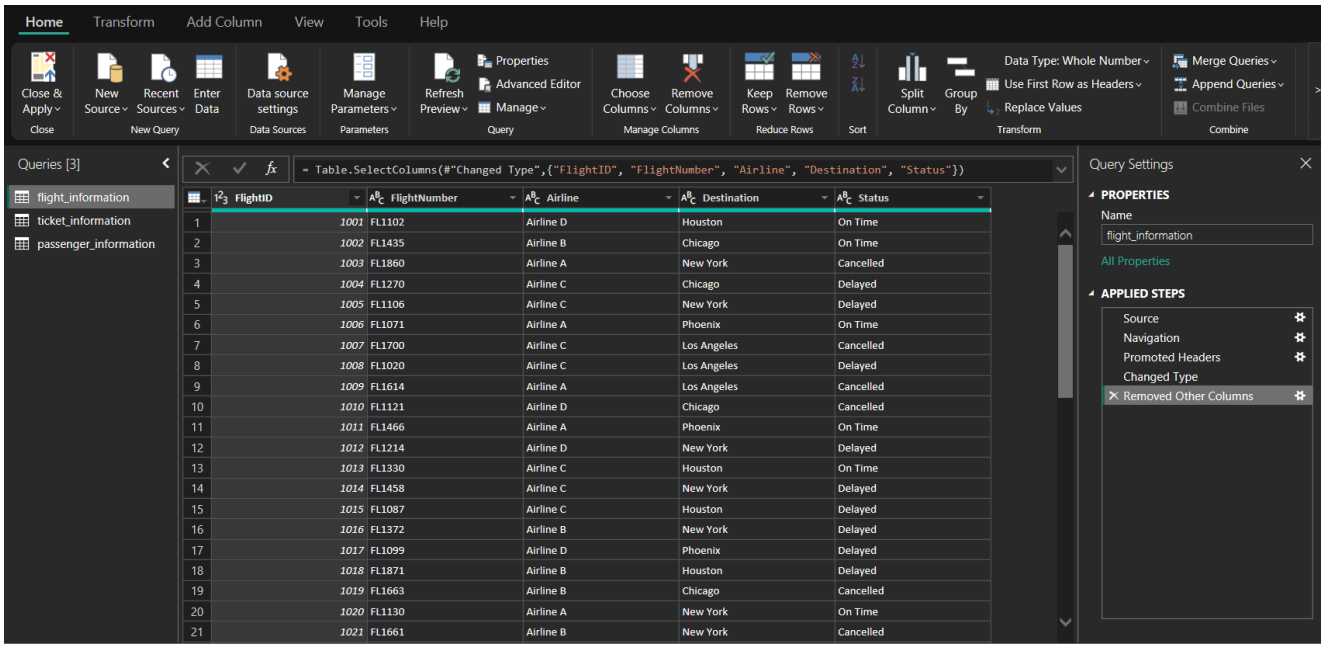
Airline Data Management and Analysis Using Power BI

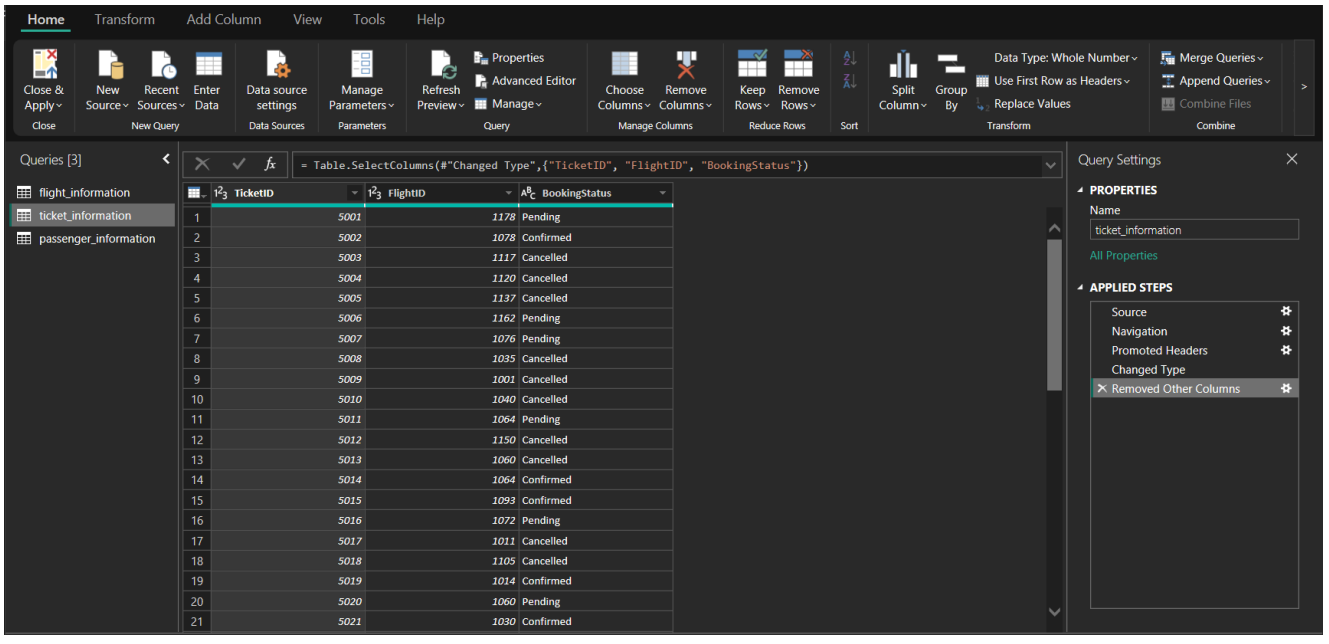
By: Hardik Verma

1. Data Preparation and Cleaning

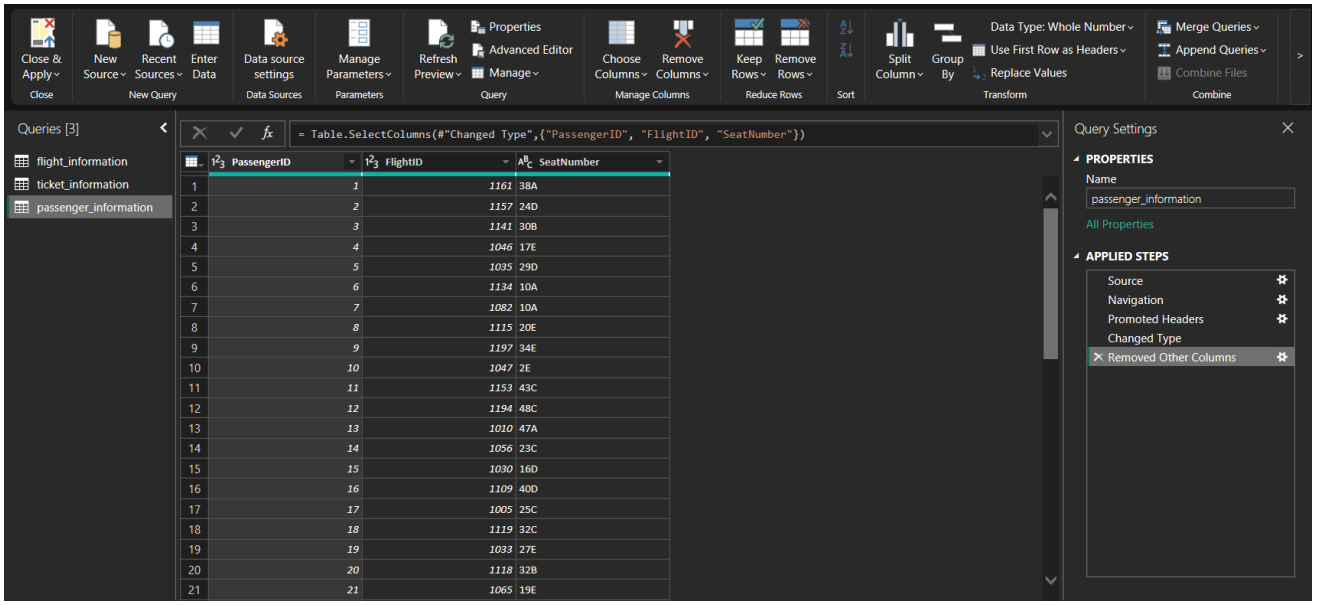
I used the **Power Query Editor** clean and prepare the data**. All the 3 datasets given had null columns which I removed using the remove columns option in the Home menu**.
I ensured that all the columns had the proper format and handled missing values and duplicates.



	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
21	1021	FL1661	Airline B	New York	Cancelled



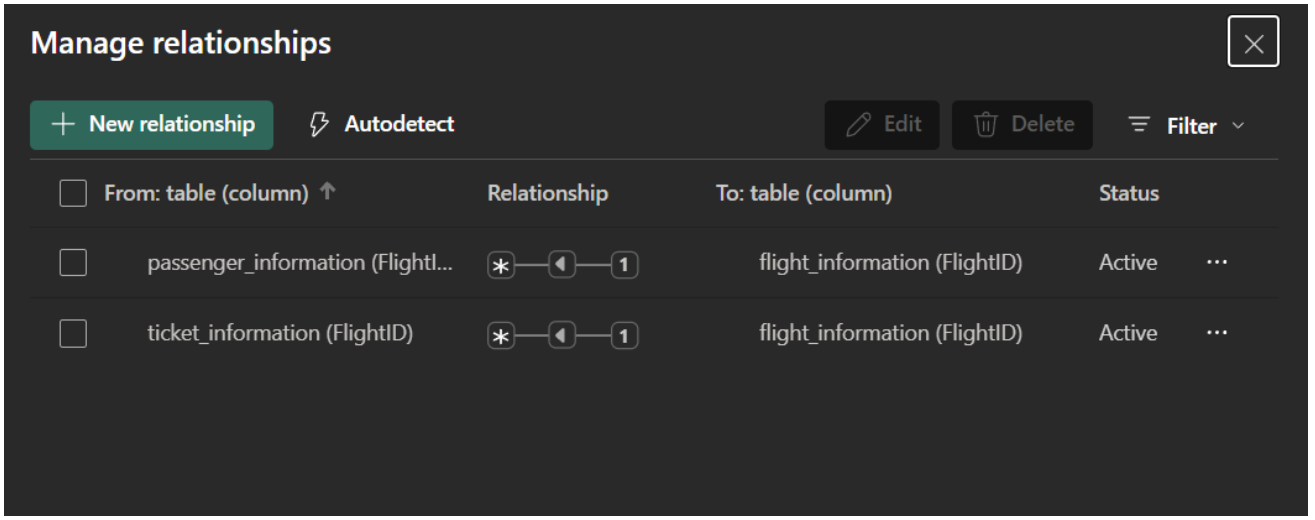
	TicketID	FlightID	BookingStatus
1	5001	1178	Pending
2	5002	1078	Confirmed
3	5003	1117	Cancelled
4	5004	1120	Cancelled
5	5005	1137	Cancelled
6	5006	1162	Pending
7	5007	1076	Pending
8	5008	1035	Cancelled
9	5009	1001	Cancelled
10	5010	1040	Cancelled
11	5011	1064	Pending
12	5012	1150	Cancelled
13	5013	1060	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
21	5021	1030	Confirmed

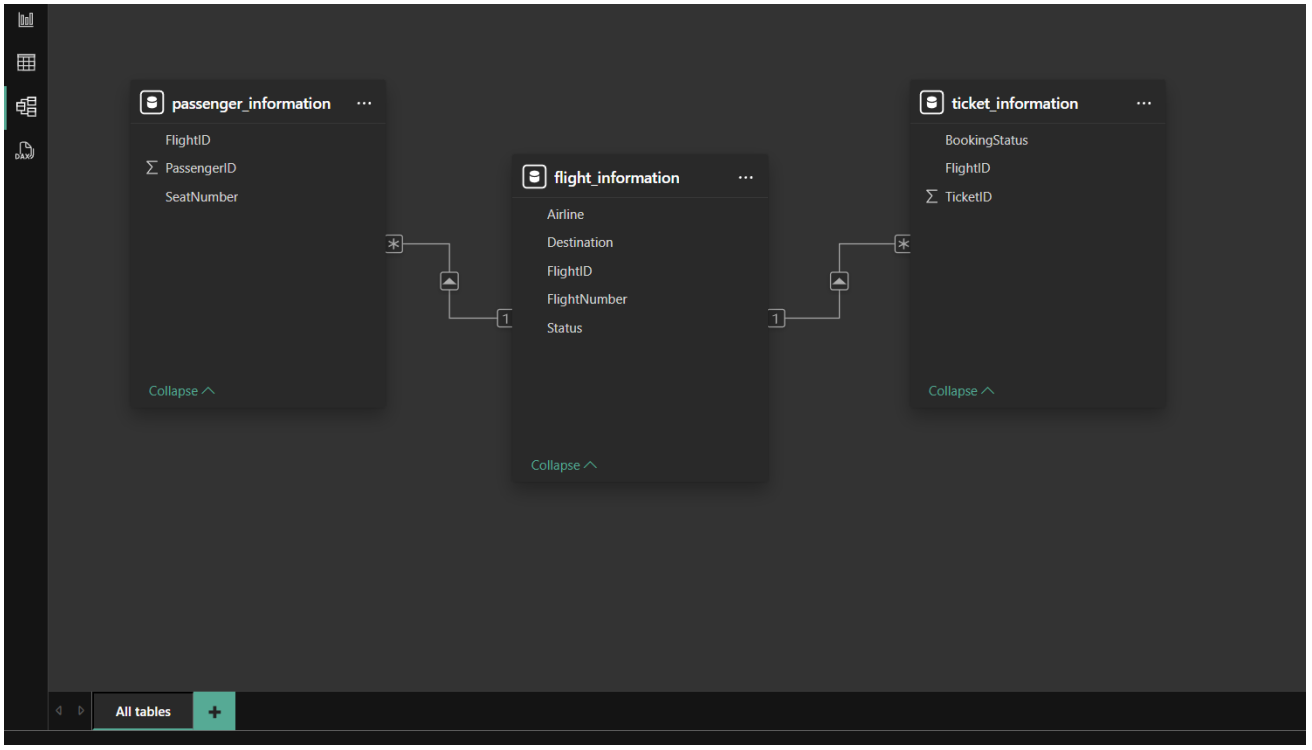


2. Data Modeling

I created **two relationships** that are as follows:

1. Between **flight_information and ticket_information on the basis of FlightID column**. There is a **1-to-many relationship from flight_information to ticket_information and FlightID is the key**.
2. Between **flight_information and passenger_information on the basis of FlightID column**. There is a **1-to-many relationship from flight_information to passenger_information and FlightID is the key**.

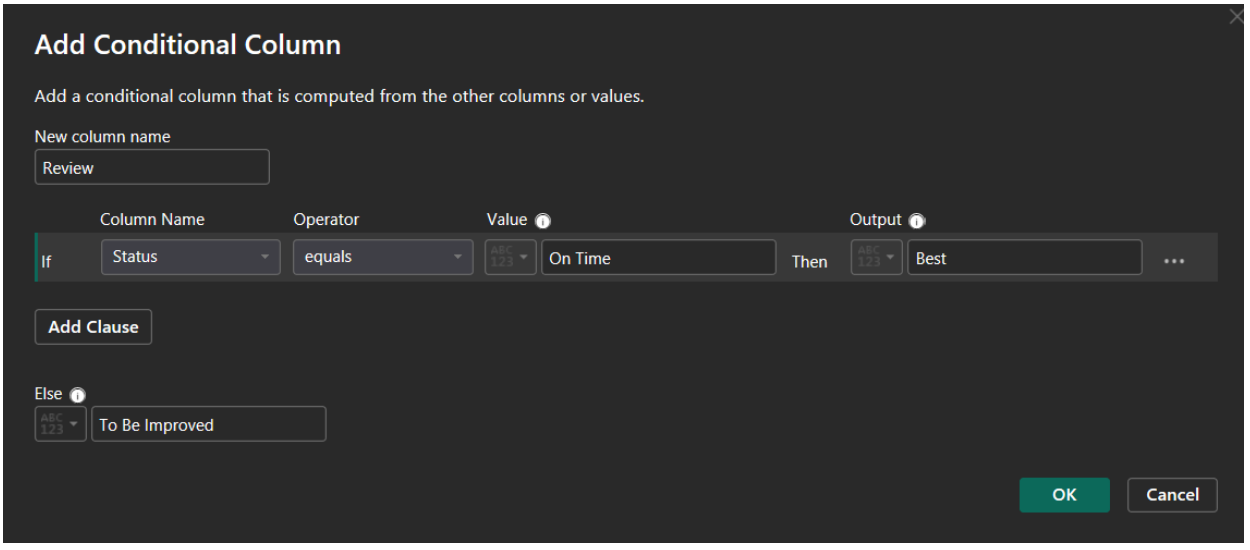


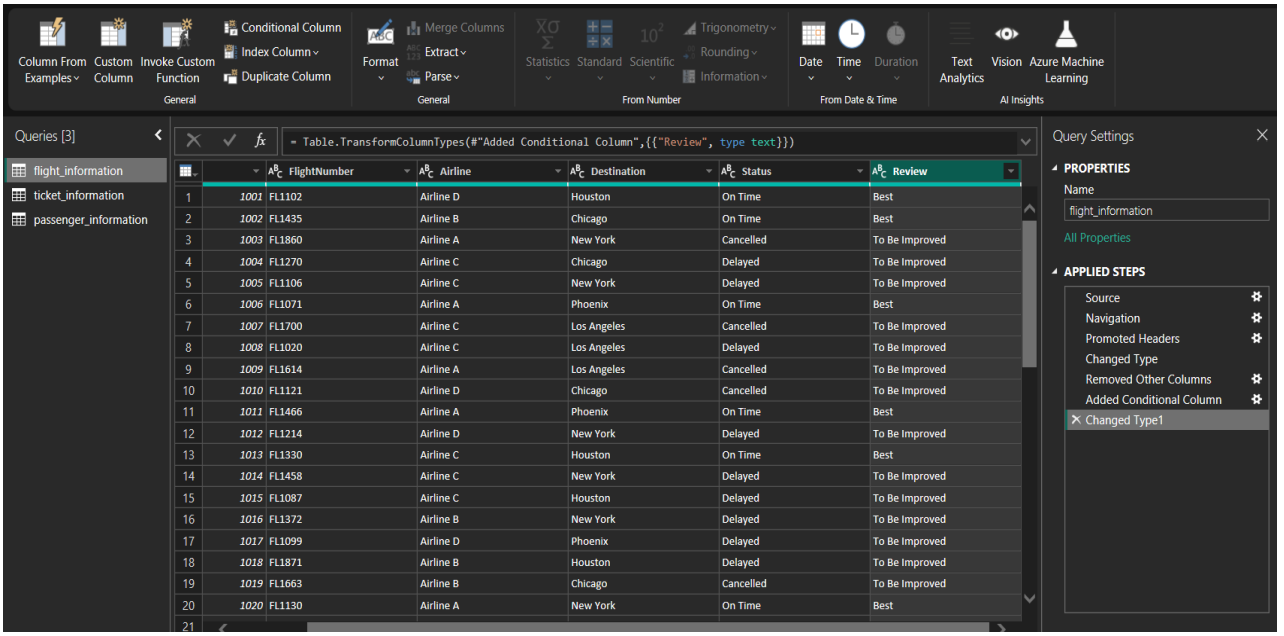


3. Enhanced Data Insights

I added a **Review** column using the **conditional column** option in the **add column** menu of the **Power Query Editor**.

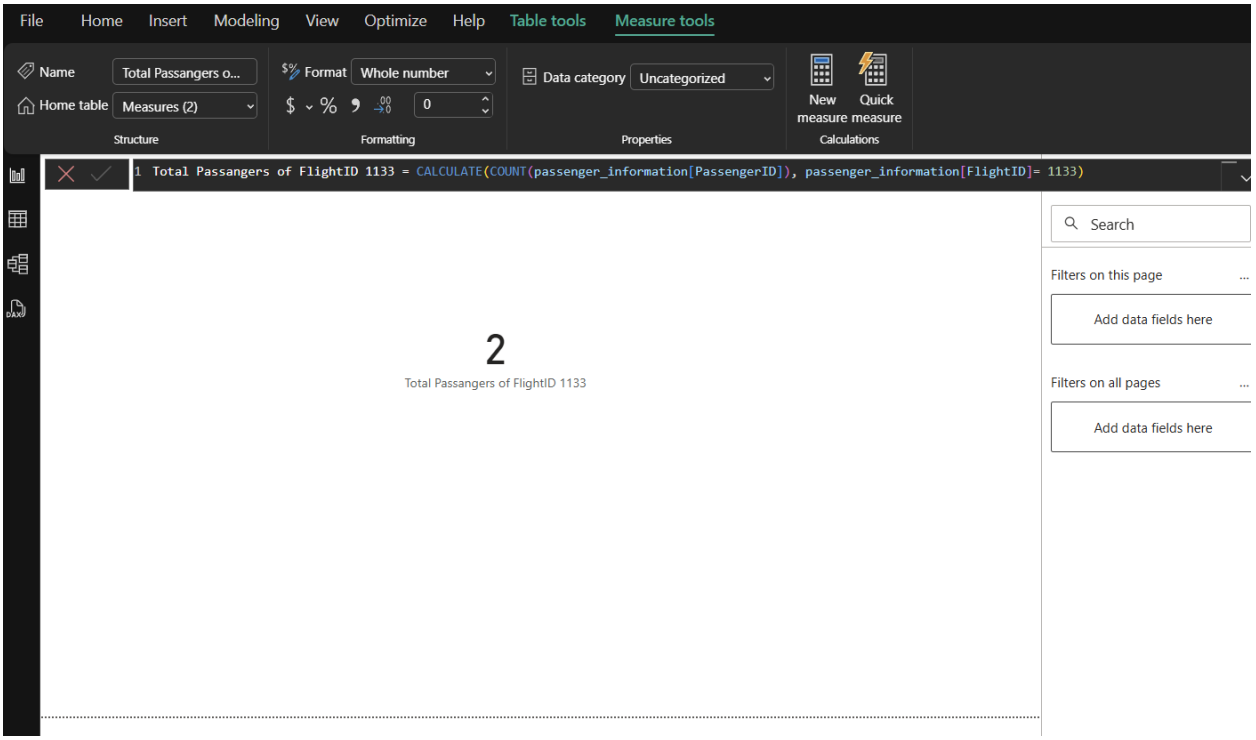
I defined the condition that if the Status is “On Time”, then that flight is “Best” and otherwise, in the case of “Delayed” and “Cancelled”, the flight is “To Be Improved”.
Finally, I corrected the format of the Review column to Text.





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

4. Calculations Using DAX

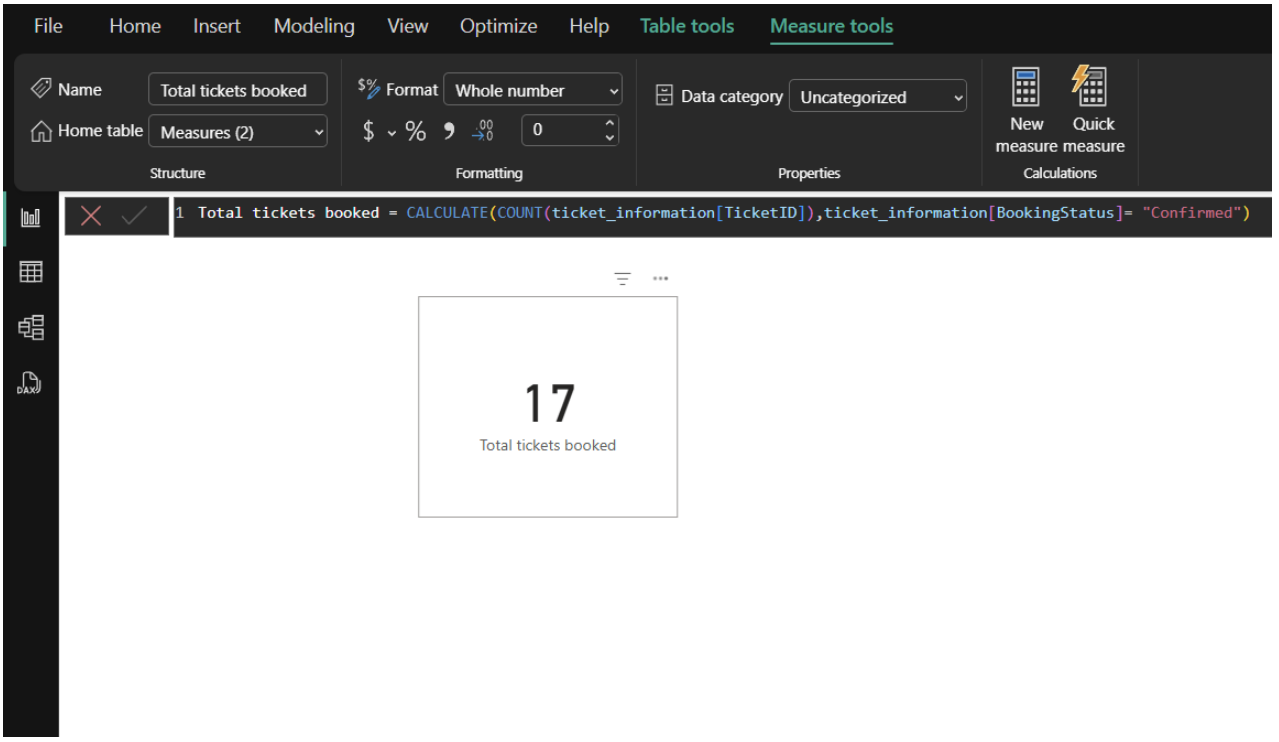


To calculate the Total passengers for the flight having **FlightID 1133**, I created a **measure** and visualized it with the **card visual** from the **visualization pane**.

I used the DAX functions **CALCULATE and COUNT** to structure the formula. The DAX formula is as follows;

Total Passangers of FlightID 1133 =
CALCULATE(COUNT(passenger_information[PassengerID]),
passenger_information[FlightID]= 1133)

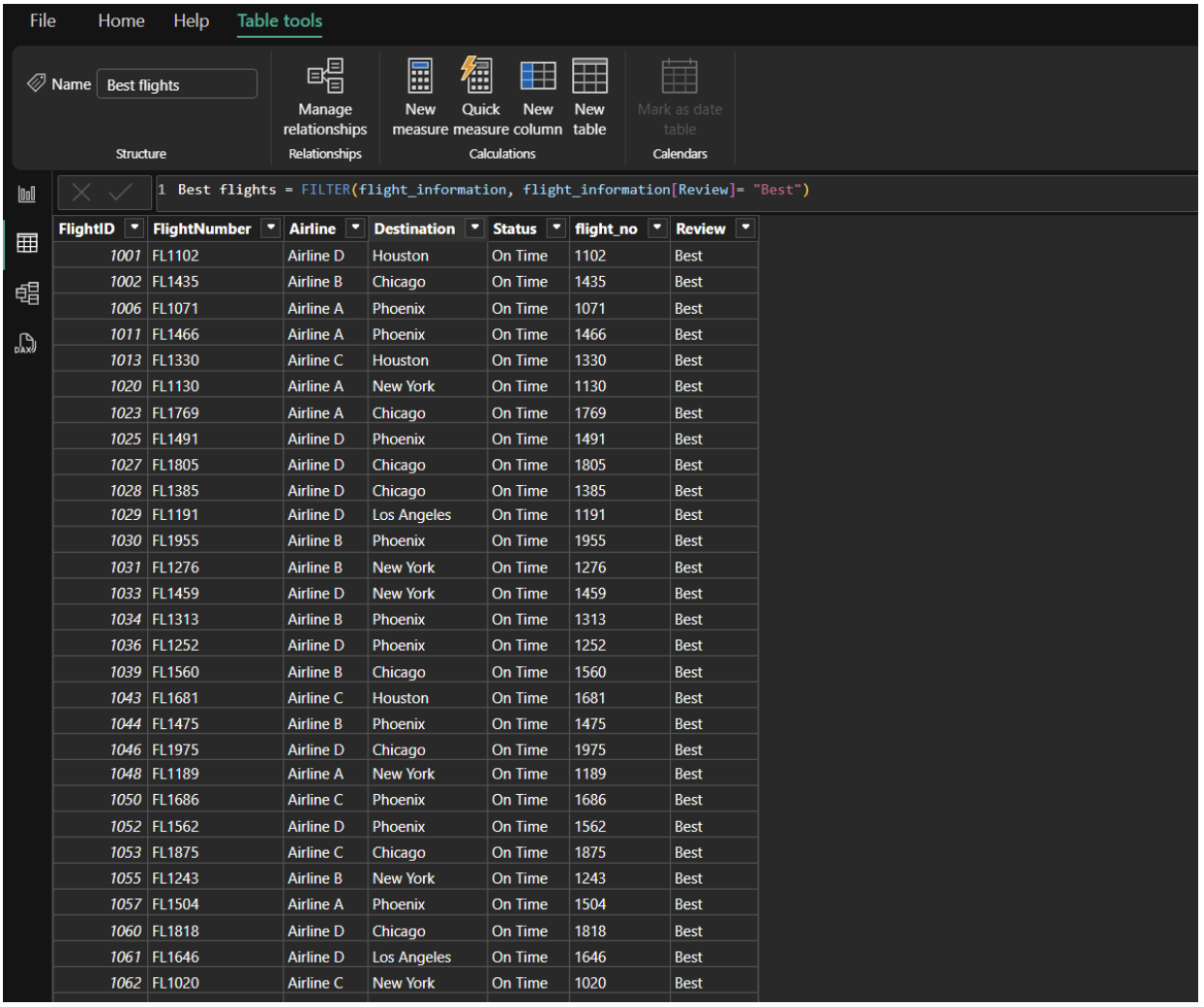
I found that we have the information of 2 passengers travelling in flight having FlightID 1133.



To calculate the Total tickets booked, I created a **measure using the CALCULATE and COUNT DAX functions.** I counted all the tickets that have the **BookingStatus as "Confirmed"**.

The DAX formula is as follows;
Total tickets booked = CALCULATE(COUNT(ticket_information[TicketID]),ticket_information[BookingStatus]= "Confirmed")

I found that there are a total of 17 tickets that are booked and confirmed out of all the ticket information we have**.**



FlightID	FlightNumber	Airline	Destination	Status	flight_no	Review
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
1050	FL1686	Airline C	Phoenix	On Time	1686	Best
1052	FL1562	Airline D	Phoenix	On Time	1562	Best
1053	FL1875	Airline C	Chicago	On Time	1875	Best
1055	FL1243	Airline B	New York	On Time	1243	Best
1057	FL1504	Airline A	Phoenix	On Time	1504	Best
1060	FL1818	Airline D	Chicago	On Time	1818	Best
1061	FL1646	Airline D	Los Angeles	On Time	1646	Best
1062	FL1020	Airline C	New York	On Time	1020	Best

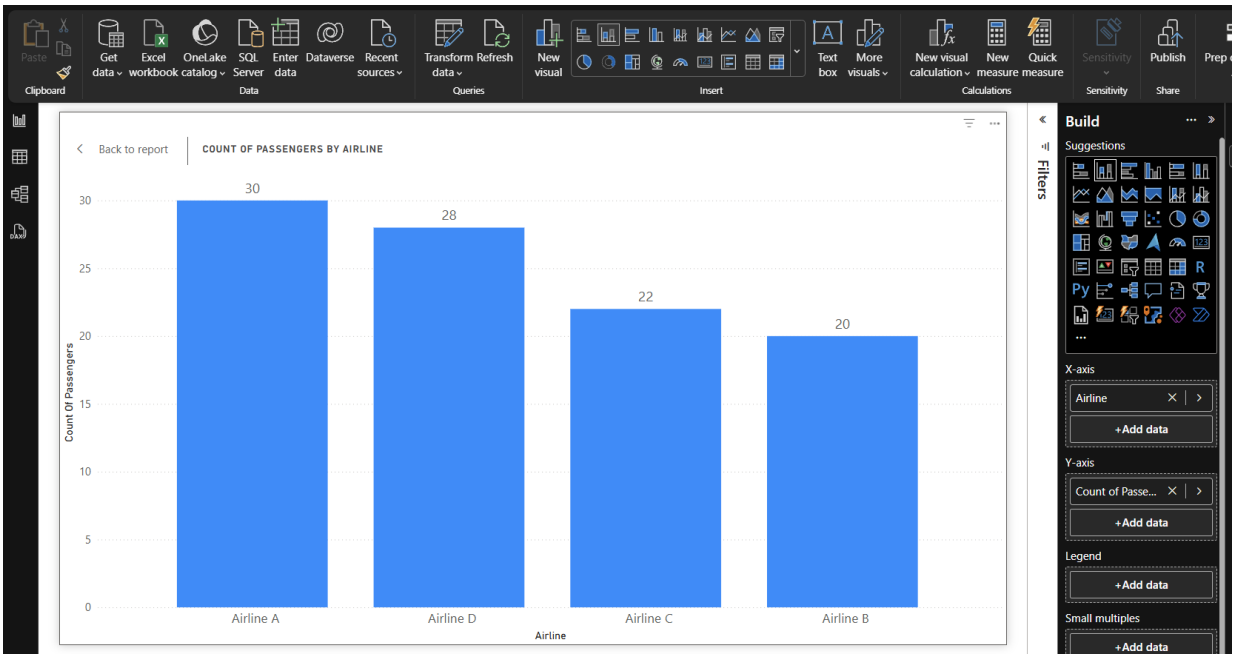
I added a new table using the **New Table** option in the **Table view**. I used the **FILTER** function to **filter the flight_information table into only best flights**.

The DAX formula is as follows;

Best flights = FILTER(flight_information, flight_information[Review]= "Best")

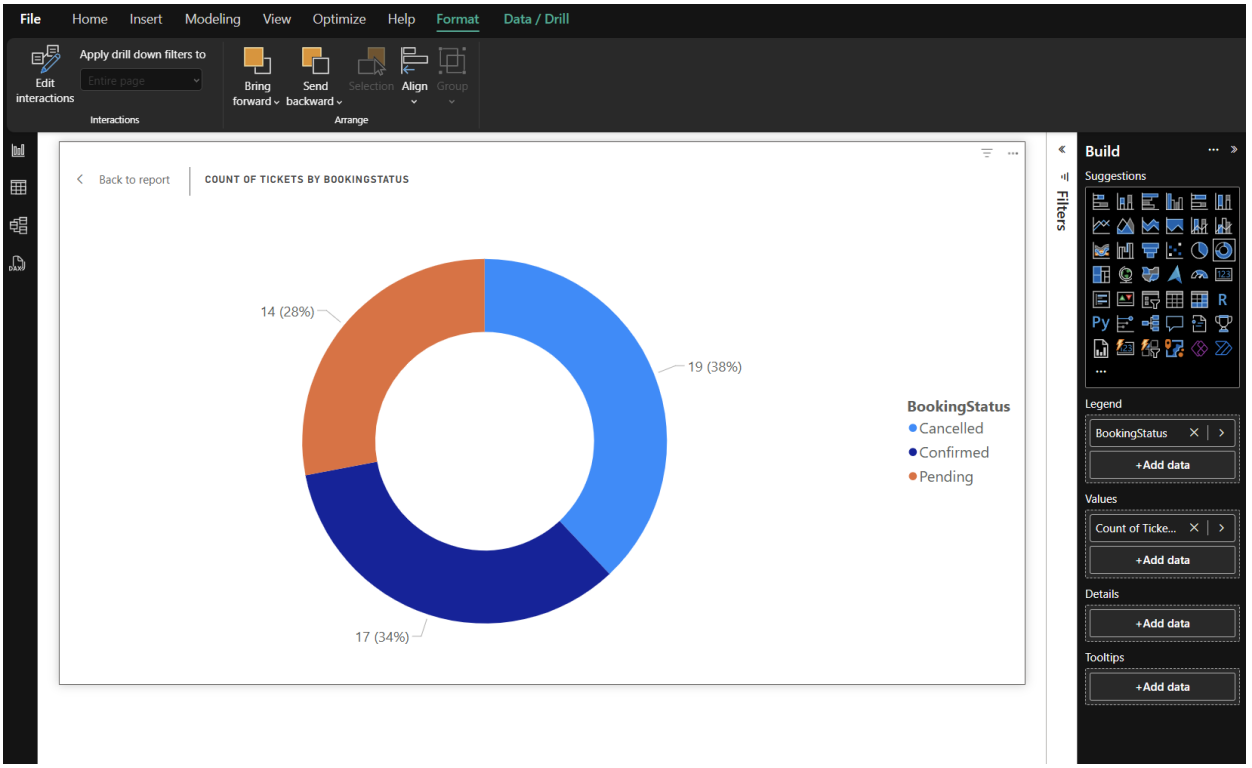
5. Visualization and Interactive Features

To show the **Passenger count by airline**, I used the **Bar chart visual from the visualisation pane**.



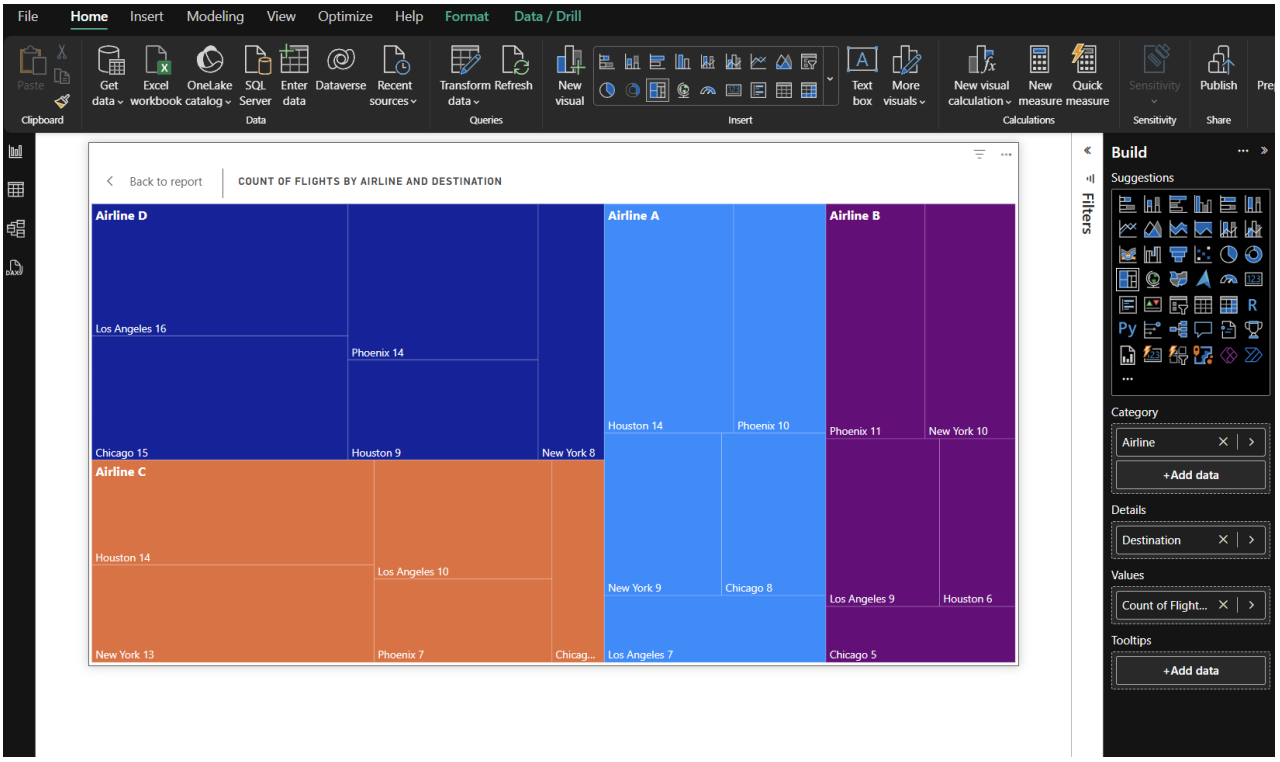
Airline A has the most number of passengers and Airline D is the close second.

To show the **Ticket booking statuses**, I used the **Donut chart from the visualization pane to visualize the number of tickets with different statuses**.



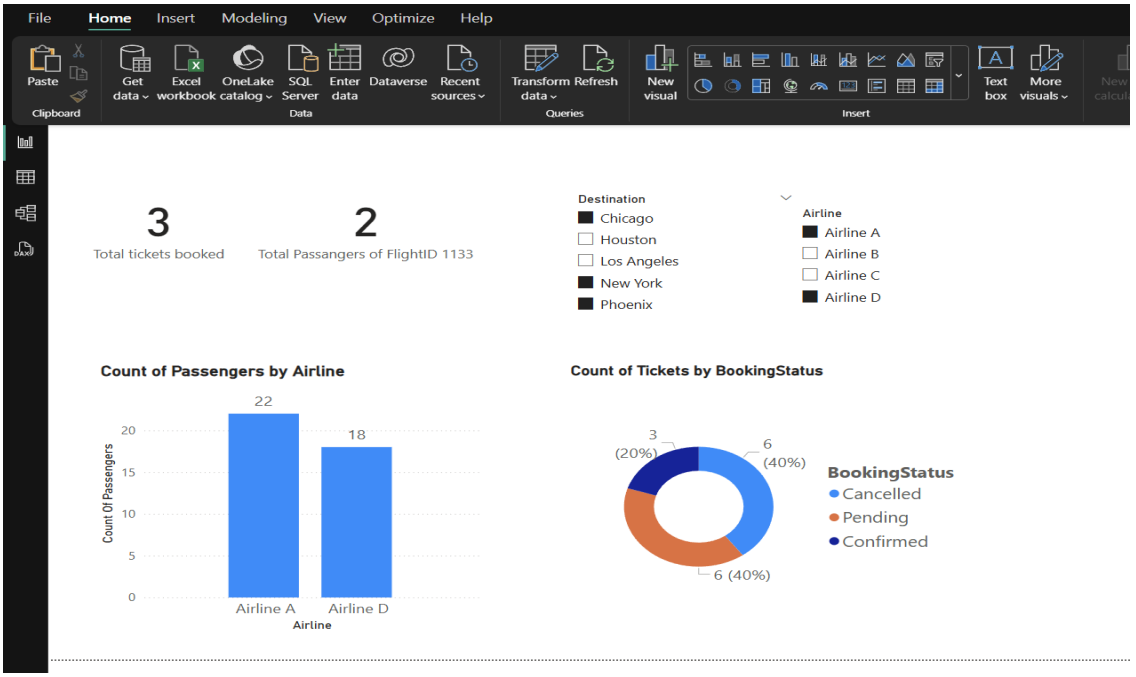
38% of the tickets are cancelled and 34% are confirmed.

To show the **Flights by airline and destination**, I used the **Treemap visual** from the **visualization pane** where **I visualized the count of flights to the different Destinations of every Airline.**

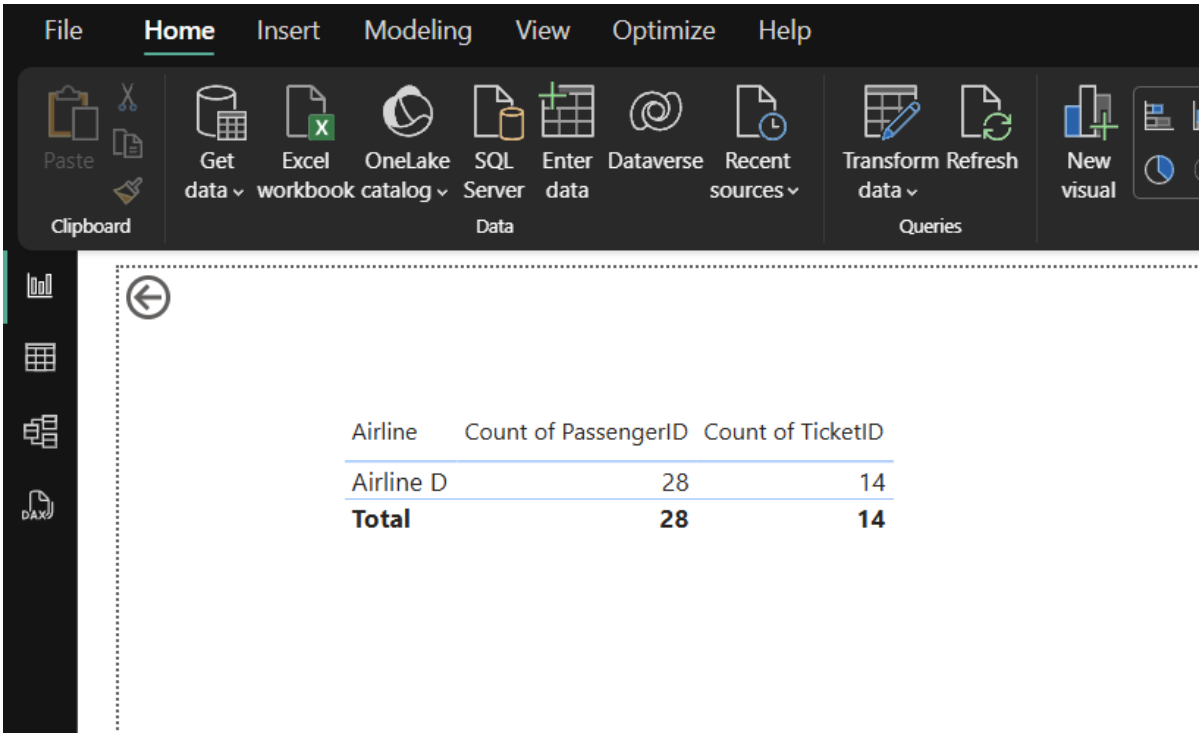


Airline D has the greatest number of flights to Los Angeles. While, both Airline A and C have the greatest number of flights to Houston. And Airline B has the greatest number of flights to Phoenix.

To make the report interactive for the user, I added interactive **slicers** for **Destination and Airline**. This gives the users, **an interactive experience** where they can **filter on the basis of Destination and Airline.**

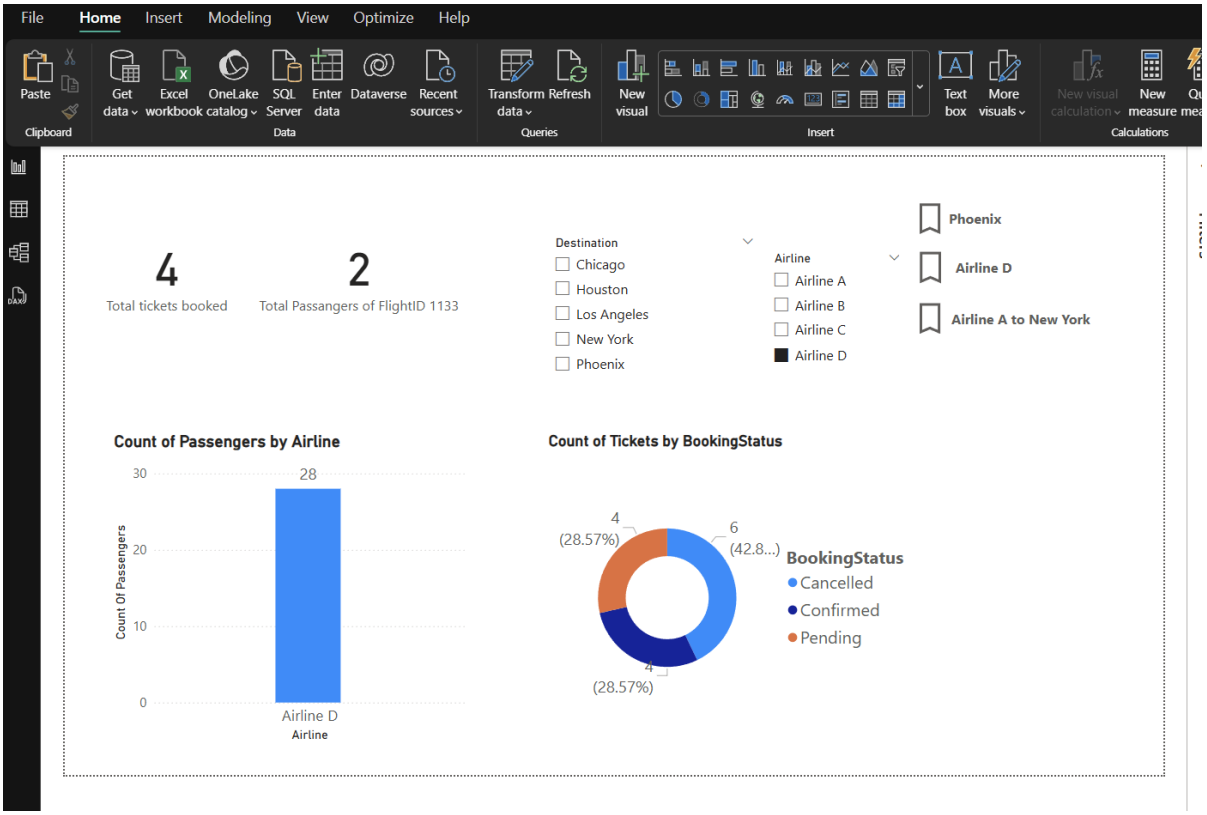


Also, I integrated **Drill through page for summarization for a specific Airline**.

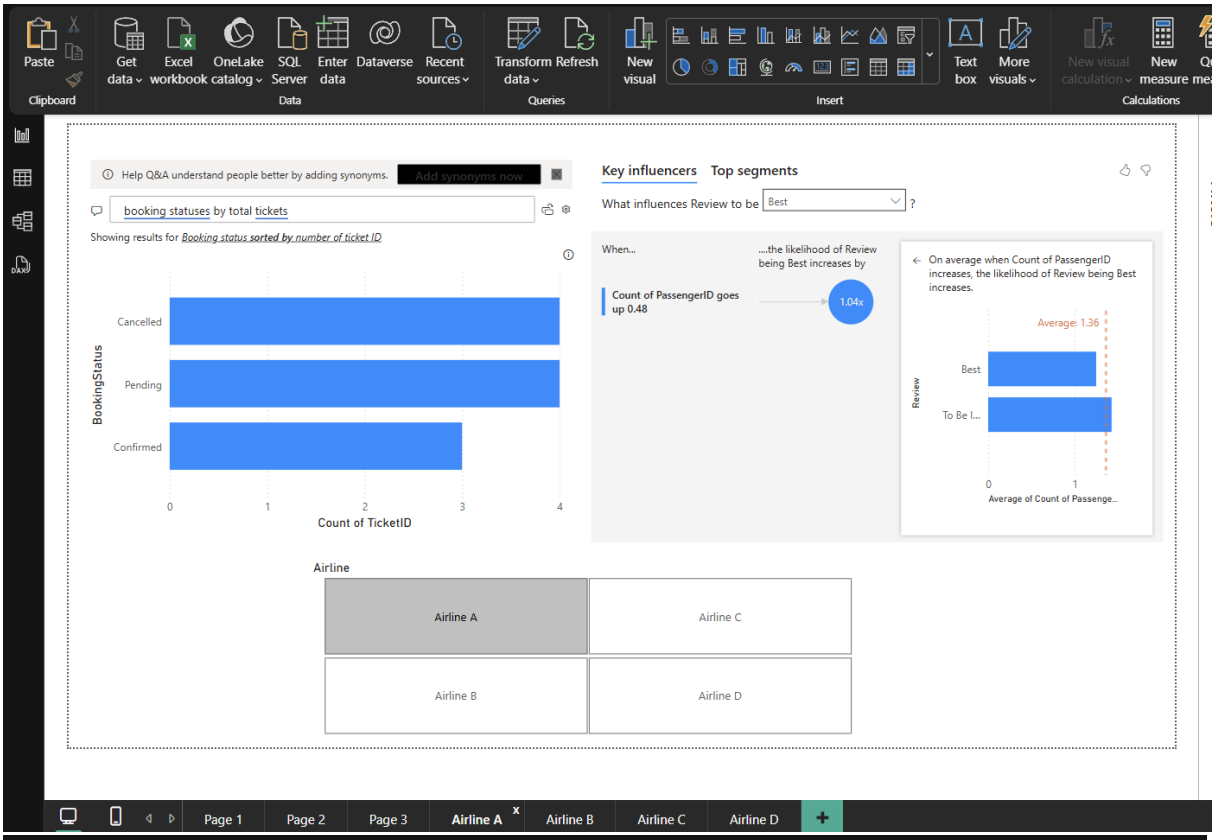


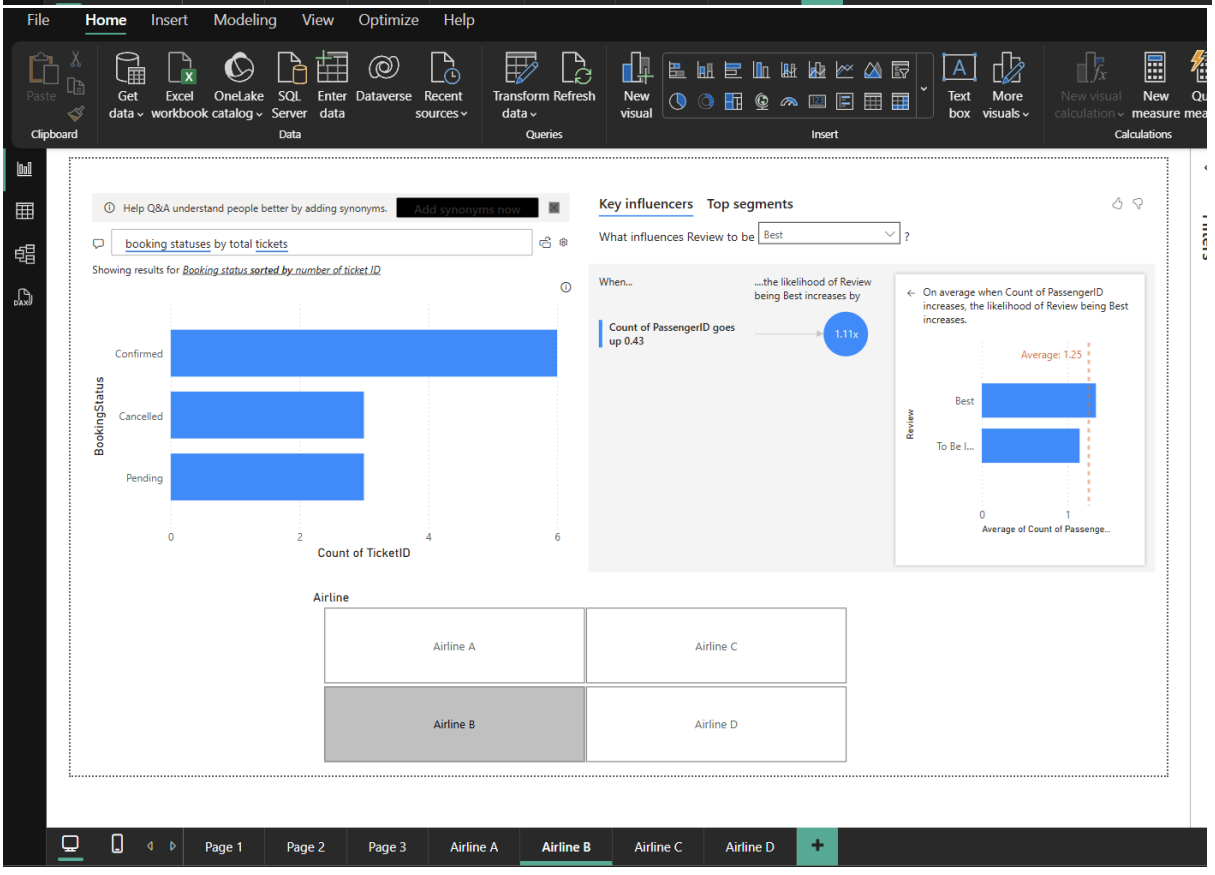
Airline	Count of PassengerID	Count of TicketID
Airline D	28	14
Total	**28**	**14**

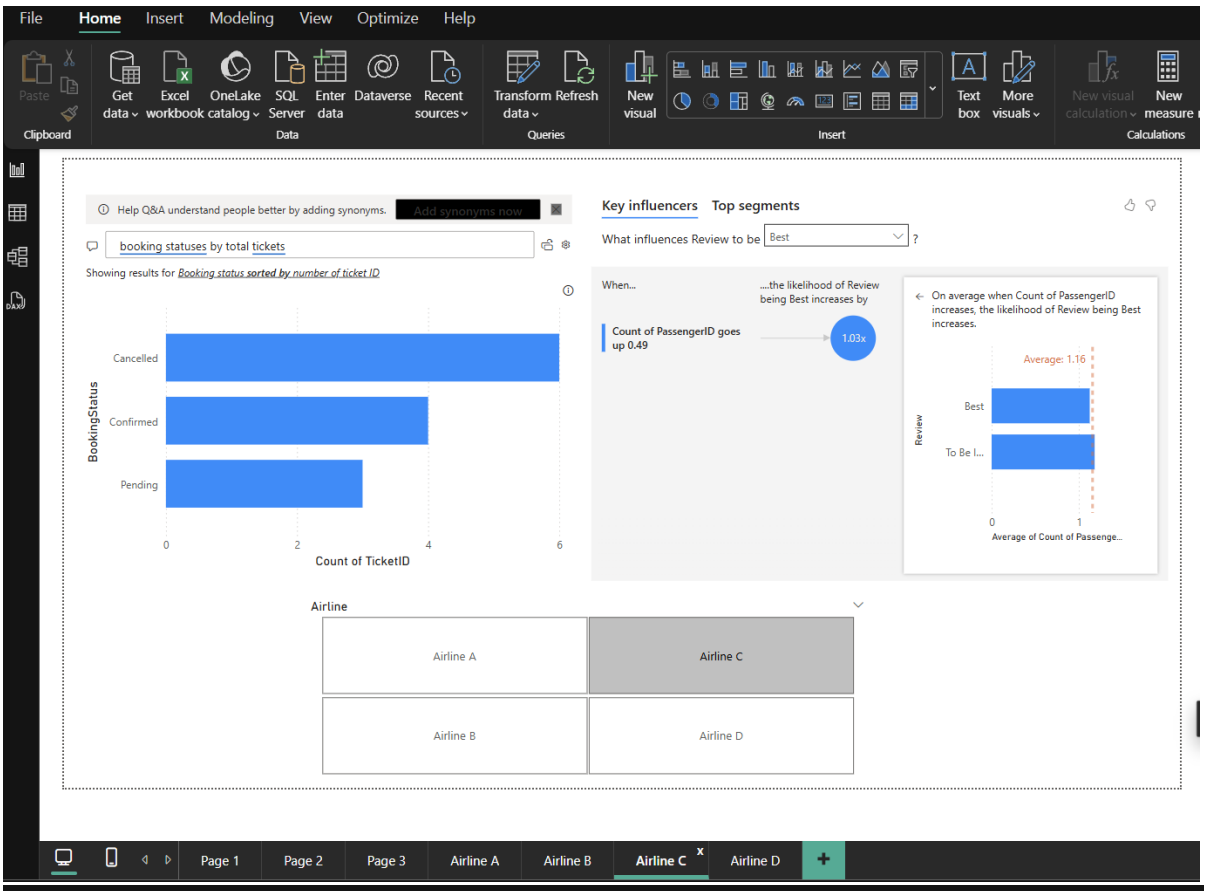
Also created bookmarks to for quick access of certain filters in the report to improve the user experience and facilitate any kind of user.

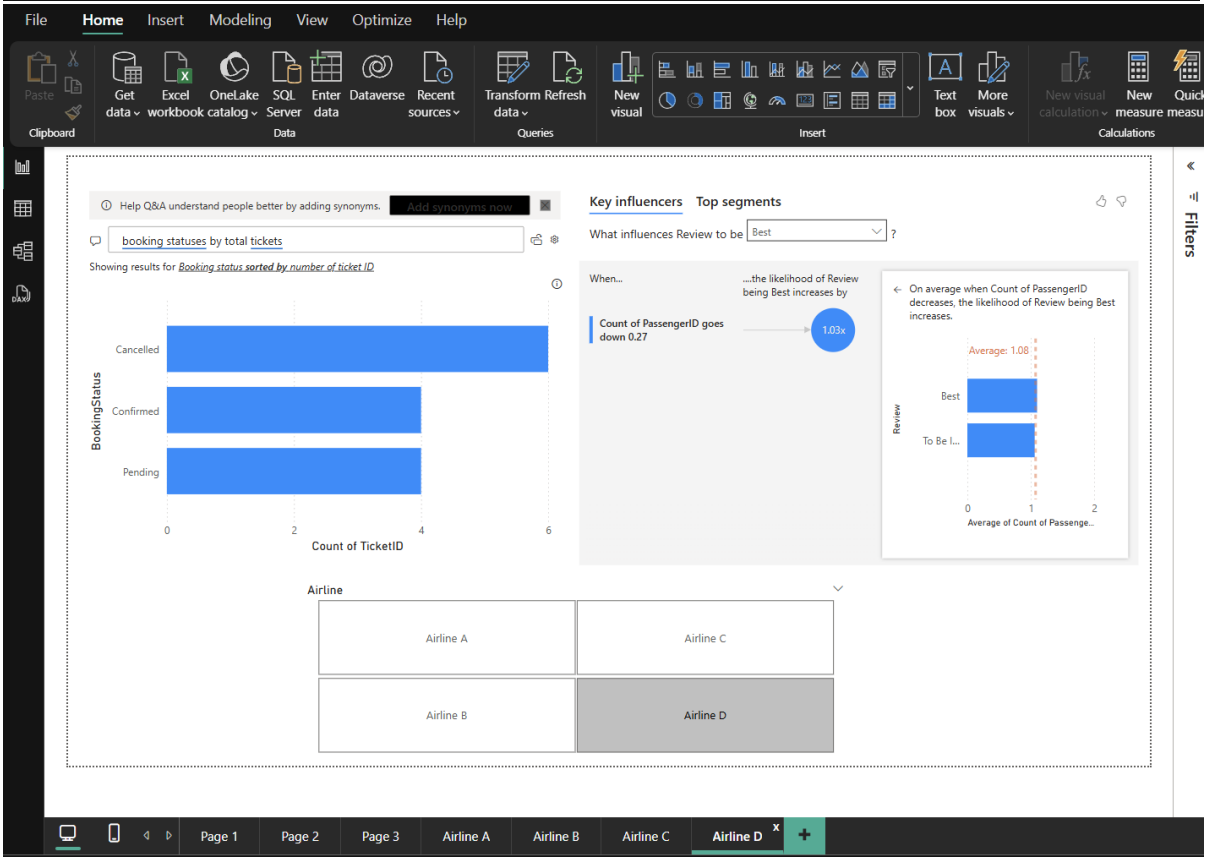


Also to improve understanding and interactivity, **I created Airline specific pages**. In these pages, **I integrated AI visuals, like Q&A and Key influencers** which improves user interaction and understanding.



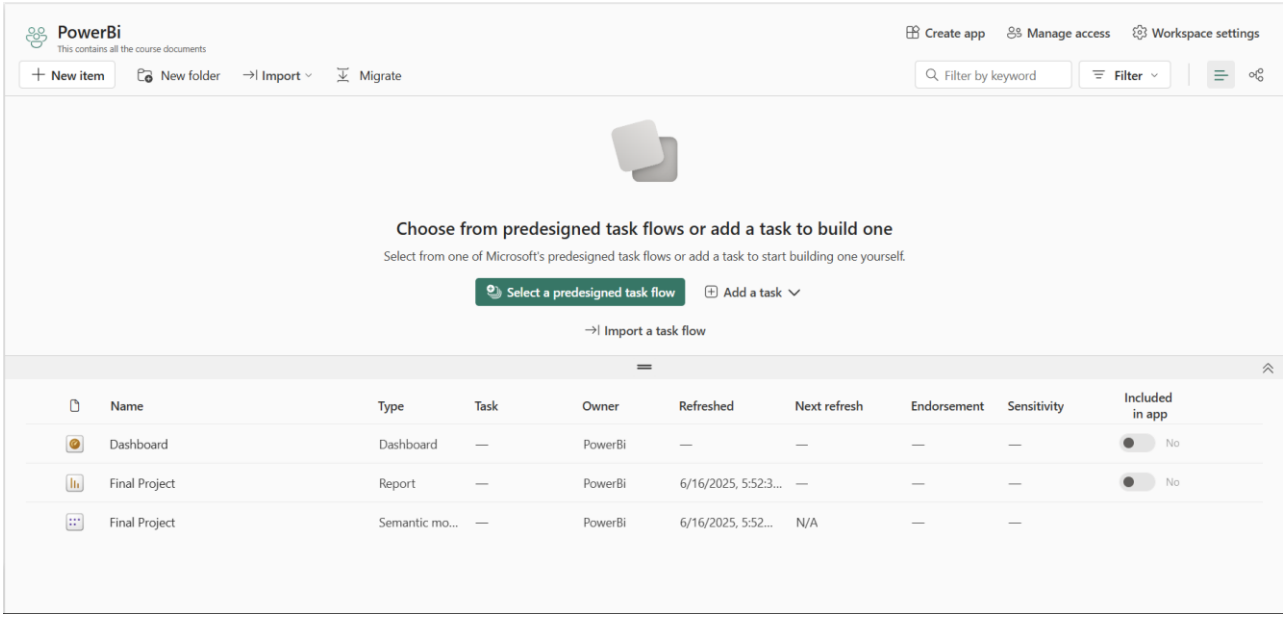



File
Home
Insert
Modeling
View
Optimize
Help
Paste
Get data
Excel workbook
OneLake catalog
SQL Server
Enter data
Dataverse
Recent sources
Transform data
Refresh
New visual
Text box
More visuals
New visual calculation
New measure
Clipboard
Data
Queries
Insert
Calculations
Help Q&A understand people better by adding synonyms.
Add synonyms now
booking statuses by total tickets
Showing results for Booking status sorted by number of ticket ID
Cancelled
Confirmed
Pending
BookingStatus
0
2
4
6
Count of TicketID
Key influencers
Top segments
What influences Review to be
Best
When...
...the likelihood of Review being Best increases by
Count of PassengerID goes up 0.49
1.03x
On average when Count of PassengerID increases, the likelihood of Review being Best increases.
Average: 1.16
Best
To Be I...
Review
0
1
Average of Count of Passenge...
Airline
Airline A
Airline C
Airline B
Airline D
Page 1
Page 2
Page 3
Airline A
Airline B
Airline C
Airline D

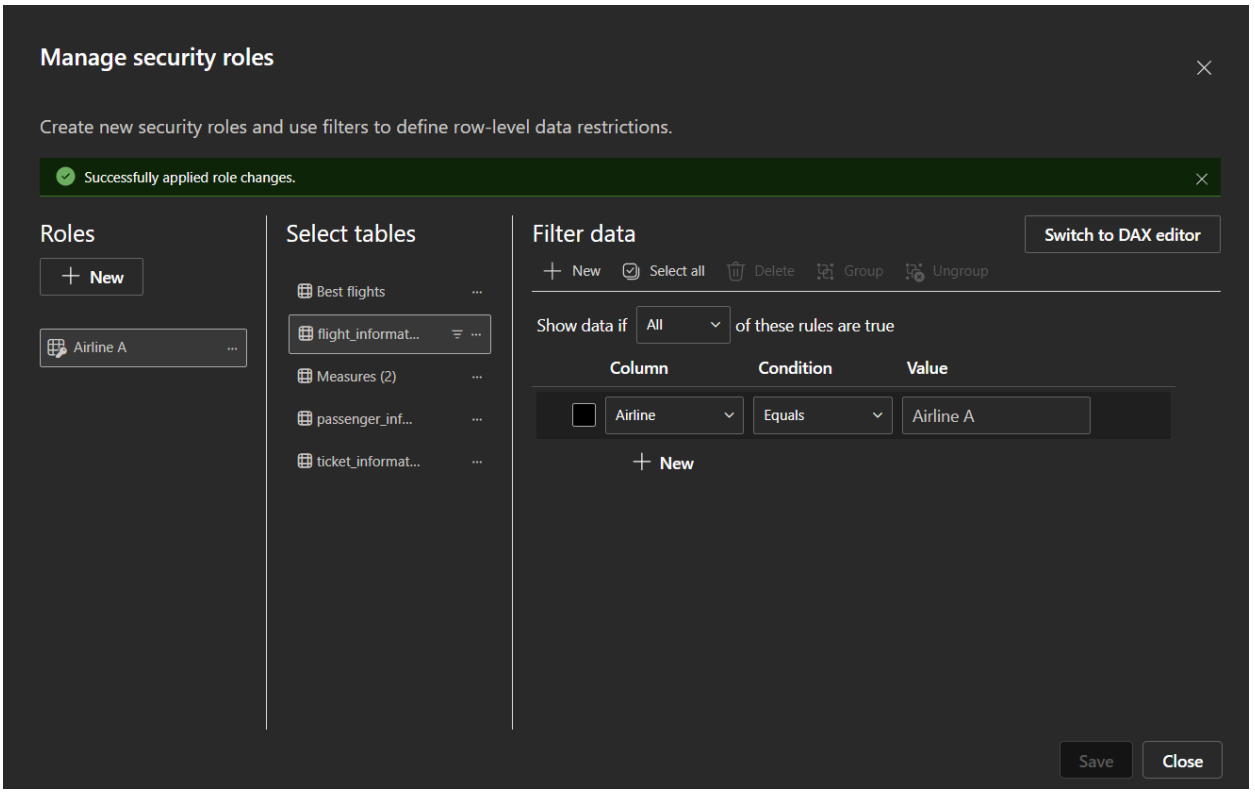

File
Home
Insert
Modeling
View
Optimize
Help
Paste
Get data
Excel workbook
OneLake catalog
SQL Server
Enter data
Dataverse
Recent sources
Transform data
Refresh
New visual
Text box
More visuals
New visual calculation
New measure
Clipboard
Data
Queries
Insert
Calculations
Help Q&A understand people better by adding synonyms.
Add synonyms now
booking statuses by total tickets
Showing results for Booking status sorted by number of ticket ID
Cancelled
Confirmed
Pending
BookingStatus
0
2
4
6
Count of TicketID
Key influencers
Top segments
What influences Review to be
Best
When...
...the likelihood of Review being Best increases by
Count of PassengerID goes down 0.27
1.03x
On average when Count of PassengerID decreases, the likelihood of Review being Best increases.
Average: 1.08
Best
To Be I...
Review
0
1
2
Average of Count of Passenge...
Airline
Airline A
Airline C
Airline B
Airline D
Filters
Page 1
Page 2
Page 3
Airline A
Airline B
Airline C
Airline D

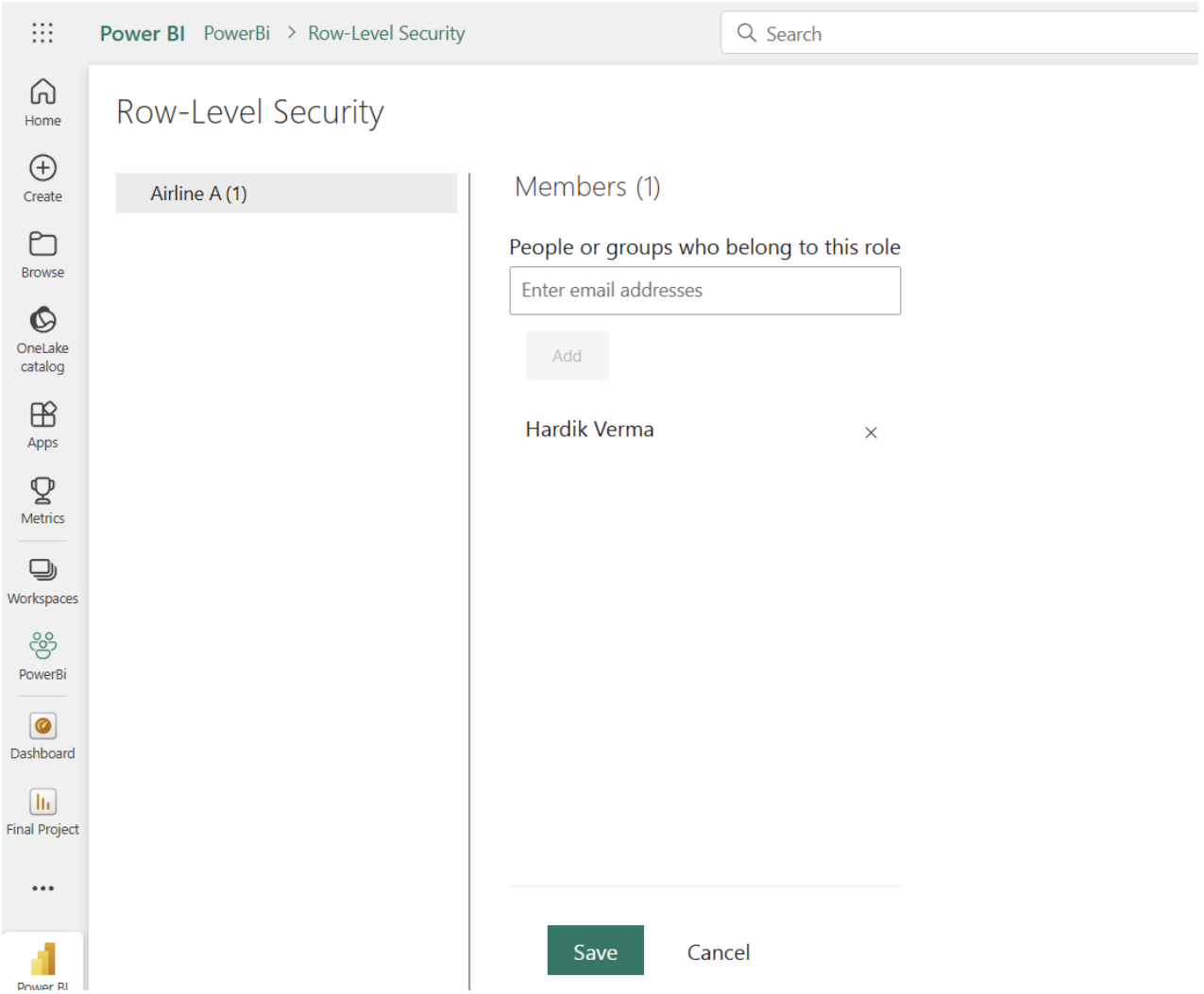
6. Final Dashboard and Power BI Service

I pinned the important visuals into the dashboard after publishing the report into a workspace **in PowerBI service.**
A dashboard can summarize the whole report in one page with the most important visuals in the report all together.



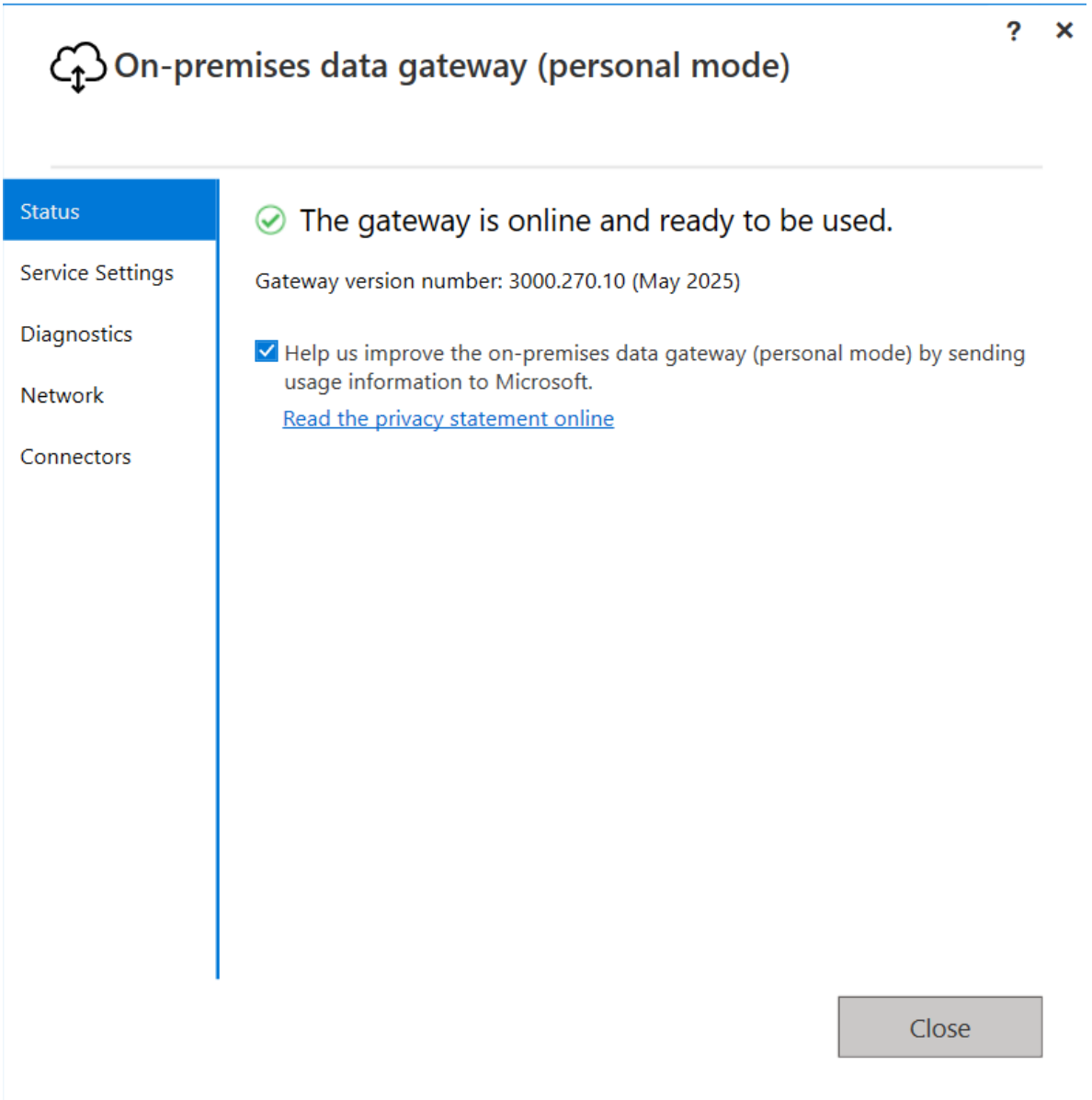
I configured a **Row-Level Security (RLS) for Airline A data** using the **Manage Roles** option in the **Security** section of the **Modeling** menu in the PowerBI Desktop.



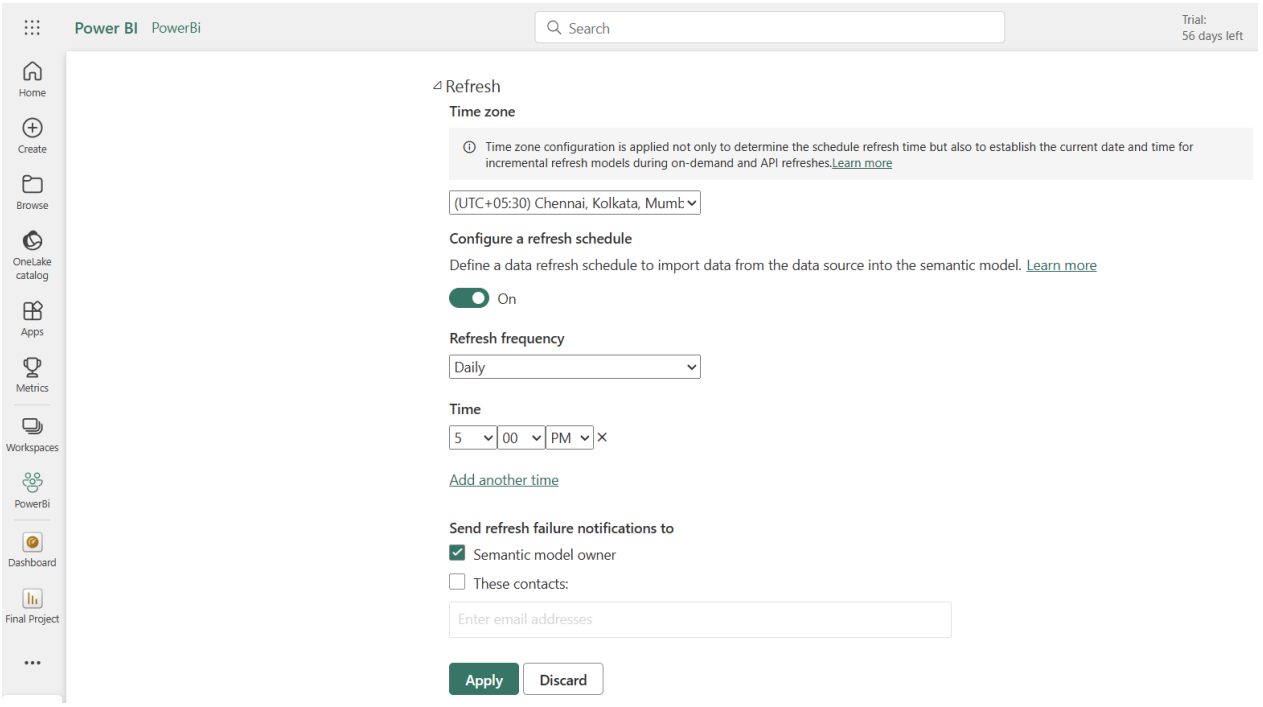


For the flight_information dataset, I created RLS on Airline A to restrict data for some users. RLS is a security feature used to restrict certain users to a filtered data usage.

I have setup the gateway from my system to the PowerBI service and also configured the data sources in the semantic model menu.



After that I scheduled a daily refresh of the datasets at 5 PM which will ensure that any changes in the data sources in the system will be updated on the PowerBI service also.



Video Explanation

https://drive.google.com/file/d/1vUbG7maPiUgKcO9TfQERTakZEy9mNrbO/view?usp=sharing