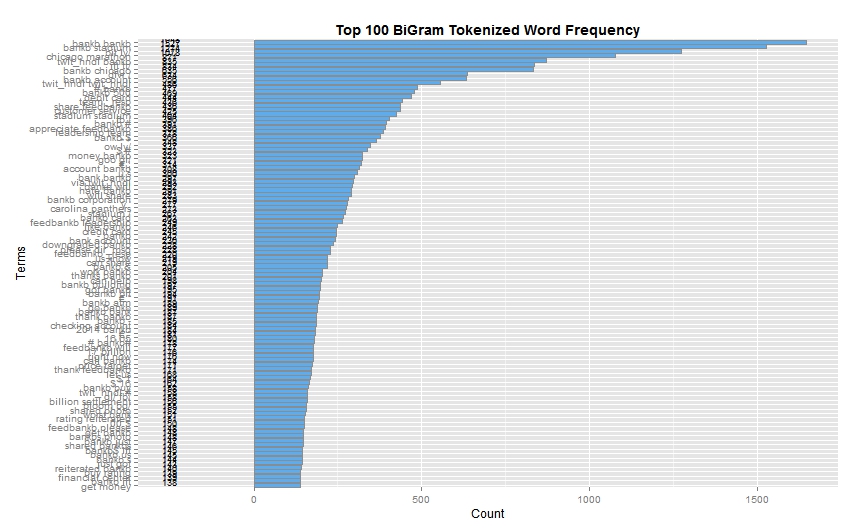
**Deliverable E**

**The data obtained is first cleaned removing junk comments and comments that cannot be useful.**

**The cleaned data is classified according to banks.**

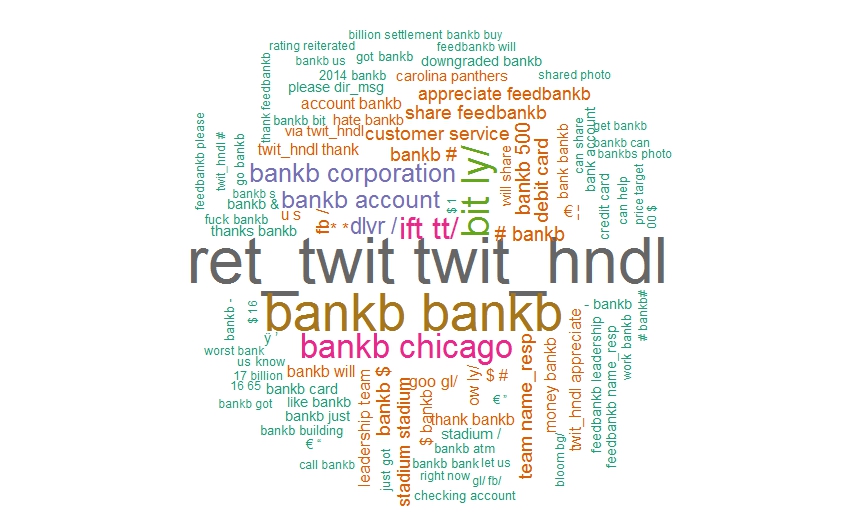
**The primary purpose is to classify the data**

**When sentiment analysis of the data is done, it reflects the direct sentiment on that bank.**

****

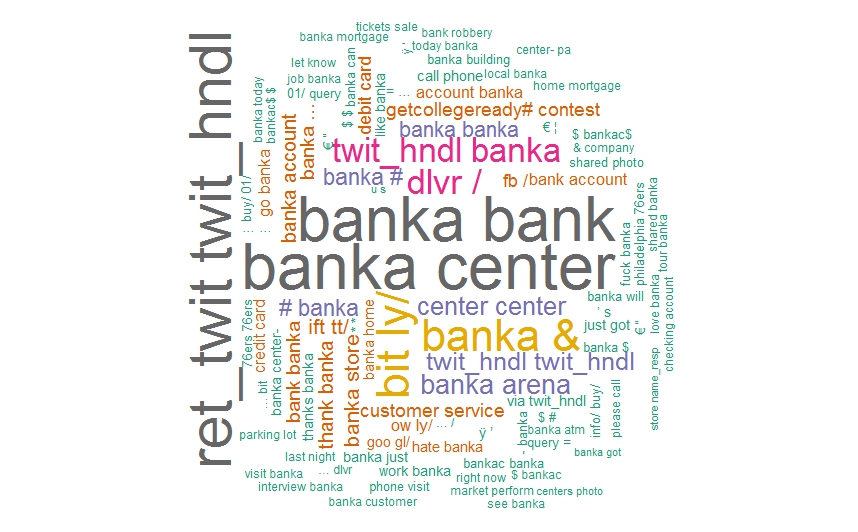
**The above gives us the insight of the top bi grams used in comments of bank B**

**Chicago marathon is one such comment which is found trendy in the comments. This could be because bankB could be the sponsor for that marathon and people are identifying that initiative. Such sponsorships can be encouraged. Such trendy lines are useful in finding overall sentiment of the bank. We can also see that negative comments such as worst bank, downgraded bankB, hate bankB can be seen here.**

****

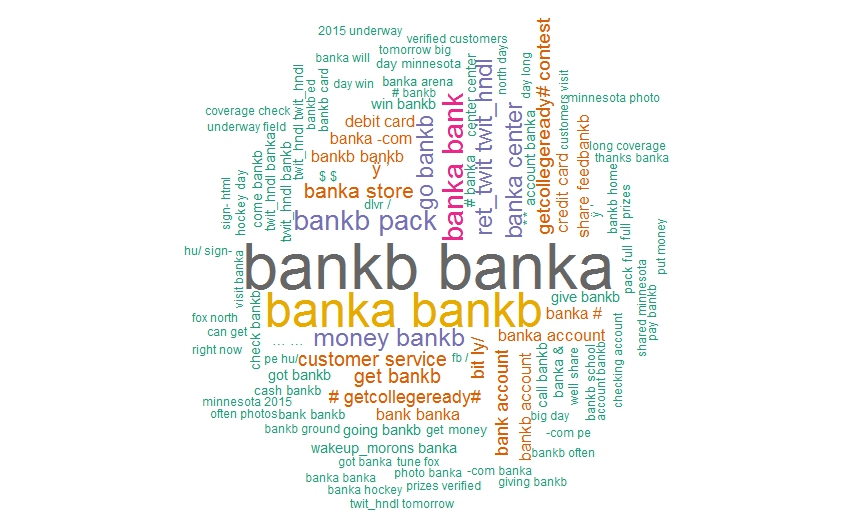
**However, further cleaning of data could remove the unwanted links which are scrubbed.**

**Similarly, For bankA**

****

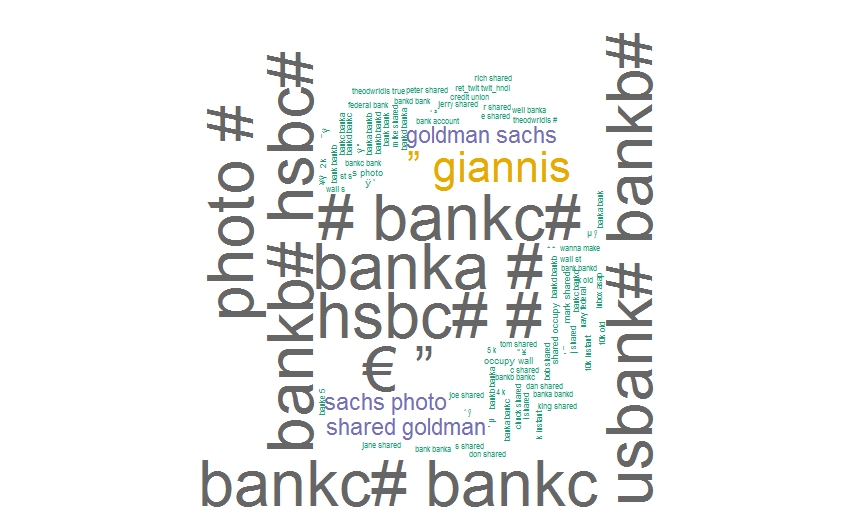
**The data containing comments of two banks would mostly compare the banks or tell us about what is common between them.**

**This is the word cloud of the data containing comments on both bank A and bank B**

****

**This gives us an insite on what made the two banks come in a single comment.**

**Wordcloud containing info from all the banks**

****

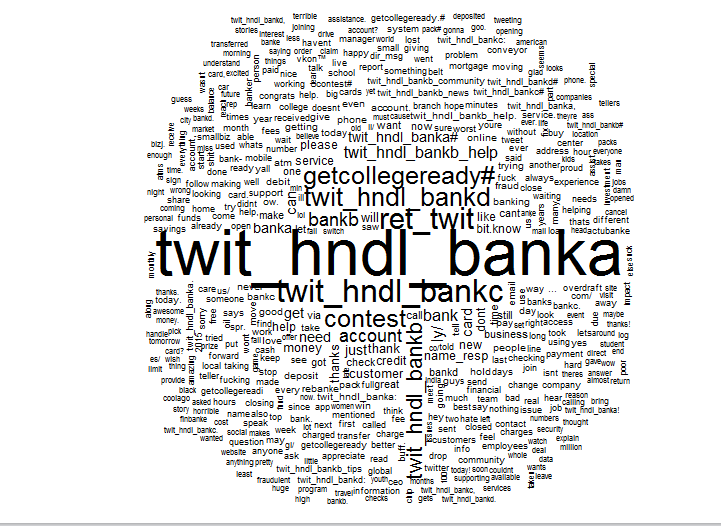
**By eliminating the bank names which is an obvious high frequent word, we can find useful incites such as occupy wall street moment, credit union etc which are common for all banks.**

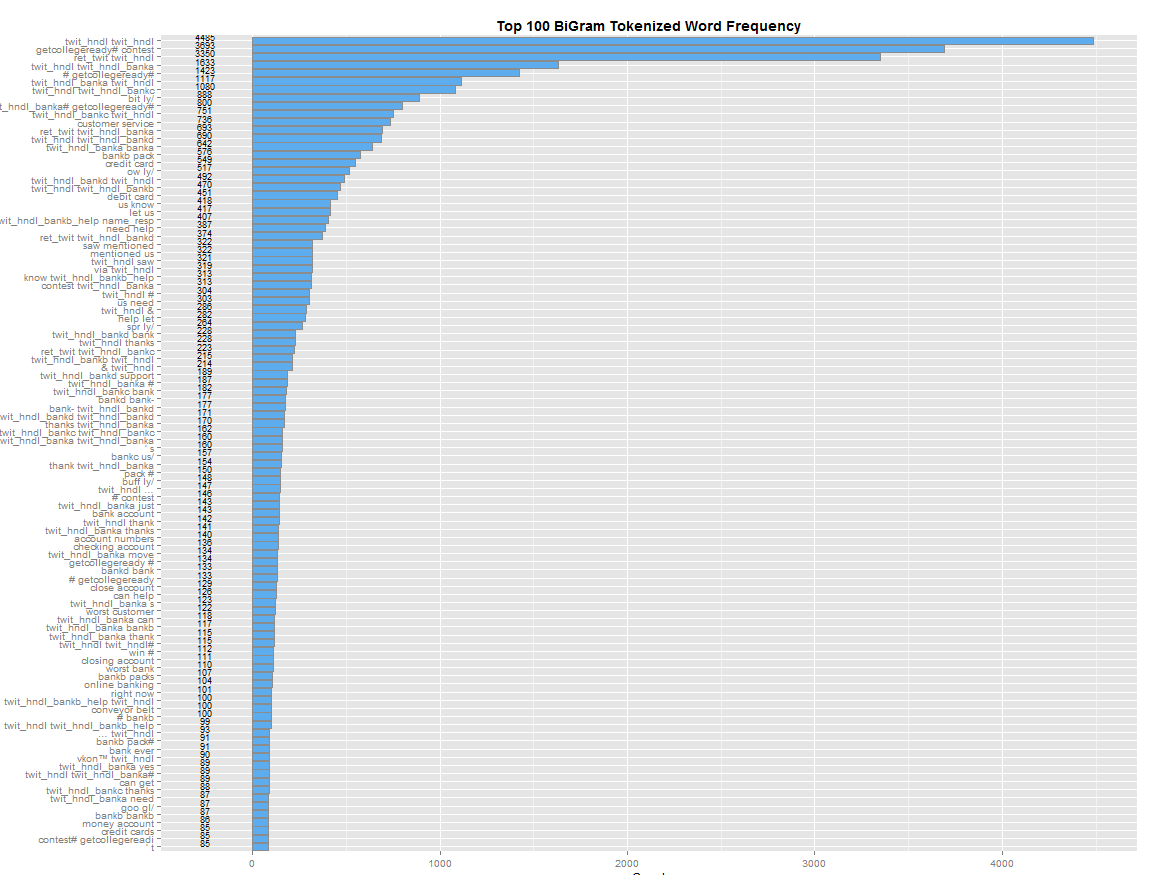
**For tweets**

**We can see which bank Is tagged the most and which hash tag is used the most.**

**Getcollegeready# contest is the most talked about. This initiative by bankA is the most hashtagged.**

**Followed by bank C**

****

****