CASE STUDY 3 HOOK DOCUMENT

DS 4002 – Fall 2024, Mina Gorani

PROJECT SOURCE REPOSITORY: https://github.com/mgorani/DS4002_CS3

INTRODUCTION

Imagine this—you're a movie critic in 2024 trying to make it in the world of freelancing. Job security is tough, and if you can't convince publishers that people will want to read your reviews then they simply won't ask you to write them. You also know that people love a little dose of confirmation bias. So what do you do? Just write reviews that mirror public sentiments to increase engagement with your writing? Perhaps that's the solution! The only catch is you typically won't get a sense for public sentiments until a good bit after the movie has been released.

In this case study, you will explore if this phenomenon is actually happening based on the average sentiments of movie reviews from audiences and critics over the course of a movie's release period.

CONTEXT

The rise in online streaming and discussion platforms has allowed for a new world of film discourse to flourish. General audiences have more power than ever to dispute and voice their opinions on the words film critics so boldly and confidently publish. While critics' initial pre-release reviews can be written without influence from online audiences, once movies are released in theaters and again when they are released for streaming, audiences flock to X, Reddit, Letterboxd, and more to share their personal thoughts. As the power of confirmation bias is on the forefront of online publishers' minds, especially as their role is increasingly being questioned, critics may skew their reviews to better match the perspectives of online audiences. The goal of this case study is to determine the extent in which this post-streaming shift is occurring amongst Rotten Tomatoes critics through the use of your data science skills.

YOUR TASK

You have been provided with two datasets and a template python notebook script in the Project Source Repository (linked above). The first dataset includes thousands of movie reviews written by critics on Rotten Tomatoes. The second dataset includes information on movies, their theater release date, streaming release data, and Rotten Tomatoes audience score. With the guidance of the provided python script, you will set a quantitative benchmark to analyze this subject, clean through the data, run sentiment analysis on the movie reviews, and create visuals to communicate your findings.

More specifically, you will employ the VADER analysis package in python that was specially designed for social media analysis. The package takes written text and gives it an associated sentiment score of -1 (most negative) to 1 (most positive), with scores falling between -0.05 and 0.05 considered as indicative of neutral sentiments.

Once you have completed your analysis through the Python Script, you will compile all your materials into your own GitHub Repository and create a slidedeck that communicates your findings. Happy exploring!