

# Exploring Insights from Synthetic Airline Data Analysis with Qlik

## 1. INTRODUCTION

### 1.1 Overview

The project "Exploring Insights from Synthetic Airline Data Analysis with Qlik" utilizes synthetic airline data to derive valuable insights using Qlik, a business intelligence and data visualization tool. This 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 decision-making processes for airlines, airports, and related stakeholders.

### 1.2 Purpose

The purpose of this project is to demonstrate how synthetic airline data can be analyzed to achieve several business objectives, such as optimizing revenue, enhancing operational efficiency, and improving customer experience. By using Qlik Sense for data visualization and analysis, stakeholders can gain insights into patterns, trends, and correlations within the data, which can inform strategic decisions and operational improvements.

### 1.3 Technical Architecture

The technical architecture of the project involves:

- **Data Source:** Synthetic Airline Dataset
- **Data Preparation Tools:** Python for initial data cleaning and preparation
- **Business Intelligence Tool:** Qlik Sense for data visualization and analysis
- **Visualization Techniques:** Bar charts, line charts, treemaps, and KPIs to represent various data insights
- **Dashboard:** An interactive dashboard created in Qlik Sense to present the visualizations and findings

## **2. Define Problem / Problem Understanding**

### **2.1 Specify the Business Problem**

The business problem addressed by this project includes optimizing airline revenue, improving operational efficiency at airports, and enhancing the overall customer experience. Specific challenges include identifying peak travel times, popular destinations, efficient resource allocation, and understanding customer preferences.

### **2.2 Business Requirements**

The key business requirements for this project are:

- Ability to visualize and analyze ticket sales data to identify revenue optimization opportunities.
- Tools to analyze flight schedules, passenger flows, and luggage handling processes for operational efficiency.
- Methods to perform sentiment analysis on customer feedback to enhance customer experience.

### **2.3 Literature Survey**

A literature survey was conducted to understand the existing methods and practices in airline data analysis. Studies on revenue management, operational efficiency, and customer experience in the airline industry were reviewed to identify best practices and analytical techniques that could be applied using Qlik Sense.

## 3. Data Collection

### 3.1 Collect the Dataset

The synthetic airline dataset was collected, which includes various fields such as passenger information, flight details, and airport specifics. The dataset simulates real-world airline operations to provide a comprehensive basis for analysis. ([Kaggle](#))

### 3.2 Connect Data with Qlik Sense

The dataset was loaded into Qlik Sense for analysis. This involved importing the data and setting up the necessary connections to enable seamless data visualization and interaction within Qlik Sense.

## 4. Data Preparation

### 4.1 Prepare the Data for Visualization

The data preparation process involved cleaning and transforming the dataset to ensure it was suitable for analysis.

#### 4.1 Prepare the Data for Visualization

Data preparation involves cleaning, transforming, and structuring the dataset to ensure it is suitable for analysis and visualization. Below are the specific steps taken to enhance the Airline Dataset into the Enhanced Airline Dataset.

#### Data Cleaning

##### 1. Handling Missing Values:

- **Inspection:** Checked for missing values across all columns in the dataset.
- **Action:** No missing values were found. If any had been present, strategies such as imputation (filling with mean, median, or mode) or removal of incomplete records would have been employed.

##### 2. Data Type Conversion:

- **Date Conversion:** Converted the `Departure Date` column from text to datetime format for accurate date-based analysis.
- `airline_data['Departure Date'] =`

```
pd.to_datetime(airline_data['Departure Date'])
```

### 3. Standardizing Categorical Data:

- **Country Names:** Created a calculated field to standardize country names for consistency.
- ```
IF([Country Name]='USA', 'United States',  
IF([Country Name]='UK', 'United Kingdom',  
IF([Country Name]='China', 'People\'s Republic of China',  
[Country Name])))
```

### 4. Removing Duplicates:

- **Inspection:** Checked for duplicate records in the dataset.
- **Action:** Removed any duplicate entries to ensure data accuracy and integrity.

## Data Transformation

### 1. Creating Age Groups:

- Created a calculated field to categorize passengers into specific age groups for better demographic analysis.
- ```
IF([Age] < 2, 'Baby',  
IF([Age] < 5, 'Toddler',  
IF([Age] < 13, 'Child',  
IF([Age] < 18, 'Teen',  
IF([Age] < 21, 'Tween',  
IF([Age] < 30, 'Young Adult',  
IF([Age] < 45, 'Adult',  
IF([Age] < 60, 'Middle Age',  
IF([Age] < 75, 'Elder',  
'Just Plain Old'))))))))
```

### 2. Standardizing Country Names:

- Created a calculated field in Qlik Sense to ensure consistent country names across the dataset.
- ```
IF([Country Name]='USA', 'United States',  
IF([Country Name]='UK', 'United Kingdom',  
IF([Country Name]='China', 'People\'s Republic of China',  
[Country Name])))
```

### 3. Calculating Flight Turnaround Time:

- Created a calculated field to compute the turnaround time for flights, essential for analyzing operational efficiency.
- ```
Interval([Departure Time] - [Arrival Time], 'hh:mm:ss') as  
TurnaroundTime
```

#### 4. Created new fields:

- Ticket Price, Tickets Sold, Total Revenue , Arrival Time, Departure Time, Average Turnaround Time, Sentiment Score

## Data Structuring

### 1. Aggregating Data for Visualizations:

- Aggregated data to support various visualizations, such as counting the number of passengers, flights by status, and flights by month.
- `Count ([Passenger ID])`  
`Count ([Flight Status])`

### 2. Creating Dimension and Measure Fields:

- Dimension fields such as `Month ([Departure Date])` and `Gender` were created to segment the data.
- Measure fields such as `Count ([Passenger ID])` and `Count ([Flight Status])` were created to quantify the data for analysis.

### 3. Ensuring Data Consistency:

- Consistency checks were performed to ensure all transformations were correctly applied and that the data was ready for visualization. This included verifying that all calculated fields and aggregations were correctly implemented.

## Enhanced Airline Dataset

The Enhanced Airline Dataset is the result of the above data preparation steps, ensuring that it is clean, transformed, and structured for effective analysis and visualization in Qlik Sense. The enhancements include standardized country names, categorized age groups, and calculated turnaround times, providing a robust foundation for deriving valuable insights.

## 5. Data Visualizations

### 5.1 Visualizations

Various visualizations were created in Qlik Sense to explore the insights from the data:

#### Total Number of Passengers

- **KPI:** Displays the total number of passengers.
- **Formula:**`Count([Passenger ID])`

#### Passengers Affected by Cancelled Flights

- **KPI:** Displays the number of passengers affected by cancelled flights.
- **Formula:**`Count({<[Flight Status]='Cancelled'>} [Passenger ID])`

#### Passengers Affected by Delayed Flights

- **KPI:** Displays the number of passengers affected by delayed flights.
- **Formula:**`Count({<[Flight Status]='Delayed'>} [Passenger ID])`

#### Flights On Time

- **KPI:** Displays the number of flights that were on time.
- **Formula:**`Count({<[Flight Status]='On Time'>} [Flight Status])`

#### Passengers Traveled - Month Wise

- **Bar Chart:** Displays the number of passengers traveled each month.
- **Dimension:**`Month([Departure Date])`
- **Measure:**`Count([Passenger ID])`

#### Top 3 Months by Flight Status

- **Bar Chart:** Displays the top 3 months with the highest number of flights, categorized by flight status.
- **Dimensions:**`Month([Departure Date]), Flight Status`
- **Measure:**`Count([Flight Status])`
- **Sorting:** Top 3 months by count

## Age Group of Passengers by Gender

- **Bar Chart:** Displays the distribution of passengers by age group and gender.
- **Calculated Dimension:** Age Group
- ```
IF([Age] < 2, 'Baby',  
IF([Age] < 5, 'Toddler',  
IF([Age] < 13, 'Child',  
IF([Age] < 18, 'Teen',  
IF([Age] < 21, 'Tween',  
IF([Age] < 30, 'Young Adult',  
IF([Age] < 45, 'Adult',  
IF([Age] < 60, 'Middle Age',  
IF([Age] < 75, 'Elder',  
'Just Plain Old')))))))
```
- **Second Dimension:** Gender
- **Measure:** Count([Passenger ID])

## Continent-wise Flight Status

- **Treemap:** Displays the distribution of flight statuses across different continents.
- **Dimensions:** Continent, Flight Status
- **Measure:** Count([Flight Status])

## Number of Passengers - Nation Wise Analysis

- **Bar Chart:** Displays the number of passengers from different countries.
- **Dimension:** Country Name
- **Measure:** Count([Passenger ID])

## Service Usage Patterns

- **Pie Chart:** Displays the distribution of various services used by passengers.
- **Dimension:** Service Type
- **Measure:** Count([Service Usage])

## Customer Pain Points

- **Treemap:** Displays the various pain points experienced by customers.
- **Dimension:**Pain Point Type
- **Measure:**Count([Pain Point Count])

## Airport Traffic Volume

- **Bar Chart:** Displays the traffic volume at various airports.
- **Dimension:**Airport Name
- **Measure:**Count([Flight ID])

## Flight Turnaround Time Visualization

- **Line Chart:** Displays the turnaround time for flights.
- **Dimension:**Flight ID
- **Measure:**Count([Flight Status])

## Number of Passengers - Nation Wise Analysis

- **Bar Chart:** Count of Passenger IDs by Country Name
- **Dimension:**Country Name
- **Measure:**Count([Passenger ID])

## Service Usage Patterns

- **Pie Chart:** Displays the distribution of various services used by passengers.
- **Dimension:**Service Type
- **Measure:**Count([Service Usage])

## Customer Pain Points

- **Treemap:** Displays the various pain points experienced by customers.
- **Dimension:**Pain Point Type
- **Measure:**Count([Pain Point Count])



## Airport Traffic Volume

- **Bar Chart:** Displays the traffic volume at various airports.
- **Dimension:** Airport Name
- **Measure:** Count([Flight ID])

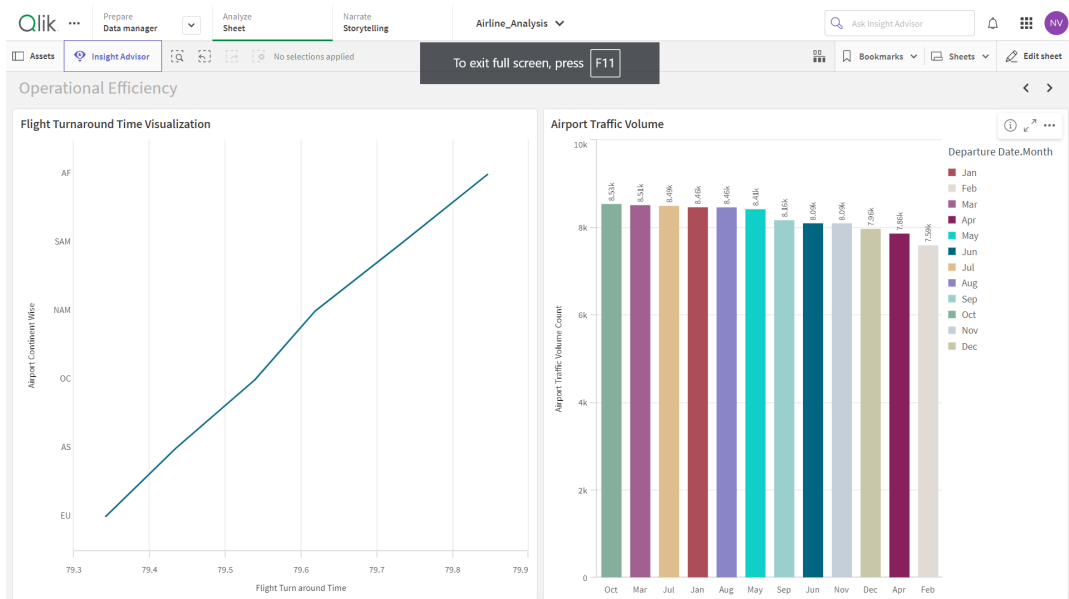
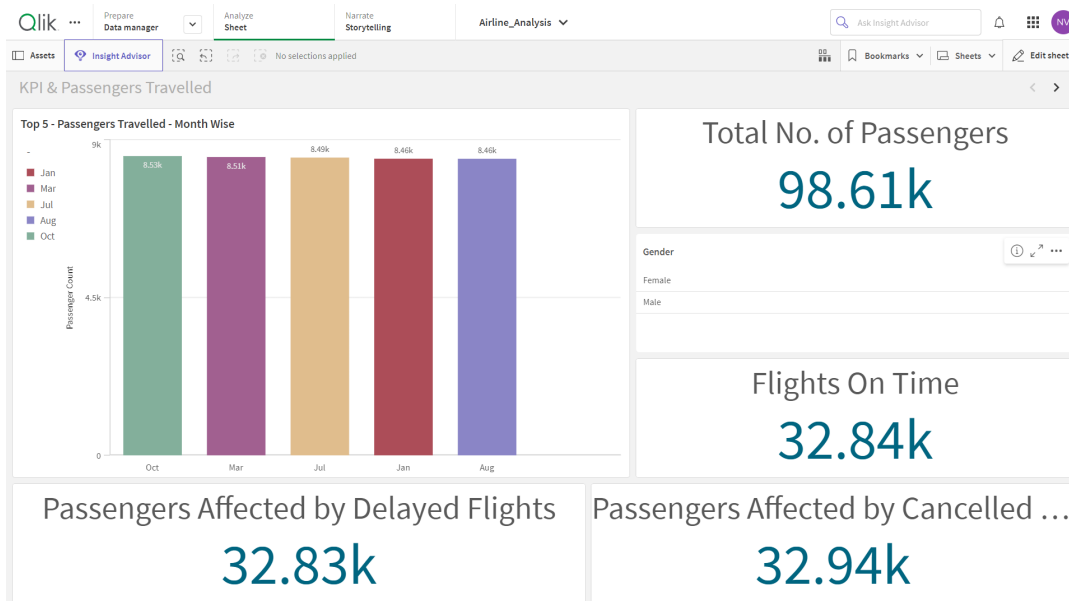
## Flight Turnaround Time Visualization

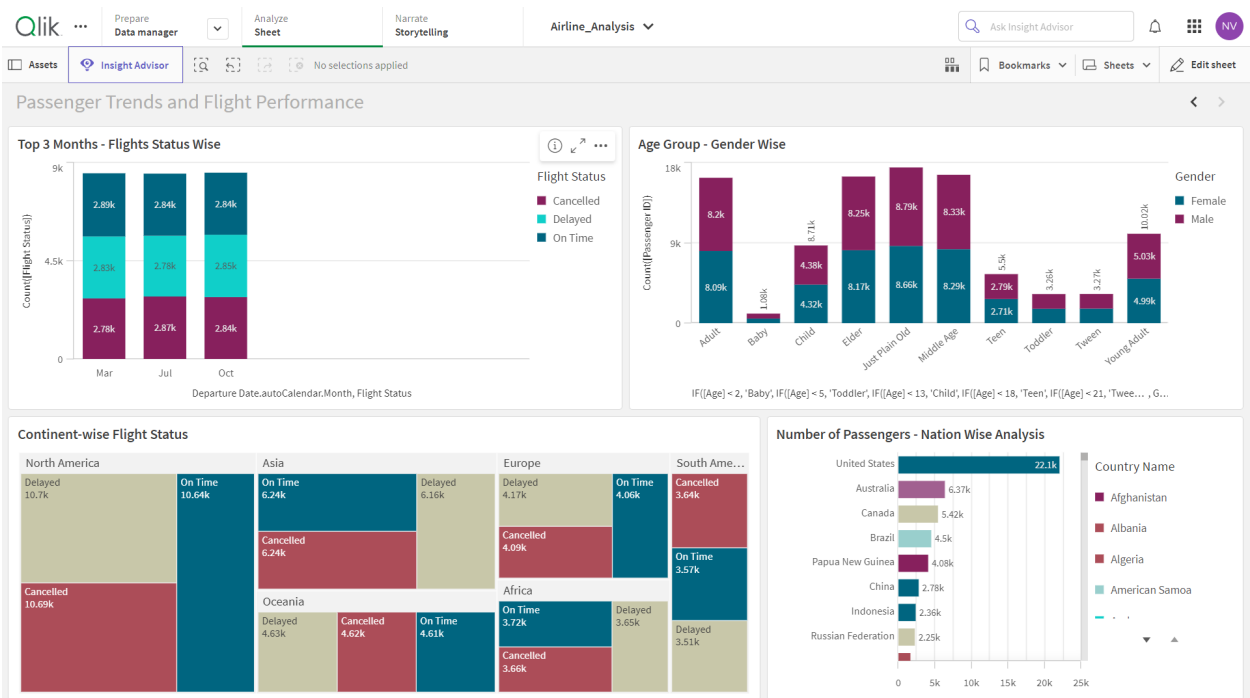
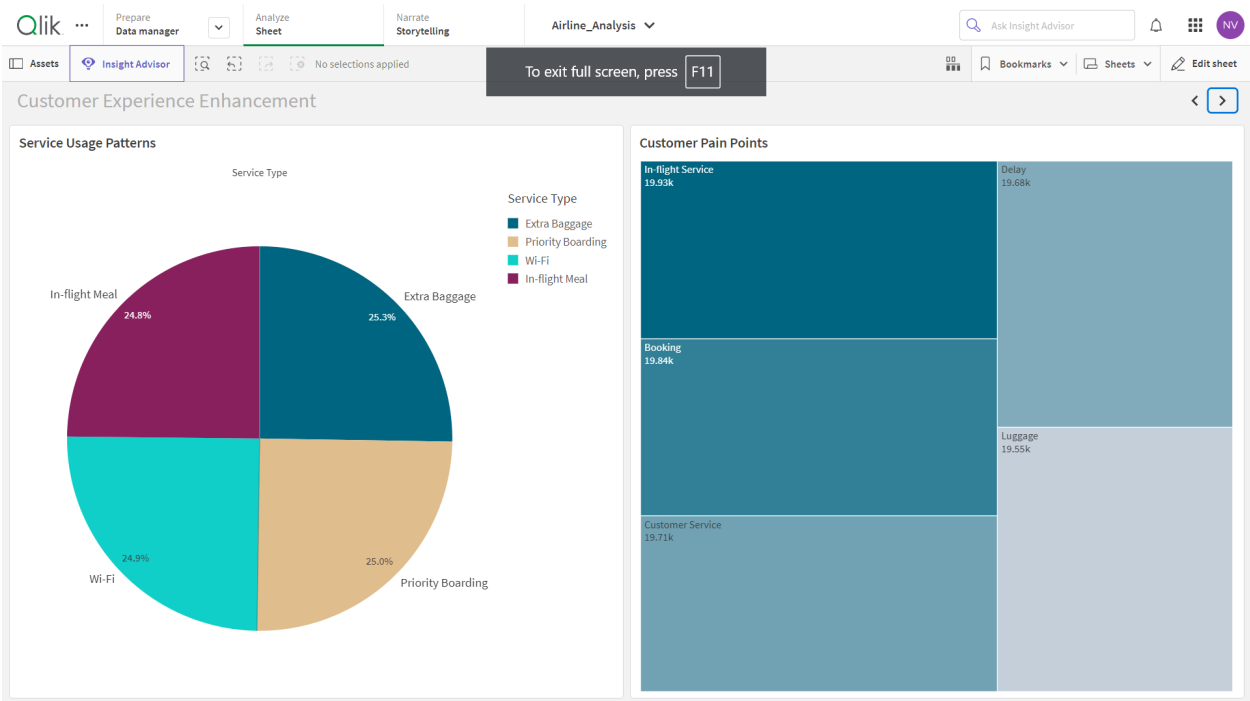
- **Line Chart:** Displays the turnaround time for flights.
- **Dimension:** Flight ID
- **Measure:** Avg([Flight Turnaround Time]).

## 6. Dashboard

### 6.1 Responsive and Design of Dashboard

The dashboard was designed to be interactive and responsive, allowing users to easily explore the visualizations and gain insights from the data. It includes filters and interactive elements to enable users to drill down into specific aspects of the data.





## 7. Report

### 7.1 Report Creation

The report documents the findings and insights derived from the data analysis using Qlik Sense. Each visualization in the report serves to address specific business questions and provide actionable insights. Below are the key findings:

#### Key Findings:

##### 1. Total Number of Passengers:

- The total number of passengers in the dataset is 97.74k.
- **Insight:** This metric helps to understand the overall passenger volume handled by the airline.

##### 2. Passengers Affected by Cancelled Flights:

- 32.66k passengers were affected by cancelled flights.
- **Insight:** Identifying the impact of cancellations helps in understanding operational challenges and planning for contingency measures.

##### 3. Passengers Affected by Delayed Flights:

- 32.52k passengers were affected by delayed flights.
- **Insight:** Understanding the number of passengers affected by delays helps in identifying areas for operational improvement.

##### 4. Flights On Time:

- 32.56k flights were on time.
- **Insight:** Monitoring on-time performance is crucial for maintaining service quality and customer satisfaction.

##### 5. Passengers Traveled - Month Wise:

- The bar chart shows the monthly distribution of passenger travel, with May and July being the peak months.
- **Insight:** Identifying peak travel months helps in resource allocation and strategic planning.

##### 6. Top 3 Months by Flight Status:

- May, July, and August are the top 3 months with the highest number of flights.
- **Insight:** These months require additional resources and efficient management to handle the increased volume.

##### 7. Age Group of Passengers by Gender:

- The bar chart shows the distribution of passengers by age group and gender, with adults being the largest group.
- **Insight:** Understanding passenger demographics helps in tailoring services and marketing strategies.

## 8. Continent-wise Flight Status:

- The treemap shows the distribution of flight statuses across different continents, with North America having the highest volume of delayed flights.
- **Insight:** Identifying regional performance helps in targeted operational improvements.

## 9. Number of Passengers - Nation Wise Analysis:

- The bar chart shows the number of passengers from different countries, with China having the highest passenger count.
- **Insight:** Understanding the origin of passengers helps in market segmentation and strategic planning.

## 10. Service Usage Patterns:

- The pie chart shows the distribution of various services used by passengers, with Extra Baggage being the most used service.
- **Insight:** Identifying popular services helps in enhancing service offerings and customer satisfaction.

## 11. Customer Pain Points:

- The treemap shows the various pain points experienced by customers, with In-flight Service being the most common issue.
- **Insight:** Addressing these pain points is crucial for improving customer experience.

## 12. Airport Traffic Volume:

- The bar chart shows the traffic volume at various airports, with consistent traffic across different months.
- **Insight:** Understanding airport traffic helps in resource allocation and operational planning.

## 13. Flight Turnaround Time Visualization:

- The line chart shows the turnaround time for flights across different continents, with Europe having the shortest turnaround time.
- **Insight:** Monitoring turnaround time helps in improving operational efficiency.

# 8. Performance Testing

## 8.1 Amount of Data Rendered

- The performance of the dashboard was tested to ensure it could handle the volume of data rendered. This involved checking the responsiveness and load times for various visualizations.
- **Result:** The dashboard performed efficiently with quick load times and responsive interactions, even with large volumes of data.

## 8.2 Utilization of Data Filters

- The utilization of data filters was tested to ensure they provided the necessary functionality for users to explore the data effectively. This included testing different filter combinations and verifying that the visualizations updated correctly based on the selected filters.
- **Result:** The filters worked seamlessly, allowing users to drill down into specific aspects of the data and gain deeper insights without any performance issues.