# Module 2: Power BI Desktop and Data Transformation

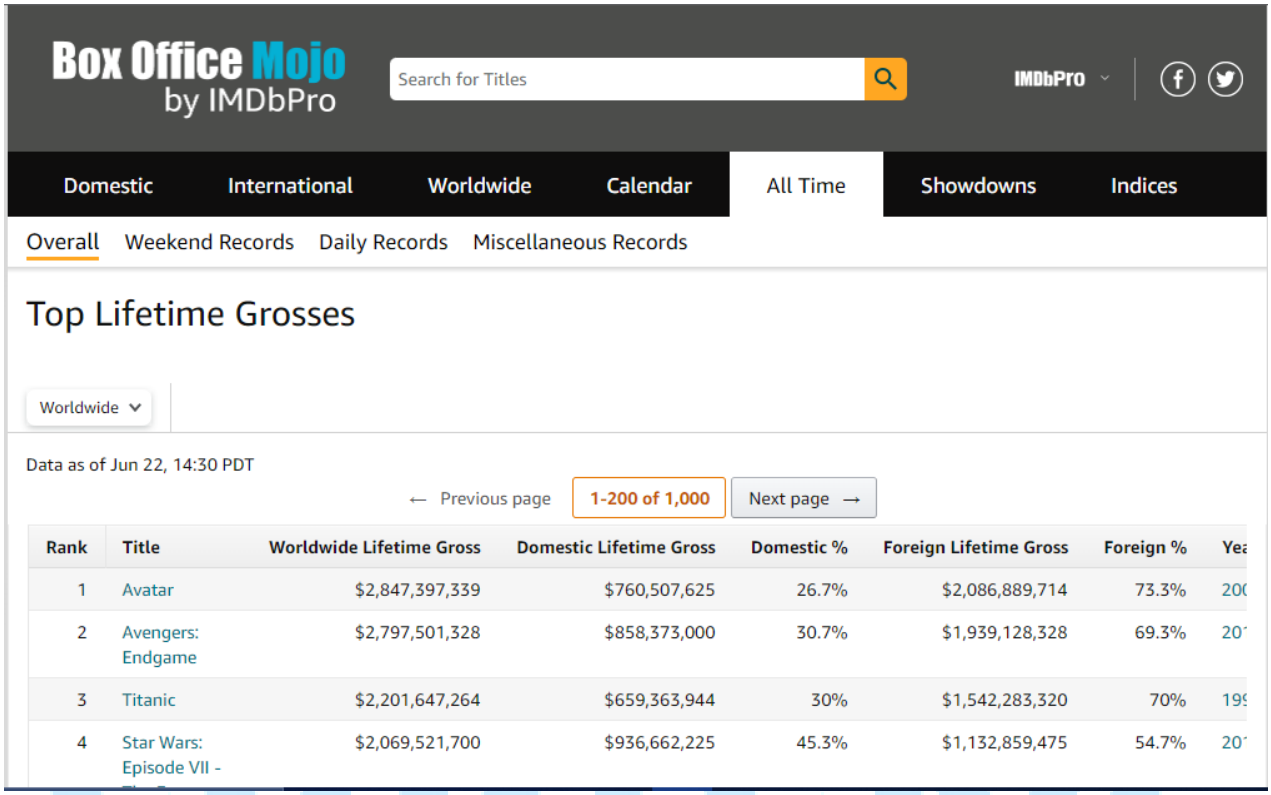
## Demo 1 – Box Office Analysis

edureka!

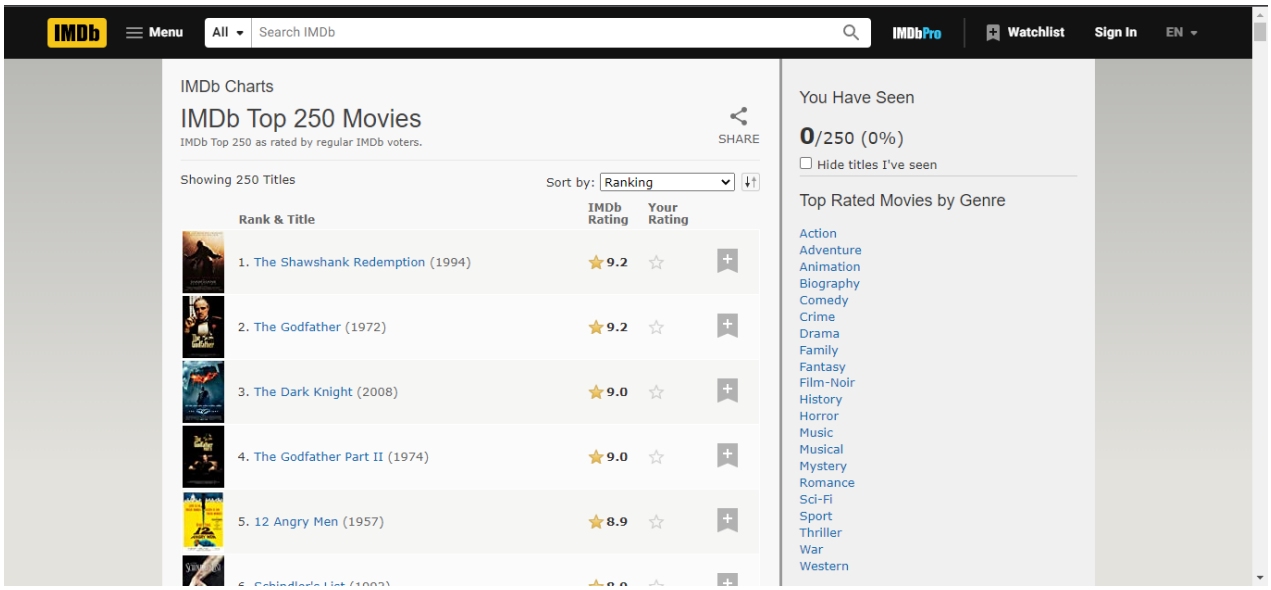
© Brain4ce Education Solutions Pvt. Ltd.

# Box Office Analysis

Part 1: Worldwide Gross Sales Information of Movies can be found here.

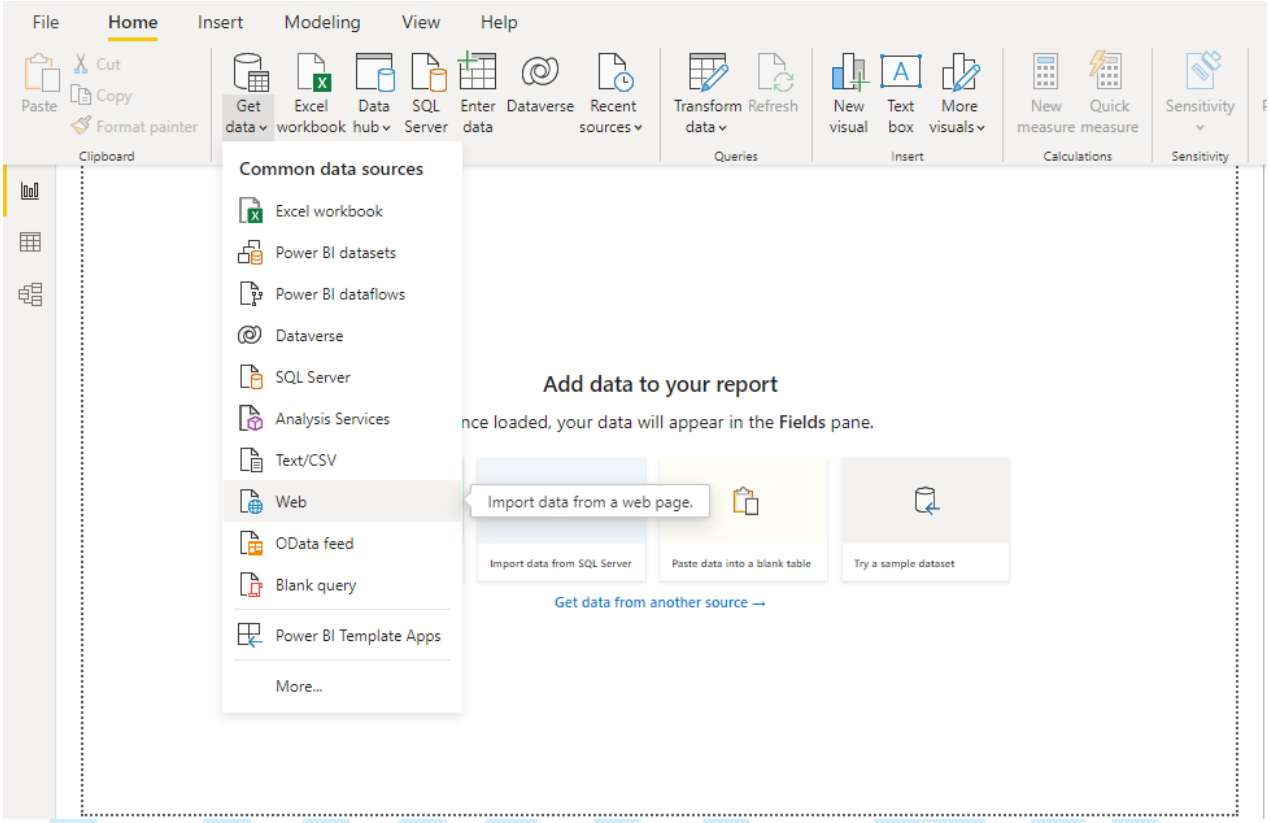



Part 2: IMDB Top 250 Movies is used here.



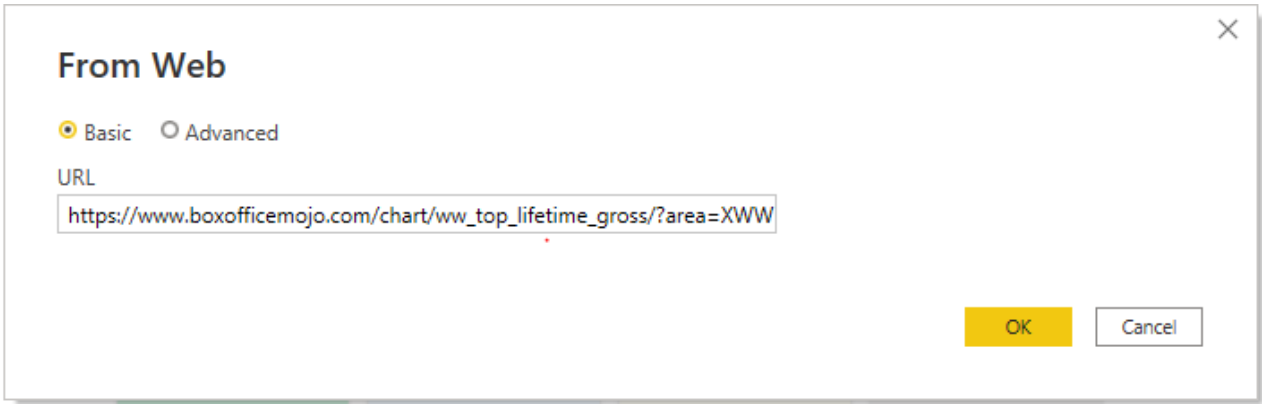
# Part 1

Step 1. Connect Data from the Web



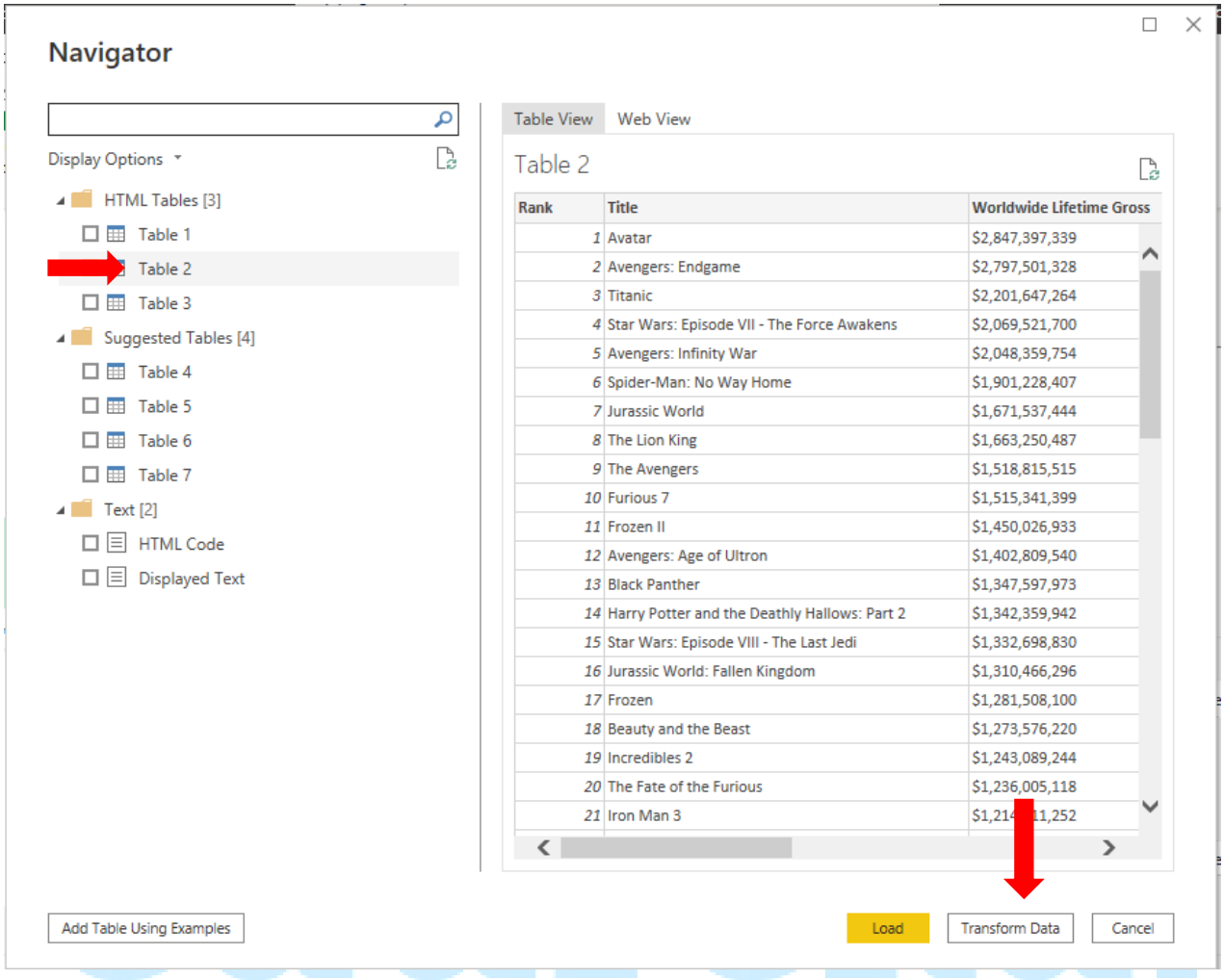

Step 2. Input the following link: https://www.boxofficemojo.com/chart/ww_top_lifetime_gross/?area=XWW

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

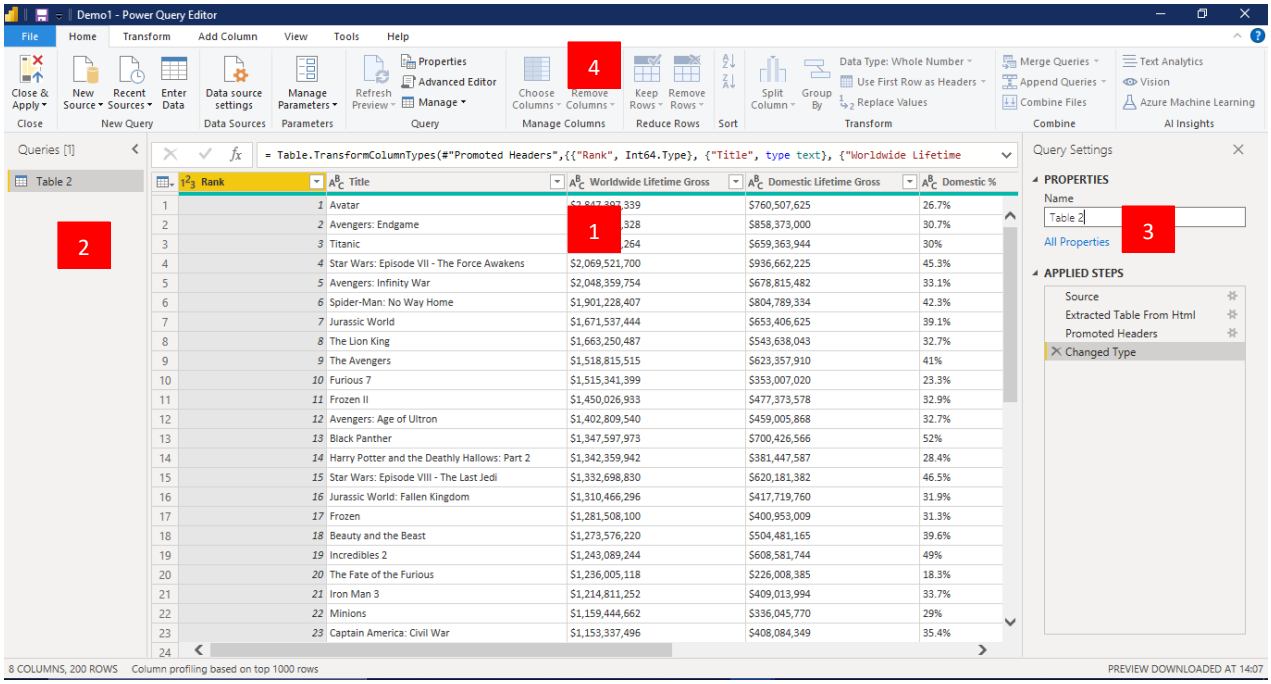



Click on **Transform Data**

**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

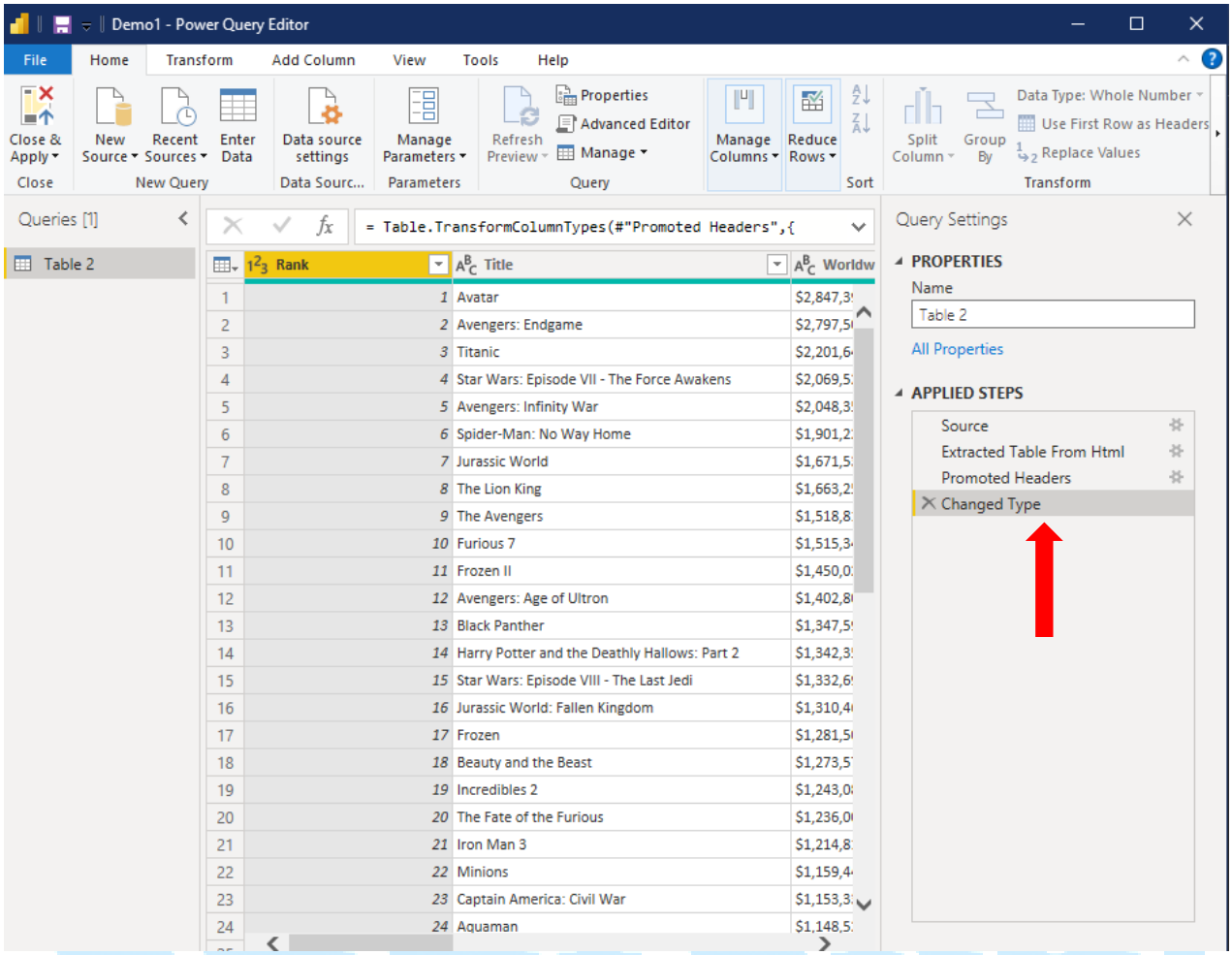
## Query Editor



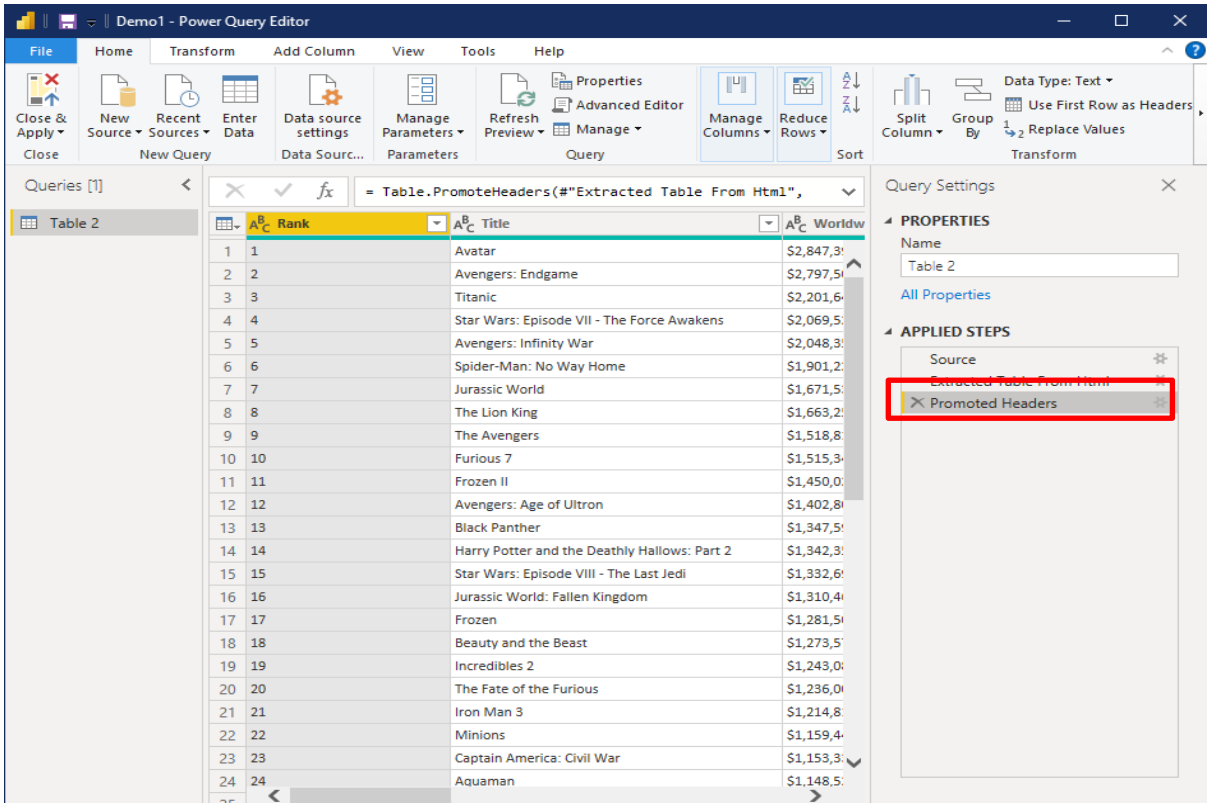

Query Editor has four main sections (numbers matched to screenshot above)

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

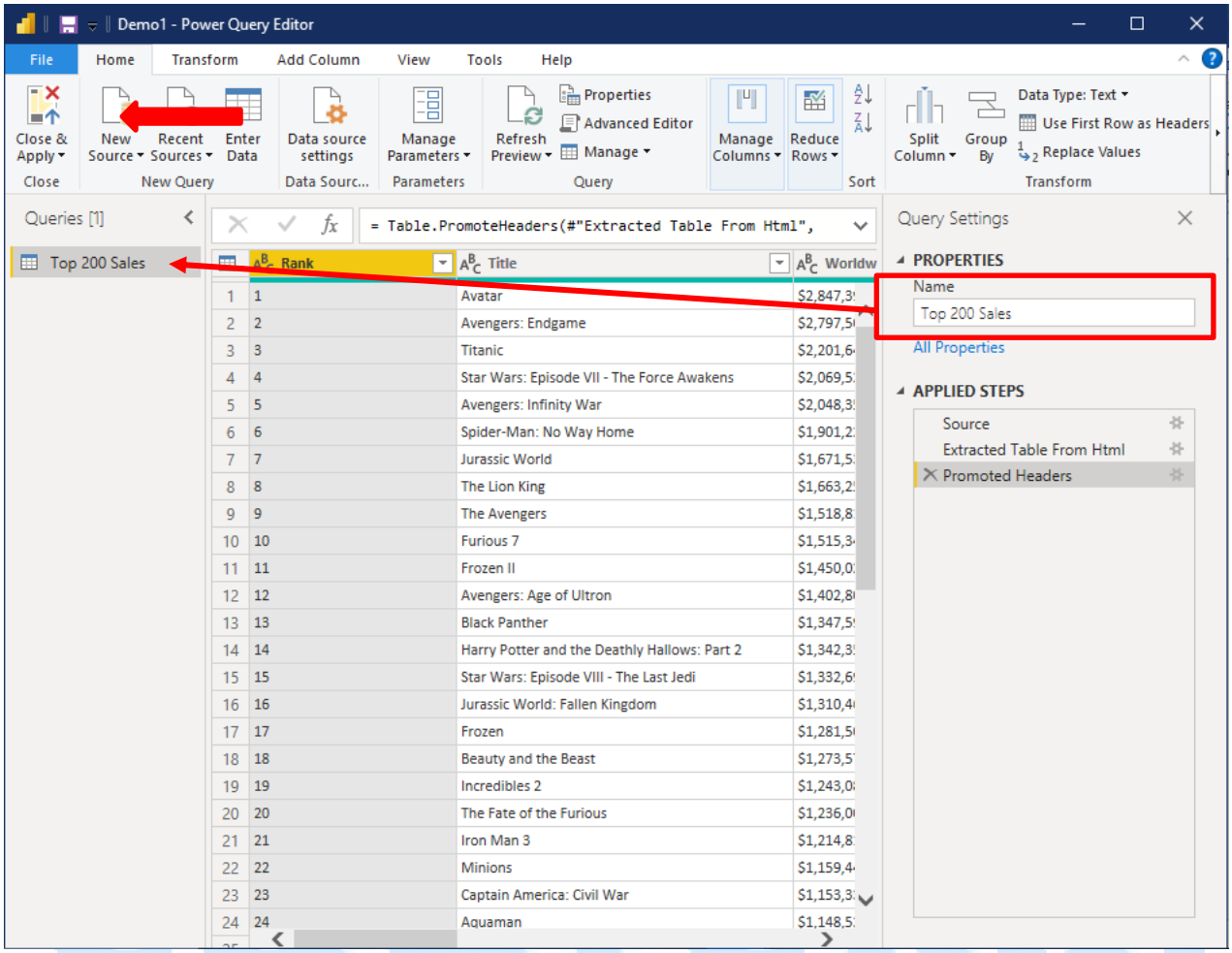



After removing the Changed Type, it will be viewed as:



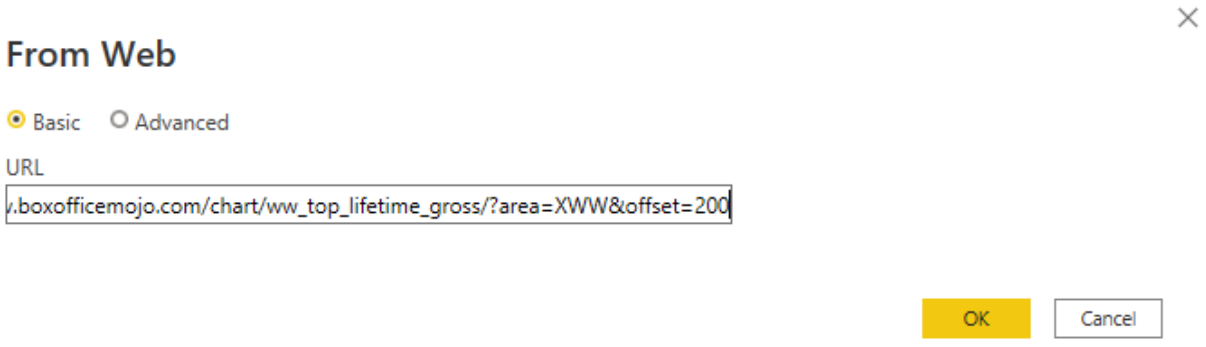

Step 5. Rename the table from Table 2 to → Top 200 Sales

edureka!



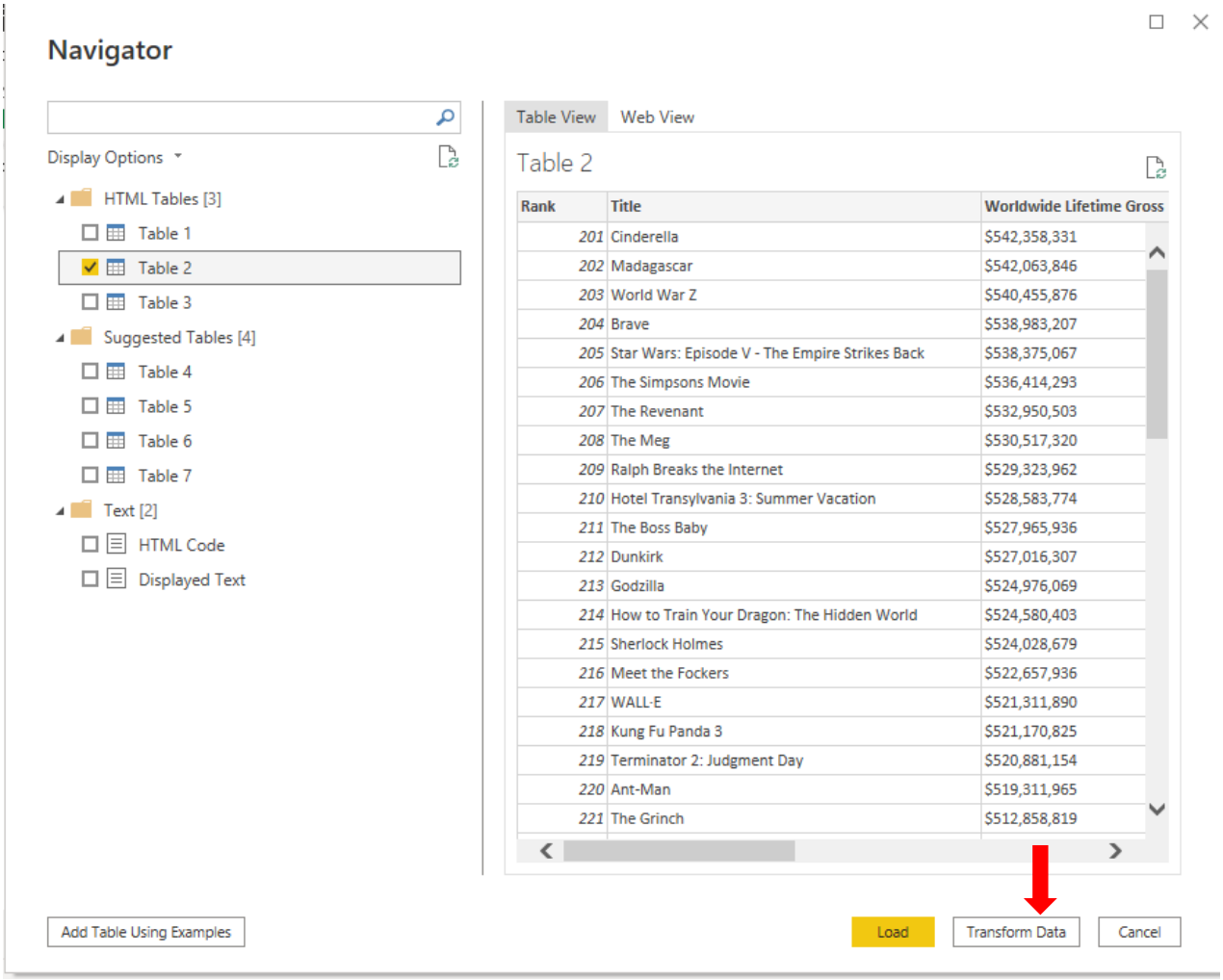

Click on **Close & Apply**

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

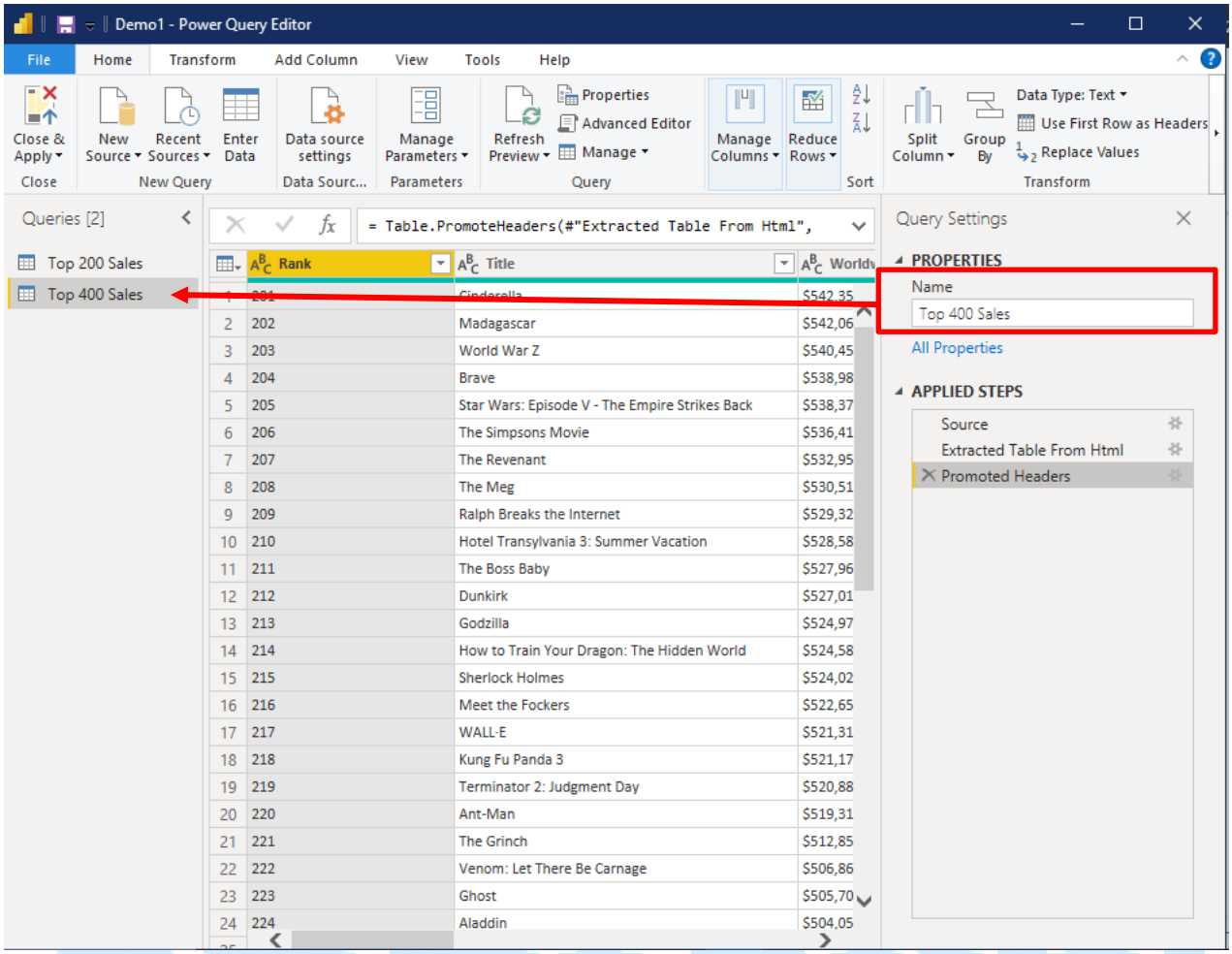



Click on Ok

Step 7. Click on Transform Data and choose appropriate table

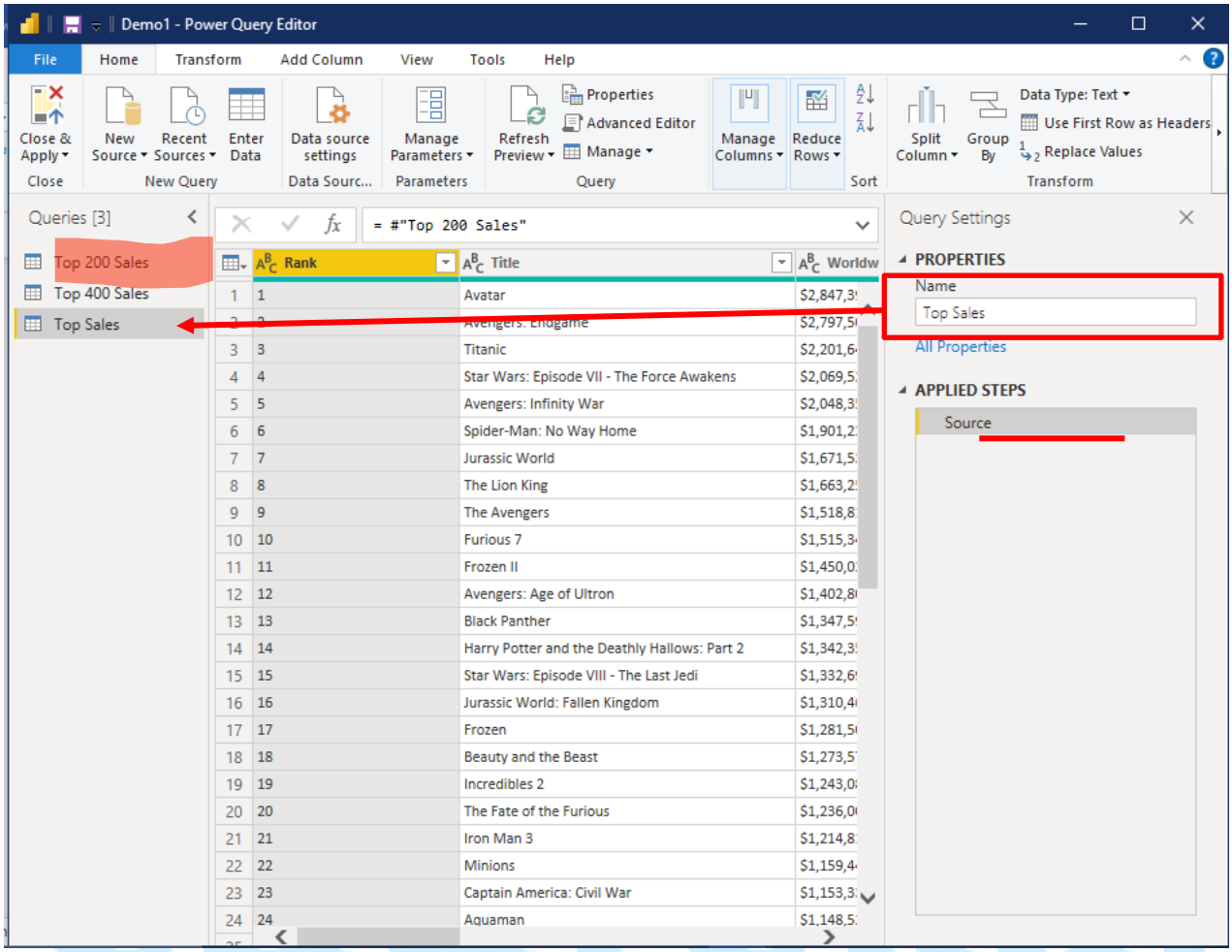



Step 8. Perform Similar above Steps



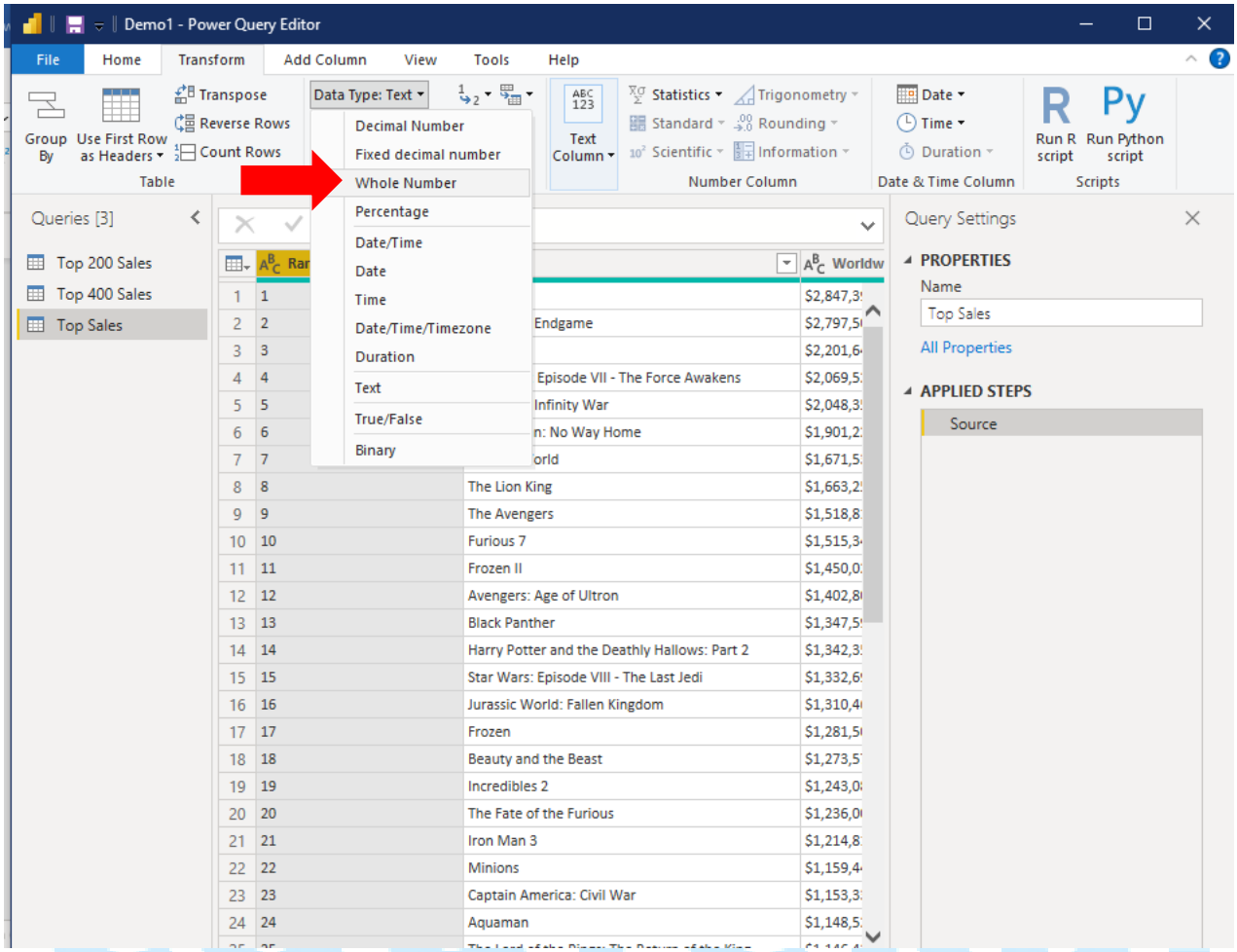

Use Query as Reference

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

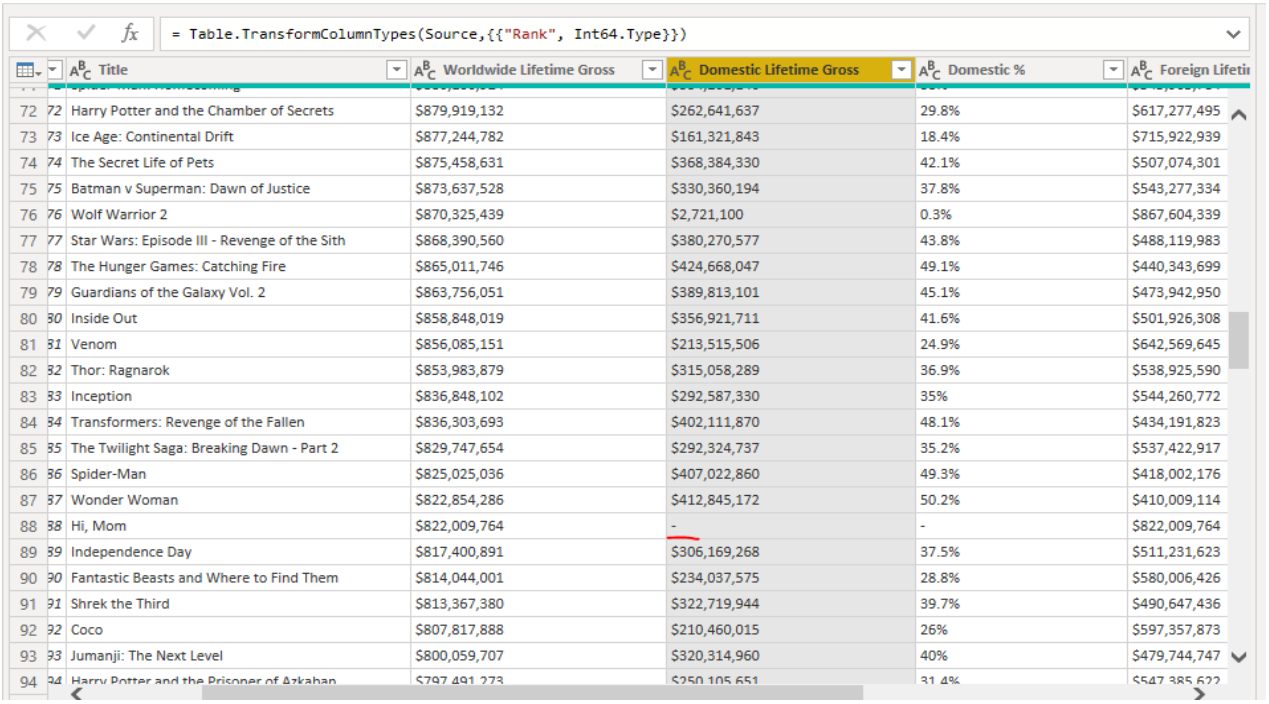



Append Queries

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

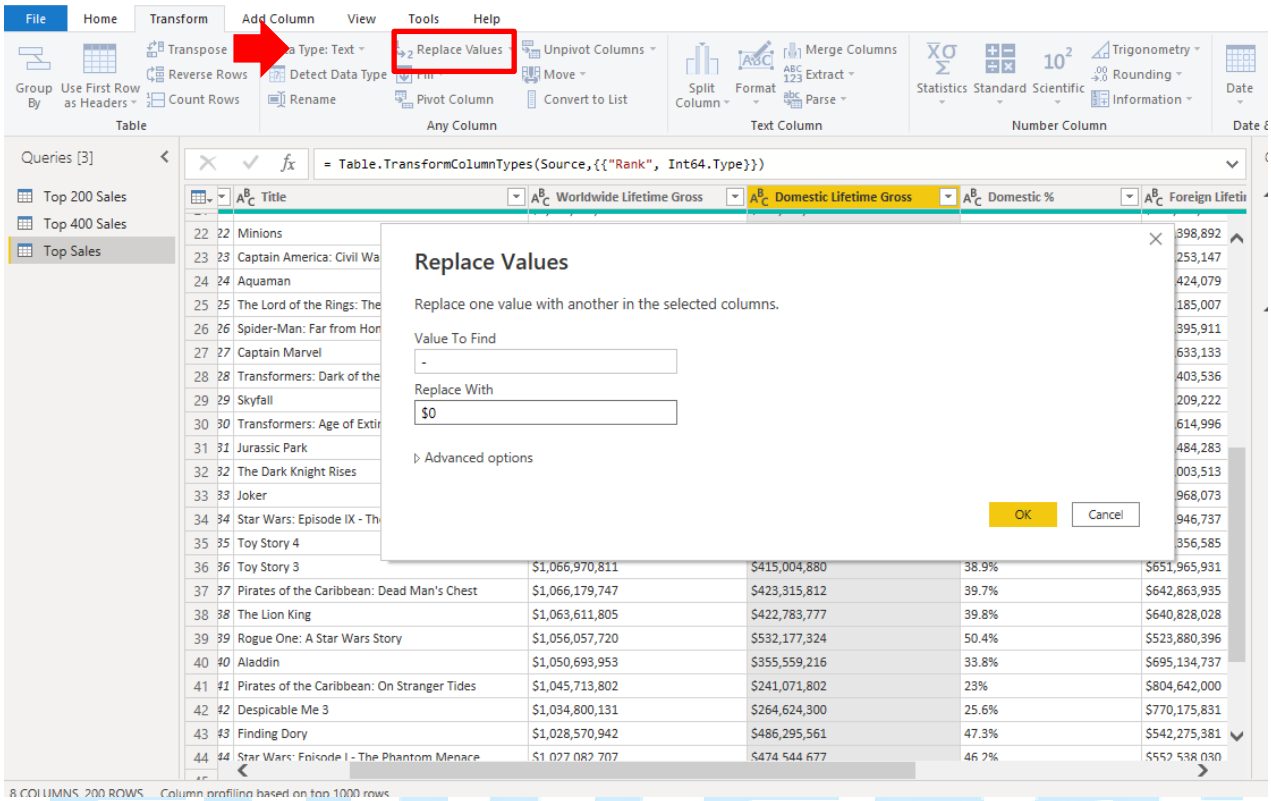



Step 11. There are some missing values in the column **Domestic Lifetime Gross**



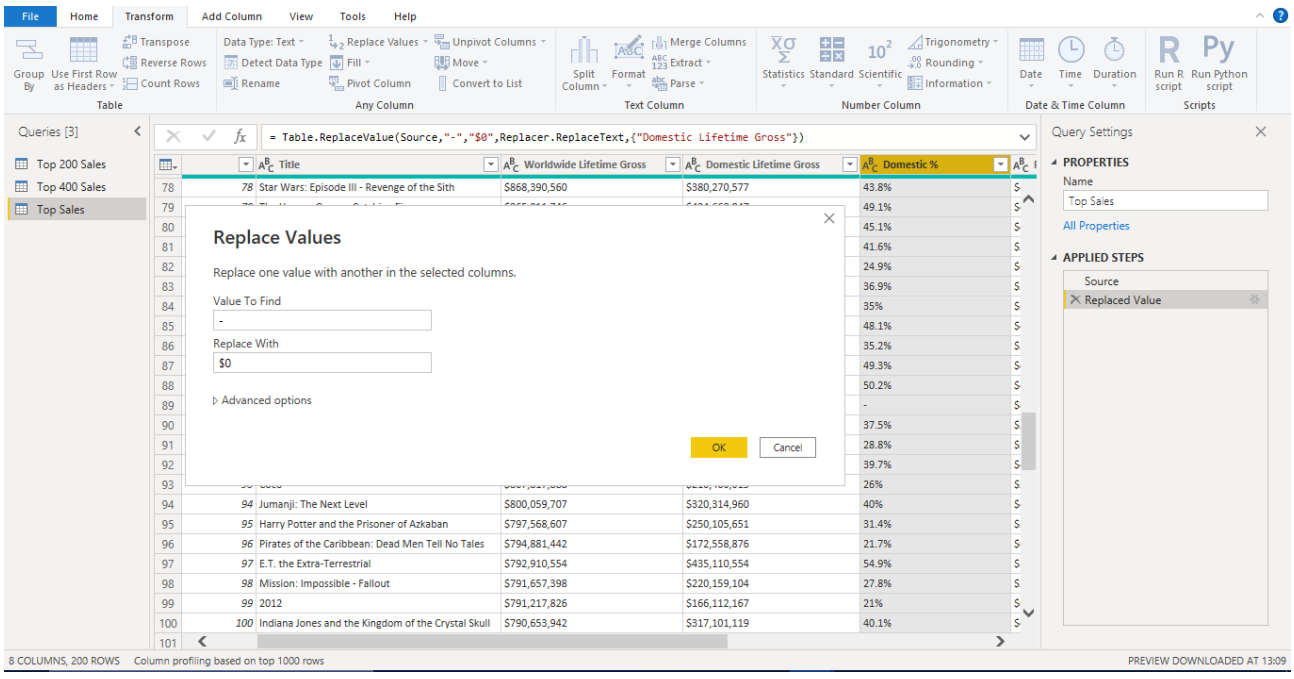

| | Rank | Title | Worldwide Lifetime Gross | Domestic Lifetime Gross | Domestic % | Foreign Lifeti |
|---|---|---|---|---|---|---|
| 72 | 72 | Harry Potter and the Chamber of Secrets | $879,919,132 | $262,641,637 | 29.8% | $617,277,495 |
| 73 | 73 | Ice Age: Continental Drift | $877,244,782 | $161,321,843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



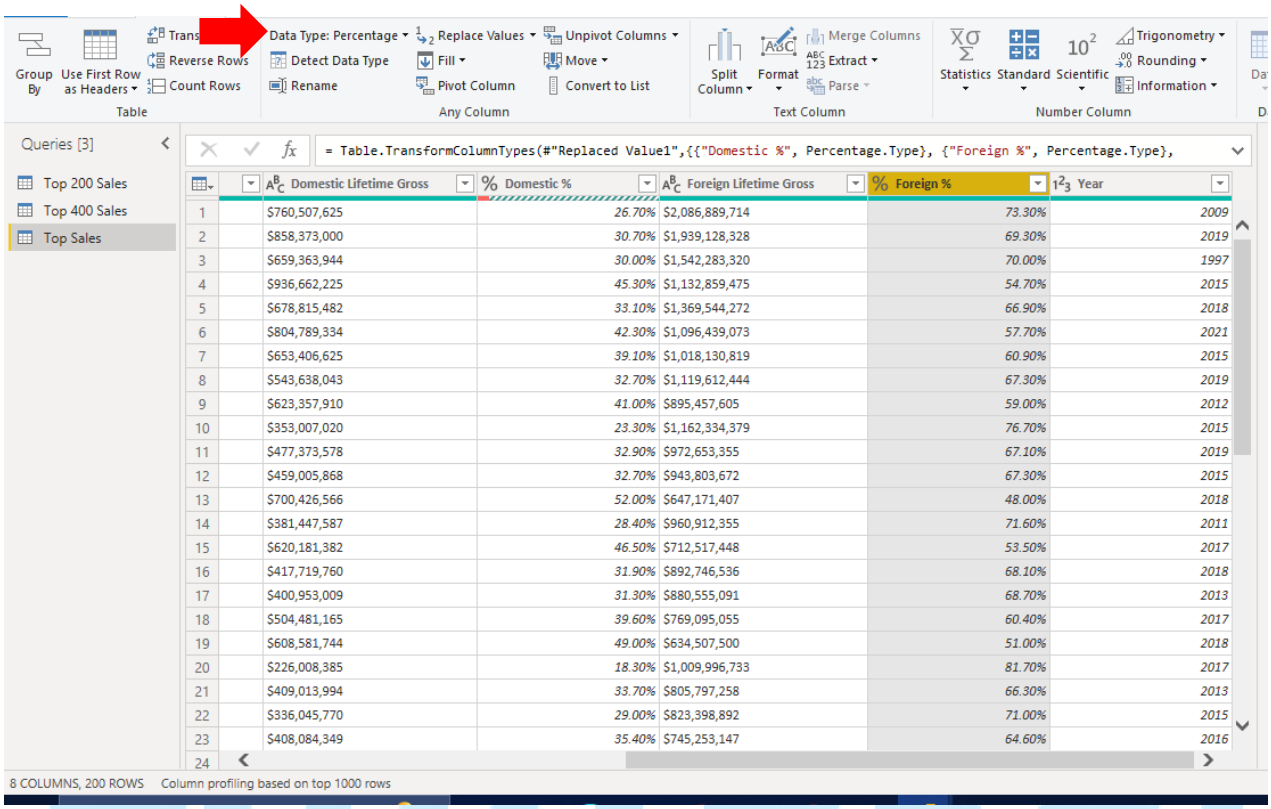

Click on Ok

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

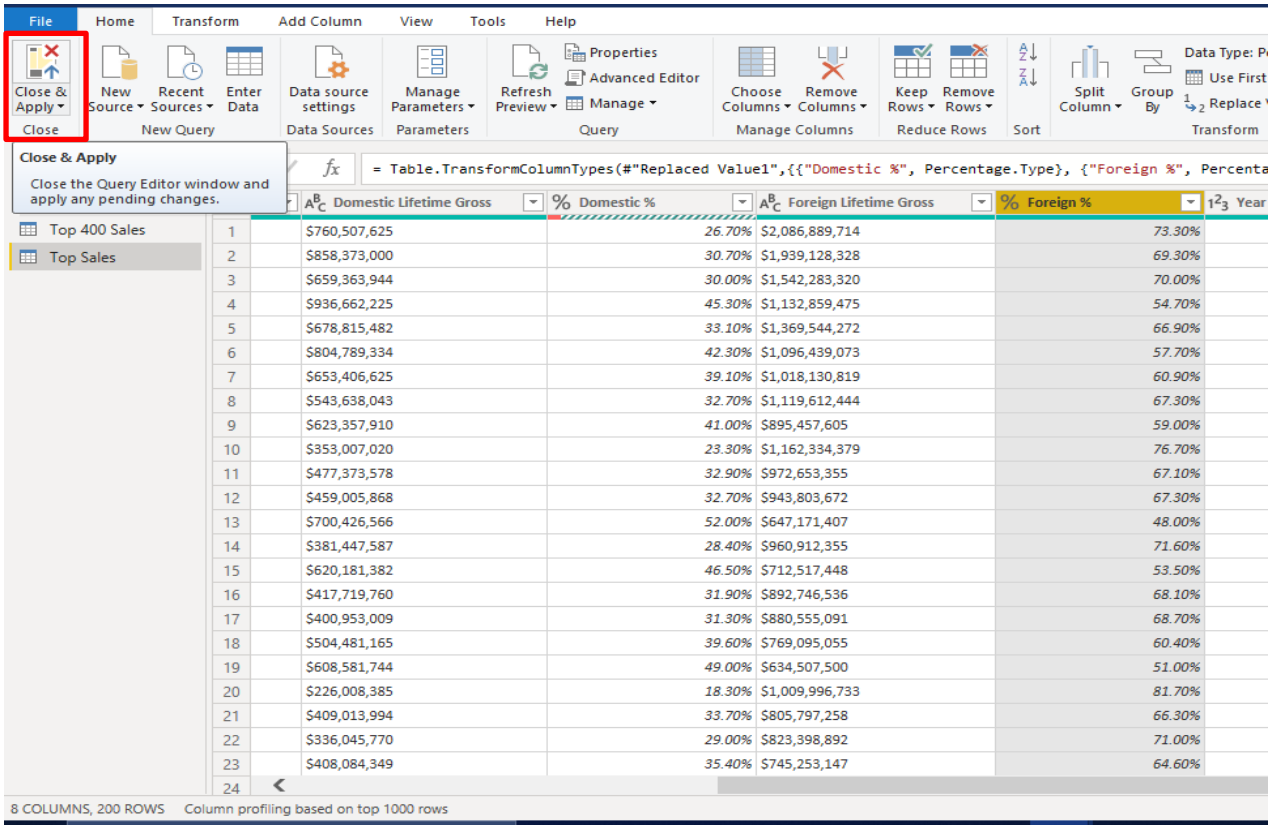



Step 14. Change the data types of the following columns

**Foreign %** column from **Text** to → Percentage
**Year** column from **Text** to→ Whole Number




Step 15. Click on **Close & Apply**

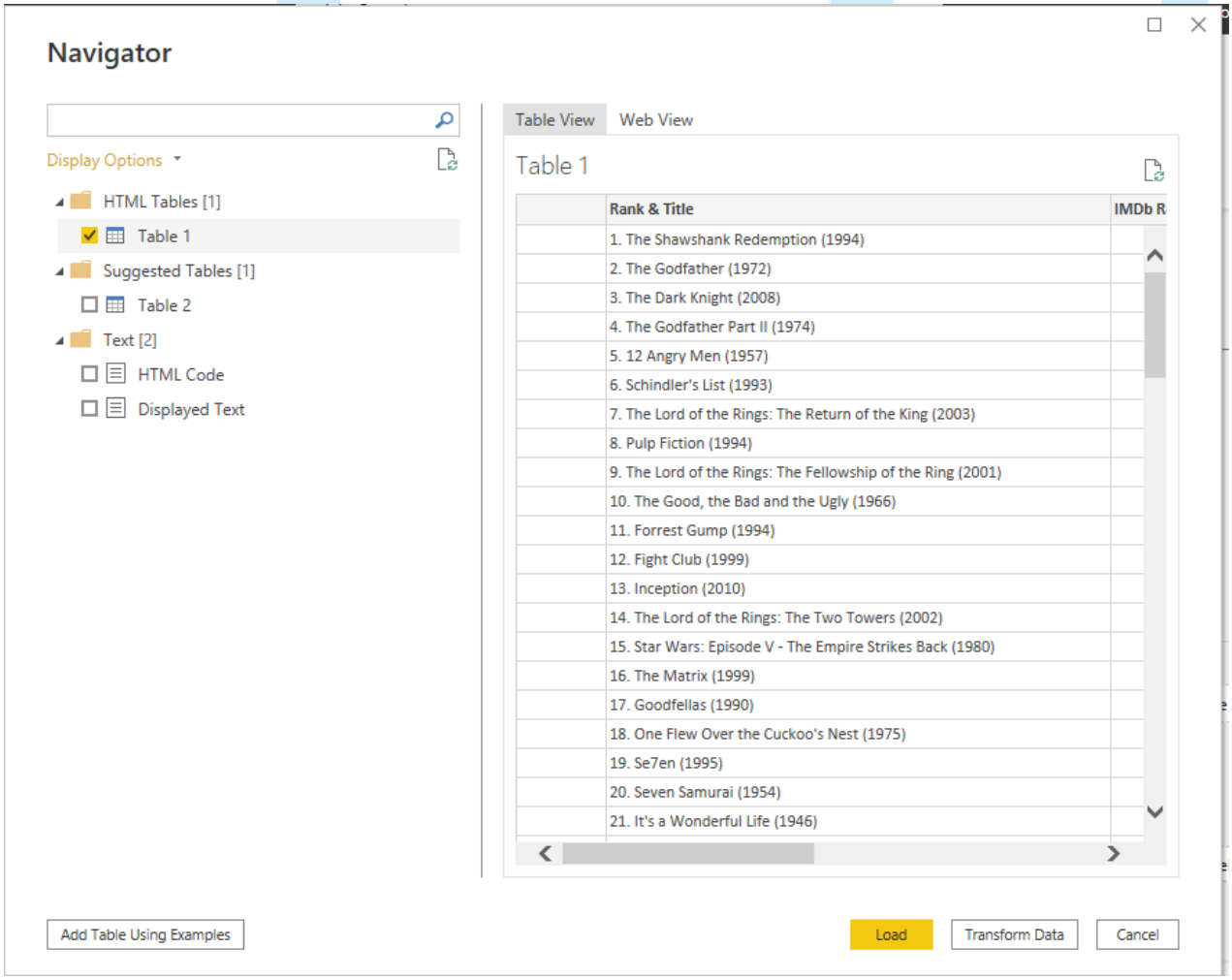


Note: The website owners over time might change the HTML page over time so do the above steps carefully if something isn't working, please try to work it out.

Part - 2

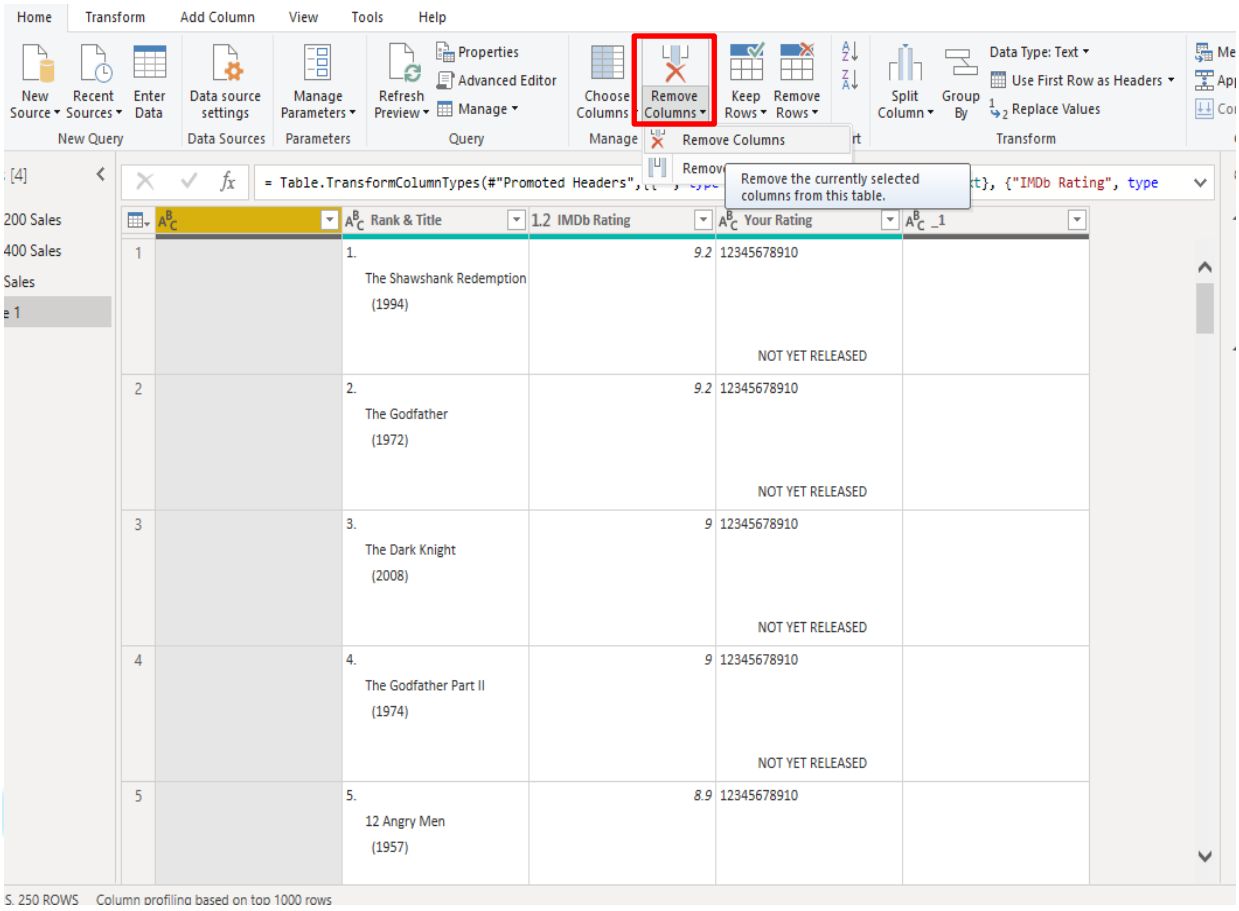
Step 16. Use the IMDB dataset found here
Get Data → Web , input the address

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



## Data Cleaning

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

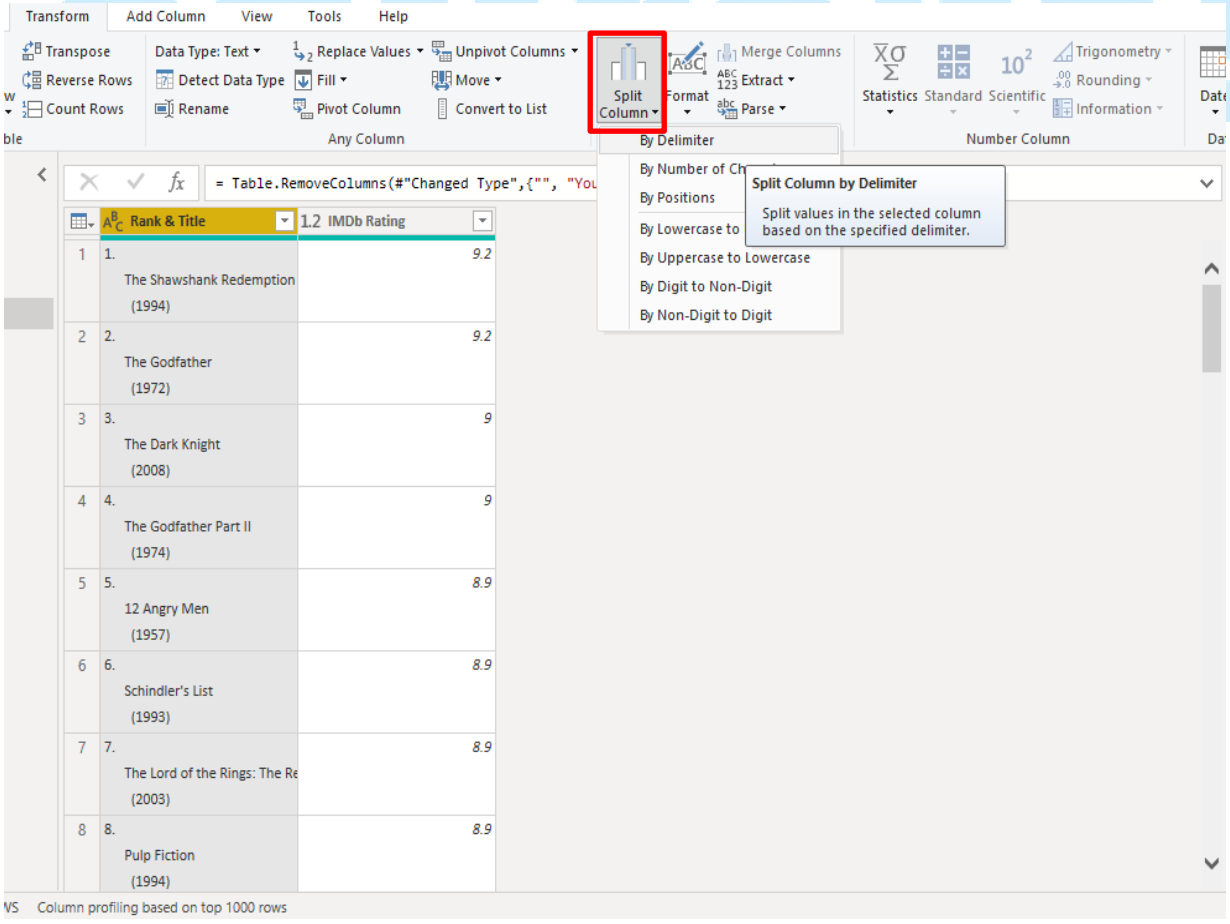



After deletion, it looks as:

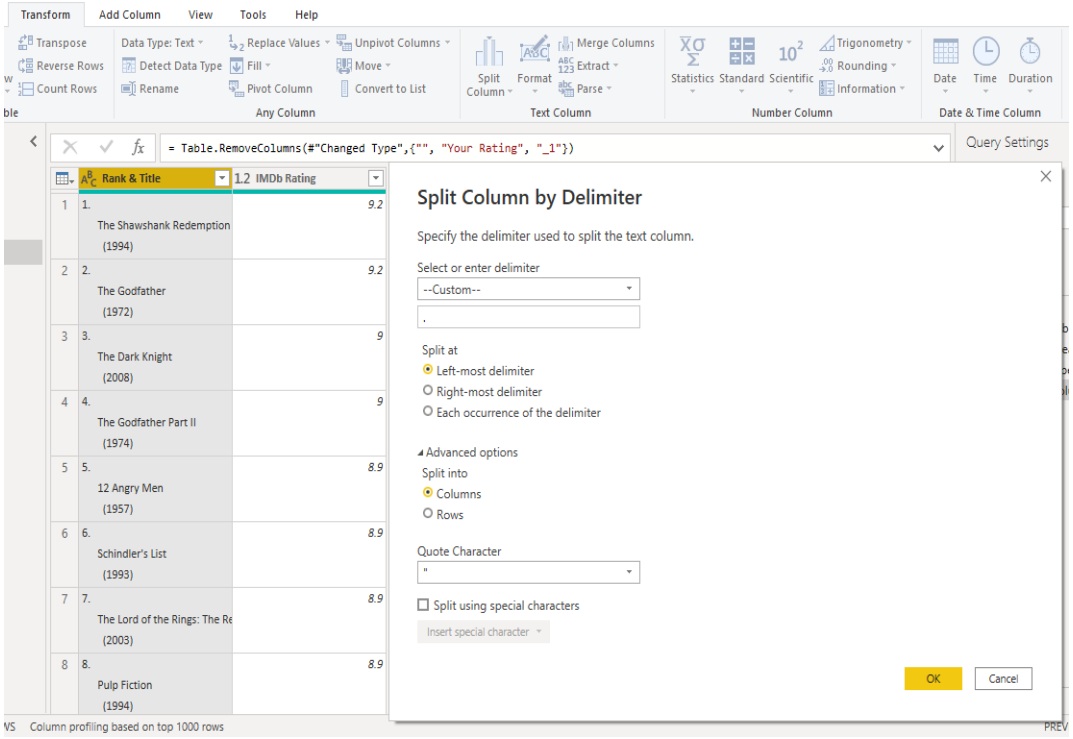
| | Rank & Title | IMDb Rating |
|---|---|---|
| 1 | 1. The Shawshank Redemption (1994) | 9.2 |
| 2 | 2. The Godfather (1972) | 9.2 |
| 3 | 3. The Dark Knight (2008) | 9 |
| 4 | 4. The Godfather Part II (1974) | 9 |
| 5 | 5. 12 Angry Men (1957) | 8.9 |
| 6 | 6. Schindler's List (1993) | 8.9 |
| 7 | 7.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

mn profiling based on top 1000 rows

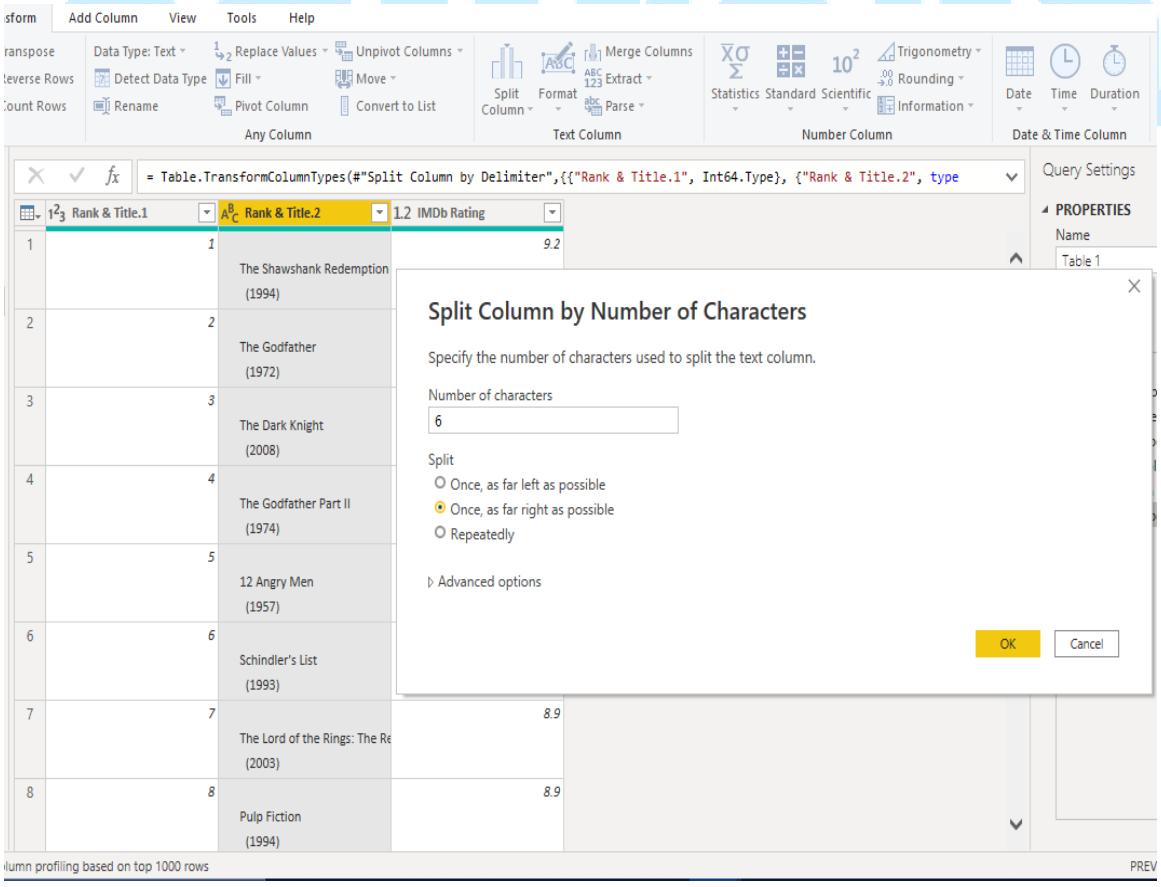
Step 19. Select the 1st Column → Click on Split Column by Delimiter

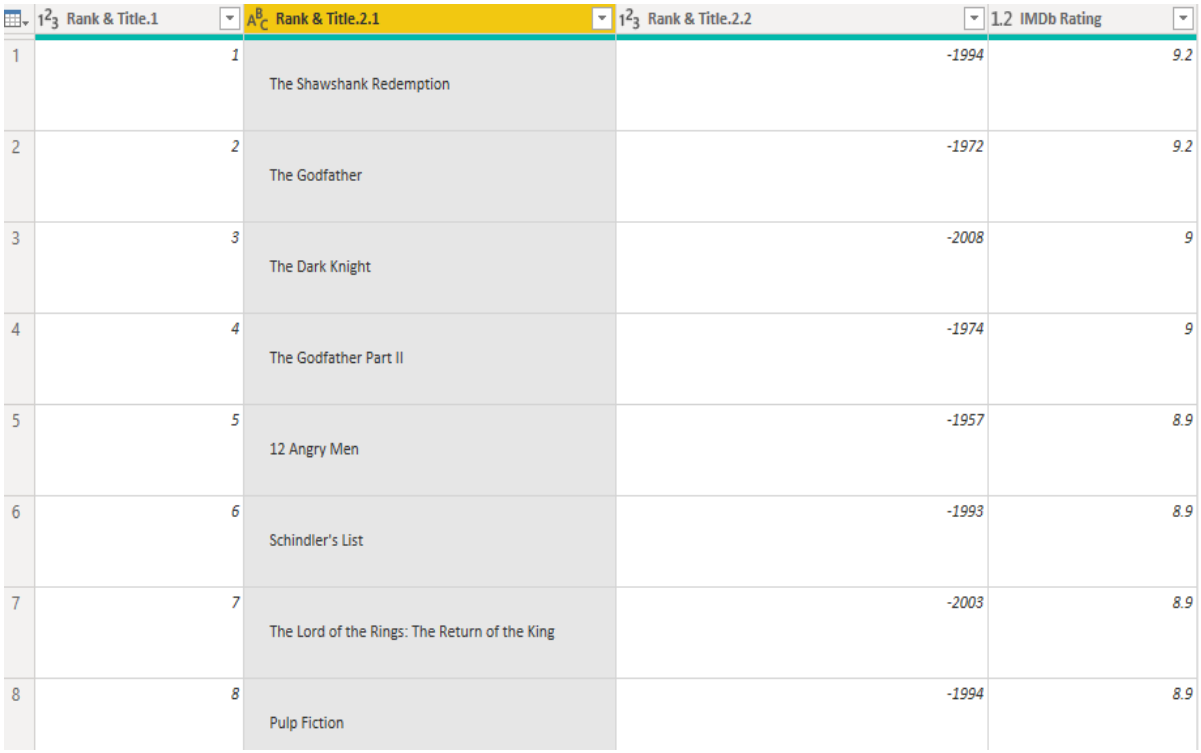


- Select **Custom** Delimiter dot (.) and Click on Ok.




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





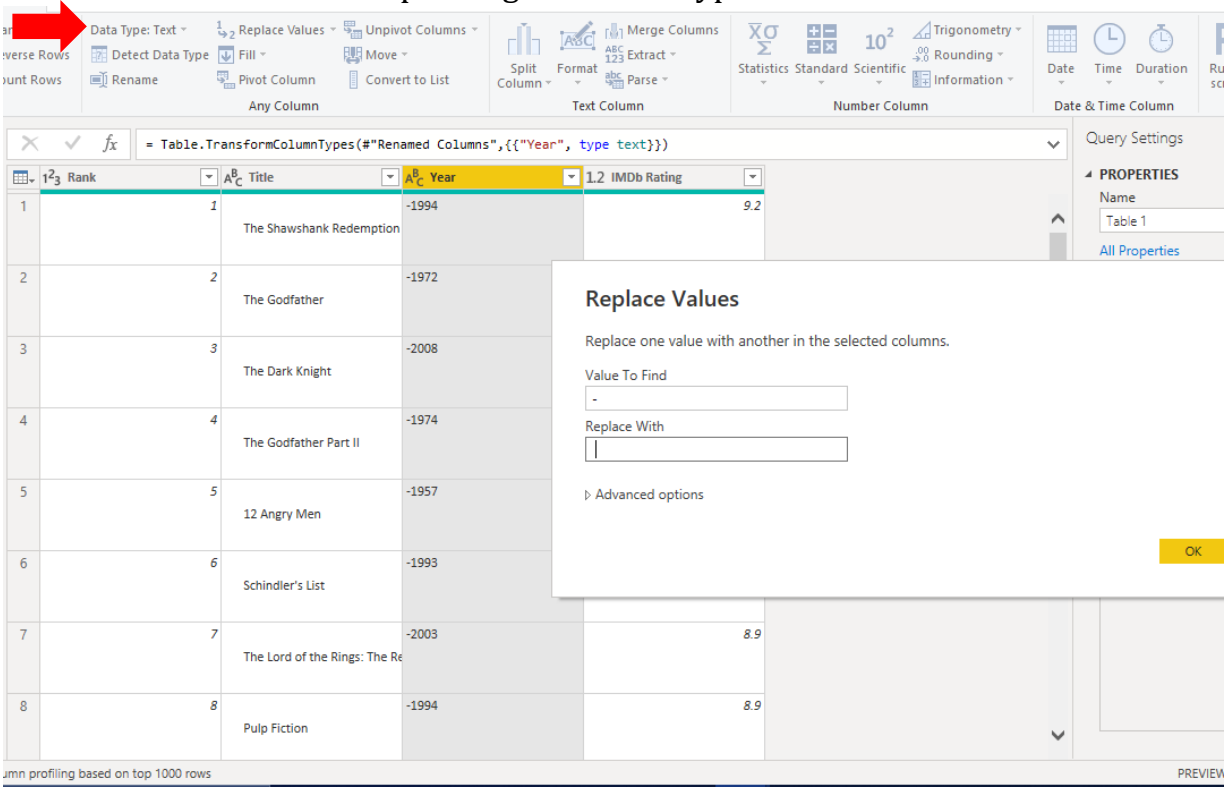
| | Rank & Title.1 | Rank & Title.2.1 | Rank & Title.2.2 | IMDb Rating |
|---|---|---|---|---|
| 1 | 1 | The Shawshank Redemption | -1994 | 9.2 |
| 2 | 2 | The Godfather | -1972 | 9.2 |
| 3 | 3 | The Dark Knight | -2008 | 9 |
| 4 | 4 | The Godfather Part II | -1974 | 9 |
| 5 | 5 | 12 Angry Men | -1957 | 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**

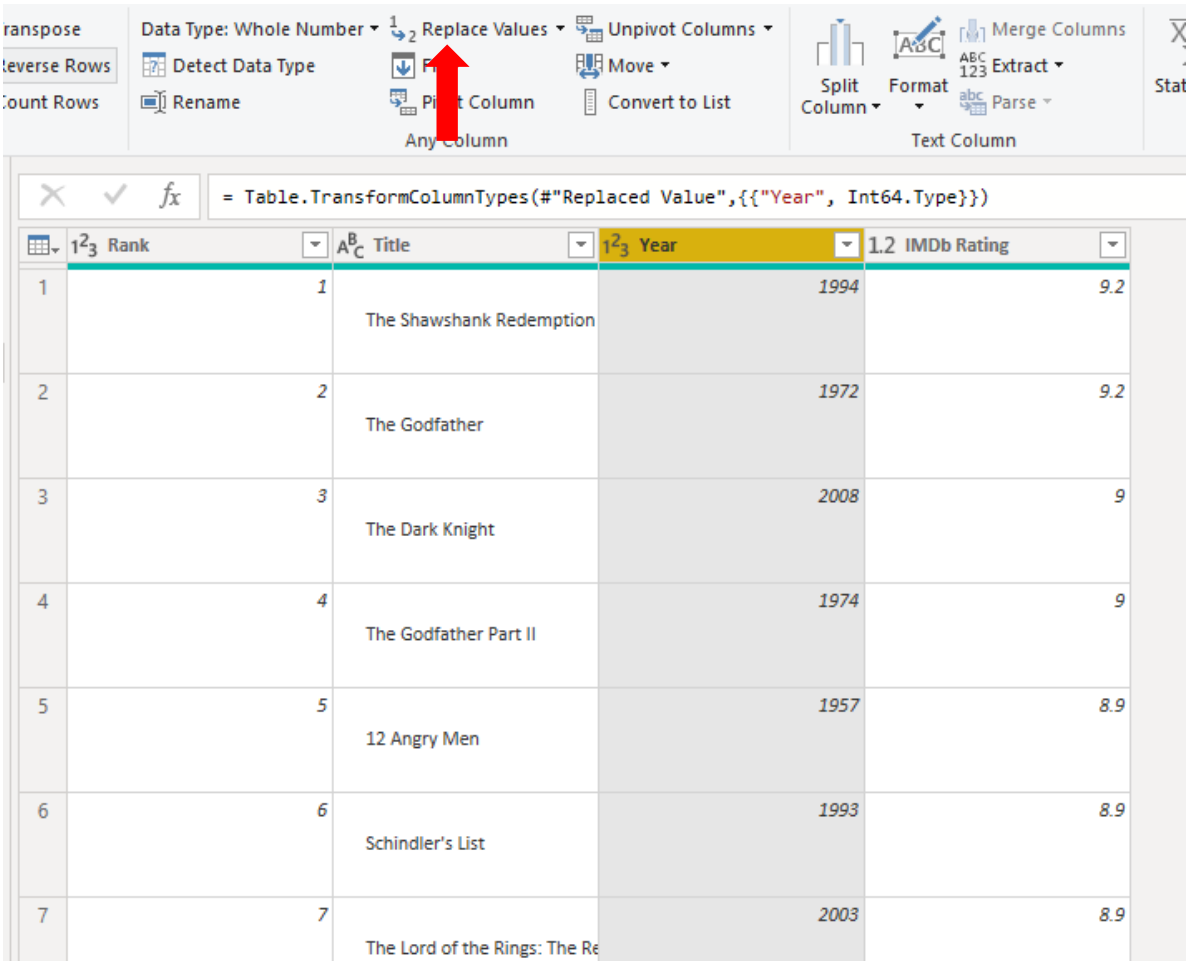
edureka!

| | 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 |
| 8 | 8 | Pulp Fiction | -1994 | 8.9 |

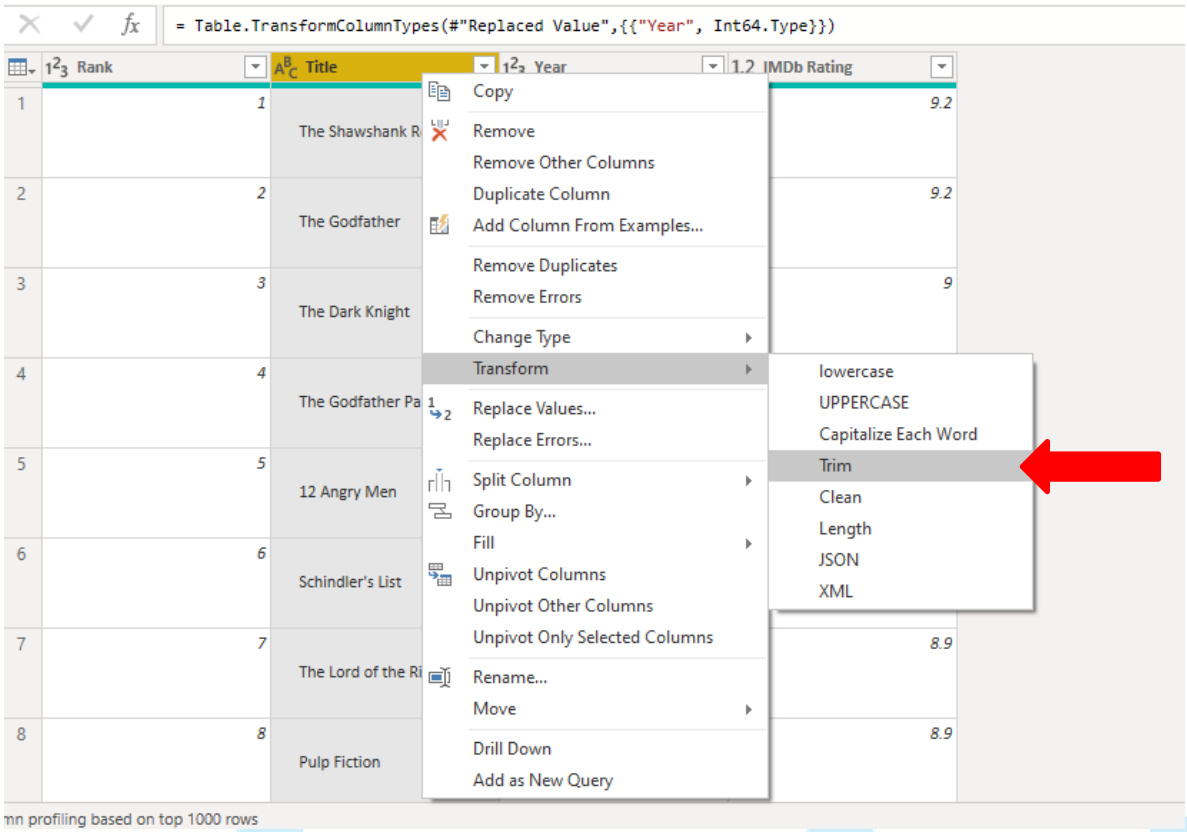
Step 22. The Column **Year** has extra Dash Replace it with space
Before the above step Change the Data Type from whole Number to Text



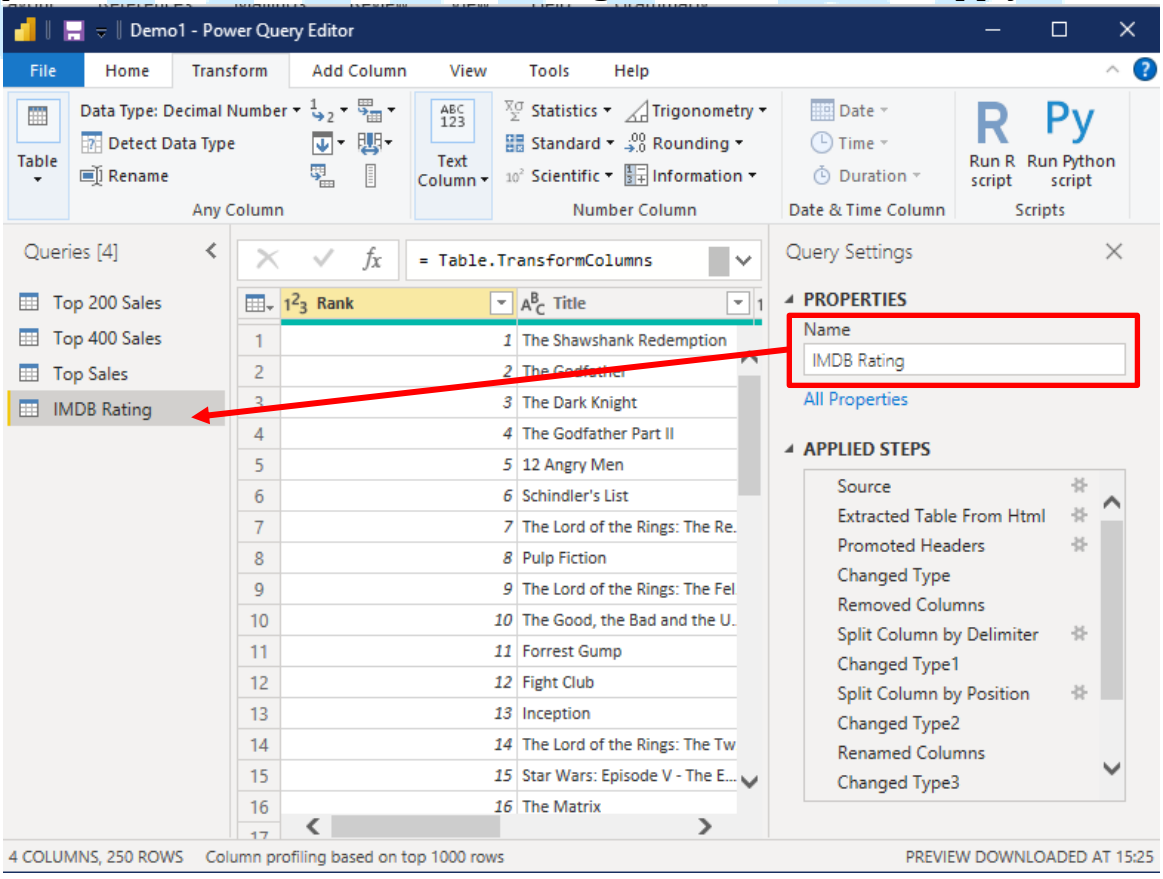
Check if the Data Type is a Whole Number




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



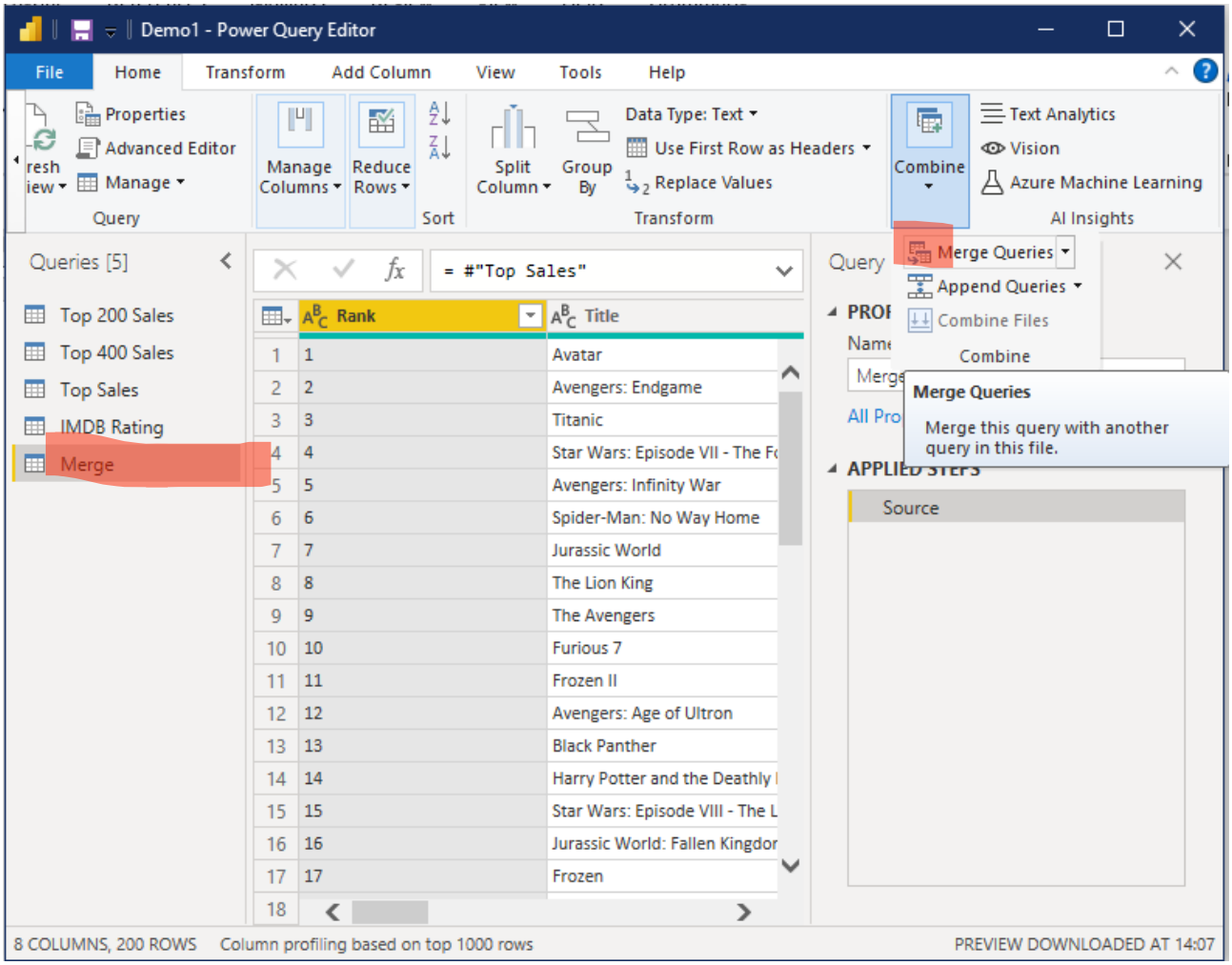

Step 24. Rename the Table to **IMDB Rating** and Click on **Close & Apply.**



Final Merge

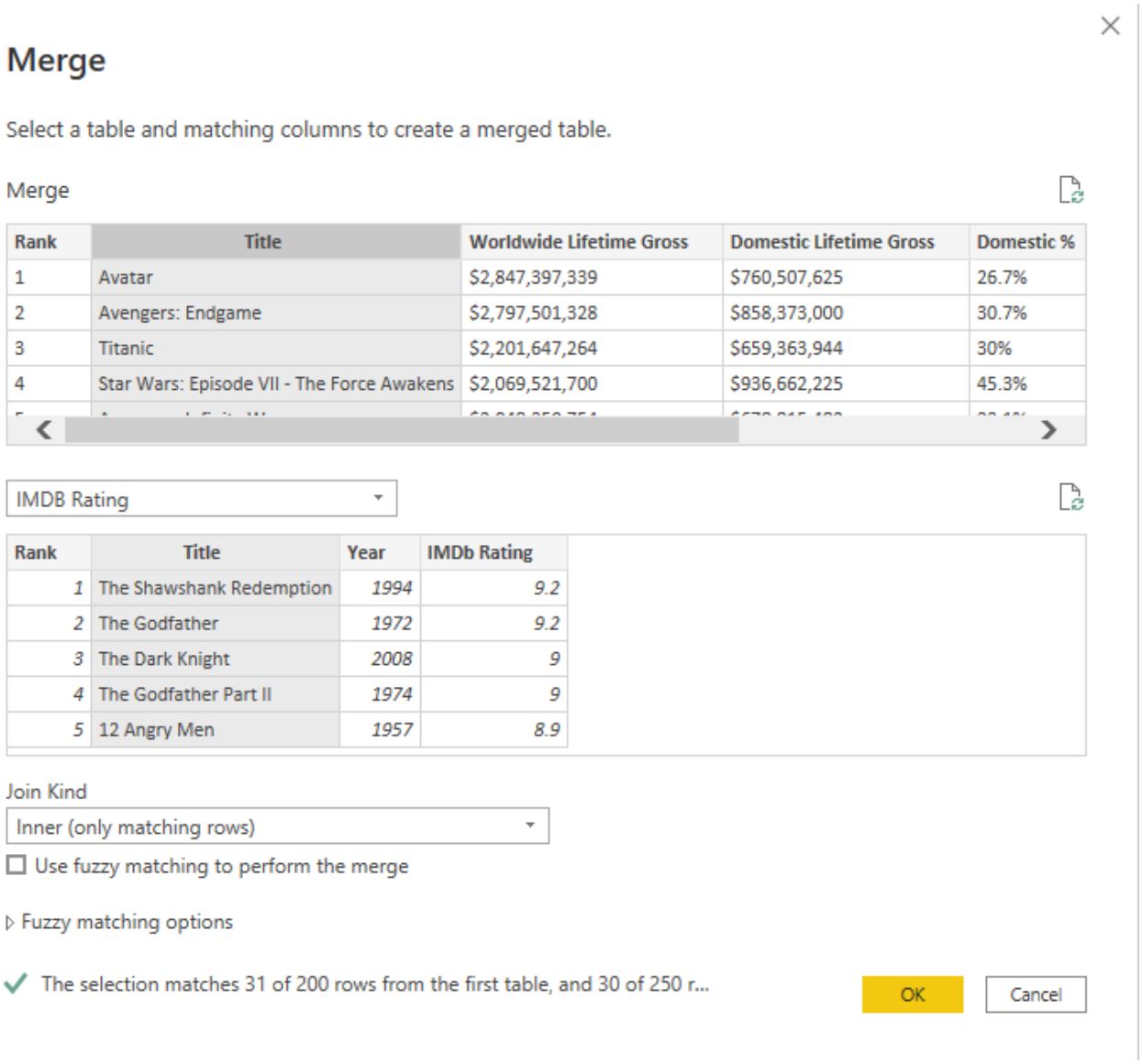
The preparation for both data sets for a final merge together to see how best-selling movies are among top user rated film. Merge these two data sets or join them. Click on the **Top Sales** query and create a Reference for it, name the new query as **Merge**. Now click on **Merge Result**, and then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

Step 25. Merge Queries

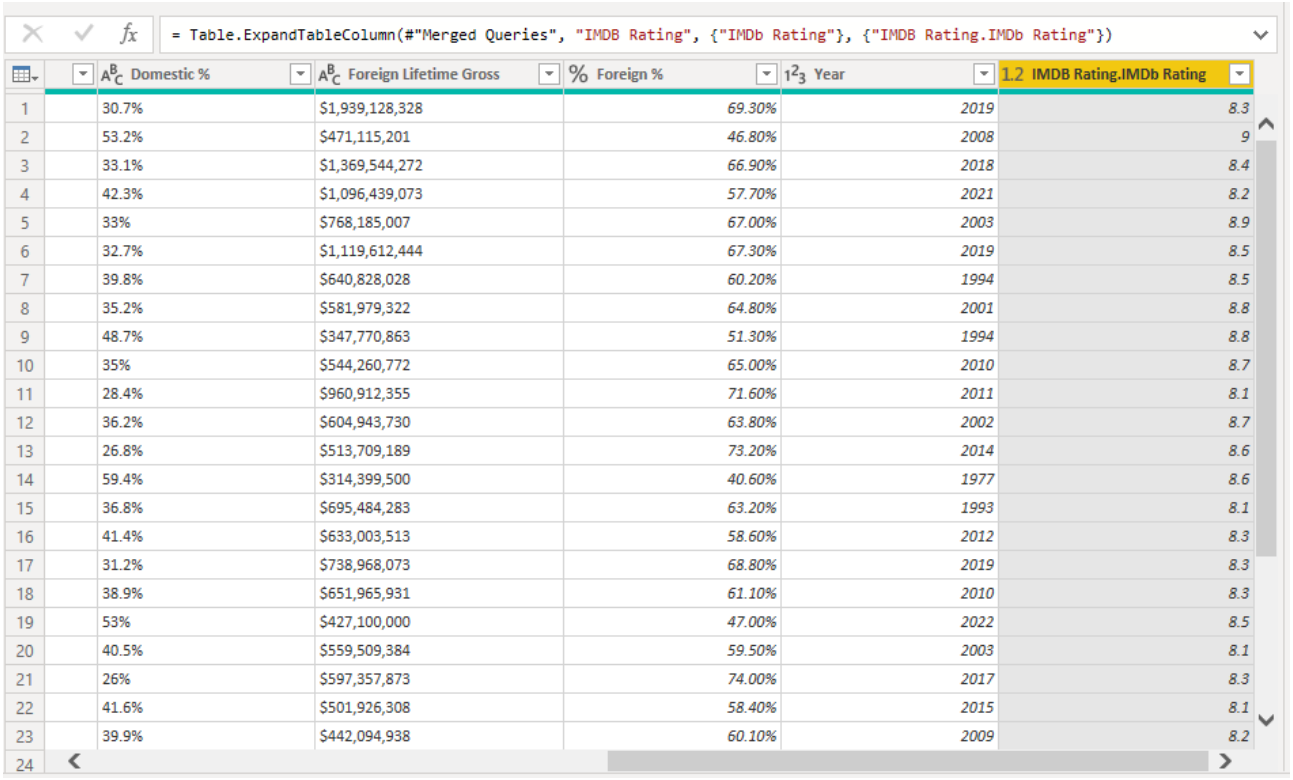



After completion of the previous step, set the following setting

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)




- 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.



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

| | Domestic % | Foreign Lifetime Gross | Foreign % | Year | 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!