PRODIGY\_DS\_04.R

anush

2024-07-30

library(tidyverse)

## Warning: package 'tidyverse' was built under R version 4.3.3

## Warning: package 'ggplot2' was built under R version 4.3.3

## Warning: package 'tidyr' was built under R version 4.3.3

## Warning: package 'readr' was built under R version 4.3.3

## Warning: package 'dplyr' was built under R version 4.3.3

## Warning: package 'forcats' was built under R version 4.3.3

## Warning: package 'lubridate' was built under R version 4.3.3

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.4 ✔ readr 2.1.5  
## ✔ forcats 1.0.0 ✔ stringr 1.5.1  
## ✔ ggplot2 3.5.1 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.3 ✔ tidyr 1.3.1  
## ✔ purrr 1.0.2   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(sentimentr)

## Warning: package 'sentimentr' was built under R version 4.3.3

# Load the dataset  
library(readxl)

## Warning: package 'readxl' was built under R version 4.3.3

data <- read\_excel("twitter\_entity\_sentiment.xlsx")  
View(data)  
  
# View the first few rows of the dataset  
head(data)

## # A tibble: 6 × 4  
## `2401` Borderlands Positive im getting on borderlands and i will murder you …¹  
## <dbl> <chr> <chr> <chr>   
## 1 2401 Borderlands Positive I am coming to the borders and I will kill you al…  
## 2 2401 Borderlands Positive im getting on borderlands and i will kill you all,  
## 3 2401 Borderlands Positive im coming on borderlands and i will murder you al…  
## 4 2401 Borderlands Positive im getting on borderlands 2 and i will murder you…  
## 5 2401 Borderlands Positive im getting into borderlands and i can murder you …  
## 6 2402 Borderlands Positive So I spent a few hours making something for fun. …  
## # ℹ abbreviated name: ¹​`im getting on borderlands and i will murder you all ,`

# Check the column names  
colnames(data)

## [1] "2401"   
## [2] "Borderlands"   
## [3] "Positive"   
## [4] "im getting on borderlands and i will murder you all ,"

# Select relevant columns (assuming 'text' is the column with tweet text)  
data <- data %>% select(`im getting on borderlands and i will murder you all ,`)  
  
# Perform sentiment analysis  
sentiment\_results <- sentiment(data$`im getting on borderlands and i will murder you all ,`)

## Warning: Each time `sentiment` is run it has to do sentence boundary disambiguation when a  
## raw `character` vector is passed to `text.var`. This may be costly of time and  
## memory. It is highly recommended that the user first runs the raw `character`  
## vector through the `get\_sentences` function.

# View the sentiment results  
head(sentiment\_results)

## Key: <element\_id, sentence\_id>  
## element\_id sentence\_id word\_count sentiment  
## <int> <int> <int> <num>  
## 1: 1 1 12 -0.21650635  
## 2: 2 1 10 -0.23717082  
## 3: 3 1 10 -0.23717082  
## 4: 4 1 11 -0.22613351  
## 5: 5 1 10 -0.23717082  
## 6: 6 1 10 0.03162278

# Summarize sentiment scores  
summary\_sentiment <- sentiment\_results %>%  
 group\_by(sentiment) %>%  
 summarize(count = n())  
  
print(summary\_sentiment)

## # A tibble: 10,568 × 2  
## sentiment count  
## <dbl> <int>  
## 1 -2.27 1  
## 2 -2.11 1  
## 3 -2.08 1  
## 4 -2.04 2  
## 5 -1.91 1  
## 6 -1.85 1  
## 7 -1.84 1  
## 8 -1.82 1  
## 9 -1.79 2  
## 10 -1.78 1  
## # ℹ 10,558 more rows

# Plot sentiment distribution  
sentiment\_results %>%  
 ggplot(aes(x = sentiment)) +  
 geom\_histogram(binwidth = 0.1) +  
 labs(title = "Sentiment Distribution", x = "Sentiment Score", y = "Frequency")

