sentiment analysis

lomibao and rabago

2024-12-05

installing the packages.

```
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("stringr")
library("lubridate")
```

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

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

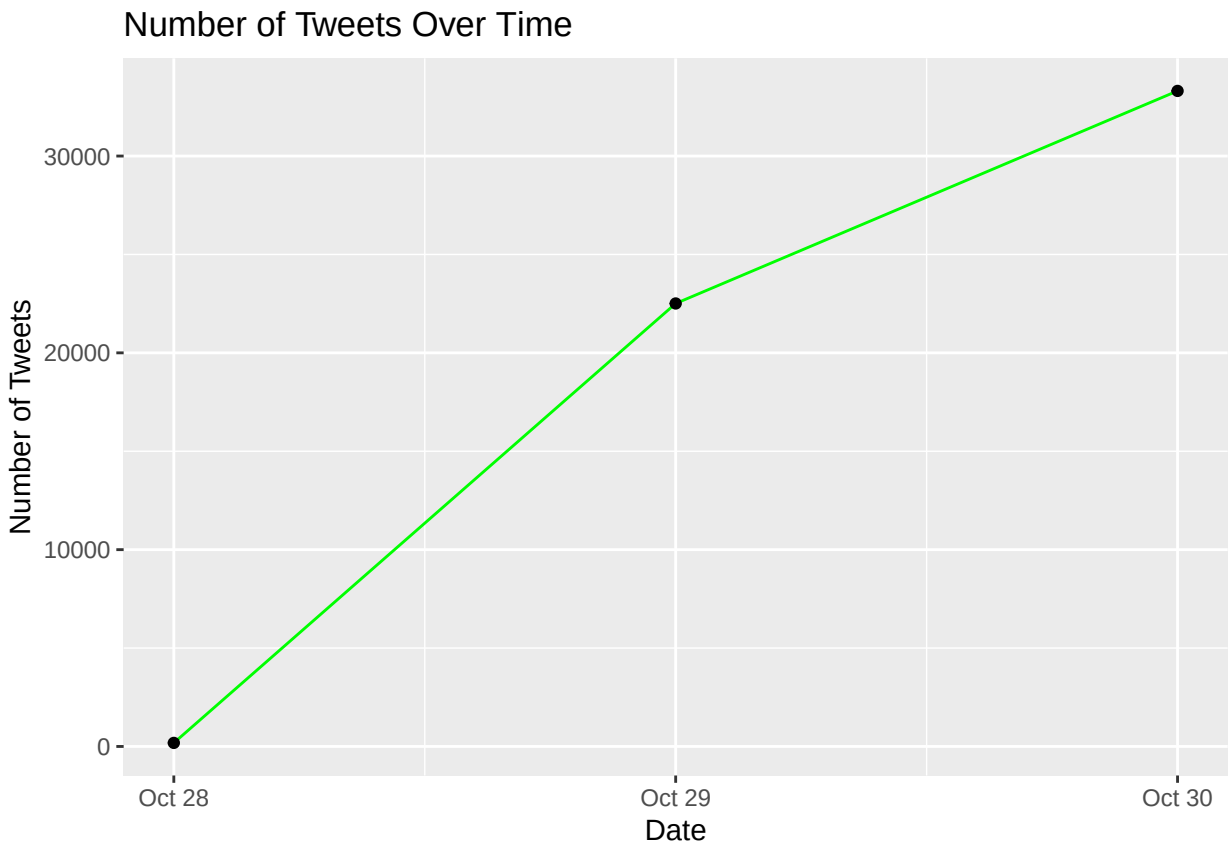
importing the csv.file.

```
dataf <- read.csv("/cloud/project/sentiment analysis/tweetsDF.csv")
```

cleaning the data from the dataset.

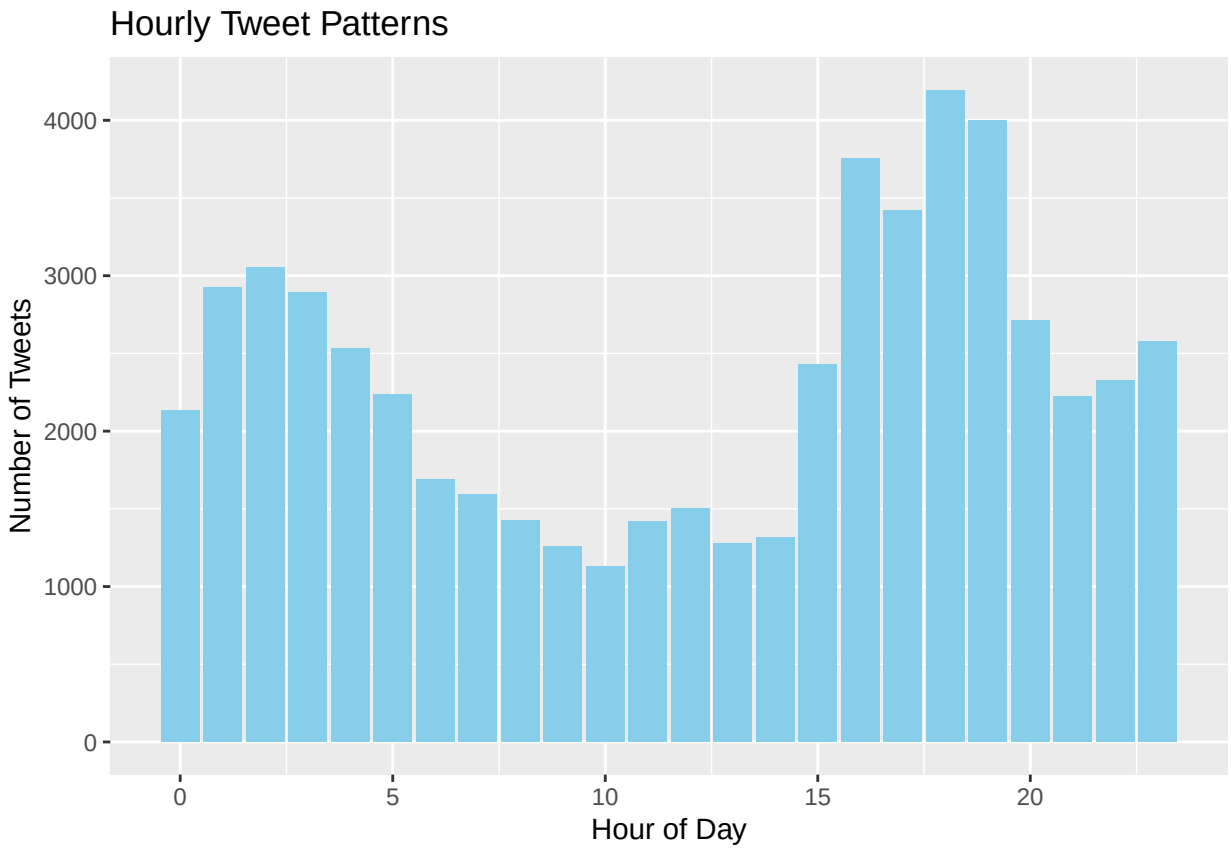
```
cleaned_tweetsDf <- dataf %>%
  select(-c( statusSource, Created_At_Round)) %>%
  mutate(created = as.POSIXct(created, format = "%Y-%m-%d %H:%M:%S"),
         date = as.Date(created),
         hour = hour(created),
         day_of_week = weekdays(created)) %>%
  distinct(text, .keep_all = TRUE)

# Trend 1: Tweet Volume Over Time
tweet_trend <- cleaned_tweetsDf %>%
  group_by(date) %>%
  summarise(tweet_count = n())

#plotting the trend
ggplot(tweet_trend, aes(x = date, y = tweet_count)) +
```

```
  geom_line(color = "green") +
  geom_point() +
  theme_gray() +
  labs(title = "Number of Tweets Over Time",
       x = "Date",
       y = "Number of Tweets")
```



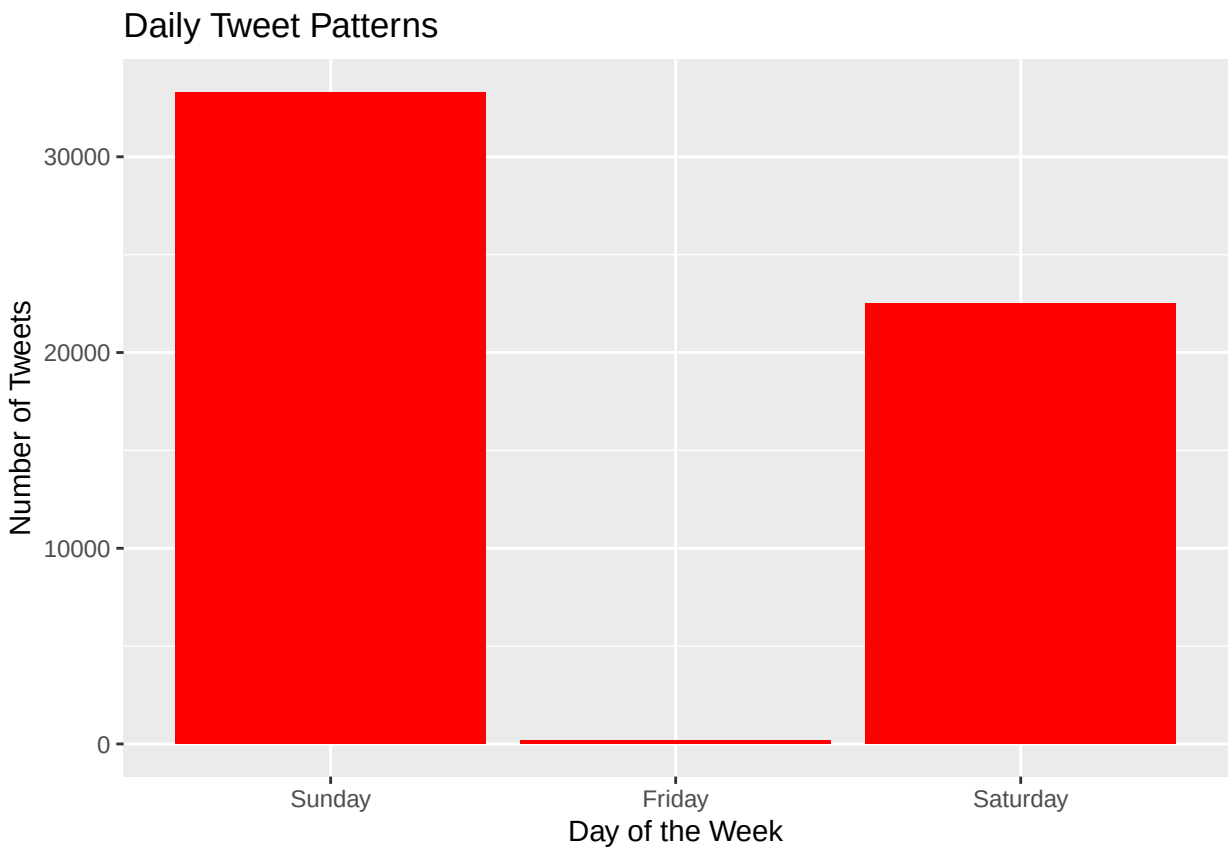
Trend volume but per hour.

```
hourly_trend <- cleaned_tweetsDf %>%
  group_by(hour) %>%
  summarise(tweet_count = n())

ggplot(hourly_trend, aes(x = hour, y = tweet_count)) +
  geom_bar(stat = "identity", fill = "skyblue") +
  theme_gray() +
  labs(title = "Hourly Tweet Patterns",
       x = "Hour of Day",
       y = "Number of Tweets")
```



Trend patterns in week;

```
daily_trend <- cleaned_tweetsDf %>%
  group_by(day_of_week) %>%
  summarise(tweet_count = n()) %>%
  mutate(day_of_week = factor(day_of_week,
                              levels = c("Sunday", "Monday", "Tuesday", "Wednesday",
                                         "Thursday", "Friday", "Saturday")))

ggplot(daily_trend, aes(x = day_of_week, y = tweet_count)) +
  geom_bar(stat = "identity", fill = "red") +
  theme_gray() +
  labs(title = "Daily Tweet Patterns",
       x = "Day of the Week",
       y = "Number of Tweets")
```


Daily Tweet Patterns
30000
20000
10000
0
Number of Tweets
Sunday
Friday
Saturday
Day of the Week