Sentiment Analysis

Barrientos, Delfin, Infiesto

2024-11-28

```
library(readxl)
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.4.2
```

```
##
## Attaching package: 'dplyr'
```

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

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

```
library(ggplot2)
library(tidyr)
library(stringr)
library(syuzhet)
```

```
## Warning: package 'syuzhet' was built under R version 4.4.2
```

```
tweetsDF <- read_excel("C:/PROJ/tweetsDF.xlsx")
```

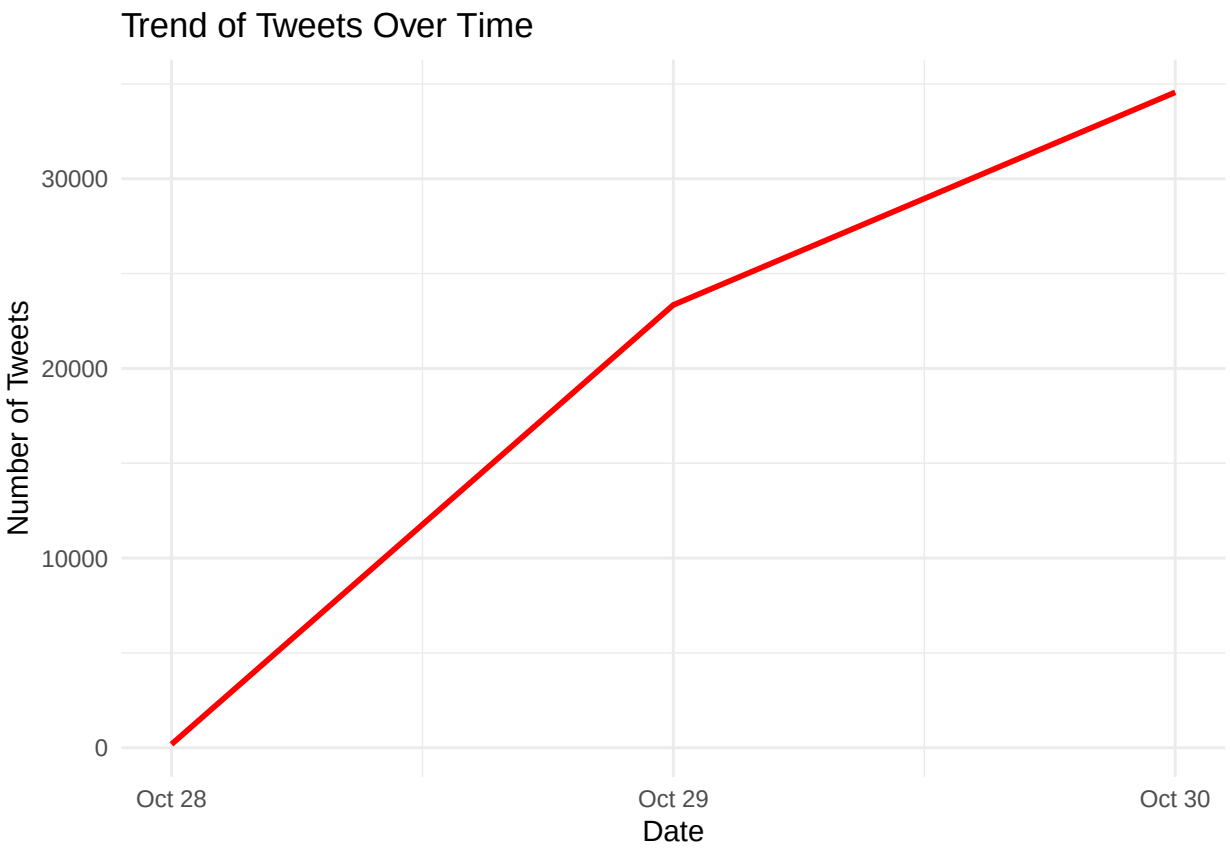
```
## New names:
## * `` -> `...1`
```

```
View(tweetsDF)
#Data Cleaning

tweetsDF <- tweetsDF %>%
  rename(
    screen_name = screenName,
    tweet = text,
    created_at = created,
    source = statusSource,
    rounded_time = Created_At_Round,
```

```
    tweet_source = tweetSource
  )

  tweetsDF <- tweetsDF %>%
    filter(!is.na(tweet), !is.na(created_at))

  tweetsDF$created_at <- as.POSIXct(tweetsDF$created_at, format = "%Y-%m-%d %H:%M:%S")

  str(tweetsDF)
```

```
## tibble [58,084 x 7] (S3: tbl_df/tbl/data.frame)
##  $ ...1        : num [1:58084] 1 2 3 4 5 6 7 8 9 10 ...
##  $ screen_name : chr [1:58084] "whourj31" "nnainot" "febry_sri_M" "telehuntwatch" ...
##  $ tweet       : chr [1:58084] "A soldier angry at the support fund consolation money for the bereave
##  $ created_at  : POSIXct[1:58084], format: "2022-10-30 23:59:43" "2022-10-30 23:59:32" ...
##  $ source      : chr [1:58084] "<a href=\"https://www.fs-poster.com/\" rel=\"nofollow\">FS_Poster_App
##  $ rounded_time: POSIXct[1:58084], format: "2022-10-31 00:00:00" "2022-10-31 00:00:00" ...
##  $ tweet_source: chr [1:58084] "others" "android" "android" "others" ...
```

```
  # number of tweets over time
tweets_trend <- tweetsDF %>%
  mutate(date = as.Date(created_at)) %>%
  group_by(date) %>%
  summarise(tweet_count = n())


ggplot(tweets_trend, aes(x = date, y = tweet_count)) +
  geom_line(color = "red", linewidth = 1) +
  labs(
    title = "Trend of Tweets Over Time",
    x = "Date",
    y = "Number of Tweets"
  ) +
  theme_minimal()
```




```
print(tweets_trend)
```

```
## # A tibble: 3 x 2
##   date       tweet_count
##   <date>           <int>
## 1 2022-10-28         181
## 2 2022-10-29       23347
## 3 2022-10-30       34556
```

The graph titled "Trend of Tweets Over Time" shows how many tweets were posted each day over a certain time period. The dates are on the x-axis, and the number of tweets is on the y-axis. The red line connects the daily tweet counts, making it easy to see when there were more tweets or fewer tweets. Some days have a lot more tweets, which could mean something important happened, while other days have fewer tweets. This graph helps us see patterns in how people were posting tweets over time.

```
library(readxl)
library(dplyr)
library(ggplot2)
library(tidyr)
library(stringr)
library(syuzhet)

tweetsDF <- read_excel("C:/PROJ/tweetsDF.xlsx")
```

```
## New names:
## * `` -> `...1`
```

```r
tweetsDF <- tweetsDF %>%
  rename(
    screen_name = screenName,
    tweet = text,
    created_at = created,
    source = statusSource,
    rounded_time = Created_At_Round,
    tweet_source = tweetSource
  )
```