Project-3

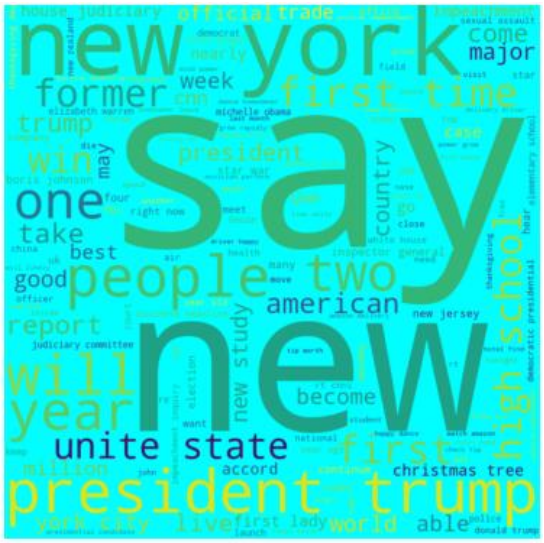

Figure 1: "Good"

Word Cloud for "Good" tweets contain words with positive sentiments. For example, words like "new", "major", "good", "first", "win" have higher frequency and hence more important.

Figure 2: "Neutral"

Word Cloud for "Neutral" tweets contain words which have neither positive nor negative sentiment. The frequency of these words is higher in this figure. For example, "house", "meet", "people", "committee".

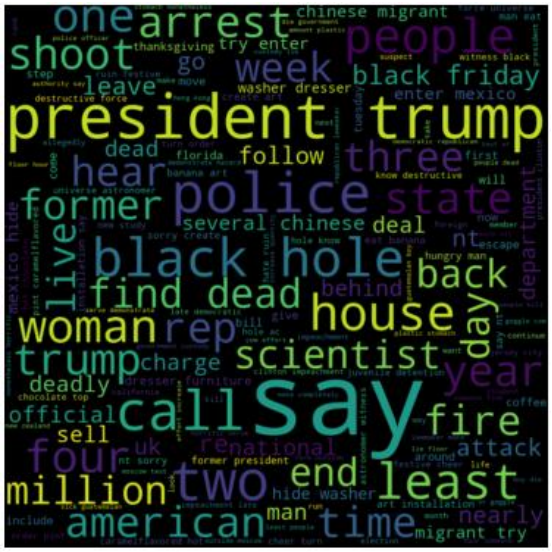

Figure 3: "Bad"

Word Cloud for "Bad" tweets contain words with negative sentiments and these words have higher frequency in this figure. For example, "dead", "arrest", "shoot", "fire".

Note: One thing is noteworthy. President Trump appears in all the word clouds!

Reasons for using GCP vs. a local environment for this project:

1) We are dealing with Twitter data which is an enormous and unstructured dataset (Big Data). With the help of DataProc (managed Spark and Hadoop) processing of such large datasets becomes faster and easier. We don't have to rely on our local system's computational power to process such huge datasets.
2) We can use the IaaS service that provides virtual machines on the fly hosted on Google's infrastructure. We don't have to worry about local environment hardware or configurations.
3) We can allocate memory and CPU resources to the clusters as per our need and can run multiple processes without weighing down the performance.