

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 with the "Passenger_Information" query selected. The query preview shows a table with three columns: PassengerID, FlightID, and SeatNumber. The "Applied Steps" pane on the right indicates that the "Removed Duplicates" step was applied.

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 58
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 with the "Ticket_Information" query selected. The query preview shows a table with three columns: TicketID, FlightID, and BookingStatus. The "Applied Steps" pane on the right indicates that the "Removed Duplicates" step was applied.

TicketID	FlightID	BookingStatus
5001	1178	Pending
5002	1078	Confirmed
5003	1117	Cancelled
5004	1120	Cancelled
5005	1137	Cancelled
5006	1162	Pending
5007	1076	Pending
5008	1035	Cancelled
5009	1001	Cancelled
5010	1040	Cancelled
5011	1064	Pending
5012	1150	Cancelled
5013	1060	Cancelled
5014	1064	Confirmed
5015	1093	Confirmed
5016	1072	Pending
5017	1011	Cancelled
5018	1105	Cancelled
5019	1014	Confirmed
5020	1060	Pending
5021	1030	Confirmed
5022	1035	Confirmed
5023	1165	Confirmed
5024	1005	Confirmed
5025	1083	Cancelled
5026	1123	Cancelled
5027	1078	Confirmed
5028	1154	Pending

FLIGHT INFORMATION

Untitled - Power Query Editor

File Home Transform Add Column View Tools Help

Close & Apply Data source settings Parameters Refresh Preview Manage Advanced Editor Properties Choose Columns Remove Columns Keep Rows Remove Rows Split Column Group By Replace Values Data Type: Whole Number Use First Row as Headers Merge Queries Append Queries Text Analytics Vision Combine Files Azure Machine Learning Close New Query Data Sources Parameters Query Manage Columns Sort Transform Combine AI Insights

Queries [3]

FlightID	FlightNumber	Airline	Destination	Status
1	1001	Airline D	Houston	On Time
2	1002	Airline B	Chicago	On Time
3	1003	Airline A	New York	Cancelled
4	1004	Airline C	Chicago	Delayed
5	1005	Airline C	New York	Delayed
6	1006	Airline A	Phoenix	On Time
7	1007	Airline C	Los Angeles	Cancelled
8	1008	Airline C	Los Angeles	Delayed
9	1009	Airline A	Los Angeles	Cancelled
10	1010	Airline D	Chicago	Cancelled
11	1011	Airline A	Phoenix	On Time
12	1012	Airline D	New York	Delayed
13	1013	Airline C	Houston	On Time
14	1014	Airline C	New York	Delayed
15	1015	Airline C	Houston	Delayed
16	1016	Airline B	New York	Delayed
17	1017	Airline D	Phoenix	Delayed
18	1018	Airline B	Houston	Delayed
19	1019	Airline B	Chicago	Cancelled
20	1020	Airline A	New York	On Time
21	1021	Airline B	New York	Cancelled
22	1022	Airline A	Houston	Delayed
23	1023	Airline A	Chicago	On Time
24	1024	Airline B	Chicago	Delayed
25	1025	Airline D	Phoenix	On Time
26	1026	Airline D	Chicago	Cancelled
27	1027	Airline D	Chicago	On Time
28	1028	Airline D	Chicago	On Time

5 COLUMNS, 200 ROWS Column profiling based on top 1000 rows

PREVIEW DOWNLOADED AT 21:11

Query Settings

- PROPERTIES**
 - Name: Flight_Information - flight_information
 - All Properties
- APPLIED STEPS**
 - Source
 - Use First Row as Headers
 - Change Type
 - Removed Duplicates
 - Changed Type

TASK 2: DATA MODELING

RELATIONSHIP BETWEEN DATASETS:

MANY TO ONE RELATION

The screenshot shows the Power BI Data Model Editor interface. At the top, there's a ribbon with various data sources and tools. Below the ribbon, two tables are displayed: 'Ticket_Information' and 'Flight_Information'. A many-to-one relationship is established between them, with 'TicketID' in 'Ticket_Information' pointing to 'FlightID' in 'Flight_Information'. The 'Flight_Information' table has a primary key constraint on 'FlightID'. On the right side, the 'Edit relationship' dialog is open, showing the selected tables and columns for the relationship. It also includes settings for cardinality ('Many to one (*:1)'), cross-filter direction ('Single'), and security options.

Edit relationship

Select tables and columns that are related.

From table

Passenger_Information - passenger_informati... ▾

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) ▾ Single

Make this relationship active

Assume referential integrity

The screenshot shows the Power BI Data Model Editor interface. At the top, there's a ribbon with various data model management tools. Below the ribbon, two tables are displayed: 'Passenger_Information' and 'Flight_Information'. A many-to-one relationship is established between them, with 'FlightID' in 'Flight_Information' pointing to 'FlightID' in 'Passenger_Information'. The 'Passenger_Information' table has a primary key constraint on 'FlightID'. On the right side, the 'Edit relationship' dialog is open, showing the selected tables and columns for the relationship. It also includes settings for cardinality ('Many to one (*:1)'), cross-filter direction ('Single'), and security options.

TASK 3: ENHANCED DATA INSIGHTS

Two columns added:

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

The screenshot shows the Microsoft Power BI Data Editor interface. The top ribbon has tabs like File, Home, Transform, Add Column, View, Tools, and Help. The main area displays a table titled "Table.RenameColumns#" with 200 rows and 7 columns. The columns are: FlightNumber, Airline, Destination, Status, Flight_Performance, and Flight_Code. The Flight_Code column contains values like "Best", "To Be Improved", and "Delayed". The status bar at the bottom indicates "7 COLUMNS, 200 ROWS" and "Column profiling based on top 1000 rows". On the right side, there are two panes: "Query Settings" and "APPLIED STEPS". The "APPLIED STEPS" pane lists several steps: "Source", "Use First Row as Headers", "Change Type", "Removed Duplicates", "Changed Type", "Added Conditional Column", "Changed Type1", and "Inserted Text After Delimiter". The "Renamed Columns" step is highlighted.

FlightNumber	Airline	Destination	Status	Flight_Performance	Flight_Code
1001 FL1102	Airline D	Houston	On Time	Best	1102
1002 FL1435	Airline B	Chicago	On Time	Best	1435
1003 FL1860	Airline A	New York	Cancelled	To Be Improved	1860
1004 FL1270	Airline C	Chicago	Delayed	To Be Improved	1270
1005 FL1106	Airline C	New York	Delayed	To Be Improved	1106
1006 FL1071	Airline A	Phoenix	On Time	Best	1071
1007 FL1700	Airline C	Los Angeles	Cancelled	To Be Improved	1700
1008 FL1020	Airline C	Los Angeles	Delayed	To Be Improved	1020
1009 FL1614	Airline A	Los Angeles	Cancelled	To Be Improved	1614
1010 FL1121	Airline D	Chicago	Cancelled	To Be Improved	1121
1011 FL1466	Airline A	Phoenix	On Time	Best	1466
1012 FL1214	Airline D	New York	Delayed	To Be Improved	1214
1013 FL1330	Airline C	Houston	On Time	Best	1330
1014 FL1458	Airline C	New York	Delayed	To Be Improved	1458
1015 FL1087	Airline C	Houston	Delayed	To Be Improved	1087
1016 FL1372	Airline B	New York	Delayed	To Be Improved	1372
1017 FL1099	Airline D	Phoenix	Delayed	To Be Improved	1099
1018 FL1871	Airline B	Houston	Delayed	To Be Improved	1871
1019 FL1663	Airline B	Chicago	Cancelled	To Be Improved	1663
1020 FL1130	Airline A	New York	On Time	Best	1130
1021 FL1661	Airline B	New York	Cancelled	To Be Improved	1661
1022 FL1308	Airline A	Houston	Delayed	To Be Improved	1308
1023 FL1769	Airline A	Chicago	On Time	Best	1769
1024 FL1343	Airline B	Chicago	Delayed	To Be Improved	1343
1025 FL1491	Airline D	Phoenix	On Time	Best	1491
1026 FL1413	Airline D	Chicago	Cancelled	To Be Improved	1413
1027 FL1805	Airline D	Chicago	On Time	Best	1805
1028 FL1295	Airline D	Chicago	On Time	Best	1295

TASK 4: CALCULATIONS USING DAX

TOTAL PASSENGERS FOR A SPECIFIC FLIGHT

The screenshot shows the Power BI DAX Editor interface. At the top, there are icons for 'Model', 'Table', and 'DAX'. Below the table, there are buttons for 'Run' and 'Update model with changes (0)'. A 'Share feedback' button is also present. The DAX code is as follows:

```
7 //SUMMARIZE('Passenger_Information - passenger_information','Passenger_Information - passenger_information'[FlightID],"TOTAL  
8 PASSENGERS",DISTINCTCOUNT('Passenger_Information - passenger_information'[PassengerID]))  
9  
10 --1)Total passengers for a specific flight.  
11  
12 EVALUATE  
13 SELECTCOLUMNS('Passenger_Information - passenger_information',"FLIGHT ID",'Passenger_Information - passenger_information'  
[FlightID],"TOTAL PASSENGERS",CALCULATE(DISTINCTCOUNT('Passenger_Information - passenger_information'[PassengerID])))  
14  
15
```

The results table shows 11 rows of data with columns [FLIGHT ID] and [TOTAL PASSENGERS]. The data is as follows:

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

At the bottom, there is a success message: Success (34.3 ms) Query 1 of 1 Result 1 of 1 2 columns, 100 rows.

TOTAL TICKETS BOOKED

The screenshot shows the Power BI DAX Editor interface. At the top, there are icons for 'Model', 'Table', and 'DAX'. Below the table, there are buttons for 'Run' and 'Update model with changes (0)'. A 'Share feedback' button is also present. The DAX code is as follows:

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

The results table shows 1 row of data with column [TOTAL TICKETS BOOKED]. The data is as follows:

[TOTAL TICKETS BOOKED]
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)

```
24  
25 --Filtered table showing "Best" flights only.  
26  
27 EVALUATE  
28 FILTER('Flight_Information - flight_information','Flight_Information - flight_information'[Flight_Performance]="Best")  
29  
30
```

Results | Result 1 of 1 | Copy ▾

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

Run Update model with changes (0)

```
18  
19 --Filtered table showing "Best" flights only.  
20  
21 EVALUATE  
22 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 ▾

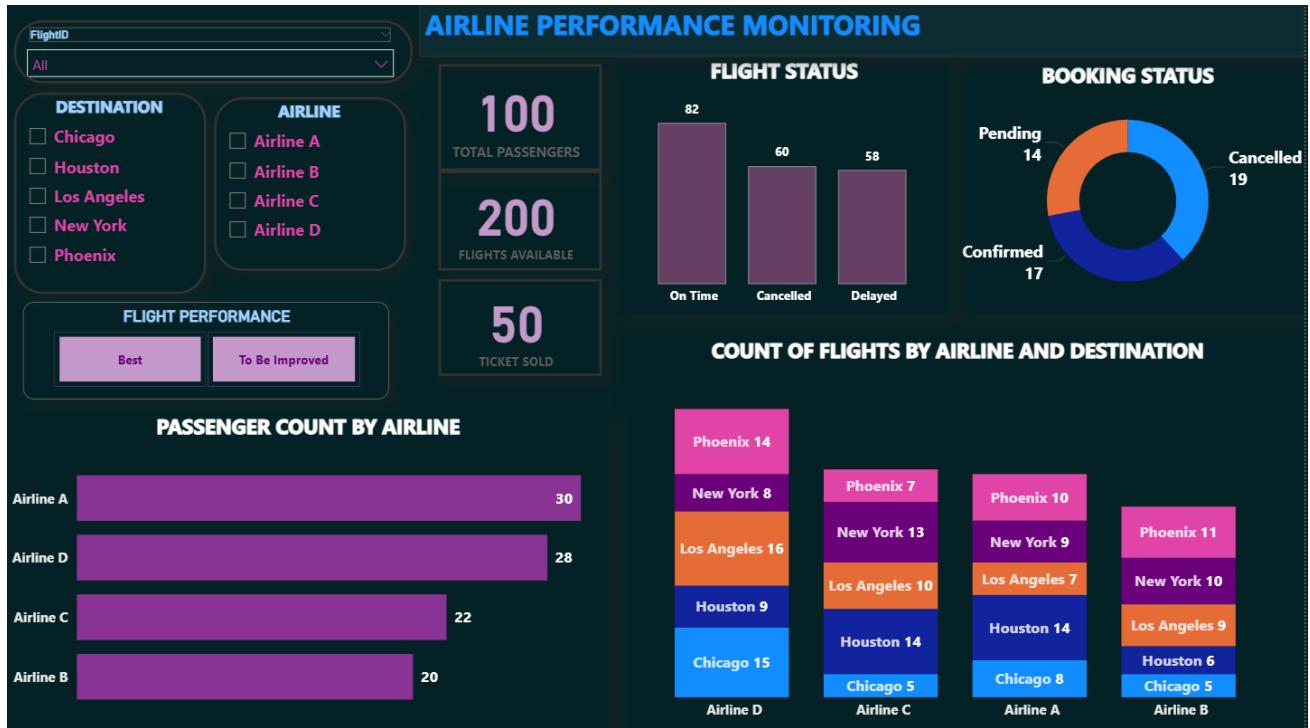
	[FLIGHT ID]	[FLIGHT PERFORMANCE]
1	1001	Best
2	1002	Best
3	1006	Best
4	1011	Best
5	1013	Best
6	1020	Best
7	1023	Best
8	1025	Best
9	1027	Best
10	1028	Best
11	1029	Best
12	1030	Best
13	1031	Best

Query 1 +

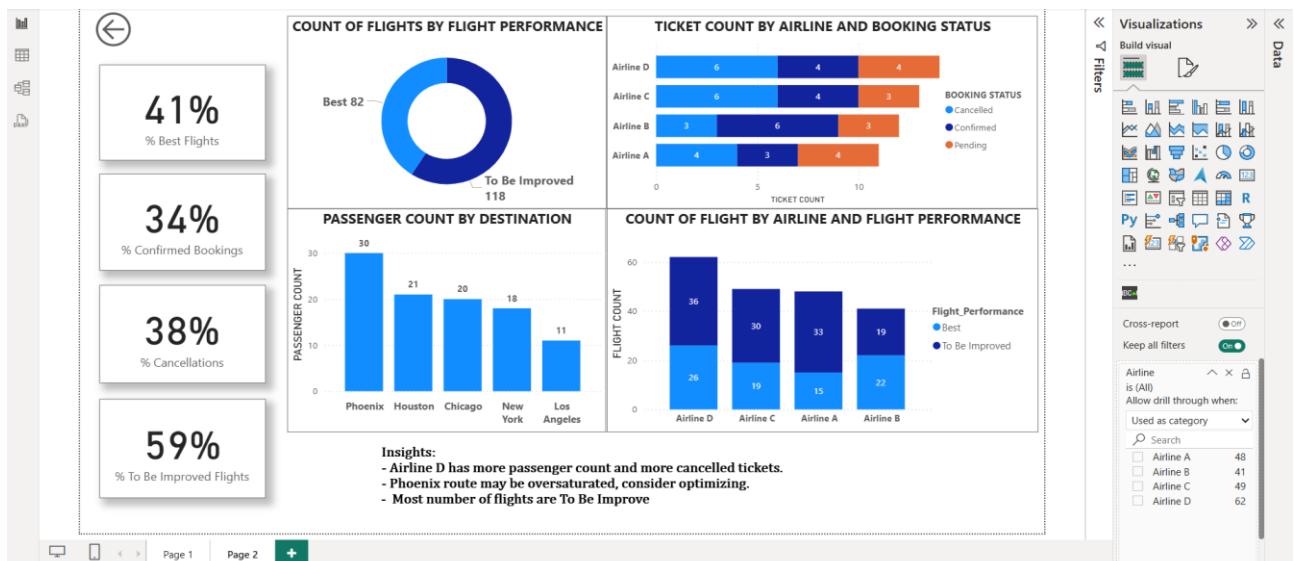
Success (37.0 ms) | Query 1 of 1 | Result 1 of 1 | 2 columns, 82 rows

TASK 5: VISUALIZATION AND INTERACTIVE FEATURES

DASHBOARD

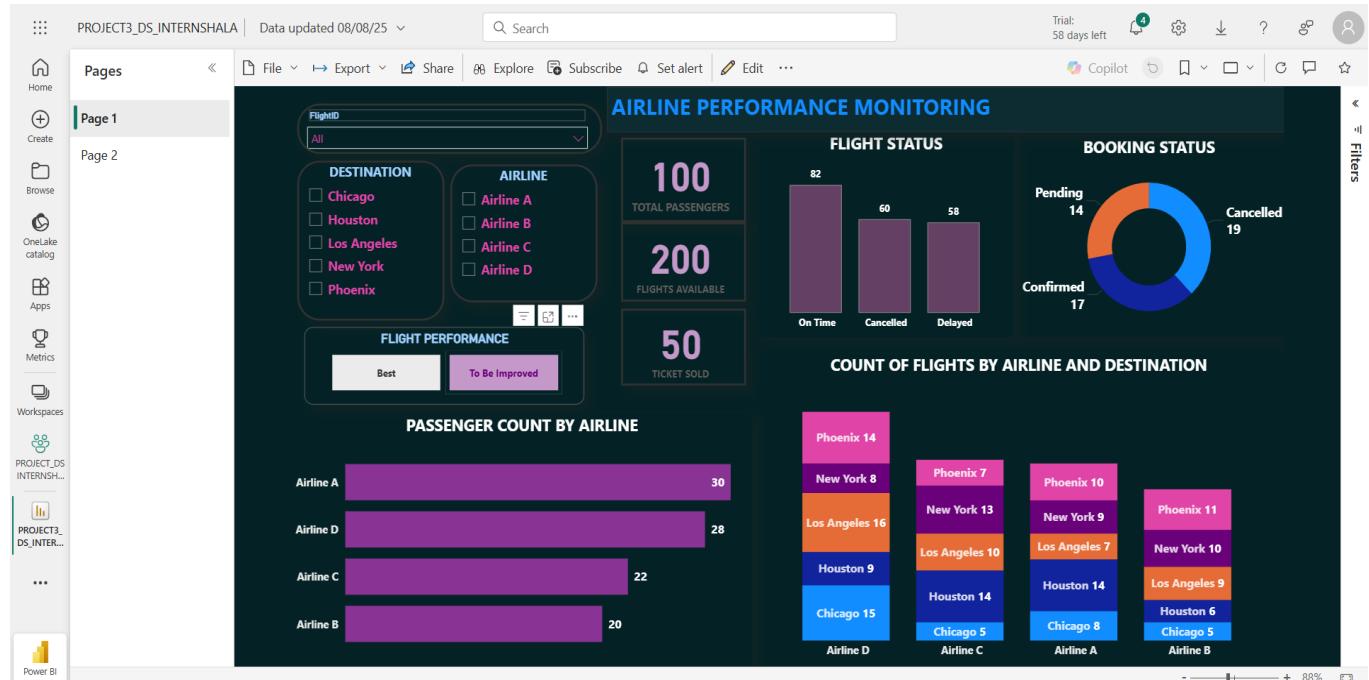


- Created Bar Chart for analysing the count of passengers by Airline.
- Created Donut Chart for analysing the Booking Status.
- Created Stacked column chart for Number of Flights by airline and destination.
- Created Stacked column chart for Flight Status.
- Created 3 Cards for 'Total Passengers', 'Total 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:

The 'Manage security roles' screen shows the following configuration:

- Roles:** Airline A
- Select tables:** Flight_Information, Passenger_Information, Ticket_Information
- Filter data:** Show data if **All** of these rules are true:

Column	Condition	Value
Airline	Equals	Airline A

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 x

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 ▾ x

[Add another time](#)

Send refresh failure notifications to

Semantic model owner
 These contacts:

Devina vas x

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 -