AeroInsights: Optimizing Airlines with Qlik Analytics

In today's competitive aviation business, data analytics are critical for airlines and airports to make better decisions and increase efficiency. The project "Exploring Insights from Synthetic Airline Data Analysis with Qlik" uses synthetic airline data with Qlik, a strong business intelligence platform, to discover useful insights. This project demonstrates how Qlik's strong analytical capabilities may be applied to numerous areas of airline operations, including aircraft schedules, passenger demographics and performance metrics.

This project develops a thorough platform for investigating critical airline business scenarios by modeling real-world airline data. The synthetic data used in this study builds up key aspects of airline operations, allowing for an exhaustive review of patterns, trends, and correlations that can inform strategic decisions for airlines, airports, and other stakeholders.

Problem Understanding:

Airlines and airport authorities in the highly competitive aviation business are continuously looking for methods to optimise their operations, improve customer experiences, and maximise revenue. However, getting useful insights from massive amounts of operational data can be difficult. Traditional data analysis methods frequently fail to detect hidden patterns and trends that can help with strategic decision-making. This project seeks to address these issues by utilizing synthetic airline data and Qlik's advanced data visualization and analytical capabilities to discover significant insights.

- ### Specify the business problem

The key business concerns addressed in this project are revenue optimization, operational efficiency enhancement, and customer experience improvement in the airline industry. Airlines must determine peak travel hours, popular locations, and successful pricing techniques in order to optimize profits. Simultaneously, airports expose problems with operations that cause delays and increased expenses, demanding analysis of flight schedules, passenger flows, and luggage handling operations in order to predict peak traffic periods and optimize resource allocation. Furthermore, improving customer happiness and loyalty necessitates a thorough grasp of client preferences and feedback in order to personalize services and solve issue areas.

- ### Business requirements

To tackle these commercial issues, the project outlines several critical requirements. To ensure data accuracy and completeness, synthetic airline data such as flight schedules, passenger demographics, ticket sales, and performance indicators must first be integrated and preprocessed. Revenue analysis requires displaying revenue trends, determining peak travel hours, and designing pricing strategies based on customer behavior. Analyzing aircraft schedules, researching passenger patterns, and projecting peak traffic all contribute to increased operational efficiency. Finally, improving the customer experience necessitates sentiment analysis of customer input, segmenting customers based on preferences, and tailoring services to create loyalty and satisfaction.

- ### Literature Survey

The literature review for the present work demonstrates the importance of data analytics in enhancing operational efficiency and customer happiness in the aviation industry. Studies stress the use of predictive analytics and machine learning models to forecast demand, optimize pricing, and improve operational planning. Qlik's outstanding data visualization and business intelligence capabilities have been well documented, with case studies illustrating their successful use in a variety of industries, including aviation. Airlines have used Qlik to acquire insights into customer behavior, manage flight schedules, and increase operational efficiency, resulting in cost savings and improved customer experiences. The inquiry builds on established approaches and creative analytical tools to meet the unique issues that airlines and airport authorities face.

Data Collection :

- ### Collect the dataset

The process of obtaining and evaluating information on relevant variables in a predetermined, methodical manner is known as data collection. This makes it possible for researchers to assess results, test hypotheses, respond to research questions, and draw conclusions from the data. To gather information for this project, a synthetic airline dataset must be obtained and understood in order to study many facets of airline operations.

- The dataset can be downloaded using the following link:Dataset-Link

This dataset contains detailed meta-information regarding the columns described in the CSV file. Below is a description of each column included in 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, canceled).

- ### Connect Data with Qlik Sense

Connecting the data to Qlik Sense is the next step after gathering and analyzing the dataset. The data must be uploaded to Qlik Cloud Analytics after it has been downloaded. We can preprocess the data in Qlik Sense to make sure it is clear and prepared for analysis. This covers operations like eliminating duplicates, dealing with missing values, and altering data as needed. After that, the dataset will be visualized and analyzed using Qlik Sense, providing insightful information that tackles the previously mentioned business issues. In order to maximize income, boost operational effectiveness, and enhance customer experiences in the airline business, we may design interactive dashboards and carry out sophisticated analysis using Qlik Sense

Data Preparation:

One of the most important steps in making sure the data is easily comprehensible and ready for the creation of visuals to provide insights into performance and efficiency is preparing the data for visualization. Many crucial tasks are involved in this process:

- ✓ Cleaning the Data: In order to ensure consistency, this stage eliminates any unnecessary or missing data and fixes any mistakes.
- ✓ Transforming the Data: Modify the format to make it easier to visualize. This covers data aggregation, value normalization, and the creation of new computed fields.
- ✓ Studying the Data: Find patterns and trends to comprehend how variables relate to one another, which will inform the visualization styles selected.
- ✓ Filtering the Data: To obtain specific insights, concentrate on particular subsets, such as time periods, locations, or client profiles.
- ✓ Getting Ready for Visualization Software: Make sure all required fields are available and that the data is properly formatted for Qlik Sense.
- ✓ Ensuring Integrity and Accuracy: Use validation and cross-checking to confirm that the data is accurate and comprehensive.

In order to get the dataset ready for analysis in Qlik Sense, a number of preprocessing procedures were followed. To make it easier to analyze patterns related to seasons, we extracted the month from the "Departure Date" column and generated a new "Month" column. In order to improve gender-based analysis, we standardized the "Gender" field to ensure equivalent marking for "male" and "female."Concatenating the "First Name" and "Last Name" columns into a single "Full Name" column was one of the additional improvements that helped with passenger identification. Additionally, we combined the "Country" and "Continent" columns to form a "Airport Address" column, which gives each airport a comprehensive location identity. These actions enhance the dataset's usability and structure, preparing it for in-depth Qlik Sense analysis and visualization.

Data Visualizations

The collection of data and information are represented graphically in data visualization. Data visualization tools offer an easy-to-use means of seeing and comprehending trends, outliers, and patterns in data through the use of visual elements like as charts, graphs, and maps. Data visualization is essential to business intelligence because it helps with decision-making, helps with the interpretation of large, complicated data sets, and effectively communicates findings.

The project efficiently analyzes and presents the artificial airline data by utilizing a variety of data visualization approaches. In the first set of visualizations, country_name is plotted on the X-axis and the count of airport_name is plotted on the Y-axis to show the number of flights that originate from each nation. Furthermore, by counting passenger_id, a KPI indicator displays the total number of unique passengers. Using departure_month as the dimension, a combination chart shows monthly flight counts and the average passenger age. The bar measure counts departure_date, while the line measure averages passenger age. Using a box plot with gender as the dimension and age as the measure, one may examine the age distribution of male and female passengers.

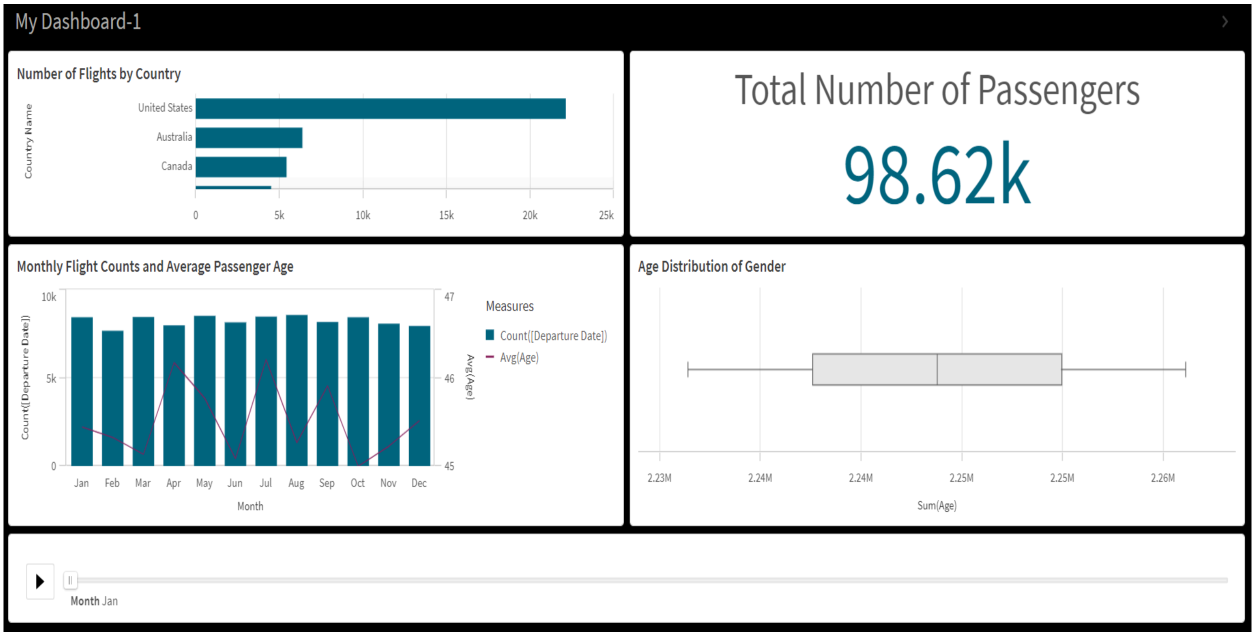
The percentage of flights with the status "On Time" is shown by an on-time flight percentage KPI in the second set. The distribution of flights by status is displayed as a pie chart, where the count of departure dates serves as the measure and flight status serves as the dimension. Using departure_date on the X-axis and age on the Y-axis, a scatter plot illustrates the link between the passenger's age and departure date. In addition, by plotting departure_date and counting the occurrences, a line chart illustrates the pattern of the number of flights over time.

With a specific focus on metrics like passenger ages broken down by gender, the third set uses natural language insights (NL Insights) to automatically provide summaries of important trends and patterns. The percentage of on-time flights is indicated on a gauge that shows the performance of the on-time flight. Users can explore data based on parameters including gender, flight status, departure date, and country name using a filter pane and a distribution plot that displays the gender distribution of age. The fourth set of data consists of a pivot table and a Mekko chart that show the distribution of flights by status and country. The measure is the number of departure dates, while the dimensions are the names of the countries and flight statuses.

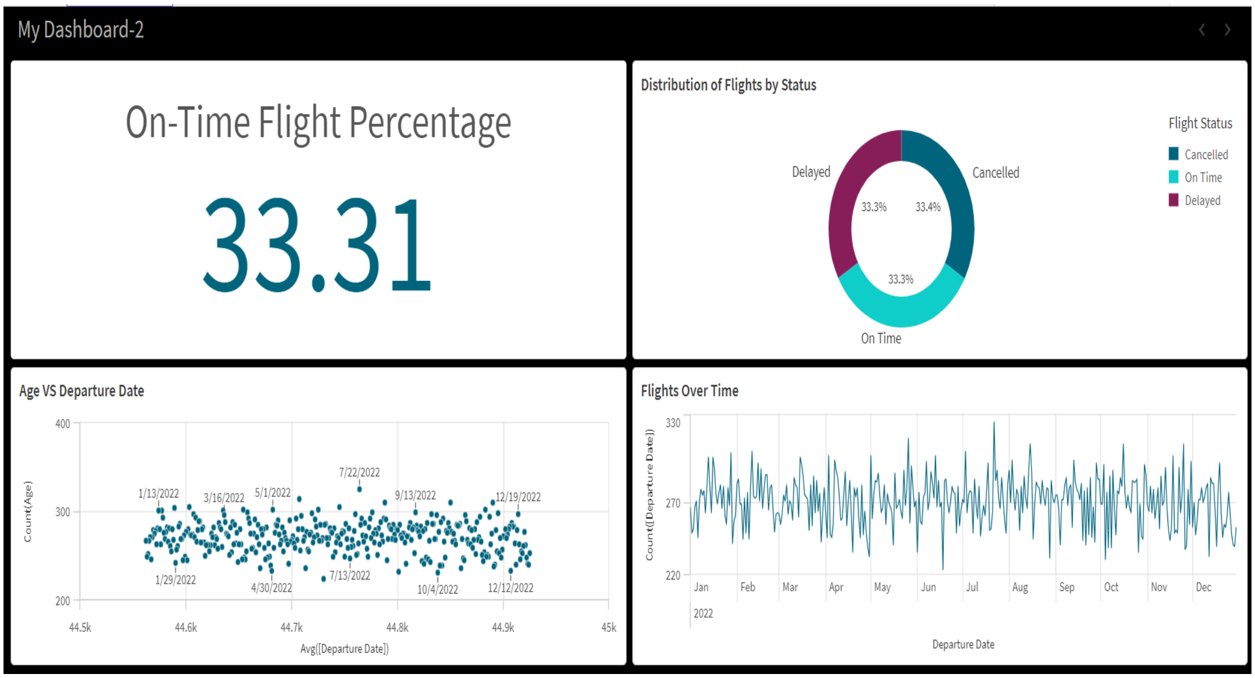
Finally, the project includes a treemap that depicts the hierarchy of flights from countries down to particular airports, with country and airport names as dimensions and departure dates as the measure. Text and images add context and explanation to the visualizations, while a button allows you to navigate between sheets or do activities like reloading data.

Dashboard:

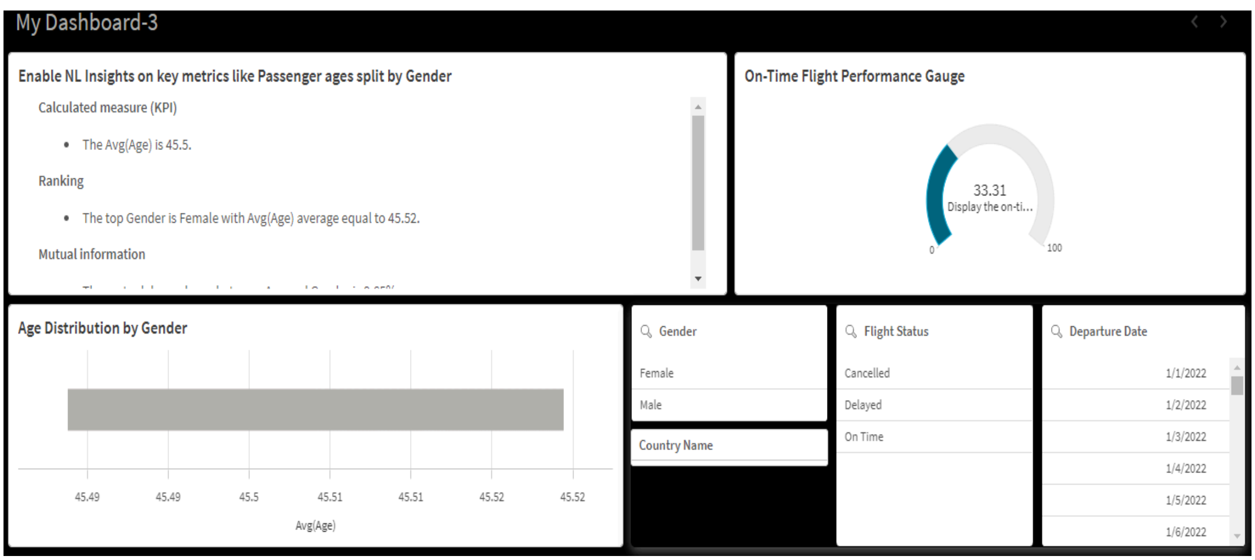
Using a variety of data visualization approaches, I have developed five efficient dashboards for this project in order to efficiently examine the simulated airline data.



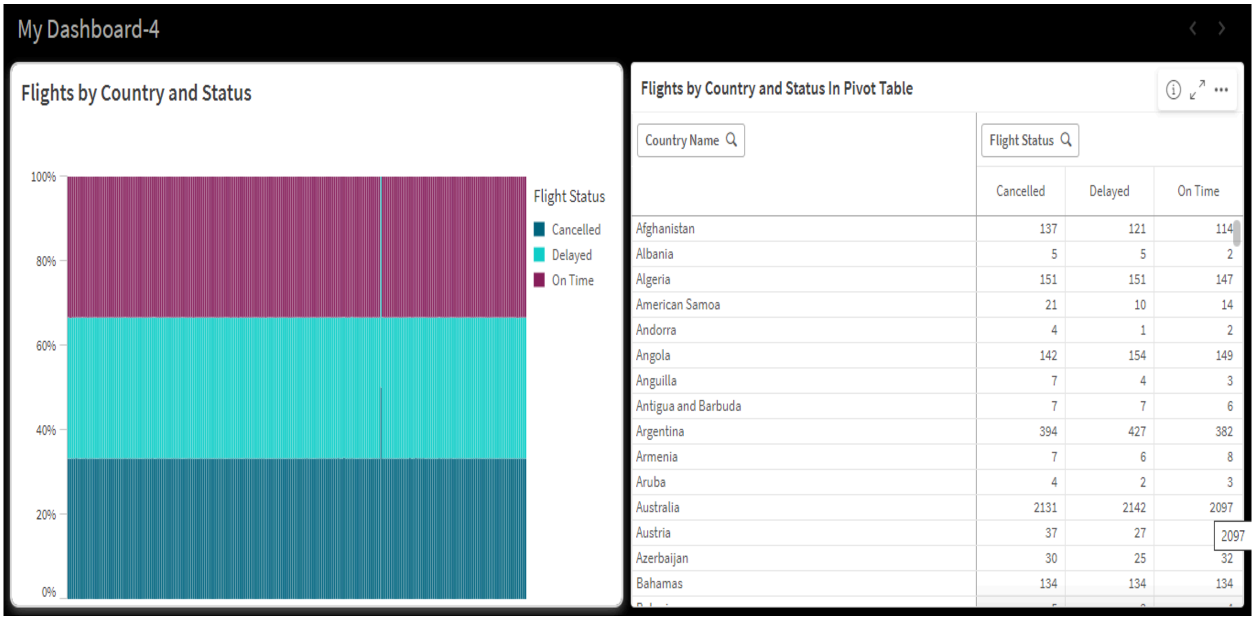
This dashboard displays the number of flights by country using bar plots, shows the total number of unique passengers with a KPI, and illustrates monthly flight counts and average passenger age with a combo chart. Additionally, it compares the age distribution by gender using a box plot.



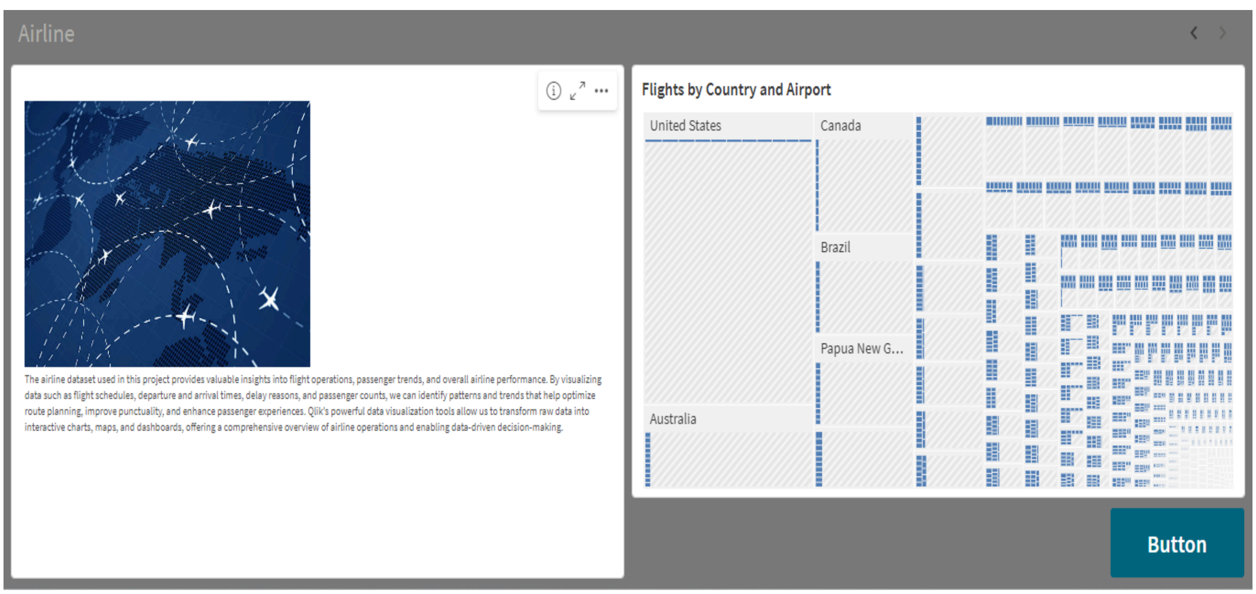
In this dashboard, the percentage of on-time flights is displayed with a KPI, while flight distribution by status is shown using a pie chart. It also reveals the relationship between passenger age and departure date with a scatter plot, and shows flight trends over time with a line chart.



This dashboard generates key trend summaries with NL Insights and displays on-time flight performance with a gauge. It also shows age distribution by gender using a distribution plot, and allows for data exploration with a filter pane.

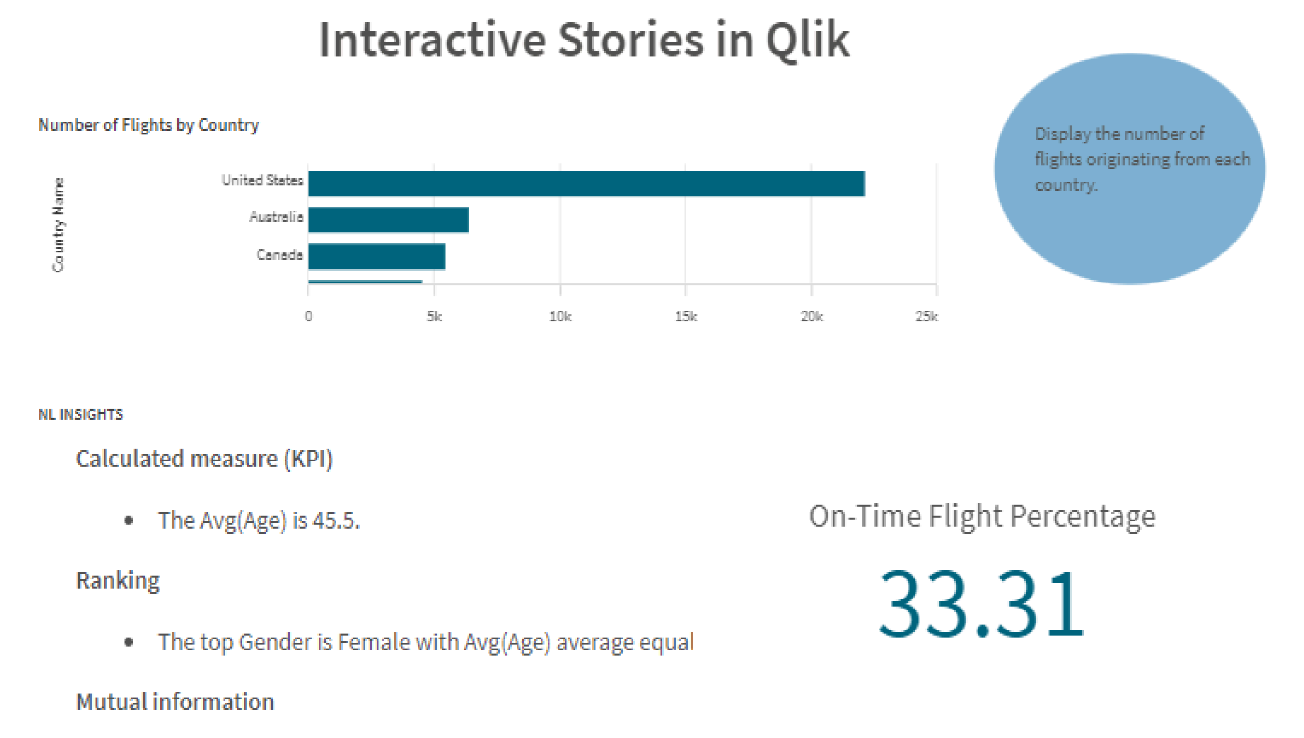


The fourth dashboard displays flight distribution by country and status with a Mekko chart and breaks down flights by country and status using a pivot table, providing a comprehensive view of these metrics.

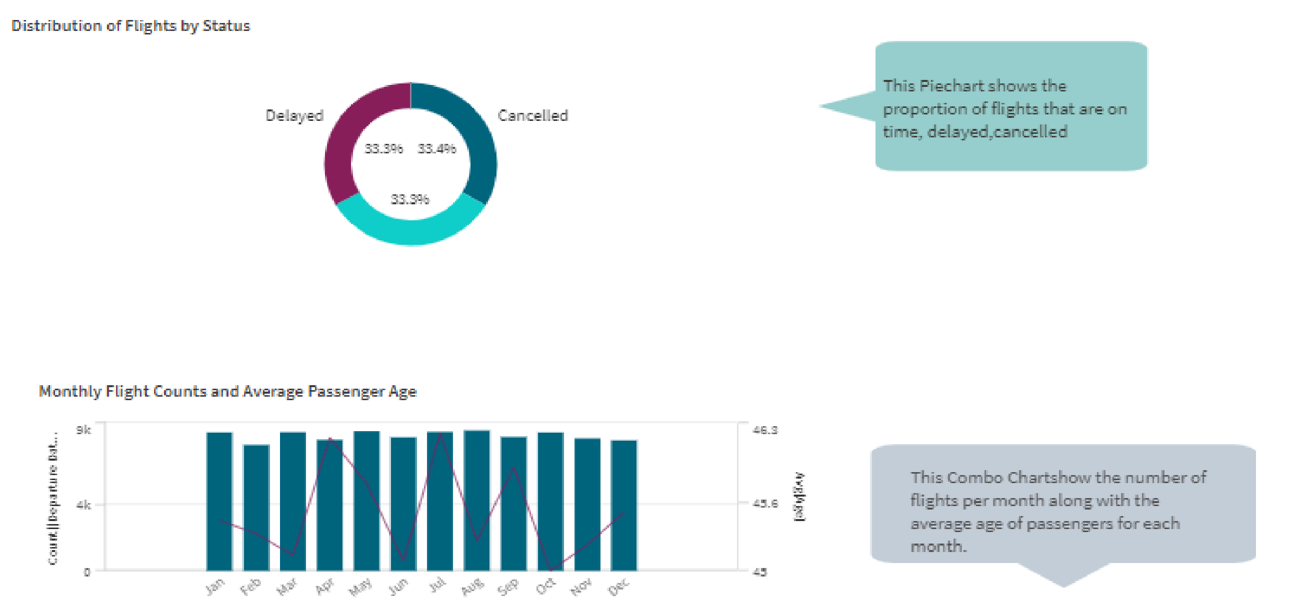


The final dashboard shows the hierarchy of flights from countries to individual airports with a treemap. It provides context and explanations with text and images and includes a button for navigation between sheets, enhancing the user experience.

Story:



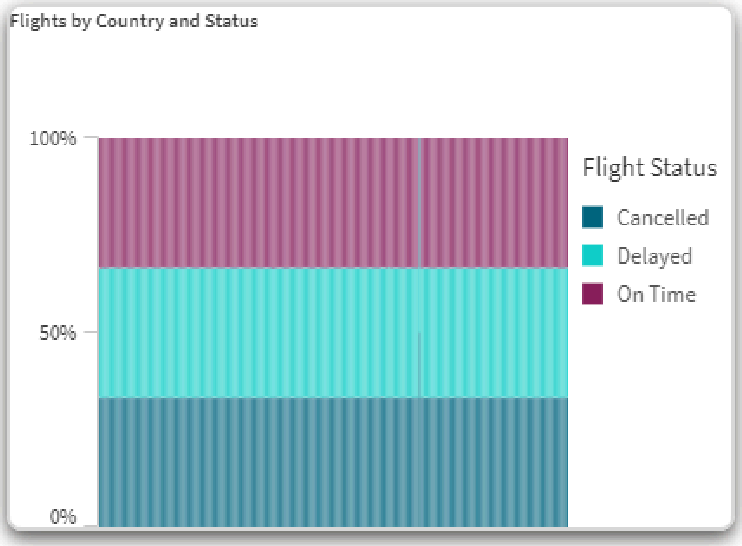
Interactive Stories in Qlik
Number of Flights by Country
Country Name
United States
Australia
Canada
0
5k
10k
15k
20k
25k
Display the number of flights originating from each country.
NL INSIGHTS
Calculated measure (KPI)
• The Avg(Age) is 45.5.
Ranking
• The top Gender is Female with Avg(Age) average equal
Mutual information
On-Time Flight Percentage
33.31



Distribution of Flights by Status
Delayed
Cancelled
33.3%
33.4%
33.3%
This Piechart shows the proportion of flights that are on time, delayed,cancelled
Monthly Flight Counts and Average Passenger Age
9k
4k
0
46.3
45.6
45
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
This Combo Chartshow the number of flights per month along with the average age of passengers for each month.

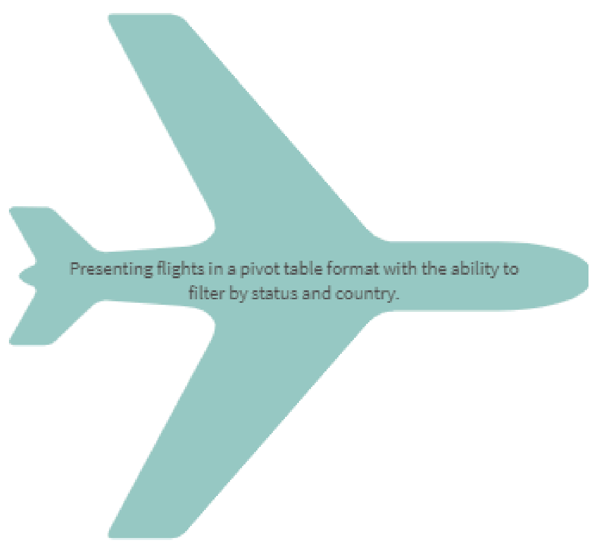
On-Time Flight Performance Gauge

Displaying the on-time flight percentage in a gauge format.



Flights by Country and Status In Pivot Table

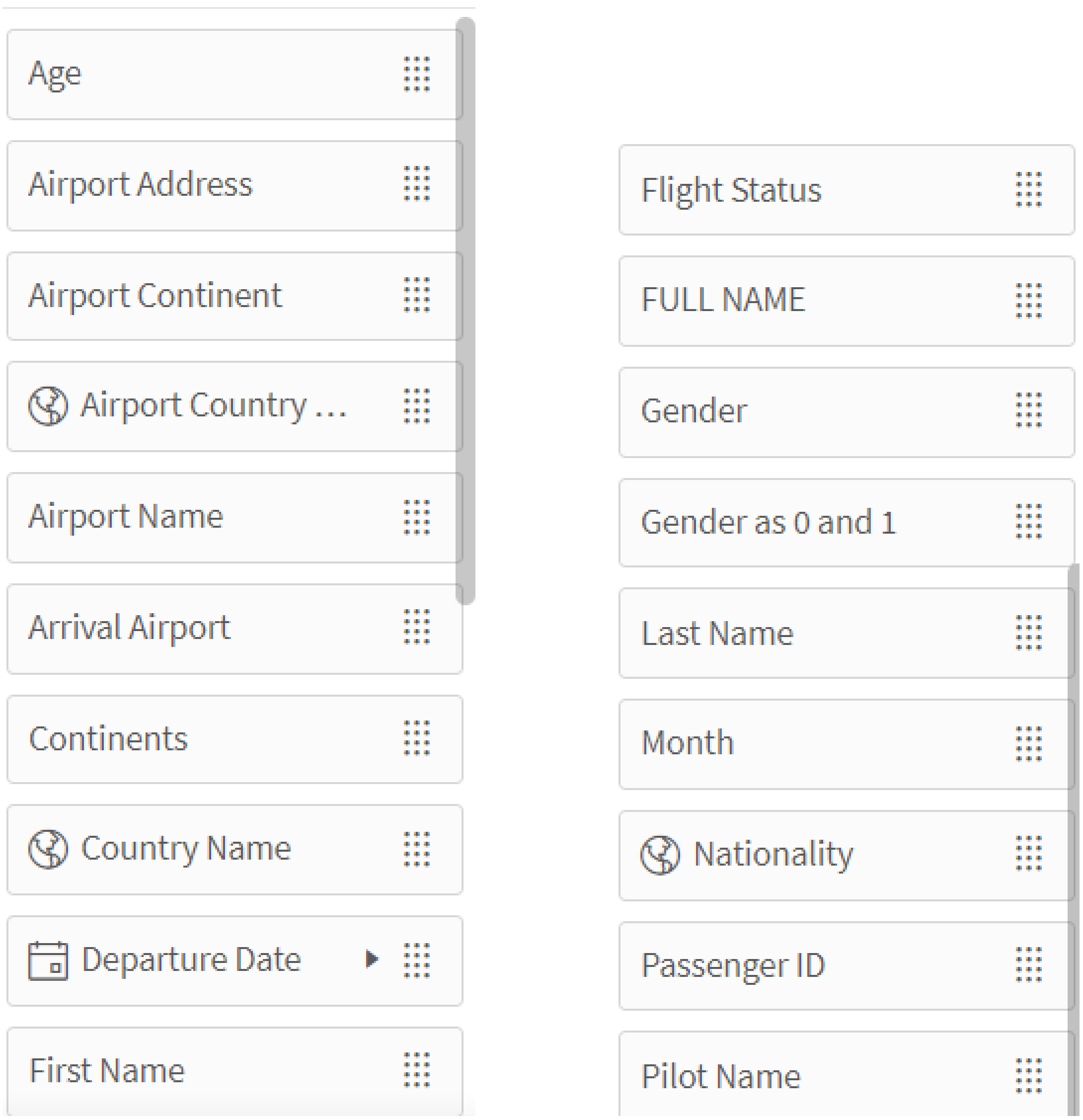
Country Name	Cancelled	Delayed	On Time
Afghanistan	137	121	114
Albania	5	5	2
Algeria	131	131	147
American Samoa	21	10	14
Andorra	4	1	2
Angola	142	134	149
Anguilla	7	4	3
Antigua and Barbuda	7	7	6
Argentina	394	427	382
Armenia	7	6	8
Aruba	4	2	3
Australia	2131	2142	2097
Austria	37	27	27
Azerbaijan	30	25	32
Bahamas	134	134	134
Bahrain	5	3	4



Performance Testing:

- Amount Of Data Loaded:

The term "Amount of Data Loaded" describes the quantity of data that is brought into a program or system to be processed and analyzed. It gauges the volume of data that has been effectively made available for system usage.



- ## Data Pre-Processing – Qlik Sense Script:

```
SET ThousandSep=',';
SET DecimalSep='.';
SET MoneyThousandSep=',';
SET MoneyDecimalSep='.';
SET MoneyFormat='$ ###0.00;-$ ###0.00';
SET TimeFormat='h:mm:ss TT';
SET DateFormat='M/D/YYYY';
SET TimestampFormat='M/D/YYYY h:mm:ss[.fff] TT';
SET FirstWeekDay=6;
SET BrokenWeeks=1;
SET ReferenceDay=0;
SET FirstMonthOfYear=1;
SET CollationLocale='en-US';
SET CreateSearchIndexOnReload=1;
SET MonthNames='Jan;Feb;Mar;Apr;May;Jun;Jul;Aug;Sep;Oct;Nov;Dec';
SET LongMonthNames='January;February;March;April;May;June;July;August;September;October;November;December';
SET DayNames='Mon;Tue;Wed;Thu;Fri;Sat;Sun';
SET LongDayNames='Monday;Tuesday;Wednesday;Thursday;Friday;Saturday;Sunday';
SET NumericalAbbreviation='3:k;6:M;9:G;12:T;15:P;18:E;21:Z;24:Y;-3:m;-6:μ;-9:n;-12:p;-15:f;-18:a;-21:z;-24:y';
```

```
Set dataManagerTables = '','Airline Dataset','Airline Dataset Updated';
//This block renames script tables from non generated section which conflict with the names of managed tables

For each name in $(dataManagerTables)
    Let index = 0;
    Let currentName = name;
    Let tableNumber = TableNumber(name);
    Let matches = 0;
    Do while not IsNull(tableNumber) or (index > 0 and matches > 0)
        index = index + 1;
        currentName = name & '-' & index;
        tableNumber = TableNumber(currentName)
        matches = Match('$(currentName)', $(dataManagerTables));
    Loop
    If index > 0 then
            Rename Table '$(name)' to '$(currentName)';
    EndIf;
Next;
Set dataManagerTables = ;


Unqualify *;

__countryAliasesBase:
LOAD
    Alias AS [__Country],
    ISO3Code AS [__ISO3Code]
FROM [lib://DataFiles/countryAliases.qvd]
(qvd);
```

```
__countryGeoBase:
LOAD
    ISO3Code AS [__ISO3Code],
    ISO2Code AS [__ISO2Code],
    Polygon AS [__Polygon]
FROM [lib://DataFiles/countryGeo.qvd]
(qvd);

__countryName2IsoThree:
MAPPING LOAD
    __Country,
    __ISO3Code
RESIDENT __countryAliasesBase;

__countryCodeIsoThree2Polygon:
MAPPING LOAD
    __ISO3Code,
    __Polygon
RESIDENT __countryGeoBase;

__countryCodeIsoTwo2Polygon:
MAPPING LOAD
    __ISO2Code,
    __Polygon
RESIDENT __countryGeoBase;

[Airline Dataset]:
LOAD
```

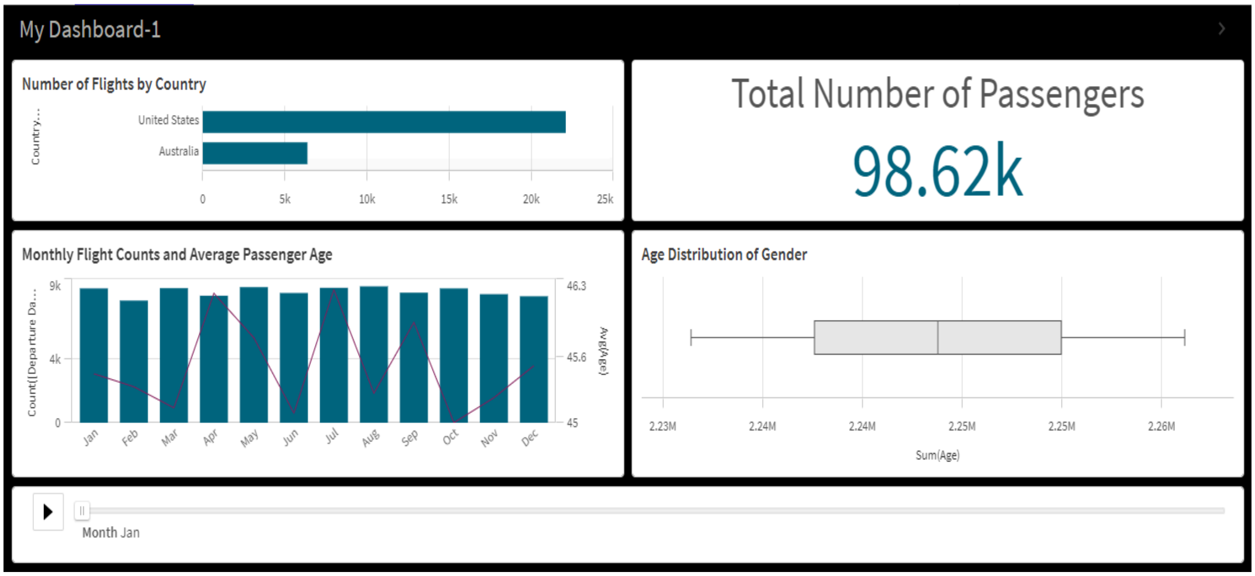
```
    [Passenger ID],
    [First Name],
    [Last Name],
    [Gender],
    [Age],
    [Nationality],
    [Airport Name],
    [Airport Country Code],
    [Country Name],
    [Airport Continent],
    [Continents],
    Date([Departure Date] ) AS [Departure Date],
    [Arrival Airport],
    [Pilot Name],
    [Flight Status],
    APPLYMAP( '__countryCodeIsoThree2Polygon', APPLYMAP( '__countryName2IsoThree', LOWER([Nationality])), '-') AS [@6_GeoInfo],
    APPLYMAP( '__countryCodeIsoThree2Polygon', APPLYMAP( '__countryName2IsoThree', LOWER([Country Name])), '-') AS [Airline Dataset.@9_GeoInfo],
    APPLYMAP( '__countryCodeIsoTwo2Polygon', UPPER([Airport Country Code]), '-') AS [Airline Dataset.Airport Country Code_GeoInfo],
    Month((Date([Departure Date] ))) AS [Month],
    if([Gender] = 'Female',1,0) AS [Gender as 0 and 1],
    [First Name] & ' ' & [Last Name]
 AS [FULL NAME],
    [Country Name] & ' ' & [Continents]
 AS [Airport Address]
 FROM [lib://DataFiles/Airline Dataset.csv]
(txt, utf8, embedded labels, delimiter is ',', msq);

[Airline Dataset Updated]:
LOAD
```

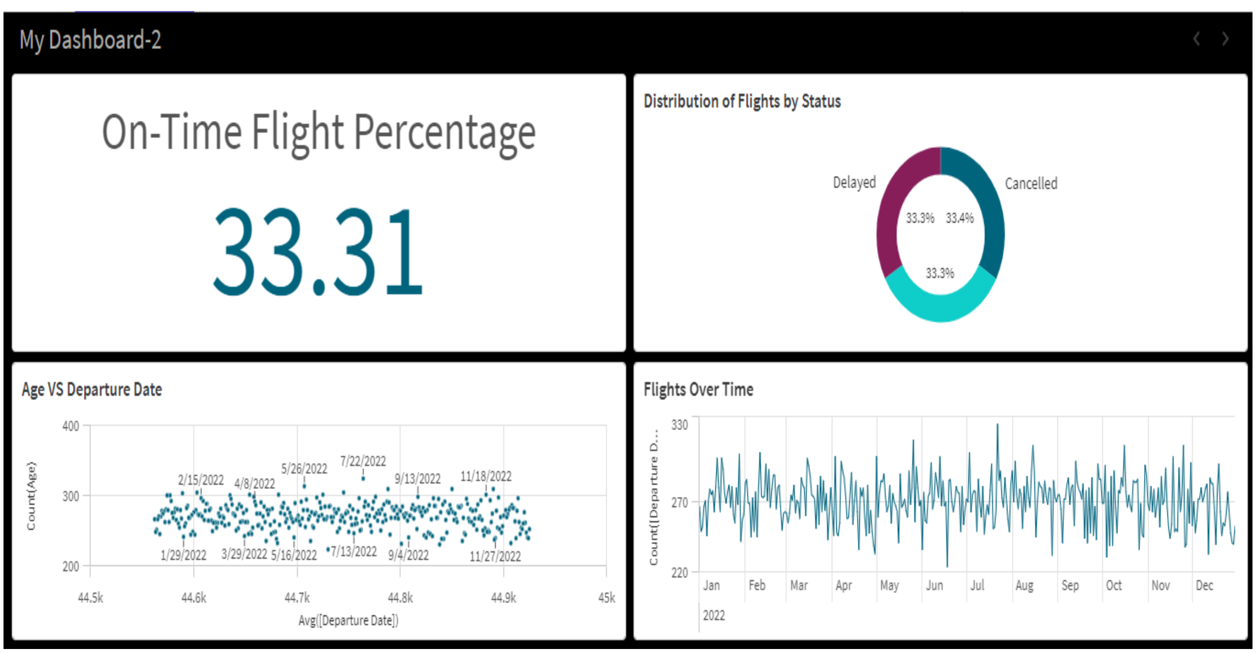
```
    [Passenger ID] AS [Airline Dataset Updated.Passenger ID],
    [First Name] AS [Airline Dataset Updated.First Name],
    [Last Name] AS [Airline Dataset Updated.Last Name],
    [Gender] AS [Airline Dataset Updated.Gender],
    [Age] AS [Airline Dataset Updated.Age],
    [Nationality] AS [Airline Dataset Updated.Nationality],
    [Airport Name] AS [Airline Dataset Updated.Airport Name],
    [Airport Country Code] AS [Airline Dataset Updated.Airport Country Code],
    [Country Name] AS [Airline Dataset Updated.Country Name],
    [Airport Continent] AS [Airline Dataset Updated.Airport Continent],
    [Continents] AS [Airline Dataset Updated.Continents],
    Date([Departure Date] ) AS [Airline Dataset Updated.Departure Date],
    [Arrival Airport] AS [Airline Dataset Updated.Arrival Airport],
    [Pilot Name] AS [Airline Dataset Updated.Pilot Name],
    [Flight Status] AS [Airline Dataset Updated.Flight Status],
    APPLYMAP( '__countryCodeIsoThree2Polygon', APPLYMAP( '__countryName2IsoThree', LOWER([Nationality])), '-') AS [@6_GeoInfo],
    APPLYMAP( '__countryCodeIsoThree2Polygon', APPLYMAP( '__countryName2IsoThree', LOWER([Country Name])), '-') AS [Airline Dataset Updated.@9_GeoInfo],
    APPLYMAP( '__countryCodeIsoTwo2Polygon', UPPER([Airport Country Code]), '-') AS [Airline Dataset Updated.Airline Dataset Updated.Airport Country Code_GeoInfo]
 FROM [lib://DataFiles/Airline Dataset Updated.csv]
(txt, utf8, embedded labels, delimiter is ',', msq);
```

```
TAG FIELD [Nationality] WITH '$geoname', '$relates_@6_GeoInfo';
TAG FIELD [@6_GeoInfo] WITH '$geopolygon', '$hidden', '$relates_Nationality';
TAG FIELD [Country Name] WITH '$geoname', '$relates_Airline Dataset.@9_GeoInfo';
TAG FIELD [Airline Dataset.@9_GeoInfo] WITH '$geopolygon', '$hidden', '$relates_Country Name';
TAG FIELD [Airport Country Code] WITH '$geoname', '$relates_Airline Dataset.Airport Country Code_GeoInfo';
TAG FIELD [Airline Dataset.Airport Country Code_GeoInfo] WITH '$geopolygon', '$hidden', '$relates_Airport Country Code';

TAG FIELD [Airline Dataset Updated.Nationality] WITH '$geoname', '$relates_@6_GeoInfo';
TAG FIELD [@6_GeoInfo] WITH '$geopolygon', '$hidden', '$relates_Airline Dataset Updated.Nationality';
TAG FIELD [Airline Dataset Updated.Country Name] WITH '$geoname', '$relates_Airline Dataset Updated.@9_GeoInfo';
TAG FIELD [Airline Dataset Updated.@9_GeoInfo] WITH '$geopolygon', '$hidden', '$relates_Airline Dataset Updated.Country Name';
TAG FIELD [Airline Dataset Updated.Airport Country Code] WITH '$geoname', '$relates_Airline Dataset Updated.Airline Dataset Updated.Airport Country Code_GeoInfo';
TAG FIELD [Airline Dataset Updated.Airline Dataset Updated.Airport Country Code_GeoInfo] WITH '$geopolygon', '$hidden', '$relates_Airline Dataset Updated.Airport Country Code';
```

```
  Dual(Year($1), YearStart($1)) AS [Year] Tagged ('$axis', '$year'),
  Dual('Q'&Num(Ceil(Num(Month($1))/3)),Num(Ceil(NUM(Month($1))/3),00)) AS [Quarter] Tagged ('$quarter', '$cyclic'),
  Dual(Year($1)&'-Q'&Num(Ceil(Num(Month($1))/3)),QuarterStart($1)) AS [YearQuarter] Tagged ('$yearquarter', '$qualified'),
  Dual('Q'&Num(Ceil(Num(Month($1))/3)),QuarterStart($1)) AS [_YearQuarter] Tagged ('$yearquarter', '$hidden', '$simplified'),
  Month($1) AS [Month] Tagged ('$month', '$cyclic'),
  Dual(Year($1)&'-'&Month($1), monthstart($1)) AS [YearMonth] Tagged ('$axis', '$yearmonth', '$qualified'),
  Dual(Month($1), monthstart($1)) AS [_YearMonth] Tagged ('$axis', '$yearmonth', '$simplified', '$hidden'),
  Dual('W'&Num(Week($1),00), Num(Week($1),00)) AS [Week] Tagged ('$weeknumber', '$cyclic'),
  Date(Floor($1)) AS [Date] Tagged ('$axis', '$date', '$qualified'),
  Date(Floor($1), 'D') AS [_Date] Tagged ('$axis', '$date', '$hidden', '$simplified'),
  If (DayNumberOfYear($1) <= DayNumberOfYear(Today()), 1, 0) AS [InYTD] ,
  Year(Today())-Year($1) AS [YearsAgo] ,
  If (DayNumberOfQuarter($1) <= DayNumberOfQuarter(Today()),1,0) AS [InQTD] ,
  4*Year(Today())+Ceil(Month(Today())/3)-4*Year($1)-Ceil(Month($1)/3) AS [QuartersAgo] ,
  Ceil(Month(Today())/3)-Ceil(Month($1)/3) AS [QuarterRelNo] ,
  If(Day($1)<=Day(Today()),1,0) AS [InMTD] ,
  12*Year(Today())+Month(Today())-12*Year($1)-Month($1) AS [MonthsAgo] ,
  Month(Today())-Month($1) AS [MonthRelNo] ,
  If(WeekDay($1)<=WeekDay(Today()),1,0) AS [InWTD] ,
  (WeekStart(Today())-WeekStart($1))/7 AS [WeeksAgo] ,
  Week(Today())-Week($1) AS [WeekRelNo] ;

DERIVE FIELDS FROM FIELDS [Departure Date], [Airline Dataset Updated.Departure Date] USING [autoCalendar] ;
```

- Utilization Of Filters:





No Of Visualizations/ Graphs:

- Number of Flights by Country
- Total Number of Passengers
- Monthly Flight Counts and Average Passenger Age
- Age Distribution by Gender
- On-Time Flight Percentage
- Distribution of Flights by Status
- Age vs. Departure Date
- Flights Over Time
- Enable NL Insights on key metrics like Passenger ages split by gender
- On-Time Flight Performance Gauge
- Age Distribution by Gender
- Filter Options for Data Exploration
- Flights by Country and Status
- Flights by Country and Status
- Airline Flight Data Analysis
- Flights by Country and Airport