How Deadly is Joining the Avengers?

CMSE 381 Final Report

Kaylen Doyle

Introduction

The Marvel Cinematic Universe is a well-known and well-loved franchise of superheroes and supervillains. According to Time Magazine, *Avengers: Endgame (2019)* was the most successful Marvel film, grossing over $2.7 billion worldwide. Based on a comic book series, the Marvel Universe is complex, and has lots of characters who die and return from the dead. The goal of this project is to replicate the work done by Walt Hickey in his article *Joining the Avengers is as Deadly as Jumping Off a Four-Story Building*, and to build a model using the same data to see just how deadly joining the Avengers is, and if there are any factors that will predict if a superhero will live or die.

Related Work

I replicated the results in Hickey's article using the attached code in R.

Model

I modeled the data with a logistic regression classification to explore which variables had the most impact in predicting whether or not a character would die their first death. I chose Death1 to be the variable because usually the first death of a character is shocking, and the audience doesn't necessarily expect them to return. My code is below.

I also made a decision tree model to see if a nonparametric model would have better results.

Results

Based on the coefficients of the logistic regression, the Honorary variable is the biggest predictor of whether or not a character will die. The regression coefficient of Full Avengers membership is 2.26, whereas the regression coefficient for Honorary members is 1.77, and for Probationary members is 1.68. This makes sense logically, more of the main characters will die for shock value. When only using Honorary as the predictor, the logistic regression model had a 60% accuracy.

The decision tree did not fit the data well, it had a 37% accuracy rate.

Discussion and Conclusions

Logistic regression fit the data best, and the decision tree model did not fit the data well at all. In the future, other classification models can be used to predict if an Avenger will die the first time, or the same models can be used to predict if an Avenger will "stay dead", meaning the number of their deaths is greater than their number of returns.

References

Greenspan, R. (2019, July 21). Here are the highest-grossing marvel movies by box office. Retrieved April 18, 2021, from https://time.com/5523398/highest-grossing-marvel-movies/

Hickey, W. (2015, May 12). Joining the avengers is as deadly as jumping off a four-story building. Retrieved April 18, 2021, from https://fivethirtyeight.com/features/avengers-death-comics-age-of-ultron/