**Executive Summary**

# Data collection:

* 1. As twitter is described "the SMS of the Internet" with 100 million users posted 340 million tweets a day, and the service handled an average of 1.6 billion search queries per day. In 2013 it was one of the ten most-visited websites, so we picked twitter as the platform for capturing the emotions of people
  2. We used various Hashtags to collect data through twitter API for respective states.
  3. Tweets were collected by using state specific election hashtags during election period.
  4. Ensured all columns filling up as expected while scarping Twitter data.
  5. Tweets which are scrapped using the twitter API are stored into CSV files.
  6. Scrapped information metadata (Tweet Text, created at, Location, verified, description, followers count, Tweet Date, id, statuses count).

## Tools & Packages used for Collecting data:

Tweepy Library, Python, csv, anaconda

# Data cleaning and Preparation for analysis:

* 1. Scrapped tweets data which was stored in csv had been loaded into the python environment(Jupyter) for cleaning the tweets as required for further analysis
  2. Tweet Text is the Colum where we have tweet documents of users
  3. We used custom regex expression for cleaning data and removing unwanted pattern from tweet text.
  4. We created a logic which will assign tweets to respective political parties by inspecting from given custom dictionary of words.
  5. Using text blob polarity, we analyzed the sentiment of each tweet and assigned sentiment scores of 1, -1 and 0
  6. Where 1 is positive, -1 is negative and 0 is neutral tweet
  7. We added a flag to the data frame for further processing and analyzing in tableau to differentiate different states data.
  8. Then we removed stop words from cleaned tweets and have split into words for analyzing most frequent words and further added state tag to the data frame
  9. Stored both analyzed tweets information with columns of sentiment score and party in single csv and hashtag frequency data frame into another csv for analyzing the data in tableau

## 2.1 Tools & Packages used for preparation and cleaning:

1. NLTK, Pandas, NumPy, Textblob, Matplot, Python, csv, anaconda

# Analysis:

The processed data from the python is loaded into tableau for visualization and analysis purpose.

We have published tableau dashboard at below location

[https://public.tableau.com/profile/satwick.bodduna#!/vizhome/Elections2018\_0/GetStarted](https://public.tableau.com/profile/satwick.bodduna%23!/vizhome/Elections2018_0/GetStarted)