Lab 11

Prepare the movie release dataset of all the movies released in the last 5 years using IMDB.

(a) Find all movies that were released in the last 5 years.

(b) Generate a file containing URLs for the top 50 movies every year on IMDB.

(c) Read in the URL's IMDB page and scrape the following information:

Producer(s), Director(s), Star(s), Taglines, Genres, (Partial) Storyline, Box office budget, and Box office gross.

(d) Make a table out of these variables as columns with movie name being the first variable.

(e) Using **Power BI**, analyze the movie-count for every Genre. See if you can come up with some interesting hypotheses. For example, you could hypothesize that "Action Genres occur significantly more often than Drama in the top-250 list." or that "Action movies gross higher than Romance movies in the top-250 list."

(f) Write a markdown doc with your code and explanation. See if you can storify your hypotheses.

Note:
- You can web-scrape with the rest package in R or Python.
- Don't use any scrapping tool.

Instructions:

- **This is a graded assignment (Max 5).**

- You can make a group of 2 students.

- Due Date: **11.55 pm Thursday, 17th December 2020.**

- **No extensions will be given**: Penalties are applied to late submissions. Note that delays caused by computer downtime cannot be accepted as a valid reason for a late submission without penalty. Students must plan their work to allow for both scheduled and unscheduled downtime.