Congratulations!

Your team has been hired by Lionsgate to complete a data science consulting job for them. Lionsgate has struggled recently and the CEO said they are “looking at ways to better align our company with our industry's evolving landscape.” Therefore your team is charged with doing data analysis and creating a presentation that explores what type of films are currently doing the best at the box office. You must then translate those findings into actionable insights that the CEO can use when deciding what type of films they should be creating.

Methodology

You want to examine the profitability of films and compare films across variables. Some examples of these variables are movie types (Thriller, Drama, Comedy, etc.), movie ratings, budget, social media discussion, and critic or user reviews. Your team must use its knowledge of probability distributions (mean, standard deviation, correlation etc.) to determine where there are significant differences in the profitability of films. You can compare films across variables (Are action movies more profitable than dramas?) or across time periods (Scar movies have been more profitable this decade than last). Your team is expected to use git as a collaborative tool for this project to manage version control and history.

Data Sources

Lionsgate does not have any data, so your team can use your web scraping and API skills to pull your data. Some great sites to pull this data are Box Office Mojo, IMDB, Rotten Tomatoes, and TheMovieDB.org. You are not limited to these sites, and can pull your data from any reputable site.

Deliverables

Your team must prepare a 4-5 minute presentation that gives the CEO insights as to what type of films they should be creating to meet consumer demand. Your presentation should outline the process you went through and use at least 4 meaningful data visualizations to help illustrate your findings.

Project Checklist:

- ❏ Pull data from at least two sources
    - ❏ Establish naming conventions for variables and datasets
    - ❏ Clean dataset & **record parameters** used to clean the data
        - ❏ You may use Pandas or Python functions
    - ❏ Put the data into a **sqlite database**

- ❏ Use Pandas to get useful statistics for comparing films

- ❏ Posted to git repo:
    - ❏ A readme listing lab members, project goals, etc
    - ❏ A Jupyter notebook of **clean and commented code** so an independent party can replicate your analysis
    - ❏ Your final joined and cleaned dataset that was used for analysis
    - ❏ A **narrative** Jupyter notebook that provides:
        - ❏ The purpose of your analysis and why it matters
        - ❏ 4 **well annotated** visualizations created using Matplotlib/Seaborn
        - ❏ 4 meaningful summary tables
        - ❏ Give at least two actionable insights (What type of films should they be looking to produce? What should the budget requirements be? Should they recruit certain actors for their films?)

- ❏ Instructors and TCFs should be able to view a commented commit history for each element in the git repo