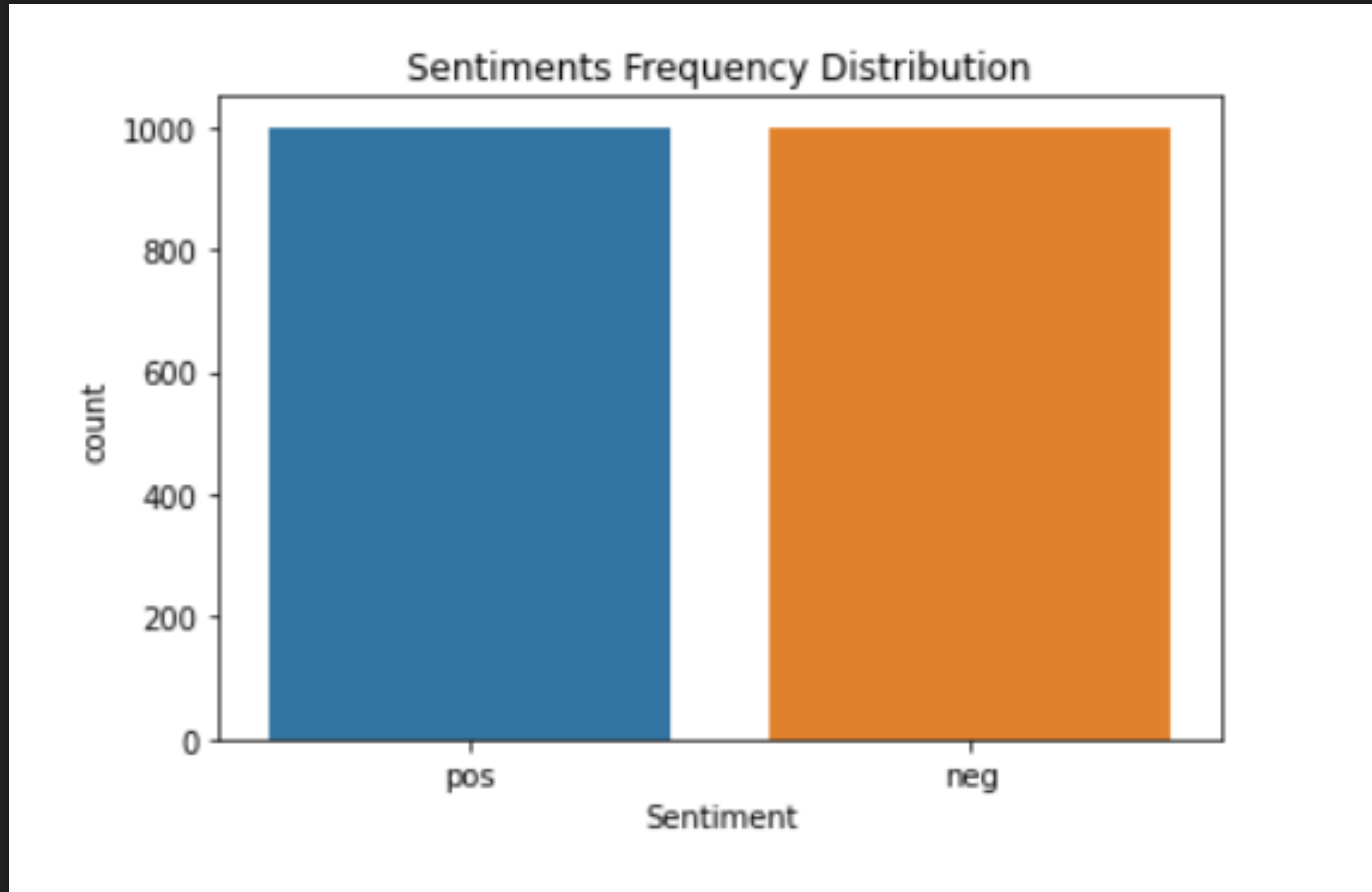# Upgrad Assignment

Movie Review Analysis

By: Sumit Pandey

Problem Statement

○ Introduction: This data set is a set of movie reviews which is similar to what data analysts at IMDB handle. Solving this assignment will give you an idea about how problems are systematically solved using NLP.

○ Business Understanding: Movie reviews are an important way to gauge the performance of a movie. While providing a numerical/stars rating to a movie tells us about the success or failure of a movie quantitatively, a collection of movie reviews is what gives us a deeper qualitative insight into different aspects of the movie.

○ **Business Objectives**: In this assignment, we aim to use Sentiment Analysis on a set of movie reviews given by reviewers and try to understand what their overall reaction to the movie was, i.e. if they liked the movie or they hated it. We aim to utilize the relationships of the words in the review to predict the overall polarity of the review.
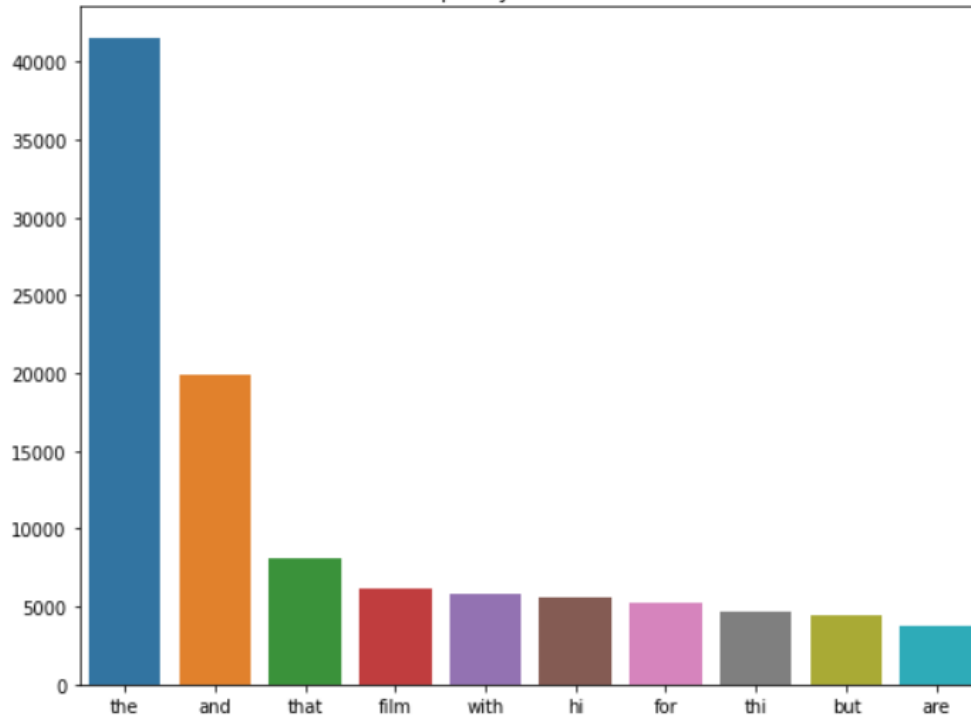
Data Exploration:

Insights: Equal Distribution of both labels.

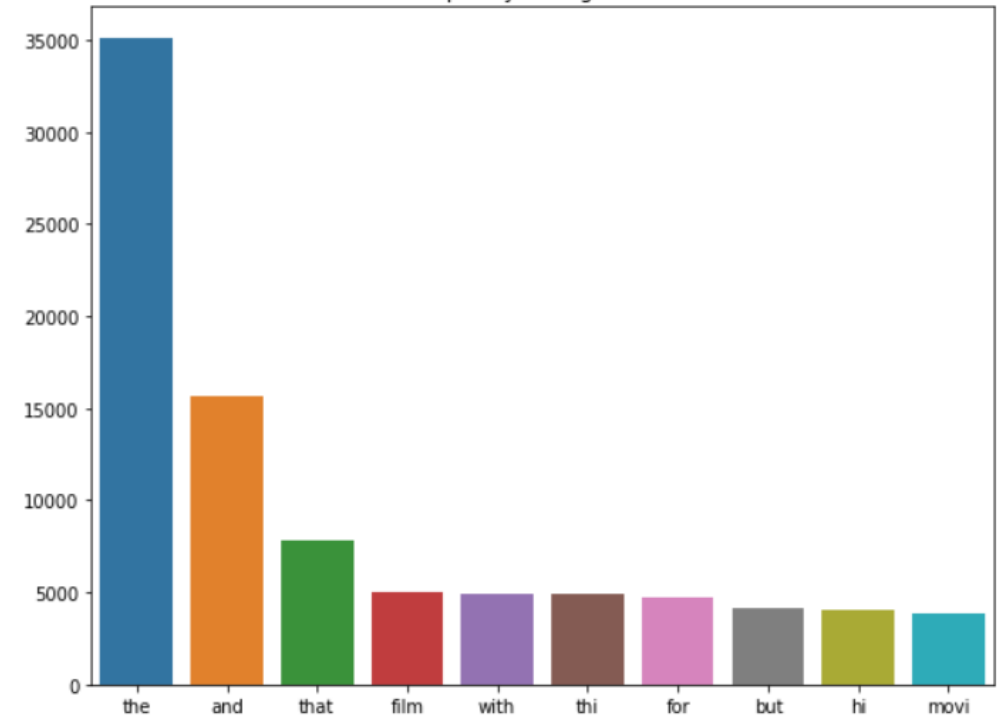# Applying Basic Lexical Processing:

- Normalizing data
- Removing stop words
- Applying stemming
- Applying tf-idf
- Vectorization

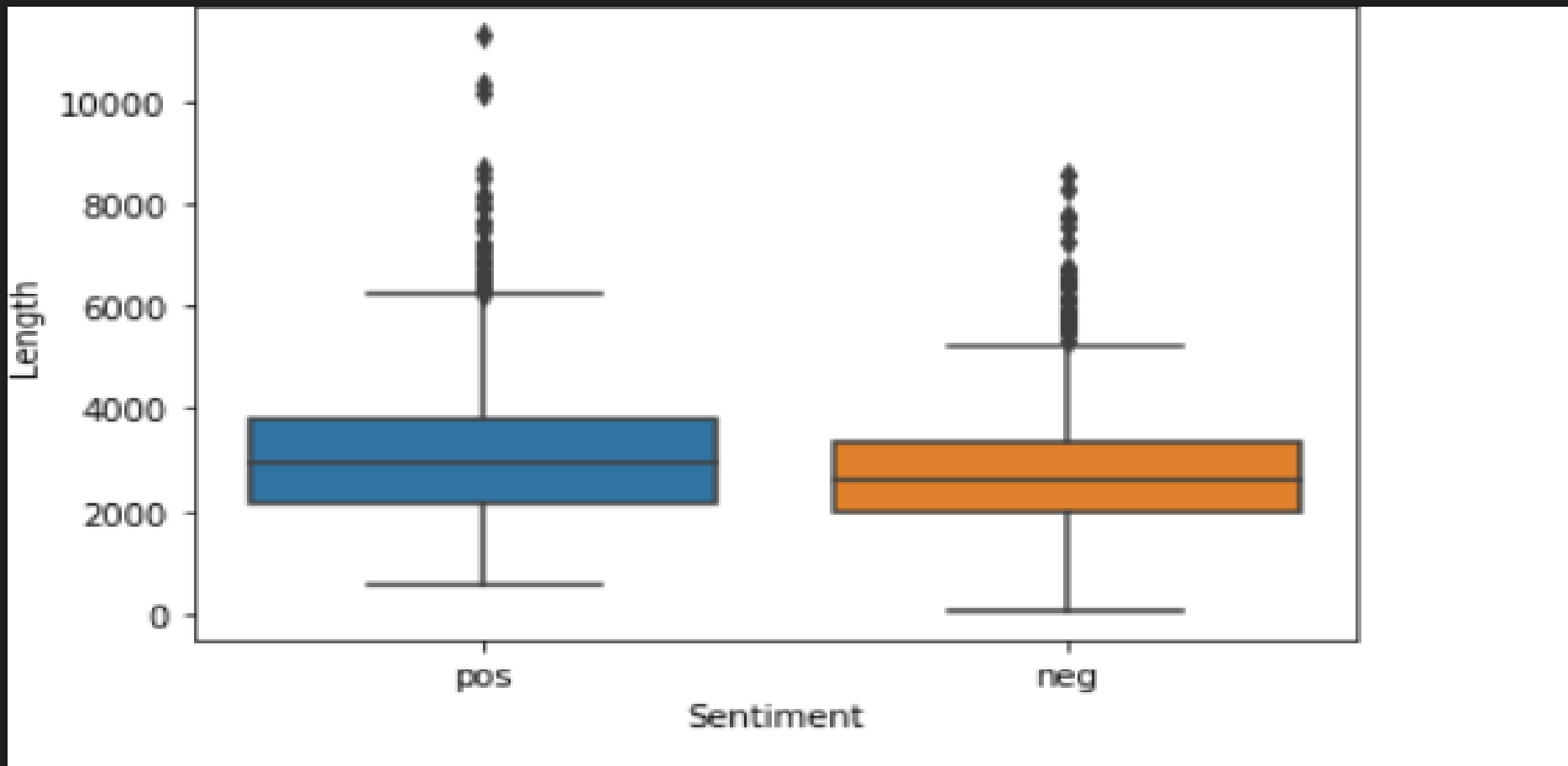# Word frequency of Positive and Negative words:

Review Length Box Plot:

Confusion Matrix: