

AIRLINE DATA MANAGEMENT AND ANALYSIS

POWER BI PROJECT

ABSTRACT

To analyse and visualize airline data for operational insights, passenger management, and ticket booking trends using Power BI.

Devina M Vasudevan

DATA SCIENCE PGC JUNE 1ST BATCH

TASK 1: DATA PREPARATION AND CLEANING

PASSENGER INFORMATION

The screenshot shows the Power Query Editor interface. The main area displays a table with 100 rows and 3 columns: PassengerID, FlightID, and SeatNumber. The data is transformed using the formula: `Table.TransformColumnTypes(#"Removed Duplicates",{{"SeatNumber", type text}, {"FlightID", Int64.Type}, {"PassengerID", Int64.Type}})`. The right-hand pane shows the Query Settings for "Passenger Information - passenger_infor", with the "APPLIED STEPS" list containing: Source, Use First Row as Headers, Change Type, Removed Duplicates, and Changed Type. The status bar at the bottom indicates "3 COLUMNS, 100 ROWS" and "Column profiling based on top 1000 rows".

PassengerID	FlightID	SeatNumber
1	1	1161 38A
2	2	1157 24D
3	3	1141 30B
4	4	1046 17E
5	5	1035 29D
6	6	1134 10A
7	7	1082 10A
8	8	1115 20E
9	9	1197 34E
10	10	1047 2E
11	11	1153 43C
12	12	1194 48C
13	13	1010 47A
14	14	1056 23C
15	15	1030 16D
16	16	1109 40D
17	17	1005 25C
18	18	1119 32C
19	19	1033 27E
20	20	1118 32B
21	21	1065 19E
22	22	1146 5B
23	23	1177 28B
24	24	1011 22E
25	25	1085 6A
26	26	1026 5A
27	27	1063 12B
28	28	1086 46B

TICKET INFORMATION

The screenshot shows the Power Query Editor interface. The main area displays a table with 50 rows and 3 columns: TicketID, FlightID, and BookingStatus. The data is transformed using the formula: `Table.TransformColumnTypes(#"Removed Duplicates",{{"TicketID", Int64.Type}, {"FlightID", Int64.Type}, {"BookingStatus", type text}})`. The right-hand pane shows the Query Settings for "Ticket Information - ticket_information", with the "APPLIED STEPS" list containing: Source, Use First Row as Headers, Change Type, Removed Duplicates, and Changed Type. The status bar at the bottom indicates "3 COLUMNS, 50 ROWS" and "Column profiling based on top 1000 rows".

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
22	5022	1035 Confirmed
23	5023	1165 Confirmed
24	5024	1005 Confirmed
25	5025	1083 Cancelled
26	5026	1123 Cancelled
27	5027	1078 Confirmed
28	5028	1154 Pending

FLIGHT INFORMATION

FileHomeTransformAdd ColumnViewToolsHelp

Close & Apply

New Source

Recent Sources

Enter Data

Data source settings

Manage Parameters

Refresh Preview

Properties

Advanced Editor

Manage

Choose Columns

Remove Columns

Keep Rows

Remove Rows

Sort

Split Column

Group By

Replace Values

Data Type: Whole Number

Use First Row as Headers

Combine Files

Merge Queries

Append Queries

Text Analytics

Vision

Azure Machine Learning

AI Insights

Close

New Query

Data Sources

Parameters

Query

Manage Columns

Reduce Rows

Transform

Combine

Table.TransformColumnTypes(#"Removed Duplicates",{{"FlightID", Int64.Type}, {"FlightNumber", type text}, {"Airline", type text}, {"Destination", type text}, {"Status", type text}})

	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
22	1022	FL1308	Airline A	Houston	Delayed
23	1023	FL1769	Airline A	Chicago	On Time
24	1024	FL1343	Airline B	Chicago	Delayed
25	1025	FL1491	Airline D	Phoenix	On Time
26	1026	FL1413	Airline D	Chicago	Cancelled
27	1027	FL1805	Airline D	Chicago	On Time
28	1028	FL1385	Airline D	Chicago	On Time

Query Settings

PROPERTIES

Name

Flight_Information - flight_information

All Properties

APPLIED STEPS

Source

Use First Row as Headers

Change Type

Removed Duplicates

Changed Type

5 COLUMNS, 200 ROWS

Column profiling based on top 1000 rows

PREVIEW DOWNLOADED AT 21:11

TASK 2: DATA MODELING

RELATIONSHIP BETWEEN DATASETS:

MANY TO ONE RELATION

Edit relationship

Select tables and columns that are related.

From table
Ticket_Information - ticket_information

BookingStatus	FlightID	TicketID
Pending	1178	5001
Confirmed	1078	5002
Cancelled	1117	5003

To table
Flight_Information - flight_information

Airline	Destination	FlightID	FlightNumber	Status
Airline D	Houston	1001	FL1102	On Time
Airline B	Chicago	1002	FL1435	On Time
Airline A	Phoenix	1006	FL1071	On Time

Cardinality
Many to one (*:1)

Cross-filter direction
Single

☒ Make this relationship active

☐ Apply security filter in both directions

☐ Assume referential integrity

Save **Cancel**

Edit relationship

Select tables and columns that are related.

From table
Passenger_Information - passenger_information

FlightID	PassengerID	SeatNumber
1161	1	38A
1157	2	24D
1141	3	30B

To table
Flight_Information - flight_information

Airline	Destination	FlightID	FlightNumber	Status
Airline D	Houston	1001	FL1102	On Time
Airline B	Chicago	1002	FL1435	On Time
Airline A	Phoenix	1006	FL1071	On Time

Cardinality
Many to one (*:1)

Cross-filter direction
Single

☒ Make this relationship active

☐ Apply security filter in both directions

☐ Assume referential integrity

Save **Cancel**

TASK 3: ENHANCED DATA INSIGHTS

Two columns added:

Flight Performance with Conditional Column and Flight code with Column From Example

The screenshot displays a data transformation tool interface. The main workspace shows a table with 7 columns and 28 rows of flight data. The columns are: ID, FlightNumber, Airline, Destination, Status, Flight_Performance, and Flight_Code. The data includes flight numbers, airlines, destinations, and status updates. The right sidebar contains a 'Query Settings' panel with 'PROPERTIES' and 'APPLIED STEPS' sections. The 'APPLIED STEPS' list includes: Source, Use First Row as Headers, Change Type, Removed Duplicates, Added Conditional Column, Changed Type1, Inserted Text After Delimiter, and Renamed Columns. The bottom status bar indicates '7 COLUMNS, 200 ROWS' and 'Column profiling based on top 1000 rows'.

ID	FlightNumber	Airline	Destination	Status	Flight_Performance	Flight_Code
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
21	1021	FL1661	Airline B	New York	Cancelled	To Be Improved
22	1022	FL1308	Airline A	Houston	Delayed	To Be Improved
23	1023	FL1769	Airline A	Chicago	On Time	Best
24	1024	FL1343	Airline B	Chicago	Delayed	To Be Improved
25	1025	FL1491	Airline D	Phoenix	On Time	Best
26	1026	FL1413	Airline D	Chicago	Cancelled	To Be Improved
27	1027	FL1805	Airline D	Chicago	On Time	Best
28	1028	FL1385	Airline D	Chicago	On Time	Best

TASK 4: CALCULATIONS USING DAX

TOTAL PASSENGERS FOR A SPECIFIC FLIGHT

DAX queries will be saved to your model. They won't be visible when published in the Power BI service. [Learn more](#)

Run Update model with changes (0) [Share feedback](#)

```
7 //SUMMARIZE('Passenger_Information - passenger_information','Passenger_Information - passenger_information'[FlightID],"TOTAL
8 PASSENGERS",DISTINCTCOUNT('Passenger_Information - passenger_information'[PassengerID]))
9 --1)Total passengers for a specific flight.
10
11 EVALUATE
12 SELECTCOLUMNS('Passenger_Information - passenger_information',"FLIGHT_ID",'Passenger_Information - passenger_information'
13 [FlightID],"TOTAL_PASSENGERS",CALCULATE(DISTINCTCOUNT('Passenger_Information - passenger_information'[PassengerID])))
```

Results Result 1 of 1 [Copy](#)

	[FLIGHT ID]	[TOTAL PASSENGERS]
1	1161	1
2	1157	1
3	1141	1
4	1046	1
5	1035	1
6	1134	1
7	1082	1
8	1115	1
9	1197	1
10	1047	1
11	1153	1

Query 1 +

Success (34.3 ms) Query 1 of 1 Result 1 of 1 2 columns, 100 rows

TOTAL TICKETS BOOKED

DAX queries will be saved to your model. They won't be visible when published in the Power BI service. [Learn more](#)

Run Update model with changes (0)

```
13
14
15
16 EVALUATE
17 ROW("TOTAL TICKETS BOOKED",DISTINCTCOUNT('Ticket_Information - ticket_information'[TicketID]))
18
19
20
21
```

Results Result 1 of 1 [Copy](#)

	[TOTAL TICKETS BOOKED]
1	50

FILTERED TABLE SHOWING ALL INFORMATION OF BEST FLIGHTS ONLY

DAX queries will be saved to your model They won't be visible when published in the Power BI service. [Learn more](#)

Run Update model with changes (0)

```
--Filtered table showing "Best" flights only.
```

EVALUATE
 FILTER('Flight_Information' - flight_information', 'Flight_Information' - flight_information'[Flight_Performance]="Best")

Results Result 1 of 1 Copy

	Flight_Information - flig...	Flight_Information - flig...	Flight_Information - flig...	Flight_Information - flig...	Flight_Information - flig...	Flight_Information - flig...	Flight_Information - flig...
1	1001	FL1102	Airline D	Houston	On Time	Best	1102
2	1002	FL1435	Airline B	Chicago	On Time	Best	1435
3	1006	FL1071	Airline A	Phoenix	On Time	Best	1071
4	1011	FL1466	Airline A	Phoenix	On Time	Best	1466
5	1013	FL1330	Airline C	Houston	On Time	Best	1330
6	1020	FL1130	Airline A	New York	On Time	Best	1130
7	1023	FL1769	Airline A	Chicago	On Time	Best	1769
8	1025	FL1491	Airline D	Phoenix	On Time	Best	1491
9	1027	FL1805	Airline D	Chicago	On Time	Best	1805
10	1028	FL1385	Airline D	Chicago	On Time	Best	1385
11	1029	FL1191	Airline D	Los Angeles	On Time	Best	1191
12	1030	FL1955	Airline B	Phoenix	On Time	Best	1955
13	1031	FL1276	Airline B	New York	On Time	Best	1276

Query 1 +

Success (45.4 ms) Query 1 of 1 Result 1 of 1 7 columns, 82 rows

FILTERED TABLE SHOWING BEST FLIGHTS AND FLIGHT ID'S ONLY

DAX queries will be visible to your model. They won't be visible when published in the Power BI service. [Learn more](#)

Run

Update model with changes (0)

--Filtered table showing "Best" flights only.

EVALUATE
SELECTCOLUMNS(FILTER('Flight_Information' - flight_information', 'Flight_Information' - flight_information'[Flight_Performance] = "Best"), "FLIGHT_ID", 'Flight_Information' - flight_information'[FlightID], "FLIGHT PERFORMANCE", 'Flight_Information' - flight_information'[Flight_Performance])

Results

Result 1 of 1

Copy

	[FLIGHT ID]	[FLIGHT PERFORMANCE]
1	1001	Best
2	1002	Best
3	1006	Best
4	1011	Best
5	1013	Best
6	1020	Best
7	1023	Best
8	1025	Best
9	1027	Best
10	1028	Best
11	1029	Best
12	1030	Best
13	1031	Best

<

>

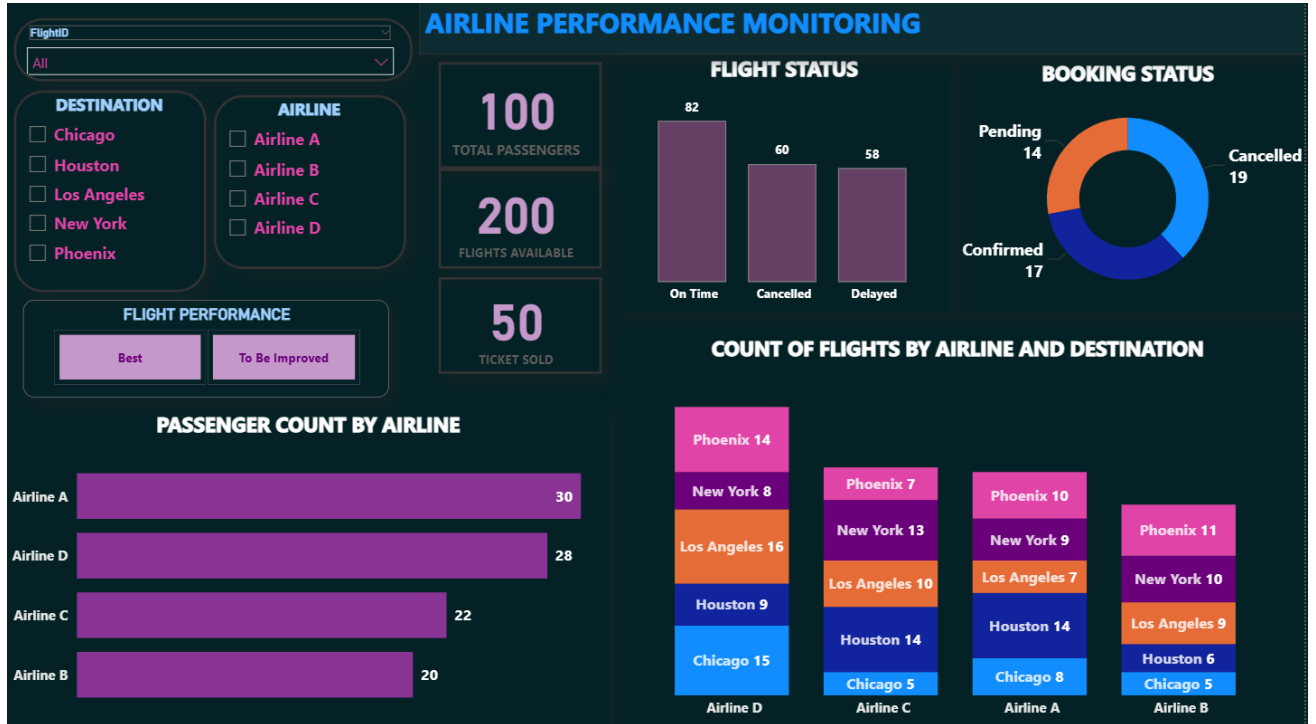
Query 1

+

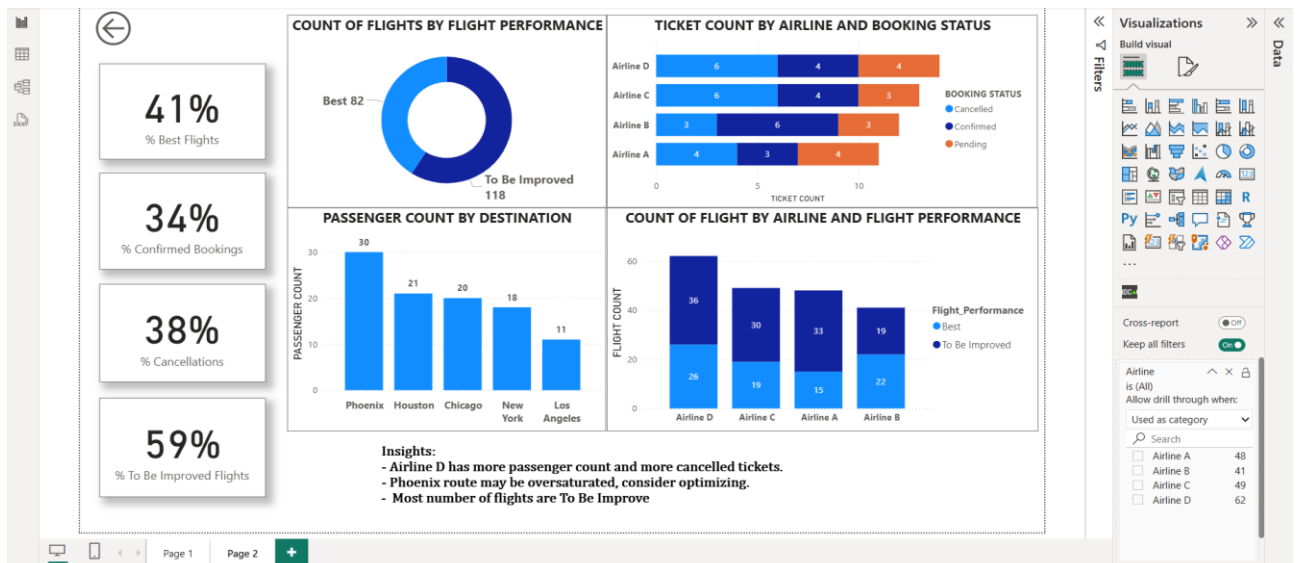
Success (37.0 ms) | Query 1 of 1 | Result 1 of 1 | 2 columns, 82 rows

TASK 5: VISUALIZATION AND INTERACTIVE FEATURES

DASHBOARD

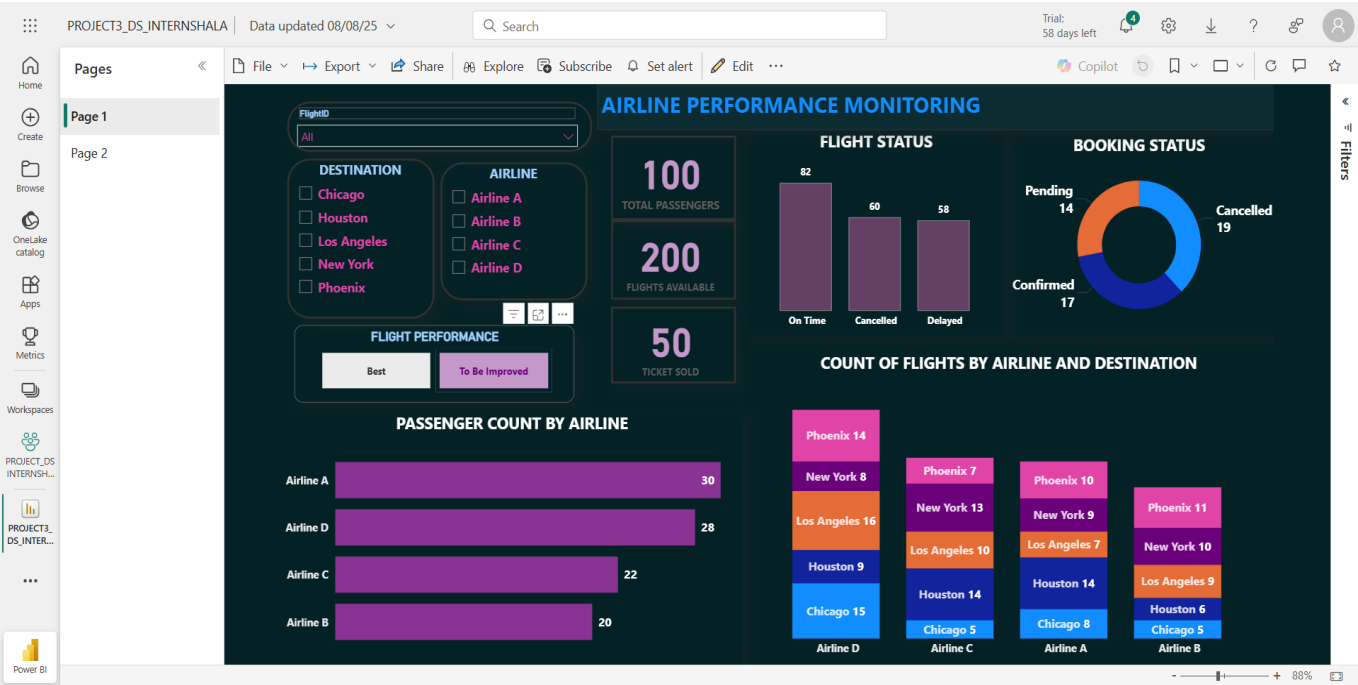


- Created Bar Chart for analysing the count of passengers by Airline.
- Created Donut Chart for analysing the Booking Status.
- Created Stacked column chart for Number of Flights by airline and destination.
- Created Stacked column chart for Flight Status.
- Created 3 Cards for 'Total Passengers', 'Total Flights Available', 'Ticket Sold'.
- Created 4 Slicers for Airline, Destination, Flight id and Flight Performance

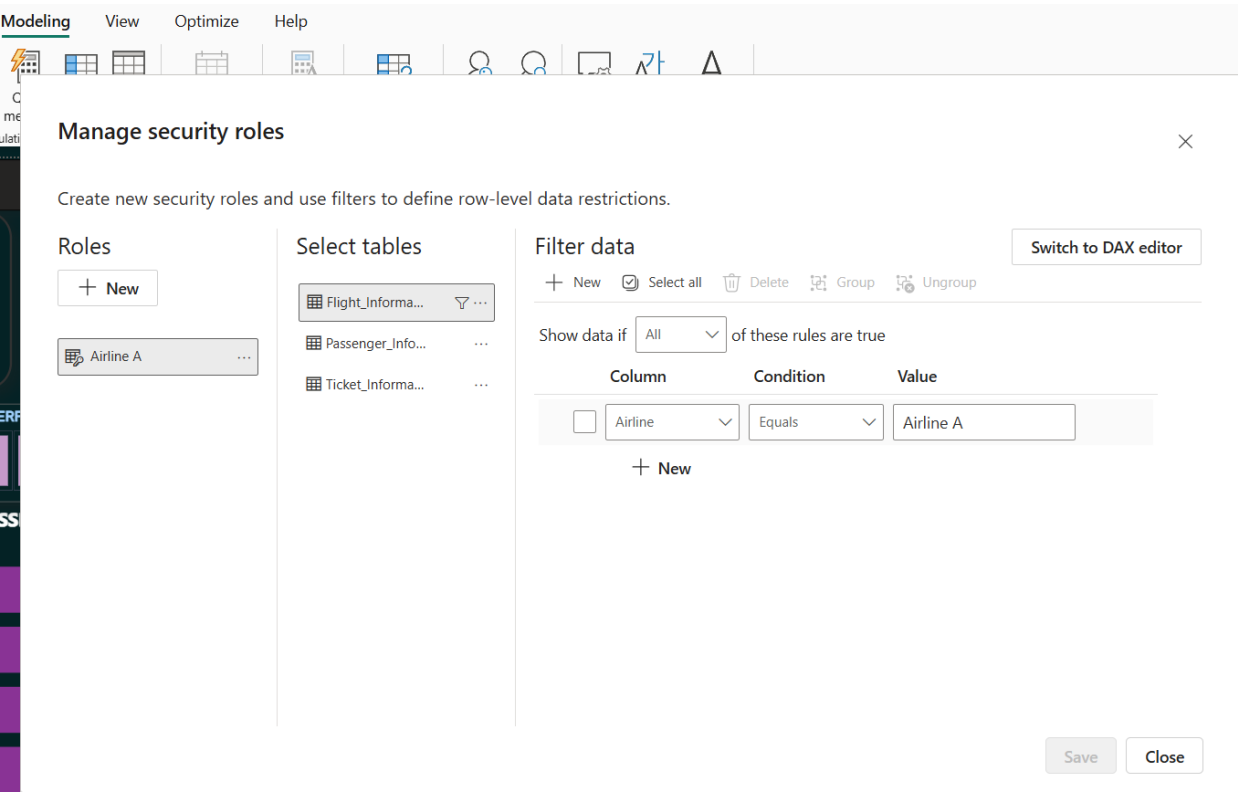


TASK 6: FINAL DASHBOARD AND POWER BI SERVICE

DESIGNED A COMPREHENSIVE DASHBOARD WITH KEY VISUALS AND INSIGHTS IN POWER BI SERVICE



CONFIGURED ROW-LEVEL SECURITY FOR AIRLINE A DATA AND ASSIGN IT TO A USER:



SET UP ROW LEVEL SECURITY IN POWER BI SERVICE:

Row-Level Security

Airline A (1)

Members (1)

People or groups who belong to this role

Enter email addresses

Add

Devina vas ×

Save

Cancel

SET UP A SCHEDULE REFRESH AT 5 PM DAILY:

▾ Data source credentials

Flight_Information - flight_information.csv	Edit credentials	Show in lineage view
Passenger_Information - passenger_information.csv	Edit credentials	Show in lineage view
Ticket_Information - ticket_information.csv	Edit credentials	Show in lineage view

▾ Parameters

▾ Refresh

Time zone

ⓘ Time zone configuration is applied not only to determine the schedule refresh time but also to establish the current date and time for incremental refresh models during on-demand and API refreshes. [Learn more](#)

(UTC) Coordinated Universal Time ▾

Configure a refresh schedule

Define a data refresh schedule to import data from the data source into the semantic model. [Learn more](#)

☒ On

Refresh frequency

Daily ▾

Time

5 ▾ 00 ▾ PM ▾ ×

[Add another time](#)

Send refresh failure notifications to

☒ Semantic model owner

☒ These contacts:

Devina vas ×

CONCLUSION

- **Flight Performance:** Only 41% of flights are marked as 'Best', with the majority needing improvement.
- **Booking Status:** A significant portion of tickets (38%) are cancelled, and only 34% are confirmed, highlighting inefficiencies in the booking process. The ticket booking process needs enhancement to reduce cancellations.
- **Passenger Trends:** The route to Phoenix shows maximum passenger traffic and flight frequency, requiring better traffic distribution.
- **Airline D** has a greater number of passengers and flights but also records more cancellations and underperformance, requiring urgent system review.
- The data reveals an irregularity, where **ticket purchase** figures are either higher or lower than the actual number of passengers traveling.

- THANK YOU -