

SentimentAnalysis

Group6

2024-12-18

```
# Load necessary libraries
library(dplyr)

##
## 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(lubridate)

##
## Attaching package: 'lubridate'
## The following objects are masked from 'package:base':
##
##   date, intersect, setdiff, union
library(syuzhet) # For sentiment analysis

# Read the CSV file
tweets_df <- read.csv("/cloud/project/SentimentAnalysis/tweetsDF.csv", stringsAsFactors = FALSE)

# DATA CLEANING
tweets_df$created <- as.POSIXct(tweets_df$created, format = "%Y-%m-%d %H:%M:%S")

tweets_df <- tweets_df %>% distinct()

colSums(is.na(tweets_df))

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

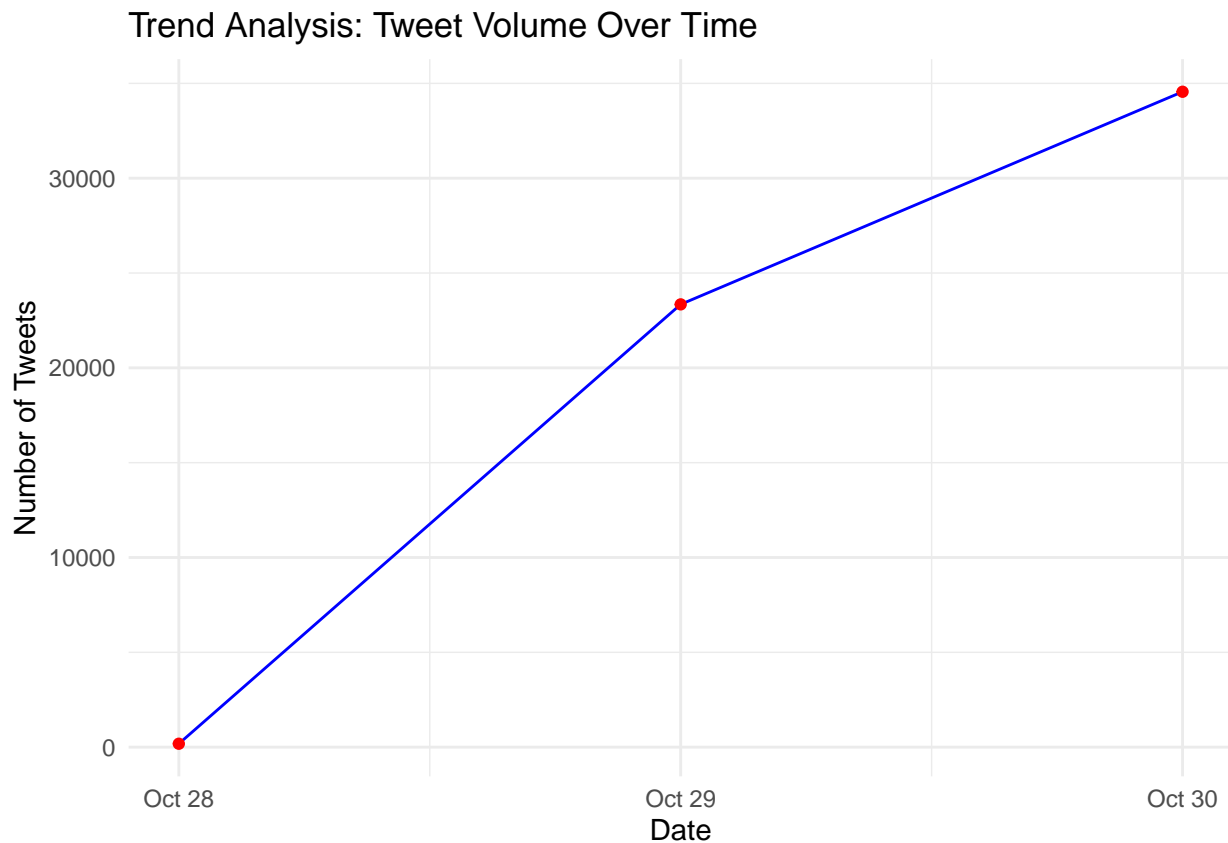
#TREND ANALYSIS
tweets_per_day <- tweets_df %>%
  mutate(date = as.Date(created)) %>%
  group_by(date) %>%
```

```

summarise(tweet_count = n())

ggplot(tweets_per_day, aes(x = date, y = tweet_count)) +
  geom_line(color = "blue") +
  geom_point(color = "red") +
  labs(title = "Trend Analysis: Tweet Volume Over Time",
       x = "Date",
       y = "Number of Tweets") +
  theme_minimal()

```



```

#SENTIMENT ANALYSIS
tweets_df$sentiment <- get_sentiment(tweets_df$text, method = "syuzhet")

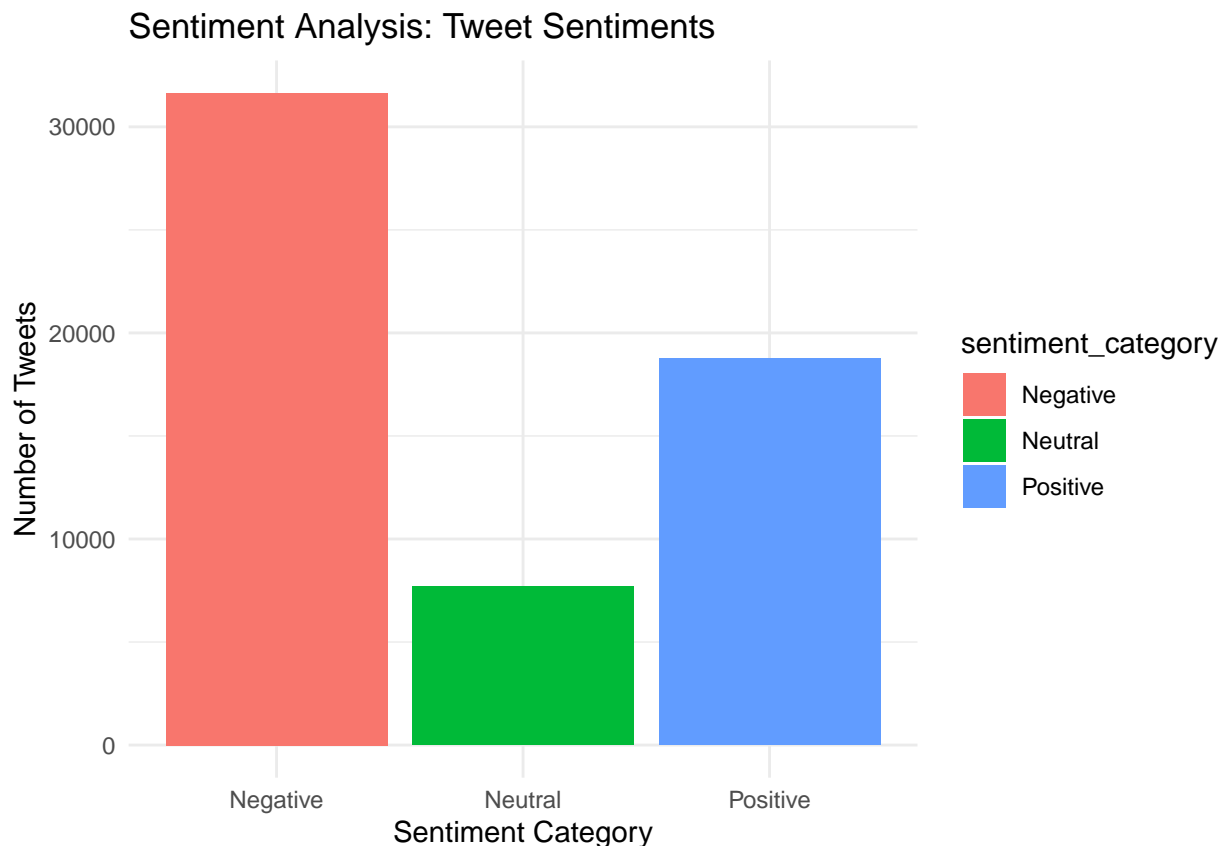
tweets_df$sentiment_category <- ifelse(tweets_df$sentiment > 0, "Positive",
                                       ifelse(tweets_df$sentiment < 0, "Negative", "Neutral"))

sentiment_counts <- tweets_df %>%
  group_by(sentiment_category) %>%
  summarise(count = n())

ggplot(sentiment_counts, aes(x = sentiment_category, y = count, fill = sentiment_category)) +
  geom_bar(stat = "identity") +
  labs(title = "Sentiment Analysis: Tweet Sentiments",

```

```
x = "Sentiment Category",
y = "Number of Tweets" +
theme_minimal()
```



#2. Present your Use Case on what you will do with the dataset.

#-The dataset contains tweets, including the tweet text, date, and source. The use case is to analyze trends in the volume of tweets over time and perform sentiment analysis to understand whether people's opinions in the tweets are positive, negative, or neutral. This can help identify patterns in user behavior, sentiment shifts during specific dates, and overall public opinion trends.

#3. Make sure to describe each graph you have created. Give insights.

#Trend Analysis

#-The trend analysis shows how the number of tweets changes over time. By counting tweets for each day, we can see when there was an increase or decrease in activity. For example, certain days might have a higher number of tweets, which could be due to important events, trending topics, or news that people discussed more on those days. This helps identify patterns in user engagement and the timing of important discussions.

#Sentiment Analysis

#-The sentiment analysis categorizes tweets as Positive, Negative, or Neutral based on the words used in each tweet. This gives an idea of the overall mood or opinion of people. If most tweets are Positive, it suggests a favorable public opinion. If there are more Negative tweets, it indicates dissatisfaction or criticism. A high number of Neutral tweets means the content is more factual or without strong emotion. This analysis helps understand how people feel about specific topics or events discussed in the tweets.