

=PROJECT

Airline Data Management and Analysis Using Power BI

Task-1:

Data Preparation and Cleaning (10 Marks)

- Extract and transform data in Power Query.
- Clean data: remove duplicates, handle missing values, and format columns.

For flight_information dataset:

The screenshot displays the Microsoft Power BI Desktop interface. The main window shows a table with 22 rows and 5 columns: FlightID, FlightNumber, Airline, Destination, and Status. The data is filtered to show the first 22 rows. The ribbon at the top includes tabs for Home, Transform, Add Column, View, Tools, and Help. The 'Transform' tab is active, showing various data manipulation options. On the right side, the 'Query Settings' pane is open, showing the 'APPLIED STEPS' section with 'Removed Duplicates' selected. The status bar at the bottom indicates '5 COLUMNS, 200 ROWS' and 'Column profiling based on top 1000 rows'.

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

For ticket_information:

The screenshot shows the Power BI Query Editor interface. The main view displays a table with 3 columns and 50 rows. The columns are TicketID, FlightID, and BookingStatus. The data is filtered to show rows 1 through 22. The 'Removed Duplicates' step is highlighted in the 'APPLIED STEPS' pane on the right. The 'Query Settings' pane on the right shows the 'Name' as 'ticket_information' and the 'Applied Steps' as 'Removed Duplicates 1'.

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

For passenger_information dataset:

The screenshot shows the Power BI Query Editor interface. The main view displays a table with 3 columns and 100 rows. The columns are PassengerID, FlightID, and SeatNumber. The data is filtered to show rows 1 through 22. The 'Removed Duplicates' step is highlighted in the 'APPLIED STEPS' pane on the right. The 'Query Settings' pane on the right shows the 'Name' as 'passenger_information' and the 'Applied Steps' as 'Removed Duplicates'.

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

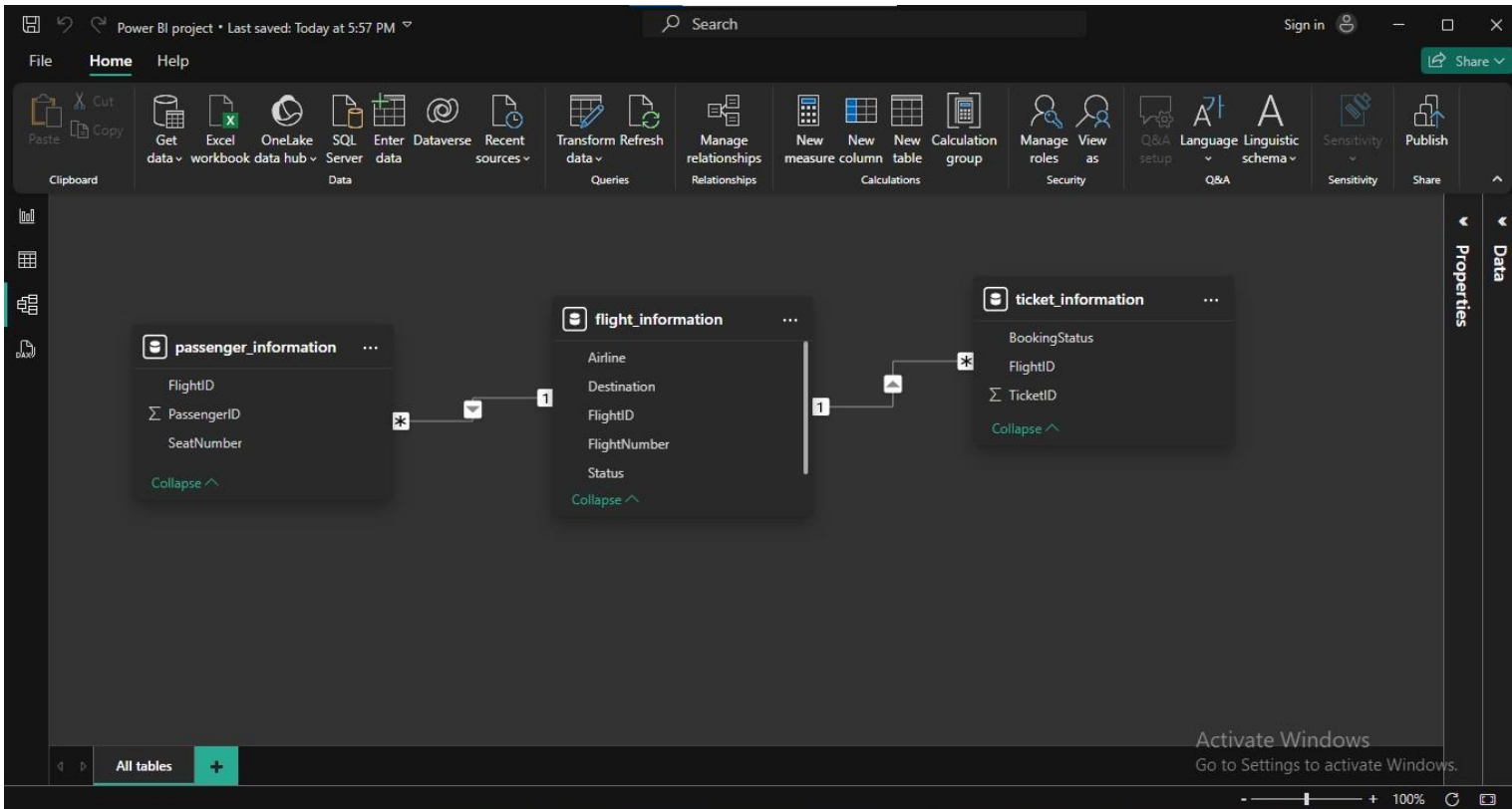
Steps Involved:

Power query editor > Select the dataset using ctrl + A > Home menu > Remove Duplicate Rows and Empty rows > changed the format of columns accordingly

Task-2:

Data Modeling

- Create relationships between datasets (FlightID as the key).
- Understand cardinality and configure the model appropriately



- Flight_information is forming one to many relationship with passenger_information and ticket_information.

Steps involved:

Model view > arrange the blocks in correct order> drag the flightID column from passenger_information and ticket_information and drop in flight_information

Task-3:

Enhanced Data Insights

- Add a conditional column to classify flights as "Best" or "To Be Improved" based on status.
- Use "Column from Examples" to extract the flight number from FlightNumber.

Power BI project

Home Transform **Add Column** View Tools Help

Column From Examples Custom Column Invoke Custom Function Conditional Column Index Column Duplicate Column

Format Extract Parse Merge Columns

Statistics Standard Scientific Trigonometry Rounding

Date Time Duration From Date & Time

Text Analytics Vision Azure Machine Learning AI Insights

Queries [3]

passenger_information ticket_information **flight_information**

fx = Table.ReorderColumns(#"Renamed Columns",{"FlightID", "FlightNumber", "Extracted_Flight_Number", "Airline", "Status", "Flight_Remark"})

	Extracted_Flight_Number	Airline	Destination	Status	Flight_Remark
1	1102	Airline D	Houston	On Time	Best
2	1435	Airline B	Chicago	On Time	Best
3	1860	Airline A	New York	Cancelled	To Be Improved
4	1270	Airline C	Chicago	Delayed	To Be Improved
5	1106	Airline C	New York	Delayed	To Be Improved
6	1071	Airline A	Phoenix	On Time	Best
7	1700	Airline C	Los Angeles	Cancelled	To Be Improved
8	1020	Airline C	Los Angeles	Delayed	To Be Improved
9	1614	Airline A	Los Angeles	Cancelled	To Be Improved
10	1121	Airline D	Chicago	Cancelled	To Be Improved
11	1466	Airline A	Phoenix	On Time	Best
12	1214	Airline D	New York	Delayed	To Be Improved
13	1330	Airline C	Houston	On Time	Best
14	1458	Airline C	New York	Delayed	To Be Improved
15	1087	Airline C	Houston	Delayed	To Be Improved
16	1372	Airline B	New York	Delayed	To Be Improved
17	1099	Airline D	Phoenix	Delayed	To Be Improved
18	1871	Airline B	Houston	Delayed	To Be Improved
19	1663	Airline B	Chicago	Cancelled	To Be Improved
20	1130	Airline A	New York	On Time	Best
21	1661	Airline B	New York	Cancelled	To Be Improved
22	1308	Airline A	Houston	Delayed	To Be Improved

Query Settings

PROPERTIES

Name: flight_information

All Properties

APPLIED STEPS

- Source
- Navigation
- Promoted Headers
- Changed Type
- Removed Other Columns
- Removed Duplicates
- Added Conditional Column
- Inserted Text After Delimiter
- Renamed Columns
- Reordered Columns**

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

PREVIEW DOWNLOADED AT 5:53 PM

Steps involved:

Power query editor > Flight_information > Select Status column > Add column menu > conditional column > If Status = "On Time" then output = "Best" Otherwise = "To Be Improved" > Select Flight_Number column > column from example > 1102 > 1435 > Rename the column > change the format of the column accordingly

Task-4:

Calculations Using DAX

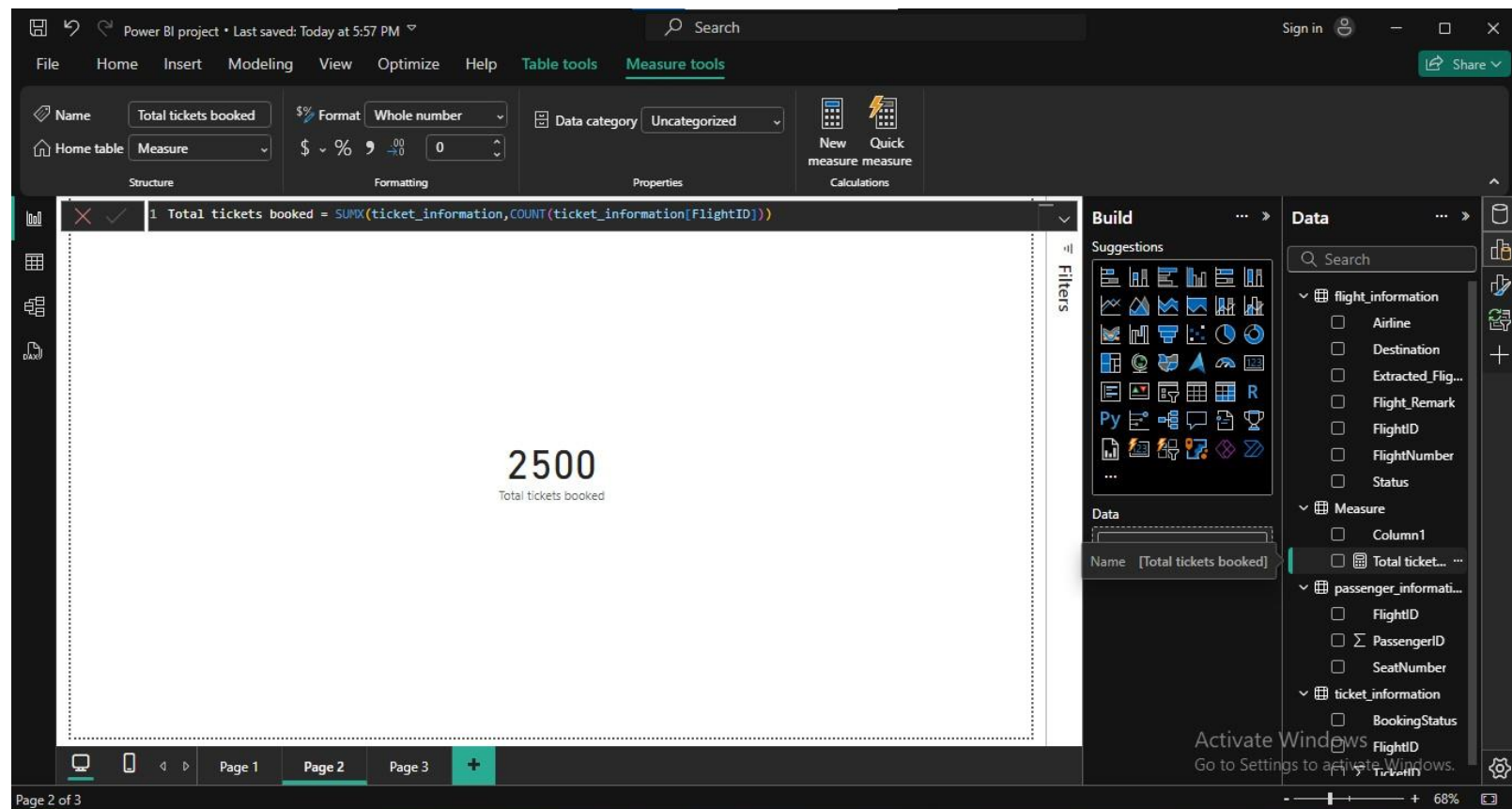
- Calculate:
 - Total passengers for a specific flight.



Steps involved:

Report view > Right click on measure > New measure > Total passengers =
DISTINCTCOUNT(passenger_information[PassengerID]) > Select table visual > Add data =
FlightID, Total passengers.

o Total tickets booked.



Steps Involved:

Report view > Right click on measure > New measure > Total tickets booked = `SUMX(ticket_information,COUNT(ticket_information[FlightID]))` > select single-row card > add data = Total tickets booked

o Filtered table showing "Best" flights only

The screenshot displays the Microsoft Power BI Desktop interface. The main workspace shows a table visual with the following data:

FlightID	Filtered Flights
1001	Best
1002	Best
1006	Best
1011	Best
1013	Best
1020	Best
1023	Best
1025	Best
1027	Best
1028	Best
1029	Best
1030	Best
1031	Best
1033	Best
1034	Best
1036	Best
1039	Best
1043	Best
1044	Best

The interface includes a top ribbon with tabs: File, Home, Insert, Modeling, View, Optimize, Help, Format, Data / Drill, Table tools, and Column tools. The 'Column tools' tab is active, showing options for Name, Data type, Format, Summarization, Data category, Sort, Data groups, Manage relationships, and New column. A right-hand pane shows the 'Format' tab with various styling options. A bottom status bar indicates 'Page 3 of 8' and a zoom level of 68%.

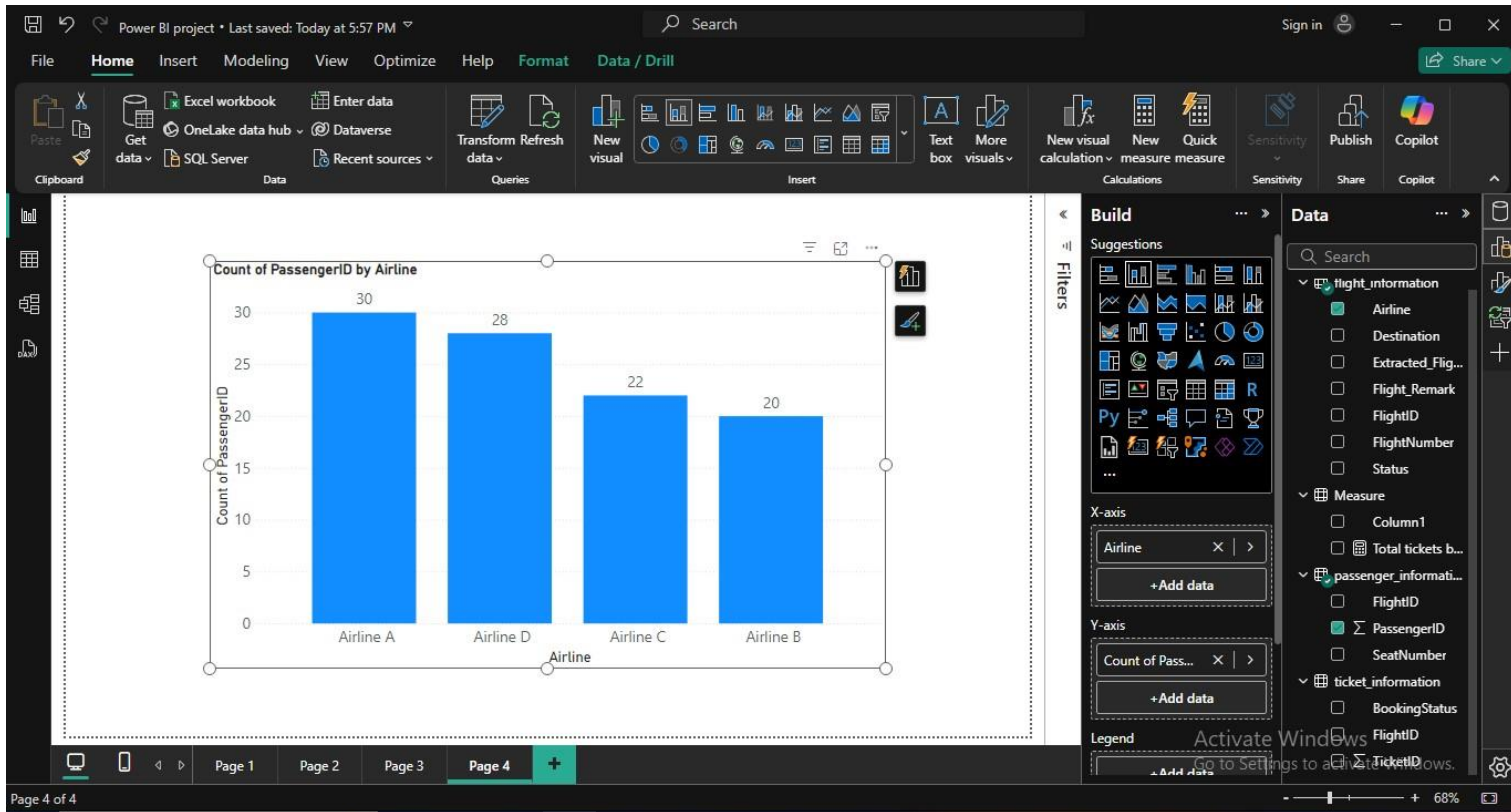
Steps Involved:

Report view > Right click on measure > New measure > Filtered Flights =
FILTER(Flight_information, Flight_information[Flight_Remarks]="Best") >Table View > add
Data = FlightID, Filtered Flights

Task-5:

Visualization and Interactive Features

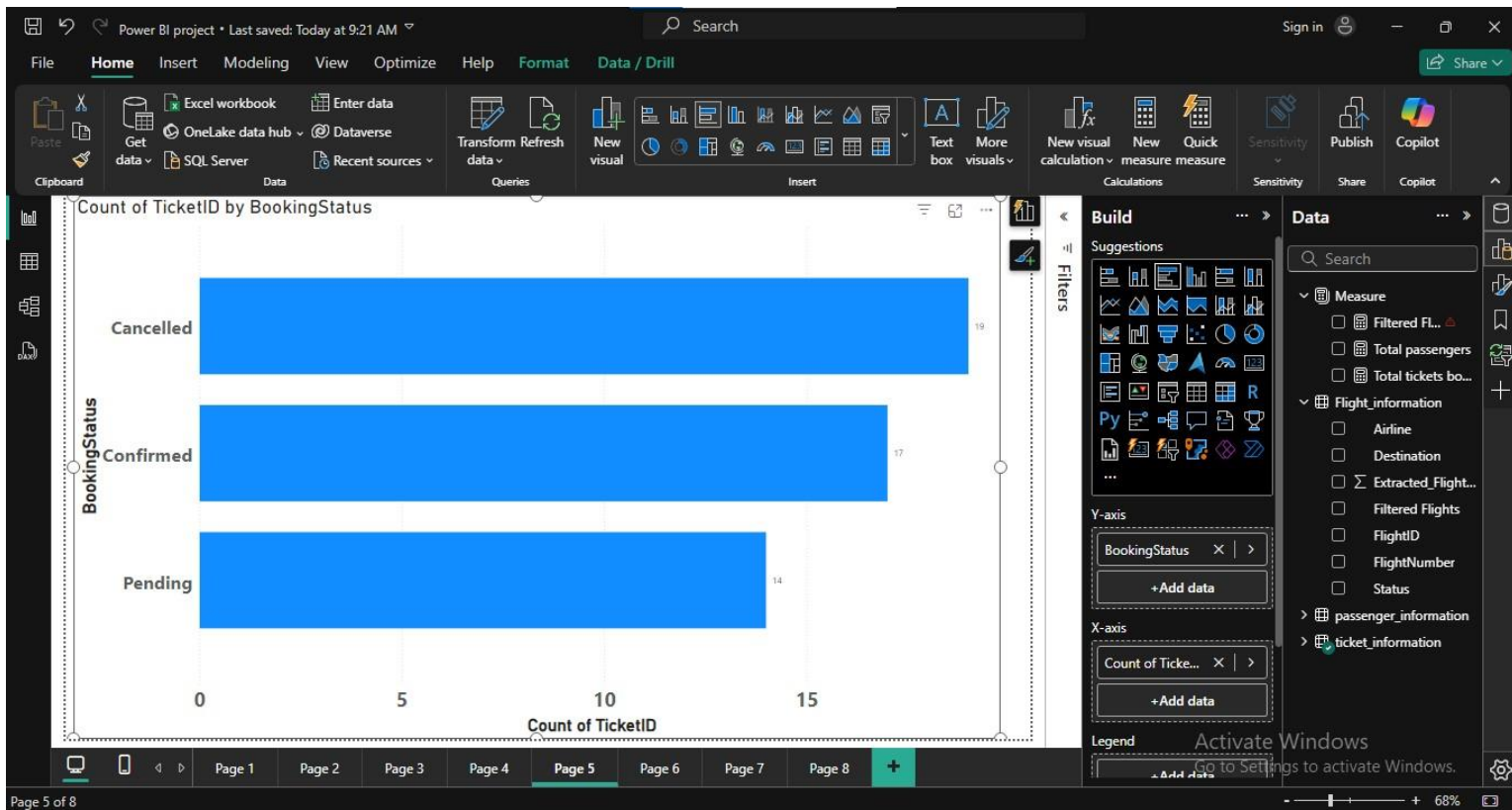
- Create visuals for:
 - Passenger count by airline.



Steps involved:

Report view > Select stacked bar chart > X-axis = Airlines > Y-axis= Count of passengers

o Ticket booking statuses.



Steps involved:

Report view > Select clustered bar chart > Y-axis= BookingStatus > X-axis = Count of TicketID

o Flights by airline and destination.

The screenshot displays the Microsoft Power BI Desktop interface. The main area shows a table titled "Flights by airline and destination" with the following data:

Airline	Destination	FlightID
Airline A	Chicago	1023
Airline A	Chicago	1037
Airline A	Chicago	1068
Airline A	Chicago	1095
Airline A	Chicago	1124
Airline A	Chicago	1146
Airline A	Chicago	1173
Airline A	Chicago	1182
Airline A	Houston	1022
Airline A	Houston	1042
Airline A	Houston	1047
Airline A	Houston	1054
Airline A	Houston	1056
Airline A	Houston	1086
Airline A	Houston	1092
Airline A	Houston	1111
Airline A	Houston	1115
Airline A	Houston	1125
Airline A	Houston	1156
Airline A	Houston	1168

The right-hand pane is the "Build" pane, which is currently showing the "Data" section. It includes a search bar and a list of fields that can be added to the report. The fields are organized into categories:

- Measure
 - ☐ Filtered FL...
 - ☐ Total passengers
 - ☐ Total tickets bo...
- Flight_information
 - ☐ Airline
 - ☐ Destination
 - ☐ Σ Extracted_Flight...
 - ☐ Filtered Flights
 - ☐ FlightID
 - ☐ FlightNumber
 - ☐ Status
- passenger_information
- ticket_information

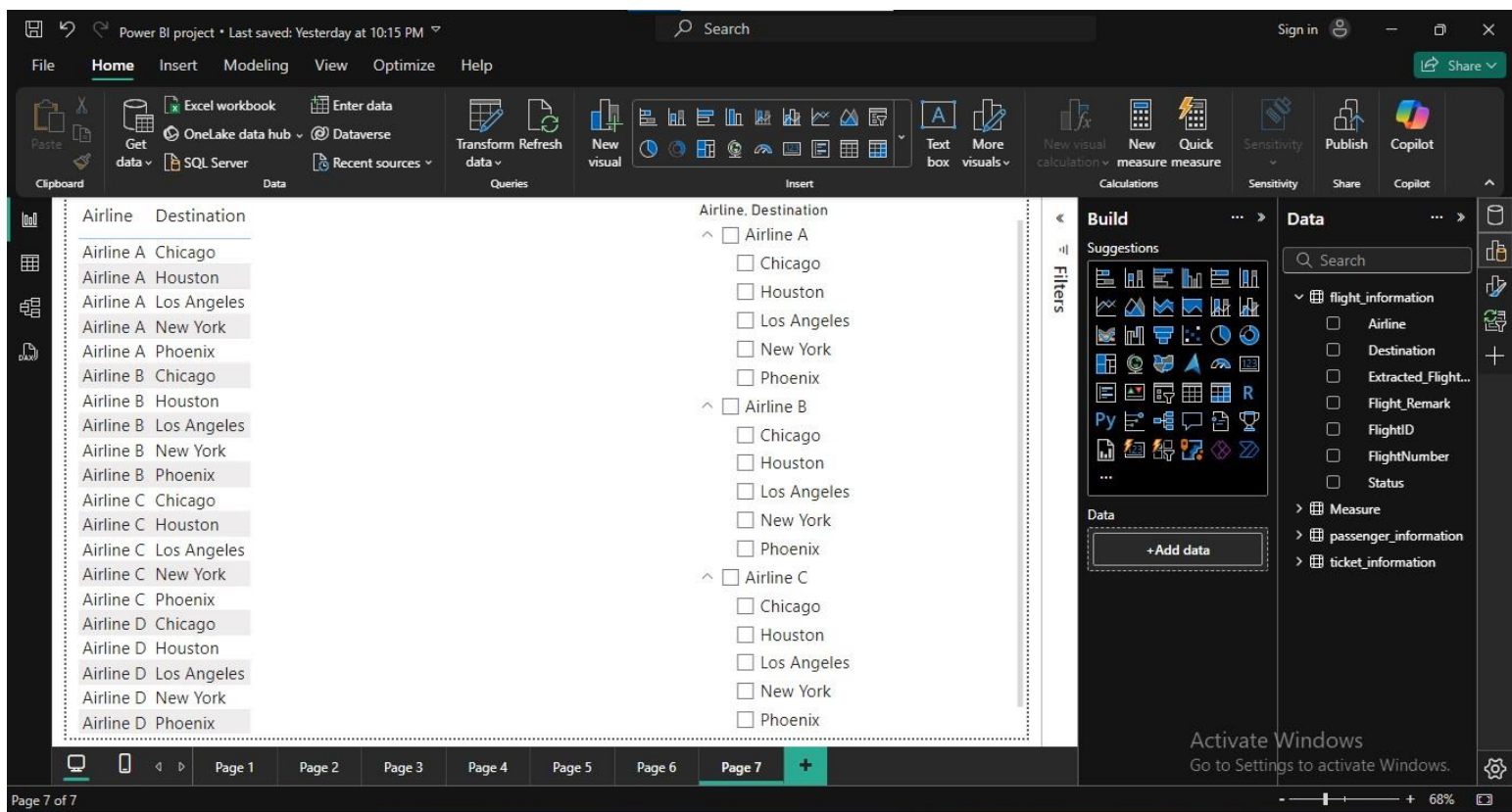
The bottom of the screen shows the "Page 6 of 8" indicator and the "Activate Windows" watermark.

Steps involved:

Report view > Select table > add data = Airline, Destination, FlightID

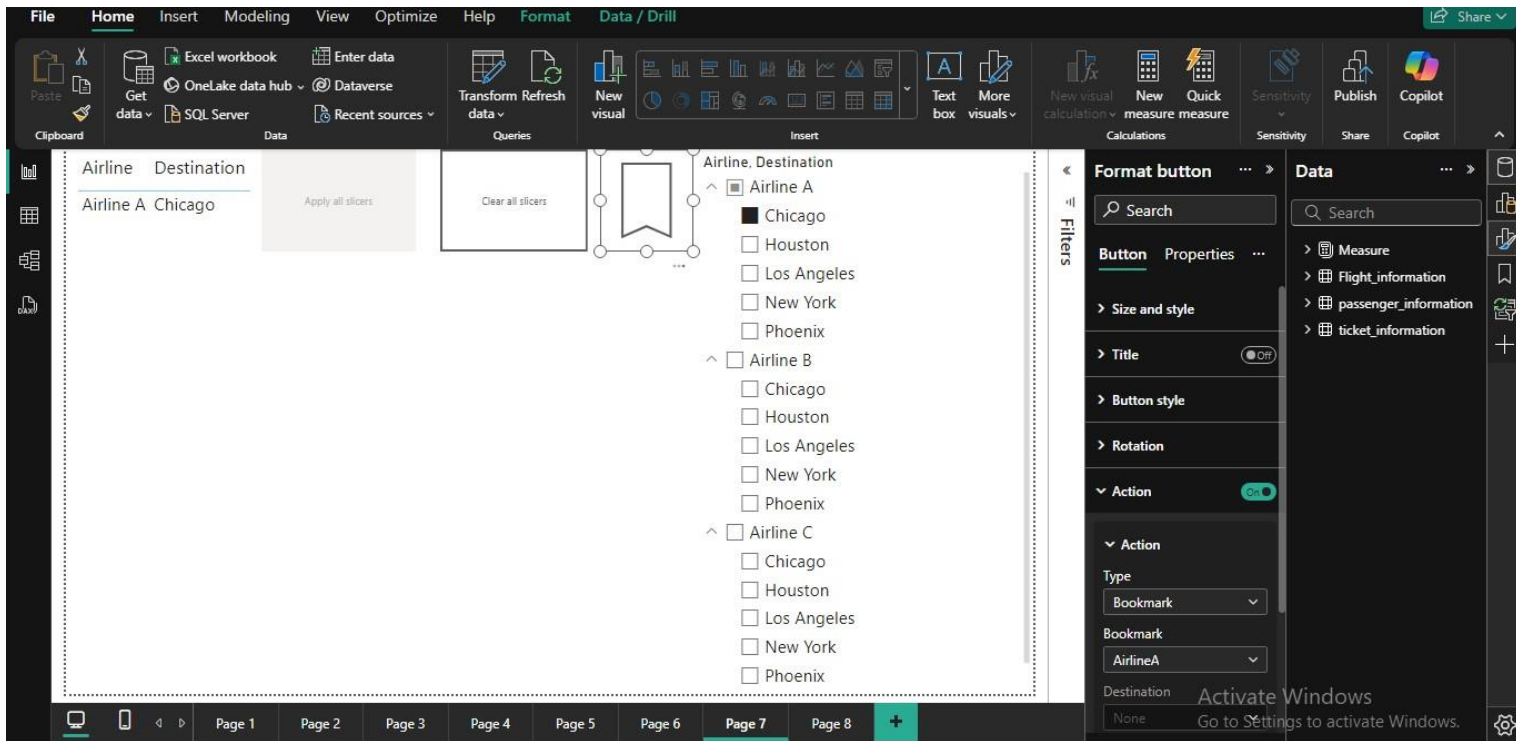
- Add interactive features for:

○ Destination and Airline.



Report view > select Table visual > Add data = Airlines, Destination > Select slicer visual > fields = Airlines, Destination

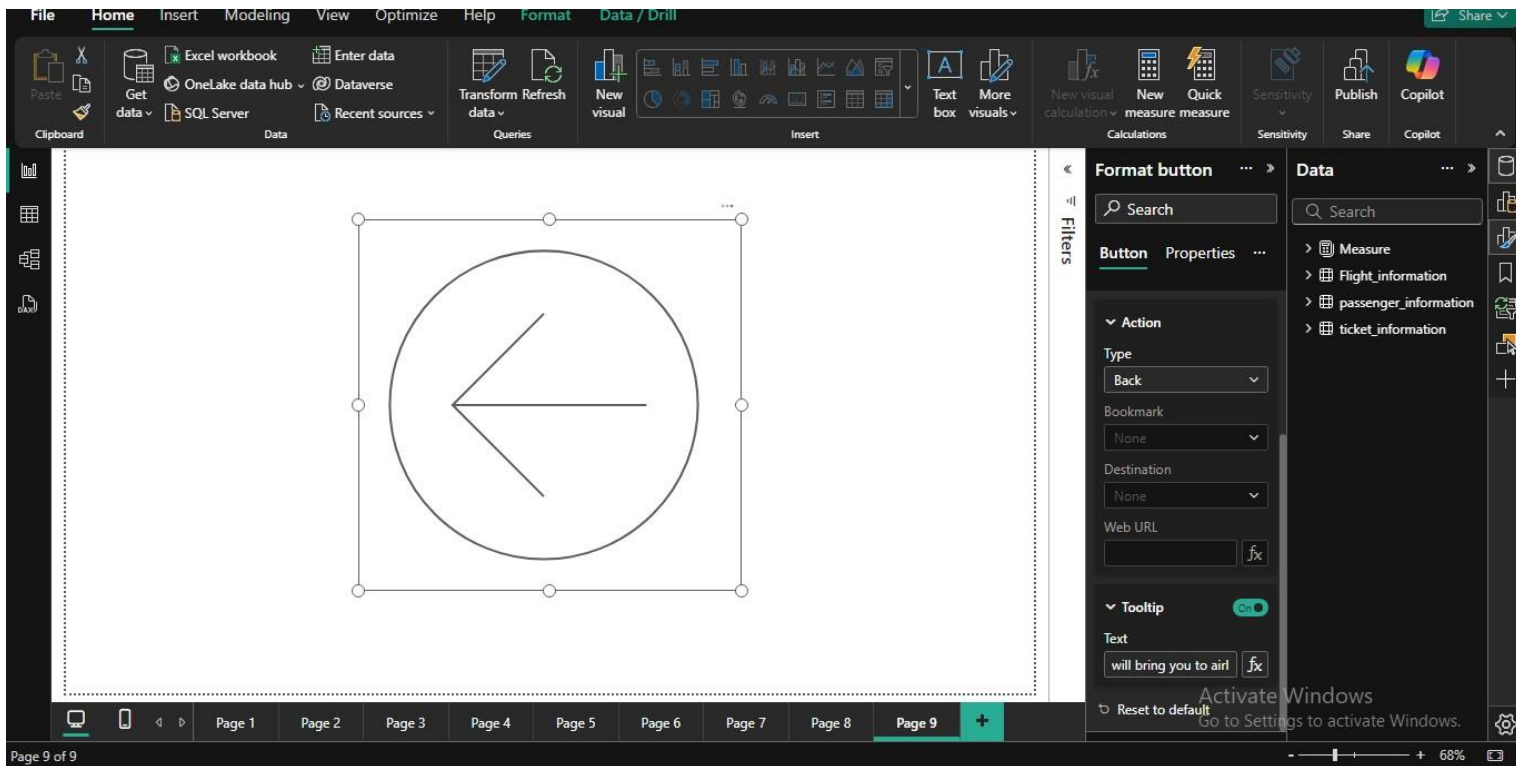
o Quick views.



Steps involved:

Report view > Select table visual > Add data > Airlines, Destination > Select slicer visual > fields = Airlines, Destination > Select Chicago in Airline > view menu > Bookmark > Add > Rename to AirlineA > Insert menu > button > Bookmark > format pane of bookmark > bookmark = AirlineA > Button > Apply all slicer, Clear all slicers

o Airline-specific pages



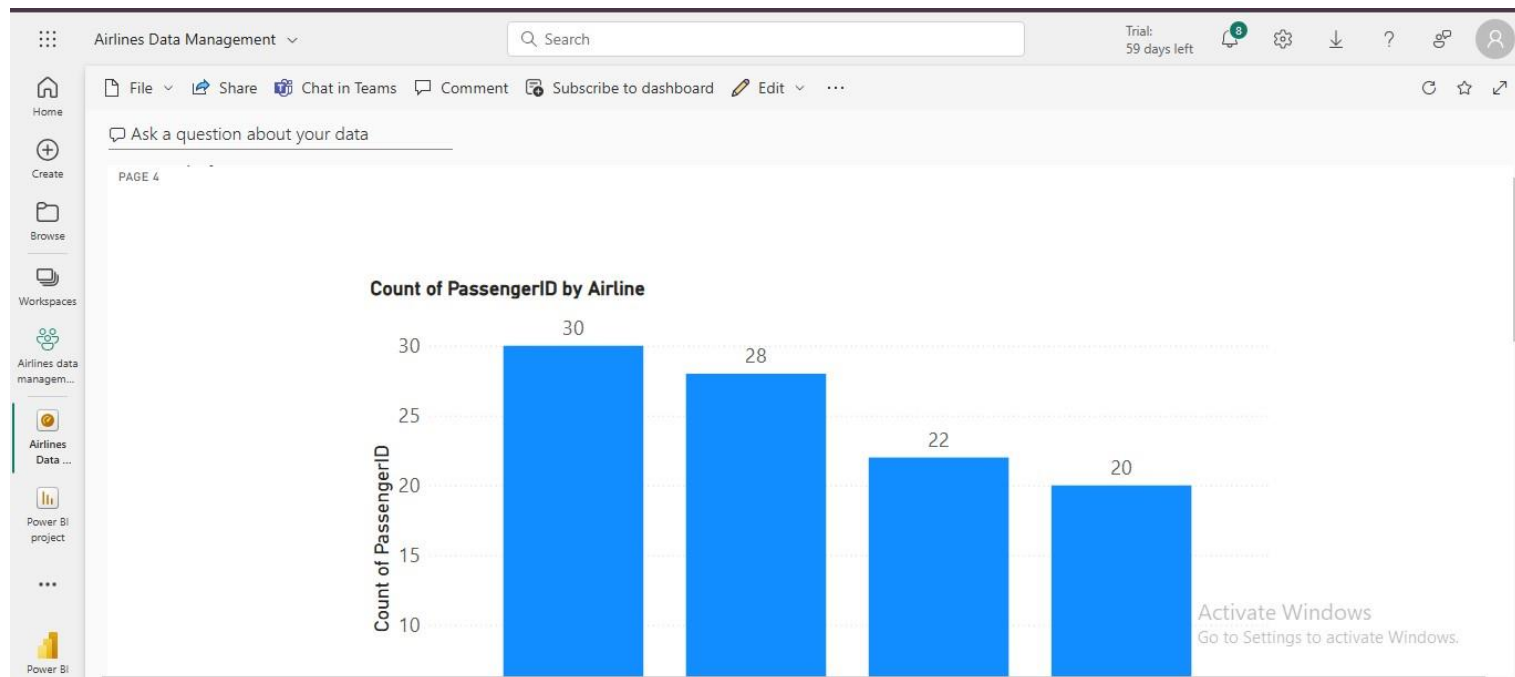
Steps involved:

Report view > Insert menu > button > back > format pane > tooltip > text = will bring you to airline specific pages.

Task-6:

Final Dashboard and Power BI Service

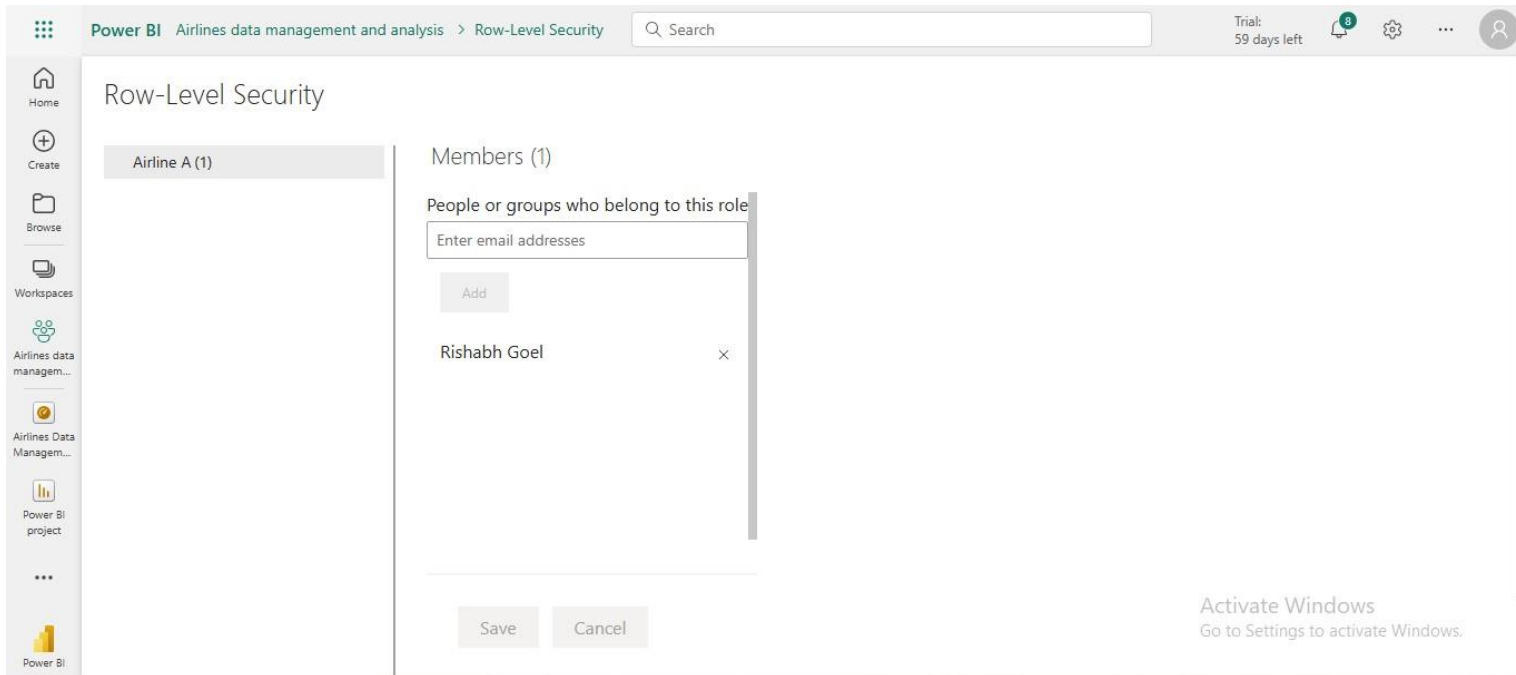
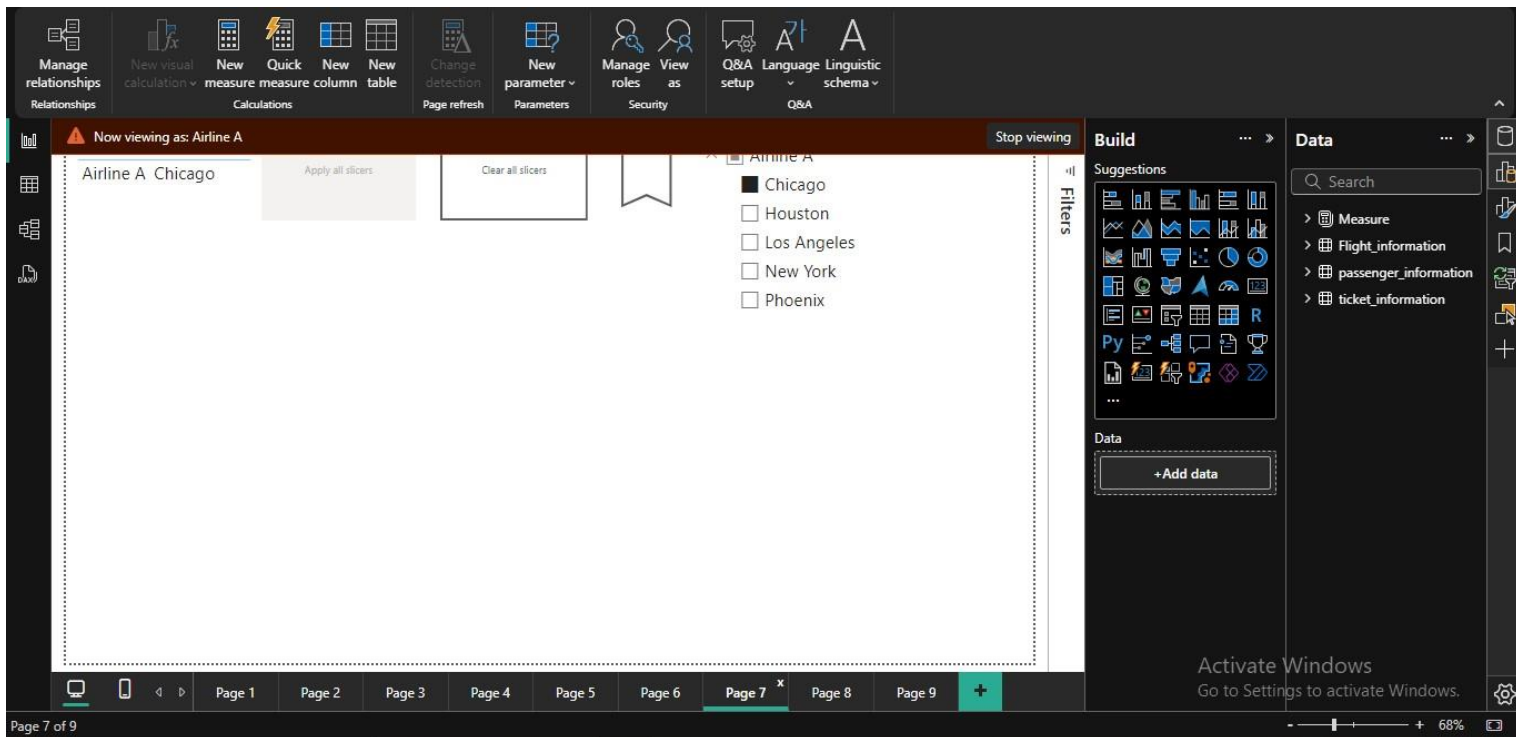
- Design a comprehensive dashboard with key visuals and insights.



Steps Involved:

PowerBI Service > Workspace > Report > ... > Pin to a Dashboard > New Dashboard > Airlines Data Management > Pin Live

- Configure Row-Level Security (RLS) for Airline A data and assign it to a user.



Steps Involved:

PowerBI Desktop > Modelling > Manage Roles > Airline A > Home > Publish > Airlines Data Management and analysis > select

PowerBI Service > Wrokspace > Semantic View of dataset > Security > Airline A > add the mail address of the person need to assign the file

- Set up a schedule refresh at 5 PM daily.

Power BI Airlines data management and analysis

Search

Trial: 59 days left

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+05:30) Chennai, Kolkata, Mumt

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:

Enter email addresses

Apply Discard

Activate Windows
Go to Settings to activate Windows.

PowerBI Service > Create New Workspace = Airlines data management and analysis > Installed Data gateway > Settings > Semantic Model > Refresh Frequency = Daily > Time = 5:00 PM

Link for the video:

https://drive.google.com/file/d/1ZxvGKyK7EWk3-KqHmW7RTO0RSE_tL9ij/view?usp=sharing