TrendAnalysisSentimentAnalysisProject(Berja, Bibit, Buenvenida)

Bibit, Berja, Buenvenida

2024-12-16

The goal of this research is to investigate popular sentiment toward BLACKPINK and its members throughout a period overlapping with both their Houston concert and the Itaewon tragedy. In particular in which, we want to:

1. Recognize how outside events affect mood and activity on social media of the people.

2. Give BLACKPINK's management team useful information on public opinion and involvements.

3. Determine important patterns and feelings to guide public relations management tactics during emergencie situations.

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----------------------- tidyverse 2.0.0 --
## v dplyr     1.1.4      v readr     2.1.5
## v forcats   1.0.0      v stringr   1.5.1
## v ggplot2   3.5.1      v tibble    3.2.1
## v lubridate 1.9.4      v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ----------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become error:
```

```
library(lubridate)
library(ggplot2)
library(tidytext)
library(sentimentr)
```

– Data Loading and Inspection –

```
# Load the dataset
data <- read.csv("/cloud/project/tweetsDF.csv")

# Inspect the structure of the data
str(data)
```

```
## 'data.frame':    58086 obs. of  7 variables:
## $ X              : int  1 2 3 4 5 6 7 8 9 10 ...
## $ screenName     : chr  "whourj31" "nnainot" "febry_sri_M" "telehuntwatch" ...
## $ text           : chr  "A soldier angry at the support fund consolation money for the bereaved far
## $ created        : chr  "2022-10-30 23:59:43" "2022-10-30 23:59:32" "2022-10-30 23:59:31" "2022-10-
## $ statusSource   : chr  "<a href=\"https://www.fs-poster.com/\" rel=\"nofollow\">FS_Poster_App</a>"
## $ Created_At_Round: chr  "2022-10-31 00:00:00" "2022-10-31 00:00:00" "2022-10-31 00:00:00" "2022-10-
## $ tweetSource    : chr  "others" "android" "android" "others" ...
```

```r
# Check for missing values
summary(data)
```

```
##        X            screenName            text              created
##  Min.   :    1   Length:58086       Length:58086       Length:58086
##  1st Qu.:14522   Class :character   Class :character   Class :character
##  Median :29044   Mode  :character   Mode  :character   Mode  :character
##  Mean   :29044
##  3rd Qu.:43565
##  Max.   :58086
##  statusSource       Created_At_Round    tweetSource
##  Length:58086       Length:58086       Length:58086
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##
##
##
```

```r
# Display the first few rows of the dataset
head(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 Itaeu
## 2                                                                                        Nah this Itae
## 3
## 4 TRANSLATION :\nSeoul residents lay flowers at a makeshift memorial near the site of the crush in I
## 5  The Itaewon stampede incident really caught me off guard. Makes me notice how important it is to
## 6  "What to do about my child? What to do about my child?" Park Ga-young's mother, Choi Seon-mi, sai
##             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
```

– Data Cleaning –

```r
# Data Cleaning
# Remove duplicates and handle missing values
data <- data %>% distinct() %>% drop_na()

# Ensure the dataset contains a 'created' column
if (!"created" %in% colnames(data)) {
  stop("The dataset does not contain a 'created' column. Please ensure the dataset includes a timestamp
}

# Rename 'created' to 'datetime' for consistency
data <- data %>% rename(datetime = created)

# Convert 'datetime' to a proper date-time format
data <- data %>% mutate(datetime = ymd_hms(datetime))

# Text preprocessing: Remove URLs, hashtags, mentions, and special characters
data <- data %>%
  mutate(text = str_remove_all(text, "http[s]?://\\S+|#\\w+|@\\w+|[^\\w\\s]"))

# Filter tweets mentioning Blackpink or its members
keywords <- c("Blackpink", "Rose", "Lisa", "Jennie", "Jisoo", "concert")
data_filtered <- data %>% filter(str_detect(tolower(text), paste(keywords, collapse = "|")))
```

– Trend Analysis –

```r
# Restrict data to October 28-30, 2022
data_filtered <- data_filtered %>%
  filter(datetime >= as.POSIXct("2022-10-28 00:00:00") & datetime <= as.POSIXct("2022-10-30 23:59:59"))

# Count tweets over time
time_trend <- data_filtered %>%
  mutate(date = as.Date(datetime)) %>%
  group_by(date) %>%
  summarise(tweet_count = n())

# Ensure all dates in the range are included, even with zero counts
date_range <- seq(as.Date("2022-10-28"), as.Date("2022-10-30"), by = "day")
time_trend <- time_trend %>%
  complete(date = date_range, fill = list(tweet_count = 0))

# Plot tweet trends over time
ggplot(time_trend, aes(x = date, y = tweet_count)) +
  geom_line(color = "blue", size = 1) +
  labs(title = "Tweet Trends for Blackpink and Members (Oct 28-30)", x = "Date", y = "Number of Tweets")
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5, size = 16), axis.title = element_text(size = 12))
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

Tweet Trends for Blackpink and Members (Oct 28–30)