

DATA SCIENCE POWER BI PROJECT

Task 1. Data Preparation and Cleaning

Steps taken :

Step 1 : Home

Step 2: Remove duplicates/Blank rows/Format

File

Home

Transform

Add Column

View

Tools

Help

Close & Apply

New Query

Data Source Settings

Parameters

Data Source

Parameters

Refresh Preview

Advanced Editor

Choose Columns

Remove Columns

Keep Rows

Remove Rows

Sort

Column

Group By

Transform

Merge Queries

Append Queries

Combine Files

Combine

Text Analytics

Vision

Azure Machine Learning

AI Insights

Queries [3]

flight_information

passenger_information

ticket_information

Table.SelectRows(*Removed Duplicates*, each true)

FlightID

FlightNumber

Airline

Destination

Status

1

1001

FL1102

Airline D

Houston

On Time

2

1002

FL1435

Airline B

Chicago

On Time

3

1003

FL1805

Airline A

New York

Cancelled

4

1004

FL1270

Airline C

Chicago

Delayed

5

1005

FL11306

Airline C

New York

Delayed

6

1006

FL11071

Airline A

Phoenix

On Time

7

1007

FL17000

Airline C

Los Angeles

Cancelled

8

1008

FL10200

Airline C

Los Angeles

Delayed

9

1009

FL11824

Airline A

Los Angeles

Cancelled

10

1010

FL11221

Airline D

Chicago

Cancelled

11

1011

FL1466

Airline A

Phoenix

On Time

12

1012

FL12214

Airline D

New York

Delayed

13

1013

FL11330

Airline C

Houston

On Time

14

1014

FL1458

Airline C

New York

Delayed

15

1015

FL1087

Airline C

Houston

Delayed

16

1016

FL11372

Airline B

New York

Delayed

17

1017

FL1099

Airline D

Phoenix

Delayed

18

1018

FL11871

Airline B

Houston

Delayed

19

1019

FL1463

Airline B

Chicago

Cancelled

20

1020

FL11130

Airline A

New York

On Time

21

1021

FL11661

Airline B

New York

Cancelled

22

1022

FL1308

Airline A

Houston

Delayed

23

1023

FL17809

Airline A

Chicago

On Time

24

1024

FL1343

Airline B

Chicago

Delayed

25

1025

FL1491

Airline D

Phoenix

On Time

26

1026

FL14113

Airline D

Chicago

Cancelled

27

1027

FL13809

Airline D

Chicago

On Time

28

1028

FL11385

Airline D

Chicago

On Time

5 COLUMNS, 200 ROWS

Column profiling based on top 1000 rows

Properties

Name

flight_information

All Properties

Applied Steps

Source

Navigation

Promoted Headers

Changed Type

Removed Other Columns

Removed Duplicates

Filtered Rows

PREVIEW DOWNLOADED AT 01:26 PM

35°F

Haze

Search

Table: Table.SelectRows(*Removed Duplicates*, each true)

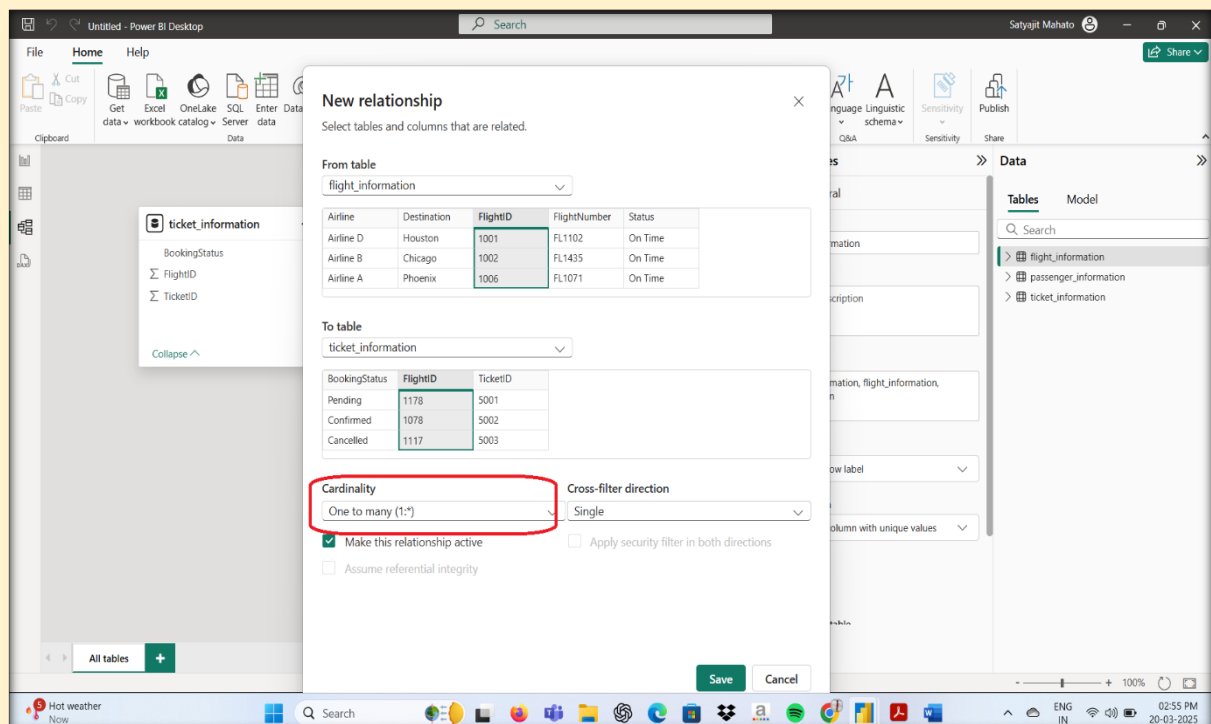
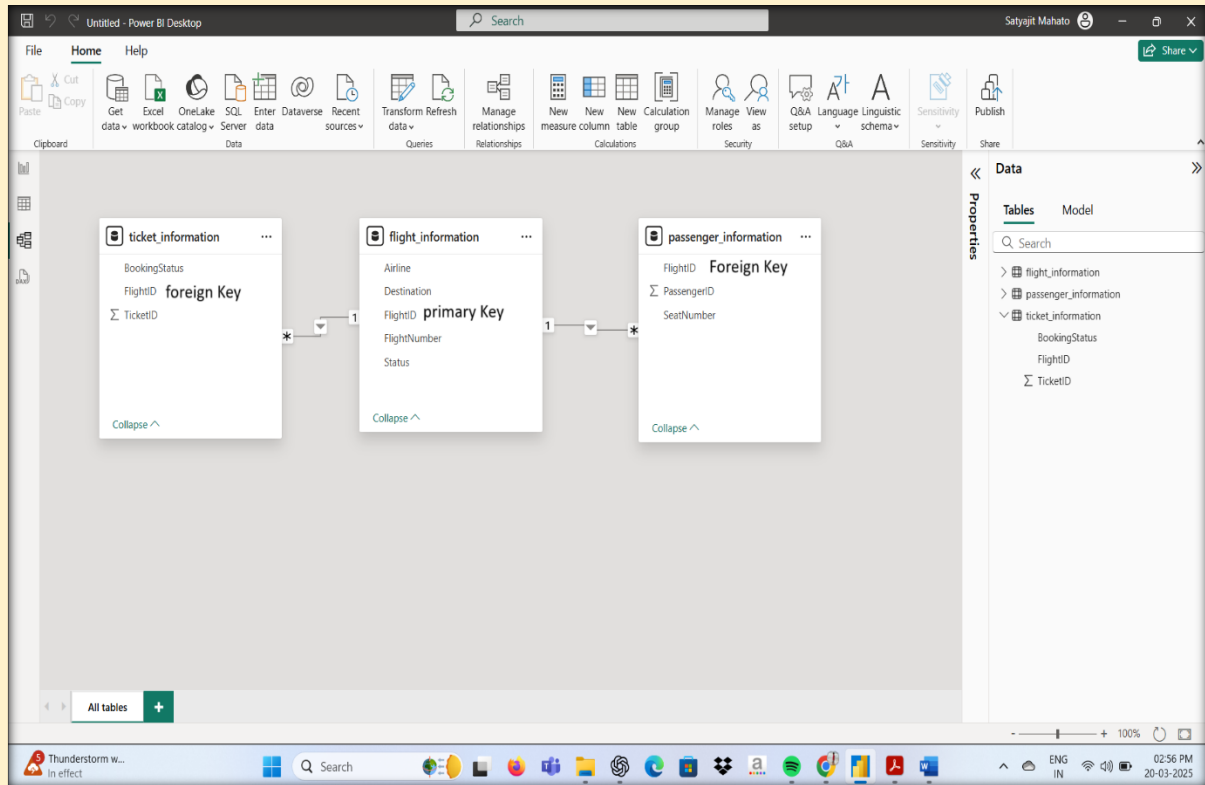
PassengerID	FlightID	SeatNumber
1	1	1161 38A
2	2	1157 24D
3	3	1141 30B
4	4	1046 17E
5	5	1035 29D
6	6	1134 10A
7	7	1082 10A
8	8	1115 20E
9	9	1197 84E
10	10	1047 2E
11	11	1154 48C
12	12	1194 48C
13	13	1010 47A
14	14	1056 23C
15	15	1030 16D
16	16	1109 40D
17	17	1008 25C
18	18	1119 32C
19	19	1032 27E
20	20	1118 32D
21	21	1065 19C
22	22	1146 5B
23	23	1177 28B
24	24	1011 22E
25	25	1085 6A
26	26	1026 5A
27	27	1068 12B
28	28	1089 46B

Table: Table.SelectRows(*Removed Duplicates*, each true)

TicketID	FlightID	BookingStatus
1	1001	1178 Pending
2	1002	1078 Confirmed
3	1003	1117 Cancelled
4	1004	1120 Cancelled
5	1005	1157 Cancelled
6	1006	1102 Pending
7	1007	1070 Pending
8	1008	1035 Cancelled
9	1009	1001 Cancelled
10	1010	1040 Cancelled
11	1011	1094 Pending
12	1012	1150 Cancelled
13	1013	1060 Cancelled
14	1014	1064 Confirmed
15	1015	1093 Confirmed
16	1016	1072 Pending
17	1017	1011 Cancelled
18	1018	1105 Cancelled
19	1019	1014 Confirmed
20	1020	1060 Pending
21	1021	1030 Confirmed
22	1022	1015 Confirmed
23	1023	1105 Confirmed
24	1024	1005 Confirmed
25	1025	1083 Cancelled
26	1026	1173 Cancelled
27	1027	1078 Confirmed
28	1028	1134 Pending

Task 2. Data Modeling

Flight_ID column is unique in flight_information which means all IDs are unique whereas flight_ID column in ticket_information and passenger_information query has repeated flight_ID values.



Task 3: Enhanced Data Insights

- Conditional Column displaying classification of flights as “Best” or “To be improved”
- Used “column from example” option to extract flight number from flight number

FileHomeTransform

Conditional Column

Index Column

Duplicate Column

Merge Columns

Extract

Format

Parse

Statistics

Standard Scientific

Information

Date

Time

Duration

Test Analytics

Vision

Azure Machine Learning

Help

Trigonometry

Rounding

10²

From Date & Time

AI Insights

Column From Examples

Custom Column

Invoke Custom Function

General

Queries [3]

flight_information

passenger_information

ticket_information

Table.TransformColumns(Table.TransformColumnTypes(#"Renamed Columns",{"FlightNumberOnly", Int64.Type}, {"Flight Classification", type Text})),

	FlightNumber	Airline	Destination	Status	Flight Classification	FlightNumberOnly	
1	1001	FL1102	Airline D	Houston	On Time	Best	1102
2	1002	FL1435	Airline B	Chicago	On Time	Best	1435
3	1003	FL1860	Airline A	New York	Cancelled	To Be Improved	1860
4	1004	FL1270	Airline C	Chicago	Delayed	To Be Improved	1270
5	1005	FL1108	Airline C	New York	Delayed	To Be Improved	1108
6	1006	FL1071	Airline A	Phoenix	On Time	Best	1071
7	1007	FL1700	Airline C	Los Angeles	Cancelled	To Be Improved	1700
8	1008	FL1020	Airline C	Los Angeles	Delayed	To Be Improved	1020
9	1009	FL1614	Airline A	Los Angeles	Cancelled	To Be Improved	1614
10	1010	FL1121	Airline D	Chicago	Cancelled	To Be Improved	1121
11	1011	FL1466	Airline A	Phoenix	On Time	Best	1466
12	1012	FL1214	Airline D	New York	Delayed	To Be Improved	1214
13	1013	FL1330	Airline C	Houston	On Time	Best	1330
14	1014	FL1458	Airline C	New York	Delayed	To Be Improved	1458
15	1015	FL1087	Airline C	Houston	Delayed	To Be Improved	1087
16	1016	FL1372	Airline B	New York	Delayed	To Be Improved	1372
17	1017	FL1099	Airline D	Phoenix	Delayed	To Be Improved	1099
18	1018	FL1871	Airline B	Houston	Delayed	To Be Improved	1871
19	1019	FL1963	Airline B	Chicago	Cancelled	To Be Improved	1963
20	1020	FL1130	Airline A	New York	On Time	Best	1130
21	1021	FL1461	Airline B	New York	Cancelled	To Be Improved	1461
22	1022	FL1308	Airline A	Houston	Delayed	To Be Improved	1308
23	1023	FL1769	Airline A	Chicago	On Time	Best	1769
24	1024	FL1343	Airline B	Chicago	Delayed	To Be Improved	1343
25	1025	FL1491	Airline D	Phoenix	On Time	Best	1491
26	1026	FL1413	Airline D	Chicago	Cancelled	To Be Improved	1413
27	1027	FL1805	Airline D	Chicago	On Time	Best	1805
28							

Query Settings

PROPERTIES

Name

flight_information

All Properties

APPLIED STEPS

Source

Navigation

Promoted Headers

Changed Type

Removed Other Columns

Removed Duplicates

Filtered Rows

Added Conditional Column

Inserted Rows1

Inserted Text After Delimiter

Renamed Columns

Changed Type1

7 COLUMNS, 200 ROWS

Column profiling based on top 1000 rows

PREVIEW DOWNLOADED AT 01:26 PM

33°C Haze

Search

5:13 PM 20-03-2025

Task 4 :Calculations Using DAX

- Total passengers for a specific flight.

DAX function

Total Passenger of flight 1003 =

CALCULATE(COUNT(passenger_information[FlightID]),passenger_information[FlightID]=1003)

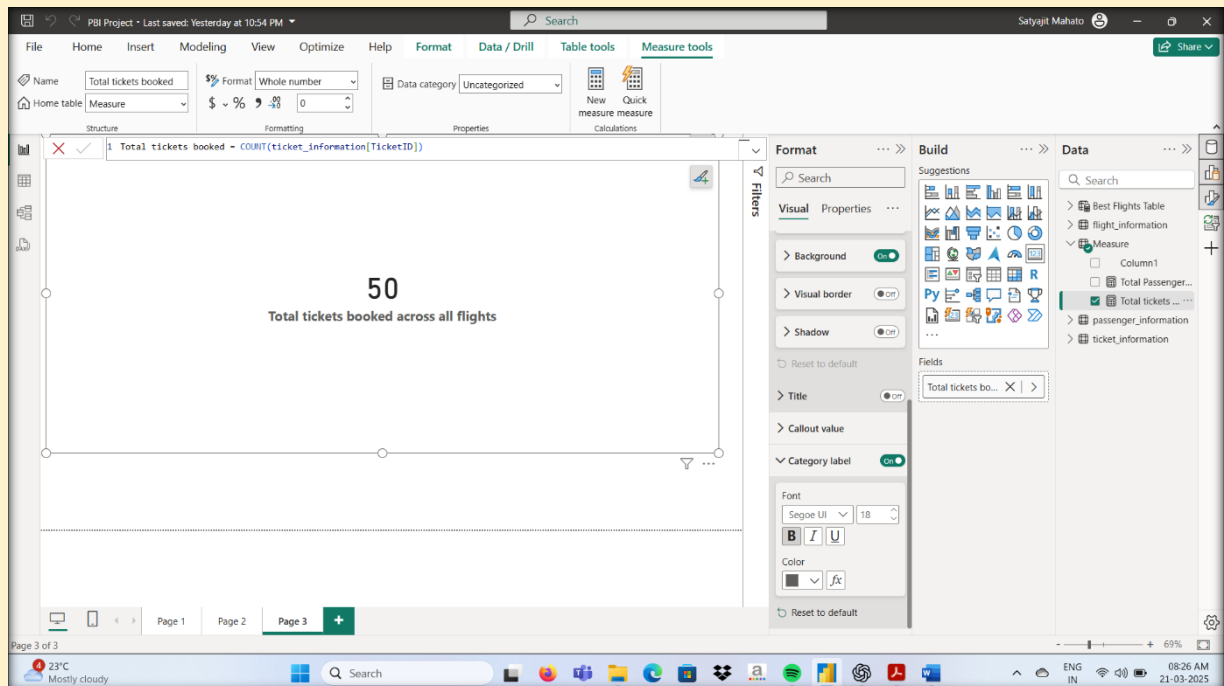
The screenshot shows a Power BI Desktop window with a DAX measure calculated. The measure is named 'Total Passenger of flight 1003' and is calculated using the formula: **CALCULATE(COUNT(passenger_information[FlightID]),passenger_information[FlightID]=1003)**. The result of the calculation is 2.

Measure	Value
Total Passenger of flight 1003	2

- Total tickets booked across all flights.

DAX function

Total tickets booked = COUNT(ticket_information[TicketID])



- Filtered table showing "Best" flights only.

DAX Function

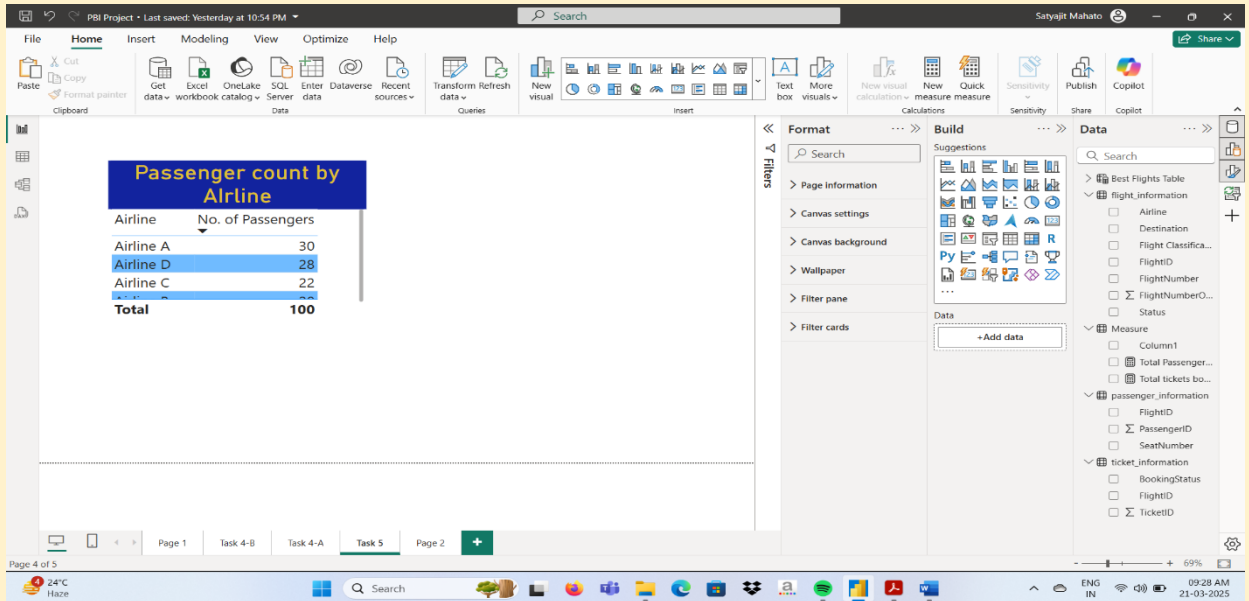
Best Flights Table = FILTER(flight_information,flight_information[Flight Classification]="Best")

The screenshot shows the Power BI Desktop interface with a table view of the 'Best Flights Table'. The table contains 82 rows of flight data, filtered by 'Flight Classification' = 'Best'. The columns are: FlightID, FlightNumber, Airline, Destination, Status, Flight Classification, and FlightNumberOnly. The bottom status bar indicates 'Table: Best Flights Table (82 rows)' and '24°C Mostly cloudy'.

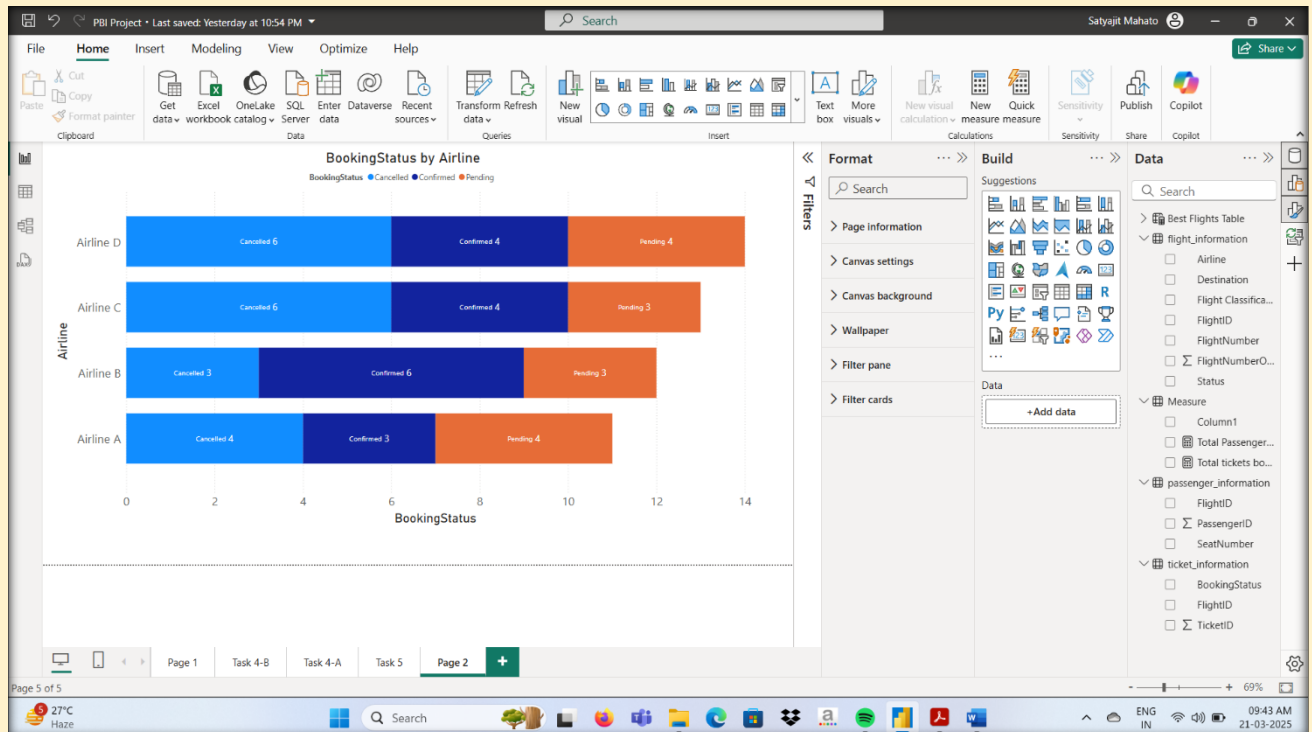
FlightID	FlightNumber	Airline	Destination	Status	Flight Classification	FlightNumberOnly
1001	FL1102	Airline D	Houston	On Time	Best	1102
1002	FL1435	Airline B	Chicago	On Time	Best	1435
1006	FL1071	Airline A	Phoenix	On Time	Best	1071
1011	FL1466	Airline A	Phoenix	On Time	Best	1466
1013	FL1330	Airline C	Houston	On Time	Best	1330
1020	FL1130	Airline A	New York	On Time	Best	1130
1023	FL1769	Airline A	Chicago	On Time	Best	1769
1025	FL1491	Airline D	Phoenix	On Time	Best	1491
1027	FL1805	Airline D	Chicago	On Time	Best	1805
1028	FL1385	Airline D	Chicago	On Time	Best	1385
1029	FL1191	Airline D	Los Angeles	On Time	Best	1191
1030	FL1955	Airline B	Phoenix	On Time	Best	1955
1031	FL1276	Airline B	New York	On Time	Best	1276
1033	FL1459	Airline D	New York	On Time	Best	1459
1034	FL1313	Airline B	Phoenix	On Time	Best	1313
1036	FL1252	Airline D	Phoenix	On Time	Best	1252
1039	FL1560	Airline B	Chicago	On Time	Best	1560
1043	FL1681	Airline C	Houston	On Time	Best	1681
1044	FL1475	Airline B	Phoenix	On Time	Best	1475
1046	FL1975	Airline D	Chicago	On Time	Best	1975
1048	FL1189	Airline A	New York	On Time	Best	1189
1050	FL1686	Airline C	Phoenix	On Time	Best	1686
1052	FL1562	Airline D	Phoenix	On Time	Best	1562
1053	FL1875	Airline C	Chicago	On Time	Best	1875
1055	FL1243	Airline B	New York	On Time	Best	1243
1057	FL1504	Airline A	Phoenix	On Time	Best	1504
1060	FL1818	Airline D	Chicago	On Time	Best	1818
1061	FL1646	Airline D	Los Angeles	On Time	Best	1646

Task 5. Visualization and Interactive Features

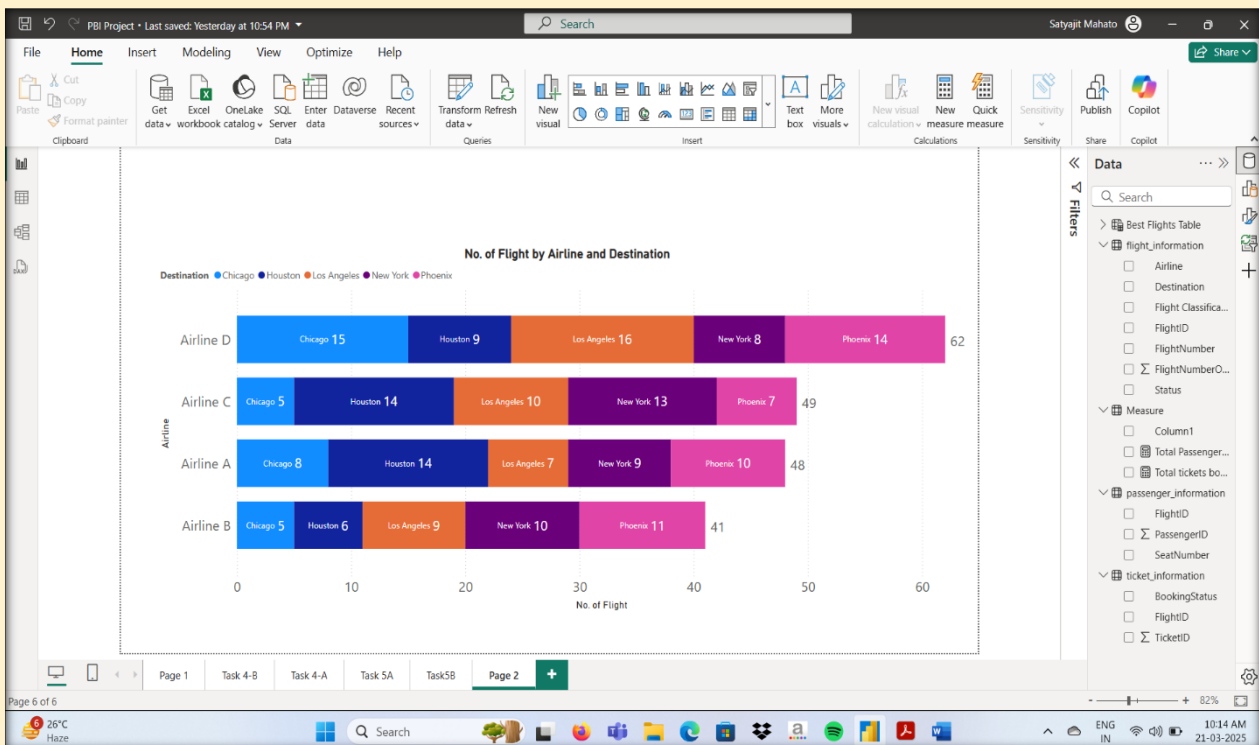
○ Passenger count by airline.



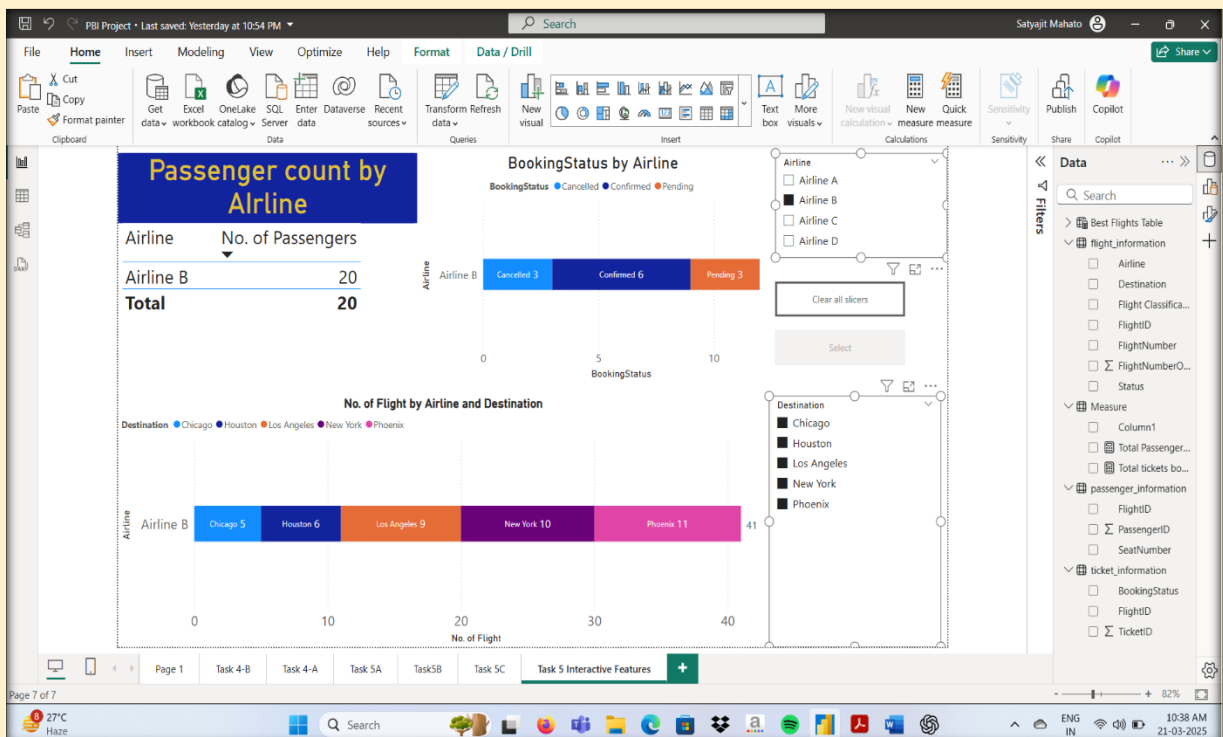
○ Ticket booking statuses.



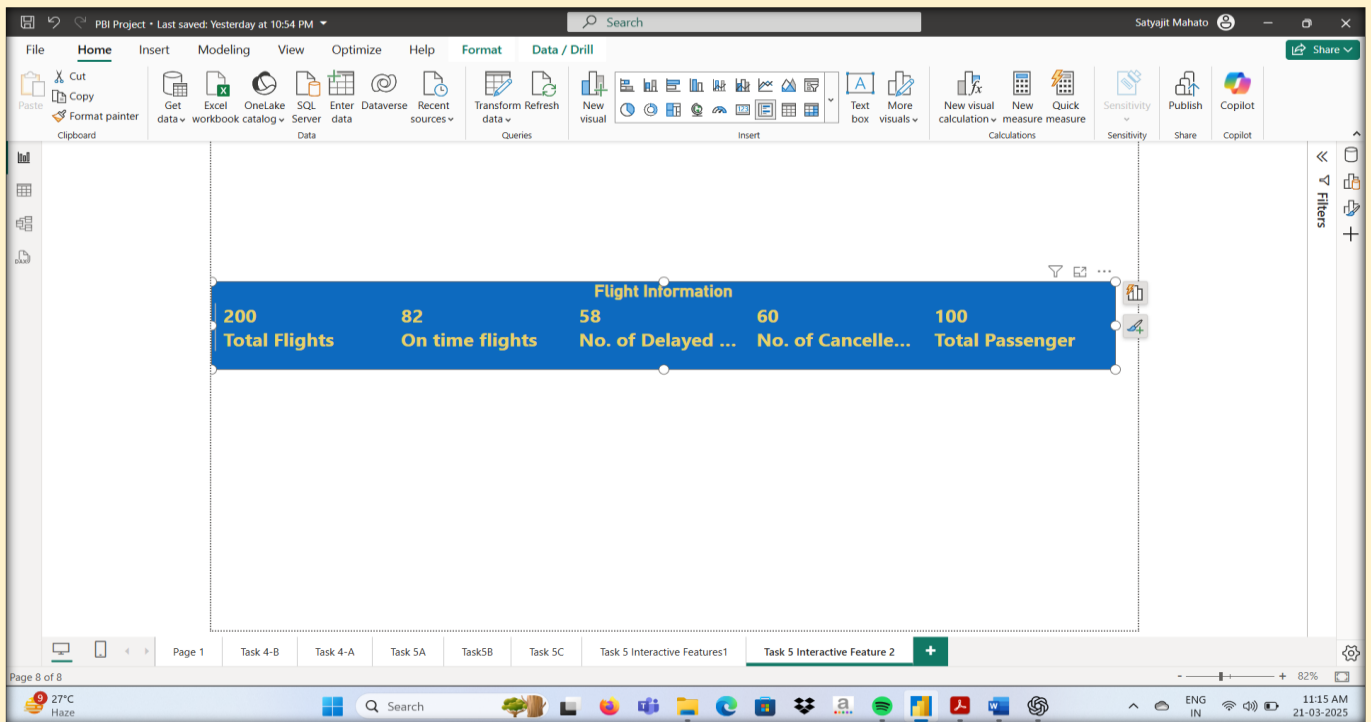
○ Flights by airline and destination.



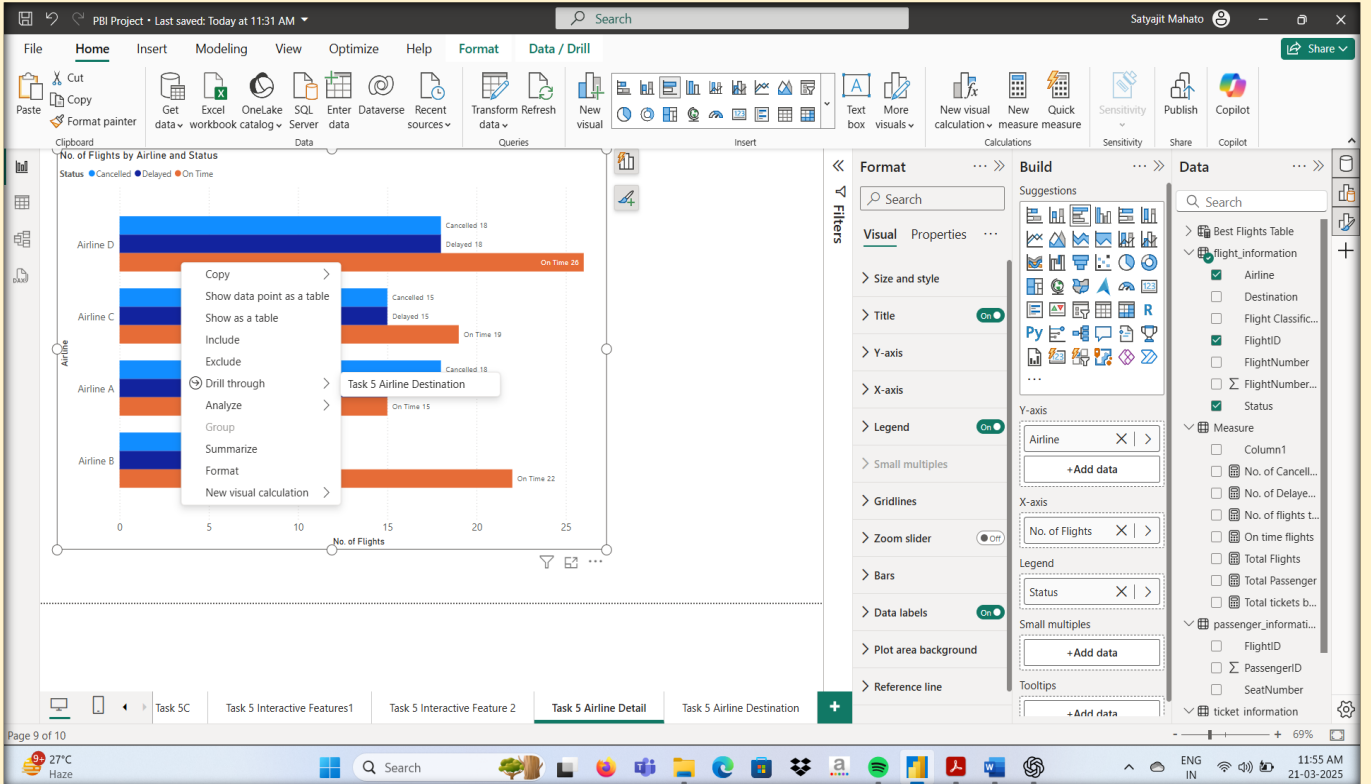
Interactive Feature Destination and Airline



Quick Views

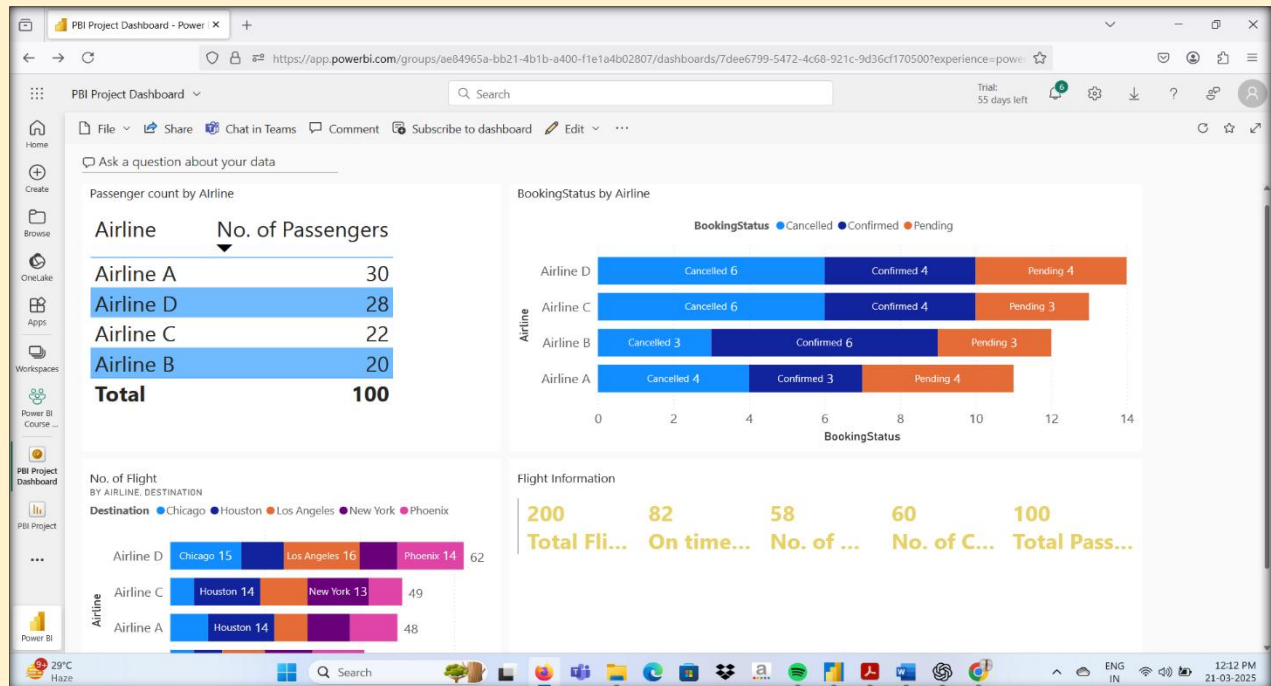


Airline Specific View using Drillthrough



Task 6. Final Dashboard and Power BI Service

● Dashboard with key visuals and insights.



● Configuration of Row-Level Security (RLS) for Airline A data

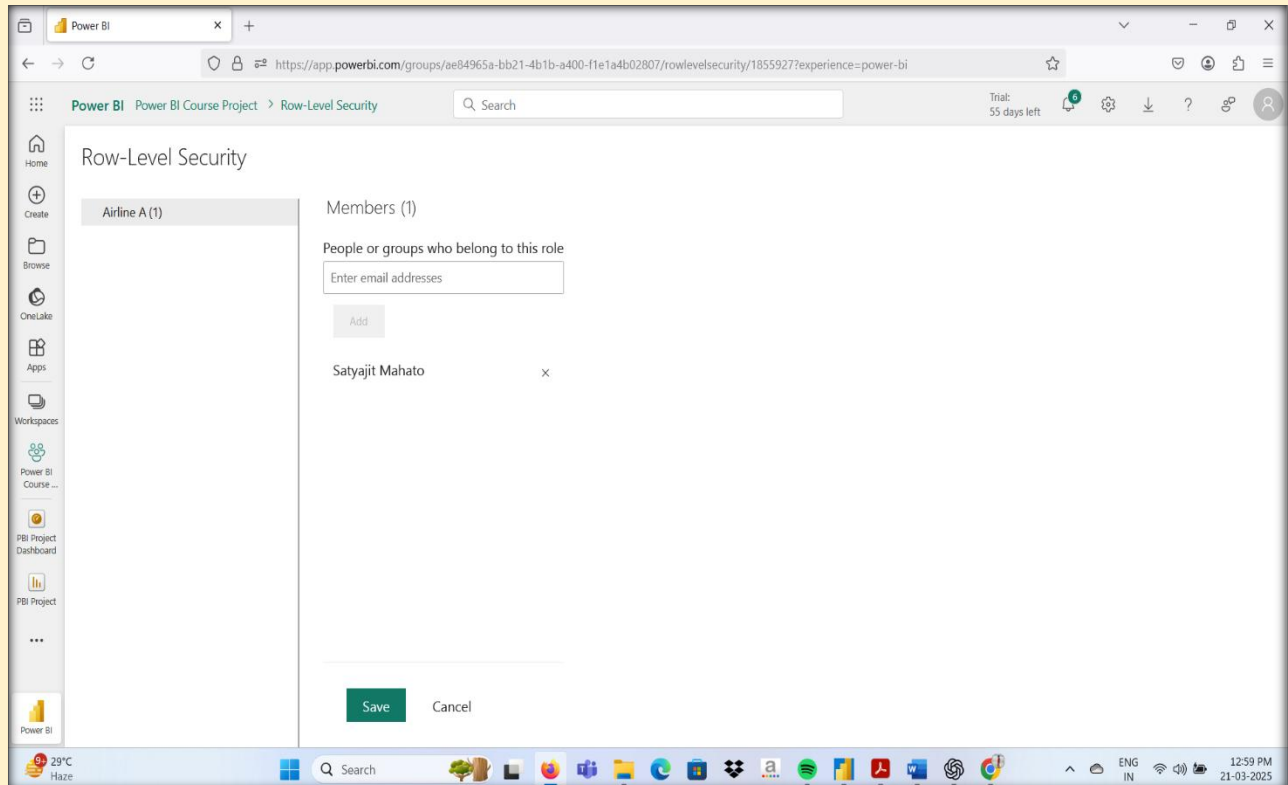
The 'Manage security roles' dialog box is open, showing the configuration for Row-Level Security (RLS) for Airline A data. The dialog includes the following sections:

- Roles:** A list of roles with a '+ New' button. The role 'Airline A' is selected.
- Select tables:** A list of tables with checkboxes. The table 'flight_informat...' is selected.
- Filter data:** A section for defining filters. The filter is set to 'Show data if All of these rules are true'. The rule is defined as:

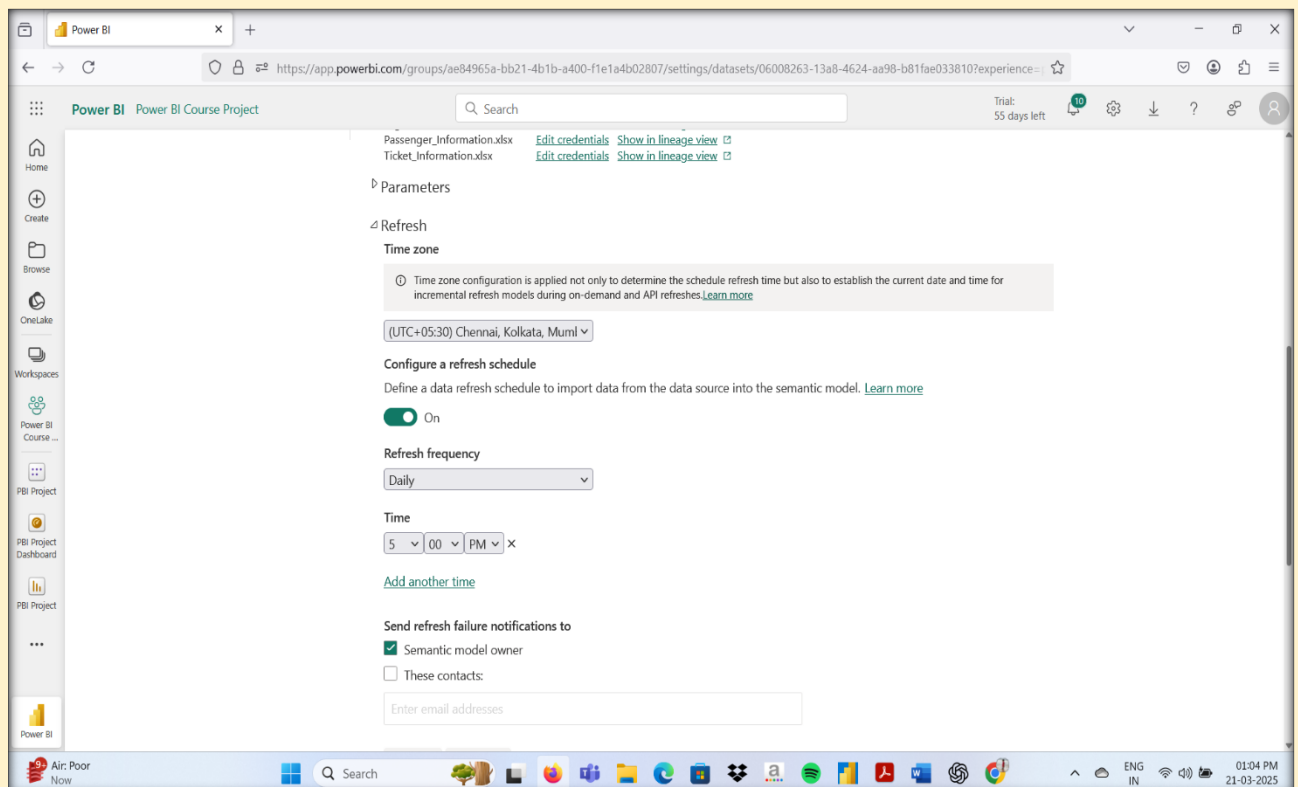
Column	Condition	Value
Airline	Equals	Airline A

The dialog also includes a 'Switch to DAX editor' button and 'Save' and 'Close' buttons at the bottom.

● Assigned to a user



● Set up a schedule refresh at 5 PM daily.



Key Findings:

1. Airline D operates highest no. of flights i.e. 62
2. 40 % flights to Los Angeles and 39 % flights to Chicago were cancelled
3. 48 % flights to Houston were delayed.
3. Airline A and D has highest number of cancellations
4. Total number of passengers were 100
5. Total tickets booked across all flights are 50
6. Total 19 tickets were cancelled
7. Total flights operated were 200
8. 82 flights were on time across all flights
9. 58 flights were delayed across all flights
10. 60 flights were cancelled across all flights
11. 53% flights of airline B are on time.
12. 37.5% flights of Airline A are cancelled
13. More than 25% flights are delayed on an average across all flights

Recommendations:

1. Several flights were delayed and cancelled adding to customer dissatisfaction. Investigate the reasons and Reduce flight delays and cancellation overall to improve customer satisfaction.
2. Investigate the reason behind ticket cancellation
3. Find out why Airline A and D has highest cancellations.
4. Find out the reasons behind delayed flights to Houston
5. Enhance customer satisfaction by improving on-time performance
6. Optimize flight scheduling based popular routes and occupancy rates

In conclusion, this project highlights how Power BI can revolutionize airline data management.

By leveraging data analytics, airlines can:

- ✓ Reduce delays and cancellations
- ✓ Improve ticketing efficiency
- ✓ Enhance passenger satisfaction

Video Explanation Link

https://drive.google.com/file/d/1D6LzT3i9e3lfmMdPR22TGWnAbXx35oqS/view?usp=drive_link