Congratulations!

Microsoft sees all the big companies creating original video content, and they want to get in on the fun. They have decided to create a new entertainment studio, but the problem is they don't know anything about creating content. They have hired you to help them better understand the movie/television industry.

Your team is charged with doing data analysis and creating a presentation that explores what type of content is currently the most popular and successful. You must then translate those findings into **actionable insights** that the CEO can use when deciding what type of shows or films they should be creating. You can define popularity and success as you see fit: critic ratings, user ratings, tv ratings, box office earnings, actors, etc. You will want to decide first whether you want to look at television shows or movies. From there you can further narrow your analysis by time frame, genres, etc.

Data Sources

Microsoft 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, Metacritic, and TheMovieDB.org. You are not limited to these sites, and can pull your data from any reputable site. As you are acquiring your data, you need to put in a place where all team members can access it. To do so, you must set up a DB and push all of your data into the DB so that others can access it.

Methodology

Some areas you can look to examine are movie/show genres (Thriller, Drama, Comedy, etc.), ratings, budget, social media discussion, and critic or user reviews. Your team gets to define its own questions about the entertainment industry and then use its knowledge of descriptive statistics to answer those questions.

Deliverables

Your team must prepare a 5 minute presentation that gives the CEO insights as to what type of content 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. Your team is expected to use git as a collaborative tool for this project to manage version control and history, as well as an AWS DB instance to hold your data.

Project Checklist:

- ❏ Pull data from at least 2/3 sources (1 data source per team member).
- ❏ Create a database on **your current AWS instance** to hold your data
- ❏ Develop an ETL pipeline for your data
- ❏ Generate descriptive statistics to compare the success of different films
- ❏ A slideshow presentation that entails the following:
    - ❏ The purpose of your analysis and why it matters
    - ❏ 4 **well labeled** visualizations created using Matplotlib/Seaborn
    - ❏ 2 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?)

- ❏ Posted to git repo:
    - ❏ A *readme.md* listing lab members, project goals, etc
    - ❏ Python files and/or Jupyter notebook or of **clean and commented code** so an independent party can replicate your analysis
    - ❏ A link to your Google slide show

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