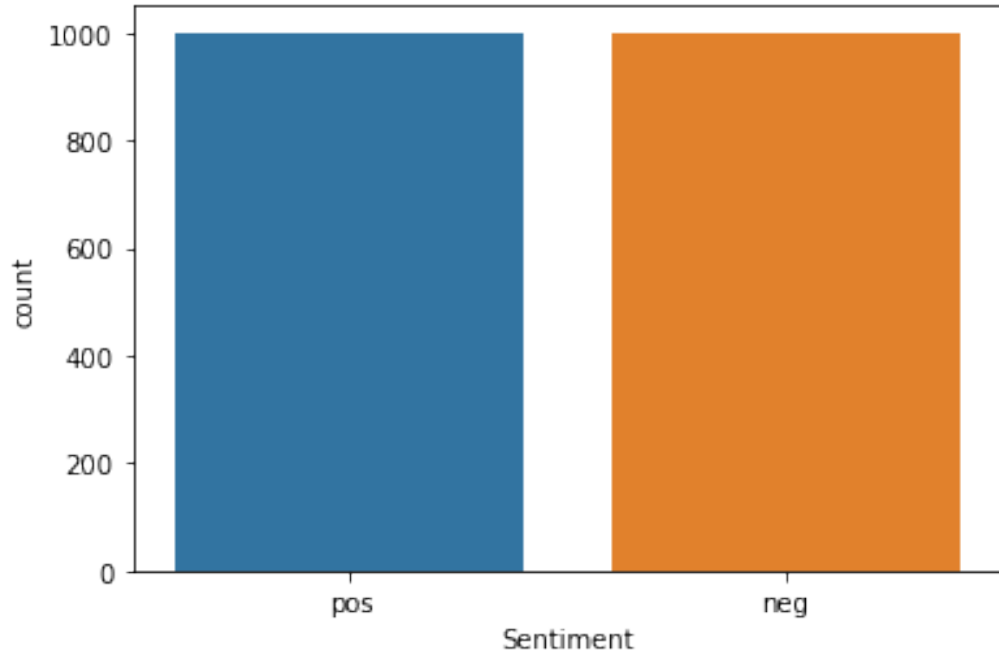


IMDB Movie Review Classification

Business Objectives

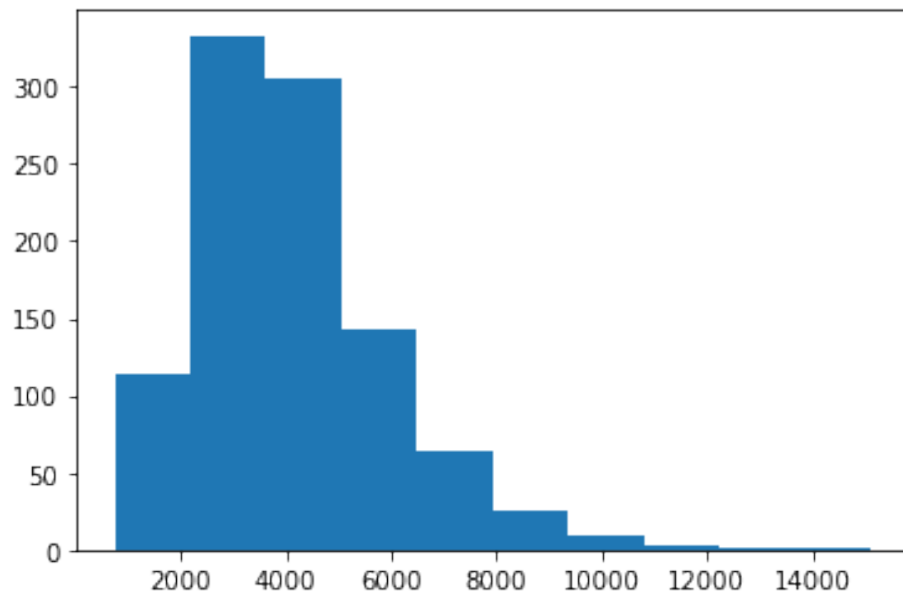
- Movie reviews are an important way to gauge the performance of a movie. While providing a numerical/stars rating to a movie tells us about the success or failure of a movie quantitatively, a collection of movie reviews is what gives us a deeper qualitative insight into different aspects of the movie.
- A textual movie review tells us about the strong and weak points of the movie and a deeper analysis of a movie review can tell us if the movie, in general, meets the expectations of the reviewer.
- 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.

About Data



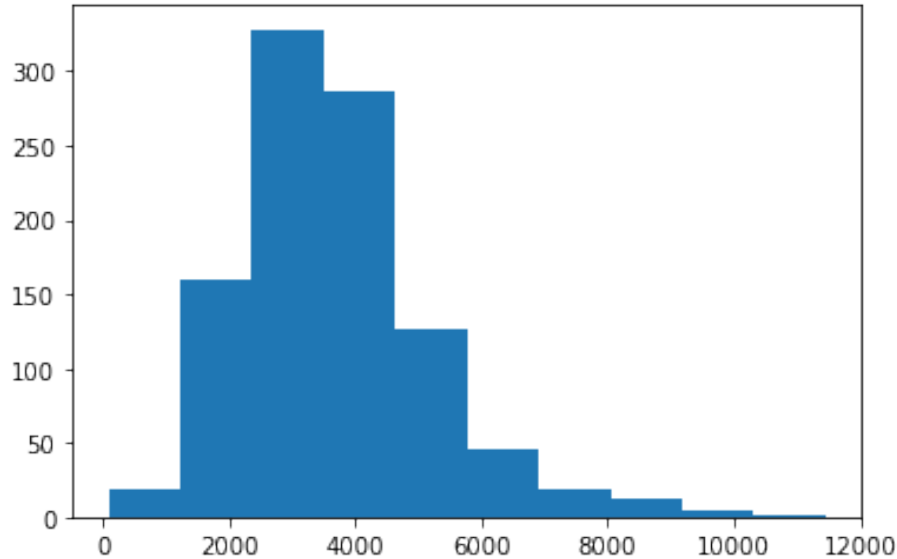
- We have equal number of classes, so don't need to worry about imbalanced data.

Exploring positive and Negative Review



- This histogram represents length of positive review.
- The average length of positive review is 4134 letters.

Exploring positive and Negative Review



- This histogram represents length of negative review.
- The average length of negative review is 3673 letters.
- by visualising both graphs it is visible that review of length shorter than length 2000 is likely to be negative.

Model Evaluation

- PRECISION SCORE : 0.8023715415019763
- RECALL SCORE : 0.796078431372549
- F1 SCORE : 0.7992125984251968
- AREA UNDER CURVE : 0.8722849139655864

ROC - Curve

