

CASE STUDY ON AIRLINES ANALYSIS (SPRINT 1)

Batch: BI V7 with MS Azure PT Sep 6th Batch1

Group No.: 4

Group Members:

1. Harshit Mudgal (46263678)
2. Abhaya Shankar (46263748)
3. Armaandeep Sandhu (46265960)
4. Ashitosh Joshi (46265287)
5. Abhishek Baghel (46263842)

INTRODUCTION

Analysis Services is an analytical data engine used in decision support and business analytics. It provides enterprise-grade semantic data model capabilities for business intelligence (BI), data analysis, and reporting applications such as Power BI, Excel, Reporting Services, and other data visualization tools.

Analysis Services is available in different platforms:

1. Azure Analysis Services
2. Power BI Premium
3. SQL Server Analysis Services

In our project scope we have worked upon **SQL Server Analysis Services** and created an end to end flow of data.

Installed as an on-premises server instance, SQL Server Analysis Services supports tabular models at all compatibility levels (depending on version), multidimensional models, data mining, and Power Pivot for SharePoint.

Workflow of SQL Service Analysis Services:

- ❖ A typical implementation workflow includes installing a SQL Server Analysis Services instance, creating a tabular or multidimensional data model. Our scope works on Tabular Data Model.
- ❖ Deploying the model as a database to a server instance, processing the database to load it with data, and then assigning permissions to allow data access. When ready to go, the data model can be accessed by any client application supporting Analysis Services as a data source.
- ❖ To create a model, use Visual Studio with Analysis Services projects extension, also known as SQL Server Data Tools or simply SSDT, choosing a Tabular project template. The project template contains folders for all of the objects needed in a model.

- ❖ Models specify query objects, such as calculations, relationships and DAX Queries.
- ❖ To use a model, it's deployed to a server instance that runs databases in a particular server mode, making the data available to authorized users who connect through Excel or other applications like SQL Server Management Studio.

PROBLEM STATEMENT

We need to create an end-to-end flow for deployment of Tabular model in the SQL Server Analysis Services.

Files attached here with all schema and data creation script:



Tasks to be performed:

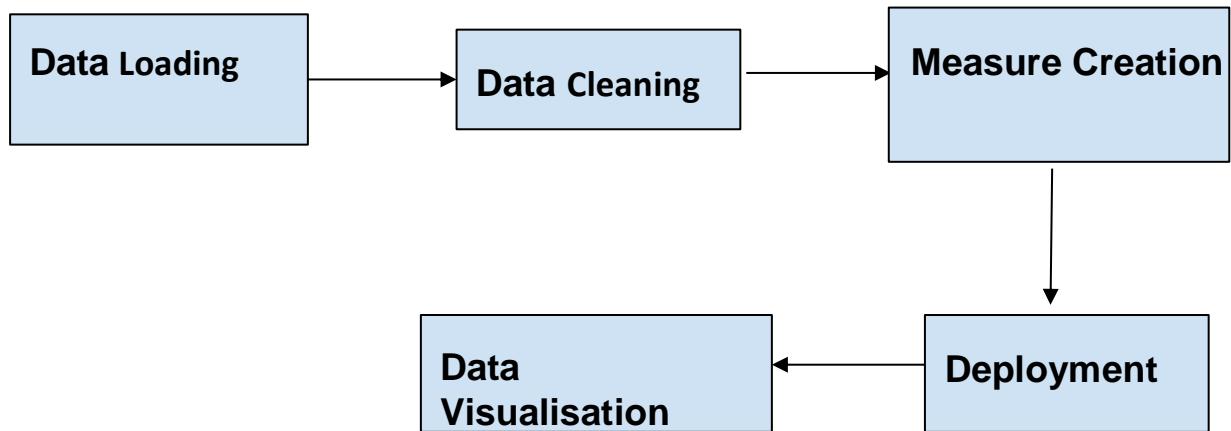
- Run the script “Script.sql” attached in the zipped folder “Airlines Data.7z” to create following tables:
 - Hub_Airport,
 - Hub_Flight1,
 - Link_Flight_Airport1
 - Sat_Airport and
 - Sat_Flight1.
- Load that data into the **Analysis Services Tabular model** project from SQL DB.
- Clean the dataset, remove any non-required columns, and import the data into the tabular model project.
- Create the relationship between the tables.
- Create some measures/Calculated Columns using DAX queries in SQL Service Analysis Services:
 1. Total No. of different flights running.
 2. Create Hierarchy:
 - Country
 - State
 - City
 3. Flights going to particular country
 4. Flights going to state
 5. Flights going to a city.
 6. Perform calculation to Identify Regional/International Airport.
 7. Flights going to every State.
 8. Find Different Airlines Available
 9. Which Airline has the maximum running flights
 10. No of Airports in the state AZ, DE, and NY.
 11. Create Hierarchy for Airlines and its flight numbers.
- After successfully creating all measures and calculated columns, deploy the tabular model and browse the model in SSMS or in Excel and Get all the possible insights.

PREREQUISITES

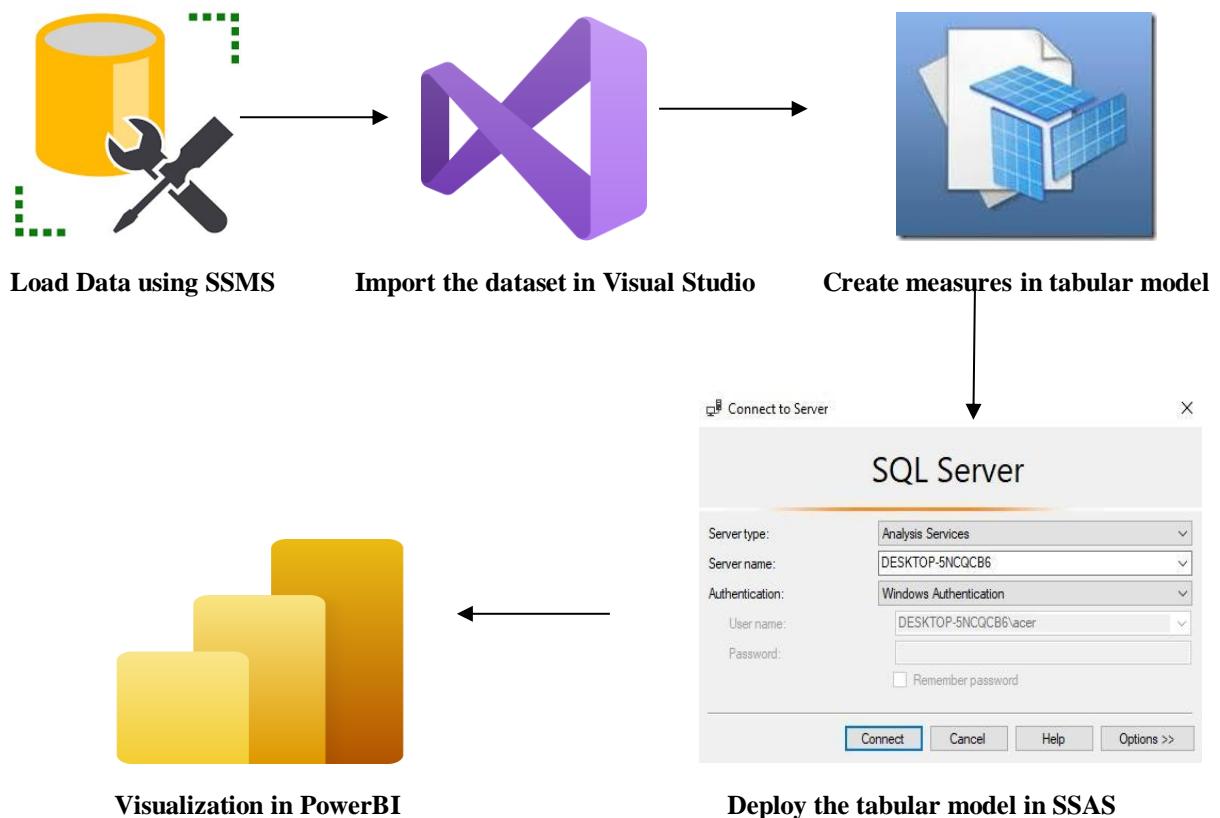
1. SQL Server 2017 (or higher) with SQL Server Analysis Services Implemented
2. Visual Studio 2017 (or higher) with Tabular Project Template
3. Power BI Desktop to create Visualizations
4. SQL Server Management Studio to Load and Deploy the model

PROJECT DESIGN

High Level Design:



Low Level Design:



1.Run Script file (Data Source) in SSMS to create database.



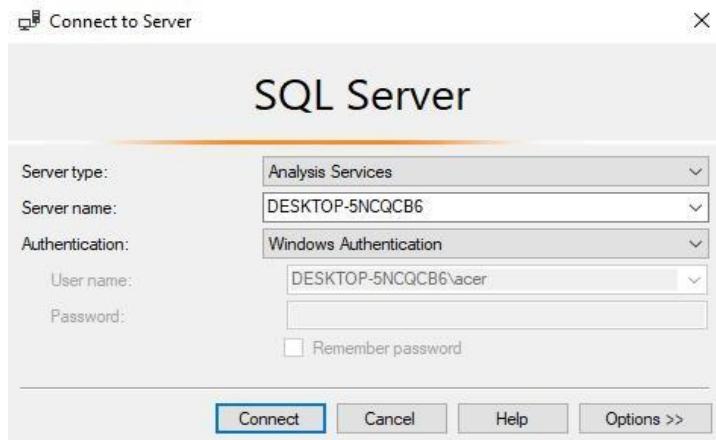
2. Import the data into Visual studio to create Tabular Model Project.



3. Data cleaning, building relationships and measures creation is performed in Tabular Model Project.



4. Tabular Model Project is deployed in SSAS(SQL Server Analysis Service) - On premise.



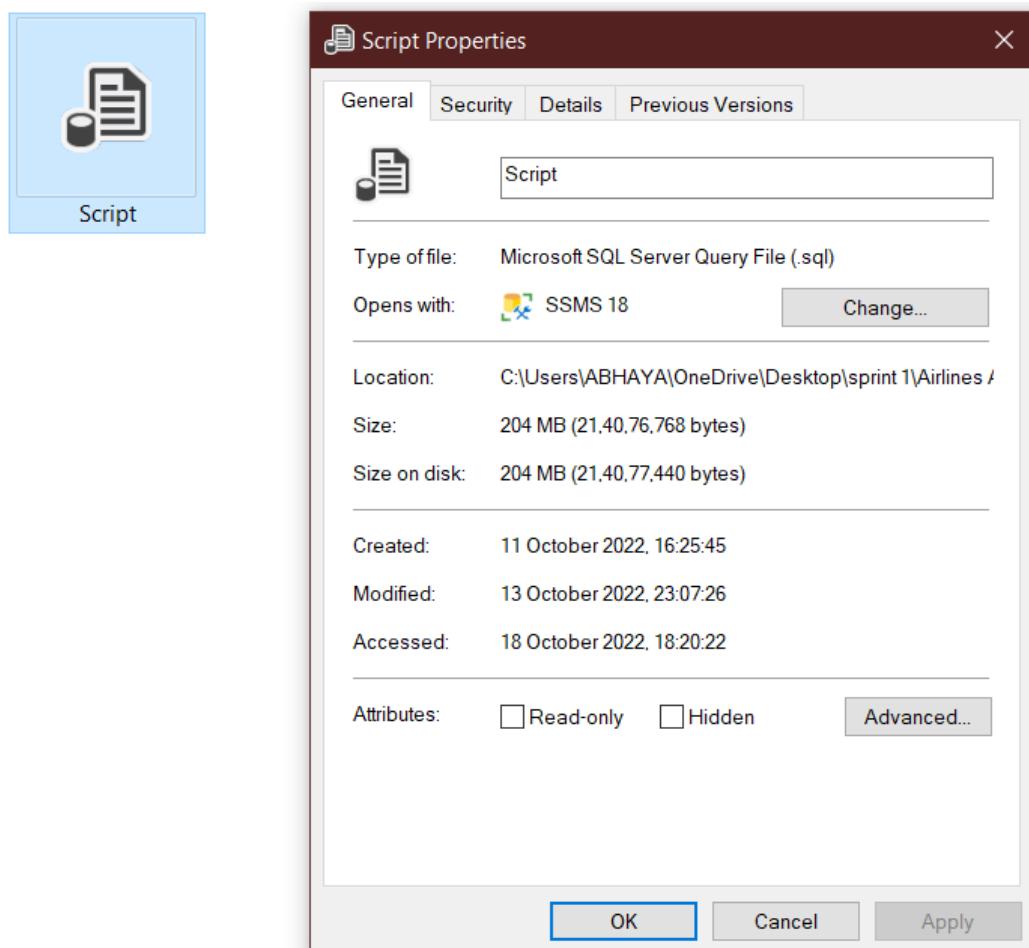
5. Microsoft Power BI desktop is used for visualisation.



IMPLEMENTATION STEPS

LOADING DATA INTO SQL SERVER:

1. Run the script as given in AIRLINE SERVICES Sprint Case Study.



2. Create Database and run the code to apply tables and values into the database.

Object Explorer                                                             <img alt="New dbo.Hub_Airport icon" data-bbox="7262 113 7276 125

3. Clean the datasets for finding better performance and less errors. (Example: Removing duplicate and unwanted columns)

7	7	2020-03-04 06:35:59.167	NULL	Flight_Data.csv	LAS	MSP	25	526
8	506	2020-03-04 06:35:59.167	NULL	Flight_Data.csv	LAS	MSP	25	526

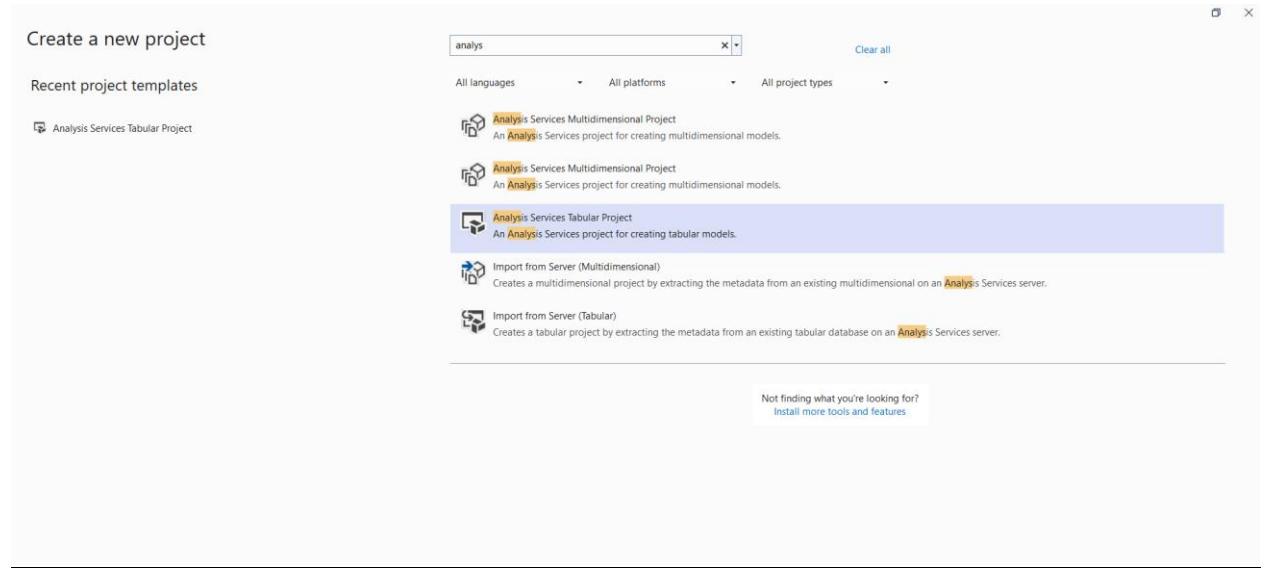
43	42	2020-03-04 06:35:59.167	NULL	Flight_Data.csv	PHX	DFW	159	502
44	330	2020-03-04 06:35:59.167	NULL	Flight_Data.csv	PHX	DFW	159	502

333	42	2020-03-04 06:35:59.167	NULL	Flight_Data.csv	DFW	FLL	600	939
334	330	2020-03-04 06:35:59.167	NULL	Flight_Data.csv	DFW	FLL	600	939

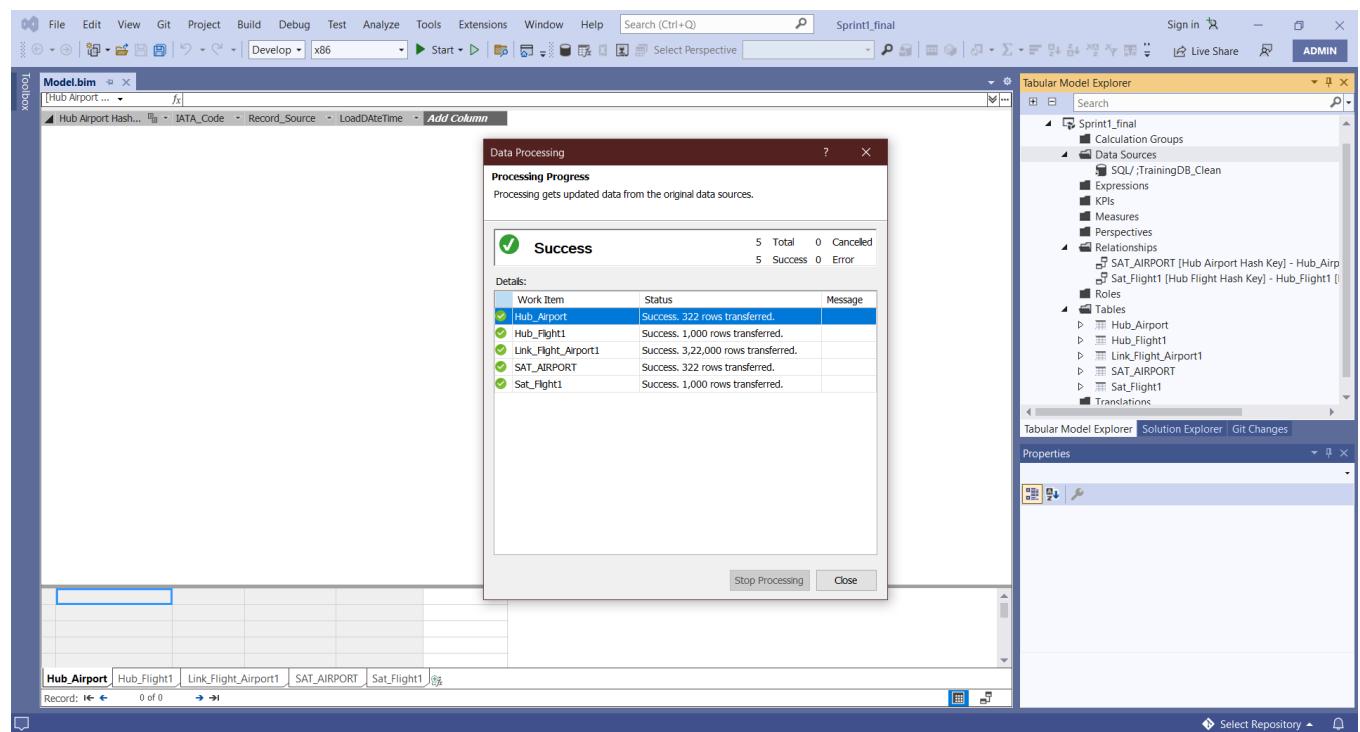
510	7	2020-03-04 06:35:59.167	NULL	Flight_Data.csv	MSP	ORD	616	745
511	506	2020-03-04 06:35:59.167	NULL	Flight_Data.csv	MSP	ORD	616	745

LOADING DATABASE INTO VISUAL STUDIO:

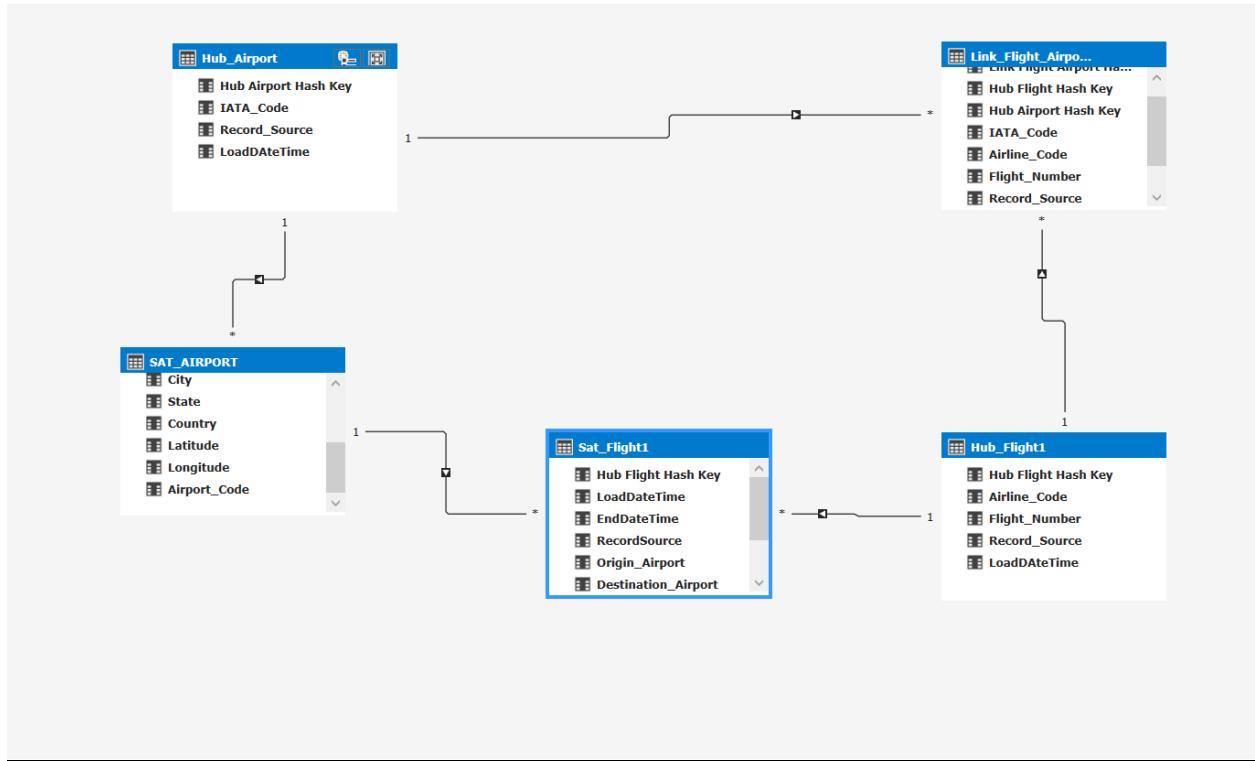
1. Open and run Visual Studio as Administrator.
2. Create a new project and choose Analysis Services Tabular Project.



3. Connect to SQL Server and transform data.



4. Create Relationships.



5. Apply DAX Queries to create calculated columns and measures

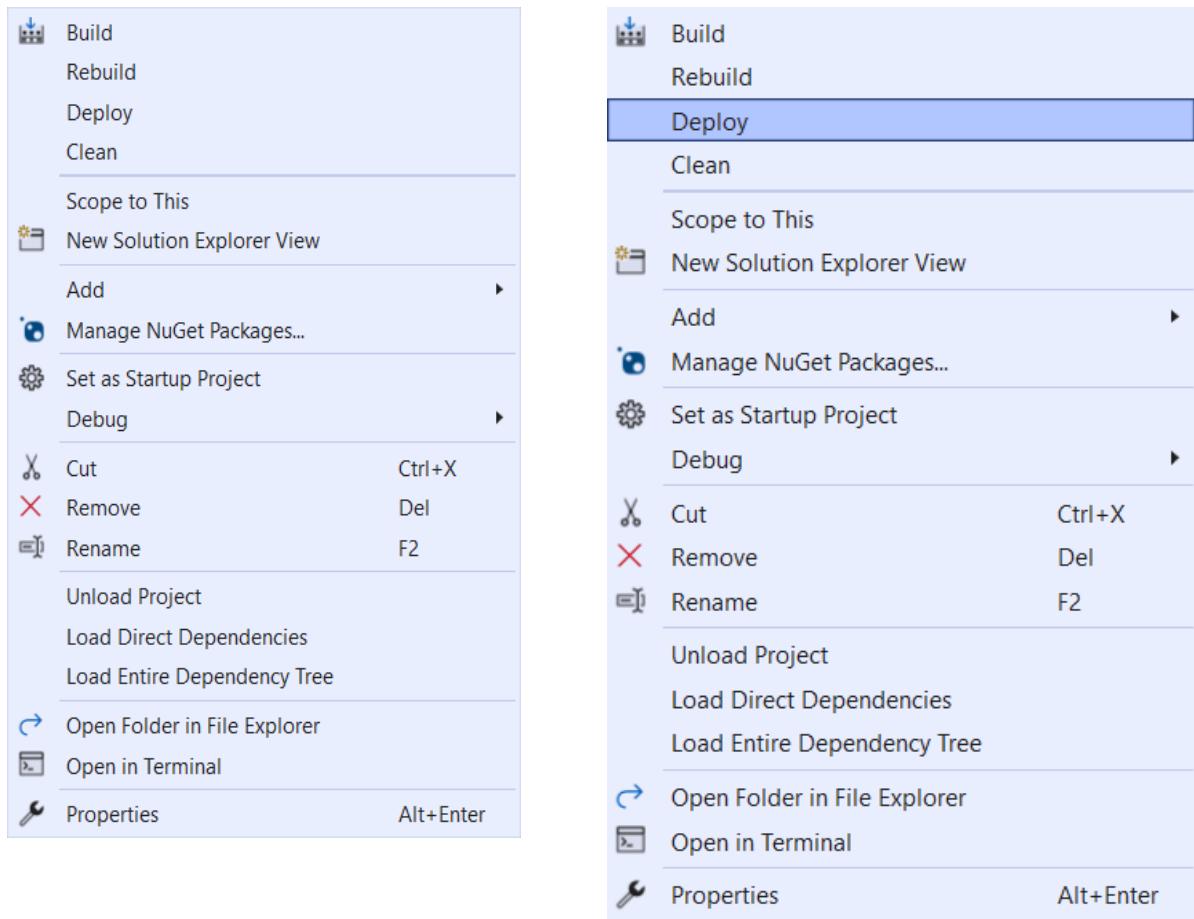
Model.bim* [Country]

fx|TotalFlightsGoingToUSA:= CALCULATE(COUNT(SAT_AIRPORT[Hub Airport Hash Key]), FILTER(SAT_AIRPORT,[Country])="USA"))

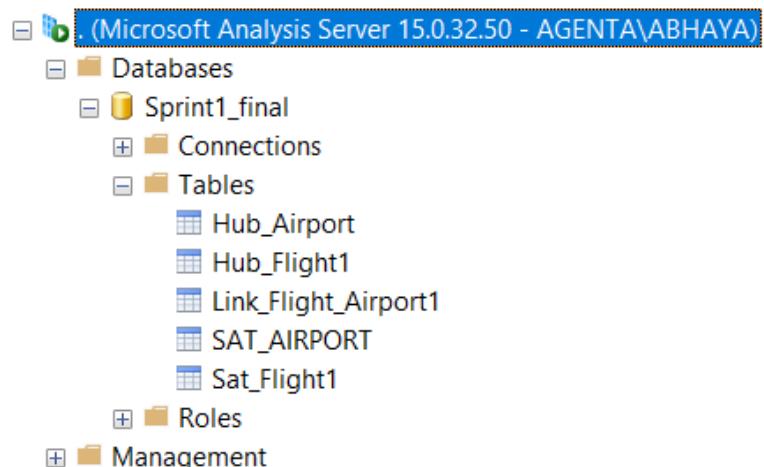
	Hub Airport Has...	LoadDateTime	EndDateTime	RecordSource	AirportName	City	State	Country	Latitude	Longitude	Airport_...	Add C...
1	1	03-03-2020 10:...		airports.csv	Lehigh Valley I...	Allent...	PA	USA	40.65236	-75.4404	ABE	
2	2	03-03-2020 10:...		airports.csv	Abilene Region...	Abilene	TX	USA	32.41132	-99.6819	ABI	
3	3	03-03-2020 10:...		airports.csv	Albuquerque I...	Albuquerque	NM	USA	35.04022	-106.60919	ABQ	
4	4	03-03-2020 10:...		airports.csv	Aberdeen Regi...	Aberd...	SD	USA	45.44906	-98.42183	ABR	
5	5	03-03-2020 10:...		airports.csv	Southwest Ge...	Albany	GA	USA	31.53552	-84.19447	ABY	
6	6	03-03-2020 10:...		airports.csv	Nantucket Me...	Nantu...	MA	USA	41.25305	-70.06018	ACK	
7	7	03-03-2020 10:...		airports.csv	Waco Regiona...	Waco	TX	USA	31.61129	-97.23052	ACT	
8	8	03-03-2020 10:...		airports.csv	Arcata Airport	Arcat...	CA	USA	40.97812	-124.10862	ACV	
9	9	03-03-2020 10:...		airports.csv	Atlantic City In...	Atlanti...	NJ	USA	39.45758	-74.57717	ACY	
10	10	03-03-2020 10:...		airports.csv	Adak Airport	Adak	AK	USA	51.87796	-176.64603	ADK	
11	11	03-03-2020 10:...		airports.csv	Kodiak Airport	Kodiak	AK	USA	57.74997	-152.49386	ADQ	
12	12	03-03-2020 10:...		airports.csv	Alexandria Inte...	Alexa...	LA	USA	31.32737	-92.54856	AEX	
13	13	03-03-2020 10:...		airports.csv	Augusta Regio...	Augusta	GA	USA	33.36996	-81.9645	AGS	
14	14	03-03-2020 10:...		airports.csv	King Salmon Ai...	King S...	AK	USA	58.6768	-156.64922	AKN	
15	15	03-03-2020 10:...		airports.csv	Albany Interna...	Albany	NY	USA	42.74812	-73.80298	ALB	
16	16	03-03-2020 10:...		airports.csv	Waterloo Regi...	Water...	IA	USA	42.55708	-92.40034	ALO	
17	17	03-03-2020 10:...		airports.csv	Rick Husband ...	Amarillo	TX	USA	35.21937	-101.70593	AMA	
18	18	03-03-2020 10:...		airports.csv	Ted Stevens A...	Anch...	AK	USA	61.17432	-149.99619	ANC	
19	19	03-03-2020 10:...		airports.csv	Alpena County...	Alpena	MI	USA	45.07807	-83.56029	APN	
20	20	03-03-2020 10:...		airports.csv	Aspen-Pitkin C...	Aspen	CO	USA	39.22316	-106.86885	ASE	
21	21	03-03-2020 10:...		airports.csv	Hartsfield-Jack...	Atlanta	GA	USA	33.64044	-84.42694	ATL	
22	22	03-03-2020 10:...		airports.csv	Appleton Inter...	Applet...	WI	USA	44.25741	-88.51948	ATW	
23	23	03-03-2020 10:...		airports.csv	Austin-Bergstr...	Austin	TX	USA	30.19453	-97.66987	AUS	
24	24	03-03-2020 10:...		airports.csv	Asheville Regio...	Asheville	NC	USA	35.43619	-82.54181	AVL	
25	25	03-03-2020 10:...		airports.csv	Wilkes-Barre/S...	Wilkes...	PA	USA	41.33815	-75.72427	AVP	
26	26	03-03-2020 10:...		airports.csv	Kalamazoo/Bat...	Kalam...	MI	USA	42.23488	-85.55206	AZO	
27	27	03-03-2020 10:...		airports.csv	Bradley Intern...	Winds...	CT	USA	41.93887	-72.68323	BDL	
28	28	03-03-2020 10:...		airports.csv	Bethel Airport	Bethel	AK	USA	60.77978	-161.838	BET	
29	29	03-03-2020 10:...		airports.csv	Meadows Field	Baker...	CA	USA	35.4336	-119.05677	BFL	

TotalFlightsGoingToUSA: 322

6. Then Finally, Build and Deploy Project to Analysis services

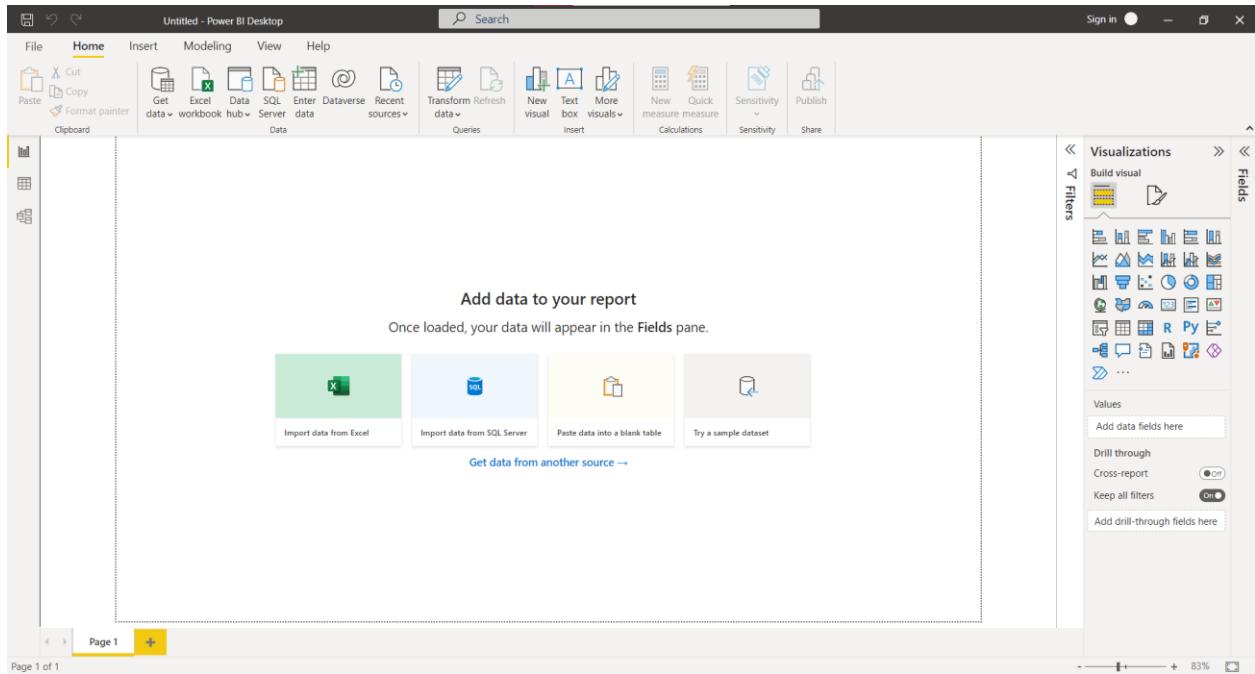


7. Connect Analysis Services in SQL Server.

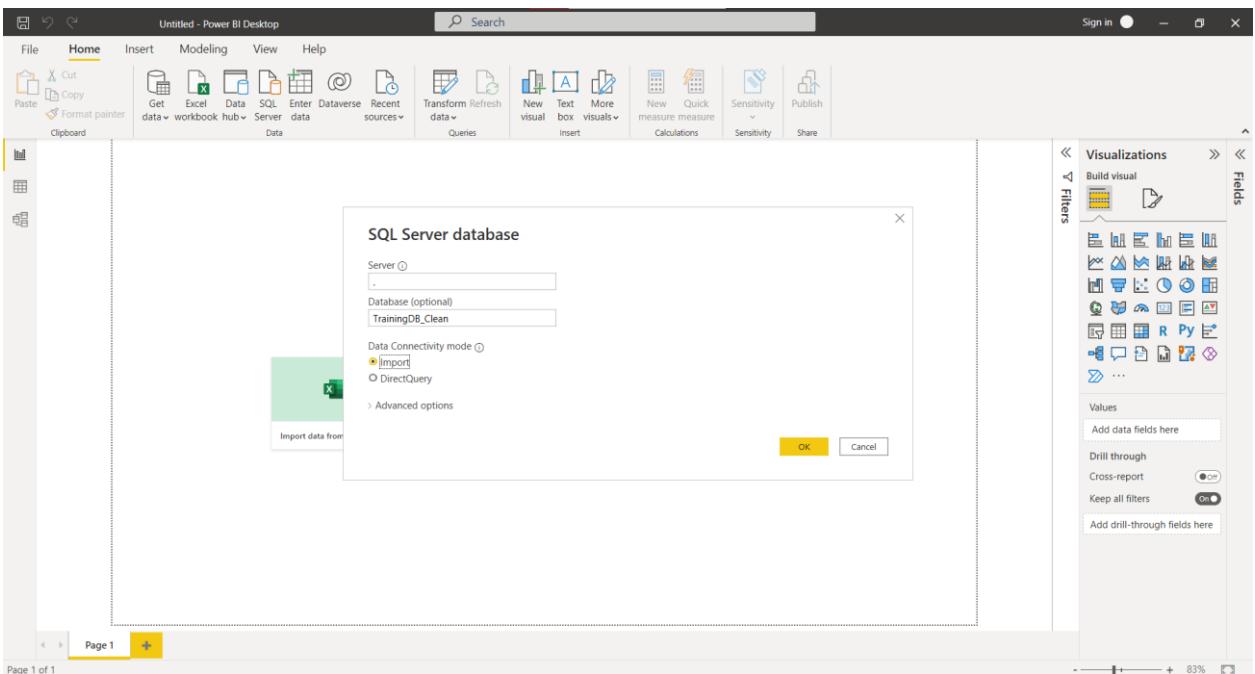


ANALYSIS OF FINAL TRANSFORMED DATA:

1. Open PowerBI Desktop and Import data from SQL Server.



2. Connect to SQL Server Database.



3. Load and transform data.

The screenshot shows the Power BI Desktop interface with the 'Navigator' pane open. The 'Sat_Flight1' table is selected, displaying a list of 23 rows with columns: Hub Flight Hash Key, LoadDateTime, EndDateTime, and RecordSource. All rows have a LoadDateTime of 04-03-2020 06:35:59 and a RecordSource of 'null Flight_Data.csv'. The 'Navigator' pane also lists other tables: Hub_Airport, Hub_Flight1, Link_Flight_Airport1, SAT_AIRPORT, and another 'Sat_Flight1' table. The 'Visualizations' pane on the right shows various chart and report icons. The 'Fields' pane on the far right lists the fields for each table, such as Hub_Flight1 (Hub_Flight Hash Key, Airline_Code, Flight_Number, LoadDateTime, Record_Source), Link_Flight_Airport1 (Flight_Number, Hub_Airport_Hash_Key, IATA_Code, Link_Flight_Airport_Hash_Key, LoadDateTime, Record_Source), and SAT_AIRPORT (AirportName, City, Country, EndDateTime, Hub_Airport_Hash_Key, Latitude).

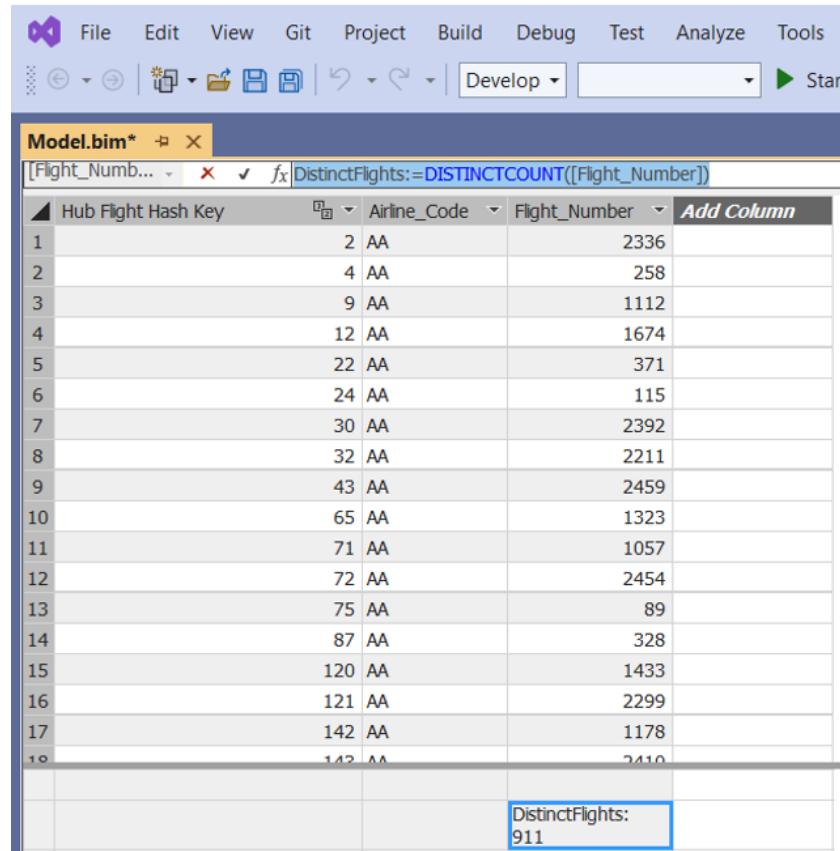
4. Apply different charts and pie-charts to get Visualization to analyze data.

The screenshot shows the Power BI Desktop interface with a bar chart titled 'Count of Hub Flight Hash Key by Airline_Code' on the left. The chart displays the count of flights for various airline codes: AA (135), DL (121), UA (118), B6 (103), EV (91), WN (78), MQ (68), AS (60), US (59), NK (40), P9 (34), HA (17), and VX (4). The 'Visualizations' pane on the right shows various chart and report icons. The 'Fields' pane on the far right lists the fields for each table, including Hub_Flight1, Link_Flight_Airport1, and SAT_AIRPORT, with detailed sub-fields like Airline_Code, Flight_Number, and Latitude.

DAX QUERIES AND SCREENSHOTS

1) Total No. of different flights running

DistinctFlights:=DISTINCTCOUNT([Flight_Number])

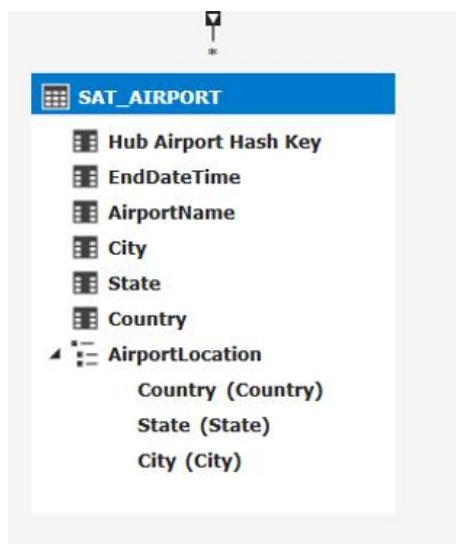


The screenshot shows the Power BI Model view with a table named 'Hub Flight Hash Key'. The table has four columns: 'Hub Flight Hash Key', 'Airline_Code', 'Flight_Number', and 'Add Column'. The 'Add Column' column contains the DAX formula 'DistinctFlights:=DISTINCTCOUNT([Flight_Number])'. The calculated column shows the value '911'.

Hub Flight Hash Key	Airline_Code	Flight_Number	Add Column
1	2 AA	2336	
2	4 AA	258	
3	9 AA	1112	
4	12 AA	1674	
5	22 AA	371	
6	24 AA	115	
7	30 AA	2392	
8	32 AA	2211	
9	43 AA	2459	
10	65 AA	1323	
11	71 AA	1057	
12	72 AA	2454	
13	75 AA	89	
14	87 AA	328	
15	120 AA	1433	
16	121 AA	2299	
17	142 AA	1178	
18	143 AA	2410	

2) Create Hierarchy:

- i) Country
- ii) State
- iii) City



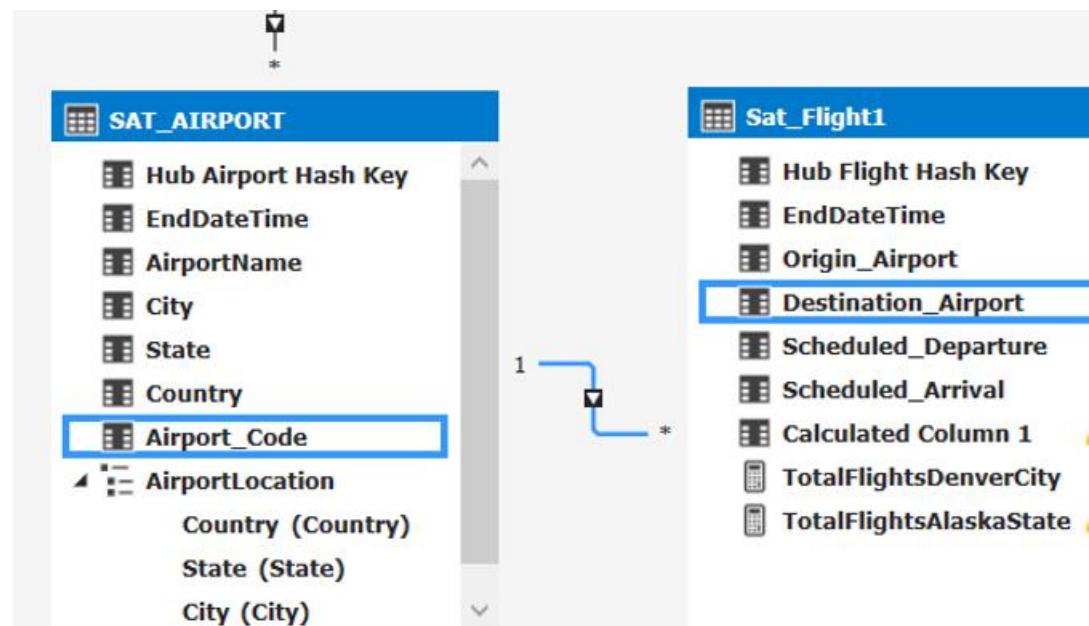
3) Flights going to particular country (USA)

A calculated column ‘Airport_Code’ is created in ‘SAT_AIRPORT’ table

=RELATED('Hub_Airport'[IATA_Code])

	Hub Airport Hash Key	EndDateTime	AirportName	City	State	Country	Airport_Code	Add
1		1	Lehigh Valley International Airport	Allentown	PA	USA	ABE	
2		2	Abilene Regional Airport	Abilene	TX	USA	ABI	
3		3	Albuquerque International Sunport	Albuquerque	NM	USA	ABQ	
4		4	Aberdeen Regional Airport	Aberdeen	SD	USA	ABR	
5		5	Southwest Georgia Regional Airport	Albany	GA	USA	ABY	
6		6	Nantucket Memorial Airport	Nantucket	MA	USA	ACK	
7		7	Waco Regional Airport	Waco	TX	USA	ACT	
8		8	Arcata Airport	Arcata/Eureka	CA	USA	ACV	
9		9	Atlantic City International Airport	Atlantic City	NJ	USA	ACY	
10		10	Adak Airport	Adak	AK	USA	ADK	
11		11	Kodiak Airport	Kodiak	AK	USA	ADQ	
12		12	Alexandria International Airport	Alexandria	LA	USA	AEX	
13		13	Augusta Regional Airport (Bush Field)	Augusta	GA	USA	AGS	

A relationship is created between tables ‘Sat_Airport’ and ‘Sat_Flight1’ through columns ‘Airport_Code’ and ‘Destination_Airport’



Edit Relationship

?

X

Table 1

Sat_Flight1

Table 1 Columns:

- Calculated Column 1
- Destination_Airport**
- EndDateTime
- Hub Flight Hash Key
- Origin_Airport
- Scheduled_Arrival
- Scheduled_Departure

Cardinality: Many to One (*:1)

Filter Direction: << To Sat_Flight1

Active

Row Level Security

Table 2

SAT_AIRPORT

Table 2 Columns:

- Airport_Code**
- AirportName
- City
- Country
- EndDateTime
- Hub Airport Hash Key
- State

A calculated column 'Country' is created in 'Sat_Flight1'
 $=\text{RELATED}(\text{SAT_AIRPORT}[\text{Country}])$

File Edit View Git Project Build Debug Test Analyze Tools Extensions Window Help Search (Ctrl+Q) Airlines_Analysis

Model.bim* Develop Start Select Perspective <Default>

[Country] f_x = RELATED(SAT_AIRPORT[Country])

Hub Flight Hash Key	EndDateTime	Origin_Airport	Destination_Airport	Scheduled_Departure	Scheduled_Arrival	State	Country
1	173	MSP	DEN	600	719	CO	USA
2	174	PBI	IAH	600	750	TX	USA
3	175	PHX	IAH	600	928	TX	USA
4	176	RSW	IAH	600	736	TX	USA
5	177	EWR	MCO	600	853	FL	USA
6	178	OMA	ORD	600	737	IL	USA
7	179	SMF	PHX	600	850	AZ	USA
8	180	OAK	PHX	600	855	AZ	USA
9	181	PVD	CLT	600	819	NC	USA
10	182	RNO	DEN	600	910	CO	USA
11	183	PIT	IAH	600	815	TX	USA
12	184	DSM	IAH	600	631	TX	USA

Then total number of flights going to USA is calculated using the query:

```
TotalFlightsUSACountry:=CALCULATE (
    COUNT ( 'Sat_Flight1'[Hub Flight Hash Key] ),
    FILTER ('Sat_Flight1', [Country]="USA" )
)
```

Model.bim* X

[Destination_Airport] fx

```
TotalFlightsUSACountry:=CALCULATE (
    COUNT ('Sat_Flight1'[Hub Flight Hash Key]),
    FILTER ('Sat_Flight1', [Country]="USA")
)
```

Hub Flight Hash Key	EndDateTime	Origin_Airport	Destination_Airport	Scheduled_Departure	Scheduled_Arrival	State	Country
1	173	MSP	DEN	600	719	CO	USA
2	174	PBI	IAH	600	750	TX	USA
3	175	PHX	IAH	600	928	TX	USA
4	176	RSW	IAH	600	736	TX	USA
5	177	EWR	MCO	600	853	FL	USA
6	178	OMA	ORD	600	737	IL	USA
7	179	SMF	PHX	600	850	AZ	USA
8	180	OAK	PHX	600	855	AZ	USA
9	181	PVD	CLT	600	819	NC	USA
10	182	RNO	DEN	600	910	CO	USA
11	183	PIT	IAH	600	815	TX	USA
12	184	DEN	TAN	600	922	TX	USA

TotalFlightsDenverCity:
68

TotalFlightsAlaskaState:
4

TotalFlightsUSACountry
1006

A calculated table 'FlightsGoingToUSACountry' is created to display details of flights going to USA

= FILTER('Sat_Flight1', 'Sat_Flight1'[Country]="USA")

Model.bim* X

[Country] fx

= FILTER('Sat_Flight1', 'Sat_Flight1'[Country]="USA")

Hub Flight Hash Key	EndDateTime	Origin_Airport	Destination_Airport	Scheduled_Departure	Scheduled_Arrival	State	Country	Flight_Number
1	173	MSP	DEN	600	719	CO	USA	210
2	174	PBI	IAH	600	750	TX	USA	214
3	175	PHX	IAH	600	928	TX	USA	247
4	176	RSW	IAH	600	736	TX	USA	256
5	177	EWR	MCO	600	853	FL	USA	319
6	178	OMA	ORD	600	737	IL	USA	374
7	179	SMF	PHX	600	850	AZ	USA	635
8	180	OAK	PHX	600	855	AZ	USA	646
9	181	PVD	CLT	600	819	NC	USA	1941
10	182	RNO	DEN	600	910	CO	USA	422
11	183	PIT	IAH	600	815	TX	USA	544
12	184	DEN	IAH	600	922	TX	USA	1016
13	185	BOS	BWI	600	745	MD	USA	4122
14	186	LAS	BWI	600	1330	MD	USA	4822
15	187	ABQ	HOU	600	900	TX	USA	2215

4) Flights going to a state (Alaska)

A calculated column 'State' is created in 'Sat_Flight1' table

=RELATED('SAT_AIRPORT'[State])

	Hub Flight Hash Key	EndDateTime	Origin_Airport	Destination_Airport	Scheduled_Departure	Scheduled_Arrival	State	Add
1	173		MSP	DEN	600	719	CO	
2	174		PBI	IAH	600	750	TX	
3	175		PHX	IAH	600	928	TX	
4	176		RSW	IAH	600	736	TX	
5	177		EWR	MCO	600	853	FL	
6	178		OMA	ORD	600	737	IL	
7	179		SMF	PHX	600	850	AZ	
8	180		OAK	PHX	600	855	AZ	
9	181		PVD	CLT	600	819	NC	
10	182		RNO	DEN	600	910	CO	
11	183		PIT	IAH	600	815	TX	
12	184		DEN	IAH	600	922	TX	
13	185		BOS	BWI	600	745	MD	

Then total number of flights going to Alaska state is calculated using the query:

```
TotalFlightsAlaskaState:=CALCULATE (
    COUNT ( 'Sat_Flight1'[Hub Flight Hash Key] ),
    FILTER ('Sat_Flight1', [State]="AK" )
)
```

	Hub Flight Hash Key	EndDateTime	Origin_Airport	Destination_Airport	Scheduled_Departure	Scheduled_Arrival	State	
1	173		MSP	DEN	600	719	CO	
2	174		PBI	IAH	600	750	TX	
3	175		PHX	IAH	600	928	TX	
4	176		RSW	IAH	600	736	TX	
5	177		EWR	MCO	600	853	FL	
6	178		OMA	ORD	600	737	IL	
7	179		SMF	PHX	600	850	AZ	
8	180		OAK	PHX	600	855	AZ	
9	181		PVD	CLT	600	819	NC	
10	182		RNO	DEN	600	910	CO	
11	183		PIT	IAH	600	815	TX	
12	184		DEN	IAH	600	922	TX	
13	185		BOS	BWI	600	745	MD	

TotalFlightsDenverCity: 68
TotalFlightsAlaskaState: 4

A calculated table 'FlightsGoingToAlaskaState' is created to display flight details going to Alaska State

```
= FILTER('Sat_Flight1', 'Sat_Flight1'[State]="AK")
```

Hub Flight Hash Key	EndDateTime	Origin_Airport	Destination_Airport	Scheduled_Departure	Scheduled_Arrival	State	Country	Flight_Number
1	228	SEA	ANC	600	854	AK	USA	81
2	229	FAI	ANC	600	713	AK	USA	162
3	5	SEA	ANC	25	320	AK	USA	135
4	601	ANC	OTZ	630	812	AK	USA	157

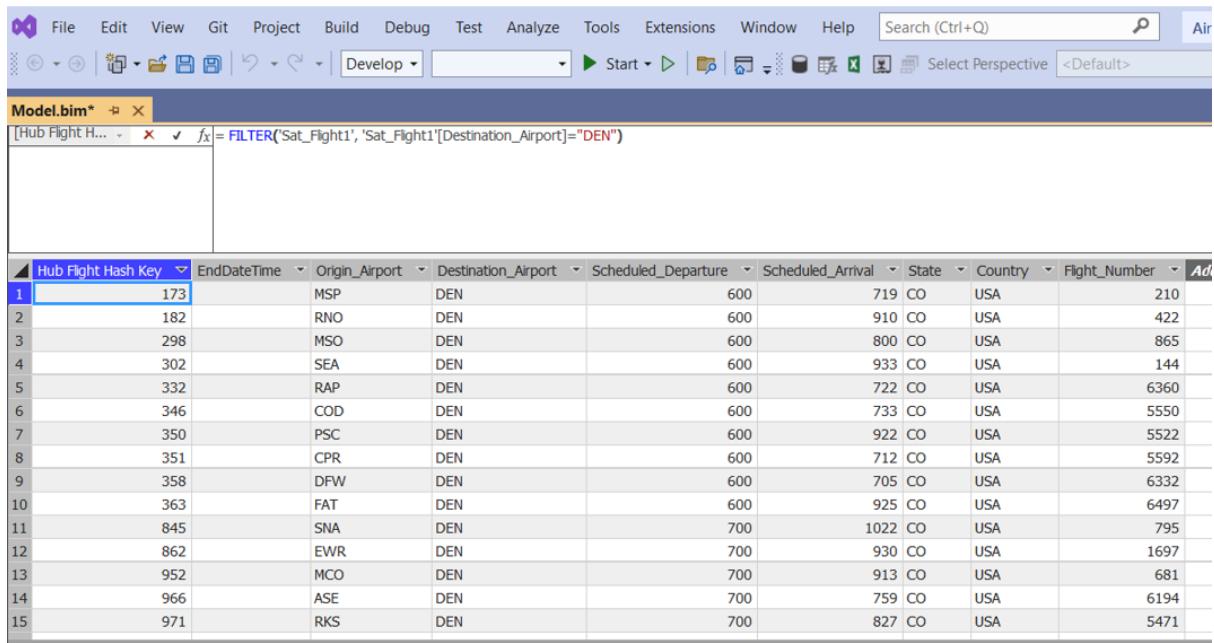
5) Flights Going to a city (Denver)

```
TotalFlightsDenverCity:=CALCULATE (
    COUNT ( 'Sat_Flight1'[Hub Flight Hash Key] ),
    FILTER ( 'Sat_Flight1',[Destination_Airport]="DEN" )
)
```

Hub Flight Hash Key	EndDateTime	Origin_Airport	Destination_Airport	Scheduled_Departure	Scheduled_Arrival	TotalFlightsDenverCity:
1	173	MSP	DEN	600	719	68
2	174	PBI	IAH	600	750	68
3	175	PHX	IAH	600	928	68
4	176	RSW	IAH	600	736	68
5	177	EWR	MCO	600	853	68
6	178	OMA	ORD	600	737	68
7	179	SMF	PHX	600	850	68
8	180	OAK	PHX	600	855	68
9	181	PVD	CLT	600	819	68
10	182	RNO	DEN	600	910	68
11	183	PIT	IAH	600	815	68
12	184	DEN	IAH	600	922	68
13	185	BOS	BWI	600	745	68

A calculated table 'FlightsGoingToDenverCity' is created to find Flights going to Denver city

```
= FILTER('Sat_Flight1', 'Sat_Flight1'[Destination_Airport] = "DEN")
```



The screenshot shows the Power BI Model view with a table titled 'Hub Flight Hash Key'. The table contains 15 rows of flight data. The columns are: Hub Flight Hash Key, EndDateTime, Origin_Airport, Destination_Airport, Scheduled_Departure, Scheduled_Arrival, State, Country, Flight_Number, and Ad. A filter is applied to the Destination_Airport column, specifically for flights to DEN. The first row is selected, showing a flight from MSP to DEN at 600 UTC, arriving at 719 UTC.

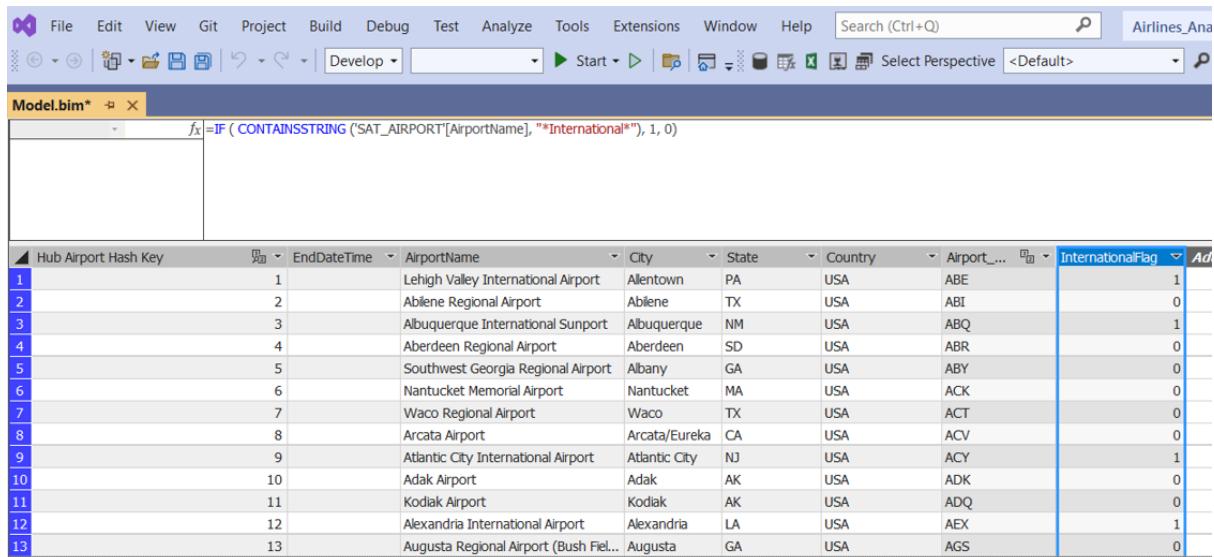
Hub Flight Hash Key	EndDateTime	Origin_Airport	Destination_Airport	Scheduled_Departure	Scheduled_Arrival	State	Country	Flight_Number	Ad
1	173	MSP	DEN	600	719	CO	USA	210	
2	182	RNO	DEN	600	910	CO	USA	422	
3	298	MSO	DEN	600	800	CO	USA	865	
4	302	SEA	DEN	600	933	CO	USA	144	
5	332	RAP	DEN	600	722	CO	USA	6360	
6	346	COD	DEN	600	733	CO	USA	5550	
7	350	PSC	DEN	600	922	CO	USA	5522	
8	351	CPR	DEN	600	712	CO	USA	5592	
9	358	DFW	DEN	600	705	CO	USA	6332	
10	363	FAT	DEN	600	925	CO	USA	6497	
11	845	SNA	DEN	700	1022	CO	USA	795	
12	862	EWR	DEN	700	930	CO	USA	1697	
13	952	MCO	DEN	700	913	CO	USA	681	
14	966	ASE	DEN	700	759	CO	USA	6194	
15	971	RKS	DEN	700	827	CO	USA	5471	

6) Perform calculation to Identify Regional/International Airport.

To find international airports

A calculated column 'InternationalFlag' is created in 'SAT_AIRPORT' table

=IF (CONTAINSSTRING ('SAT_AIRPORT'[AirportName], "*International*"), 1, 0)

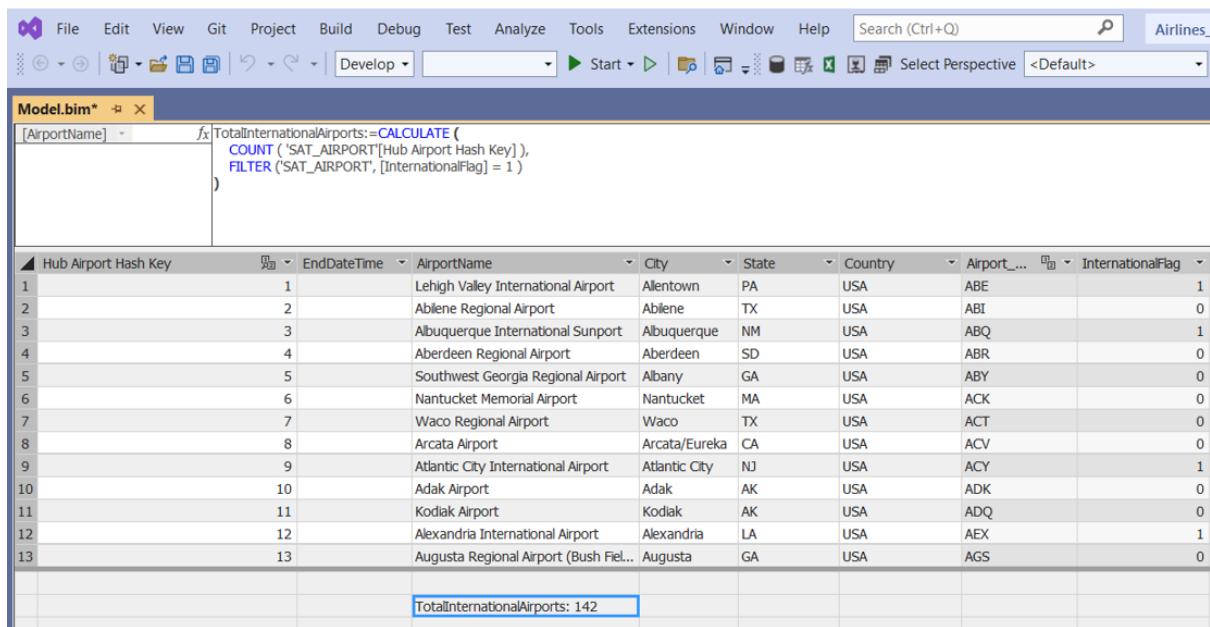


The screenshot shows the Power BI Model view with a table titled 'Hub Airport Hash Key'. The table contains 13 rows of airport data. The columns are: Hub Airport Hash Key, EndDateTime, AirportName, City, State, Country, and InternationalFlag. A calculated column 'InternationalFlag' is present, which is 1 for international airports and 0 for regional airports. The first row is selected, showing Lehigh Valley International Airport in Allentown, PA.

Hub Airport Hash Key	EndDateTime	AirportName	City	State	Country	InternationalFlag
1	1	Lehigh Valley International Airport	Allentown	PA	USA	1
2	2	Abilene Regional Airport	Abilene	TX	USA	0
3	3	Albuquerque International Sunport	Albuquerque	NM	USA	1
4	4	Aberdeen Regional Airport	Aberdeen	SD	USA	0
5	5	Southwest Georgia Regional Airport	Albany	GA	USA	0
6	6	Nantucket Memorial Airport	Nantucket	MA	USA	0
7	7	Waco Regional Airport	Waco	TX	USA	0
8	8	Arcata Airport	Arcata/Eureka	CA	USA	0
9	9	Atlantic City International Airport	Atlantic City	NJ	USA	1
10	10	Adak Airport	Adak	AK	USA	0
11	11	Kodiak Airport	Kodiak	AK	USA	0
12	12	Alexandria International Airport	Alexandria	LA	USA	1
13	13	Augusta Regional Airport (Bush Field)	Augusta	GA	USA	0

Then total number of international airports is calculated using the query:

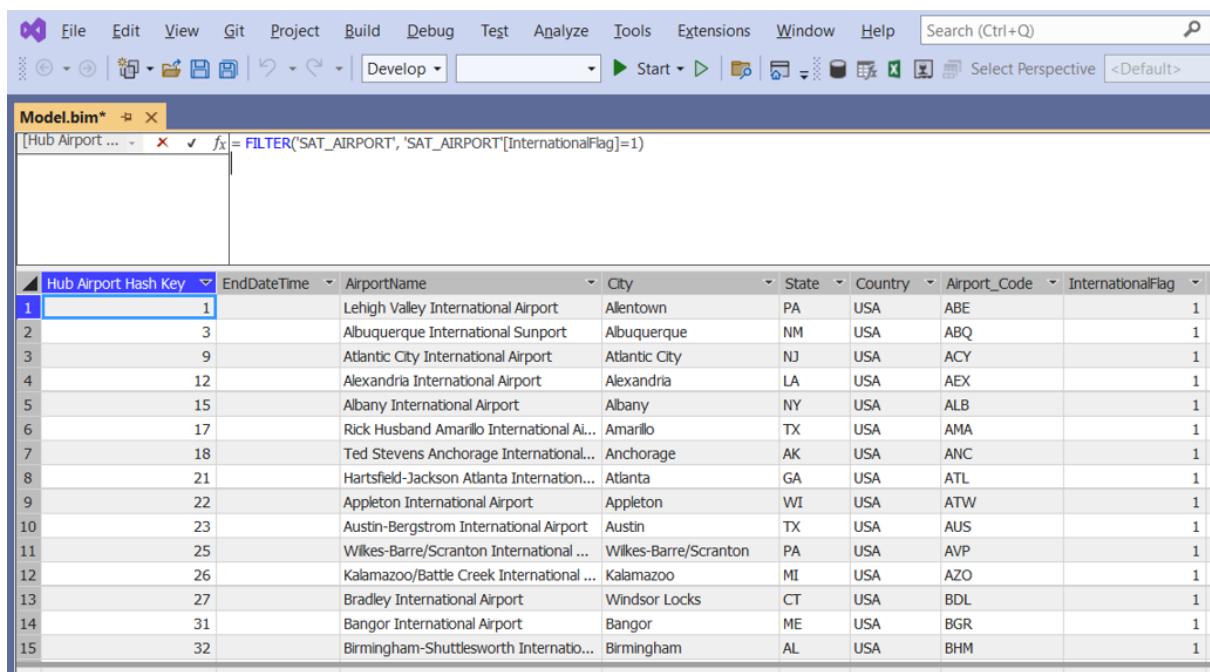
```
TotalInternationalAirports:=CALCULATE (
    COUNT ( 'SAT_AIRPORT'[Hub Airport Hash Key] ),
    FILTER ('SAT_AIRPORT', [InternationalFlag] = 1 )
)
```



The screenshot shows the Power BI Model view with a calculated table named 'InternationalAirports'. The table has 142 rows and 9 columns. The columns are: Hub Airport Hash Key, EndDateTime, AirportName, City, State, Country, Airport_Code, InternationalFlag, and a calculated column 'InternationalFlag' which is set to 1. The calculated column formula is: `=CALCULATE (COUNT ('SAT_AIRPORT'[Hub Airport Hash Key]), FILTER ('SAT_AIRPORT', [InternationalFlag] = 1))`. The table also contains a footer cell with the value 'TotalInternationalAirports: 142'.

A calculated table 'InternationalAirports' is created to show details of international airports

`= FILTER('SAT_AIRPORT', 'SAT_AIRPORT'[InternationalFlag]=1)`



The screenshot shows the Power BI Model view with a calculated table named 'Hub Airports'. The table has 32 rows and 9 columns. The columns are: Hub Airport Hash Key, EndDateTime, AirportName, City, State, Country, Airport_Code, InternationalFlag, and a calculated column 'InternationalFlag' which is set to 1. The calculated column formula is: `= FILTER('SAT_AIRPORT', 'SAT_AIRPORT'[InternationalFlag]=1)`. The table contains a footer cell with the value 'TotalHubAirports: 32'.

To find regional airports

A calculated column 'RegionalFlag' is created in 'SAT_AIRPORT' table

`=IF (CONTAINSSTRING ('SAT_AIRPORT'[AirportName], "*Regional*"), 1, 0)`

Model.bim*  Airlines_Analysis                                                       

Model.bim* ✚ ✚ ✖ fx = FILTER('SAT_AIRPORT', 'SAT_AIRPORT[RegionalFlag]=1')

Hub Airport Hash Key	EndDateTime	AirportName	City	State	Country	Airport_Code	InternationalFlag	RegionalFlag	Add
1	2	Abilene Region...	Abilene	TX	USA	ABI	0	1	
2	4	Aberdeen Regi...	Aberd...	SD	USA	ABR	0	1	
3	5	Southwest Ge...	Albany	GA	USA	ABY	0	1	
4	7	Waco Regiona...	Waco	TX	USA	ACT	0	1	
5	13	Augusta Regio...	Augusta	GA	USA	AGS	0	1	
6	16	Waterloo Regi...	Water...	IA	USA	ALO	0	1	
7	19	Alpena County...	Alpena	MI	USA	APN	0	1	
8	24	Asheville Regio...	Asheville	NC	USA	AVL	0	1	
9	35	Bemidji Region...	Bemidji	MN	USA	BJI	0	1	
10	37	Central Illinois ...	Bloom...	IL	USA	BMI	0	1	
11	41	Jack Brooks R...	Beau...	TX	USA	BPT	0	1	
12	44	Brainerd Lakes...	Brainerd	MN	USA	BRD	0	1	
13	55	Akron-Canton ...	Akron	OH	USA	CAK	0	1	
14	56	Cedar City Re...	Cedar...	UT	USA	CDC	0	1	
15	72	Yellowstone R...	Cody	WY	USA	COD	0	1	

7) Flights going to every State.

A calculated column 'Flight_Numbers' is created in 'Sat_Flight' table

=RELATED(Hub_Flight1[Flight_Number])

Model.bim* ✚ ✚ ✖ fx =RELATED(Hub_Flight1[Flight_Number])

Hub Flight Hash Key	Flight_Number	EndDateTime	Origin_Airport	Destination_Airport	Scheduled_Departure	Scheduled_Arrival	State	Country
1	173	210	MSP	DEN	600	719	CO	USA
2	174	214	PBI	IAH	600	750	TX	USA
3	175	247	PHX	IAH	600	928	TX	USA
4	176	256	RSW	IAH	600	736	TX	USA
5	177	319	EWR	MCO	600	853	FL	USA
6	178	374	OMA	ORD	600	737	IL	USA
7	179	635	SMF	PHX	600	850	AZ	USA
8	180	646	OAK	PHX	600	855	AZ	USA
9	181	1941	PVD	CLT	600	819	NC	USA
10	182	422	RNO	DEN	600	910	CO	USA
11	183	544	PIT	IAH	600	815	TX	USA
12	184	1016	DEN	IAH	600	922	TX	USA
13	185	4122	BOS	BWI	600	745	MD	USA
14	186	4822	LAS	BWI	600	1330	MD	USA

Then a calculated table 'FlightsGoingToEveryState' is created to find number of flights going to each state

=GROUPBY('Sat_Flight1', 'Sat_Flight1[State]', "NumberOfRunningFlights", COUNTX(CURRENTGROUP(), COUNT(Sat_Flight1[Flight_Number])))

Model.bim*

[Sat_Flight1_State] fx =GROUPBY('Sat_Flight1', 'Sat_Flight1'[State], "NumberOfRunningFlights", COUNTX(CURRENTGROUP()), COUNT(Sat_Flight1[Flight_Number])))

Sat_Flight1_State	NumberOfRunningFlights	Add Column
1 CO	68	
2 TX	198	
3 FL	121	
4 IL	70	
5 AZ	41	
6 NC	29	
7 MD	11	
8 UT	17	
9 GA	76	
10 NJ	33	
11 CA	93	
12 PA	6	
13 NY	54	
14 AK	4	
15 WA	29	

A calculated table 'FlightsGoingToEveryStateDetails' is created to show flight numbers of flights going to every state

=GROUPBY('Sat_Flight1', 'Sat_Flight1'[State], 'Sat_Flight1'[Flight_Number])

Model.bim*

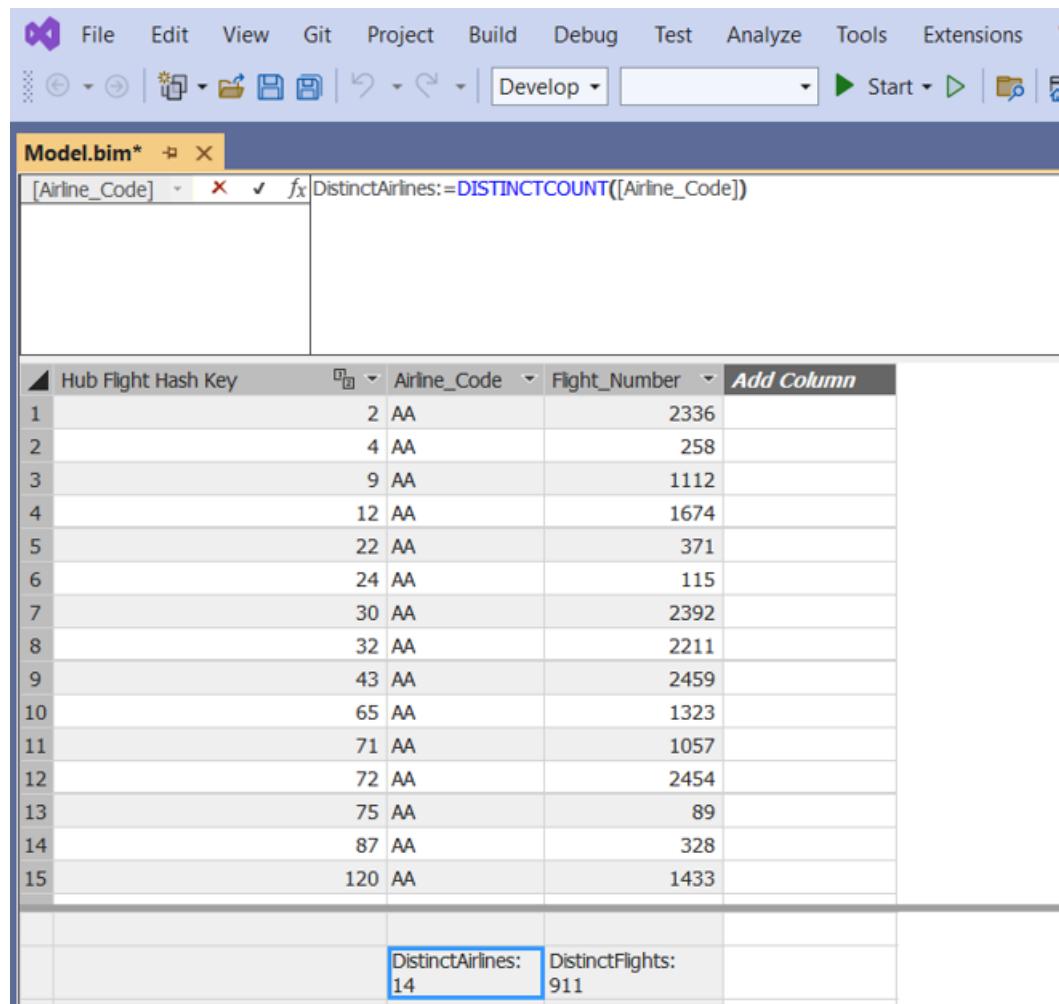
[State] fx =GROUPBY('Sat_Flight1', 'Sat_Flight1'[State], 'Sat_Flight1'[Flight_Number])

State	Flight_Number	Add Column
1 TX	214	
2 TX	247	
3 TX	256	
4 TX	544	
5 TX	1016	
6 TX	2215	
7 TX	1188	
8 TX	304	
9 TX	175	
10 TX	281	
11 TX	1103	
12 TX	1039	
13 TX	1055	
14 TX	1180	
15 TX	1345	

8) Find Different Airlines Available

Total no. of different airlines

DistinctAirlines:=DISTINCTCOUNT([Airline_Code])



The screenshot shows the Power BI desktop application. The ribbon menu at the top includes File, Edit, View, Git, Project, Build, Debug, Test, Analyze, Tools, and Extensions. The 'Model.bim*' file is open in the workspace. A calculated column 'DistinctAirlines' is defined with the formula =DISTINCTCOUNT([Airline_Code]). The data view shows a table with three columns: 'Hub Flight Hash Key', 'Airline_Code', and 'Flight_Number'. The 'Airline_Code' column contains 14 distinct values: AA (repeated 13 times) and 87. The 'Flight_Number' column contains 14 distinct values. At the bottom of the data view, there are summary values: 'DistinctAirlines: 14' and 'DistinctFlights: 911'.

Hub Flight Hash Key	Airline_Code	Flight_Number
1	2 AA	2336
2	4 AA	258
3	9 AA	1112
4	12 AA	1674
5	22 AA	371
6	24 AA	115
7	30 AA	2392
8	32 AA	2211
9	43 AA	2459
10	65 AA	1323
11	71 AA	1057
12	72 AA	2454
13	75 AA	89
14	87 AA	328
15	120 AA	1433

To get names of different airlines create new calculated table 'NamesOfDifferentAirlines' is created

=DISTINCT('Hub_Flight1'[Airline_Code])

Model.bim*

[Airline_Code]	$f_x = \text{DISTINCT}(\text{Hub_Flight1}[Airline_Code])$
1 AS	
2 AA	
3 US	
4 DL	
5 NK	
6 UA	
7 HA	
8 B6	
9 OO	
10 EV	
11 MQ	
12 F9	
13 WN	
14 VX	

9) Which Airline has the maximum running flights

A calculated table 'AirlinesWithFlights' is created

```
=GROUPBY('Hub_Flight1', 'Hub_Flight1'[Airline_Code], "NumberOfRunningFlights",
COUNTX(CURRENTGROUP(), COUNT(Hub_Flight1[Flight_Number])))
```

Model.bim*

[Hub_Flight1_Airline_Code]	NumberOfRunningFlights	Add Column
1 AA	135	
2 OO	121	
3 DL	118	
4 UA	103	
5 B6	91	
6 EV	78	
7 WN	72	
8 MQ	68	
9 AS	60	
10 US	59	
11 NK	40	
12 HA	17	
13 F9	34	
14 VX	4	

A calculated table 'AirlineWithMaximumFlights' is created to find airline with Max number of Running Flights

```
= FILTER('AirlinesWithFlights',
'AirlinesWithFlights'[NumberOfRunningFlights]=MAX('AirlinesWithFlights'[NumberofRunningFlights]))
```

The screenshot shows the Power BI Model view. The ribbon at the top has 'File', 'Edit', 'View', 'Git', 'Project', 'Build', 'Debug', 'Test', 'Analyze', 'Tools', 'Extensions', 'Window', 'Help', and 'Search (Ctrl+Q)'. The 'Develop' tab is selected. The main area shows a table named 'Hub_Flight1_Airline_Code'. The formula bar at the top has the formula: `= FILTER('AirlinesWithFlights', 'AirlinesWithFlights'[NumberOfRunningFlights]=MAX('AirlinesWithFlights'[NumberofRunningFlights]))`. The table has two columns: 'Hub_Flight1_Airline_Code' and 'NumberOfRunningFlights'. There is one row with the value 'AA' in the first column and '135' in the second column.

10) No of Airports in the state AZ, DE, and NY.

Number of airports for individual cities:

```
TotalAirportsAZ:=CALCULATE (
COUNT ( 'SAT_AIRPORT'[Hub Airport Hash Key] ),
FILTER ('SAT_AIRPORT', [State] = "AZ" )
)
```

The screenshot shows the Power BI Model view. The ribbon at the top has 'File', 'Edit', 'View', 'Git', 'Project', 'Build', 'Debug', 'Test', 'Analyze', 'Tools', 'Extensions', 'Window', 'Help', and 'Search (Ctrl+Q)'. The 'Develop' tab is selected. The main area shows a table named 'TotalAirportsAZ' with the formula: `TotalAirportsAZ:=CALCULATE (COUNT ('SAT_AIRPORT'[Hub Airport Hash Key]), FILTER ('SAT_AIRPORT', [State] = "AZ"))`. Below it is a data grid showing airports for state AZ. The grid has columns: 'Hub Airport Hash Key', 'EndDateTime', 'AirportName', 'City', 'State', 'Country', and 'Airport'. The data includes rows for Lehigh Valley International Airport, Abilene Regional Airport, Albuquerque International Sunport, Aberdeen Regional Airport, Southwest Georgia Regional Airport, Nantucket Memorial Airport, Waco Regional Airport, Arcata Airport, Atlantic City International Airport, Adak Airport, Kodiak Airport, Alexandria International Airport, and Augusta Regional Airport (Bush Field). The last two rows show summary counts: 'TotalInternationalAirports: 142' and 'TotalRegionalAirports: 83'.

```
TotalAirportsDE:=CALCULATE (
COUNT ( 'SAT_AIRPORT'[Hub Airport Hash Key] ),
FILTER ('SAT_AIRPORT', [State] = "DE" )
)
```

File Edit View Git Project Build Debug Test Analyze Tools Extensions Window Help Search (Ctrl+Q) Airlines_Analysis

Model.bim* fx

```
[State] fx TotalAirportsDE:=CALCULATE (
  COUNT ('SAT_AIRPORT'[Hub Airport Hash Key]),
  FILTER ('SAT_AIRPORT', [State] = "DE")
)
```

Hub Airport Hash Key	EndDateTime	AirportName	City	State	Country	Airport...
1	1	Lehigh Valley International Airport	Allentown	PA	USA	ABE
2	2	Abilene Regional Airport	Abilene	TX	USA	ABI
3	3	Albuquerque International Sunport	Albuquerque	NM	USA	ABQ
4	4	Aberdeen Regional Airport	Aberdeen	SD	USA	ABR
5	5	Southwest Georgia Regional Airport	Albany	GA	USA	ABY
6	6	Nantucket Memorial Airport	Nantucket	MA	USA	ACK
7	7	Waco Regional Airport	Waco	TX	USA	ACT
8	8	Arcata Airport	Arcata/Eureka	CA	USA	ACV
9	9	Atlantic City International Airport	Atlantic City	NJ	USA	ACY
10	10	Adak Airport	Adak	AK	USA	ADK
11	11	Kodiak Airport	Kodiak	AK	USA	ADQ
12	12	Alexandria International Airport	Alexandria	LA	USA	AEX
13	13	Augusta Regional Airport (Bush Fiel...	Augusta	GA	USA	AGS

TotalInternationalAirports: 142 TotalAirportsAZ: 4
TotalRegionalAirports: 83 TotalAirportsDE: 1

```
TotalAirportsNY:=CALCULATE (
  COUNT ('SAT_AIRPORT'[Hub Airport Hash Key]),
  FILTER ('SAT_AIRPORT', [State] = "NY")
)
```

File Edit View Git Project Build Debug Test Analyze Tools Extensions Window Help Search (Ctrl+Q) Airlines_Analysis

Model.bim* fx

```
[State] fx TotalAirportsNY:=CALCULATE (
  COUNT ('SAT_AIRPORT'[Hub Airport Hash Key]),
  FILTER ('SAT_AIRPORT', [State] = "NY")
)
```

Hub Airport Hash Key	EndDateTime	AirportName	City	State	Country
1	1	Lehigh Valley International Airport	Allentown	PA	USA
2	2	Abilene Regional Airport	Abilene	TX	USA
3	3	Albuquerque International Sunport	Albuquerque	NM	USA
4	4	Aberdeen Regional Airport	Aberdeen	SD	USA
5	5	Southwest Georgia Regional Airport	Albany	GA	USA
6	6	Nantucket Memorial Airport	Nantucket	MA	USA
7	7	Waco Regional Airport	Waco	TX	USA
8	8	Arcata Airport	Arcata/Eureka	CA	USA
9	9	Atlantic City International Airport	Atlantic City	NJ	USA
10	10	Adak Airport	Adak	AK	USA
11	11	Kodiak Airport	Kodiak	AK	USA
12	12	Alexandria International Airport	Alexandria	LA	USA
13	13	Augusta Regional Airport (Bush Fiel...	Augusta	GA	USA
14	14	King Salmon Airport	King Salmon	AK	USA

TotalInternationalAirports: 142 TotalAirportsAZ: 4
TotalRegionalAirports: 83 TotalAirportsDE: 1
TotalAirportsNY: 14

Number of airports for three cities combined:

```
TotalAirportsAZDENY:=CALCULATE (
  COUNT ('SAT_AIRPORT'[Hub Airport Hash Key]),
  FILTER ('SAT_AIRPORT', [State] = "AZ" || [State] = "DE" || [State] = "NY")
)
```

File Edit View Git Project Build Debug Test Analyze Tools Extensions Window Help Search (Ctrl+Q) Ai

Develop Start Select Perspective <Default>

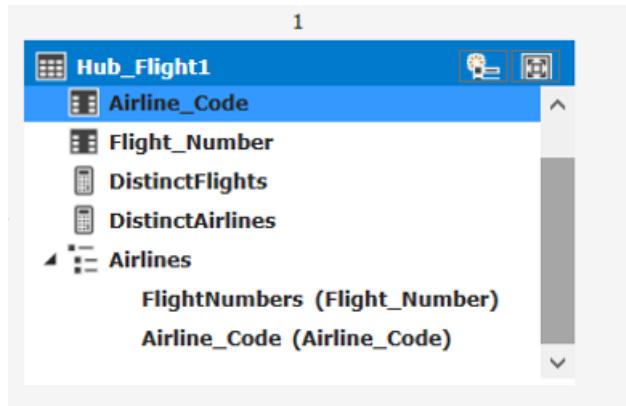
Model.bim*

```
[State] x ✓ fx TotalAirportsAZDENY:=CALCULATE (
    COUNT ("SAT_AIRPORT[Hub Airport Hash Key]"),
    FILTER ("SAT_AIRPORT", [State] = "AZ" || [State] = "DE" || [State] = "NY")
)
```

Hub Airport Hash Key	EndDateTime	AirportName	City	State	Country
1	1	Lehigh Valley International Airport	Allentown	PA	USA
2	2	Abilene Regional Airport	Abilene	TX	USA
3	3	Albuquerque International Sunport	Albuquerque	NM	USA
4	4	Aberdeen Regional Airport	Aberdeen	SD	USA
5	5	Southwest Georgia Regional Airport	Albany	GA	USA
6	6	Nantucket Memorial Airport	Nantucket	MA	USA
7	7	Waco Regional Airport	Waco	TX	USA
8	8	Arcata Airport	Arcata/Eureka	CA	USA
9	9	Atlantic City International Airport	Atlantic City	NJ	USA
10	10	Adak Airport	Adak	AK	USA
11	11	Kodiak Airport	Kodiak	AK	USA
12	12	Alexandria International Airport	Alexandria	LA	USA
13	13	Augusta Regional Airport (Bush Fiel...	Augusta	GA	USA
14	14	King Salmon Airport	King Salmon	AK	USA

TotalInternationalAirports: 142
 TotalRegionalAirports: 83
 TotalAirportsAZ: 4
 TotalAirportsDE: 1
 TotalAirportsNY: 14
 TotalAirportsAZDENY: 19

11) Create Hierarchy for Airlines and its flight numbers.



VISUALIZATION OF DATA

❖ Tools/ Technologies used:

- Power BI Desktop

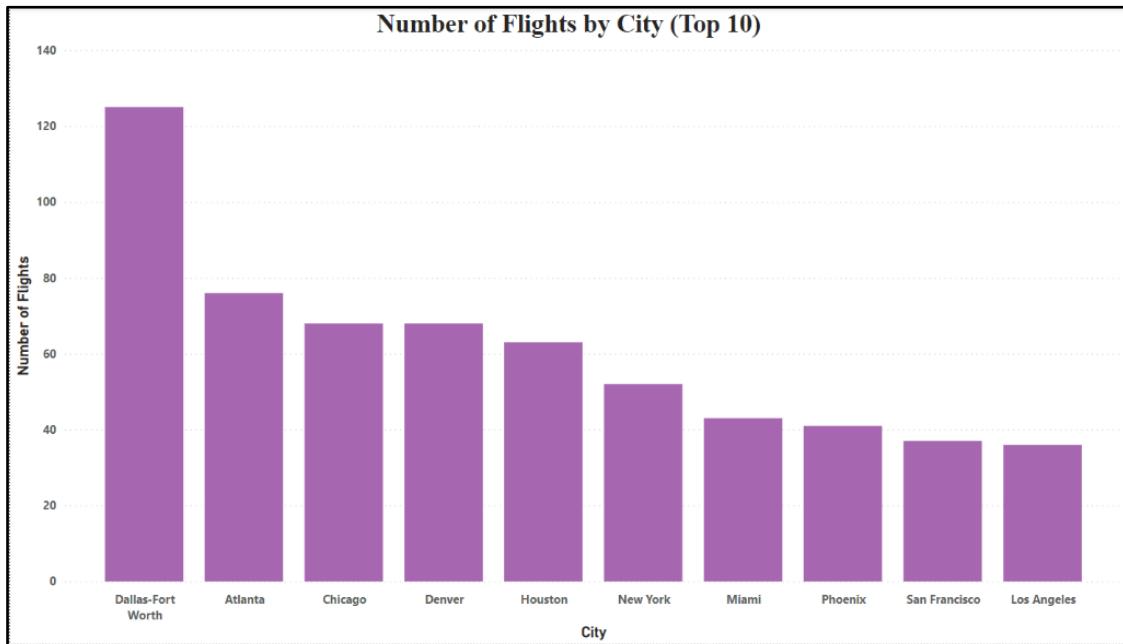


❖ Visualisations:

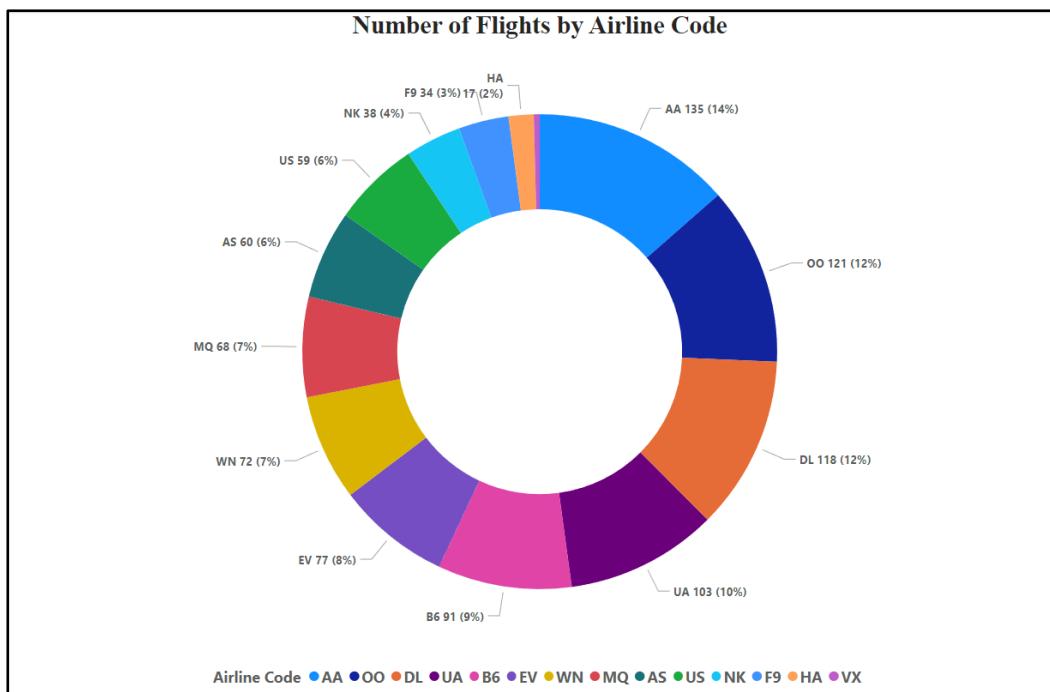
1. State-wise location of different airports



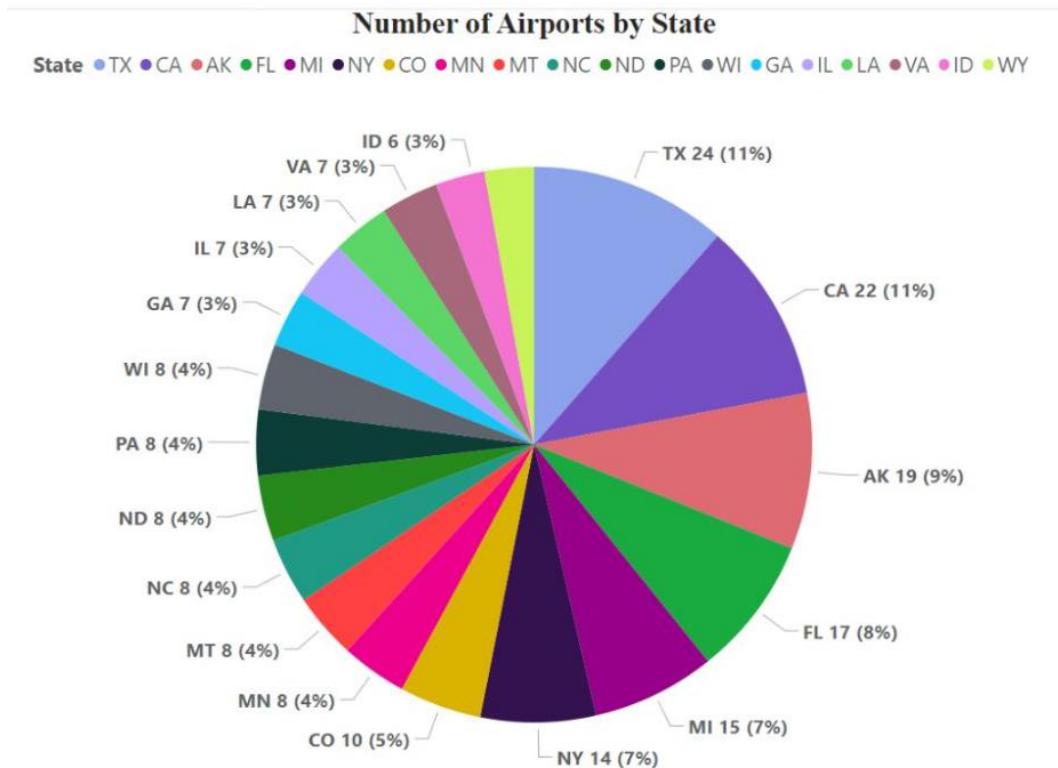
2. Top 10 cities with maximum number of flights running



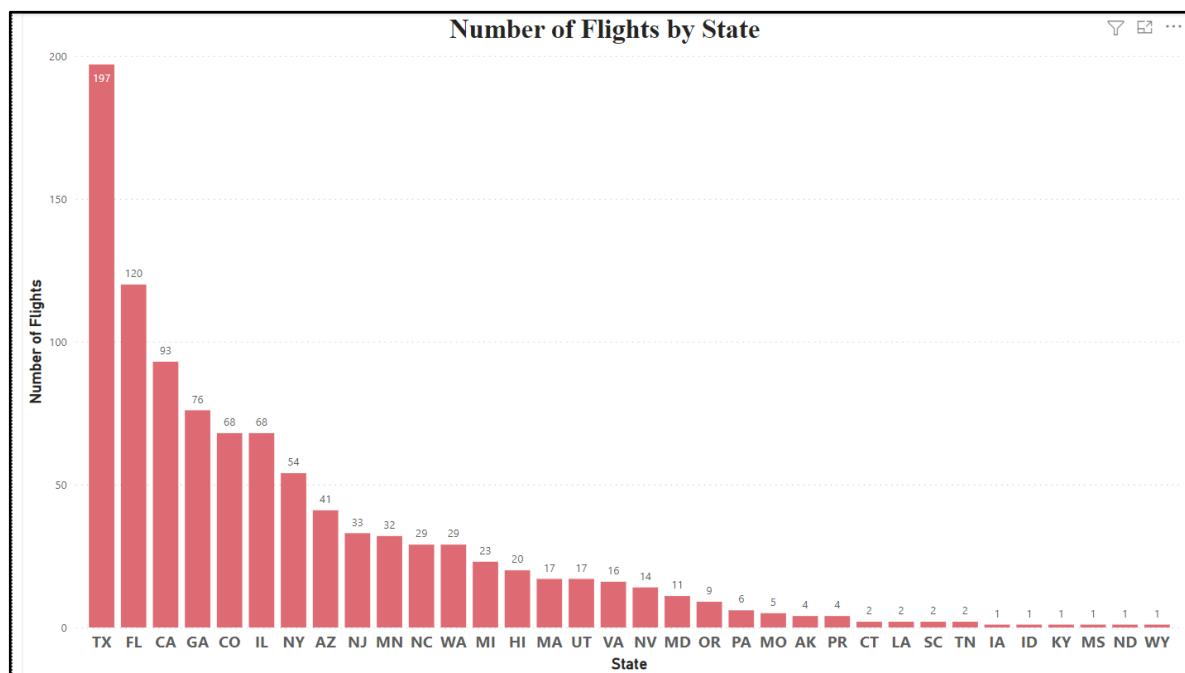
3. Airline-wise number of flights running



4. State-wise number of airports



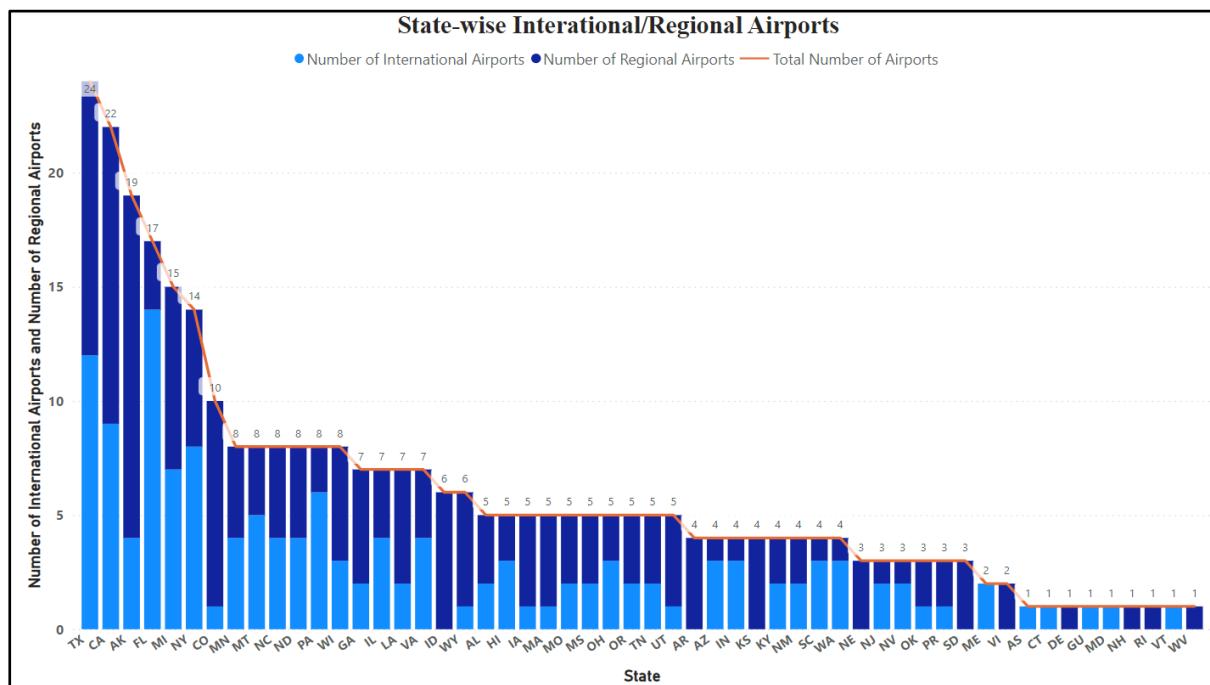
5. State-wise number of flights running



6. Region-wise airport hierarchy table

Row Labels
<input type="checkbox"/> ABE
<input type="checkbox"/> Lehigh Valley International Airport
<input type="checkbox"/> Allentown
PA
<input type="checkbox"/> ABI
<input type="checkbox"/> Abilene Regional Airport
<input type="checkbox"/> Abilene
TX
<input type="checkbox"/> ABQ
<input type="checkbox"/> Albuquerque International Sunport
<input type="checkbox"/> Albuquerque
NM
<input type="checkbox"/> ABR
<input type="checkbox"/> Aberdeen Regional Airport
<input type="checkbox"/> Aberdeen
SD
<input type="checkbox"/> ABY
<input type="checkbox"/> Southwest Georgia Regional Airport
<input type="checkbox"/> Albany
GA
<input type="checkbox"/> ACK
<input type="checkbox"/> Nantucket Memorial Airport
<input type="checkbox"/> Nantucket
MA
<input type="checkbox"/> ACT
<input type="checkbox"/> Waco Regional Airport
<input type="checkbox"/> Waco
TX

7. State-wise number of International/Regional airports



CHALLENGES

- DAX queries was a new concept, so we found it initially difficult to implement. But we referred the DAX documentation and understood it.
- Data cleaning also was very tricky. We had to run the queries in SSMS before loading to tabular model to check for duplicate entries. We also checked unwanted columns and removed them before loading.

SUMMARY

The learning outcomes from this case study are:

1. Successfully created tabular model project.
2. Built relationships between tables
3. Implemented DAX queries to create required measures and hierarchies.
4. Deployed the model to SQL Server Analysis Services
5. Created possible visualizations of data

APPENDIX

DAX Queries Documentation:

<https://dax.guide/>

<https://learn.microsoft.com/en-us/analysis-services/ssas-overview?view=asallproducts-allversions>

Microsoft Support Link:

<https://learn.microsoft.com/en-us/dax/dax-queries>

Github Link:

https://github.com/aj1497/Airlines_Analysis