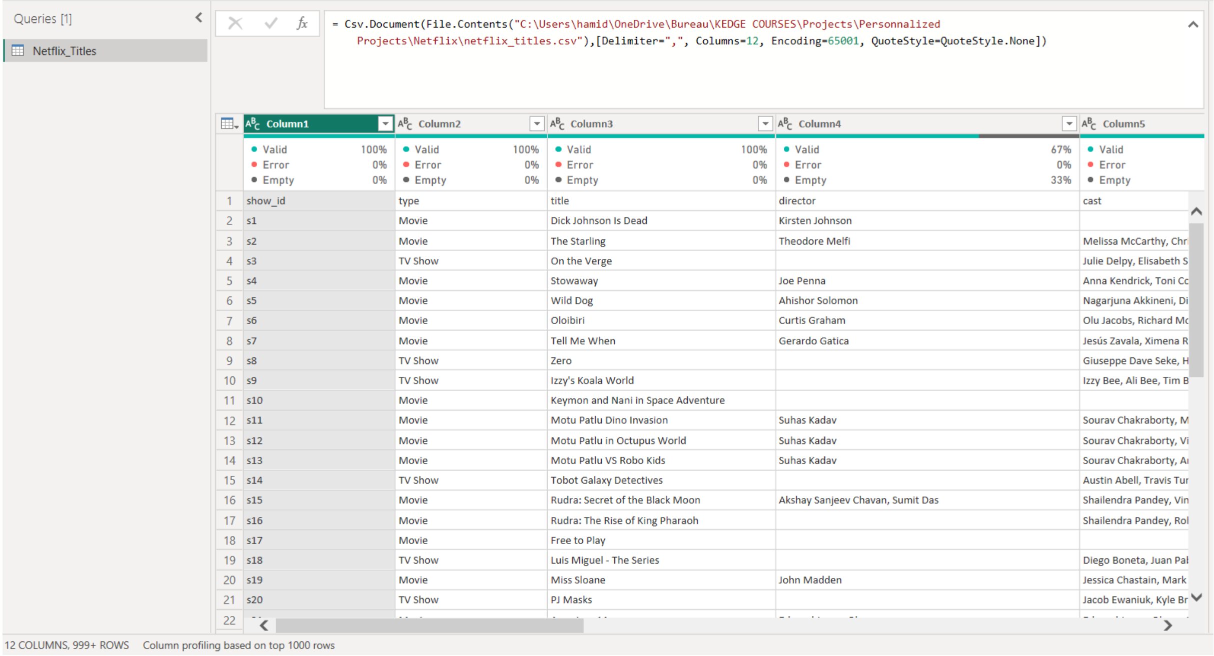
NETFLIX CLEANING AND DATA ANALYTICS

Power BI

1. DATA CLEANING AND PRE-PROCESSING

A- LOAD DATA

First, we must loads the csv in Power BI.



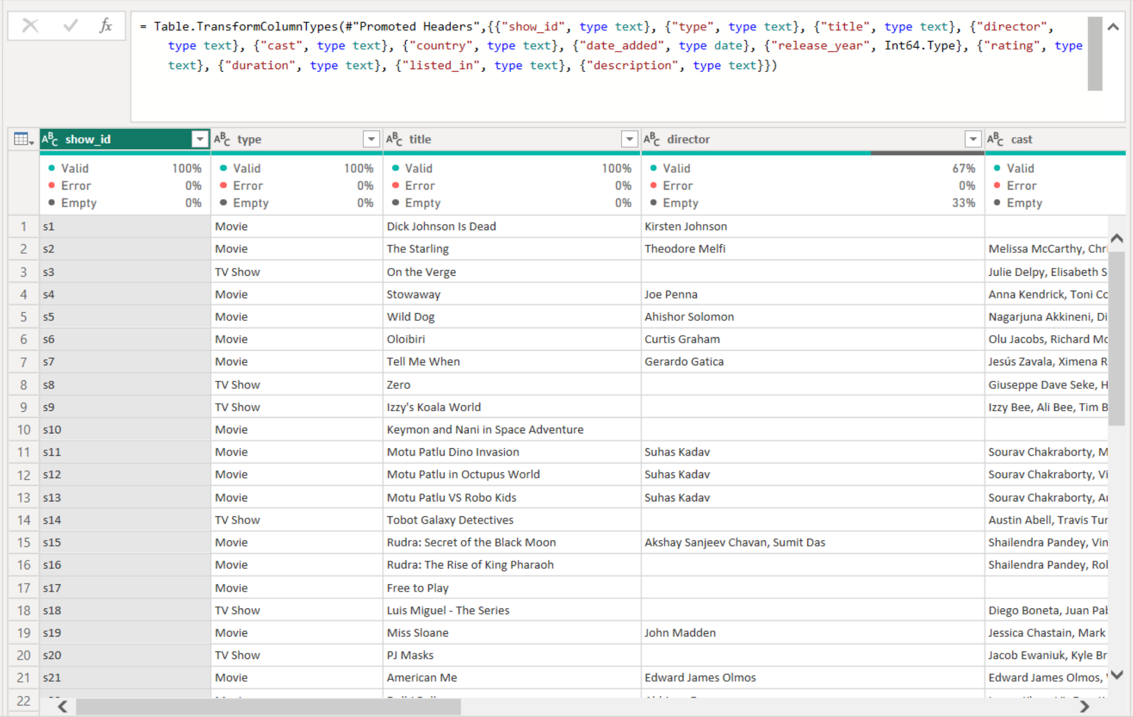
```
= Csv.Document(File.Contents("C:\Users\hamid\OneDrive\Bureau\KEDGE COURSES\Projects\Personnalized
   Projects\Netflix\netflix_titles.csv"),[Delimiter=",", Columns=12, Encoding=65001, QuoteStyle=QuoteStyle.None])
```

	Column1	Column2	Column3	Column4	Column5
	● Valid 100% ● Error 0% ● Empty 0%	● Valid 100% ● Error 0% ● Empty 0%	● Valid 100% ● Error 0% ● Empty 0%	● Valid 67% ● Error 0% ● Empty 33%	● Valid ● Error ● Empty
1	show_id	type	title	director	cast
2	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	
3	s2	Movie	The Starling	Theodore Melfi	Melissa McCarthy, Chr
4	s3	TV Show	On the Verge		Julie Delpy, Elisabeth S
5	s4	Movie	Stowaway	Joe Penna	Anna Kendrick, Toni Cc
6	s5	Movie	Wild Dog	Ahishor Solomon	Nagarjuna Akkineni, Di
7	s6	Movie	Oloibiri	Curtis Graham	Olu Jacobs, Richard Mc
8	s7	Movie	Tell Me When	Gerardo Gatica	Jesús Zavala, Ximena R
9	s8	TV Show	Zero		Giuseppe Dave Seke, H
10	s9	TV Show	Izzy's Koala World		Izzy Bee, Ali Bee, Tim B
11	s10	Movie	Keymon and Nani in Space Adventure		
12	s11	Movie	Motu Patlu Dino Invasion	Suhas Kadav	Sourav Chakraborty, M
13	s12	Movie	Motu Patlu in Octupus World	Suhas Kadav	Sourav Chakraborty, Vi
14	s13	Movie	Motu Patlu VS Robo Kids	Suhas Kadav	Sourav Chakraborty, Ai
15	s14	TV Show	Tobot Galaxy Detectives		Austin Abell, Travis Tur
16	s15	Movie	Rudra: Secret of the Black Moon	Akshay Sanjeev Chavan, Sumit Das	Shailendra Pandey, Vin
17	s16	Movie	Rudra: The Rise of King Pharaoh		Shailendra Pandey, Rol
18	s17	Movie	Free to Play		
19	s18	TV Show	Luis Miguel - The Series		Diego Boneta, Juan Pal
20	s19	Movie	Miss Sloane	John Madden	Jessica Chastain, Mark
21	s20	TV Show	PJ Masks		Jacob Ewaniuk, Kyle Br

12 COLUMNS, 999+ ROWS   Column profiling based on top 1000 rows

B- COLUMNS NAMES AND DATA TYPES

Let’s now promote the columns header and check for all the data types:

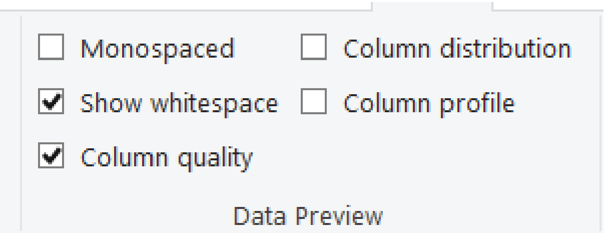


```
= Table.TransformColumnTypes(#"Promoted Headers",{{"show_id", type text}, {"type", type text}, {"title", type text}, {"director",
    type text}, {"cast", type text}, {"country", type text}, {"date_added", type date}, {"release_year", Int64.Type}, {"rating", type
    text}, {"duration", type text}, {"listed_in", type text}, {"description", type text}})
```

	show_id	type	title	director	cast
	● Valid 100% ● Error 0% ● Empty 0%	● Valid 100% ● Error 0% ● Empty 0%	● Valid 100% ● Error 0% ● Empty 0%	● Valid 67% ● Error 0% ● Empty 33%	● Valid ● Error ● Empty
1	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	
2	s2	Movie	The Starling	Theodore Melfi	Melissa McCarthy, Chr
3	s3	TV Show	On the Verge		Julie Delpy, Elisabeth S
4	s4	Movie	Stowaway	Joe Penna	Anna Kendrick, Toni Cc
5	s5	Movie	Wild Dog	Ahishor Solomon	Nagarjuna Akkineni, Di
6	s6	Movie	Oloibiri	Curtis Graham	Olu Jacobs, Richard Mc
7	s7	Movie	Tell Me When	Gerardo Gatica	Jesús Zavala, Ximena R
8	s8	TV Show	Zero		Giuseppe Dave Seke, H
9	s9	TV Show	Izzy's Koala World		Izzy Bee, Ali Bee, Tim B
10	s10	Movie	Keymon and Nani in Space Adventure		
11	s11	Movie	Motu Patlu Dino Invasion	Suhas Kadav	Sourav Chakraborty, M
12	s12	Movie	Motu Patlu in Octupus World	Suhas Kadav	Sourav Chakraborty, Vi
13	s13	Movie	Motu Patlu VS Robo Kids	Suhas Kadav	Sourav Chakraborty, Ai
14	s14	TV Show	Tobot Galaxy Detectives		Austin Abell, Travis Tur
15	s15	Movie	Rudra: Secret of the Black Moon	Akshay Sanjeev Chavan, Sumit Das	Shailendra Pandey, Vin
16	s16	Movie	Rudra: The Rise of King Pharaoh		Shailendra Pandey, Rol
17	s17	Movie	Free to Play		
18	s18	TV Show	Luis Miguel - The Series		Diego Boneta, Juan Pal
19	s19	Movie	Miss Sloane	John Madden	Jessica Chastain, Mark
20	s20	TV Show	PJ Masks		Jacob Ewaniuk, Kyle Br
21	s21	Movie	American Me	Edward James Olmos	Edward James Olmos, '

C -NULL VALUE

We must fill the null value from each column with column quality:



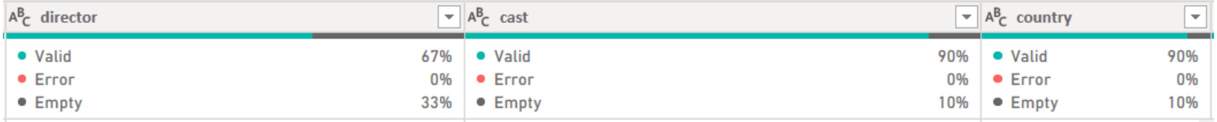
☐ Monospaced ☐ Column distribution

☑ Show whitespace ☐ Column profile

☑ Column quality

Data Preview

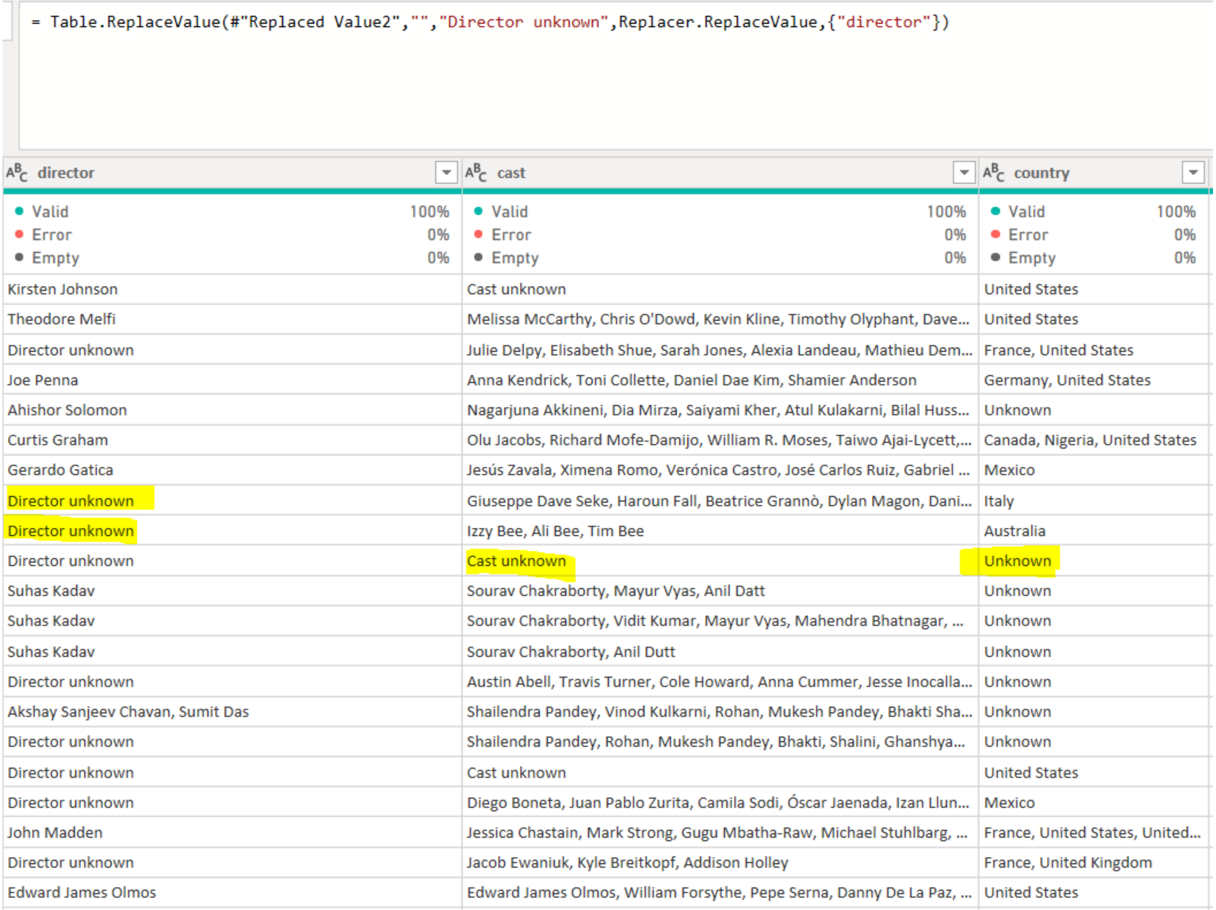
We can now see the columns with empty value:



director		cast		country	
● Valid	67%	● Valid	90%	● Valid	90%
● Error	0%	● Error	0%	● Error	0%
● Empty	33%	● Empty	10%	● Empty	10%

Go to Transform and "Replaces Values to change the empty values:

Do this for all the columns:



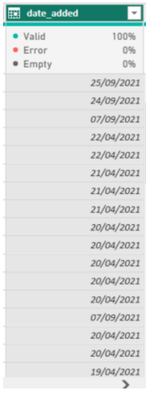
```
= Table.ReplaceValue(#"Replaced Value2","","Director unknown",Replacer.ReplaceValue,{"director"})
```

director		cast		country	
● Valid	100%	● Valid	100%	● Valid	100%
● Error	0%	● Error	0%	● Error	0%
● Empty	0%	● Empty	0%	● Empty	0%
Kirsten Johnson		Cast unknown		United States	
Theodore Melfi		Melissa McCarthy, Chris O'Dowd, Kevin Kline, Timothy Olyphant, Dave...		United States	
Director unknown		Julie Delpy, Elisabeth Shue, Sarah Jones, Alexia Landeau, Mathieu Dem...		France, United States	
Joe Penna		Anna Kendrick, Toni Collette, Daniel Dae Kim, Shamier Anderson		Germany, United States	
Ahishor Solomon		Nagarjuna Akkineni, Dia Mirza, Saiyami Kher, Atul Kulakarni, Bilal Huss...		Unknown	
Curtis Graham		Olu Jacobs, Richard Mofe-Damijo, William R. Moses, Taiwo Ajai-Lycett,...		Canada, Nigeria, United States	
Gerardo Gatica		Jesús Zavala, Ximena Romo, Verónica Castro, José Carlos Ruiz, Gabriel ...		Mexico	
Director unknown		Giuseppe Dave Seke, Haroun Fall, Beatrice Grannò, Dylan Magon, Dani...		Italy	
Director unknown		Izzy Bee, Ali Bee, Tim Bee		Australia	
Director unknown		Cast unknown		Unknown	
Suhas Kadav		Sourav Chakraborty, Mayur Vyas, Anil Datt		Unknown	
Suhas Kadav		Sourav Chakraborty, Vidit Kumar, Mayur Vyas, Mahendra Bhatnagar, ...		Unknown	
Suhas Kadav		Sourav Chakraborty, Anil Dutt		Unknown	
Director unknown		Austin Abell, Travis Turner, Cole Howard, Anna Cummer, Jesse Inocalla...		Unknown	
Akshay Sanjeev Chavan, Sumit Das		Shailendra Pandey, Vinod Kulkarni, Rohan, Mukesh Pandey, Bhakti Sha...		Unknown	
Director unknown		Shailendra Pandey, Rohan, Mukesh Pandey, Bhakti, Shalini, Ghanshya...		Unknown	
Director unknown		Cast unknown		United States	
Director unknown		Diego Boneta, Juan Pablo Zurita, Camila Sodi, Óscar Jaenada, Izan Llun...		Mexico	
John Madden		Jessica Chastain, Mark Strong, Gugu Mbatha-Raw, Michael Stuhlbarg, ...		France, United States, United...	
Director unknown		Jacob Ewaniuk, Kyle Breitkopf, Addison Holley		France, United Kingdom	
Edward James Olmos		Edward James Olmos, William Forsythe, Pepe Serna, Danny De La Paz, ...		United States	

D-DATES (YEAR)

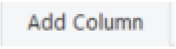
Let’s extract the Year of date_added:

- ***Click on the column date_added:***

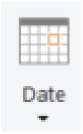


date_added	
● Valid	100%
● Error	0%
● Empty	0%
	25/09/2021
	24/09/2021
	07/09/2021
	22/04/2021
	22/04/2021
	21/04/2021
	21/04/2021
	21/04/2021
	20/04/2021
	20/04/2021
	20/04/2021
	20/04/2021
	20/04/2021
	07/09/2021
	20/04/2021
	20/04/2021
	19/04/2021

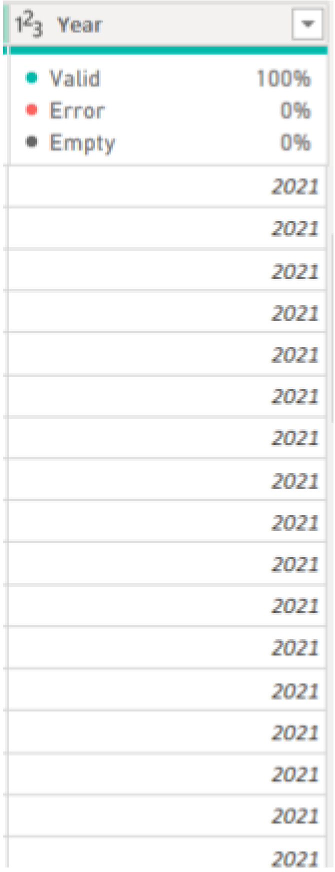
- ***Go to Add column in the ribbon***



- ***Go to From Date & Time then click on Date***



- ***And select Year:***

- 

123 Year

● Valid 100%

● Error 0%

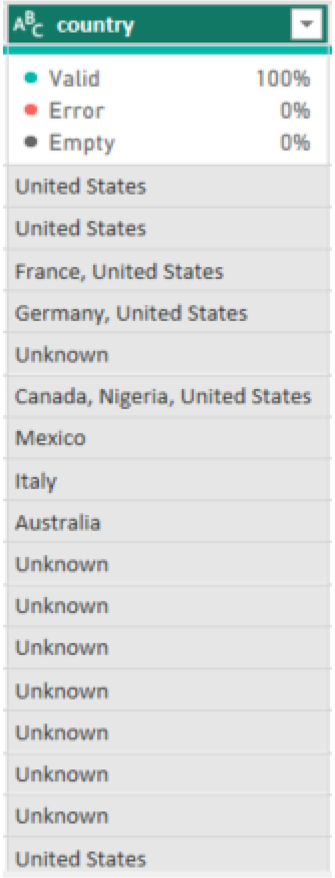
● Empty 0%

Year
2021
2021
2021
2021
2021
2021
2021
2021
2021
2021
2021
2021
2021
2021
2021
2021
2021

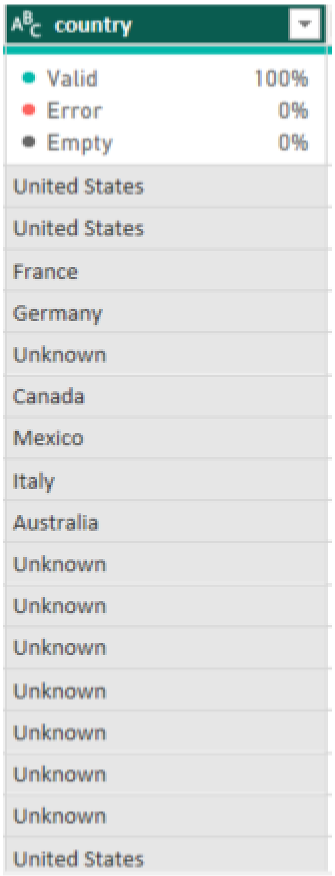
E- STANDARDIZATION

Extract the first country before the comma:

For this, go to Transfom and then Extract : ABC 123 Extract ▾
And select "," for the Delimiter:



country	
● Valid	100%
● Error	0%
● Empty	0%
United States	
United States	
France, United States	
Germany, United States	
Unknown	
Canada, Nigeria, United States	
Mexico	
Italy	
Australia	
Unknown	
Unknown	
Unknown	
Unknown	
Unknown	
Unknown	
Unknown	
United States	

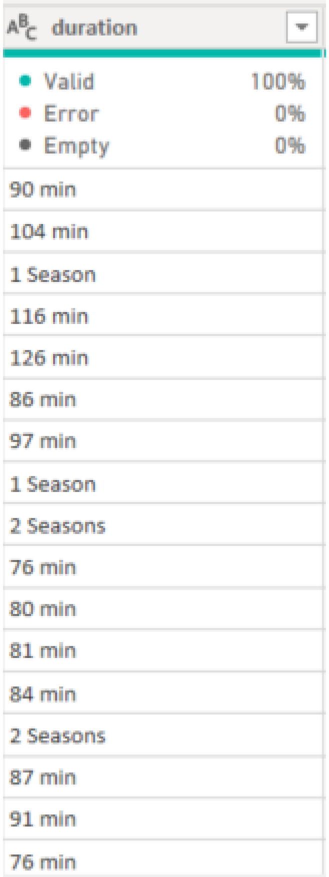


country	
● Valid	100%
● Error	0%
● Empty	0%
United States	
United States	
France	
Germany	
Unknown	
Canada	
Mexico	
Italy	
Australia	
Unknown	
Unknown	
Unknown	
Unknown	
Unknown	
Unknown	
Unknown	
United States	

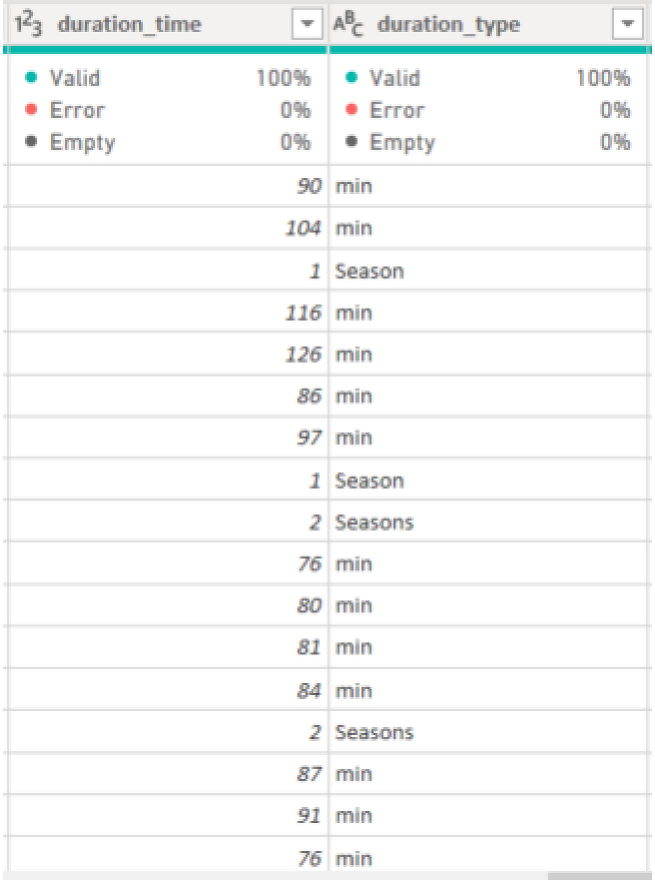
Do the same for the column listed_in.

F- DURATION

For a better visualisation of the duration, we must create a group of duration. To do this, split the column duration with a space delimiter to separate the number and the type.

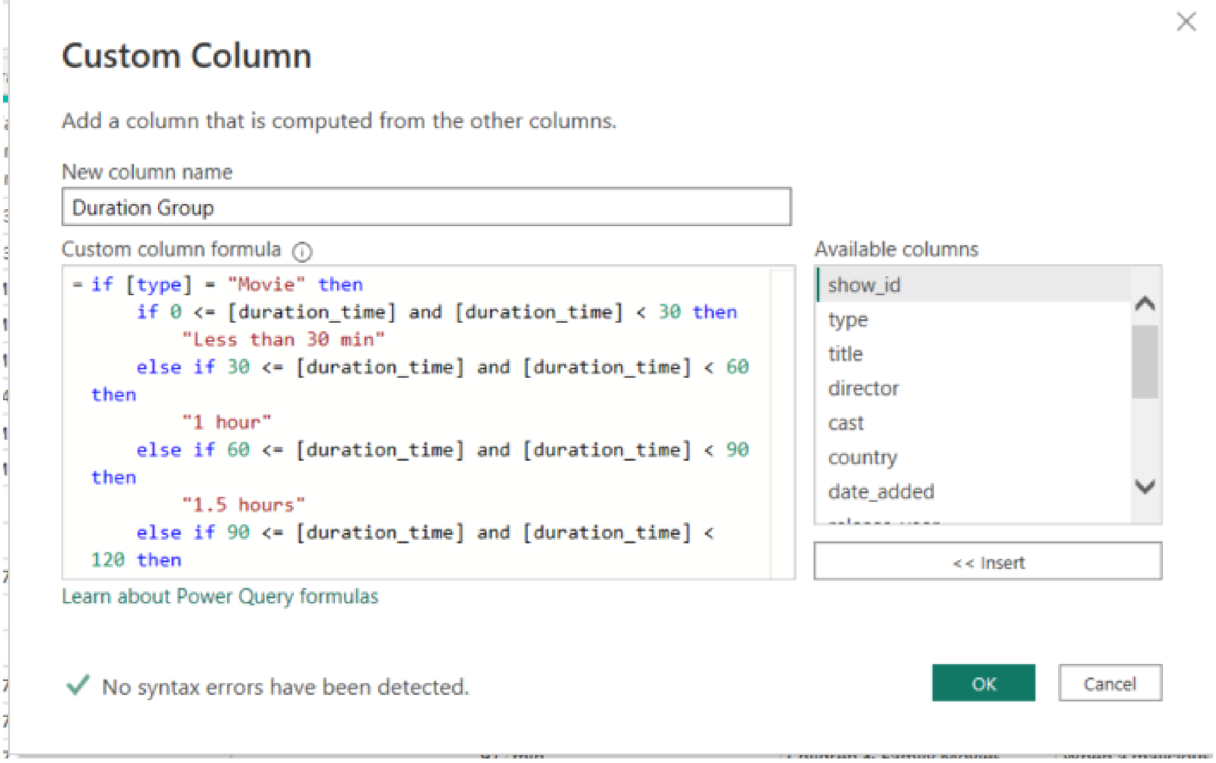


duration	
● Valid	100%
● Error	0%
● Empty	0%
90 min	
104 min	
1 Season	
116 min	
126 min	
86 min	
97 min	
1 Season	
2 Seasons	
76 min	
80 min	
81 min	
84 min	
2 Seasons	
87 min	
91 min	
76 min	



duration_time		duration_type	
● Valid	100%	● Valid	100%
● Error	0%	● Error	0%
● Empty	0%	● Empty	0%
	90	min	
	104	min	
	1	Season	
	116	min	
	126	min	
	86	min	
	97	min	
	1	Season	
	2	Seasons	
	76	min	
	80	min	
	81	min	
	84	min	
	2	Seasons	
	87	min	
	91	min	
	76	min	

Now we can create our new Duration group column with this M formula in Add column and Custom Column:



Custom Column

Add a column that is computed from the other columns.

New column name

Duration Group

Custom column formula ⓘ

```
= if [type] = "Movie" then
      if 0 <= [duration_time] and [duration_time] < 30 then
          "Less than 30 min"
      else if 30 <= [duration_time] and [duration_time] < 60
  then
          "1 hour"
      else if 60 <= [duration_time] and [duration_time] < 90
  then
          "1.5 hours"
      else if 90 <= [duration_time] and [duration_time] <
  120 then
```

✓ No syntax errors have been detected.

The following formula:

```
if [type] = "Movie" then
   if 0 <= [duration_time] and [duration_time] < 30 then
      "Less than 30 min"
   else if 30 <= [duration_time] and [duration_time] < 60 then
      "1 hour"
   else if 60 <= [duration_time] and [duration_time] < 90 then
      "1.5 hours"
   else if 90 <= [duration_time] and [duration_time] < 120 then
      "2 hours"
   else if 120 <= [duration_time] and [duration_time] < 180 then
      "3 hours"
   else if 180 <= [duration_time] and [duration_time] < 400 then
      "More than 3 hours"
   else
      Text.From([duration_time]) & " " & [duration_type]
else
   Text.From([duration_time]) & " " & [duration_type]
```

2. EDA (EXPLORATORY DATA ANALYSIS)