

# Interactive Website for Harry Potter Movie Script Character Analysis

Toshiki Kato (group project with Raphael Li and Zyn Yee Ang)

## Hypothesis

The following hypothesis were evaluated. 1: There exists no correlation between the number of times a character appears and their popularity among the fanbase. 2: There exists a positive correlation between the frequency that two characters share a scene and the presence of a romantic relationship between them. 3: There exists a positive correlation between the frequency a character speaks and the number of characters they share scenes with in total. We also applied a 4: linear regression between number of lines vs. number of mentions, and 5: number of lines vs. popularity vote for characters for additional insight.

## Data

The dataset was collected through webscraping the Warner Bros Entertainment Wiki page for the 8 movies. We also webscraped a IMDb website on a fan vote for the Harry Potter movie characters, along with all couple pairings. This resulted in us having 4 databases: one to store cleaned names, one to store movie script data, one to store popularity votes, and one to store romantic couple pairings. Names have been standardized for consistency across databases and tables. Additionally, a few movies had unfinished scripts along with one completely missing which should be considered as we are evaluating a sample pool and not the population.

## Findings

Within the hypothesis questions, we found 1 and 2 to have no correlation, while 3 had a positive, statistically significant, correlation. Additionally, we found the linear regressions between number of lines vs. number of mentions to also be statistically significant, though their MSE and errors were high. Meanwhile, the linear regression between number of lines vs. number of votes varied depending on which data was for training and testing.

## Capstone Extension

The capstone component consisted of extending the findings and code to include an interactive element, whether it be through terminal commands, web apps, etc. I decided to design and implement a React web app to share our findings in a more organized manner that is easier to digest for readers. The contents are largely similar to what is displayed in a research poster, but each hypothesis and machine learning model is isolated in being displayed through an interactive drop down menu.