بسم الله الرحمن الرحیم

تحلیل احساسات توییت‌ها با R

استاد راهنما : دکتر حامد لروند

تهیه کننده : محمد امین طاوسی

مقدمه و توضیح کلی درباره پروژه

این پروژه با عنوان "استخراج متن و تحلیل احساسات توییت‌ها" به بررسی و تحلیل توییت های رد و بدل شده در میان پشتیبانی شرکت های و مشتری ها در برنامه X یا توییتر سابق می‌پردازد. هدف اصلی این پروژه استخراج بینش‌های ارزشمند از متن‌های کاربران و تحلیل احساسات نهفته در توییت‌ها است. این تحلیل به ما کمک می‌کند تا دیدگاه‌ها، حالات احساسی و رفتار کاربران را در زمینه‌های مختلف بررسی کنیم.

اهمیت تحلیل احساسات و استخراج متن

تحلیل احساسات یکی از کاربردی‌ترین تکنیک‌ها در علم داده و پردازش زبان طبیعی (NLP) است. این روش به ما این امکان را می‌دهد که داده‌های متنی بزرگ، نظیر توییت‌ها، بررسی شوند و اطلاعاتی مانند احساسات مثبت، منفی یا خنثی استخراج شود. تحلیل متن می‌تواند در زمینه‌های زیر کاربرد داشته باشد:

- تجزیه و تحلیل بازخورد مشتریان: شرکت‌ها می‌توانند از تحلیل احساسات برای درک نظرات مشتریان استفاده کنند.
- پیش‌بینی روندها و رفتارها: تحلیل توییت‌ها می‌تواند برای پیش‌بینی رفتارهای اجتماعی یا تغییرات بازار مفید باشد.
- شناسایی موضوعات مهم: کمک به فهم موضوعات پررنگ و کلمات کلیدی در مکالمات آنلاین.

اهداف پروژه

1. استخراج و پاک‌سازی داده‌ها: جمع‌آوری داده‌های توییتر و آماده‌سازی آن‌ها برای تحلیل.
2. تحلیل کلمات پرتکرار: شناسایی واژه‌هایی که بیشتر در مکالمات استفاده می‌شوند.

3. تحلیل احساسات: استفاده از فرهنگ لغت‌های مختلف (مانند AFINN، Bing و NRC) برای تعیین احساسات مثبت و منفی کاربران.
4. بصری‌سازی نتایج: نمایش نتایج تحلیل به صورت نمودارها برای درک بهتر روندها.

ساختار پروژه

پروژه طبق این مراحل پیش می‌رود:

1. بارگذاری داده‌ها و بسته‌های مورد نیاز.
2. پاک‌سازی داده‌ها از لینک‌ها، کلمات زشت و نویسه‌های غیرضروری.
3. تحلیل زمان‌بندی انتشار توییت‌ها.
4. استخراج کلمات پرتکرار و حذف کلمات غیر مفید (Stop Words).
5. انجام تحلیل احساسات با استفاده از روش‌های مختلف.
6. رسم نمودارها و ارائه نتایج.

در ادامه، به بررسی دقیق‌تر مراحل این پروژه خواهیم پرداخت.

مرحله اول : جمع آوری داده ها

دریافت api توییتر از سایت https://developer.x.com که سایت رسمی توییتر برای گرفتن api هست و استخراج دیتاست مورد نیاز .

داده‌های پروژه و ویژگی‌های آن

داده‌های این پروژه شامل مجموعه‌ای از توییت‌های کاربران است که اطلاعات ارزشمندی را برای تحلیل فراهم می‌کنند. این داده‌ها به صورت یک فایل اکسل یا CSV در دسترس

هستند و ساختار آن شامل چندین ستون کلیدی است که هر کدام نماینده جنبه‌ای از اطلاعات توییت‌ها هستند.

ویژگی‌های اصلی داده‌ها

1. شناسه توییت (“tweet_id”):

- شناسه منحصربه‌فرد برای هر توییت.
- این ستون برای شناسایی و پیگیری هر توییت در مجموعه داده استفاده می‌شود.

2. شناسه نویسنده (“author_id”):

- شناسه کاربری که توییت را ارسال کرده است.
- این ویژگی مشخص می‌کند که هر توییت توسط کدام کاربر نوشته شده است.

3. وضعیت ورودی/خروجی (“inbound”):

- نشان‌دهنده جهت ارتباط توییت (ورودی یا خروجی).
- مقدار TRUE نشان‌دهنده توییت‌های ورودی (از سمت کاربران) و FALSE برای توییت‌های خروجی (پاسخ به کاربران).

4. زمان انتشار (“created_at”):

- تاریخ و زمان دقیق ارسال هر توییت.
- این ویژگی به تحلیل زمانی رفتار کاربران کمک می‌کند (مانند شناسایی بیشترین زمان فعالیت).

5. متن توییت ("text"):

- این ستون شامل محتوای اصلی توییت‌ها است.
- داده‌ها ممکن است شامل لینک‌ها، ایموجی‌ها و سایر نویسه‌های متنی باشند که در مراحل پاک‌سازی حذف می‌شوند.

6. شناسه پاسخ ("response_tweet_id"):

- شناسه توییتی که پاسخ داده شده است.
- این ستون مشخص می‌کند که آیا یک توییت پاسخ به توییت دیگر است یا خیر.

7. شناسه مرجع ("in_response_to_tweet_id"):

- شناسه توییتی که این توییت به آن پاسخ داده است.
- برای ردیابی مکالمات و روابط بین توییت‌ها استفاده می‌شود.

مرحله دوم : وارد کردن داده‌ها در محیط R ، نصب و فراخوانی تابع های مورد نیاز

وارد کردن داده ها :

```
# step 2

Tweets<-choose.files()
tweets_df <- Tweets
```

نصب و فراخوانی کتابخانه ها :

```
install.packages("dplyr")
install.packages("ggplot2")
install.packages("lubridate")
install.packages("stringr")
install.packages("tidytext")
install.packages("textdata")
install.packages("sentimentr")
install.packages("lexicon")
install.packages("tm")
install.packages("tibble")
install.packages("afinn")
```

```
library(dplyr)
library(ggplot2)
library(lubridate)
library(stringr)
library(tidytext)
library(textdata)
library(sentimentr)
library(lexicon)
library(tm)
library(tibble)
library(afinn)
```

توضیح کتابخانه ها :

در پروژه از کتابخانه‌های زیر استفاده شده است. این کتابخانه‌ها ابزارهای مختلفی برای پردازش داده، تحلیل متن و رسم نمودار فراهم می‌کنند. در ادامه، کاربرد هر کدام با جزئیات بیشتری توضیح داده می‌شود:

1. dplyr

کاربرد: مدیریت و پردازش داده‌ها.

در پروژه: انتخاب ستون‌های کلیدی و مرتب‌سازی داده‌ها.

فیلتر کردن توییت‌ها بر اساس معیارهای خاص مانند وضعیت ورودی یا خروج بودن.

2. ggplot2

کاربرد: رسم نمودارهای بصری.

در پروژه: نمایش توزیع فراوانی کلمات پرتکرار.

رسم نمودارهای مرتبط با احساسات (مثبت و منفی).

3. lubridate

کاربرد: پردازش زمان و تاریخ.

در پروژه: استخراج اطلاعات زمانی مثل تاریخ و ساعت انتشار توییت‌ها.

شناسایی روندهای زمانی در داده‌ها.

4. stringr

کاربرد: پردازش رشته‌های متنی.

در پروژه: حذف لینک‌ها، ایموجی‌ها و نویسه‌های اضافی از متن توییت‌ها.

5. tidytext

کاربرد: تحلیل متن به صورت ساختارمند.

در پروژه: تجزیه متن به کلمات مستقل برای تحلیل کلمات پرتکرار و احساسات.

حذف کلمات بی‌معنی که تاثیری در تحلیل ندارند.

6. textdata

کاربرد: ارائه مجموعه داده‌های از پیش آماده برای تحلیل.

در پروژه: بارگذاری فرهنگ لغت‌های مربوط به تحلیل احساسات (AFINN، Bing، NRC).

7. sentimentr

کاربرد: تحلیل احساسات در سطح متن.

در پروژه: شناسایی کلمات مثبت و منفی در توییت‌ها.

تولید امتیاز کلی احساسات هر توییت.

8. lexicon

کاربرد: دسترسی به لیست‌های متنی خاص.

در پروژه: حذف کلمات زشت یا غیر مفید برای بهبود کیفیت تحلیل.

9. tm (Text Mining)

کاربرد: انجام پیش‌پردازش‌های متنی.

در پروژه: ساخت مجموعه متن (Corpus) و حذف علائم نگارشی یا اعداد از متن.

10. tibble

کاربرد: مدیریت داده‌ها در قالب ساده‌تر.

در پروژه: ساخت داده‌های ساختار یافته برای تحلیل راحت‌تر.

11. afinn

کاربرد: تحلیل احساسات با مقیاس عددی.

در پروژه: امتیازدهی به کلمات و تعیین میزان مثبت یا منفی بودن آن‌ها

مرحله سوم : پاک سازی داده‌ها

```
# Step 3

# creating a new data frame
clean_tweets_df <- tweets_df
```

اهداف این بخش :

- کاهش نویز در داده‌ها.
- افزایش دقت تحلیل متن و احساسات.
- آماده‌سازی داده‌ها برای مرحله تجزیه و تحلیل معنایی.

مراحل پاک‌سازی داده‌ها

1. حذف لینک‌ها

کد مرتبط:

```
# Removing URLs
clean_tweets_df$text <- str_remove_all(clean_tweets_df$text,
"\\bhttps?://[[:alnum:]]+\\.[[:alnum:]]+(/[[:alnum:].,_/-]*)?")
```

- هدف: لینک‌های موجود در متن، اطلاعاتی درباره محتوا ارائه نمی‌دهند و می‌توانند دقت تحلیل را کاهش دهند. این کد تمامی لینک‌های شروع‌شده با http یا https را از متن حذف می‌کند.

2. حذف ایموجی‌ها و کاراکترهای غیرضروری

کد:

```
# Removing emojis
clean_tweets_df$text <- str_remove_all
(clean_tweets_df$text, "\\p{Emoji}")
```

- هدف: ایموجی‌ها یا کاراکترهای گرافیکی ممکن است برای تحلیل احساسات مفید نباشند، بنابراین حذف می‌شوند.
-

3. حذف کلمات زشت

کد مرتبط:

```
# removing Profanity
#Profanity list using 3 in the provided list.

profanity_list <- c(lexicon::profanity_alvarez,
lexicon::profanity_banned, lexicon::profanity_arr_bad)

words_in_tweets_df <- clean_tweets_df %>%
  dplyr::select(text) %>%
  unnest_tokens(word, text) %>%
  anti_join(stop_words) %>%
  #anti_join(data.frame(profanity_list)) %>%
  filter(!word %in% c("rt", "t.co", profanity_list))
```

- سه لیست از کلمات زشت و نامناسب (profanity_alvarez، profanity_banned و profanity_arr_bad) از کتابخانه lexicon ترکیب شده و به یک لیست واحد به نام profanity_list تبدیل می‌شود.
- این لیست شامل کلماتی است که در توییت‌ها ممکن است نامناسب باشند و باید حذف شوند.
- کلماتی مثل "of"، "the"، "and" و سایر کلمات بی‌معنی که در تحلیل معنایی مفید نیستند، حذف می‌شوند و این کار با استفاده از یک لیست پیش‌فرض به نام stop_words انجام می‌شود.
- در انتها کلماتی که در لیست زیر هستند با دستور filter حذف می‌شوند:
  - rt: نشان‌دهنده ریتوییت‌ها (Retweets) که برای تحلیل احساسات مفید نیستند.
  - t.co: لینک‌های کوتاه‌شده که توسط توییتر به‌طور خودکار ایجاد می‌شوند.

- profanity_list: لیست کلمات زشت و نامناسب که در ابتدا تعریف شده است.

مرحله چهارم : استخراج و تحلیل جزئیات زمانی از داده‌ها

استاندارد سازی زمان :

کد مرتبط :

```
# reformatting the created_at field of new data frame using POSIXct
clean_tweets_df$created_at <- as.POSIXct(clean_tweets_df$created_at,
format = "%a %b %d %H:%M:%S %z %Y")
```

این کد ستون created_at در دیتا فریم clean_tweets_df را به فرمت استاندارد زمانی تبدیل می‌کند. این تغییر باعث می‌شود بتوانیم تاریخ و زمان را به‌راحتی تحلیل کرده یا عملیات‌هایی مانند مرتب‌سازی یا استخراج الگوهای زمانی را انجام دهیم.

```
# step 4
# creating a new data frame with additional 6 columns from crated_at

clean_tweets_addition <- clean_tweets_df %>%
  mutate(
    year = year(created_at),
    month = month(created_at),
    day = day(created_at),
    hour = hour(created_at),
    date = as.Date(created_at),
    weekday = weekdays(as.Date(created_at))
  )
```

این کد اطلاعات زمانی توییت‌ها را از ستون created_at استخراج می‌کند و به دیتا فریم جدید اضافه می‌کند. به این ترتیب، جزئیاتی مثل سال، ماه، روز، ساعت، تاریخ (بدون زمان) و روز هفته به‌صورت ستون‌های جداگانه ذخیره می‌شوند. این کار به ما کمک می‌کند الگوهای زمانی رفتار کاربران را شناسایی کنیم، مثلاً بفهمیم کاربران در چه روزهایی از هفته یا چه ساعاتی از روز بیشتر فعالیت دارند. این مرحله یک پیش‌نیاز مهم برای تحلیل‌های زمانی و استخراج بینش‌های دقیق‌تر از داده‌ها است.

مرحله پنجم : حذف کلمات بی‌معنی و استخراج کلمات معنادار از متن

- این بخش از کد تمرکز بر حذف کلمات بی‌معنی (Stop Words) دارد. هدف اصلی این است که فقط کلمات معنادار و مهم در توییت‌ها باقی بمانند. کلمات بی‌معنی مثل "is"، "and"، "the" که معمولاً در هر متنی دیده می‌شوند اما ارزش تحلیلی ندارند، از متن حذف می‌شوند.

1. تعریف لیست کلمات بی‌معنی (Stop Words):

کد مرتبط :

```
# Get a list of English stop words
stop_words <- tm::stopwords("en")
```

- یک لیست پیش‌ساخته از کلمات بی‌معنی انگلیسی با استفاده از کتابخانه tm تعریف می‌شود.

- این کلمات شامل موارد رایج مثل "is"، "the"، "and"، و غیره هستند.

2. تعریف تابع remove_stopwords:

کد مرتبط :

```
# Function to remove stop words from a text
remove_stopwords <- function(text) {
  words <- unlist(strsplit(as.character(text), " "))
  meaningful_words <- words[!(words %in% stop_words)]
  return(meaningful_words)
}
```

- متن هر توییت به کلمات جداگانه (Tokens) تقسیم می‌شود.
- کلمات موجود در لیست کلمات بی‌معنی حذف می‌شوند.
- فقط کلمات معنادار باقی می‌مانند.

3. اعمال تابع روی توییت‌ها:

```
# Apply the function to create a new data frame with non-stop words
tweets_non_stopwords <- clean_tweets_df %>%
  mutate(cleaned_text = sapply(text, remove_stopwords))
```

- تابع remove_stopwords روی هر توییت اعمال می‌شود.
- ستون جدیدی به نام cleaned_text به دیتافریم اضافه می‌شود که شامل نسخه پاک‌سازی‌شده توییت‌ها است (بدون کلمات بی‌معنی).

4. ساخت دیتا فریم جدید با کلمات معنادار:

```
# Create a new data frame containing all non-stop words
non_stop_words <- clean_tweets_df %>%
  unnest_tokens(word, text) %>%
  anti_join(stop_words)
```

- متن توییت‌ها به کلمات جداگانه (Tokens) شکسته می‌شود.
- کلمات بی‌معنی با استفاده از anti_join(stop_words) حذف می‌شوند.
- دیتا فریم جدید non_stop_words شامل فقط کلمات معنادار است.

مرحله ششم : شناسایی و بصری‌سازی کلمات پرتکرار در توییت‌ها

(قسمت a)

1. شناسایی 20 کلمه پرتکرار:

کد مرتبط :

```
# Find the top 20 unique words
top_words <- words_in_tweets_df %>%
  unnest_tokens(word, word) %>%
  count(word, sort = TRUE) %>%
  top_n(20)
```

- unnest_tokens(word, word):

متن توییت‌ها به کلمات جداگانه شکسته می‌شود.

- count(word, sort = TRUE):

تعداد تکرار هر کلمه در توییت‌ها محاسبه می‌شود و بر اساس تعداد مرتب می‌شود.

- top_n(20): 20

کلمه با بیشترین تعداد تکرار انتخاب می‌شود .

2. رسم نمودار کلمات پرتکرار:

کد مرتبط :

```
# Visualizing the top 20 unique words in a plot
ggplot(top_words, aes(x = reorder(word, n), y = n, fill = word)) +
  geom_col() +
  xlab("Word") +
  ylab("Frequency") +
  ggtitle("Top 20 Unique Words in Tweets") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  guides(fill=FALSE)
```

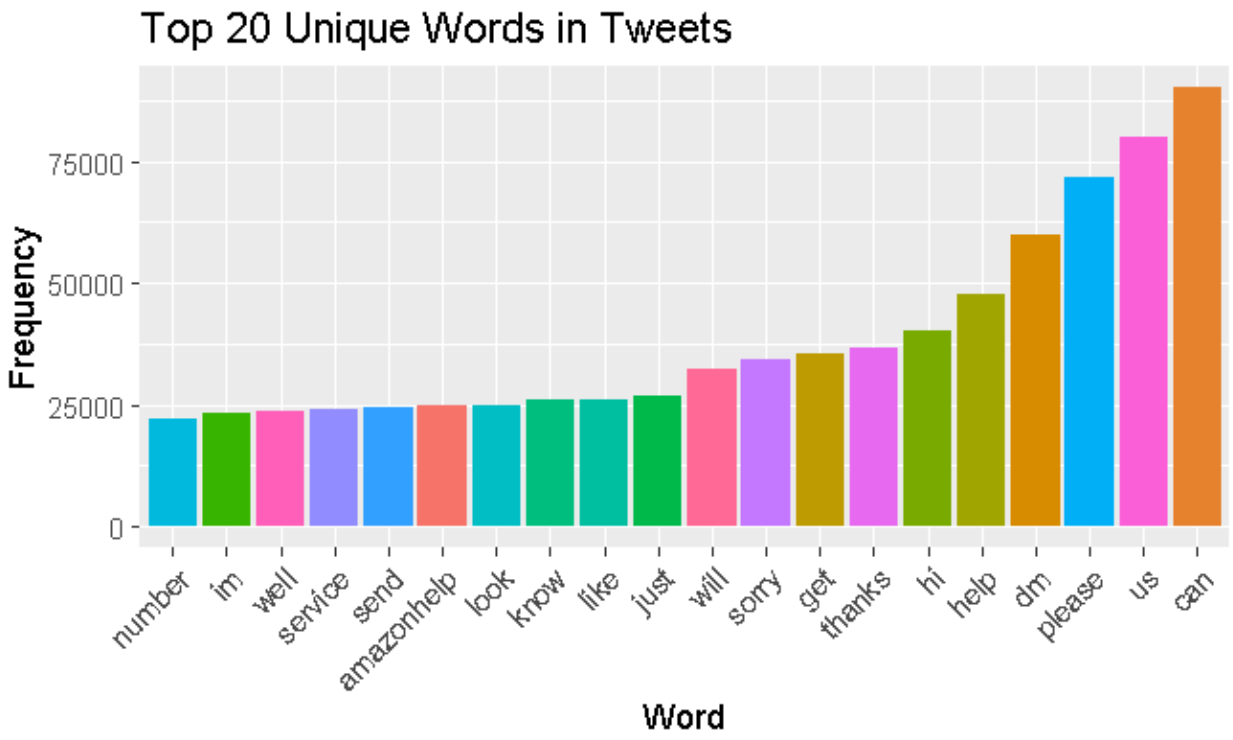
- ggplot(): یک نمودار با استفاده از داده‌های top_words ایجاد می‌شود.
- aes(x = reorder(word, n), y = n, fill = word):
  - محور x: کلمات پرتکرار، بر اساس فراوانی مرتب‌شده.
  - محور y: تعداد دفعات تکرار هر کلمه.
  - fill: هر کلمه با رنگ متفاوت نمایش داده می‌شود.
- geom_col(): نمودار ستونی (Bar Plot) رسم می‌کند.
- xlab و ylab: برچسب‌های محورهای x و y مشخص می‌شوند.
- ggtitle(): عنوان نمودار تنظیم می‌شود.
- theme(axis.text.x = element_text(angle = 45, hjust = 1)):

- متن محور x به صورت زاویه‌دار (45 درجه) نمایش داده می‌شود تا خوانایی بهتری داشته باشد.

- guides(fill=FALSE):

رنگ‌های نمودار در قسمت راهنما نمایش داده نمی‌شوند.

نمودار مربوط :



(قسمت b)

حذف نویسه‌های غیر استاندارد (غیر ASCII) از داده‌ها :

کد مرتبط :

```
# 6.b

# Filtering out all non-ASCII rows from the data frame
non_ASCII_df <- words_in_tweets_df %>%
  filter(!Encoding(word) == "known")
```

این کد کلماتی را که از کاراکترهای غیر استاندارد (غیر ASCII) تشکیل شده‌اند، از داده‌ها حذف می‌کند. هدف از این کار، پاک‌سازی داده‌ها از نویسه‌هایی است که ممکن است باعث مشکلاتی در تحلیل شوند یا در متن‌های استاندارد وجود نداشته باشند.

1. فیلتر کردن کلمات:

- تابع (Encoding(word نوع رمزگذاری (Encoding) هر کلمه را بررسی می‌کند.
- اگر نوع رمزگذاری کلمه "known" نباشد (یعنی از استاندارد ASCII نباشد)، آن کلمه حذف می‌شود.
- ASCII استانداردی برای کاراکترهای متنی (مانند حروف انگلیسی و اعداد) است. کاراکترهای غیر ASCII شامل ایموجی‌ها، حروف زبان‌های دیگر (مثل فارسی، چینی و ...) یا کاراکترهای ناشناخته هستند.

2. ذخیره در دیتا فریم جدید:

نتیجه فیلتر کردن در دیتا فریم جدیدی به نام non_ASCII_df ذخیره می‌شود. این دیتافریم فقط شامل کلماتی است که از کاراکترهای استاندارد ASCII تشکیل شده‌اند.

(قسمت c)

کد مرتبط :

```
# 6.c

top_words2 <- non_ASCII_df %>%
  unnest_tokens(word, word) %>%
  count(word, sort = TRUE) %>%
  top_n(20)

ggplot(top_words2, aes(x = reorder(word, n), y = n, fill = word)) +
  geom_col() +
  xlab("Word") +
  ylab("Frequency") +
  ggtitle("Top 20 Words after filtering out all non-ASCII in Tweets") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  guides(fill=FALSE)
```

این کد 20 کلمه پرتکرار را پس از حذف کلمات حاوی نویسه‌های غیر ASCII شناسایی کرده و در قالب یک نمودار بصری نمایش می‌دهد. هدف این مرحله، تحلیل کلمات پرتکرار در داده‌های تمیز شده است که فقط شامل کاراکترهای استاندارد ASCII هستند.

1. شناسایی 20 کلمه پرتکرار:

```
top_words2 <- non_ASCII_df %>%
  unnest_tokens(word, word) %>%
  count(word, sort = TRUE) %>%
  top_n(20)
```

- unnest_tokens(word, word):

کلمات موجود در ستون word را به توکن‌های جداگانه تقسیم می‌کند.

- count(word, sort = TRUE):

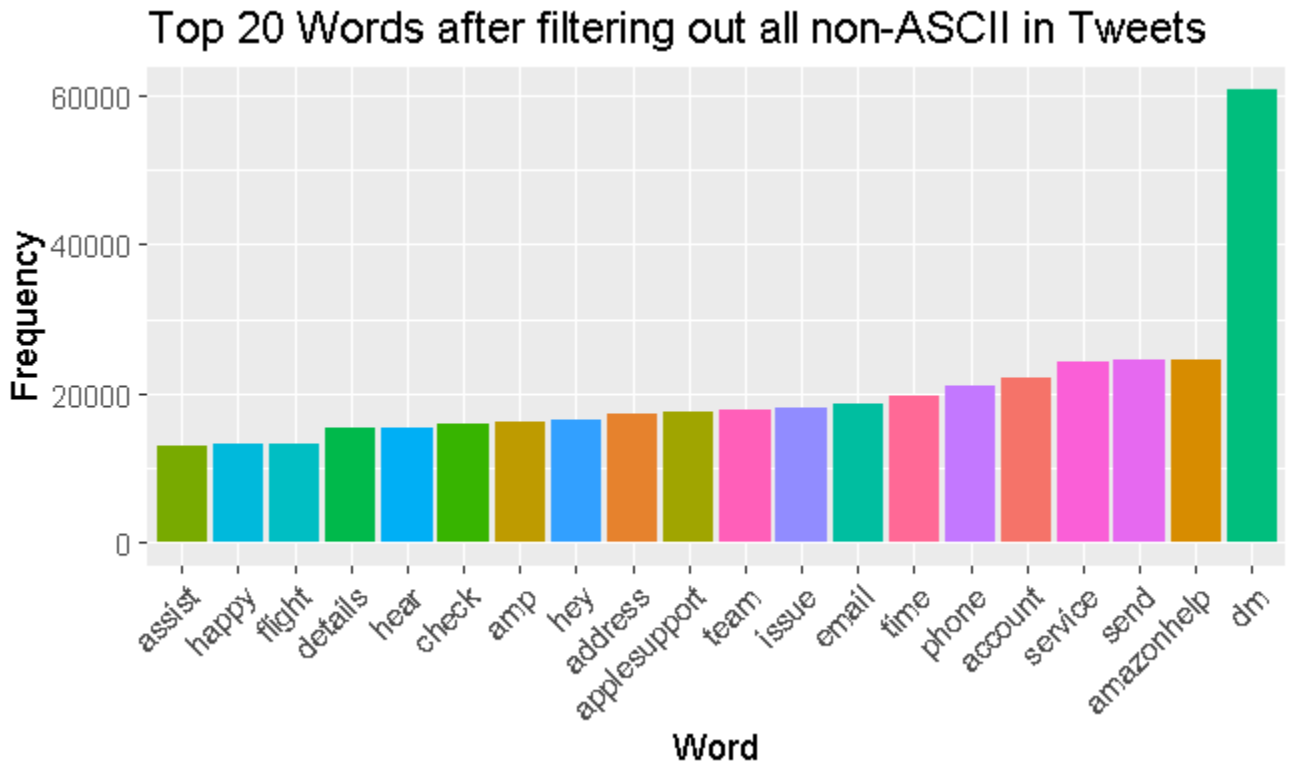
تعداد تکرار هر کلمه را محاسبه کرده و کلمات را بر اساس تعداد تکرار مرتب می‌کند.

- top_n(20): 20

کلمه با بیشترین تعداد تکرار انتخاب می‌شود

2. رسم نمودار کلمات پرتکرار:

مانند بخش قبلی :



مرحله هفتم : تحلیل و بصری‌سازی کلمات پرتکرار در توییت‌های ورودی

(قسمت a)

کد مرتبط :

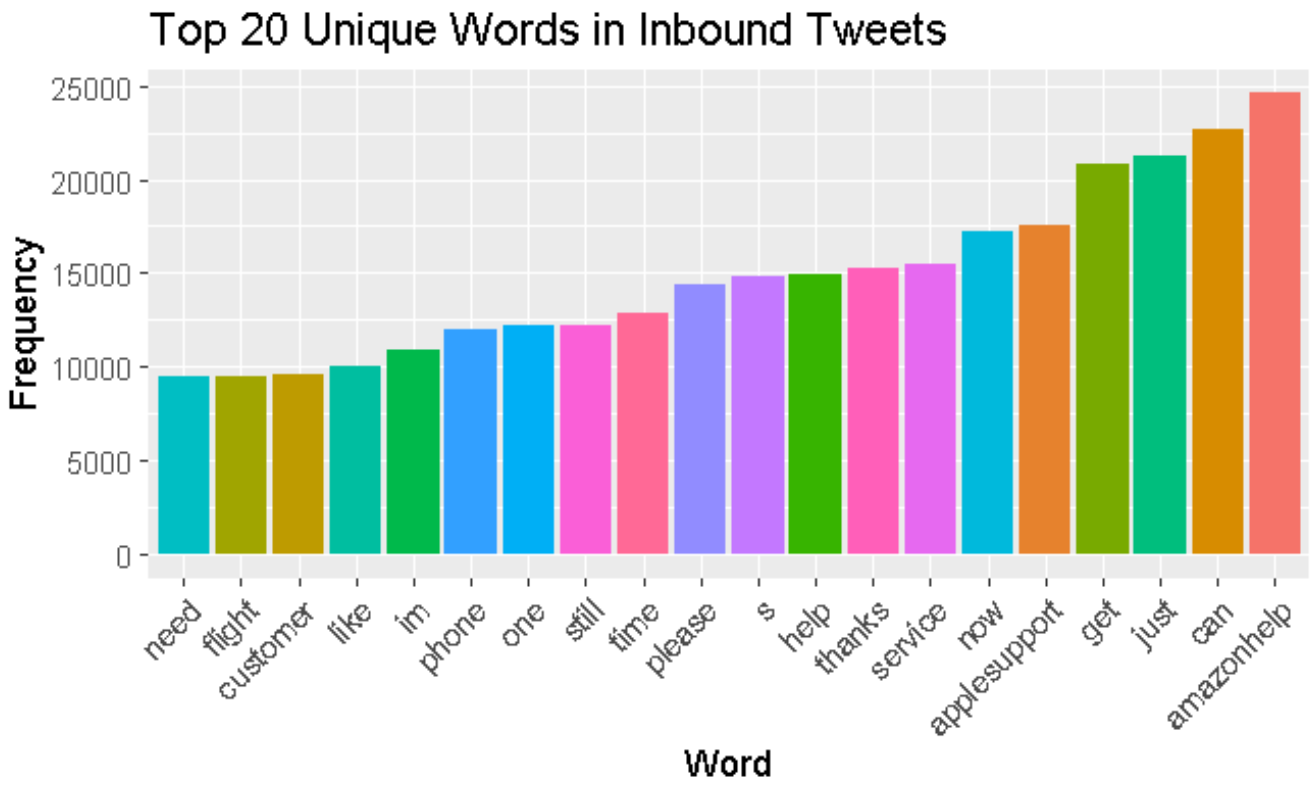
```
# step7
# 7.a

# Filter the inbound tweets
inbound_tweets <- clean_tweets_df %>%
  filter(inbound == "True")

# Create a Corpus
tweet_corpus <- Corpus(VectorSource(inbound_tweets$text))

# Preprocessing: Convert to lower case, remove punctuation, numbers, and stop words
tweet_corpus <- tm_map(tweet_corpus, content_transformer(tolower))
tweet_corpus <- tm_map(tweet_corpus, removePunctuation)
tweet_corpus <- tm_map(tweet_corpus, removeNumbers)
tweet_corpus <- tm_map(tweet_corpus, removeWords, stopwords("english"))

# Convert the Corpus back to a character vector
cleaned_tweets <- sapply(tweet_corpus, as.character)

# Replace the tweets in the dataset with the cleaned tweets
inbound_tweets$text <- cleaned_tweets
```

```r
# Find the top 20 unique words for inbound tweets
top_inbound_words <- inbound_tweets %>%
  unnest_tokens(word, text) %>%
  count(word, sort = TRUE) %>%
  top_n(20)

# Visualize the top 20 unique words for inbound tweets in a plot
ggplot(top_inbound_words, aes(x = reorder(word, n), y = n, fill = word)) +
  geom_col() +
  xlab("Word") +
  ylab("Frequency") +
  ggtitle("Top 20 Unique Words in Inbound Tweets") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  guides(fill=FALSE)
```

این کد ابتدا توییت‌های ورودی (inbound tweets) را از داده‌ها جدا می‌کند و سپس متن آن‌ها را تمیزسازی می‌کند. فرآیند تمیزسازی شامل حذف علائم نگارشی، اعداد، کلمات بی‌معنی، و تبدیل متن به حروف کوچک است. بعد از پاک‌سازی، کلمات موجود در این توییت‌ها به واحدهای جداگانه تقسیم می‌شوند و پرتکرارترین کلمات شناسایی می‌شوند. در پایان، 20 کلمه پرتکرار در قالب یک نمودار ستونی نمایش داده می‌شوند تا بتوانیم به‌طور بصری موضوعات اصلی و پرتکرار در توییت‌های ورودی را درک کنیم.

این مرحله برای شناسایی کلمات کلیدی در پیام‌های دریافتی کاربران طراحی شده و به تحلیل دیدگاه‌ها و موضوعات مورد توجه آن‌ها کمک می‌کند. اگر تصویری از نمودار وجود داشته باشد، می‌توان بینش بیشتری در مورد محتوای توییت‌های ورودی ارائه داد.

نمودار مرتبط :



تفسیر نمودار:

نمودار نشان‌دهنده 20 کلمه پرتکرار در توییت‌های ورودی است که پس از انجام فرآیند تمیزسازی استخراج شده‌اند. در محور افقی (Word)، کلمات پرتکرار نمایش داده شده‌اند و در محور عمودی (Frequency) تعداد تکرار هر کلمه مشخص است.

بینش‌های کلیدی:

1. کلمات مرتبط با درخواست کمک: کلماتی مانند "support"، "please"، "help"، و "service" نشان می‌دهند که کاربران بیشتر به دنبال دریافت خدمات یا کمک هستند.
2. کلمات مرتبط با موضوعات خاص: کلمات خاصی مانند "amazonhelp" یا "applesupport" نشان می‌دهند که بخشی از توییت‌ها مربوط به درخواست‌های پشتیبانی از شرکت‌های خاص هستند.
3. کلمات رایج در توییت‌ها: کلماتی مانند "can"، "need"، و "get" نشان‌دهنده درخواست‌ها و نیازهای کاربران است.
4. حجم بالای ارتباطات مشتری: کلمات "customer" و "service" به این موضوع اشاره می‌کنند که بسیاری از توییت‌های ورودی به خدمات مشتریان اختصاص دارند.

(قسمت b)

تحلیل و بصری‌سازی کلمات پرتکرار در توییت‌های خروجی شرکت

کد مرتبط :

```
# 7.b

# Filter the outbound tweets
outbound_tweets <- clean_tweets_df %>%
  filter(inbound == "False")

tweet_corpus <- Corpus(VectorSource(outbound_tweets$text))

# Preprocessing: Convert to lower case, remove punctuation, numbers, and stop words
tweet_corpus <- tm_map(tweet_corpus, content_transformer(tolower))
tweet_corpus <- tm_map(tweet_corpus, removePunctuation)
tweet_corpus <- tm_map(tweet_corpus, removeNumbers)
tweet_corpus <- tm_map(tweet_corpus, removeWords, stopwords("english"))

# Convert the Corpus back to a character vector
cleaned_tweets <- sapply(tweet_corpus, as.character)

# Replace the tweets in the dataset with the cleaned tweets
outbound_tweets$text <- cleaned_tweets
```

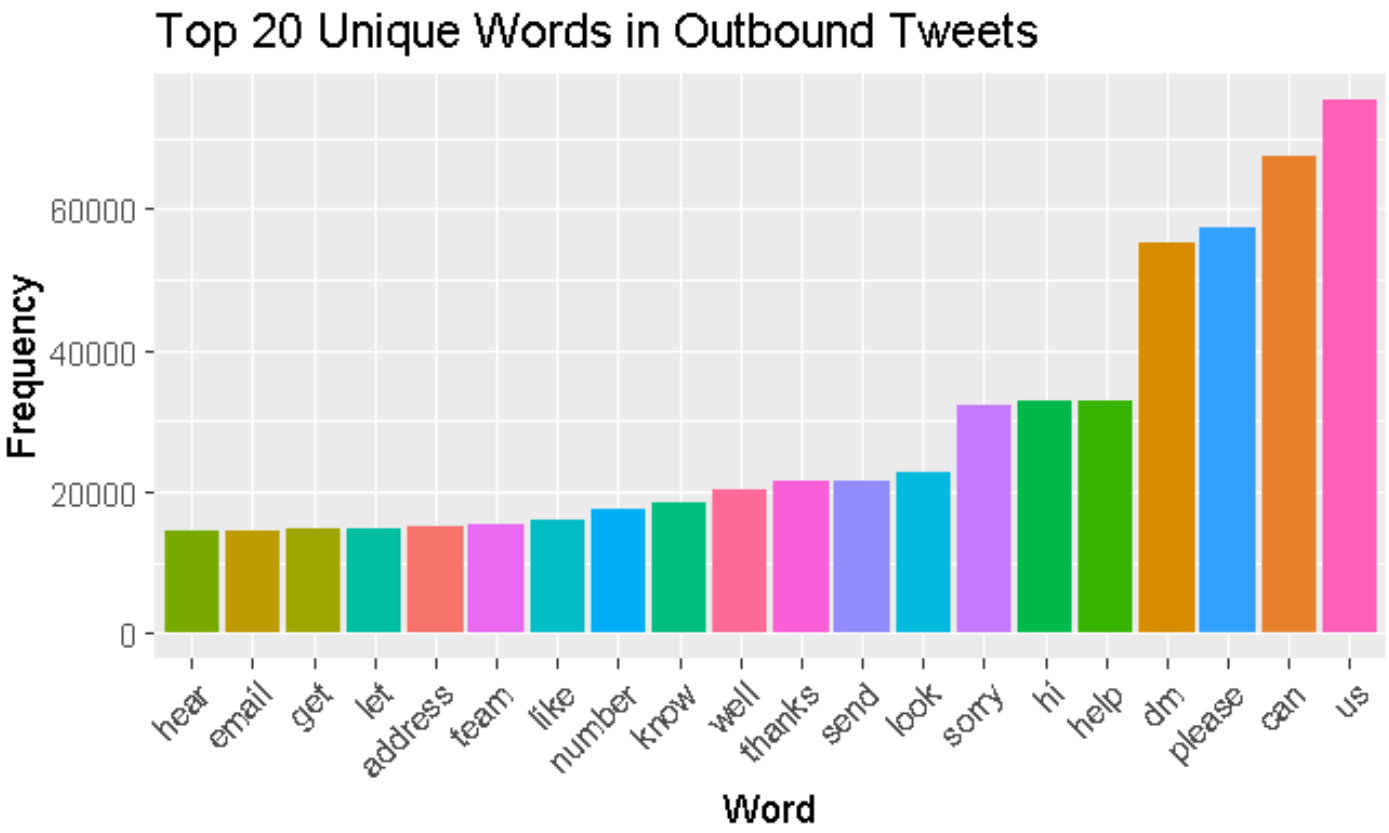
```r
# Find the top 20 unique words for outbound tweets
top_outbound_words <- outbound_tweets %>%
  unnest_tokens(word, text) %>%
  count(word, sort = TRUE) %>%
  top_n(20)


# Visualize the top 20 unique words for outbound tweets in a plot
ggplot(top_outbound_words, aes(x = reorder(word, n), y = n, fill = word)) +
  geom_col() +
  xlab("Word") +
  ylab("Frequency") +
  ggtitle("Top 20 Unique Words in Outbound Tweets") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  guides(fill = FALSE)
```

این کد برای تحلیل و نمایش کلمات پرتکرار در توییت‌های خروجی (Outbound Tweets) استفاده می‌شود. توییت‌های خروجی، پیام‌هایی هستند که شرکت یا سازمان به کاربران ارسال کرده است. مراحل این کد شامل موارد زیر است:

1. فیلتر کردن توییت‌های خروجی:
   - توییت‌هایی که مقدار ستون inbound آن‌ها برابر "False" است، انتخاب می‌شوند.
2. پاک‌سازی متن توییت‌ها:
   - متن توییت‌ها با تبدیل به حروف کوچک، حذف اعداد، علائم نگارشی و کلمات بی‌معنی تمیز می‌شود.
3. استخراج کلمات پرتکرار:
   - کلمات در توییت‌ها به واحدهای جداگانه شکسته می‌شوند، تعداد تکرار آن‌ها محاسبه شده و 20 کلمه پرتکرار شناسایی می‌شوند.
4. نمایش بصری کلمات پرتکرار:

- یک نمودار ستونی ایجاد می‌شود که نشان می‌دهد کدام کلمات در پیام‌های خروجی شرکت بیشتر استفاده شده‌اند.

نمودار مرتبط :



تفسیر نمودار :

این نمودار نشان‌دهنده 20 کلمه پرتکرار در توییت‌های خروجی شرکت است. در محور افقی (Word) کلمات پرتکرار و در محور عمودی (Frequency) تعداد دفعات تکرار آن‌ها نشان داده شده است.

بینش‌های کلیدی:

1. کلمات نشان‌دهنده تعامل مستقیم:
   - کلماتی مانند "help"، "please"، "can"، "us"، و "hi" نشان می‌دهند که شرکت به‌طور مستقیم با کاربران ارتباط برقرار می‌کند و رویکرد مودبانه‌ای دارد.
   - استفاده از DM(" Direct Message") نشان می‌دهد که شرکت کاربران را برای حل مشکلات به پیام‌های خصوصی هدایت می‌کند.
2. تمرکز بر اطلاعات و پیگیری:
   - کلماتی مانند "number"، "address"، "email"، و "send" نشان می‌دهند که شرکت درخواست اطلاعات بیشتری از کاربران دارد تا مسائل آن‌ها را حل کند.
3. تشکر و عذرخواهی:
   - کلماتی مثل "thanks" و "sorry" نشان‌دهنده رفتار مثبت و تلاش شرکت برای ایجاد ارتباط خوب با کاربران است.
4. حضور تیم پشتیبانی:
   - کلمه "team" نشان می‌دهد که شرکت به تیم پشتیبانی اشاره می‌کند، احتمالاً برای ایجاد اطمینان در کاربران.

تأثیر برای تحلیل:

این نمودار نشان می‌دهد که شرکت در توییت‌های خروجی خود:

- روی کمک به کاربران و رفع مشکلات تمرکز دارد.
- سعی می‌کند ارتباطی مودبانه و حرفه‌ای برقرار کند.
- از کاربران اطلاعات بیشتری برای حل مسائل درخواست می‌کند.

مرحله هشتم : تحلیل احساسات توییت‌ها با استفاده از فرهنگ لغت AFINN

(قسمت a)

کد مرتبط :

```
# step8
# 8.a

get_sentiments("afinn")

afinn_word_counts <- words_in_tweets_df %>%
  inner_join(get_sentiments("afinn")) %>%
  mutate(sentiment = if_else(value>0, 'positive', 'negative')) %>%
  count(word, value, sentiment, sort = T) %>%
  ungroup()

afinn_word_counts %>%
  group_by(sentiment) %>%
  top_n(15) %>%
  ungroup() %>%
  mutate(word = reorder(word,n)) %>%
  ggplot(aes(word, n, fill=sentiment))+
  geom_col(show.legend = F)+
  facet_wrap(~sentiment, scales = "free_y")+
  labs(y="sentiment", x= "Words",
       title = "Afinn sentiment analysis on tweets")+
  coord_flip()
```

این کد از فرهنگ لغت AFINN برای تحلیل احساسات توییت‌ها استفاده می‌کند. کلمات موجود در توییت‌ها با امتیازهای احساسی AFINN (مثبت یا منفی) مقایسه می‌شوند. پس از این، کلمات بر اساس امتیاز مثبت یا منفی دسته‌بندی شده و پرتکرارترین کلمات در هر دسته شناسایی و در یک نمودار نمایش داده می‌شوند.

مراحل کد :

بارگذاری فرهنگ لغت AFINN:

- کلمات موجود در فرهنگ لغت AFINN هرکدام امتیاز احساسی مثبت یا منفی دارند که برای تحلیل احساسات استفاده می‌شوند.

تطبیق کلمات توییت‌ها با AFINN:

- کلمات توییت‌ها با کلمات فرهنگ لغت AFINN مقایسه می‌شوند و فقط کلمات موجود در این فرهنگ لغت باقی می‌مانند.

دسته‌بندی احساسات:

- کلمات با امتیاز مثبت به دسته "احساس مثبت" و کلمات با امتیاز منفی به دسته "احساس منفی" تخصیص داده می‌شوند.

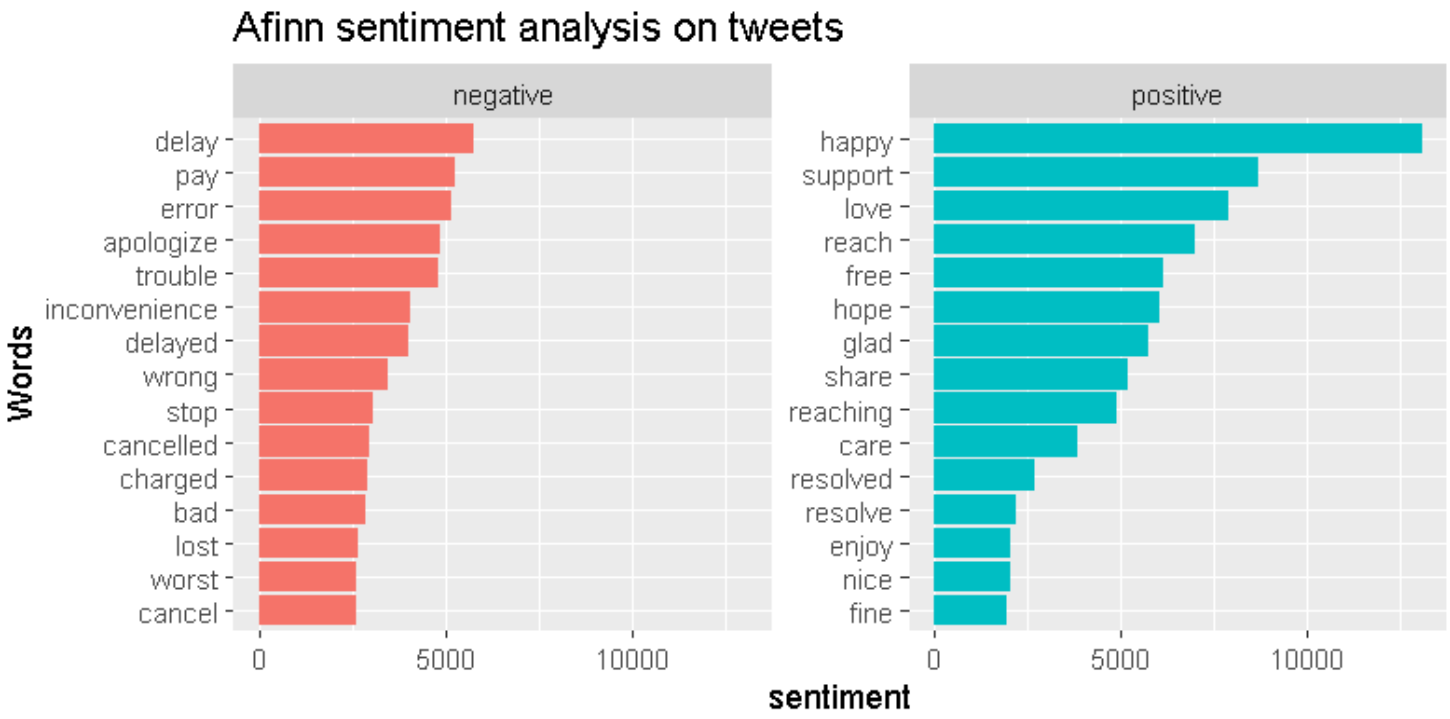
شناسایی کلمات پرتکرار:

- تعداد تکرار کلمات مثبت و منفی محاسبه می‌شود و پرتکرارترین کلمات هر دسته مشخص می‌شوند.

بصری‌سازی احساسات:

- یک نمودار ستونی برای نمایش کلمات پرتکرار مثبت و منفی رسم می‌شود. این نمودار به دو بخش تقسیم می‌شود: یکی برای احساسات مثبت و دیگری برای احساسات منفی، و کلمات به ترتیب فراوانی مرتب می‌شوند.

نمودار مرتبط :



تفسیر نمودار:

این نمودار نتایج تحلیل احساسات توییت‌ها را با استفاده از فرهنگ لغت AFINN نشان می‌دهد. نمودار به دو بخش تقسیم شده است:

1. احساسات منفی (Negative):
   - کلمات پرتکرار در این دسته شامل "error"، "cancel"، "delay"، و "apologize" هستند.

- این کلمات نشان می‌دهند که کاربران اغلب در مورد مشکلاتی مانند تأخیر، لغو، اشتباهات یا ناراحتی‌های دیگر توییت می‌کنند.

2. احساسات مثبت (Positive):
   - کلمات پرتکرار در این دسته شامل "love"، "support"، "happy" و "resolved" هستند.
   - این کلمات نشان‌دهنده رضایت کاربران یا قدردانی آن‌ها از خدمات شرکت است، مانند تجربه‌های خوب یا حل موفقیت‌آمیز مشکلات.

بینش‌های کلیدی:

1. احساسات منفی:
   - کاربران معمولاً زمانی توییت‌های منفی ارسال می‌کنند که با مشکلاتی مانند تاخیر یا لغو خدمات مواجه می‌شوند.
   - کلماتی مانند "apologize" و "error" نشان می‌دهند که شرکت‌ها به طور فعال به مشکلات پاسخ می‌دهند.
2. احساسات مثبت:
   - احساسات مثبت اغلب حول موضوعاتی مانند حمایت ("support")، حل مشکلات ("resolved")، و رضایت کاربران ("happy") شکل می‌گیرند.
   - این نشان می‌دهد که شرکت‌ها در ایجاد تجربیات مثبت برای مشتریان موفق عمل می‌کنند.

اهمیت تحلیل:

- این نمودار به شرکت‌ها کمک می‌کند تا نقاط قوت و ضعف خود را شناسایی کنند:
  - نقاط ضعف: مشکلاتی که باعث احساسات منفی می‌شوند.

- نقاط قوت: جنبه‌هایی که باعث ایجاد رضایت و احساسات مثبت در مشتریان می‌شوند.

(قسمت b)

تحلیل و نمایش کلمات مثبت و منفی پرتکرار در توییت‌های ورودی با AFINN

کد مرتبط :

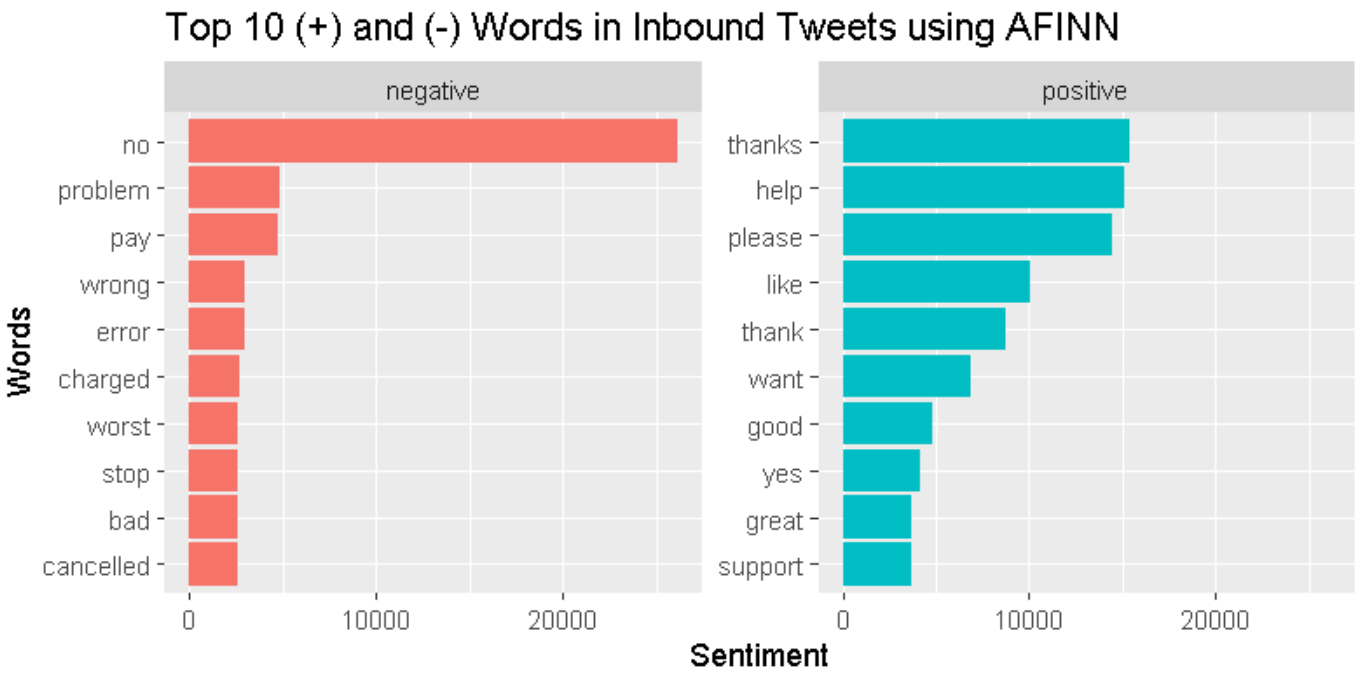
```
# 8.b

# Filter the inbound tweets
inbound_tweets2 <- clean_tweets_df %>%
  filter(inbound == "True")

# Perform AFINN sentiment analysis on inbound tweets and find the top 10 positive and negative words
afinn_word_counts_inbound <- inbound_tweets2 %>%
  unnest_tokens(word, text) %>%
  inner_join(get_sentiments("afinn")) %>%
  mutate(sentiment = if_else(value > 0, 'positive', 'negative')) %>%
  count(word, value, sentiment, sort = TRUE) %>%
  ungroup()

  # Visualize and save the plot for the top 10 positive and negative words in inbound tweets
afinn_word_counts_inbound %>%
  group_by(sentiment) %>%
  top_n(10) %>%
  ungroup() %>%
  mutate(word = reorder(word, n)) %>%
  ggplot(aes(word, n, fill = sentiment)) +
  geom_col(show.legend = FALSE) +
  facet_wrap(~sentiment, scales = "free_y") +
  labs(y = "Sentiment", x = "Words", title = "Top 10 (+) and (-) Words in Inbound Tweets using AFINN ") +
  coord_flip()
```

این کد بر تحلیل احساسات توییت‌های ورودی (inbound tweets) تمرکز دارد. کلمات توییت‌ها با استفاده از فرهنگ لغت AFINN تحلیل شده و به دو دسته کلمات مثبت و منفی تقسیم می‌شوند. در نهایت، 10 کلمه پرتکرار مثبت و 10 کلمه پرتکرار منفی شناسایی شده و در یک نمودار نمایش داده می‌شوند.

گام‌های کلیدی:

1. فیلتر توییت‌های ورودی:
   - فقط توییت‌هایی که توسط کاربران به شرکت ارسال شده‌اند، جدا می‌شوند.
2. تحلیل احساسات با AFINN:
   - کلمات موجود در متن توییت‌ها با کلمات فرهنگ لغت AFINN تطبیق داده می‌شوند.
   - بر اساس امتیاز مثبت یا منفی AFINN، کلمات به دو دسته مثبت و منفی تقسیم می‌شوند.
3. شناسایی کلمات پرتکرار:
   - تعداد تکرار کلمات مثبت و منفی در توییت‌ها محاسبه می‌شود.
   - پرتکرارترین کلمات در هر دسته (10 کلمه برای مثبت و 10 کلمه برای منفی) انتخاب می‌شوند.
4. بصری‌سازی:
   - یک نمودار ستونی برای نمایش کلمات مثبت و منفی رسم می‌شود.
   - نمودار به دو بخش تقسیم می‌شود: کلمات مثبت و کلمات منفی.
   - کلمات بر اساس تعداد تکرار مرتب شده و به صورت چرخشی نمایش داده می‌شوند.

نمودار مرتبط :



تفسیر نمودار:

این نمودار نشان‌دهنده 10 کلمه پرتکرار مثبت و 10 کلمه پرتکرار منفی در توییت‌های ورودی (Inbound Tweets) است که با استفاده از فرهنگ لغت AFINN تحلیل شده‌اند.

بینش‌های کلیدی:

1. احساسات منفی:

- پرتکرارترین کلمه منفی "no" است، که احتمالاً نشان‌دهنده نارضایتی کاربران یا رد درخواست‌های آن‌ها است.
- کلماتی مانند "wrong"، "pay"، "problem"، و "error" مشکلات یا دغدغه‌های کاربران را نشان می‌دهند.

- کلمات دیگری مانند "charged" و "cancelled" به مسائل مالی یا لغو خدمات اشاره دارند.

2. احساسات مثبت:

- پرتکرارترین کلمه مثبت "thanks" است، که نشان‌دهنده قدردانی کاربران است.
- کلمات "help" و "please" نشان می‌دهند که کاربران اغلب به دنبال دریافت کمک هستند و در درخواست‌های خود مودبانه عمل می‌کنند.
- کلماتی مانند "great" ،"good"، و "support" نشان‌دهنده رضایت کاربران از خدمات یا پاسخگویی هستند.

اهمیت برای تحلیل:

1. برای احساسات منفی:
   - کلمات منفی به شرکت کمک می‌کنند تا مشکلات اصلی کاربران (مانند مسائل مالی یا فنی) را شناسایی کرده و برای رفع آن‌ها اقدام کند.
2. برای احساسات مثبت:
   - کلمات مثبت به شرکت نشان می‌دهند که کاربران از چه جنبه‌هایی از خدمات یا پاسخگویی رضایت دارند و این نقاط قوت باید تقویت شوند.

(قسمت c)

تحلیل و نمایش کلمات مثبت و منفی پرتکرار در توییت‌های خروجی شرکت با AFINN

کد مرتبط :

```r
# 8.c

# Filter the outbound tweets
outbound_tweets2 <- clean_tweets_df %>%
  filter(inbound == "False")

# Perform AFINN sentiment analysis on outbound tweets and find the top 10 positive and negative words
afinn_word_counts_outbound <- outbound_tweets2 %>%
  unnest_tokens(word, text) %>%
  inner_join(get_sentiments("afinn")) %>%
  mutate(sentiment = if_else(value > 0, 'positive', 'negative')) %>%
  count(word, value, sentiment, sort = TRUE) %>%
  ungroup()

# Visualize and save the plot for the top 10 positive and negative words in outbound tweets as
afinn_word_counts_outbound %>%
  group_by(sentiment) %>%
  top_n(10) %>%
  ungroup %>%
  mutate(word = reorder(word, n)) %>%
  ggplot(aes(word, n, fill = sentiment)) +
  geom_col(show.legend = FALSE) +
  facet_wrap(~sentiment, scales = "free_y") +
  labs(y = "Sentiment", x = "Words", title = "Top 10 (+) and (-) Words in Outbound Tweets using AFINN") +
  coord_flip()
```

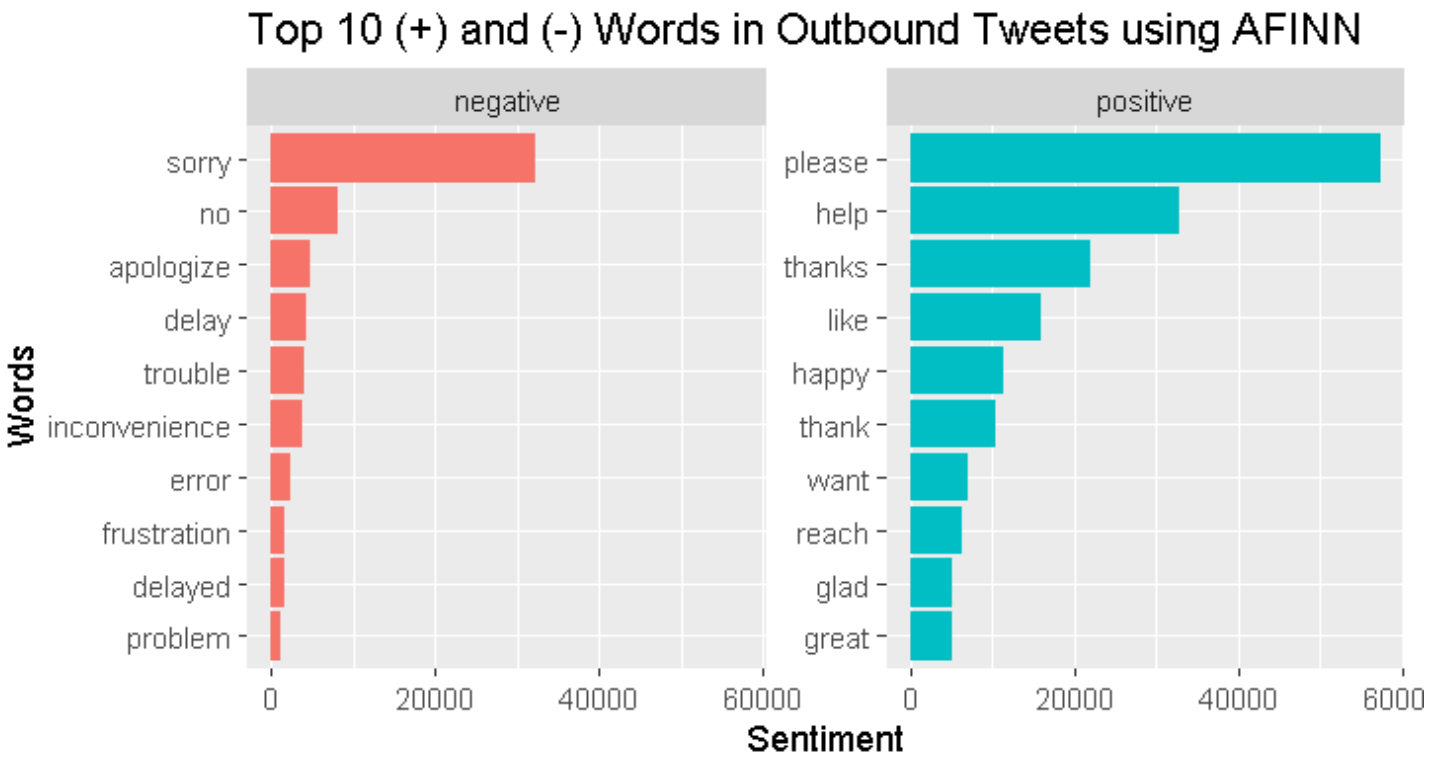
این کد به تحلیل احساسات توییت‌های خروجی (Outbound Tweets) می‌پردازد. کلمات موجود در این توییت‌ها با استفاده از فرهنگ لغت AFINN تحلیل شده و به دو دسته کلمات مثبت و کلمات منفی تقسیم‌بندی می‌شوند. سپس پرتکرارترین کلمات مثبت و منفی شناسایی و در یک نمودار نمایش داده می‌شوند.

گام‌های کلیدی:

1. فیلتر توییت‌های خروجی:
   - فقط توییت‌هایی که توسط شرکت ارسال شده‌اند جدا می‌شوند.
2. تحلیل احساسات با استفاده از AFINN:
   - کلمات توییت‌ها با کلمات فرهنگ لغت AFINN مقایسه می‌شوند.

- امتیاز مثبت یا منفی هر کلمه مشخص شده و به دسته‌بندی "مثبت" یا "منفی" اضافه می‌شود.

3. شناسایی پرتکرارترین کلمات:
   - تعداد تکرار هر کلمه در دسته‌های مثبت و منفی محاسبه می‌شود.
   - 10 کلمه پرتکرار برای هر دسته انتخاب می‌شوند.
4. بصری‌سازی نتایج:
   - یک نمودار ستونی برای نمایش پرتکرارترین کلمات مثبت و منفی رسم می‌شود.
   - نمودار به دو بخش تقسیم می‌شود: یکی برای کلمات مثبت و دیگری برای کلمات منفی.
   - کلمات بر اساس تعداد تکرار مرتب و به صورت چرخشی نمایش داده می‌شوند.



تفسیر نمودار:

این نمودار نشان‌دهنده 10 کلمه پرتکرار مثبت و 10 کلمه پرتکرار منفی در توییت‌های خروجی شرکت است که با استفاده از فرهنگ لغت AFINN تحلیل شده‌اند. توییت‌های خروجی پیام‌هایی هستند که شرکت به کاربران ارسال کرده است.

بینش‌های کلیدی:

1. احساسات منفی:

- پرتکرارترین کلمه "sorry" است، که نشان‌دهنده تلاش شرکت برای عذرخواهی از کاربران در موارد مشکل‌ساز است.
- کلماتی مانند "error"، "inconvenience"، "apologize"، و "problem" نشان می‌دهند که بخش عمده‌ای از تعاملات خروجی شرکت مربوط به رسیدگی به مشکلات کاربران است.
- این کلمات نشان‌دهنده رویکرد فعال شرکت در پاسخ به شکایات و رفع مشکلات است.

2. احساسات مثبت:

- پرتکرارترین کلمه "please" است، که نشان‌دهنده استفاده مودبانه شرکت در درخواست اطلاعات یا ارائه راهنمایی است.
- کلماتی مانند "thanks"، "help"، و "happy" نشان می‌دهند که شرکت در پاسخ‌های خود سعی دارد احساس مثبتی را به کاربران منتقل کند.
- استفاده از کلماتی مثل "great" و "glad" نشان‌دهنده تلاش شرکت برای تأکید بر جنبه‌های مثبت خدمات و تعاملات است.

اهمیت برای تحلیل:

1. برای احساسات منفی:
   - شرکت با استفاده از عبارات عذرخواهی و پذیرش مشکلات، سعی در جلب رضایت کاربران دارد.
2. برای احساسات مثبت:
   - استفاده از زبان مثبت و مودبانه نشان‌دهنده رویکرد حرفه‌ای شرکت در ارتباطات خروجی است.

نتیجه‌گیری:

این نمودار تأکید می‌کند که شرکت:

- در مواجهه با مشکلات، به طور فعال عذرخواهی کرده و پاسخگو است.
- در پیام‌های خود از زبان مثبت و احترام‌آمیز برای تقویت تعاملات و رفع نگرانی‌های کاربران استفاده می‌کند.

مرحله نهم : تحلیل احساسات توییت‌ها با استفاده از فرهنگ لغت Bing

(قسمت a)

کد مرتبط :

```r
# step9
# 9.a

get_sentiments("bing")


bing_word_counts <- words_in_tweets_df %>%
  inner_join(get_sentiments("bing")) %>%
  count(word, sentiment, sort = T) %>%
  ungroup()

bing_word_counts %>%
  group_by(sentiment) %>%
  top_n(15) %>%
  ungroup() %>%
  mutate(word = reorder(word,n)) %>%
  ggplot(aes(word, n, fill=sentiment))+
  geom_col(show.legend = F)+
  facet_wrap(~sentiment, scales = "free_y")+
  labs(y="sentiment", x= "Words",
       title = "Bing sentiment analysis on Tweets")+
  coord_flip()
```

این کد از فرهنگ لغت Bing برای تحلیل احساسات توییت‌ها استفاده می‌کند. کلمات توییت‌ها با لیست کلمات مثبت و منفی موجود در فرهنگ لغت Bing مقایسه شده و تعداد تکرار آن‌ها بر اساس نوع احساس (مثبت یا منفی) محاسبه می‌شود. در نهایت، پرتکرارترین کلمات مثبت و منفی در هر دسته شناسایی شده و در قالب یک نمودار بصری نمایش داده می‌شوند.

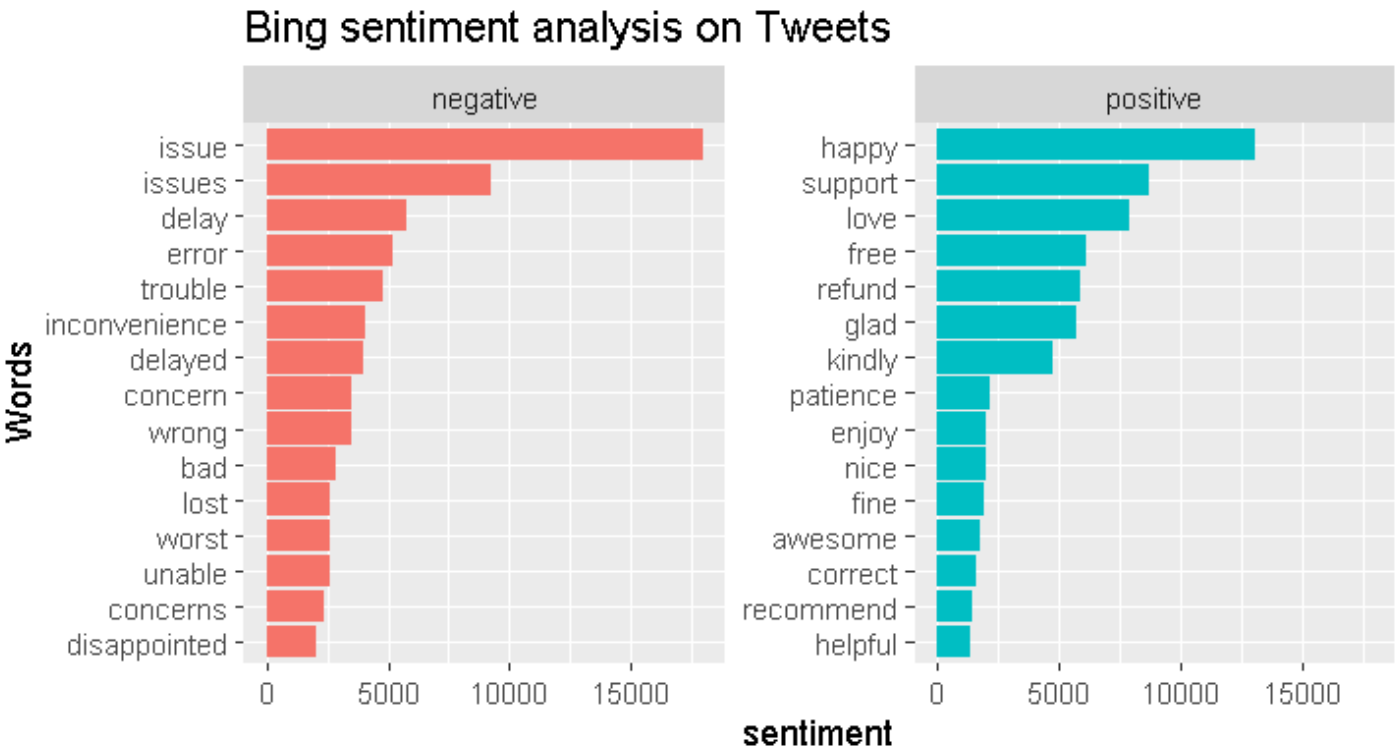
گام‌های کلیدی:

1. بارگذاری فرهنگ لغت Bing:

- این فرهنگ لغت کلمات را به دو دسته مثبت و منفی طبقه‌بندی می‌کند.

2. تطبیق کلمات توییت‌ها با Bing:
   - کلمات موجود در متن توییت‌ها با کلمات مثبت و منفی فرهنگ لغت Bing تطبیق داده می‌شوند.
   - تنها کلماتی که در Bing تعریف شده‌اند (مثبت یا منفی) برای تحلیل باقی می‌مانند.
3. محاسبه تعداد تکرار کلمات:
   - تعداد تکرار هر کلمه در هر دسته (مثبت و منفی) محاسبه شده و نتایج به ترتیب تعداد مرتب می‌شوند.
4. شناسایی پرتکرارترین کلمات:
   - 15 کلمه پرتکرار از هر دسته انتخاب می‌شوند.
5. بصری‌سازی نتایج:
   - یک نمودار ستونی برای نمایش کلمات مثبت و منفی پرتکرار رسم می‌شود.
   - کلمات مثبت و منفی در دو بخش جداگانه نمایش داده می‌شوند.
   - نمودار برای مقایسه بهتر به صورت افقی چرخانده می‌شود.

نمودار مرتبط :



تفسیر نمودار:

این نمودار نتایج تحلیل احساسات توییت‌ها با استفاده از فرهنگ لغت Bing را نمایش می‌دهد. کلمات به دو دسته منفی و مثبت تقسیم شده‌اند و 15 کلمه پرتکرار هر دسته نشان داده شده‌اند.

بینش‌های کلیدی:

1. احساسات منفی:

- پرتکرارترین کلمه منفی "issue" و "issues" است، که نشان‌دهنده مشکلات مکرر کاربران با محصولات یا خدمات است.

- کلماتی مانند "trouble"، "error"، "delay"، و "inconvenience" بیانگر نارضایتی کاربران از تاخیر، اشتباهات یا ناراحتی‌های دیگر هستند.
- کلماتی مثل "lost"، "concern"، و "disappointed" نیز نشان‌دهنده نارضایتی عمیق‌تر کاربران هستند.

2. احساسات مثبت:

- پرتکرارترین کلمه مثبت "happy" است، که نشان‌دهنده تجربه‌های مثبت کاربران است.
- کلمات دیگری مانند "love"، "support"، و "free" نشان‌دهنده رضایت از خدمات یا محصولات هستند.
- کلماتی مثل "recommend"، "refund"، و "helpful" نشان‌دهنده اقداماتی هستند که کاربران آن‌ها را به عنوان راه‌حل موفق ارزیابی می‌کنند.

(قسمت b)

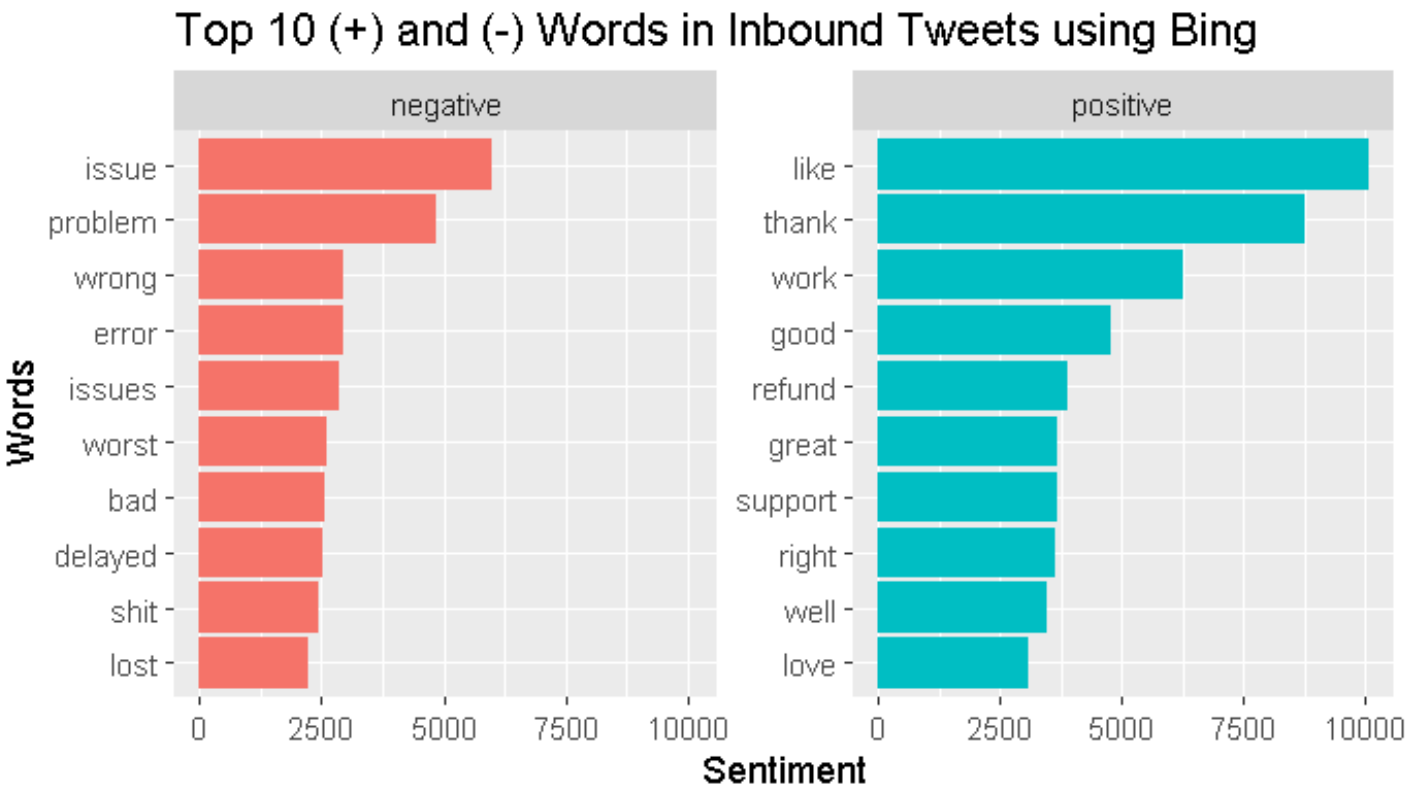
```
# 9.b

# Perform Bing sentiment analysis on inbound tweets
bing_word_counts_inbound <- clean_tweets_df %>%
  filter(inbound == "True") %>%
  unnest_tokens(word, text) %>%
  inner_join(get_sentiments("bing")) %>%
  count(word, sentiment, sort = TRUE) %>%
  ungroup()

# Visualize and save the plot for the top 10 positive and negative words
#in inbound tweets using Bing sentiment analysis
bing_word_counts_inbound %>%
  group_by(sentiment) %>%
  top_n(10) %>%
  ungroup() %>%
  mutate(word = reorder(word, n)) %>%
  ggplot(aes(word, n, fill = sentiment)) +
  geom_col(show.legend = FALSE) +
  facet_wrap(~sentiment, scales = "free_y") +
  labs(y = "Sentiment", x = "Words", title = "Top 10 (+) and (-) Words in Inbound Tweets using Bing ") +
  coord_flip()
```

این کد به تحلیل احساسات توییت‌های ورودی (Inbound Tweets) با استفاده از فرهنگ لغت Bing می‌پردازد. کلمات توییت‌های ورودی با لیست کلمات مثبت و منفی Bing مقایسه می‌شوند، و تعداد تکرار آن‌ها بر اساس نوع احساس (مثبت یا منفی) محاسبه می‌شود. سپس، 10 کلمه پرتکرار مثبت و 10 کلمه پرتکرار منفی شناسایی شده و در یک نمودار نمایش داده می‌شوند.

گام‌های کلیدی:

1. فیلتر توییت‌های ورودی:
   - فقط توییت‌هایی که از سمت کاربران به شرکت ارسال شده‌اند جدا می‌شوند.
2. تطبیق کلمات توییت‌ها با Bing:
   - کلمات موجود در توییت‌ها با لیست کلمات مثبت و منفی فرهنگ لغت Bing تطبیق داده می‌شوند.
   - تنها کلماتی که در Bing تعریف شده‌اند (مثبت یا منفی) برای تحلیل باقی می‌مانند.
3. محاسبه تعداد تکرار کلمات:
   - تعداد تکرار کلمات مثبت و منفی محاسبه شده و به ترتیب تعداد مرتب می‌شوند.
4. شناسایی پرتکرارترین کلمات:
   - 10 کلمه پرتکرار برای هر دسته مثبت و منفی انتخاب می‌شوند.
5. بصری‌سازی نتایج:
   - یک نمودار ستونی برای نمایش کلمات مثبت و منفی پرتکرار رسم می‌شود.
   - کلمات مثبت و منفی در دو بخش جداگانه نمایش داده می‌شوند.
   - نمودار به صورت افقی چرخانده می‌شود برای خوانایی بهتر.

نمودار مرتبط :



این نمودار نتایج تحلیل احساسات توییت‌های ورودی (Inbound Tweets) را با استفاده از فرهنگ لغت Bing نشان می‌دهد. کلمات مثبت و منفی پرتکرار در دو بخش جداگانه ارائه شده‌اند.

بینش‌های کلیدی:

1. احساسات منفی:

- پرتکرارترین کلمات منفی "issue" و "problem" هستند، که نشان‌دهنده نگرانی‌های اصلی کاربران در ارتباط با مشکلات است.
- کلماتی مانند "delayed" ،"error"، و "lost" مشکلات مشخصی را نشان می‌دهند که کاربران از آن شکایت می‌کنند.
- استفاده از کلمه "worst" و "shit" نشان‌دهنده شدت نارضایتی برخی کاربران است.

2. احساسات مثبت:

- پرتکرارترین کلمات مثبت "like" و "thank" هستند، که نشان‌دهنده قدردانی و رضایت کاربران است.
- کلماتی مانند "good" ،"support"، و "great" نشان‌دهنده تجربه‌های مثبت کاربران با خدمات یا محصولات است.
- کلمات "refund" و "love" نشان می‌دهند که کاربران گاهی از خدمات شرکت قدردانی می‌کنند یا پیشنهاداتی دریافت کرده‌اند که باعث خوشحالی آن‌ها شده است.

اهمیت برای تحلیل:

1. احساسات منفی:
    - این کلمات شرکت را از نگرانی‌ها و شکایات اصلی کاربران آگاه می‌کند.
    - مشکلاتی مانند تاخیر، اشتباهات، و از دست دادن کالاها باید به‌طور جدی مورد بررسی قرار گیرند.
2. احساسات مثبت:
    - این کلمات نشان‌دهنده نقاط قوت خدمات یا محصولاتی هستند که باعث جلب رضایت کاربران شده‌اند.

- استفاده از زبان قدردانی و پشتیبانی نشان‌دهنده تعامل مثبت شرکت با کاربران است.

(قسمت c)

تحلیل احساسات توییت‌های خروجی با استفاده از فرهنگ لغت Bing

کد مرتبط :

```
# 9.c

# Perform Bing sentiment analysis on outbond tweets
bing_word_counts_outbound <- clean_tweets_df %>%
  filter(inbound == "False") %>%
  unnest_tokens(word, text) %>%
  inner_join(get_sentiments("bing")) %>%
  count(word, sentiment, sort = TRUE) %>%
  ungroup()

# Visualize and save the plot for the top 10 positive and negative words
#in outbound tweets using Bing sentiment analysis
bing_word_counts_outbound %>%
  group_by(sentiment) %>%
  top_n(10) %>%
  ungroup() %>%
  mutate(word = reorder(word, n)) %>%
  ggplot(aes(word, n, fill = sentiment)) +
  geom_col(show.legend = FALSE) +
  facet_wrap(~sentiment, scales = "free_y") +
  labs(y = "Sentiment", x = "Words", title = "Top 10 (+) and (-) Words in outbound Tweets using Bing ") +
  coord_flip()
```

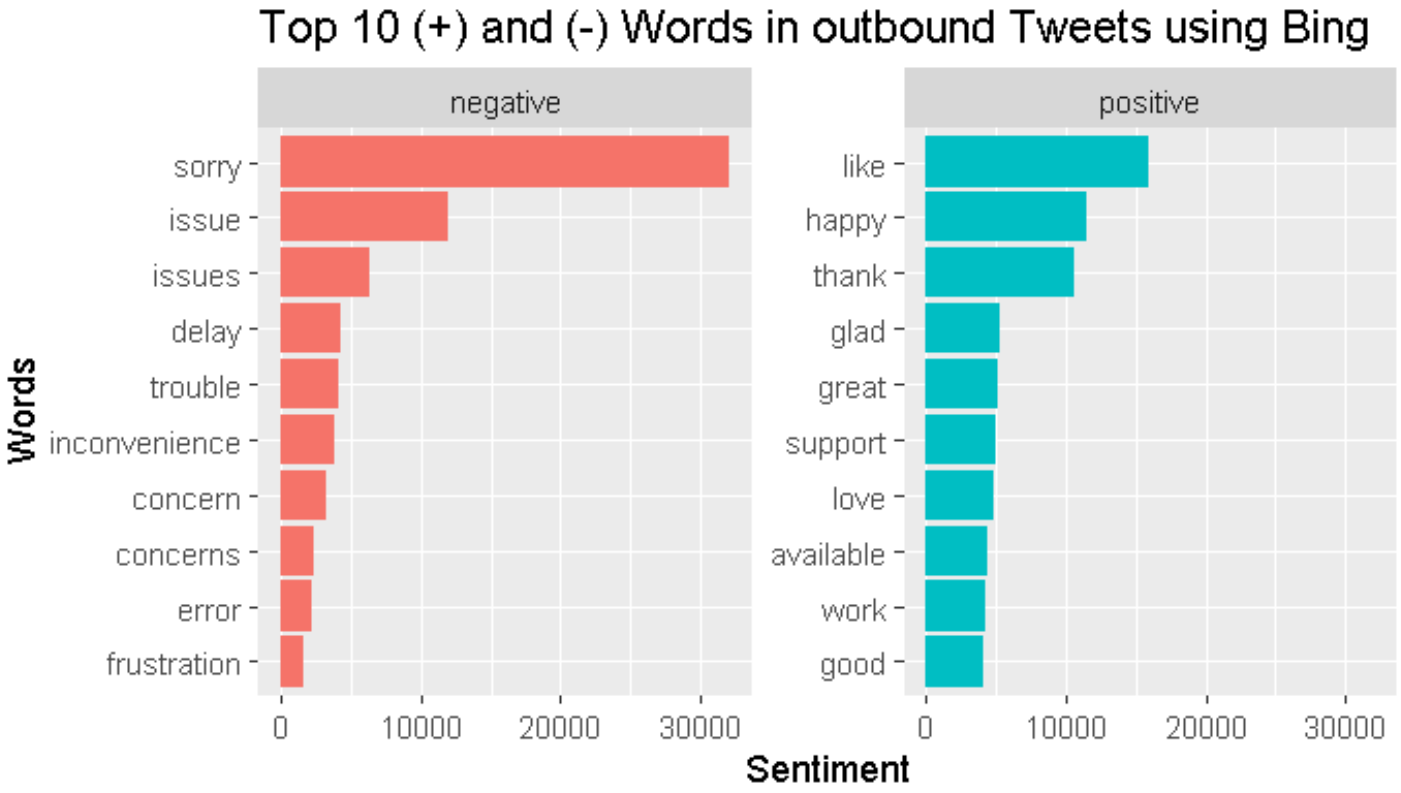
این کد به تحلیل احساسات توییت‌های خروجی (Outbound Tweets) با استفاده از فرهنگ لغت Bing می‌پردازد. کلمات موجود در این توییت‌ها بر اساس مثبت یا منفی بودن آن‌ها دسته‌بندی شده و پرتکرارترین کلمات مثبت و منفی شناسایی می‌شوند. در

نهایت، 10 کلمه پرتکرار مثبت و 10 کلمه پرتکرار منفی در یک نمودار نمایش داده می‌شوند.

گام‌های کلیدی:

1. فیلتر توییت‌های خروجی:
   - فقط توییت‌هایی که از طرف شرکت ارسال شده‌اند جدا می‌شوند.
2. تحلیل احساسات با Bing:
   - کلمات توییت‌ها با فرهنگ لغت Bing تطبیق داده می‌شوند.
   - کلمات مثبت و منفی مشخص می‌شوند و برای هر دسته تعداد تکرار محاسبه می‌شود.
3. شناسایی کلمات پرتکرار:
   - پرتکرارترین کلمات مثبت و منفی در توییت‌های خروجی شناسایی می‌شوند.
4. بصری‌سازی نتایج:
   - یک نمودار ستونی برای نمایش 10 کلمه پرتکرار مثبت و منفی رسم می‌شود.
   - کلمات مثبت و منفی در دو بخش جداگانه (Facets) نمایش داده می‌شوند و نمودار به صورت افقی چرخانده می‌شود برای خوانایی بهتر.

نمودار مرتبط :



تفسیر نمودار:

این نمودار نتایج تحلیل احساسات توییت‌های خروجی (Outbound Tweets) را با استفاده از فرهنگ لغت Bing نشان می‌دهد. کلمات مثبت و منفی پرتکرار در دو بخش جداگانه ارائه شده‌اند.

بینش‌های کلیدی:

1. احساسات منفی:

- پرتکرارترین کلمه منفی "sorry" است، که نشان‌دهنده تلاش شرکت برای عذرخواهی از کاربران در شرایط مشکل‌ساز است.

- کلماتی مانند "issues"، "issue"، و "delay" مشکلاتی را نشان می‌دهند که شرکت در حال رسیدگی به آن‌ها است.
- کلمات "error"، "concern"، و "frustration" نشان‌دهنده نگرانی‌ها و مشکلاتی است که مشتریان گزارش کرده‌اند و شرکت سعی در رفع آن‌ها دارد.

2. احساسات مثبت:

- پرتکرارترین کلمات مثبت شامل "happy"، "like"، و "thank" هستند، که نشان‌دهنده تلاش شرکت برای ایجاد احساس مثبت و قدردانی در کاربران است.
- کلماتی مانند "great"، "support"، و "love" نشان‌دهنده تأکید شرکت بر پشتیبانی خوب و ایجاد تجربه مثبت برای کاربران هستند.
- کلماتی مانند "available" و "work" ممکن است به حل مشکلات یا ارائه راه‌حل‌های مناسب اشاره داشته باشند.

اهمیت برای تحلیل:

1. احساسات منفی:
   - این کلمات نشان می‌دهند که شرکت در مواجهه با مشکلات کاربران مسئولانه رفتار می‌کند و از زبان عذرخواهی و همدلی استفاده می‌کند.
2. احساسات مثبت:
   - تأکید بر کلمات مثبت نشان‌دهنده تلاش شرکت برای ارائه پاسخ‌های دوستانه و افزایش اعتماد کاربران است.

مرحله دهم : تحلیل احساسات توییت‌ها با دسته‌بندی احساسی فرهنگ لغت NRC

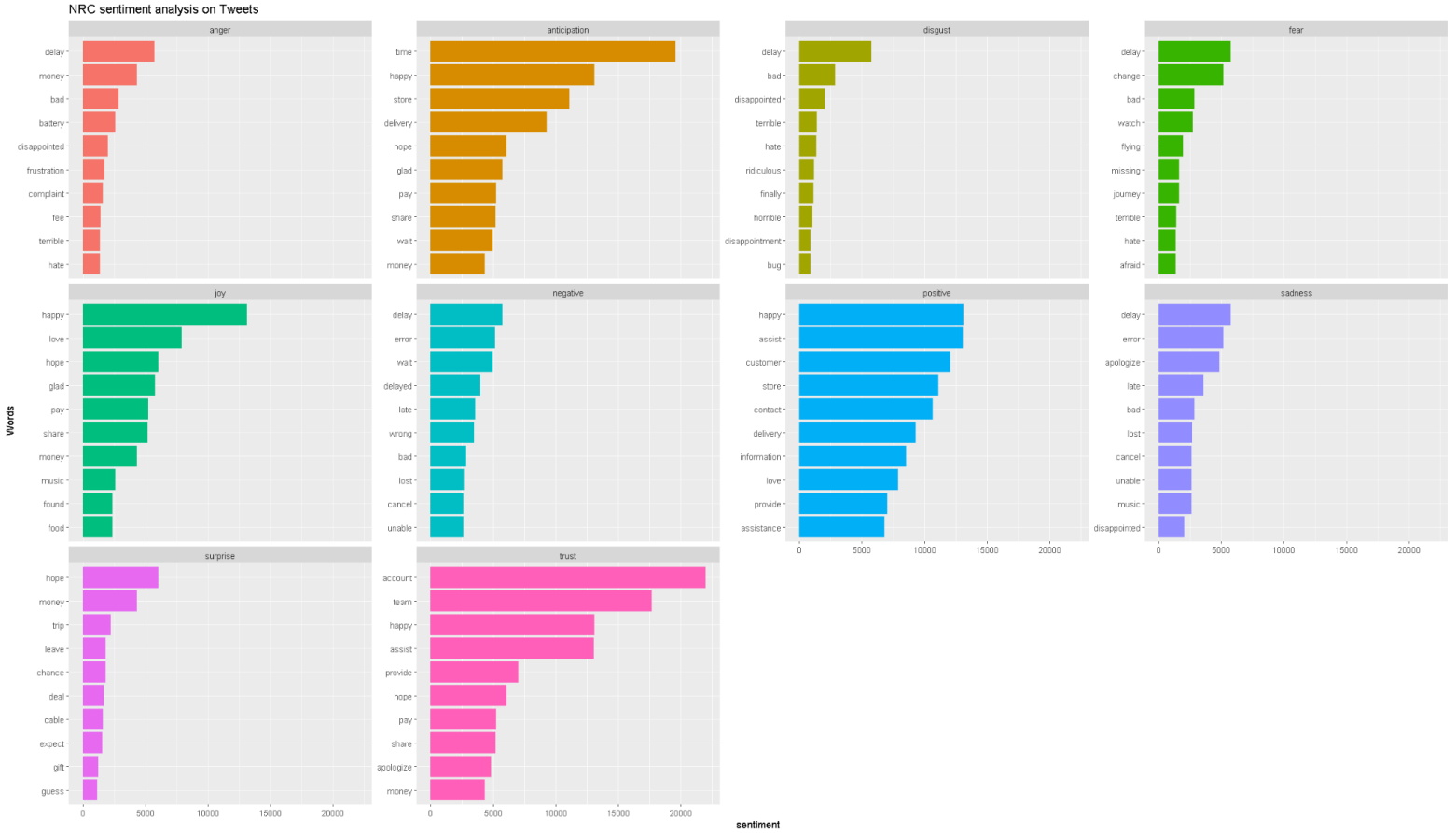
(قسمت a)

```
# step10
# 10.a

get_sentiments("nrc")

nrc_word_counts <- words_in_tweets_df %>%
  inner_join(get_sentiments("nrc")) %>%
  count(word, sentiment, sort = T) %>%
  ungroup()


nrc_word_counts %>%
  group_by(sentiment) %>%
  top_n(10) %>%
  ungroup() %>%
  mutate(word = reorder(word,n)) %>%
  ggplot(aes(word, n, fill=sentiment))+
  geom_col(show.legend = F)+
  facet_wrap(~sentiment, scales = "free_y")+
  labs(y="sentiment", x= "Words",
       title = "NRC sentiment analysis on Tweets")+
  coord_flip()
```

این کد از فرهنگ لغت NRC برای تحلیل احساسات توییت‌ها استفاده می‌کند. برخلاف فرهنگ لغت‌های دیگر، NRC احساسات را به دسته‌های مختلفی مانند خشم، ترس، شادی، غم، اعتماد، و غیره تقسیم می‌کند. کلمات توییت‌ها با این دسته‌بندی‌ها تطبیق داده شده و تعداد تکرار آن‌ها برای هر احساس محاسبه می‌شود. در نهایت، پرتکرارترین کلمات برای هر دسته احساسی در یک نمودار نمایش داده می‌شوند.

گام‌های کلیدی:

1. بارگذاری فرهنگ لغت NRC:
   - این فرهنگ لغت کلمات را به چندین دسته احساسی مانند "شادی"، "اعتماد"، "ترس"، "خشم"، و غیره طبقه‌بندی می‌کند.
2. تطبیق کلمات توییت‌ها با NRC:
   - کلمات توییت‌ها با دسته‌های احساسی NRC تطبیق داده می‌شوند.
   - فقط کلماتی که در NRC تعریف شده‌اند، برای تحلیل باقی می‌مانند.
3. محاسبه تعداد تکرار کلمات:
   - تعداد تکرار کلمات برای هر دسته احساسی محاسبه شده و به ترتیب تعداد مرتب می‌شود.
4. شناسایی پرتکرارترین کلمات:
   - 10 کلمه پرتکرار برای هر دسته احساسی انتخاب می‌شوند.
5. بصری‌سازی نتایج:
   - یک نمودار ستونی برای نمایش کلمات پرتکرار در هر دسته احساسی رسم می‌شود.
   - هر احساس به‌صورت جداگانه در یک بخش نمایش داده می‌شود.
   - نمودار به صورت افقی چرخانده می‌شود برای خوانایی بهتر.


NRC sentiment analysis on Tweets
anger
delay
money
bad
battery
disappointed
frustration
complaint
fee
terrible
hate
anticipation
time
happy
store
delivery
hope
glad
pay
share
wait
money
disgust
delay
bad
disappointed
terrible
hate
ridiculous
finally
horrible
disappointment
bug
fear
delay
change
bad
watch
flying
missing
journey
terrible
hate
afraid
joy
happy
love
hope
glad
pay
share
money
music
found
food
negative
delay
error
wait
delayed
late
wrong
bad
lost
cancel
unable
positive
happy
assist
customer
store
contact
delivery
information
love
provide
assistance
sadness
delay
error
apologize
late
bad
lost
cancel
unable
music
disappointed
surprise
hope
money
trip
leave
chance
deal
cable
expect
gift
guess
trust
account
team
happy
assist
provide
hope
pay
share
apologize
money
0
5000
10000
15000
20000
Words
sentiment

تفسیر نمودار:

این نمودار نتایج تحلیل احساسات توییت‌ها را با استفاده از فرهنگ لغت NRC نمایش می‌دهد. کلمات توییت‌ها بر اساس دسته‌بندی‌های احساسی مختلف مانند خشم، شادی، غم، ترس، اعتماد و غیره تحلیل شده‌اند. هر بخش از نمودار نشان‌دهنده 10 کلمه پرتکرار در یکی از این دسته‌های احساسی است.

بینش‌های کلیدی:

1. خشم (Anger):

- پرتکرارترین کلمات شامل "disappointed"، "bad"، "delay"، و "frustration" هستند.
- این کلمات نشان‌دهنده احساس خشم کاربران ناشی از تأخیرها، مشکلات خدمات، و نارضایتی‌ها است.

2. شادی (Joy):

- پرتکرارترین کلمات شامل "love"، "happy"، و "hope" هستند.
- این کلمات نشان‌دهنده تجربه‌های مثبت کاربران از خدمات یا تعاملات شرکت است.

3. ترس (Fear):

- کلماتی مانند "change"، "delay"، و "flying" نشان‌دهنده نگرانی کاربران در ارتباط با تغییرات، تأخیرها، یا شرایط غیرمنتظره است.

4. اعتماد (Trust):

- کلماتی مانند "assist"، "happy"، و "provide" نشان‌دهنده اعتماد کاربران به خدمات و پاسخگویی شرکت است.

5. غم (Sadness):

- پرتکرارترین کلمات شامل "error"، "delay"، و "apologize" هستند، که به مشکلات کاربران و احساس ناامیدی اشاره دارند.

6. شگفتی (Surprise):

- کلمات پرتکرار مانند "trip"، "hope"، و "chance" نشان‌دهنده تجربه‌های غیرمنتظره و مثبت یا منفی کاربران است.

7. انزجار (Disgust):

- کلمات پرتکرار شامل "bad"، "delay"، و "hate" هستند که نشان‌دهنده واکنش‌های منفی شدید کاربران است.

8. مثبت (Positive):

- کلمات پرتکرار مانند "support"، "happy"، و "customer" نشان‌دهنده تجربه‌های مثبت کاربران از خدمات است.

9. منفی (Negative):

- کلماتی مانند "error"، "delay"، و "lost" نشان‌دهنده مشکلات کاربران است که باعث نارضایتی شده است.

نتیجه‌گیری:

این نمودار نشان می‌دهد که کاربران:

1. در برخی مواقع احساسات منفی شدیدی مانند خشم، ترس، یا انزجار را تجربه کرده‌اند.
2. در مواردی دیگر اعتماد، شادی، و احساسات مثبت به خدمات شرکت ابراز کرده‌اند.

این اطلاعات برای بهبود

(قسمت b)

تحلیل احساسات کاربران در توییت‌های ورودی با دسته‌بندی احساسی NRC

کد مرتبط :

```r
# 10.b

# Perform NRC sentiment analysis on inbound tweets
nrc_word_counts_inbound <- clean_tweets_df %>%
  filter(inbound == "True") %>%
  unnest_tokens(word, text) %>%
  inner_join(get_sentiments("nrc")) %>%
  count(word, sentiment, sort = TRUE) %>%
  ungroup()

# Visualize and save the plot for the top 10 words for each emotion in inbound tweets using NRC sentiment analysis
nrc_word_counts_inbound %>%
  group_by(sentiment) %>%
  top_n(10) %>%
  ungroup() %>%
  mutate(word = reorder(word, n)) %>%
  ggplot(aes(word, n, fill = sentiment)) +
  geom_col(show.legend = FALSE) +
  facet_wrap(~ sentiment, scales = "free_y") +
  labs(y = "Sentiment", x = "Words", title = "Top 10 Words for Each Emotion in Inbound Tweets using NRC Analysis") +
  coord_flip()
```

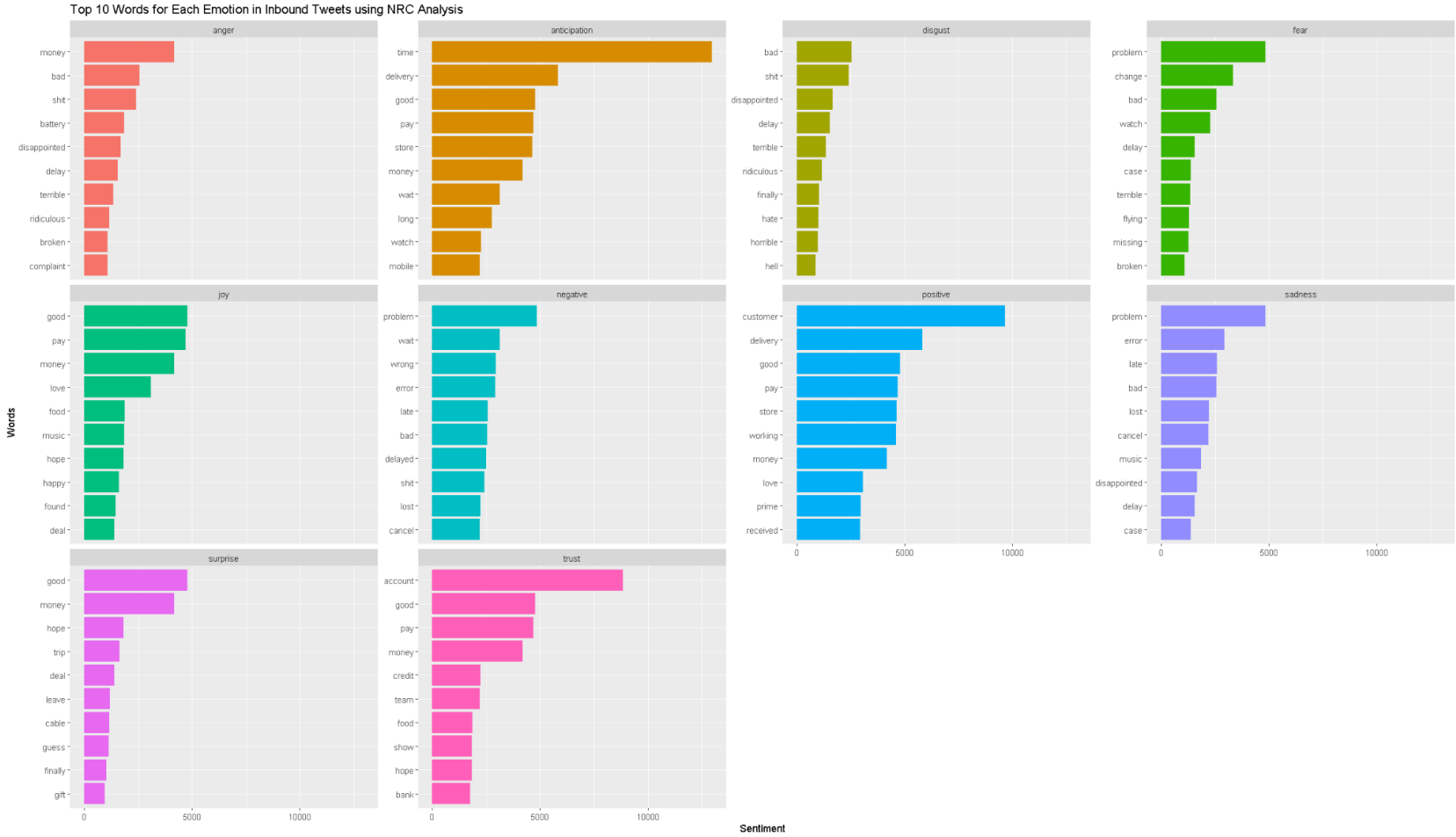
این کد تحلیل احساسات توییت‌های ورودی (Inbound Tweets) را با استفاده از فرهنگ لغت NRC انجام می‌دهد. در این تحلیل، کلمات توییت‌های ورودی به دسته‌های احساسی مختلف مانند خشم، شادی، غم، ترس، اعتماد، انزجار و غیره طبقه‌بندی

می‌شوند. سپس پرتکرارترین کلمات برای هر دسته احساسی شناسایی شده و در یک نمودار نمایش داده می‌شوند.

گام‌های کلیدی:

1. فیلتر توییت‌های ورودی:
   - تنها توییت‌هایی که از سمت کاربران به شرکت ارسال شده‌اند برای تحلیل انتخاب می‌شوند.
2. تطبیق کلمات با NRC:
   - کلمات موجود در توییت‌ها با فرهنگ لغت NRC مقایسه می‌شوند.
   - هر کلمه به دسته احساسی مربوطه مانند شادی، ترس، یا خشم تخصیص داده می‌شود.
3. محاسبه تعداد تکرار کلمات:
   - تعداد تکرار هر کلمه در هر دسته احساسی محاسبه شده و به ترتیب تعداد مرتب می‌شود.
4. شناسایی پرتکرارترین کلمات:
   - 10 کلمه پرتکرار برای هر دسته احساسی انتخاب می‌شوند.
5. بصری‌سازی نتایج:
   - یک نمودار ستونی برای هر دسته احساسی رسم می‌شود.
   - نمودار به صورت جداگانه برای هر احساس نمایش داده می‌شود.
   - چیدمان نمودار به صورت افقی است تا خوانایی بهتری داشته باشد.

نمودار مرتبط :


Top 10 Words for Each Emotion in Inbound Tweets using NRC Analysis
anger
money
bad
shit
battery
disappointed
delay
terrible
ridiculous
broken
complaint
anticipation
time
delivery
good
pay
store
money
wait
long
watch
mobile
disgust
bad
shit
disappointed
delay
terrible
ridiculous
finally
hate
horrible
hell
fear
problem
change
bad
watch
delay
case
terrible
flying
missing
broken
joy
good
pay
money
love
food
music
hope
happy
found
deal
negative
problem
wait
wrong
error
late
bad
delayed
shit
lost
cancel
positive
customer
delivery
good
pay
store
working
money
love
prime
received
0
5000
10000
sadness
problem
error
late
bad
lost
cancel
music
disappointed
delay
case
0
5000
10000
surprise
good
money
hope
trip
deal
leave
cable
guess
finally
gift
0
5000
10000
trust
account
good
pay
money
credit
team
food
show
hope
bank
0
5000
10000
Words
Sentiment

این نمودار نتایج تحلیل احساسات توییت‌های ورودی را با استفاده از فرهنگ لغت NRC نمایش می‌دهد. کلمات در دسته‌بندی‌های مختلف احساسی مانند خشم، شادی، غم، ترس، و غیره طبقه‌بندی شده و پرتکرارترین کلمات هر دسته نمایش داده شده‌اند.

بینش‌های کلیدی:

1. خشم (Anger):

- پرتکرارترین کلمات شامل "bad"، "money"، و "delay" هستند.
- این کلمات نشان‌دهنده نارضایتی کاربران در ارتباط با مسائل مالی، خدمات نامناسب و تأخیرها است.

2. شادی (Joy):

- کلمات پرتکرار مانند "love"، "good"، و "money" نشان می‌دهند که برخی کاربران تجربه‌های مثبتی را از خدمات به اشتراک گذاشته‌اند.

3. ترس (Fear):

- کلماتی مانند "change"، "problem"، و "delay" نشان‌دهنده نگرانی کاربران درباره مسائل غیرمنتظره و مشکلات خدماتی است.

4. اعتماد (Trust):

- کلماتی مانند "good"، "account"، و "pay" نشان‌دهنده اطمینان کاربران به خدمات و پاسخگویی شرکت است.

5. غم (Sadness):

- کلمات پرتکرار مانند "error"، "problem"، و "late" مشکلات کاربران را منعکس می‌کنند که باعث احساس ناراحتی یا ناامیدی شده‌اند.

6. شگفتی (Surprise):

- کلمات "hope"، "good"، و "trip" نشان‌دهنده احساسات غیرمنتظره مثبت یا حتی منفی در بین کاربران است.

7. انزجار (Disgust):

- کلماتی مانند "disappointed"، "bad"، و "delay" نشان‌دهنده واکنش‌های منفی شدید کاربران نسبت به برخی خدمات است.

8. مثبت (Positive):

- کلمات پرتکرار شامل "delivery"، "customer"، و "love" نشان‌دهنده بازخوردهای مثبت کاربران در مورد خدمات یا محصولات است.

9. منفی (Negative):

- کلمات پرتکرار مانند "error"، "problem"، و "wrong" مشکلات کاربران را منعکس می‌کنند که نیازمند توجه شرکت است.

نتیجه‌گیری:

این نمودار نشان می‌دهد که:

1. کاربران اغلب مشکلات مالی، تأخیرها، و خطاها را گزارش می‌دهند که منجر به احساسات منفی می‌شود.

2. احساسات مثبت حول تجربه‌های خوب مشتری، ارسال موفق، و کیفیت خدمات شکل می‌گیرد.

(قسمت c)

تحلیل احساسات توییت‌های خروجی شرکت با دسته‌بندی NRC

کد مرتبط :

```
# 10.c

# Perform NRC sentiment analysis on outbound tweets
nrc_word_counts_outbound <- clean_tweets_df %>%
  filter(inbound == "False") %>%
  unnest_tokens(word, text) %>%
  inner_join(get_sentiments("nrc")) %>%
  count(word, sentiment, sort = TRUE) %>%
  ungroup()

# Visualize and save the plot for the top 10 words for each emotion in outbound tweets using NRC sentiment analysis
nrc_word_counts_outbound %>%
  group_by(sentiment) %>%
  top_n(10) %>%
  ungroup() %>%
  mutate(word = reorder(word, n)) %>%
  ggplot(aes(word, n, fill = sentiment)) +
  geom_col(show.legend = FALSE) +
  facet_wrap(~ sentiment, scales = "free_y") +
  labs(y = "Sentiment", x = "Words", title = "Top 10 Words for Each Emotion in outbound Tweets using NRC Analysis") +
  coord_flip()
```

این کد تحلیل احساسات توییت‌های خروجی (Outbound Tweets) را با استفاده از فرهنگ لغت NRC انجام می‌دهد. در این تحلیل، کلمات توییت‌های خروجی به دسته‌های احساسی مختلف مانند خشم، شادی، غم، ترس، اعتماد، و غیره طبقه‌بندی شده و پرتکرارترین کلمات هر دسته شناسایی و نمایش داده می‌شوند.

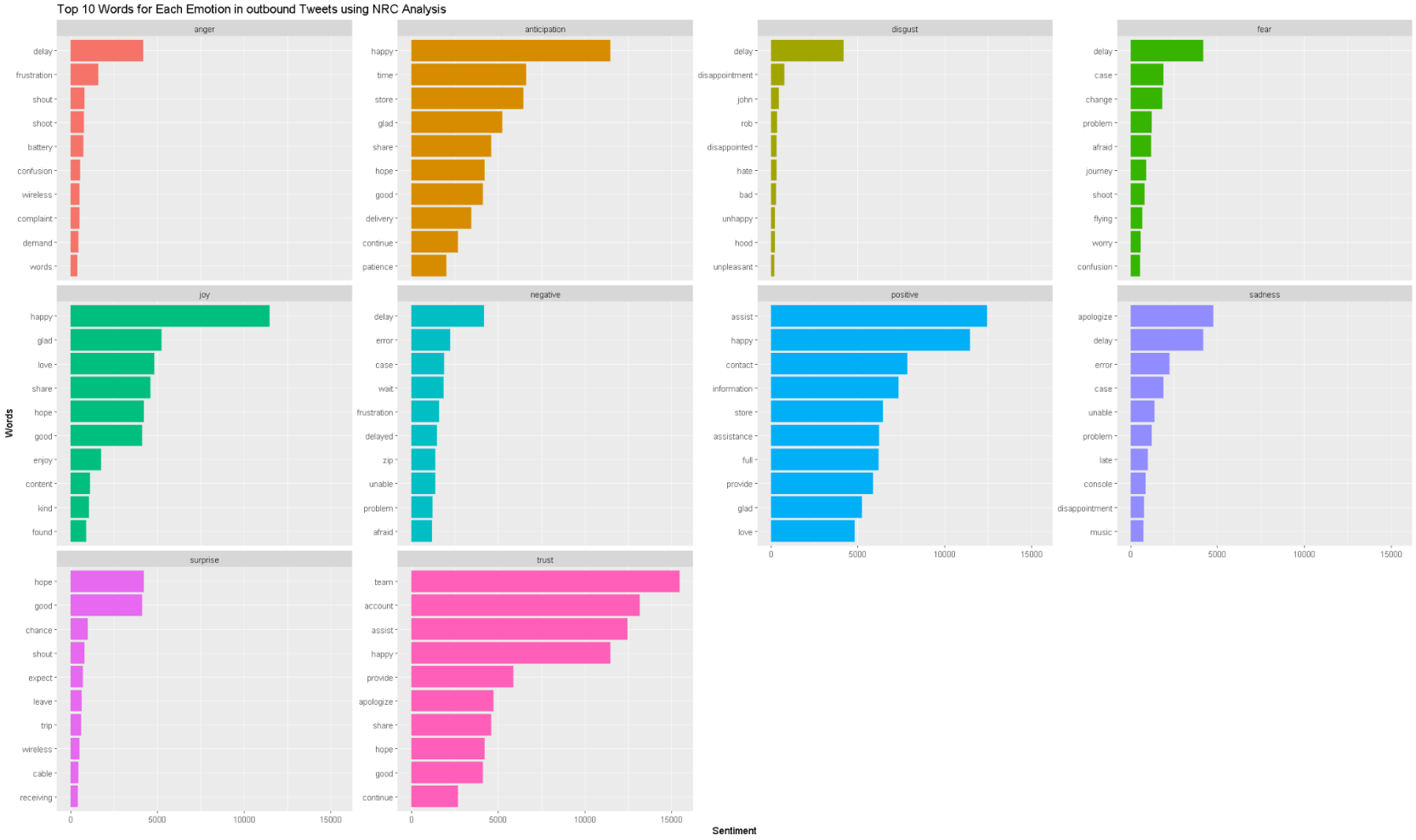
گام‌های کلیدی:

1. فیلتر توییت‌های خروجی:

- فقط توییت‌هایی که از سمت شرکت به کاربران ارسال شده‌اند برای تحلیل انتخاب می‌شوند.

2. تطبیق کلمات با NRC:
   - کلمات موجود در توییت‌ها با فرهنگ لغت NRC مقایسه می‌شوند.
   - هر کلمه به دسته احساسی مرتبط مانند شادی، ترس، یا اعتماد تخصیص داده می‌شود.
3. محاسبه تعداد تکرار کلمات:
   - تعداد تکرار هر کلمه در هر دسته احساسی محاسبه شده و به ترتیب تعداد مرتب می‌شود.
4. شناسایی پرتکرارترین کلمات:
   - 10 کلمه پرتکرار برای هر دسته احساسی انتخاب می‌شوند.
5. بصری‌سازی نتایج:
   - یک نمودار ستونی برای هر دسته احساسی رسم می‌شود.
   - هر احساس به‌صورت جداگانه در یک بخش نمایش داده می‌شود.
   - نمودار به صورت افقی است تا خوانایی بهتری داشته باشد.

نمودار مرتبط :


Top 10 Words for Each Emotion in outbound Tweets using NRC Analysis
anger
delay
frustration
shout
shoot
battery
confusion
wireless
complaint
demand
words
anticipation
happy
time
store
glad
share
hope
good
delivery
continue
patience
disgust
delay
disappointment
john
rob
disappointed
hate
bad
unhappy
hood
unpleasant
fear
delay
case
change
problem
afraid
journey
shoot
flying
worry
confusion
joy
happy
glad
love
share
hope
good
enjoy
content
kind
found
negative
delay
error
case
wait
frustration
delayed
zip
unable
problem
afraid
positive
assist
happy
contact
information
store
assistance
full
provide
glad
love
sadness
apologize
delay
error
case
unable
problem
late
console
disappointment
music
surprise
hope
good
chance
shout
expect
leave
trip
wireless
cable
receiving
trust
team
account
assist
happy
provide
apologize
share
hope
good
continue
0
5000
10000
15000
Words
Sentiment

این نمودار نتایج تحلیل احساسات توییت‌های خروجی شرکت را با استفاده از فرهنگ لغت NRC نمایش می‌دهد. کلمات بر اساس دسته‌بندی‌های احساسی مانند خشم، شادی، اعتماد، ترس، و غیره سازماندهی شده و پرتکرارترین کلمات هر دسته نشان داده شده‌اند.

بینش‌های کلیدی:

1. خشم (Anger):

- پرتکرارترین کلمات شامل "frustration"، "delay"، و "complaint" هستند.
- این کلمات نشان‌دهنده تلاش شرکت برای رسیدگی به مشکلاتی است که باعث نارضایتی کاربران شده است.

2. شادی (Joy):

- کلمات "glad"، "happy"، و "love" نشان می‌دهند که شرکت در پیام‌های خروجی خود سعی دارد حس مثبتی را به کاربران منتقل کند.

3. ترس (Fear):

- کلماتی مانند "problem"، "delay"، و "case" نشان‌دهنده توجه شرکت به نگرانی‌ها و مسائل غیرمنتظره کاربران است.

4. اعتماد (Trust):

- پرتکرارترین کلمات شامل "assist"، "team"، و "provide" هستند که نشان‌دهنده تأکید شرکت بر پشتیبانی و ایجاد اعتماد است.

5. غم (Sadness):

- کلماتی مانند "delay"، "apologize"، و "error" نشان می‌دهند که شرکت از زبان عذرخواهی برای کاهش ناراحتی کاربران استفاده می‌کند.

6. شگفتی (Surprise):

- کلماتی مانند "good"، "hope"، و "chance" نشان‌دهنده تلاش شرکت برای انتقال تجربه‌های مثبت و غیرمنتظره است.

7. انزجار (Disgust):

- کلماتی مانند "disappointment"، "delay"، و "bad" نشان‌دهنده تلاش شرکت برای حل مشکلات کاربران و کاهش واکنش‌های منفی است.

8. مثبت (Positive):

- کلماتی مانند "contact"، "assist"، و "happy" نشان می‌دهند که شرکت سعی دارد در پیام‌های خود حس مثبت ایجاد کند و بر کمک‌رسانی تأکید دارد.

9. منفی (Negative):

- کلماتی مانند "error"، "delay"، و "frustration" نشان‌دهنده مشکلات گزارش‌شده کاربران است که شرکت به آن‌ها رسیدگی می‌کند.

اهمیت برای تحلیل:

- برای احساسات منفی: شرکت از زبان همدلی و عذرخواهی استفاده می‌کند تا واکنش‌های منفی را مدیریت کند.
- برای احساسات مثبت: شرکت روی ایجاد اعتماد و انتقال پیام‌های مثبت به کاربران تمرکز دارد.

مرحله یازدهم : توزیع تعداد توییت‌ها بر اساس روزهای هفته

کد مرتبط :

```r
# step11

# Extract the day of the week from the 'created_at' column
tweets_df$day_of_week <- weekdays(clean_tweets_addition$created_at)

# Plotting the number of tweets by day of the week
tweets_df %>%
  ggplot(aes(x = day_of_week)) +
  geom_bar(fill = "#ff5733") +
  labs(x = "Day of the Week", y = "Number of Tweets", title = "Number of Tweets by Day of the Week") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))  # Rotate x-axis labels for better visibility
```

این کد برای تحلیل و نمایش توزیع توییت‌ها بر اساس روزهای هفته طراحی شده است. ابتدا روز هفته هر توییت از ستون created_at استخراج می‌شود و سپس تعداد توییت‌ها برای هر روز از هفته محاسبه و در قالب یک نمودار ستونی نمایش داده می‌شود.

گام‌های کلیدی:

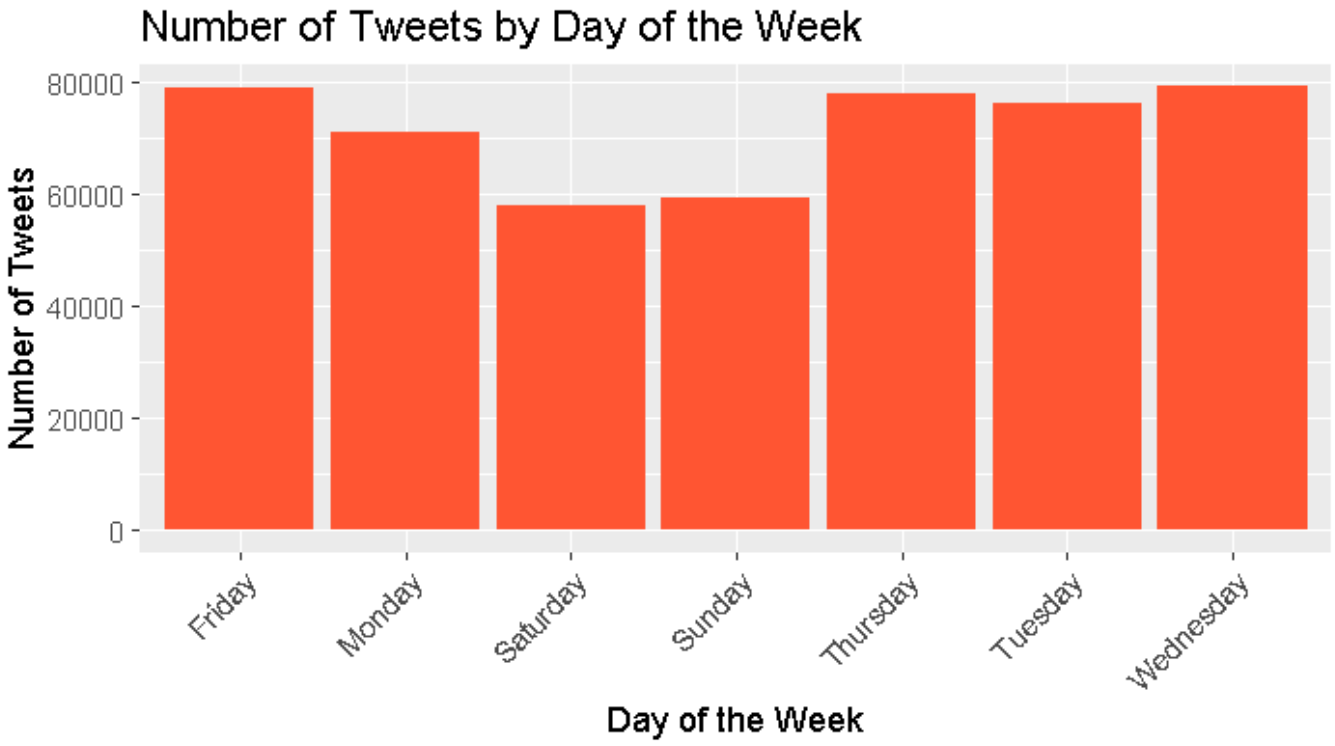
1. استخراج روز هفته:
   - با استفاده از تابع weekdays، روز هفته (مانند دوشنبه، سه‌شنبه) از تاریخ و زمان هر توییت در ستون created_at استخراج و در یک ستون جدید به نام day_of_week ذخیره می‌شود.
2. محاسبه تعداد توییت‌ها:
   - تعداد توییت‌ها برای هر روز هفته محاسبه می‌شود.

3. بصری‌سازی توزیع توییت‌ها:
   - یک نمودار ستونی با استفاده از کتابخانه ggplot2 رسم می‌شود.
   - محور X روزهای هفته و محور Y تعداد توییت‌ها را نشان می‌دهد.
   - برچسب‌های محور X برای خوانایی بهتر کمی چرخانده می‌شوند.

هدف کد:

- شناسایی الگوی فعالیت کاربران در روزهای مختلف هفته.
- بررسی توزیع زمانی توییت‌ها برای برنامه‌ریزی بهتر پاسخگویی یا تحلیل بیشتر.

نمودار مرتبط :



تفسیر نمودار:

این نمودار توزیع تعداد توییت‌ها بر اساس روزهای هفته را نمایش می‌دهد. محور X نشان‌دهنده روزهای هفته و محور Y تعداد توییت‌ها در هر روز است.

بینش‌های کلیدی:

1. روزهای پر ترافیک:
   - روز جمعه بیشترین تعداد توییت‌ها را دارد که ممکن است نشان‌دهنده افزایش فعالیت کاربران در انتهای هفته باشد.
   - روزهای پنج‌شنبه و چهارشنبه نیز فعالیت بالایی را نشان می‌دهند.
2. روزهای کم‌ترافیک:
   - شنبه و یکشنبه کمترین تعداد توییت‌ها را دارند، که ممکن است به دلیل کاهش فعالیت کاربران در تعطیلات آخر هفته باشد.
3. فعالیت متعادل در هفته:
   - تعداد توییت‌ها در باقی روزهای هفته تقریباً ثابت و متعادل است.

اهمیت برای تحلیل:

- این الگو می‌تواند به شرکت‌ها کمک کند تا در روزهای پرترافیک، منابع بیشتری برای پاسخگویی اختصاص دهند.
- روزهای کم‌ترافیک می‌توانند زمان مناسبی برای انجام بهینه‌سازی‌ها یا فعالیت‌های دیگر باشند.

نتیجه‌گیری:

این تحلیل نشان می‌دهد که:

1. کاربران در انتهای هفته فعالیت بیشتری در توییتر دارند.
2. شرکت‌ها می‌توانند استراتژی‌های پاسخگویی خود را با توجه به این الگو تنظیم کنند.

جمع‌بندی و پایان‌بندی پروژه

این پروژه با هدف تحلیل متون و احساسات توییت‌ها طراحی و اجرا شد. با استفاده از داده‌های توییت‌های ورودی و خروجی، مجموعه‌ای از تکنیک‌های پاک‌سازی داده، تحلیل احساسات، و بصری‌سازی اجرا شد تا بینش‌های عمیقی از تعاملات کاربران و پاسخ‌های شرکت به دست آید.

مراحل کلیدی پروژه:

1. پاک‌سازی داده‌ها:
   - حذف نویزهای متنی مانند لینک‌ها، علائم نگارشی، و کلمات نامناسب برای افزایش کیفیت داده‌ها و آماده‌سازی آن‌ها برای تحلیل.
2. تحلیل احساسات:
   - با استفاده از فرهنگ لغت‌های مختلف مانند Bing، AFINN، و NRC، احساسات کاربران در دسته‌های مثبت و منفی یا جزئی‌تر مانند شادی، خشم، اعتماد و غیره طبقه‌بندی شد.
3. توزیع زمانی و پرتکرارترین کلمات:
   - تحلیل‌های زمانی نشان داد که کاربران در روزهای خاصی از هفته (مانند جمعه) فعالیت بیشتری دارند.

- بررسی کلمات پرتکرار در توییت‌های ورودی و خروجی بینش‌هایی از دغدغه‌های کاربران و نحوه پاسخگویی شرکت فراهم کرد.

4. بصری‌سازی داده‌ها:
   - تمامی تحلیل‌ها با استفاده از نمودارها ارائه شد تا بینش‌ها به صورت گرافیکی و قابل‌فهم‌تر منتقل شوند.

نتایج کلیدی:

- کاربران معمولاً در مواجهه با تاخیرها، مشکلات فنی، و مسائل مالی احساسات منفی مانند خشم و غم را ابراز می‌کنند.
- شرکت‌ها با استفاده از زبان همدلانه و مثبت، از جمله عذرخواهی و ارائه پشتیبانی، سعی در کاهش تأثیر احساسات منفی دارند.
- تجربه‌های مثبت کاربران شامل رضایت از خدمات مشتری، ارسال به‌موقع، و پاسخگویی سریع بوده که موجب افزایش اعتماد و شادی آن‌ها شده است.

کاربردهای پروژه:

- شرکت‌ها می‌توانند از این تحلیل‌ها برای بهبود تعاملات خود با کاربران و افزایش کیفیت خدمات استفاده کنند.
- با شناسایی مشکلات پرتکرار و نقاط قوت، امکان تنظیم استراتژی‌های پاسخگویی بهتر و ارائه تجربه مشتری بهینه وجود دارد.

پایان‌بندی:

این پروژه نشان داد که تحلیل داده‌های متنی و احساسات می‌تواند به شرکت‌ها در درک بهتر مشتریان و بهبود استراتژی‌های تعامل با آن‌ها کمک کند. با استفاده از این رویکرد، سازمان‌ها می‌توانند روابط قوی‌تری با کاربران خود ایجاد کرده و تجربه‌ای مثبت و ارزشمند برای مشتریان فراهم کنند.