Project Title: Marvel v DC

Team Members: Nikki Alfrey, Drew Bollinger, Nathan Savage, Frenci De la Cruz

Project Description/Outline:

Research Questions to Answer:

1. Golden age within years?
2. IMDB score correlation to length of show/movie/minutes
   1. Animated vs live action
   2. Movies vs TV shows
3. Profits/costs
   1. Box office revenues – budgets vs profits
   2. Opening “weekend” profits
4. Directors – do they have higher scores, or do they net higher profits
5. Worldmap – API – did the movies/tv shows net higher dollars in US or Asia (examples)

Dataset(s) to be used:

<https://www.kaggle.com/datasets/willianoliveiragibin/marvel-vs-dc>

<https://www.kaggle.com/datasets/jcraggy/marvel-vs-dc-imdb-rotten-tomatoes>

<https://www.kaggle.com/datasets/darrylljk/marvel-cinematic-universe-films-box-office>

<https://www.kaggle.com/datasets/davidgdong/marvel-cinematic-universe-box-office-dataset>

<https://www.kaggle.com/datasets/mdtoomey/box-office-of-dc-and-marvel-superhero-movies>

<https://www.kaggle.com/code/cliomaciel/eda-dc-marvel/notebook>

Rough breakdown of Tasks:

Andrew –$$ finances

Nikki – ppt/slide deck

Nathan – wording/summarization

Frenci – ace

Timeline/tasks

Nikki to send out zoom for 10/14 meeting to work on project

Gather appropriate datasets (**filter data to 10-15 years 2004-2019** and filter out COVID years to avoid data bias) and clearly define questions (at least 3) *TARGET COMPLETION: 10/14/2024*

Matplot lib

Numpy

Scipy

Pprint

Geoapify

hvplot

Merge *TARGET COMPLETION: 10/15/2024*

Clean data set *TARGET COMPLETION: 10/16/2024*

Define at least 3-4 clear questions *TARGET COMPLETION: 10/19/2024*

Have 2 visualizations per question

Save png’s of visualizations *TARGET COMPLETION: 10/19/2024*

Written summaries for major findings *TARGET COMPLETION: 10/20/2024*

Assemble the presentation *TARGET COMPLETION: 10/21/2024*

Dropping TV Shows, to focus on movies.

Visuals: Box Office DC vs. Marvel

Drew – Budget (column F) vs Box office gross worldwide (column E) with Franchise (column M)

Frenci – Length column L vs rating column Q

Nate – Year (column S) vs score (column Q) with Franchise (column M) – scatter plot

Nikki – Character (malevsfemale led – column R) and rating and/or budget vs character family