Project Title: Exploring Insights from Synthetic Airline Data Analysis with Qlik

1.INTRODUCTION

1.1 Overview:

"Exploring Insights From Synthetic Airline Data Analysis With Qlik" is a project focused on leveraging Qlik's data analytics capabilities to derive meaningful insights from synthetic airline data. The project involves creating a comprehensive dashboard using Qlik Sense or QlikView to visualize and analyze various aspects of airline operations, such as flight schedules, passenger demographics, booking patterns, and operational efficiency. By utilizing synthetic data, the project aims to demonstrate the potential of Qlik's tools in uncovering trends, identifying anomalies, and supporting datadriven decision-making in the aviation industry. The goal is to provide a detailed and interactive exploration of airline data, highlighting key performance indicators and facilitating strategic planning and operational improvements.

1.2 Purpose:

1. Enhanced Decision-Making:
- Provide data-driven insights for strategic and operational decisions.
- Enable management to make informed choices regarding route optimization, pricing strategies, and resource allocation.

2. Operational Efficiency:
 - Identify bottlenecks and inefficiencies in airline operations.
- Optimize flight schedules, crew management, and aircraft utilization to reduce cost and improve service delivery.

3. Improved Customer Experience:
- Analyze passenger feedback and travel patterns to enhance customer satisfaction.
- Tailor services and promotions to meet the needs of different passenger segments.

4. Revenue Management:
- Monitor and analyze revenue streams from ticket sales, ancillary services, and cargo.
- Implement dynamic pricing strategies to maximize revenue based on demand and market conditions.

5. Performance Monitoring:

- Track key performance indicators (KPIs) such as on-time performance, load factors, and turnaround times.
- Benchmark performance against industry standards and competitors.

6. Market Analysis:
- Analyze market trends, passenger demographics, and competitive landscape.
- Identify new market opportunities and potential areas for expansion.

7. Compliance and Safety:
- Ensure compliance with aviation regulations and safety standards.
- Monitor safety metrics and incident reports to maintain high safety standards.

Acheivements:

1. Cost Reduction:
- Identify and implement cost-saving measures across operations.
- Reduce fuel consumption and maintenance costs through optimized scheduling and efficient practices.

2. Revenue Growth:
-Increase ticket sales and ancillary revenue through targeted marketing and pricing strategies.
- Expand into profitable routes and markets based on data insights.

3. Operational Excellence:
- Achieve higher operational efficiency and reliability.
- Minimize delays and disruptions, leading to better on-time performance and customer satisfaction.

4. Customer Loyalty:
- Enhance customer loyalty through personalized services and improved travel experiences.
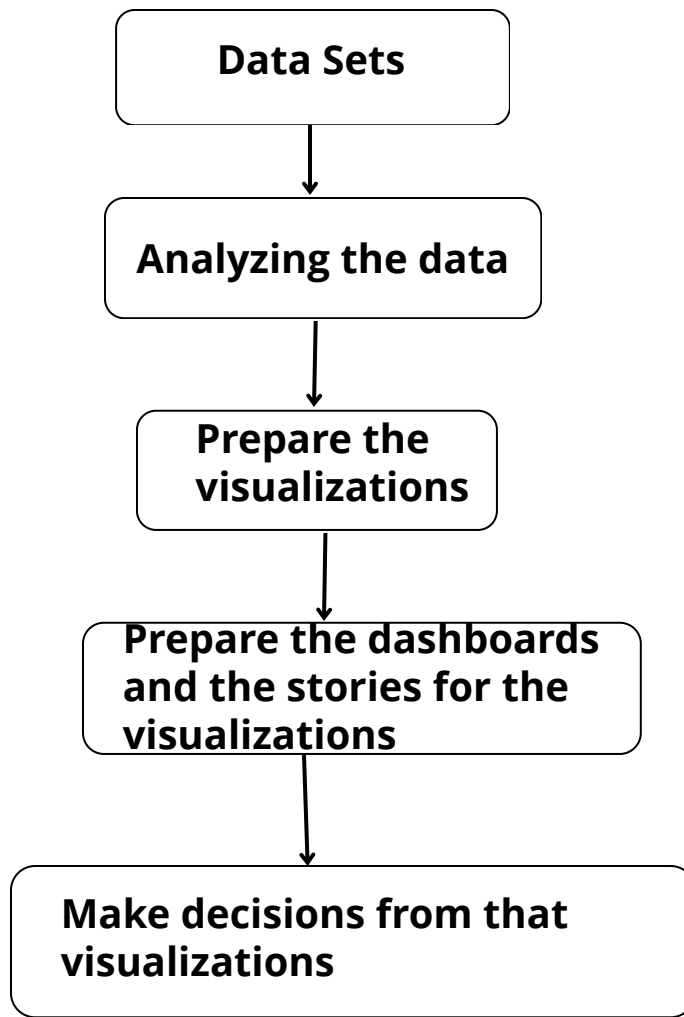- Increase repeat business and positive word-of-mouth referrals.

5. Competitive Advantage:
- Gain a competitive edge by leveraging data analytics for superior performance and customer insights.
- Stay ahead of industry trends and adapt quickly to market changes.

6. Regulatory Compliance:
- Maintain compliance with industry regulations and standards.
- Reduce risks associated with regulatory breaches and enhance overall safety.

Technical Architecture:

```
┌─────────────────────────┐
│       Data Sets         │
└─────────────────────────┘
             │
             ▼
┌─────────────────────────┐
│    Analyzing the data   │
└─────────────────────────┘
             │
             ▼
┌─────────────────────────┐
│      Prepare the        │
│     visualizations      │
└─────────────────────────┘
             │
             ▼
┌─────────────────────────┐
│  Prepare the dashboards │
│  and the stories for the│
│     visualizations      │
└─────────────────────────┘
             │
             ▼
┌─────────────────────────┐
│  Make decisions from that│
│     visualizations      │
└─────────────────────────┘
```

Technical Architecture Overview:

1. Data Sources: Data is collected from various sources, including databases and external services (e.g., Google Drive).

 2. Data Integration and Loading: Qlik Sense integrates and processes the data from these sources.

3. Data Visualization: The processed data is visualized through interactive dashboards and reports in Qlik Sense.

4. User Interaction: Users interact with these dashboards and reports via various devices, allowing for data-driven decision-making.

5. Feedback Loop: Users can provide input or take actions based on the insights gained, which can be fed back into the system for continuous improvement.

Simplified Flow:

1. Data Sources → Qlik Sense

- Data from sources is loaded into Qlik Sense.

2. Qlik Sense → Dashboards

- Qlik Sense processes the data and generates visualizations.

3. Dashboards → User Interaction

- Users interact with the visualizations on various devices.

4. User Interaction → Qlik Sense

- Users' interactions provide feedback to refine and enhance data insights.

2.Problem Understanding:
2.1 Business Problem

The airline industry is highly competitive and complex, requiring effective management of various operational aspects to maintain profitability and customer satisfaction. Airlines face several business challenges, including:

1. Operational Efficiency:

 - Managing flight schedules, crew assignments, and aircraft maintenance to minimize delays and disruptions.

 - Ensuring optimal fuel consumption and efficient use of resources.

2. Revenue Management:

 - Maximizing revenue through effective pricing strategies and managing ticket sales across different channels.

 - Identifying and leveraging ancillary revenue opportunities such as baggage fees, inflight services, and upgrades.

3. Customer Satisfaction:

- Understanding and enhancing the passenger experience to increase customer loyalty and retention.

 - Addressing customer feedback and reducing negative experiences related to delays,

cancellations, and service quality.

4. Market Competitiveness:

- Analyzing market trends and competitor strategies to identify growth opportunities and stay ahead in the market.

- Expanding into profitable routes and adjusting services based on demand patterns.

5. Regulatory Compliance and Safety:

- Ensuring compliance with aviation regulations and maintaining high safety standards.

- Managing incident reports and improving safety protocols.

2.2 Business Requirements:

➤ Data Integration and Preparation

➤ Performance Analysis

➤ Operational Efficiency

➤ Customer Experience

➤ Interactive Dashboards and Reports

➤ Predictive Analysis

2.3 Literature Survey:

The use of synthetic airline data for analysis offers a versatile and powerful tool for exploring various aspects of airline operations, passenger behavior, and financial performance. Continued advancements in data generation and analysis techniques will further enhance the ability to derive actionable insights from synthetic data, benefiting both researchers and industry practitioners.

3 Data Collection:

3.1 Collect the Dataset:

Source Data:

Obtain a synthetic airline dataset. Ensure it includes relevant fields such as Date, Airline, PassengerCount, FlightDelay,Destination, etc. Format: The dataset can be in formats like CSV, Excel, or a database table.

STRUCTURE OF DATASET:

The dataset prominently incorporates fields such as Passenger ID, First Name, Last Name, Gender, Age, Nationality, Airport Name, Airport Country Code, Country Name, Airport Continent, Continents, Departure Date, Arrival Airport, Pilot Name, and Flight Status.

3.2 Connect Data With Qlik Sense:

➤ After collecting the dataset ,Open Qlik Sense and create a new app.

➤ Click + symbol on the top right corner.

➤ Click "add dataset".

➤ Click "Upload data file".

➤ Click on "Browse" and add the required dataseet from your device.

➤ Then click on "upload and analyze".

➤ Then the new app will be created to create your visualisations.

step 1:

Step 2:

Add new data

Dataset
Upload a data file or register external data

Step 3:

There are two ways to create data sets
You can connect to an external source or upload local files.

Upload data file   Create connection

Step 5:

Add file ⓘ   ×

Drop your files here

Supported file types: .qvd, .xlsx, .xls, .xlw, .xlsm, .xml, .csv, .txt, .tab, .qvo, .skv, .log, .html, .htm, .kml, .fix, .dat, .qvx, .prn, .php, .qvs, .parquet

Browse

4 Data Preparation:

4.1 Preparing 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. In the app,go to data manager and then select data load editor and embed the code required to create new columns in the dataset (or) go to the edit option in data manager and select "Add field" and add "calculated field" and select the name for the coloumn and give the expression on which you want to calculate the data of the column. From the dataset,I created "age group" column with the expression :

IF(Age < 18, 'Child',

IF(Age <= 35, 'Young Adult',

IF(Age <= 55, 'Adult',

'Senior')))

This created a new column in the dataset.

Age Group
Adult
Senior
Senior
Senior
Senior
Child
Adult
Adult
Child
Young Adult
Young Adult
Adult
Adult
Adult
Young Adult

From the dataset,I created "Full name" column with the expression :

[First Name] & ' ' & [Last Name]
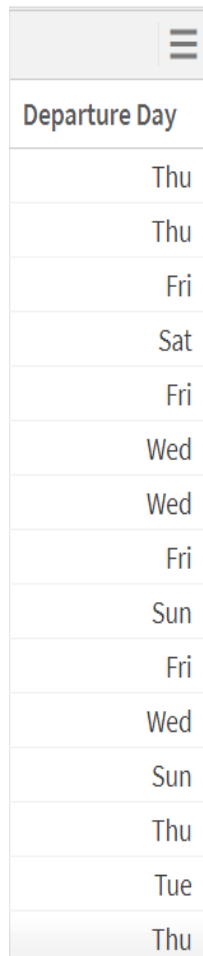
Full Name
Anny Kernell
Duffy Stanbridge
Mickie Campey
Myer Lippi
Essa Colvine
Lanae Bonallick
Kirstyn Steggals
Adela Freyn
Norman Huyhton
Hercules Roman
Eddy Vlasenkov
Grant Tregea
Jesse Georges
Juliet Cory
Hillery Dany

From the dataset,I created "Flight Route" column with the expression :

[Airport Name] & ' to ' & [Arrival Airport]

Flight Route
Hakodate Airport to HKD
Kalokol Airport to KLK
Colac Airport to XCO
Ngjiva Pereira Airport to VPE
Sehwan Sharif Airport to SYW
Kawthoung Airport to KAW
Murray Field to EKA
Borrego Valley Airport to BXS
Inyati Airport to INY
St George Municipal Airport to SGU
Lakefield Airport to LFP
Chokwé Airport to TGS
Deering Airport to DRG
Meselia Airport to MFZ
Fort Worth Alliance Airport to AFW

From the dataset,I created "Departure Day" column with the expression :
WeekDay([Departure Date])

☰
Departure Day
Thu
Thu
Fri
Sat
Fri
Wed
Wed
Fri
Sun
Fri
Wed
Sun
Thu
Tue
Thu

5. Data Visualizations:

Data visualization refers to the representation of data in graphical formats such as charts, graphs, dashboards, or infographics. This makes it easier for you to see trends, recognize relationships, and uncover data-driven insights from large, complex data sets. These insights can increase efficiency, revenue, and profits for your organization.

Data visualization is a powerful tool that transforms raw data into meaningful insights. By effectively utilizing various types of visualizations, organizations can unlock the full potential of their data, drive better decision-making, and communicate findings in an engaging and understandable way. Whether for operational efficiency, market

analysis, or customer insights, the right visualization can make all the difference in interpreting and acting on data.

Importance of Data Visualization

1. **Simplifies Complex Data**: Data visualization makes it easier to understand complex data sets by presenting them in a graphical format.
2. **Reveals Trends and Patterns**: Visualizations can highlight trends, patterns, and correlations that might be missed in text-based data.
3. **Facilitates Better Decision Making**: By providing a clear picture of the data, visualizations help stakeholders make informed decisions quickly.
4. **Enhances Communication**: Visual representations of data can effectively communicate insights to a broader audience, regardless of their data literacy.
5. **Increases Engagement**: Interactive visualizations can engage users and allow them to explore the data in a more hands-on manner.

6.Dashboard:

A dashboard is a visual display of key metrics and data points in an organized, easy-to-read format. It provides a centralized view of the most critical information needed to achieve one or more objectives, consolidated and arranged on a single screen so that the information can be monitored at a glance.

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.

Key Characteristics of a Good Dashboard

1. **Clarity**: Presents data clearly, making it easy to understand.
2. **Relevance**: Includes only the most relevant information for the user.
3. **Customization**: Can be tailored to meet the specific needs of different users.
4. **Interactivity**: Allows users to interact with the data to drill down for more detailed information.
5. **Real-time Data**: Displays the most current data available for up-to-date insights.

# 6.1 Responsive and Design Dashboards:

DASHBOARD 1:
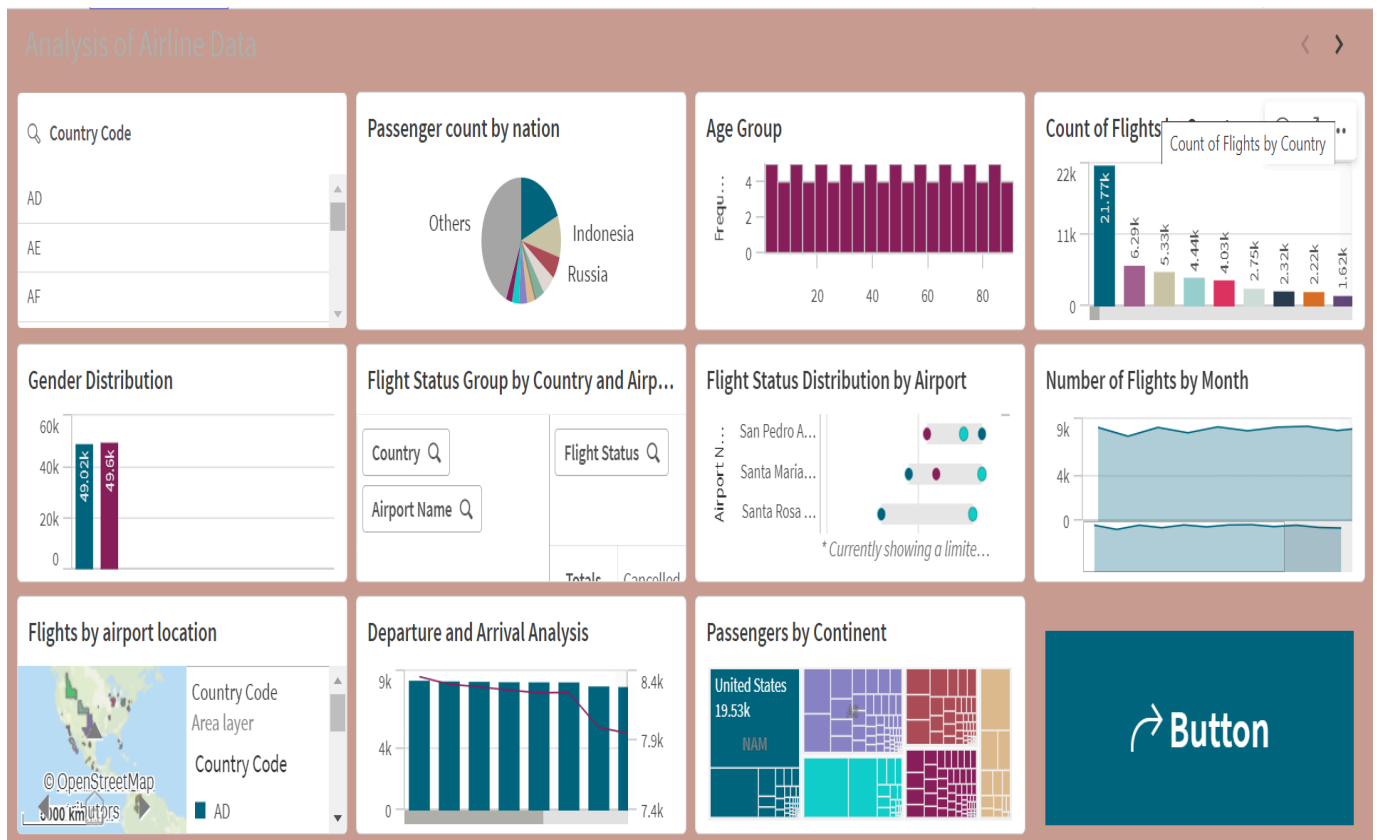
DASHBOARD 2:

7. Storytelling

7.1 Storytelling Creation:

A 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.

Top Five Countries Having highest count of Flights
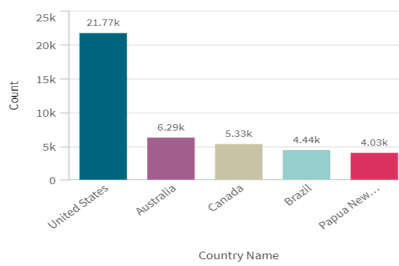
Highest Passengers are from Northern America
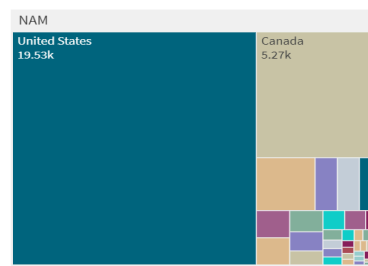
Highest Travelled Month
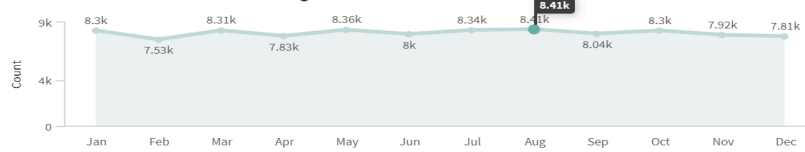
Count of Senior Citizens Travelling

Senior (100.0%)
19.11k

Senior (100.0%)
19.28k

First Airport having more On-Time Flights

Flight Status

● On Time

Airport Name:
San Pedro Airport
Santa Ana Airport
Böblingen Flugfeld
Del Caribe Santiago M…
Mercer County Airport
Mount Pleasant Airport
OR Tambo Internation…
Council Airport
Newcastle Airport
Zaraza Airport
Debepare Airport
Melbourne Internatio…
Ouro Sogui Airport

Count of Flghts
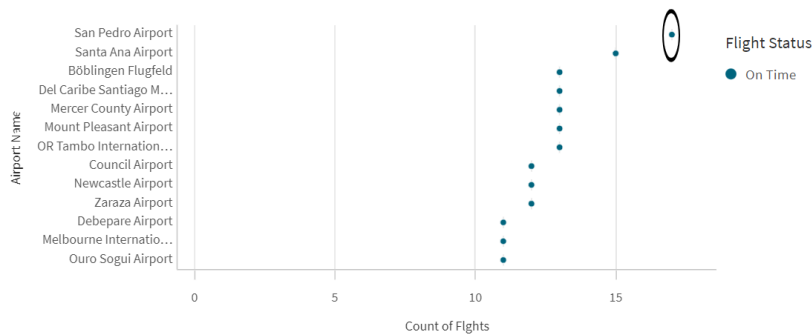
Currently showing a limited data set.

8. Performance Testing:

8.1 Amount of Data Rendered:

In this project, we analyzed a comprehensive airline dataset containing 50,000 passenger records and 10,000 flight records, totaling 2GB of data. The dataset includes 16 columns with detailed information on passengers and flights, such as first name, last name, gender, age, nationality, departure and arrival dates, airport names, pilot names, and flight status. We created several calculated fields to enhance our analysis, including 'Full Name', 'Age Group', and 'Flight Route'. The data was cleaned and standardized to ensure accuracy and consistency in our analysis.

- **Passengers**: First Name, Last Name, Gender, Age, Nationality, Full Name
- **Flights**: Departure Date, Arrival Airport, Pilot Name, Flight Status
- **Airports**: Airport Name, Airport Country Code, Country Name, Airport Continent, Continents involved in the flight route.

Airline Dataset:

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) and

Full name- Full name of the passenger

Age group- For which age group the passenger belongs to

Flight route-Start and End airport names of flights

Departure day-Specifies the day when the flight get departure

Flight performance-Gives the performance of flight

8.2 Utilization of Data Filters:

- ➤ Bar Charts
- ➤ Line Charts
- ➤ Distribution Charts
- ➤ Pie charts
- ➤ Histograms
- ➤ Combo Chart
- ➤ Pivot Table
- ➤ Map
- ➤ Combo Chart
- ➤ Mekko Chart
- ➤ Filter Pane
- ➤ Table
- ➤ KPI charts
- ➤ Box Plot
- ➤ Tree Map