Tweet Sentimental Analysis

Importing Required Libraries

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.0.4
```

```
## -- Attaching packages ------------------------------------- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.0.6     v dplyr   1.0.5
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1
```

```
## Warning: package 'ggplot2' was built under R version 4.0.4
```

```
## Warning: package 'tibble' was built under R version 4.0.3
```

```
## Warning: package 'tidyr' was built under R version 4.0.4
```

```
## Warning: package 'readr' was built under R version 4.0.4
```

```
## Warning: package 'purrr' was built under R version 4.0.4
```

```
## Warning: package 'dplyr' was built under R version 4.0.4
```

```
## Warning: package 'stringr' was built under R version 4.0.3
```

```
## Warning: package 'forcats' was built under R version 4.0.4
```

```
## -- Conflicts ----------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(tidytext)
```

```
## Warning: package 'tidytext' was built under R version 4.0.5
```

```
library(glue)
```

```
## Warning: package 'glue' was built under R version 4.0.3
```

```
##
## Attaching package: 'glue'
```

```
## The following object is masked from 'package:dplyr':
##
##      collapse
```

```
library(stringr)
library(SnowballC)
```

```
## Warning: package 'SnowballC' was built under R version 4.0.3
```

```
library(tm)
```

```
## Warning: package 'tm' was built under R version 4.0.5
```

```
## Loading required package: NLP
```

```
## Warning: package 'NLP' was built under R version 4.0.3
```

```
##
## Attaching package: 'NLP'
```

```
## The following object is masked from 'package:ggplot2':
##
##      annotate
```

```
library(twitteR)
```

```
## Warning: package 'twitteR' was built under R version 4.0.5
```

```
##
## Attaching package: 'twitteR'
```

```
## The following objects are masked from 'package:dplyr':
##
##      id, location
```

```
library(syuzhet)
```

```
## Warning: package 'syuzhet' was built under R version 4.0.5
```

Reading Dataset and assigning column names

```
df = read.csv('D:/My Files/Sem-2/R Programming-Akhil Sir/tweets.csv', nrows = 50000, header =
TRUE, sep = ',')
attach(df)
df = subset(df, select = c(-X0))
colnames(df) = c('id', 'date', 'flag', 'user','tweet')
attach(df)
head(df)
```

```
##              id                        date        flag            user
## 1 1467810672 Mon Apr 06 22:19:49 PDT 2009 NO_QUERY scotthamilton
## 2 1467810917 Mon Apr 06 22:19:53 PDT 2009 NO_QUERY       mattycus
## 3 1467811184 Mon Apr 06 22:19:57 PDT 2009 NO_QUERY        ElleCTF
## 4 1467811193 Mon Apr 06 22:19:57 PDT 2009 NO_QUERY         Karoli
## 5 1467811372 Mon Apr 06 22:20:00 PDT 2009 NO_QUERY       joy_wolf
## 6 1467811592 Mon Apr 06 22:20:03 PDT 2009 NO_QUERY        mybirch
##
tweet
## 1 is upset that he can't update his Facebook by texting it... and might cry as a result  S
chool today also. Blah!
## 2                      @Kenichan I dived many times for the ball. Managed to save 50%  Th
e rest go out of bounds
## 3                                                        my whole body feels itch
y and like its on fire
## 4 @nationwideclass no, it's not behaving at all. i'm mad. why am i here? because I can't s
ee you all over there.
## 5                                                                               @Kwesi
dei not the whole crew
## 6
Need a hug
```

First 10 tweets from the Dataset.

```
for(i in seq(1:10)){
print(df$tweet[i])
}
```

```
## [1] "is upset that he can't update his Facebook by texting it... and might cry as a result
School today also. Blah!"
## [1] "@Kenichan I dived many times for the ball. Managed to save 50%  The rest go out of bo
unds"
## [1] "my whole body feels itchy and like its on fire "
## [1] "@nationwideclass no, it's not behaving at all. i'm mad. why am i here? because I ca
n't see you all over there. "
## [1] "@Kwesidei not the whole crew "
## [1] "Need a hug "
## [1] "@LOLTrish hey  long time no see! Yes.. Rains a bit ,only a bit  LOL , I'm fine thanks
, how's you ?"
## [1] "@Tatiana_K nope they didn't have it "
## [1] "@twittera que me muera ? "
## [1] "spring break in plain city... it's snowing "
```

creating a function to clean tweets

```r
library(tidyverse)
clean_tweets <- function(x) {
            x %>%
                    str_remove_all(" ?(f|ht)(tp)(s?)(://)(.*)[.|/](.*)") %>%
                    str_replace_all("&amp;", "and") %>%
                    str_remove_all("[[:punct:]]") %>%
                    str_remove_all("^RT:? ") %>%
                    str_remove_all("@[[:alnum:]]+") %>%
                    str_remove_all("#[[:alnum:]]+") %>%
                    str_replace_all("\\\n", " ") %>%
                    str_to_lower() %>%
                    str_trim("both")
    return (x)
}
```

applying the cleaning function to tweet column

```r
df$tweet = sapply(df$tweet, clean_tweets)
```

Getting the sentiments of each tweets

```r
get_sents = function(tweet){
sent.value = get_sentiment(tweet)
return (sent.value)
}
```

applying the get_sents function to tweet column

```r
df$sentiment_score = sapply(df$tweet, get_sents)
head(df$sentiment_score)
```

```
## [1] -2.00  0.50 -0.25 -0.75  0.00  0.75
```

Final Dataset

```r
head(df)
```

```
##          id                          date      flag          user
## 1 1467810672 Mon Apr 06 22:19:49 PDT 2009 NO_QUERY scotthamilton
## 2 1467810917 Mon Apr 06 22:19:53 PDT 2009 NO_QUERY      mattycus
## 3 1467811184 Mon Apr 06 22:19:57 PDT 2009 NO_QUERY        ElleCTF
## 4 1467811193 Mon Apr 06 22:19:57 PDT 2009 NO_QUERY         Karoli
## 5 1467811372 Mon Apr 06 22:20:00 PDT 2009 NO_QUERY       joy_wolf
## 6 1467811592 Mon Apr 06 22:20:03 PDT 2009 NO_QUERY        mybirch
##
tweet
## 1 is upset that he can't update his Facebook by texting it... and might cry as a result  S
chool today also. Blah!
## 2                        @Kenichan I dived many times for the ball. Managed to save 50%  Th
e rest go out of bounds
## 3                                                             my whole body feels itch
y and like its on fire
## 4 @nationwideclass no, it's not behaving at all. i'm mad. why am i here? because I can't s
ee you all over there.
## 5                                                                                  @Kwesi
dei not the whole crew
## 6
Need a hug
##    sentiment_score
## 1           -2.00
## 2            0.50
## 3           -0.25
## 4           -0.75
## 5            0.00
## 6            0.75
```

Creating two different dataframe

```
positive.tweets = df[df$sentiment_score > 0,]
negative.tweets = df[df$sentiment_score < 0,]
```

Assigning labels to tweets

```
to_sent_labels = function(x){
  if (x >0){
    'positive'
  }
  else if (x <0){
    'negative'
  }else{
    'neutral'
  }
}
```

```
df$sent_labels = sapply(df$sentiment_score , to_sent_labels)
head(df[c('id','user','sentiment_score','sent_labels')])
```

```
##           id            user sentiment_score sent_labels
## 1 1467810672 scotthamilton           -2.00    negative
## 2 1467810917       mattycus            0.50    positive
## 3 1467811184         ElleCTF           -0.25    negative
## 4 1467811193          Karoli           -0.75    negative
## 5 1467811372        joy_wolf            0.00     neutral
## 6 1467811592         mybirch            0.75    positive
```