

SentimentAnalysisProject_Sobusa_Tamonan_Delgado.Rmd

Nexon Sobusa

2024-12-14

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

# Remove duplicates
tweets_df <- tweets_df %>%
  distinct()

missing_values <- colSums(is.na(tweets_df))

# Clean the text column
tweets_df$text <- tweets_df$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_df))

print(missing_values)

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

# Trend Analysis
install.packages("lubridate")

## Installing package into 'C:/Users/kurts/AppData/Local/R/win-library/4.4'
## (as 'lib' is unspecified)
```

```
## Warning in install.packages :
##   unable to access index for repository https://cloud.r-project.org/src/contrib:
##   cannot open URL 'https://cloud.r-project.org/src/contrib/PACKAGES'
## Warning in install.packages :
##   package 'lubridate' is not available for this version of R
##
## A version of this package for your version of R might be available elsewhere,
## see the ideas at
## https://cran.r-project.org/doc/manuals/r-patched/R-admin.html#Installing-packages
## Warning in install.packages :
##   unable to access index for repository https://cloud.r-project.org/bin/windows/contrib/4.4:
##   cannot open URL 'https://cloud.r-project.org/bin/windows/contrib/4.4/PACKAGES'
```

```
library(dplyr)
library(ggplot2)
```

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

```
library(lubridate)
```

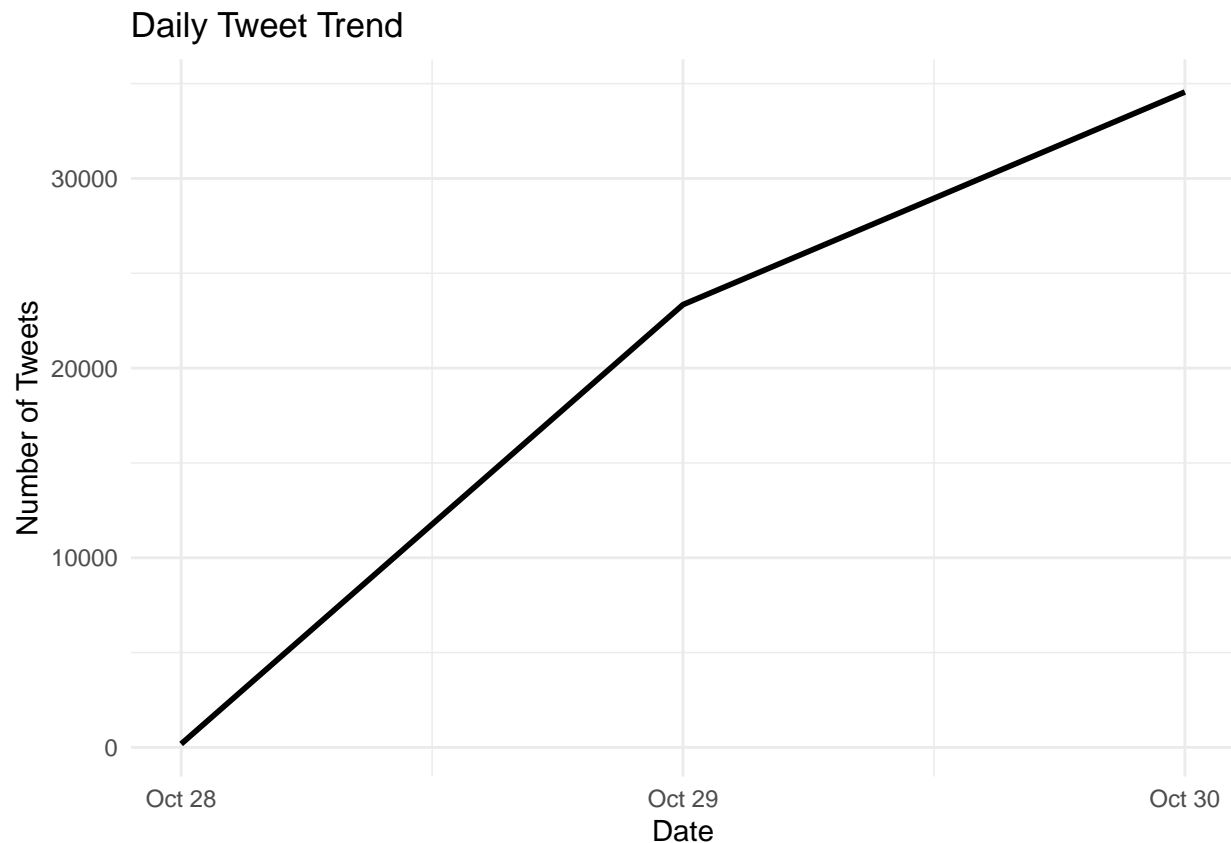
```
## Warning: package 'lubridate' was built under R version 4.4.2
```

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

```
tweets_df$created <- ymd_hms(tweets_df$created)
```

```
daily_trend <- tweets_df %>%
  mutate(date = as_date(created)) %>%
  group_by(date) %>%
  summarise(tweet_count = n())
```

```
ggplot(daily_trend, aes(x = 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")
```

```
## Installing package into 'C:/Users/kurts/AppData/Local/R/win-library/4.4'
## (as 'lib' is unspecified)

## Warning in install.packages :
##   unable to access index for repository https://cloud.r-project.org/src/contrib:
##   cannot open URL 'https://cloud.r-project.org/src/contrib/PACKAGES'
## Warning in install.packages :
##   package 'tidytext' is not available for this version of R
##
## A version of this package for your version of R might be available elsewhere,
## see the ideas at
## https://cran.r-project.org/doc/manuals/r-patched/R-admin.html#Installing-packages
## Warning in install.packages :
##   unable to access index for repository https://cloud.r-project.org/bin/windows/contrib/4.4:
##   cannot open URL 'https://cloud.r-project.org/bin/windows/contrib/4.4/PACKAGES'
```

```
library(dplyr)
library(tidytext)
```

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

```
library(ggplot2)
```

```
bing_lexicon <- get_sentiments("bing")
```

```
tokenized_tweets <- tweets_df %>%
  unnest_tokens(word, text) %>%
```

```
anti_join(stop_words)
```

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

```
sentiment_analysis <- tokenized_tweets %>%  
  inner_join(bing_lexicon, by = "word") %>%  
  count(sentiment) %>%  
  mutate(percent = n / sum(n) * 100)
```

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

