**Assignment 3**

Data link: <https://app.box.com/s/vjbv27r081a2d8doy3o1ar27j3zfnjpc>

1. (20 points) The data file contains tweets that have been pulled from Twitter. In this dataset use the text data in the “OriginalTweet” column and perform the following:
2. Convert the text corpus into tokens.

Below is the screenshot of first 50 array elements of the tokens

A screen shot of a computer code

Description automatically generated with low confidence

b) Perform stop word removal.

Below is the screenshot of first 50 array elements of the after-filtering stop words from the tokens array.

A screen shot of a computer

Description automatically generated with low confidence

c) Count Word frequencies

A picture containing text, font, line, white

Description automatically generated

d) Create word clouds.

A picture containing text, font, poster, graphics

Description automatically generated