Airline Data Management and Analysis Using Power BI

Introduction

The airline industry is highly dynamic, requiring efficient data management and insightful analysis to enhance operational efficiency and customer satisfaction. This project leverages Power BI to analyze airline operations, focusing on flight schedules, passenger management, and ticketing systems.

Objective

The primary goal of this project is to analyze and visualize airline data to gain operational insights, improve passenger management, and track booking trends effectively.

Datasets Used

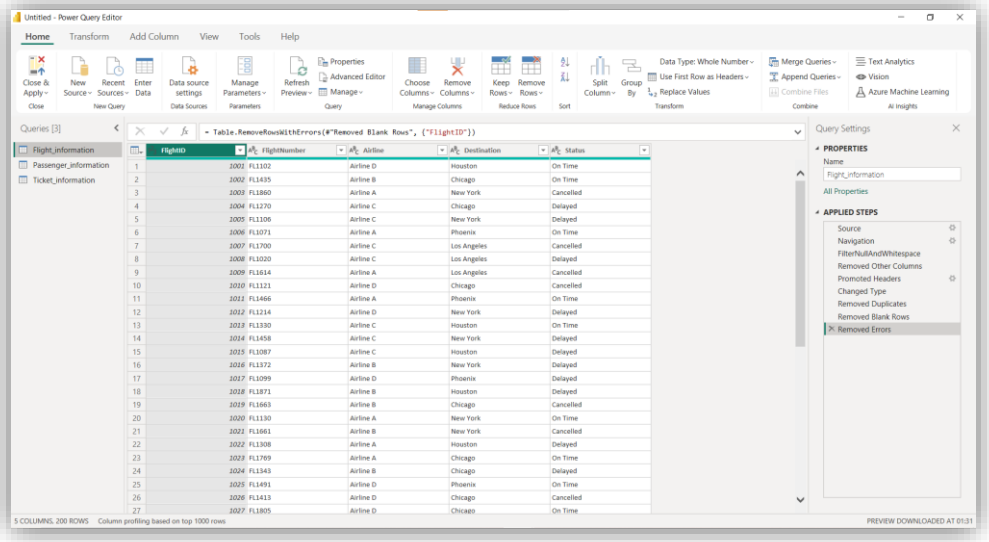
Three datasets were utilized in this project:

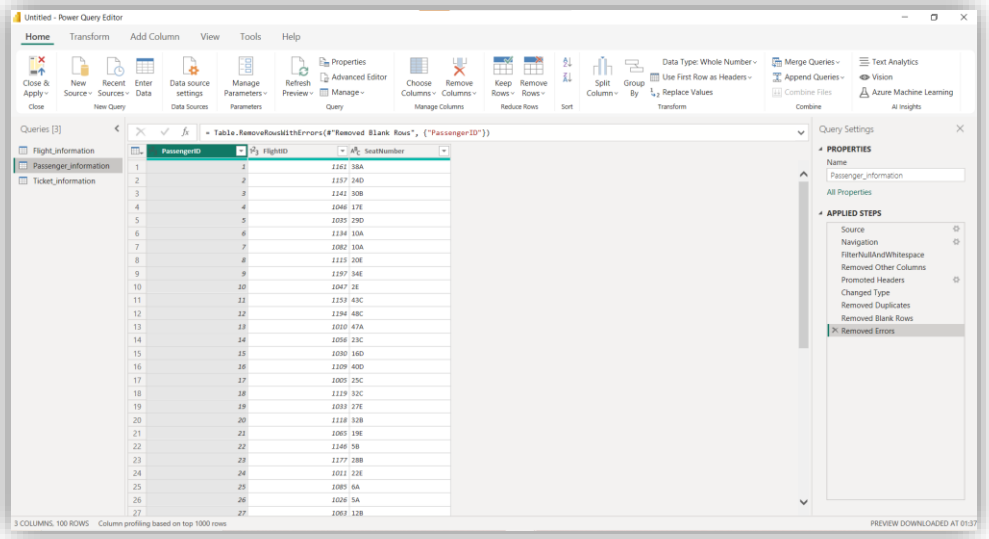
1. **Flight Information**: Contains details such as FlightID, FlightNumber, Airline, Destination, and Status.
2. **Passenger Information**: Includes PassengerID, FlightID, and SeatNumber.
3. **Ticket Information**: Comprises TicketID, FlightID, and BookingStatus.

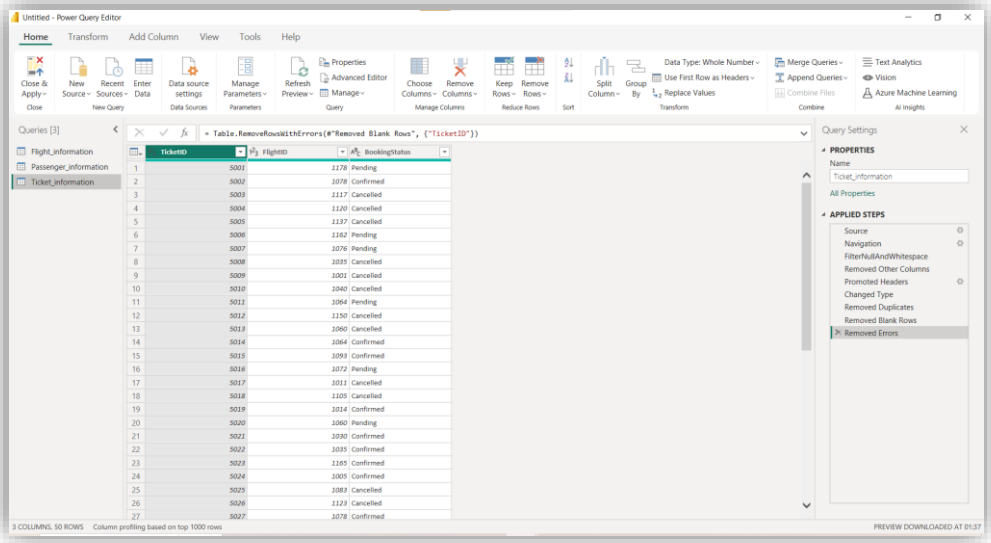
Task Performed

1. Data Preparation and Cleaning

- In this task, I have Extracted and transformed data using Power Query by importing raw datasets.
- Removed duplicate entries and checked any missing values for appropriate replacement.
- Formatted columns to ensure consistency in data types, such as converting numerical data into number formats and ensuring text fields are properly categorized.
- Applied filters to detect and correct anomalies in the dataset.

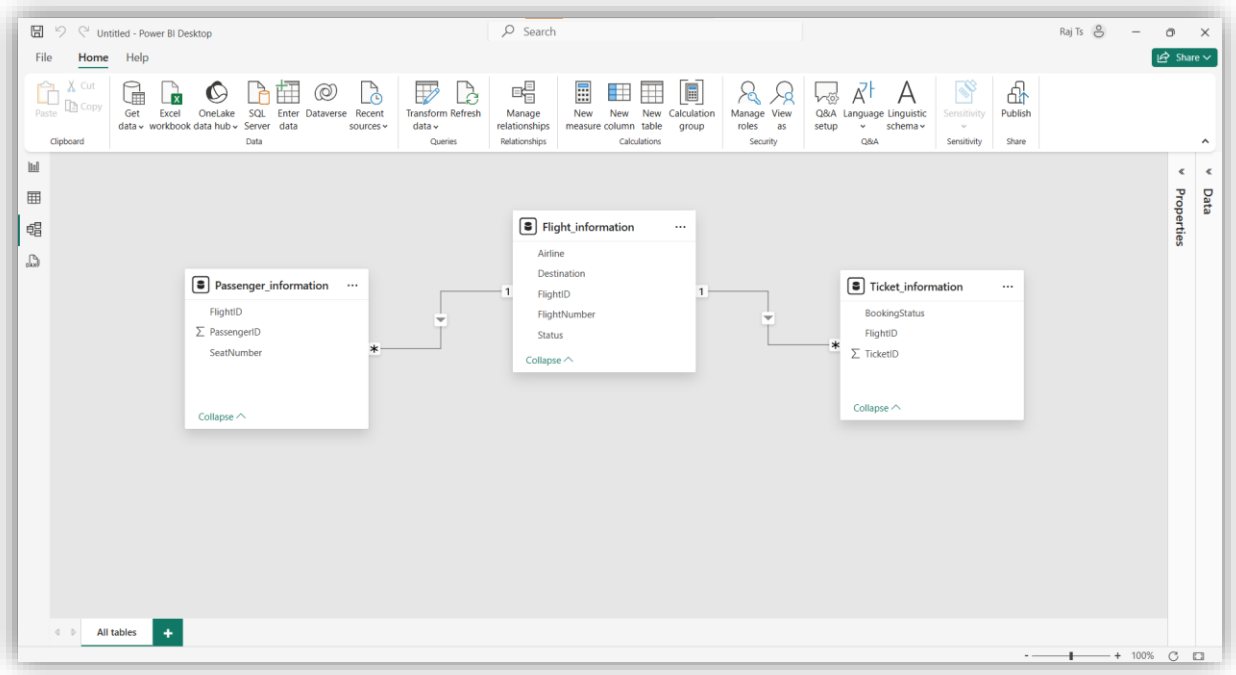

Untitled - Power Query Editor
Home
Transform
Add Column
View
Tools
Help
Queries [3]
Flight_information
Passenger_information
Ticket_information
= Table.RemoveRowsWithErrors(#"Removed Blank Rows", {"FlightID"})
Query Settings
PROPERTIES
Name
Flight_information
All Properties
APPLIED STEPS
Source
Navigation
FilterNullAndWhitespace
Removed Other Columns
Promoted Headers
Changed Type
Removed Duplicates
Removed Blank Rows
Removed Errors
5 COLUMNS, 200 ROWS
Column profiling based on top 1000 rows
PREVIEW DOWNLOADED AT 01:31


Untitled - Power Query Editor
Home
Transform
Add Column
View
Tools
Help
Queries [3]
Flight_information
Passenger_information
Ticket_information
= Table.RemoveRowsWithErrors(#"Removed Blank Rows", {"PassengerID"})
Query Settings
PROPERTIES
Name
Passenger_information
All Properties
APPLIED STEPS
Source
Navigation
FilterNullAndWhitespace
Removed Other Columns
Promoted Headers
Changed Type
Removed Duplicates
Removed Blank Rows
Removed Errors
3 COLUMNS, 100 ROWS
Column profiling based on top 1000 rows
PREVIEW DOWNLOADED AT 01:37


Untitled - Power Query Editor
Home
Transform
Add Column
View
Tools
Help
Queries [3]
Flight_information
Passenger_information
Ticket_information
= Table.RemoveRowsWithErrors(#"Removed Blank Rows", {"TicketID"})
Query Settings
PROPERTIES
Name
Ticket_information
All Properties
APPLIED STEPS
Source
Navigation
FilterNullAndWhitespace
Removed Other Columns
Promoted Headers
Changed Type
Removed Duplicates
Removed Blank Rows
Removed Errors
3 COLUMNS, 50 ROWS
Column profiling based on top 1000 rows
PREVIEW DOWNLOADED AT 01:37

2. Data Modeling

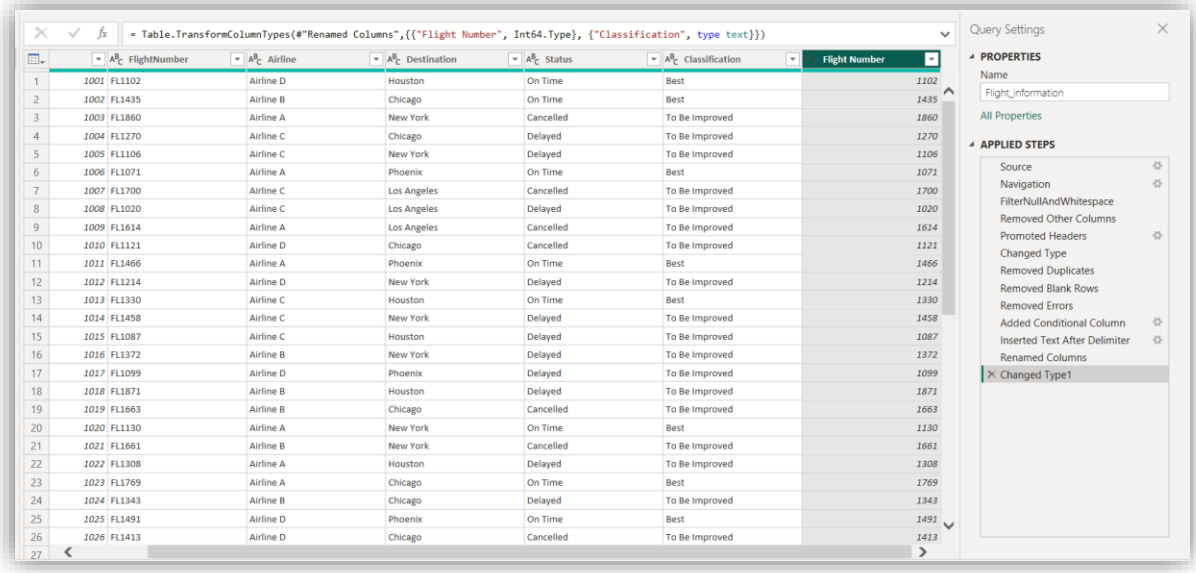
- For this model, I created relationships between datasets using **FlightID** as the primary key, ensuring seamless data integration.
- Then I configured appropriate **cardinality** for each relationship to maintain referential integrity. For which I explored datasets in Table view and noticed unique and repetitive entries of records.
- The ‘Flight_information’ table were having unique values where as ‘Passenger_information’ and ‘Ticket_information’ was with multiple entries which means it will be a One-to-Many relationship between Flight_information to other two tables and also single cross filter direction
- Finally, I ensured the correct data flow and successfully established relationship between tables.



3. Enhanced Data Insights

- I have Added a **conditional column** to classify flights as **"Best"** for on-time departures and **"To Be Improved"** for delayed or canceled flights and named it as “Classification”.
- Further, I have Used **Column from Examples** in Power Query to efficiently extract flight numbers from FlightNumber fields.
- Then I Applied transformation functions to split and restructure data where necessary. And verified data integrity by cross-checking transformed columns with the original dataset.

FlightID	FlightNumber	Airline	Destination	Status	Classification
1001	FL1102	Airline D	Houston	On Time	Best
1002	FL1435	Airline B	Chicago	On Time	Best
1003	FL1860	Airline A	New York	Cancelled	To Be Improved
1004	FL1270	Airline C	Chicago	Delayed	To Be Improved
1005	FL1106	Airline C	New York	Delayed	To Be Improved
1006	FL1071	Airline A	Phoenix	On Time	Best
1007	FL1700	Airline C	Los Angeles	Cancelled	To Be Improved
1008	FL1020	Airline C	Los Angeles	Delayed	To Be Improved
1009	FL1614	Airline A	Los Angeles	Cancelled	To Be Improved
1010	FL1121	Airline D	Chicago	Cancelled	To Be Improved
1011	FL1466	Airline A	Phoenix	On Time	Best
1012	FL1214	Airline D	New York	Delayed	To Be Improved
1013	FL1330	Airline C	Houston	On Time	Best
1014	FL1458	Airline C	New York	Delayed	To Be Improved
1015	FL1087	Airline C	Houston	Delayed	To Be Improved
1016	FL1372	Airline B	New York	Delayed	To Be Improved
1017	FL1099	Airline D	Phoenix	Delayed	To Be Improved
1018	FL1871	Airline B	Houston	Delayed	To Be Improved
1019	FL1663	Airline B	Chicago	Cancelled	To Be Improved
1020	FL1130	Airline A	New York	On Time	Best
1021	FL1661	Airline B	New York	Cancelled	To Be Improved
1022	FL1308	Airline A	Houston	Delayed	To Be Improved
1023	FL1769	Airline A	Chicago	On Time	Best
1024	FL1343	Airline B	Chicago	Delayed	To Be Improved
1025	FL1491	Airline D	Phoenix	On Time	Best



4. Calculations Using DAX

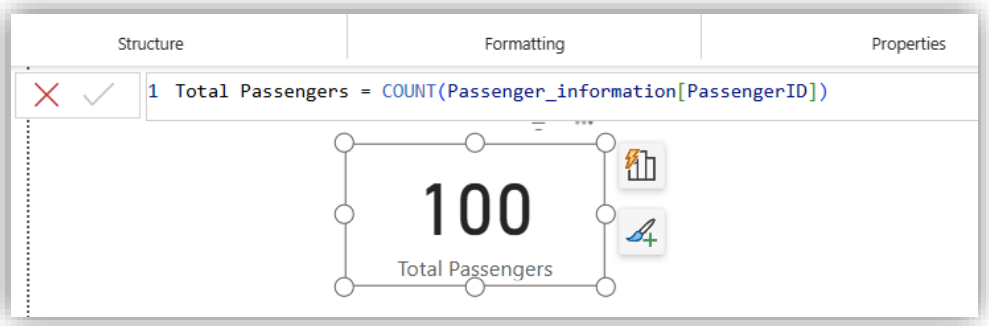
- I have Created the following **DAX** calculations:

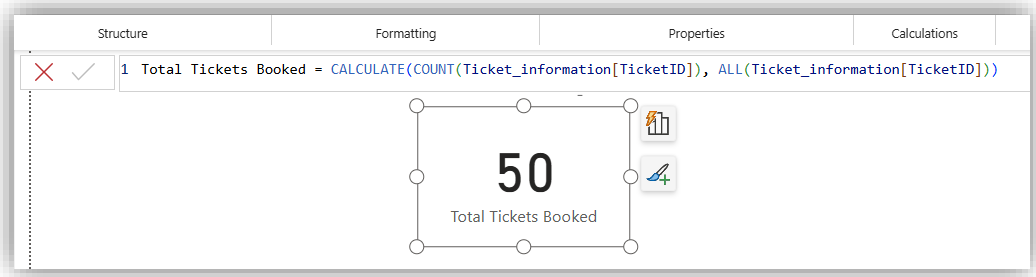
  a) **Total Passengers**: Count of passengers per flight, calculated using COUNT function.
  Total Passengers = COUNT(Passenger_information[PassengerID])

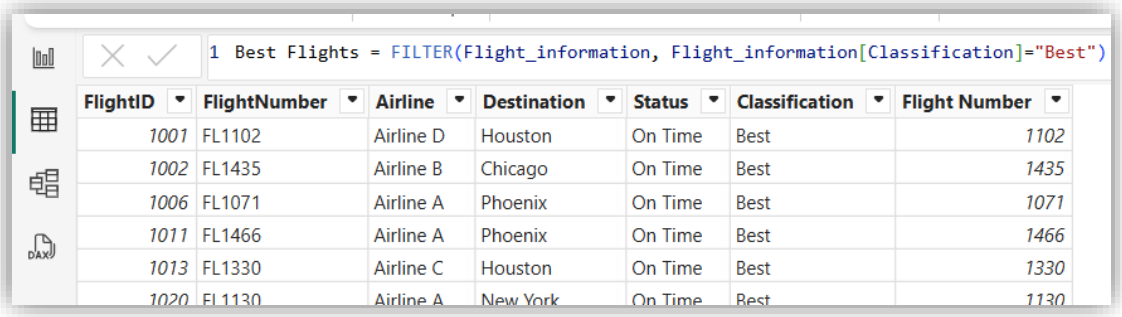
  b) **Total Tickets Booked**: Count of tickets booked across all flights ignoring any filter using COUNT.
  Total Tickets Booked = CALCULATE(COUNT(Ticket_information[TicketID]), ALL(Ticket_information[TicketID]))

  c) **Filtered Table for "Best" Flights**: Applied a FILTER function to extract only flights classified as "Best".
  Best Flights = FILTER(Flight_information, Flight_information[Classification]="Best")

- Optimized calculations using DAX to enhance performance.







FlightID	FlightNumber	Airline	Destination	Status	Classification	Flight Number
1001	FL1102	Airline D	Houston	On Time	Best	1102
1002	FL1435	Airline B	Chicago	On Time	Best	1435
1006	FL1071	Airline A	Phoenix	On Time	Best	1071
1011	FL1466	Airline A	Phoenix	On Time	Best	1466
1013	FL1330	Airline C	Houston	On Time	Best	1330
1020	FL1130	Airline A	New York	On Time	Best	1130

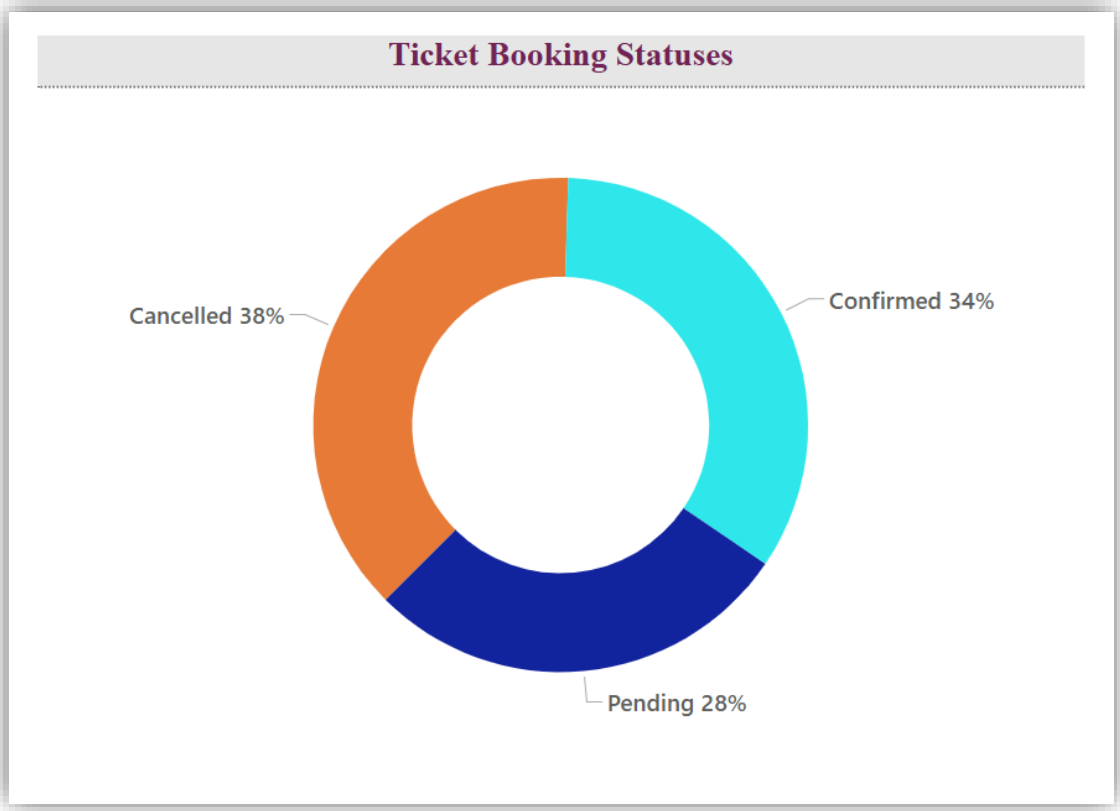
5. Visualization and Interactive Features

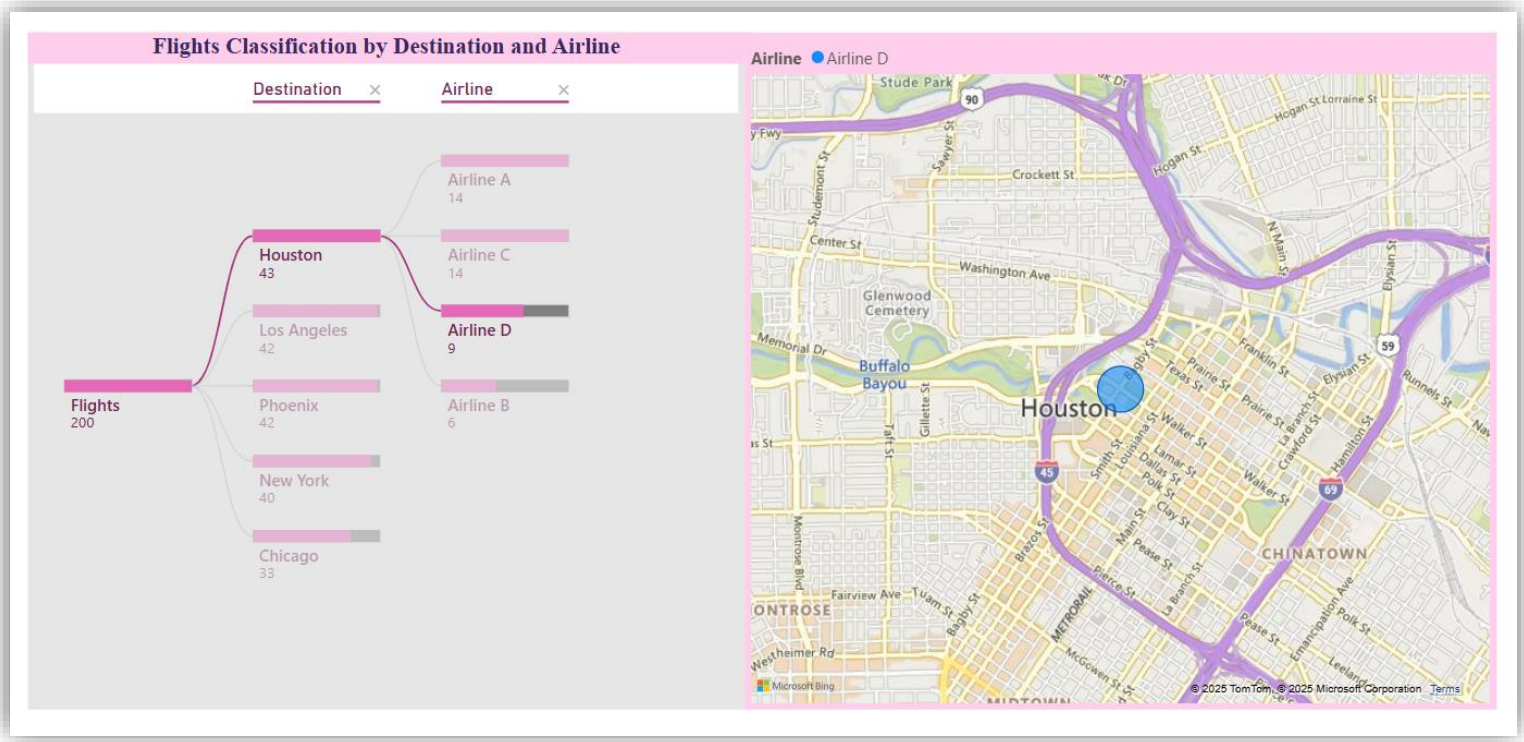
To complete this task I have developed various **visual representations** using Power BI, including:

a) **Passenger count by airline**: Bar chart showing total passenger distribution.

b) **Ticket booking statuses**: Pie chart representing confirmed, pending, and canceled bookings.

c) **Flights by airline and destination**: Decomposition Tree for classified view and Map visualization illustrating flight destinations.

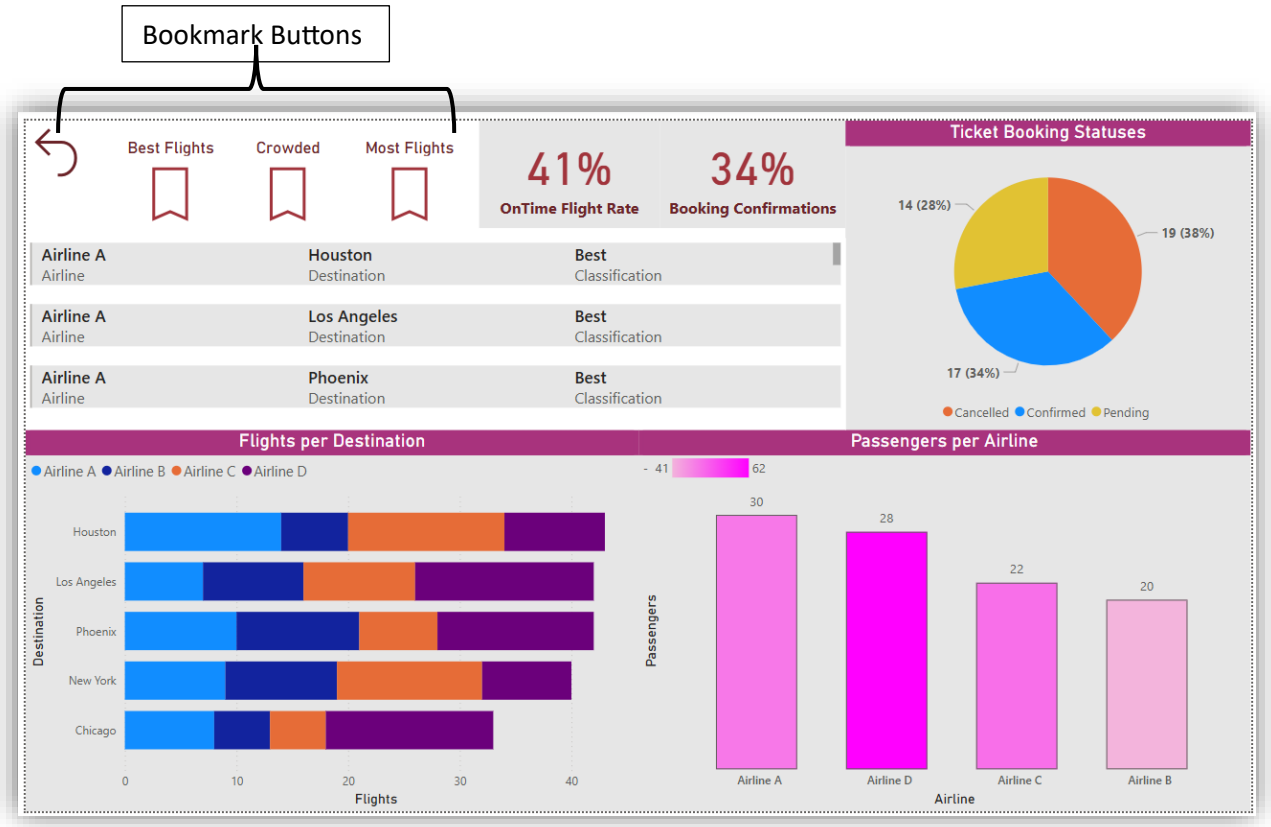

Ticket Booking Statuses
Cancelled 38%
Confirmed 34%
Pending 28%


Flights Classification by Destination and Airline
Destination
Airline
Flights
200
Houston
43
Los Angeles
42
Phoenix
42
New York
40
Chicago
33
Airline A
14
Airline C
14
Airline D
9
Airline B
6
Airline
Airline D
Houston
CHINATOWN
Microsoft Bing
© 2025 TomTom, © 2025 Microsoft Corporation Terms

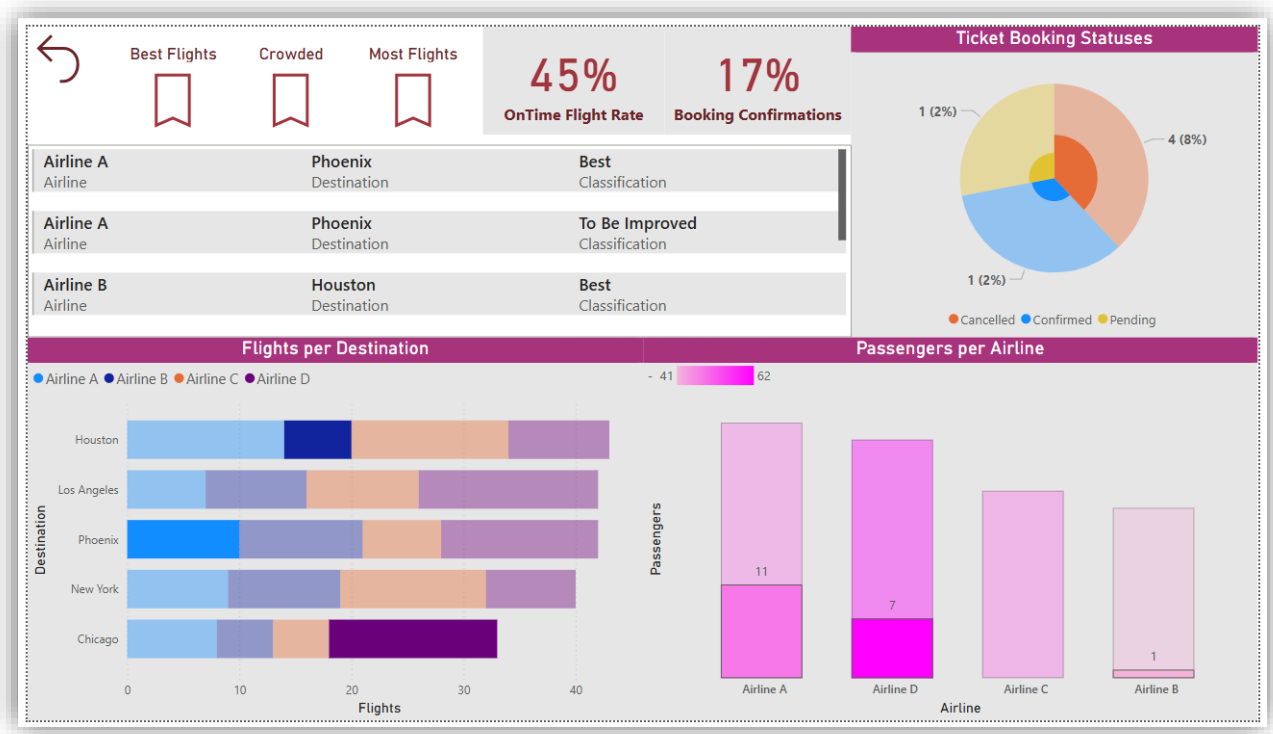
Interactive Features used are as,

- Implemented interactive **slicers** and **filters** with **Reset** button to allow users to explore data dynamically.
- Inserted important bookmarked pages assigned to '**Bookmark**' button and also filter '**Reset**' button.
- Created **drill-through** reports for deeper insights into airline-specific operations.

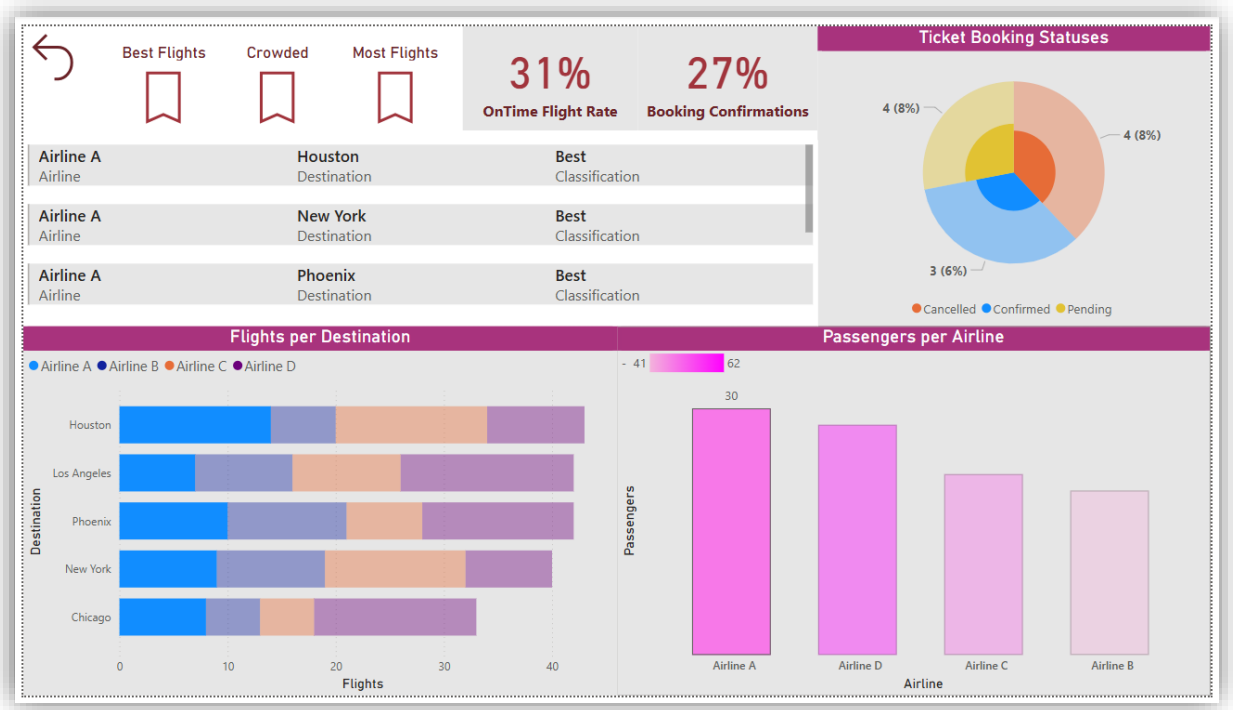
Use of Bookmark for Quick Views:



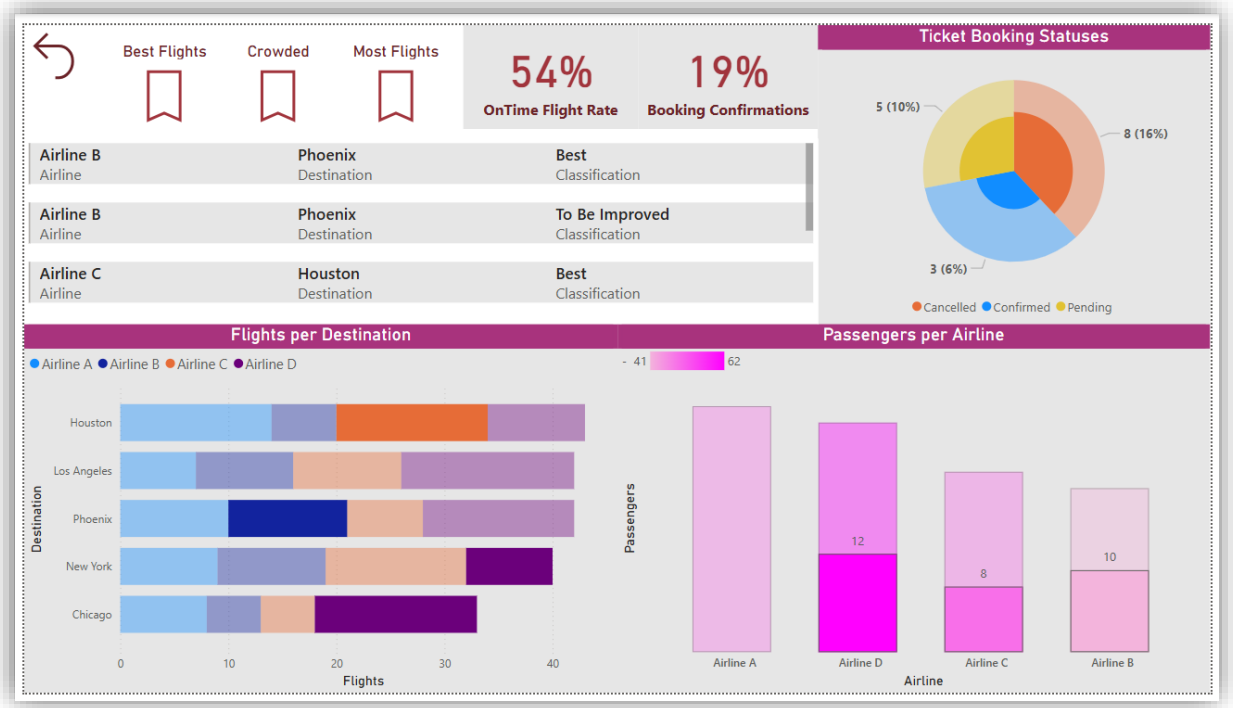
Best Flights:



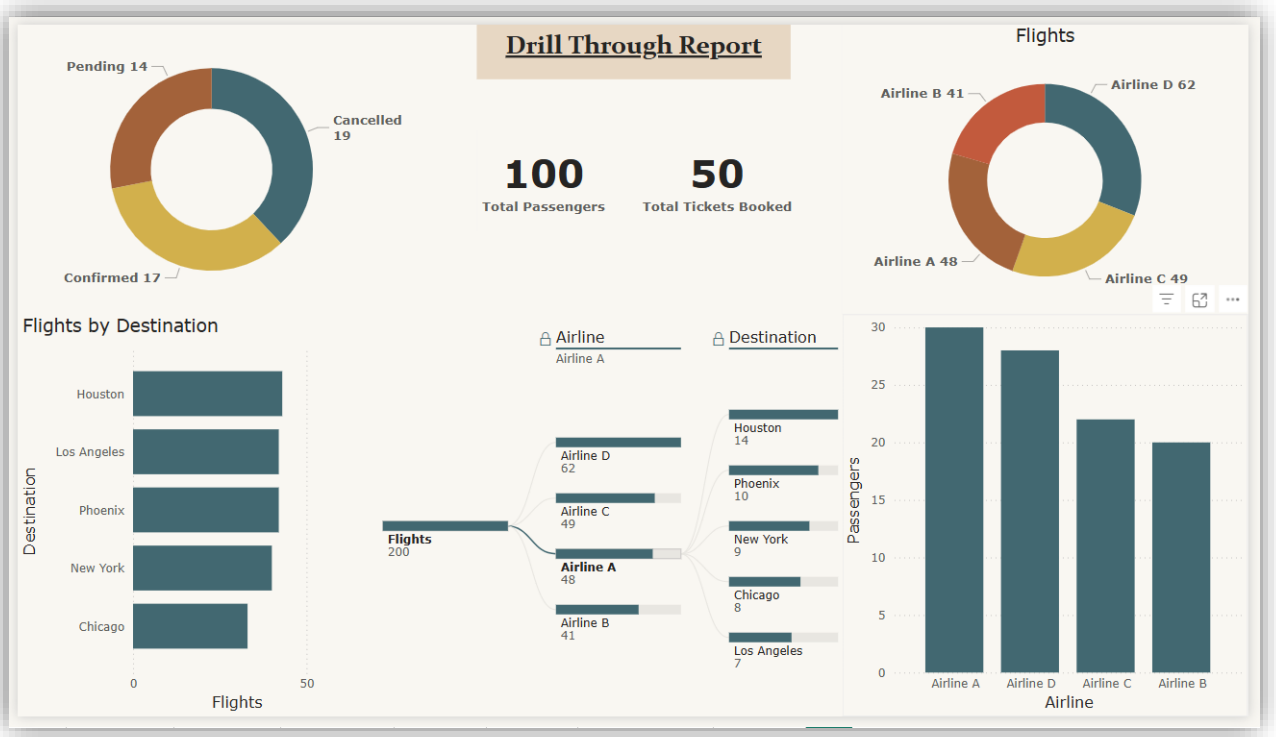
Crowded Flights (Most Passengers):



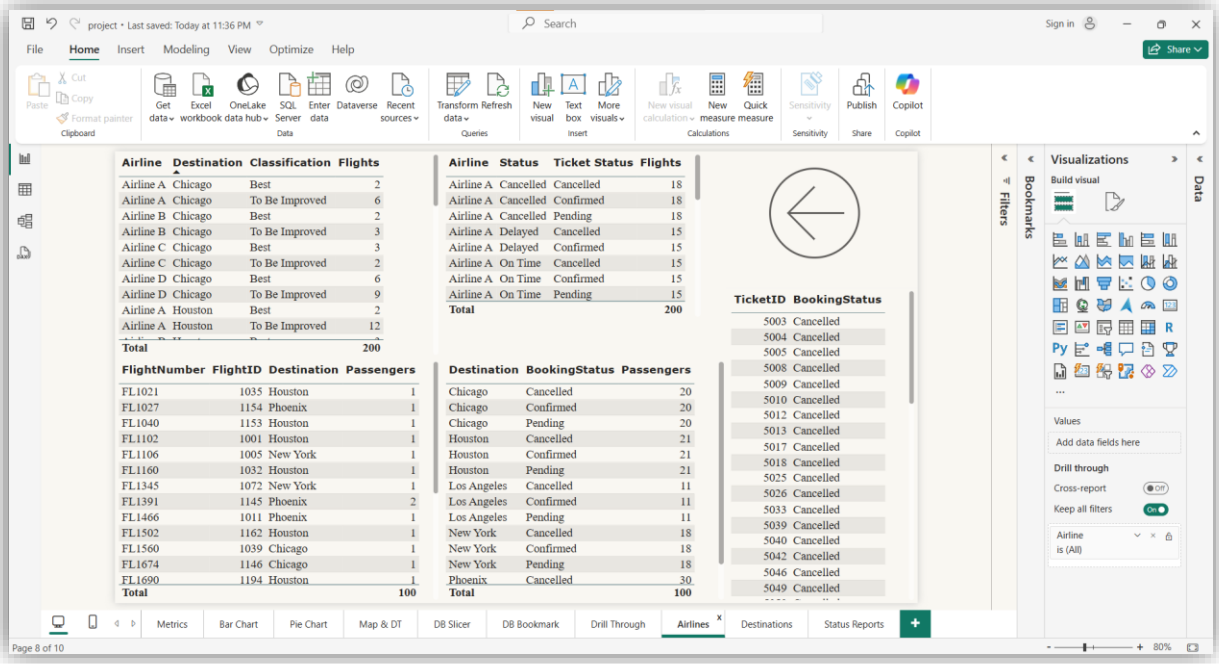
Airlines with most flights:

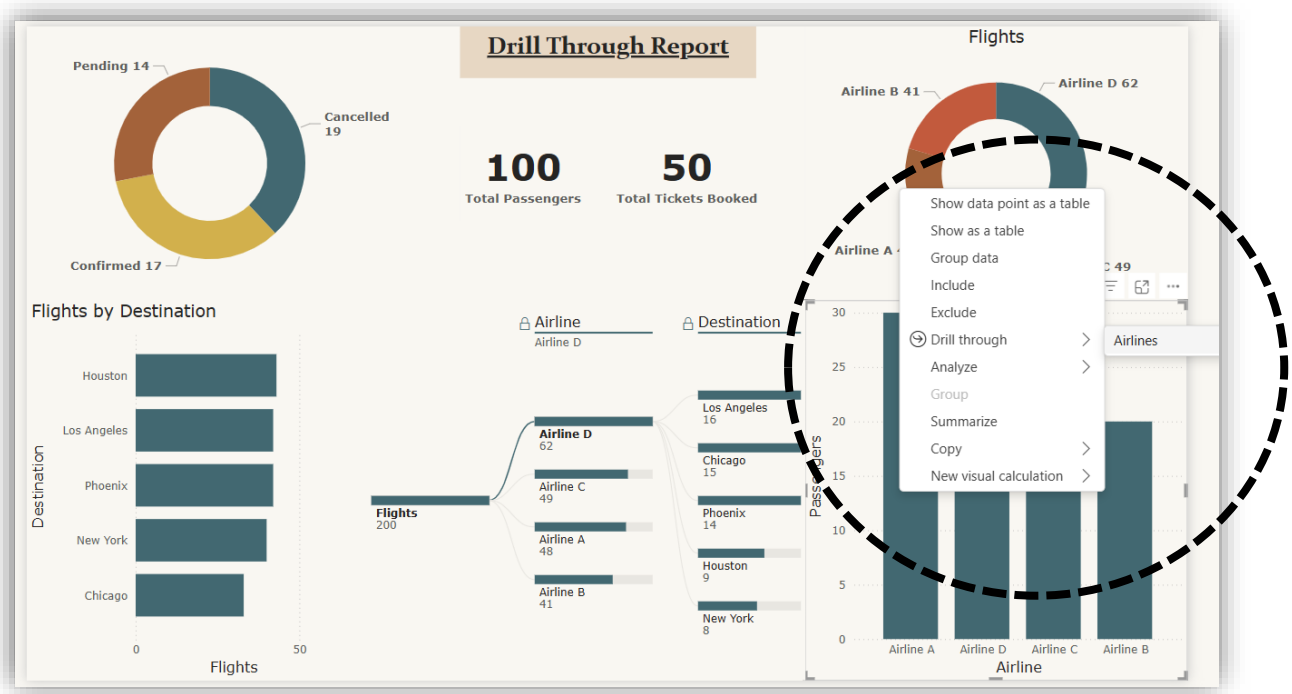


Drill Through Report with Airline Specific pages:



The report is interactive and connected to specific pages where we find the granular details of each category.





Airline	Destination	Classification	Flights
Airline A	Chicago	Best	2
Airline A	Chicago	To Be Improved	6
Airline A	Houston	Best	2
Airline A	Houston	To Be Improved	12
Airline A	Los Angeles	Best	1
Airline A	Los Angeles	To Be Improved	6
Airline A	New York	Best	5
Airline A	New York	To Be Improved	4
Airline A	Phoenix	Best	5
Airline A	Phoenix	To Be Improved	5
Total			**48**

Airline	Status	Ticket Status	Flights
Airline A	Cancelled	Cancelled	18
Airline A	Cancelled	Confirmed	18
Airline A	Cancelled	Pending	18
Airline A	Delayed	Cancelled	15
Airline A	Delayed	Confirmed	15
Airline A	On Time	Cancelled	15
Airline A	On Time	Confirmed	15
Airline A	On Time	Pending	15
Total			**48**

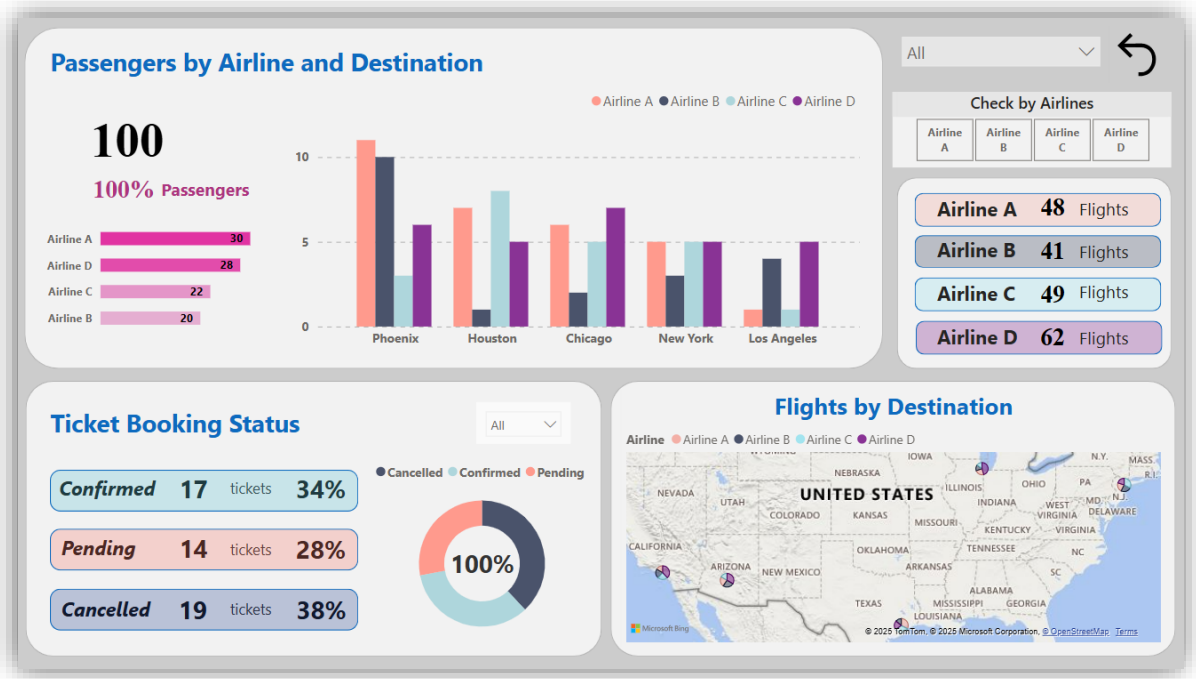
TicketID	BookingStatus
5010	Cancelled
5017	Cancelled
5042	Cancelled
5049	Cancelled
5032	Confirmed
5035	Confirmed
5045	Confirmed
5001	Pending
5016	Pending
5034	Pending
5036	Pending

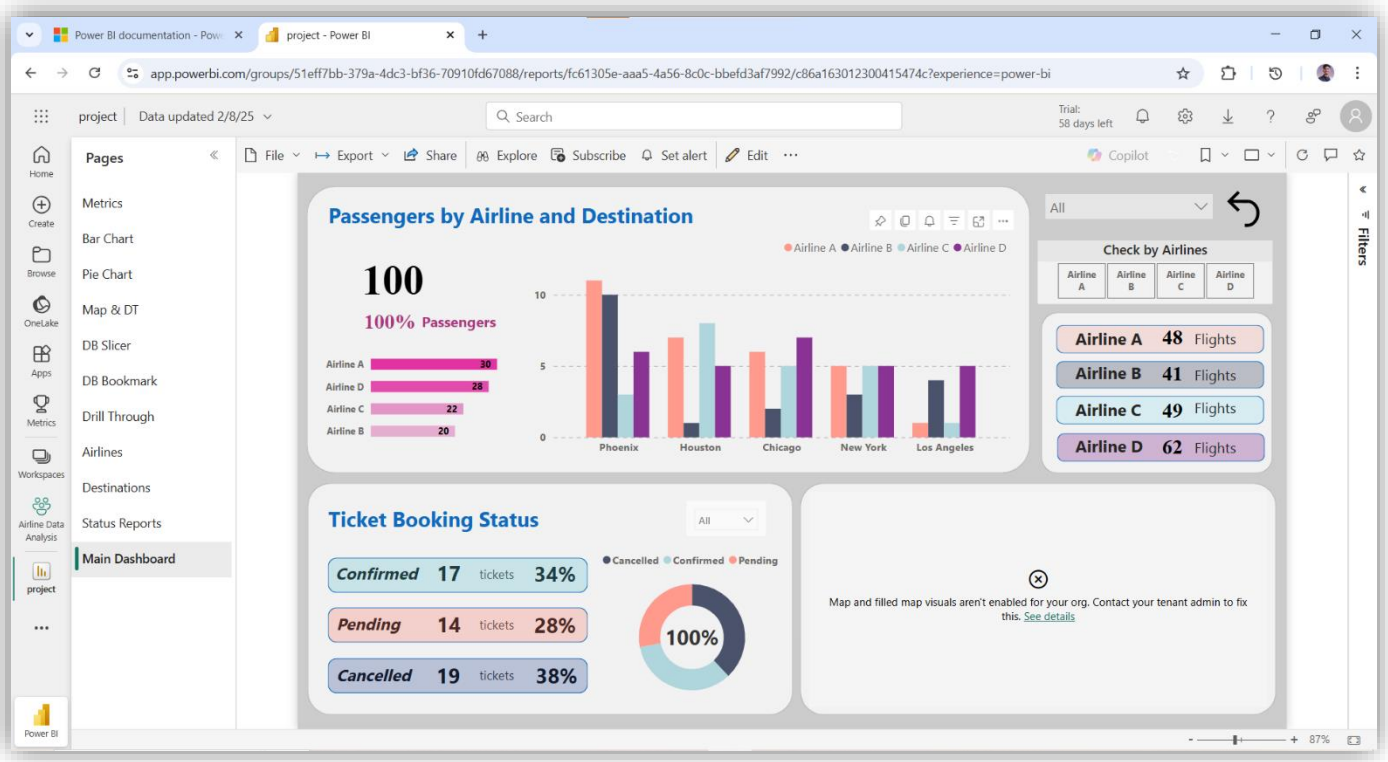
FlightNumber	FlightID	Destination	Passengers
FL1345	1072	New York	1
FL1391	1145	Phoenix	2
FL1466	1011	Phoenix	1
FL1674	1146	Chicago	1
FL1769	1023	Chicago	1
Total			**30**

Destination	BookingStatus	Passengers
Chicago	Confirmed	6
Chicago	Pending	6
Houston	Cancelled	7
Houston	Confirmed	7
Houston	Pending	7
New York	Cancelled	5
New York	Confirmed	5
New York	Pending	5
Phoenix	Cancelled	11
Total		**30**

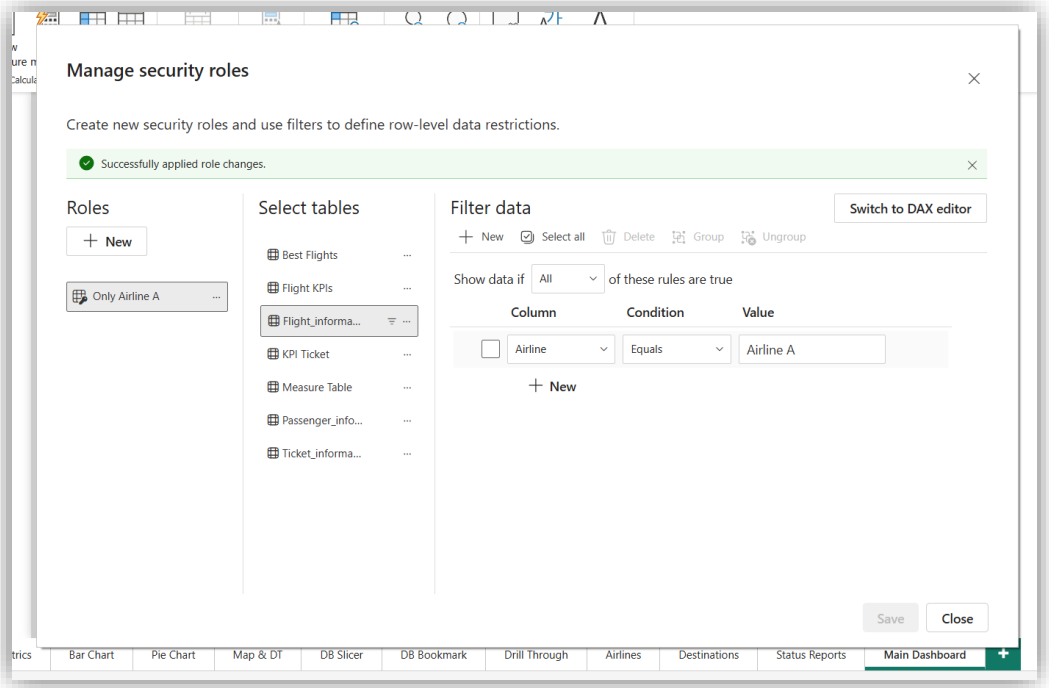
6. Final Dashboard and Power BI Service

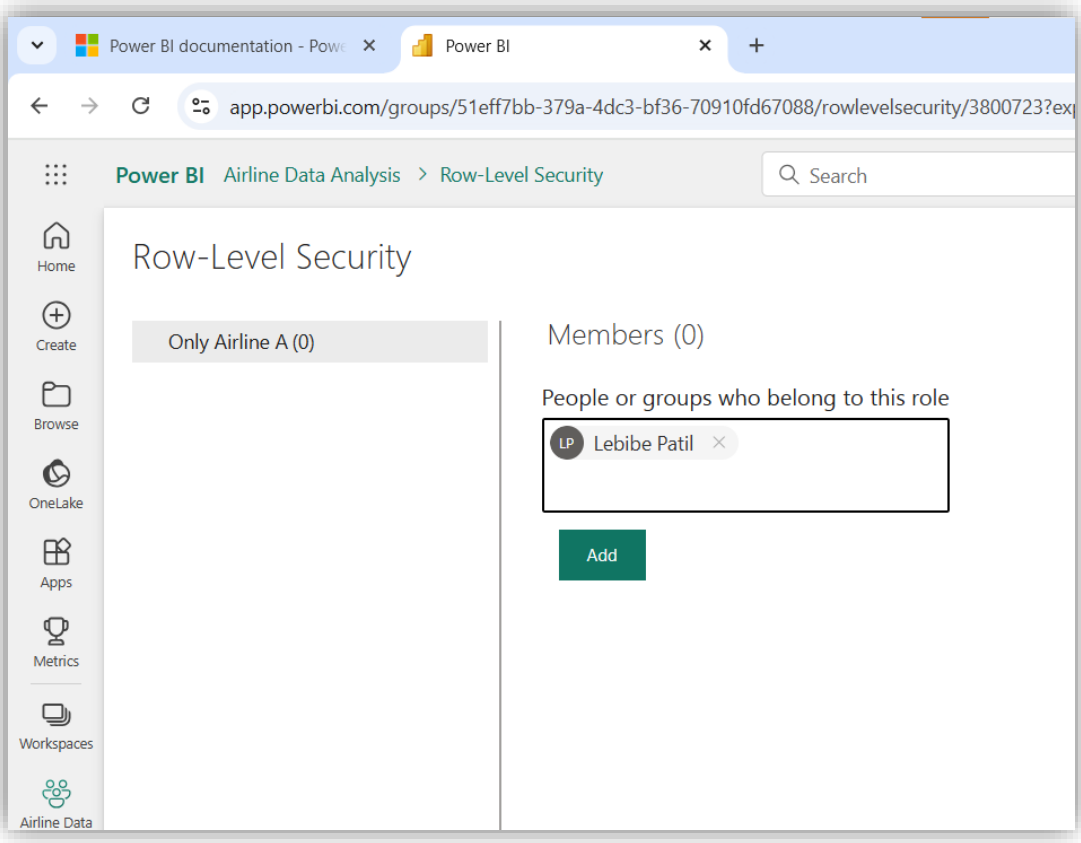
- I have designed a **comprehensive dashboard** summarizing key insights using various interactive visuals and appropriate filters. The slicers and buttons are also available for ease of viewer's access.
- Then I published the dashboard to Power BI service in new workspace (Airline Data Analysis) for seamless sharing but the Map visual got user restricted
- **(I tried creating new service account but the following Map Visual Error repetitively occurred)** but the dashboard is successfully published.

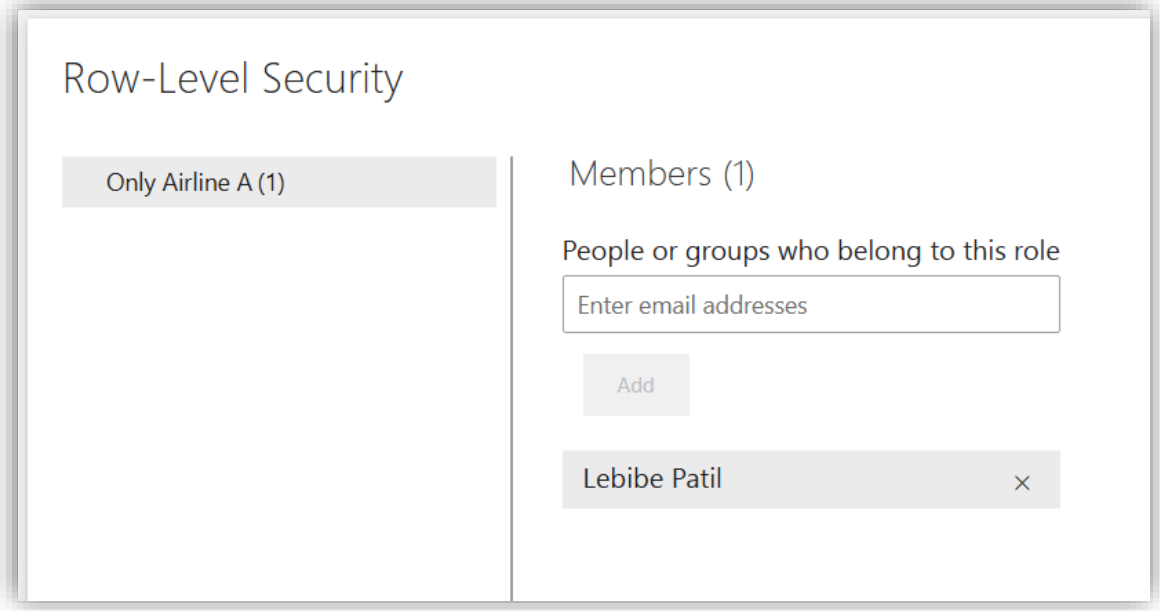




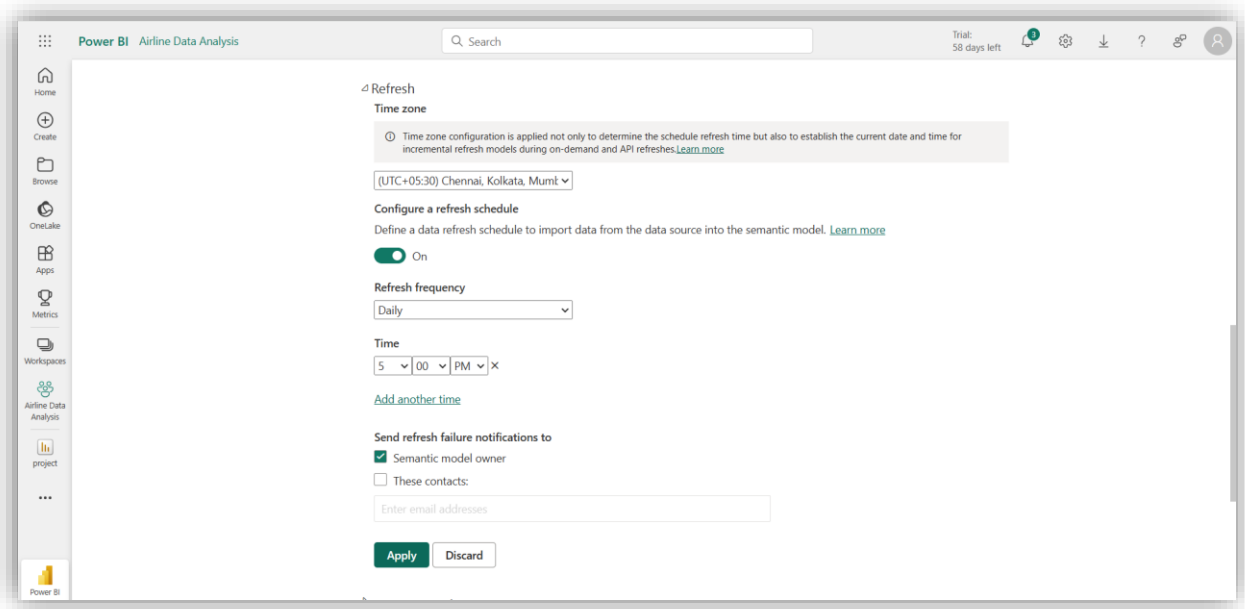
- Then I configured **Row-Level Security (RLS)** for Airline A, ensuring restricted data access to authorized users only. And from Power BI Service account managed the RLS by assigning to other user email.

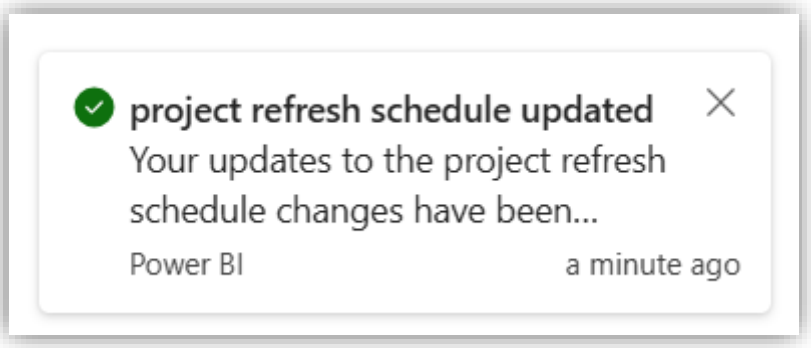



Power BI documentation - Powe
Power BI
app.powerbi.com/groups/51eff7bb-379a-4dc3-bf36-70910fd67088/rowlevelsecurity/3800723?exp
Power BI
Airline Data Analysis
Row-Level Security
Search
Home
Create
Browse
OneLake
Apps
Metrics
Workspaces
Airline Data
Row-Level Security
Only Airline A (0)
Members (0)
People or groups who belong to this role
LP
Lebibe Patil
Add


Row-Level Security
Only Airline A (1)
Members (1)
People or groups who belong to this role
Enter email addresses
Add
Lebibe Patil

- Then I have set up a **scheduled data refresh** at **5 PM daily** to ensure up-to-date reporting and live updating of datasets.





Key Findings & Insights:

1. **Flight Status Analysis:**

   We were provided with airline data of 200 flights out of which 82 flights were On Time, 60 flights were Cancelled, indicating a significant disruption and 58 flights were Delayed, which may impact customer satisfaction.

2. **Ticket Booking Trends:**

   There were 50 ticket transactions out of which 19 bookings were Cancelled, higher than Confirmed which were 17 and Pending 14.

   A high cancellation rate suggests operational inefficiencies or customer dissatisfaction.

3. **Passenger Load Per Flight Status:**

   On average, Cancelled flights had 1.28 passengers, indicating last-minute cancellations or operational issues. Whereas Delayed flights had 1.17 passengers on average, possibly due to poor scheduling.

   On-Time flights had 1.2 passengers, showing better efficiency but still room for improvement in occupancy rates.

Flight Cancellations by Airline and Destination:

- Airline A and Airline D had the highest cancellations with 18 cancellation records. Airline B had the lowest cancellations, suggesting better reliability.
- Los Angele had the highest cancellation followed by New York and Chicago.

Flight Delays by Airline:

- Airline D had the highest delays which is 18 Flights, matching its cancellation trend.
- Houston had the most delays with 21 flights followed by Los Angeles and Phoenix which had 11 delays each.

Key Takeaways:

- Airline D and Airline A need operational improvements to reduce cancellations and delays.
- Los Angeles and Houston are the most problematic destinations, requiring further investigation.
- Airline B appears the most reliable, with the fewest cancellations and delays.

Explanation : Video Link

https://drive.google.com/file/d/1SL4uVQmg8ZF6jqmPlbt8LdOug5yBUXSB/view?usp=sharing

(It is better to copy the link and paste it in browser if not worked by clicking)

Thank You !

: Ruturaj T. Saravane