DS\_04.R

anany

2024-06-16

library(rtweet)

## Warning: package 'rtweet' was built under R version 4.3.3

# Authenticate with Twitter API  
create\_token(  
 app = "your\_app\_name",  
 consumer\_key = "your\_consumer\_key",  
 consumer\_secret = "your\_consumer\_secret",  
 access\_token = "your\_access\_token",  
 access\_secret = "your\_access\_secret"  
)

## Warning: `create\_token()` was deprecated in rtweet 1.0.0.  
## ℹ See vignette('auth', package = 'rtweet') for details  
## This warning is displayed once every 8 hours.  
## Call `lifecycle::last\_lifecycle\_warnings()` to see where this warning was  
## generated.

## Saving auth to  
## 'C:\Users\anany\AppData\Roaming/R/config/R/rtweet/create\_token.rds'

# Search for tweets based on a keyword  
tweets <- search\_tweets(q = "your\_keyword", n = 1000)

## Warning: `render()` was deprecated in rmarkdown 2.0.0.  
## This function might work until your bot/app/user is blocked or fail randomly!  
## ! The API has been deprecated and the new API v2.2 requires subscriptions for  
## most endpoints.  
## ℹ See updates of function and API: help('rtweet-deprecated', 'rtweet' )  
## ℹ The deprecated feature was likely used in the rtweet package.  
## Please report the issue at <https://github.com/ropensci/rtweet/issues>.  
## This warning is displayed once every 8 hours.  
## Call `lifecycle::last\_lifecycle\_warnings()` to see where this warning was  
## generated.

#loading the necessary libraries  
library(tidyverse)

## Warning: package 'tidyverse' was built under R version 4.3.3

## Warning: package 'ggplot2' was built under R version 4.3.3

## Warning: package 'tidyr' was built under R version 4.3.3

## Warning: package 'readr' was built under R version 4.3.3

## Warning: package 'purrr' was built under R version 4.3.3

## Warning: package 'dplyr' was built under R version 4.3.3

## Warning: package 'forcats' was built under R version 4.3.3

## Warning: package 'lubridate' was built under R version 4.3.3

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.4 ✔ readr 2.1.5  
## ✔ forcats 1.0.0 ✔ stringr 1.5.1  
## ✔ ggplot2 3.5.0 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.3 ✔ tidyr 1.3.1  
## ✔ purrr 1.0.2   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ purrr::flatten() masks rtweet::flatten()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(tm)

## Warning: package 'tm' was built under R version 4.3.3

## Loading required package: NLP  
##   
## Attaching package: 'NLP'  
##   
## The following object is masked from 'package:ggplot2':  
##   
## annotate

library(stringr)  
library(tidytext)

## Warning: package 'tidytext' was built under R version 4.3.3

library(syuzhet)

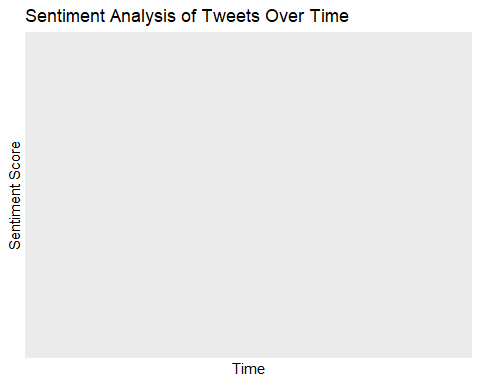
## Warning: package 'syuzhet' was built under R version 4.3.3

##   
## Attaching package: 'syuzhet'  
##   
## The following object is masked from 'package:rtweet':  
##   
## get\_tokens

library(ggplot2)  
  
#searching for tweets  
tweets<-search\_tweets(q="artificial intelligence",n=1000)  
  
#cleaning the data   
clean\_tweets <- tweets %>%  
 select(text) %>%  
 mutate(text = str\_replace\_all(text, "[^[:graph:]]", " ")) %>% # Remove emojis and non-graphic characters  
 mutate(text = str\_replace\_all(text, "https?://\\S+\\s?", "")) %>% # Remove URLs  
 mutate(text = str\_replace\_all(text, "#\\S+", "")) %>% # Remove hashtags  
 mutate(text = str\_replace\_all(text, "@\\S+", "")) %>% # Remove mentions  
 mutate(text = str\_replace\_all(text, "\\bRT\\b", "")) %>% # Remove RT (retweet)  
 mutate(text = str\_replace\_all(text, "\\s+", " ")) %>% # Remove extra white spaces  
 mutate(text = str\_trim(text)) # Trim leading and trailing white spaces  
  
#check the structure of clean\_tweets  
str(clean\_tweets)

## tweets [0 × 1] (S3: tweets/tbl\_df/tbl/data.frame)  
## $ text: chr(0)   
## - attr(\*, "users")= tibble [1 × 23] (S3: tbl\_df/tbl/data.frame)  
## ..$ id : int NA  
## ..$ id\_str : chr NA  
## ..$ name : chr NA  
## ..$ screen\_name : chr NA  
## ..$ location : chr NA  
## ..$ derived :List of 1  
## .. ..$ : NULL  
## .. ..- attr(\*, "class")= chr "AsIs"  
## ..$ url : chr NA  
## ..$ description : chr NA  
## ..$ protected : logi NA  
## ..$ verified : logi NA  
## ..$ followers\_count : int NA  
## ..$ friends\_count : int NA  
## ..$ listed\_count : chr NA  
## ..$ favourites\_count : int NA  
## ..$ statuses\_count : int NA  
## ..$ created\_at : chr NA  
## ..$ profile\_banner\_url : chr NA  
## ..$ profile\_image\_url\_https: chr NA  
## ..$ default\_profile : logi NA  
## ..$ default\_profile\_image : logi NA  
## ..$ withheld\_in\_countries :List of 1  
## .. ..$ : NULL  
## .. ..- attr(\*, "class")= chr "AsIs"  
## ..$ entities :List of 1  
## .. ..$ : NULL  
## .. ..- attr(\*, "class")= chr "AsIs"  
## ..$ withheld\_scope : logi NA

#performing sentiment analysis  
sentiment\_scores <- get\_sentiment(clean\_tweets$text, method = "syuzhet")  
clean\_tweets$sentiment\_score <- sentiment\_scores  
  
#visualization of sentiment patterns  
clean\_tweets$time <- as.Date(tweets$created\_at)  
  
ggplot(clean\_tweets, aes(x = time, y = sentiment\_scores)) +  
 geom\_point() +  
 geom\_smooth(method = "loess", se = FALSE) +  
 labs(title = "Sentiment Analysis of Tweets Over Time",  
 x = "Time",  
 y = "Sentiment Score")



#preparing histogram of sentiment scores  
ggplot(clean\_tweets, aes(x = sentiment\_scores)) +  
 geom\_histogram(binwidth = 1, fill = "blue", color = "black") +  
 labs(title = "Distribution of Sentiment Scores",  
 x = "Sentiment Score",  
 y = "Frequency")

