**Goal** – connect from PowerBI to the imdb website and collect information about the top 250 movies **https://www.imdb.com/chart/top**

## Suggested Steps

* 1. Start new PowerBI project
  2. Get data – Web
  3. Paste URL of the web page, OK, Connect
  4. On the left, find a table that contains your data and select it (tip : you may encounter duplicate tables detected, so choose the table which will require the least manipulation to isolate the useful data)
  5. We wish to create a single table with the movie rank (1-250), movie title, year of release and user rating
  6. Notice that table structure needs editing as it contains empty / unusable columns and at least one column needs splitting, so click Transform (not Load). Other columns may need cleaning
  7. Add transformation steps to remove any not needed columns
  8. Highlight the column which contains movie title and rank and use Split Column – By delimiter- to capture the data into separate columns, retaining the column you want.
  9. Clean and modify the other columns as required using available methods in powerbi
  10. Rename columns by double-clicking each header (**IMDB rank, Movie title, Release year, IMDB rating**)
  11. Build simple powerbi visuals to demonstrate the data you have collected (examples are provided below)
  12. Now we want to obtain similar information about the top-rated English language movies and we will see two methods to identify movies found in both lists.
  13. Repeat the same steps as above for the url «imdb top 250 english movies»
  14. URL: <https://www.imdb.com/chart/top-english-movies>
  15. Name this table EnglishMovies
  16. Create a method to distinguish which of the top 250 movies overall are in English, by using this second table as a reference. You can solve this using a DAX formula to create a new column with DAX (when in Report mode only) in the top 250 movies table or create a merged query table. A movie is considered English if it can be found in the new table EnglishMovies, and non-English if it cannot be found there. Of course, we must assume no spelling errors, as the only way to link the tables is by the field title (of movie).
      1. DAX examples :

**CONTAINS(lookup\_table, lookup\_table[lookup\_field], target\_table[target\_field])**

**IF(target\_table[target\_field] IN DISTINCT(lookup\_table[lookup\_field]),"label\_true","label\_false")**

* + 1. Merged query :

From the Transform data (powerquery) window, use the Combine options to create a merged query as new, linking your two existing movie queries.

Choose the merge (join) type which prioritises the top250 movies and returns matching results from the englishlanguage movies.

Identify a method (this may occur over two steps) to replace null values with a label ‘non-English language’ and all remaining values with ‘english language’ (hint: add conditional column)

**Example Visuals for top 250 movies**

Chart, line chart

Description automatically generated

Chart, background pattern

Description automatically generated

**Sample Visual for English Language Movies**

**Chart, pie chart

Description automatically generated**