Project 2 - Bechdel Movie Test

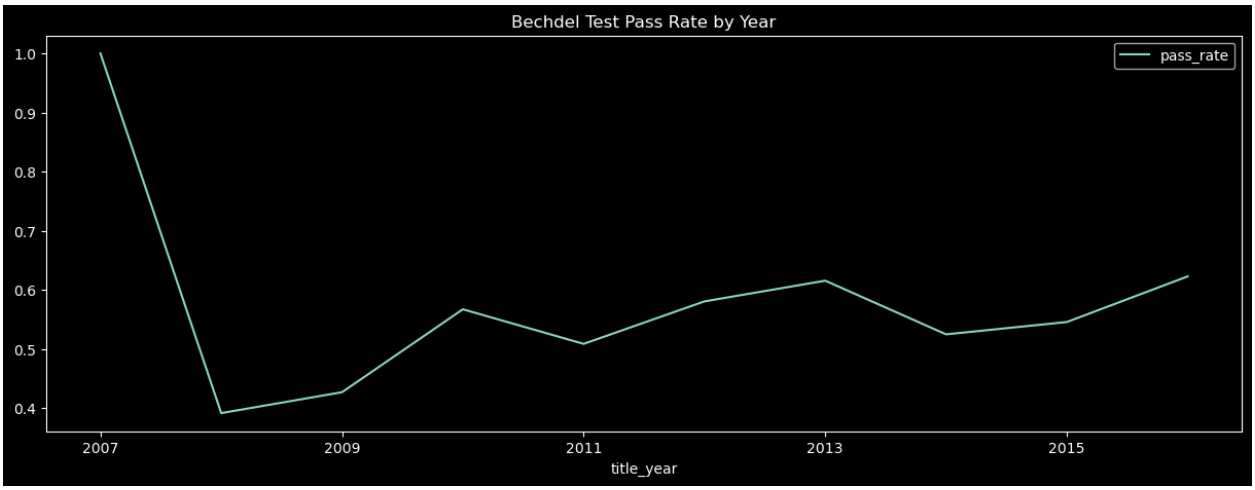
Names of the Team: Michelle Sinani, Sophie Chance, Marty Fromuth

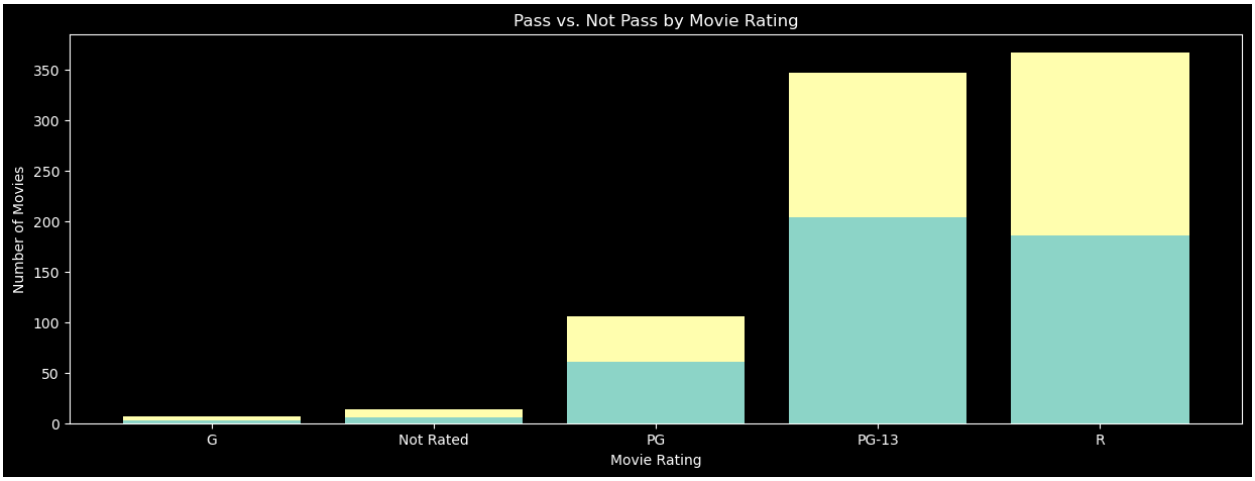
Name of the Team GitHub Repo: Project 2_Chance_Fromuth_Sinani
https://github.com/UC-Berkeley-I-School/Project2_Chance_Fromuth_Sinani

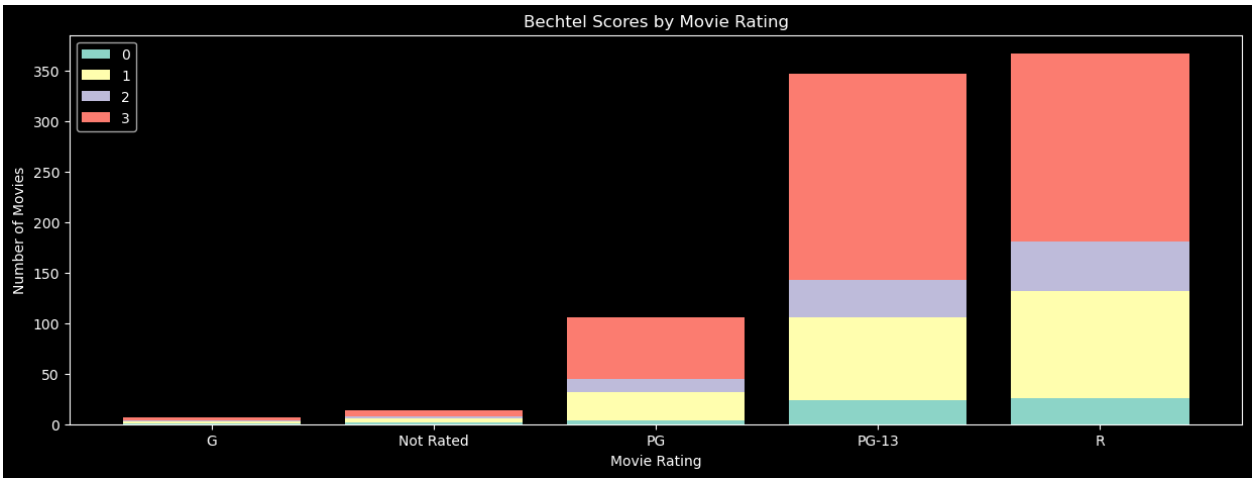
Primary Datasets:
Movie Bechdel Test Scores | Kaggle
(https://www.kaggle.com/datasets/alisonyao/movie-bechdel-test-scores)

IMDB 5000 Movie Dataset | Kaggle
(https://www.kaggle.com/datasets/carolzhangdc/imdb-5000-movie-dataset)

Initial Plots, Figures, or Tables:







Variables to Explore & Insights:

Ultimately the objective or question of this project will be to identify what attributes influence the Bechdel score. We will do so by looking at the following variables from the two datasets above as well as using those variables to conduct exploratory and descriptive analysis on trends and correlations. Through that analysis, we will attempt to answer the subsequent list of sub-questions to support our ultimate question.

Variables: Name (Dataset Title / Column-field Name / Description)

- Bechdel Score (Movie Bechdel Test Score / rating / integer - 0 to 3)
- IMDB ID (Bechdel Score (Movie Bechdel Test Score / imdbid / float - 1 to 15943414)*
- IMDB ID (IMDB 5000 Movie Dataset / movie_imdb_link / string - includes tt number that is the IMDB ID)*
- Revenue (IMDB 5000 Movie Dataset / gross / integer - null to 760505847)
- Rating (IMDB 5000 Movie Dataset / content_rating / string - R, PG-13, PG, Not Rated, G)
- Genre (IMDB 5000 Movie Dataset / genre / string - Drama, Comedy, Comedy|Drama, Comedy|Drama|Romance, Comedy|Romance)
- Budget (IMDB 5000 Movie Dataset / budget / integer - 218 to 12.2Bill)
- IMDB Score (IMDB 5000 Movie Dataset / imbd_score / float - 1.6 to 9.5)
- Directors (IMDB 5000 Movie Dataset / director_name / string)
- #1 Actor (IMDB 5000 Movie Dataset / actor_1_name / string)
- Year Released (IMDB 5000 Movie Dataset / title_year / integer - 1916 to 2016)
- Plot Keywords (IMDB 5000 Movie Dataset / plot_keywords / string)

Insights:

- General Trends:
    - Changes in Bechdel Over Time
    - Directors with Top 5 Bechtel Scores

    - Actors with Top 5 Bechtel Scores
    - Plot Topics with the Top 5 Bechtel Scores
- Correlation Between - Is there a correlation or relationship between ...
    - Revenue & Bechdel Score
    - Rating & Bechdel Score
    - Genre & Bechdel Score
    - Budget & Bechdel Score
    - IMDB Score & Bechdel Score

Potential Sub-Questions:

- What is the average IMDB score of a movie that passes the Bechdel test? And a movie that fails the test?
- What is the difference in gross revenue for a movie that passes the Bechdel test v fails the test?
- Which directors have the highest percentage of Bechdel passing movies? Which directors have the lowest percentage?
- What is the trend of movies passing the Bechdel test over the years? Do more movies pass the Bechdel test now than 10 years ago?
- What are the most common plot keywords in movies that pass the Bechdel test?

Potential Supplemental Datasets:

- 25k IMDb Movie Dataset | Kaggle (https://www.kaggle.com/datasets/utsh0dey/25k-movie-dataset)

Report Topics & Organization:

- Overview & Summary of Conclusions
  Summary of the findings of the analysis
- Description of the Dataset
  Description of the data, discussion on the strengths and weaknesses of the data, explanation of what actions were taken to merge the datasets, how did we clean or transform the data prior to the analysis
- Discussion of Findings: Trends and Correlation
  Approach for our data analysis, general analysis of trends for variables of interest, strengths and weaknesses of the trend analysis, general analysis for the correlation analysis, strengths and weaknesses of the correlation analysis
- Conclusions
  What variables did the data highlight had the most significant influence on the Bechtel score for the film; which ones had the least impact