A close up of a newspaper

Description automatically generated

This World cloud of Positive reviews, it has most words that are related to character, time, makes, picture, plot, takes, performance which are positive sentiment emotions.

A close up of text on a white background

Description automatically generated  
This is a word cloud of negative reviews, where it has words like movie, time, plot, story, performance, few words same as positive reviews. Comparing both the word clouds, if you look at uncommon words such as potential, enjoyable, stunning, highly in positive reviews, you can tell the word cloud belongs to positive reviews.

A screenshot of a map

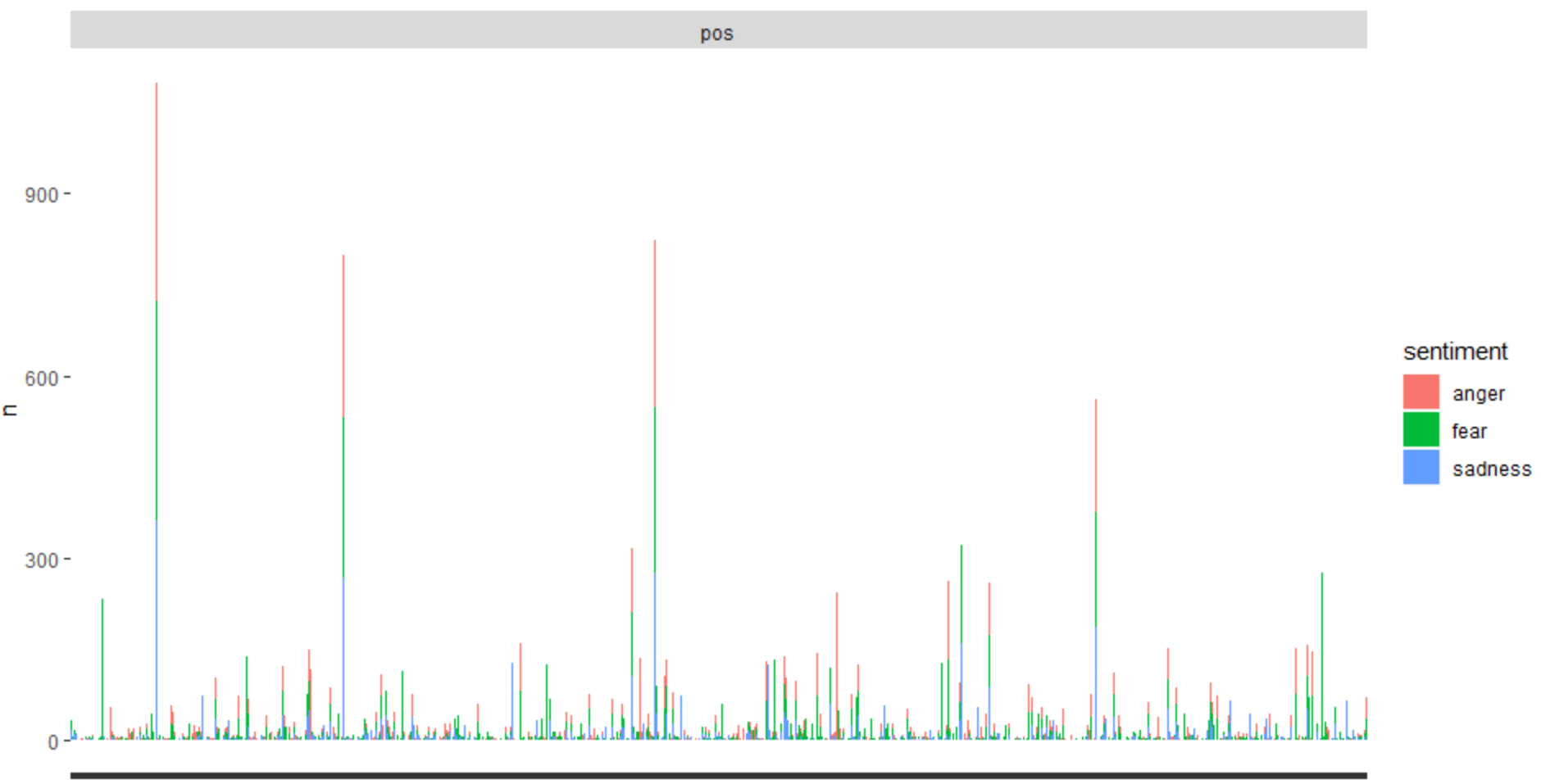
Description automatically generated

The graph above is Jitter Scatter plot where Positive and negative reviews almost similar words on those dotted lines. As we can see most of them are concentrate so closely to origin and along the line. This means that reviews are almost connected but by different sentiments such as bad, dull, fails, bye etc., are on negative side whereas outstanding, dread, legal, mature, sweet, true etc., are on positive side carrying positive emotions.

A screenshot of a cell phone

Description automatically generatedA screenshot of a cell phone

Description automatically generated  
If you have look at both positive sentiment and negative sentiment, you can see that positive reviews has diff value higher to be positive, showing the positive sentiment, where negative sentiment has lowest value since it is negative sentiment.

   
As you can see the The sentiment such fear is seen more, on top of it is Anger, bottom is sadness. If we take a closer look, we can see the same pattern repeating itself. Words that have higher n value tend to have higher anger, lower n value has mixed sentiments. A screenshot of a cell phone

Description automatically generated

If you look at Negative NRC Sentiment Lexicons, we can see that there is only one high at the beginning if you can see, which shows that these words have negative sentiment in beginning. As also you can see there is higher n value, which is actually showing the negative sentiment to be higher than positive one.

A screenshot of a cell phone

Description automatically generatedA screenshot of a social media post

Description automatically generated  
As we can’t really see close enough in this Lexical Dispersion plot but having a look at it, we can see there are dispersed in several ways, which is not uniform token index. As you can see top one is negative reviews, whereas bottom one is positive reviews. This shows that one word elaborate has different token index in each text, which means it has several meaning in each text, which is kind of what we wanted to see. The token index is closer to each other in positive reviews since they had almost same lexical sense in each text showing almost same relative token index. A screenshot of a cell phone

Description automatically generatedA screenshot of a cell phone

Description automatically generated  
If we look at the words in Topic 1, time, scene, real, plot, director, characters, bad etc., these are negative lexicons, which comes under Negative reviews. In fact, if we below at the tile plot, we can see that Topic 1 has dark red gamma indication which is a proof.   
Similarly, with Topic 2 words like bad are negative, which is making the topic 2 belonging to Negative reviews.   
Take an example of Topic 6, it does not have negative emotions or lexicons to describe, hence it belongs to Positive reviews.   
Topic 20 in above picture, has words that are not negative, therefore it is positive review.   
There are few topics like topic 3, 5, 13, 17, 19 which are not either positive or negative because it has both mixed positive and negative emotions which is making gamma percentage to be lower.

A screenshot of a cell phone

Description automatically generated

In Question 7,  
After the Training and Testing set of 75% vs 25%, We get a confusion matrix   
Which gives us outcomes possible to calculate its accuracy.   
reviews\_pred neg pos

neg 218 56

pos 32 194

The percentage we get out of this matrix would be 82.4%

In Question 8,  
If we look at the visualization of bigrams that are two words especially words like special effects, Star Wars, real life, science fiction, writer-director, romantic comedy, almost we can see more than 25 in the top 50. Which makes more sense as the reviews are about movies?