

# Writeup

December 3, 2023

```
[1]: print("""
This is the completed version of my final project. I left short comments for_
    ↳each thing I'm doing - but in short,
it's doing some pre-processing stuff as EDA and then getting into it.

Then it does:
TF-IDF into LDA
NER
WordClouds for a couple reviews for fun
Translation via TextBlob into Spanish
POS tagging
Word2Vec similarity checks
FreqDist from NLTK to find most common words
Sentiment Analysis using pre-trained BertTokenizer

The data is taken from a Rotten Tomatoes review compilation, as a text file.
""")
```

This is the completed version of my final project. I left short comments for each thing I'm doing - but in short, it's doing some pre-processing stuff as EDA and then getting into it.

Then it does:  
TF-IDF into LDA  
NER  
WordClouds for a couple reviews for fun  
Translation via TextBlob into Spanish  
POS tagging  
Word2Vec similarity checks  
FreqDist from NLTK to find most common words  
Sentiment Analysis using pre-trained BertTokenizer

The data is taken from a Rotten Tomatoes review compilation, as a text file.

```
[ ]:
```