So today I am going to demonstrate the power of a simple and intuitive method called ‘Term Frequency’ in data analysis. This post is an extension of an older post where I talked about what motivated me to try out the simpler methods first. If you want to read up on that though not required, you can do so by clicking here.

Term Frequency:

It is the frequency of any term in a given document. For e.g. in the following text:

‘Hi my name is Shashank. My interests include data analytics and machine learning. ‘

The frequency of the word ‘my’ is 2. So basically the number of occurrences. Term frequency is mainly used in the field of information retrieval and it is one of the oldest methods used to automatically extract information from a set of text documents.

Dataset:

The dataset used in this post can be obtained from [here](https://www.kaggle.com/deepmatrix/imdb-5000-movie-dataset). It consists of