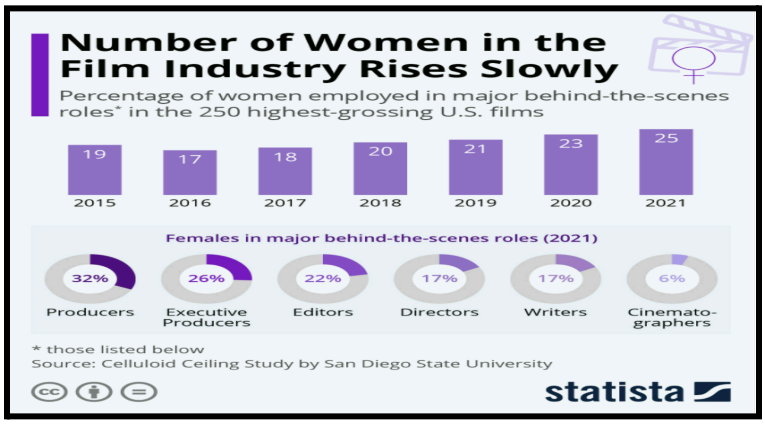
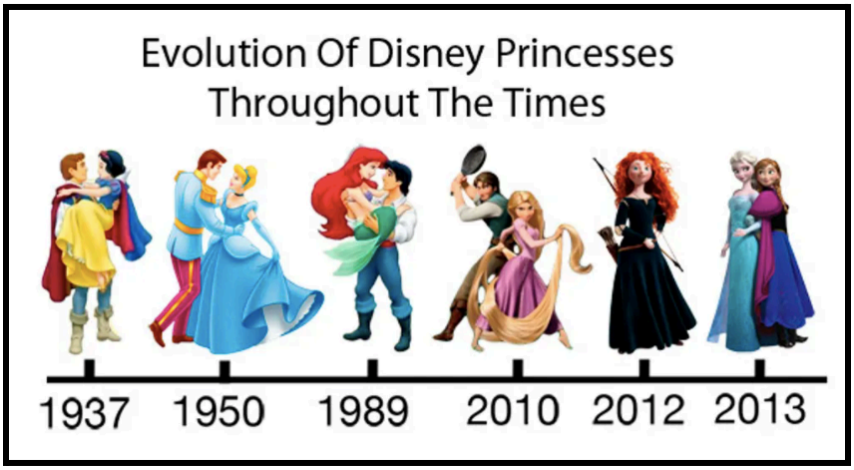
Blog Document
Taylor Toth
DS4002
12/9/2024

Investigating Evolving Gender Representations in Film



As the world of cinema has become more widespread and diverse in recent decades, there have been changes in film that correlate to gender-based changes seen culturally today. After movements like #MeToo and Time's Up, more institutional-wide discussions about gender equality and representation in film started taking place. Female-led films like 'Wonder Woman" and "Captain Marvel" which introduce female protagonists, are examples of blockbuster movies that have created a variability of more women-centered narratives. Characters have even begun to challenge more traditional gender-based stereotypes, showing more varied versions of masculinity, femininity, and sexuality in women. With this increasing number of female directors, producers, and films with female-centered plot lines, it becomes essential to investigate if there are any correlations between movie reviews across the different decades.



This case study is going to involve collecting sentiment scores from positive and negative movie reviews across several decades and investigating connections between time and these gender-based changes. To include the gender-based aspect of the study, the data set will include Bechdel-Test scores for each movie. The Bechdel Test is based on a pass-fail metric and is used to evaluate the representation of women in film and other narratives. A movie that is considered to pass the Bechdel Test has at least two names of female characters, and these women have a conversation about something other than male. This serves as a good metric to investigate the gender-based differences seen in cinema over time because we will be able to see if there has been an increase or decrease in movies that pass or fail the Bechdel test over time and how the sentiment scores from this change fluctuate. For sentiment analysis, sentiment scores will be produced using a VADER model from a collection of positive and negative reviews scraped from the IMBd website on the most recorded popular movies over time. Using this analytical model is essential to make informed future decisions to increase business revenue and to look at current societal changes that are taking place that are also impacting the film industry. An important question to be asking here is if these societal changes in gender roles and representation impacted how well movies that pass the Bechdel Test perform when it comes to the reviews. Performing this case study is aimed at giving real-world applications to data and giving practice in creating predictive models that can continue to be used in the future.

Replication Process Summary

1. **Data Collection:**
    - The initial data set contains 9,000 movies from various decades, and each has a correlating Bechdel test rating. The data first needs to be cleaned, removing any unnecessary variables and preparing it for analysis. The current data set has no sentiment score information and will be added in the following steps following scraping the IMBd website. The resulting data set should be ready for data analysis, except for the sentiment scores.
2. **Scrape IMDb movie review data:**
    - About 15-20 reviews per movie will be scraped from the IMDb website with a balance of both positive and negative reviews. These are the reviews that will be used to generate the sentiment scores for analysis.
3. **Sentiment Analysis:**

- Each review will be analyzed using the VADER sentiment analysis Python package. In total, 200,000 movie reviews will be analyzed to generate sentiment scores. The analysis will return positive, negative, neutral, and compound sentiment scores for each review. For this analysis, the compound sentiment score will be the key value added to the dataset for further evaluation.

4. **Data Compilation and Early Data Analysis**
    - The information containing the Bechdel test data and the average sentiment score from the movie reviews will now be combined into a single data set. Now perform exploratory analysis to look at possible trends in the data that may be important to investigate prior to hypothesis testing.
5. **Hypothesis Testing**
    - Hypothesis testing will be done by performing a logistic regression to allow for the evaluation of the current trends of the Bechdel Test against time and the creation of a predictive model that can help determine if a movie that passes the Bechdel Test is more or less likely to have a positive, negative, or neutral review. The goal is to have a statistically significant p-value of less than 0.05.

Here are some links that could be beneficial when completing this project:

- This article discusses current studies investigating the impact of female representation in film over time.
  https://dl.acm.org/doi/fullHtml/10.1145/3411213
- Explanations and examples for the VADER model
  https://pypi.org/project/vaderSentiment/

Good luck :)