CS3 Hook Document
Naylor Stone
DS 4002
28 April 2025

Using VADER and NRC Emotion Lexicon Packages For Sentiment Analysis Across Star Wars Films

Since *A New Hope* released in 1977, the series Star Wars has been one of the most recognizable and loved franchises across the globe. The series has many spin off movies, television shows, and other media, but the recognized canon of film consists of nine movies across three trilogies, known as the prequels, the originals, and the sequels. The first trilogy was released between the years of 1977 and 1983, the second was released between 1999 and 2005 (prequels), and the most recent trilogy was released in the years between 2015 and 2019 (sequels). While the series is loved by people of all ages, the movies are most likely marketed to children and young adults, and since Star Wars spans almost 50 years, each trilogy is marketed to the generation of children during each time. This means that the original trilogy was made with Generation X in mind, the prequels for Millenials, and the sequels for Generation Z. Tracking how generations change over the years is the basis of this project.

The book *Generations: The Real Differences Between Gen Z, Millennials, Gen X, Boomers, and Silents—and What They Mean for America's Future* by Jean Twenge states that, socially, each generation is affected by technology and historical moments, leading to distinct differences between each one. For example, Twenge finds members of Generation X to be more skeptical, millennials to be more purpose driven, and members of Generation Z to be more pragmatic. Furthermore, Twenge sees the rise of personal computers in Generation X, the creation of the Internet in the millennial generation, and the rise of smartphones and social media in Generation Z as having distinct effects on the social attitudes of each generation.

In this case study, your job is to utilize both VADER (Valence Aware Dictionary and sEntiment Reasoning) and NRC Emotion Lexicon packages in order to create a sentiment analysis of each Star Wars movie. These packages score words to levels of ten different emotions. Tracking these levels across each trilogy and looking for trends will be the goal of this case study. Comparing each trilogy will show character differences across three generations of people, and provide interesting insight into how screenwriters view the people they market to.

GitHub Repository: https://github.com/naylorstone/CS3-DS4002/tree/main