EXPLORING INSIGHTS FROM SYNTHETIC AIRLINE DATA ANALYSIS WITH QLIK

PROJECT REPORT

DONE BY:

PAIDI MANEESH SAI

1 INTRODUCTION

1.1 Overview: A brief description of the project

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. 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.

1.2 Purpose: The use of this project and what can be achieved using this

Scenario 1: Revenue Optimization
An airline wants to optimize its revenue by analysing 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 behaviour, and adjust pricing strategies accordingly to maximize profitability.

Scenario 2: Operational Efficiency
An airport authority aims to enhance operational efficiency by analysing 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.

Scenario 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 UNDERSTANDING

2.1 Specify the business Problem

The project "Exploring Insights from Synthetic Airline Data Analysis with Qlik" addresses multiple business problems for airlines and related stakeholders, focusing on optimizing revenue, improving operational efficiency, and enhancing customer experience. Specifically:

2.2 Business Requirements

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

1. Data Integration and Management:

- Consolidate synthetic airline data, including flight schedules, passenger demographics, ticket sales, performance metrics, and customer feedback.

- Ensure data accuracy, consistency, and completeness for reliable analysis.

2. Revenue Optimization:

- Develop Qlik dashboards to visualize revenue trends over time.

- Segment customers based on purchasing behavior and identify peak travel times and popular destinations.

- Implement pricing strategy models to simulate and adjust pricing for maximizing profitability.

3. Operational Efficiency:

- Create Qlik dashboards to monitor flight schedules, passenger flows, and luggage handling processes.

- Identify operational bottlenecks and predict peak traffic periods using historical data analysis.

- Develop resource allocation models to streamline processes and improve overall efficiency.

4. Customer Experience Enhancement:

- Conduct sentiment analysis on customer feedback data integrated with Qlik.

- Visualize customer preferences, satisfaction levels, and pain points.

- Identify areas for improvement and personalize services based on customer insights.

- Tailor marketing campaigns to meet customer needs and foster loyalty.

2.3 Literature Survey

Revenue Optimization in Airlines

1. Revenue Management: Talluri, K.T., & Van Ryzin, G.J. (2004). "The Theory and Practice of Revenue Management." Kluwer Academic Publishers. This book provides a comprehensive framework for understanding revenue management techniques, which can be applied to optimize airline ticket sales.

2. Dynamic Pricing Strategies: Phillips, R. (2005). "Pricing and Revenue Optimization." Stanford University Press. This literature discusses dynamic pricing strategies and models that can be used to maximize airline revenue by adjusting prices based on demand and other factors.

Operational Efficiency in Airports

1. Airport Operations Management: de Neufville, R., & Odoni, A. (2003). "Airport Systems: Planning, Design, and Management." McGraw-Hill. This book covers various aspects of airport operations management, including resource allocation and process optimization.

2. Passenger Flow Analysis: Yeh, C.H., & Kuo, Y.H. (2003). "Evaluating passenger services of Asia-Pacific international airports." Transportation Research Part E: Logistics and Transportation Review, 39(1), 35-48. This paper discusses methodologies for evaluating and improving passenger services and flow within airports.

Customer Experience in Airlines

1. Customer Satisfaction and Loyalty: Hallowell, R. (1996). "The relationships of customer satisfaction, customer loyalty, and profitability: an empirical study." International Journal of Service Industry Management, 7(4), 27-42. This study explores the link between customer satisfaction, loyalty, and profitability, providing insights for enhancing customer experience in airlines.

2. Sentiment Analysis: Liu, B. (2012). "Sentiment Analysis and Opinion Mining." Morgan & Claypool Publishers. This book covers techniques for sentiment analysis, which can be applied to analyze customer feedback and improve service quality.

3. Personalized Marketing: Kotler, P., & Armstrong, G. (2017). "Principles of Marketing." Pearson. This textbook provides foundational knowledge on personalized marketing strategies that can be used to tailor services and campaigns to individual customer preferences.

By leveraging these literatures, the project can effectively address the business problems and meet the outlined requirements using Qlik's analytical and visualization capabilities.

3 DATA COLLECTION

3.1 Collect the dataset

Download the dataset

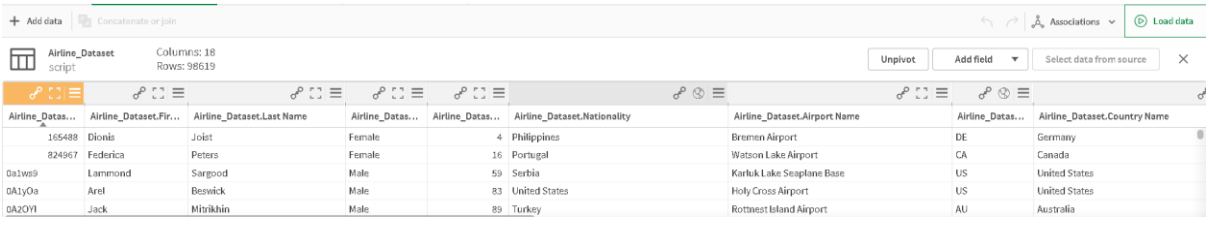
Link -
https://drive.google.com/file/d/1G7ZnBIhjkpn2R1e5J0lbB2Dnkcu6dw/view?usp=drive_link

Understand the dataset

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, cancelled)

3.2 Connect the Data with Qlik Sense

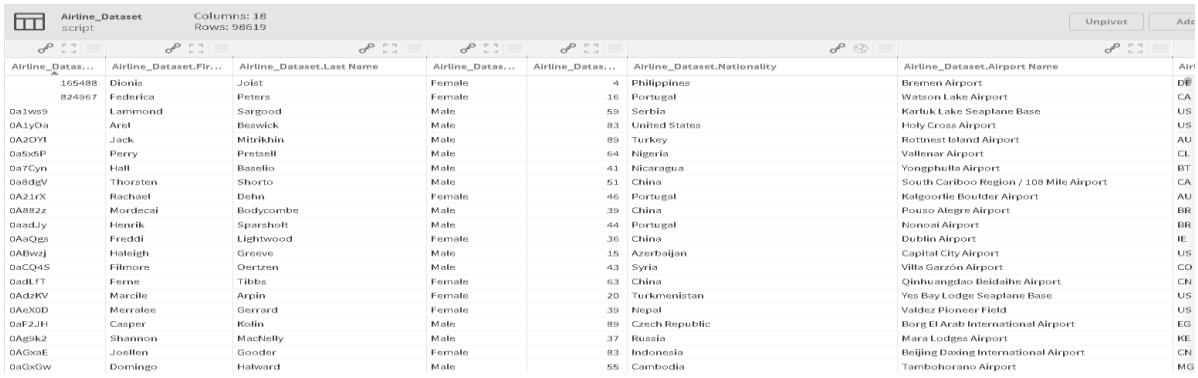



```
[Airline_Dataset]:

LOAD *,
if(Age >= 0 AND Age <= 1, 'Baby',
    if(Age > 1 AND Age <= 3, 'Toddler',
    if(Age > 3 AND Age <= 9, 'Child',
    if(Age > 9 AND Age <= 12, 'Tween',
    if(Age > 12 AND Age <= 19, 'Teen',
    if(Age > 19 AND Age <= 24, 'Young Adult',
    if(Age > 24 AND Age <= 39, 'Adult',
    if(Age > 39 AND Age <= 54, 'Middle',
    if(Age > 54 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
RESIDENT [Airline_Dataset]
WHERE NOT ([Arrival Airport] = '0' OR [Arrival Airport] = '-');
```

4 DATA PREPARATION

4.1 Prepare the data for Visualization

Data Loading

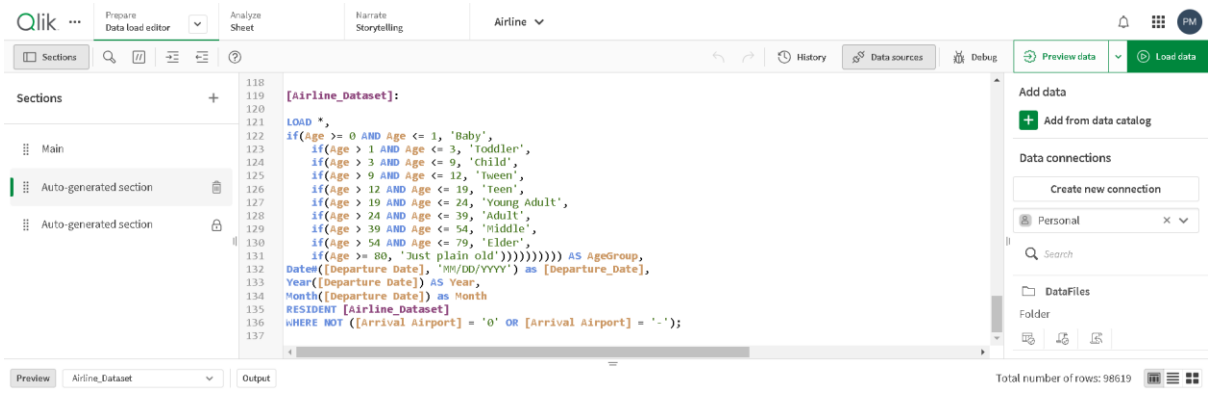


Airline_Dataset
script
Columns: 18
Rows: 98619

Unpivot

Airline_Datas...	Airline_Dataset.Fir...	Airline_Dataset.Last Name	Airline_Datas...	Airline_Datas...	Airline_Dataset.Nationality	Airline_Dataset.Airport Name	Air
165488	Dionis	Joist	Female	4	Philippines	Bremen Airport	DE
824967	Federica	Peters	Female	16	Portugal	Watson Lake Airport	CA
0a1ws9	Lammond	Sargood	Male	59	Serbia	Karluk Lake Seaplane Base	US
0A1yDa	Arel	Beswick	Male	83	United States	Holy Cross Airport	US
0A2OY1	Jack	Mitrikhin	Male	89	Turkey	Rottnest Island Airport	AU
0a5x5P	Perry	Pretsell	Male	64	Nigeria	Vallenar Airport	CL
0a7Cyn	Hall	Baselio	Male	41	Nicaragua	Yongphulla Airport	BT
0a8dgV	Thorsten	Shorto	Male	51	China	South Cariboo Region / 108 Mile Airport	CA
0A21rX	Rachael	Dehn	Female	46	Portugal	Kalgoorlie Boulder Airport	AU
0A882z	Mordecai	Bodycombe	Male	39	China	Pouso Alegre Airport	BR
0aadJy	Henrik	Sparsholt	Male	44	Portugal	Nonoai Airport	BR
0AaQga	Freddi	Lightwood	Female	36	China	Dublin Airport	IE
0ABwzj	Haleigh	Greeve	Male	15	Azerbaijan	Capital City Airport	US
0aCQ4S	Filmore	Oertzen	Male	43	Syria	Villa Garzón Airport	CO
0adLFT	Ferne	Tibbs	Female	63	China	Qinhuangdao Beidaihe Airport	CN
0AdzKV	Marcile	Arpin	Female	20	Turkmenistan	Yes Bay Lodge Seaplane Base	US
0AeX0D	Merralee	Gerrard	Female	39	Nepal	Valdez Pioneer Field	US
0aF2JH	Casper	Kolin	Male	89	Czech Republic	Borg El Arab International Airport	EG
0Ag9k2	Shannon	MacNelly	Male	37	Russia	Mara Lodges Airport	KE
0AGxaE	Joellen	Gooder	Female	83	Indonesia	Beijing Daxing International Airport	CN
0aGxGw	Domingo	Halward	Male	55	Cambodia	Tambohorano Airport	MG

Data Cleaning and Pre-Processing



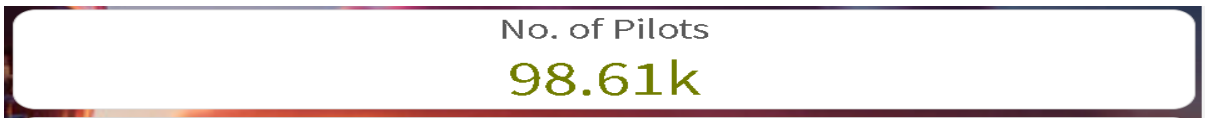
Qlik
Prepare: Data load editor
Analyze: Sheet
Narrate: Storytelling
Airline
Sections
History
Data sources
Debug
Preview data
Load data
Sections
Main
Auto-generated section
Auto-generated section

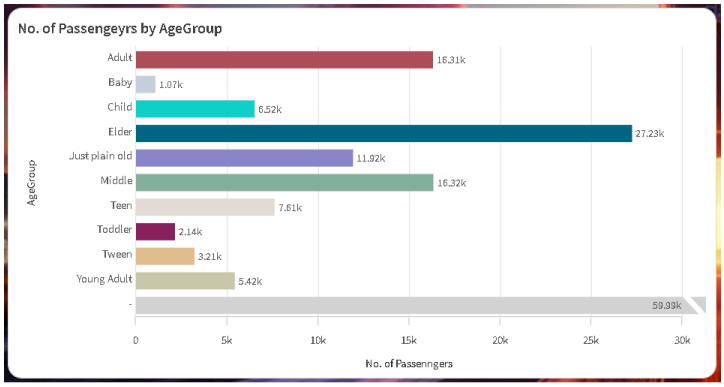
```
[Airline_Dataset]:

LOAD *,
if(Age >= 0 AND Age <= 1, 'Baby',
    if(Age > 1 AND Age <= 3, 'Toddler',
    if(Age > 3 AND Age <= 9, 'Child',
    if(Age > 9 AND Age <= 12, 'Tween',
    if(Age > 12 AND Age <= 19, 'Teen',
    if(Age > 19 AND Age <= 24, 'Young Adult',
    if(Age > 24 AND Age <= 39, 'Adult',
    if(Age > 39 AND Age <= 54, 'Middle',
    if(Age > 54 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
RESIDENT [Airline_Dataset]
WHERE NOT ([Arrival Airport] = '0' OR [Arrival Airport] = '-');
```

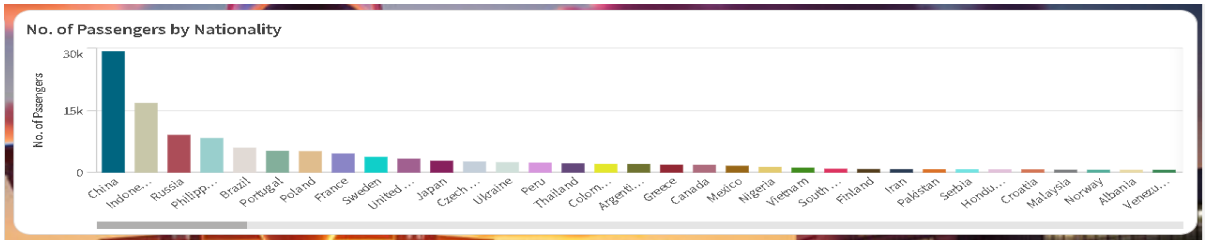
Add data
Add from data catalog
Data connections
Create new connection
Personal
Search
DataFiles
Folder
Preview
Airline_Dataset
Output
Total number of rows: 98619

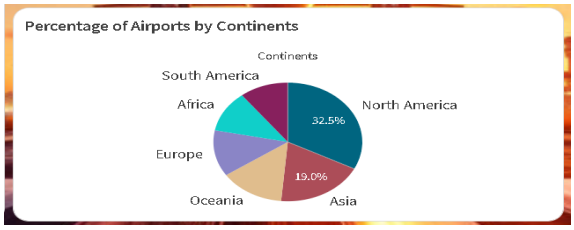
5 DATA VISUALISATIONS

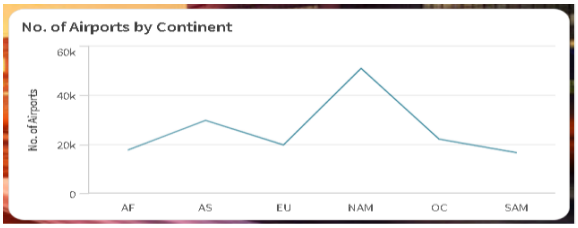
5.1 Visualisations







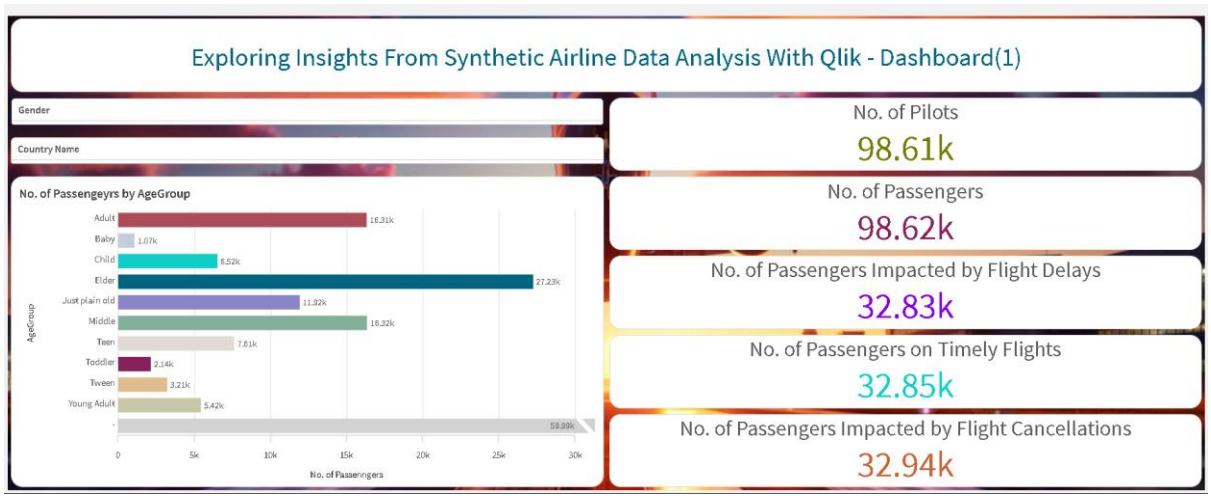




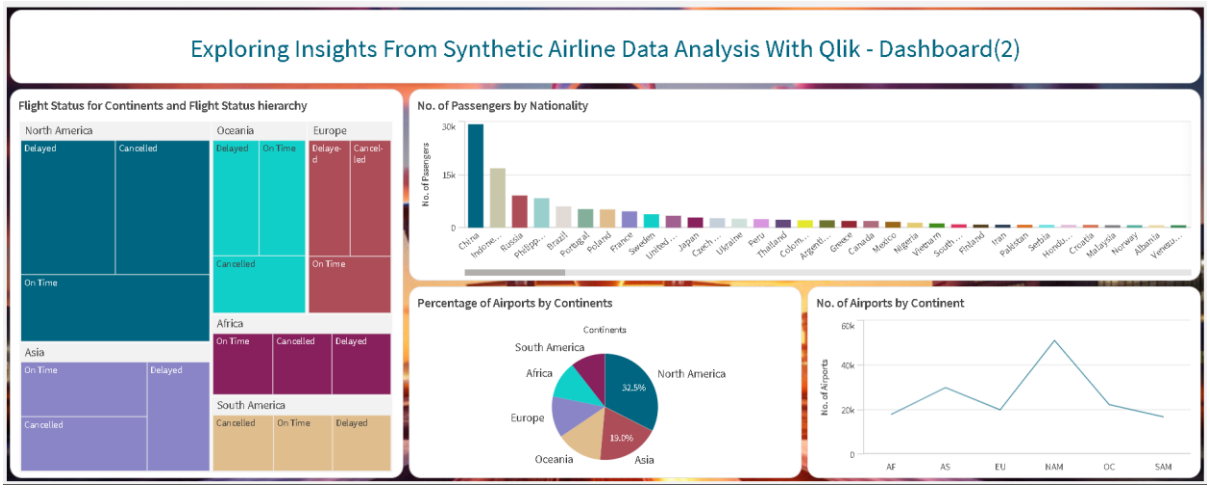
6 DASHBOARDS

6.1 Responsive and Design of Dashboard

Dashboard-1



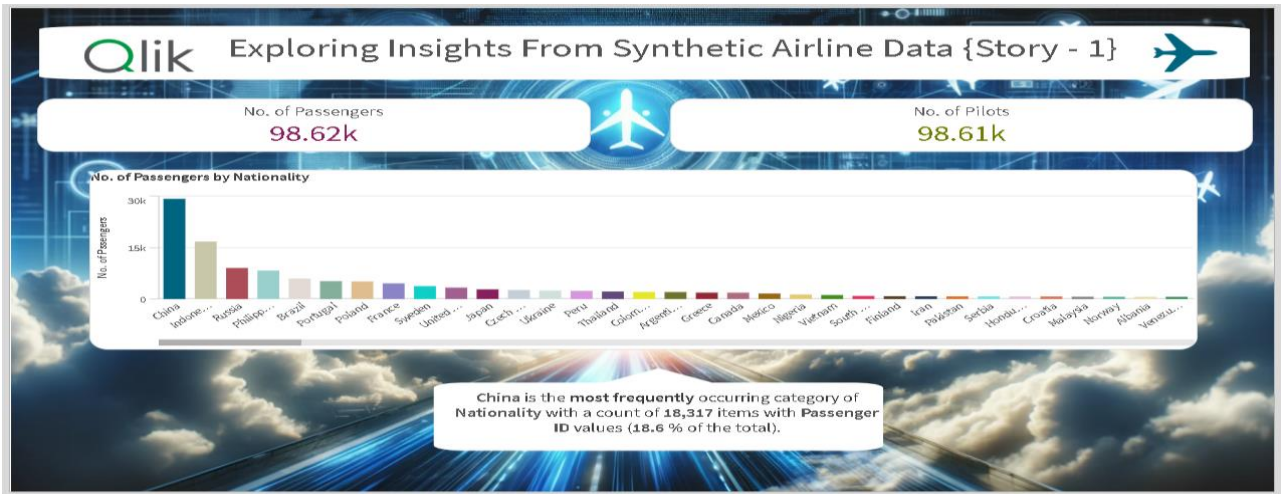
Dashboard-2



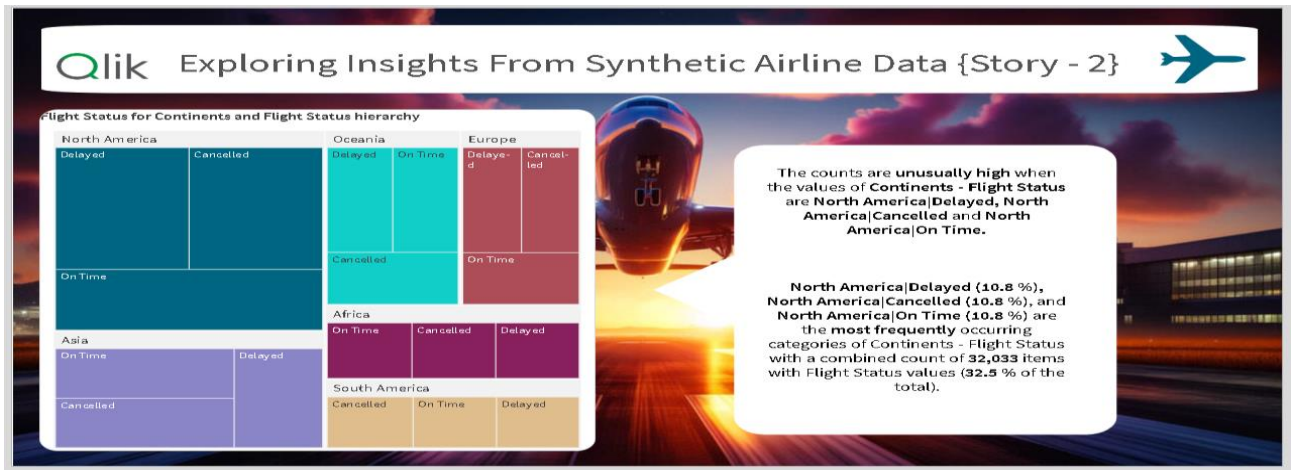
7 STORY

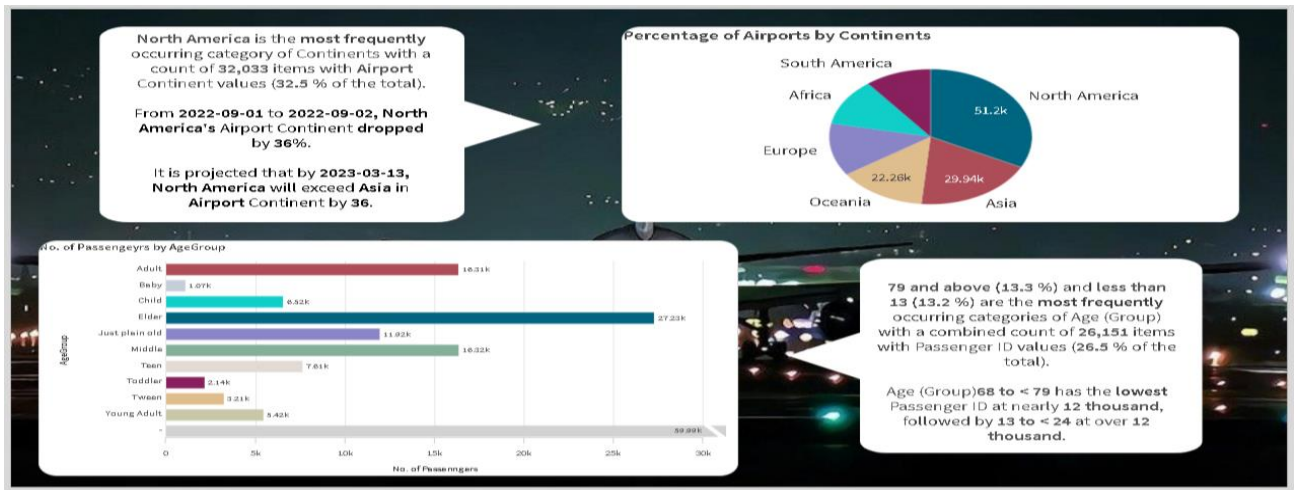
7.1 Design of Story

Story-1



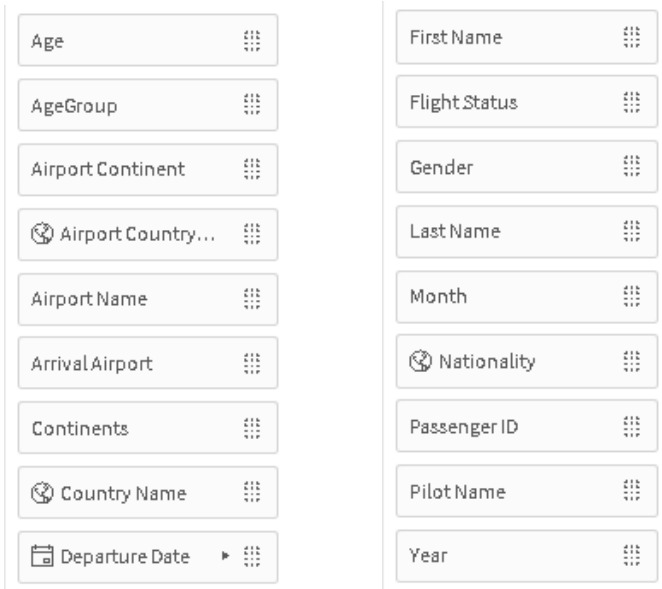
Story-2





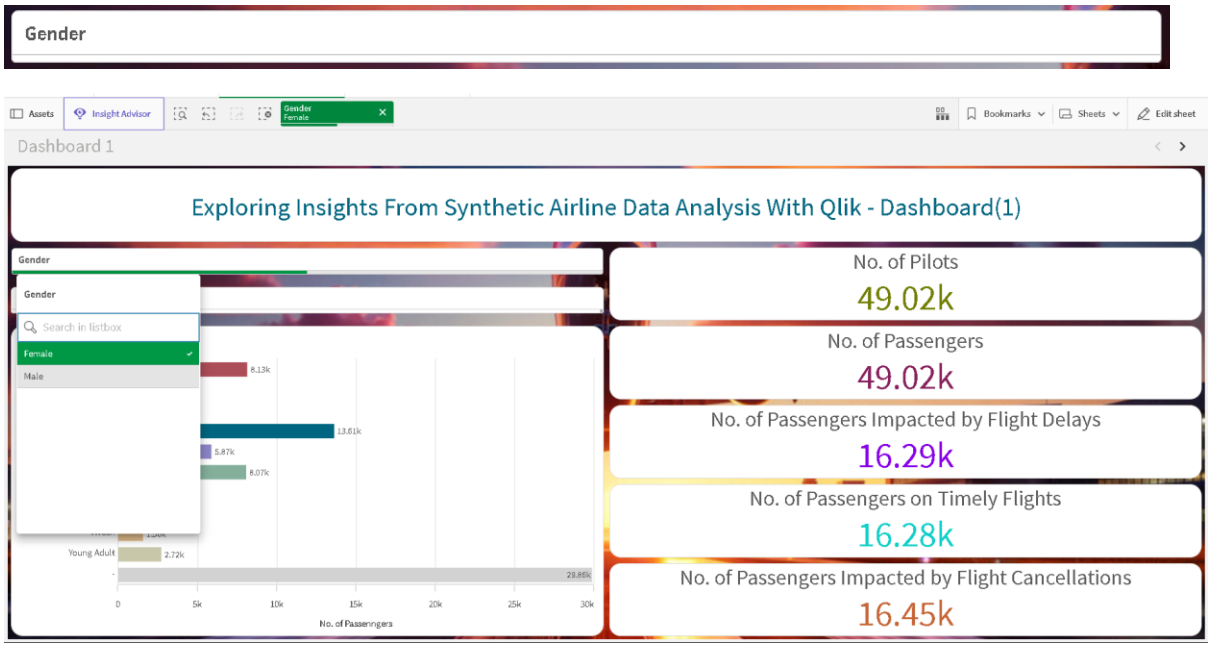
8 PERFORMANCE TESTING

8.1 Amount of Data Rendered



8.2 Utilisation of Data Filters

Filter-1



Filter-2