

Lessons Learned From My Power BI Journey

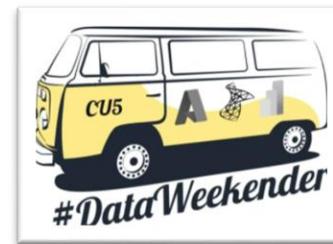
Pragati Jain



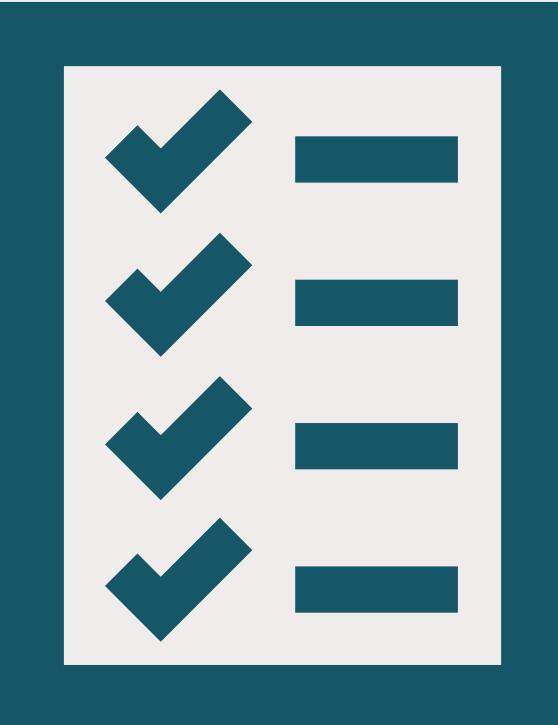
Pragati Jain

- Microsoft Data Platform MVP
- Superuser at Microsoft Fabric Community
- 10+ years of industry experience
- Analytics Manager at Avanade, London
- Organizer at DataWeekender conference
- Manager at Microsoft Fabric UK User Group

Fun Fact About Me:
I am a painter and proudly call myself
'TechieArtist' 😊



AGENDA



Power Query



Data Modelling



Visualization

POWER QUERY EDITOR

*BRING
ONLY
WHAT
YOU NEED*

The image displays three separate tables stacked vertically, representing different data sources:

- Top Table:** Shows a list of countries with their ISO codes, names, regions, and administrative regions. It includes columns for `ABC_Iso2Code`, `ABC_Country`, `ABC_Region`, `ABC_Region.Attribute:id`, `ABC_Region.Attribute:Iso2code`, and `ABC_adminregion.Attribute:id`.
- Middle Table:** Shows income levels with corresponding codes and lending types. It includes columns for `ABC_adminregion.Attribute:Iso2code`, `ABC_incomeLevel.Element:Text`, `ABC_incomeLevel.Attribute:id`, `ABC_incomeLevel.Attribute:Iso2code`, `ABC_lendingType.Element:Text`, and `ABC_lendingType.Attribute:id`.
- Bottom Table:** Shows capital cities with their coordinates and links to their flags. It includes columns for `ABC_capitalCity.Element:Text`, `1.2_longitude.Element:Text`, `1.2_latitude.Element:Text`, `ABC_Attribute:id`, and `ABC_Custom`.

A large dark blue downward-pointing arrow is positioned between the top two tables and above the bottom table.

The bottom table shows data for Afghanistan across various years, specifically focusing on population values. It includes columns for `ABC_iso_code`, `1^23 year`, `ABC_Attribute`, `1.2_Value`, `ABC_Country`, `ABC_CO2 Emission Type`, and `ABC_Region`.

	ABC iso_code	1^23 year	ABC Attribute	1.2_Value	ABC Country	ABC CO2 Emission Type	ABC Region	
1	AFG		1924	population	9204707	Afghanistan	Other	South Asia
2	AFG		1925	population	8721710	Afghanistan	Other	South Asia
3	AFG		1926	population	8264058	Afghanistan	Other	South Asia
4	AFG		1927	population	7830419	Afghanistan	Other	South Asia
5	AFG		1928	population	7419535	Afghanistan	Other	South Asia
6	AFG		1929	population	7099916	Afghanistan	Other	South Asia
7	AFG		1930	population	6866773	Afghanistan	Other	South Asia
8	AFG		1931	population	6715569	Afghanistan	Other	South Asia

Transform Upstream As Possible

”

Data should be transformed as far upstream as possible, and as far downstream as necessary.

- Roche's Maxim of Transformation – ssbipolar.com

Should I add Transformation in DAX or Power Query?

=> Power Query

Should I add Transformation in Power BI or Data Warehouse?

=> Data Warehouse

Making sure to choose the right Datatypes

"For example: Fixed decimal number datatype can cause performance issues."

Fixed decimal number datatype stores value up to 4 decimal places. It's better to set precision in Power Query editor for columns like rounding them up to just 2 decimal places!

Executing Vertipaq Analyzer in Tabular Editor, I can see size of the column = 8,259,736

Emission Values (billion tons) is Fixed decimal number

Name	Cardinality	Table Size	Col Size	Data Size	Dict Size	Hier Size	Encoding	Data Type
Country	180	30,882	30,882	752	27,218	2,912		
Region	7	18,206	18,206	400	17,710	96		
Emission Type	9	19,356	19,356	400	18,796	160		
Year	99	4,036	4,036	352	2,884	800		
Emission Sub-Type	50	20,752	20,752	480	19,648	624		
Emissions	561,818	9,032,260	9,031,916	1,882,456	5,741,396	1,408,064		
RowNumber-2662979B-1795-4F...	0	264	128	136	0 VALUE	Int64		
year	98	503,072	499,528	2,744	800 HASH	Int64		
Country Key	179	163,220	156,664	5,116	1,440 HASH	Int64		
Region Key	7	2,628	1,208	1,356	64 HASH	Int64		
Emission Type Key	9	1,732	288	1,364	80 HASH	Int64		
Emission Sub-Type Key	50	101,264	99,320	1,528	416 HASH	Int64		
Emission Value (billion tons)	175,657	8,259,736	1,125,320	5,729,152	1,405,264 HASH	Decimal		
world population	26,569	3,353,371	3,353,371	541,960	2,012,859	798,552		

Emission Value (billion tons)

- Copy
- Remove
- Remove Other Columns
- Duplicate Column
- Add Column From Examples...
- Remove Duplicates
- Remove Errors
- Change Type
- Transform
- Replace Values...
- Replace Errors...
- Group By...

Properties

Name: Emissions
All Properties

Applied Steps

- Source
- Reordered Columns
- Renamed Column
- Merged Queries1

Round

Specify how many decimal places to round to.

Decimal Places: 2

OK Cancel

VertiPaq Analyzer

Name	Cardinality	Table Size	Col Size	Data Size	Dict Size	Hier Size	Encoding	Data Type
Country	180	30,882	30,882	752	27,218	2,912		
Region	7	18,206	18,206	400	17,710	96		
Emission Type	9	19,356	19,356	400	18,796	160		
Year	99	4,036	4,036	352	2,884	800		
Emission Sub-Type	50	20,752	20,752	480	19,648	624		
Emissions	561,818	5,213,472	5,213,128	1,737,192	2,866,192	609,744		
RowNumber-2662979B-1795-4F74...	0		264	128	136	0 VALUE	Int64	
year	98		503,072	499,528	2,744	800 HASH	Int64	
Country Key	179		208,596	202,040	5,116	1,440 HASH	Int64	
Region Key	7		2,596	1,176	1,356	64 HASH	Int64	
Emission Type Key	9		1,732	288	1,364	80 HASH	Int64	
Emission Sub-Type Key	50		142,008	140,064	1,528	416 HASH	Int64	
Emission Value (billion tons)	75,867		4,354,860	893,968	2,853,948	606,944 HASH	Decimal	
world population	26,569		3,353,371	3,353,371	541,960	2,012,859	798,552	

Refreshing model in Vertipaq Analyzer, I can see that the updated size of the column is reduced to = 4,354,860

Emission Values (billion tons) is rounded to 2 decimal places

Column size is reduced by ~ 53%

Use Views Over Tables

View is basically the result of a query over a database

View doesn't store any physical data – key difference between table & view

Views reduce complexity over querying the data – for e.g. joins over 3 tables

Views contain a sub-set of data – can be referred to filtered data over a table with relevant query. For e.g., sales for EMEA region only

As views don't store physical data, they take less space on your database - (take up very little memory for query definition)

Views can also help limit access to users to underlying data

Usage of Views can help with performance of reports – end users query only relevant data via views rather than underlying big tables

Don't Ignore Query Folding



"Query Folding is the ability for a Power Query query to generate a single statement to retrieve and transform source data."



When Power Query generates this single statement, it means the query folds!

Query Folding is supported by => databases that support querying language like MySQL, SQL Server Database, etc.

Query Folding is not supported by => flat files, excel/csv files, etc.

Transformations That Support Query Folding

Removing columns

Renaming columns

Filtering rows

Grouping & Summarizing

Expanding record columns

Appending foldable queries (same source)

Merging foldable queries (same source)

Custom columns with simple logic

Transformations That Don't Support Query Folding

Appending queries (different sources)

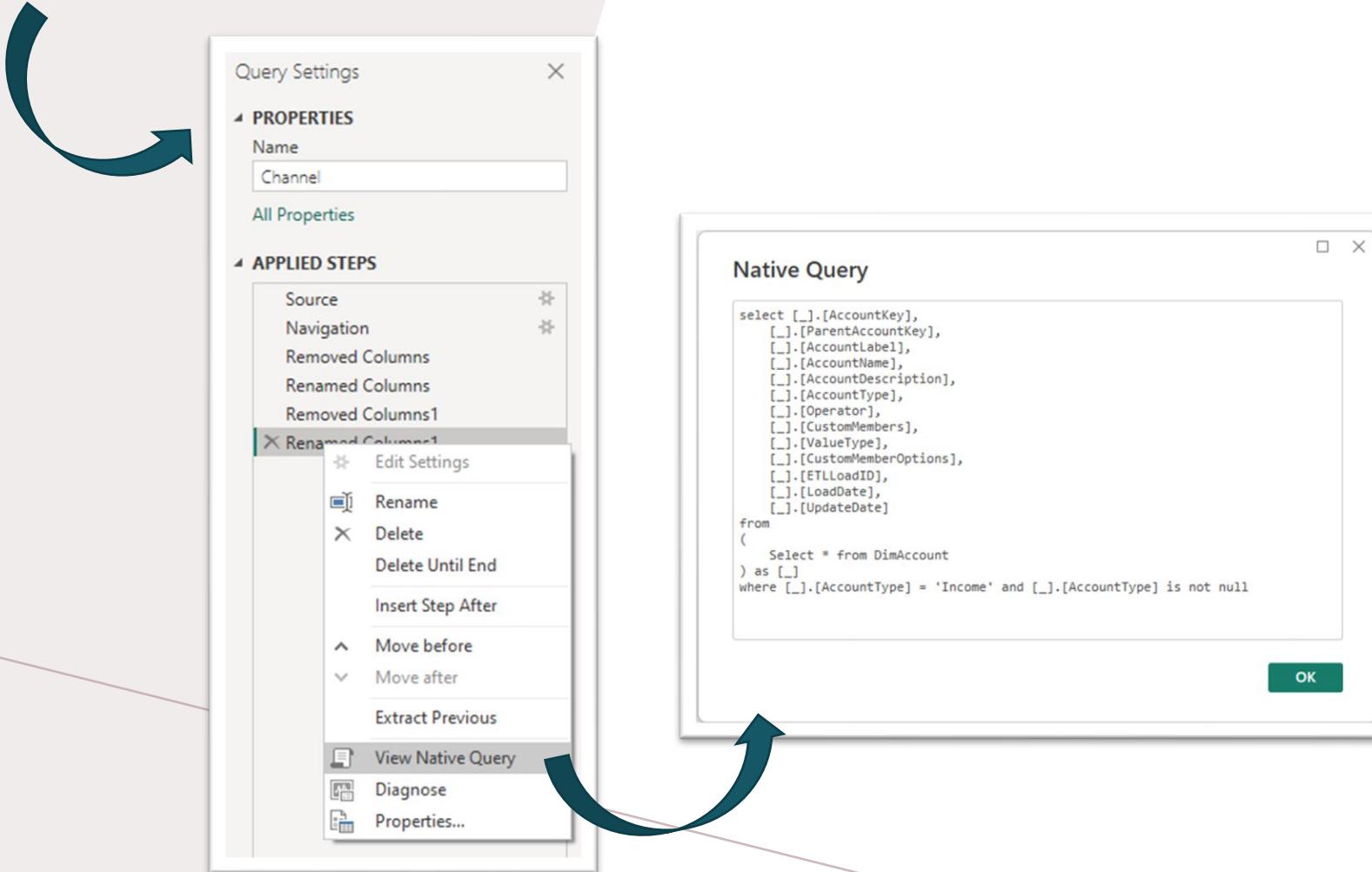
Merging queries (different sources)

Custom columns with complex logic

Adding Index columns

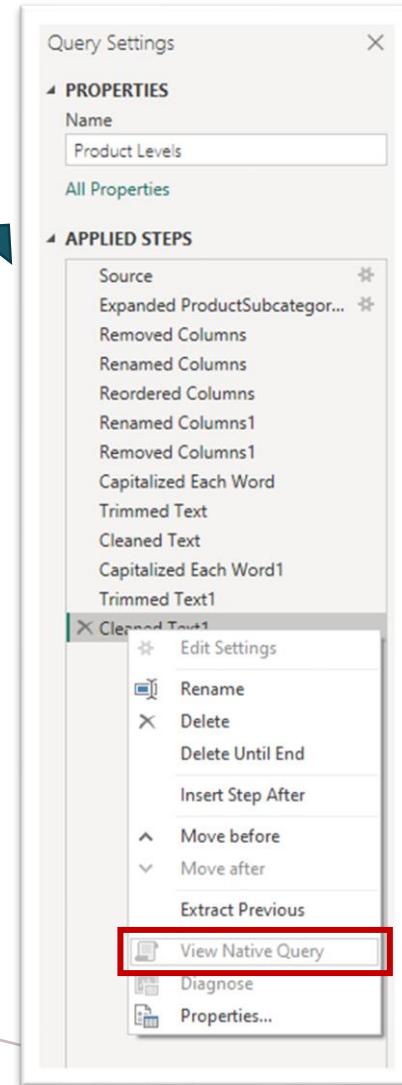
Is Query Folding Happening?

Checking if Query Folding is Happening



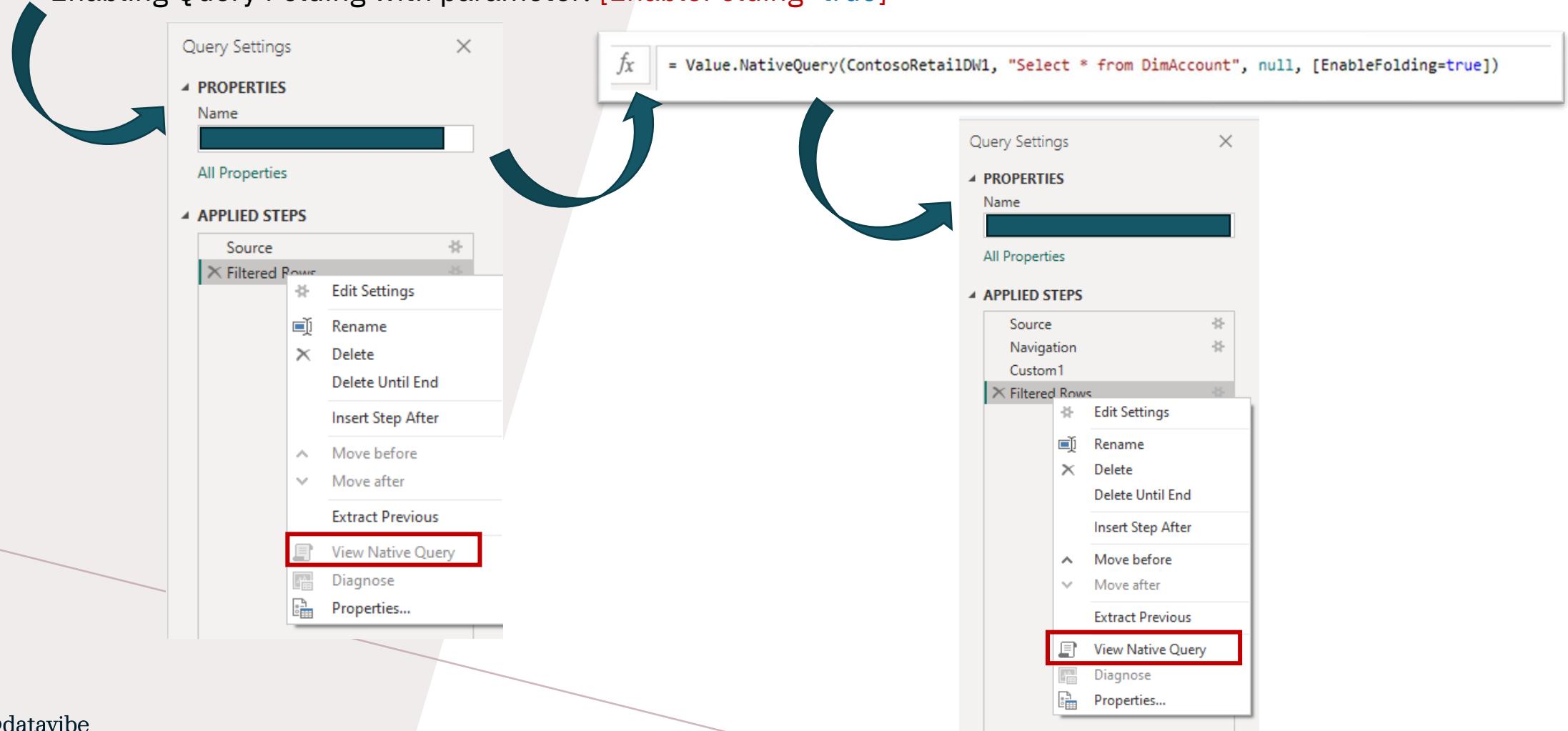
If Query Folding Is NOT Happening

Checking if Query Folding is NOT Happening



Enabling Query Folding

Enabling Query Folding with parameter: [EnableFolding=true]



Advantages Of Query Folding

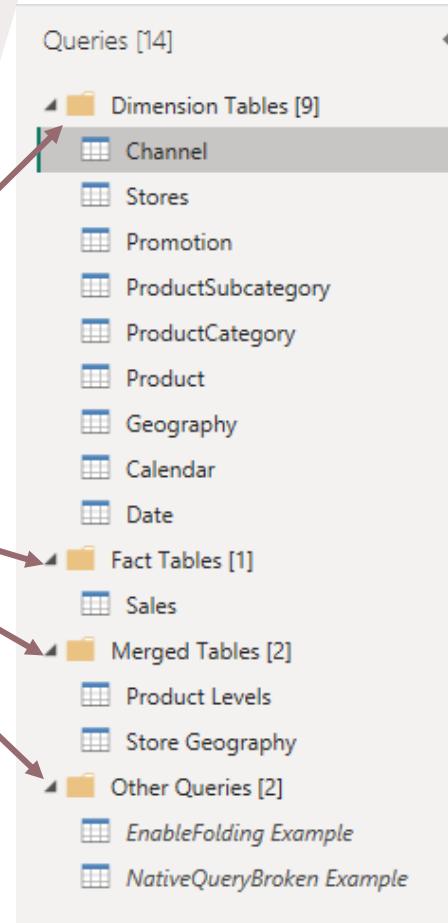
Efficient data
refresh for import
model tables

Efficient resource
utilization by
Power BI engine

Avoids
performance
issues

Organize Folders In Power Query Editor

Tables/Views are properly organized –
Dimension, Fact, Merged, etc.



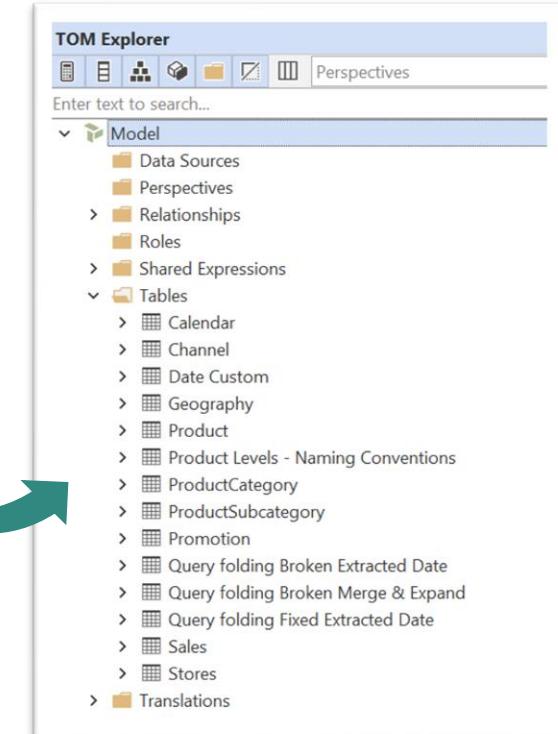
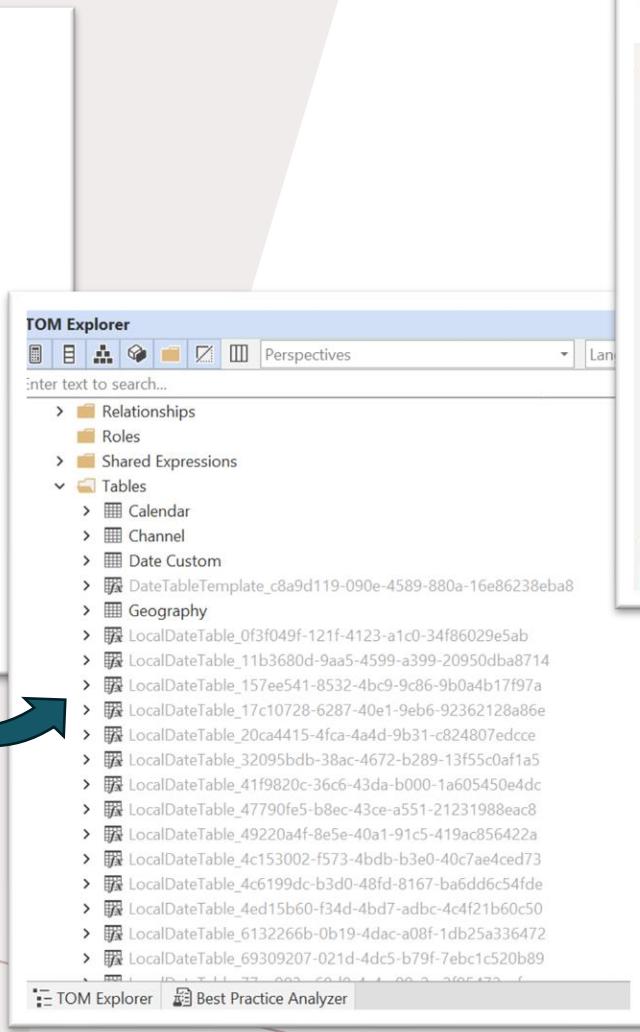
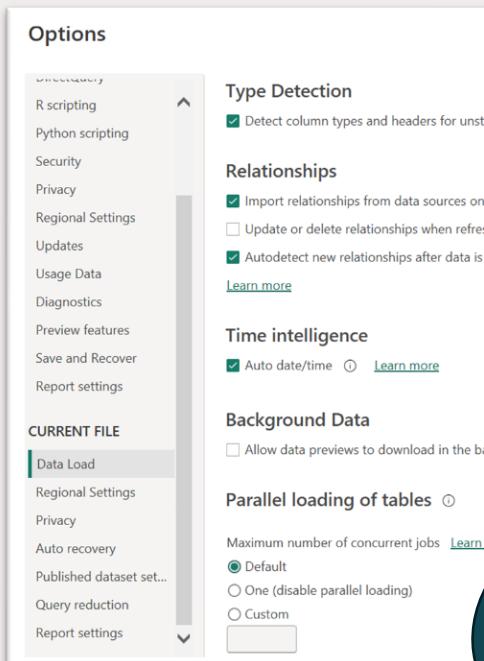
POWER QUERY

Things To Remember (In Addition)

Rename steps (easy identification)	Advantage: Avoid the repetition that occurs with auto naming How: Select step to rename: right click and rename, double click step or press F2
Avoid spaces, replace by Upper letters or _	Advantage: Facilitates InteliSense and/or Advance Editor coding Example: Format_Delivery_Date or FormatDeliveryDate
Look for repetition of steps	Advantages: Avoid duplication of names Reduce number of steps
Define datatype for all columns	Advantage: Avoid issues when loading the query How: View column header for icon ABC/123 43 or Home tab – Data Type equals Any
Datatypes transformation in the last step and only once	Turn off the Type detection on the Options of Power BI desktop Detect datatype is standard based on the top 1000 rows
Organize Queries by using folders in Power Query editor	Advantage: Clarity when working with Power BI file later and easy for other developers to understand How: Select multiple queries => right-click => create folders

DATA MODELLING

Turn Off Auto Datetime & Time Intelligence



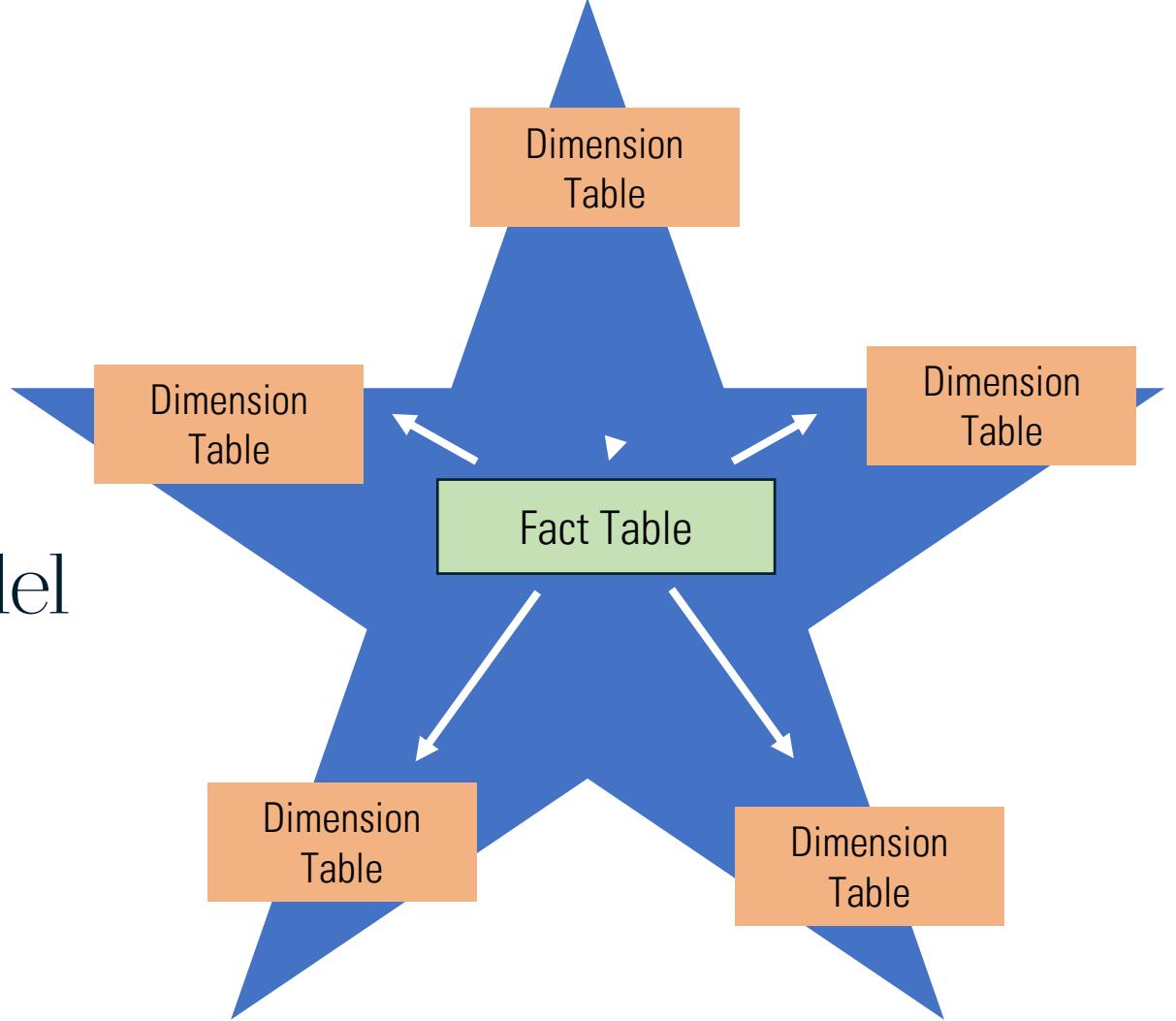
The image shows three screenshots illustrating how to turn off the 'Auto date/time' feature in Power BI.

Screenshot 1 (Left): Shows the 'Options' dialog with the 'Data Load' section selected. The 'Auto date/time' checkbox is checked (selected). A large blue circular arrow points from this screen towards the center screen.

Screenshot 2 (Center): Shows the same 'Options' dialog, but the 'Data Load' section is deselected. The 'Auto date/time' checkbox is now unchecked (not selected). A large blue circular arrow points from this screen towards the right screen.

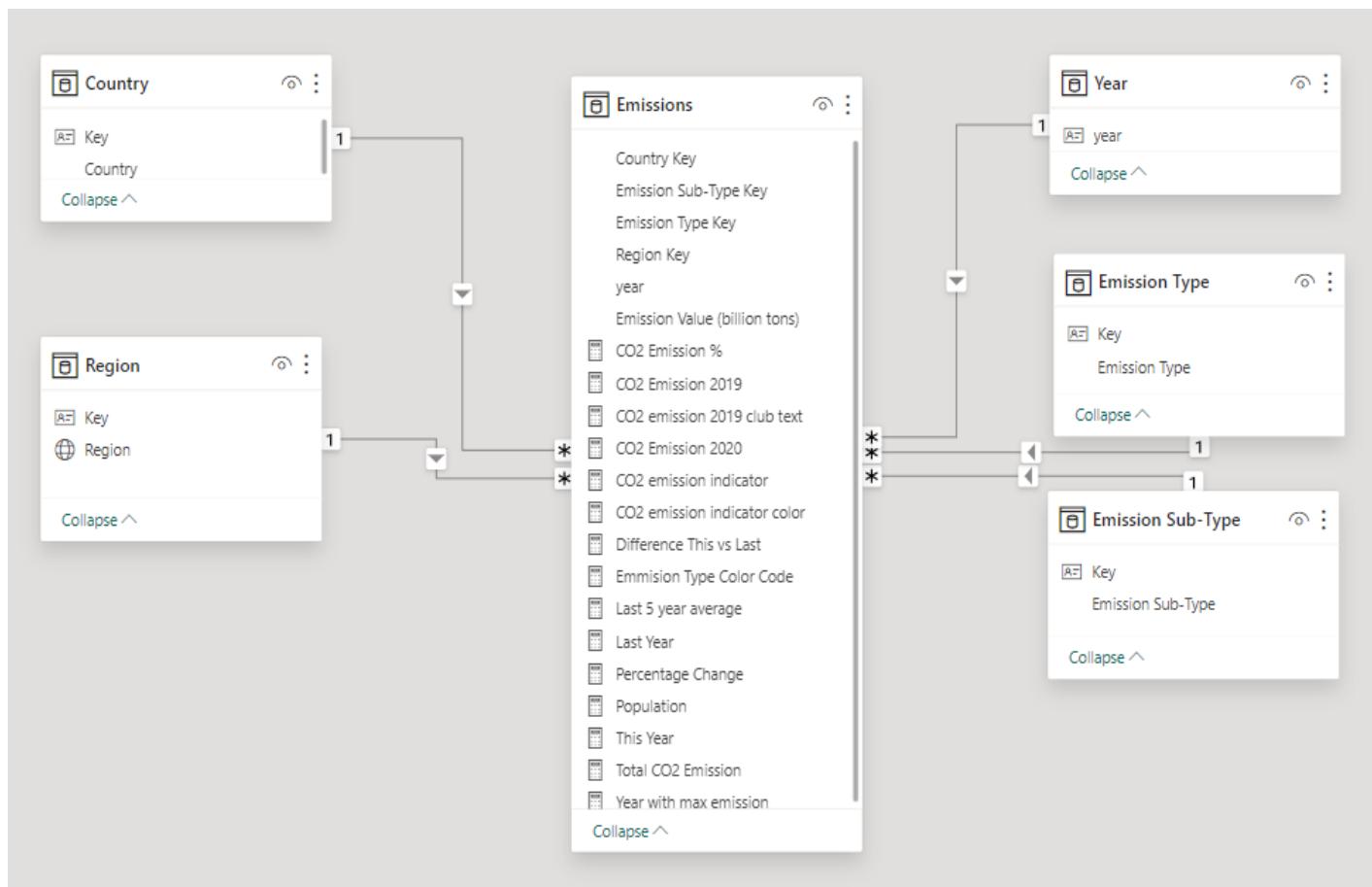
Screenshot 3 (Right): Shows the 'TOM Explorer' interface with the 'Model' node expanded. It lists various data models and tables, including Calendar, Channel, Date Custom, Geography, Product, and Sales.

Go For Star Schema Data Model

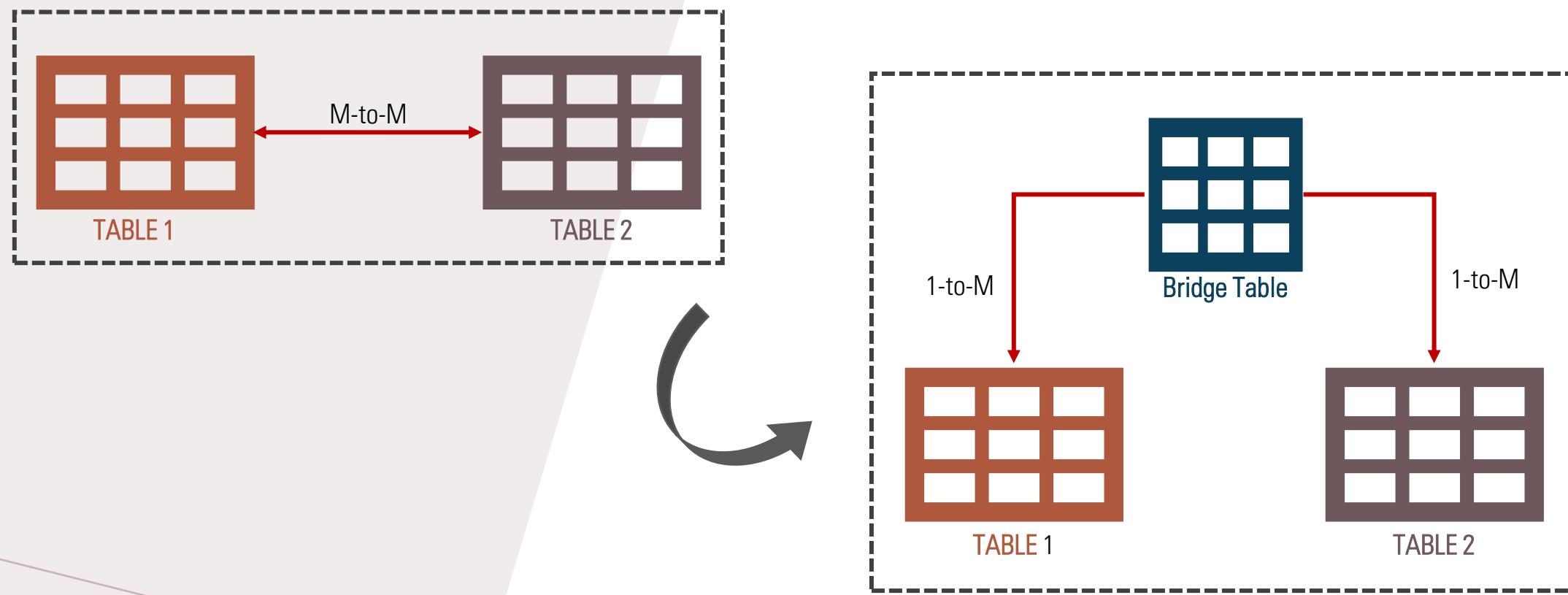


Star Schema Approach: Benefits

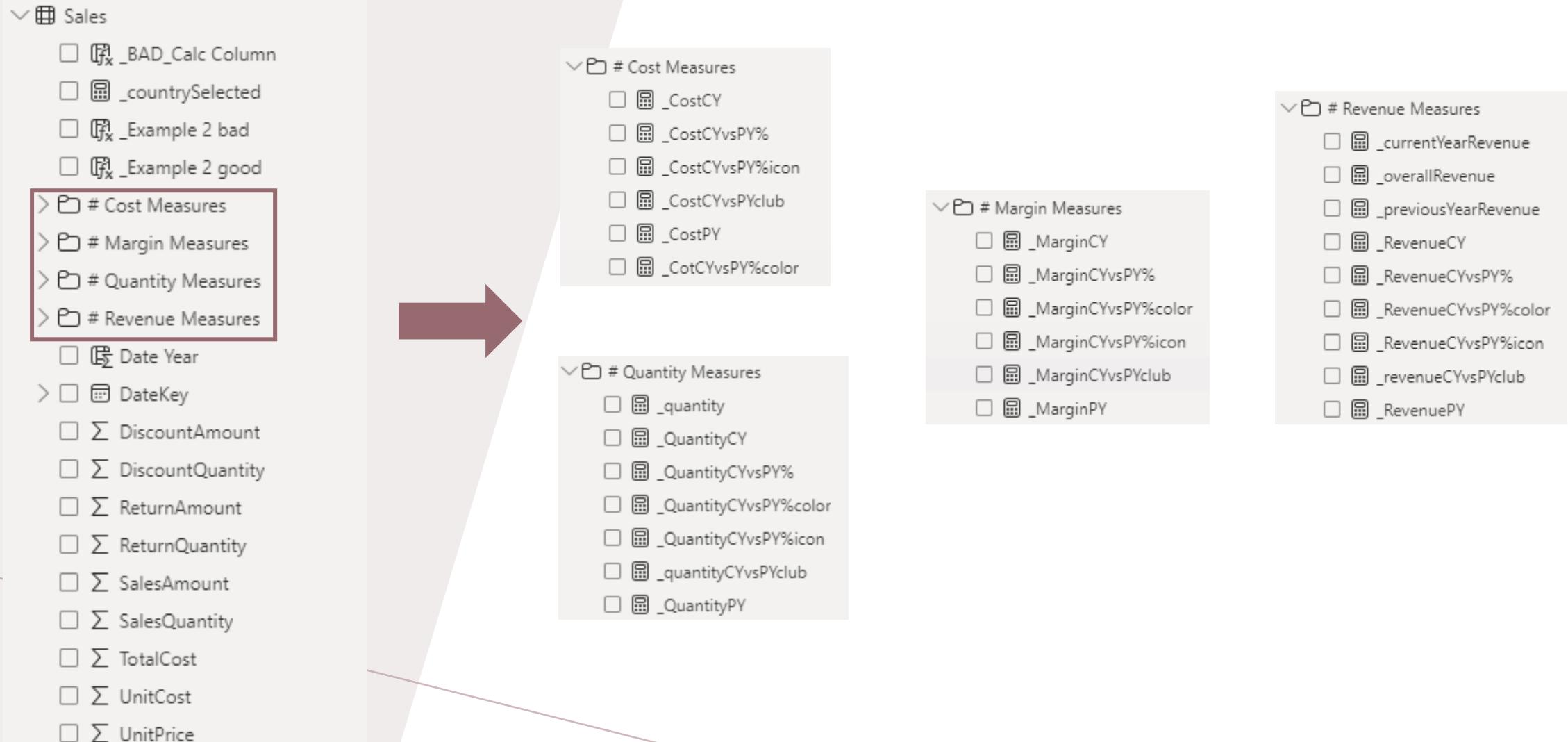
- Data is more organized
- Helps with slicing of data properly
- A good data model avoids complex DAX
- Star schema data models are designed to handle large volume of data efficiently
- Improves performance and data refreshes



Think Before Creating Many-to-many Relationships



Display Folders For Measures



Data Modelling Best Practices



BRING ONLY WHAT YOU
NEED – THE RIGHT COLUMNS



BE CAUTIOUS ABOUT THE
FORMAT OF THE COLUMNS –
DATE, NUMBER, ETC.



SEE IF YOU CAN FOLLOW THE
STAR SCHEMA APPROACH



TURN-OFF AUTO-DATE TIME



MAKE SURE TO CHECK YOUR
RELATIONSHIPS (AUTO-
DETECT !!!)



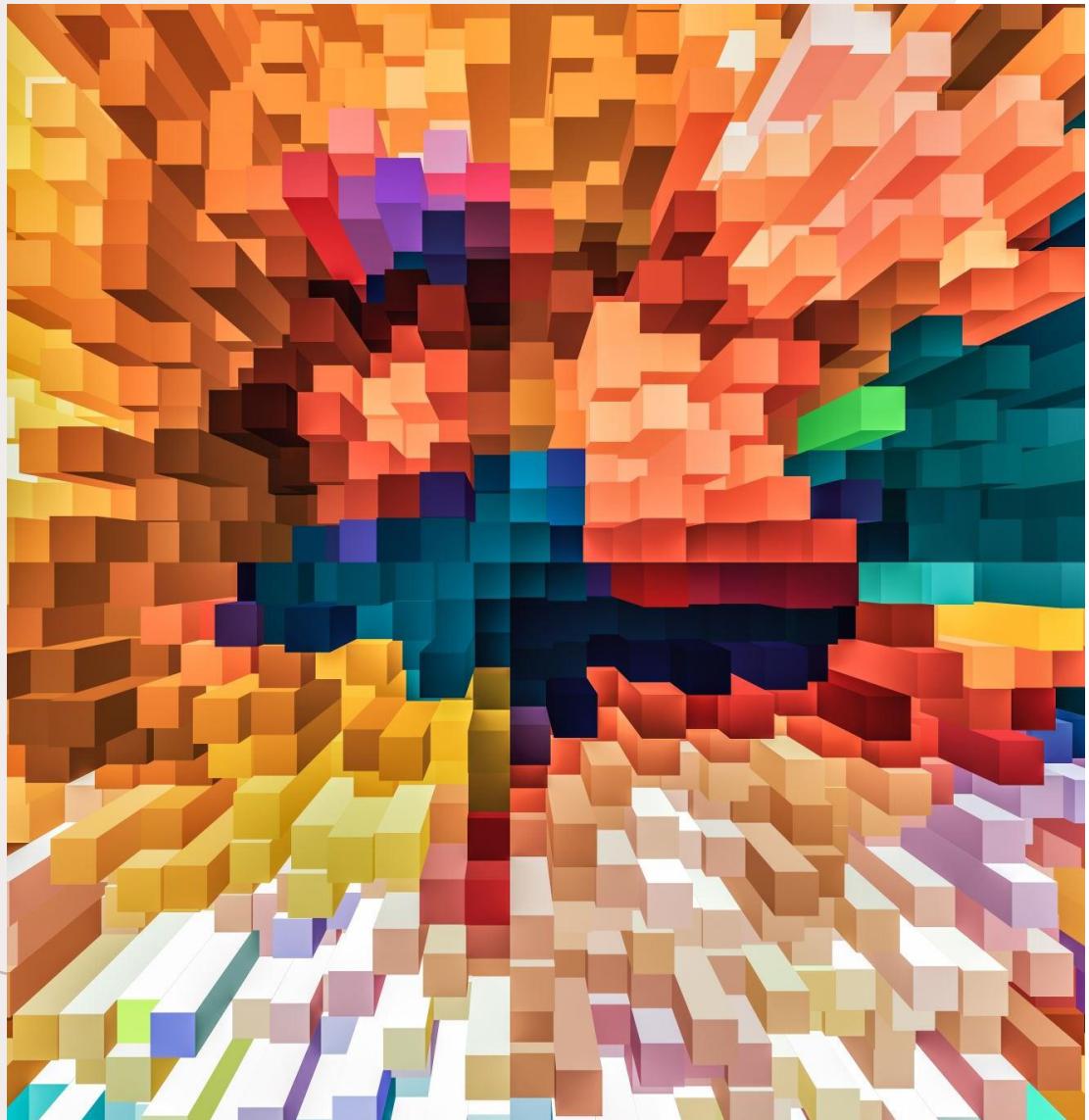
AVOID MANY-MANY
RELATIONSHIP. USE BRIDGE
TABLES!



CLEAN YOUR DATA ON THE
DATA-SOURCE SIDE. LET'S
TRY TO AVOID BURDEN ON
POWER BI SIDE

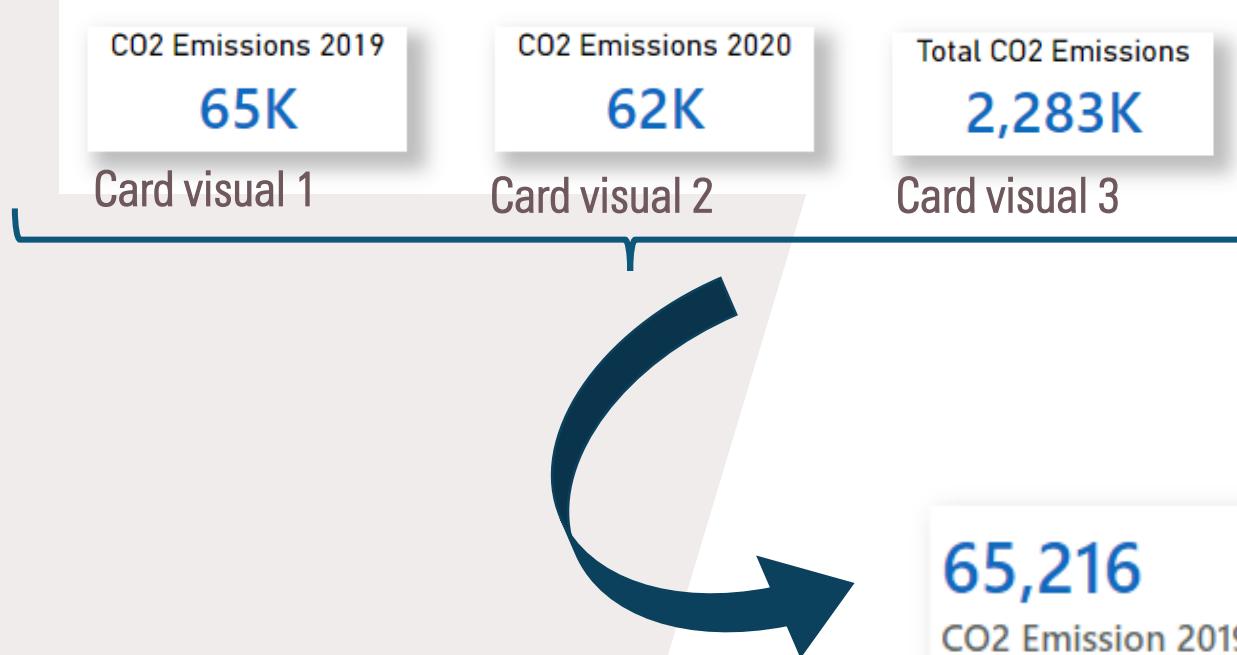


THINK ABOUT CALCULATED
COLUMNS, USE MEASURES
WHEREVER POSSIBLE



VISUALIZATION

Performance Impact by Number of Visuals



Name	Duration (ms)
⌚ Recording started (08/01/2024 10:26:23)	-
⌚ Refreshed visual	-
+ CO2 Emissions 2019	71
+ CO2 Emissions 2020	71
+ Total CO2 Emissions	71
+ Multi-row card	72

SINGLE Multi-Card visual

Map Visual Performance

population by city



© 2023 Microsoft Corporation. All rights reserved. Microsoft, the Microsoft logo, Microsoft Bing, and Microsoft OpenDAX are either registered trademarks or trademarks of Microsoft Corporation in the United States and/or other countries.

population by lat and lng



© 2023 Microsoft Corporation. All rights reserved. Microsoft, the Microsoft logo, Microsoft Bing, and Microsoft OpenDAX are either registered trademarks or trademarks of Microsoft Corporation in the United States and/or other countries.

Performance analyzer

Start recording Refresh visuals Stop

Filters

Clear Export

Name	Duration (ms)
Recording started (31/05/2023 21:45:12)	-
Refreshed visual	-
population by city	2976
DAX query	14
Visual display	260
Other	2702
population by lat and lng	978
DAX query	15
Visual display	162
Other	801

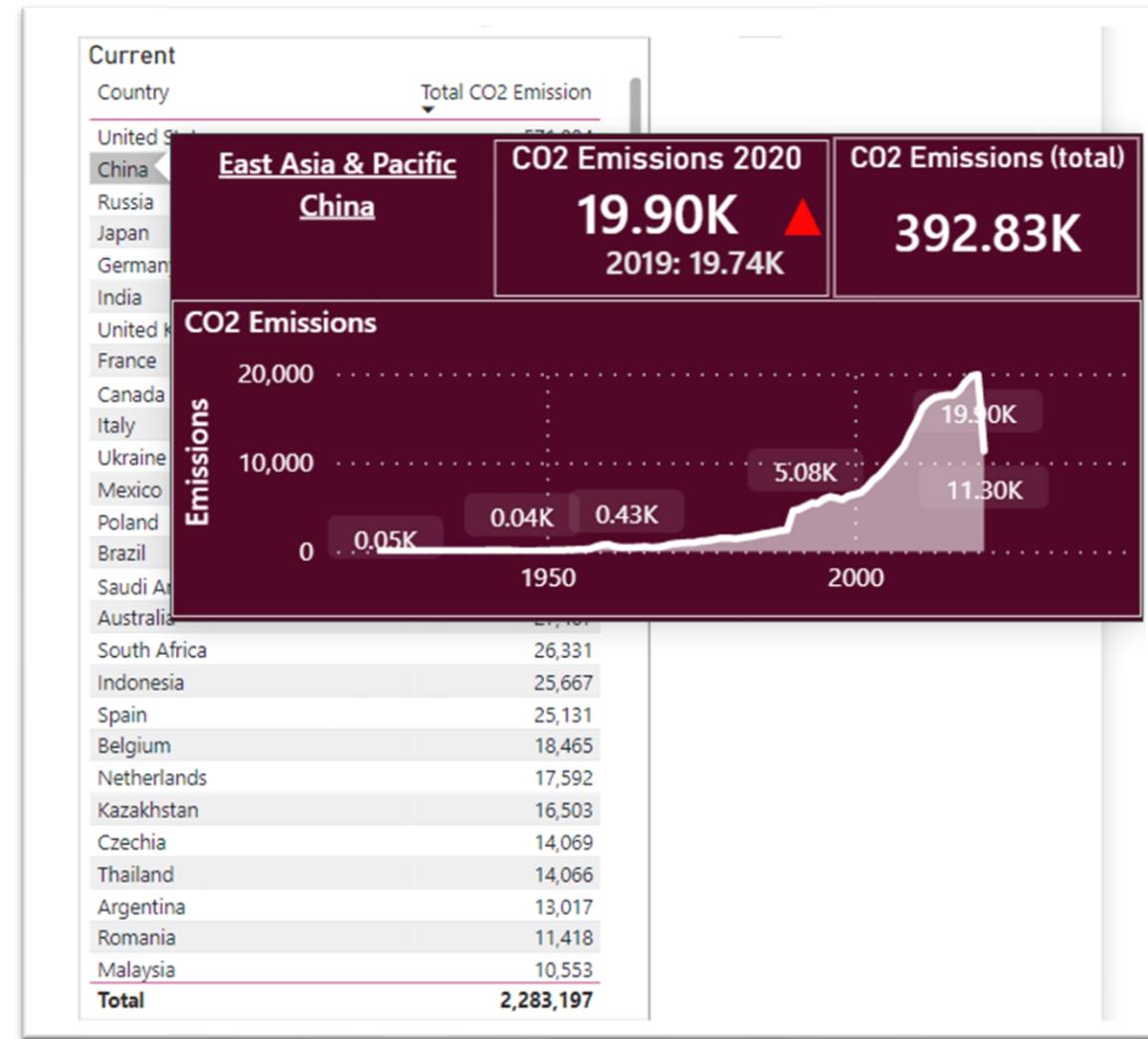
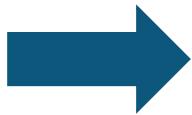
Copy query

Learn more about optimizing your report's performance on [the Microsoft Power BI documentation](#).

Use Custom Tooltips For Additional Insights

Previous

Region	Country	Emission Type	Total CO2 Emission	CO2 Emission 2019	CO2 Emission 2020
East Asia & Pacific	China	Coal	186,610	7,543	7,606
North America	United States	Consumption	180,879	5,692	5,197
East Asia & Pacific	China	Consumption	172,903	9,959	10,033
North America	United States	Oil	158,437	2,359	2,049
North America	United States	Coal	140,665	1,071	879
North America	United States	Gas	77,815	1,674	1,654
Europe & Central Asia	Germany	Coal	48,773	240	197
East Asia & Pacific	Japan	Consumption	44,622	1,250	1,187
Europe & Central Asia	Russia	Coal	44,466	390	367
South Asia	India	Consumption	40,985	2,468	2,277
Europe & Central Asia	Russia	Consumption	40,021	1,380	1,360
South Asia	India	Coal	36,380	1,678	1,588
Europe & Central Asia	Russia	Gas	35,054	804	781
East Asia & Pacific	China	Oil	34,492	1,560	1,628
East Asia & Pacific	Japan	Oil	33,120	416	388
Europe & Central Asia	Russia	Oil	32,761	398	381
Europe & Central Asia	Germany	Consumption	31,752	843	769
Europe & Central Asia	United Kingdom	Coal	30,231	25	23
East Asia & Pacific	Japan	Coal	22,965	432	399
Europe & Central Asia	United Kingdom	Consumption	20,439	514	465
Europe & Central Asia	Poland	Coal	20,068	178	164
Europe & Central Asia	Germany	Oil	19,759	271	251
Sub-Saharan Africa	South Africa	Coal	18,145	394	373
North America	Canada	Consumption	16,961	527	491
Europe & Central Asia	Italy	Consumption	16,833	434	389
Total			2,283,197	65,216	62,296

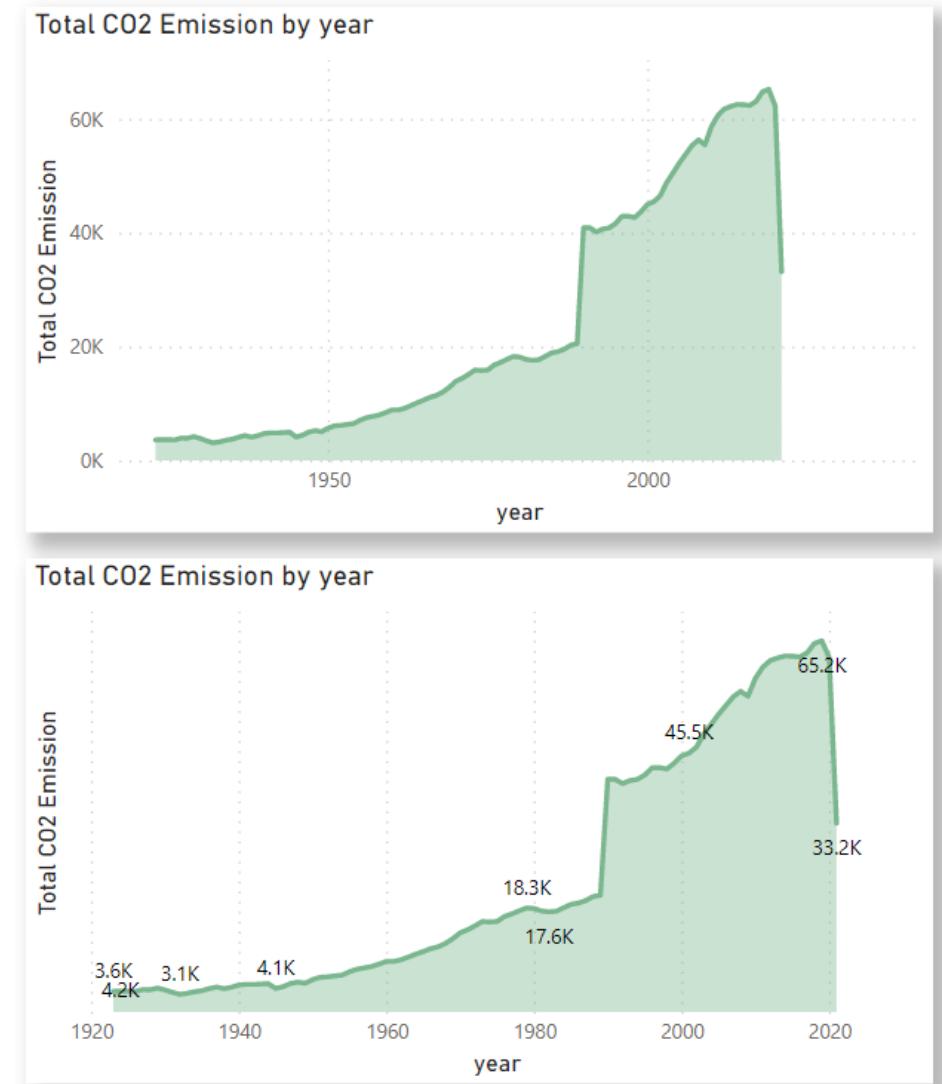


Axis Vs Labels On A Visual

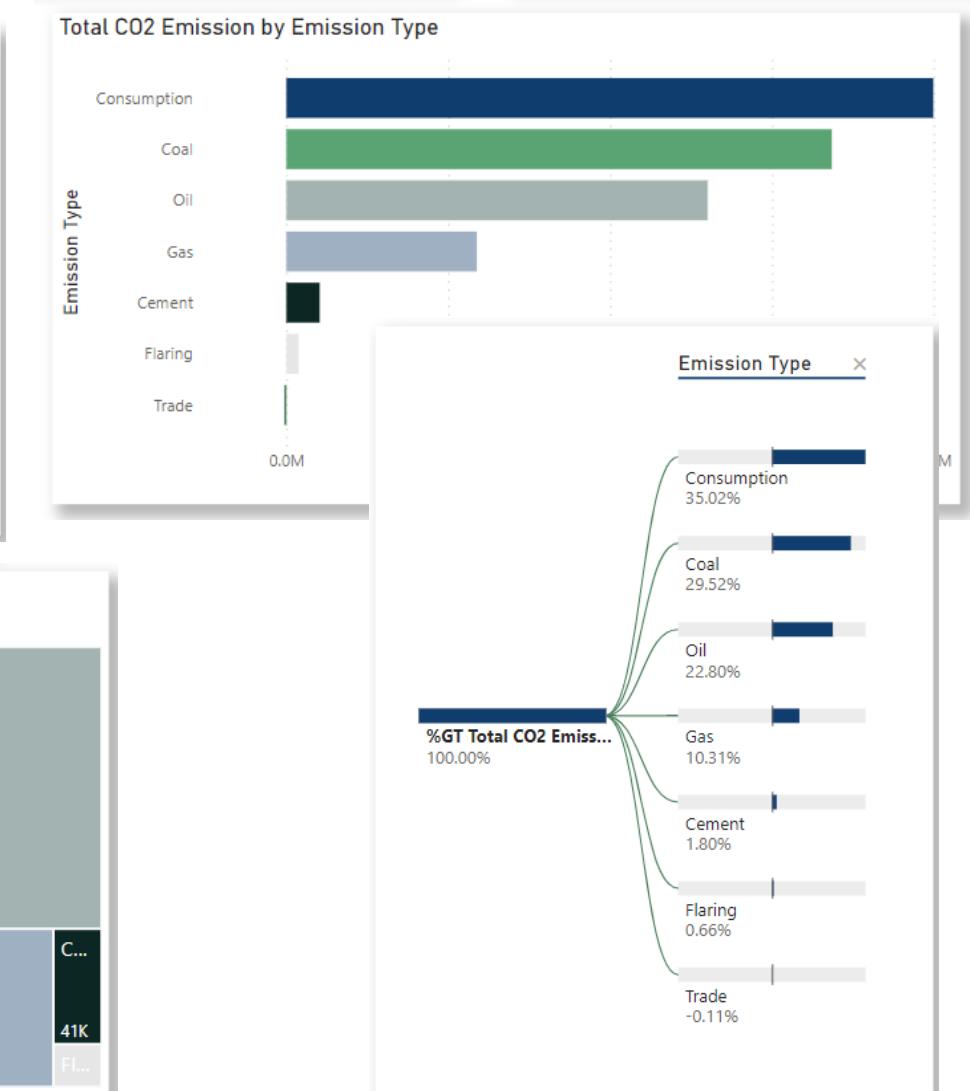
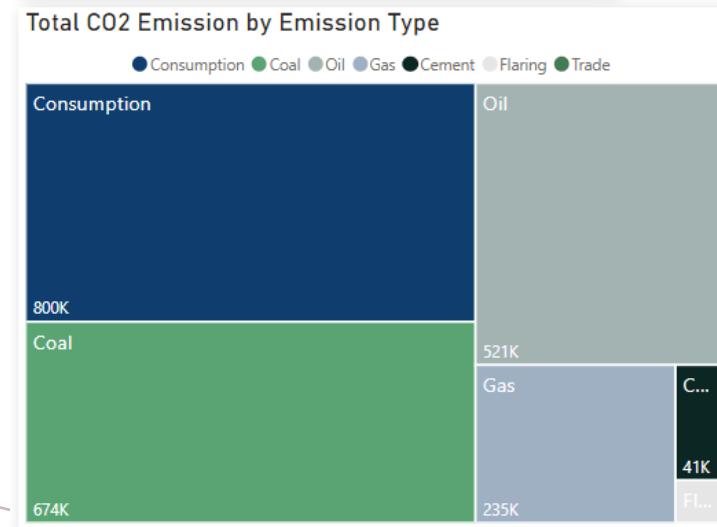
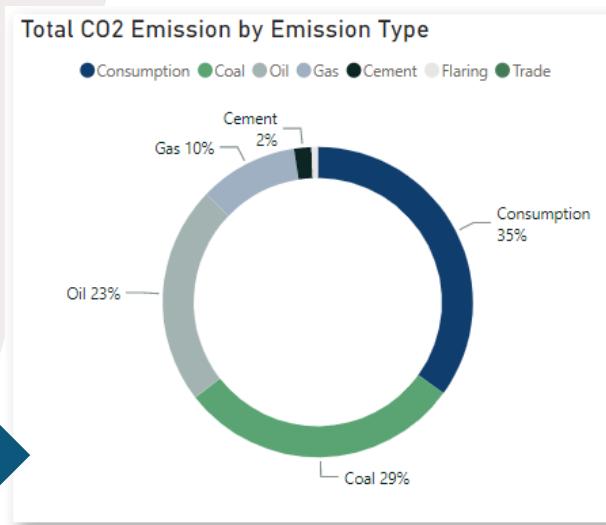
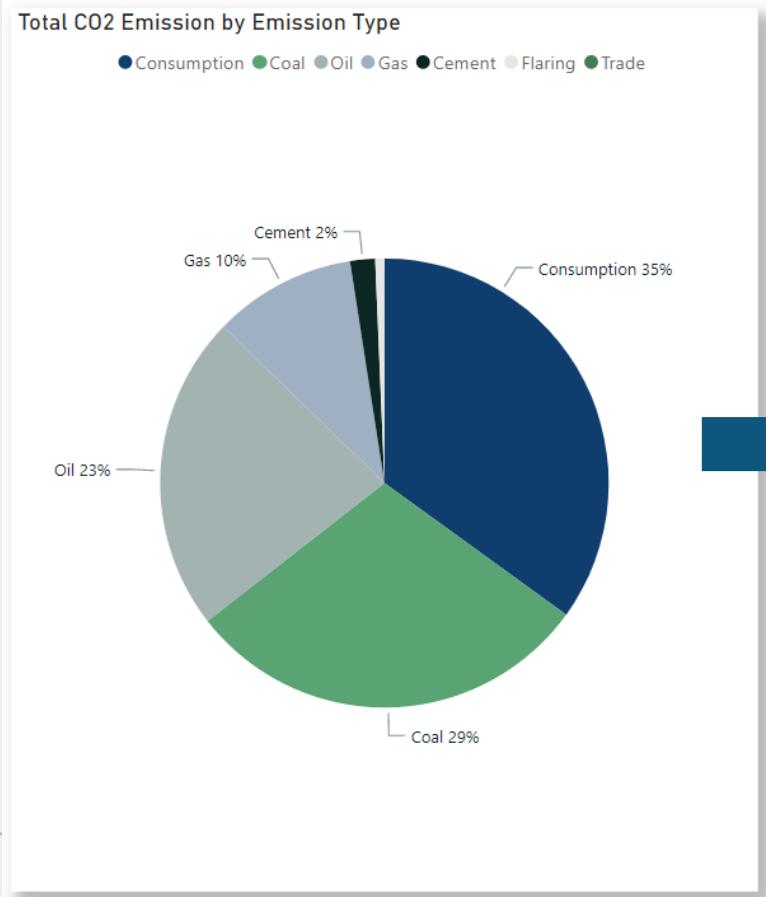
Axis AND Data Labels



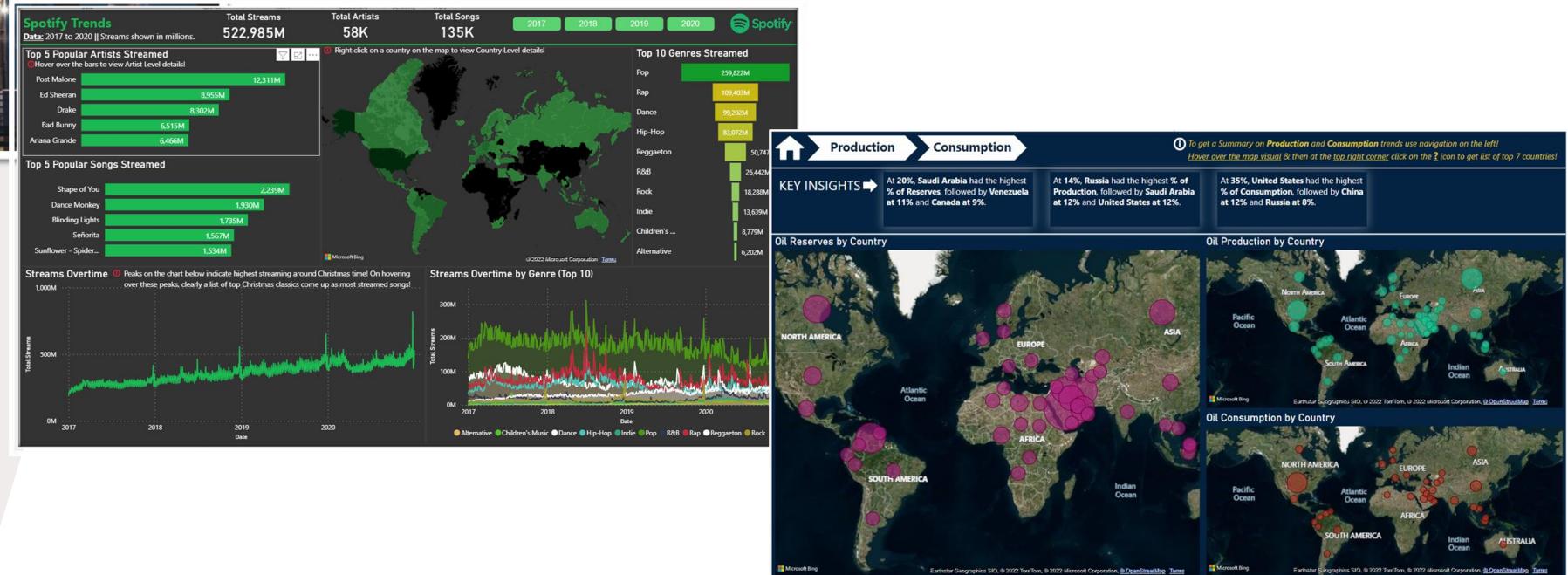
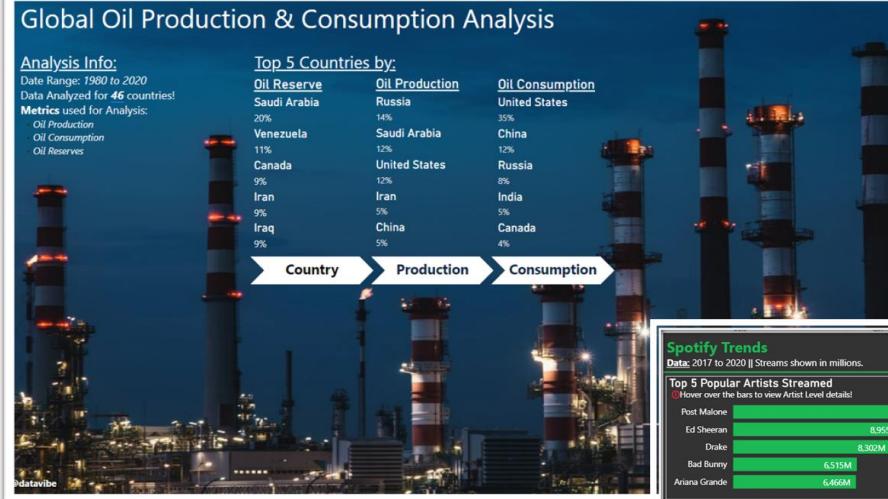
Axis OR Data Labels



Think Before Choosing Pie-charts!



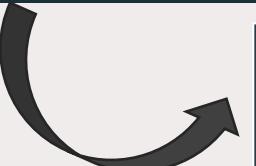
Adding Summary, Key Insights & Commentary Helps



This will help users to understand the right flow of the data insights generated.

Try to focus on standard colors to show the impact.
Use icons/indicators where they fit.

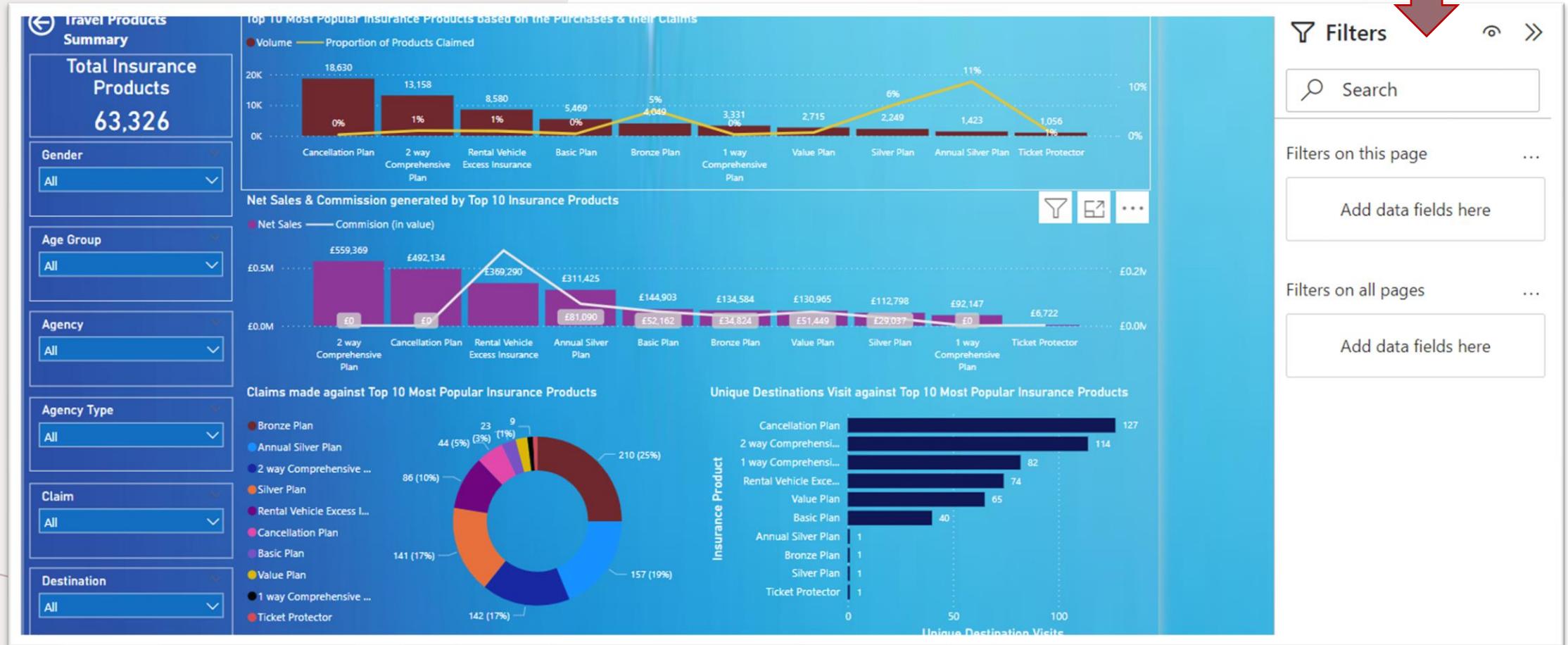
Evolution of CO2 Emissions - Last 10 Yrs									
Name	2010	2011	2012	2013	2014	2015	2016	2017	2018
Russia									
Germany									
United Kingdom									
Italy									
France									
Turkey									
Poland									
Spain									
Ukraine									
Kazakhstan									
Netherlands									
Belgium									



Evolution of CO2 Emissions - Last 10 Yrs									
Name	2010	2011	2012	2013	2014	2015	2016	2017	2018
Russia									
Germany									
United Kingdom									
Italy									
France									
Turkey									
Poland									
Spain									
Ukraine									
Kazakhstan									
Netherlands									
Belgium									

Payment	Product line	Total price
Ewallet	Home and lifestyle	£21,290
Cash	Electronic accessories	£20,730
Credit card	Food and beverages	£20,235
Ewallet	Fashion accessories	£19,357
Cash	Food and beverages	£19,211
Cash	Sports and travel	£18,874
Cash	Home and lifestyle	£18,589
	Sports and travel	£18,433
	Electronic accessories	£18,180
	Sports and travel	£17,816
	Fashion accessories	£17,614
	Fashion accessories	£17,335
	Health and beauty	£17,189
	Food and beverages	£16,699
	Health and beauty	£16,035
	Health and beauty	£15,969
	Electronic accessories	£15,428
	Home and lifestyle	£13,983
		£322,967
Total		7.03

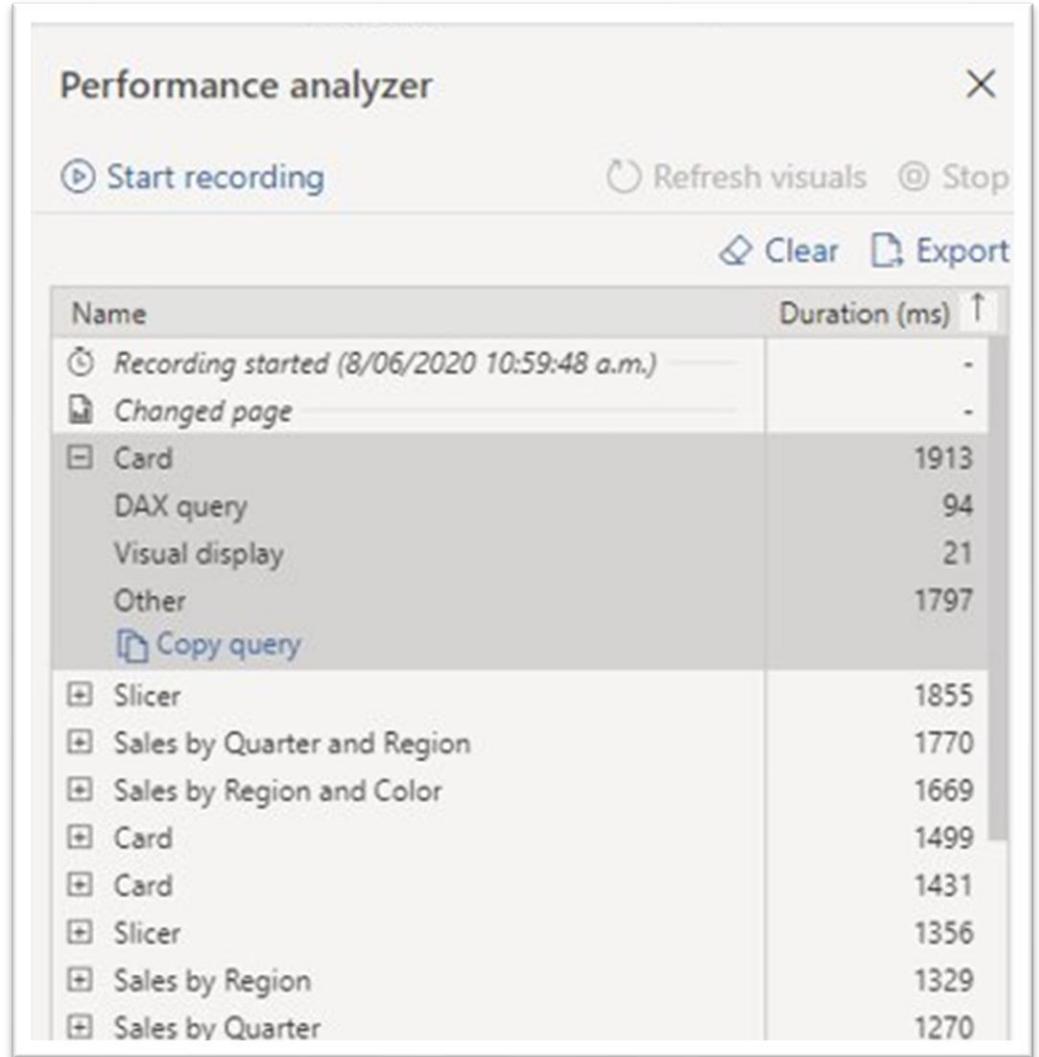
Make use of Filter Pane to add multiple filters rather than multiple slicer visuals.



Don't Forget Performance Analyzer!

- Find out how each report element is performing.
- Measure performance of report elements during user interaction.
- Detect which aspects are least or most resource intensive.

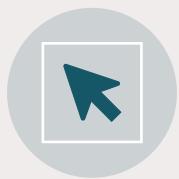
"Review Performance Results & identify the bottlenecks behind the report performance!"



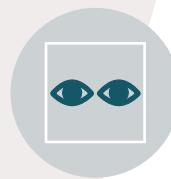
The screenshot shows the 'Performance analyzer' window in Power BI. It has a toolbar with 'Start recording', 'Refresh visuals', 'Stop', 'Clear', and 'Export'. The main area is a table with columns 'Name' and 'Duration (ms)'. The table lists various operations and their execution times:

Name	Duration (ms)
⌚ Recording started (8/06/2020 10:59:48 a.m.)	-
📄 Changed page	-
🕒 Card	1913
DAX query	94
Visual display	21
Other	1797
📄 Copy query	
🕒 Slicer	1855
🕒 Sales by Quarter and Region	1770
🕒 Sales by Region and Color	1669
🕒 Card	1499
🕒 Card	1431
🕒 Slicer	1356
🕒 Sales by Region	1329
🕒 Sales by Quarter	1270

Visualization Best Practices!



Less is More! Think about the number of visuals on the page/tab



Keep it simple – use the right visuals to represent the right information



Make sure to use the right themes/colours – standard colours make a very big difference



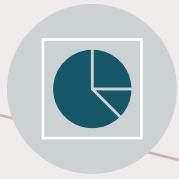
Look for Accessibility friendly colours



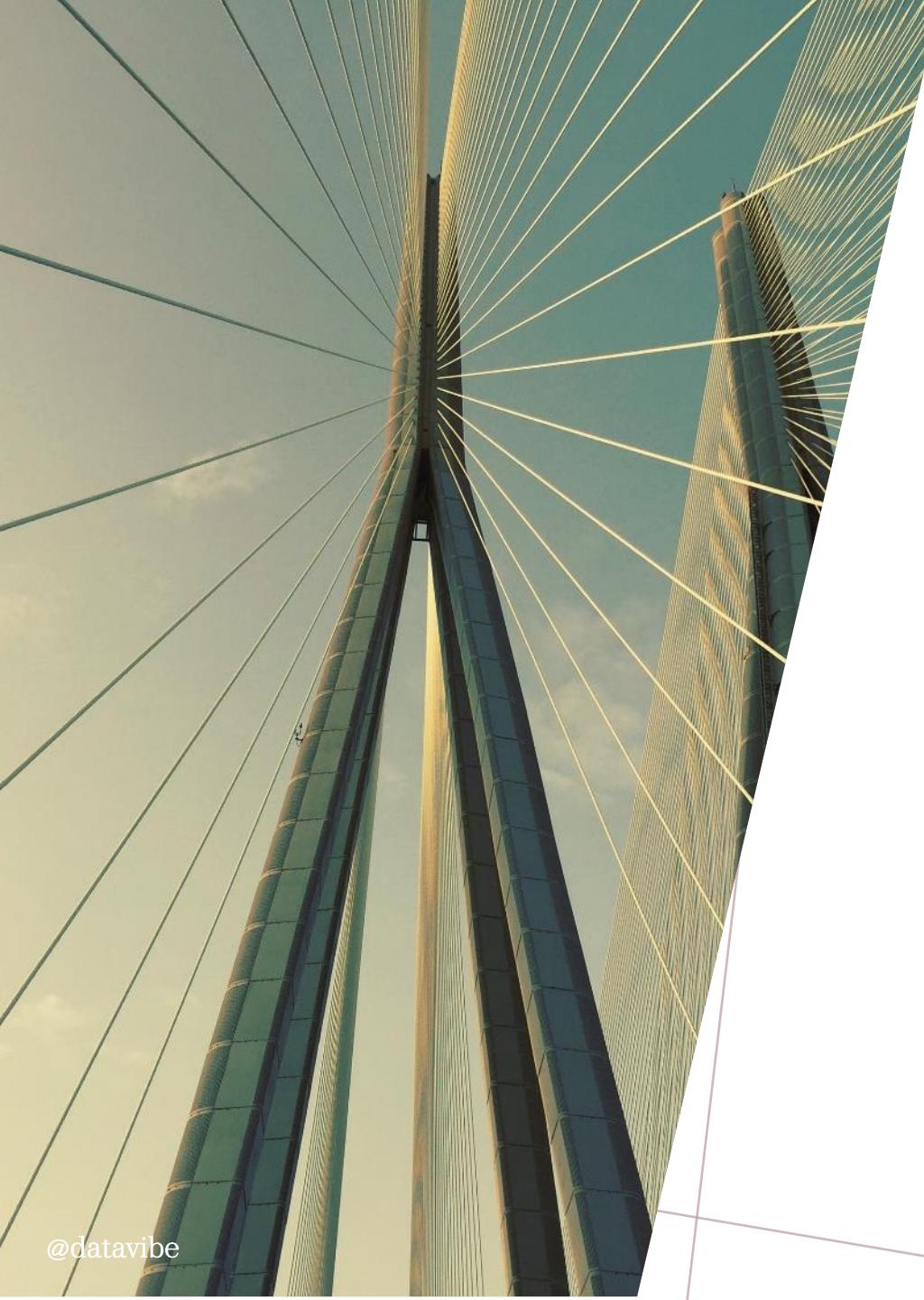
Try to tell a story with the visuals



Use conditional formatting, icons, where relevant



Think before adding Pie-charts/Donut-charts to your report



Thank You
Happy To Answer
Any Questions!

