Project Guideline:

1. How we will complete the project – PROCESS OF PROJECT
   1. Data Preparation
      1. Create a “investors.txt" with all investors tweeter screen name
      2. Write a program to get all tweets from these all screen names. Say for example, you have 100 screen names in “investors.txt” and suppose we have almost 2000 tweets per screen name, total we will be having 200000 tweets to process.
      3. Make these all tweets in chronological (sorted based on timestamp of the tweet, which mean when this tweet was recorded) order. Just take text, screenname, description and timestamp parts of the tweet.
      4. Store them in Jason format to “data.txt”.
   2. Computation on data
      1. Read month wise data from “data.txt”, say for example, after sorting we have the tweets from March 2015 to November 2015 stored in json format in “data.txt” => so we will read data iteratively, first March-2015 then April-2015 and so on till November-2015.
      2. Each time we will create a counter to calculate the most frequent terms in this month wise tweets
      3. Based on these frequent terms we will just select top 10 or 20 terms
      4. These can be considered as hot terms in investors for that particular month
   3. Identifying Areas
      1. For now, we will sit and look at these frequent terms and assign a sector to that, which can be considered as the sector grabbing investor’s attention.
      2. We will look into the changes in these sector for each month.