External Links:

Exploratory Data Analysis:
https://colab.research.google.com/drive/1ZBO8EsHZJIxj-U37F4wwN8KV8EmY02zM

Machine Learning Code:
https://colab.research.google.com/drive/1fwhOH_CY1UxiWkWwpEedmhRJlnbw8snI#scrollTo=c
KUX8x5BAavg

GitHub:
https://github.com/tmccormack165/McCormack_Final_IMDb/tree/main

Data Source: The source of this data was the Stanford AI Lab. The data as well as the documentation can be found here. The title identification values correlated perfectly with the IMDb Movies dataset, located here. The joining of these two data sources allowed access to the meta-information for features like the genre of the film, or the start year, in correspondence with the textual review. The Stanford AI Lab dataset contains 25,000 training and validation movie reviews. Each set is perfectly balanced with 12,500 positive and negative reviews. A negative review is defined as having a rating from 1-5, and a positive review has a rating from 6-10.

File Descriptions (In Codebase Folder):
Machine_Learning_III_McCormack.ipynb: The notebook with all of the pytorch models and visualizations
Machine_Learning_II_McCormack.ipynb: The notebook with all of the exploratory data analysis
Data_Munging_I.ipynb: The notebook that aggregates multiple data sources. Requires access to cs1 school server, please send me an email if you do not have an account and you can borrow my user account.