

Sentiment Analysis

Soumya Paul

29/07/2021

Tokyo Olympics 2021(2020) tweets

Monitoring tweets on Olympics -Updated at least once daily.

- Search queries are - 'Olympics'
- collected using tweepy Python and Twitter API

Last updated - 29th July,2021

Number of tweets - 140,000+

```
library(readxl)
Olympics_Tokyo_tweets <- read_excel("/Users/soumyapaul/Dropbox/R works/Projects/Olympics_Tokyo_tweets.x

## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i = sheet, :
## Expecting numeric in A58586 / R58586C1: got '#After | #Covid | #Coronavirus |
## #Tokyo | #China | #Biden | #Olympics | #Up | #Police | #Olympic,Ã¶ https://t.co/
## EcmDip6rOk'

## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i = sheet, :
## Coercing boolean to numeric in E58586 / R58586C5

## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i = sheet, :
## Expecting numeric in G58586 / R58586C7: got 'Addressed to the Inhabitants of the
## World - https://t.co/RNwVwFZSKy'

## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i = sheet, :
## Coercing numeric to date I58586 / R58586C9

## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i = sheet, :
## Expecting numeric in K58586 / R58586C11: got a date

## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i = sheet, :
## Expecting date in L58586 / R58586C12: got 'en'

## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i = sheet, :
## Expecting numeric in A131160 / R131160C1: got '#Covid | #Capitol | #Police |
## #After | #Tokyo | #Olympics | #Cases | #Vaccinated | #People |,Ã¶ https://t.co/
## TOAPmcnGYC'

## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i = sheet, :
## Coercing boolean to numeric in E131160 / R131160C5

## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i = sheet, :
## Expecting numeric in G131160 / R131160C7: got 'Addressed to the Inhabitants of
## the World - https://t.co/RNwVwFZSKy'
```

```

## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i = sheet, :
## Coercing numeric to date I131160 / R131160C9

## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i = sheet, :
## Expecting numeric in K131160 / R131160C11: got a date

## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i = sheet, :
## Expecting date in L131160 / R131160C12: got 'en'

library(twitterR)
library(purrr)
library(dplyr)

##
## Attaching package: 'dplyr'

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

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

require("ROAuth")

## Loading required package: ROAuth

require("RCurl")

## Loading required package: RCurl

library(plyr)

## -----

## You have loaded plyr after dplyr - this is likely to cause problems.
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:
## library(plyr); library(dplyr)

## -----

##
## Attaching package: 'plyr'

## The following objects are masked from 'package:dplyr':
##
##   arrange, count, desc, failwith, id, mutate, rename, summarise,
##   summarize

## The following object is masked from 'package:purrr':
##
##   compact

## The following object is masked from 'package:twitterR':
##
##   id

```

```

library(stringr)
library(ggplot2)
#Sentiment function
score.sentiment=function(sentences,pos.words,neg.words,.progress="none")
{
  require(plyr)
  require(stringr)
  scores=lapply(sentences, function(sentence,pos.words,neg.words){
    sentence=gsub('[[[:punct:]]',' ',sentence)
    sentence=gsub('[[[:cntrl:]]',' ',sentence)
    sentence=gsub('\\d+',' ',sentence)
    sentence=tolower(sentence)
    word.list=strsplit(sentence,'\\s+')
    words=unlist(word.list)
    pos.matches=match(words,pos.words)
    neg.matches=match(words,neg.words)
    pos.matches=!is.na(pos.matches)
    neg.matches=!is.na(neg.matches)
    score=sum(pos.matches)-sum(neg.matches)
    return(score)
  },pos.words,neg.words,.progress= .progress)
  scores.df=data_frame(score=scores,text=sentences)
  return(scores.df)
}

pos.words=scan('/Users/soumyapaul/Dropbox/R works/Projects/positive-words.txt',what='character',comment='')
neg.words=scan('/Users/soumyapaul/Dropbox/R works/Projects/negative-words.txt',what='character',comment='')
Olympicscore=score.sentiment(Olympics_Tokyo_tweets$text,pos.words,neg.words,.progress = 'text')

## |

```

```

ggplot(NULL,aes(x=Olympicscore$score))+geom_histogram(binwidth=1,col="darkblue",fill="lightblue")+scale_x_continuous(breaks=c(-10,0,10))

```

