

Airline Data Management and Analysis Using Power BI

By: Hardik Verma

1. Data Preparation and Cleaning

I used the **Power Query Editor** to clean and prepare the data. All the 3 datasets given had null columns which I removed using the **remove columns** option in the Home menu.

I ensured that all the columns had the proper format and handled missing values and duplicates.

The screenshot shows the Power Query Editor interface with the 'flight_information' query selected. The main pane displays a table with columns: FlightID, FlightNumber, Airline, Destination, and Status. The 'Applied Steps' pane on the right shows the 'Removed Other Columns' step, indicating that other columns were removed during the preparation process. The 'Properties' pane shows the name is set to 'flight_information'.

The screenshot shows the Power Query Editor interface with the 'ticket_information' query selected. The main pane displays a table with columns: TicketID, FlightID, and BookingStatus. The 'Applied Steps' pane on the right shows the 'Removed Other Columns' step, indicating that other columns were removed during the preparation process. The 'Properties' pane shows the name is set to 'ticket_information'.

Queries [3]

flight_information

ticket_information

passenger_information

Table.SelectColumns(#"Changed Type", {"PassengerID", "FlightID", "SeatNumber"})

	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

Query Settings

Properties

Name: passenger_information

Applied Steps

- Source
- Navigation
- Promoted Headers
- Changed Type
- Removed Other Columns

2. Data Modeling

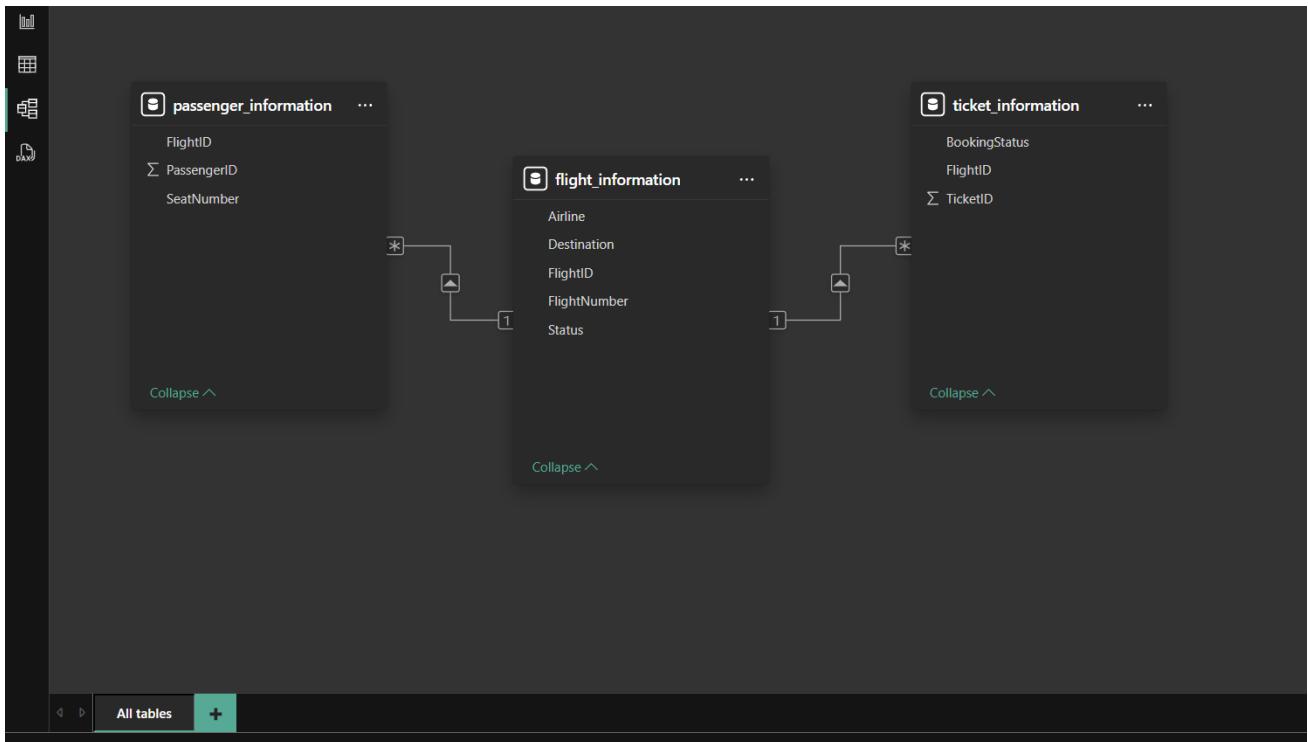
I created **two relationships** that are as follows:

1. Between **flight_information** and **ticket_information** on the basis of **FlightID column**. There is a **1-to-many relationship** from **flight_information** to **ticket_information** and **FlightID** is the key.
2. Between **flight_information** and **passenger_information** on the basis of **FlightID column**. There is a **1-to-many relationship** from **flight_information** to **passenger_information** and **FlightID** is the key.

Manage relationships

+ New relationship Autodetect Edit Delete Filter

From: table (column)	Relationship	To: table (column)	Status
passenger_information (FlightID)	* —> 1	flight_information (FlightID)	Active
ticket_information (FlightID)	* —> 1	flight_information (FlightID)	Active

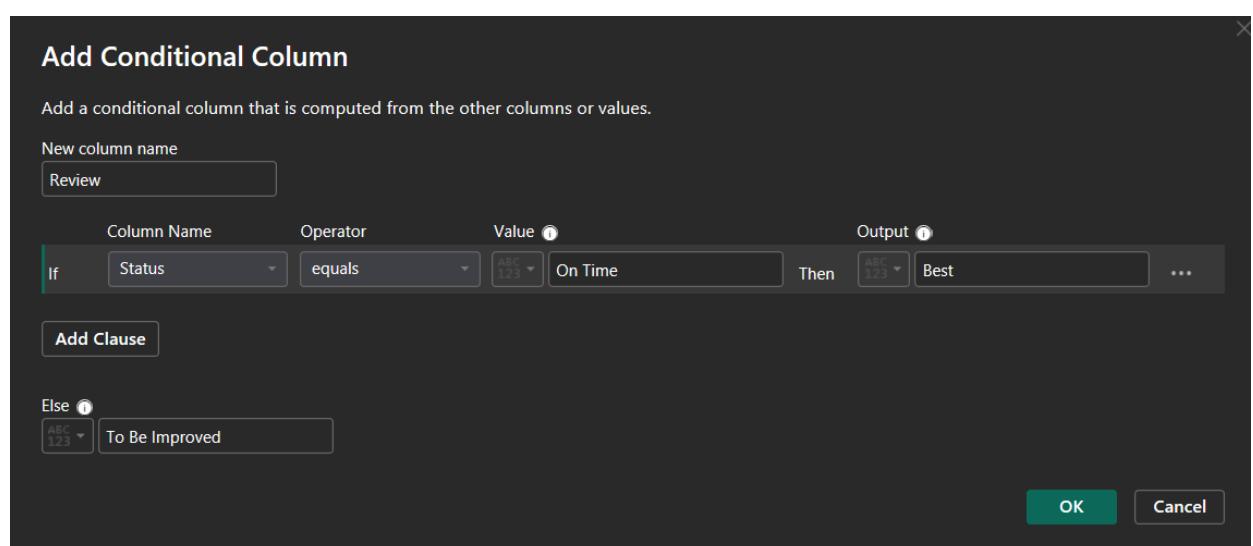


3. Enhanced Data Insights

I added a **Review** column using the **conditional column** option in the **add column** menu of the **Power Query Editor**.

I defined the condition that if the Status is “On Time”, then that flight is “Best” and otherwise, in the case of “Delayed” and “Cancelled”, the flight is “To Be Improved”.

Finally, I corrected the format of the Review column to Text.



Queries [3]

	FlightNumber	Airline	Destination	Status	Review
1	1001	Airline D	Houston	On Time	Best
2	1002	Airline B	Chicago	On Time	Best
3	1003	Airline A	New York	Cancelled	To Be Improved
4	1004	Airline C	Chicago	Delayed	To Be Improved
5	1005	Airline C	New York	Delayed	To Be Improved
6	1006	Airline I	Phoenix	On Time	Best
7	1007	Airline C	Los Angeles	Cancelled	To Be Improved
8	1008	Airline C	Los Angeles	Delayed	To Be Improved
9	1009	Airline A	Los Angeles	Cancelled	To Be Improved
10	1010	Airline D	Chicago	Cancelled	To Be Improved
11	1011	Airline A	Phoenix	On Time	Best
12	1012	Airline D	New York	Delayed	To Be Improved
13	1013	Airline C	Houston	On Time	Best
14	1014	Airline C	New York	Delayed	To Be Improved
15	1015	Airline C	Houston	Delayed	To Be Improved
16	1016	Airline B	New York	Delayed	To Be Improved
17	1017	Airline D	Phoenix	Delayed	To Be Improved
18	1018	Airline B	Houston	Delayed	To Be Improved
19	1019	Airline B	Chicago	Cancelled	To Be Improved
20	1020	Airline A	New York	On Time	Best

4. Calculations Using DAX

Name: Total Passangers of FlightID 1133

Home table: Measures (2)

Format: Whole number

Data category: Uncategorized

Measure:

```
Total Passangers of FlightID 1133 = CALCULATE(COUNT(passenger_information[PassengerID]), passenger_information[FlightID]= 1133)
```

Properties:

Calculated Column:

Search:

Filters on this page:

Add data fields here

Filters on all pages:

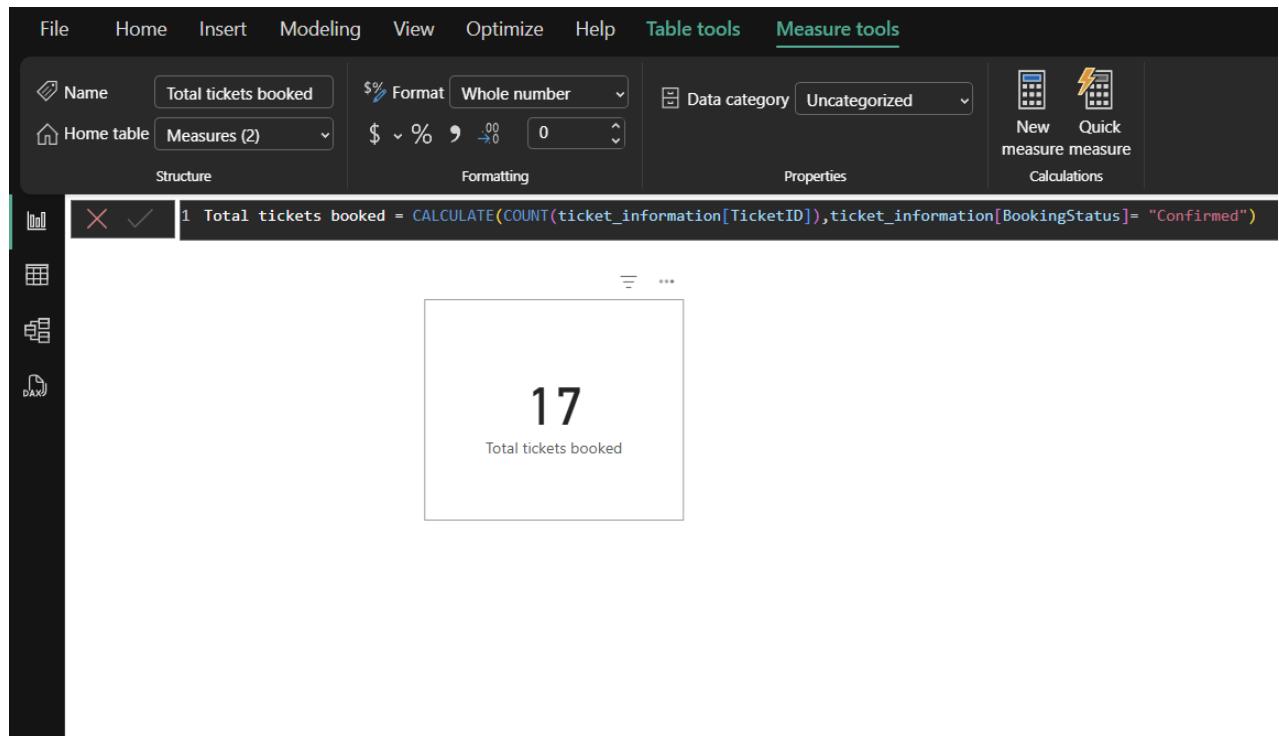
Add data fields here

To calculate the Total passengers for the flight having **FlightID 1133**, I created a **measure** and visualized it with the **card visual** from the **visualization pane**.

I used the DAX functions **CALCULATE** and **COUNT** to structure the formula. The DAX formula is as follows;

Total Passangers of FlightID 1133 =
CALCULATE(COUNT(passenger_information[PassengerID]),
passenger_information[FlightID]= 1133)

I found that we have the information of 2 passengers travelling in flight having FlightID 1133.



To calculate the Total tickets booked, I created a **measure using the CALCULATE and COUNT DAX functions**. I counted all the tickets that have the **BookingStatus** as "Confirmed".

The DAX formula is as follows;

Total tickets booked =

CALCULATE(COUNT(ticket_information[TicketID]),ticket_information[BookingStatus]= "Confirmed")

I found that there are a total of 17 tickets that are booked and confirmed out of all the ticket information we have.

File Home Help **Table tools**

Name: Best flights

Structure

Manage relationships Relationships

New measure New table Calculations

Mark as date table Calendars

1 Best flights = FILTER(flight_information, flight_information[Review]= "Best")

FlightID	FlightNumber	Airline	Destination	Status	flight_no	Review
1001	FL1102	Airline D	Houston	On Time	1102	Best
1002	FL1435	Airline B	Chicago	On Time	1435	Best
1006	FL1071	Airline A	Phoenix	On Time	1071	Best
1011	FL1466	Airline A	Phoenix	On Time	1466	Best
1013	FL1330	Airline C	Houston	On Time	1330	Best
1020	FL1130	Airline A	New York	On Time	1130	Best
1023	FL1769	Airline A	Chicago	On Time	1769	Best
1025	FL1491	Airline D	Phoenix	On Time	1491	Best
1027	FL1805	Airline D	Chicago	On Time	1805	Best
1028	FL1385	Airline D	Chicago	On Time	1385	Best
1029	FL1191	Airline D	Los Angeles	On Time	1191	Best
1030	FL1955	Airline B	Phoenix	On Time	1955	Best
1031	FL1276	Airline B	New York	On Time	1276	Best
1033	FL1459	Airline D	New York	On Time	1459	Best
1034	FL1313	Airline B	Phoenix	On Time	1313	Best
1036	FL1252	Airline D	Phoenix	On Time	1252	Best
1039	FL1560	Airline B	Chicago	On Time	1560	Best
1043	FL1681	Airline C	Houston	On Time	1681	Best
1044	FL1475	Airline B	Phoenix	On Time	1475	Best
1046	FL1975	Airline D	Chicago	On Time	1975	Best
1048	FL1189	Airline A	New York	On Time	1189	Best
1050	FL1686	Airline C	Phoenix	On Time	1686	Best
1052	FL1562	Airline D	Phoenix	On Time	1562	Best
1053	FL1875	Airline C	Chicago	On Time	1875	Best
1055	FL1243	Airline B	New York	On Time	1243	Best
1057	FL1504	Airline A	Phoenix	On Time	1504	Best
1060	FL1818	Airline D	Chicago	On Time	1818	Best
1061	FL1646	Airline D	Los Angeles	On Time	1646	Best
1062	FL1020	Airline C	New York	On Time	1020	Best

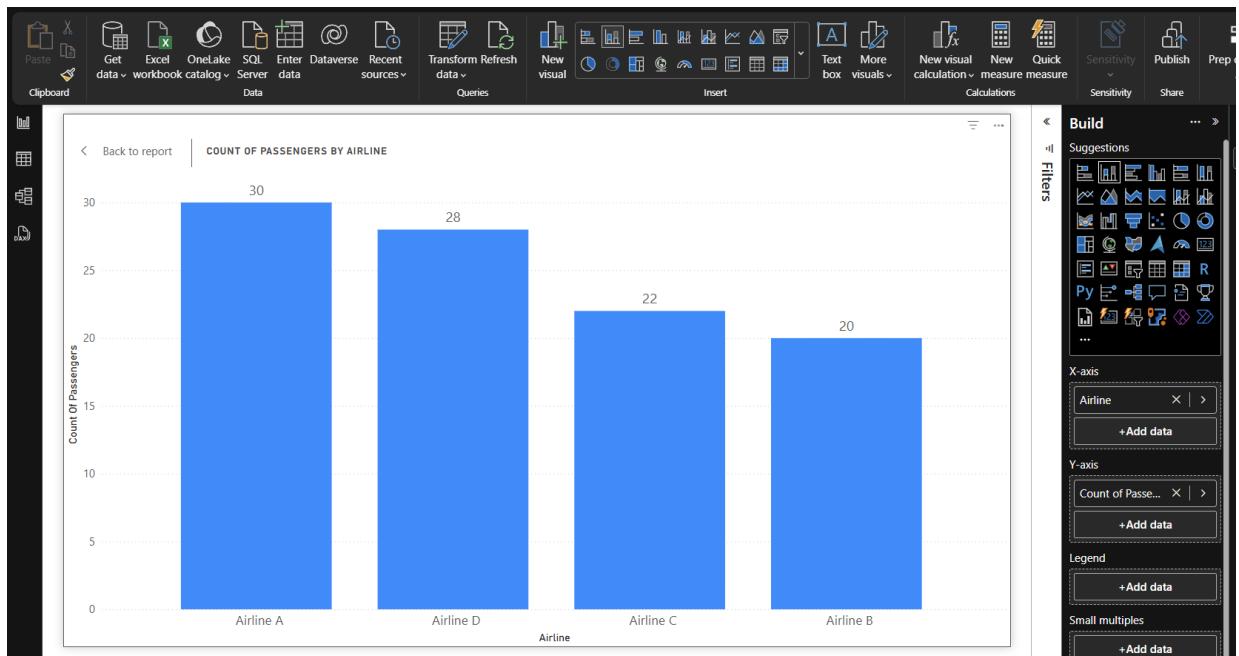
I added a new table using the **New Table** option in the **Table view**. I used the **FILTER** function to **filter the flight_information table into only best flights**.

The DAX formula is as follows;

Best flights = FILTER(flight_information, flight_information[Review]= "Best")

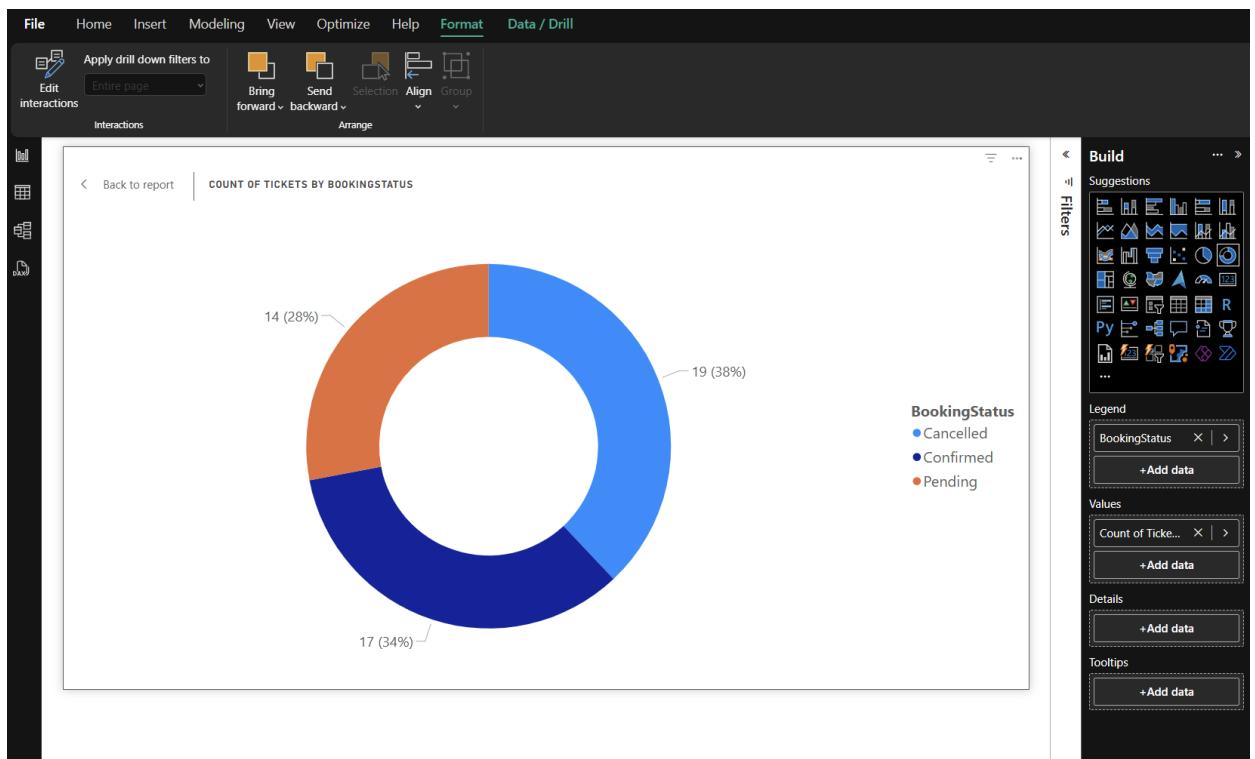
5. Visualization and Interactive Features

To show the Passenger count by airline, I used the Bar chart visual from the visualisation pane.



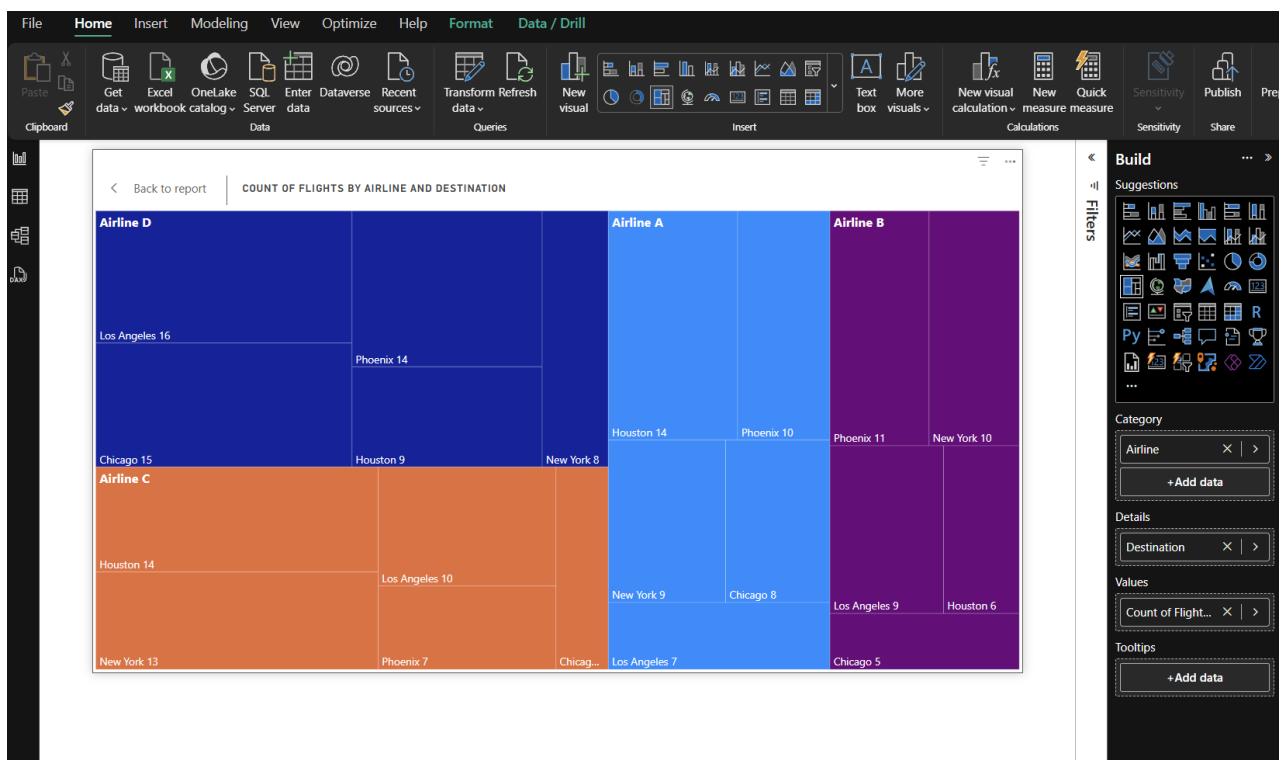
Airline A has the most number of passengers and Airline D is the close second.

To show the Ticket booking statuses, I used the Donut chart from the visualization pane to visualize the number of tickets with different statuses.



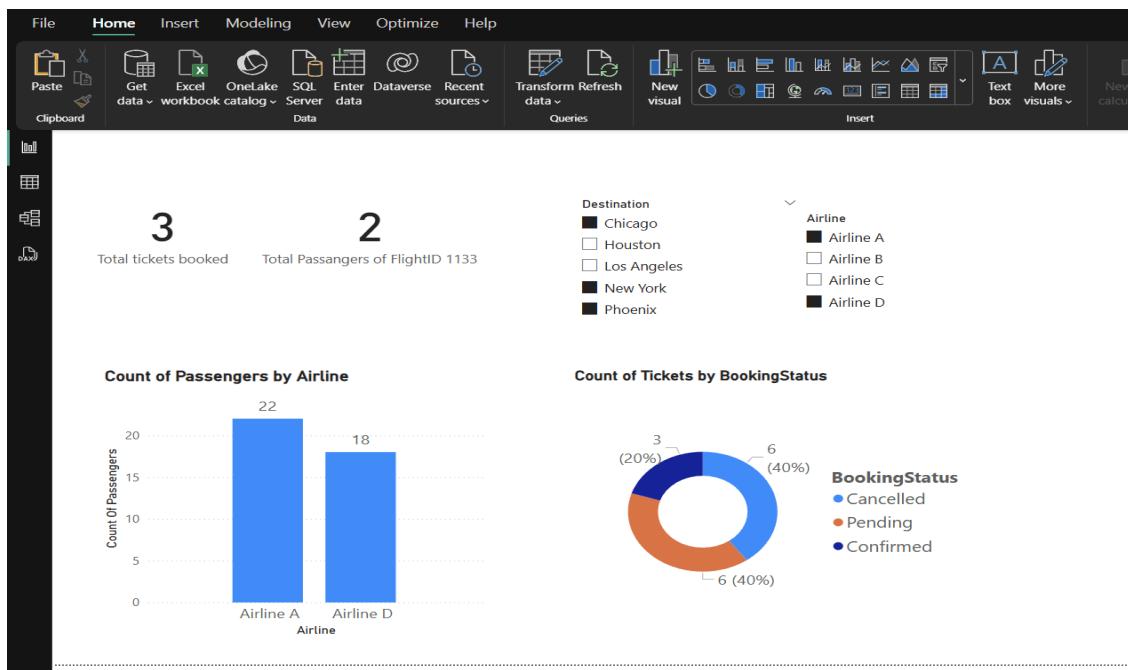
38% of the tickets are cancelled and 34% are confirmed.

To show the **Flights by airline and destination**, I used the **Treemap visual** from the **visualization pane** where I visualized the count of flights to the different **Destinations of every Airline**.



Airline D has the greatest number of flights to Los Angeles. While, both Airline A and C have the greatest number of flights to Houston. And Airline B has the greatest number of flights to Phoenix.

To make the report interactive for the user, I added interactive **slicers** for **Destination** and **Airline**. This gives the users, an **interactive experience** where they can filter on the basis of **Destination** and **Airline**.



Also, I integrated Drill through page for summarization for a specific Airline.

The screenshot shows the Power BI desktop interface with the 'Home' tab selected. On the left, there's a ribbon of icons for clipboard operations, data sources like OneLake, SQL Server, and Data, and queries. The main area displays a table with three columns: 'Airline', 'Count of PassengerID', and 'Count of TicketID'. The data shows one row for 'Airline D' with values 28 and 14 respectively, and a total row also showing 28 and 14. A vertical navigation bar on the far left contains icons for report, dashboard, and DAX.

Airline	Count of PassengerID	Count of TicketID
Airline D	28	14
Total	28	14

Also created bookmarks to for quick access of certain filters in the report to improve the user experience and facilitate any kind of user.

The screenshot shows a more complex Power BI report. At the top, there are two large numbers: '4' and '2'. Below them are the labels 'Total tickets booked' and 'Total Passangers of FlightID 1133'. To the right, there are two filter panes: 'Destination' (Chicago, Houston, Los Angeles, New York, Phoenix) and 'Airline' (Airline A, Airline B, Airline C, Airline D). On the far right, there are three bookmark icons labeled 'Phoenix', 'Airline D', and 'Airline A to New York'. Below these filters, there are two visualizations: a bar chart titled 'Count of Passengers by Airline' showing a single bar for 'Airline D' with a value of 28, and a donut chart titled 'Count of Tickets by BookingStatus' showing segments for 'Cancelled', 'Confirmed', and 'Pending' status categories.

Also to improve understanding and interactivity, I created Airline specific pages. In these pages, I integrated AI visuals, like Q&A and Key influencers which improves user interaction and understanding.

The screenshot displays two separate Power BI pages, each dedicated to a different airline. Both pages have a similar layout with a top navigation bar and a bottom ribbon menu.

Top Page (Airline A):

- Top Navigation:** Includes standard Power BI icons for Paste, Get data, OneLake, SQL Server, Enter data, Data, Transform Refresh data, New visual, Insert, and Calculations.
- Search Bar:** Shows the query "booking statuses by total tickets".
- Key influencers:** A section titled "Key influencers Top segments" showing that "Review" is the best segment. It includes a statement: "...the likelihood of Review being Best increases by Count of PassengerID goes up 0.48" with a multiplier of "1.04x".
- Review Summary:** A bar chart titled "Review" showing the average count of passengers for "Best" and "To Be ...".
- Airline Matrix:** A 2x2 grid labeled "Airline" with four cells: Airline A (top-left), Airline C (top-right), Airline B (bottom-left), and Airline D (bottom-right).
- Bottom Navigation:** Shows tabs for Page 1, Page 2, Page 3, and Airline A, with Airline A currently selected.

Bottom Page (Airline B):

- Top Navigation:** Similar to the top page, with standard Power BI icons.
- Search Bar:** Shows the query "booking statuses by total tickets".
- Key influencers:** A section titled "Key influencers Top segments" showing that "Review" is the best segment. It includes a statement: "...the likelihood of Review being Best increases by Count of PassengerID goes up 0.43" with a multiplier of "1.11x".
- Review Summary:** A bar chart titled "Review" showing the average count of passengers for "Best" and "To Be ...".
- Airline Matrix:** A 2x2 grid labeled "Airline" with four cells: Airline A (top-left), Airline C (top-right), Airline B (bottom-left), and Airline D (bottom-right).
- Bottom Navigation:** Shows tabs for Page 1, Page 2, Page 3, and Airline A, with Airline B currently selected.

Top Segments

Showing results for *Booking status sorted by number of ticket ID*

BookingStatus	Count of TicketID
Cancelled	6
Confirmed	4
Pending	3

Key Influencers

What influences Review to be **Best** ?

When... ...the likelihood of Review being Best increases by

Count of PassengerID goes up 0.49 → 1.03x

On average when Count of PassengerID increases, the likelihood of Review being Best increases.

Review	Average: Count of PassengerID
Best	1.16
To Be I...	1.07

Airline

Airline A	Airline C
Airline B	Airline D

Bottom Segments

Showing results for *Booking status sorted by number of ticket ID*

BookingStatus	Count of TicketID
Cancelled	6
Confirmed	4
Pending	3

Key Influencers

What influences Review to be **Best** ?

When... ...the likelihood of Review being Best increases by

Count of PassengerID goes down 0.27 → 1.03x

On average when Count of PassengerID decreases, the likelihood of Review being Best increases.

Review	Average: Count of PassengerID
Best	1.08
To Be I...	0.81

Airline

Airline A	Airline C
Airline B	Airline D

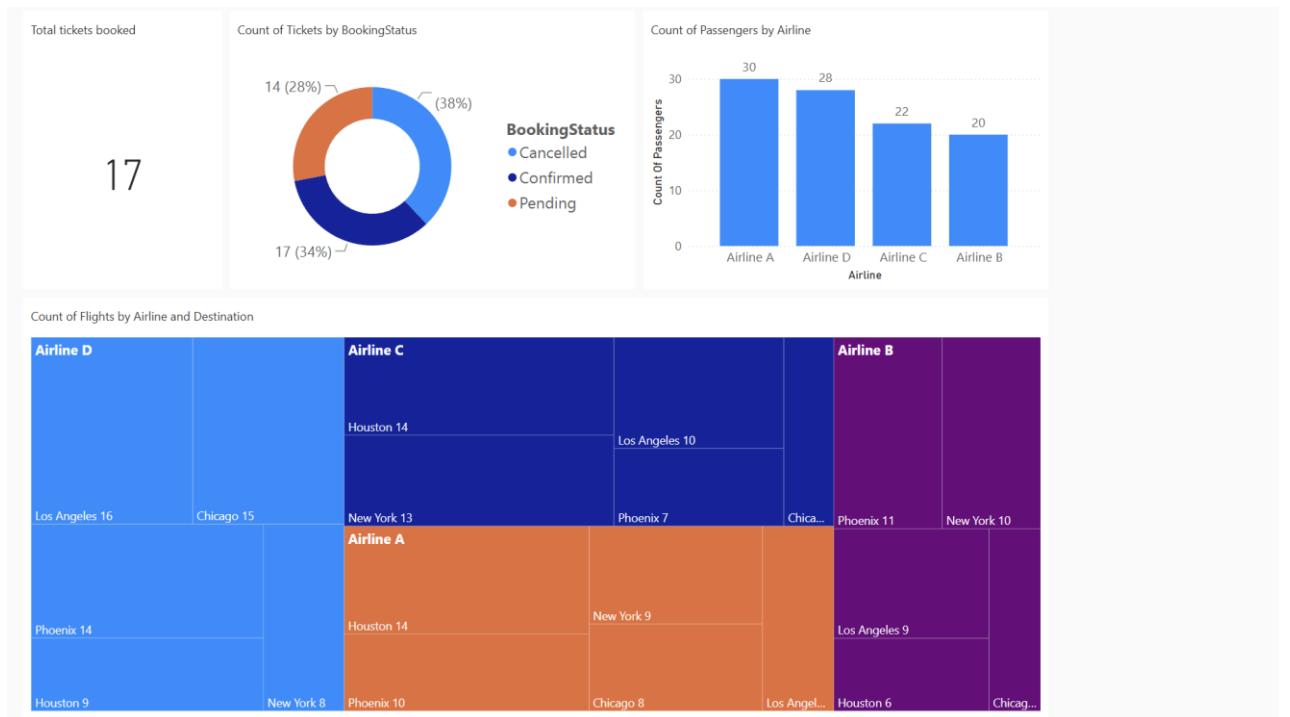
6. Final Dashboard and Power BI Service

I pinned the important visuals into the dashboard after publishing the report into a workspace in PowerBI service.

A dashboard can summarize the whole report in one page with the most important visuals in the report all together.

The screenshot shows the Microsoft Power BI service workspace. At the top, there are navigation links for 'New item', 'New folder', 'Import', 'Migrate', and search/filter options. Below this is a placeholder area with a large gray square icon. The main content area is titled 'Choose from predesigned task flows or add a task to build one' with a sub-instruction 'Select from one of Microsoft's predesigned task flows or add a task to start building one yourself.' There are two buttons: 'Select a predesigned task flow' and 'Add a task'. Below these buttons is a link 'Import a task flow'. The bottom half of the screen displays a table of task flows:

Name	Type	Task	Owner	Refreshed	Next refresh	Endorsement	Sensitivity	Included in app
Dashboard	Dashboard	—	PowerBI	—	—	—	—	No
Final Project	Report	—	PowerBI	6/16/2025, 5:52:3...	—	—	—	No
Final Project	Semantic mo...	—	PowerBI	6/16/2025, 5:52...	N/A	—	—	



I configured a **Row-Level Security (RLS)** for Airline A data using the **Manage Roles** option in the **Security** section of the **Modeling** menu in the PowerBI Desktop.

Manage security roles

Create new security roles and use filters to define row-level data restrictions.

Successfully applied role changes.

Roles

+ New

Airline A

Select tables

Best flights

flight_informat...

Measures (2)

passenger_infor...

ticket_informat...

Filter data

Show data if All of these rules are true

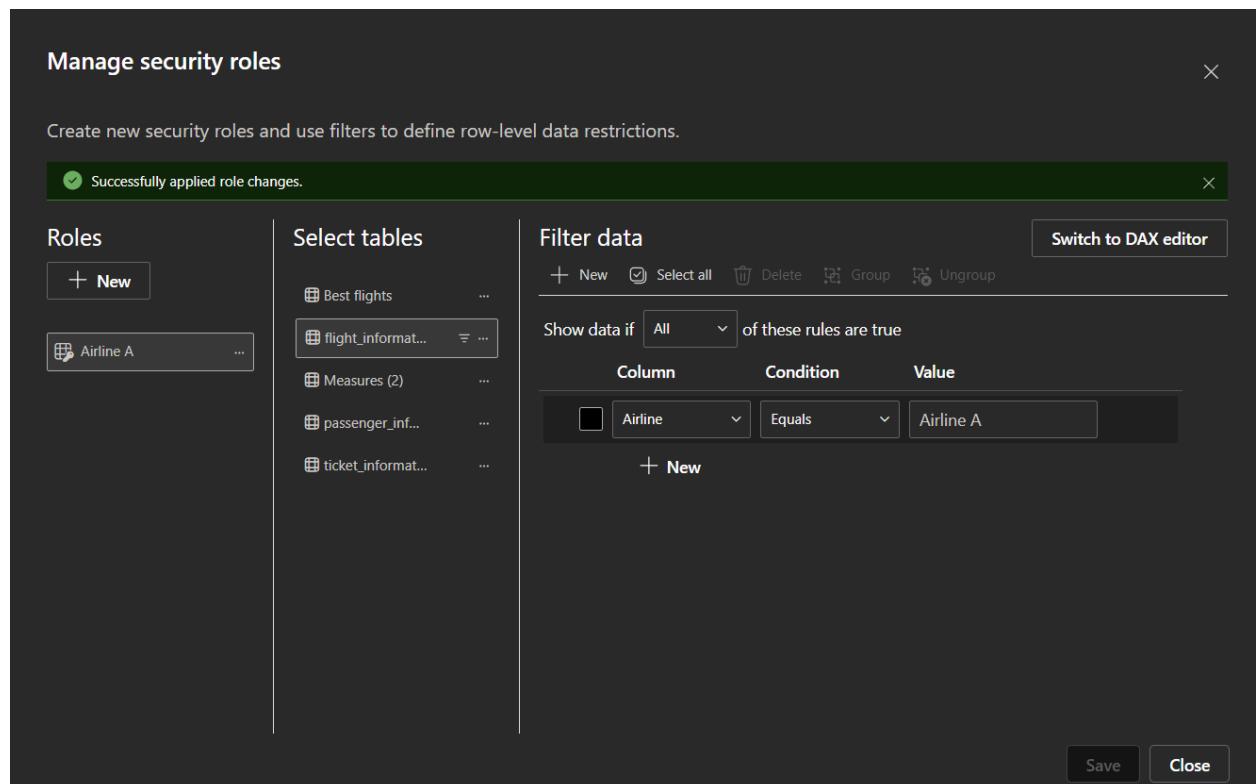
Column Condition Value

Airline Equals Airline A

+ New

Switch to DAX editor

Save Close



Power BI PowerBi > Row-Level Security

Search

Home

Create

Browse

OneLake catalog

Apps

Metrics

Workspaces

PowerBI

Dashboard

Final Project

...

Power BI

Row-Level Security

Airline A (1)

Members (1)

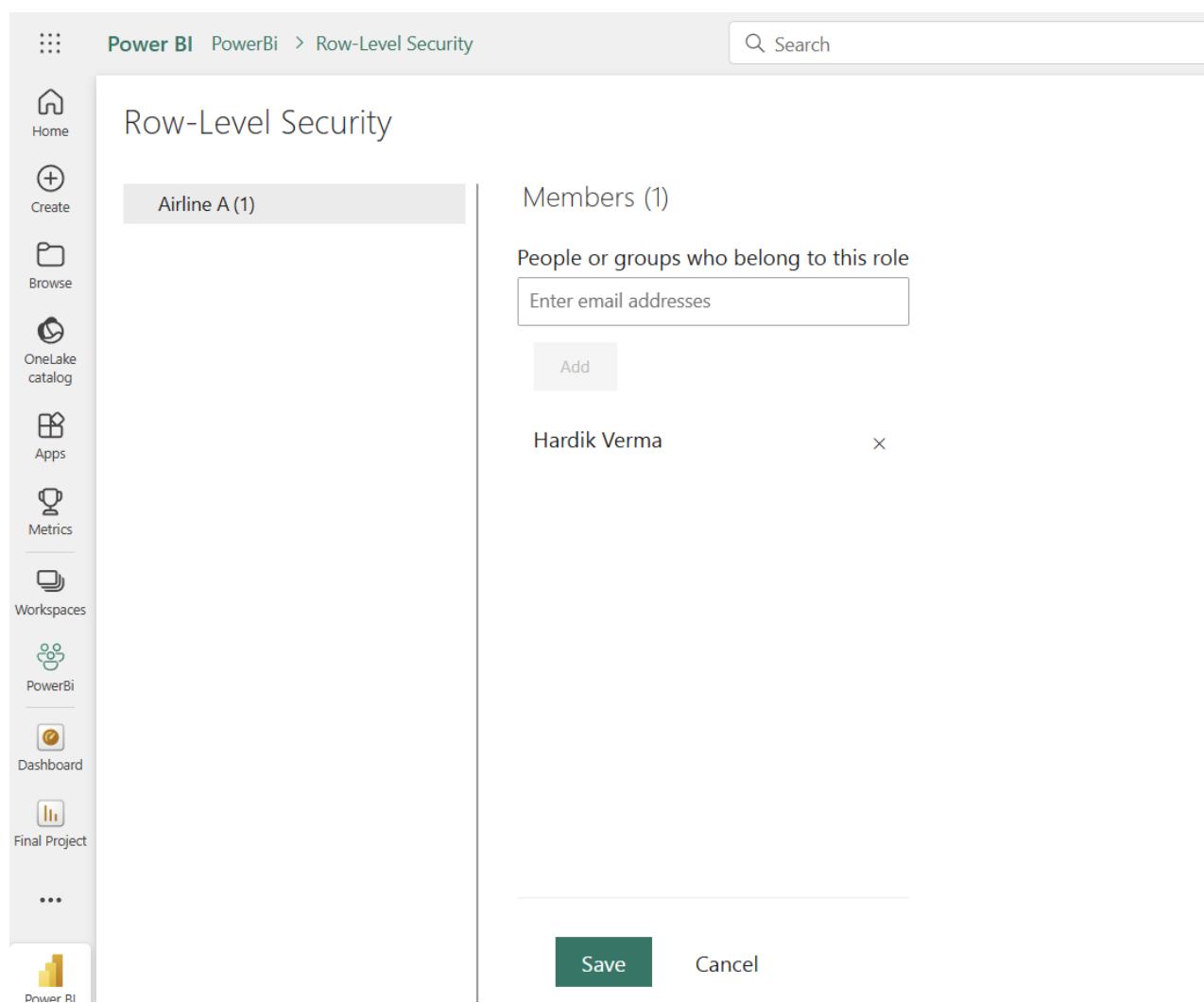
People or groups who belong to this role

Enter email addresses

Add

Hardik Verma

Save Cancel



For the flight_information dataset, I created RLS on Airline A to restrict data for some users. RLS is a security feature used to restrict certain users to a filtered data usage.

I have setup the gateway from my system to the PowerBI service and also configured the data sources in the semantic model menu.

The screenshot shows a window titled "On-premises data gateway (personal mode)". The left sidebar has a blue header labeled "Status" and contains links for "Service Settings", "Diagnostics", "Network", and "Connectors". The main pane displays the following information:

- A green checkmark icon followed by the text: "The gateway is online and ready to be used."
- The text: "Gateway version number: 3000.270.10 (May 2025)"
- An unchecked checkbox followed by the text: "Help us improve the on-premises data gateway (personal mode) by sending usage information to Microsoft." Below this is a link: "Read the privacy statement online".

In the bottom right corner of the main pane, there is a "Close" button.

After that I scheduled a daily refresh of the datasets at 5 PM which will ensure that any changes in the data sources in the system will be updated on the PowerBI service also.

The screenshot shows the 'Refresh' configuration page in the Power BI service. On the left is a sidebar with navigation links: Home, Create, Browse, OneLake catalog, Apps, Metrics, Workspaces, PowerBI, Dashboard, Final Project, and three dots for more. The main area has a search bar at the top right. Below it, the 'Time zone' section is expanded, showing a note about determining schedule refresh time and current date/time for incremental refresh models. It lists '(UTC+05:30) Chennai, Kolkata, Mumbai'. The 'Configure a refresh schedule' section is also expanded, showing the 'On' toggle is turned on. Under 'Refresh frequency', a dropdown menu is set to 'Daily'. The 'Time' field shows '5 00 PM'. There's a link to 'Add another time'. The 'Send refresh failure notifications to' section includes a checked checkbox for 'Semantic model owner' and an unchecked checkbox for 'These contacts', with a text input field for email addresses. At the bottom are 'Apply' and 'Discard' buttons.

Video Explanation

<https://drive.google.com/file/d/1vUbG7maPiUgKcO9TfQERTakZEy9mNrbO/view?usp=sharing>