**Data Clean-Up Process**

A CSV of Movie Data from The Numbers was provided as an academic sample of films.

The first step in the clean-up process was importing the CSV and reading it into a DataFrame.

I then added two columns to the DataFrame:

* **Total Box Office** which took the sum of the Domestic and International Revenue.
* **ROI** which divided the Total Box Office by the Production Budget, with a multiplier of 100 for a percentage value.

API Tests:

* Sample movie name from the CSV DataFrame.
* Test to pull indexed values

I then converted the movie\_name column from the CSV DataFrame into a list for movie titles.

I printed to check length of values.

Ran the titles through an API Loop.

Then created a new cell to append the necessary API values into lists for each movie title:

* Title, Country, Released, Production, imdbRating, MetaScore, and Rotten Tomatoes.

Note: For all the films it was not able to find stats for I created an exception to print as Not Found. I then imported these into a list and ran through IMDB to find the matching title and correct typos or foreign translations and then corrected in the original movie data CSV. There were about 111 missing titles, this was brought down to 34.

Then imported API list values into a DataFrame.

Ran counts. 1931 x 7

Dropped Null Values. 1913 x 7

Merged with CSV DataFrame on=”movie\_name” 1826 x 21

Reduced Columns for necessary analysis. 1826 x 13

Dropped Null Values. 1800 X 13

And then saved DataFrame to CSV in Resources folder to use in following analysis: