Airline Data Management and Analysis

Abhishek Ambasta | Data Science

Task 1: Data Preparation and Cleaning (10 Marks)

Question:

● Extract and transform data in Power Query.

● Clean data: remove duplicates, handle missing values, and format columns.

● Deliverables: Screenshot of Power Query Editor showing cleaned data.

Answer: Followed the following steps:

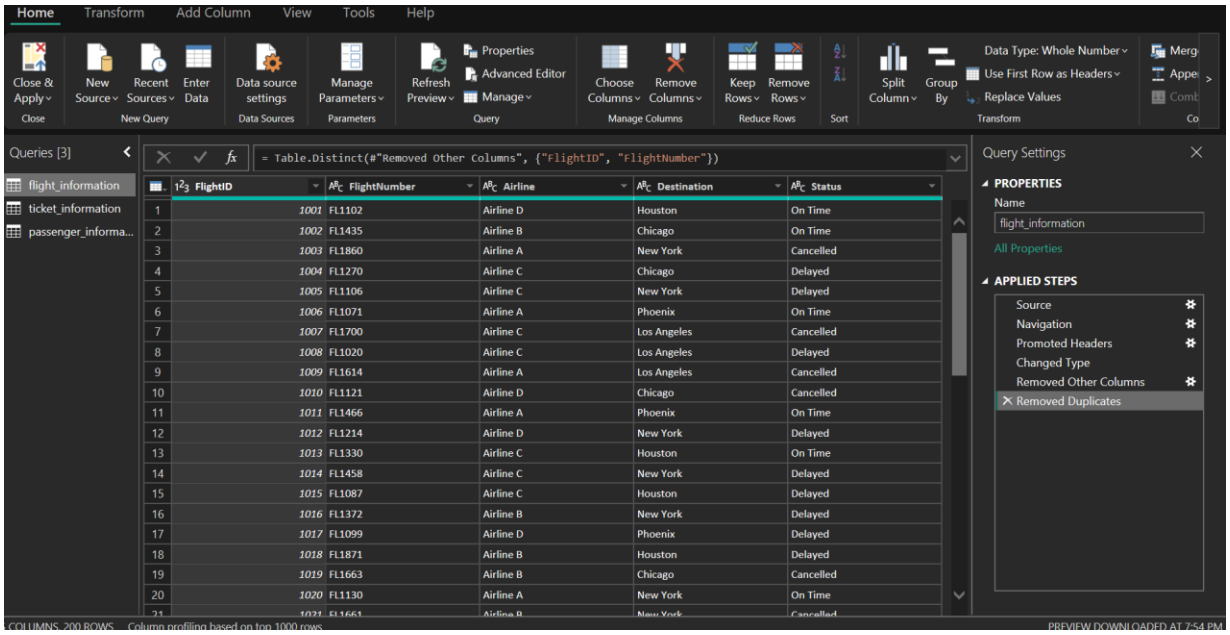
1. First, we will load the data in the power query editor → open Power BI desktop → Under Home → Transform Data → Opens Power Query editor.
2. Go to home → New Source → Excel Workbook → Load the 3 data set →Ticket Information, Passenger Information and Flight Information.
3. There are many columns which shows null values We will remove them first → Select the columns from each data set that are necessary like:

For Flight information data set:

A. Flight information: Includes → FlightID, FlightNumber, Airline, Destination, and Status.

   Select these by pressing Ctrl + click on each column with holding the ctrl button → Home → Remove columns → Select remove other columns. (This will keep the selected columns and remove other that shows the null values).

B. Removed the Duplicates that are there in the FlightID Column → select the column → right click → Remove duplicates.

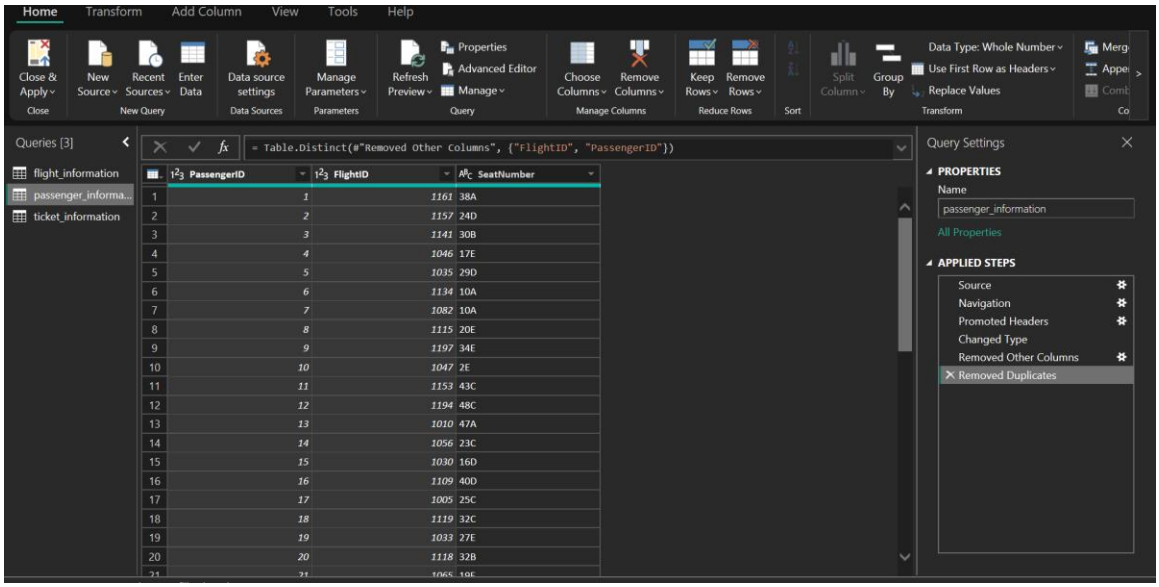


	FlightID	FlightNumber	Airline	Destination	Status
1	1001	FL1102	Airline D	Houston	On Time
2	1002	FL1435	Airline B	Chicago	On Time
3	1003	FL1860	Airline A	New York	Cancelled
4	1004	FL1270	Airline C	Chicago	Delayed
5	1005	FL1106	Airline C	New York	Delayed
6	1006	FL1071	Airline A	Phoenix	On Time
7	1007	FL1700	Airline C	Los Angeles	Cancelled
8	1008	FL1020	Airline C	Los Angeles	Delayed
9	1009	FL1614	Airline A	Los Angeles	Cancelled
10	1010	FL1121	Airline D	Chicago	Cancelled
11	1011	FL1466	Airline A	Phoenix	On Time
12	1012	FL1214	Airline D	New York	Delayed
13	1013	FL1330	Airline C	Houston	On Time
14	1014	FL1458	Airline C	New York	Delayed
15	1015	FL1087	Airline C	Houston	Delayed
16	1016	FL1372	Airline B	New York	Delayed
17	1017	FL1099	Airline D	Phoenix	Delayed
18	1018	FL1871	Airline B	Houston	Delayed
19	1019	FL1663	Airline B	Chicago	Cancelled
20	1020	FL1130	Airline A	New York	On Time

The above image shows the clean data of Flight Information table.

For Passenger Information table, taken the following steps to clean the data set:

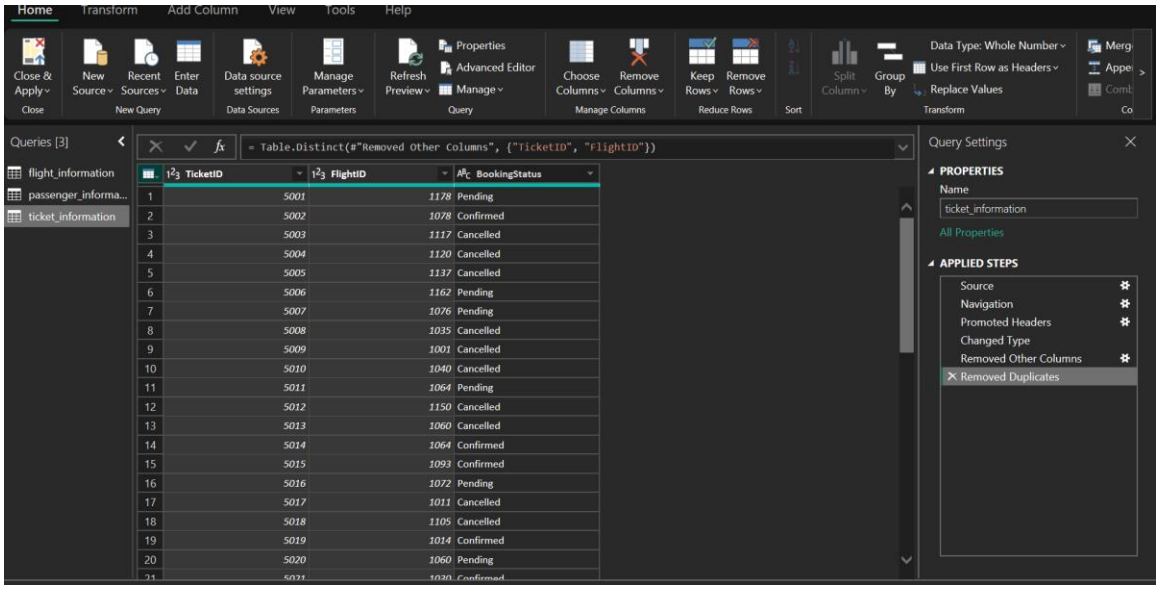
1. Passenger Information: Includes PassengerID, FlightID, and SeatNumber.
2. Select these by pressing Ctrl + click on each column with holding the ctrl button → Home → Remove columns → Select remove other columns. (This will keep the selected columns and remove other that shows the null values).
3. Remove the Duplicates from PassengerID and FlightID → select both and then right click → remove duplicates (We can also do this by clicking on Home → Remove Rows → Remove Duplicates).



The above image shows the clean data of Passenger Information table.

For Ticket Information table, taken the following steps to clean the data set:

1. Ticket Information: Includes TicketID, FlightID, and BookingStatus.
2. Select these by pressing Ctrl + click on each column with holding the ctrl button → Home → Remove columns → Select remove other columns. (This will keep the selected columns and remove other that shows the null values).
3. Remove the Duplicates from TicketID and FlightID → select both and then right click → remove duplicates (We can also do this by clicking on Home → Remove Rows → Remove Duplicates).



	TicketID	FlightID	BookingStatus
1	5001	1178	Pending
2	5002	1078	Confirmed
3	5003	1117	Cancelled
4	5004	1120	Cancelled
5	5005	1137	Cancelled
6	5006	1162	Pending
7	5007	1076	Pending
8	5008	1035	Cancelled
9	5009	1001	Cancelled
10	5010	1040	Cancelled
11	5011	1064	Pending
12	5012	1150	Cancelled
13	5013	1060	Cancelled
14	5014	1064	Confirmed
15	5015	1093	Confirmed
16	5016	1072	Pending
17	5017	1011	Cancelled
18	5018	1105	Cancelled
19	5019	1014	Confirmed
20	5020	1060	Pending

The above image shows the clean data of Ticket Information table.

Then Load all the data by clicking on → close and apply.

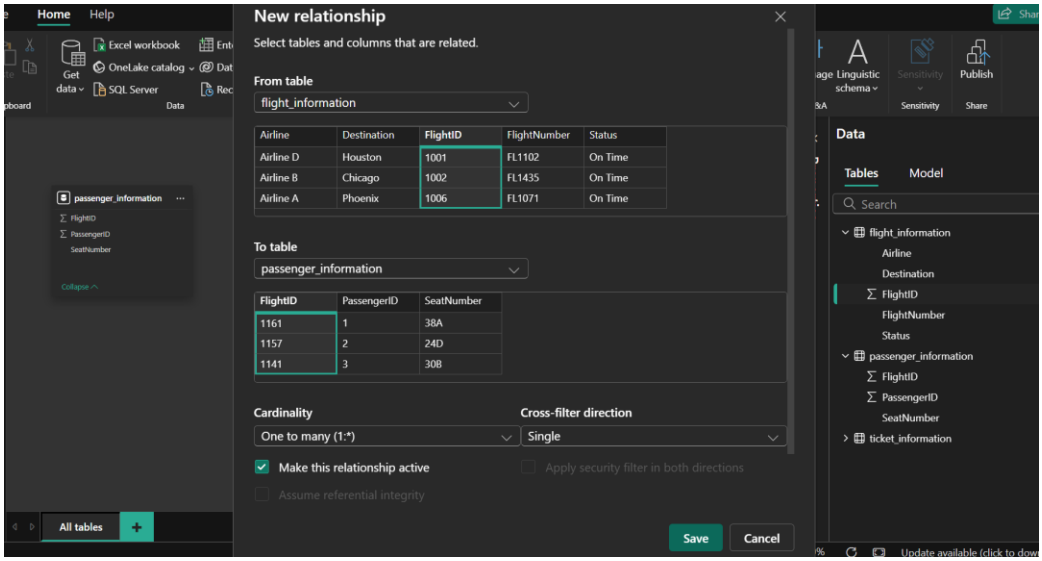
TASK 2: DATA MODELING (10 MARKS)

- CREATE RELATIONSHIPS BETWEEN DATASETS (FLIGHT ID AS THE KEY).
- UNDERSTAND CARDINALITY AND CONFIGURE THE MODEL APPROPRIATELY.

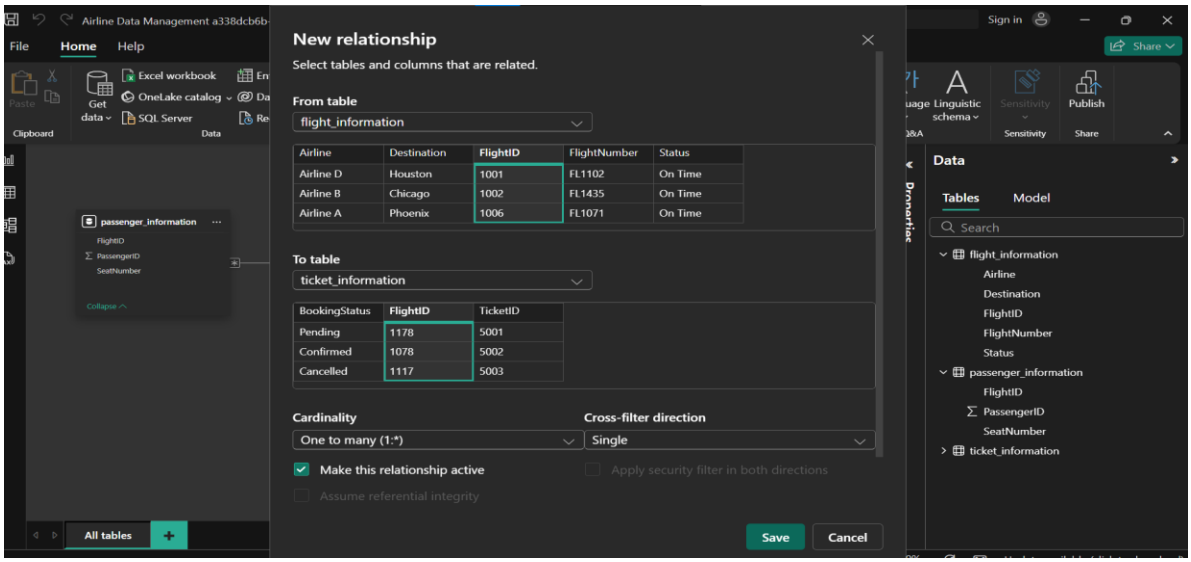
Answer: Followed the following steps:

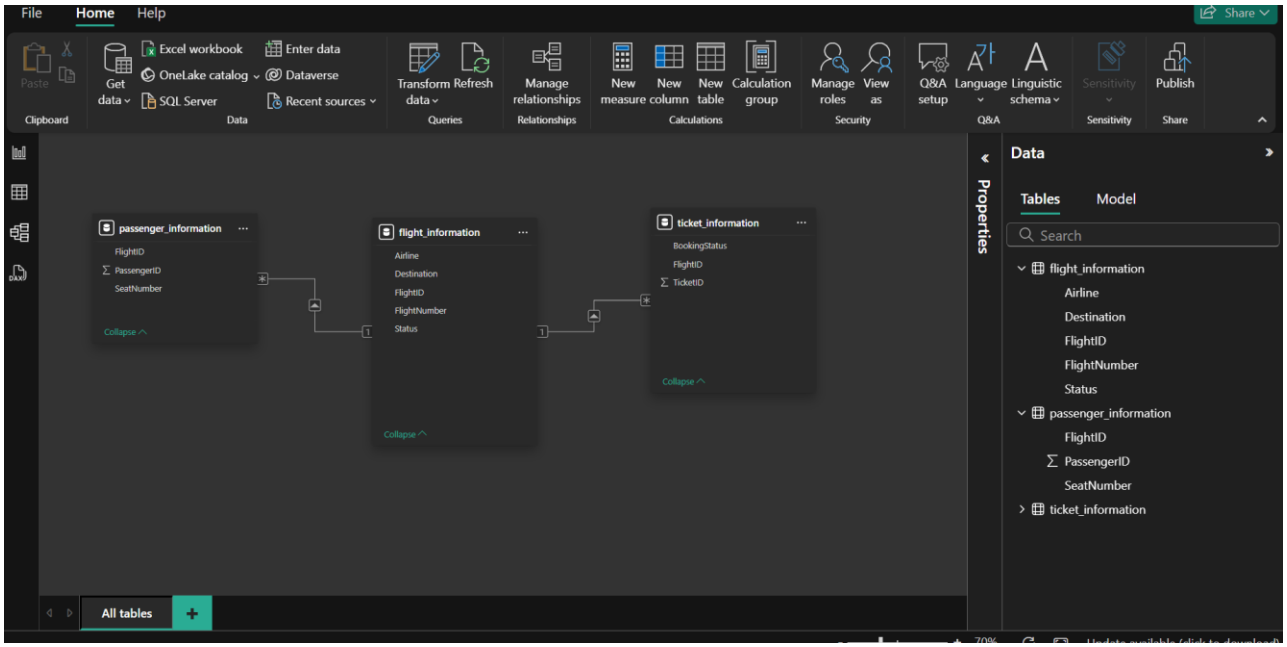
1. On the left-hand side go to the Mode View.

2. Keep Flight_information at the center (fact/dimension hybrid), acting as the bridge for both:
   (a) Passenger_information (1 → many): By dragging the FlightID column from Flight_Information table to the Passenger Information table (FlightID column), this will create one to many relationships.



   (b) Table_information (1 → many): By dragging the FlightID column from Flight_Information table to the Table_Information table (FlightID column), this will create one to many relationships.





The above image shows that:

1. **One-to-many** relationships are set with correct cardinality and direction.
2. Arrows confirm **single-directional filtering**, which is ideal for performance and logic.

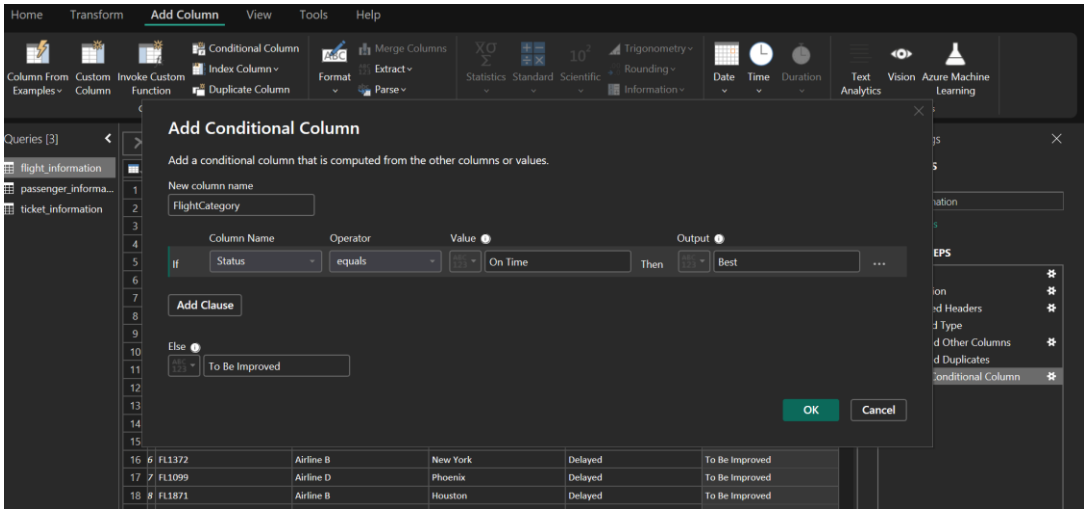
TASK 3: ENHANCED DATA INSIGHTS (10 MARKS)

Question 1: Add a conditional column to classify flights as "Best" or "To Be Improved" based on status.

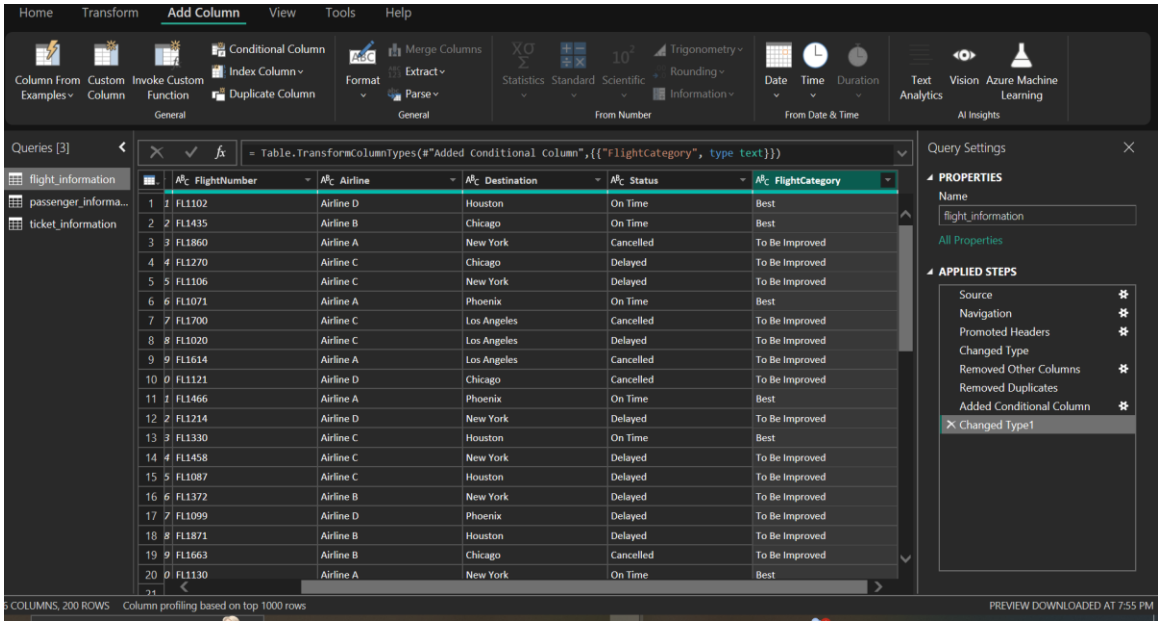
Answer: Followed the following steps:

1. Open power query editor by clicking on home → Transform data.

2. From Flight information table → Click on the status column → Add column → Conditional Column → Name it as Flight Category →If column name → Status Equals to On Time then make Output as Best, Else To be Improved.



After this change the column type as text.

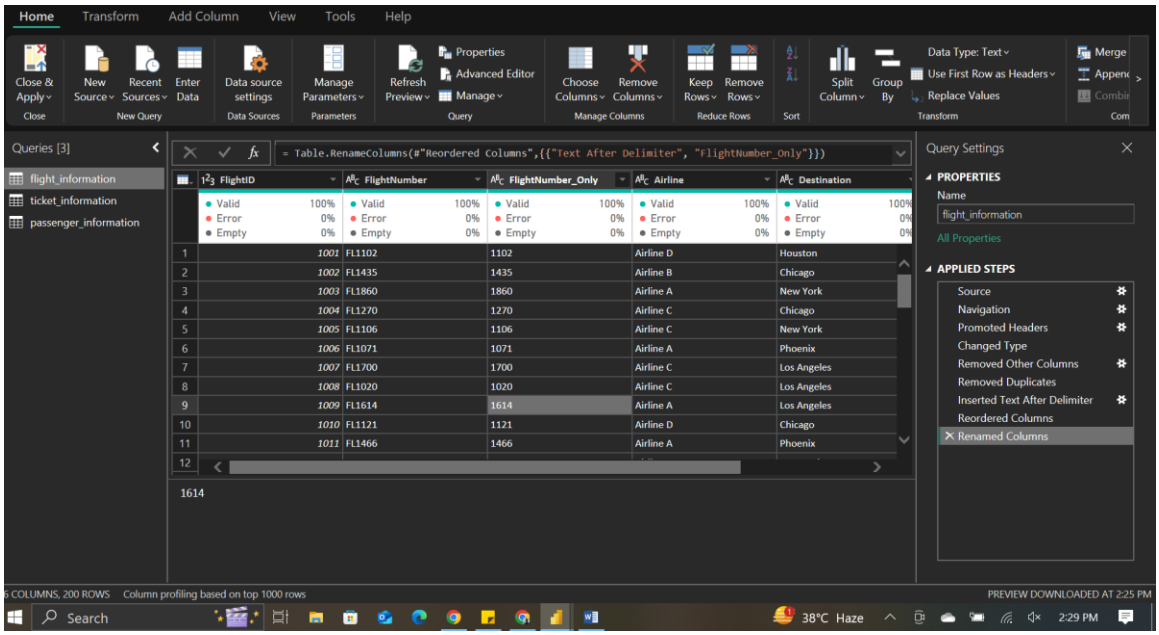


This shows On Time flights is the “Best” and rest are in “To be Improved” Category.

Question 2: Use "Column from Examples" to extract the flight number from FlightNumber.

Answer: Followed the following steps:

1. Select the column FlightNumber → click on Add Column → Column from Example → From Selection.
2. Type only the flight Number as 1102, and this will extract all the other rows flight numbers → Rename the column as FlightNumber_Only.



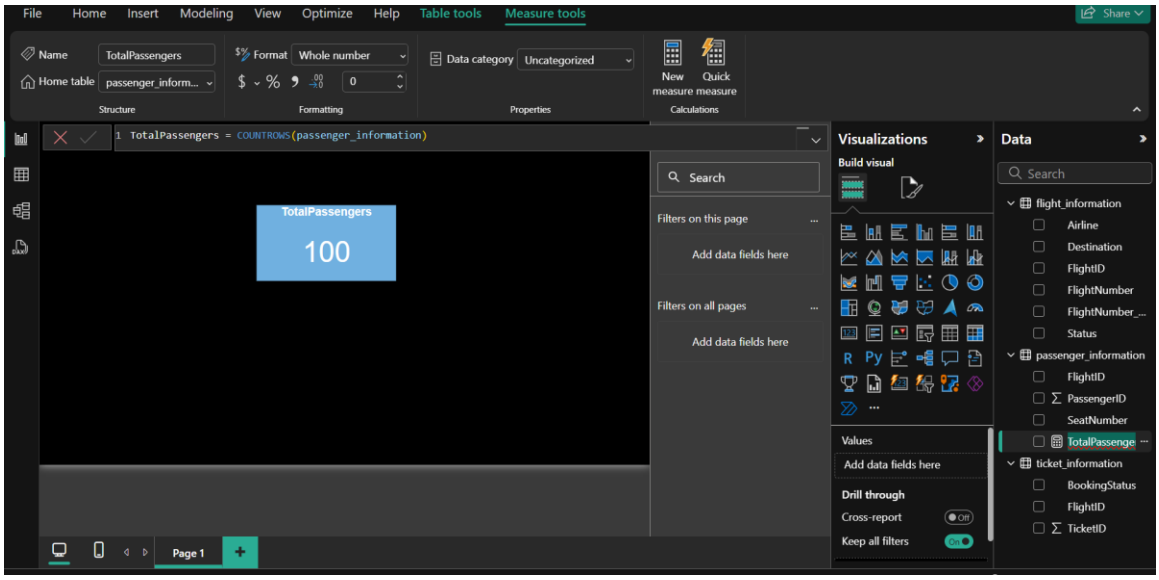
In the above image, I have chosen Column from Example.

TASK 4: CALCULATIONS USING DAX (10 MARKS)

Question 1:
Calculate: Total passengers for a specific flight.

Answer: Followed the following steps:

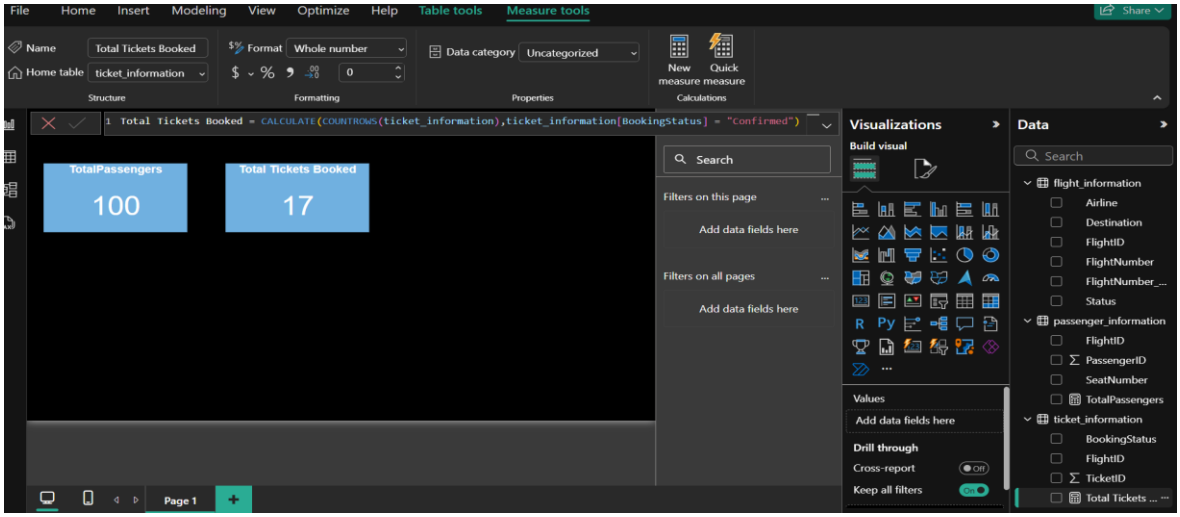
1. Under the Passenger_information table on the Data Pane → right click → add new measure → Write the DAX formula as:
   TotalPassengers = COUNTROWS(passenger_information)
2. Press Enter.



The above image shows how we have calculated total passengers.

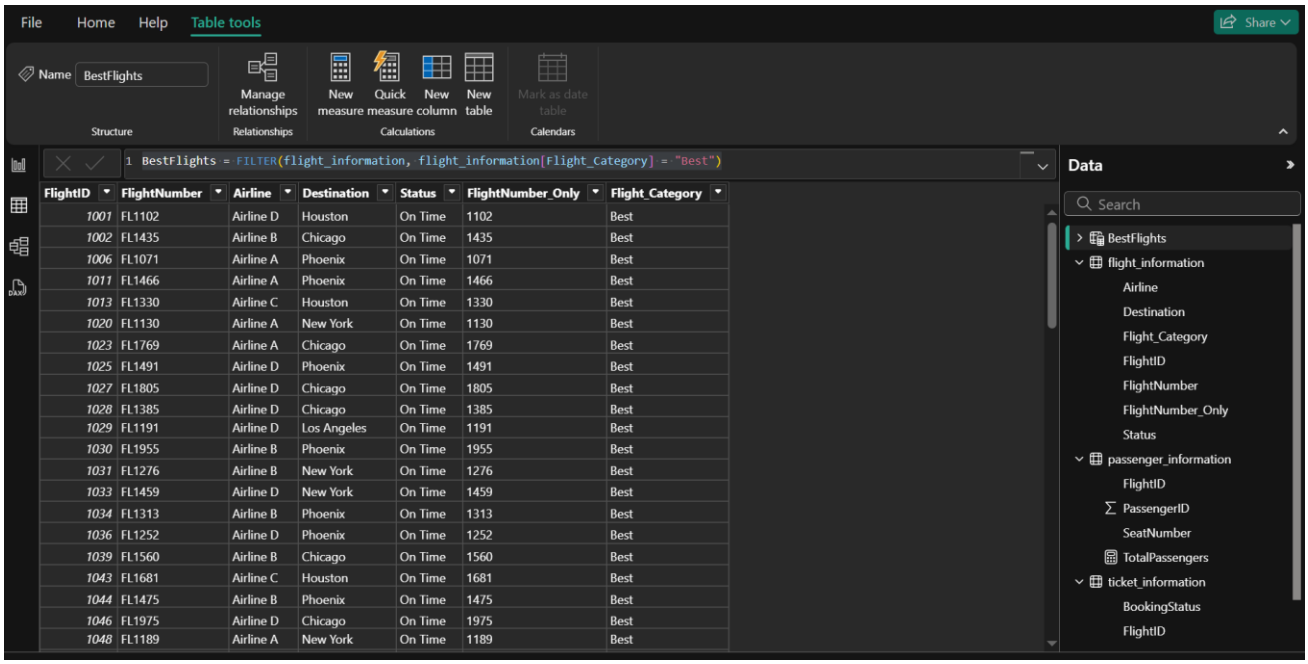
Question 2: Using DAX Calculate Total tickets booked?
Answer: Followed the following steps:

1. Under the Ticket_information table on the Data Pane → right click → add new measure → Write the DAX formula as:
   Total Tickets Booked = CALCULATE(COUNTROWS(ticket_information),ticket_information[BookingStatus] = "Confirmed")
2. In this we are only counting the numbers of Tickets that are booked and status is Confirmed.



Question 3: Using DAX Calculate Filtered table showing "Best" flights only.
Answer: Followed the following steps:

1. Go to the table View → Under table tool click on New Table.
2. As we already have the column which shows the Flight_Category as “Best” and “To be Improved” (From Task 3.1 the conditional column).
3. Write a DAX Formula:
   BestFlights = FILTER
   (flight_information,flight_information[Flight_Category] = "Best").



FlightID	FlightNumber	Airline	Destination	Status	FlightNumber_Only	Flight_Category
1001	FL1102	Airline D	Houston	On Time	1102	Best
1002	FL1435	Airline B	Chicago	On Time	1435	Best
1006	FL1071	Airline A	Phoenix	On Time	1071	Best
1011	FL1466	Airline A	Phoenix	On Time	1466	Best
1013	FL1330	Airline C	Houston	On Time	1330	Best
1020	FL1130	Airline A	New York	On Time	1130	Best
1023	FL1769	Airline A	Chicago	On Time	1769	Best
1025	FL1491	Airline D	Phoenix	On Time	1491	Best
1027	FL1805	Airline D	Chicago	On Time	1805	Best
1028	FL1385	Airline D	Chicago	On Time	1385	Best
1029	FL1191	Airline D	Los Angeles	On Time	1191	Best
1030	FL1955	Airline B	Phoenix	On Time	1955	Best
1031	FL1276	Airline B	New York	On Time	1276	Best
1033	FL1459	Airline D	New York	On Time	1459	Best
1034	FL1313	Airline B	Phoenix	On Time	1313	Best
1036	FL1252	Airline D	Phoenix	On Time	1252	Best
1039	FL1560	Airline B	Chicago	On Time	1560	Best
1043	FL1681	Airline C	Houston	On Time	1681	Best
1044	FL1475	Airline B	Phoenix	On Time	1475	Best
1046	FL1975	Airline D	Chicago	On Time	1975	Best
1048	FL1189	Airline A	New York	On Time	1189	Best

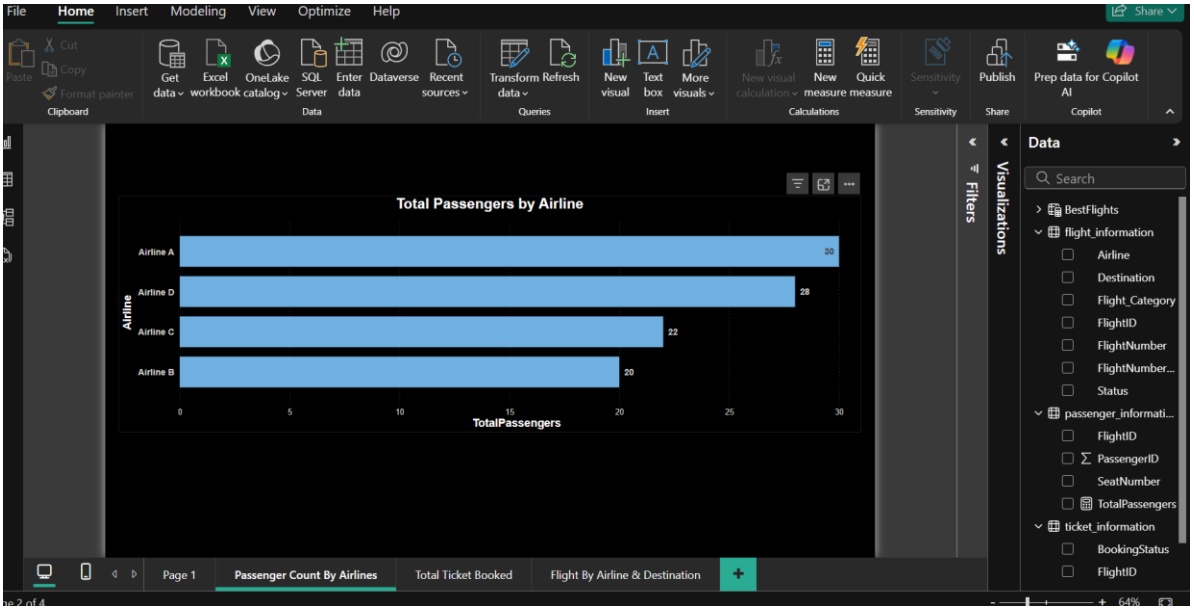
The above image shows that we have created a filtered table that only shows the flight those falls under “Best” Category.

TASK 5: VISUALIZATION AND INTERACTIVE FEATURES (20 MARKS)

Question: Create visuals for:
1. Passenger count by airline
Answer: Followed the following steps:

1. Add a New Page and Rename it as Passenger Count By Airlines.

2. From the Visualization pane use **Clustered Bar Chart**.
3. In Y- Axis Drag → Airline from Flight_Information table.
4. In the X- Axis Drag → TotalPassengers (Measure that earlier created).
5. Click on Format your Visual under the Visualization Pane → Turn on the Data label and format it according to readability.

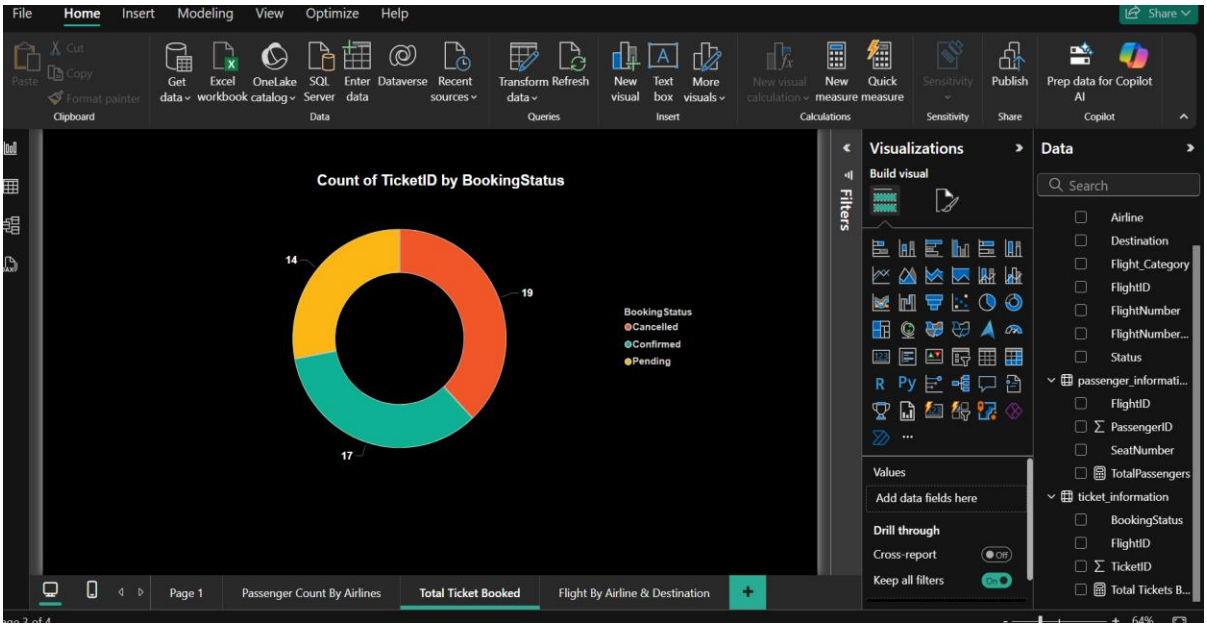


Airline A has the highest number of passengers (30), followed closely by Airline D (28), Airline B carried the fewest passengers (20).

Question 2: Ticket booking statuses.

Answer: Followed the following steps:

1. Add a New Page and Rename it as Total Ticket Booked.
2. Add a **Donut chart** from the Visualization Pane.
3. Under Legends Drag → BookingStatus from Ticket Information table.
4. Under Values Drag → TicketID (Set that as Count) from the same table.
5. From the Format your visual → Turn on the detailed Label and then Under Label Content → set as Data Value.
6. Format the chart according to the readability.

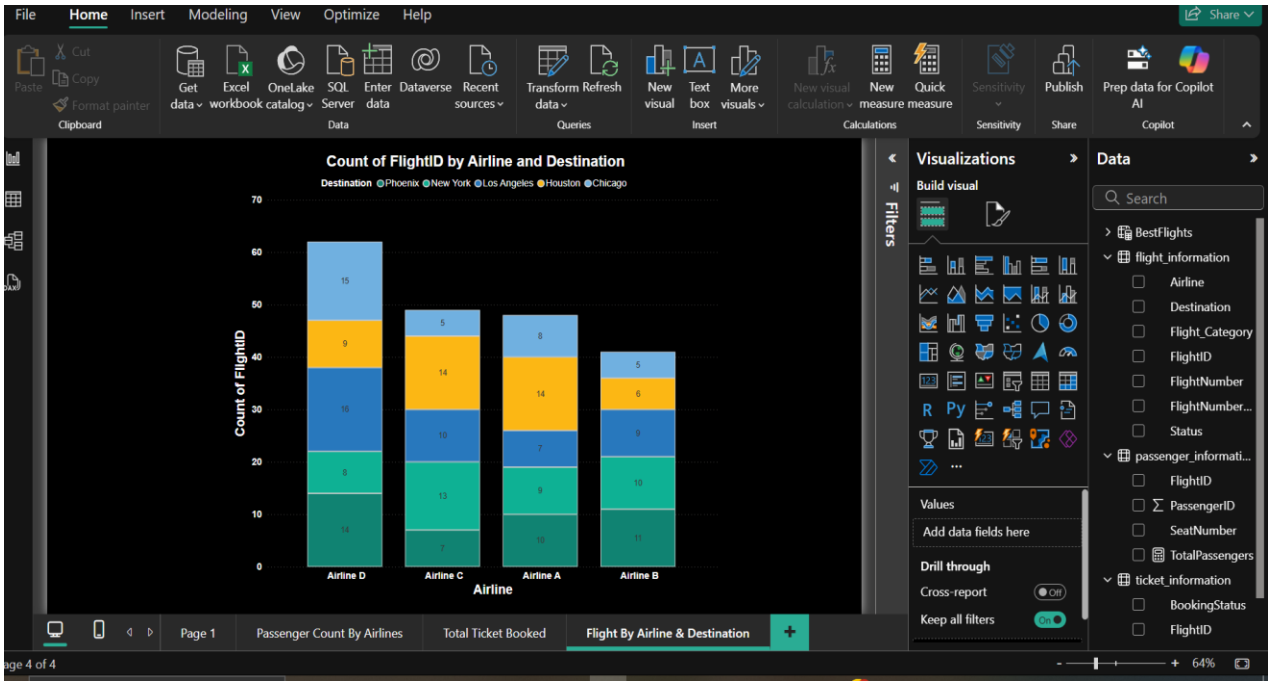


The above image shows the count of Ticket those are confirmed are 17, 19 tickets are cancelled and 14 is under Pending state.

Question 3: Flights by airline and destination.

Answer: Followed the following steps:

1. Add a New Page and Rename it as Flight by Airline & Destination.
2. From the Visualization pane select **Stacked Column Chart**.
3. Under X- Axis → Drag Airline from Flight Information Table.
4. Under Y-Axis → Count of FlightID from the same table.
5. Under Legends → Keep Destination.
6. Go to Format your Visual → Turn On Data Label.
7. Change the font size and the title of X and Y axis as per readability.



The above image shows that Airline D has the highest total flights, particularly to New York and Phoenix, while Airline B has the lowest number of total flights distributed fairly evenly across destinations.

Question 4: Add interactive features for:

➔ Destination and Airline

Answer: Followed the following steps:

For doing this first copy all the visuals from the 3 pages:

1. Passengers count by Airlines.
2. Total Ticket Booked.
3. Flight by Airline & destination.

After this we will rename the page as Dashboard, and then arrange all these visuals as per readability.

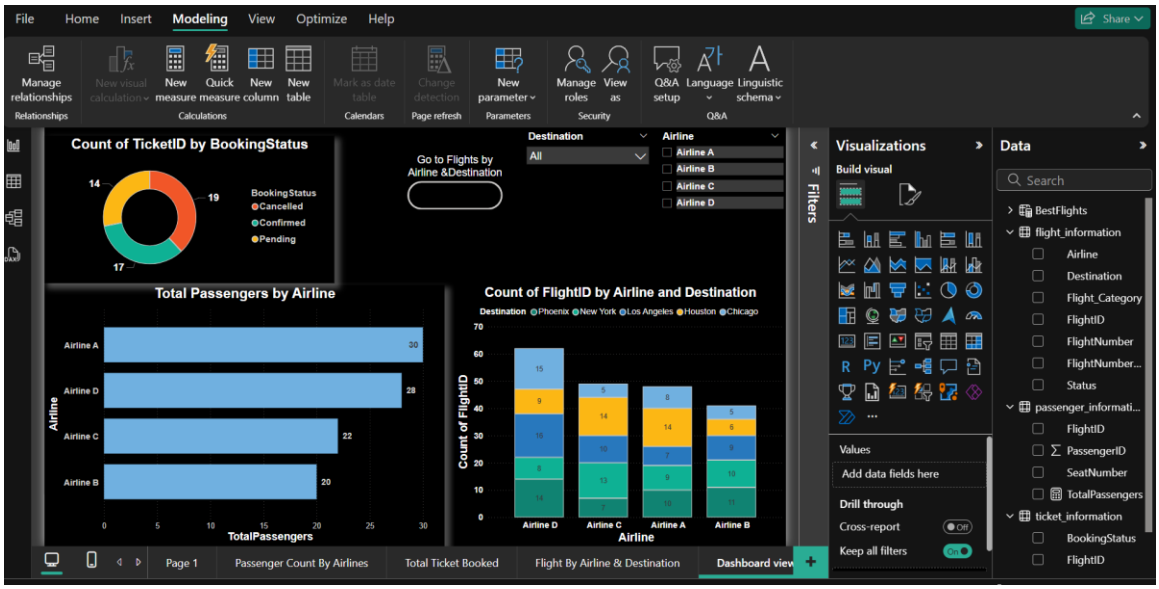
Now to Add Interactive feature we will use **Slicer** from the Visualization pane.

1. Add 1st slicer and under Field Drag → Airline from Flight Information table.
2. Add 2nd slicer and under Field Drag → Destination from Flight Information table → Go to format your visual and then slicer setting and set the style as dropdown.
3. Place the slicer as per convenience.

Question 2: → Quick views.

Answer: Followed the following steps:

1. For this we can Add a Quick View Button for Viewing the previous Page "Flight by Airline & Destination".
2. Go to Insert → Buttons → Blank.
3. Click on the button and Add a title from Format your Visual → general → Title → "Go to Flight By Airline and Destination".
4. Come back to button view in the same and then turn on Action → Type → set it as Page Navigation.
5. Under destination Choose Flight By Airline & Destination.



The above image shows that, it is a Dashboard where we can use the slicers to select the destination and the Airline and according to that it will show the data, also created a button that will navigate the user to go to the previous page of Flight By Airline & Destination.

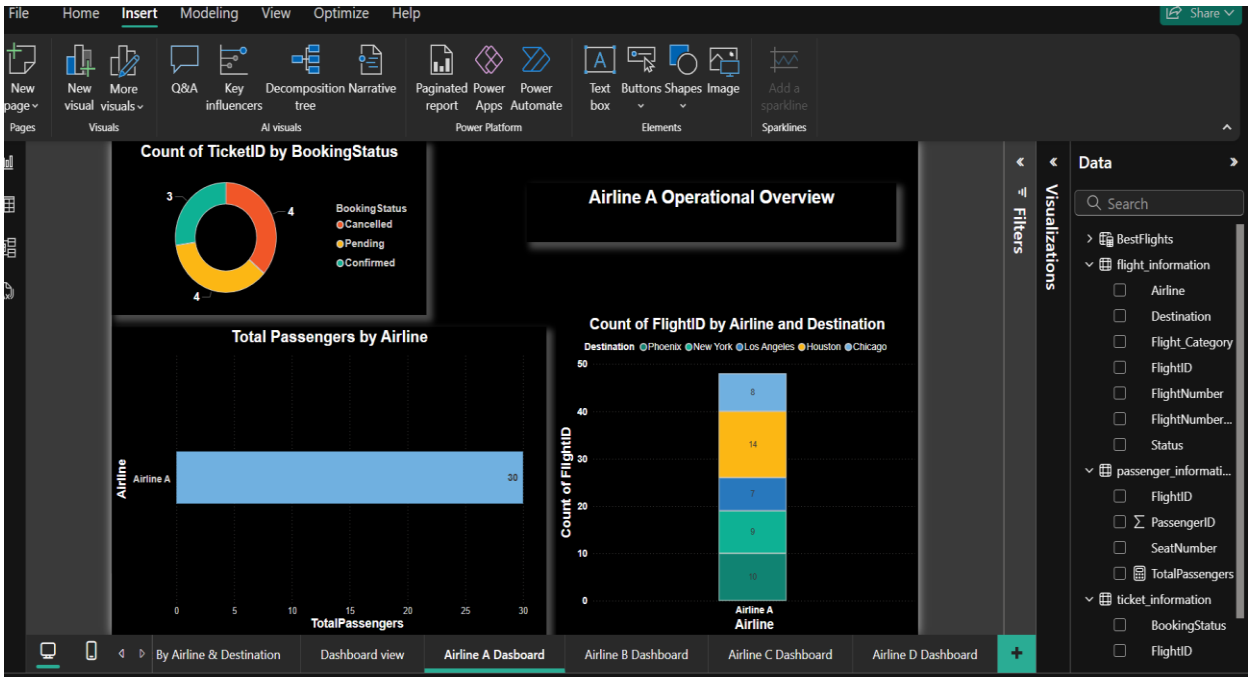
Question: Airline-specific pages.

Answer: Followed the following steps:

Create a Dedicated Page for Airline A

1. Duplicate an Existing Page:
   - Right-click on your main dashboard page ("Dashboard view") in the tabs at the bottom.
   - Click Duplicate Page.
   - Rename the new page to:
     → Airline A Dashboard
2. Apply Page-Level Filter:
   - Go to the Filters pane (on the right side).
   - Drag the Airline field from flight_information into the Page Filters area.
   - Select only:
     - Airline A
3. Customize the Title:
   - Add a text box (Insert → Text Box) and write:
     → "Airline A Operational Overview"
4. Confirm Filtering:
   - All visuals on this page will now only show data related to Airline A.

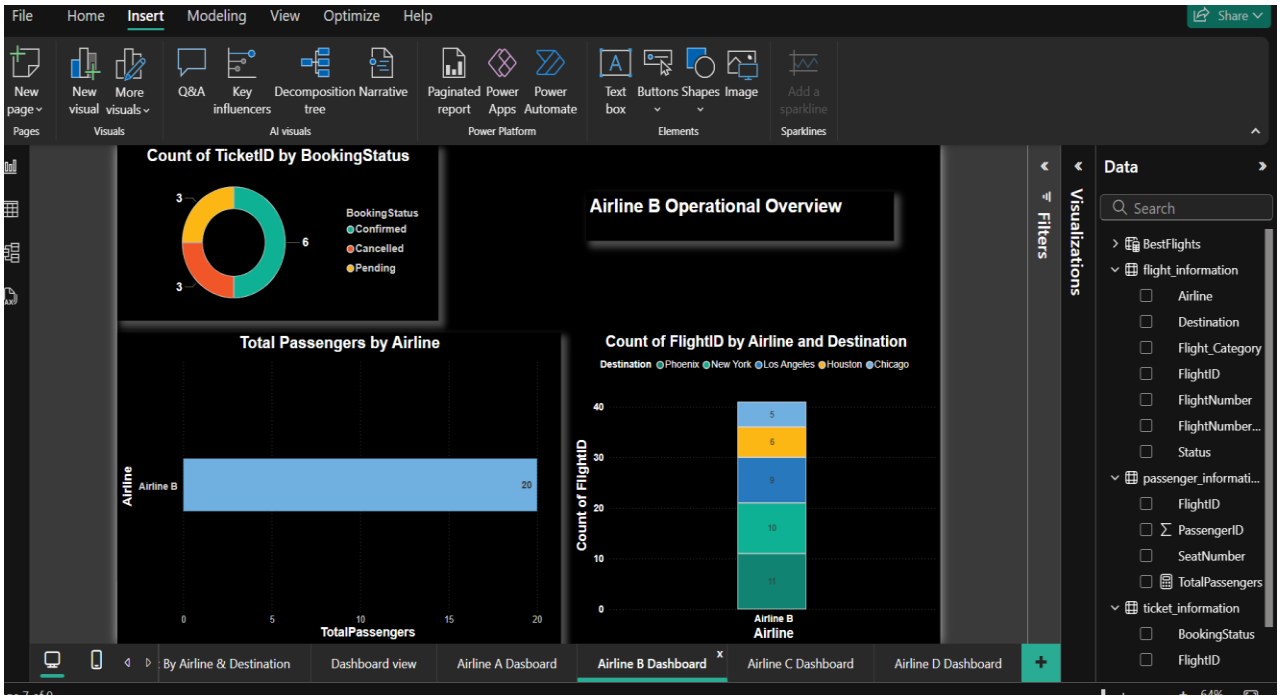
Note: Removed the slicers and the Button as it requires no filtering after this.



Create a Dedicated Page for Airline B:

1. Duplicate an Existing Page:
    - Right-click on your main dashboard page ("Dashboard view") in the tabs at the bottom.
    - Click Duplicate Page.
    - Rename the new page to:
      → Airline B Dashboard
2. Apply Page-Level Filter:
    - Go to the Filters pane (on the right side).
    - Drag the Airline field from flight_information into the Page Filters area.
    - Select only:
        - Airline B
3. Customize the Title:
    - Add a text box (Insert → Text Box) and write:
      → "Airline B Operational Overview"
4. Confirm Filtering:
    - All visuals on this page will now only show data related to Airline B.

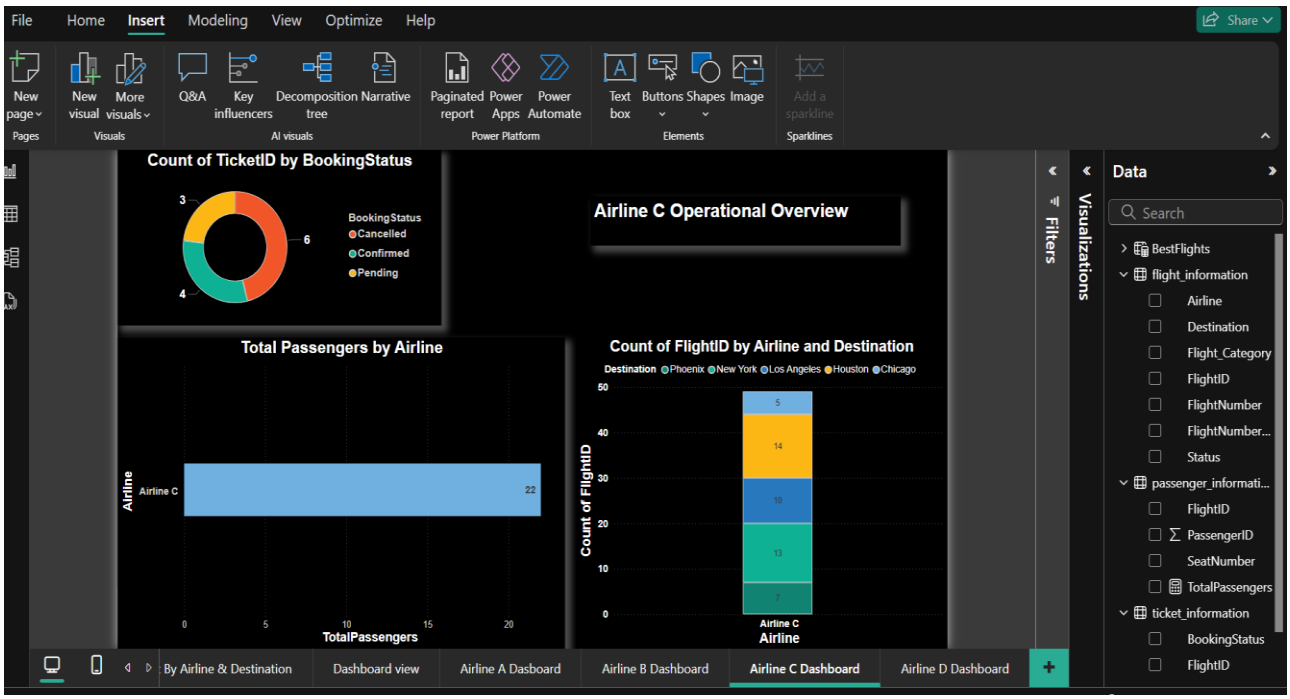
Note: Removed the slicers and the Button as it requires no filtering after this.



Create a Dedicated Page for Airline C

1. Duplicate an Existing Page:
    - Right-click on your main dashboard page (“Dashboard view”) in the tabs at the bottom.
    - Click Duplicate Page.
    - Rename the new page to:
      → Airline C Dashboard
2. Apply Page-Level Filter:
    - Go to the Filters pane (on the right side).
    - Drag the Airline field from flight_information into the Page Filters area.
    - Select only:
        - Airline C
3. Customize the Title:
    - Add a text box (Insert → Text Box) and write:
      → "Airline C Operational Overview"
4. Confirm Filtering:
    - All visuals on this page will now only show data related to Airline C.

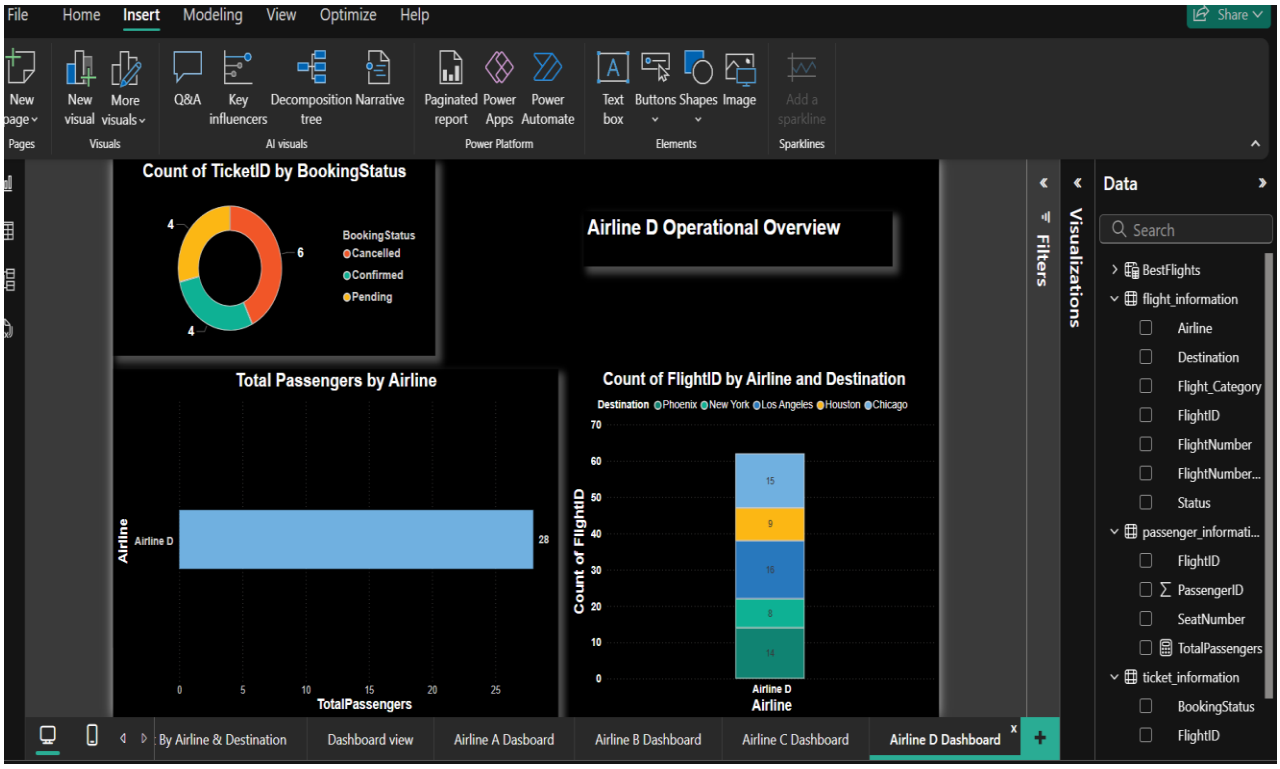
Note: Removed the slicers and the Button as it requires no filtering after this.



Create a Dedicated Page for Airline D

1. Duplicate an Existing Page:
    - Right-click on your main dashboard page ("Dashboard view") in the tabs at the bottom.
    - Click Duplicate Page.
    - Rename the new page to:
      → Airline D Dashboard
2. Apply Page-Level Filter:
    - Go to the Filters pane (on the right side).
    - Drag the Airline field from flight_information into the Page Filters area.
    - Select only:
        - Airline D
3. Customize the Title:
    - Add a text box (Insert → Text Box) and write:
      → "Airline D Operational Overview"
4. Confirm Filtering:
    - All visuals on this page will now only show data related to Airline D.

Note: Removed the slicers and the Button as it requires no filtering after this.



TASK 6: FINAL DASHBOARD AND POWER BI SERVICE (20 MARKS)

Question 1: Design a comprehensive dashboard with key visuals and insights.

Answer: Followed the following steps:

Note: Already Created the Dashboard in Task5 (Question 4), But writing these steps again to do that:

For doing this first copy all the visuals from the 3 pages:

1. Passengers count by Airlines.
2. Total Ticket Booked.
3. Flight by Airline & destination.

After this we will rename the page as Dashboard, and then arrange all these visuals as per readability.

Now to Add Interactive feature we will use **Slicer** from the Visualization pane.

4. Add 1st slicer and under Field Drag → Airline from Flight Information table.
5. Add 2nd slicer and under Field Drag → Destination from Flight Information table → Go to format your visual and then slicer setting and set the style as dropdown.
6. Place the slicer as per convenience.

For → Quick views.

7. For this we can Add a Quick View Button for Viewing the previous Page "Flight by Airline & Destination".
8. Go to Insert → Buttons → Blank.
9. Click on the button and Add a title from Format your Visual → general → Title → "Go to Flight By Airline and Destination".
10. Come back to button view in the same and then turn on Action → Type → set it as Page Navigation.
11. Under destination Choose Flight By Airline & Destination.

The final dashboard integrates all critical visualizations and interactivity features, providing users with dynamic insights into airline performance and operations.

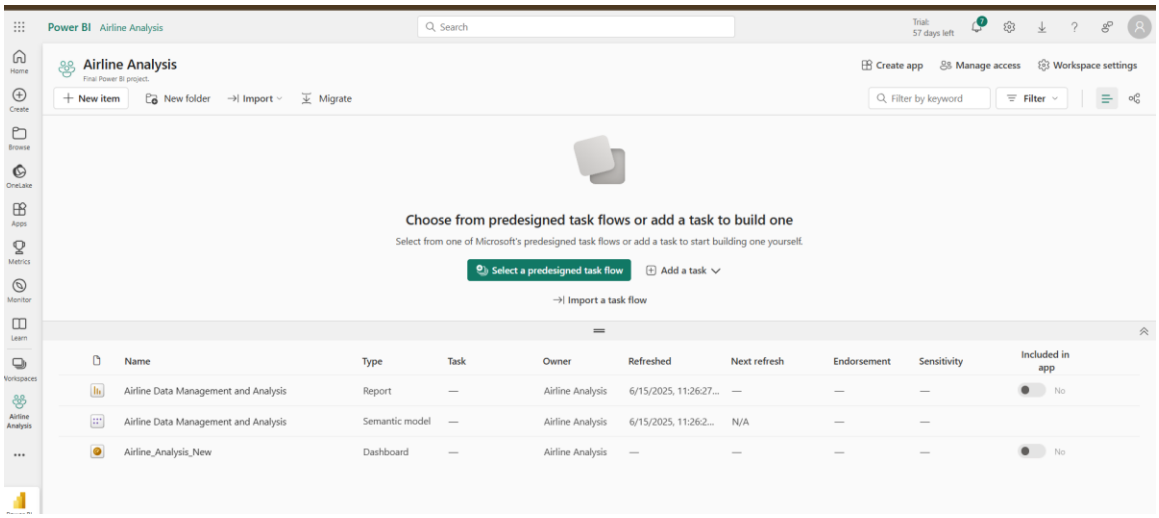
Features:

- **Passenger Analysis**: Bar chart to show total passengers per airline.
- **Ticket Booking Overview**: Donut chart to visualize booking statuses.
- **Flight Distribution**: Stacked column chart by airline and destination.
- **Slicers**: Airline and Destination filters for focused insights.
- **Navigation**: Button added for quick access to detailed views.
- **Dedicated Page**: An Airline A-specific page created for role-based insights.

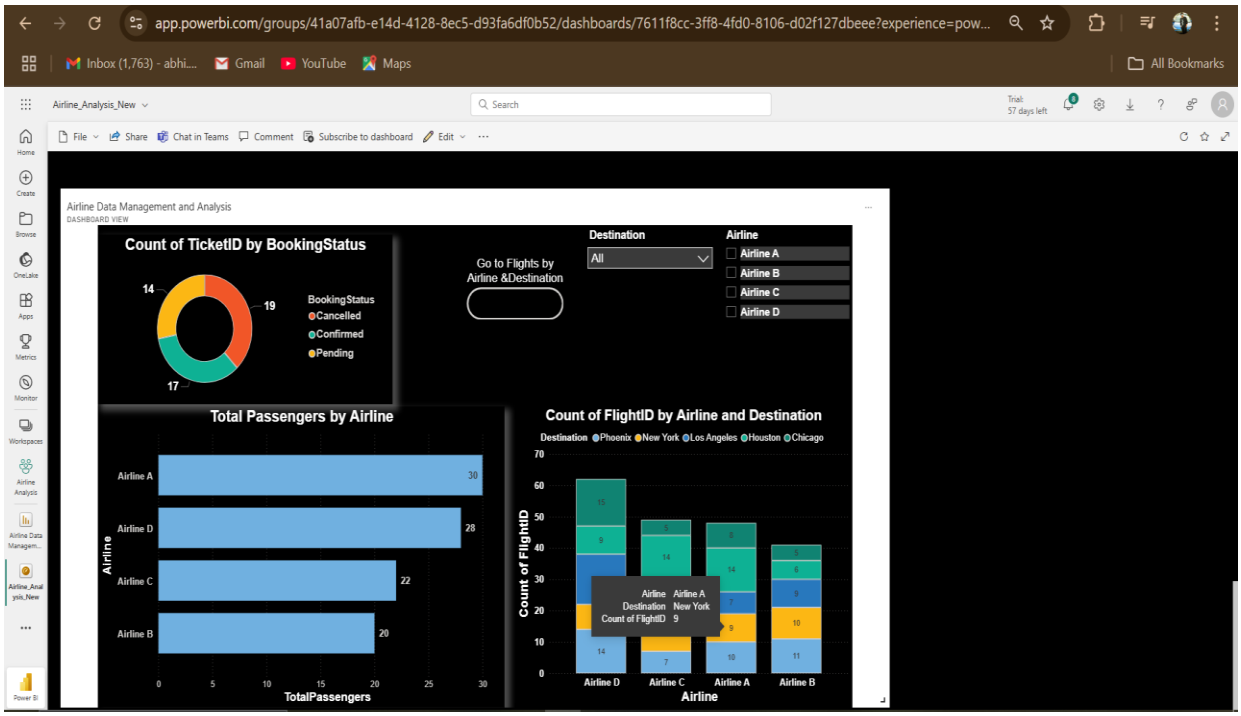
This dashboard serves as the central view for data-driven decision-making.

We can also create a Dashboard from Power BI service → In the power BI service create a New Workspace named as Airline Analysis and from Power BI Desktop → Go to Publish and select the Airline Analysis workspace.
In that once its published in Power BI service → Go to the report → Click on the Three dots for more options and then select Pin to Dashboard → you can either pin to existing Dashboard or can create a new → Rename it as

Airline_Analysis_New and then pin other report pages to the same.
Click again on the Workspace named as Airline_Analysis.
It will show the dashboard → click on that and you can format that accordingly.



Before clicking on the dashboard, it will show like this.



It will show like this if we will pin the dashboard report page to this.

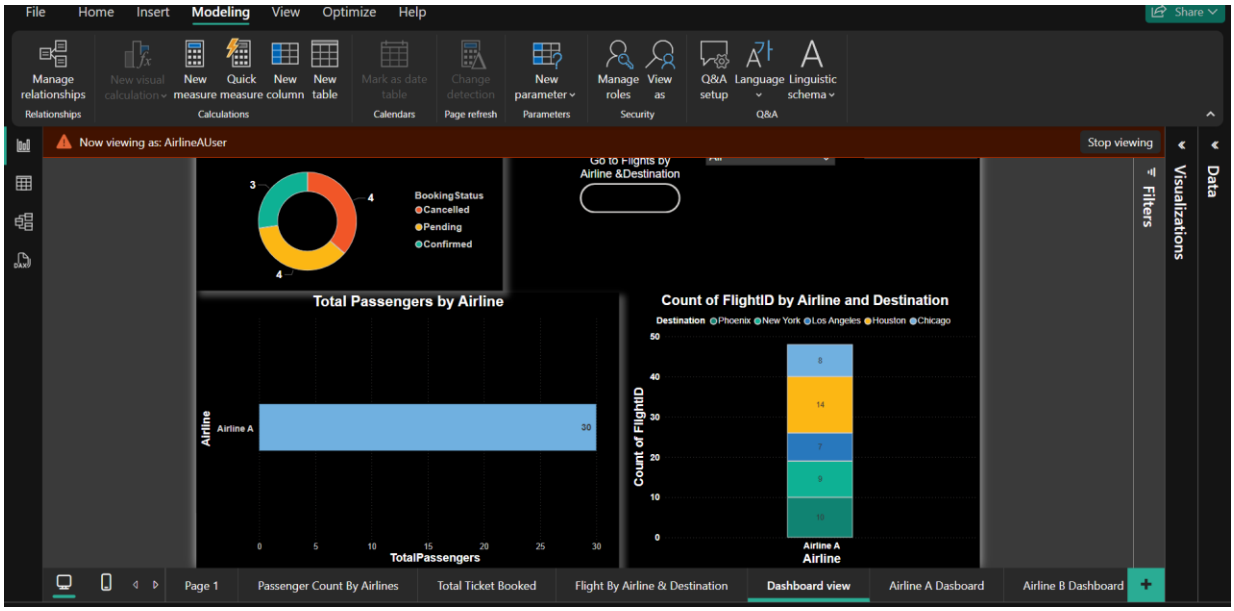
Question 2: Configure Row-Level Security (RLS) for Airline A data and assign it to a user.

Answer: Followed the following steps:

1. In Power BI Desktop:
   Go to **Modeling → Manage Roles**
2. Click **Create** and name the role:
   ➤ AirlineAUser
3. Select the flight_information table
4. Add this DAX filter:
   [Airline] = "Airline A"
5. Click on Save.

To test it:

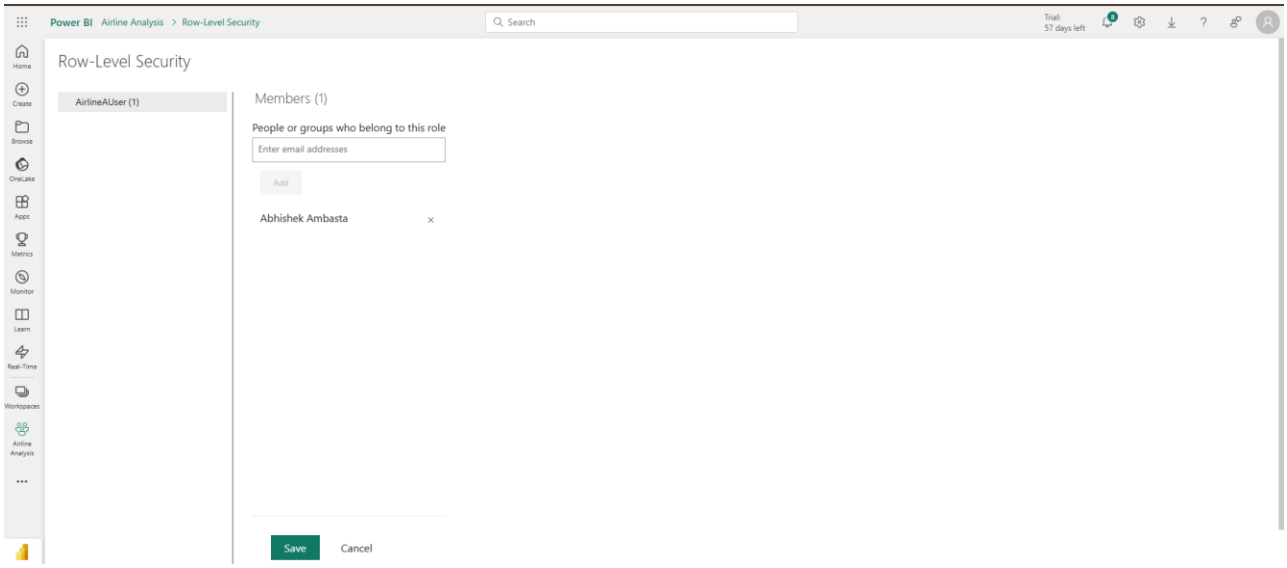
1. Go to Modeling → View as Roles
2. Select AirlineAUser → visuals should now only show Airline A data.



Now save this and Publish it to the Airline Analysis Workspace.

To Assign it to the User:

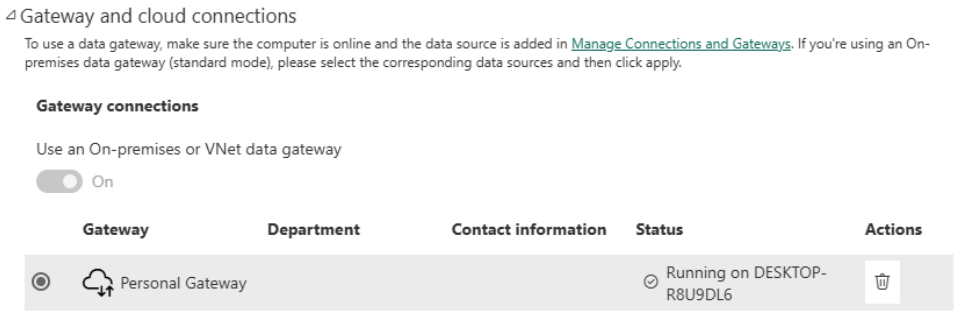
1. Go to Power BI Service.
2. In your dataset → click Security.
3. Assigned my email address to the role AirlineAUser.



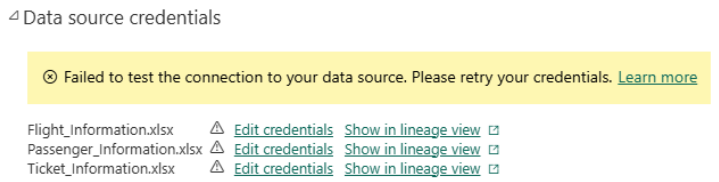
Question 3: Set up a schedule refresh at 5 PM daily.

Answer: Followed the following steps:

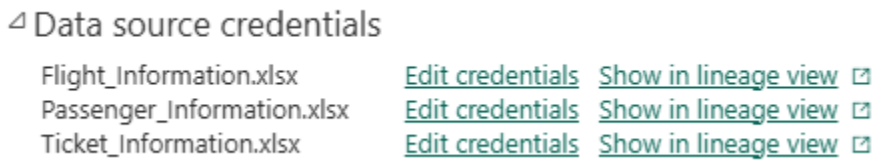
1. Go to setting → Power BI setting → Click on Semantic Models → expand Gateway and cloud connection → this will show like:



2. Under Data Source connect it will show like this:

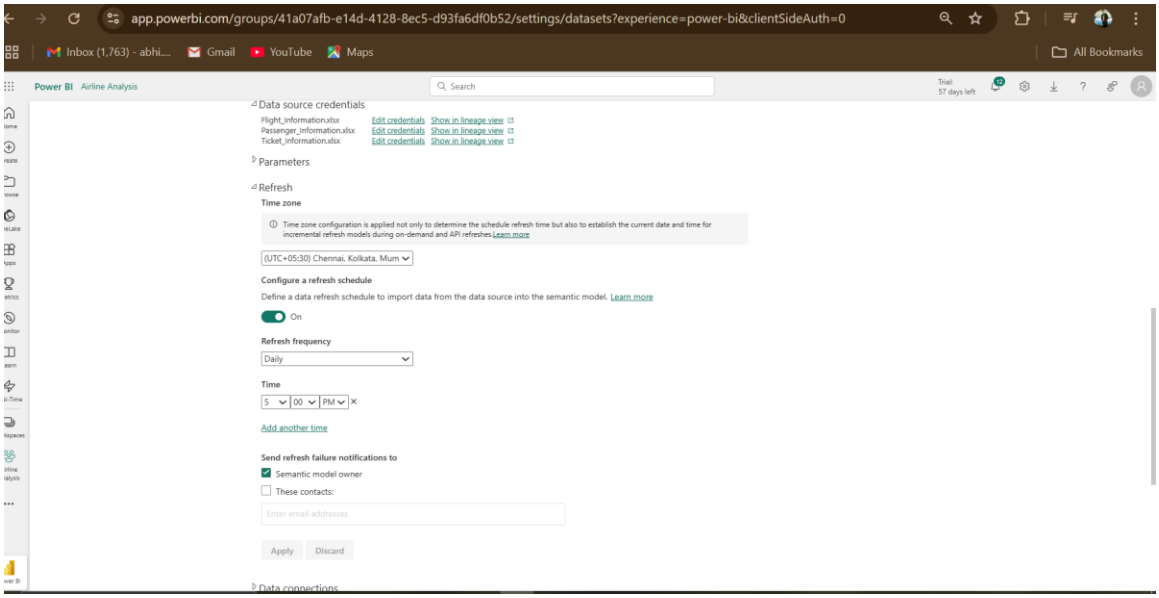


3. To resolve this, we will click on the edit credentials → click sign in do it for 3 times.



This will resolve the error.

1. By scrolling down, you will find Refresh option → expand that → Select time Zone as India UTC +05:30 → click on add time and set the time as 05:00 PM.
2. Click on Apply.



This will set the daily refresh time as 05:00 PM as per Indian time zone.