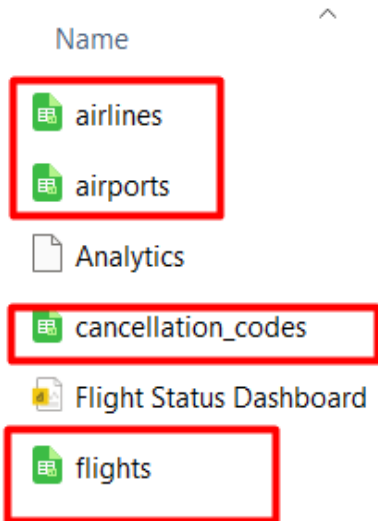




**Generating intelligence and insight
from Airline flight Data
2015**

Introduction

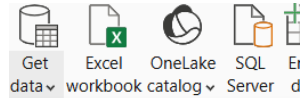


This analysis is based on a set of aviation data which includes the following files: -

- “**Airlines.csv**” which consists detailed records of airline operations.
- “**Airport.csv**” which comprises of airport codes,
- “**flight.csv**” which comprises of flight information and,
- “**cancellation codes.csv**” which comprises of cancelation description and code

Our aviation dataset captures the operational dynamics between airlines and airports, with a particular focus on flight volumes and cancellation patterns. This dataset enables us to determine a structured assessment of airline performance, airport activity levels, and the underlying reasons for flight cancellations. This foundational overview sets the stage for deeper insights into operational efficiency, disruption causes, and potential areas for systemic improvement in air transport services.

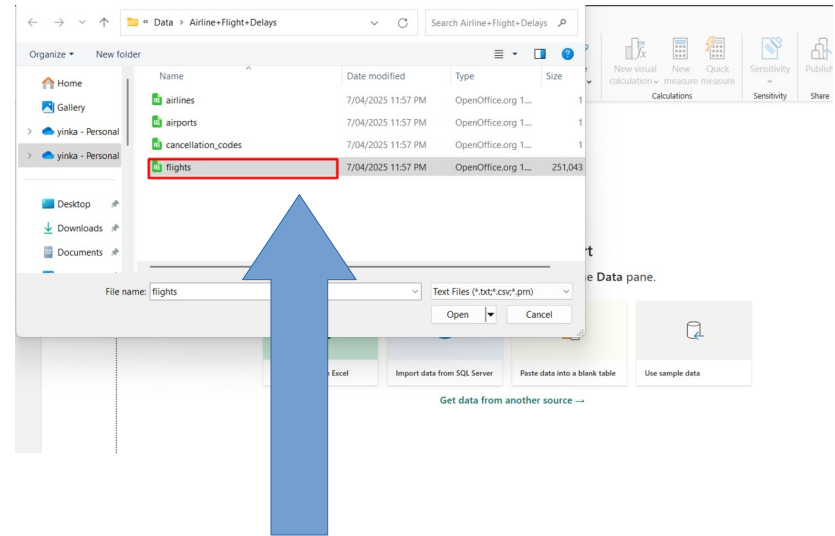
Reading our Dataset



Common data sources

- Excel workbook
- Power BI semantic models
- Dataflows
- Dataverse
- SQL Server
- Analysis Services
- Text/CSV**
- Web
- OData feed
- Blank query
- Power BI Template Apps
- More...

To analyse the data provided using PowerBI we read our CSV files “Text/CSV” option using the “Get Data”



To analyse the data provided using PowerBI we read our CSV files “Text/CSV” option using the “Get Data”

Analyzing and transforming our Dataset

flights.csv

File Origin: 1252: Western European (Windows) | Delimiter: Comma | Data Type Detection: Based on first 200 rows

	YEAR	MONTH	DAY	DAY_OF_WEEK	AIRLINE	FLIGHT_NUMBER	TAIL_NUMBER	ORIGIN_AIRPORT	DESTINATION_AIRPORT
1	2015	1	1		AA	2336	N3KUAA	LAX	PBI
2	2015	1	1		US	840	N171US	SFO	CLT
3	2015	1	1		AA	258	N3HYAA	LAX	MIA
5	2015	1	1		DL	806	N3730B	SFO	MSP
6	2015	1	1		NK	612	N635NK	LAS	MSP
7	2015	1	1		US	2013	N584UW	LAX	CLT
8	2015	1	1		AA	1112	N31AAA	SFO	DFW
9	2015	1	1		DL	1173	N826DN	LAS	ATL
10	2015	1	1		DL	2336	N958DN	DEN	ATL
11	2015	1	1		AA	1674	N853AA	LAS	MIA
12	2015	1	1		DL	1434	N547US	LAX	MSP
17	2015	1	1		UA	1197	N7844B	SFO	IAH
20	2015	1	1		NK	520	N525NK	LAS	MCI
22	2015	1	1		NK	214	N632NK	LAS	DFW
23	2015	1	1		AA	115	N3CTAA	LAX	MIA
24	2015	1	1		DL	1450	N671DN	LAS	MSP
25	2015	1	1		UA	1545	N76517	LAX	IAH
27	2015	1	1		NK	597	N528NK	MSP	FLL
28	2015	1	1		US	413	N571UW	LAS	CLT
29	2015	1	1		AA	2392	N3HRAA	DEN	MIA

Buttons: Load, Transform Data, Cancel



Add Conditional Column

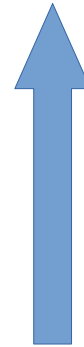
Add a conditional column that is computed from the other columns or values.

New column name: Status

	Column Name	Operator	Value	Output
If	CANCELLED	equals	ABC 123 * 1	Then ABC 123 Canceled
Else If	DEPARTURE_DELAY	equals	ABC 123 * 0	Then ABC 123 Delayed ...

Add Clause

Else On-Time

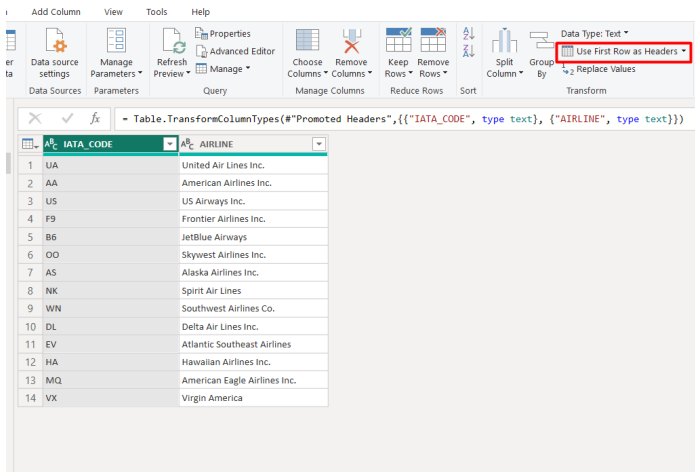


OK Cancel

Reviewing our dataset allows us to clean our dataset or carry out operation such as turning our rows into headers. This activity can be carried out using our “Transform data” button

We can modify and add columns using conditional statements

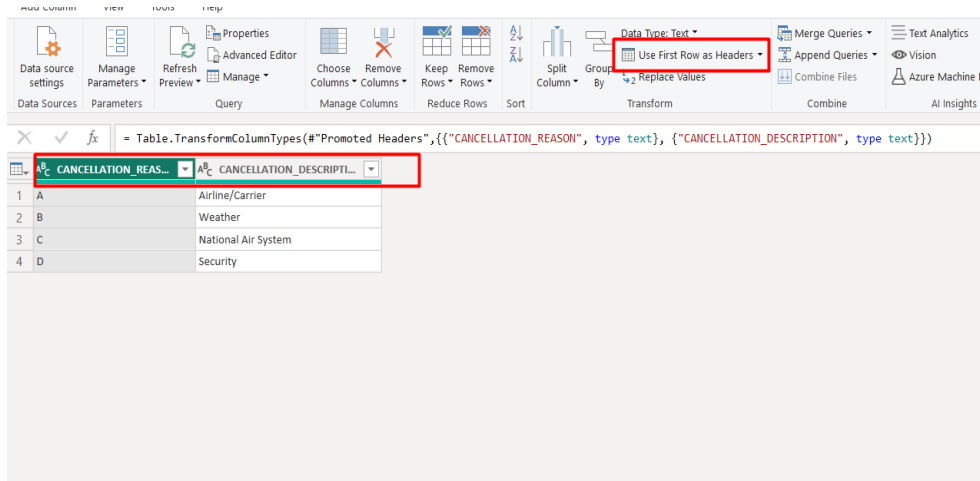
Analyzing and transforming our Dataset



The screenshot shows the Power Query Editor interface. The 'Data Type: Text' dropdown menu is open, and the 'Use First Row as Headers' option is highlighted with a red box. The formula bar shows the following M code:

```
= Table.TransformColumnTypes(#"Promoted Headers",{{"IATA_CODE", type text}, {"AIRLINE", type text}})
```

IATA_CODE	AIRLINE
1 UA	United Air Lines Inc.
2 AA	American Airlines Inc.
3 US	US Airways Inc.
4 F9	Frontier Airlines Inc.
5 B6	JetBlue Airways
6 OO	Skywest Airlines Inc.
7 AS	Alaska Airlines Inc.
8 NK	Spirit Air Lines
9 WN	Southwest Airlines Co.
10 DL	Delta Air Lines Inc.
11 EV	Atlantic Southeast Airlines
12 HA	Hawaiian Airlines Inc.
13 MQ	American Eagle Airlines Inc.
14 VX	Virgin America

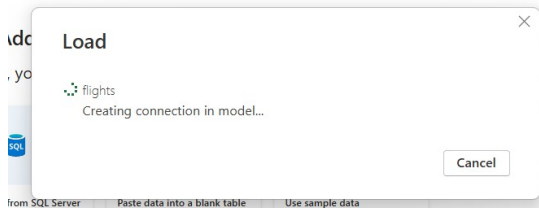


The screenshot shows the Power Query Editor interface. The 'Data Type: Text' dropdown menu is open, and the 'Use First Row as Headers' option is highlighted with a red box. The formula bar shows the following M code:

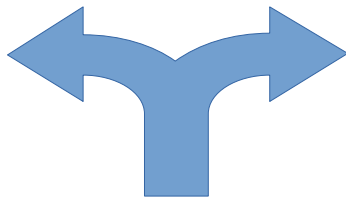
```
= Table.TransformColumnTypes(#"Promoted Headers",{{"CANCELLATION_REASON", type text}, {"CANCELLATION_DESCRIPTION", type text}})
```

CANCELLATION_REASON	CANCELLATION_DESCRIPTION
1 A	Airline/carrier
2 B	Weather
3 C	National Air System
4 D	Security

Reviewing and loading our Dataset



Column1	Column2
IATA_CODE	AIRLINE
UA	United Air Lines Inc.
AA	American Airlines Inc.
US	US Airways Inc.
F9	Frontier Airlines Inc.
B6	JetBlue Airways
OO	Skywest Airlines Inc.
AS	Alaska Airlines Inc.
NK	Spirit Air Lines
WN	Southwest Airlines Co.
DL	Delta Air Lines Inc.
EV	Atlantic Southeast Airlines
HA	Hawaiian Airlines Inc.
MQ	American Eagle Airlines Inc.
VX	Virgin America



After reviewing and confirming the validity of our dataset, we can load our data for further analysis

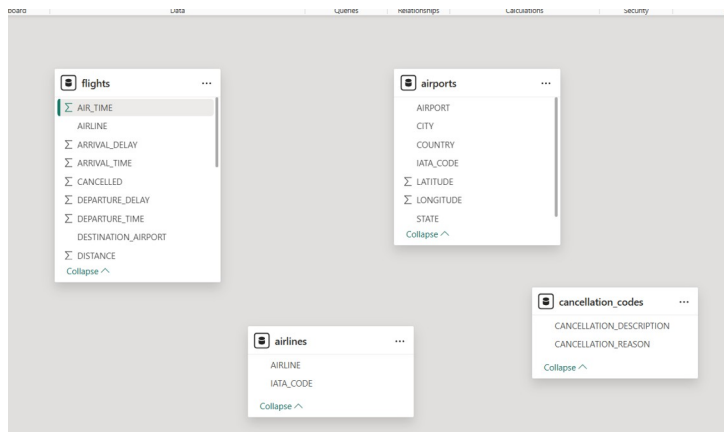
airports.csv

File Origin: 1252: Western European (Windows) | Delimiter: Comma | Data Type Detection: Based on first 200 rows

IATA_CODE	AIRPORT	CITY	STATE	COUNTRY	LATITUDE	LONGITUDE
ATL	Hartsfield-Jackson Atlanta International Airport	Atlanta	GA	USA	33.64044	-84.42694
DEN	Denver International Airport	Denver	CO	USA	39.85841	-104.667
DFW	Dallas/Fort Worth International Airport	Dallas-Fort Worth	TX	USA	32.89595	-97.0372
IAH	George Bush Intercontinental Airport	Houston	TX	USA	29.98047	-95.33972
LAS	McCarran International Airport	Las Vegas	NV	USA	36.08036	-115.15233
LAX	Los Angeles International Airport	Los Angeles	CA	USA	33.94254	-118.40807
MSP	Minneapolis-Saint Paul International Airport	Minneapolis	MN	USA	44.88055	-93.21692
ORD	Chicago O'Hare International Airport	Chicago	IL	USA	41.9796	-87.90446
PHX	Phoenix Sky Harbor International Airport	Phoenix	AZ	USA	33.43417	-112.00806
SFO	San Francisco International Airport	San Francisco	CA	USA	37.619	-122.37484

Extract Table Using Examples | Load | Transform Data | Cancel

Entity Relationship Mapping



Our dataset is structured around four primary entities:

- Airlines,
- Airports,
- Flights, and
- Cancellation Codes.

These entities are interconnected as follows:

- Flights serve as the central fact table, containing foreign keys referencing both Airlines and Airports.
- Each flight record links to a specific Airline (via airline code) and a Departure and Arrival Airport (via airport codes).
- Cancellation Codes are associated with canceled flights, providing categorical reasons for cancellation.
- Airlines and Airports function as dimension tables, offering descriptive metadata for reporting and analysis.

Entity Relationship Mapping...

New relationship

Select tables and columns that are related.

From table

cancellation_codes

CANCELLATI...	CANCELLATI...
Airline/Carrier	A
Weather	B
National Air S...	C

To table

flights

ARRIVAL_TIME	CANCELLATI...	CANCELLED	DEPARTURE_...	DEPARTURE_T...	DESTINATION...	DISTAN
1146			-3	1022	IND	432
1354			-3	1202	SLC	1590
1329			-3	1202	RSW	515

Cardinality

One to many (1:*)

Cross-filter direction

Single

☒ Make this relationship active

☐ Apply security filter in both directions

☐ Assume referential integrity

Save

Cancel

Edit relationship

Select tables and columns that are related.

From table

flights

AIR_SYSTEM_...	AIR_TIME	AIRLINE	AIRLINE_DELAY	ARRIVAL_DEL...	ARRIVAL_TIME	CANCELLA
null	67	DL	null	-20	1146	
null	215	DL	null	-31	1354	
null	70	DL	null	-18	1329	

To table

airlines

AIRLINE	IATA_CODE
United Air Lin...	UA
American Airli...	AA
US Airways Inc.	US

Cardinality

Many to one (*:1)

Cross-filter direction

Single

☒ Make this relationship active

☐ Apply security filter in both directions

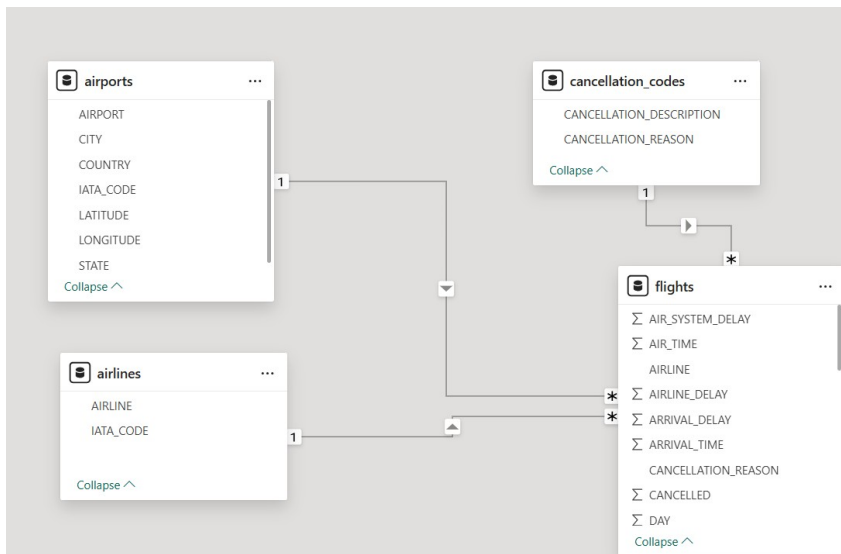
☐ Assume referential integrity

Save

Cancel

Mapping and Editing
Relationship across
different tables allows us
to map unique identifiers in
one table to its
corresponding foreign key
in another table

Entity Relationship mapping...



This relational structure enables efficient querying and supports multidimensional analysis across time, carriers, locations, and disruption causes.

Using Dax Statements...

```
1 % Canceled Flights = FORMAT(DIVIDE([Canceled Flight], [Total Flight], "-") * 100, "0.00") & "%"
1 % Delayed Flights = FORMAT(DIVIDE([Delayed Flight], [Total Flight], "-") * 100, "0.00") & "%"
1 % On-Time Flights = FORMAT( DIVIDE([On-Time Flight], [Total Flight], "-") * 100, "0.00") & "%"
1 Canceled Flight = CALCULATE([Total Flight], flights[Status] = "Canceled")
```



DAX (Data Analysis Expressions) is used for calculations. It helps us to leverage the use of expressions to simulate or utilize our entity relationship in defining measures and calculated columns.

Flight Performance Report

Total Flight

Total Flight and % On-Time Flights by MONTH

200K

150K

0

MONTH

Total Flight
✈️
2M

Canceled Flight

Total Flight and % Canceled Flights by MONTH

200K

150K

0

MONTH

Canceled Flight
✈️✕
29K

Delayed Flight

Total Flight and % Delayed Flights by MONTH

200K

150K

0

10

Delayed Flight
✈️🕒
790K

Canceled Flight by CITY

CITY

Chicago
Dallas-Fort Wo...
Atlanta
Los Angeles

0K

5K

10K

Total Flight by AIRLINE

AIRLINE

Delta Air Lines...
American Airli...
Southwest Air...
United Air Line...

0.0M

0.2M

0.4M

Total Flight

Total Flight by Status

Status ● Canceled ● Delayed ● On-Time



0% 50% 100%

Total Flight

✈️✕ % Canceled Flights

1.47%

✈️🕒 % Delayed Flights

40.54%

On-Time Flight by CITY



Conclusion...

From our dashboard, we can infer the following insights:

Total Flight
✈️ 2M

We had about two million flights scheduled across all airports surveyed for the year 2015.

Canceled Flight
✈️ 29K

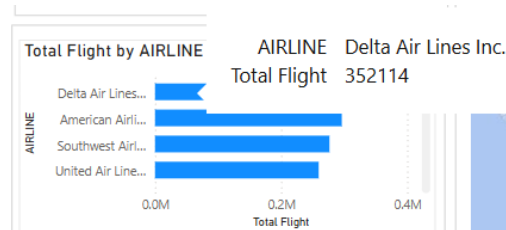
We had about twenty nine thousand of those flights making 1.47% of the total were cancelled across all airports in 2015.

Delayed Flight
✈️ 790K

We had over seven hundred and ninety thousand flights scheduled which were delayed making 40.54% of the total.



Chicago airport recorded the highest number of cancelled flight by city with 8,548



Delta Airline has the highest record of flights at over three hundred and fifty thousand for the year – 2015

Conclusion...

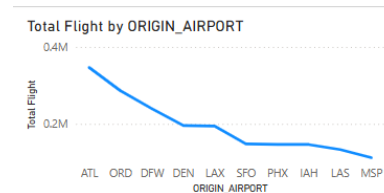
From our dashboard, we can infer the following: -

% Delayed Flights
40.54%

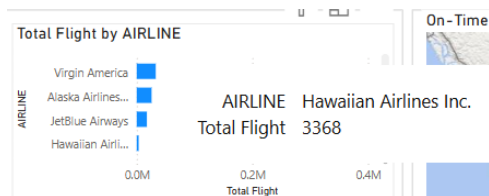
Out of the total flights scheduled, we had one million and a hundred flights were on time making 50.99%

On-Time Flight
1M

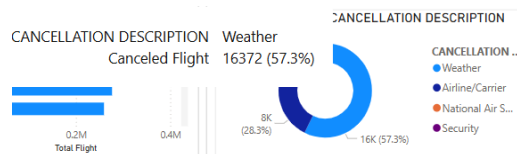
Out of the total flights scheduled, we had one million and a hundred flights were on time making 50.99% of the total number of flights across all airports



We can see a downward trend in total flights across Origin Airport



Hawaiian Airlines had the least number of flights with a total number of flights at 3,368



Cancellation due to weather reasons constitutes about 57.3% of total Cancellation