Exploring Insights from Synthetic Airline Data Analysis with Qlik

submitted by : Dhana Rajeswari Sunkara

1. Introduction:

1.1 Overview:

- The project "Exploring Insights from Synthetic Airline Data Analysis with Qlik" involves utilizing synthetic airline data to derive valuable insights using Qlik, a business intelligence and data visualization tool.
- In this project, the synthetic airline data simulates various aspects of airline operations, including flight schedules, passenger demographics, ticket sales, and performance metrics.

1.2 purpose:

The objective is to leverage Qlik's analytical capabilities to uncover patterns, trends, and correlations within this data, aiding in decision-making processes for airlines, airports, and related stakeholders.

purpose 1: Revenue Optimization

An airline wants to optimize its revenue by analyzing historical ticket sales data, identifying peak travel times, popular destinations, and pricing strategies. Using Qlik, they can visualize revenue trends over time, segment customers based on purchasing behavior, and adjust pricing strategies accordingly to maximize profitability

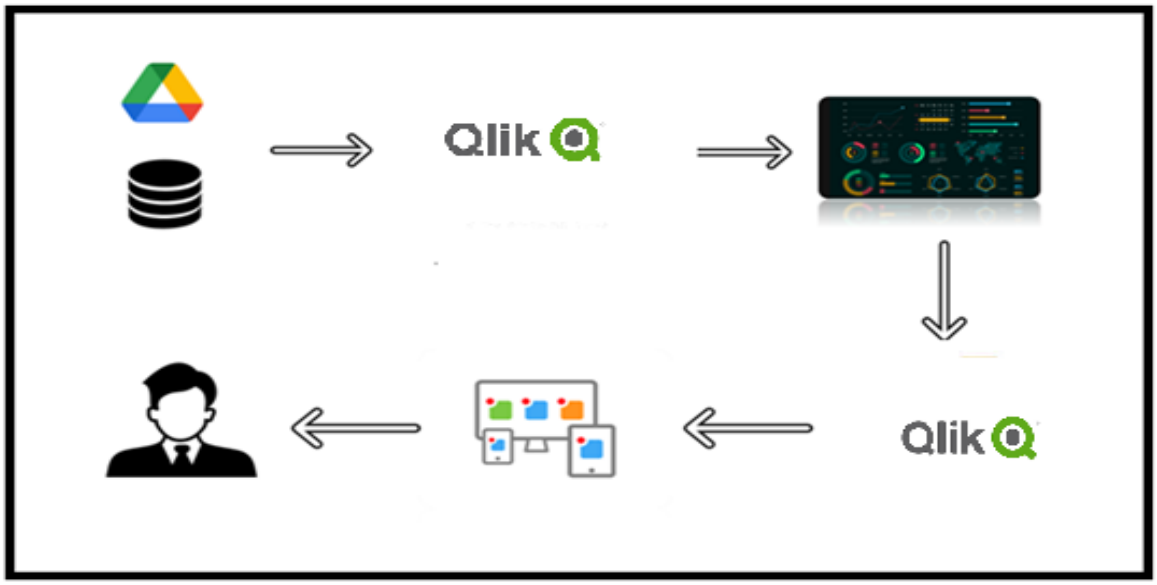
purpose 2: Operational Efficiency

An airport authority aims to enhance operational efficiency by analyzing flight schedules, passenger flows, and luggage handling processes. By integrating Qlik with synthetic airline data, they can identify bottlenecks in airport operations, predict peak traffic periods, and allocate resources effectively to streamline processes and improve overall efficiency.

purpose 3: Customer Experience Enhancement

Airlines are keen to enhance the passenger experience by understanding customer preferences, satisfaction levels, and pain points. Through sentiment analysis on customer feedback data integrated with Qlik, airlines can identify areas for improvement, personalize services, and tailor marketing campaigns to better meet customer needs, ultimately fostering loyalty and satisfaction.

1.3 Technical Architecture:



2. Define Problem / Problem:

2.1 Specify the business problem

"Exploring Insights From Synthetic Airline Data Analysis With Qlik" could address several business problems within the airline industry, including:

1.Operational Efficiency:

Airlines deal with vast amounts of data related to flight schedules, crew management, maintenance, and passenger information. Analyzing this data with tools like Qlik can uncover inefficiencies in operations, such as flight delays, crew scheduling conflicts, or underutilized resources.

2.Customer Experience Enhancement:

Understanding passenger behavior, preferences, and trends can help airlines personalize services, improve on-board experiences, and optimize pricing strategies. Analyzing synthetic data with Qlik can reveal insights into customer satisfaction, loyalty drivers, and areas for service enhancement.

3.Revenue Optimization:

Airlines need to maximize revenue from ticket sales, ancillary services, and partnerships. Synthetic data analysis with Qlik can identify revenue-generating opportunities, such as dynamic pricing adjustments, targeted marketing campaigns, or strategic alliances based on demand patterns and market trends.

4.Safety and Compliance:

Safety is paramount in the airline industry, and compliance with regulations is mandatory. Analyzing synthetic data can help airlines monitor safety metrics, identify potential risks, and ensure compliance with regulatory standards, leading to safer operations and reduced liabilities.

5.Cost Reduction:

Airlines face significant cost pressures from fuel prices, labor expenses, and operational overheads. Analyzing synthetic data with Qlik can uncover cost-saving opportunities, such as optimizing fuel consumption, streamlining maintenance schedules, or renegotiating vendor contracts based on historical data and predictive analytics.

6.Risk Management:

Airlines are exposed to various risks, including economic downturns, geopolitical instability, and natural disasters. Analyzing synthetic data can help identify and mitigate risks by providing insights into market trends, demand fluctuations, and operational vulnerabilities, enabling proactive risk management strategies.

By addressing these business problems through the exploration of insights from synthetic airline data analysis with Qlik, airlines can enhance their competitiveness, improve operational performance, and drive sustainable growth in a dynamic and challenging industry landscape.

2.2 Business Requirements:

When exploring insights from synthetic airline data analysis with Qlik, the following business requirements should be considered:

1.Data Integration:

Ensure seamless integration of diverse data sources such as flight schedules, passenger demographics, crew information, maintenance logs, and weather data into the Qlik platform.

2.Data Quality Assurance:

Implement robust data validation and cleansing processes within Qlik to ensure the accuracy, completeness, and consistency of the synthetic airline data.

3.Visualization and Dashboarding:

Develop interactive dashboards and visualizations in Qlik that provide intuitive and insightful representations of the synthetic data, allowing users to explore trends, correlations, and anomalies effectively.

4.Advanced Analytics Capabilities:

Leverage Qlik's advanced analytics features, such as predictive modeling, clustering, and trend analysis, to derive deeper insights from the synthetic airline data and support data-driven decision-making.

5.Self-Service Analytics:

Empower business users with self-service analytics capabilities in Qlik, enabling them to explore and analyze synthetic airline data independently without extensive technical expertise.

6.Real-Time Data Analysis:

Enable real-time or near-real-time analysis of synthetic airline data in Qlik to support dynamic decision-making and proactive management of operational challenges such as flight delays, crew scheduling issues, or equipment failures.

7.Mobile Access:

Provide mobile access to Qlik dashboards and reports, allowing stakeholders to access critical insights and make informed decisions while on the go.

8.Security and Governance:

Implement robust security measures and data governance policies within Qlik to protect sensitive airline data and ensure compliance with regulatory requirements such as GDPR, HIPAA, or industry-specific standards.

9.Collaboration and Sharing:

Facilitate collaboration and knowledge sharing among stakeholders by enabling the sharing of Qlik apps, dashboards, and insights across teams and departments within the airline organization.

10.Performance Optimization:

Optimize the performance of Qlik applications and dashboards to ensure fast response times and smooth user experience, even when analyzing large volumes of synthetic airline data.

By addressing these business requirements, organizations can effectively leverage Qlik for exploring insights from synthetic airline data analysis, driving operational excellence, enhancing customer experiences, and gaining a competitive edge in the aviation industry.

2.3 Literature Survey:

A literature survey exploring insights from synthetic airline data analysis would likely cover a range of topics related to data analysis techniques, tools, and applications specific to the aviation industry. Here's an outline of what such a survey might include:

1. Introduction to Synthetic Data Analysis:

- Definition of synthetic data and its role in data analysis.
- Overview of the importance and benefits of using synthetic data in the aviation sector.

2. Aviation Data Sources and Characteristics:

- Discussion on various types of data sources available in the aviation industry, including flight data, passenger data, crew data, maintenance logs, and external factors like weather and air traffic.
- Analysis of the unique characteristics and challenges associated with aviation data, such as high volume, variety, velocity, and complexity.

3.Data Analysis Techniques:

- Review of traditional and advanced data analysis techniques applicable to synthetic airline data, including descriptive statistics, predictive modeling, clustering, anomaly detection, and time series analysis.
- Exploration of machine learning algorithms and artificial intelligence methods commonly used for aviation data analysis.

4.Tools and Platforms for Aviation Data Analysis:

- Overview of data analytics tools and platforms suitable for analyzing synthetic airline data, with a focus on Qlik and other industry-standard solutions.
- Comparative analysis of features, capabilities, and use cases of different analytics platforms for aviation data analysis.

5.Applications of Synthetic Data Analysis in Aviation:

- Examination of various applications and use cases of synthetic data analysis in the aviation industry, such as flight optimization, route planning, crew scheduling, predictive maintenance, and passenger experience enhancement.
- Case studies and real-world examples illustrating successful implementations of synthetic data analysis techniques in aviation organizations.

6. Challenges and Opportunities:

- Identification and discussion of challenges and limitations associated with synthetic airline data analysis, such as data quality issues, regulatory compliance, and privacy concerns.
- Exploration of emerging trends, opportunities, and future directions in aviation data analysis, including the integration of IoT, big data, and cloud computing technologies.

7.Best Practices and Recommendations:

- Compilation of best practices, methodologies, and guidelines for conducting synthetic data analysis in the aviation sector, ensuring data integrity, security, and regulatory complian

- Recommendations for aviation organizations seeking to leverage synthetic data analysis for operational improvement, cost reduction, and strategic decision-making.

8.Conclusion and Future Directions:

- Summary of key findings from the literature survey and implications for the aviation industry.
- Discussion of potential future research directions and areas for further exploration in the field of synthetic airline data analysis.

By conducting a comprehensive literature survey on exploring insights from synthetic airline data analysis, researchers and practitioners can gain valuable insights into the state-of-the-art techniques, tools, and applications shaping the future of data-driven decision-making in the aviation sector.

3.Data Collection:

3.1 Collect the Dataset

Data collection is the process of gathering and measuring information on variables of interest, in an established systematic fashion that enables one to answer stated research questions, test hypotheses, evaluate outcomes and generate insights from the data.

Data set Contains the Following :

- 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

connect Data With Qlik Sense you have upload the dataset in it.create app then upload the data in it you can add the columns to it by using the add field then click the load data then its automatically connect data with qlik. Data contains all the meta information regarding the columns in the CSV

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. Since the data is already cleaned, we can move to visualization.

Data Loading:

First, you have to create the app.In the add new option contain the multiple option from that options you select the "New Analytic App" .click on it then give name to the app then create after creation you have to upload the data set into the app.

Data Cleaning and Pre-Processing:

In the data cleaning and pre-processing you have to add the columns to the uploaded data set .

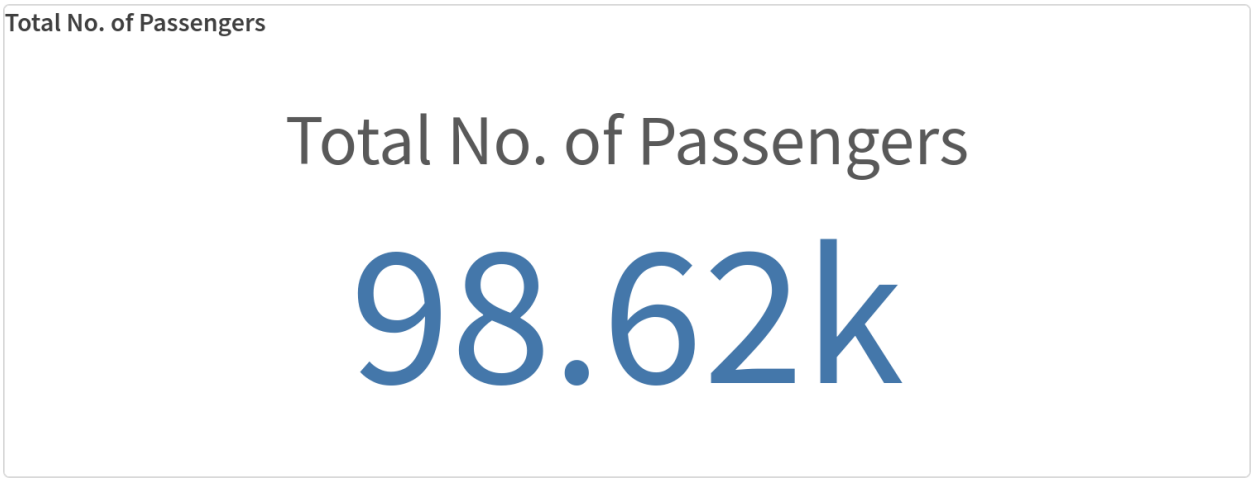
```
if(Age >= 0 AND Age <= 1, 'Baby',
if(Age >= 1 AND Age <= 3, 'Toddler',
if(Age >= 4 AND Age <= 9, 'Child',
if(Age >= 10 AND Age <= 12, 'Tween',
if(Age >= 13 AND Age <= 19, 'Teen',
if(Age >= 20 AND Age <= 24, 'Young Adul',
if(Age >= 25 AND Age <= 39, 'Adult',
if(Age >= 40 AND Age <= 54, 'Middle',
if(Age >= 55 AND Age <= 79, 'Elder',
if(Age >= 80,'Just Plain Old'))))))))) As AgeGroup,
Date#([Departure Date], 'MM/DD/YYYY') as [Departure_Date],
Year([Departure Date]) AS Year,
MOnth([Departure Date]) as Month
```

From the Above you have to add the age goup , date , year , month columns to the uploaded data set.

5.Data 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.I have done some visualization according to Airline Data Analysis

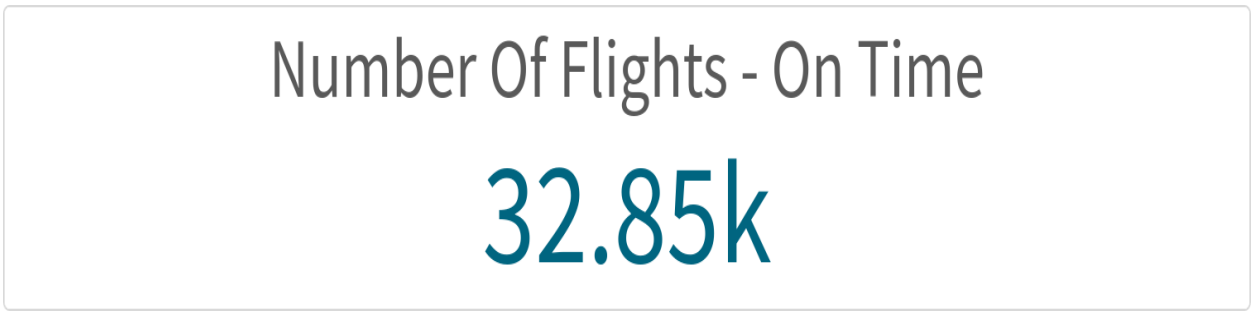
1.Total No. of Passengers



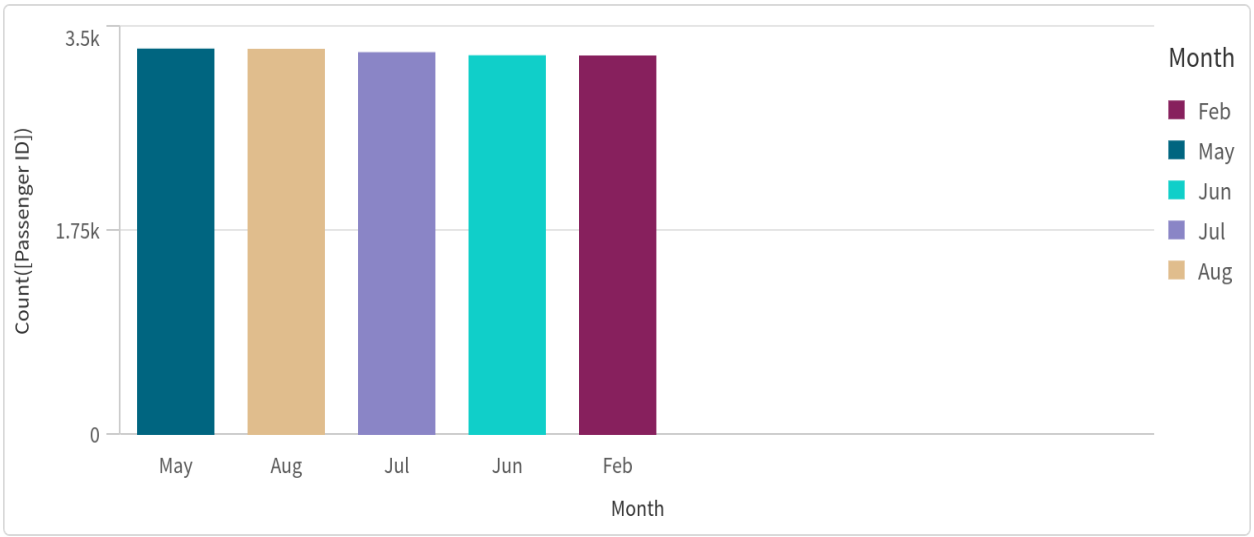
2.Number of Passengers effected by cancelled flights

3.No. of Passengers Effected by delay of flights

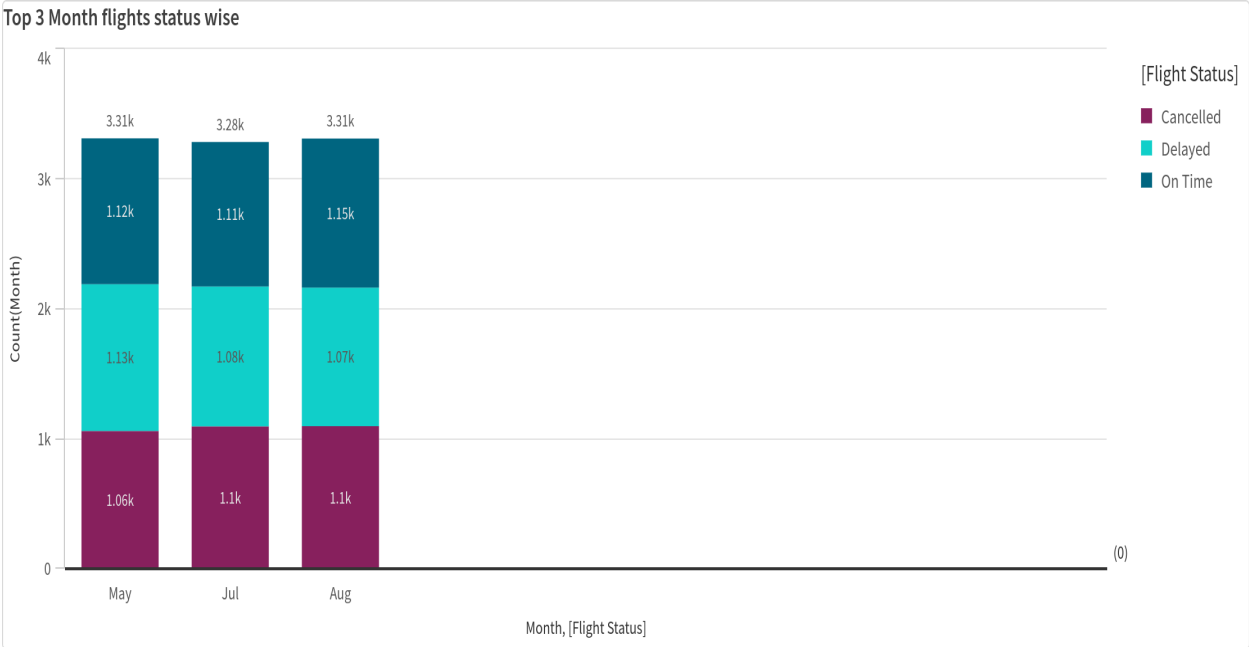
4. No of Flights on Time



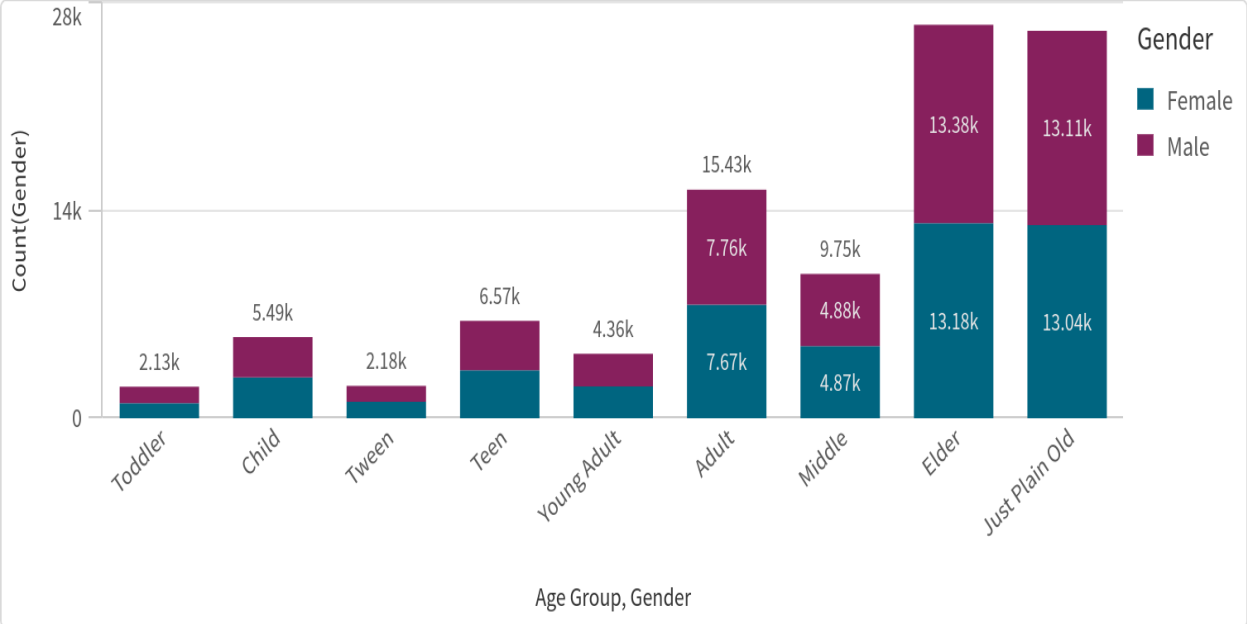
5.No of Passengers travelled- Month Wise



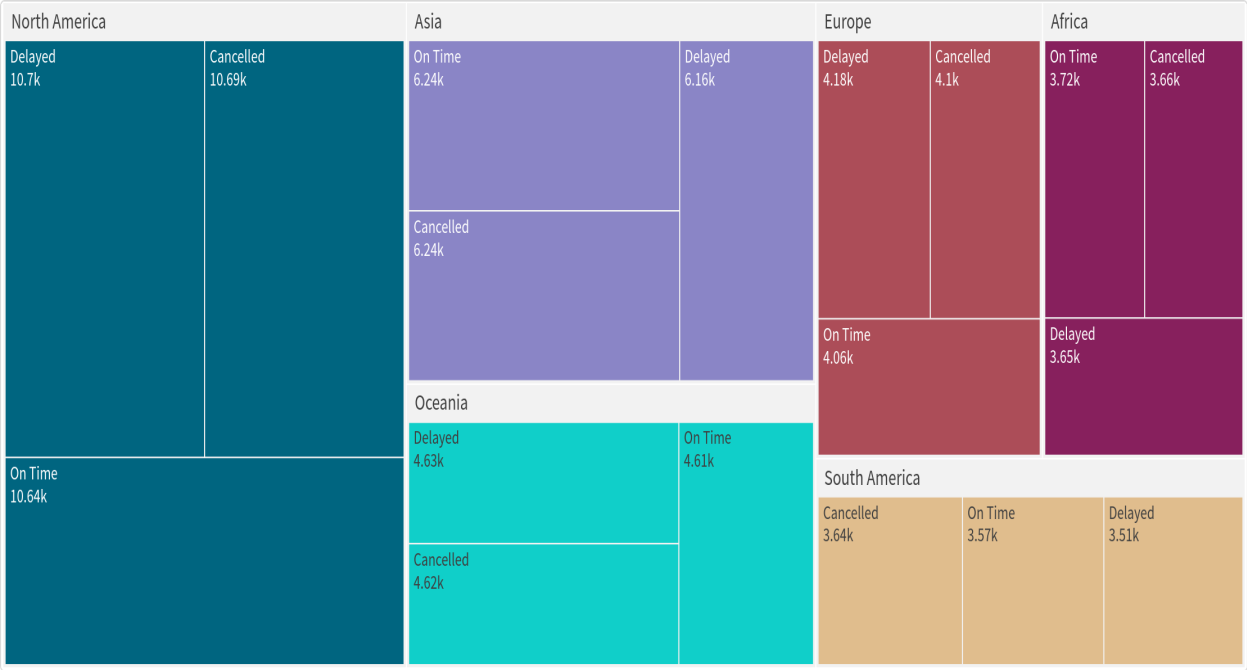
6.Top 3 Month flights status wise



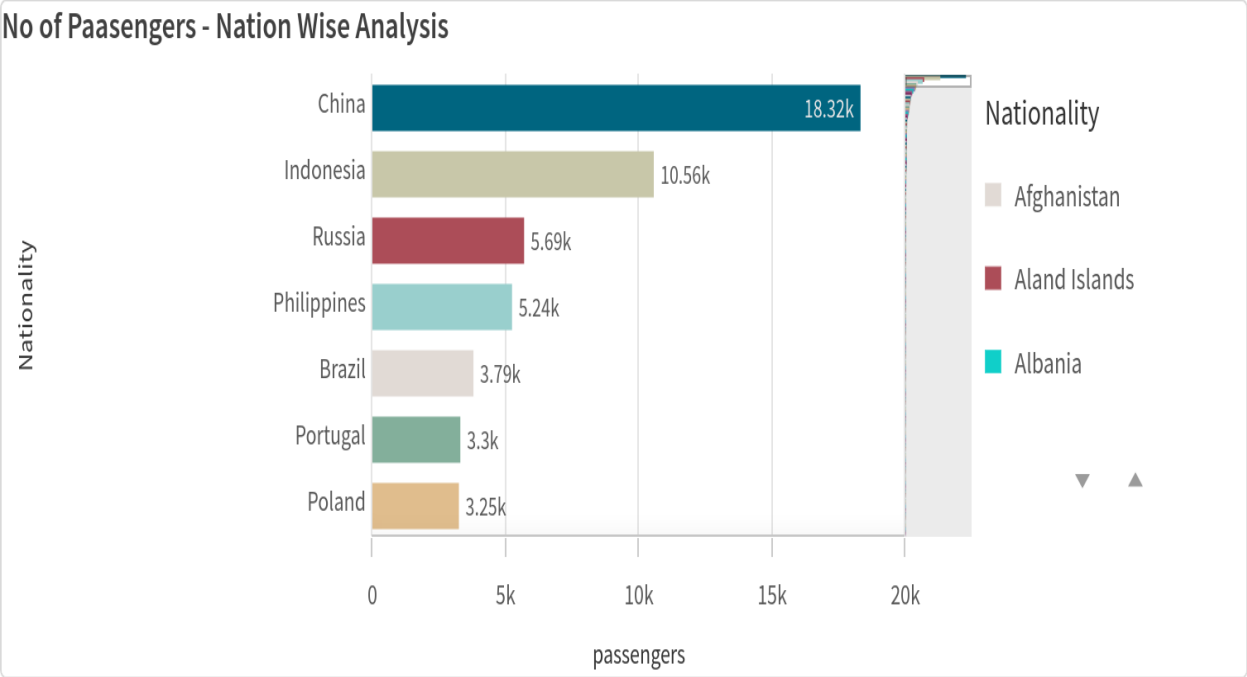
7.Age group of passengers as per gender wise



8.Continent wise flight status



9.No of Paasengers - Nation Wise Analysis



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.I have done some dashboards.

Dashboard1: Link:

https://evfwbvkodmf83ja.sg.qlikcloud.com/sense/app/0820ac19-34ca-405a-9601-f50ed64d755b/sheet/ApqDbM/state/analysis

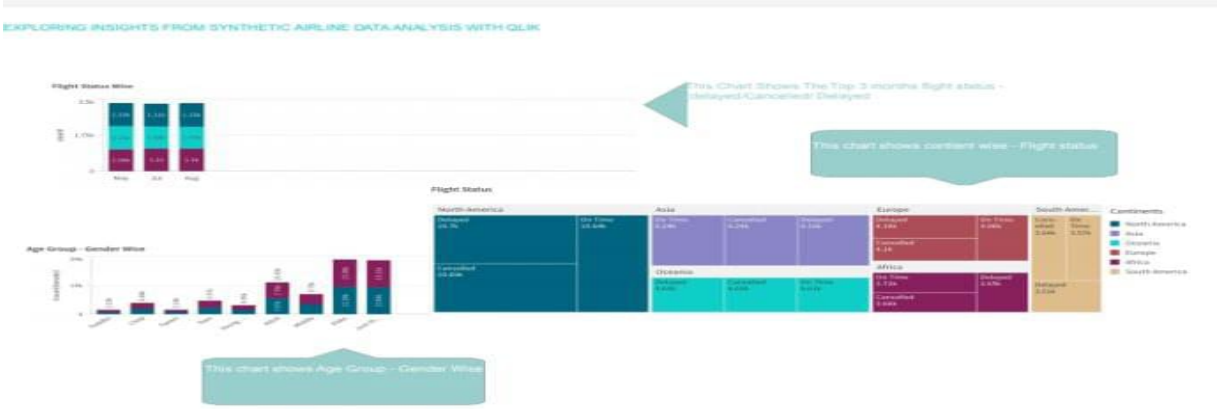
Dashboard 2: Link

https://evfwbvkodmf83ja.sg.qlikcloud.com/sense/app/0820ac19-34ca-405a-9601-f50ed64d755b/sheet/JjvQRQ/state/analysis

7 .Report:

7.1 Report Creation

In the Report Creation Story is the one of the roleA data story is a way of presenting data and analysis in a narrative format, with the goal of making the information more engaging and easier to understand. A data story typically includes a clear introduction that sets the stage and explains the context for the data, a body that presents the data and analysis in a logical and systematic way, and a conclusion that summarizes the key findings and highlights their implications. Data stories can be told using a variety of mediums, such as reports, presentations, interactive visualizations, and videos.



8.PerformanceTesting:

8.1 Amount Of Data Rendered

link for data loaded:
https://evfwbvkodmf83ja.sg.qlikcloud.com/sense/app/0820ac19-34ca-405a-9601-f50ed64d755b

Airline Dataset Updated - v2
Airline Dataset Updated - v2.csv
Columns: 18
Rows: 98619

Unpivot | Add field | Select data from source

Airport Count...	Country Name	Airport Conti...	Continents	Departure Date	Arrival Airport	Pilot Name	Flight Status	Year	Month	Age Group
BR	Brazil	SAM	South America	-	0	Bert Insworth	Cancelled	-	-	Elder
IE	Ireland	EU	Europe	12/8/2022	DUB	Griswold Hugli	Cancelled	2022	Dec	Middle
US	United States	NAM	North America	-	HAR	Robenia Willford	On Time	-	-	Teen
CO	Colombia	SAM	South America	-	VGZ	Shae Hercock	On Time	-	-	Elder
CN	China	AS	Asia	11/4/2022	BPE	Nessi Clandillon	Cancelled	2022	Nov	Elder
US	United States	NAM	North America	9/9/2022	WYB	Odelia Dunbabin	Delayed	2022	Sep	Adult
US	United States	NAM	North America	-	VDZ	Norma Jellico	Delayed	-	-	Middle
EG	Egypt	AF	Africa	10/2/2022	HBE	Joy Brose	Cancelled	2022	Oct	Just Plain Old
KE	Kenya	AF	Africa	-	MRE	Samaria Woolway	On Time	-	-	Middle
CN	China	AS	Asia	-	PKX	Terrel Carillo	On Time	-	-	Just Plain Old
MG	Madagascar	AF	Africa	-	WTA	Zollie Keynd	On Time	-	-	Elder
PG	Papua New Guinea	OC	Oceania	-	TLP	Hermione Giovannacc@i	Cancelled	-	-	Middle
ID	Indonesia	AS	Asia	-	RSK	Zulema Rosson	On Time	-	-	Just Plain Old
AU	Australia	OC	Oceania	5/10/2022	RPM	Theresa Huddleston	On Time	2022	May	Just Plain Old
IS	Iceland	EU	Europe	12/5/2022	VEY	Antonia Kulis	On Time	2022	Dec	Elder
ZA	South Africa	AF	Africa	-	OVG	Collie Abbes	On Time	-	-	Just Plain Old
AZ	Azerbaijan	AS	Asia	-	LLK	Francois Tolan	Cancelled	-	-	Adult
LR	Liberia	AF	Africa	7/9/2022	VOI	Harris Costock	On Time	2022	Jul	Middle
US	United States	NAM	North America	-	STE	Babita Schwand	On Time	-	-	Just Plain Old
AR	Argentina	SAM	South America	-	CNT	Iolande Grocock	Delayed	-	-	Adult
TD	Chad	AF	Africa	-	OUT	Elicia Pisculli	Delayed	-	-	Adult

8.2 utilization of Data filters