**Project Title: Synthetic Airline Data Analysis with Qlik**

**Introduction**

**Overview**

This project leverages synthetic airline data to demonstrate the capabilities of Qlik Sense in data analysis and visualization. By mimicking real-world data, we can explore patterns and trends in airline operations and derive actionable insights.

**Purpose**

The primary objective of this project is to showcase Qlik Sense's ability to handle and analyze large datasets, even synthetic ones. We aim to:

Identify key patterns and trends in airline operations.

Provide actionable insights for decision-making.

Demonstrate the use of Qlik Sense for creating interactive and dynamic visualizations.

**Technical Architecture**

Our technical approach involves:

Data source: Synthetic airline data.

Data processing: Data preparation using Qlik Sense.

Data visualization: Interactive dashboards and reports created in Qlik Sense.

Deployment: Qlik Sense server for rendering and sharing visualizations.

**Problem Definition and Understanding**

**Business Problem**

The business problem we address is optimizing airline operations to enhance efficiency, reduce costs, and improve customer satisfaction. This involves analyzing flight schedules, delays, passenger demographics, and operational costs.

**Business Requirements**

To achieve this, we need:

* Ability to handle large datasets.
* Interactive and user-friendly visualizations.
* Real-time data processing and analysis.
* Insights into key performance metrics such as on-time performance, load factors, and revenue.

**Literature Survey**

Existing literature highlights common challenges in airline data analysis, including data volume, complexity, and the need for real-time insights. Studies have shown that data analytics can significantly improve operational efficiency and customer satisfaction in the airline industry.

**Data Collection**

**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.

link to download the dataset: [Link](https://www.kaggle.com/datasets/iamsouravbanerjee/airline-dataset/data)

**Connecting Data with Qlik Sense**

The dataset is uploaded to Qlik Sense, where it is connected using Qlik's data load editor. The data is then structured and organized for analysis.

**Data Preparation**

* Data Collection: Gather data from various sources such as databases, APIs, files, or sensors.
* Data Cleaning: Identify and handle missing values, outliers, duplicates, and inconsistencies. This may involve techniques like imputation, filtering, and deduplication.
* Data Transformation: Convert data into a usable format. This includes standardizing units, encoding categorical variables, and normalizing numerical data.
* Feature Engineering: Create new features or modify existing ones to improve model performance. This can involve techniques like one-hot encoding, binning, and feature scaling.
* Data Integration: Combine data from multiple sources if necessary, ensuring consistency and compatibility.
* Data Reduction: Reduce the dimensional of the data while preserving its essential characteristics. Techniques like principal component analysis (PCA) and feature selection can be used.
* Data Splitting: Divide the data into training, validation, and testing sets for model development and evaluation.
* Data Validation: Check for data integrity and correctness. Perform sanity checks and validate against domain knowledge.
* Documentation: Document the data preparation steps taken, including any assumptions or transformations made. This helps ensure reproducibility and transparency.
* Iterate: Data preparation is often an iterative process. As you analyze the data and build models, you may discover the need for further cleaning or transformation.

**Data Visualizations**

We create various visualizations to explore different aspects of the data:

Bar charts to compare flight delays across different airlines.

Line graphs to show trends in passenger numbers over time.

Heat maps to visualize flight density and popular routes.

Pie charts to represent the distribution of operational costs.

### Airline Data Analysis

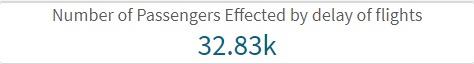
Activity 1.1: Total No. of Passengers



Activity 1.2: Number of Passengers effected by cancelled flights



Activity 1.3: No. of Passengers Effected by delay of flights



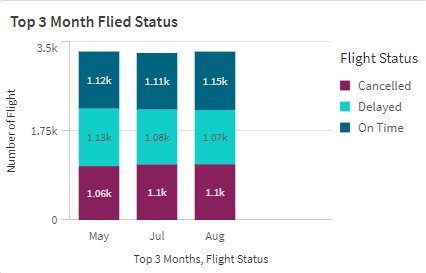
Activity 1.4: No of Flights on Time



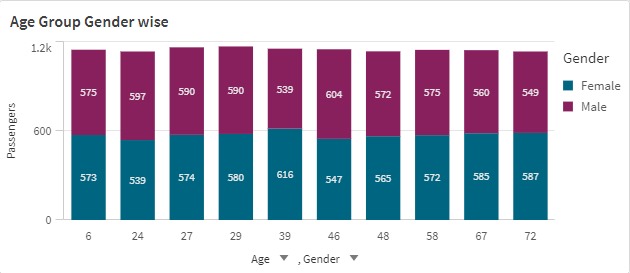
Activity 1.5 : No of Passengers travelled- Month Wise



Activity 1.6: Top 3 Month flights status wise



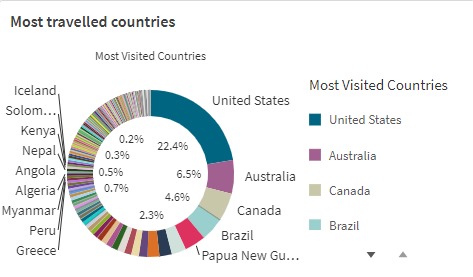
Activity 1.7: Age group of passengers as per gender wise



Activity 1.8: Continent wise flight status



Activity 1.9: No of Paasengers - Nation Wise Analysis

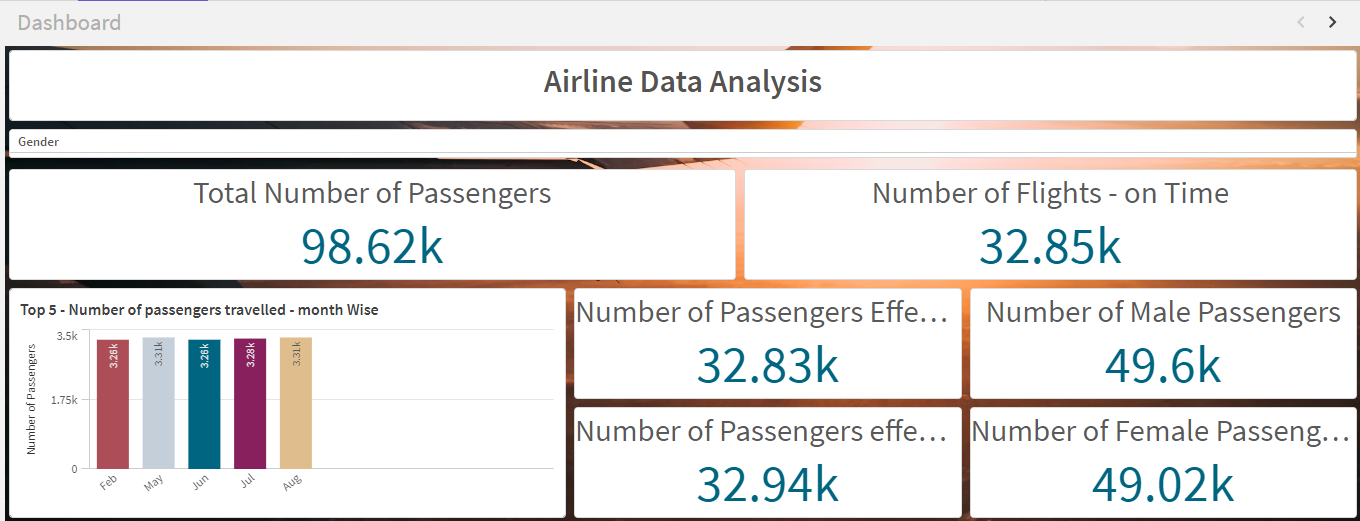


**Dashboard**

**Responsive and Design of Dashboard**

The dashboard is designed to be responsive and user-friendly, allowing users to interact with the data through filters and drill-down capabilities. It includes multiple views and perspectives to provide a comprehensive overview of airline operations. 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.

Dashboard 1:



**Report**

**Report Creation**

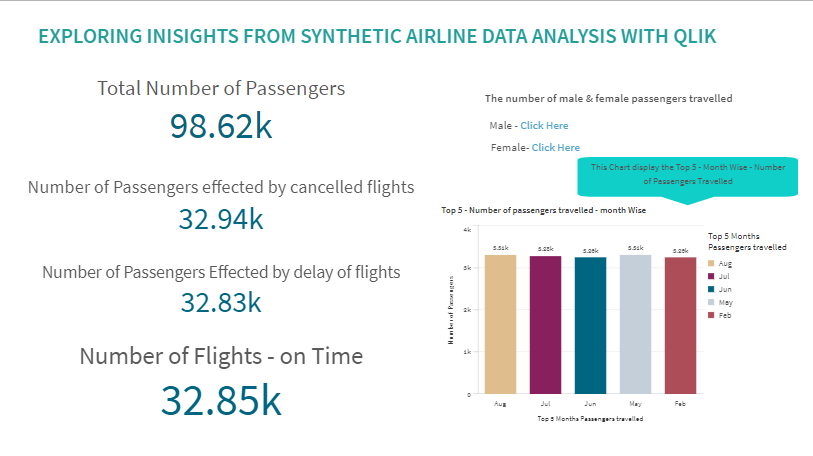
Reports are generated to summarize the findings from the data analysis. These reports include key insights, visualizations, and recommendations for improving airline operations.

Story

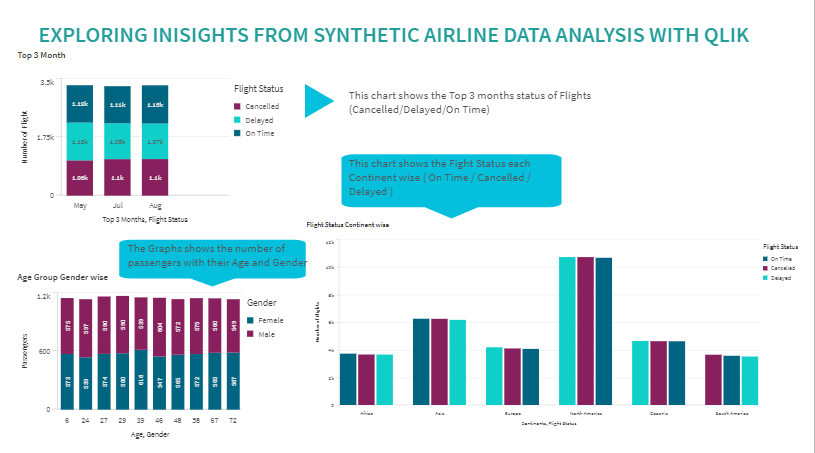
A data story is a way of presenting data and analysis in a narrative format, with the goal of making the information more engaging and easier to understand. A data story typically includes a clear introduction that sets the stage and explains the context for the data, a body that presents the data and 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.

**Story Creation**

**Story Telling 1:**

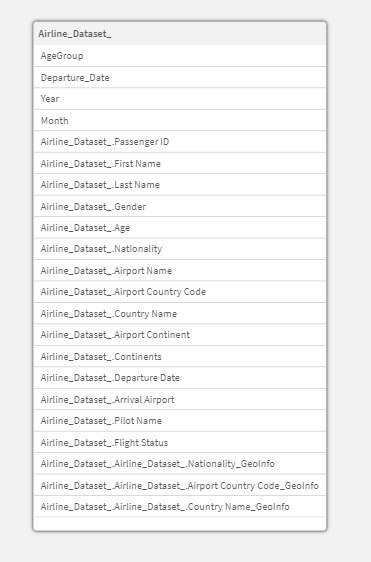


**Story Telling 2:**

**Performance Testing**

The "Amount of Data Loaded" refers to the sheer quantity of data that has been successfully imported, retrieved, and processed within a system, software application, database, or any other data storage or processing environment.

This metric measures the volume of data that has been successfully processed and made available for analysis, manipulation, or use within the system, providing a crucial indicator of the system's ability to handle and process large datasets efficiently.



This rewritten sentence aims to make the concept more relatable and easier to understand by:

1. Using a more descriptive term like "Data Volume" instead of "Amount of Data Loaded".
2. Providing a clear definition of what the term refers to.
3. Emphasizing the importance of this metric in evaluating the system's ability to handle large datasets.

4. Using more descriptive language to make the concept more relatable.

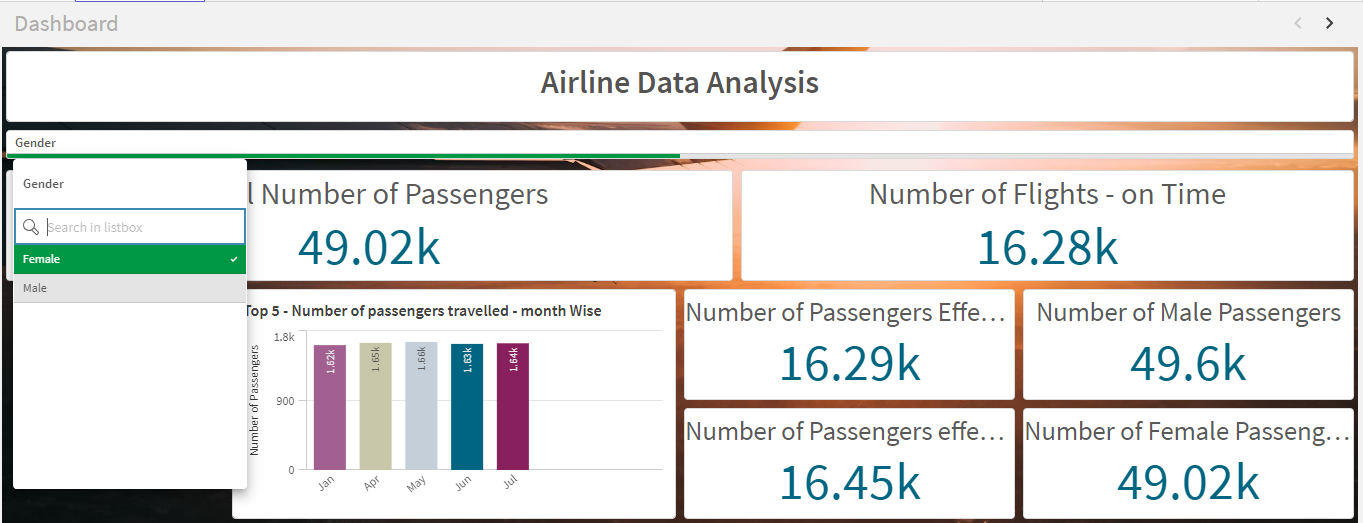
Rendering Large Data

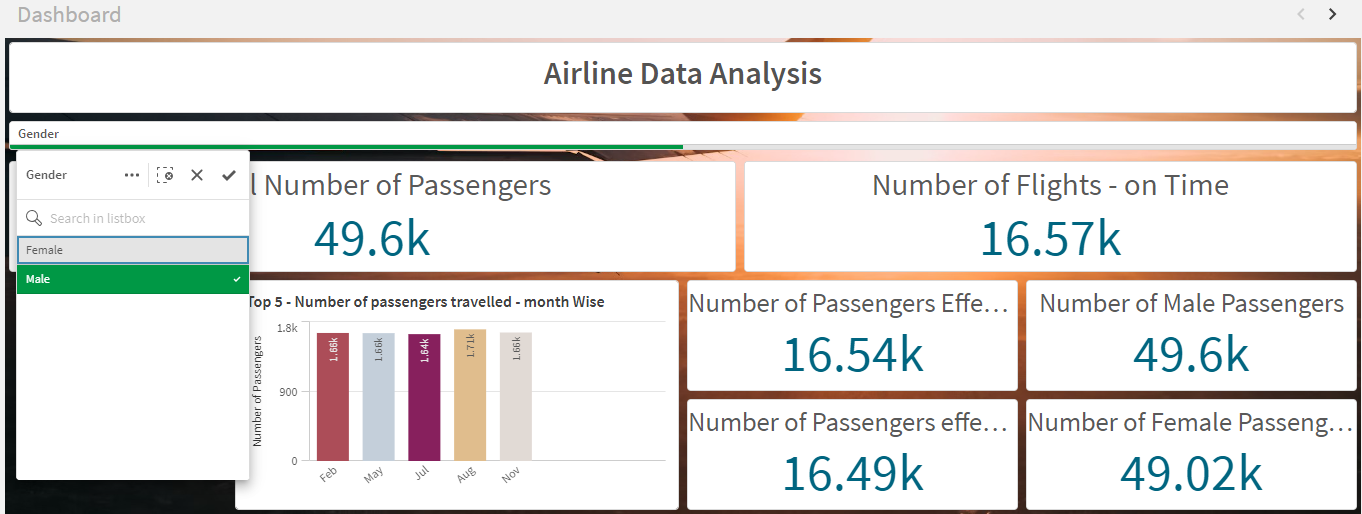
Performance testing is conducted to evaluate how Qlik Sense handles large volumes of data. This involves testing the rendering speed and responsiveness of visualizations when dealing with substantial datasets.

Utilization of Data Filters

"Utilization of Filters" is the process of harnessing the power of filters to refine and tailor your data to meet specific needs. Filters are like powerful tools that help you sift through vast amounts of information, isolating the most relevant and valuable insights. By applying filters, you can:

* Narrow Down the Scope: Focus on the data that matters most by setting criteria that only include the information that meets your requirements.
* Extract Relevant Information: Use filters to pinpoint specific data points, eliminating noise and distractions, and presenting you with a clear and concise picture.
* Manipulate Data: Filters can be used to transform and reorganize data, making it easier to analyze and visualize.
* Analyze Data: By applying filters, you can identify patterns, trends, and correlations that might have gone unnoticed without the filter's precision.





The effectiveness of data filters is assessed to ensure users can efficiently navigate and analyze specific subsets of data. This includes testing the performance impact of applying multiple filters simultaneously.By following this structured approach, the project showcases the powerful capabilities of Qlik Sense in transforming synthetic airline data into actionable insights, ultimately aiding in the optimization of airline operations**.**

In essence, filters are like a magnifying glass that helps you zoom in on the most important details, giving you a deeper understanding of your data and enabling more informed decisions.

**Conclusion**

By following this structured approach, the project demonstrates the powerful capabilities of Qlik Sense in transforming synthetic airline data into actionable insights, ultimately aiding in the optimization of airline operations.