

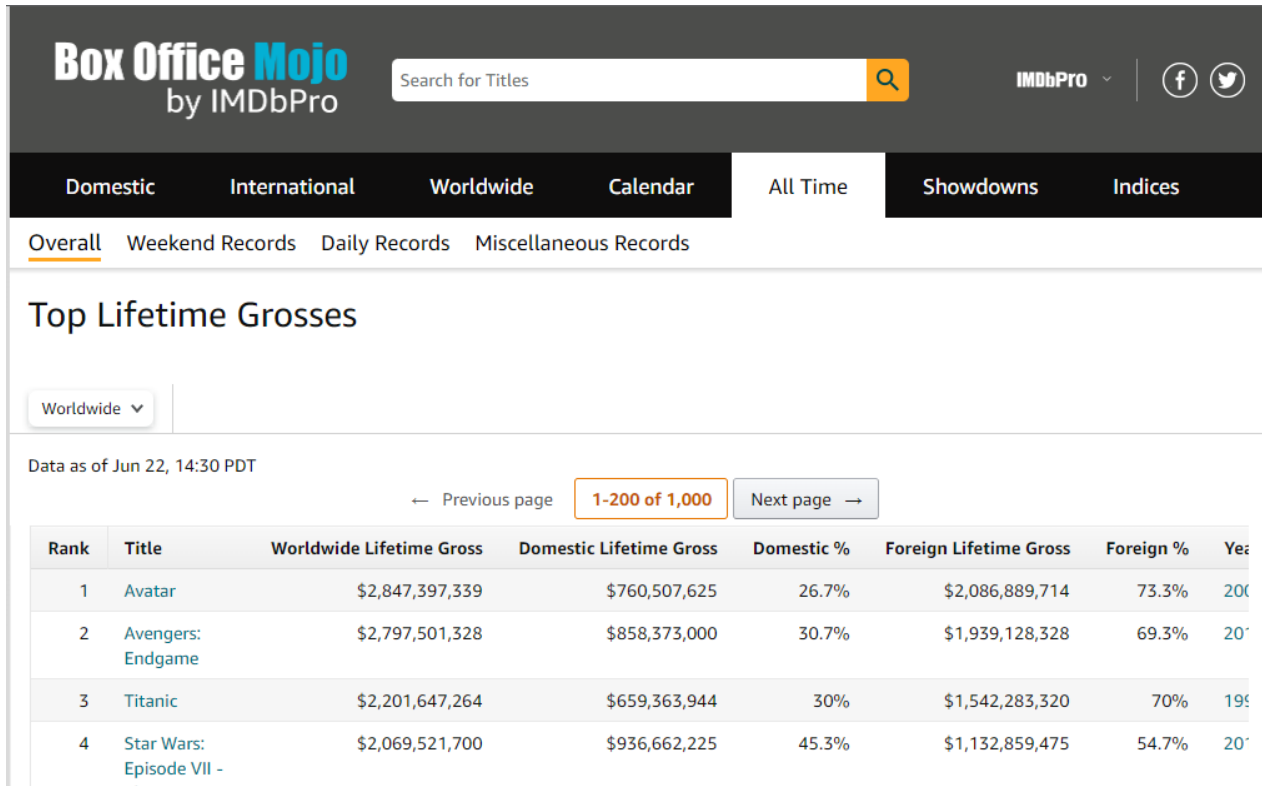
Module 2: Power BI Desktop and Data Transformation

Demo 1 – Box Office Analysis

edureka!

Box Office Analysis

Part 1: Worldwide Gross Sales Information of Movies can be found [here](#).



Box Office Mojo by IMDbPro

Search for Titles

IMDbPro

Domestic International Worldwide Calendar All Time Showdowns Indices

Overall Weekend Records Daily Records Miscellaneous Records

Top Lifetime Grosses

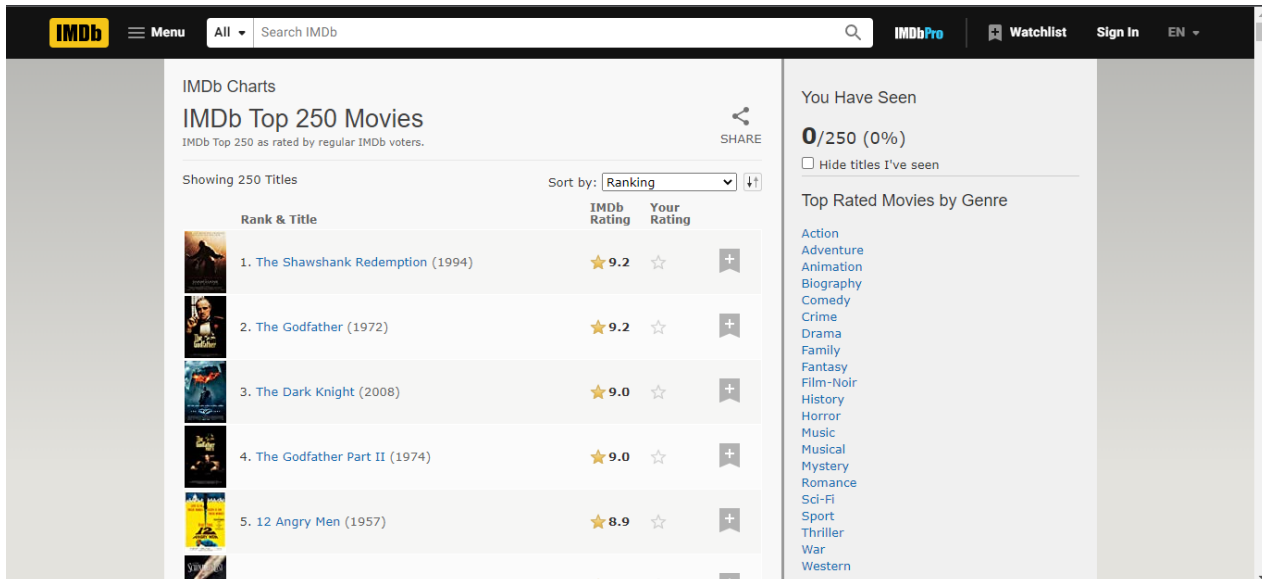
Worldwide

Data as of Jun 22, 14:30 PDT

← Previous page 1-200 of 1,000 Next page →

Rank	Title	Worldwide Lifetime Gross	Domestic Lifetime Gross	Domestic %	Foreign Lifetime Gross	Foreign %	Year
1	Avatar	\$2,847,397,339	\$760,507,625	26.7%	\$2,086,889,714	73.3%	2009
2	Avengers: Endgame	\$2,797,501,328	\$858,373,000	30.7%	\$1,939,128,328	69.3%	2019
3	Titanic	\$2,201,647,264	\$659,363,944	30%	\$1,542,283,320	70%	1997
4	Star Wars: Episode VII - The Force Awakens	\$2,069,521,700	\$936,662,225	45.3%	\$1,132,859,475	54.7%	2015

Part 2: IMDB Top 250 Movies is used [here](#).



IMDb

Menu All Search IMDb

IMDbPro Watchlist Sign In EN

IMDb Charts

IMDb Top 250 Movies

IMDb Top 250 as rated by regular IMDb voters.

Showing 250 Titles Sort by: Ranking

Rank & Title	IMDb Rating	Your Rating
1. The Shawshank Redemption (1994)	★ 9.2	☆
2. The Godfather (1972)	★ 9.2	☆
3. The Dark Knight (2008)	★ 9.0	☆
4. The Godfather Part II (1974)	★ 9.0	☆
5. 12 Angry Men (1957)	★ 8.9	☆
6. Schindler's List (1993)	★ 8.9	☆

You Have Seen 0/250 (0%)

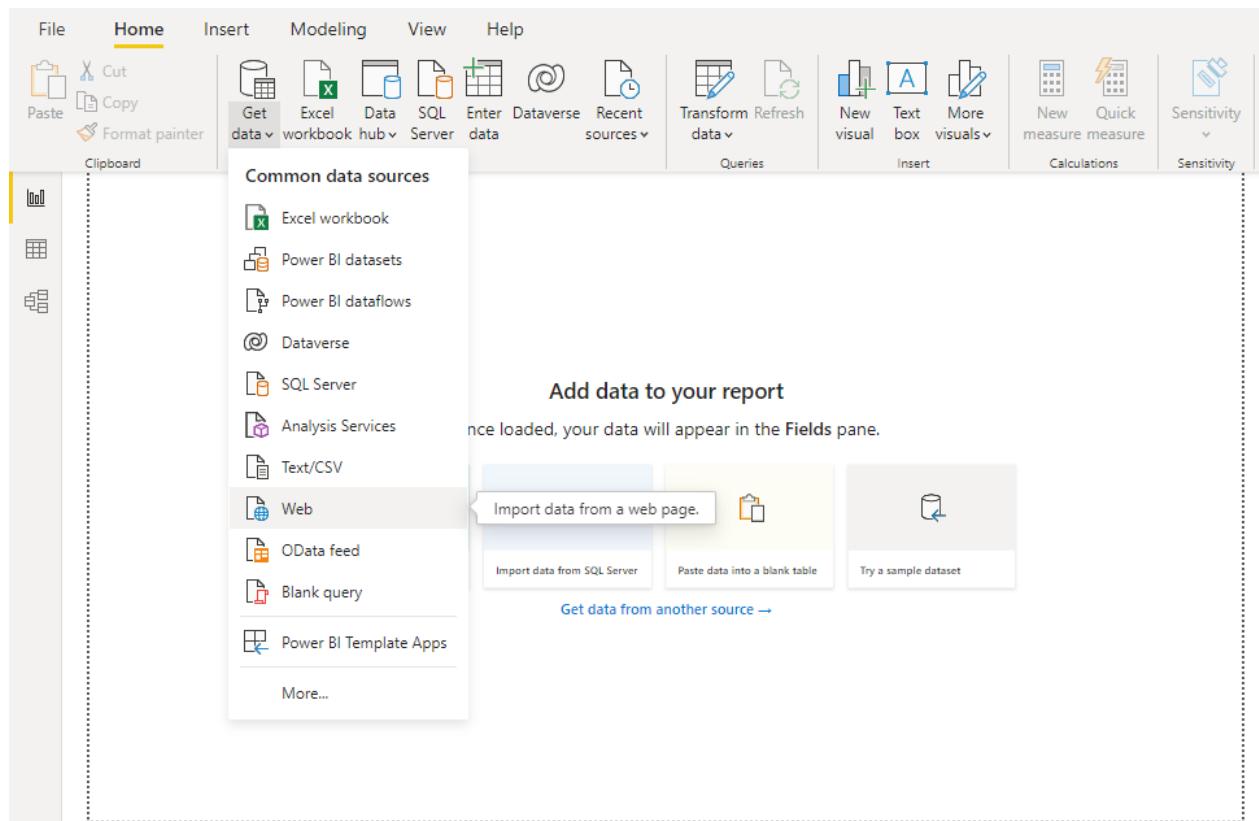
☐ Hide titles I've seen

Top Rated Movies by Genre

- Action
- Adventure
- Animation
- Biography
- Comedy
- Crime
- Drama
- Family
- Fantasy
- Film-Noir
- History
- Horror
- Music
- Musical
- Mystery
- Romance
- Sci-Fi
- Sport
- Thriller
- War
- Western

Part 1

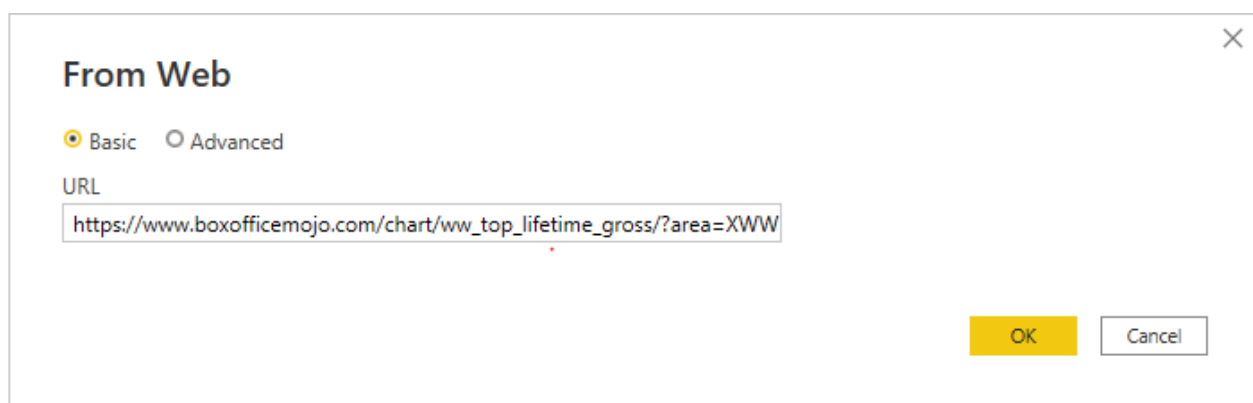
Step 1. Connect Data from the Web



Step 2. Input the following link:

https://www.boxofficemojo.com/chart/ww_top_lifetime_gross/?area=XWW

in the **URL** section tab, which is opened:



Click on **OK**

Step 3. Select the **Table** which shows the correct information from the Preview tab

Navigator

Display Options ▾

- HTML Tables [3]
 - Table 1
 - Table 2**
 - Table 3
- Suggested Tables [4]
 - Table 4
 - Table 5
 - Table 6
 - Table 7
- Text [2]
 - HTML Code
 - Displayed Text

Table 2

Rank	Title	Worldwide Lifetime Gross
1	Avatar	\$2,847,397,339
2	Avengers: Endgame	\$2,797,501,328
3	Titanic	\$2,201,647,264
4	Star Wars: Episode VII - The Force Awakens	\$2,069,521,700
5	Avengers: Infinity War	\$2,048,359,754
6	Spider-Man: No Way Home	\$1,901,228,407
7	Jurassic World	\$1,671,537,444
8	The Lion King	\$1,663,250,487
9	The Avengers	\$1,518,815,515
10	Furious 7	\$1,515,341,399
11	Frozen II	\$1,450,026,933
12	Avengers: Age of Ultron	\$1,402,809,540
13	Black Panther	\$1,347,597,973
14	Harry Potter and the Deathly Hallows: Part 2	\$1,342,359,942
15	Star Wars: Episode VIII - The Last Jedi	\$1,332,698,830
16	Jurassic World: Fallen Kingdom	\$1,310,466,296
17	Frozen	\$1,281,508,100
18	Beauty and the Beast	\$1,273,576,220
19	Incredibles 2	\$1,243,089,244
20	The Fate of the Furious	\$1,236,005,118
21	Iron Man 3	\$1,214,11,252

Add Table Using Examples

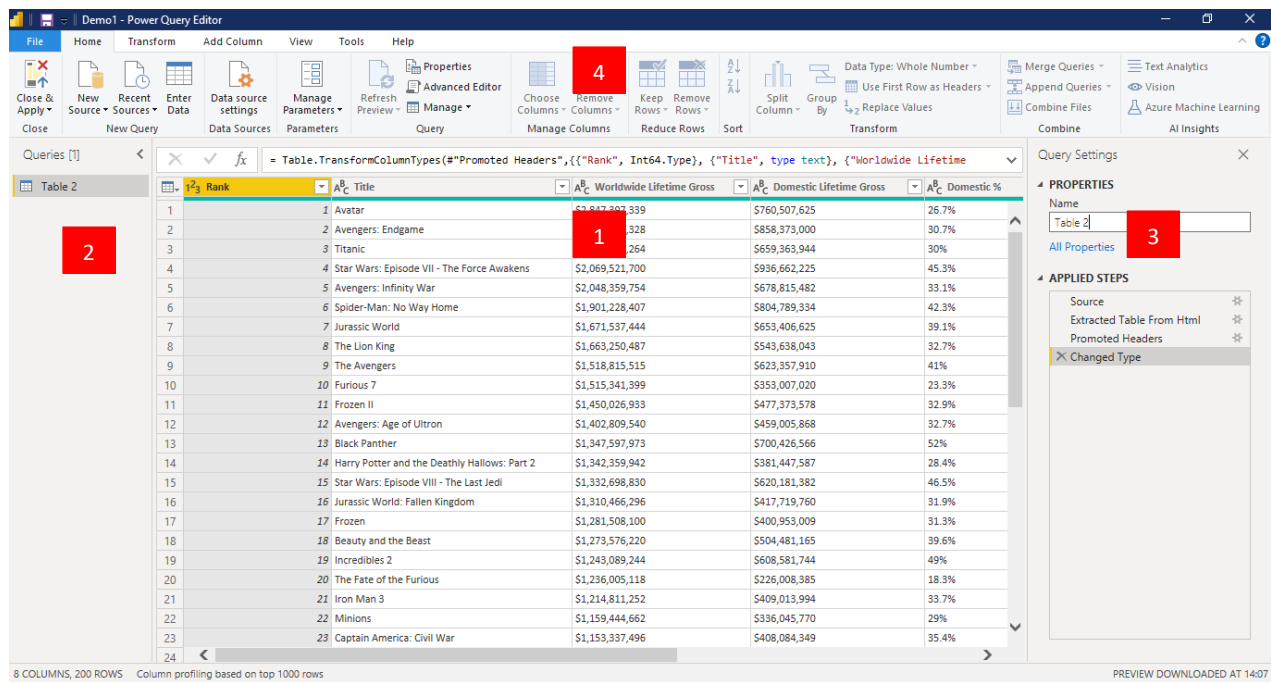
Load Transform Data Cancel

Click on **Transform Data**

Note 1: Due to the HTML formatting of the Page multiple tables might be visible choose the table which is appropriate for analysis and is viewed in the Webpage

Note 2: This demo is speed oriented sometimes due to session expiration of the URL the changes applied during data transformation throws some error

Query Editor



Query Editor has four main sections (numbers matched to screenshot above)

- 1) Main dataset pane: This is the central area where the result set will be displayed as a preview with a limited number of rows
- 2) List of Queries; Left-hand side pane will show a list of all queries in this solution or file
- 3) Query settings pane; Properties such as the name of the query can be set here. Also, a list of all applied steps to the current query is visible in this pane.
- 4) Transformations menu; The Power Query editor has many transformations options in GUI that are available through the menu in the top section

Step 4. Remove the Changed Type from the applied Steps

The screenshot shows the Microsoft Power Query Editor interface. The ribbon at the top includes 'File', 'Home', 'Transform', 'Add Column', 'View', 'Tools', and 'Help'. The 'Query Settings' pane on the right shows the query name 'Table 2' and the applied step 'Changed Type', which is highlighted with a red arrow. The main data table has columns 'Rank', 'Title', and 'Worldw'.

Rank	Title	Worldw
1	Avatar	\$2,847,3
2	Avengers: Endgame	\$2,797,5
3	Titanic	\$2,201,6
4	Star Wars: Episode VII - The Force Awakens	\$2,069,5
5	Avengers: Infinity War	\$2,048,3
6	Spider-Man: No Way Home	\$1,901,2
7	Jurassic World	\$1,671,5
8	The Lion King	\$1,663,2
9	The Avengers	\$1,518,8
10	Furious 7	\$1,515,3
11	Frozen II	\$1,450,0
12	Avengers: Age of Ultron	\$1,402,8
13	Black Panther	\$1,347,5
14	Harry Potter and the Deathly Hallows: Part 2	\$1,342,3
15	Star Wars: Episode VIII - The Last Jedi	\$1,332,6
16	Jurassic World: Fallen Kingdom	\$1,310,4
17	Frozen	\$1,281,5
18	Beauty and the Beast	\$1,273,5
19	Incredibles 2	\$1,243,0
20	The Fate of the Furious	\$1,236,0
21	Iron Man 3	\$1,214,8
22	Minions	\$1,159,4
23	Captain America: Civil War	\$1,153,3
24	Aquaman	\$1,148,5

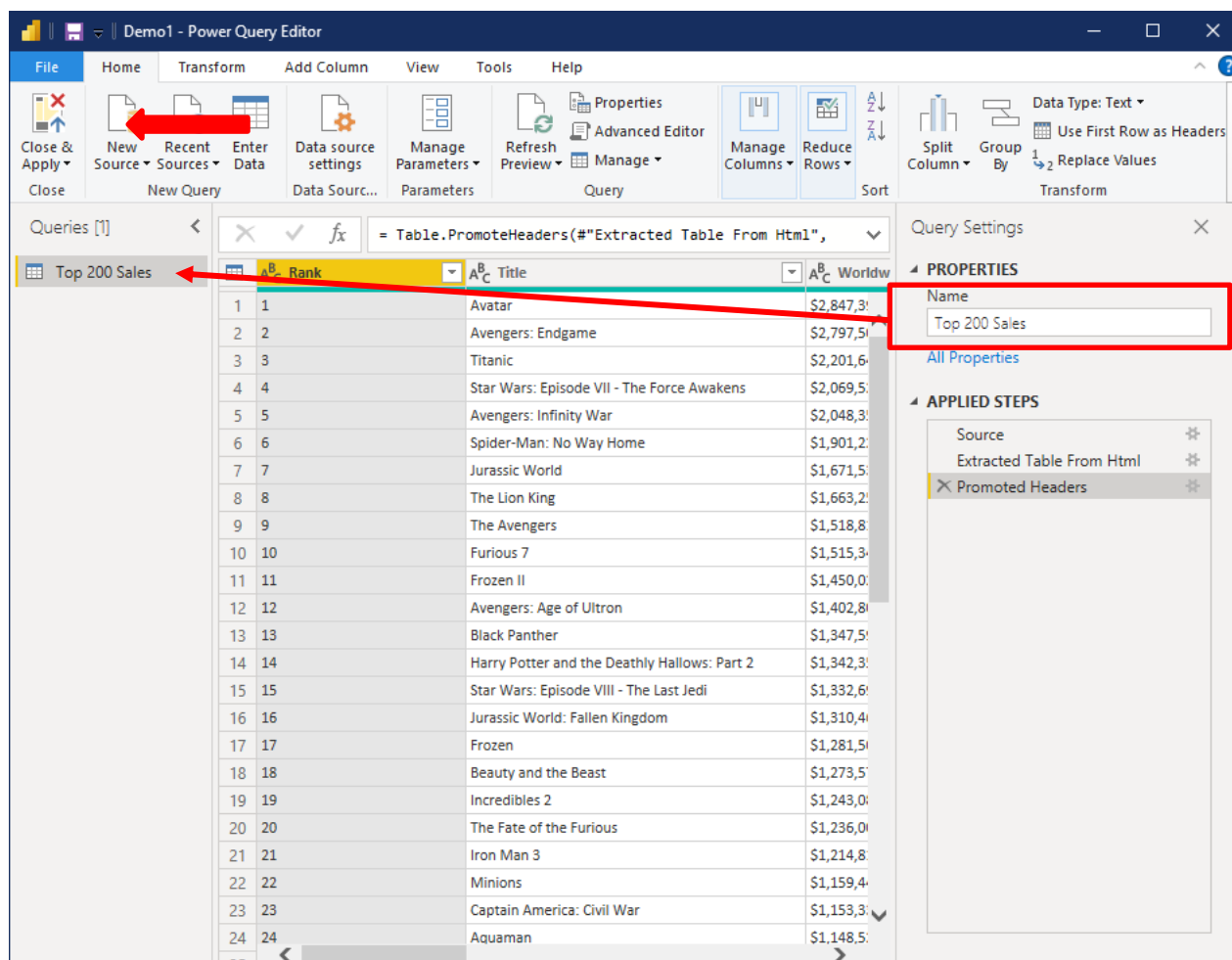
After removing the Changed Type, it will be viewed as:

The screenshot shows the Power Query Editor window titled 'Demo1 - Power Query Editor'. The main area displays a table with columns 'Rank', 'Title', and 'Worldwide Gross'. The table contains 24 rows of movie data. On the right, the 'Query Settings' pane is open, showing the 'Name' as 'Table 2'. Below it, the 'APPLIED STEPS' pane lists two steps: 'Source' and 'Promoted Headers'. The 'Promoted Headers' step is highlighted with a red box.

Rank	Title	Worldwide Gross
1	Avatar	\$2,847,300
2	Avengers: Endgame	\$2,797,500
3	Titanic	\$2,201,600
4	Star Wars: Episode VII - The Force Awakens	\$2,069,500
5	Avengers: Infinity War	\$2,048,300
6	Spider-Man: No Way Home	\$1,901,200
7	Jurassic World	\$1,671,500
8	The Lion King	\$1,663,200
9	The Avengers	\$1,518,800
10	Furious 7	\$1,515,300
11	Frozen II	\$1,450,000
12	Avengers: Age of Ultron	\$1,402,800
13	Black Panther	\$1,347,500
14	Harry Potter and the Deathly Hallows: Part 2	\$1,342,300
15	Star Wars: Episode VIII - The Last Jedi	\$1,332,600
16	Jurassic World: Fallen Kingdom	\$1,310,400
17	Frozen	\$1,281,500
18	Beauty and the Beast	\$1,273,500
19	Incredibles 2	\$1,243,000
20	The Fate of the Furious	\$1,236,000
21	Iron Man 3	\$1,214,800
22	Minions	\$1,159,400
23	Captain America: Civil War	\$1,153,300
24	Aquaman	\$1,148,500

Step 5. Rename the table from Table 2 to → Top 200 Sales

edureka!



Click on **Close & Apply**

Step 6. Click on Get Data → Web and include the following URL:

https://www.boxofficemojo.com/chart/ww_top_lifetime_gross/?area=XWW&offset=200

From Web

☒ Basic ☐ Advanced

URL

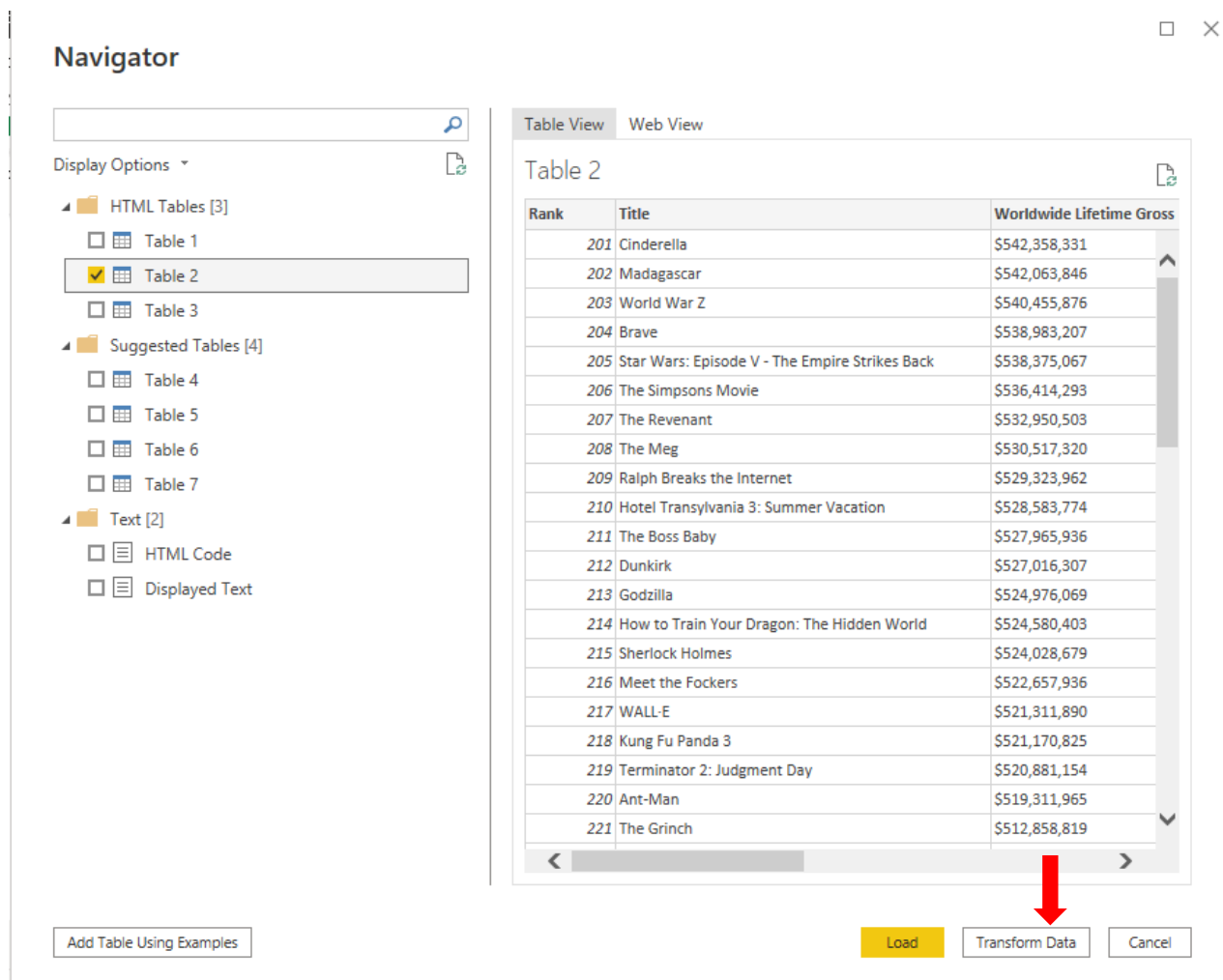
www.boxofficemojo.com/chart/ww_top_lifetime_gross/?area=XWW&offset=200

OK

Cancel

Click on Ok

Step 7. Click on Transform Data and choose appropriate table



The screenshot shows the Power BI Navigator window. On the left, under 'HTML Tables [3]', 'Table 2' is selected with a checkmark. The main pane displays 'Table 2' with the following data:

Rank	Title	Worldwide Lifetime Gross
201	Cinderella	\$542,358,331
202	Madagascar	\$542,063,846
203	World War Z	\$540,455,876
204	Brave	\$538,983,207
205	Star Wars: Episode V - The Empire Strikes Back	\$538,375,067
206	The Simpsons Movie	\$536,414,293
207	The Revenant	\$532,950,503
208	The Meg	\$530,517,320
209	Ralph Breaks the Internet	\$529,323,962
210	Hotel Transylvania 3: Summer Vacation	\$528,583,774
211	The Boss Baby	\$527,965,936
212	Dunkirk	\$527,016,307
213	Godzilla	\$524,976,069
214	How to Train Your Dragon: The Hidden World	\$524,580,403
215	Sherlock Holmes	\$524,028,679
216	Meet the Fockers	\$522,657,936
217	WALL-E	\$521,311,890
218	Kung Fu Panda 3	\$521,170,825
219	Terminator 2: Judgment Day	\$520,881,154
220	Ant-Man	\$519,311,965
221	The Grinch	\$512,858,819

At the bottom right, a red arrow points to the 'Transform Data' button. Other buttons visible are 'Load' and 'Cancel'.

Step 8. Perform Similar above Steps

The screenshot shows the Power Query Editor window titled 'Demo1 - Power Query Editor'. The ribbon includes 'File', 'Home', 'Transform', 'Add Column', 'View', 'Tools', and 'Help'. The 'Queries' pane on the left shows two queries: 'Top 200 Sales' and 'Top 400 Sales'. The main table view displays a list of movies with columns 'Rank', 'Title', and 'Worldwide Gross'. The 'Properties' pane on the right shows the query name 'Top 400 Sales' highlighted with a red box. A red arrow points from the 'Top 400 Sales' query in the list to the 'Top 400 Sales' query in the main table.

Rank	Title	Worldwide Gross
1	Cinderella	\$542,35
2	Madagascar	\$542,06
3	World War Z	\$540,45
4	Brave	\$538,98
5	Star Wars: Episode V - The Empire Strikes Back	\$538,37
6	The Simpsons Movie	\$536,41
7	The Revenant	\$532,95
8	The Meg	\$530,51
9	Ralph Breaks the Internet	\$529,32
10	Hotel Transylvania 3: Summer Vacation	\$528,58
11	The Boss Baby	\$527,96
12	Dunkirk	\$527,01
13	Godzilla	\$524,97
14	How to Train Your Dragon: The Hidden World	\$524,58
15	Sherlock Holmes	\$524,02
16	Meet the Fockers	\$522,65
17	WALL-E	\$521,31
18	Kung Fu Panda 3	\$521,17
19	Terminator 2: Judgment Day	\$520,88
20	Ant-Man	\$519,31
21	The Grinch	\$512,85
22	Venom: Let There Be Carnage	\$506,86
23	Ghost	\$505,70
24	Aladdin	\$504,05

Use Query as Reference

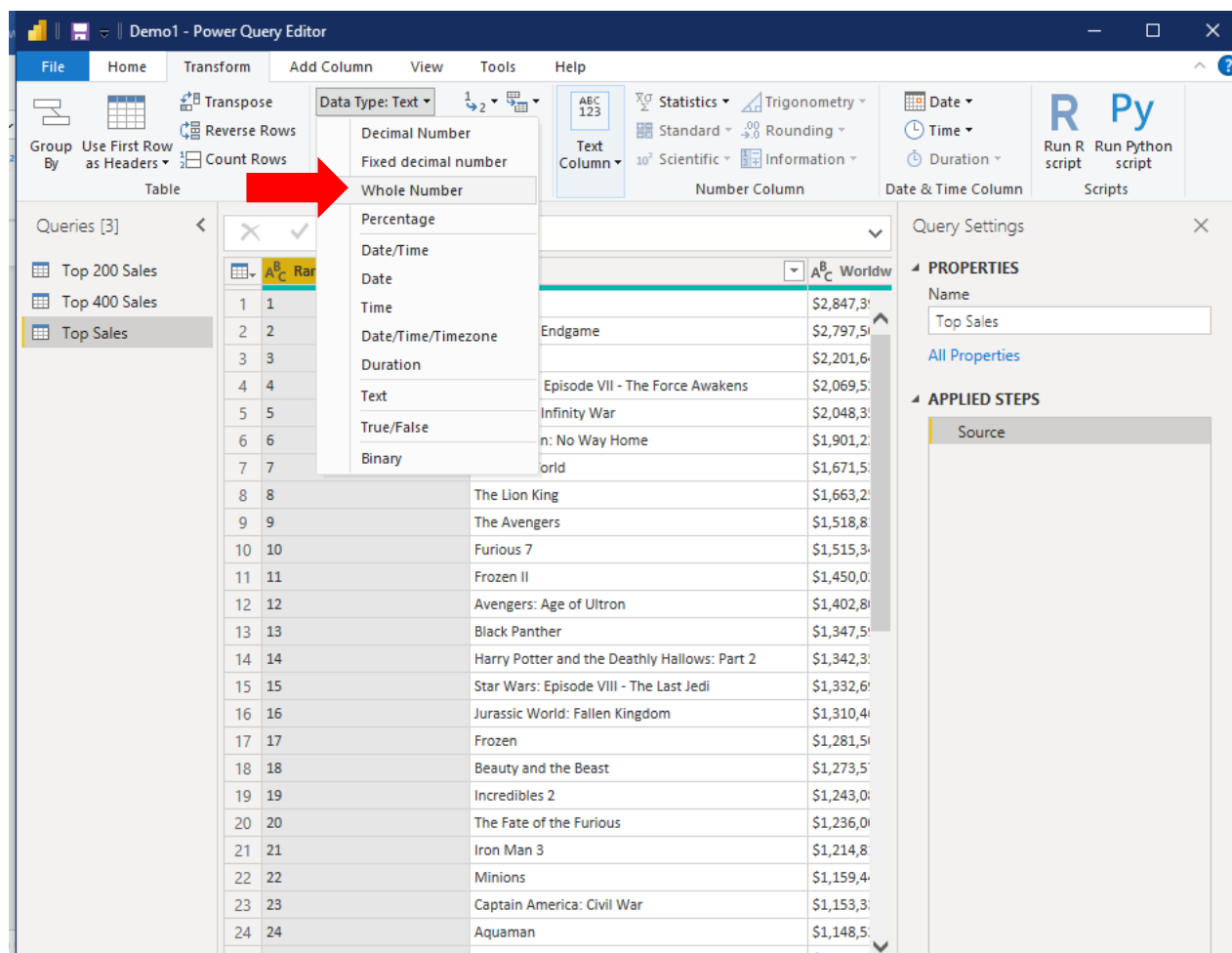
Step 9. Right click on **Top 200 Sales** and select **Reference** and Rename the new Table generated as: **Top Sales**

The screenshot shows the Power Query Editor interface. The main area displays a table with columns: Rank, Title, and Worldw. The 'Rank' column is highlighted. The 'Properties' pane on the right shows the 'Name' property set to 'Top Sales'. A red arrow points from the 'Rank' column header to the 'Name' property in the 'Properties' pane.

Rank	Title	Worldw
1	Avatar	\$2,847,3
2	Avengers: Endgame	\$2,797,5
3	Titanic	\$2,201,6
4	Star Wars: Episode VII - The Force Awakens	\$2,069,5
5	Avengers: Infinity War	\$2,048,3
6	Spider-Man: No Way Home	\$1,901,2
7	Jurassic World	\$1,671,5
8	The Lion King	\$1,663,2
9	The Avengers	\$1,518,8
10	Furious 7	\$1,515,3
11	Frozen II	\$1,450,0
12	Avengers: Age of Ultron	\$1,402,8
13	Black Panther	\$1,347,5
14	Harry Potter and the Deathly Hallows: Part 2	\$1,342,3
15	Star Wars: Episode VIII - The Last Jedi	\$1,332,6
16	Jurassic World: Fallen Kingdom	\$1,310,4
17	Frozen	\$1,281,5
18	Beauty and the Beast	\$1,273,5
19	Incredibles 2	\$1,243,0
20	The Fate of the Furious	\$1,236,0
21	Iron Man 3	\$1,214,8
22	Minions	\$1,159,4
23	Captain America: Civil War	\$1,153,3
24	Aquaman	\$1,148,5

Append Queries

Step 10. Click on **Rank** column and change its Data Type from **Text** to **Whole Number**



Step 11. There are some missing values in the column **Domestic Lifetime Gross**

fx = Table.TransformColumnTypes(Source,{{"Rank", Int64.Type}})

	A _C Title	A _C Worldwide Lifetime Gross	A _C Domestic Lifetime Gross	A _C Domestic %	A _C Foreign Lifetime Gross
72	Harry Potter and the Chamber of Secrets	\$879,919,132	\$262,641,637	29.8%	\$617,277,495
73	Ice Age: Continental Drift	\$877,244,782	\$161,321,843	18.4%	\$715,922,939
74	The Secret Life of Pets	\$875,458,631	\$368,384,330	42.1%	\$507,074,301
75	Batman v Superman: Dawn of Justice	\$873,637,528	\$330,360,194	37.8%	\$543,277,334
76	Wolf Warrior 2	\$870,325,439	\$2,721,100	0.3%	\$867,604,339
77	Star Wars: Episode III - Revenge of the Sith	\$868,390,560	\$380,270,577	43.8%	\$488,119,983
78	The Hunger Games: Catching Fire	\$865,011,746	\$424,668,047	49.1%	\$440,343,699
79	Guardians of the Galaxy Vol. 2	\$863,756,051	\$389,813,101	45.1%	\$473,942,950
80	Inside Out	\$858,848,019	\$356,921,711	41.6%	\$501,926,308
81	Venom	\$856,085,151	\$213,515,506	24.9%	\$642,569,645
82	Thor: Ragnarok	\$853,983,879	\$315,058,289	36.9%	\$538,925,590
83	Inception	\$836,848,102	\$292,587,330	35%	\$544,260,772
84	Transformers: Revenge of the Fallen	\$836,303,693	\$402,111,870	48.1%	\$434,191,823
85	The Twilight Saga: Breaking Dawn - Part 2	\$829,747,654	\$292,324,737	35.2%	\$537,422,917
86	Spider-Man	\$825,025,036	\$407,022,860	49.3%	\$418,002,176
87	Wonder Woman	\$822,854,286	\$412,845,172	50.2%	\$410,009,114
88	Hi, Mom	\$822,009,764	-	-	\$822,009,764
89	Independence Day	\$817,400,891	\$306,169,268	37.5%	\$511,231,623
90	Fantastic Beasts and Where to Find Them	\$814,044,001	\$234,037,575	28.8%	\$580,006,426
91	Shrek the Third	\$813,367,380	\$322,719,944	39.7%	\$490,647,436
92	Coco	\$807,817,888	\$210,460,015	26%	\$597,357,873
93	Jumanji: The Next Level	\$800,059,707	\$320,314,960	40%	\$479,744,747
94	Harry Potter and the Prisoner of Azkaban	\$797,491,273	\$250,105,651	31.4%	\$547,385,622

Step 12. Make sure the **Data Type**: Text (since the replace values will use the sign is “-“in the row, click on **Replace values**, enter the **Values To Find** – (Dash) and **Replace With** as \$0 (Space)

The screenshot shows the Power BI Desktop interface. The 'Transform' tab is active, and the 'Replace Values' button is highlighted with a red box. A 'Replace Values' dialog box is open, showing the 'Value To Find' as '-' and 'Replace With' as '\$0'. The background table displays movie titles and their gross values.

Title	Worldwide Lifetime Gross	Domestic Lifetime Gross	Domestic %	Foreign Lifetime Gross
22 Minions				398,892
23 Captain America: Civil War				253,147
24 Aquaman				424,079
25 The Lord of the Rings: The Two Towers				185,007
26 Spider-Man: Far from Home				395,911
27 Captain Marvel				633,133
28 Transformers: Dark of the Moon				403,536
29 Skyfall				209,222
30 Transformers: Age of Extinction				614,996
31 Jurassic Park				484,283
32 The Dark Knight Rises				100,513
33 Joker				968,073
34 Star Wars: Episode IX - The Rise of Skywalker				946,737
35 Toy Story 4				356,585
36 Toy Story 3	\$1,066,970,811	\$415,004,880	38.9%	\$651,965,931
37 Pirates of the Caribbean: Dead Man's Chest	\$1,066,179,747	\$423,315,812	39.7%	\$642,863,935
38 The Lion King	\$1,063,611,805	\$422,783,777	39.8%	\$640,828,028
39 Rogue One: A Star Wars Story	\$1,056,057,720	\$532,177,324	50.4%	\$523,880,396
40 Aladdin	\$1,050,693,953	\$355,559,216	33.8%	\$695,134,737
41 Pirates of the Caribbean: On Stranger Tides	\$1,045,713,802	\$241,071,802	23%	\$804,642,000
42 Despicable Me 3	\$1,034,800,131	\$264,624,300	25.6%	\$770,175,831
43 Finding Dory	\$1,028,570,942	\$486,295,561	47.3%	\$542,275,381
44 Star Wars: Episode I - The Phantom Menace	\$1,027,082,707	\$474,544,677	46.2%	\$552,538,030

Click on Ok

Step 13. Do the same with **Domestic %** column

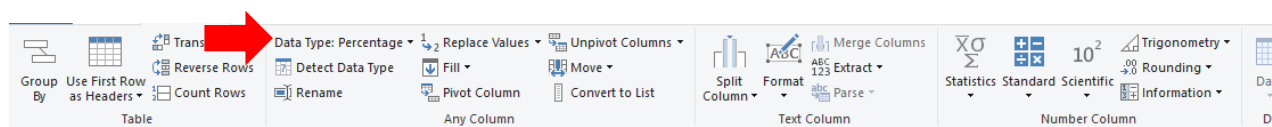
The screenshot shows the Power BI Desktop interface. The 'Transform' tab is active, and the 'Replace Values' button is highlighted with a red box. A 'Replace Values' dialog box is open, showing the 'Value To Find' as '-' and 'Replace With' as '\$0'. The background table displays movie titles and their gross values.

Title	Worldwide Lifetime Gross	Domestic Lifetime Gross	Domestic %	Foreign Lifetime Gross
78 Star Wars: Episode III - Revenge of the Sith	\$688,390,560	\$380,270,577	43.8%	
79 Star Wars: Episode II - Attack of the Clones	\$669,654,642	\$373,365,835	49.1%	
80 Star Wars: Episode IV - A New Hope	\$669,654,642	\$373,365,835	45.1%	
81 Star Wars: Episode V - The Empire Strikes Back	\$669,654,642	\$373,365,835	41.6%	
82 Star Wars: Episode VI - Return of the Jedi	\$669,654,642	\$373,365,835	24.9%	
83 Star Wars: Episode I - The Phantom Menace	\$669,654,642	\$373,365,835	36.9%	
84 Star Wars: Episode II - Attack of the Clones	\$669,654,642	\$373,365,835	35%	
85 Star Wars: Episode III - Revenge of the Sith	\$669,654,642	\$380,270,577	48.1%	
86 Star Wars: Episode IV - A New Hope	\$669,654,642	\$373,365,835	35.2%	
87 Star Wars: Episode V - The Empire Strikes Back	\$669,654,642	\$373,365,835	49.3%	
88 Star Wars: Episode VI - Return of the Jedi	\$669,654,642	\$373,365,835	50.2%	
89 Star Wars: Episode I - The Phantom Menace	\$669,654,642	\$373,365,835	-	
90 Star Wars: Episode II - Attack of the Clones	\$669,654,642	\$373,365,835	37.5%	
91 Star Wars: Episode III - Revenge of the Sith	\$688,390,560	\$380,270,577	28.8%	
92 Star Wars: Episode IV - A New Hope	\$669,654,642	\$373,365,835	39.7%	
93 Star Wars: Episode V - The Empire Strikes Back	\$669,654,642	\$373,365,835	26%	
94 Jumanji: The Next Level	\$800,059,707	\$320,314,960	40%	
95 Harry Potter and the Prisoner of Azkaban	\$797,568,607	\$250,105,651	31.4%	
96 Pirates of the Caribbean: Dead Men Tell No Tales	\$794,881,442	\$172,558,876	21.7%	
97 E.T. the Extra-Terrestrial	\$792,910,554	\$435,110,554	54.9%	
98 Mission: Impossible - Fallout	\$791,657,398	\$220,159,104	27.8%	
99 2012	\$791,217,826	\$166,112,167	21%	
100 Indiana Jones and the Kingdom of the Crystal Skull	\$790,653,942	\$317,101,119	40.1%	

Step 14. Change the data types of the following columns

Foreign % column from Text to → Percentage

Year column from Text to → Whole Number



Queries [3]

Top 200 Sales
Top 400 Sales
Top Sales

fx = Table.TransformColumnTypes(#"Replaced Value1",{{"Domestic %", Percentage.Type}, {"Foreign %", Percentage.Type}, {"Year", Whole.Number.Type})

	Domestic Lifetime Gross	% Domestic %	Foreign Lifetime Gross	% Foreign %	Year
1	\$760,507,625	26.70%	\$2,086,889,714	73.30%	2009
2	\$858,373,000	30.70%	\$1,939,128,328	69.30%	2019
3	\$659,363,944	30.00%	\$1,542,283,320	70.00%	1997
4	\$936,662,225	45.30%	\$1,132,859,475	54.70%	2015
5	\$678,815,482	33.10%	\$1,369,544,272	66.90%	2018
6	\$804,789,334	42.30%	\$1,096,439,073	57.70%	2021
7	\$653,406,625	39.10%	\$1,018,130,819	60.90%	2015
8	\$543,638,043	32.70%	\$1,119,612,444	67.30%	2019
9	\$623,357,910	41.00%	\$895,457,605	59.00%	2012
10	\$353,007,020	23.30%	\$1,162,334,379	76.70%	2015
11	\$477,373,578	32.90%	\$972,653,355	67.10%	2019
12	\$459,005,868	32.70%	\$943,803,672	67.30%	2015
13	\$700,426,566	52.00%	\$647,171,407	48.00%	2018
14	\$381,447,587	28.40%	\$960,912,355	71.60%	2011
15	\$620,181,382	46.50%	\$712,517,448	53.50%	2017
16	\$417,719,760	31.90%	\$892,746,536	68.10%	2018
17	\$400,953,009	31.30%	\$880,555,091	68.70%	2013
18	\$504,481,165	39.60%	\$769,095,055	60.40%	2017
19	\$608,581,744	49.00%	\$634,507,500	51.00%	2018
20	\$226,008,385	18.30%	\$1,009,996,733	81.70%	2017
21	\$409,013,994	33.70%	\$805,797,258	66.30%	2013
22	\$336,045,770	29.00%	\$823,398,892	71.00%	2015
23	\$408,084,349	35.40%	\$745,253,147	64.60%	2016

8 COLUMNS, 200 ROWS Column profiling based on top 1000 rows

Step 15. Click on Close & Apply

File Home Transform Add Column View Tools Help

Close & Apply (highlighted)

Close the Query Editor window and apply any pending changes.

fx = Table.TransformColumnTypes(#"Replaced Value1",{{"Domestic %", Percentage.Type}, {"Foreign %", Percentage.Type}, {"Year", Whole.Number.Type})

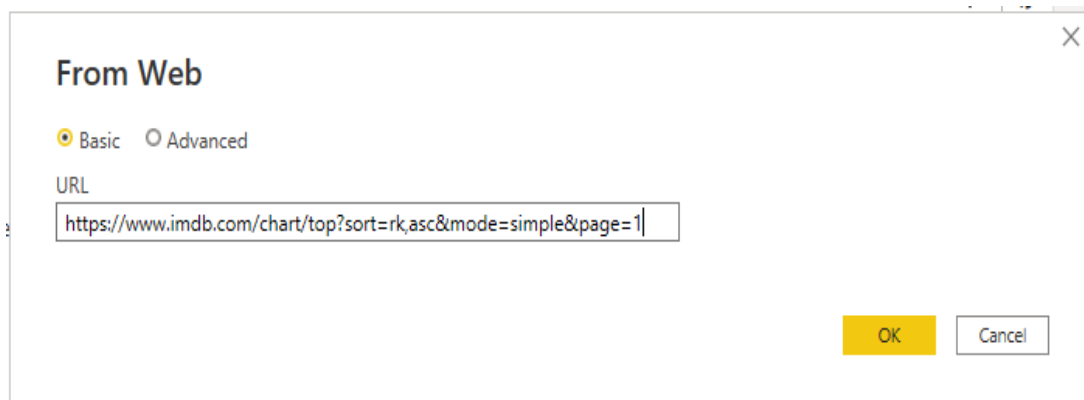
	Domestic Lifetime Gross	% Domestic %	Foreign Lifetime Gross	% Foreign %	Year
1	\$760,507,625	26.70%	\$2,086,889,714	73.30%	2009
2	\$858,373,000	30.70%	\$1,939,128,328	69.30%	2019
3	\$659,363,944	30.00%	\$1,542,283,320	70.00%	1997
4	\$936,662,225	45.30%	\$1,132,859,475	54.70%	2015
5	\$678,815,482	33.10%	\$1,369,544,272	66.90%	2018
6	\$804,789,334	42.30%	\$1,096,439,073	57.70%	2021
7	\$653,406,625	39.10%	\$1,018,130,819	60.90%	2015
8	\$543,638,043	32.70%	\$1,119,612,444	67.30%	2019
9	\$623,357,910	41.00%	\$895,457,605	59.00%	2012
10	\$353,007,020	23.30%	\$1,162,334,379	76.70%	2015
11	\$477,373,578	32.90%	\$972,653,355	67.10%	2019
12	\$459,005,868	32.70%	\$943,803,672	67.30%	2015
13	\$700,426,566	52.00%	\$647,171,407	48.00%	2018
14	\$381,447,587	28.40%	\$960,912,355	71.60%	2011
15	\$620,181,382	46.50%	\$712,517,448	53.50%	2017
16	\$417,719,760	31.90%	\$892,746,536	68.10%	2018
17	\$400,953,009	31.30%	\$880,555,091	68.70%	2013
18	\$504,481,165	39.60%	\$769,095,055	60.40%	2017
19	\$608,581,744	49.00%	\$634,507,500	51.00%	2018
20	\$226,008,385	18.30%	\$1,009,996,733	81.70%	2017
21	\$409,013,994	33.70%	\$805,797,258	66.30%	2013
22	\$336,045,770	29.00%	\$823,398,892	71.00%	2015
23	\$408,084,349	35.40%	\$745,253,147	64.60%	2016

8 COLUMNS, 200 ROWS Column profiling based on top 1000 rows

Note: The website owners over time might change the HTML page over time so do the above steps carefully if something isn't working, please try to work it out.

Part - 2

Step 16. Use the IMDB dataset found [here](https://www.imdb.com/chart/top?sort=rank,asc&mode=simple&page=1)
Get Data → Web , input the address



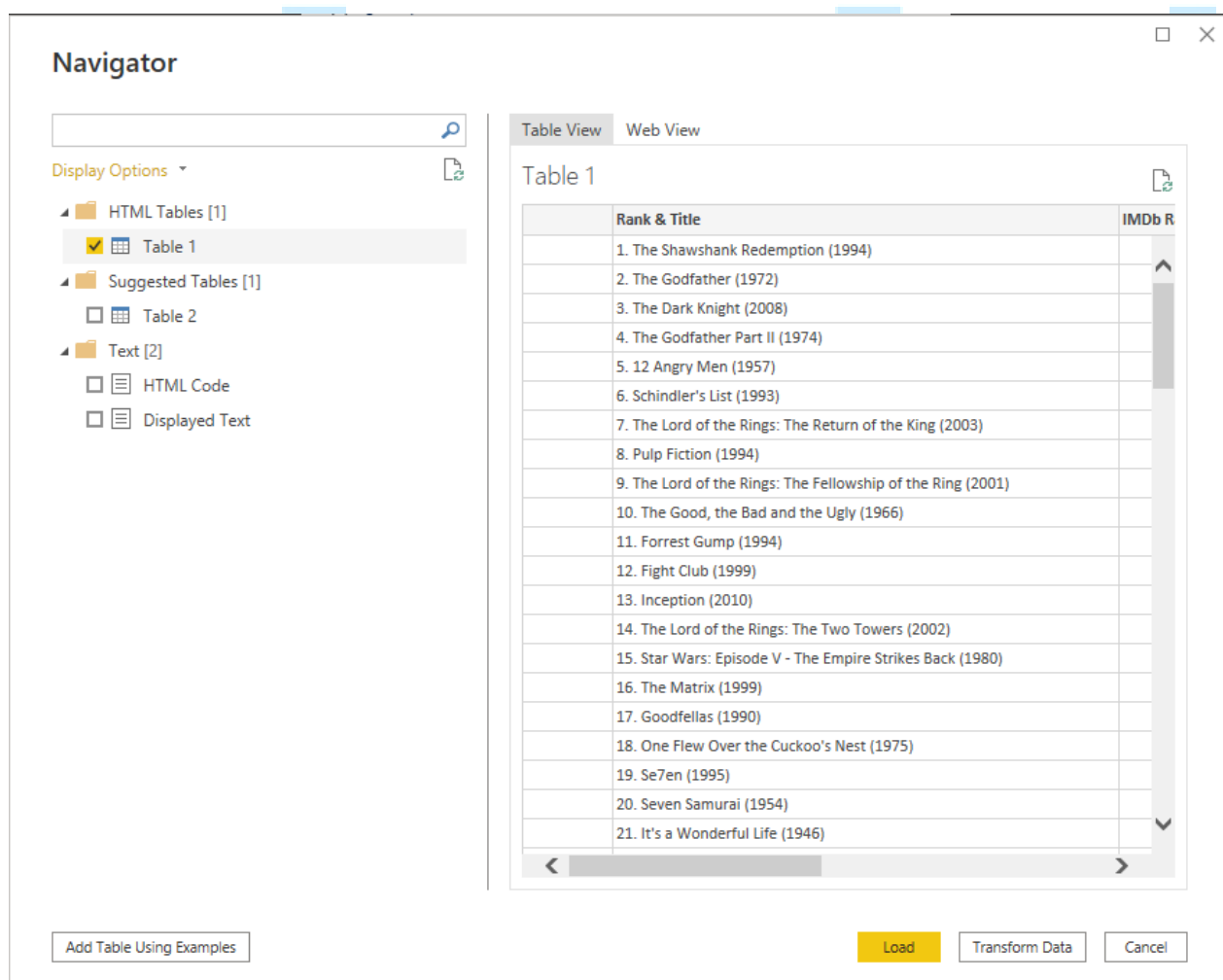
From Web

☒ Basic ☐ Advanced

URL

OK Cancel

Step 17. Select the appropriate table and click on **Transform Data**



Navigator

Display Options

- HTML Tables [1]
 - ☒ Table 1
- Suggested Tables [1]
 - ☐ Table 2
- Text [2]
 - ☐ HTML Code
 - ☐ Displayed Text

Table View Web View

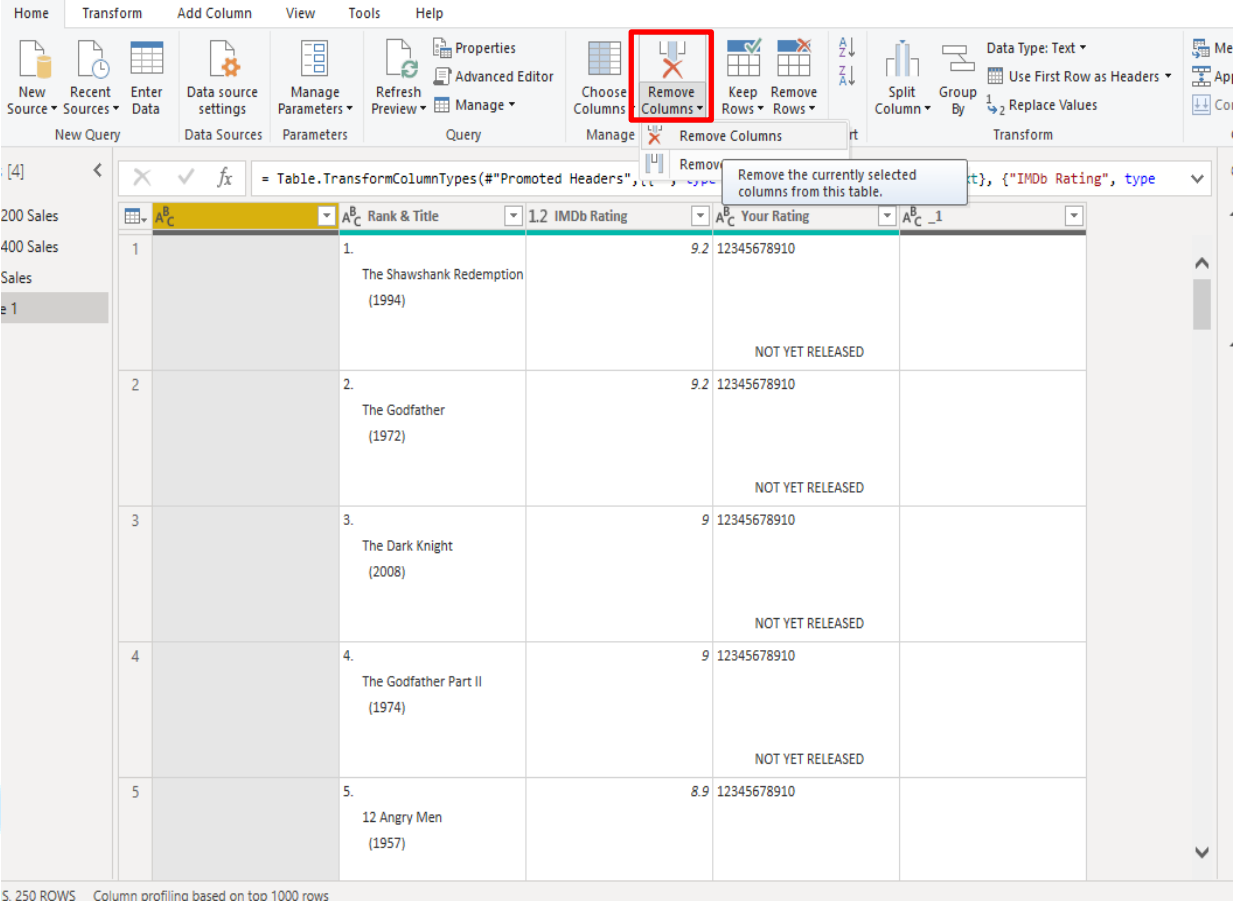
Table 1

Rank & Title	IMDb R
1. The Shawshank Redemption (1994)	
2. The Godfather (1972)	
3. The Dark Knight (2008)	
4. The Godfather Part II (1974)	
5. 12 Angry Men (1957)	
6. Schindler's List (1993)	
7. The Lord of the Rings: The Return of the King (2003)	
8. Pulp Fiction (1994)	
9. The Lord of the Rings: The Fellowship of the Ring (2001)	
10. The Good, the Bad and the Ugly (1966)	
11. Forrest Gump (1994)	
12. Fight Club (1999)	
13. Inception (2010)	
14. The Lord of the Rings: The Two Towers (2002)	
15. Star Wars: Episode V - The Empire Strikes Back (1980)	
16. The Matrix (1999)	
17. Goodfellas (1990)	
18. One Flew Over the Cuckoo's Nest (1975)	
19. Se7en (1995)	
20. Seven Samurai (1954)	
21. It's a Wonderful Life (1946)	

Add Table Using Examples Load Transform Data Cancel

Data Cleaning

Step 18. Delete the first and last two columns from the Query



After deletion, it looks as:

	A ^B C Rank & Title	1.2 IMDb Rating
1	1. The Shawshank Redemption (1994)	9.2
2	2. The Godfather (1972)	9.2
3	3. The Dark Knight (2008)	9
4	4. The Godfather Part II (1974)	9
5	5. 12 Angry Men (1957)	8.9
6	6. Schindler's List (1993)	8.9
7	7. The Lord of the Rings: The Re (2003)	8.9
8	8. Pulp Fiction (1994)	8.9

Column profiling based on top 1000 rows

Step 19. Select the 1st Column → Click on Split Column by Delimiter

The screenshot shows the Power BI Desktop interface with the 'Transform' ribbon selected. The 'Split Column' button is highlighted with a red box. A dropdown menu is open, showing the 'Split Column by Delimiter' option. The background shows the same table as in the previous image, with the 'Rank & Title' column selected.

Transform Add Column View Tools Help

Transpose Reverse Rows Count Rows Data Type: Text Detect Data Type Rename Replace Values Fill Pivot Column Unpivot Columns Move Convert to List

Split Column

By Delimiter

Split Column by Delimiter

Split values in the selected column based on the specified delimiter.

By Number of Characters

By Positions

By Lowercase to Uppercase

By Uppercase to Lowercase

By Digit to Non-Digit

By Non-Digit to Digit

	A ^B C Rank & Title	1.2 IMDb Rating
1	1. The Shawshank Redemption (1994)	9.2
2	2. The Godfather (1972)	9.2
3	3. The Dark Knight (2008)	9
4	4. The Godfather Part II (1974)	9
5	5. 12 Angry Men (1957)	8.9
6	6. Schindler's List (1993)	8.9
7	7. The Lord of the Rings: The Re (2003)	8.9
8	8. Pulp Fiction (1994)	8.9

VS Column profiling based on top 1000 rows

- Select **Custom** Delimiter dot (.) and Click on Ok.

The screenshot shows the Power BI Desktop interface with the 'Transform' ribbon selected. A table with 8 rows is visible, containing 'Rank & Title' and 'IMDb Rating' columns. The 'Split Column by Delimiter' dialog box is open, allowing the user to specify a custom delimiter (a dot '.') to split the 'Rank & Title' column. The dialog also offers options for splitting at the left-most, right-most, or each occurrence of the delimiter, and advanced options for splitting into columns or rows and using special characters.

Step 20. Select the second column and split it by **Number of Character**

The screenshot shows the Power BI Desktop interface with the 'Transform' ribbon selected. The table now has three columns: 'Rank & Title.1', 'Rank & Title.2', and 'IMDb Rating'. The 'Split Column by Number of Characters' dialog box is open, allowing the user to specify the number of characters (6) used to split the 'Rank & Title.2' column. The dialog also offers options for splitting once as far left or right as possible, or repeatedly. The 'Advanced options' section is currently collapsed.

1 ² Rank & Title.1	A ^B Rank & Title.2.1	1 ² Rank & Title.2.2	1.2 IMDb Rating
1	1 The Shawshank Redemption	-1994	9.2
2	2 The Godfather	-1972	9.2
3	3 The Dark Knight	-2008	9
4	4 The Godfather Part II	-1974	9
5	5 12 Angry Men	-1957	8.9
6	6 Schindler's List	-1993	8.9
7	7 The Lord of the Rings: The Return of the King	-2003	8.9
8	8 Pulp Fiction	-1994	8.9

Step 21. Rename the first column as **Rank**
Rename the second column as **Title**
Rename the third column as **Year**

Rank	Title	Year	IMDb Rating
1	The Shawshank Redemption	-1994	9.2
2	The Godfather	-1972	9.2
3	The Dark Knight	-2008	9
4	The Godfather Part II	-1974	9
5	12 Angry Men	-1957	8.9
6	Schindler's List	-1993	8.9
7	The Lord of the Rings: The Re	-2003	8.9
8	Pulp Fiction	-1994	8.9

Step 22. The Column **Year** has extra Dash Replace it with space
Before the above step Change the Data Type from whole Number to Text

Replace Values

Replace one value with another in the selected columns.

Value To Find
-

Replace With

Advanced options

OK

Query Settings

PROPERTIES

Name
Table 1

All Properties

Column profiling based on top 1000 rows

PREVIEW

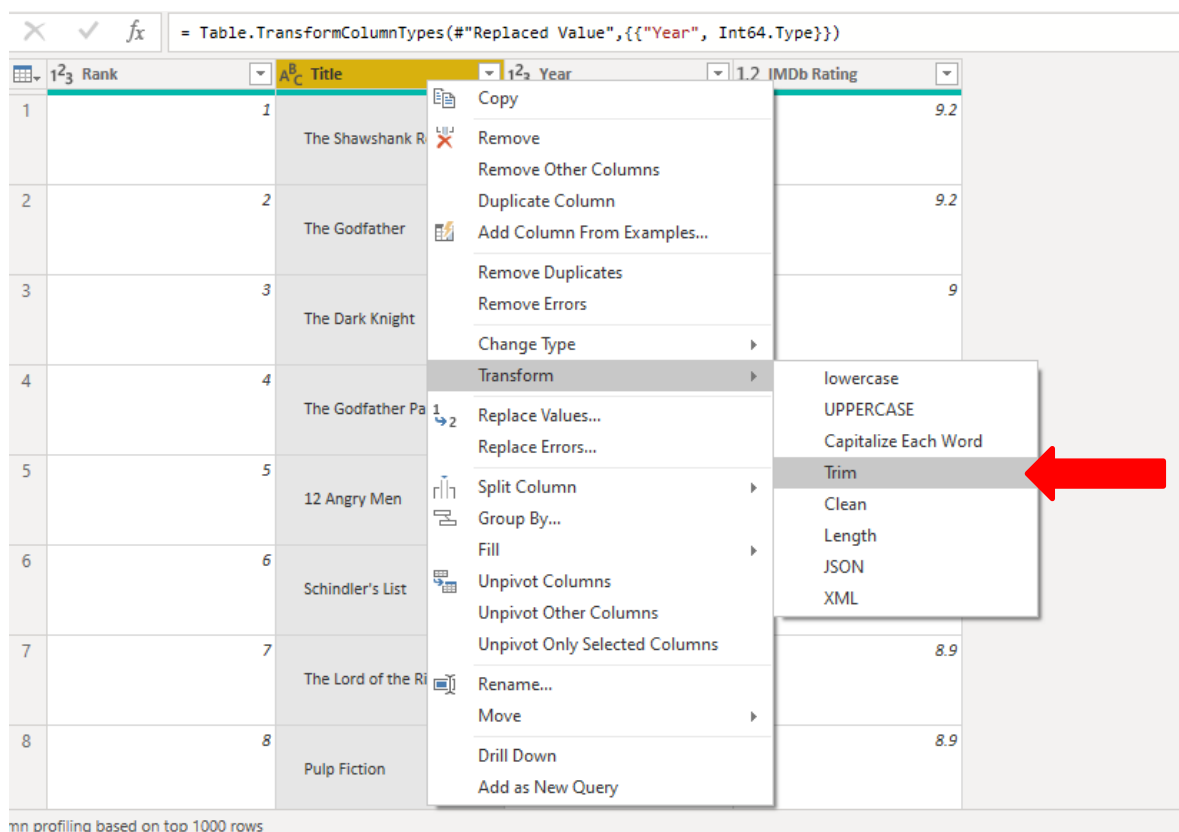
Check if the Data Type is a Whole Number

Transpose
Reverse Rows
Count Rows
Data Type: Whole Number
Detect Data Type
Rename
Replace Values
Pivot Column
Any Column
Unpivot Columns
Move
Convert to List
Split Column
Format
Text Column
Merge Columns
Extract
Parse

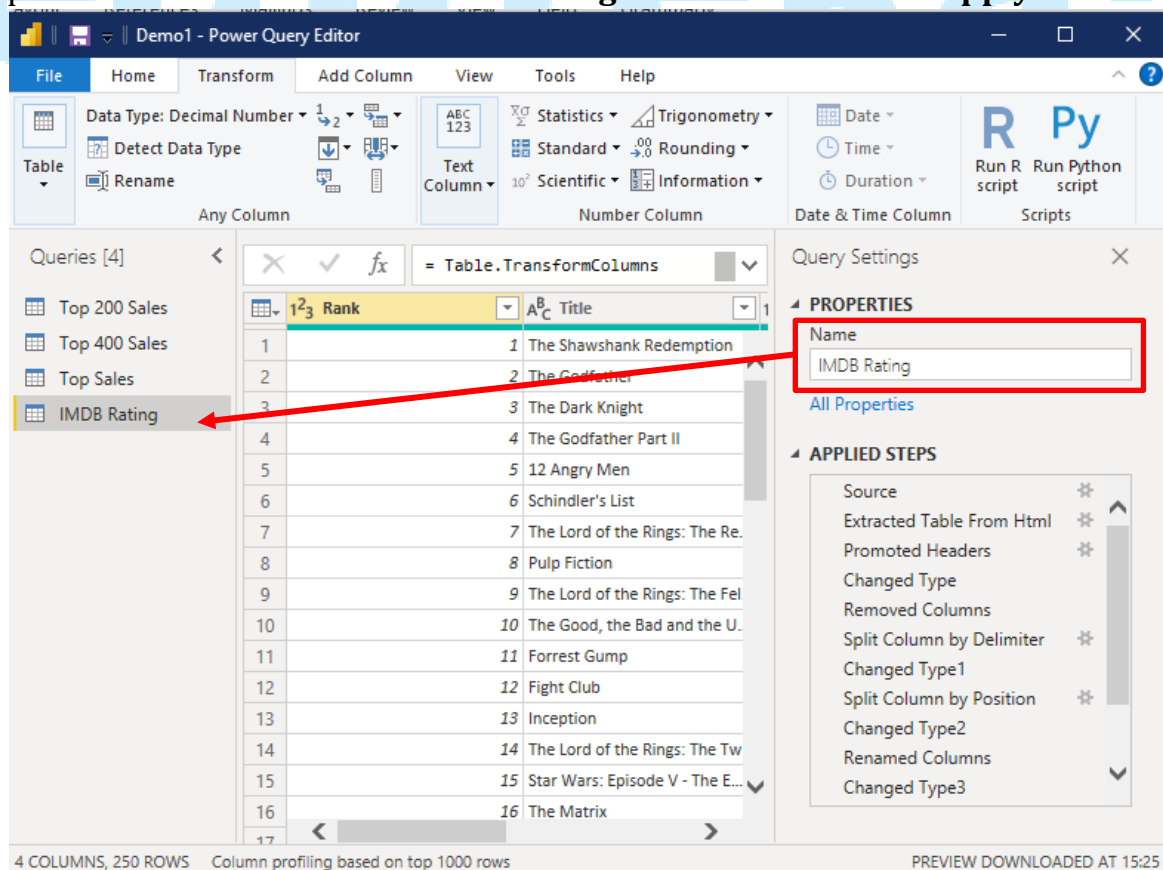
✕
✓
fx
= Table.TransformColumnTypes(#"Replaced Value",{{"Year", Int64.Type}})

	Rank	Title	Year	IMDb Rating
1	1	The Shawshank Redemption	1994	9.2
2	2	The Godfather	1972	9.2
3	3	The Dark Knight	2008	9
4	4	The Godfather Part II	1974	9
5	5	12 Angry Men	1957	8.9
6	6	Schindler's List	1993	8.9
7	7	The Lord of the Rings: The Re	2003	8.9

Step 23. Right Click on second column **Title** and **TRIM** the second column **Title**



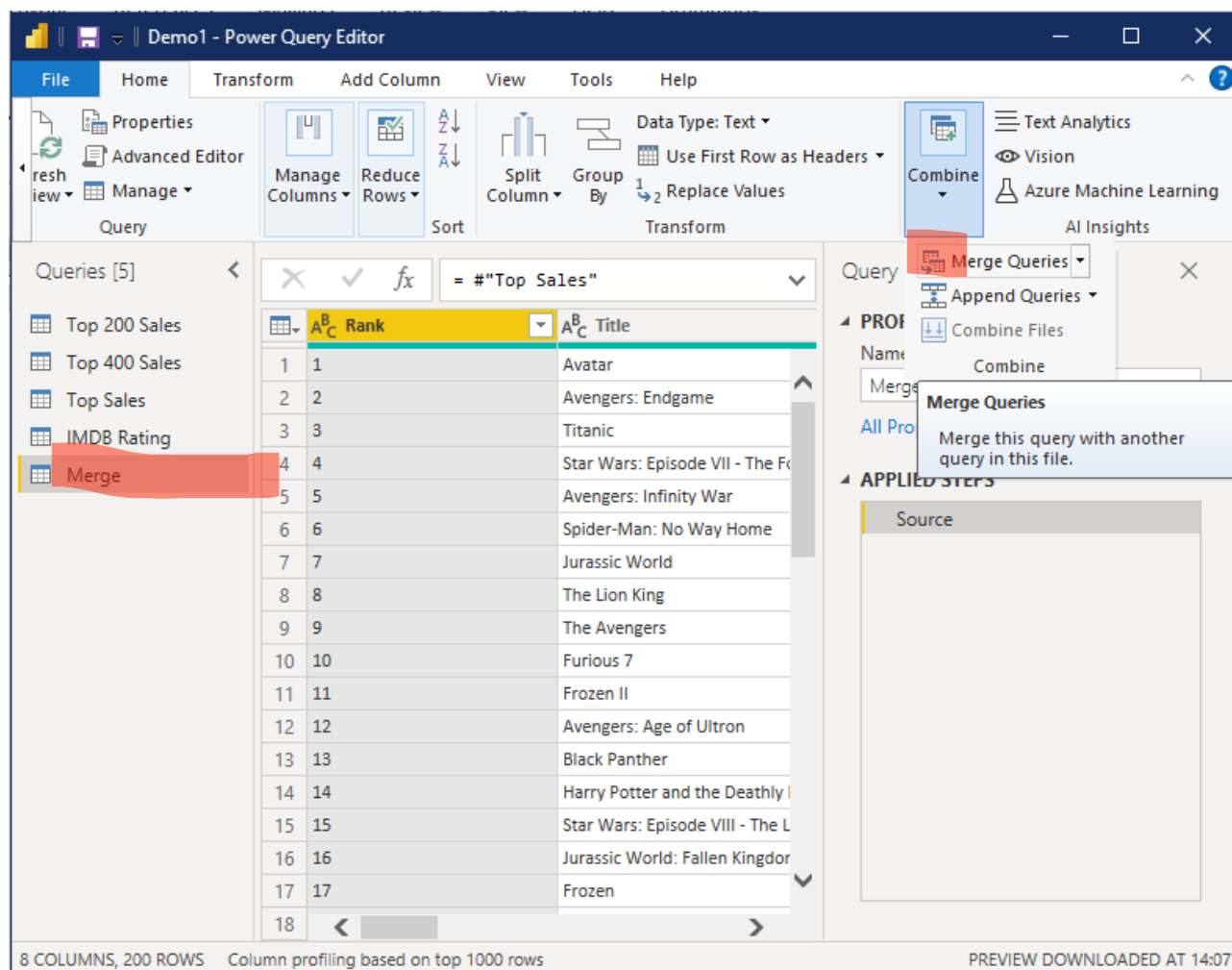
Step 24. Rename the Table to **IMDB Rating** and Click on **Close & Apply**.



Final Merge

The preparation for both data sets for a final merge together to see how best-selling movies are among top user rated film. Merge these two data sets or join them. Click on the **Top Sales** query and create a Reference for it, name the new query as **Merge**. Now click on **Merge Result**, and then from the **Home** tab, under **Combine** choose **Merge Queries**

Step 25. Merge Queries



After completion of the previous step, set the following setting

Merge Queries is equivalent to Join in SQL or database terminology. Merging queries requires two queries; the first query is the query that you are currently on it (**Merge**), and you can choose the second query in the Merge dialog box. Choose the second query as **IMDB Rating**. Now select joining keys as Title (you can also choose multiple joining columns by pressing the **ctrl** keyboard key + **Left click of the mouse**). Use Left outer join (this will only select all records from the first query with matched rows of that from the second query)

Merge

Select a table and matching columns to create a merged table.

Merge

Rank	Title	Worldwide Lifetime Gross	Domestic Lifetime Gross	Domestic %
1	Avatar	\$2,847,397,339	\$760,507,625	26.7%
2	Avengers: Endgame	\$2,797,501,328	\$858,373,000	30.7%
3	Titanic	\$2,201,647,264	\$659,363,944	30%
4	Star Wars: Episode VII - The Force Awakens	\$2,069,521,700	\$936,662,225	45.3%

IMDB Rating

Rank	Title	Year	IMDb Rating
1	The Shawshank Redemption	1994	9.2
2	The Godfather	1972	9.2
3	The Dark Knight	2008	9
4	The Godfather Part II	1974	9
5	12 Angry Men	1957	8.9

Join Kind

Inner (only matching rows)

☐ Use fuzzy matching to perform the merge

▸ Fuzzy matching options

✓ The selection matches 31 of 200 rows from the first table, and 30 of 250 r...

OK

Cancel

- Notice in the screenshot above that merge dialog mentioned only 31 records out of 200 movies matched! It means only 31 of best-seller movies are among top user-rated list! (Such a pity).
- The screenshot showed only the **Inner Join** result, but you choose **Left Outer** and then click on OK to look at the data.
- Joining experience in Power Query is a bit different from database tables. As a result of joining, you will get the first table with a new column for the new table. This new column holds table values that need to be expanded. If you click on the column header icon, you can choose which columns of the nested table you want to expand.

fx = Table.ExpandTableColumn(#"Merged Queries", "IMDB Rating", {"IMDb Rating"}, {"IMDB Rating.IMDb Rating"})

	A _C ^B Domestic %	A _C ^B Foreign Lifetime Gross	% Foreign %	1 ₃ ² Year	1.2 IMDB Rating.IMDb Rating
1	30.7%	\$1,939,128,328	69.30%	2019	8.3
2	53.2%	\$471,115,201	46.80%	2008	9
3	33.1%	\$1,369,544,272	66.90%	2018	8.4
4	42.3%	\$1,096,439,073	57.70%	2021	8.2
5	33%	\$768,185,007	67.00%	2003	8.9
6	32.7%	\$1,119,612,444	67.30%	2019	8.5
7	39.8%	\$640,828,028	60.20%	1994	8.5
8	35.2%	\$581,979,322	64.80%	2001	8.8
9	48.7%	\$347,770,863	51.30%	1994	8.8
10	35%	\$544,260,772	65.00%	2010	8.7
11	28.4%	\$960,912,355	71.60%	2011	8.1
12	36.2%	\$604,943,730	63.80%	2002	8.7
13	26.8%	\$513,709,189	73.20%	2014	8.6
14	59.4%	\$314,399,500	40.60%	1977	8.6
15	36.8%	\$695,484,283	63.20%	1993	8.1
16	41.4%	\$633,003,513	58.60%	2012	8.3
17	31.2%	\$738,968,073	68.80%	2019	8.3
18	38.9%	\$651,965,931	61.10%	2010	8.3
19	53%	\$427,100,000	47.00%	2022	8.5
20	40.5%	\$559,509,384	59.50%	2003	8.1
21	26%	\$597,357,873	74.00%	2017	8.3
22	41.6%	\$501,926,308	58.40%	2015	8.1
23	39.9%	\$442,094,938	60.10%	2009	8.2
24					

Click on **Close & Apply**

edureka!