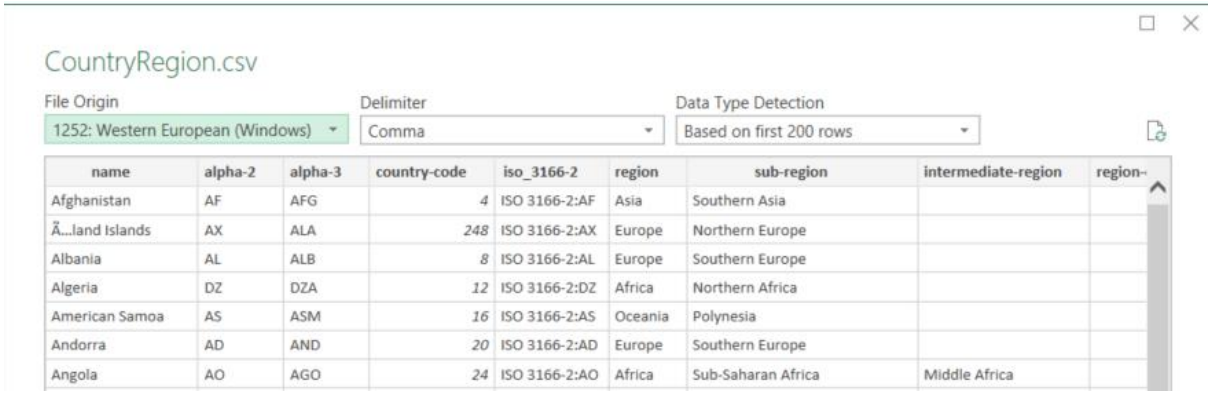
Sample Exercise Instructions SOLUTIOIN

1. Locate and download data sets you have been given access to
    a. CountryRegion.csv
    b. imf-dm-export-GDP.csv
    c. imf-dm-export-Population.csv
    d. imf_Countries

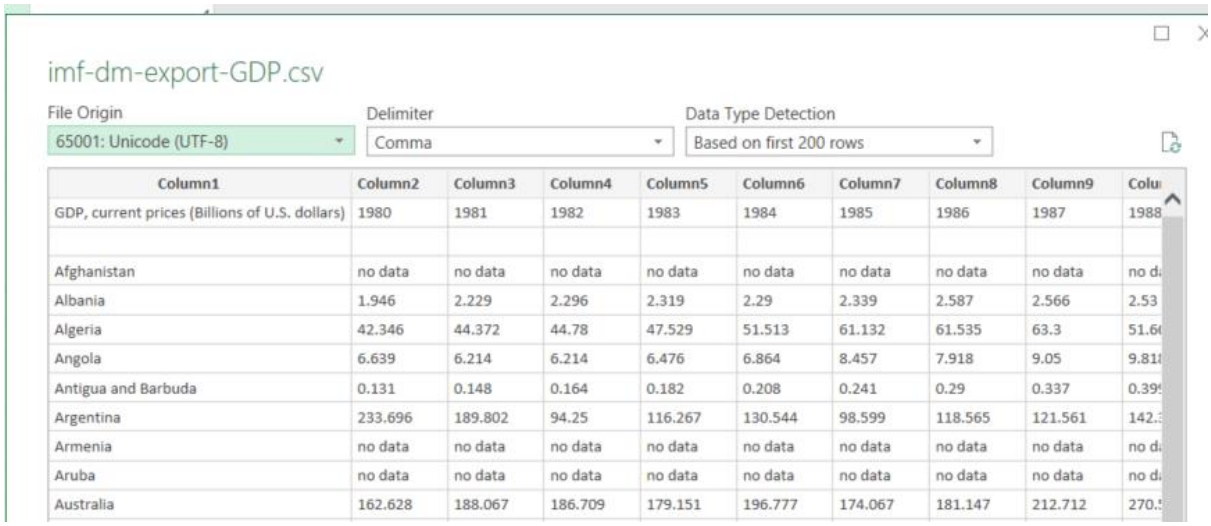
2. Open the data sets and review the structure of the data. (column and row headings). Hint: *Country names used by IMF do not match exactly the names used in the CountryRegion dataset so a cross reference file exists to link the two spellings via a Country Code column*



CountryRegion.csv

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

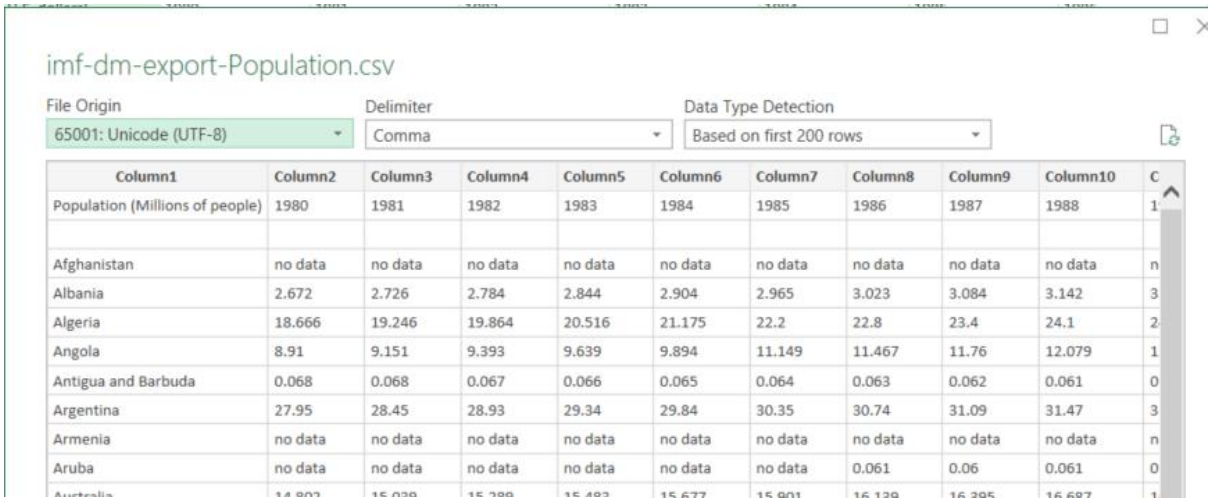
name	alpha-2	alpha-3	country-code	iso_3166-2	region	sub-region	intermediate-region	region-
Afghanistan	AF	AFG	4	ISO 3166-2:AF	Asia	Southern Asia		
Ã…land Islands	AX	ALA	248	ISO 3166-2:AX	Europe	Northern Europe		
Albania	AL	ALB	8	ISO 3166-2:AL	Europe	Southern Europe		
Algeria	DZ	DZA	12	ISO 3166-2:DZ	Africa	Northern Africa		
American Samoa	AS	ASM	16	ISO 3166-2:AS	Oceania	Polynesia		
Andorra	AD	AND	20	ISO 3166-2:AD	Europe	Southern Europe		
Angola	AO	AGO	24	ISO 3166-2:AO	Africa	Sub-Saharan Africa	Middle Africa	



imf-dm-export-GDP.csv

File Origin: 65001: Unicode (UTF-8) | Delimiter: Comma | Data Type Detection: Based on first 200 rows

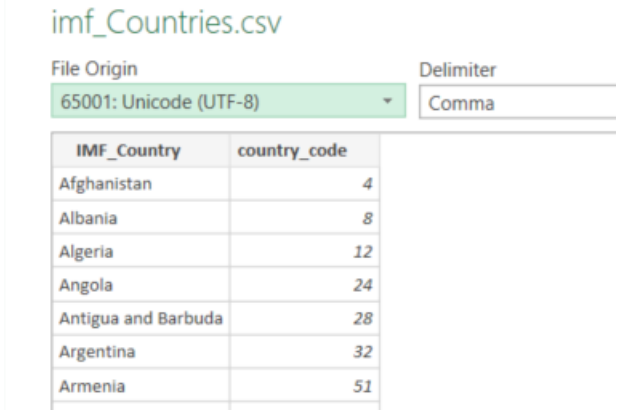
Column1	Column2	Column3	Column4	Column5	Column6	Column7	Column8	Column9	Colu
GDP, current prices (Billions of U.S. dollars)	1980	1981	1982	1983	1984	1985	1986	1987	1988
Afghanistan	no data	no data	no data	no data	no data	no data	no data	no data	no d
Albania	1.946	2.229	2.296	2.319	2.29	2.339	2.587	2.566	2.53
Algeria	42.346	44.372	44.78	47.529	51.513	61.132	61.535	63.3	51.6
Angola	6.639	6.214	6.214	6.476	6.864	8.457	7.918	9.05	9.81
Antigua and Barbuda	0.131	0.148	0.164	0.182	0.208	0.241	0.29	0.337	0.39
Argentina	233.696	189.802	94.25	116.267	130.544	98.599	118.565	121.561	142.
Armenia	no data	no data	no data	no data	no data	no data	no data	no data	no d
Aruba	no data	no data	no data	no data	no data	no data	no data	no data	no d
Australia	162.628	188.067	186.709	179.151	196.777	174.067	181.147	212.712	270.



imf-dm-export-Population.csv

File Origin: 65001: Unicode (UTF-8) | Delimiter: Comma | Data Type Detection: Based on first 200 rows

Column1	Column2	Column3	Column4	Column5	Column6	Column7	Column8	Column9	Column10	C
Population (Millions of people)	1980	1981	1982	1983	1984	1985	1986	1987	1988	1
Afghanistan	no data	no data	no data	no data	no data	no data	no data	no data	no data	n
Albania	2.672	2.726	2.784	2.844	2.904	2.965	3.023	3.084	3.142	3
Algeria	18.666	19.246	19.864	20.516	21.175	22.2	22.8	23.4	24.1	2
Angola	8.91	9.151	9.393	9.639	9.894	11.149	11.467	11.76	12.079	1
Antigua and Barbuda	0.068	0.068	0.067	0.066	0.065	0.064	0.063	0.062	0.061	0
Argentina	27.95	28.45	28.93	29.34	29.84	30.35	30.74	31.09	31.47	3
Armenia	no data	no data	no data	no data	no data	no data	no data	no data	no data	n
Aruba	no data	no data	no data	no data	no data	no data	0.061	0.06	0.061	0



imf_Countries.csv

File Origin: 65001: Unicode (UTF-8) | Delimiter: Comma

IMF_Country	country_code
Afghanistan	4
Albania	8
Algeria	12
Angola	24
Antigua and Barbuda	28
Argentina	32
Armenia	51

CountryRegion.csv contains all the names of countries with their codes and regions and sub regions. So ideal if you need to look up any categorical information like Europe, Asia for Region and North America, Sub-Sahara Africa for Sub Regions.

3. Using your skills create a data model for Country, Population and GDP

4. To achieve this you may need to combine the data sets, cleanse and filter.

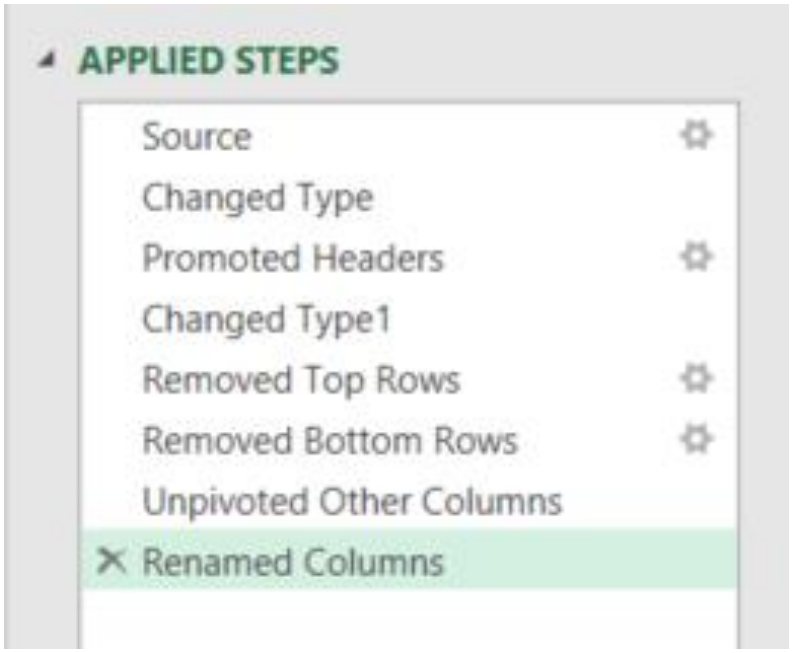
The imf-dm-export-GDP and Population files need to be converted into this format

	Country	Year	GDP
1	Afghanistan	1980	no data
2	Afghanistan	1981	no data
3	Afghanistan	1982	no data
4	Afghanistan	1983	no data
5	Afghanistan	1984	no data
6	Afghanistan	1985	no data
7	Afghanistan	1986	no data
8	Afghanistan	1987	no data
9	Afghanistan	1988	no data
10	Afghanistan	1989	no data
11	Afghanistan	1990	no data
12	Afghanistan	1991	no data
13	Afghanistan	1992	no data
14	Afghanistan	1993	no data
15	Afghanistan	1994	no data
16	Afghanistan	1995	no data
17	Afghanistan	1996	no data
18	Afghanistan	1997	no data
19	Afghanistan	1998	no data
20	Afghanistan	1999	no data
21	Afghanistan	2000	no data
22	Afghanistan	2001	no data
23	Afghanistan	2002	4.367
24	Afghanistan	2003	4.553
25	Afghanistan	2004	5.146
26	Afghanistan	2005	6.167
27	Afghanistan	2006	6.925
28	Afghanistan	2007	8.556
29	Afghanistan	2008	10.297
30	Afghanistan	2009	12.066

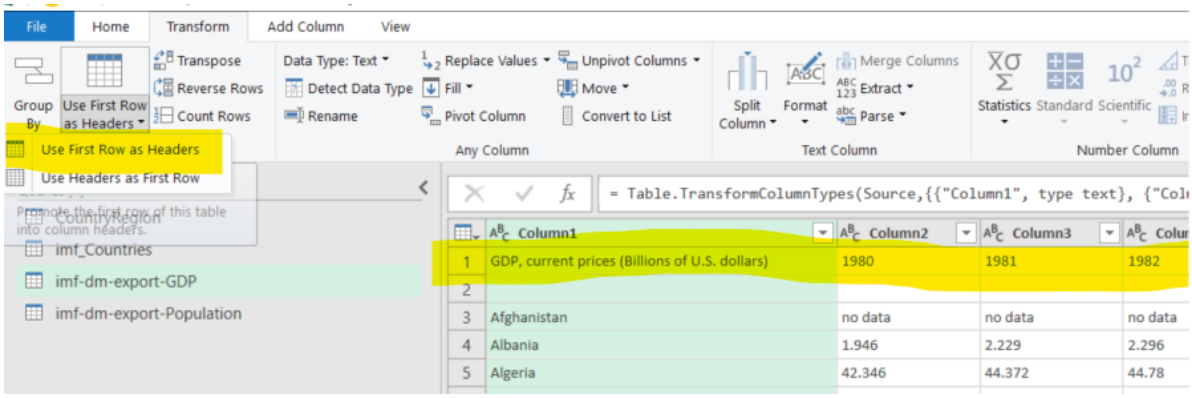
	Country	Year	Population
1	Afghanistan	1980	no data
2	Afghanistan	1981	no data
3	Afghanistan	1982	no data
4	Afghanistan	1983	no data
5	Afghanistan	1984	no data
6	Afghanistan	1985	no data
7	Afghanistan	1986	no data
8	Afghanistan	1987	no data
9	Afghanistan	1988	no data
10	Afghanistan	1989	no data
11	Afghanistan	1990	no data
12	Afghanistan	1991	no data
13	Afghanistan	1992	no data
14	Afghanistan	1993	no data
15	Afghanistan	1994	no data
16	Afghanistan	1995	no data
17	Afghanistan	1996	no data
18	Afghanistan	1997	no data
19	Afghanistan	1998	no data
20	Afghanistan	1999	no data
21	Afghanistan	2000	no data
22	Afghanistan	2001	17.887
23	Afghanistan	2002	18.707
24	Afghanistan	2003	19.477
25	Afghanistan	2004	20.237
26	Afghanistan	2005	20.947
27	Afghanistan	2006	21.596
28	Afghanistan	2007	22.426
29	Afghanistan	2008	22.997

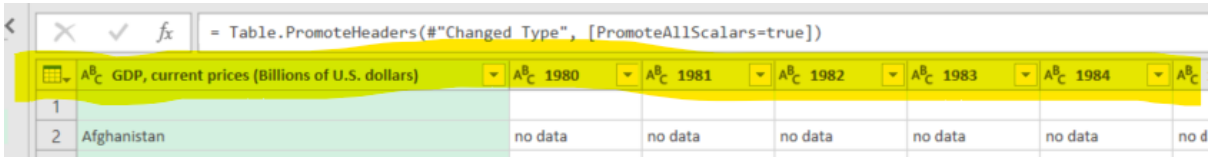
To achieve this various steps need to be performed culminating in unpivoting all the columns except the country name

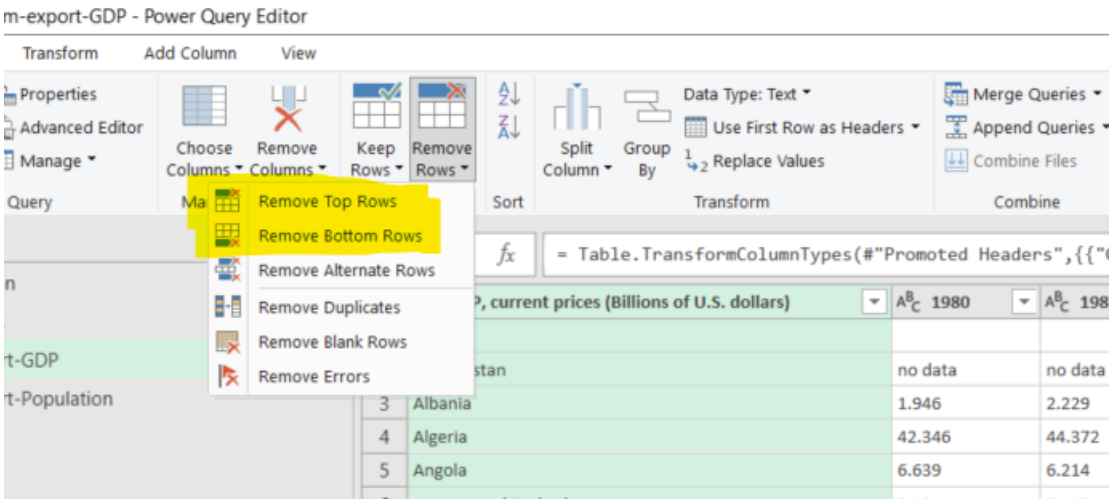
PowerQuery is the ideal tool to complete these tasks for example



So let us complete each step in turn

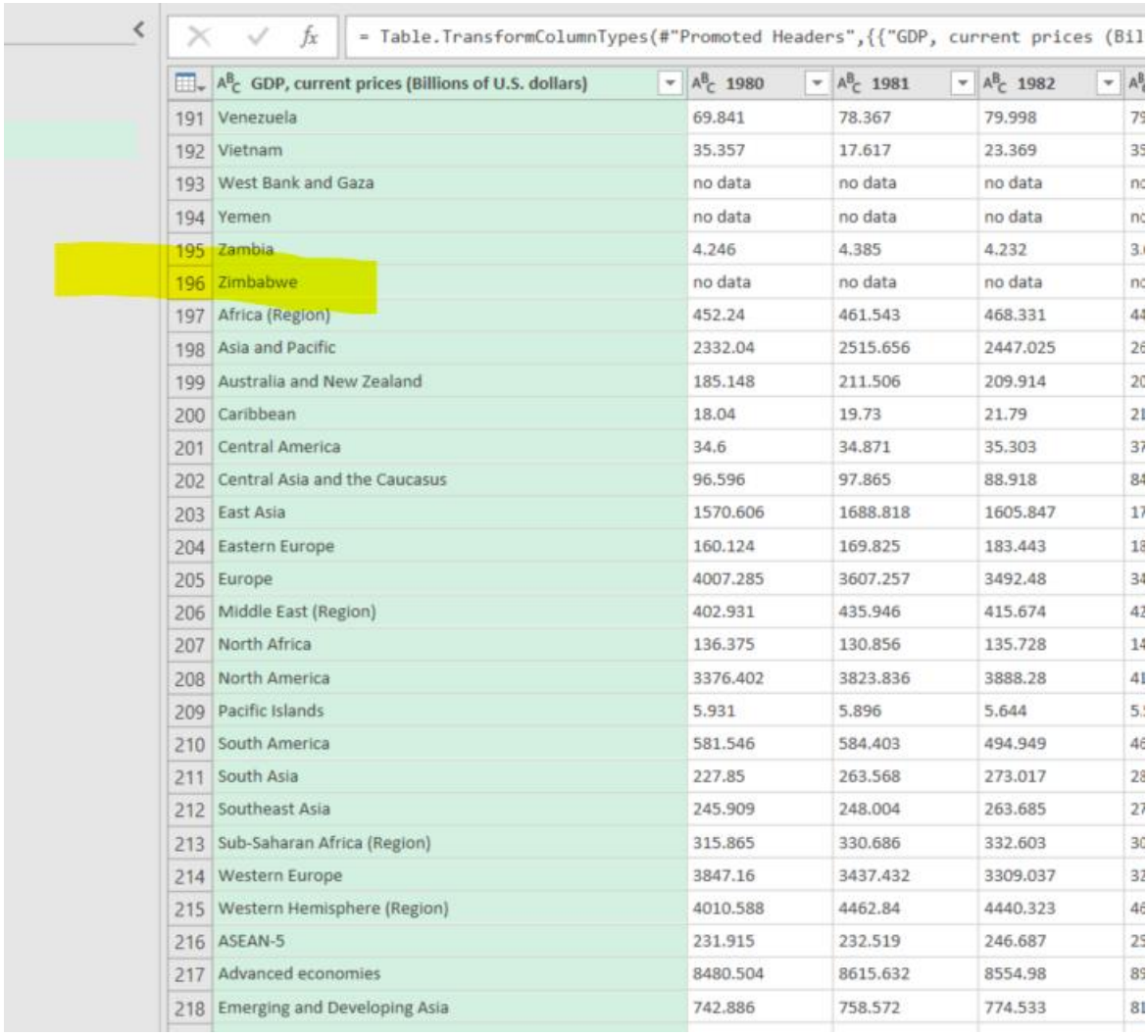






Remove top row and also the bottom 34 rows that include summary information for the regions. The last row should be Zimbabwe

Before

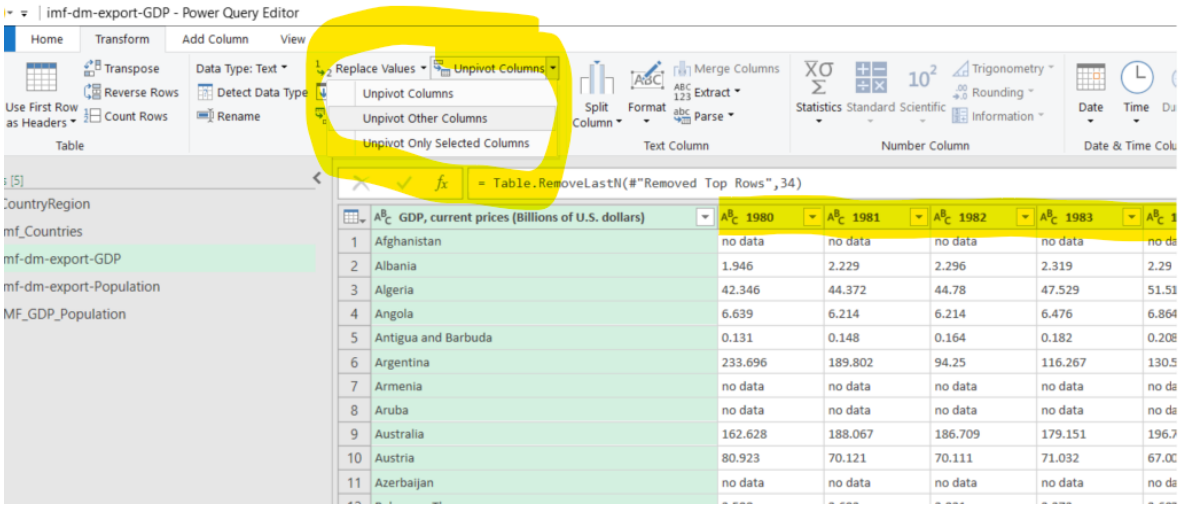


After

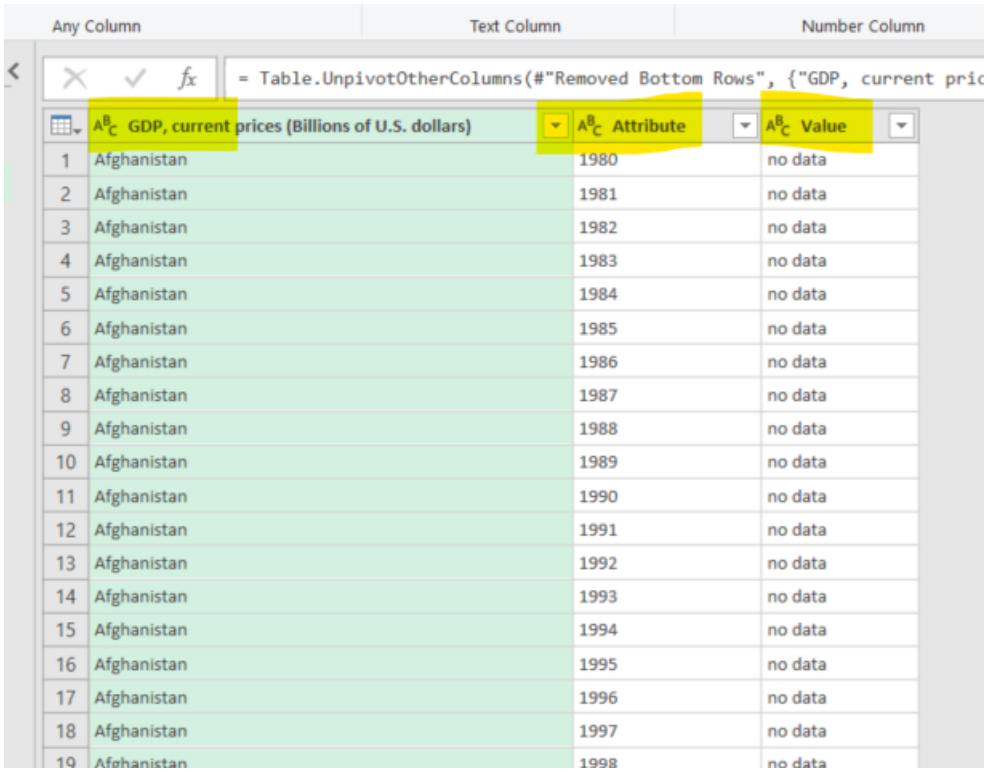
183	Ukraine	no data	no data	no data
184	United Arab Emirates	40.415	45.002	41.848
185	United Kingdom	603.983	587.652	558.72
186	United States	2857.325	3207.025	3343.8
187	Uruguay	12.165	13.578	11.1
188	Uzbekistan	no data	no data	no data
189	Vanuatu	0.121	0.114	0.114
190	Venezuela	69.841	78.367	79.998
191	Vietnam	35.357	17.617	23.369
192	West Bank and Gaza	no data	no data	no data
193	Yemen	no data	no data	no data
194	Zambia	4.246	4.385	4.232
195	Zimbabwe	no data	no data	no data

Now to do any analysis on this data we need to be able to differentiate the year associated with the GDPO and Populastion. Currenlty the Year is represented as a column so there are the years 1980 to 2026 as 46 columns one for each year in the range.

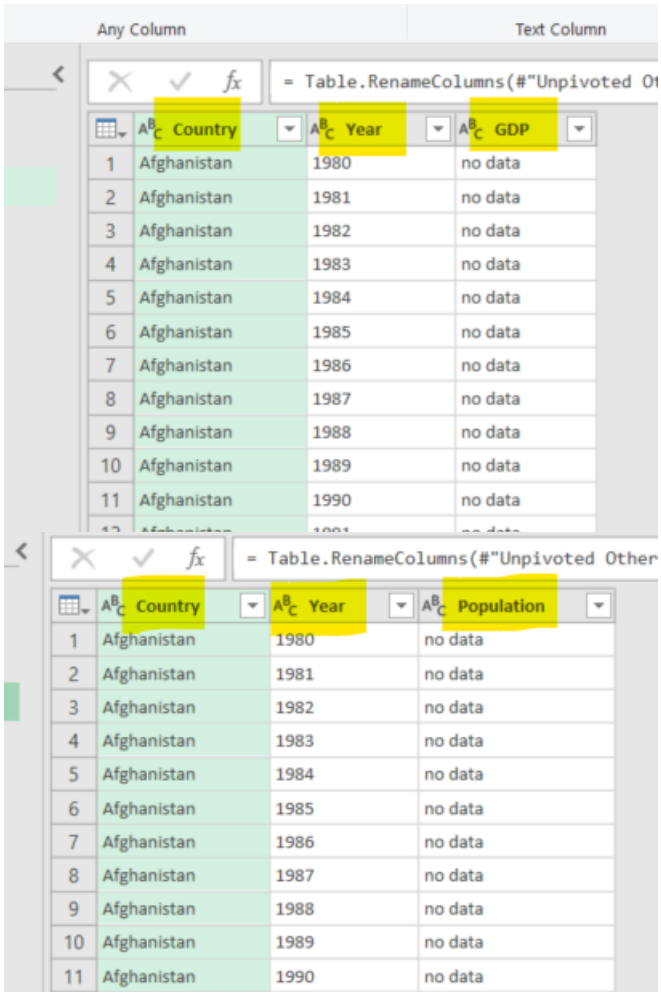
To convert these columns to rows we use the Unpivot option in the Transform tab.



This results in



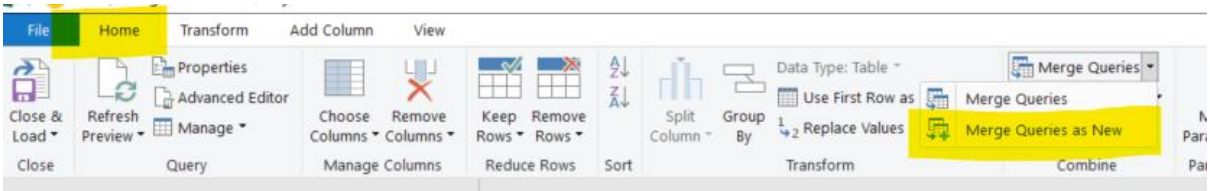
So now we just rename the columns and do exactly the same steps for the Population data set.

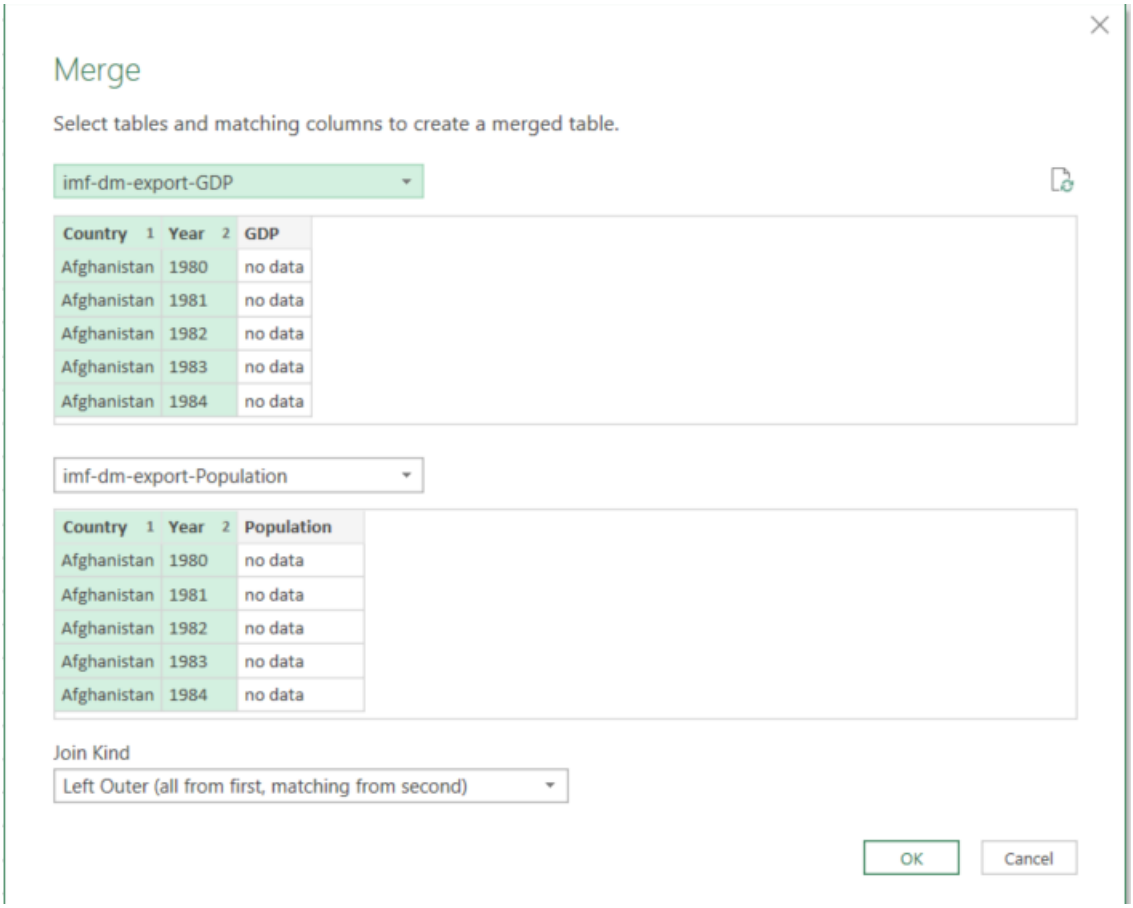


We now MERGE these two data sets together to produce one dataset consisting of

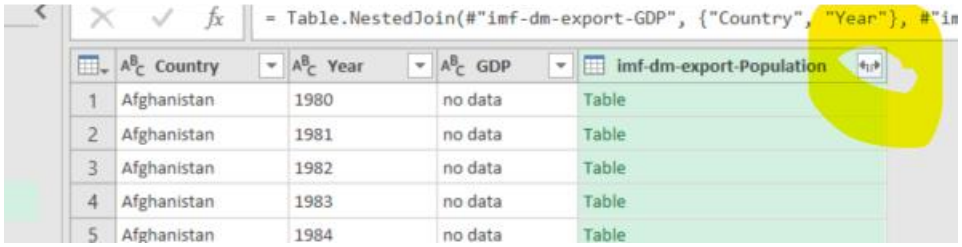
	Country	Year	GDP	Population
1	Afghanistan	1980	no data	no data
2	Afghanistan	1981	no data	no data
3	Afghanistan	1982	no data	no data
4	Afghanistan	1983	no data	no data
5	Afghanistan	1984	no data	no data
6	Afghanistan	1985	no data	no data
7	Afghanistan	1986	no data	no data
8	Afghanistan	1987	no data	no data
9	Afghanistan	1988	no data	no data
10	Afghanistan	1989	no data	no data
11	Afghanistan	1990	no data	no data
12	Afghanistan	1991	no data	no data
13	Afghanistan	1992	no data	no data
14	Afghanistan	1993	no data	no data
15	Afghanistan	1994	no data	no data
16	Afghanistan	1995	no data	no data
17	Afghanistan	1996	no data	no data
18	Afghanistan	1997	no data	no data
19	Afghanistan	1998	no data	no data
20	Afghanistan	1999	no data	no data
21	Afghanistan	2000	no data	no data
22	Afghanistan	2001	no data	17.887
23	Afghanistan	2002	4.367	18.707
24	Afghanistan	2003	4.553	19.477
25	Afghanistan	2004	5.146	20.237
26	Afghanistan	2005	6.167	20.947
27	Afghanistan	2006	6.925	21.596
28	Afghanistan	2007	8.556	22.426
29	Afghanistan	2008	10.297	22.997
30	Afghanistan	2009	12.066	23.596
31	Afghanistan	2010	15.325	24.269
32	Afghanistan	2011	17.89	25.031
33	Afghanistan	2012	20.293	25.864
34	Afghanistan	2013	20.17	26.736
35	Afghanistan	2014	20.635	27.601
36	Afghanistan	2015	20.22	28.425

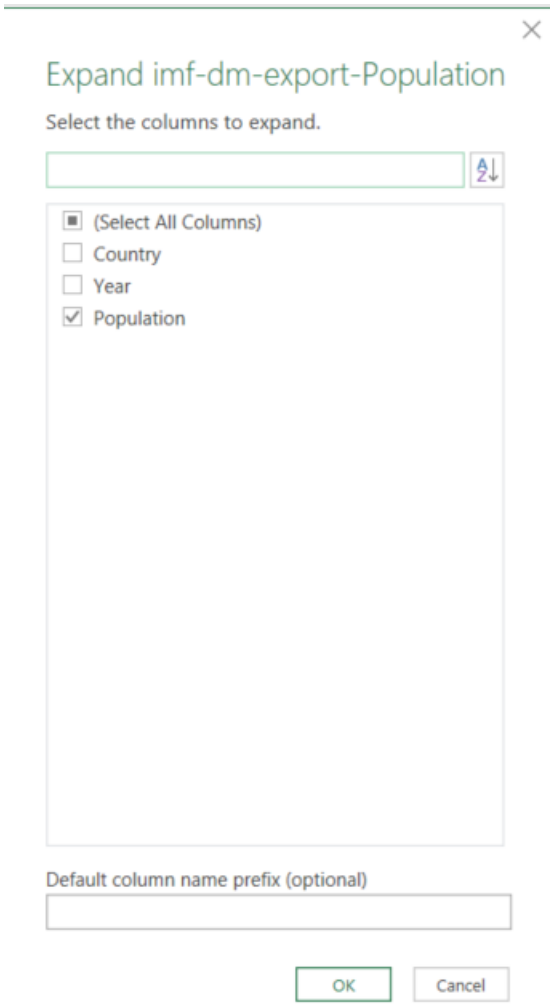
To complete this we use the Merge as New option and join on the Country name and Year this should say 195 rows match.





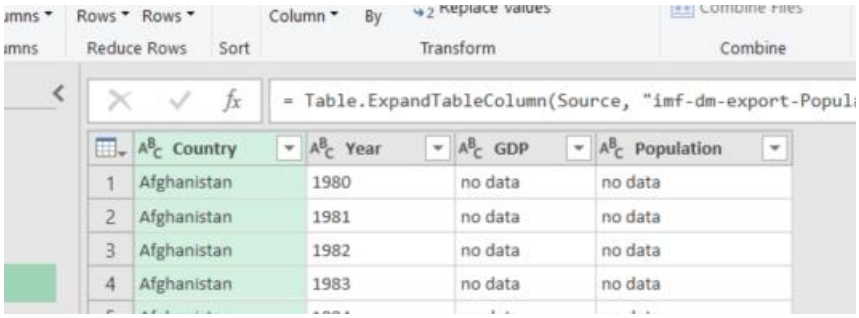
Make sure we expand the Table reference which is the second table we merged imf-dm-export-Population



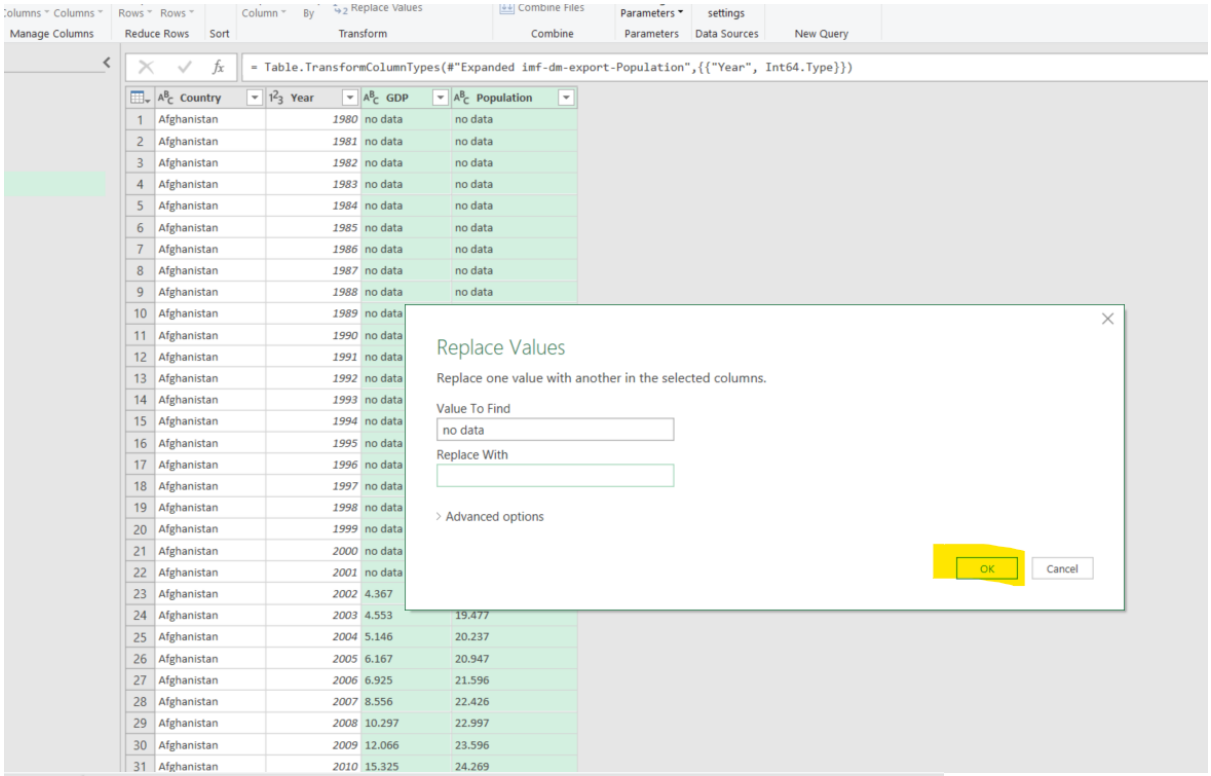


We only need to add the Population column as we already have the name and year.

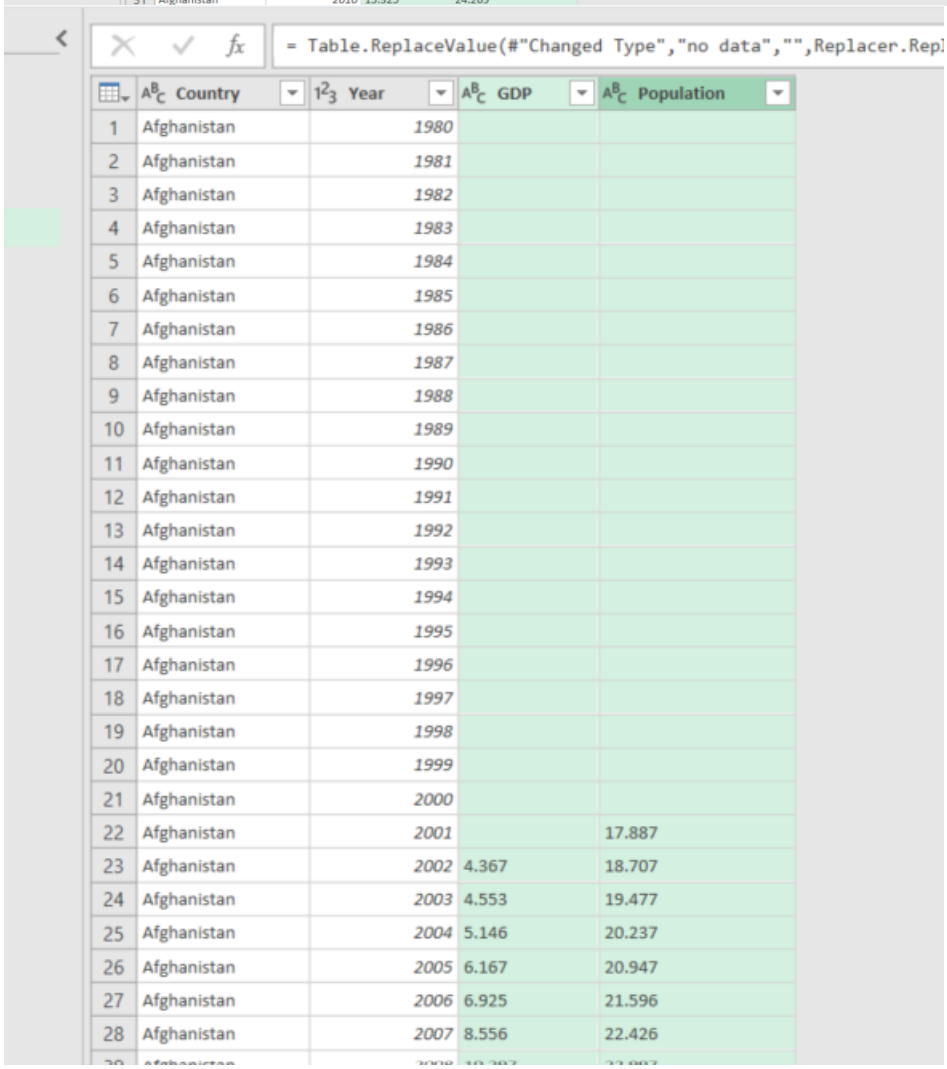
Data types of our columns are very important



Currently Year, GDP and Population are all Text columns. This is because the GDP and Population columns include the text “no data” from the original IMF csv files. We need to replace these with *null* to be able to convert them to decimal columns.

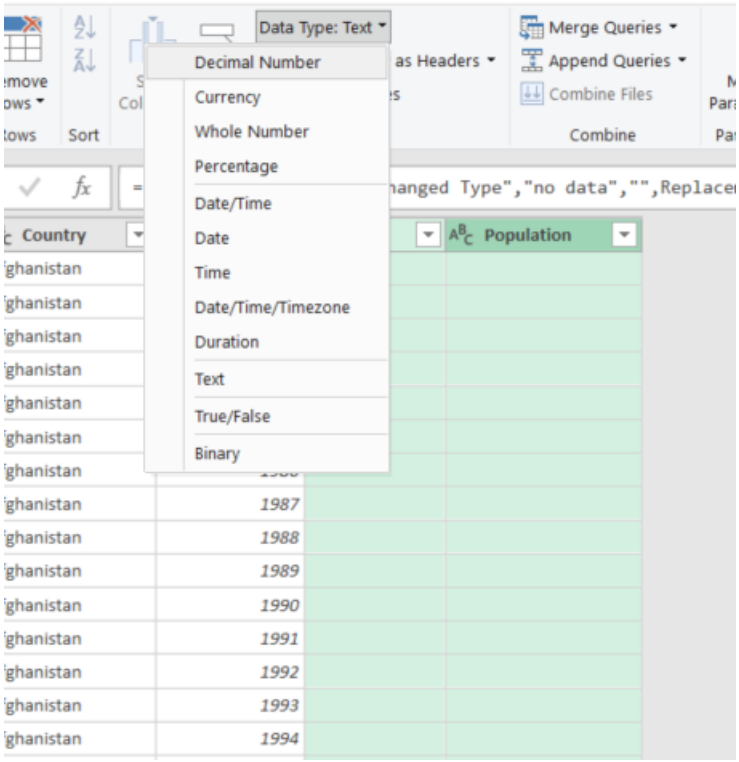


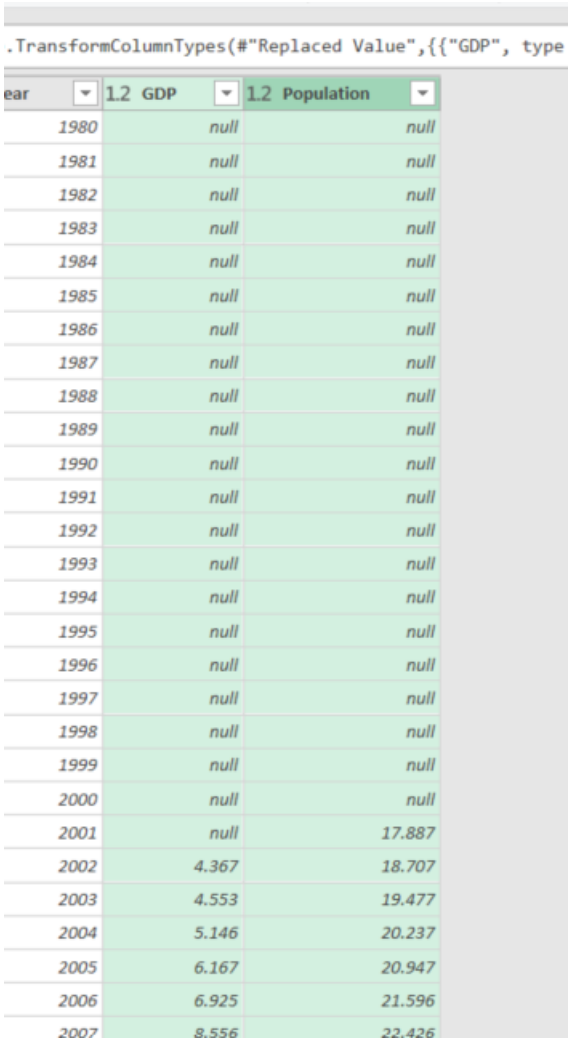
	Country	Year	GDP	Population
1	Afghanistan	1980	no data	no data
2	Afghanistan	1981	no data	no data
3	Afghanistan	1982	no data	no data
4	Afghanistan	1983	no data	no data
5	Afghanistan	1984	no data	no data
6	Afghanistan	1985	no data	no data
7	Afghanistan	1986	no data	no data
8	Afghanistan	1987	no data	no data
9	Afghanistan	1988	no data	no data
10	Afghanistan	1989	no data	
11	Afghanistan	1990	no data	
12	Afghanistan	1991	no data	
13	Afghanistan	1992	no data	
14	Afghanistan	1993	no data	
15	Afghanistan	1994	no data	
16	Afghanistan	1995	no data	
17	Afghanistan	1996	no data	
18	Afghanistan	1997	no data	
19	Afghanistan	1998	no data	
20	Afghanistan	1999	no data	
21	Afghanistan	2000	no data	
22	Afghanistan	2001	no data	
23	Afghanistan	2002	4.367	
24	Afghanistan	2003	4.553	19.477
25	Afghanistan	2004	5.146	20.237
26	Afghanistan	2005	6.167	20.947
27	Afghanistan	2006	6.925	21.596
28	Afghanistan	2007	8.556	22.426
29	Afghanistan	2008	10.297	22.997
30	Afghanistan	2009	12.066	23.596
31	Afghanistan	2010	15.325	24.269



	Country	Year	GDP	Population
1	Afghanistan	1980		
2	Afghanistan	1981		
3	Afghanistan	1982		
4	Afghanistan	1983		
5	Afghanistan	1984		
6	Afghanistan	1985		
7	Afghanistan	1986		
8	Afghanistan	1987		
9	Afghanistan	1988		
10	Afghanistan	1989		
11	Afghanistan	1990		
12	Afghanistan	1991		
13	Afghanistan	1992		
14	Afghanistan	1993		
15	Afghanistan	1994		
16	Afghanistan	1995		
17	Afghanistan	1996		
18	Afghanistan	1997		
19	Afghanistan	1998		
20	Afghanistan	1999		
21	Afghanistan	2000		
22	Afghanistan	2001		17.887
23	Afghanistan	2002	4.367	18.707
24	Afghanistan	2003	4.553	19.477
25	Afghanistan	2004	5.146	20.237
26	Afghanistan	2005	6.167	20.947
27	Afghanistan	2006	6.925	21.596
28	Afghanistan	2007	8.556	22.426

Now with GDP and Populatiopn columns selected use the Data Type option on the Home Tab


Data Type: Text
Decimal Number
Currency
Whole Number
Percentage
Date/Time
Date
Time
Date/Time/Timezone
Duration
Text
True/False
Binary
as Headers
Merge Queries
Append Queries
Combine Files
Combine
Sort
Country
Population
ghanistan
1987
1988
1989
1990
1991
1992
1993
1994



.TransformColumnTypes(#"Replaced Value",{{"GDP", type

ear	1.2 GDP	1.2 Population
1980	null	null
1981	null	null
1982	null	null
1983	null	null
1984	null	null
1985	null	null
1986	null	null
1987	null	null
1988	null	null
1989	null	null
1990	null	null
1991	null	null
1992	null	null
1993	null	null
1994	null	null
1995	null	null
1996	null	null
1997	null	null
1998	null	null
1999	null	null
2000	null	null
2001	null	17.887
2002	4.367	18.707
2003	4.553	19.477
2004	5.146	20.237
2005	6.167	20.947
2006	6.925	21.596
2007	8.556	22.426

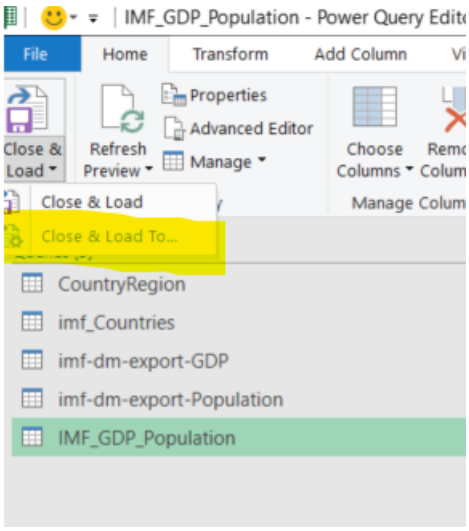
Rename the merged dataset as IMF_GDP_Population

1. Some suggested analysis

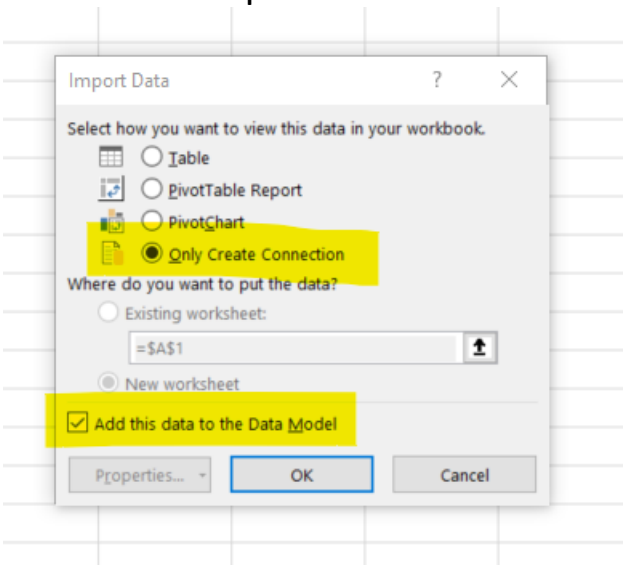
    a. What are the average GDPs and Populations for all countries?

Best way of achieving this in Excel is with a Pivot Table

When leaving Power Query using the option

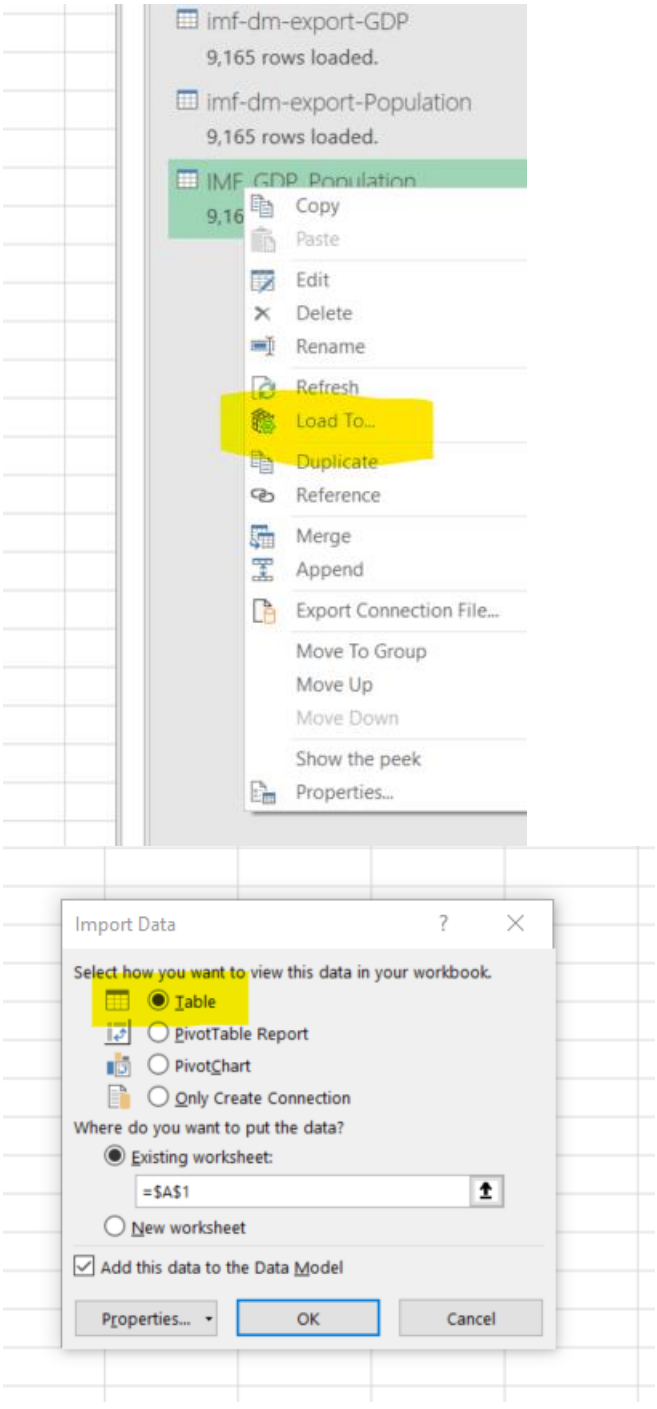


Select these options



By default you end up with worksheets for each of the datasets. We only need one data set to start with and that is the IMF_GDP_Population dataset

To create a Worksheet with that dataset as a table use the following option when right clicking on the relevant connection

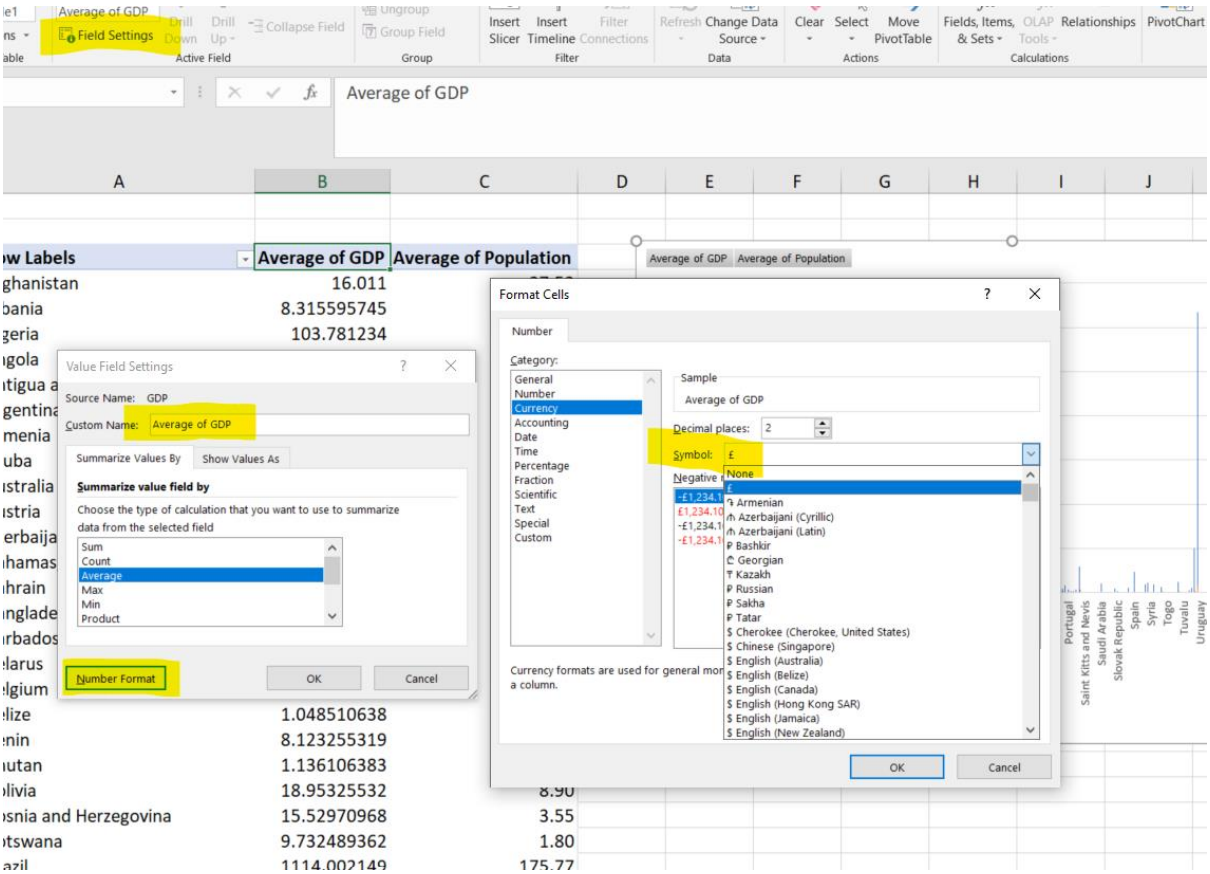

imf-dm-export-GDP
9,165 rows loaded.
imf-dm-export-Population
9,165 rows loaded.
IMF_GDP_Population
9,16
Copy
Paste
Edit
Delete
Rename
Refresh
Load To...
Duplicate
Reference
Merge
Append
Export Connection File...
Move To Group
Move Up
Move Down
Show the peek
Properties...

Import Data
Select how you want to view this data in your workbook.
Table
PivotTable Report
PivotChart
Only Create Connection
Where do you want to put the data?
Existing worksheet:
=A1
New worksheet
Add this data to the Data Model
Properties...
OK
Cancel

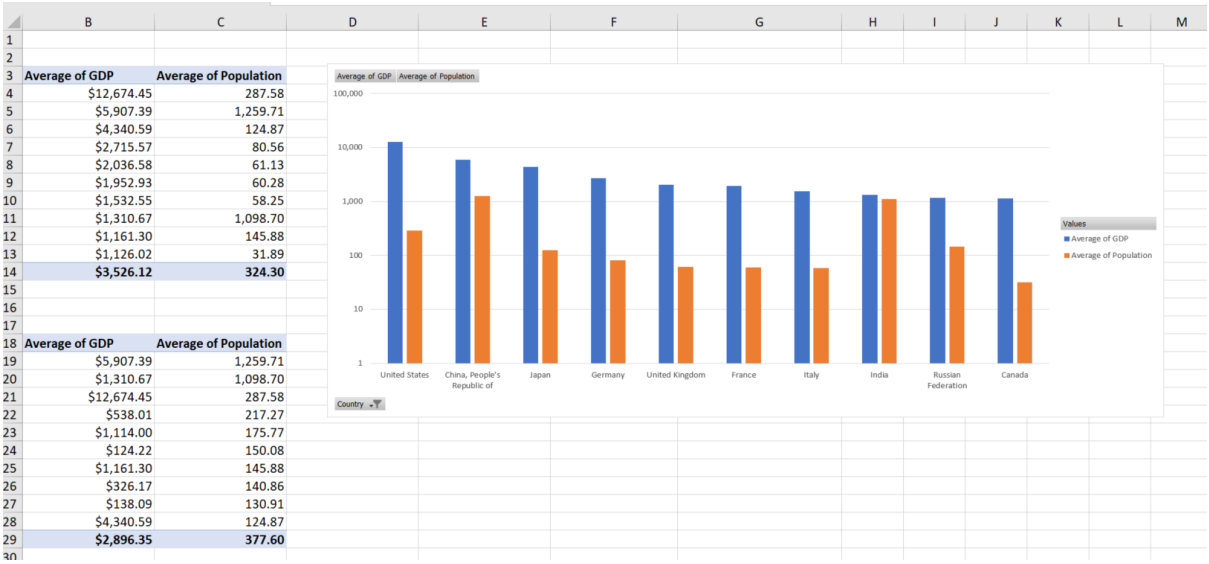
	A	B	C	D
1	Country	Year	GDP	Population
2	Afghanistan	1980		
3	Afghanistan	1981		
4	Afghanistan	1982		
5	Afghanistan	1983		
6	Afghanistan	1984		
7	Afghanistan	1985		
8	Afghanistan	1986		
9	Afghanistan	1987		
10	Afghanistan	1988		
11	Afghanistan	1989		
12	Afghanistan	1990		
13	Afghanistan	1991		
14	Afghanistan	1992		
15	Afghanistan	1993		
16	Afghanistan	1994		
17	Afghanistan	1995		
18	Afghanistan	1996		
19	Afghanistan	1997		
20	Afghanistan	1998		
21	Afghanistan	1999		
22	Afghanistan	2000		
23	Armenia	1980		
24	Armenia	1981		
25	Armenia	1982		
26	Armenia	1983		
27	Armenia	1984		
28	Armenia	1985		
29	Armenia	1986		
30	Armenia	1987		
31	Armenia	1988		

9148	Samoa	1992	0.188
9149	Samoa	1993	0.193
9150	Samoa	1994	0.154
9151	Samoa	1995	0.236
9152	Samoa	1996	0.251
9153	Samoa	1997	0.273
9154	Somalia	2011	3.499
9155	Suriname	1980	1.192
9156	Suriname	1981	1.333
9157	Suriname	1982	1.372
9158	Suriname	1983	1.325
9159	Suriname	1984	1.296
9160	Suriname	1985	1.309
9161	Suriname	1986	1.336
9162	Suriname	1987	1.469
9163	Suriname	1988	1.741
9164	Suriname	1989	2.034
9165	Tuvalu	2000	0.014
9166	Tuvalu	2001	0.013

Using the table create a Pivot Table to summarise GDP and Population



Top 10 countries by GDP and Population



Average of GDP	Average of Population
$12,674.45	287.58
$5,907.39	1,259.71
$4,340.59	124.87
$2,715.57	80.56
$2,036.58	61.13
$1,952.93	60.28
$1,532.55	58.25
$1,310.67	1,098.70
$1,161.30	145.88
$1,126.02	31.89
$3,526.12	**324.30**

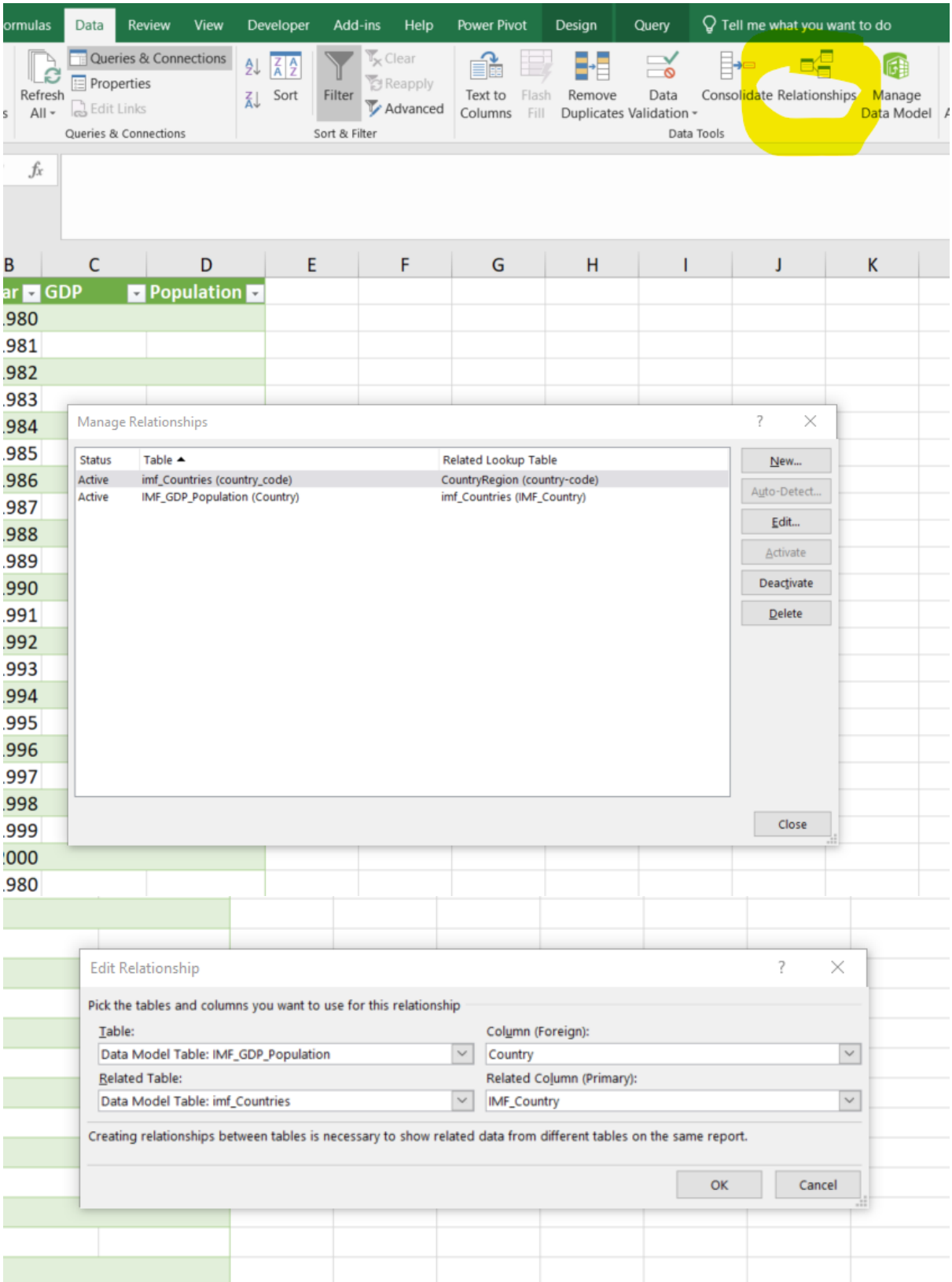
Average of GDP	Average of Population
$5,907.39	1,259.71
$1,310.67	1,098.70
$12,674.45	287.58
$538.01	217.27
$1,114.00	175.77
$124.22	150.08
$1,161.30	145.88
$326.17	140.86
$138.09	130.91
$4,340.59	124.87
$2,896.35	**377.60**

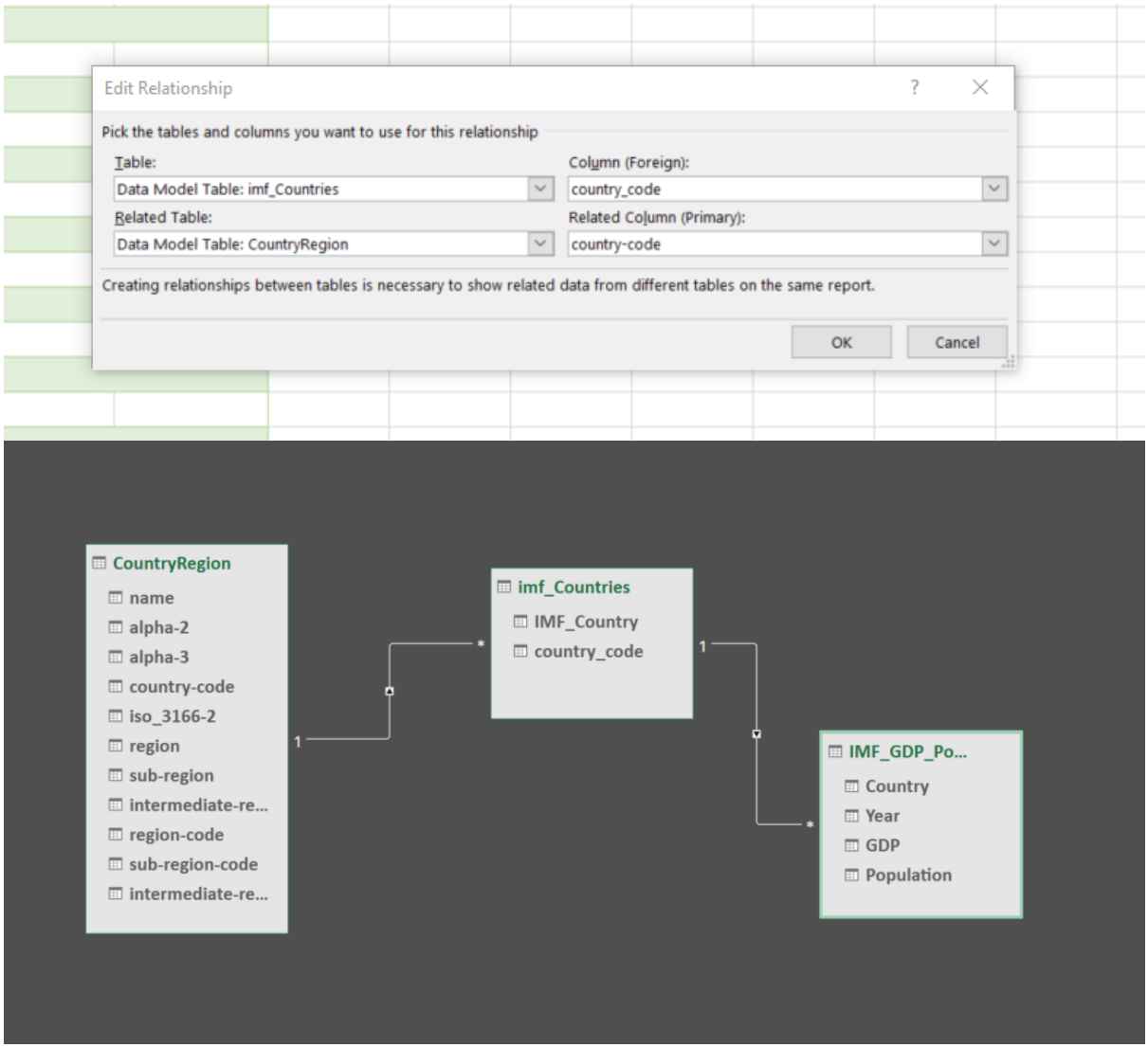
2. As an enhancement link the Region and Sub Region to the country so summary statistics for GDP and Population can be shown for the region information.

To achieve this we need to link the CountryRegion dataset to the IMF_GDP_Population dataset. We can do this using Power Pivot and first set up the relationships in the Data tab where we link

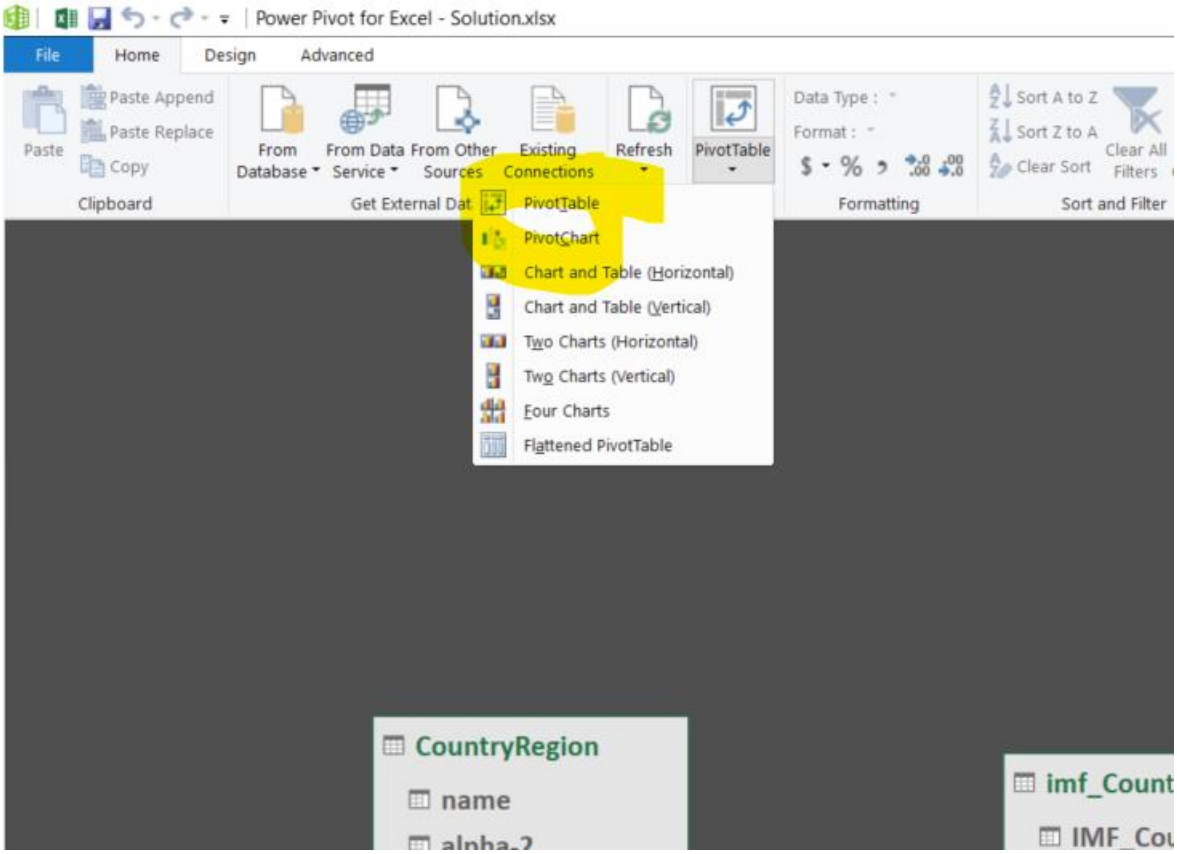
Country field in ***IMF_GDP_Population*** with ***imf_Countries IMF_Country*** field

We then link the ***imf_Countries*** field ***country_code*** with ***CountryRegion country-code*** field (hyphen rather than underscore)

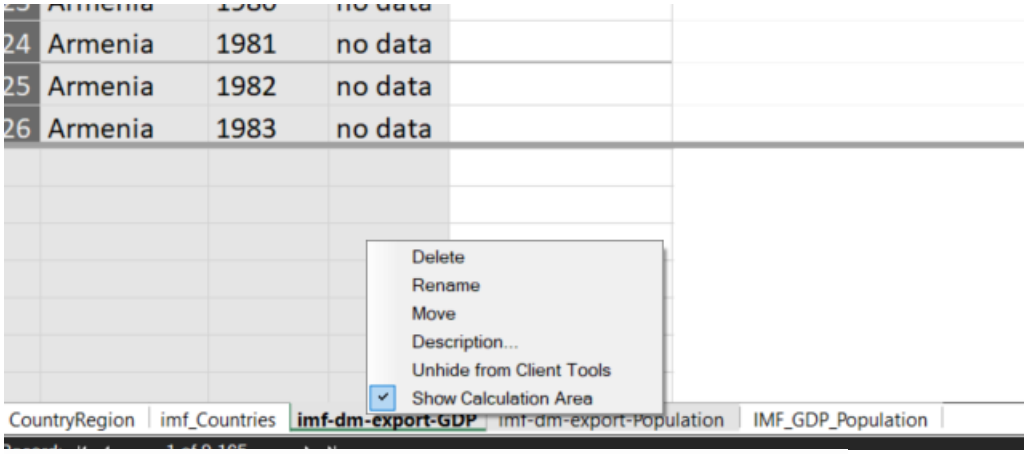

ormulas
Data
Review
View
Developer
Add-ins
Help
Power Pivot
Design
Query
Tell me what you want to do
Refresh All
Queries & Connections
Properties
Edit Links
Queries & Connections
Sort
Filter
Clear
Reapply
Advanced
Sort & Filter
Text to Columns
Flash Fill
Remove Duplicates
Data Validation
Consolidate
Relationships
Manage Data Model
Data Tools
GDP
Population
Manage Relationships
Status
Table
Related Lookup Table
Active
imf_Countries (country_code)
CountryRegion (country-code)
Active
IMF_GDP_Population (Country)
imf_Countries (IMF_Country)
New...
Auto-Detect...
Edit...
Activate
Deactivate
Delete
Close
Edit Relationship
Pick the tables and columns you want to use for this relationship
Table:
Data Model Table: IMF_GDP_Population
Column (Foreign):
Country
Related Table:
Data Model Table: imf_Countries
Related Column (Primary):
IMF_Country
Creating relationships between tables is necessary to show related data from different tables on the same report.
OK
Cancel

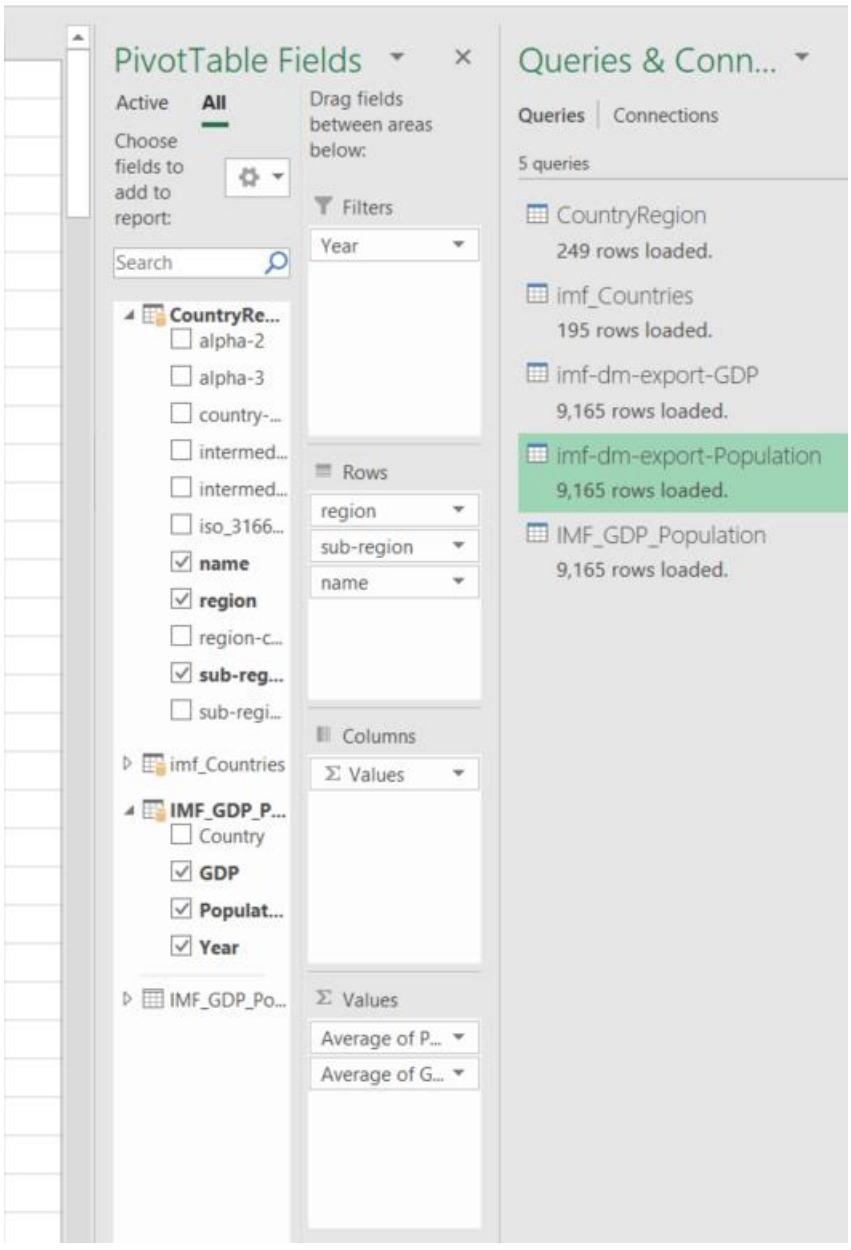


The resultant Pivot Table now has access to all the tables listed above.


Power Pivot for Excel - Solution.xlsx
File
Home
Design
Advanced
Paste
Paste Append
Paste Replace
Copy
Clipboard
From Database
From Data Service
From Other Sources
Existing Connections
Refresh
PivotTable
Get External Dat
Data Type :
Format :
Formatting
Sort A to Z
Sort Z to A
Clear Sort
Clear All Filters
Sort and Filter
PivotTable
PivotChart
Chart and Table (Horizontal)
Chart and Table (Vertical)
Two Charts (Horizontal)
Two Charts (Vertical)
Four Charts
Flattened PivotTable
CountryRegion
name
alpha-2
imf_Count
IMF_Cou

The imf-dm_export_GDP and imf-dem-export-Population tables have been hidden from the client tools





	A	B	C	D	E	F	G	H
		Year	All					
		region	sub-region	name	Average of Population	Average of GDP		
		Africa	Northern Africa		30.36	$74.32		
			Sub-Saharan Africa		15.69	$21.73		
		Americas	Latin America and the Caribbe		15.59	$95.00		
			Northern America		159.74	$6,900.23		
		Asia	Central Asia		12.73	$39.49		
			Eastern Asia		230.68	$1,794.93		
0			South-eastern Asia		51.56	$155.68		
1			Southern Asia		176.02	$235.75		
2			Western Asia		11.59	$108.95		
3		Europe	Eastern Europe		29.57	$215.84		
4			Northern Europe		10.30	$371.29		
5			Southern Europe		12.49	$263.22		
6			Western Europe		26.23	$920.09		
7		Oceania	Australia and New Zealand		12.32	$472.18		
8			Melanesia		1.85	$4.15		
9			Micronesia		0.06	$0.19		
0			Polynesia		0.10	$0.30		
1		Grand Total			34.42	$286.14		
2								
3								
4								
5								

We can now summarise based on Regions and Sub Regions