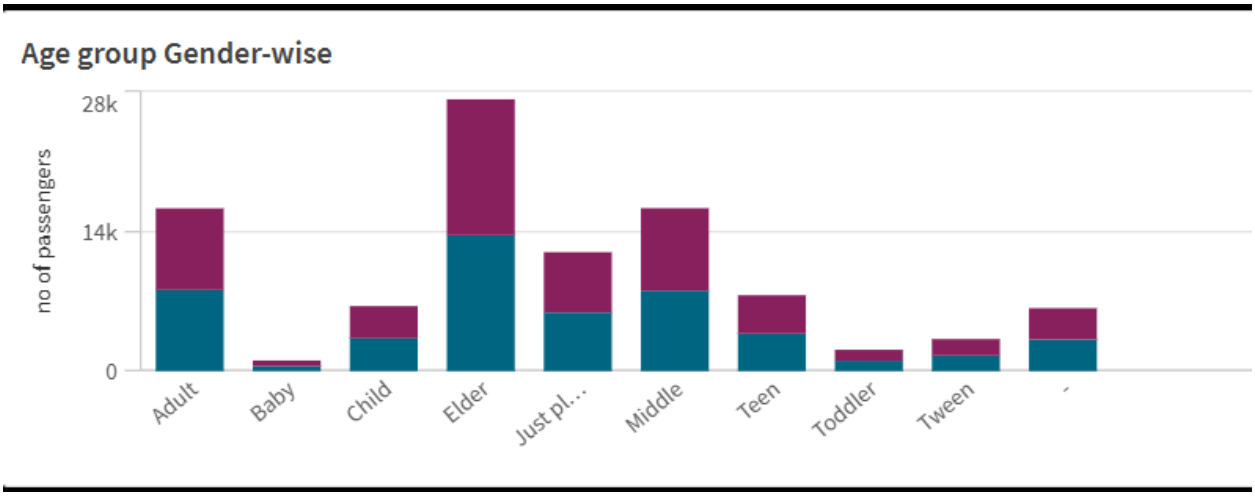
The line chart illustrates the variation in cancelled flights over departure months, with the number of flights (y-axis) depicted against the months (x-axis). Each point on the line represents the average or total number of cancelled flights for a particular month. This visualization facilitates the identification of seasonal patterns or trends in flight cancellations throughout the year. It enables stakeholders to discern any recurring patterns in cancellations and implement proactive measures to minimize disruptions during specific months, ultimately improving operational efficiency and passenger satisfaction.

- **Number of passengers month-wise:**



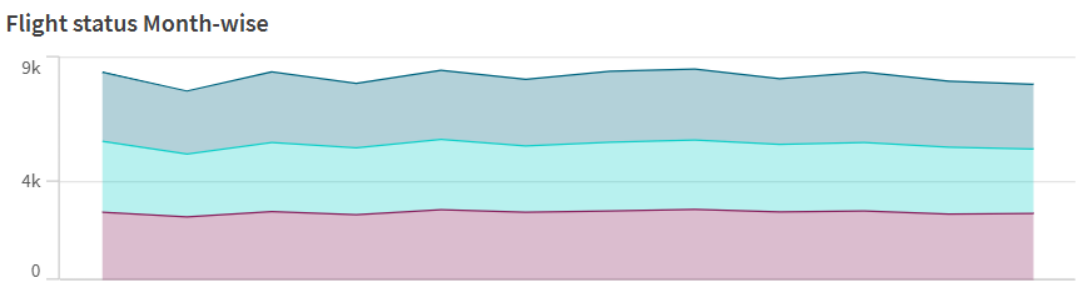
The bar chart showcases the fluctuation in the number of passengers month-wise, plotting the passenger count (y-axis) against the months (x-axis). The bars represent the total number of passengers for a specific month. This visualization offers insights into seasonal variations or trends in passenger traffic throughout the year, aiding airlines and airports in capacity planning, resource allocation, and marketing strategies.

- **Age group gender- wise**:



The visualization presents the distribution of age groups by gender, offering a concise overview of demographics within the dataset. Each age group is segmented by gender, providing insight into the composition of different demographic segments. This visualization aids in understanding the diversity of passengers based on age and gender, informing targeted marketing strategies, service offerings, and operational decisions tailored to specific demographics.

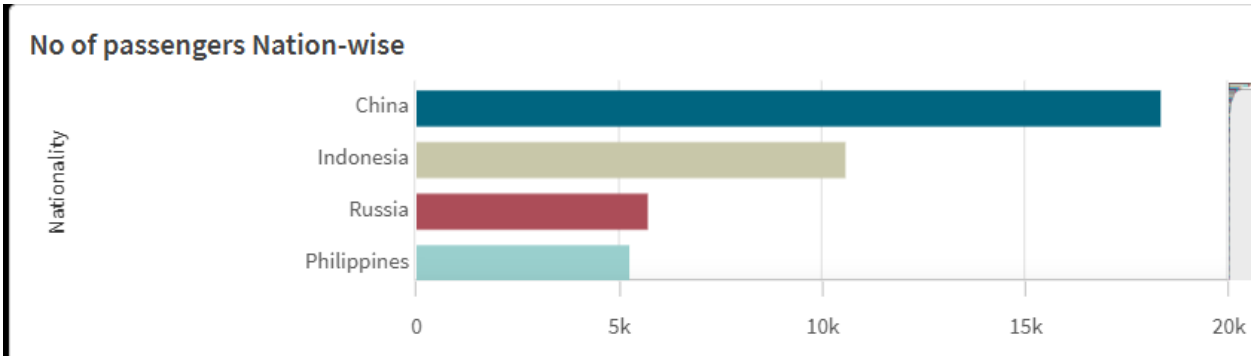
- **Flight Status Month-wise:**



:

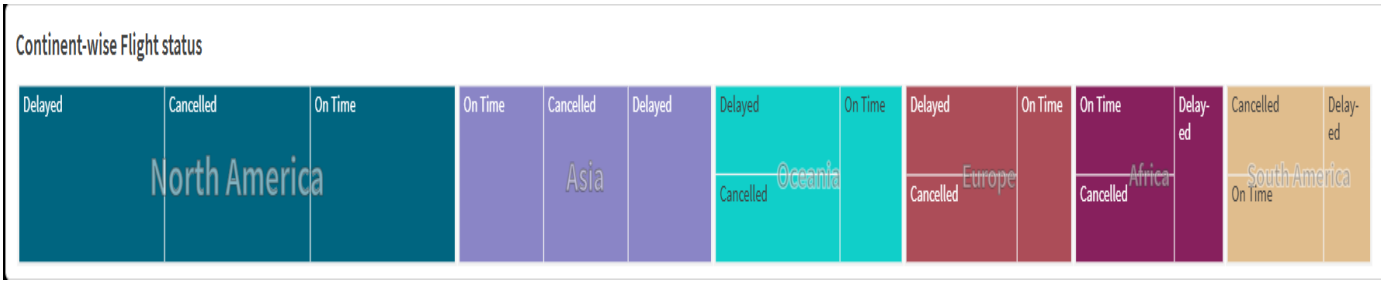
The visualization illustrates the flight status (such as delayed, cancelled, or on-time) month-wise, depicting the distribution of flights across different statuses over each month. This visualization enables stakeholders to identify trends or patterns in flight statuses over time, facilitating proactive measures to address operational challenges, minimize disruptions, and optimize overall flight performance.

- **Number of Passengers Nation-wise:**



The visualization portrays the number of passengers by nationality, offering a succinct depiction of passenger distribution across different countries. Each bar represents the total number of passengers from a specific nation. This visualization facilitates understanding of the geographic distribution of passengers, aiding in targeted marketing efforts, route planning, and catering to diverse passenger needs based on nationality.

- **Continent- wise Flight status:**

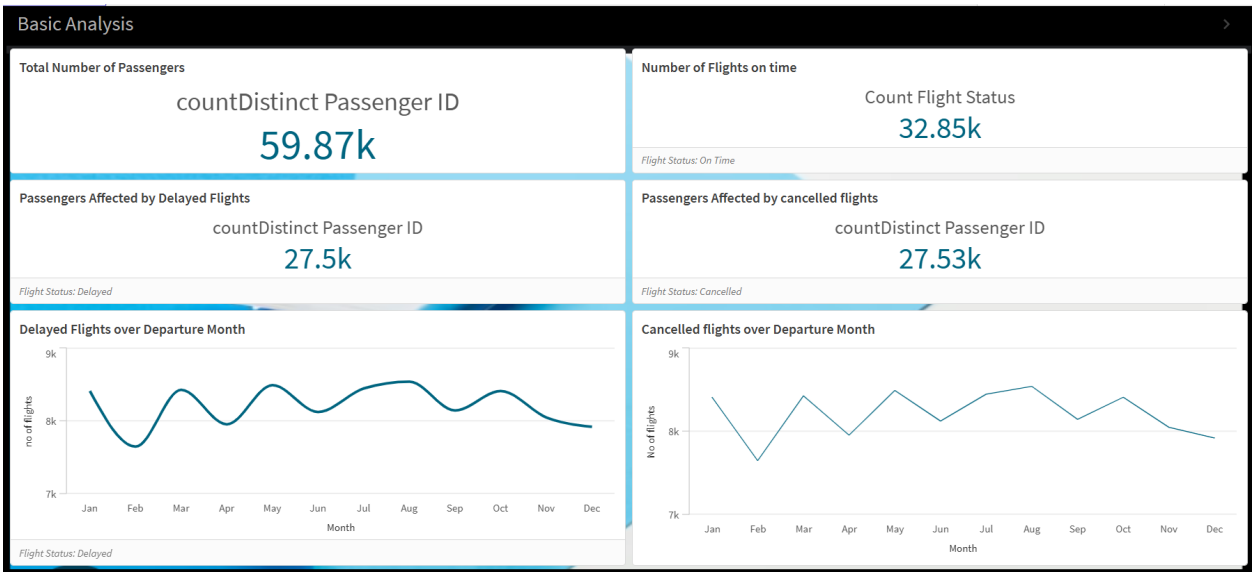


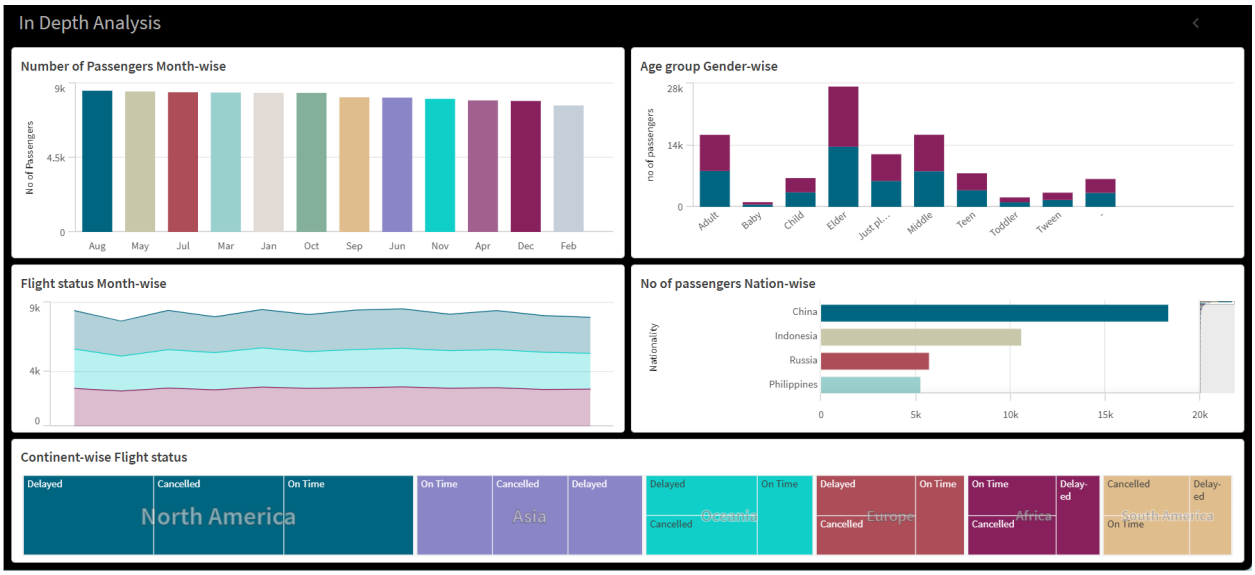
The visualization depicts the flight status (such as delayed, cancelled, or on-time)

categorized by continent, providing an overview of flight performance across different regions. Each segment represents the proportion of flights with a specific status within each continent. This visualization aids in identifying regional trends in flight punctuality and disruptions, enabling airlines to implement targeted strategies for improving operational efficiency and customer satisfaction on a continental scale.

6) DashBoard:

6.1 Responsive and Design of Dashboard:





The dashboard is designed to adapt seamlessly across devices, comprising two sheets: Basic Analysis and In-depth Analysis. The Basic Analysis sheet provides essential insights, including visualizations on delayed and cancelled flights. The In-depth Analysis sheet delves into detailed demographic trends. Notable features include fluid layout, scalable elements, and rigorous cross-device testing to ensure user-friendly navigation and optimal viewing experience.

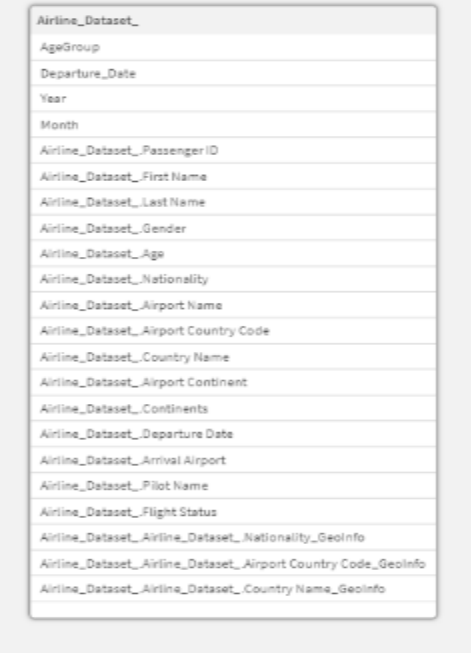
7) Report:

7.1 Report Creation:

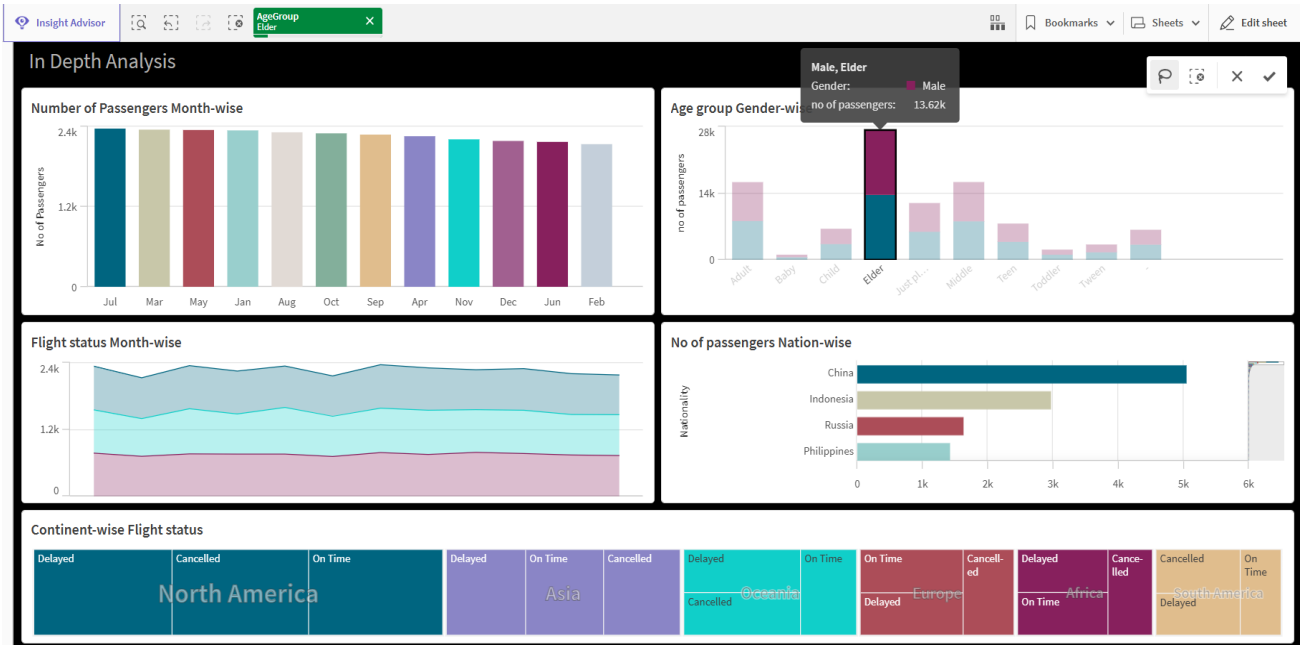
In creating the report for this project, the objective is to distill insights from the analysis of synthetic airline data with Qlik into a structured and actionable format. This involves summarizing key findings on trends in flight delays, cancellations, passenger demographics, revenue optimization, and operational efficiencies. Visualizations such as line charts, bar graphs, and geographic maps are carefully selected to effectively communicate these insights. The report is structured with clear sections covering introduction, methodology, findings, recommendations, and conclusion, with narratives crafted to provide context and interpretation of the data. After thorough review for accuracy and coherence, the finalized report is distributed to stakeholders, including airline executives and airport authorities, through various channels for informed decision-making aimed at enhancing airline operations and customer satisfaction.

8) Performance testing:

8.1 Amount of data rendered:



8.2 Utilization of Data fliters:



"Utilization of Filters" involves applying filters within a system, software application, or data processing pipeline to selectively extract, manipulate, or analyze data based on specific criteria or conditions. Filters are employed to refine the dataset, concentrating solely on pertinent information that satisfies predefined criteria.