# Text Analysis Executive Summary

This report describes the Text Analysis done on a Book to carry out various Text analysis techniques. K-Mean clustering was used to cluster the words in the book to find out which work comes in which cluster to give the conclusion on if the we can do any further analysis on the same. K-Mean cluster worked perfectly with 3 clusters and the cluster created gave us an idea of the words most used in the book. From this method we could say that the word “Elizabeth” was used in the word most number of time. So, there might be a chance the this is the main character in the book. We could also see in the word cloud created that this word was used by the author very often.

The steps we carried out in this are as follows:

1. Removing the “-” from the text
2. Removing the repeated letters words
3. Creating a corpus to further clean the data by removing the punctuations, White space etc.
4. Then we created the Term Document Matrix
5. Defined the TDM
6. Created word cloud to find which was the most used word
7. Created a barplot to see the frequency of the words
8. Calculated the best K value for K-Means
9. Conducted K mean clustering on the TDM

After working with all the steps, the learning from this would be how to clean the data.

How to create the Term Document Matrix

How to Define TDM

How to create the attractive looking Word cloud

How to conduct K- Mean Clustering on the TDM to understand the pattern in the words.