

# Project Report

Date	9th June 2024
Name	Simran Namdev
Project Name	Exploring Insights From Synthetic Airline Data Analysis With Qlik

## 1. INTRODUCTION

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.

### **Scenario 1: Revenue Optimization**

An airline wants to optimize its revenue by analyzing 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 behavior, and adjust pricing strategies accordingly to maximize profitability.

### **Scenario 2: Operational Efficiency**

An airport authority aims to enhance operational efficiency by analyzing 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.

## 2. DATA COLLECTION AND EXTRACTION FROM DATABASE

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.

The down loaded data is from kaggle website:

<https://www.kaggle.com/datasets/iamsouravbanerjee/airline-dataset?resource=download>

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Passenger	First Name	Last Name	Gender	Age	Nationality	Airport Na	Airport Co	Country N	Airport Co	Continents	Departure	Arrival	Airline	Pilot Name
2	ABVWig	Edith	Leggis	Female	62	Japan	Coldfoot	US	United Sta	NAM	North Ame	6/28/2022	CXF	Fransisco	I On Time
3	jkXXAX	Elwood	Catt	Male	62	Nicaragua	Kugluktuk	CA	Canada	NAM	North Ame	12/26/2022	YCO	Marla	Pars On Time
4	CdUz2g	Darby	Felgate	Male	67	Russia	Grenoble	IFR	France	EU	Europe	1/18/2022	GNB	Rhonda	Ar On Time
5	BRS38V	Dominica	Pyle	Female	71	China	Ottawa	/ CCA	Canada	NAM	North Ame	9/16/2022	YND	Kacie	Com Delayed
6	9kvTLo	Bay	Pencost	Male	21	China	Gillespie	Fl US	United Sta	NAM	North Ame	2/25/2022	SEE	Ebonee	Tr On Time
7	rMJKVh	Lora	Durbann	Female	55	Brazil	Coronel	Hé BR	Brazil	SAM	South Ame	#####	LEC	Inglis	Dolle On Time
8	8IPFPE	Rand	Bram	Male	73	Ivory Coas	Duxford	A GB	United Kin	EU	Europe	10/30/2022	QFO	Stanislas	T Cancelled
9	pqixbY	Perceval	Dalloso	Male	36	Vietnam	Maestro	BR	Brazil	SAM	South Ame	#####	STM	Sharyl	East Cancelled
10	QNAs2R	Aleda	Pigram	Female	35	Palestinian	Venice	Ma IT	Italy	EU	Europe	8/20/2022	VCE	Daryn	Bart On Time
11	3jmudz	Burlie	Schustl	Male	13	Thailand	Vermilion	CA	Canada	NAM	North Ame	#####	YVG	Alameda	C On Time
12	2P41gZ	Porty	Jori	Male	39	Tunisia	Nuevo Cas	MX	Mexico	NAM	North Ame	5/27/2022	NCG	Rasia	Fidel Cancelled
13	sBF524	Briant	De La Hay	Male	71	Russia	Ruben	Can PA	Panama	NAM	North Ame	#####	SYP	Alina	Flood Delayed
14	PlwjZT	Kalie	Scoble	Female	47	Sweden	Loralai	Air PK	Pakistan	AS	Asia	3/19/2022	LRG	Madelena	Delayed
15	iU75x3	Catrina	Beaumont	Female	77	Russia	Cudal	Airp AU	Australia	OC	Oceania	3/24/2022	CUG	Margie	Be Delayed
16	GUTa6R	Amberly	Handling	Female	32	China	Farmingto	US	United Sta	NAM	North Ame	#####	FAM	Lothaire	E Delayed
17	8qA80a	Dyna	De Vere - I	Female	22	China	Oudtshoorn	ZA	South Afric	AF	Africa	7/18/2022	OUH	Neila	Gier Cancelled
18	2haCDu	Janella	Hardaker	Female	28	Colombia	Zaraza	Air VE	Venezuela	SAM	South Ame	9/23/2022	ZRZ	Shaylynn	A On Time
19	WIVL8o	Alvin	Wenzel	Male	12	Greece	Enshi	Airpc CN	China	AS	Asia	3/29/2022	ENH	Alfie	MacN Delayed
20	e0H5LI	Jerrine	Peeters	Female	87	Philippines	Thompson	CA	Canada	NAM	North Ame	#####	YTH	Chandra	D Cancelled
21	nL8kyD	Warner	Driutti	Male	62	China	Guilin	Lian CN	China	AS	Asia	#####	KWL	Marita	Hol Cancelled
22	9iT79e	Paige	Hayhow	Male	24	Sweden	Crested	Bt US	United Sta	NAM	North Ame	2/19/2022	CSE	Clyde	Wini Delayed
23	kEARqp	Dorisca	Skill	Female	19	Ukraine	St Augustin	CA	Canada	NAM	North Ame	#####	YIF	Ilyse	Bartk Cancelled
24	dx3NWh	Bobbye	Patmore	Female	45	China	Port Berg	MG	Madagascar	AF	Africa	8/24/2022	WPB	Stella	Pitt On Time
25	Dpfafly	Jayne	Dairton	Female	29	Indonesia	Hato	Coro CO	Colombia	SAM	South Ame	#####	HTZ	Kaye	Clew Delayed
26	WV3aXu	Sayre	Stroyan	Male	87	Indonesia	Elkhart	Mt US	United Sta	NAM	North Ame	3/20/2022	EKI	Austine	Cr Cancelled

### 3. DATA PREPARATION

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. Since the data is already cleaned, we can move to visualization.

Exploring Insights from Synthetic Airline Data Analysis  
Data last loaded: No data loaded

Get started adding data to your app.

**Data catalog**  
Access data that's available to you

**Files and other sources**  
Drag and drop files or click to browse files and connections

**Data load editor**  
Load data and perform transformations.

From the above page we load the data by selecting the left option in the image.

Add data to Exploring Insights from Synthetic Airline Data Analysis

File format: Delimited Field names: No field names Delimiter: Comma Character set: 65001 (UTF-8)

	@1	@2	@3	@4	@5	@6	@7	@8	@9
Passenger ID	First Name	Last Name	Gender	Age	Nationality	Airport Name	Airport Country Code	Country Name	
ABVWlg	Edith	Leggis	Female	62	Japan	Coldfoot Airport	US	United States	
jXkXXA	Elwood	Catt	Male	62	Nicaragua	Kugluktuk Airport	CA	Canada	
CdfUz2g	Darby	Felgate	Male	67	Russia	Grenoble-Isère Airport	FR	France	
BRs38V	Dominica	Pyle	Female	71	China	Ottawa / Gatineau Airport	CA	Canada	
9kvTLo	Bay	Pencost	Male	21	China	Gillespie Field	US	United States	
mJKVh	Lora	Durban	Female	55	Brazil	Coronel Horácio de Mattos Airport	BR	Brazil	
8IPFPE	Rand	Bram	Male	73	Ivory Coast	Duxford Aerodrome	GB	United Kingdom	
pqxbY	Perceval	Dalloso	Male	36	Vietnam	Maestro Wilson Fonseca Airport	BR	Brazil	
QNaS2R	Aleda	Pigram	Female	35	Palestinian Territory	Venice Marco Polo Airport	IT	Italy	

Our uploaded data does not have columns name to change the first row data into column named change field names to embedded.

Add data to Exploring Insights from Synthetic Airline Data Analysis

File format: Delimited, Field names: Embedded field names, Delimiter: Comma, Quoting: MSQ, Header size: 0, Character set: 65001 (UTF-8), Ignore End-Of-File character?

Passeng...	First N...	Last N...	Ge...	Age	Nationality	Airport Name	Airport Country
ABVWlg	Edithe	Leggis	Female	62	Japan	Coldfoot Airport	US
jkXXAX	Elwood	Catt	Male	62	Nicaragua	Kugluktuk Airport	CA
CdUz2g	Darby	Felgate	Male	67	Russia	Grenoble-Isère Airport	FR
BR538V	Dominica	Pyle	Female	71	China	Ottawa / Gatineau Airport	CA
9kvtLo	Bay	Pencost	Male	21	China	Gillespie Field	US
hMJKWh	Lora	Durbann	Female	55	Brazil	Coronel Horácio de Mattos Airport	BR
BiPPFE	Rand	Bram	Male	73	Ivory Coast	Duxford Aerodrome	GB
pqjxbY	Perceval	Dalloso	Male	36	Vietnam	Maestro Wilson Fonseca Airport	BR
QNAszR	Aleda	Pigram	Female	35	Palestinian Territory	Venice Marco Polo Airport	IT
3jmudz	Burtsie	Schustl	Male	13	Thailand	Vermilion Airport	CA

Now the data column names are displayed , so click next.

```

SET ThousandSep(',');
SET DecimalSep '.';
SET MoneyThousandSep(',');
SET MoneyDecimalSep '.';
SET MoneyFormat'$ ##0.00;-$ ##0.00';
SET TimeFormat'hh:mm:ss TT';
SET DateFormat'M/D/YYYY';
SET TimestampFormat'M/D/YYYY h:mm:ss[.ffff] TT';
SET FirstWeekDay=6;
SET BrokenWeeks=1;
SET ReferenceDay=0;
SET FirstMonthInYear=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:u;-9:n;-12:p;-15:f;-18:a;-21;z;-24:y';

```

Total number of rows: 98619

Pass...	First ...	Last ...	Gender	Age	Natio...	Airpo...	Airpo...	Coun...	Airpo...	Contl...	Depa...	Arriv...	Pilot ...
ABVWlg	Edithe	Leggis	Female	62	Japan	Coldfoot Airport	US	United States	NAM	North America	-	CXF	Fransisco Ha
jkXXAX	Elwood	Catt	Male	62	Nicaragua	Kugluktuk Airpor	CA	Canada	NAM	North America	-	YCO	Marla Parsor
CdUz2g	Darby	Felgate	Male	67	Russia	Grenoble-Isère Ai	FR	France	EU	Europe	-	GNR	Rhonda Am

This the the format of data stored in the database.

Airline Dataset Updated - v2

Total tables: 1  
Unassociated tables: 1  
Recommendations: 0

To make associations manually, you can drag one table onto another.

Fields: 15

Passenger ID	First Name	Last Name	Gender	Age	Nationality	Airport Name	Airport Cou...	Country Name
165488	Dionis	Joist	Female	4	Philippines	Bremen Airport	DE	Germany
824967	Federica	Peters	Female	16	Portugal	Watson Lake Airport	CA	Canada
0a1ws9	Lammmond	Sargood	Male	59	Serbia	Karluk Lake Seaplane Base	US	United States
0A1y0a	Arel	Beswick	Male	83	United States	Holy Cross Airport	US	United States
0A2OYI	Jack	Mitrikhin	Male	89	Turkey	Rottnest Island Airport	AU	Australia
0a5xSP	Perry	Pretsell	Male	64	Nigeria	Vallenar Airport	CL	Chile

After successfully loading the dataset we have one relation (table) named "Airline Dataset Update-v2" in our database.

## 4. DATA VISUALIZATION

Data visualization is the process of creating graphical representations of data to help people understand and explore the information. The goal of data visualization is to make complex data sets more accessible, intuitive, and easier to interpret. By using visual elements such as charts, graphs, and maps, data visualizations can help people quickly identify patterns, trends, and outliers in the data.

### Activity 1.1: Total No. of Passengers

My new sheet

Click to add title

Total Number of Passengers

98.62k

Chart suggestions

Measures

First KPI

Total Number of Passengers

Expression

Count([Passenger ID])

Label

Total Number of Passengers

Number formatting

Auto

Master item

Add new

Delete

**Activity 1.2: Number of Passengers effected by cancelled flights**

## Number of Passengers effected by cancelled flights

# 32.85k

① OK  
=Count({<'Flight Status'='On Time'>} `Flight Status`)

Cancel Apply

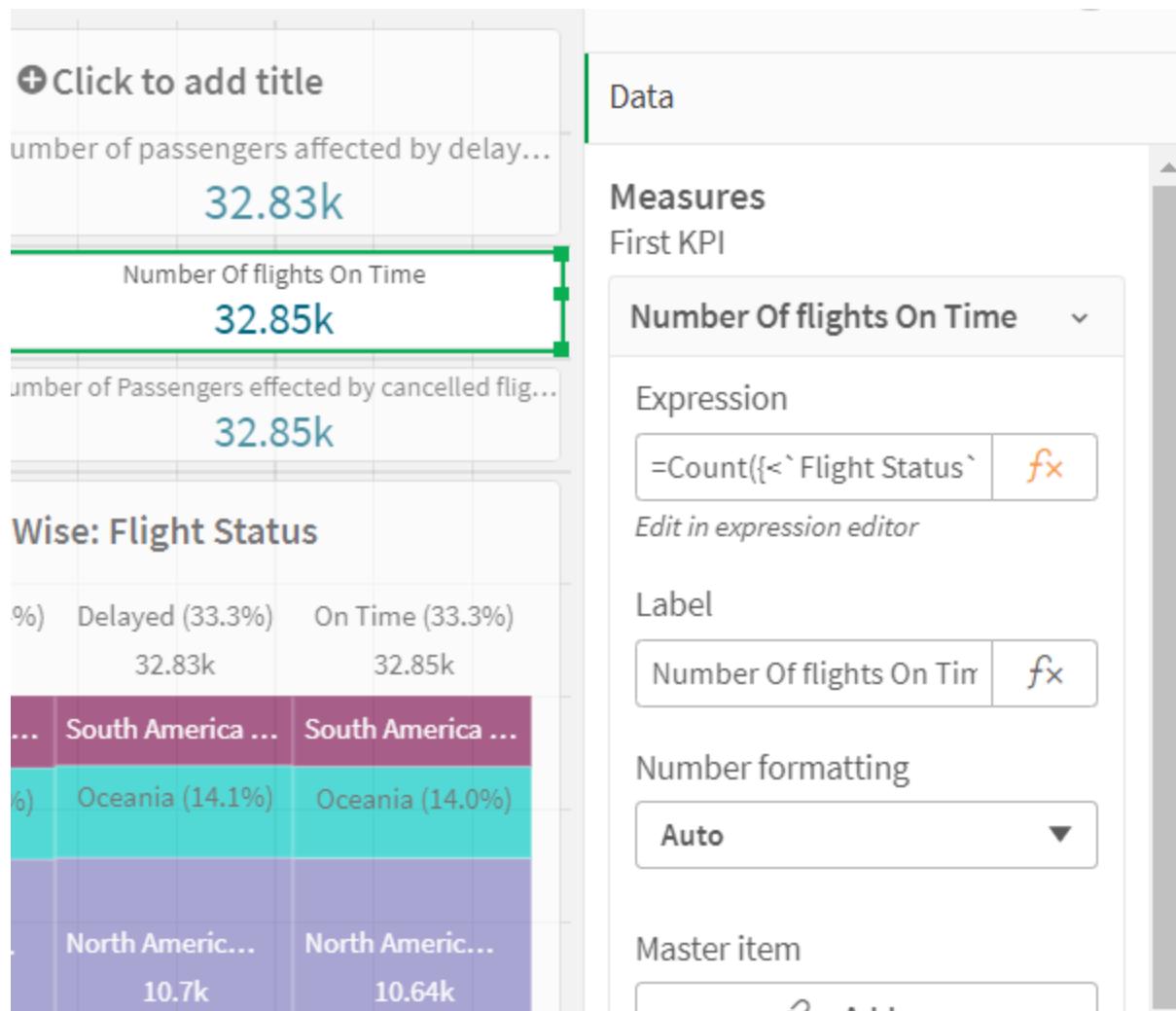
The screenshot shows a data visualization tool's interface. At the top, there is a large text box containing the title "Number of Passengers effected by cancelled flights" and a large numerical value "32.85k". Below this is a smaller text box with the expression "=Count({<'Flight Status'='On Time'>} `Flight Status`)" and buttons for "OK", "Cancel", and "Apply". A green border surrounds the main title and value. At the bottom right of the main area, there is a coordinate indicator "(0,8) 24 x 4". The background of the interface includes various toolbars and a sidebar labeled "Edit expression" with sections for "Fields", "Aggregation functions", and "Set Expression".

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

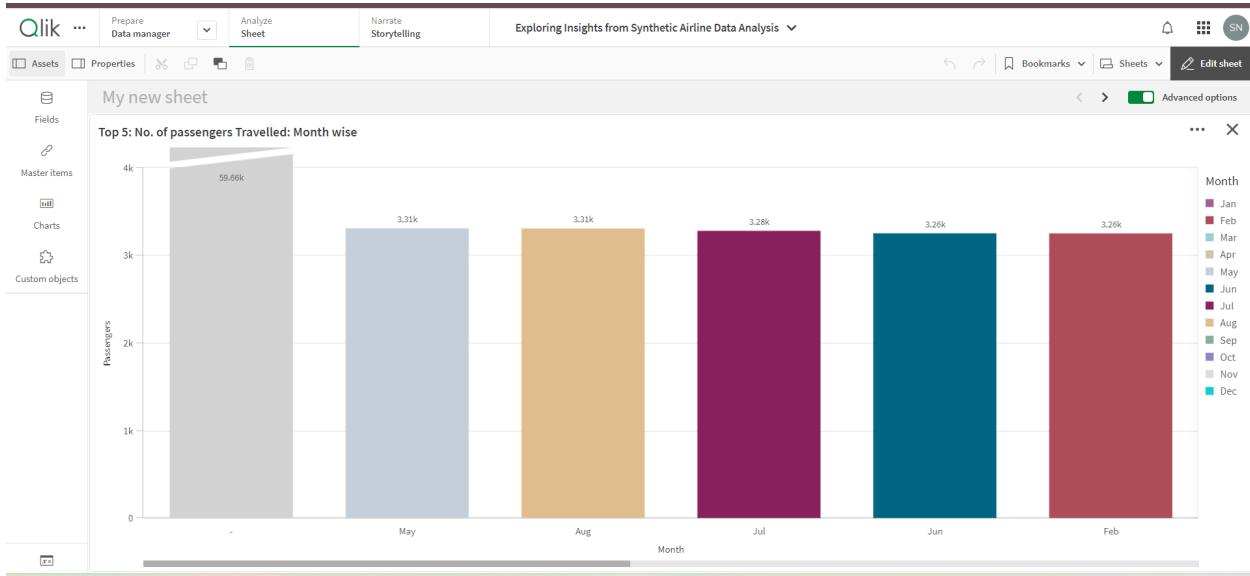
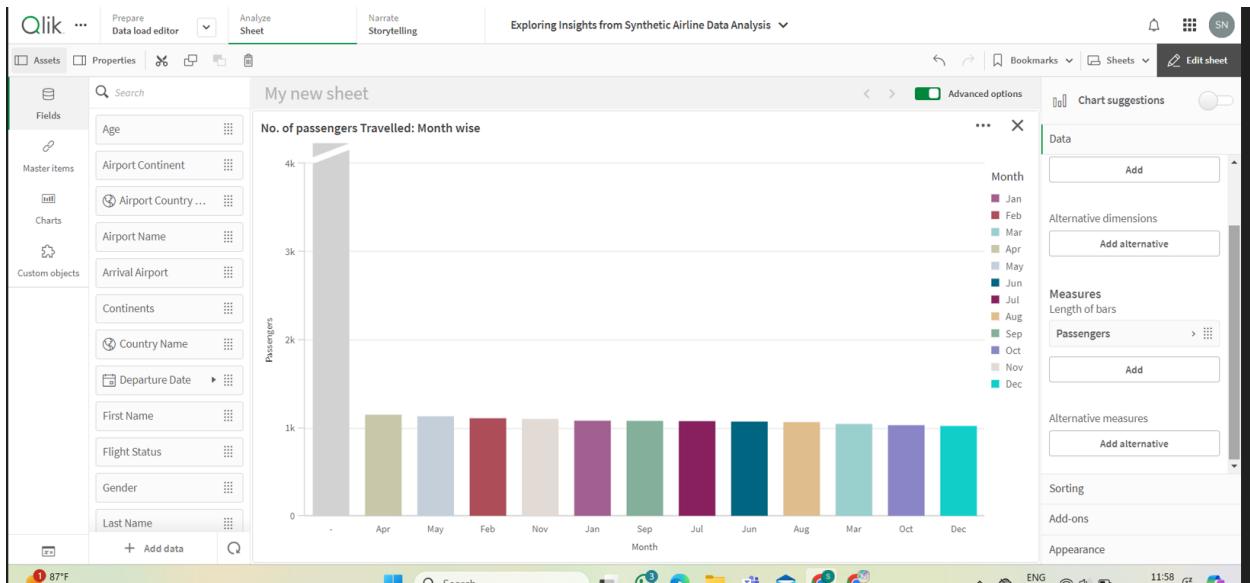
The screenshot shows the Qlik Sense interface with the following details:

- Top Bar:** Qlik, Prepare, Data manager, Analyze Sheet (selected), Narrate Storytelling, Exploring Insights from Synthetic Airline Data Analysis, Advanced options, Edit sheet.
- Left Sidebar:** Fields, Master items (Charts selected), Custom objects.
- Central Area:** My new sheet, title "Number of passengers affected by delayed flight", value "32.83k".
- Right Sidebar:** #1 Chart suggestions, Data (First KPI), Number of passengers aff..., Expression (=Count({<'Flight Status'='Delayed'})), Label (Number of passengers af...), Number formatting (Auto), Master item (Add new), Second KPI, Add-ons, Appearance.
- Bottom Window:** Edit expression, showing the expression =Count({<'Flight Status'='Delayed'}), with a note: "OK =Count({<'Flight Status'='Delayed'}), 'Flight Status')". The right pane of the expression editor shows fields (Flight Status), filter by table (All tables), aggregation functions (Count, Total checked), and other options like Expression generator, Functions, and Set expressions.

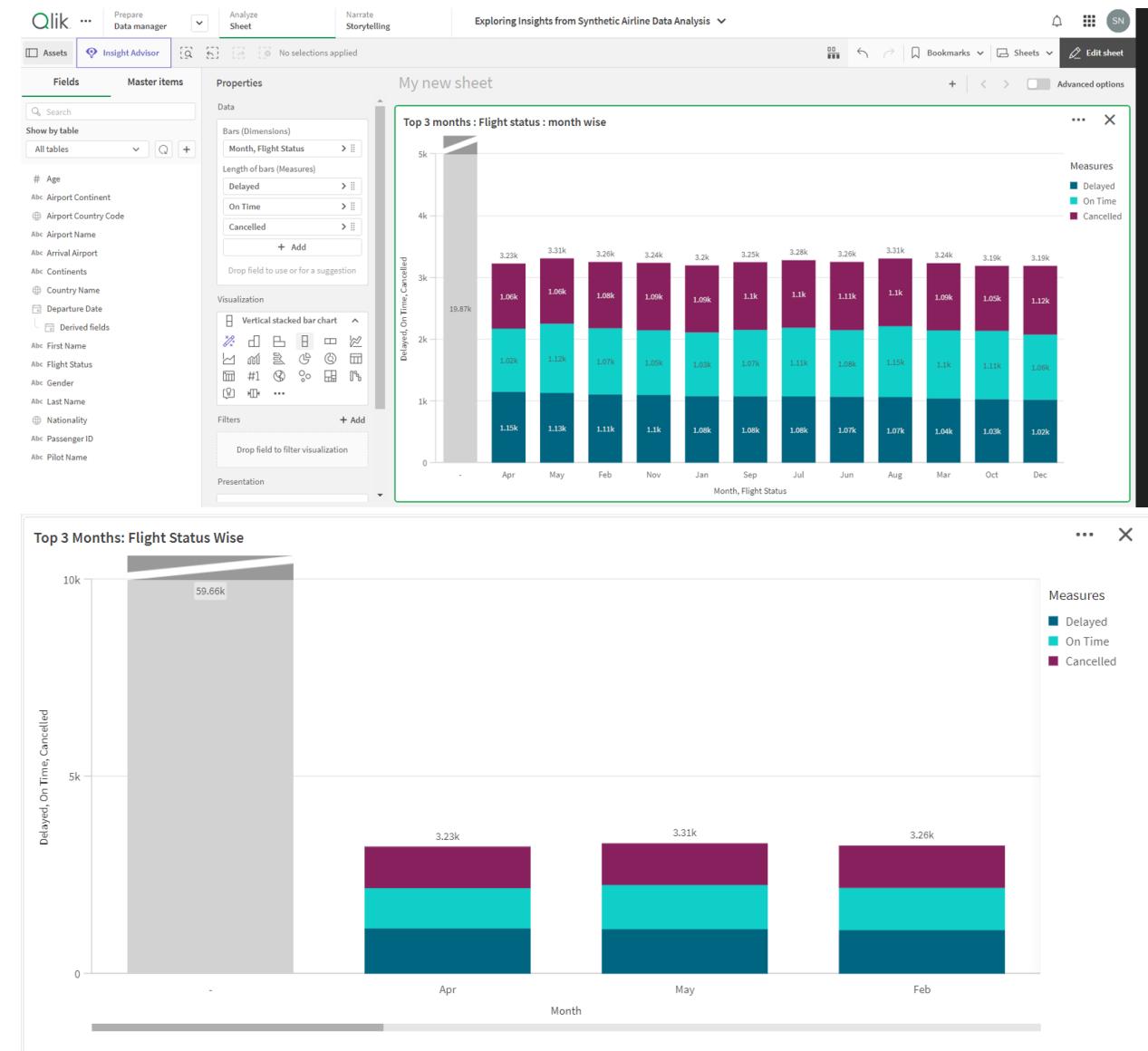
#### Activity 1.4: Number of Flights on Time



### Activity 1.5: No of Passengers travelled- Month Wise



### Activity 1.6: Top 3 Month flights status wise



**Measures**

Length of bars

Delayed > [grid icon]

On Time > [grid icon]

Cancelled > [grid icon]

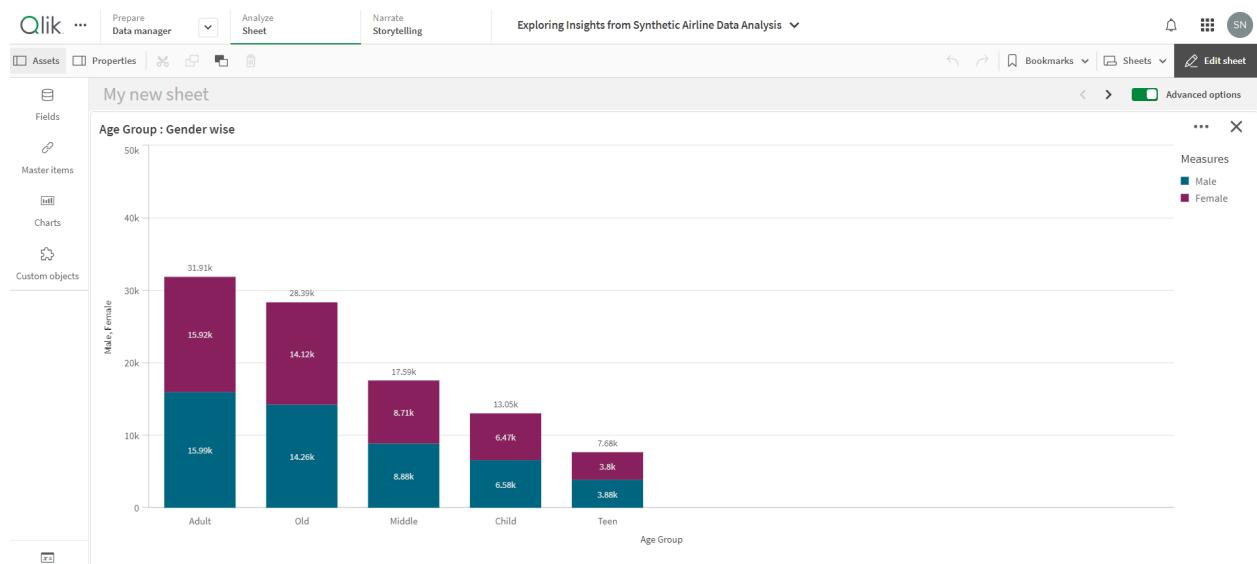
Add

**Edit expression**

```

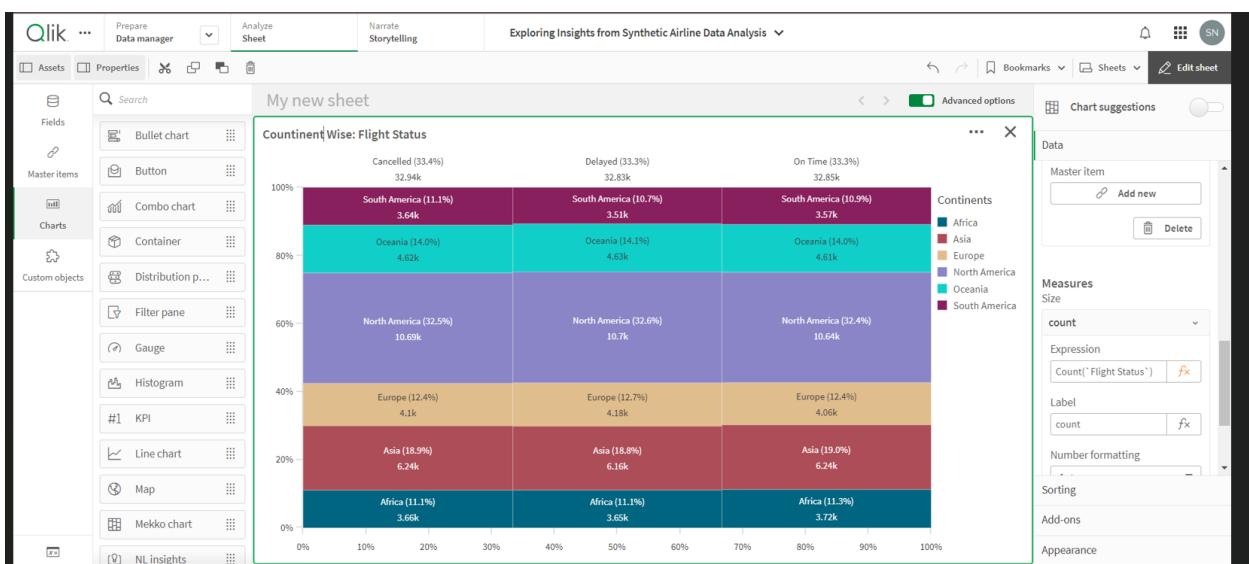
1 =Count({<`Flight Status`='Delayed'}>`Flight Status`)
2
  
```

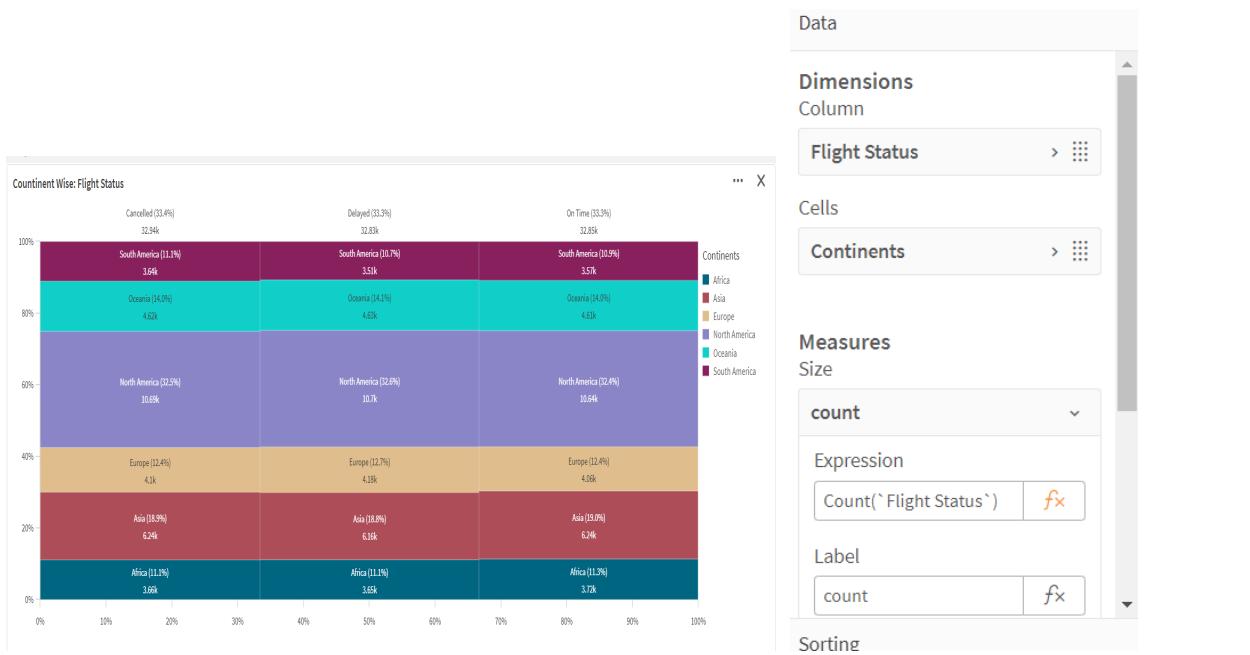
### Activity 1.7: Age group of passengers as per gender wise



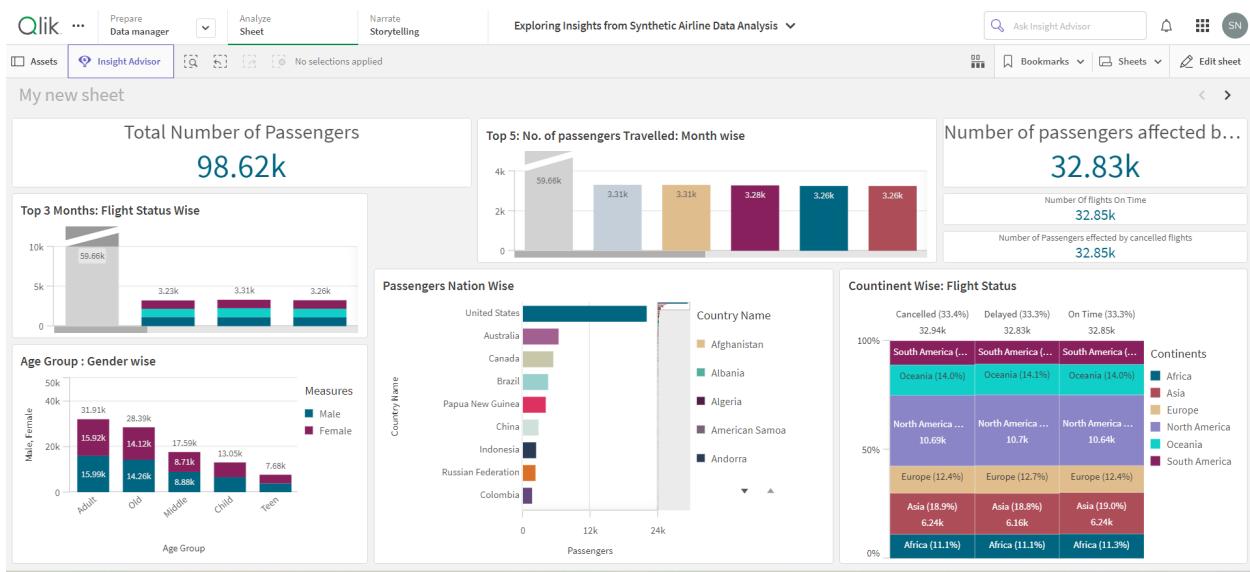


## Activity 1.8: Continent wise flight status





## Final sheet result:

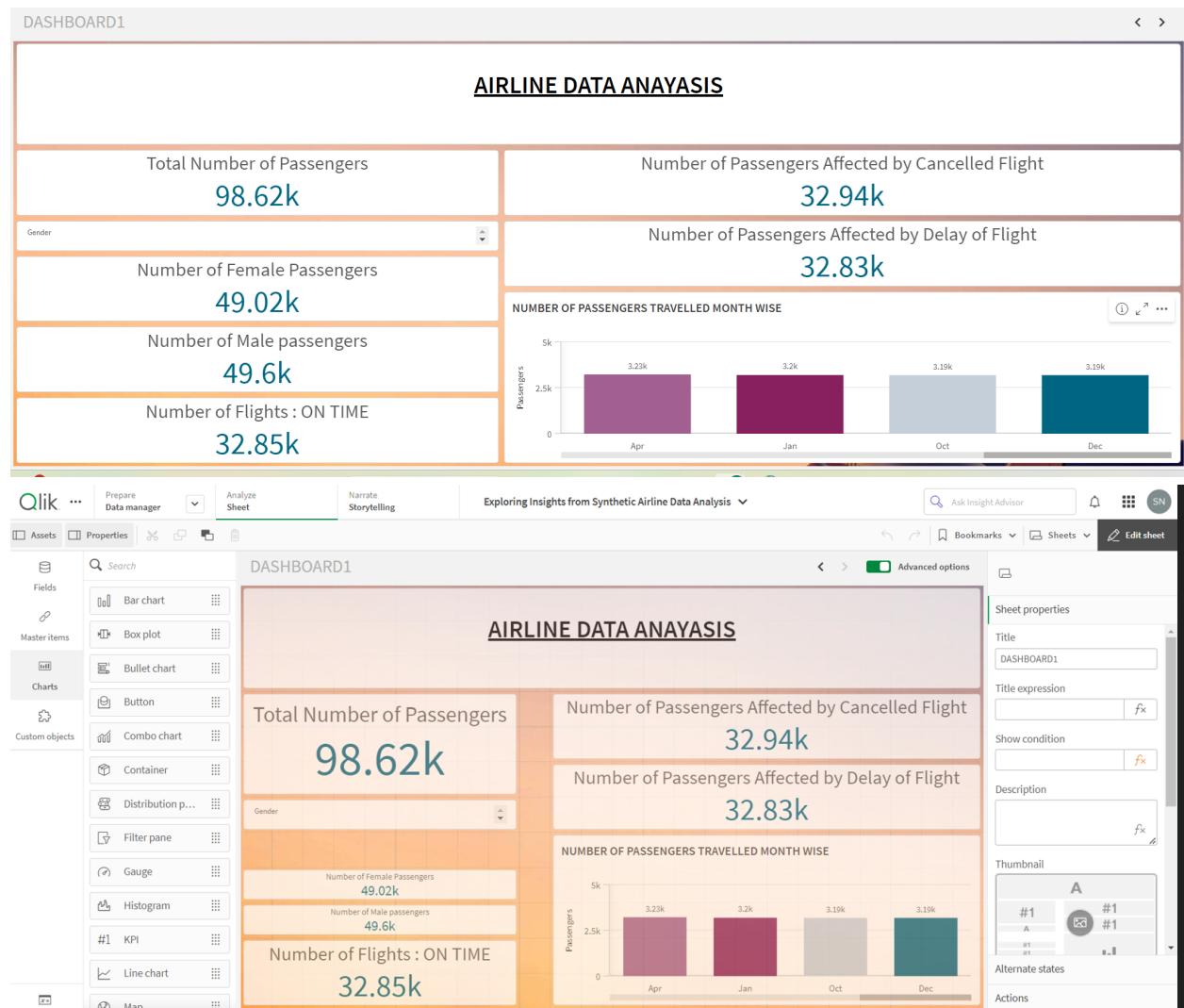


## 5. DASHBOARD

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.

### DASHBOARD 1:



**DASHBOARD1**

**Properties**

**Data**

- First KPI (Measures)
  - Count Gender
- Second KPI (Measures)
  - + Add

Drop field to use or for a suggestion

**Visualization**

#1 KPI

Gender

Selected 1 of 2

Female

Male

Remove filter

Total Number of Passengers  
98.62k

Count(Gender)  
49.02k

Count(Gender)  
49.6k

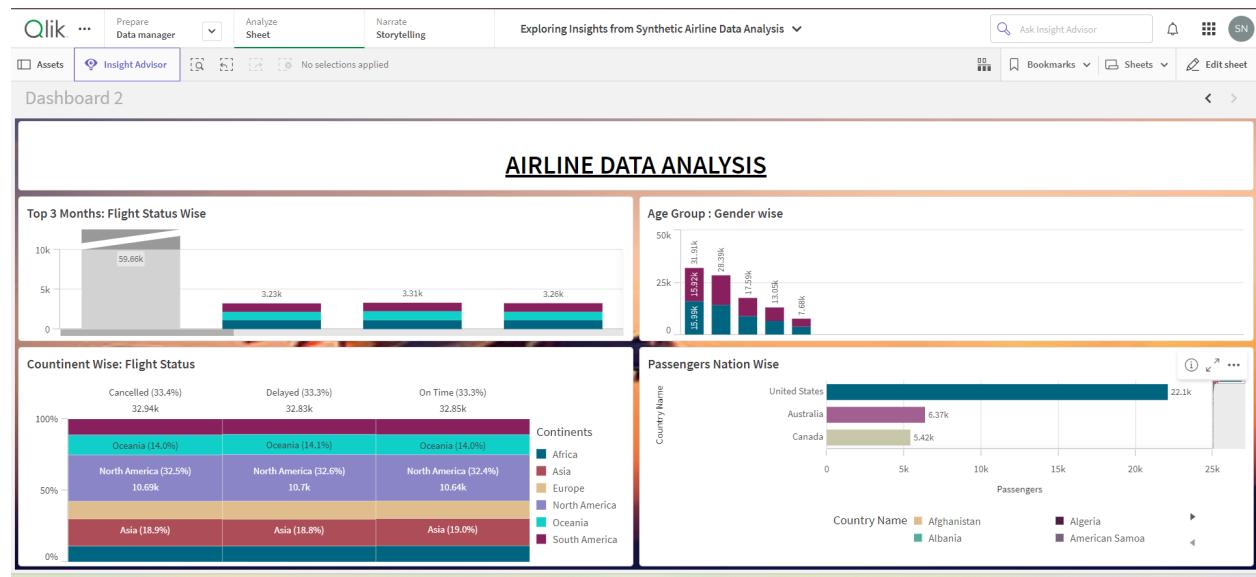
Add measure

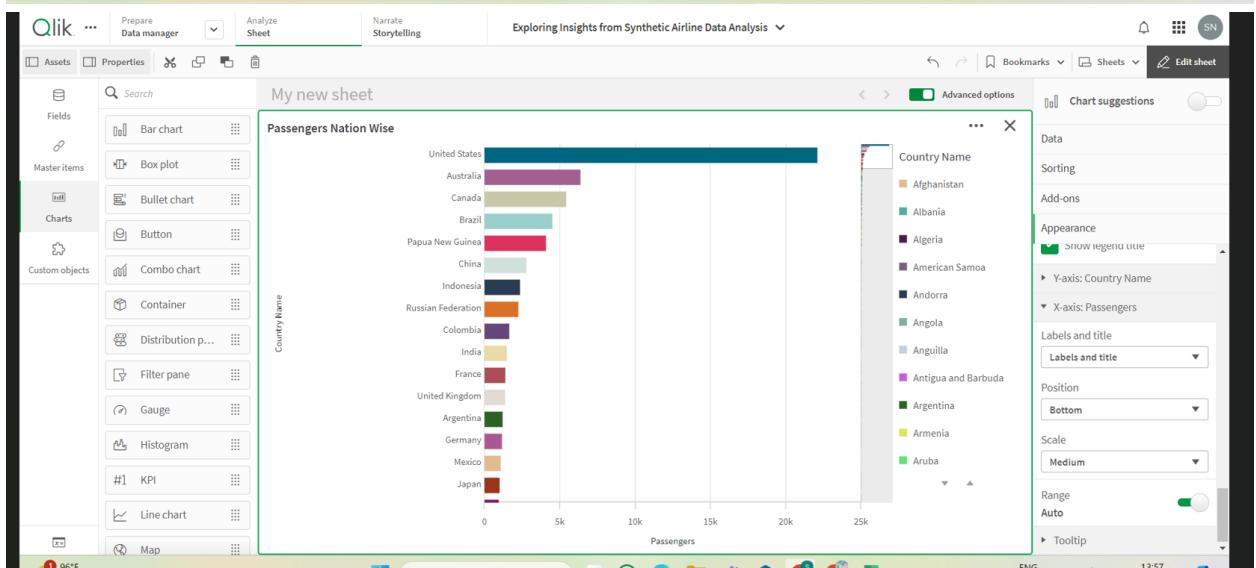
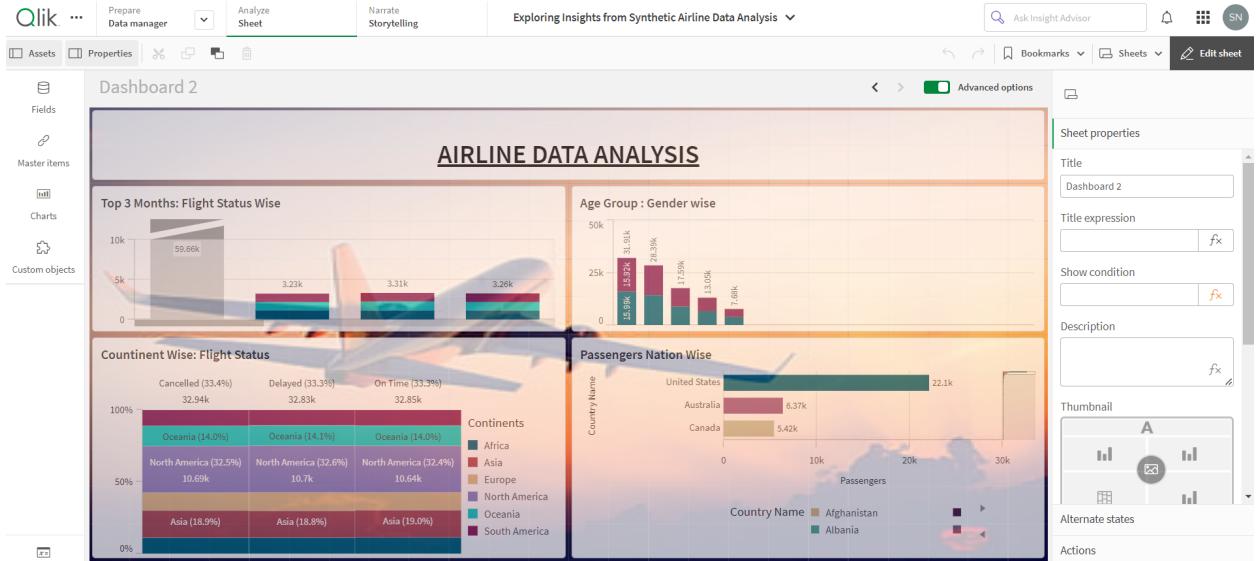
Last N... Gender Age Nation... Airport... Airport... Count

Joist	Female	4	Philippines	Bremen Airport	DE	Gern
Peters	Female	16	Portugal	Watson Lake Airpor	CA	Can
Sargood	Male	59	Serbia	Karluk Lake Seapla	US	Unit
Beswick	Male	83	United States	Holy Cross Airport	US	Unit
Mitrikhin	Male	89	Turkey	Rottnest Island Airp	AU	Aust
Pretsell	Male	64	Nigeria	Vallenar Airport	CL	Chile

Updated - v9 Expand table Rows per page: 100 1 - 100 of 98619

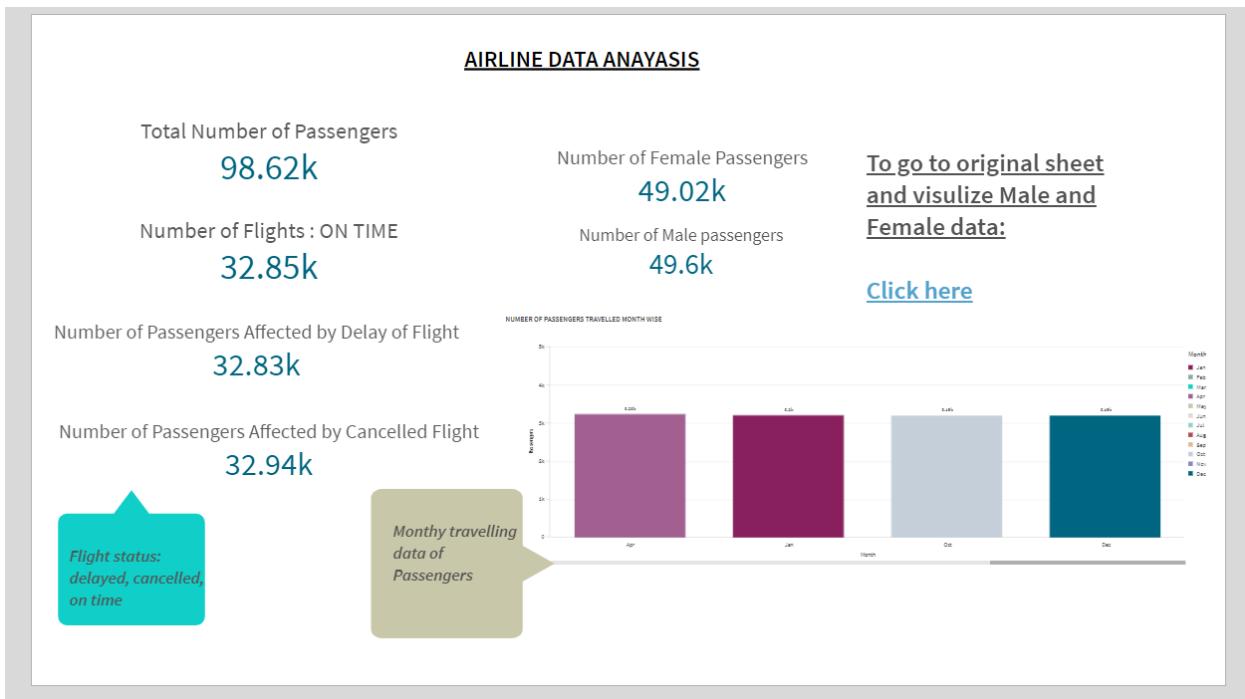
## DASHBOARD 2:

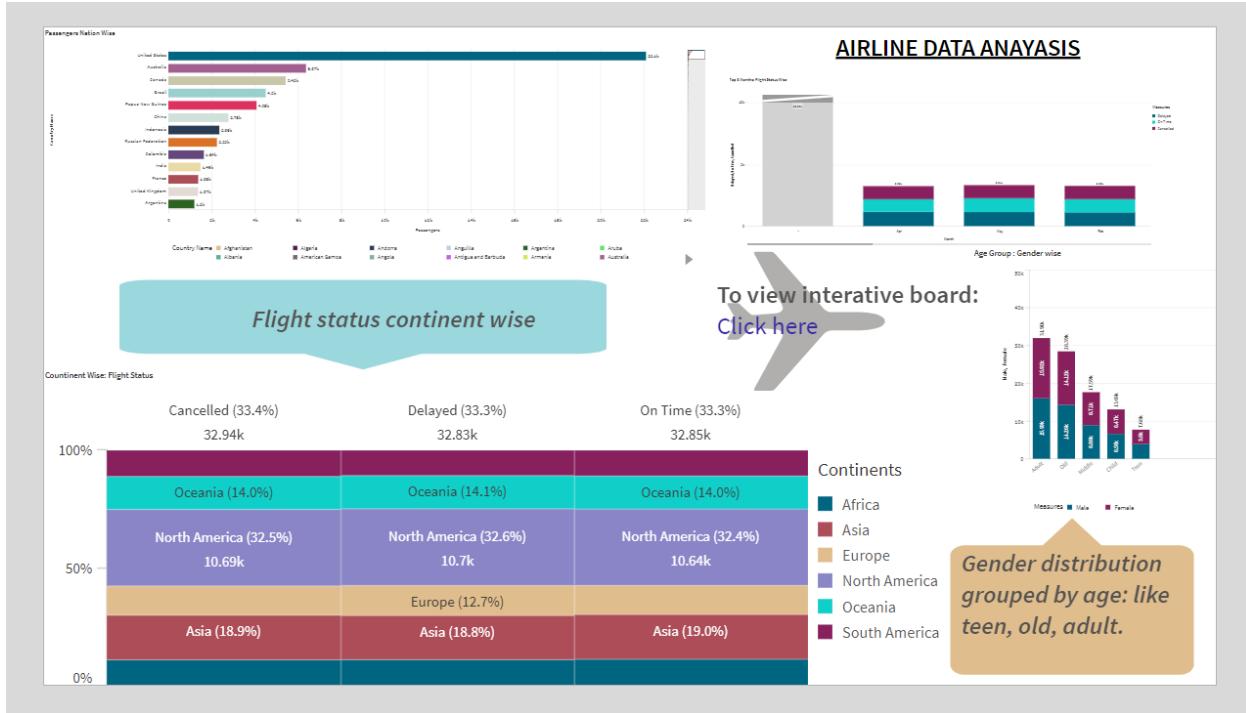




## 6. STORY

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.





## 7. PERFORMANCE TESTING

### Amount of data loaded:

"Amount of Data Loaded" refers to the quantity or volume of data that has been imported, retrieved, or loaded into a system, software application, database, or any other data storage or processing environment. It's a measure of how much data has been successfully processed and made available for analysis, manipulation, or use within the system.

Airline Dataset Updated - v2
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
Flight Status
Airline Dataset Updated - v2.Nationality_GeoInfo
Airline Dataset Updated - v2.Airport Country Code_GeoInfo
Airline Dataset Updated - v2.Country Name_GeoInfo

## Data preprocessing:

```
[Airline_Dataset]:
Load *;

// Remove rows with '0' and '--' from ArrivalAirport column
[Airline_Dataset]:
NOCONCATENATE LOAD *,
if(Age >= 0 AND Age <= 1, 'Baby',
if(Age >= 1 AND Age <= 3, 'Toddler',
if(Age >= 4 AND Age <= 9, 'Child',
if(Age >= 10 AND Age <= 12, 'Tween',
if(Age >= 13 AND Age <= 19, 'Teen',
if(Age >= 20 AND Age <= 24, 'Young Adult',
if(Age >= 25 AND Age <= 39, 'Adult',
if(Age >= 40 AND Age <= 54, 'Middle',
if(Age >= 55 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] = '--');
```

## Utilization of filters:

**DASHBOARD1**

## AIRLINE DATA ANALYSIS

**Properties**

**Data**

- First KPI (Measures)
  - Number of Female Passengers > ...
- Second KPI (Measures)
  - + Add

Drop field to use or for a suggestion

**Visualization**

#1 KPI

Gender

Values

Selected 1 of 2

Search

Female

Male

Remove filter

**Presentation**

T Labels

Tooltip

Total Number of Passengers: 98.62k

Number of Passengers Affected by Cancelled Flight: 32.94k

Number of Passengers Affected by Delay of Flight: 32.83k

**NUMBER OF PASSENGERS TRAVELED MONTH WISE**

Last Name	Gender	Age	Nation...	Airport...	Airport...	Cou...
Joist	Female	4	Philippines	Bremen Airport	DE	Gerr...
Peters	Female	16	Portugal	Watson Lake Airpor	CA	Can...
Sargood	Male	59	Serbia	Karuk Lake Seapla	US	Unit...
Beswick	Male	83	United States	Holy Cross Airport	US	Unit...
Mitrikhin	Male	89	Turkey	Rottnest Island Airp	AU	Aust...
Pretsell	Male	64	Nigeria	Vallenar Airport	CL	Chile...

## Number of Visualizations/ Graphs:

- Total Number of Passengers

Exploring Insights from Synthetic Airline Data Analysis

My new sheet

Click to add title

Total Number of Passengers

98.62k

Advanced options

Chart suggestions

Data

Measures

First KPI

Total Number of Passengers >

Second KPI

Add

Add-ons

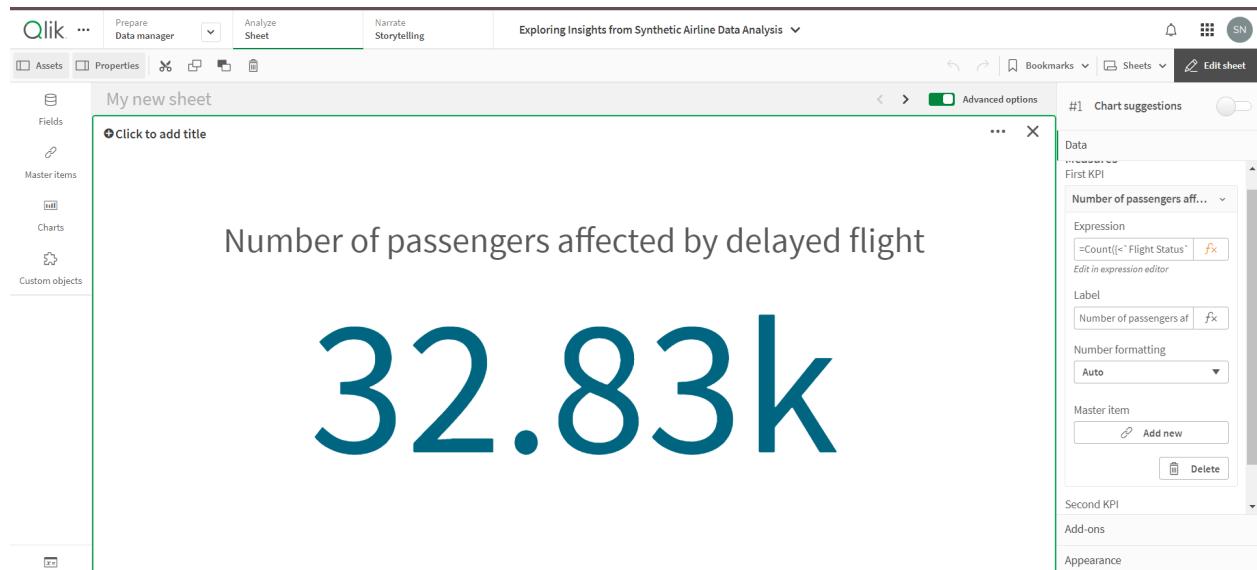
Appearance

- Number of Passengers effected by Cancelled Flights

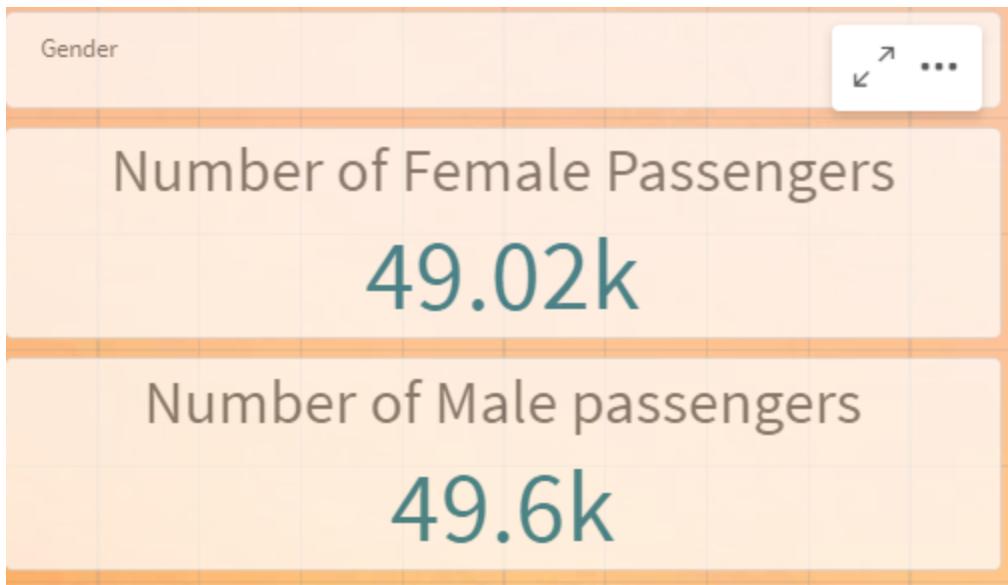
Number of Passengers effected by cancelled flights  
32.85k

⊕ (0,8) 24 x 4

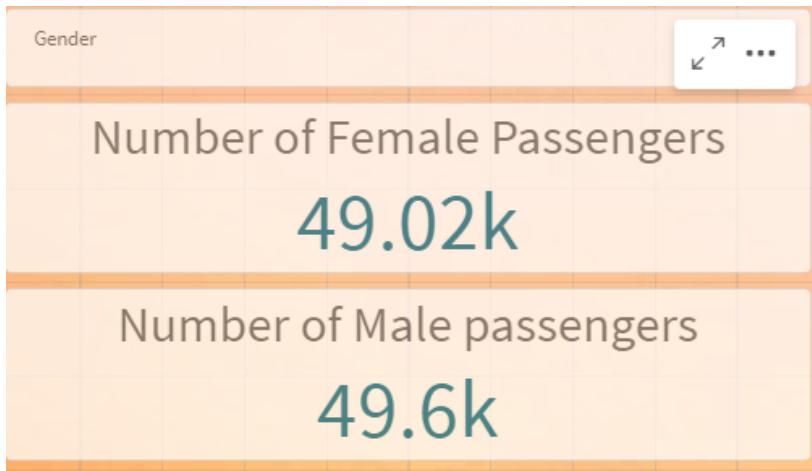
- Number of Passengers effected by Delay of Flights



- Number of Male Passengers



- Number of Female Passengers



- Number of Flights-On-Time

**Click to add title**

umber of passengers affected by delay...

**32.83k**

Number Of flights On Time  
**32.85k**

umber of Passengers effected by cancelled flig...  
**32.85k**

**Wise: Flight Status**

%)	Delayed (33.3%)	On Time (33.3%)
	32.83k	32.85k
...	South America ...	South America ...
6)	Oceania (14.1%)	Oceania (14.0%)
	10.7k	10.64k

**Data**

**Measures**

First KPI

**Number Of flights On Time**

Expression  
=Count({<`Flight Status`}) **f**

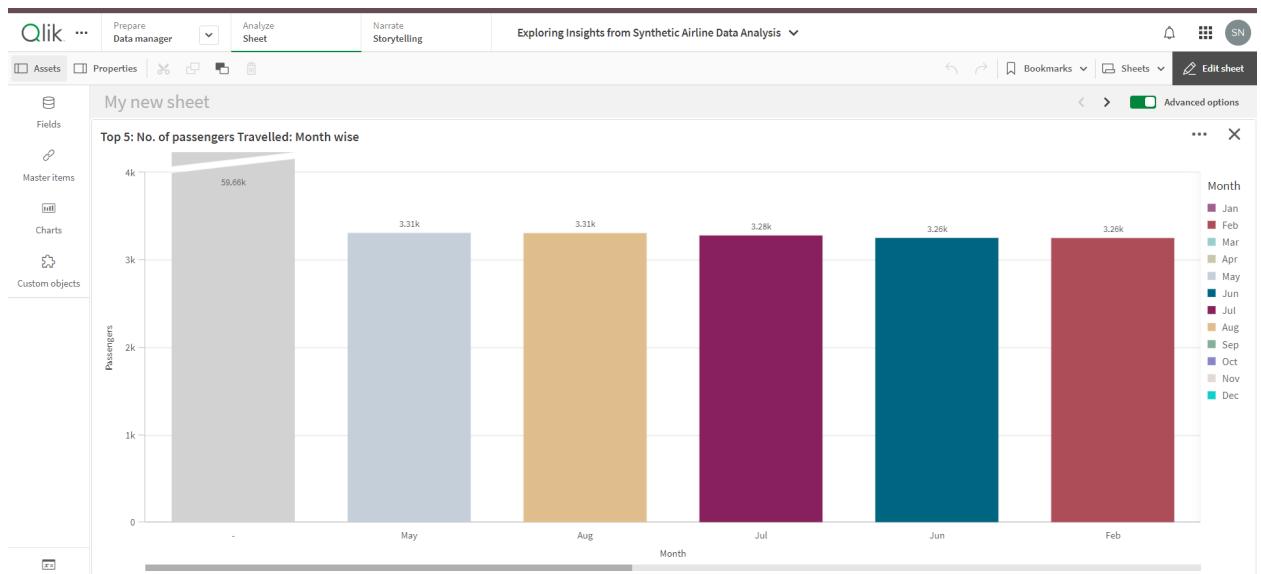
Edit in expression editor

Label  
Number Of flights On Tim **f**

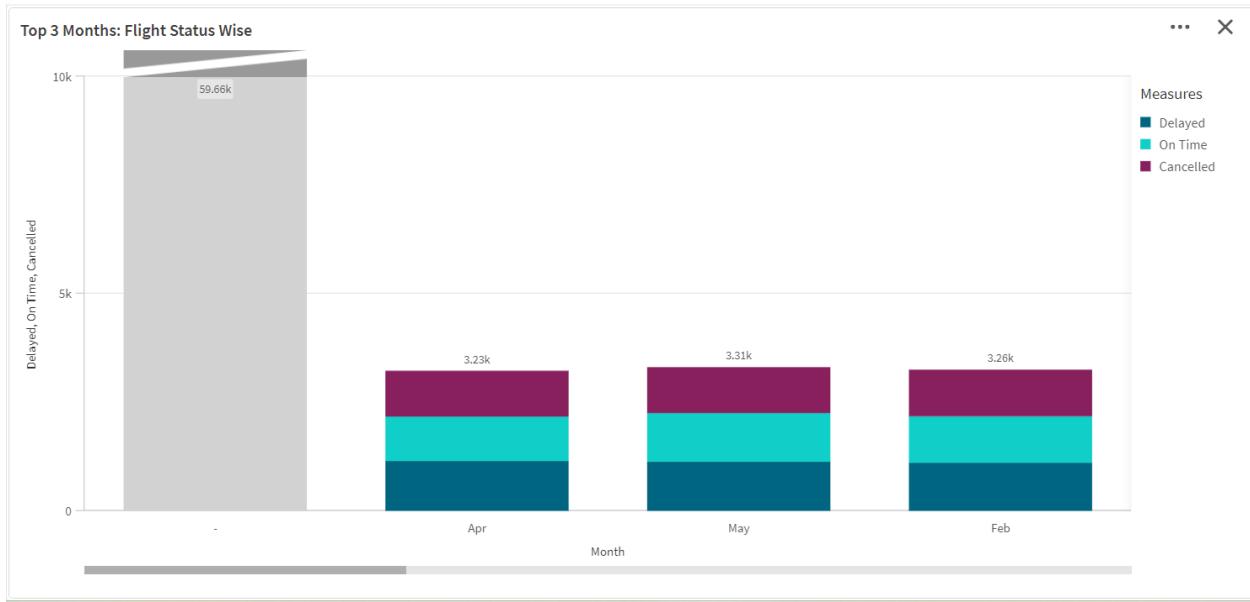
Number formatting  
Auto

Master item

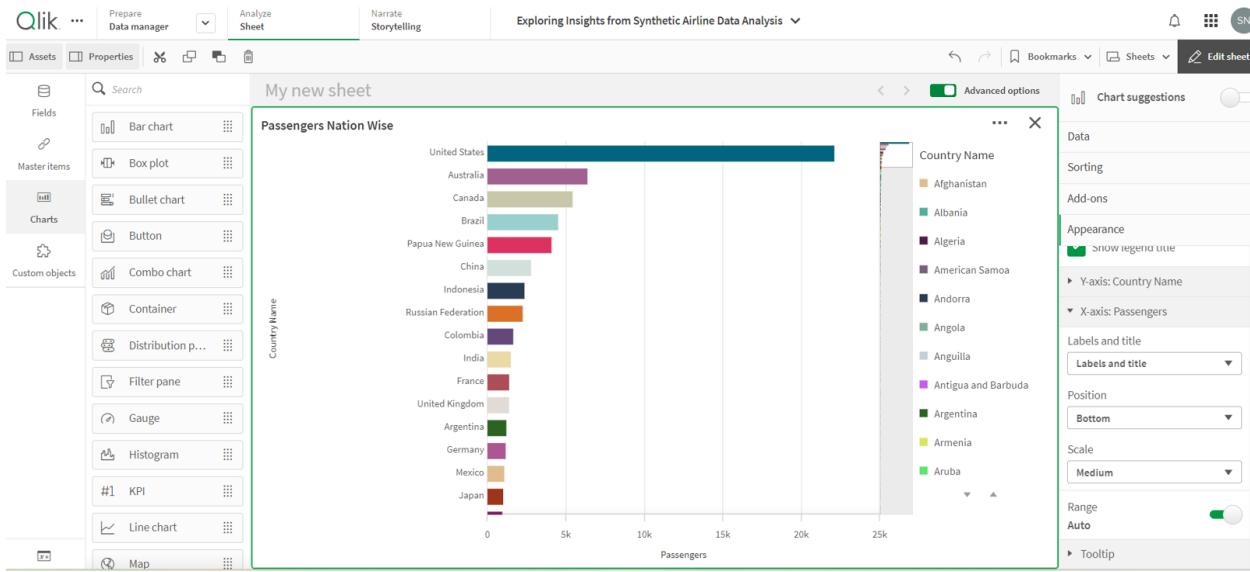
- Top 5 Months where Passengers traveled the most



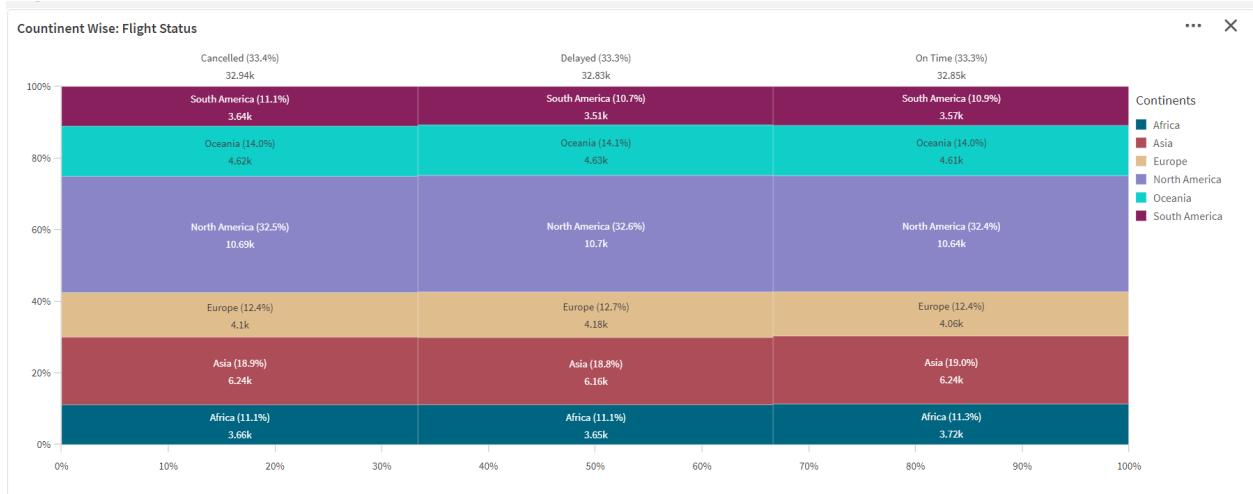
- Top 3 Months - Flight status Wise (Delayed/Cancelled/On-Time)



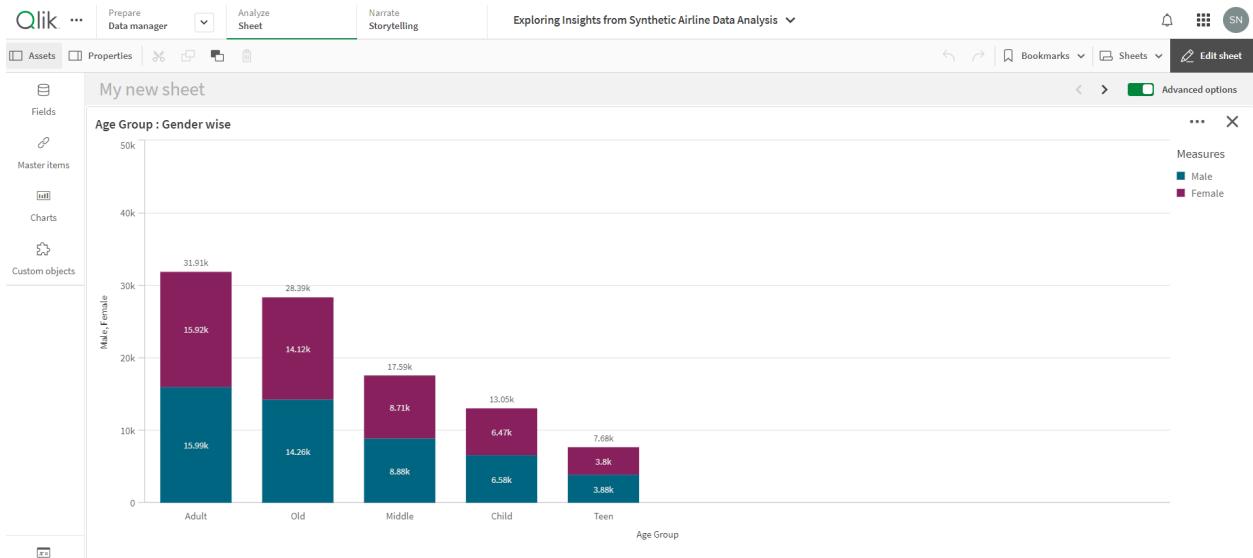
- Number of Passengers Nationality Wise



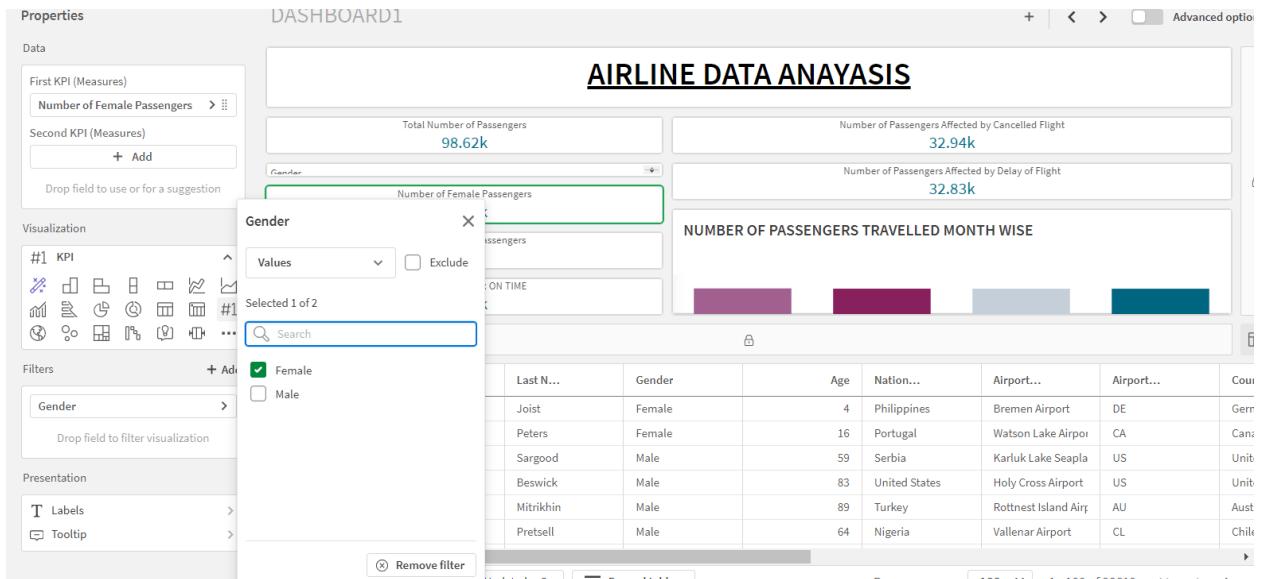
- Continent wise – Flight Status



### ● Age Group – Gender Wise



- A filter pane consisting of Gender(M/F)



## 8. APPENDIX

GITHUB REPO: <https://github.com/Simran-Namdev/Exploring-Insights-from-Synthetic-Airline-Data-Analysis-with-Qlik>