Exercise 4

** Click on following link to get all the datasets required for this exercises

https://github.com/PurpleGrad/2203/blob/main/Dataset%20for%20Exercise%204.zip

Exercise on Data Source :

1)Create a new Power BI file, and from the above folder load the Excel workbook containing a list of the **Disney Princesses.**

Hide a couple of columns that we're not interested in, so that the field well looks like this:

Create a table showing the number of films by feminist rating:

Disney Princess	
Feminist rating	**Count of Name**
Drippy	4
Feisty	2
No idea	4
Role model	3
Total	**13**

You'll need to change the summarisation option to do this.

Save this Power BI file with the name **Disney Girls**, then exit this instance of Power BI Desktop.

2)Import the **Events, Category and Country** tables from the **WorldEvents.xlsx** file, found in the above folder. Look in **Relationship View** to see what has happened.

Annoyingly, no relationships have been created between the tables. There are two ways we can handle this.

To create a join. In this case ***CategoryID to*** *the* **EventCategoryID** ***EventCountryID*** *and* ***CountryID*** *pair up!*

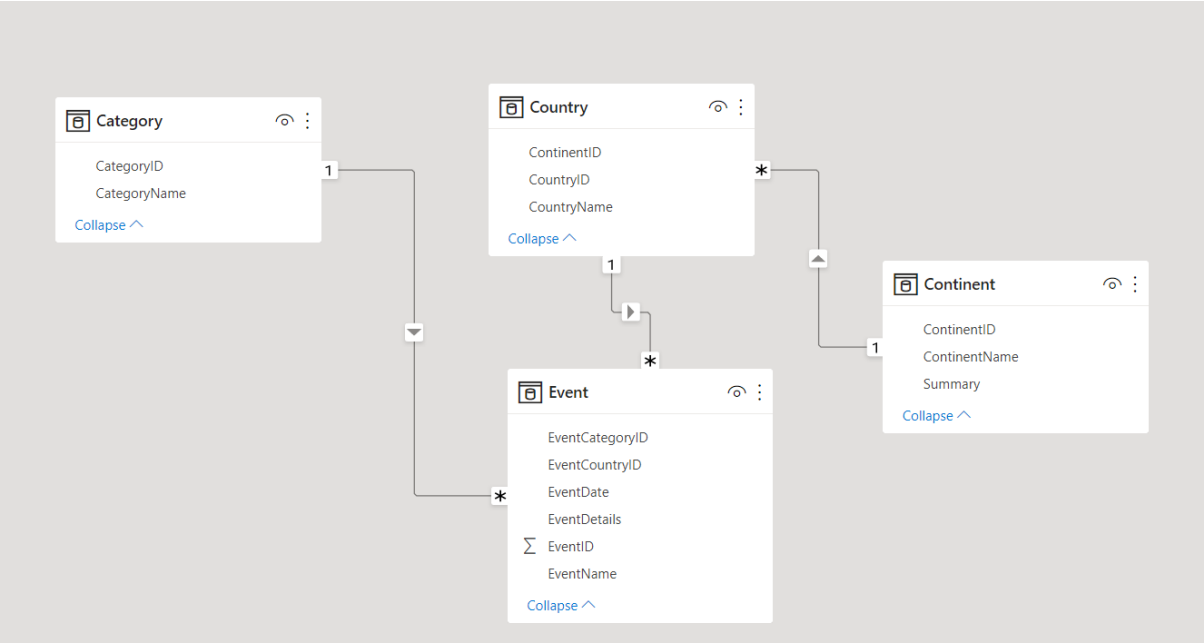
Thankfully most of the time the names of columns make it clear what the overlap is. Modern databases have a tendency to even use the exact same column name.

The danger of this method is if you miss. To check which columns have been joined let your mouse hover over the line you have just created. Power BI will highlight the two joined columns:

*The **1** means that **CategoryID** only has unique values while the * means **EventCategoryID** has duplicates (ie more than one event can be in the same category).*

How daft, there was another table we wanted from that excel document!

Bring in the **Continent** table - this should automatically create a join too:



A relationship has been auto created since they have the same field names, data types and one is unique. Excellent!

Optionally save this as **Creating connections.pbix** and close it down.

Exercise on Query Editor

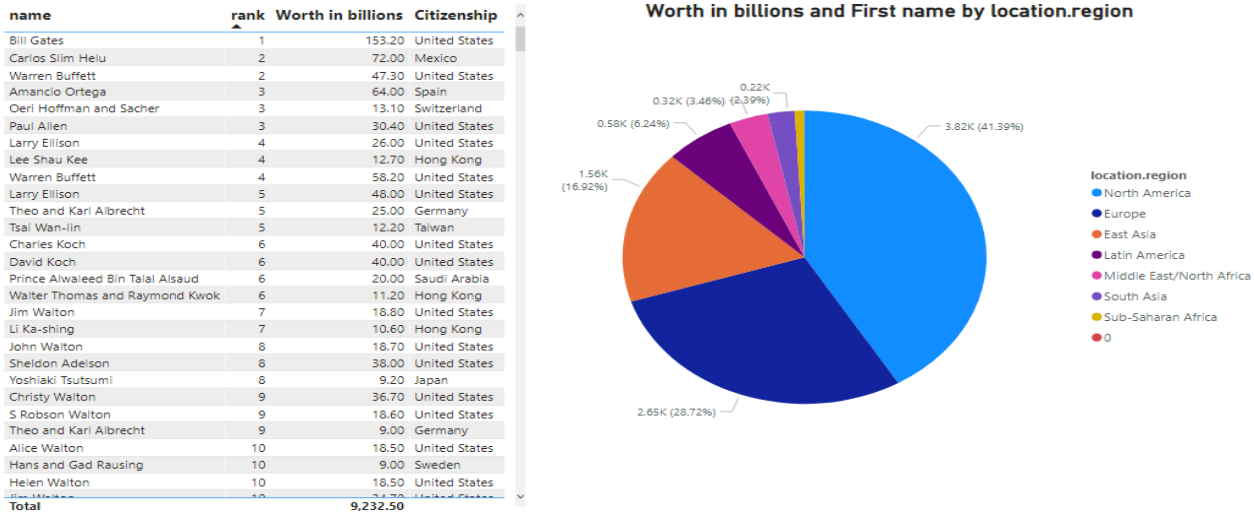
1)Create a new Power BI report, and load in the data from **billionaires** in the above workbook:

The raw data - your task is to massage it into something more presentable!

Use **Query Editor** to make this data look better:

Some columns have been renamed , Company , Name , Citizenship , Rank ,Sector and worth in billions

Bring the data back into your Power BI report, and use it to create a simple chart and list of all richest people:



name	rank	Worth in billions	Citizenship
Bill Gates	1	153.20	United States
Carlos Slim Helu	2	72.00	Mexico
Warren Buffett	2	47.30	United States
Amancio Ortega	3	64.00	Spain
Oeri Hoffman and Sacher	3	13.10	Switzerland
Paul Allen	3	30.40	United States
Larry Ellison	4	26.00	United States
Lee Shau Kee	4	12.70	Hong Kong
Warren Buffett	4	58.20	United States
Larry Ellison	5	48.00	United States
Theo and Karl Albrecht	5	25.00	Germany
Tsai Wan-lin	5	12.20	Taiwan
Charles Koch	6	40.00	United States
David Koch	6	40.00	United States
Prince Alwaleed Bin Talal Alsaud	6	20.00	Saudi Arabia
Walter Thomas and Raymond Kwok	6	11.20	Hong Kong
Jim Walton	7	18.80	United States
Li Ka-shing	7	10.60	Hong Kong
John Walton	8	18.70	United States
Sheldon Adelson	8	38.00	United States
Yoshiaki Tsutsumi	8	9.20	Japan
Christy Walton	9	36.70	United States
S Robson Walton	9	18.60	United States
Theo and Karl Albrecht	9	9.00	Germany
Alice Walton	10	18.50	United States
Hans and Gad Rausing	10	9.00	Sweden
Helen Walton	10	18.50	United States
Total		9,232.50	

Create a simple chart just to prove that the billions really are being treated as numbers.

Save this report as **But are they happy**, then close it down.

2) Create a new Power BI report.

Load the data from FTSE workbook in the above folder - you should see something like this:

In Query Editor, carry out some transforms:

Rename the remaining columns as shown below.

You should now be looking at something like this:

Company	Current price	Last closing
3i Group PLC	965.70	955.40
Admiral Group PLC	2197.00	2193.00
Anglo American PLC	2022.50	2006.00
Antofagasta PLC	955.10	941.60

The remaining table of data.

Your prices are left-aligned, which suggests that Power BI is treating them as text. Change the data types of the last two columns

Company	Current price	Last closing
3i Group PLC	965.7	955.4
Admiral Group PLC	2197	2193
Anglo American PLC	2022.5	2006

Better - now we can do arithmetic on these columns ...

Create a column which get difference between two price, rename it and sort by it to get:

Company	Current price	Last closing	Change
NMC Health PLC	2922	2856	66
Unilever PLC	4055	3999.5	55.5
AstraZeneca PLC	6267	6212	55
Reckitt Benckiser Group PLC	5951	5899	52
Ferguson PLC	5358	5316	42

*It's been a good day today for holders of shares in **NMC Health PLC**.*

Save this report as **Equity**, then close it down.

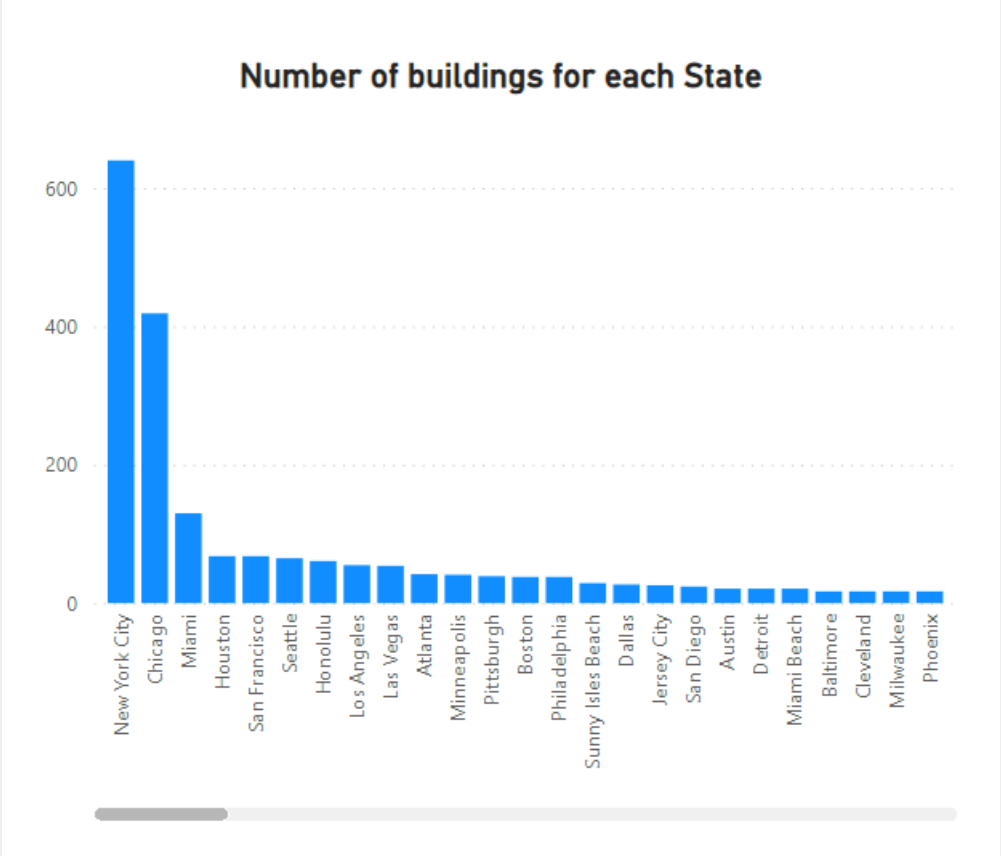
3) Create a new Power BI Desktop file, and load a list of the world's tallest buildings from **Skyscraper** Centre.

Add steps to your query so that it creates results like this:

	id	material	name	State	Country
1	12	steel	The Illinois	Chicago	US
2	62	composite	Chicago Spire	Chicago	US
3	68	composite	Miglin-Beitler Skyneedle	Chicago	US
4	98	composite	One World Trade Center	New York City	US
5	14269	concrete	Central Park Tower	New York City	US
6	16334	steel	World Trade Building	New York City	US
7	139	concrete	7 South Dearborn	Chicago	US
8	169	steel	Willis Tower	Chicago	US
9	14320	concrete	111 West 57th Street	New York City	US
10	187	composite	Bank of Southwest Tower	Houston	US
11	13227	concrete	432 Park Avenue	New York City	US
12	203	concrete	Trump International Hotel & Tower	Chicago	US
13	200	steel	One World Trade Center	New York City	US
14	202	steel	Two World Trade Center	New York City	US
15	209	composite	2 World Trade Center	New York City	US
16	19928	composite	2 World Trade Center	New York City	US
17	13325	steel/concrete	30 Hudson Yards	New York City	US
18	261	steel	Empire State Building	New York City	US
19	282	steel	Larkin Building	New York City	US

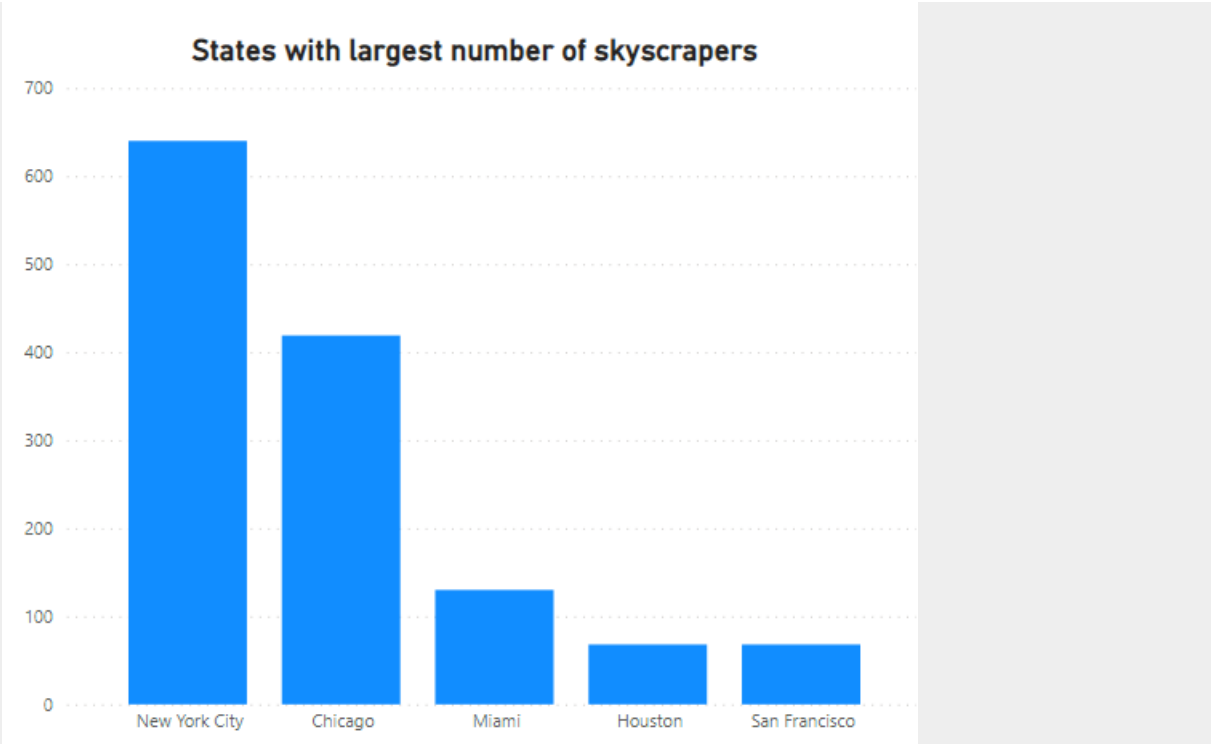
*We've split the **State** and **Country** , and renamed some other columns.*

Create a chart comparing the number of skyscrapers for each country:



The number of buildings for each State.

Apply a filter to the visualisation within Power BI Desktop itself to show only states having largest number of skyscrapers:



*No wonder New York is called ....**The Empire State** ...*

Save this file as **Skyscrapers**, and close down the Power BI instance that created it.

4) Load **Crouching Scyther, hidden Porygon** dataset and create Power bi report, With more *Pokémon* . it's time to clean up our data (and clear out the trash)!

Load this into the query editor and get rid of a couple of columns:

*The **Notes** column is gumpf and the idea of **Type(s) Secondary** is offensive. We've also moved the **National Pokedex number** column.*

Next get rid of any *Pokémon* who didn't first appear in **Red and Green**.

Finally to make our list completely free of any non-originals do a find and replace to remove any reference to **Mega Evolution**:

Since we no longer need the **First appeared column,** delete it. The column was only there to remove non-essential *Pokémon*.

	[hide]English name	Japanese name	Type(s) Primary	Evolves into
1	Bulbasaur	Fushigidane	Grass	Ivysaur (#002)
2	Ivysaur	Fushigisou	Grass	Venusaur (#003)
3	Venusaur	Fushigibana	Grass	Does not evolve
4	Charmander	Hitokage	Fire	Charmeleon (#005)
5	Charmeleon	Lizardo	Fire	Charizard (#006)
6	Charizard	Lizardon	Fire	Does not evolve
7	Squirtle	Zenigame	Water	Wartortle (#008)
8	Wartortle	Kameil	Water	Blastoise (#009)
9	Blastoise	Kamex	Water	Does not evolve
10	Caterpie	Caterpie	Bug	Metapod (#011)

We millennials might not have houses, but we do have avocado toast and PokÃ©mon!

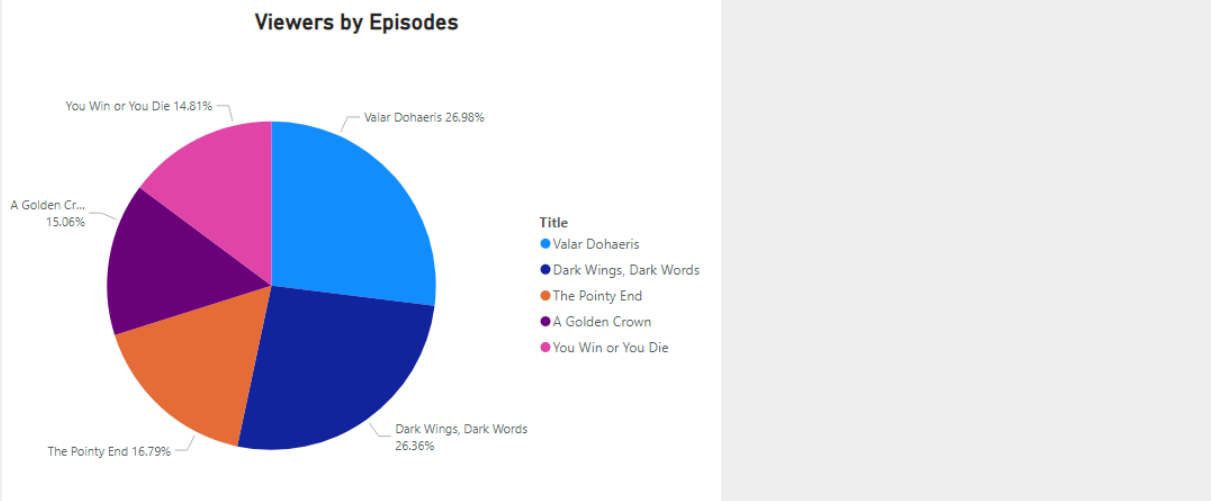
Save this as ***PokÃ©mon!***, then close it down.

5) Create a new Power BI report, and load data from **Episodes** workbook. You should see two tables (**Episodes** and **Series**).

Use Query Editor to tidy up the **Episodes** table so that it looks something like this:

Number	Title	Director	Author 1	Author 2	First aired	Us viewers (m)
6	A Golden Crown	Daniel Minahan	Various		22 May 2011	2.44
7	You Win or You Die	Daniel Minahan	David Benioff	D. B. Weiss	29 May 2011	2.4
8	The Pointy End	Daniel Minahan	George R. R. Martin		05 June 2011	2.72
21	Valar Dohaeris	Daniel Minahan	David Benioff	D. B. Weiss	31 March 2013	4.37
22	Dark Wings, Dark Words	Daniel Minahan	Vanessa Taylor		07 April 2013	4.27

Create a pie chart based on this table showing viewing figures by episode:



Your pie chart should show how many people watched each episode directed by Daniel Minahan.

Save this report as **Danny Boy**, then close it down.

6) Create a new Power BI report. Load data from the Excel workbook

Make the first row the header row and rename and remove columns to get:

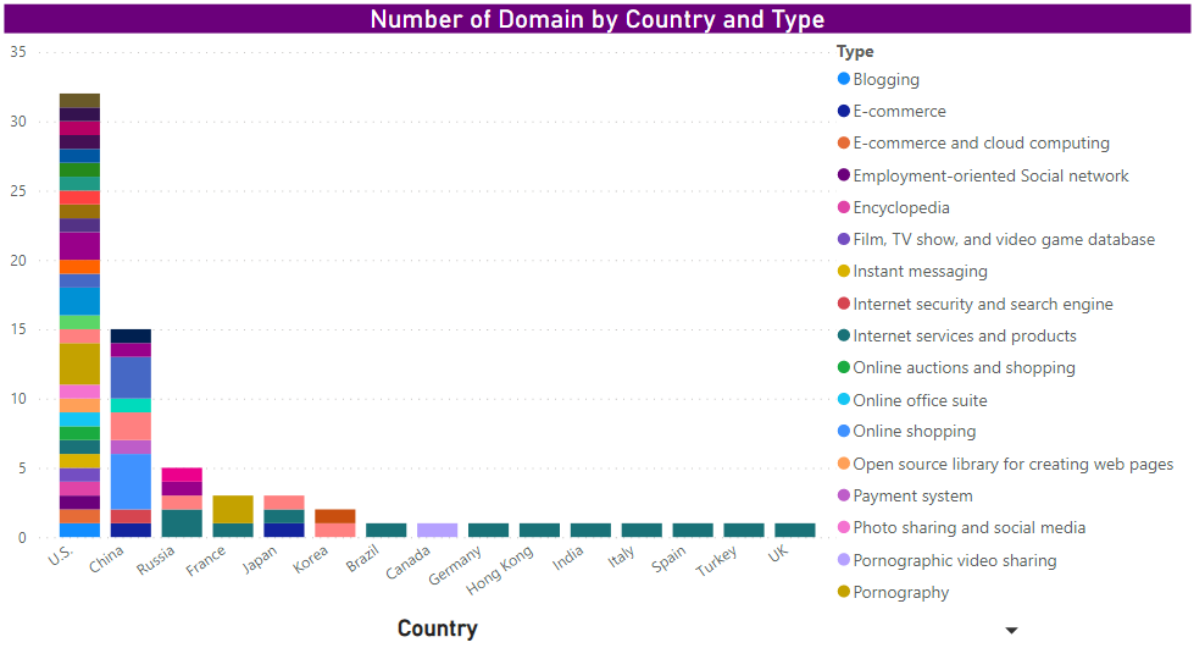
Site	Domain	Alexa rating	Type	Country
Google	google.com	1 ()	Internet services and products	U.S.
YouTube	youtube.com	2 ()	Video sharing	U.S.
Facebook	facebook.com	3 ()	Social network	U.S.

Better - now the fun begins!

Add in a column to number each row, and rename it to **Ranking**:

Use splitting columns to derive the current Alexa rank, keeping the number but losing the bit in brackets after it:

Load this data back into Power BI, and use it to create a chart something like this:



This chart shows how many domains there are for each country and type.

Save this report as **Google rules**, then close it down.

7) Create a new Power BI report, and load data from **Wiki buildings** file

Tidy the data up in **Query Editor** to get these columns:

Building	Year	City	Country	Floors	Metres	Rank	Average floor height (m)
23 Marina	2012	Dubai	UAE	89	395	25	£4.4382
432 Park Avenue	2015	New York City	United States	88	426	19	£4.8409
Abeno Harukas	2014	Osaka	Japan	60	300	129	£5
Abraj Al-Bait Clock To	2012	Mecca	Saudi Arabia	120	601	3	£5.0083
ADNOC Headquarter	2015	Abu Dhabi	UAE	76	342	52	£4.5
Ahmed Abdul Rahim	2015	Dubai	UAE	76	342	52	£4.5

Back in Query Editor, use Column from Examples to add another column giving a description of each building:

Description
23 Marina - Dubai (395 metres)
432 Park Avenue - New York City (426 metres)
Abeno Harukas - Osaka (300 metres)
Abraj Al-Bait Clock Tower - Mecca (601 metres)

Save this report as **Reach for the sky**, then close it down.

8) This exercise covers unpivoting data, changing data sources and more besides! To begin with, create a new report and load the data in the **First Half.xlsx** workbook.

User Query Editor to tidy up this data so that you have something like this:

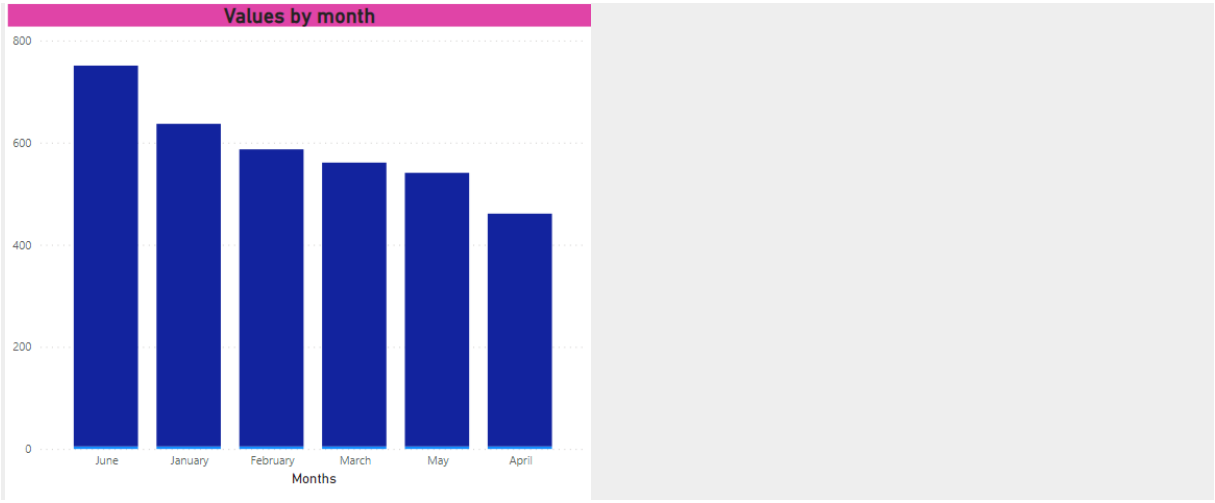
Task	January	February	March	April	May	June
Stuffing	114	77	150	40	90	192
Skinning	154	200	29	36	146	168
Furring	119	95	130	117	90	122
Accessorising	153	181	126	110	92	109
Packing	92	29	121	153	118	155

Unpivot the monthly data:

After a bit of renaming, you should have this table in Power BI:

Task	Months	Value
Accessorising	April	110
Accessorising	February	181
Accessorising	January	153
Accessorising	June	109
Accessorising	March	126
Accessorising	May	92
Furring	April	117
Furring	February	95
Furring	January	119
Furring	June	122
Furring	March	130
Furring	May	90
Packing	April	153
Packing	February	29
Packing	January	92

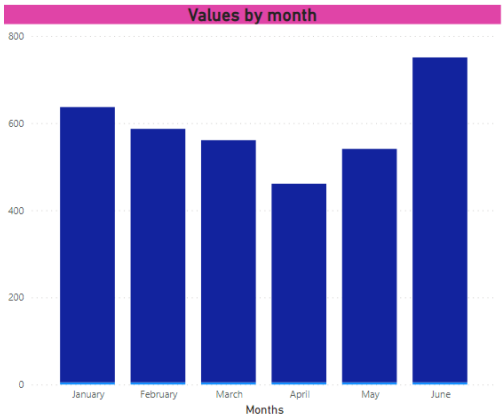
Use this to create a chart showing value by month:



The option to add a conditional column, returning the correct month number for each row.

You can now choose to sort your month names by number:

After checking your chart is sorting by the correct column, you should now have this!



The same chart, but with the correct sort order.

Now, Change your M to refer to the file called **Second half.xlsx** ,then check your chart still works:

Save your report as **The knees of a bee**, then close it down.