

# Airline Data Management and Analysis Using Power BI

## 1. Problem Statement

The airline industry has many moving parts — flight schedules, passenger details, and ticket bookings. Managing all this data efficiently is important to improve operations and customer satisfaction.

This project uses Power BI to connect, clean, and analyze airline data to provide useful insights for better decision-making.

## 2. Datasets Used

1. [Flight\\_Information.xlsx](#) – FlightID, FlightNumber, Airline, Destination, Status.
2. [Passenger\\_Information.xlsx](#) – PassengerID, FlightID, SeatNumber.
3. [Ticket\\_Information.xlsx](#) – TicketID, FlightID, BookingStatus.

## 3.Task-wise Execution

### Task 1 – Data Preparation and Cleaning

Loaded all three Excel files into Power BI and cleaned them in Power Query by:

- Removing duplicates and empty rows.
- Fixing data types (ensuring FlightID matched across all tables).
- Trimming and cleaning text columns.

Table: Table.Distinct(#"Removed Other Columns", {"FlightID"})

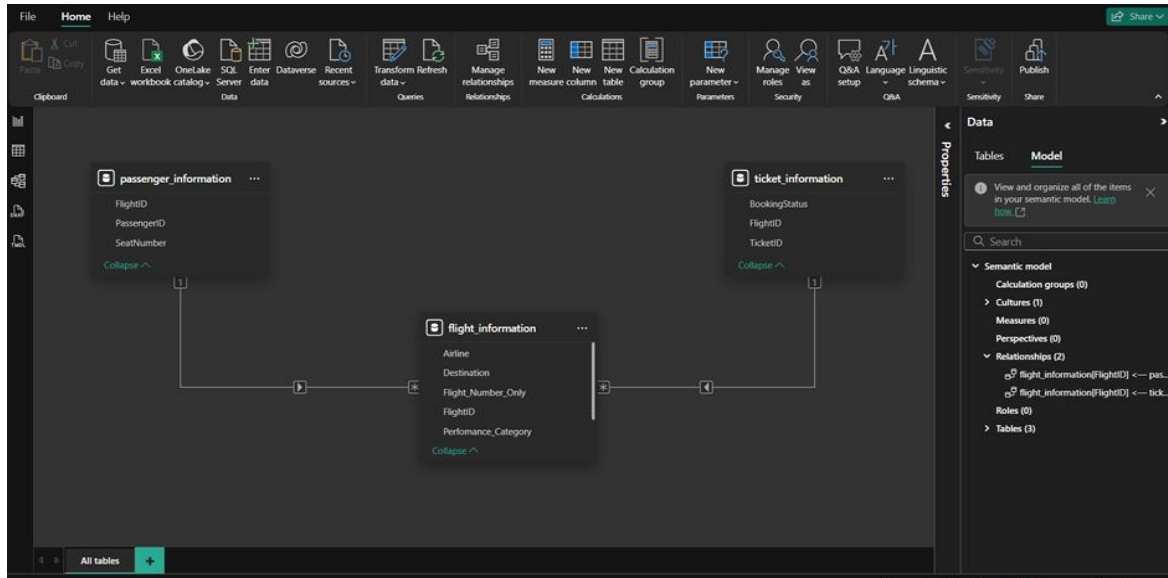
	FlightID	FlightNumber	Airline	Destination	Status
1	1001	FL1102	Airline D	Houston	On Time
2	1002	FL1435	Airline B	Chicago	On Time
3	1003	FL1860	Airline A	New York	Cancelled
4	1004	FL1270	Airline C	Chicago	Delayed
5	1005	FL1106	Airline C	New York	Delayed
6	1006	FL1071	Airline A	Phoenix	On Time
7	1007	FL1700	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

5 COLUMNS, 200 ROWS Column profiling based on top 1000 rows PREVIEW DOWNLOADED AT 04:05

## Task 2 – Data Modeling

Linked the datasets using **FlightID** as the primary key:

- Flights → Passengers (One-to-Many)
- Flights → Tickets (One-to-Many)



## Task 3 – Enhanced Data Insights

1. **Performance Category:** Created a conditional column:
  - If Status = “On Time” or “Arrived” → “Best”
  - Else → “To Be Improved”
2. **Flight Number Only:** Extracted just the number from FlightNumber using Column from Examples.

	flight_id	flight_number_only	airline	destination	status	performance_category
1	1001 1102	1102	Airline D	Houston	On Time	Best
2	1002 1435	1435	Airline B	Chicago	On Time	Best
3	1003 1860	1860	Airline A	New York	Cancelled	To Be Improved
4	1004 1270	1270	Airline C	Chicago	Delayed	To Be Improved
5	1005 1306	1306	Airline C	New York	Delayed	To Be Improved
6	1006 1071	1071	Airline A	Phoenix	On Time	Best
7	1007 1700	1700	Airline C	Los Angeles	Cancelled	To Be Improved
8	1008 1020	1020	Airline C	Los Angeles	Delayed	To Be Improved
9	1009 1614	1614	Airline A	Los Angeles	Cancelled	To Be Improved
10	1010 1121	1121	Airline D	Chicago	Cancelled	To Be Improved
11	1011 1466	1466	Airline A	Phoenix	On Time	Best
12	1012 1214	1214	Airline D	New York	Delayed	To Be Improved
13	1013 1330	1330	Airline C	Houston	On Time	Best
14	1014 1458	1458	Airline C	New York	Delayed	To Be Improved
15	1015 1087	1087	Airline C	Houston	Delayed	To Be Improved
16	1016 1372	1372	Airline B	New York	Delayed	To Be Improved
17	1017 1099	1099	Airline D	Phoenix	Delayed	To Be Improved
18	1018 1871	1871	Airline B	Houston	Delayed	To Be Improved
19	1019 1663	1663	Airline B	Chicago	Cancelled	To Be Improved
20	1020 1130	1130	Airline A	New York	On Time	Best
21	1021 1661	1661	Airline B	New York	Cancelled	To Be Improved
22	1022 1308	1308	Airline A	Houston	Delayed	To Be Improved
23	1023 1789	1789	Airline A	Chicago	On Time	Best
24	1024 1343	1343	Airline B	Chicago	Delayed	To Be Improved

## Task 4 – DAX Calculations

Created measures in DAX view to calculate:

- **Total Passengers** → `DISTINCTCOUNT(Passengers[PassengerID])`
- **Total Tickets** → `COUNTROWS(Tickets)`
- **Best Flights Table** → `FILTER(Flights, Flights[Performance_category] = "Best")`

The screenshot shows the Microsoft Power BI Desktop interface. The top ribbon includes 'File', 'Home', and 'Help'. The 'Home' ribbon has tabs for 'Clipboard' and 'Editing'. The 'Editing' tab is active, showing options like 'Format', 'Comment', 'Uncomment', 'Find', 'Replace', 'Command palette', and 'Copilot (preview)'. A notification bar at the top states: 'DAX queries will be saved to your model. They won't be visible when published in the Power BI service. [Learn more.](#)'

The main area displays DAX queries in the 'DAX view' tab. The queries are as follows:

```
1 // 1.Total passengers for a specific flight.
2 EVALUATE
3 SUMMARIZE('Passengers','Passengers'[FlightID],"Total Passengers", SUM(Passengers[PassengerID]))
4
5 // 2.Total Tickets Booked
6 EVALUATE
7 SUM('Total Tickets',COUNTROWS(Tickets))
8
9 // 3.Filtered table showing "Best" flights only.
10 EVALUATE
11 FILTER(Flights,Flights[Performance_Category] = "Best")
```

The 'Results' pane shows 'Result 1 of 1' with a 'Copy' button. The data table has 7 columns: 'Flights[FlightID]', 'Flights[FlightNumber]', 'Flights[Airline]', 'Flights[Destination]', 'Flights[Status]', 'Flights[Performance\_Cat...]', and 'Flights[Flight\_number\_o...]'.

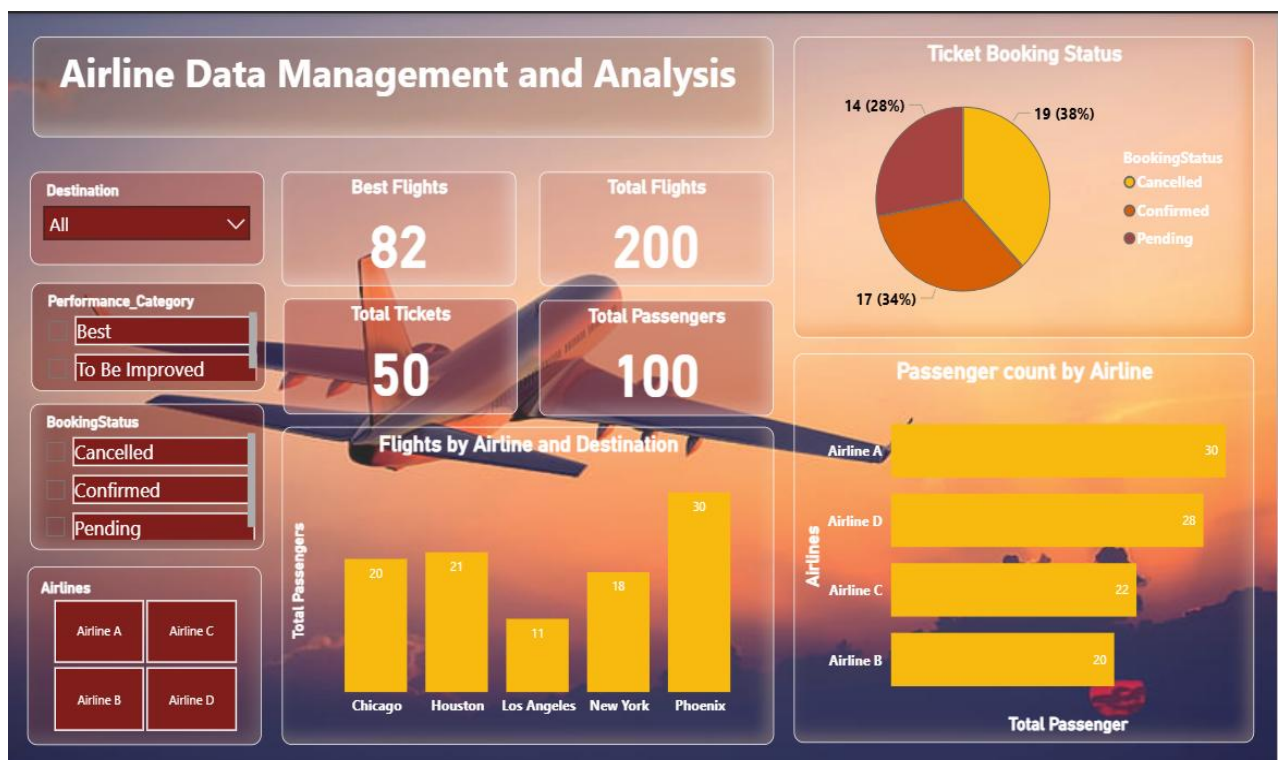
	Flights[FlightID]	Flights[FlightNumber]	Flights[Airline]	Flights[Destination]	Flights[Status]	Flights[Performance_Cat...]	Flights[Flight_number_o...]
1	1001	FL1102	Airline D	Houston	On Time	Best	1102
2	1002	FL1435	Airline B	Chicago	On Time	Best	1435
3	1006	FL1071	Airline A	Phoenix	On Time	Best	1071
4	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

The bottom status bar shows 'Success (359.5 ms)', 'Query 1 of 1', 'Result 1 of 1', '7 columns, 82 rows', and a progress indicator at 100%.

## Task 5 – Visualizations and Interactivity

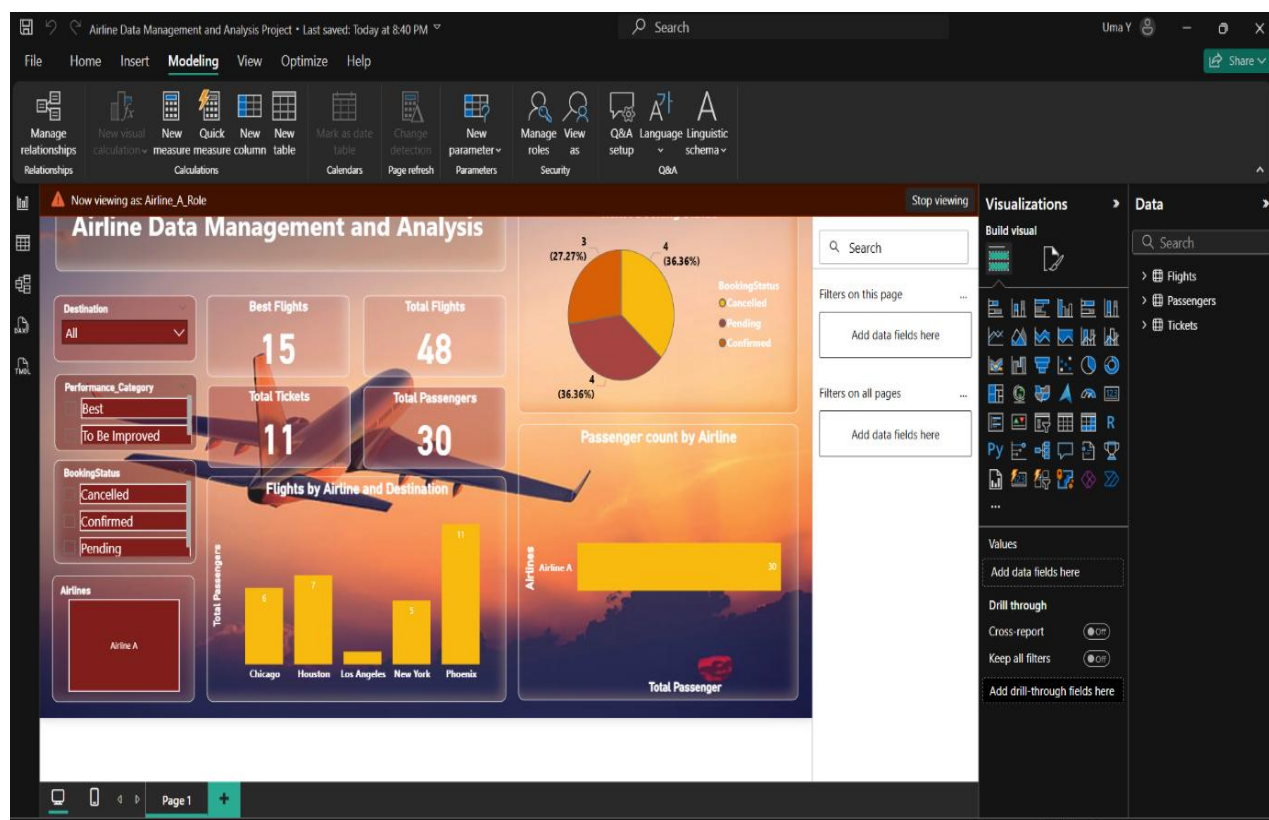
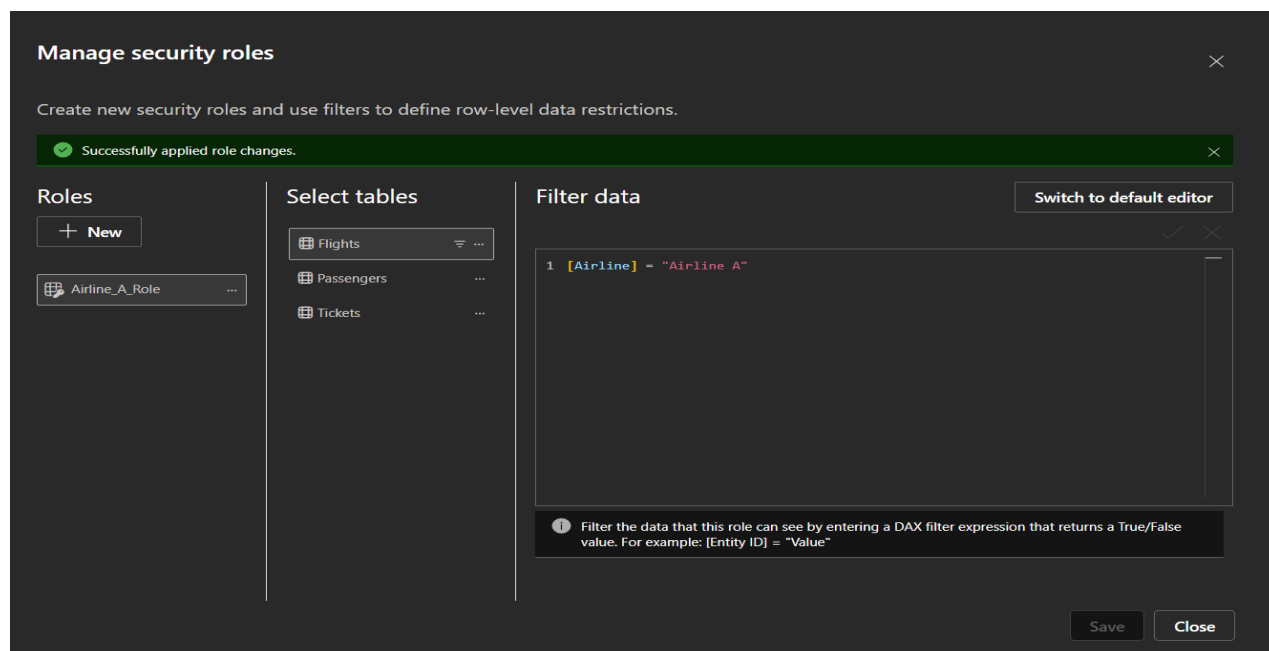
I built an interactive dashboard with:

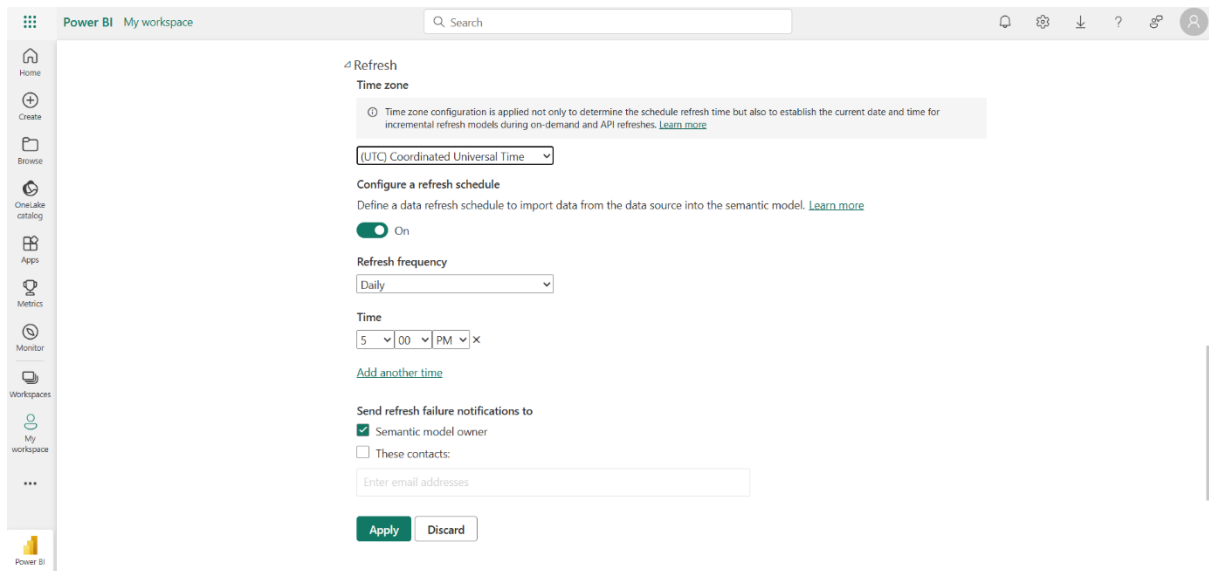
- Passenger count by airline (Bar Chart).
- Ticket booking status (Pie Chart).
- Flights by airline and destination (Stacked column Chart).
- Table with flight details and KPIs.
- Slicers for Airline, Destination, Performance category, and BookingStatus



## Task 6 – Final Dashboard & Power BI Service

- **Row-Level Security (RLS):** Created a role for Airline A so only relevant data is visible.
- **Published** the report to Power BI Service.
- **Scheduled Refresh** set to run daily at **5:00 PM**.





## 4. Key Insights

1. Airline A has the highest number of passengers.
2. Most flights are classified as “Best” based on their status.
3. Phoenix is the top destination in terms of passengers.
4. Confirmed tickets are the largest booking category.

## 5. Conclusion

This project shows how Power BI can connect multiple datasets, clean and transform data, and create interactive dashboards that make airline operations easier to monitor. With RLS and scheduled refresh, the dashboard stays updated and secure for different users.

**Video Explanation:** Please consider both videos.

<https://www.loom.com/share/6526bb5b48254ae0a564c60a58e26bfb>

<https://www.loom.com/share/d87ff68bff89484891fe74a76968ab22?sid=02bec628-5439-46e0-945c-6e571fb20b00>

**Submitted By: Uma Yadav**

**Date: 22/08/2025**