**Ai Labs DS intern evaluation**

* What are the 5 most popular genres?
* "Drama", "Comedy", "Romance Film", "Thriller", "Action"
* What words are characteristic of the movie summaries in those genres?
* Following is the word cloud generated. It contains 40 highest frequency words.
* One, back, two, film and will are characteristic of movie summaries.



* An empirical observation known as Zipf’s law is often used to describe the distribution of word  frequencies in text corpora. Do you see evidence of Zipf’s law in the summaries?

The Zipf law states that frequency of a word is inversely proportional to its rank in frequency table. We plot the following graph using Zipf\_plot on the term document matrix.

(Intercept) x

18.563848 -1.556599

We see a downward slope in the graph which shows the inverse relation. And hence the evidence of Zipf law is seen in the summaries.

