Power BI Desktop

Query Editor

Khawla Elansari

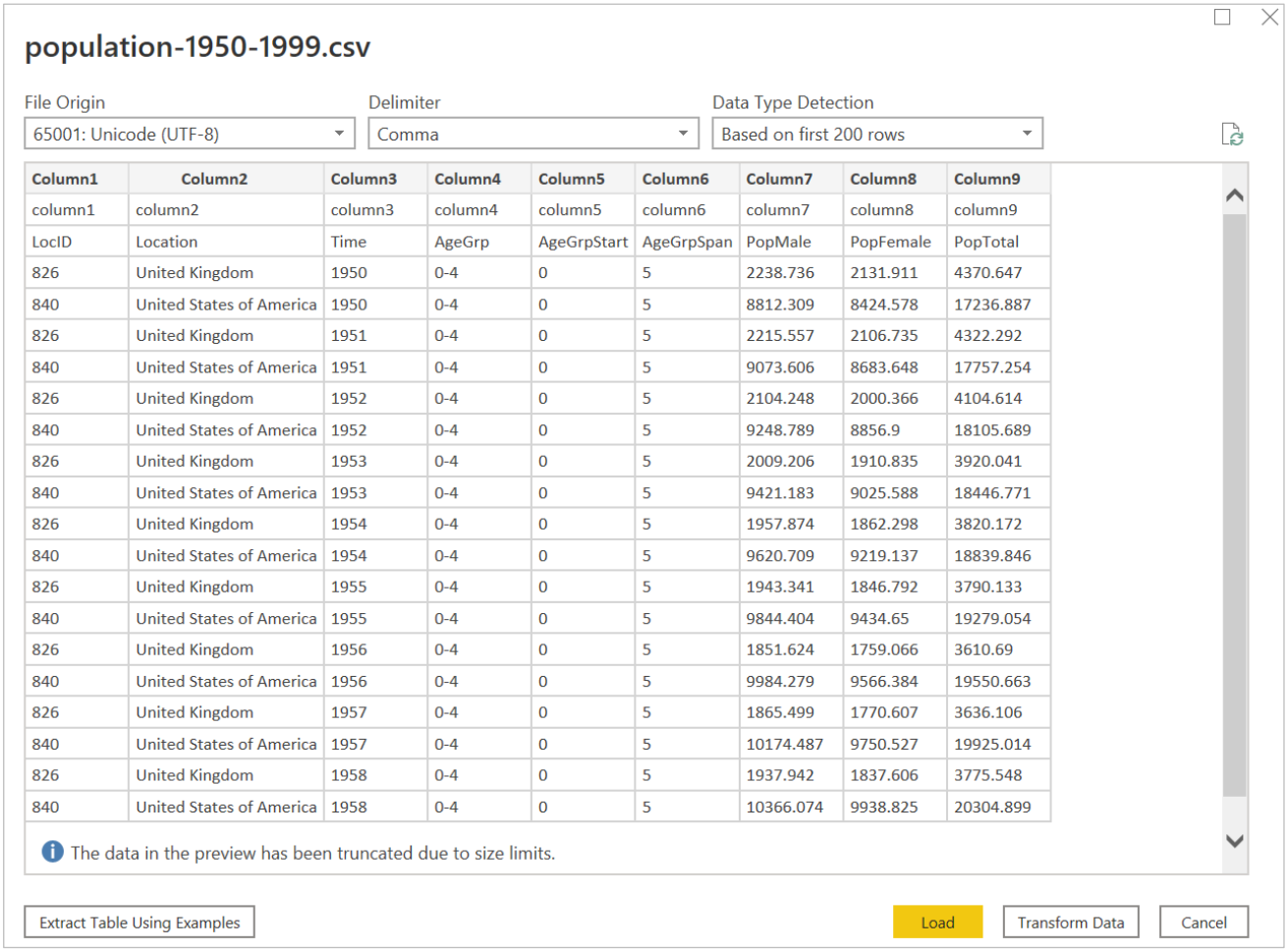
Objectif : analyser la population des USA et UK entre les années 1950 et 2100.

Nous allons utiliser 4 sources de données

- 3 fichiers csv :
  - **Population 1950 - 1999.csv**
  - **Population 2000 - 2049.csv**
  - **Population 2050 - 2100.csv**
- Un fichier Texte : **Codes-Country-Region.txt**

Etape 1 : Chargement des données

1- **Get Data** -> **csv/txt** -> (Sélectionner le 1er fichier)



Column1	Column2	Column3	Column4	Column5	Column6	Column7	Column8	Column9
column1	column2	column3	column4	column5	column6	column7	column8	column9
LocID	Location	Time	AgeGrp	AgeGrpStart	AgeGrpSpan	PopMale	PopFemale	PopTotal
826	United Kingdom	1950	0-4	0	5	2238.736	2131.911	4370.647
840	United States of America	1950	0-4	0	5	8812.309	8424.578	17236.887
826	United Kingdom	1951	0-4	0	5	2215.557	2106.735	4322.292
840	United States of America	1951	0-4	0	5	9073.606	8683.648	17757.254
826	United Kingdom	1952	0-4	0	5	2104.248	2000.366	4104.614
840	United States of America	1952	0-4	0	5	9248.789	8856.9	18105.689
826	United Kingdom	1953	0-4	0	5	2009.206	1910.835	3920.041
840	United States of America	1953	0-4	0	5	9421.183	9025.588	18446.771
826	United Kingdom	1954	0-4	0	5	1957.874	1862.298	3820.172
840	United States of America	1954	0-4	0	5	9620.709	9219.137	18839.846
826	United Kingdom	1955	0-4	0	5	1943.341	1846.792	3790.133
840	United States of America	1955	0-4	0	5	9844.404	9434.65	19279.054
826	United Kingdom	1956	0-4	0	5	1851.624	1759.066	3610.69
840	United States of America	1956	0-4	0	5	9984.279	9566.384	19550.663
826	United Kingdom	1957	0-4	0	5	1865.499	1770.607	3636.106
840	United States of America	1957	0-4	0	5	10174.487	9750.527	19925.014
826	United Kingdom	1958	0-4	0	5	1937.942	1837.606	3775.548
840	United States of America	1958	0-4	0	5	10366.074	9938.825	20304.899

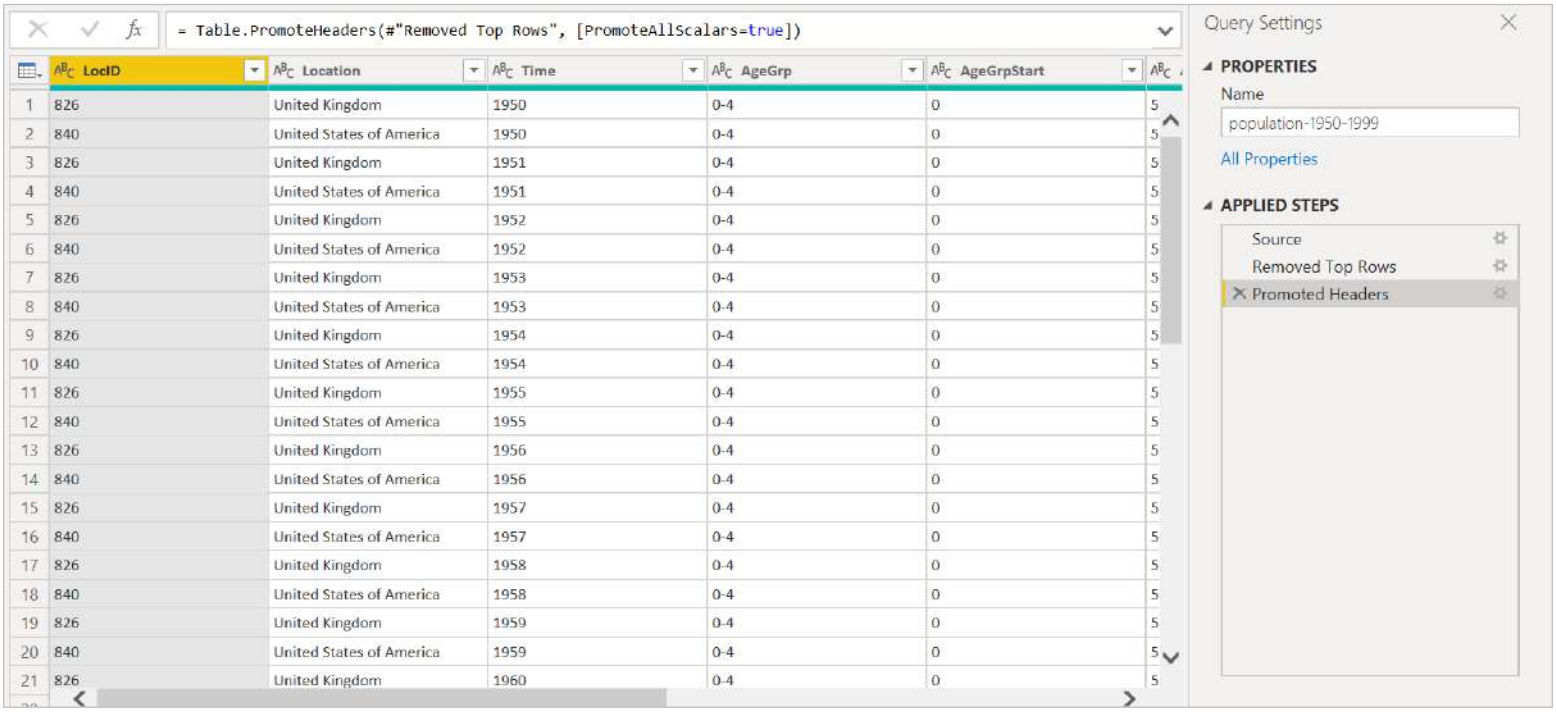
2- Cliquer sur **Load** pour charger les données**.**

3- Répéter les étapes 1 et 2 pour charger les données des 2 autres Fichiers CSV

Etape 2 : Transformation des données

1. Population 1950-1999

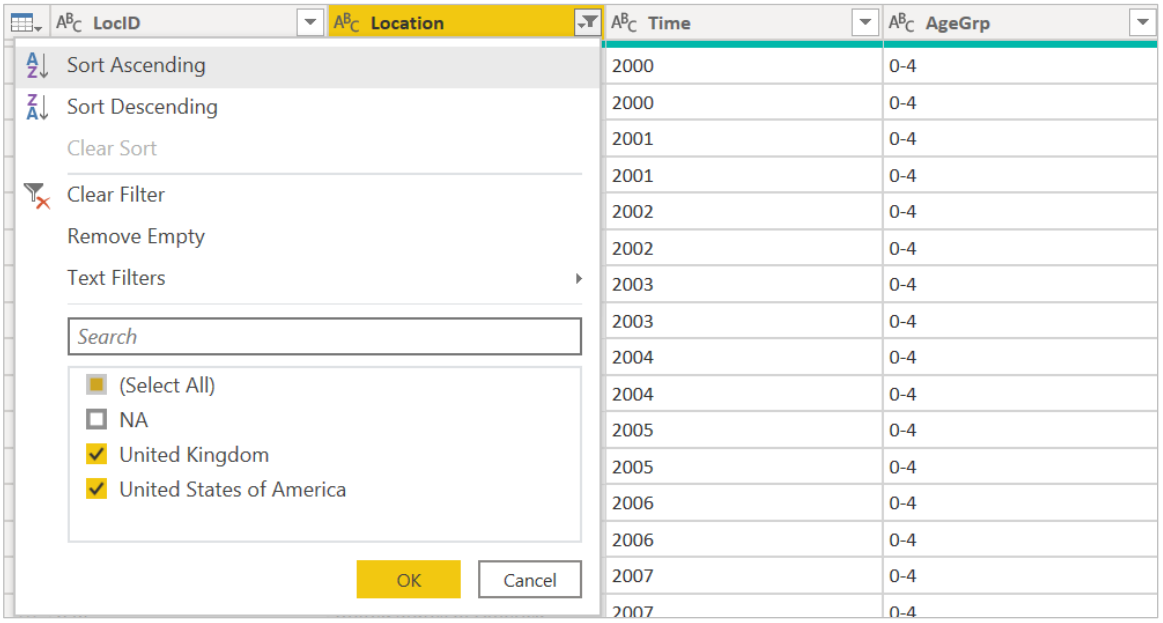
- Supprimer la première ligne: **Home** -> **Remove Rows** -> **Remove Top Rows**
- Définir la première ligne comme en-tête : **Home** -> **Use First Row As Headers**



	LocID	Location	Time	AgeGrp	AgeGrpStart	
1	826	United Kingdom	1950	0-4	0	5
2	840	United States of America	1950	0-4	0	5
3	826	United Kingdom	1951	0-4	0	5
4	840	United States of America	1951	0-4	0	5
5	826	United Kingdom	1952	0-4	0	5
6	840	United States of America	1952	0-4	0	5
7	826	United Kingdom	1953	0-4	0	5
8	840	United States of America	1953	0-4	0	5
9	826	United Kingdom	1954	0-4	0	5
10	840	United States of America	1954	0-4	0	5
11	826	United Kingdom	1955	0-4	0	5
12	840	United States of America	1955	0-4	0	5
13	826	United Kingdom	1956	0-4	0	5
14	840	United States of America	1956	0-4	0	5
15	826	United Kingdom	1957	0-4	0	5
16	840	United States of America	1957	0-4	0	5
17	826	United Kingdom	1958	0-4	0	5
18	840	United States of America	1958	0-4	0	5
19	826	United Kingdom	1959	0-4	0	5
20	840	United States of America	1959	0-4	0	5
21	826	United Kingdom	1960	0-4	0	5

2. Population 2000-2049

- Définir la première ligne comme en-tête : **Home** -> **Use First Row As Headers**
- Filtrer sur la colonne **Location** (Location <> «NA »)



3. Population 2050-2100

- Définir la première ligne comme en-tête : **Home** -> **Use First Row As Headers**
- La nouvelle requête contient des lignes vides : cliquer sur **Remove Empty** pour les supprimer (Ceci va supprimer les valeurs **Null** et **Blank** de la colonne **Location**)

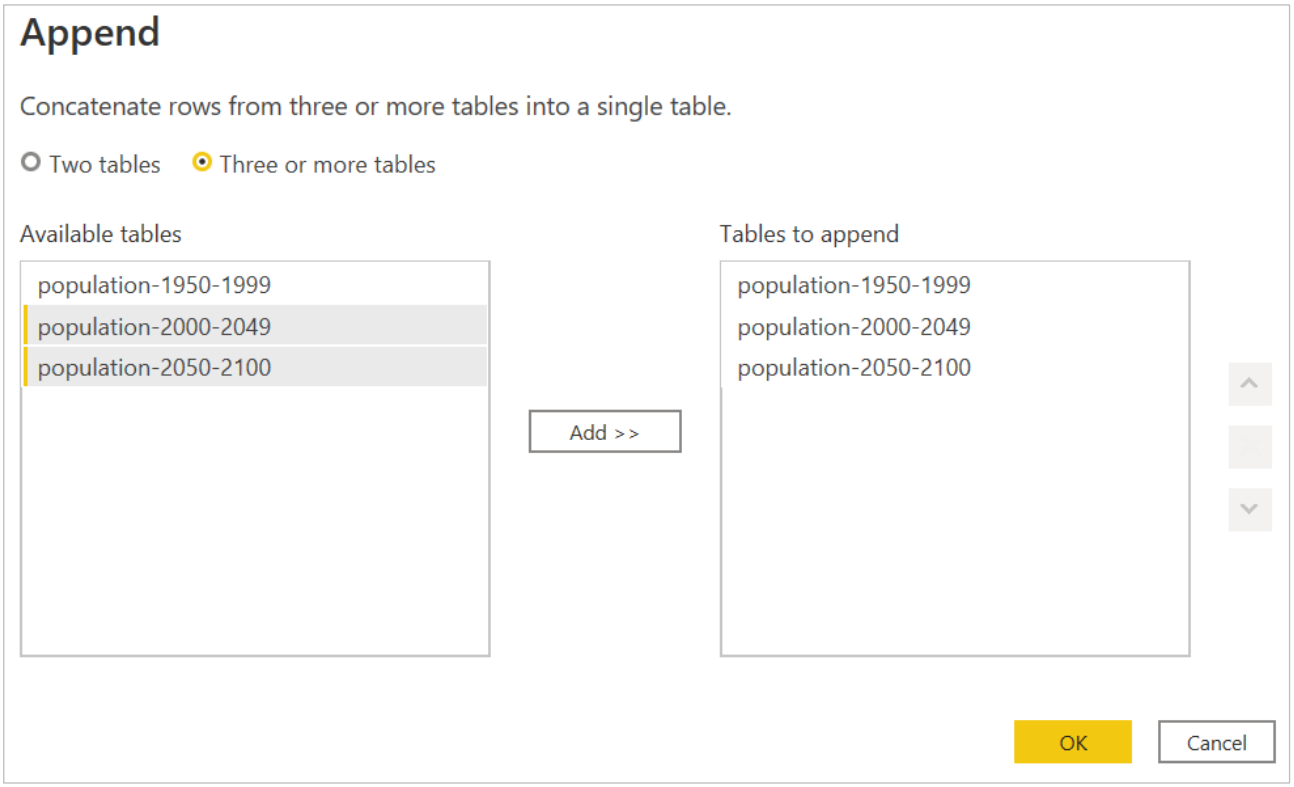
	LocID	Location	Time	AgeGrp	AgeGrpStart
1	826	United Kingdom	2050	0-4	0
2	840	United States of America	2050	0-4	0
3	826	United Kingdom	2051	0-4	0
4	840	United States of America	2051	0-4	0
5	826	United Kingdom	2052	0-4	0
6					
7					
8	840	United States of America	2052	0-4	0
9	826	United Kingdom	2053	0-4	0
10	840	United States of America	2053	0-4	0
11	826	United Kingdom	2054	0-4	0
12	840	United States of America	2054	0-4	0
13	826	United Kingdom	2055	0-4	0

	LocID	Location	Time	AgeGrp	AgeGrpStart
1	826			0-4	0
2	840			0-4	0
3	826			0-4	0
4	840			0-4	0
5	826			0-4	0
6					
7					
8	840	United States of America	2052	0-4	0
9	826	United Kingdom	2053	0-4	0
10	840	United States of America	2053	0-4	0
11	826	United Kingdom	2054	0-4	0
12	840	United States of America	2054	0-4	0

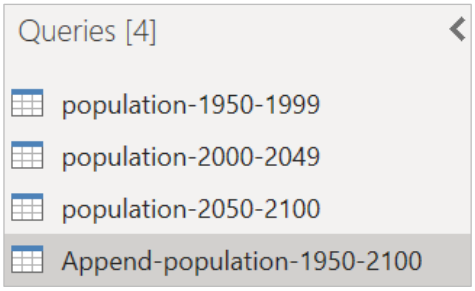
Location
998 (99%) Valid | 0 (0%) Error | 2 (< 1%) Empty
Remove Empty ...

Etape 3 : Création de la requête combinée

1- **Home** -> **Append Queries** -> **Append Queries as New** : Combiner les 3 requêtes précédentes en une seule



2- Une nouvelle requête sera créée : **Append-population-1950 – 2049**



3- Supprimer les colonnes **AgeGrpStart** et **AgeGrpSpan**

4- Renommer les colonnes comme suit :

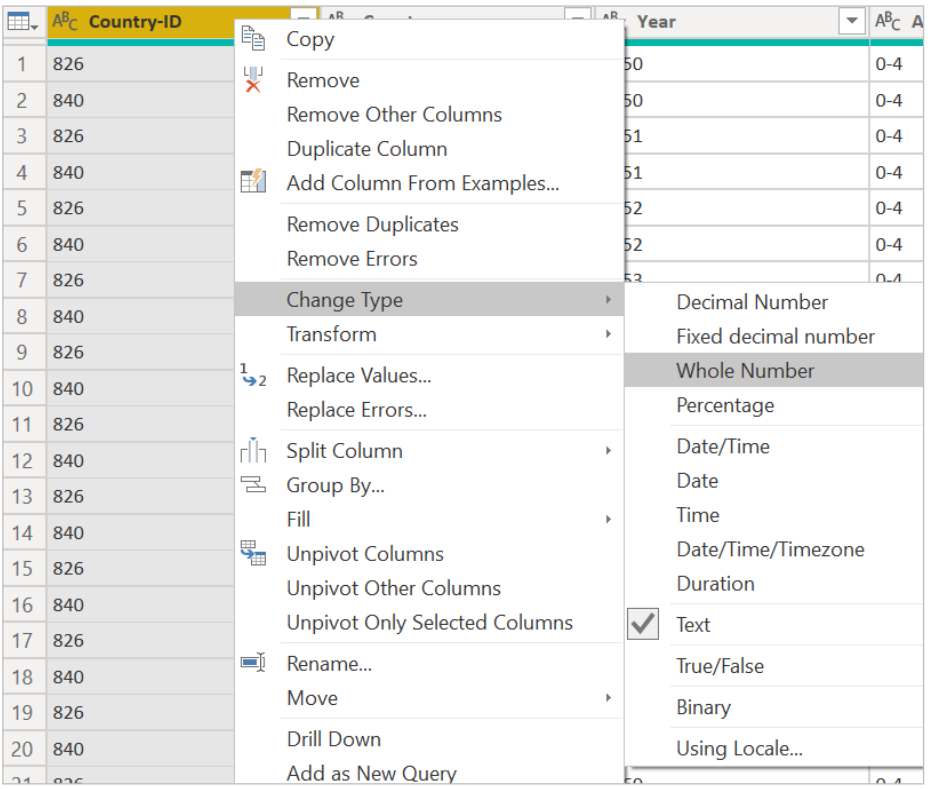
	Country-ID	Country	Year	AgeGrp	PopMale	PopFemale	PopTotal
1	826	United Kingdom	1950	0-4	2238.736	2131.911	4370.647

5- Sur la colonne **AgeGrp** remplacer les valeurs **« 14-Oct »** et **« 9-May »** par **«10-14»** et **«5-9 »** respectivement

	Country-ID	Country	Year	AgeGrp	PopMale	PopFemale
1		United Kingdom	1950	0-4		2131.911
2		United States of America	1950	0-4		8424.578
3		United Kingdom	1951	0-4		2106.735
4		United States of America	1951	0-4		8683.648
5		United Kingdom	1952	0-4		2000.366
6		United States of America	1952	0-4		8856.9
7		United Kingdom	1953	0-4		1910.835
8		United States of America	1953	0-4		9025.588
9		United Kingdom	1954	0-4		1862.298
10		United States of America	1954	0-4		9219.137
11		United Kingdom	1955	0-4		1846.792
12		United States of America	1955	0-4		9434.65
13		United Kingdom	1956	0-4		1759.066

6- Changer les types de données :

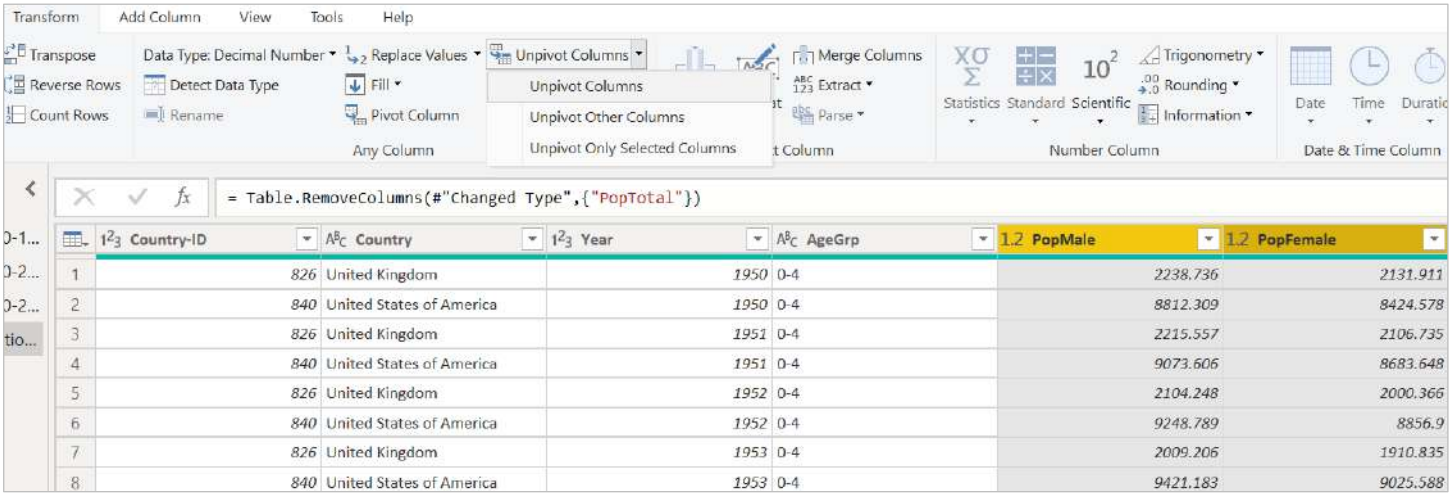
- **Country-ID**, **Year:** Whole Number
- **PopMale**, **PopFemale**: Decimal Number



7- Cliquer sur **Apply**

8- Supprimer la colonne **PopTotal**

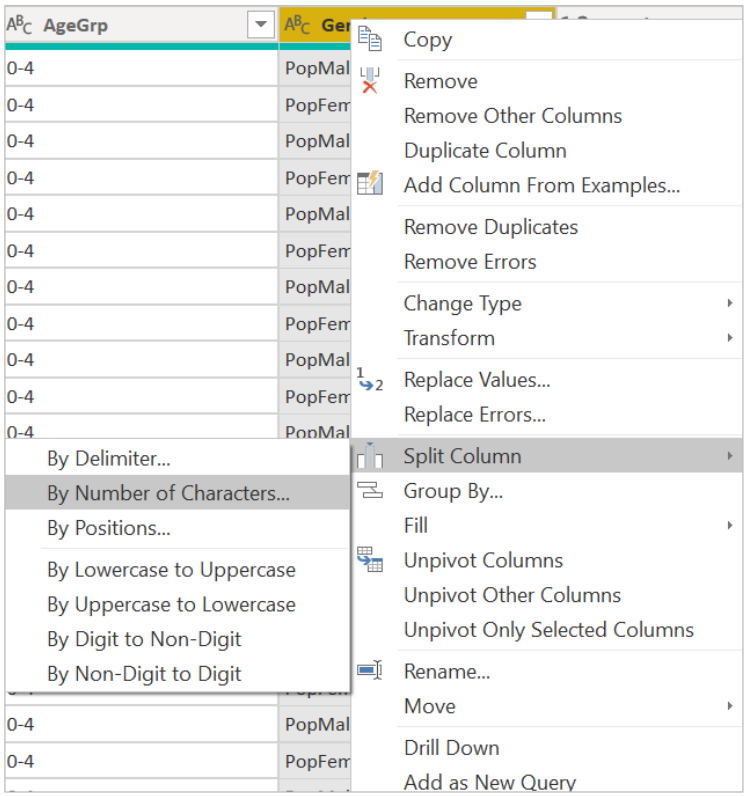
9- Appliquer le **Unpivot Columns** sur les colonnes **PopMale** et **PopFemale**

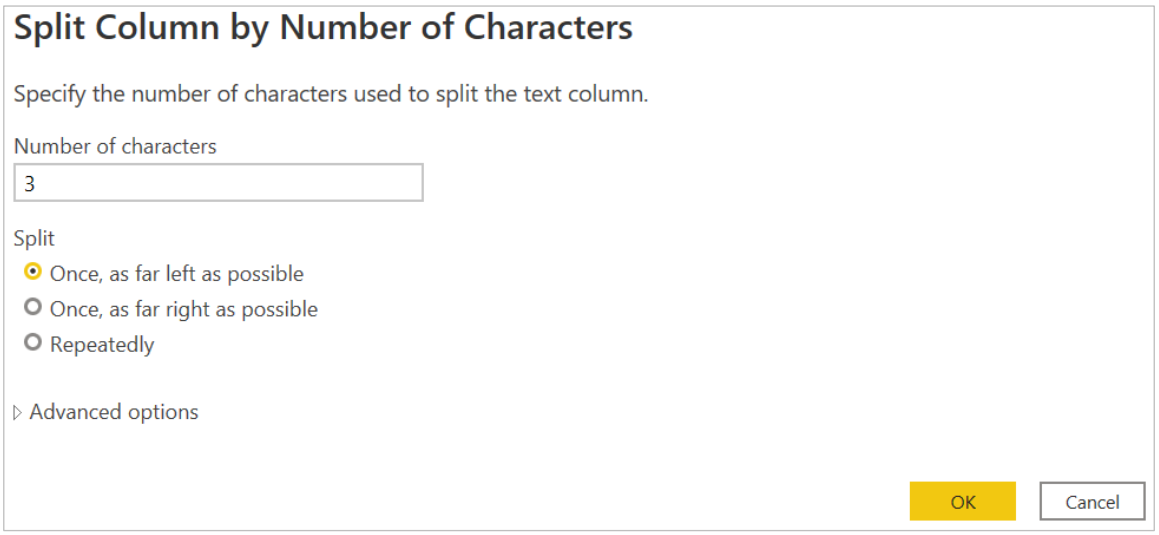


10- Renommer les colonnes (Attribute, Value) comme suit :

Year	AgeGrp	Gender	Population
1950	0-4	PopMale	2238.736
1950	0-4	PopFemale	2131.911
1950	0-4	PopMale	8812.309
1950	0-4	PopFemale	8424.578
1951	0-4	PopMale	2215.557
1951	0-4	PopFemale	2106.735
1951	0-4	PopMale	9073.606
1951	0-4	PopFemale	8683.648
1952	0-4	PopMale	2104.248

11- Appliquer un **Split** Column à la colonne **Gender** pour retourner le gender uniquement (Omettre le Pop)

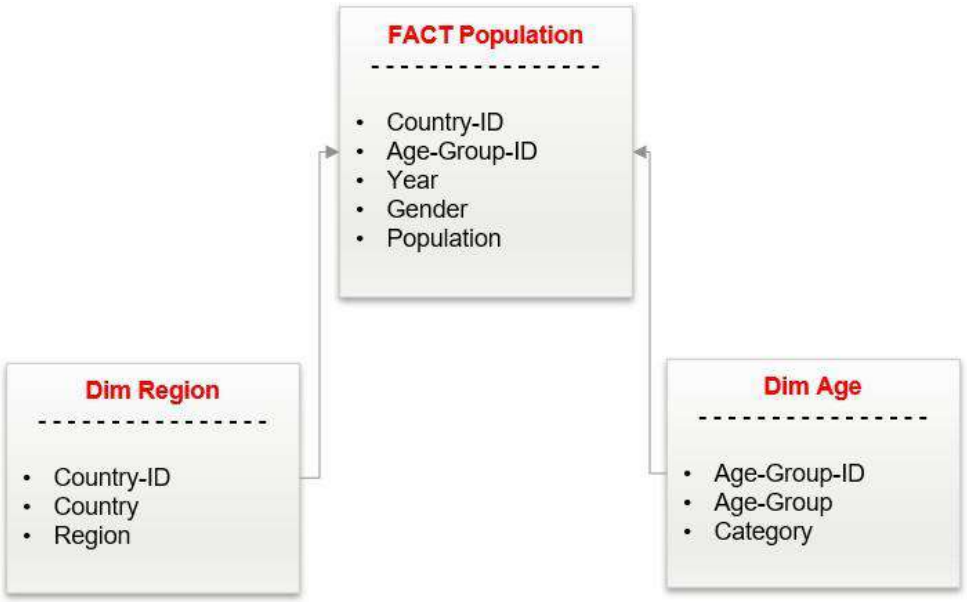




12- Supprimer la colonne **Gender.1** et renommer la colonne **Gender.2** en **Gender**

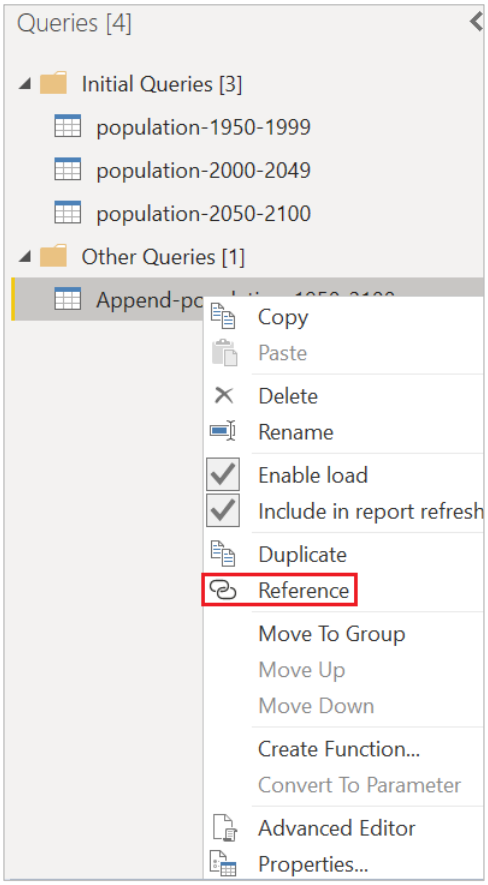
13- Créer un nouveau groupe de requêtes, le nommer « **Initial Queries** » et y ajouter les4 requêtes qui sont déjà créées

Etape 4 : Création du Data Model



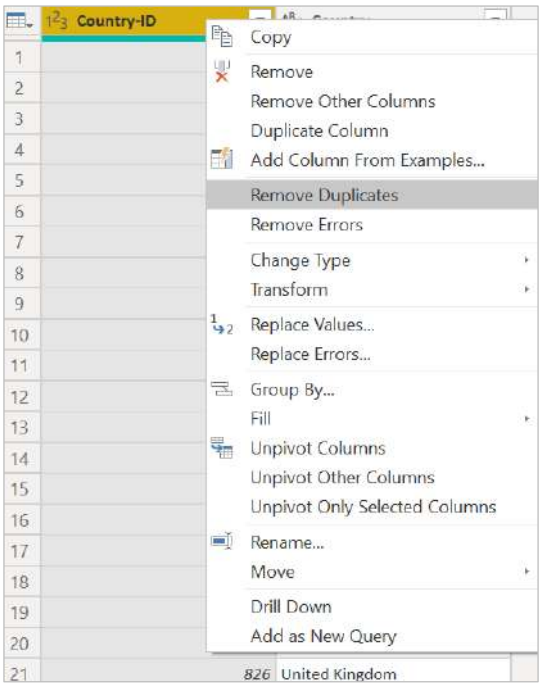
I. Dim Region

1- Créer une **référence** à la requête **Append-Population-1950-2100** et la renommer **« Dim Region »**

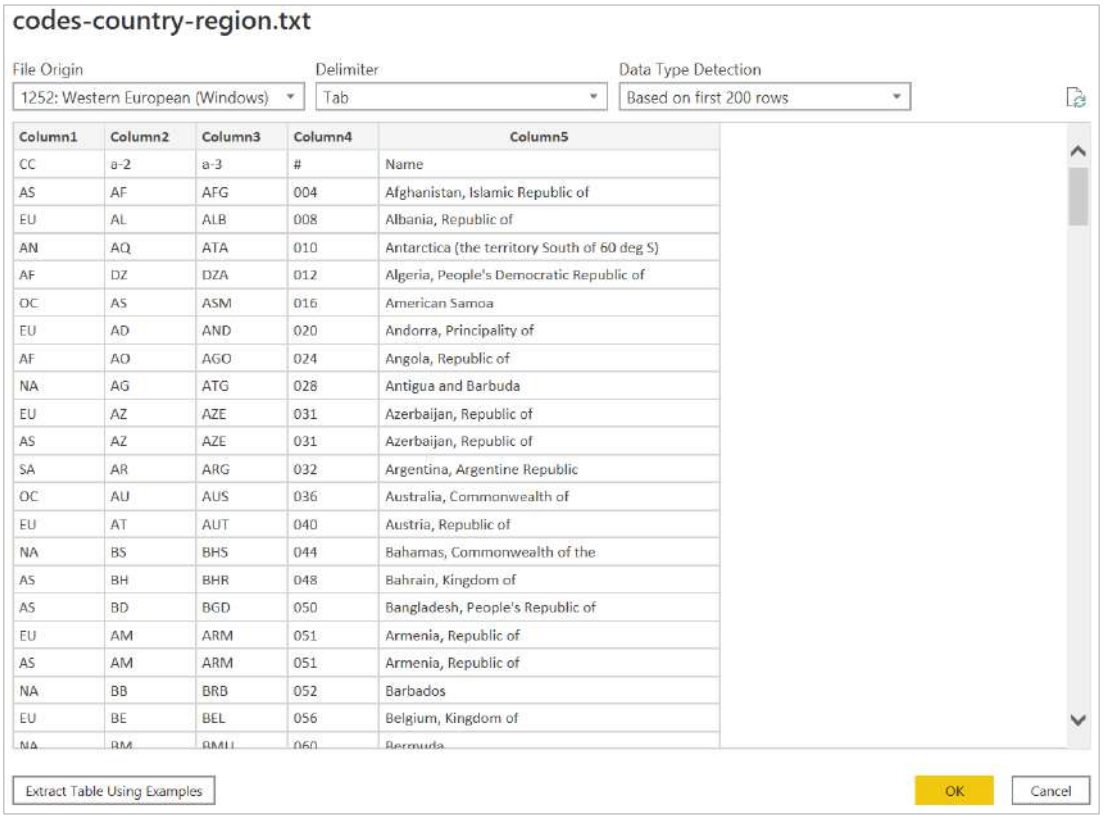


2- Supprimer toutes les colonnes sauf **Country-ID** et **Country**

3- Supprimer les doublons de la colonne **Country**



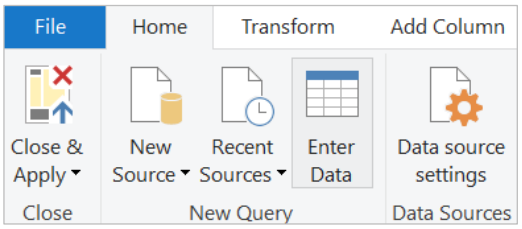
4- Charger les données du fichier **Codes-Country-Region.txt**

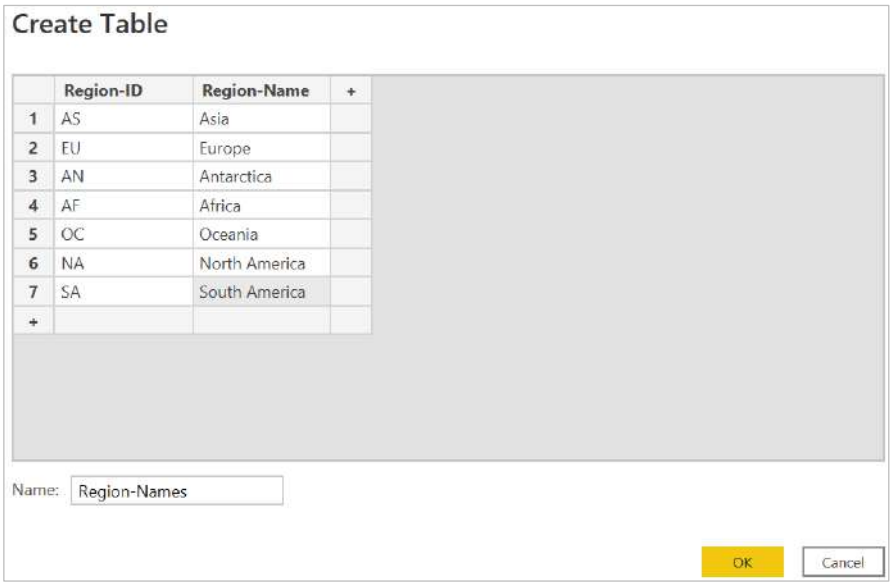


Column1	Column2	Column3	Column4	Column5
CC	a-2	a-3	#	Name
AS	AF	AFG	004	Afghanistan, Islamic Republic of
EU	AL	ALB	008	Albania, Republic of
AN	AQ	ATA	010	Antarctica (the territory South of 60 deg S)
AF	DZ	DZA	012	Algeria, People's Democratic Republic of
OC	AS	ASM	016	American Samoa
EU	AD	AND	020	Andorra, Principality of
AF	AO	AGO	024	Angola, Republic of
NA	AG	ATG	028	Antigua and Barbuda
EU	AZ	AZE	031	Azerbaijan, Republic of
AS	AZ	AZE	031	Azerbaijan, Republic of
SA	AR	ARG	032	Argentina, Argentine Republic
OC	AU	AUS	036	Australia, Commonwealth of
EU	AT	AUT	040	Austria, Republic of
NA	BS	BHS	044	Bahamas, Commonwealth of the
AS	BH	BHR	048	Bahrain, Kingdom of
AS	BD	BGD	050	Bangladesh, People's Republic of
EU	AM	ARM	051	Armenia, Republic of
AS	AM	ARM	051	Armenia, Republic of
NA	BB	BRB	052	Barbados
EU	BE	BEL	056	Belgium, Kingdom of
NA	BM	BMU	060	Bermuda

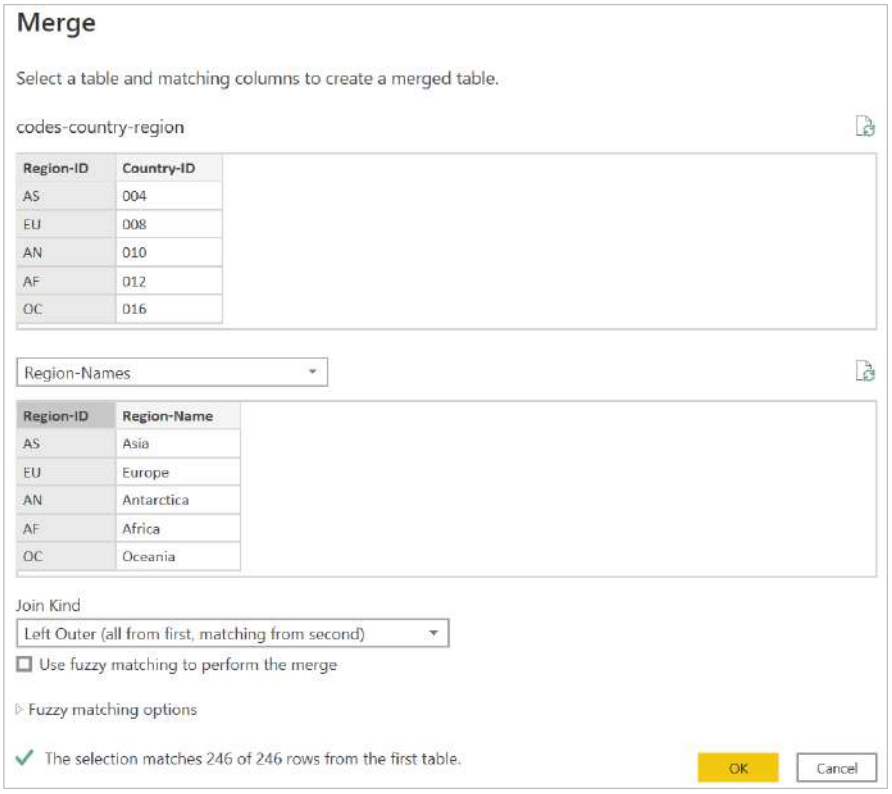
- Supprimer toutes les colonnes sauf **CC**, **#**
- Renommer la colonne **CC** en **Region-ID** et **#** en **Country-ID**
- Changer le type de données de **Country-ID** en Whole Number

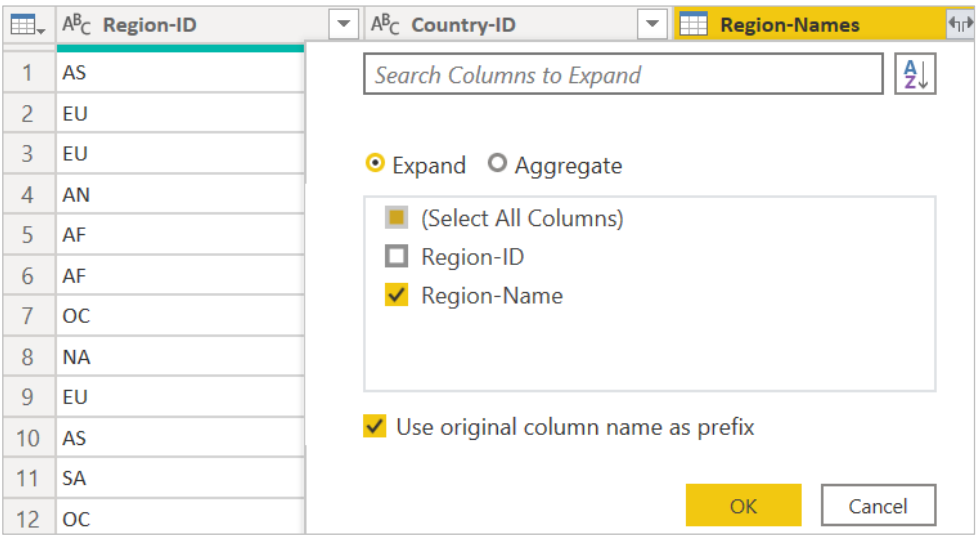
5- Créer une nouvelle **table manuellement** et la nommer **« Region-Names »**





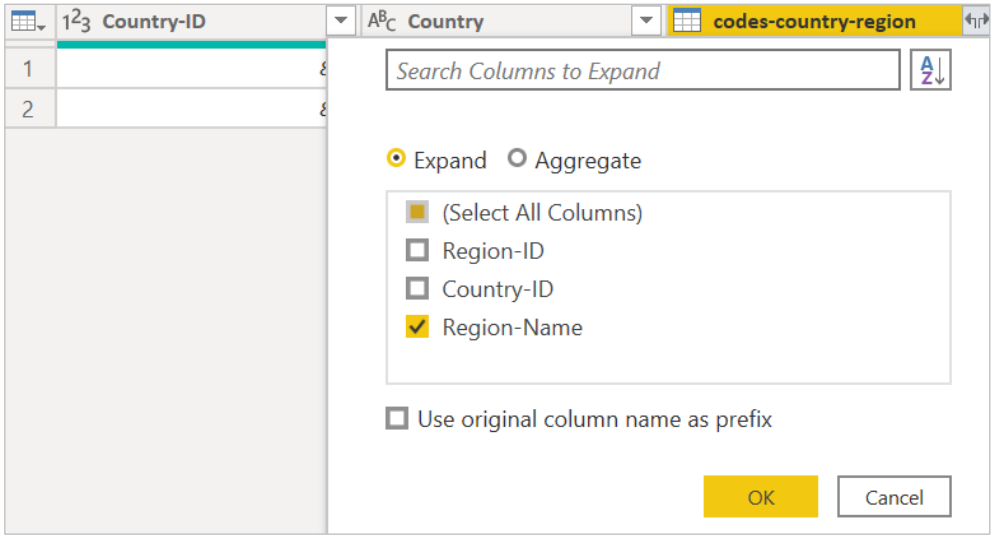
6- Effectuer un Merge entre **Codes-Country-Region** et **Region-Names**





7- Changer le type de la colonne **Country-ID** en **Whole Number**

8- Effectuer un **Left Outer join** entre **Dim-Region** et **codes-country-region**



Dim-Region :

	Country-ID	Country	Region-Name
1	826	United Kingdom	Europe
2	840	United States of America	North America

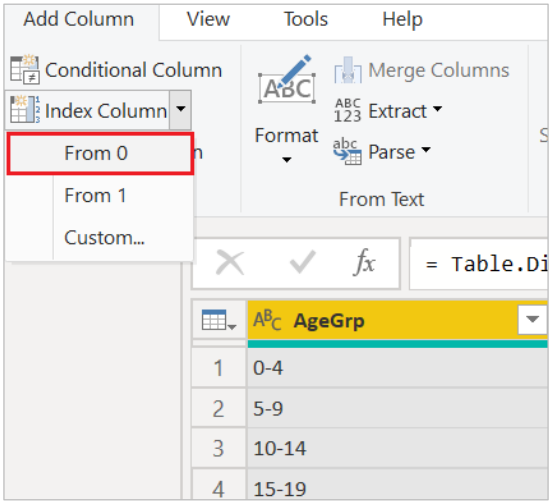
II. Dim Age

1- Créer une **référence** à la requête **Append-Population-1950-2100** et la nommer **« Dim Age »**

2- Supprimer toutes les colonnes sauf **AgeGrp**

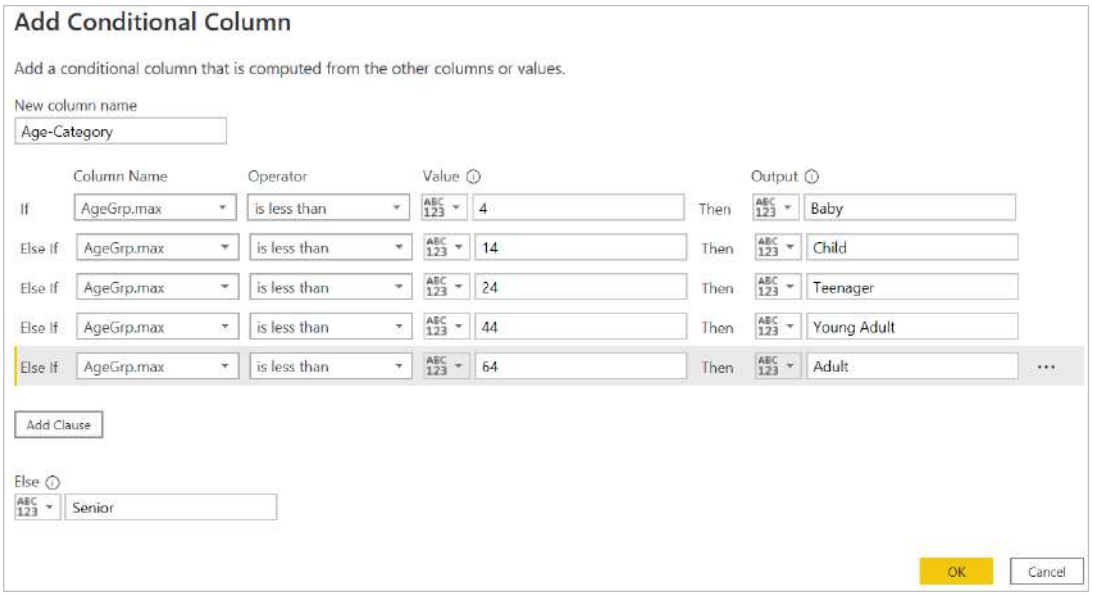
3- Supprimer les duplicates de la colonne **AgeGrp**

4- Ajouter une colonne Index, la nommer « **Age-ID** »



5- Créer une colonne **AgeGrp.max** qui retourne l'âge maximal pour chaque **AgeGrp**

6- Ajouter une colonne conditionnelle Age-Category définie comme suit :



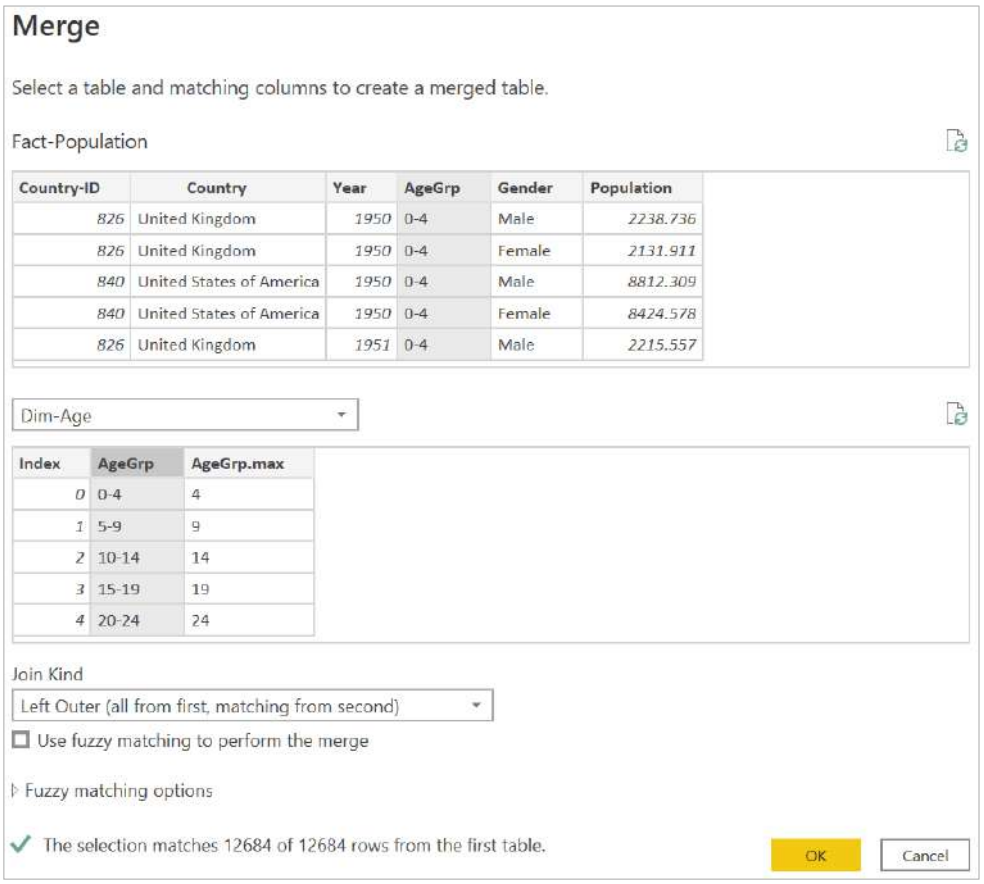
Dim-Age:

	Age-ID	AgeGrp	AgeGrp.max	Age-Category
1	0	0-4	4	Child
2	1	5-9	9	Child
3	2	10-14	14	Teenager
4	3	15-19	19	Teenager
5	4	20-24	24	Young Adult
6	5	25-29	29	Young Adult
7	6	30-34	34	Young Adult
8	7	35-39	39	Young Adult
9	8	40-44	44	Adult
10	9	45-49	49	Adult
11	10	50-54	54	Adult
12	11	55-59	59	Adult
13	12	60-64	64	Senior
14	13	65-69	69	Senior
15	14	70-74	74	Senior
16	15	75-79	79	Senior
17	16	80-84	84	Senior
18	17	85-89	89	Senior
19	18	90-94	94	Senior
20	19	95-99	99	Senior
21	20	100+	150	Senior

III. Fact Population

1- Créer une **référence** à la requête **Append-Population-1950-2100** et la nommer **« Fact Population »**

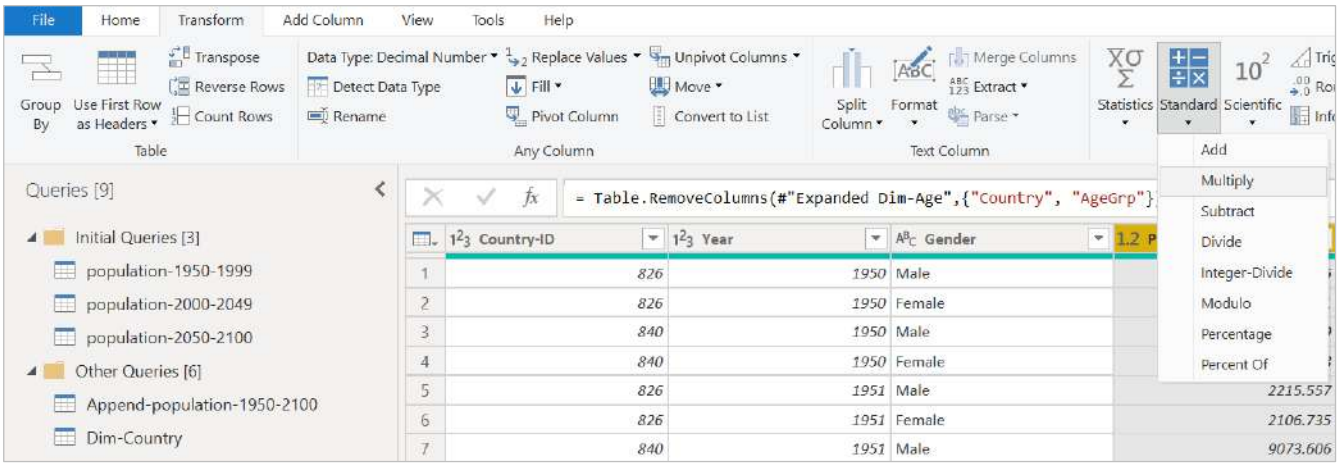
2- Effectuer un **left Outer Join** (Merge) avec **la Dim Age**



3- Ajouter le **Age-ID** à la requête

4- Supprimer les colonnes **Country** et **AgeGrp**

5- Multiplier la mesure **Population** par 1000 pour qu'elle retourne les valeurs correctes



Fact-Population :

	Country-ID	Year	Gender	Population	Age-ID
1	826	1950	Male	2238736	0
2	826	1950	Female	2131911	0
3	840	1950	Male	8812309	0
4	840	1950	Female	8424578	0
5	826	1951	Male	2215557	0
6	826	1951	Female	2106735	0
7	840	1951	Male	9073606	0
8	840	1951	Female	8683648	0
9	826	1952	Male	2104248	0
10	826	1952	Female	2000366	0
11	840	1952	Male	9248789	0
12	840	1952	Female	8856900	0
13	826	1953	Male	2009206	0
14	826	1953	Female	1910835	0
15	840	1953	Male	9421183	0
16	840	1953	Female	9025588	0
17	826	1954	Male	1957874	0
18	826	1954	Female	1862298	0
19	840	1954	Male	9620709	0