PORJECT OVERVIEW

For my project I chose to analyze IMDB review of Inception to give insight on what the general public thinks of the film. I have managed to accomplish this by analyzing the text within the reviews and trying to interpret what the frequency of these words means. I hope this will allow me to create a list of which words appear the most frequently within the reviews and how often they occur. I also hope to gain some insight on how through the reviews are by analyzing their average size.

IMPLEMENTATION

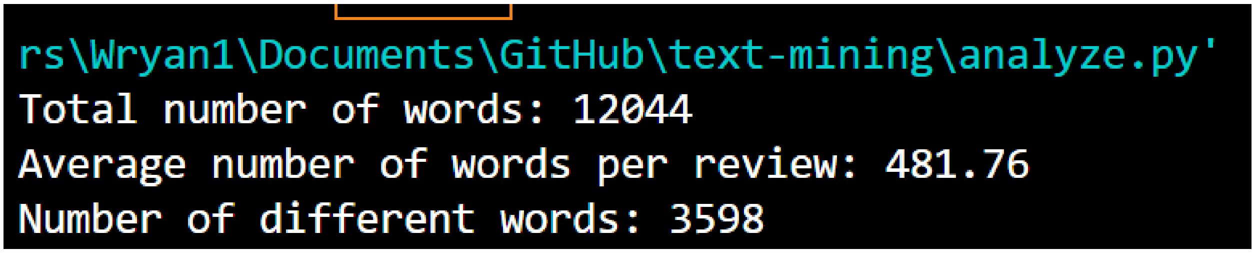
In order to analyze the review I knew I would first have to be able to pull them from the internet. Using imdbpie I was able to pull each individual review from imdb. I then combined all of the reviews into a list by appending them to each other. I then decided I would create a file out of this list so I can use it in my analysis file.

In the analysis file I broke the file contain the reviews down to lines and then down to words and added those words to a histogram that will show how many occurrences each word has. I decided to create function to tell me the total number of words so I can see the average number of words per review as well as a function to tell me the unique number of words to put my frequency of each word into perspective. I then sorted the words by those which have the highest number of occurrences or are most frequent in order to gain insights on the diction being used to describe the movie Inception.

RESULTS

I am extremely mixed about the results of the analysis. It does not appear that I am able to compile a list of adjectives used to describe the film because few of the exact same adjectives are used twice over the 25 reviews I looked at. From what I was able to see, the review themselves are very thorough. As you can see below, the average review is 481 words long which means the reviews are not just quick takeaways about the movies but actual analysis. As you can also see of the 12044 total words, 3598 of them are unique. This means there is not too much overlap in words given this contains stopwords.

Looking at the words that are the most frequent, we are left struggling to form any descriptive takeaways about the movie. The words film and movie are the 14th and 21st most popular words which allows us to know that these reviews are clearly focused on the movie and not the book or some other work of art. The next frequent word of importance is Nolan. This is referring to the director of the film which allows us to know that Nolan is being mentioned. We do not know if it is for criticism surrounding the movie or praise but we do know he is a largely discussed topic in regard to the film. The final take away is the frequency of the word dream. Dream is the 42nd most frequent word. The movie is about dreams so the mention of dreams means that the reviews have an understanding of what the film in generally about.



```
rs\Wryan1\Documents\GitHub\text-mining\analyze.py'
Total number of words: 12044
Average number of words per review: 481.76
Number of different words: 3598
```

REFLECTION

Looking back on the project there are several thing I think I could do differently. I feel as though I should remove the stop words to see what more obscure words are occurring in a high frequency. I feel as though I could also look for words of a certain length to see the more complex words being used to describe the film. I do not believe this project accomplished what it was set out to do but it accomplishes something else entirely. Originally I was aiming to see the general consensus about if the movie was good or not. Now this software seems more designed to measure if the reviews are informed about the film they're writing about. From my results I was able to know that the reviews are about a film, they mention the director by name frequently, and they discuss dreams which is one of the focal points of the movie. These signs point to the reviews being well informed about what the movie is about.