

SentimentAnalysisProject_Sobusa_Tamonan_Delgado.Rmd

Nexon Sobusa

2024-12-14

```
#install.packages("lubridate")  
#install.packages("tidytext")  
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(stringr)
```

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

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.4.2
```

```
library(lubridate)
```

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

```
##
```

```
## Attaching package: 'lubridate'
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##   date, intersect, setdiff, union
```

```
library(tidytext)
```

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

```
tweets_data <- read.csv("C:/Users/kurts/Desktop/SentimentAnalysis/tweetsDF.csv")
```

```
# Remove duplicates
```

```
tweets_data <- tweets_data %>%  
  distinct()
```

```
missing_values_count <- colSums(is.na(tweets_data))
```

```

# Clean text column
tweets_data$text <- tweets_data$text %>%
  str_replace_all("http\\S+|www\\.\\S+", "") %>% # Remove URLs
  str_replace_all("[^[:alnum:][:space:]]", "") %>% # Remove special characters
  str_squish() # Remove extra whitespaces

print(head(tweets_data))

##      X      screenName
## 1 1      whourj31
## 2 2      nnainot
## 3 3  febry_sri_M
## 4 4 telehuntwatch
## 5 5      Typing0824
## 6 6  niccijsmith
##
## 1      A soldier angry at the support fund consolation money for the bereaved family of the Itaewon
## 2                                     Nah this Itaewon tragedy really has me s
## 3                                     JIN Pray for
## 4  TRANSLATION Seoul residents lay flowers at a makeshift memorial near the site of the crush in Ita
## 5 The Itaewon stampede incident really caught me off guard Makes me notice how important it is to kno
## 6      What to do about my child What to do about my child Park Gayoungs mother Choi Seonmi said
##
##      created
## 1 2022-10-30 23:59:43
## 2 2022-10-30 23:59:32
## 3 2022-10-30 23:59:31
## 4 2022-10-30 23:59:28
## 5 2022-10-30 23:59:20
## 6 2022-10-30 23:59:04
##
##      statusSource
## 1      <a href="https://www.fs-poster.com/" rel="nofollow">FS_Poster_App</a>
## 2 <a href="http://twitter.com/download/android" rel="nofollow">Twitter for Android</a>
## 3 <a href="http://twitter.com/download/android" rel="nofollow">Twitter for Android</a>
## 4      <a href="https://ruprop.live" rel="nofollow">telehunt</a>
## 5 <a href="http://twitter.com/download/android" rel="nofollow">Twitter for Android</a>
## 6  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
##      Created_At_Round tweetSource
## 1 2022-10-31 00:00:00      others
## 2 2022-10-31 00:00:00      android
## 3 2022-10-31 00:00:00      android
## 4 2022-10-31 00:00:00      others
## 5 2022-10-31 00:00:00      android
## 6 2022-10-31 00:00:00      iphone

print(missing_values_count)

##      X      screenName      text      created
##      0      0      0      0
##      statusSource Created_At_Round      tweetSource
##      0      0      0

# Trend Analysis
install.packages("lubridate")

```

```
## Warning: package 'lubridate' is in use and will not be installed
```

```
print(missing_values_count)
```

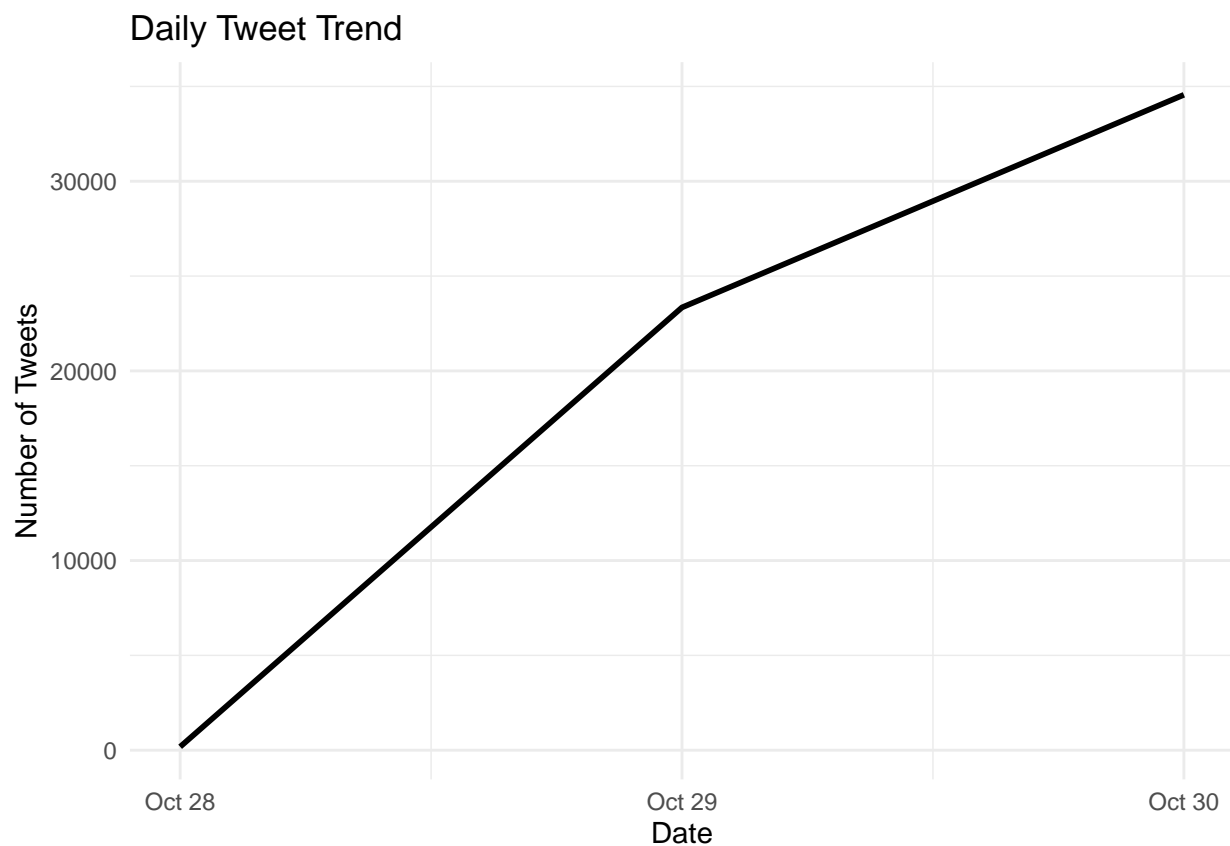
```
##           X      screenName      text      created
##           0           0           0           0
##  statusSource Created_At_Round  tweetSource
##           0           0           0
```

```
library(dplyr)
library(ggplot2)
library(lubridate)
# Trend Analysis
```

```
tweets_data$created_at <- ymd_hms(tweets_data$created)
```

```
daily_tweet_counts <- tweets_data %>%
  mutate(tweet_date = as_date(created_at)) %>%
  group_by(tweet_date) %>%
  summarise(tweet_count = n())
```

```
ggplot(daily_tweet_counts, aes(x = tweet_date, y = tweet_count)) +
  geom_line(color = "black", linewidth = 1) +
  labs(
    title = "Daily Tweet Trend",
    x = "Date",
    y = "Number of Tweets"
  ) +
  theme_minimal()
```



```

# Sentimental Analysis
install.packages("tidytext")

## Warning: package 'tidytext' is in use and will not be installed

library(dplyr)
library(tidytext)
library(ggplot2)

bing_sentiment_lexicon <- get_sentiments("bing")

# Tokenize the tweets and remove stop words
tweet_words <- tweets_data %>%
  unnest_tokens(word, text) %>%
  anti_join(stop_words)

## Joining with `by = join_by(word)`

sentiment_results <- tweet_words %>%
  inner_join(bing_sentiment_lexicon, by = "word") %>%
  count(sentiment) %>%
  mutate(percent = n / sum(n) * 100)

ggplot(sentiment_results, aes(x = sentiment, y = percent, fill = sentiment)) +
  geom_bar(stat = "identity") +
  labs(
    title = "Sentiment Analysis of Tweets",
    x = "Sentiment",
    y = "Percentage"
  ) +
  theme_minimal()

```

