

Lotte EDA

2024-03-03

I. MỤC TIÊU PHÂN TÍCH

Dữ liệu cần được phân tích là một tập tin Excel ghi lại các bài đăng và tương tác trên trang fanpage của chuỗi cửa hàng Lotteria từ năm 2012 đến năm 2022. Dựa vào tập tin đó, nhóm đã đề xuất các mục tiêu phân tích như sau:

- Hiểu rõ về sự phát triển của trang fanpage Lotteria từ 2012 đến 2022.

Trang fanpage của Lotteria là một nền tảng truyền thông quan trọng giúp thương hiệu tương tác với khách hàng. Việc nắm bắt được sự phát triển của trang này từ năm 2012 đến 2022 sẽ cung cấp cái nhìn tổng quan về cách mà sự hiện diện trực tuyến của Lotteria đã thay đổi và tiến triển qua các năm. Điều này có thể bao gồm tần suất đăng bài, nội dung bài đăng và sự tương tác từ cộng đồng mạng.

- Đánh giá hiệu suất bài đăng dựa trên tương tác của người dùng.

Việc đánh giá hiệu suất của các bài đăng trên trang fanpage Lotteria dựa trên tương tác của người dùng là một bước quan trọng để hiểu được mức độ ảnh hưởng và sự phổ biến của các nội dung được chia sẻ. Bằng cách phân tích số lượng like, comment, share và các hành động tương tác khác, chúng ta có thể xác định những bài đăng có nội dung thế nào thu hút được sự chú ý và tương tác cao nhất từ người dùng.

- Phân tích sự ảnh hưởng của xu hướng và từ khóa được sử dụng trong các bài đăng đến mức độ tương tác của người dùng.

Bằng cách phân tích các xu hướng và từ khóa được dùng chủ yếu, chúng ta có thể đánh giá được những từ khóa nào có thể kích thích sự tương tác từ phía người dùng, bao gồm cả số lượng like, comment, share và các hành động khác. Điều này giúp chúng ta hiểu rõ hơn về cách mà Lotteria tạo ra nội dung để thu hút và tương tác tích cực với cộng đồng mạng, từ đó cải thiện và tối ưu hóa chiến lược truyền thông.

II. TÌM HIỂU DỮ LIỆU

- Tải packages và các thư viện cần dùng

```
library(readxl)
library(lubridate)

## 
## Attaching package: 'lubridate'

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

```
library(ggplot2)
library(tm)

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

## Loading required package: NLP

##
## Attaching package: 'NLP'

## The following object is masked from 'package:ggplot2':
## 
##     annotate

library(readr)
library(SnowballC)
library(magrittr)
library(tidyverse)

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

## — Attaching core tidyverse packages ————— tidyverse
2.0.0 —
## ✓ dplyr    1.1.4      ✓ stringr 1.5.1
## ✓forcats 1.0.0      ✓ tibble   3.2.1
## ✓ purrr   1.0.2      ✓ tidyrr   1.3.1

## — Conflicts —————
tidyverse_conflicts() —
## X NLP::annotate()    masks ggplot2::annotate()
## X tidyrr::extract()  masks magrittr::extract()
## X dplyr::filter()    masks stats::filter()
## X dplyr::lag()       masks stats::lag()
## X purrr::set_names() masks magrittr::set_names()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all
conflicts to become errors

library(scales)

##
## Attaching package: 'scales'
##
## The following object is masked from 'package:purrr':
## 
##     discard
##
## The following object is masked from 'package:readr':
## 
##     col_factor

library(igraph)
```

```

## 
## Attaching package: 'igraph'
## 
## The following objects are masked from 'package:dplyr':
## 
##     as_data_frame, groups, union
## 
## The following objects are masked from 'package:purrr':
## 
##     compose, simplify
## 
## The following object is masked from 'package:tidyverse':
## 
##     crossing
## 
## The following object is masked from 'package:tibble':
## 
##     as_data_frame
## 
## The following objects are masked from 'package:lubridate':
## 
##     %--%, union
## 
## The following objects are masked from 'package:stats':
## 
##     decompose, spectrum
## 
## The following object is masked from 'package:base':
## 
##     union

library(wordcloud)

## Loading required package: RColorBrewer

```

- Đọc file data

```

fanpost <- read_excel("fanpost_Lotteria_Vietnam_W4.xlsx")
fanpost <- data.frame(fanpost)

stopwords_import <- read.csv('vietnamese-stopwords.txt')
for (i in stopwords_import){
  vietnamese_stopwords <- c(i)
}

```

- Số dòng, cột và cấu trúc dữ liệu

```

dim(fanpost)

## [1] 5241    16

str(fanpost)

```

```

## 'data.frame': 5241 obs. of 16 variables:
## $ STT      : num 1 2 3 4 5 6 7 8 9 10 ...
## $ create_time: chr "2022-07-25T07:49:35+0000" "2022-07-25T06:27:56+0000"
"2022-07-24T10:00:03+0000" "2022-07-24T03:00:05+0000" ...
## $ post_id   : chr "350006241681991_6080276518654906"
"350006241681991_6080149462000945" "350006241681991_6077875762228315"
"350006241681991_6077145425634682" ...
## $ message   : chr "Này bạn ơi, đừng quên chiếc hẹn \"Thanh Xuân\" tại
Lotteria với chúng tôi nha 😊\n\n🔥 Đặt hàng qua App LOTTERIA V"|
_truncated_ "Các fan Lotteria ơi, chốt đơn ngay, nhận quà liền tay nha!!!"
🎉\nMiễn phí 01 Gà rán* cho đơn hàng đầu tiên qua A"|
_truncated_ "Em chỉ
có thẻ \"say yes\" mà thôi vì em quá u mê Lotteria mất öy 😊🎉🎊\n\n🔥 Đặt
hang qua App LOTTERIA VN và nh"|
_truncated_ "Bạn thân ơi, hãy vui lên nào,
deal hời siêu hot tải app nhận quà ngay!!! \n\n🔥 Miễn phí 01 Gà rán* cho
đơn hàng"|
_truncated_ ...
## $ status     : chr "added_photos" "added_photos" "added_photos"
"added_photos" ...
## $ comment    : num 1 0 2 1 3 11 7 9 7 4 ...
## $ share      : num 0 0 1 3 3 1 3 0 11 9 ...
## $ like       : num 27 24 24 21 73 34 74 29 505 535 ...
## $ love       : num 5 1 0 1 7 2 12 3 25 12 ...
## $ haha       : num 0 0 1 0 0 0 0 0 1 0 ...
## $ wow        : num 0 0 0 0 0 0 0 0 0 0 ...
## $ sad         : num 0 0 0 0 1 0 0 0 0 1 ...
## $ angry      : num 0 0 0 0 0 0 0 0 0 0 ...
## $ thankful   : num 0 0 0 0 0 0 0 0 0 0 ...
## $ pride      : num 0 0 0 0 0 0 0 0 0 0 ...
## $ care       : num 0 0 0 0 0 0 0 0 0 0 ...

```

Data Frame gồm 16 cột và 5241 dòng với tên cột lần lượt là:

Tên cột	Kiểu dữ liệu	Mô tả
STT	num	Số thứ tự
create_time	chr	Thời gian đăng bài
post_id	chr	Mã bài đăng
message	chr	Nội dung bài đăng
status	chr	Trạng thái bài đăng
comment	num	Số lượng bình luận
share	num	Số lượt chia sẻ
like	num	Số lượt thích
love	num	Số lượt yêu thích
haha	num	Số lượt haha
wow	num	Số lượt wow
sad	num	Số lượt buồn

Tên cột	Kiểu dữ liệu	Mô tả
angry	num	Số lượt phẫn nộ
thankful	num	Số lượt biết ơn
pride	num	Số lượt tự hào
care	num	Số lượt thương thương

Để đạt được mục tiêu phân tích đã đề ra, nhóm đã xác định các biến quan trọng cần được sử dụng trong quá trình phân tích gồm “message”, “create_time” và tổng lượt tương tác. Bằng cách sử dụng những biến này, nhóm hy vọng có thể tạo ra những phân tích sâu hơn và phản ánh chính xác hơn về sự phát triển và hiệu suất của trang fanpage Lotteria qua các năm.

- Giá trị trống (Missing value)

```
# Kiểm tra giá trị trống
colSums(is.na(fanpost))
```

```
##      STT create_time    post_id message   status comment
##      0          0          0     230      85        0
## share    like       love    haha     wow      sad
##      0          0          0       0      0        0
## angry thankful    pride    care      0        0
##      0          0          0       0      0        0
```

Trong phạm vi thời gian phân tích, biến “message” có 230/5241 giá trị NA (4.39%) và biến “status” có 85/5241 giá trị NA (1.62%), trong khi các biến khác không có giá trị NA. Các giá trị NA của cả hai biến chủ yếu tập trung trong khoảng thời gian từ tháng 6/2013 đến hết năm 2015.

Ở các dòng dữ liệu có giá trị NA ở hai biến status và message, các số liệu tương tác vẫn được ghi nhận. Do đó, nhóm đã ghi nhận giá trị NA nhưng không loại bỏ chúng khi phân tích, nhằm duy trì tính toàn vẹn và chính xác của dữ liệu.

- Đổi múi giờ

Để phù hợp với bối cảnh dữ liệu, nhóm đã thực hiện việc chuyển đổi thời gian trong cột “create_time” sang múi giờ Việt Nam. Quá trình này giúp nhóm có cái nhìn toàn diện hơn về cách mà các bài đăng và hoạt động tương tác thay đổi theo thời gian và trong ngữ cảnh địa lý cụ thể của thị trường Việt Nam.

```
# Đổi sang múi giờ Việt Nam
fanpost$create_time <- as.POSIXct(fanpost$create_time, tz = "UTC", "%Y-%m-%dT%H:%M:%OS")
fanpost$create_time_VN <- with_tz(fanpost$create_time, tz = "Asia/Saigon")
```

- Thêm các cột timestamp

Để có thể phân tích xu hướng ở các mốc thời gian khác nhau, nhóm đã sử dụng thư viện lubridate để tách và thêm các cột mới như giờ, thứ, ngày, tháng, năm từ thời gian trong cột “create_time”.

```
# Tách thời gian
fanpost$Year <- year(fanpost$create_time_VN)
fanpost$Month <- month(fanpost$create_time_VN, label = TRUE)
fanpost$Day <- day(fanpost$create_time_VN)
fanpost$Weekday <- wday(fanpost$create_time_VN, label=TRUE)
fanpost$Hour <- hour(fanpost$create_time_VN)
```

- Tính điểm react và comment

Để phân tích mức độ tương tác của người dùng, nhóm đã tổng hợp số lượt tương tác ở các mục riêng lẻ về cảm xúc (như like, love, haha, wow, sad, angry, thankful, pride, care) và các tương tác khác (comment, share) thành hai biến mới là “react_score” và “comment_score”.

Biến “react_score” được tính bằng tổng số lượt tương tác từ các cảm xúc, trong đó số lượt tương tác về các cảm xúc “sad” và “angry” là số âm, phản ánh sự tiêu cực của cộng đồng mạng đối với bài đăng. Biến “comment_score” là tổng số lượt bình luận và chia sẻ.

```
# Tính điểm react_score
fanpost$react_score <- fanpost$like + fanpost$love + fanpost$haha +
fanpost$wow - fanpost$sad - fanpost$angry + fanpost$thankful + fanpost$pride
+ fanpost$care

# Tính điểm comment_score
fanpost$comment_score <- fanpost$comment + fanpost$share
```

- Dữ liệu sau khi thêm các cột

```
head(fanpost)

##   STT          create_time                  post_id
## 1   1 2022-07-25 07:49:35 350006241681991_6080276518654906
## 2   2 2022-07-25 06:27:56 350006241681991_6080149462000945
## 3   3 2022-07-24 10:00:03 350006241681991_6077875762228315
## 4   4 2022-07-24 03:00:05 350006241681991_6077145425634682
## 5   5 2022-07-23 03:00:03 350006241681991_6074517305897494
## 6   6 2022-07-22 08:00:24 350006241681991_6072339619448596
##
message
## 1
Này bạn ơi, đừng quên chiếc hẹn "Thanh Xuân" tại Lotteria với chúng tôi nha
😊\n\n🔥 Đặt hàng qua App LOTTERIA VN và nhận ngay 1 miếng Gà rán miễn phí
cho đơn đầu tiên nhé Fans ơi.\nTải ngay tại 👉
https://onelink.to/c3jzpb\n#LotteriaVietnam #MyFAVORITELOTTERIA \n-----\nLưu ý:\n(*) Chương trình khuyến mãi có thể không được áp dụng tại một số cửa
hàng, mong quý khách thông cảm\n(*) Mau mau đặt hàng qua hotline 1900.6778 để
được giao hàng miễn phí và nhận nhiều ưu đãi hấp dẫn hơn nhé!!!
## 2
Các fan Lotteria ơi, chốt đơn ngay, nhận quà liền tay nha!!! 😊\nMiễn phí 01
Gà rán* cho đơn hàng đầu tiên qua App LOTTERIA VN nha các fan ơi!!!\nMiễn phí
giao hàng cùng rất nhiều ưu đãi hấp dẫn, tải App nhanh bằng cách quét mã QR
tại hình ảnh đi kèm hoặc nhấp chọn các đường link bên dưới 👇👇\nIOS:
```

<https://apps.apple.com/vn/app/lotteria/id1423528065?l=vi> \nAndroid:
<https://play.google.com/store/apps/details...> \n*Áp dụng cho hoá đơn từ 99.000đ.\n#LotteriaVietnam #MyFAVORITELOTTERIA
3
Em chỉ có thể "say yes" mà thôi vì em quá u mê Lotteria mất ày 😍❤️🌐\n\n🔥
Đặt hàng qua App LOTTERIA VN và nhận ngay 1 miếng Gà rán miễn phí cho đơn đầu tiên nhé Fans ơi.\nTải ngay tại 
<https://onelink.to/c3jzpb>\n#LotteriaVietnam #MyFAVORITELOTTERIA \n----\nLưu ý:\n(*) Chương trình khuyến mãi có thể không được áp dụng tại một số cửa hàng, mong quý khách thông cảm\n(*) Mau mau đặt hàng qua hotline 1900.6778 để được giao hàng miễn phí và nhận nhiều ưu đãi hấp dẫn hơn nhé!!!
4
Bạn thân ơi, hãy vui lên nào, deal hời siêu hot tải app nhận quà ngay!!! \n\n🔥 Miễn phí 01 Gà rán* cho đơn hàng đầu tiên qua App LOTTERIA VN nha các fan ơi!!!\n\n⭐ Miễn phí giao hàng cùng rất nhiều ưu đãi hấp dẫn, tải App nhanh bằng cách quét mã QR tại hình ảnh đi kèm hoặc nhấp chọn các đường link bên dưới
 \niOS: <https://apps.apple.com/vn/app/lotteria/id1423528065?l=vi>
\nAndroid: <https://play.google.com/store/apps/details...> \n*Áp dụng cho hoá đơn từ 99.000đ.\n#LotteriaVietnam #MyFAVORITELOTTERIA
5 Nhanh nhanh ra tận hưởng chiếc deal siêu hời nhà Lotteria bạn ơi 😊\n👉 Combo 187k gồm 3 Gà rán + Burger bulgogi + Mì ý (L) + Khoai tây lắc + 3 Pepsi (M)\n\n🔥 Đặt hàng qua App LOTTERIA VN và nhận ngay 1 miếng Gà rán miễn phí cho đơn đầu tiên nhé Fans ơi.\nTải ngay tại 
<https://onelink.to/c3jzpb>\n#LotteriaVietnam #MyFAVORITELOTTERIA \n----\nLưu ý:\n(*) Chương trình khuyến mãi có thể không được áp dụng tại một số cửa hàng, mong quý khách thông cảm\n(*) Mau mau đặt hàng qua hotline 1900.6778 để được giao hàng miễn phí và nhận nhiều ưu đãi hấp dẫn hơn nhé!!!
6
Thì ra lăng mạn không phải là mùa hè, \nlăng mạn là mùa hè được ăn gà rán Lotteria 😊\n\n🔥 Đặt hàng qua App LOTTERIA VN và nhận ngay 1 miếng Gà rán miễn phí cho đơn đầu tiên nhé Fans ơi.\nTải ngay tại 
<https://onelink.to/c3jzpb>\n#LotteriaVietnam #MyFAVORITELOTTERIA \n----\nLưu ý:\n(*) Chương trình khuyến mãi có thể không được áp dụng tại một số cửa hàng, mong quý khách thông cảm\n(*) Mau mau đặt hàng qua hotline 1900.6778 để được giao hàng miễn phí và nhận nhiều ưu đãi hấp dẫn hơn nhé!!!
status comment share like love haha wow sad angry thankful pride care
1 added_photos 1 0 27 5 0 0 0 0 0 0 0
2 added_photos 0 0 24 1 0 0 0 0 0 0 0
3 added_photos 2 1 24 0 1 0 0 0 0 0 0
4 added_photos 1 3 21 1 0 0 0 0 0 0 0
5 added_photos 3 3 73 7 0 0 1 0 0 0 0
6 added_photos 11 1 34 2 0 0 0 0 0 0 0

```

0      create_time_VN Year Month Day Weekday Hour react_score
comment_score
## 1 2022-07-25 14:49:35 2022 Jul 25 Mon 14 32
1
## 2 2022-07-25 13:27:56 2022 Jul 25 Mon 13 25
0
## 3 2022-07-24 17:00:03 2022 Jul 24 Sun 17 25
3
## 4 2022-07-24 10:00:05 2022 Jul 24 Sun 10 22
4
## 5 2022-07-23 10:00:03 2022 Jul 23 Sat 10 79
6
## 6 2022-07-22 15:00:24 2022 Jul 22 Fri 15 36
12

```

III. Khám phá dữ liệu

- Thống kê mô tả

```
#Bảng thống kê mô tả
summary(fanpost$react_score)
```

```
##   Min. 1st Qu. Median     Mean 3rd Qu.    Max.
##   0.0   60.0  150.0   509.7  411.0 35915.0
```

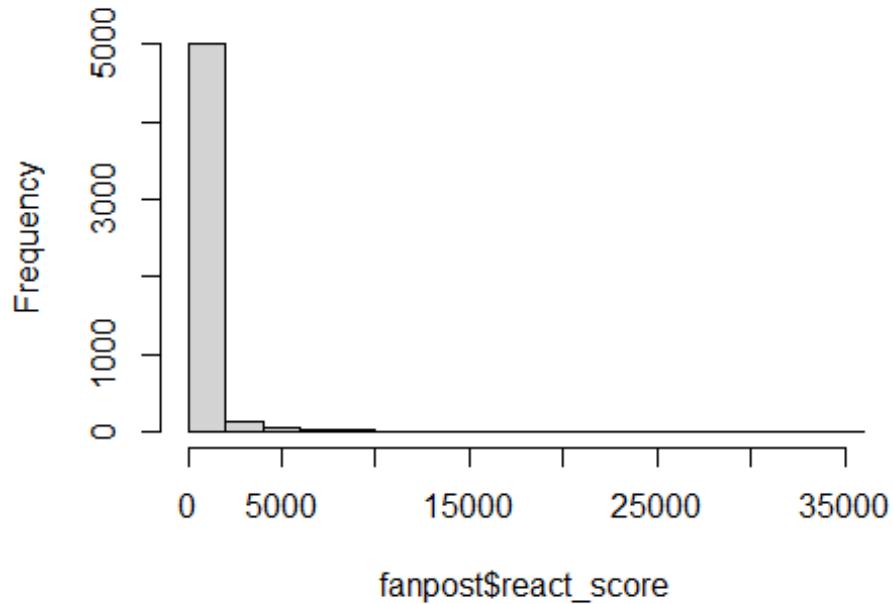
```
summary(fanpost$comment_score)
```

```
##   Min. 1st Qu. Median     Mean 3rd Qu.    Max.
##   0.00  5.00  14.00   80.81  50.00 10753.00
```

Từ bảng thống kê mô tả, các chỉ số thống kê của “react score” đều lớn hơn rất nhiều so với “comment score”, cho thấy người dùng thường chuộng tương tác thông qua việc sử dụng biểu tượng cảm xúc nhiều hơn so với việc tương tác bằng cách bình luận và chia sẻ. Điều này có thể là do tính tiện lợi và tốc độ trong việc truyền đạt ý kiến hoặc cảm xúc của họ.

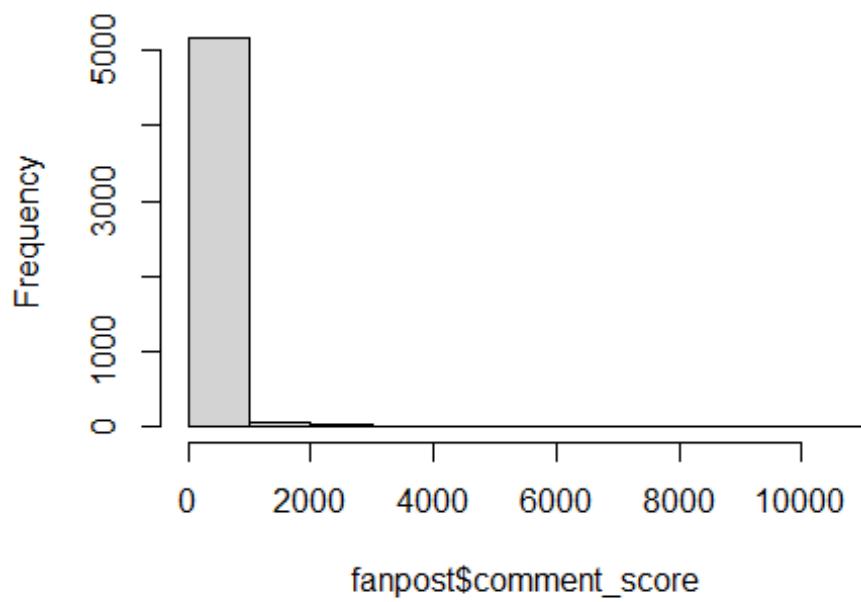
```
# Phân phối biến react_score
hist(fanpost$react_score, main="Histogram of React score")
```

Histogram of React score



```
# Phân phối biến comment_score  
hist(fanpost$comment_score, main="Histogram of React score")
```

Histogram of React score

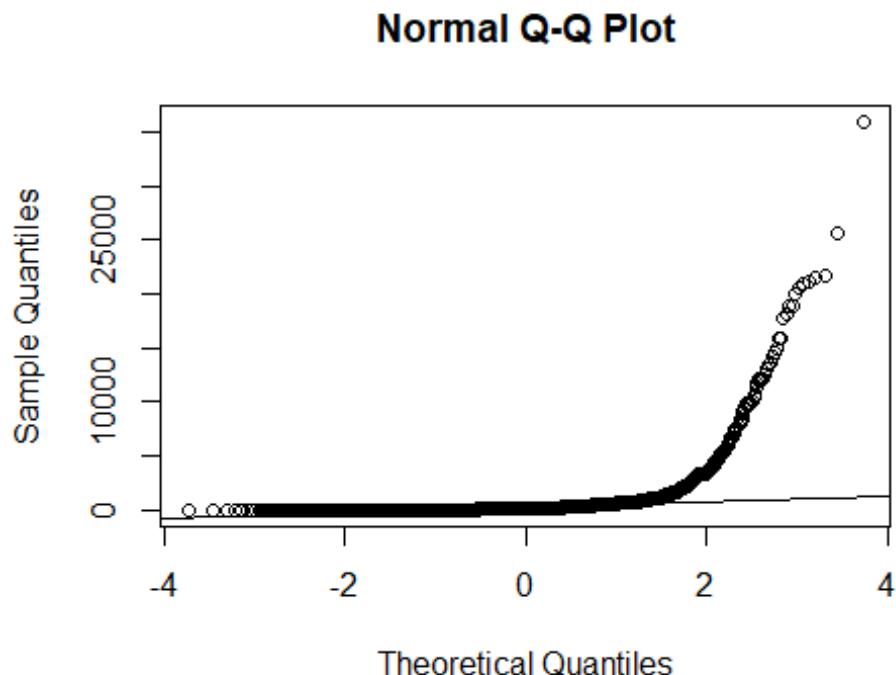


Từ bảng thống kê mô tả và biểu đồ phân phối ở cả 2 biến react_score và comment_score, giá trị mean đều lớn hơn median và các giá trị thường cao đột biến tại nửa bên trái, thấp ở nửa bên phải cho thấy dữ liệu cả 2 biến đều lệch phải (right skewed)

```
# Phương sai  
round(sd(fanpost$react_score),2)  
## [1] 1580.61  
  
round(sd(fanpost$comment_score),2)  
## [1] 306.28
```

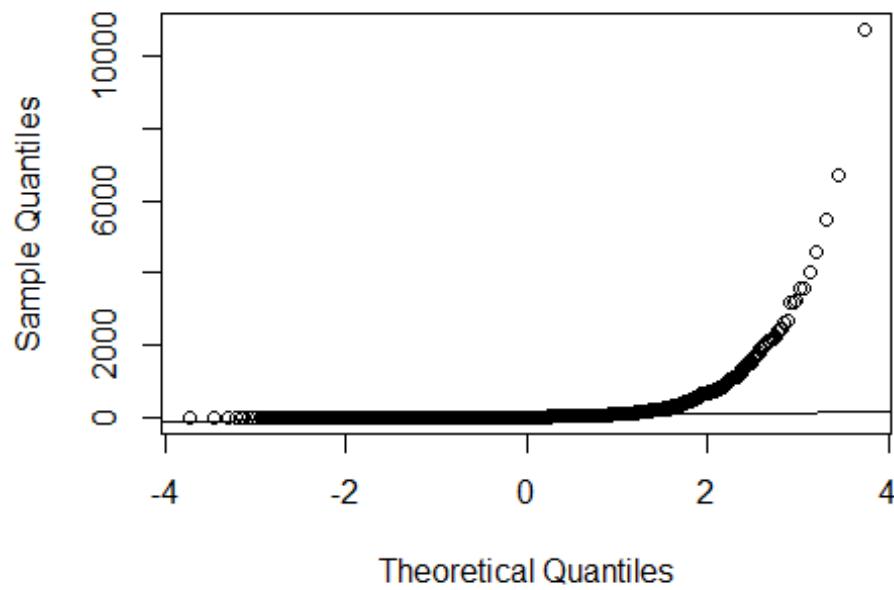
Bên cạnh đó, cả hai biến đều có khoảng biến thiên và phương sai lớn, tức là có sự biến động đáng kể giữa các giá trị dữ liệu. Tuy nhiên, giá trị mean và tứ phân vị thứ ba (75% quan sát) lại khá nhỏ so với khoảng biến thiên này. Điều này cho thấy dữ liệu có thể chứa nhiều điểm dữ liệu bất thường (giá trị ngoại lai).

```
# Biểu đồ Normal Q-Q biến react_score  
qqnorm(fanpost$react_score)  
qqline(fanpost$react_score)
```



```
# Biểu đồ Normal Q-Q biến comment_score  
qqnorm(fanpost$comment_score)  
qqline(fanpost$comment_score)
```

Normal Q-Q Plot



Kết hợp với biểu đồ Normal Q-Q, cả 2 biến đều không phân bố đồng đều gần đường xu hướng và đều có nhiều giá trị ngoại lai với giá trị lớn cần được xem xét kỹ lưỡng trong quá trình phân tích và đánh giá.

- Giá trị ngoại lai

Như đã nêu trên, cả 2 biến react_score và comment score phân phối lệch một cách rõ ràng nên nhóm quyết định chọn phương pháp IQR để xác định và loại bỏ outliers vì phương pháp này không phụ thuộc vào dạng phân phối của dữ liệu.

Do tập dữ liệu không lớn và có nhiều giá trị ngoại lai với giá trị rất cao, để tránh mất mát dữ liệu và đảm bảo tính chính xác của quá trình phân tích, nhóm chọn ghi nhận các giá trị ngoại lai lệch rất nhiều (extreme outliers) với $k=3$.

```
# Tìm outliers bằng IQR
react_score <- fanpost$react_score
comment_score <- fanpost$comment_score
# Tính Q1 và Q3 cho cả hai cột
Q1React <- quantile(react_score, 0.25)
Q3React <- quantile(react_score, 0.75)

Q1Comment <- quantile(comment_score, 0.25)
Q3Comment <- quantile(comment_score, 0.75)

# Tính IQR cho cả hai cột
IQRReact <- IQR(react_score)
IQRComment <- IQR(comment_score)
```

```

# Xác định khoảng giá trị bình thường cho cả hai cột
k <- 3
lower_bound_react <- Q1_react - k * IQR_react
upper_bound_react <- Q3_react + k * IQR_react

lower_bound_comment <- Q1_comment - k * IQR_comment
upper_bound_comment <- Q3_comment + k * IQR_comment

# Xác định outlier cho cả hai cột
outliers_react_low <- react_score[react_score < lower_bound_react]
outliers_react_high <- react_score[react_score > upper_bound_react]

outliers_comment_low <- comment_score[comment_score < lower_bound_comment]
outliers_comment_high <- comment_score[comment_score > upper_bound_comment]

# Hiển thị kết quả
print("Outliers in React score (lower)")

## [1] "Outliers in React score (lower)"

print(outliers_react_low)

## numeric(0)

print("Outliers in React score (higher)")

## [1] "Outliers in React score (higher)"

print(outliers_react_high)

## [1] 2806 1876 5817 2555 4564 1721 3080 2542 2506 3334 6571
1580
## [13] 1551 20047 12070 21430 1471 1620 1555 2274 1934 2018 1540
2091
## [25] 2018 1540 2091 2149 1505 1830 2388 3470 3828 3453 2409
2117
## [37] 2140 3643 1764 3883 3361 2392 2040 2366 2314 4080 3147
2739
## [49] 2118 1854 2187 2174 1755 1859 2043 2056 3279 1511 2876
2585
## [61] 1607 2918 3497 1517 4851 3732 1649 4304 1967 4529 3611
2259
## [73] 3982 4606 2305 9911 4504 1626 7774 1549 1602 1946 1950
5336
## [85] 2872 3418 3784 1570 1686 1482 2250 1519 3271 2079 3054
1583
## [97] 1741 1544 1688 5119 3188 1834 2019 3712 4570 5154 5667
1483
## [109] 1733 6228 5932 2392 5266 4020 3267 1645 4445 2282 2415
1532

```

```

## [121] 7739 2185 9812 2843 1523 1510 1722 1718 2902 1587 2539
1675
## [133] 2523 2346 6584 5502 1784 2243 1492 2907 2817 4129 10160
5903
## [145] 7244 3395 5683 9824 12172 9790 8347 1635 12424 1867 3016
18848
## [157] 1888 8249 1552 2752 6792 2593 3330 2479 11938 5820 5932
1588
## [169] 12160 18041 2076 18956 3193 1671 1709 2045 9998 15884 21146
4500
## [181] 1548 3398 6733 4226 7122 3049 9562 15861 10605 21636 35915
4073
## [193] 2954 10363 1781 2964 3052 14941 14544 3880 1971 2539 1701
7857
## [205] 13413 2319 4839 6550 5840 4291 2640 2153 9850 5363 3118
5087
## [217] 25683 20552 2307 6765 6786 9910 5082 2117 5287 1693 3421
2290
## [229] 2614 2275 4902 9130 1571 1976 2951 8483 7595 1987 7563
5416
## [241] 13588 13449 3617 8117 3874 1934 3808 5467 3411 3411 3411
3411
## [253] 3411 3411 3411 3411 3411 3411 3411 3411 12996 3411 3411 3411
3411
## [265] 3411 9906 7907 3411 3411 3411 3411 3411 14127 17669 3021 7431
2188
## [277] 1549 2053 1895 1749 20936 2627 1921 4516 11822 3254 1652
12137
## [289] 1683 1650 2572 8963 1565 1565 1565 1565 1565 1565 1565 1565
1565
## [301] 1565 1565 1565 11395 1498 1740 3105 2049 9418 5785 1636
2603

print("Outliers in Comment score (lower)")

## [1] "Outliers in Comment score (lower)"

print(outliers_comment_low)

## numeric(0)

print("Outliers in Comment score (higher)")

## [1] "Outliers in Comment score (higher)"

print(outliers_comment_high)

## [1] 513 704 1933 1998 3599 4023 207 579 1608 419 215
777
## [13] 877 555 455 578 880 724 202 330 693 277 259
321
## [25] 444 282 195 240 10753 334 1556 3265 200 385 536

```

648													
## [37]	1090	2650	293	310	243	3170	255	441	474	197	197		
732													
## [49]	630	271	649	777	324	354	519	347	643	328	224		
684													
## [61]	238	2104	725	198	262	1069	246	267	750	756	281		
222													
## [73]	223	328	350	1319	661	292	588	944	221	1186	4600		
189													
## [85]	642	280	1679	204	720	208	216	768	1213	270	1110		
850													
## [97]	1485	616	511	2660	343	299	874	241	212	373	234		
1092													
## [109]	353	367	272	417	459	515	617	650	824	787	691		
277													
## [121]	519	192	831	614	633	734	510	341	649	716	410		
343													
## [133]	280	642	674	302	888	209	702	261	809	311	1482		
528													
## [145]	1232	517	586	205	418	505	278	715	232	536	551		
758													
## [157]	682	540	362	293	457	243	648	403	244	333	353		
355													
## [169]	420	286	459	1379	196	269	1086	191	1733	198	229		
243													
## [181]	257	497	306	1057	1015	400	792	198	259	278	380		
1259													
## [193]	228	306	281	306	465	192	191	365	388	303	355		
1919													
## [205]	812	194	801	1498	508	219	259	378	596	2150	1416		
238													
## [217]	1090	540	2141	339	2433	210	196	455	1385	530	201		
717													
## [229]	261	719	201	286	212	433	735	807	987	1741	506		
865													
## [241]	874	1319	657	335	253	664	278	2212	1507	1687	273		
3577													
## [253]	654	232	344	211	234	227	202	198	228	186	201		
378													
## [265]	250	213	301	228	321	192	927	722	242	229	3166		
299													
## [277]	373	1069	434	251	1122	198	338	362	642	229	272		
328													
## [289]	211	231	1450	279	451	210	2394	447	304	2199	317		
244													
## [301]	244	501	244	251	244	279	455	244	244	6717	244		
352													
## [313]	244	260	243	244	244	508	244	244	244	5504	244		
244													
## [325]	241	244	206	1917	386	248	239	352	404	275	737		

```

457
## [337] 341 261 204 235 2201 2459 244 1254 780 360 327
451
## [349] 693 840 398 216 322 328 254 589 444 1165 322
471
## [361] 1494 1035 322 755 683 1814 320 230 633 491 529
1723
## [373] 522 194 322 464 377 238 376 343 223 207 354
772
## [385] 213 1099 212 210 473 510 2092 603 771 396 209
210
## [397] 1026 788 221 320 688 257 269 1049 687 213 274
1091
## [409] 672 618 233 379 255 349 1363 1057 256 254 732
348
## [421] 188 810 562 195 213 420 261 384 461 240 196
1112
## [433] 201 267 220 198 260 474 204 645 220 258 202
203
## [445] 531 276 282

```

Biến react_score có 307 giá trị ngoại lai và biến comment_score có 445 giá trị ngoại lai được ghi nhận.

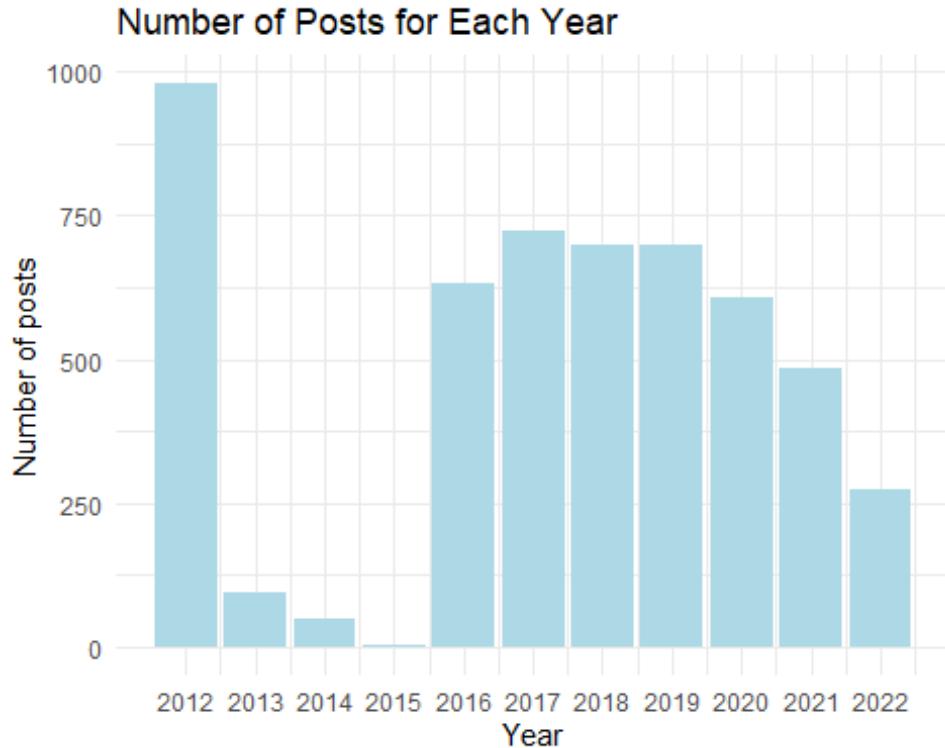
- Số lượng bài đăng theo năm

```

# số Lượng bài đăng theo năm
p1 <- ggplot(data = fanpost, aes(x = Year)) +
  geom_bar(stat='count', fill = "lightblue") +
  labs(x = "Year", y = "Number of posts",
       title = "Number of Posts for Each Year") +
  theme_minimal() + scale_x_continuous(breaks = unique(fanpost$Year))

print(p1)

```



Từ biểu đồ, có thể nhận thấy rằng số lượng bài đăng năm 2012 là cao nhất. Điều này là do trong năm đó, các bài đăng chủ yếu là các bài dự thi của các thí sinh tham gia các cuộc thi được tổ chức bởi Lotteria.

Tuy nhiên, từ năm 2012 đến 2015, số lượng bài đăng biến động mạnh và bất thường. Năm 2012 có số bài đăng cao nhất, gần 1000 bài đăng được ghi nhận. Tuy nhiên, trong những năm tiếp theo, có một sự giảm đáng kể, với khoảng 100 bài đăng vào năm 2013, khoảng 50 bài vào năm 2014 và chỉ có 2 bài được ghi nhận vào năm 2015.

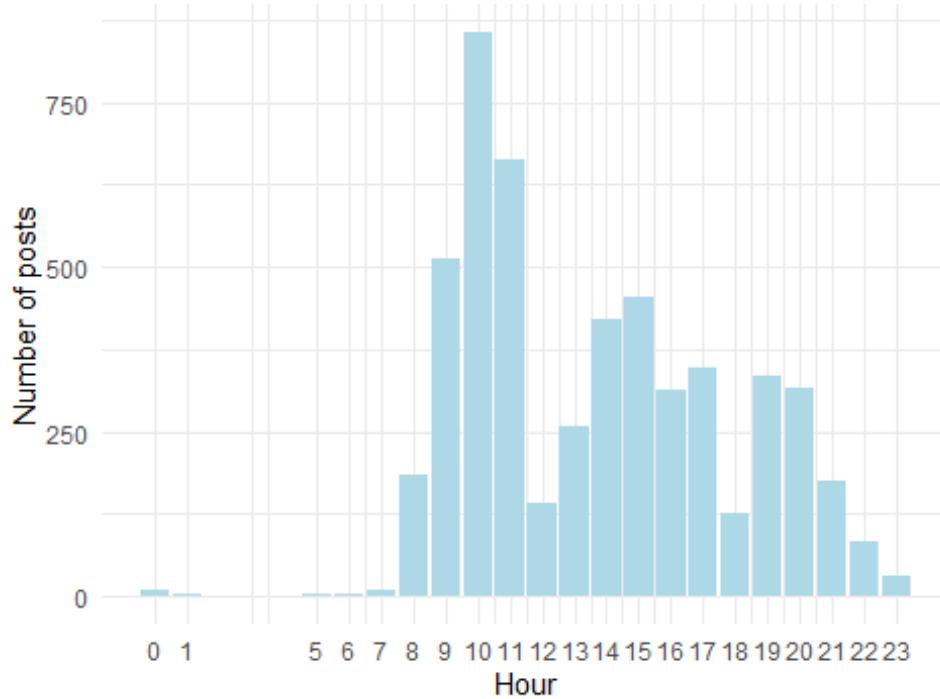
Từ giai đoạn 2016 đến 2022, số lượng bài đăng đã ổn định và có xu hướng giảm dần. Mặc dù vẫn có sự biến động trong số liệu, nhưng nó không còn quá đáng kể như trong giai đoạn trước đó.

- Số lượng bài đăng theo giờ

```
# Số Lượng bài đăng theo giờ
p2 <- ggplot(data = fanpost, aes(x = Hour)) +
  geom_bar(stat='count', fill="lightblue") +
  labs(x = "Hour", y = "Number of posts",
       title = "Number of Posts for Each Hour") +
  theme_minimal() + scale_x_continuous(breaks = unique(fanpost$Hour))

print(p2)
```

Number of Posts for Each Hour

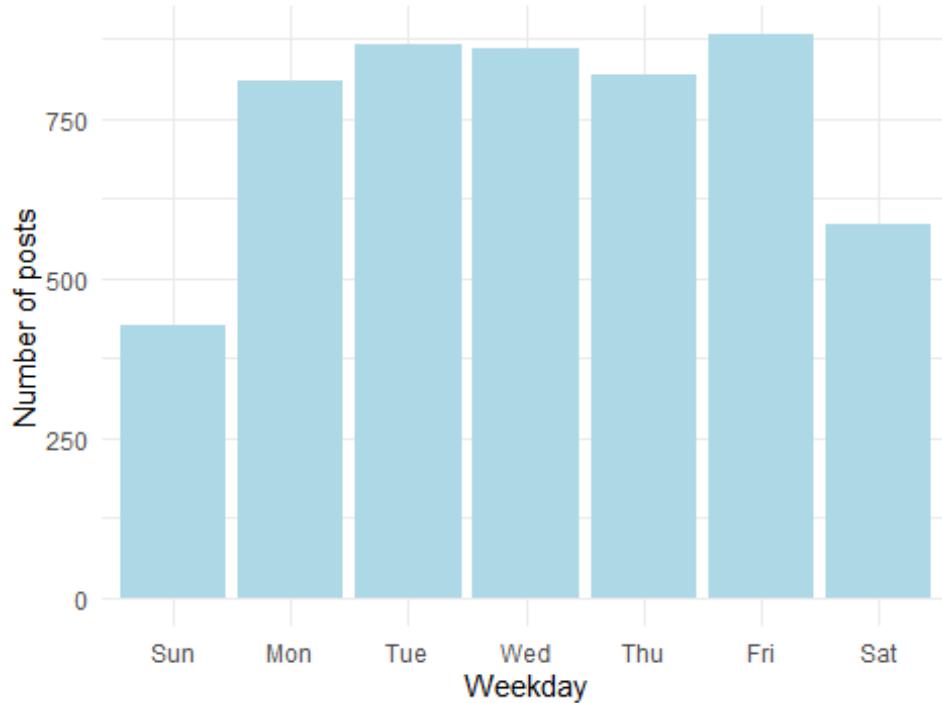


Các bài đăng thường được đăng vào các khoảng thời gian từ 9 đến 12 giờ, 14 đến 17 giờ. Đây cũng là những khoảng thời gian tốt nhất để đăng bài. (Theo: Sprout Social)

- Số lượng bài đăng theo thứ

```
p3 <- ggplot(data = fanpost, aes(x = Weekday)) +  
  geom_bar(stat='count', fill="lightblue") +  
  labs(x = "Weekday", y = "Number of posts",  
       title = "Barplot of Number of Posts for Each Weekday") +  
  theme_minimal()  
  
print(p3)
```

Barplot of Number of Posts for Each Weekday



Các bài đăng được đăng nhiều hơn vào các ngày trong tuần, với số lượng cao nhất vào thứ Sáu, trong khi có số lượng bài đăng thấp hơn nhiều vào 2 ngày cuối tuần.

- Tổng tương tác

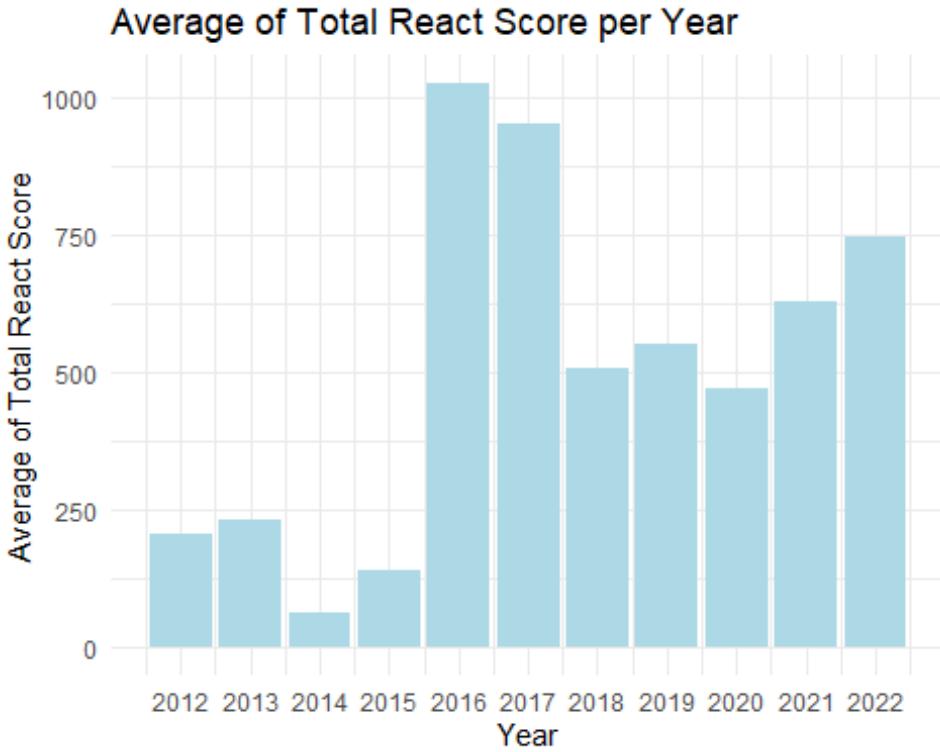
```
fanpost$total <- fanpost$react_score + fanpost$comment_score
```

- Tổng tương tác theo năm

```
average_total_score_yearly <- aggregate(fanpost$total, by =  
list(fanpost$Year), FUN = mean)
```

```
p4 <- ggplot(data = average_total_score_yearly, aes(x = Group.1, y = x)) +  
  geom_bar(stat = 'identity', fill = "lightblue") +  
  labs(x = "Year", y = "Average of Total React Score",  
       title = "Average of Total React Score per Year") +  
  theme_minimal() + scale_x_continuous(breaks =  
unique(average_total_score_yearly$Group.1))
```

```
print(p4)
```



Mặc dù năm 2012 có số lượng bài đăng cao nhất nhưng số lượt tương tác của người dùng lại thấp hơn rất nhiều so với các năm khác. Điều này có thể là do các bài đăng chủ yếu là bài dự thi của thí sinh như đã nói trên nên không quá thu hút sự quan tâm của đa số người dùng.

Trong khi đó, số lượt tương tác có xu hướng tăng trong giai đoạn từ 2018 đến 2022, mặc dù số lượng bài đăng giảm dần cho thấy hiệu quả tương tác ngày càng cao hơn ở mỗi bài đăng, đồng thời cho thấy sự chú ý và tương tác tích cực của cộng đồng người dùng trong thời gian này.

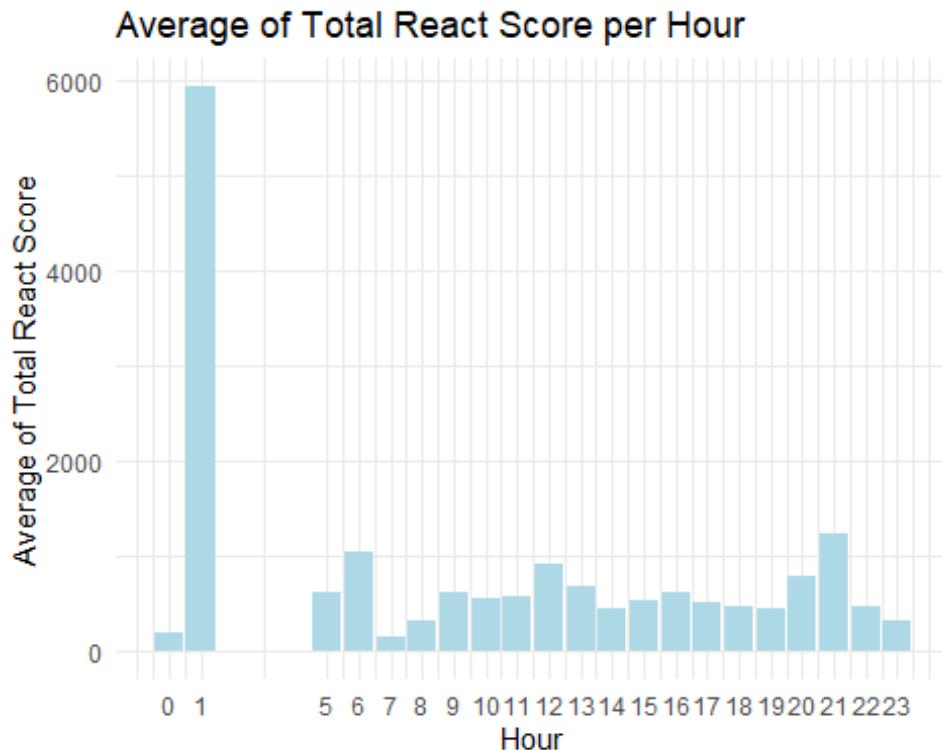
Ngoài ra, bên cạnh số lượng bài đăng, số lượt tương tác cũng ổn định hơn từ năm 2016 đến năm 2022. Vì thế, để đảm bảo tính tổng quát trong các phân tích sâu hơn tiếp theo, nhóm sẽ phân tích chủ yếu các số liệu từ năm 2016.

Tổng tương tác theo giờ

```
average_total_score_hourly <- aggregate(fanpost$total, by=
list(fanpost$Hour), FUN = mean)

p5 <- ggplot(data = average_total_score_hourly, aes(x = Group.1, y= x)) +
  geom_bar(stat='identity', fill="lightblue") +
  labs(x = "Hour", y = "Average of Total React Score",
       title = "Average of Total React Score per Hour") +
  theme_minimal() + scale_x_continuous(breaks =
unique(average_total_score_hourly$Group.1))
```

```
print(p5)
```



Có sự cao bất thường ở số lượt tương tác của người dùng vào 1 giờ do có một giá trị ngoại lai với số lượt like cao hơn hẳn (12071 lượt) so với các bài đăng còn lại, khiến cho giá trị trung bình lượt tương tác vào 1 giờ cũng cao hơn hẳn các giờ khác.

```
# Biến dữ liệu thành Corpus
fanpostCorpus <- Corpus(VectorSource(fanpost$message))

# Transform dữ liệu
fanpostCorpus <- tm_map(fanpostCorpus, content_transformer(tolower)) # Viết thường

## Warning in tm_map.SimpleCorpus(fanpostCorpus,
## content_transformer(tolower)):
## transformation drops documents

fanpostCorpus <- tm_map(fanpostCorpus, removePunctuation) # Bỏ dấu

## Warning in tm_map.SimpleCorpus(fanpostCorpus, removePunctuation):
## transformation drops documents

fanpostCorpus <- tm_map(fanpostCorpus, removeNumbers) # Bỏ những con số

## Warning in tm_map.SimpleCorpus(fanpostCorpus, removeNumbers):
## transformation
## drops documents
```

```

fanpostCorpus <- tm_map(fanpostCorpus, removeWords, vietnamese_stopwords) #  

Bỏ stopwords

## Warning in tm_map.SimpleCorpus(fanpostCorpus, removeWords,  

## vietnamese_stopwords): transformation drops documents

fanpostCorpus <- tm_map(fanpostCorpus, stripWhitespace) # Bỏ khoảng trắng

## Warning in tm_map.SimpleCorpus(fanpostCorpus, stripWhitespace):  

transformation  

## drops documents

fanpostCorpus <- tm_map(fanpostCorpus, stemDocument) # Stemming

## Warning in tm_map.SimpleCorpus(fanpostCorpus, stemDocument):  

transformation  

## drops documents

# Tạo Term Document matrix
fanpostTDM <- TermDocumentMatrix(fanpostCorpus, control =  

list(wordLengths=c(1,Inf)))
fanpostTDM <- removeSparseTerms(fanpostTDM, 0.9)
fanpostTDM <- as.matrix(fanpostTDM)

# Tìm các từ phổ biến nhất
freq <- rowSums(fanpostTDM)

print(freq)

##          chương         cảm         cửa         dụng
giao           1229          699         2832        2419
##          1598
##          gà          hàng         khuyến       lotteria
lotteriavietnam           3853          6569         1257        12604
##          997
##          lưu         miếng         miễn         mãi
nha             858          914          685         1170
##          1137
##          phí          rán          thông         trình
áp               714          1349         1098        1852
##          2152
##          quà          đơn          ưu          hình
##          1149
##          khoai         vui          burger       combo
##          1310          1637         2103        2381

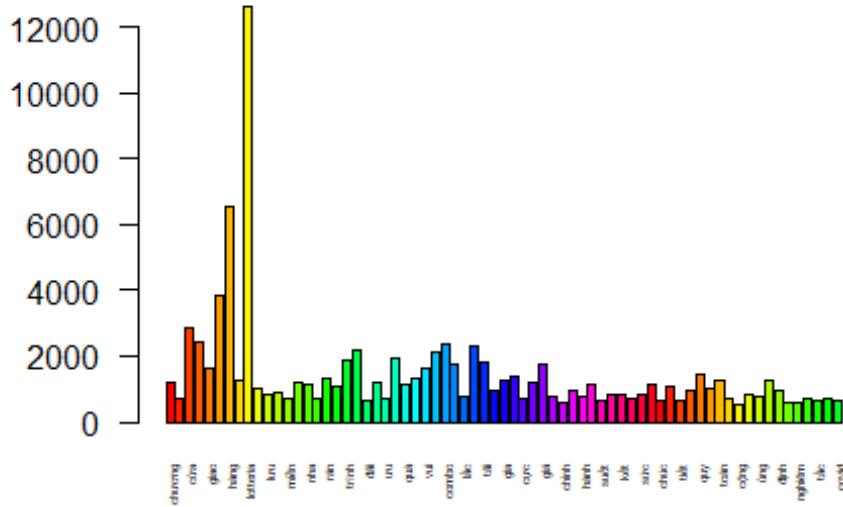
```

```

1722          lắc      pepsi       tài      mùa
##          783        2283      1801      930
1270          tặng      cực       yêu      giá
##          1387      700       1177     1778
785           chỉnh    thường   hành      mua
##           suốt
suốt          607        977       773      1117
##           món
đồng          820        823       707      855
##           chúc
1144          quy       thức      tiết      chiên
##           628        1063      621      935
1452          hộ        tham      toàn      khỏe
##           992        1271      699      539
810           nghiêm
##           747        1230      951      615
##           732        phủ       tắc      nguyên
##           660        683      covid      655

barplot(freq, las=2, col=rainbow(50), cex.names = 0.4)

```



Từ kết quả trên cho thấy một số từ có tần suất xuất hiện nhiều nhất như: lottery, hàng, cửa,... Tiếp theo, nhóm thực hiện bước tìm tổng số dòng mà mỗi từ trên xuất hiện (tức số message có các từ trên) để tìm những từ có tần suất xuất hiện nhiều nhất trong tất cả các bài đăng.

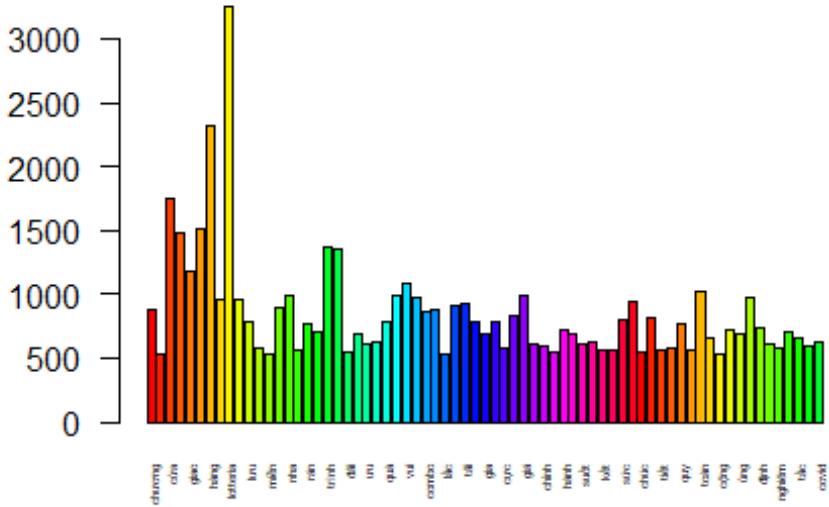
```
# Tính tổng số dòng mà mỗi từ trên xuất hiện (tức số message có các từ trên)
num_docs_with_word <- rowSums(as.matrix(fanpostTDM) > 0)
```

```
# Hiển thị kết quả
print(num_docs_with_word)
```

	chương	cảm	cửa	dụng
giao	879	525	1752	1481
##	1176			
##	gà	hàng	khuyến	lotteria
lotteriavietnam	1506	2317	960	3249
##	964			
##	lưu	miếng	miễn	mãi
nha	787	571	531	897
##	985			
##	phí	rán	thông	trình
áp	560	770	706	1364
##	1350			
##	đãi	đơn	ưu	hình

quà				
##	540	696	608	623
791				
##	đi	vui	burger	combo
khoai				
##	995	1086	977	858
885				
##	lắc	pepsi	tài	mùa
gia				
##	532	917	926	789
694				
##	tặng	cực	yêu	giá
ngon				
##	790	583	829	982
603				
##	chỉnh	thưởng	hành	mua
suốt				
##	601	554	727	692
615				
##	món	kết	điện	sức
đồng				
##	623	564	557	807
937				
##	chúc	thức	tiết	chiên
quy				
##	548	823	568	581
775				
##	tham	toàn	khỏe	cộng
hộ				
##	562	1025	663	534
723				
##	ủng	dịch	định	chấp
nghiêm				
##	693	979	742	604
582				
##	phủ	tắc	nguyên	covid
##	710	651	592	633

```
barplot(num_docs_with_word, las=2, col=rainbow(50), cex.names = 0.4)
```



Dựa trên kết quả

phân tích tần suất xuất hiện của từng từ trong các bài đăng, chúng ta có thể nhận thấy một số từ được sử dụng nhiều nhất như “lotteria”, “hàng”, “cửa”, và “gà”. Tuy nhiên, khi kết hợp cả hai kết quả trên, ta nhận thấy một số từ xuất hiện nhiều lần không chỉ trong một số bài đăng mà còn trong nhiều bài đăng khác nhau. Ví dụ, từ “lotteria” thường xuất hiện trong hashtag ở cuối mỗi bài đăng và trong nội dung của bài đăng, có thể thấy sự xuất hiện dày đặc là cách để thương hiệu “lotteria” được nhận diện bởi người dùng. Đi kèm với “lotteria”, còn có từ “lotteriavietnam”. Để tìm hiểu sâu hơn về các nhóm từ mang tính đặc trưng của từng bài đăng, nhóm quyết định loại bỏ những từ chung chung xuất hiện nhiều lần như “lotteria” và “lotteriavietnam”. Ngoài ra, từ “hàng” cũng là một từ được sử dụng nhiều thứ hai, thường kết hợp với từ “cửa” để tạo thành cụm từ “cửa hàng”. Do đó, nhóm quyết định loại bỏ từ “cửa hàng” khỏi danh sách các từ để tập trung vào những từ mang tính đặc trưng và cụ thể hơn của từng bài đăng.

```
sorted_freq <- sort(freq, decreasing = TRUE)  
print(sorted_freq)
```

## dụng	lotteria	hang	ga	cua
## 2419	12604	6569	3853	2832
## hinh	combo	pepsi	ap	burger
## 1951	2381	2283	2152	2103
## vui	trinh	tai	gia	khoai

##	1852	1801	1778	1722
1637				
##	giao	quy	tặng	rán
đi				
##	1598	1452	1387	1349
1310				
##	toàn	gia	khuyến	dịch
chương				
##	1271	1270	1257	1230
1229				
##	đơn	yêu	mãi	quà
đồng				
##	1210	1177	1170	1149
1144				
##	nha	mua	thông	thức
lotteria vietnam				
##	1137	1117	1098	1063
997				
##	tham	thưởng	định	chiên
mùa				
##	992	977	951	935
930				
##	miếng	lưu	sức	kết
món				
##	914	858	855	823
820				
##	hở	ngon	lắc	hành
ủng				
##	810	785	783	773
747				
##	phủ	ưu	phi	điện
cực				
##	732	723	714	707
700				
##	cảm	khỏe	miễn	nguyên
tắc				
##	699	699	685	683
660				
##	covid	đãi	suốt	chúc
tiết				
##	655	654	637	628
621				
##	chấp	chỉnh	nghiêm	cộng
##	615	607	595	539

Loại bỏ từ đơn, từ ghép thường xuất hiện: "Lotteria", "Lotteria vietnam", "cửa hàng",... khỏi fanpostCorpus

Danh sách các từ cần loại bỏ

```
words_to_remove <- c("lotteria", "cửa hàng", "lotteria vietnam")
```

```

# Loại bỏ các từ khỏi mỗi tài liệu trong fanpostCorpus
fanpostCorpus <- lapply(fanpostCorpus, function(doc) {
  for (word in words_to_remove) {
    doc <- gsub(word, "", doc)
  }
  return(doc)
})

# Chuyển đổi fanpostCorpus thành dataframe
fanpostData <- data.frame(text = sapply(fanpostCorpus, as.character),
stringsAsFactors = FALSE)

```

Sau khi có data của message mới đã loại bỏ các yếu tố trên, thêm cột mới có tên là new_message vào dataset fanpost để tiếp tục phân tích các nhóm từ có đặc trưng rõ ràng hơn.

```

# Thêm cột mới từ fanpostData vào fanpost
fanpost$new_message <- fanpostData$text

```

Nhóm tiến hành phân tích độ dài trung bình của các bài viết theo năm, sau đó so sánh với lượng tương tác trung bình nhằm kiểm tra sự ảnh hưởng của chiều dài bài viết tới độ tương tác.

```

# Phân cụm
positive <- subset(fanpost, react_score > 411)
negative <- subset(fanpost, react_score < 60)
neutral <- subset(fanpost, react_score >= 60 & react_score <= 411)

```

Sau đó, dựa trên react_score phân thành 3 cụm gồm: positive, negative và neutral. - Phân tích dòng

```

#Tính chiều dài của từng dòng message
fanpostData$text_length <- nchar(fanpostData$text)

```

Theo kết quả trên so sánh với biểu đồ trung bình react theo năm, có thể thấy ít sự tương quan về ảnh hưởng của độ dài bài viết tới độ tương. Vì vậy, nhóm chuyển sang hướng chia cụm dựa trên điểm tương tác. Nhóm tiến hành tính điểm tương tác qua công thức:
react_score = like + love + haha + wow - sad - angry + thankful + care. Tính điểm comment = comment + share

```

# Thêm cột mới từ fanpostData vào fanpost
fanpostData$Year <- fanpost$Year

# In kết quả
head(fanpostData)

##
text
## 1
hẹn xuân nha 🐶 🐶 hàng app vn miếng gà rán miễn phí đơn fan tải 🐶
đừng quên

```

```

https://onelinktocjzpb/vietnam/myfavorite/lưu-chương-trình-khuyến-mãi-áp-dụng
mong-quý-thông-cảm-mau-hàng-hotline-giao-hàng-miễn-phí-ưu đãi-hấp
## 2 fan-chốt đơn-quà-liền-nha 🥈 miễn-phí-gà-rán đơn-hàng
app-vn-nha-fan-miễn-phí-giao-hàng-ưu đãi-hấp-tải-app-quét-mã-qr-hình-ảnh-đi
kèm-nhấp-đường-link 🚶 io https://apps.apple.com/vn/app/id1171111111
https://play.google.com/store/apps/details?id=vietnam.myfavorite
## 3 thể-say-yes-u
mê-øy 🐦❤️🐦❤️🐦❤️🐦 hàng-app-vn-miếng-gà-rán-miễn-phí-đơn-fan-tải 🐦
https://onelinktocjzpb/vietnam/myfavorite/lưu-chương-trình-khuyến-mãi-áp-dụng
mong-quý-thông-cảm-mau-hàng-hotline-giao-hàng-miễn-phí-ưu đãi-hấp
## 4 thân-vui-deal-hời-siêu-hot-tải-app-quà 🍔 miễn-phí-gà-rán-đơn-hàng
app-vn-nha-fan ⭐ miễn-phí-giao-hàng-ưu đãi-hấp-tải-app-quét-mã-qr-hình-ảnh-đi
kèm-nhấp-đường-link 🚶 io https://apps.apple.com/vn/app/id1171111111
https://play.google.com/store/apps/details?id=vietnam.myfavorite
## 5 tận-hưởng-deal-siêu-hời 🍔combo-k-gà-rán-burger-bulgogi-mì-l-khoai
tái-lắc-pepsi-m 🍔 hàng-app-vn-miếng-gà-rán-miễn-phí-đơn-fan-tải 🍔
https://onelinktocjzpb/vietnam/myfavorite/lưu-chương-trình-khuyến-mãi-áp-dụng
mong-quý-thông-cảm-mau-hàng-hotline-giao-hàng-miễn-phí-ưu đãi-hấp
## 6 lăng-mạn-mùa-hè-lăng-mạn-mùa
hè-gà-rán 🍔 hàng-app-vn-miếng-gà-rán-miễn-phí-đơn-fan-tải 🍔
https://onelinktocjzpb/vietnam/myfavorite/lưu-chương-trình-khuyến-mãi-áp-dụng
mong-quý-thông-cảm-mau-hàng-hotline-giao-hàng-miễn-phí-ưu đãi-hấp
## text_length Year
## 1 221 2022
## 2 264 2022
## 3 227 2022
## 4 278 2022
## 5 281 2022
## 6 237 2022

# Tính trung bình tổng số từ của cột fanpostData qua từng năm
average_word_count_by_year <- aggregate(fanpostData$text_length,
by=list(fanpostData$Year), FUN = mean)

# In kết quả
print(average_word_count_by_year)

##   Group.1      x
## 1 2012 265.8417
## 2 2013 329.7812
## 3 2014 113.8800
## 4 2015  2.0000
## 5 2016 166.9335
## 6 2017 229.2465
## 7 2018 230.7468
## 8 2019 203.8336
## 9 2020 324.0312
## 10 2021 513.2422
## 11 2022 274.2088

```

4.2. Phân tích cụ thể từng cụm

- Phân tích cụm Positive
- Kiểm tra dữ liệu

```
# Kiểm tra data
dim(positive)

## [1] 1309   26

# Đếm missing values
sum(is.na(positive))

## [1] 10
```

Kiểm tra cụm positive có 10 dòng chứa giá trị NA trên tổng 1309 dòng; chiếm 0,007%. - Xử lý dữ liệu

```
# Tạo corpus cho cụm positive
positiveMessage <- positive$new_message
positiveCorpus <- Corpus(VectorSource(positiveMessage))

# Thêm thông tin năm cho positiveCorpus
meta(positiveCorpus, tag = "Create_Year") <- positive$Year
```

- Tạo TDM

```
# Tạo TDM cho cụm positive
positiveTDM <- TermDocumentMatrix(positiveCorpus, control =
list(wordLengths=c(1, Inf)))
positiveTDM <- removeSparseTerms(positiveTDM, 0.9)
inspect(positiveTDM)

## <<TermDocumentMatrix (terms: 102, documents: 1309)>>
## Non-/sparse entries: 24428/109090
## Sparsity           : 82%
## Maximal term length: 10
## Weighting          : term frequency (tf)
## Sample             :
##               Docs
## Terms    1225 1226 1228 1229 1230 1234 1236 1237 1238 1239
## áp        1     1     1     1     1     1     1     1     1     1
## combo     6     6     6     6     6     6     6     6     6     6
## dự       20    21    20    20    20    20    20    21    20    21
## dụng      1     1     1     1     1     1     1     1     1     1
## gà        0     0     0     0     0     0     0     0     0     0
## giải      16    16    16    16    16    16    16    16    16    16
## hàng      0     0     0     0     0     0     0     0     0     0
## hình      27    28    27    27    27    27    27    28    27    28
## pepsi     0     0     0     0     0     0     0     0     0     0
## thi       28    27    28    28    28    28    28    27    28    27
```

```

dim(positiveTDM)
## [1] 102 1309

```

Chọn ngưỡng tuần suất tối thiểu là 0.9, nhận được 102 từ đơn. Với đoạn mẫu hiển thị của Term Document Matrix, có thể nhận thấy các chữ đều có nghĩa, nhóm dự đoán các từ đơn này thuộc từ ghép và theo xu hướng, chủ đề nào đó.

- Xác định từ khóa phổ biến

```

# Xác định từ khóa phổ biến
term_freq <- rowSums(as.matrix(positiveTDM))
top_keywords <- head(sort(term_freq, decreasing = TRUE), 20)
print(top_keywords)

##      gà     thi    hàng   hình   combo   giải    dự   dụng    áp
pepsi    1471   1423   1321   1258   1130   1070   1029   1009   942
847
## burger     bình     giá   trình    tài   khoai    rán khuyễn   giao
comment   772    721    721    644    638    602    577    574    573
558

```

Danh sách 20 từ đơn được sử dụng nhiều nhất trong cụm positive đều là từ đơn có nghĩa.

- Kiểm tra độ tương quan giữa các từ đơn

```

# Lấy danh sách các từ khóa
keywords <- names(top_keywords)

# Tạo ma trận chứa thông tin về sự xuất hiện của từng từ khóa trong mỗi bài
viết positive
keyword_matrix <- sapply(keywords, function(keyword) grepl(keyword,
positive$new_message, ignore.case = TRUE))

# Tính ma trận tương quan giữa các từ khóa
correlation_matrix <- cor(keyword_matrix)

# In ma trận tương quan
print(correlation_matrix)

##                  gà          thi         hàng         hình         combo
## gà 1.000000000 -0.170378449  0.36286670 -0.14518992  0.42373906
## thi -0.17037845  1.000000000 -0.12078118  0.44832213 -0.02789680
## hàng 0.36286670 -0.120781178  1.00000000 -0.08865682  0.26818353
## hình -0.14518992  0.448322133 -0.08865682  1.00000000  0.01711339
## combo 0.42373906 -0.027896801  0.26818353  0.01711339  1.00000000
## giải -0.19522285  0.521895333 -0.19806580  0.48757512 -0.01712991
## dự  -0.16471432  0.525328372 -0.26260760  0.50390590  0.02016924
## dụng 0.51365860 -0.127714860  0.45349474 -0.06781277  0.50748087
## áp   0.48933483 -0.132936098  0.36643710 -0.04974448  0.47079027

```

## pepsi	0.53905821	-0.182798062	0.33887889	-0.11945962	0.57901285
## burger	0.44107569	-0.149592629	0.32762526	-0.05866353	0.48860823
## bình	-0.07886968	0.550766292	-0.01074695	0.40689203	0.06128086
## giá	0.19909098	0.169459833	0.17471177	0.20322322	0.25307214
## trình	0.32415775	-0.165627759	0.50288810	-0.18586755	0.23188582
## tài	0.51907682	-0.129009749	0.31948185	-0.11616593	0.50685697
## khoai	0.53662487	-0.145405463	0.29478413	-0.11282299	0.53045708
## rán	0.65226412	-0.161064717	0.28827179	-0.11012722	0.46103445
## khuyến	0.31066479	-0.004584332	0.32711543	-0.03229410	0.28493101
## giao	0.30906687	-0.215575428	0.69647585	-0.07530840	0.18197292
## comment	-0.15819212	0.287697070	-0.25132254	0.42883118	-0.04033738
	## giải	## dự	## dụng	## áp	## pepsi
burger					
## gà	-0.19522285	-0.16471432	0.513658597	0.48933483	0.5390582
0.44107569					
## thi	0.52189533	0.52532837	-0.127714860	-0.13293610	-0.1827981
0.14959263					
## hàng	-0.19806580	-0.26260760	0.453494739	0.36643710	0.3388789
0.32762526					
## hình	0.48757512	0.50390590	-0.067812772	-0.04974448	-0.1194596
0.05866353					
## combo	-0.01712991	0.02016924	0.507480867	0.47079027	0.5790129
0.48860823					
## giải	1.00000000	0.53858858	-0.139301163	-0.07395861	-0.2151588
0.16359641					
## dự	0.53858858	1.00000000	-0.127185593	-0.04151605	-0.1634433
0.13803848					
## dụng	-0.13930116	-0.12718559	1.000000000	0.88442416	0.4980790
0.41430831					
## áp	-0.07395861	-0.04151605	0.884424162	1.00000000	0.4602377
0.38042995					
## pepsi	-0.21515884	-0.16344331	0.498078991	0.46023769	1.0000000
0.58869312					
## burger	-0.16359641	-0.13803848	0.414308313	0.38042995	0.5886931
1.00000000					
## bình	0.54612918	0.40659812	0.005573819	-0.01001020	-0.1079771
0.09302016					
## giá	0.24604755	0.15278998	0.246325677	0.27847772	0.2470818
0.23619079					
## trình	-0.19275935	-0.18262913	0.454998647	0.40521244	0.2944203
0.25725383					
## tài	-0.19829187	-0.15649910	0.447030052	0.41036895	0.7119460
0.58719998					
## khoai	-0.19606368	-0.15411163	0.443300671	0.41388161	0.7287593
0.59941143					
## rán	-0.20592047	-0.18620071	0.457749223	0.42086246	0.5124753
0.47328411					
## khuyến	-0.03065575	-0.02921535	0.566354853	0.52800091	0.3194054
0.22607141					
## giao	-0.22283530	-0.23388318	0.406837650	0.32053338	0.3095620

```

0.28130978
## comment 0.51864135 0.61110762 -0.144619247 0.04251595 -0.2054538 -
0.09907037
##           bình      giá     trình     tài    khoai
rán
## gà      -0.078869675 0.19909098 0.32415775 0.51907682 0.53662487
0.6522641
## thi     0.550766292 0.16945983 -0.16562776 -0.12900975 -0.14540546 -
0.1610647
## hàng   -0.010746952 0.17471177 0.50288810 0.31948185 0.29478413
0.2882718
## hình   0.406892030 0.20322322 -0.18586755 -0.11616593 -0.11282299 -
0.1101272
## combo  0.061280858 0.25307214 0.23188582 0.50685697 0.53045708
0.4610344
## giải   0.546129179 0.24604755 -0.19275935 -0.19829187 -0.19606368 -
0.2059205
## dự     0.406598121 0.15278998 -0.18262913 -0.15649910 -0.15411163 -
0.1862007
## dụng  0.005573819 0.24632568 0.45499865 0.44703005 0.44330067
0.4577492
## áp     -0.010010201 0.27847772 0.40521244 0.41036895 0.41388161
0.4208625
## pepsi  -0.107977126 0.24708175 0.29442026 0.71194602 0.72875930
0.5124753
## burger -0.093020161 0.23619079 0.25725383 0.58719998 0.59941143
0.4732841
## bình  1.000000000 0.23310901 -0.08816638 -0.02325917 -0.09129437 -
0.1021589
## giá   0.233109013 1.00000000 0.02832588 0.22067110 0.22781341
0.1726308
## trình -0.088166375 0.02832588 1.00000000 0.28221366 0.29400987
0.2633579
## tài   -0.023259165 0.22067110 0.28221366 1.00000000 0.95377756
0.4999662
## khoai -0.091294369 0.22781341 0.29400987 0.95377756 1.00000000
0.5116592
## rán   -0.102158860 0.17263080 0.26335794 0.49996621 0.51165924
1.0000000
## khuyến 0.076487610 0.24375820 0.51836075 0.31154623 0.31194350
0.2964701
## giao   -0.095659207 0.04085649 0.55596699 0.28975701 0.26756761
0.2711376
## comment 0.303845061 0.20123798 -0.19126337 -0.18307571 -0.17529766 -
0.2017572
##           khuyến     giao     comment
## gà       0.310664790 0.30906687 -0.15819212
## thi     -0.004584332 -0.21557543 0.28769707
## hàng   0.327115432 0.69647585 -0.25132254
## hình   -0.032294099 -0.07530840 0.42883118

```

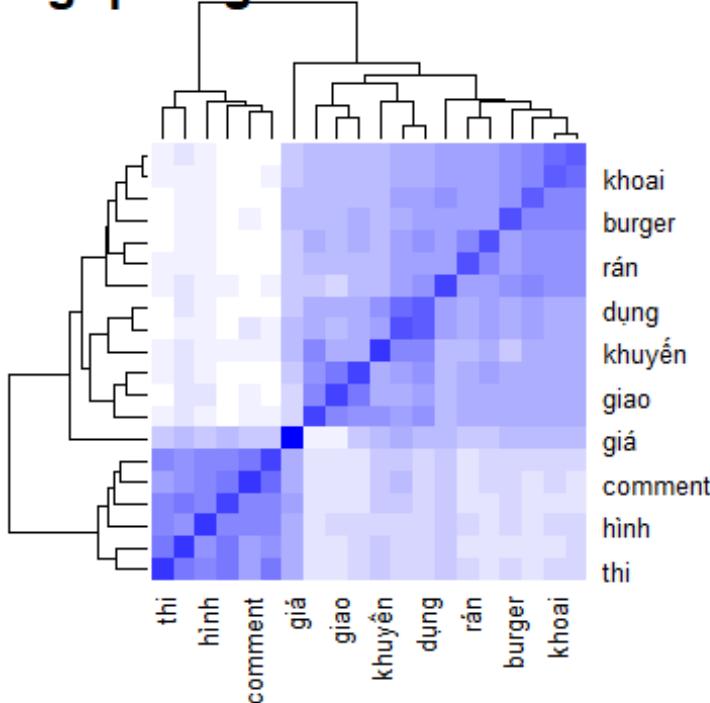
```

## combo      0.284931006  0.18197292 -0.04033738
## giải       -0.030655747 -0.22283530  0.51864135
## dự        -0.029215350 -0.23388318  0.61110762
## dụng      0.566354853  0.40683765 -0.14461925
## áp         0.528000905  0.32053338  0.04251595
## pepsi     0.319405375  0.30956198 -0.20545378
## burger    0.226071406  0.28130978 -0.09907037
## bình      0.076487610 -0.09565921  0.30384506
## giá        0.243758197  0.04085649  0.20123798
## trình     0.518360753  0.55596699 -0.19126337
## tài        0.311546226  0.28975701 -0.18307571
## khoai      0.311943503  0.26756761 -0.17529766
## rán        0.296470099  0.27113762 -0.20175715
## khuyến   1.000000000  0.32695415 -0.03266365
## giao       0.326954147  1.00000000 -0.20224474
## comment   -0.032663649 -0.20224474  1.00000000

# Vẽ biểu đồ nhiệt tương quan
heatmap(correlation_matrix,
        col = colorRampPalette(c("white", "blue"))(20),
        main = "Biểu đồ nhiệt tương quan giữa các từ khóa trong bài viết
positive")

```

Biểu đồ nhiệt tương quan giữa các từ khóa trong bài viết



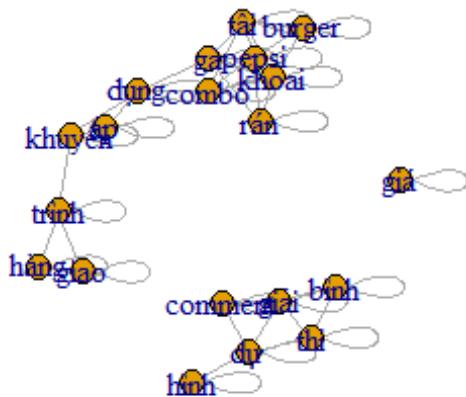
```

# Vẽ biểu đồ mạng các từ khóa có giá trị tương quan hơn 0.5
library(igraph)
keyword_network <- graph.adjacency(correlation_matrix > 0.5, mode =
"undirected")

```

```
plot(keyword_network, main = "Biểu Đồ Mạng Tương Quan Từ Khóa trong Bài Viết Positive")
```

Biểu Đồ Mạng Tương Quan Từ Khóa trong Bài Viết Positive

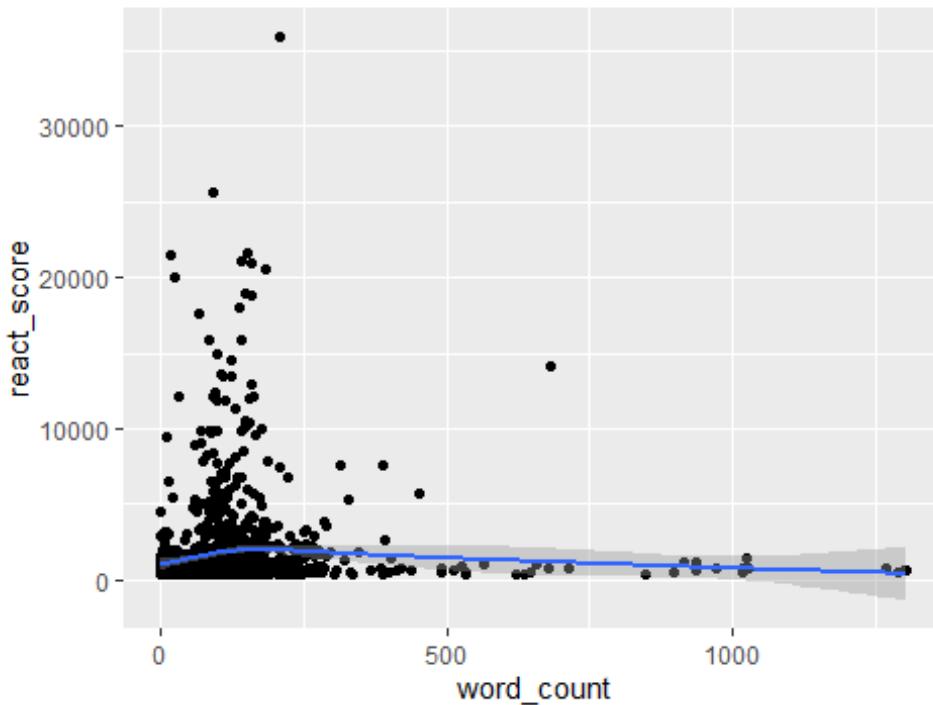


Ma trận tương quan thể hiện giá trị tương quan của 20 từ khóa nổi bật, để trực quan rõ hơn, nhóm sử dụng biểu đồ nhiệt thể hiện sử tương quan cho ma trận 20 từ khóa và biểu đồ mang tương quan để trực quan những mối quan hệ tương quan lớn hơn 0,5. Đối với biểu đồ nhiệt, có thể dễ nhìn rõ hơn rằng 20 từ đơn này có liên quan nhiều đến nhau vì có rất nhiều cụm màu đậm và ít cụm màu trắng. Biểu đồ mạng tương quan thể hiện sâu hơn các cụm chữ thường đi liền với nhau, cho thấy 2 chủ đề mà bài viết có lượng tương tác cao: Chương trình khuyến mãi (với từ khóa là: combo, khoai tây, gà rán, pepsi, chương trình, khuyến mãi, áp dụng, giao hàng...) và Cuộc thi (với từ khóa là: dự thi, giải, hình, comment, bình luận,...).

- Kiểm tra về lượng tương tác tích cực theo thời gian và độ dài caption

```
# Kiểm tra tương tác tích cực theo độ dài câu
positive$word_count <- sapply(strsplit(positive$message, "\\s+"), length)
ggplot(positive, aes(x = word_count, y = react_score)) +
  geom_point() +
  geom_smooth() +
  labs(title = "Tương tác tích cực theo độ dài câu")
## `geom_smooth()` using method = 'gam' and formula = 'y ~ s(x, bs = "cs")'
```

Tương tác tích cực theo độ dài câu



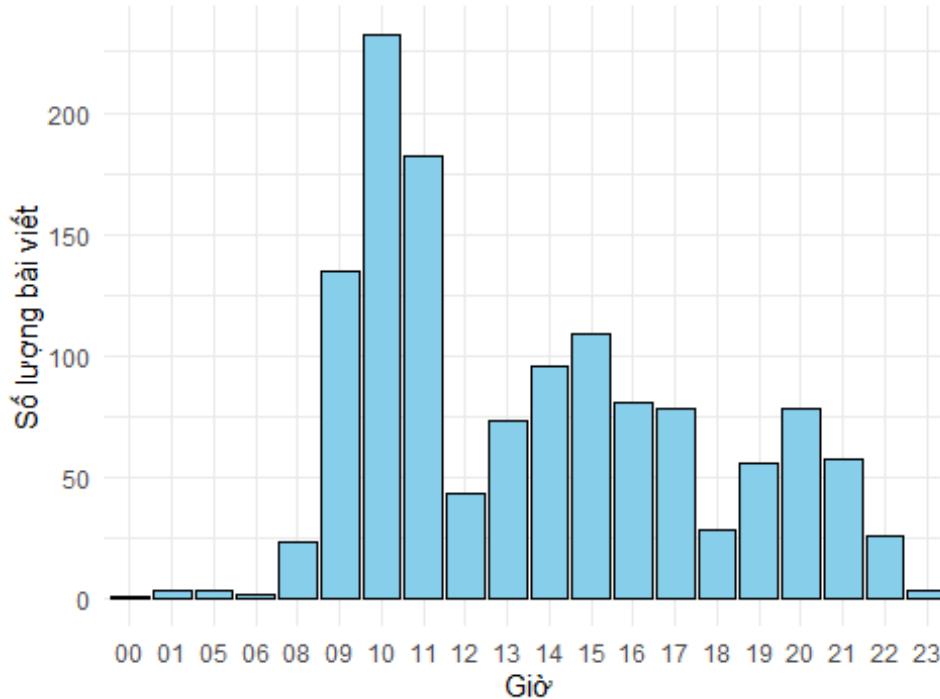
Độ dài câu thường từ dưới 500 từ đơn, và các bài viết được tương tác tích cực nằm phần lớn ở khoảng này.

- Kiểm tra thời gian fanpage đăng các bài viết được tương tác tích cực

```
# Tạo biến thời gian  
positive$hour <- format(positive$create_time_VN, format = "%H")
```

```
# Tạo biểu đồ  
ggplot(positive, aes(x = hour)) +  
  geom_bar(fill = "skyblue", color = "black") +  
  labs(title = "Phân phối thời gian đăng bài tích cực",  
       x = "Giờ",  
       y = "Số lượng bài viết") +  
  theme_minimal()
```

Phân phối thời gian đăng bài tích cực



Các bài viết nhận được tương tác cao thường được đăng tải vào 2 khung giờ: 9 giờ đến 11 giờ sáng, 14 giờ đến 16 giờ chiều. Đây là 2 khung giờ trước bữa trưa và bữa tối, thích hợp để đăng các bài truyền thông khuyến mãi, combo, giao hàng, tương thích với những từ khóa đã phân tích trước đó. Đặc biệt, khung 10 giờ đến 11 giờ có số lượng bài đăng cao nhất, có thể vì khung giờ này phù hợp để truyền thông cho các bộ phận dân văn phòng, sinh viên dùng bữa trưa bên ngoài nhiều hơn so với bữa tối. Ngoài ra, các khung giờ 19 giờ đến 21 giờ cũng có số lượng bài đăng tương đối, vì đây là thời gian nghỉ cuối ngày, và cũng dẫn đến khung giờ hiếm khi đăng bài trên fanpage là từ 23 giờ đến 7 giờ sáng.

Phần cụm Neutral

Phần preprocessing

```
# Biến dữ liệu thành Corpus
neutralMessage <- neutral$new_message
neutralCorpus <- Corpus(VectorSource(neutralMessage))

# Thêm metadata
meta(neutralCorpus, tag = "id") <- neutral$post_id
meta(neutralCorpus, tag = "create_Year") <- neutral$Year
meta(neutralCorpus, tag = "status") <- neutral$status
meta(neutralCorpus, tag = "React_score") <- neutral$react_score
meta(neutralCorpus, tag = "comment_score") <- neutral$comment_score

# Transform dữ liệu
neutralCorpus <- tm_map(neutralCorpus, content_transformer(tolower)) # Viết
```

thường

```
neutralCorpus <- tm_map(neutralCorpus, removePunctuation) # Bỏ dấu
neutralCorpus <- tm_map(neutralCorpus, removeNumbers) # Bỏ những con số
neutralCorpus <- tm_map(neutralCorpus, removeWords, vietnamese_stopwords) # Bỏ stopwords
neutralCorpus <- tm_map(neutralCorpus, stripWhitespace) # Bỏ khoảng trắng
neutralCorpus <- tm_map(neutralCorpus, stemDocument) # Stemming
```

Tạo TDM

```
# Create the Term Document Matrix
neutralTDM <- TermDocumentMatrix(neutralCorpus, control = list(wordLengths =
c(1, Inf)))
neutralTDM <- removeSparseTerms(neutralTDM, 0.9)

# Examine the TDM
inspect(neutralTDM)

## <<TermDocumentMatrix (terms: 60, documents: 2647)>>
## Non-/sparse entries: 25847/132973
## Sparsity : 84%
## Maximal term length: 7
## Weighting : term frequency (tf)
## Sample :
##          Docs
## Terms    105 107 354 360 362 370 422 425 436 530
## áp      3   2   1   1   1   3   3   3   4
## burger  1   3   4   4   4   4   6   6   6   1
## combo   5   3   4   4   5   5   9   9   8   7
## dụng   3   2   1   1   1   3   3   3   5
## gà     3   6   9   9   9   9   4   4   4   6
## hàng   2   3   4   4   4   4   3   3   2   6
## k      0   0   4   4   3   3   3   3   3   4
## khoai  3   2   4   4   4   4   5   5   5   1
## pepsi  3   2   4   4   4   4   4   4   4   3
## tai    6   5   3   3   3   3   4   4   4   1
```

Insight

Với độ sparse tầm 80%, kết quả cho thấy sự phân bổ của các từ khá là đều, không có từ nào quá nổi bật và đa số được dùng trong mọi bài đăng. Từ “gà”, “hàng”, “giao” xuất hiện với tần suất cao. Cho thấy các từ được sử dụng nhiều lần trong một bài đăng cụ thể.

Tìm những từ phổ biến nhất

```
# Find the most frequent terms
term_freq <- rowSums(as.matrix(neutralTDM))
term_freq <- sort(term_freq, decreasing = TRUE)
head(term_freq, 20) # Display the top 20 most frequent terms

##     gà    hàng      k    pepsi    dụng    burger    combo      áp    tai    khoai
trình
```

##	1991	1840	1395	1242	1157	1142	1113	1010	966	938
860										
##	vui	giá	m	giao	d	quy	đi	yêu	dịch	
##	839	827	816	808	773	750	741	690	683	

Insight

Đúng với nhận xét trên, các từ được sử dụng nhiều nhất là ‘gà’, ‘hang’,... Tuy nhiên, có vài từ không có nghĩa xuất hiện ở trên kết quả như ‘k’, ‘d’, ‘m’; Hiện tượng này xảy ra là do trong quá trình tiền xử lý dữ liệu đã bỏ mất context. Ví dụ giá tiền thường được ghi là ‘39k’, ‘4000d’, trong quá trình tiền xử lý đã bỏ mất những con số bỏ lại các ký tự. Có thể xem các ký tự là giá tiền của một sản phẩm nào đó được giới thiệu trong bài đăng.

Phân tích wordcloud

```
wordcloud(names(term_freq), term_freq, scale = c(4, 0.5), max.words = 100,  
random.order = FALSE, colors = brewer.pal(8, "Dark2"))
```



Insight Sau khi phân tích wordcloud, nhóm nhận thấy được các từ liên quan đến món ăn như là 'gà', 'khoai tây', 'burger' được sử dụng nhiều, các từ liên quan đến chương trình khuyến mãi như 'combo', 'áp - dụng' cũng được dùng khá là nhiều. Điều này cho thấy rằng Lotteria thường xuyên đăng bài về các deals, khuyến mãi, giới thiệu sản phẩm mới để tăng doanh thu.

Phân tích thể loại bài đăng theo năm

```
# Aggregate data by year and status, counting the number of posts
yearly_status_counts <- fanpost %>%
  group_by(Year, status) %>%
```

```

summarise(count = n(), .groups = 'drop')

# Calculate the total posts per year
yearly_totals <- yearly_status_counts %>%
  group_by(Year) %>%
  summarise(total = sum(count))

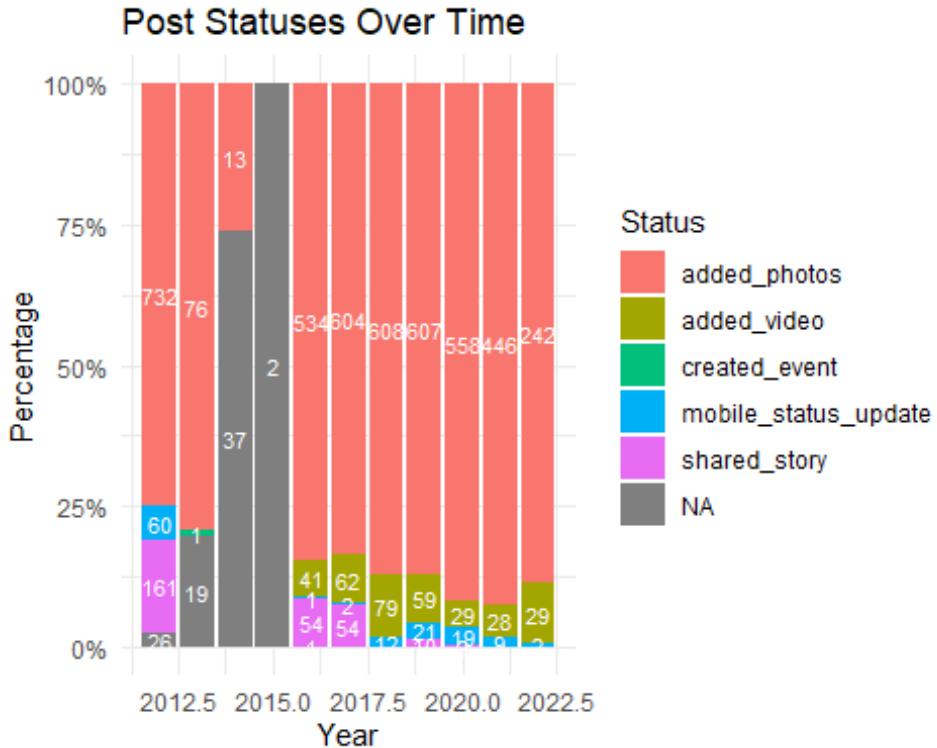
# Join the totals back to the original aggregated data
yearly_status_percentages <- yearly_status_counts %>%
  left_join(yearly_totals, by = "Year") %>%
  mutate(percentage = (count / total) * 100)

# Plot the data, showing the percentage of each status per year
# Plot with annotations
# Base plot
p <- ggplot(yearly_status_percentages, aes(x = Year, y = percentage)) +
  geom_bar(aes(fill = status), stat = "identity", position = "fill") + # Stacked percentage bars
  geom_text(aes(label = count, group = status), position =
  position_fill(vjust = 0.5), size = 3, color = "white") + # Add count annotations within bars
  scale_y_continuous(labels = percent_format()) + # Format y-axis as percentage
  labs(title = "Post Statuses Over Time",
       x = "Year",
       y = "Percentage",
       fill = "Status") +
  theme_minimal()

# Add total posts annotations at the top
p <- p + geom_text(data = yearly_totals, aes(x = Year, label = total), y =
100, vjust = 0, size = 3, color = 'red')

# Display the plot
print(p)

```



Insight

Theo xu hướng ở trên, cụm neutral cũng theo xu hướng như thế với số lượng cao vào năm 2012, thấp hoặc gần như không có vào các năm tiếp theo, và ổn định vào năm 2016 trở đi. Kết quả trên thể hiện rằng tuy có sự thay đổi trong số lượng bài đăng nhưng tỷ lệ các loại bài đăng khá là đều với tỷ lệ bài đăng kèm ảnh chiếm đa số trừ trường hợp vào 2012, 2013. Điều này cho thấy rằng xu hướng các bài đăng vẫn giữ ổn định qua các thời kỳ.

Phân tích điểm tương tác và thể loại bài đăng

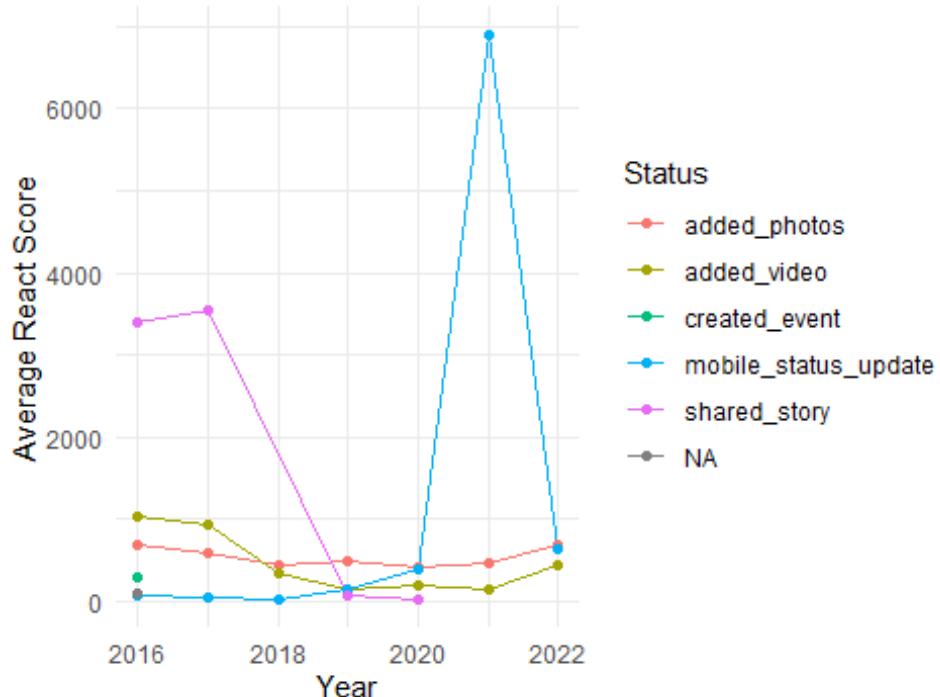
```
# Filter fanpost to include only entries from 2016 and beyond
fanpost_filtered <- fanpost %>%
  filter(Year >= 2016)

status_scores_yearly <- fanpost_filtered %>%
  group_by(Year, status) %>%
  summarise(average_react_score = mean(react_score, na.rm = TRUE),
            average_comment_score = mean(comment_score, na.rm = TRUE),
            .groups = 'drop')

ggplot(status_scores_yearly, aes(x = Year, y = average_react_score, color = status, group = status)) +
  geom_line() +
  geom_point() +
  labs(title = "Average React Score by Status Over Time",
       x = "Year",
       y = "Average React Score",
```

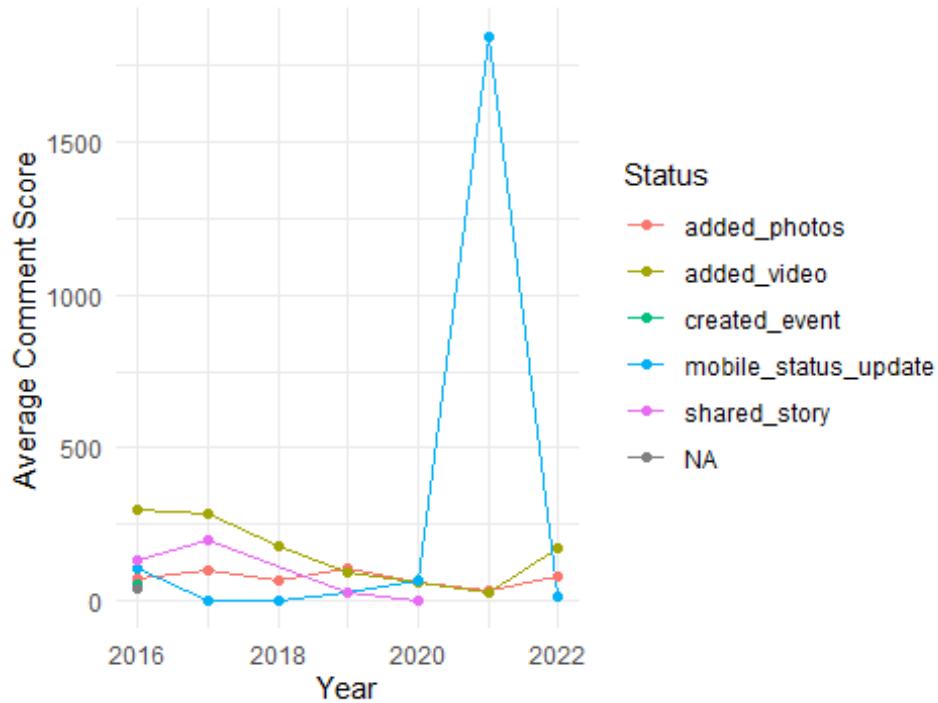
```
color = "Status") +  
theme_minimal()
```

Average React Score by Status Over Time



```
ggplot(status_scores_yearly, aes(x = Year, y = average_comment_score, color =  
status, group = status)) +  
geom_line() +  
geom_point() +  
labs(title = "Average Comment Score by Status Over Time",  
x = "Year",  
y = "Average Comment Score",  
color = "Status") +  
theme_minimal()
```

Average Comment Score by Status Over Time



Insight

Khi nhìn kết quả, có thể thấy được các điểm tương tác và comment khá ổn định ngoại trừ một số outlier làm biến đổi số lượng. Nhìn chung thì thấy được loại bài có hình ảnh và video nhận được sự tương tác trung bình cao hơn và có độ ổn định hơn

Phân tích từ phổ biến theo thời gian

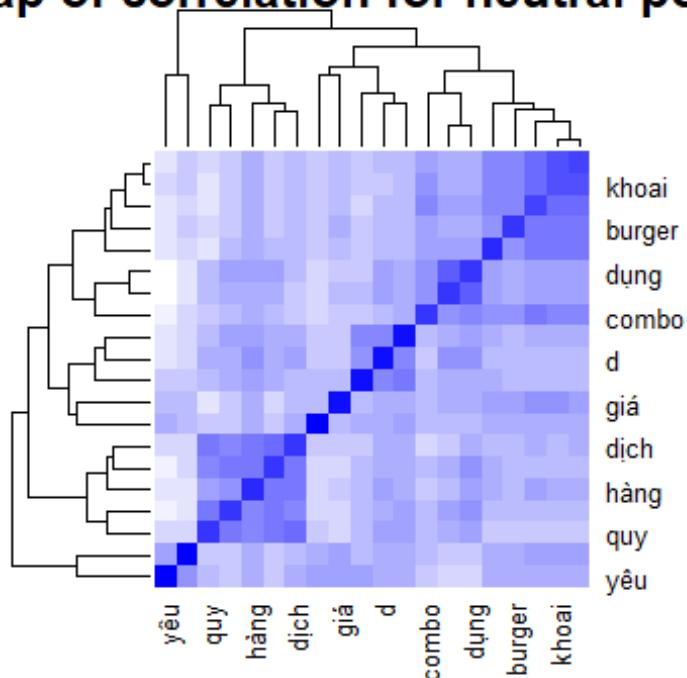
```
term_freq <- rowSums(as.matrix(neutralTDM))
top_keywords <- head(sort(term_freq, decreasing = TRUE), 20)
# Lấy danh sách các từ khóa
keywords <- names(top_keywords)

# Tạo ma trận chứa thông tin về sự xuất hiện của từng từ khóa trong mỗi bài viết
keyword_matrix <- sapply(keywords, function(keyword) grep1(keyword,
neutral$new_message, ignore.case = TRUE))

# Tính ma trận tương quan giữa các từ khóa
correlation_matrix <- cor(keyword_matrix)

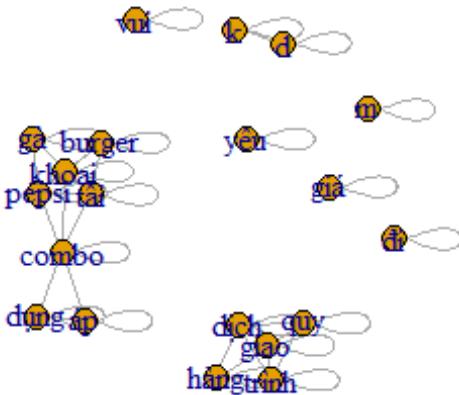
# Vẽ biểu đồ nhiệt tương quan
heatmap(correlation_matrix,
col = colorRampPalette(c("white", "blue"))(20),
main = "Heatmap of correlation for neutral post words")
```

Heatmap of correlation for neutral post words



```
# Vẽ biểu đồ mạng các từ khóa có giá trị tương quan hơn 0.5
keyword_network <- graph.adjacency(correlation_matrix > 0.5, mode =
"undirected")
plot(keyword_network, main = "Network diagram for Neutral post words")
```

Network diagram for Neutral post words



#Cụm Negative

head(negative)

```
##   STT      create_time          post_id
## 1   1 2022-07-25 07:49:35 350006241681991_6080276518654906
## 2   2 2022-07-25 06:27:56 350006241681991_6080149462000945
## 3   3 2022-07-24 10:00:03 350006241681991_6077875762228315
## 4   4 2022-07-24 03:00:05 350006241681991_6077145425634682
## 6   6 2022-07-22 08:00:24 350006241681991_6072339619448596
## 8   8 2022-07-21 10:43:57 350006241681991_6069833113032580
##
message
## 1                               Nay ban ơi, đừng quên chiếc hẹn "Thanh Xuân" tại
Lotteria với chúng tớ nha 😊\n\n🔥 Đặt hàng qua App LOTTERIA VN và nhận ngay
1 miếng Gà rán miễn phí cho đơn đầu tiên nhé Fans ơi.\nTải ngay tại 👉
https://onelink.to/c3jzpb\n#LotteriaVietnam #MyFAVORITELOTTERIA\n-----\nLưu ý:\n(*) Chương trình khuyến mãi có thể không được áp dụng tại một số cửa
hàng, mong quý khách thông cảm\n(*) Mau mau đặt hàng qua hotline 1900.6778 để
được giao hàng miễn phí và nhận nhiều ưu đãi hấp dẫn hơn nhé!!!
## 2                               Các fan Lotteria ơi, chốt đơn ngay, nhận quà liền
tay nha!!! 😊\nMiễn phí 01 Gà rán* cho đơn hàng đầu tiên qua App LOTTERIA VN
nhà các fan ơi!!!\nMiễn phí giao hàng cùng rất nhiều ưu đãi hấp dẫn, tải App
nhanh bằng cách quét mã QR tại hình ảnh đi kèm hoặc nhấp chọn các đường link
bên dưới 👇👇\niOS:
https://apps.apple.com/vn/app/lotteria/id1423528065?l=vi
\nAndroid:
https://play.google.com/store/apps/details...
\n*Áp dụng cho hoá đơn từ
99.000đ.\n#LotteriaVietnam #MyFAVORITELOTTERIA
```

3 Em chỉ có thẻ "say yes" mà thôi vì em quá u mê Lotteria mất ày 😊❤️💍 \n\n🔥 Đặt hàng qua App LOTTERIA VN và nhận ngay 1 miếng Gà rán miễn phí cho đơn đầu tiên nhé Fans ơi.\nTải ngay tại👉 <https://onelink.to/c3jzpb> \n#LotteriaVietnam #MyFAVORITELOTTERIA \n----\nLưu ý:\n(*) Chương trình khuyến mãi có thể không được áp dụng tại một số cửa hàng, mong quý khách thông cảm\n(*) Mau mau đặt hàng qua hotline 1900.6778 để được giao hàng miễn phí và nhận nhiều ưu đãi hấp dẫn hơn nhé!!!

4 Bạn thân ơi, hãy vui lên nào, deal hời siêu hot tải app nhận quà ngay!!! \n\n🔥 Miễn phí 01 Gà rán* cho đơn hàng đầu tiên qua App LOTTERIA VN nha các fan ơi!!!!\n★ Miễn phí giao hàng cùng rất nhiều ưu đãi hấp dẫn, tải App nhanh bằng cách quét mã QR tại hình ảnh đi kèm hoặc nhấp chọn các đường link bên dưới 👇👇👇 \niOS: <https://apps.apple.com/vn/app/lotteria/id1423528065?l=vi> \nAndroid: <https://play.google.com/store/apps/details...> \n*Áp dụng cho hoá đơn từ 99.000đ. \n#LotteriaVietnam #MyFAVORITELOTTERIA

6 Thì ra lăng mạn không phải là mùa hè, \nlăng mạn là mùa hè được ăn gà rán Lotteria 🍗 \n\n🔥 Đặt hàng qua App LOTTERIA VN và nhận ngay 1 miếng Gà rán miễn phí cho đơn đầu tiên nhé Fans ơi.\nTải ngay tại👉 <https://onelink.to/c3jzpb> \n#LotteriaVietnam #MyFAVORITELOTTERIA \n----\nLưu ý:\n(*) Chương trình khuyến mãi có thể không được áp dụng tại một số cửa hàng, mong quý khách thông cảm\n(*) Mau mau đặt hàng qua hotline 1900.6778 để được giao hàng miễn phí và nhận nhiều ưu đãi hấp dẫn hơn nhé!!!

8 Hic, đói bụng quá rùi nà T.T Lotteria ơi giao gà đến nhanh nào 😊 \n\n🔥 Đặt hàng qua App LOTTERIA VN và nhận ngay 1 miếng Gà rán miễn phí cho đơn đầu tiên nhé Fans ơi.\nTải ngay tại👉 <https://onelink.to/c3jzpb> \n#LotteriaVietnam #MyFAVORITELOTTERIA \n----\nLưu ý:\n(*) Chương trình khuyến mãi có thể không được áp dụng tại một số cửa hàng, mong quý khách thông cảm\n(*) Mau mau đặt hàng qua hotline 1900.6778 để được giao hàng miễn phí và nhận nhiều ưu đãi hấp dẫn hơn nhé!!!

status comment share like love haha wow sad angry thankful pride care

	added_photos	1	0	27	5	0	0	0	0	0	0
## 1 added_photos	1	0	27	5	0	0	0	0	0	0	0
## 2 added_photos	0	0	24	1	0	0	0	0	0	0	0
## 3 added_photos	2	1	24	0	1	0	0	0	0	0	0
## 4 added_photos	1	3	21	1	0	0	0	0	0	0	0
## 6 added_photos	11	1	34	2	0	0	0	0	0	0	0
## 8 added_video	9	0	29	3	0	0	0	0	0	0	0

	create_time_VN	Year	Month	Day	Weekday	Hour	react_score	comment_score
## 1	2022-07-25 14:49:35	2022	Jul	25	Mon	14	32	1
## 2	2022-07-25 13:27:56	2022	Jul	25	Mon	13	25	0

```
## 3 2022-07-24 17:00:03 2022 Jul 24 Sun 17 25
3
## 4 2022-07-24 10:00:05 2022 Jul 24 Sun 10 22
4
## 6 2022-07-22 15:00:24 2022 Jul 22 Fri 15 36
12
## 8 2022-07-21 17:43:57 2022 Jul 21 Thu 17 32
9
## total
## 1 33
## 2 25
## 3 28
## 4 26
## 6 48
## 8 41
##
new_message
## 1
đừng quên hẹn
xuân nha 🐶 🐶 hàng app vn miễn phí đơn fan tải 🐶
https://onelinktocjzpb.vietnam-myfavorite.luu/chuong-trinh/khuyen-mai-ap-dung-mong-quy-thong-cam-mau-mau-hang-hotline-giao-hang-mien-phi-uu-dai-hap
## 2
fan chốt đơn quà liền nha 🛍 miễn phí gà rán đơn hàng
app vn nha fan miễn phí giao hàng ưu đãi hấp tải app quét mã qr hình ảnh đi kèm nhấp đường link 🛍 io https://play.google.com/store/apps/detail áp dụng hoá đơn đ vietnam myfavorite
## 3
thể say yes u mê
ò y 🐦 🐦 ❤️ 🐦 🐦 🐦 hàng app vn miễn phí đơn fan tải 🐦
https://onelinktocjzpb.vietnam-myfavorite.luu/chuong-trinh/khuyen-mai-ap-dung-mong-quy-thong-cam-mau-mau-hang-hotline-giao-hang-mien-phi-uu-dai-hap
## 4
thân vui deal hời siêu hot tải app quà 🎁 miễn phí gà rán đơn hàng app
vn nha fan ⭐ miễn phí giao hàng ưu đãi hấp tải app quét mã qr hình ảnh đi kèm nhấp đường link 🎁 io https://play.google.com/store/apps/detail áp dụng hoá đơn đ vietnam myfavorite
## 6
lãng mạn hè lãng mạn mùa
hè gà rán 🎉 🎉 hàng app vn miễn phí đơn fan tải 🎉
https://onelinktocjzpb.vietnam-myfavorite.luu/chuong-trinh/khuyen-mai-ap-dung-mong-quy-thong-cam-mau-mau-hang-hotline-giao-hang-mien-phi-uu-dai-hap
## 8
hic đói bụng rùi nà tt
giao gà 🎉 🎉 hàng app vn miễn phí đơn fan tải 🎉
https://onelinktocjzpb.vietnam-myfavorite.luu/chuong-trinh/khuyen-mai-ap-dung-mong-quy-thong-cam-mau-mau-hang-hotline-giao-hang-mien-phi-uu-dai-hap
nrow(negative)
## [1] 1285
```

Cụm Negative chứa tổng cộng 1285 dòng. Vì đây là cụm có tổng điểm tương tác (react_score) thấp nhất, tức khả năng cao đây là cụm chứa những bài đăng có nội dung tiêu cực nên mới nhận lại những phản ứng như sad, angry (mang dấu âm trong công thức tính).

điểm). Tuy nhiên đây mới là suy đoán ban đầu nên nhóm sẽ tiến hành phân tích đặc điểm các bài đăng của cụm này để hiểu rõ lý do thật sự dẫn đến tổng điểm tương tác thấp.

```
# Tạo corpus cho cụm positive
NegativeMessage <- negative$new_message
NegativeCorpus <- Corpus(VectorSource(NegativeMessage))

# Thêm metadata
meta(NegativeCorpus, tag = "id") <- negative$post_id
meta(NegativeCorpus, tag = "create_Year") <- negative$Year
meta(NegativeCorpus, tag = "status") <- negative$status
meta(NegativeCorpus, tag = "React_score") <- negative$react_score
meta(NegativeCorpus, tag = "comment_score") <- negative$comment_score

# Transform dữ liệu
NegativeCorpus <- tm_map(NegativeCorpus, content_transformer(tolower)) # Viết thường

## Warning in tm_map.SimpleCorpus(NegativeCorpus,
content_transformer(tolower)):
## transformation drops documents

NegativeCorpus <- tm_map(NegativeCorpus, removePunctuation) # Bỏ dấu

## Warning in tm_map.SimpleCorpus(NegativeCorpus, removePunctuation):
## transformation drops documents

NegativeCorpus <- tm_map(NegativeCorpus, removeNumbers) # Bỏ những con số

## Warning in tm_map.SimpleCorpus(NegativeCorpus, removeNumbers):
transformation
## drops documents

NegativeCorpus <- tm_map(NegativeCorpus, removeWords, vietnamese_stopwords) # Bỏ stopwords

## Warning in tm_map.SimpleCorpus(NegativeCorpus, removeWords,
## vietnamese_stopwords): transformation drops documents

NegativeCorpus <- tm_map(NegativeCorpus, stripWhitespace) # Bỏ khoảng trắng

## Warning in tm_map.SimpleCorpus(NegativeCorpus, stripWhitespace):
transformation
## drops documents

NegativeCorpus <- tm_map(NegativeCorpus, stemDocument) # Stemming

## Warning in tm_map.SimpleCorpus(NegativeCorpus, stemDocument):
transformation
## drops documents
```

Tạo TDM

```
# Create the Term Document Matrix
negativeTDM <- TermDocumentMatrix(neutralCorpus, control = list(wordLengths =
c(1, Inf)))
negativeTDM <- removeSparseTerms(negativeTDM, 0.9)

# Examine the TDM
inspect(negativeTDM)

## <<TermDocumentMatrix (terms: 60, documents: 2647)>>
## Non-/sparse entries: 25847/132973
## Sparsity : 84%
## Maximal term length: 7
## Weighting : term frequency (tf)
## Sample :
##          Docs
## Terms    105 107 354 360 362 370 422 425 436 530
## áp      3    2    1    1    1    1    3    3    3    4
## burger  1    3    4    4    4    4    6    6    6    1
## combo   5    3    4    4    5    5    9    9    8    7
## dung   3    2    1    1    1    1    3    3    3    5
## gà     3    6    9    9    9    9    4    4    4    6
## hàng   2    3    4    4    4    4    3    3    2    6
## k      0    0    4    4    3    3    3    3    3    4
## khoai  3    2    4    4    4    4    5    5    5    1
## pepsi  3    2    4    4    4    4    4    4    4    3
## tài    6    5    3    3    3    3    4    4    4    1
```

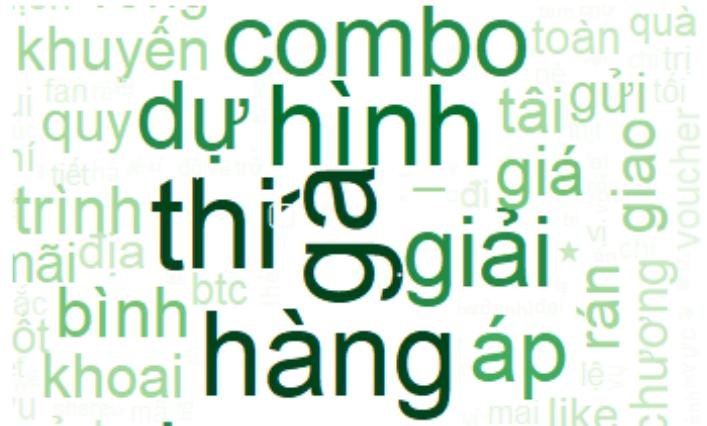
Wordcloud so sánh các từ xuất hiện thường xuyên của 3 cụm

```
# Tạo wordCloud cho negative
wordcloudNegative<-wordcloud( NegativeCorpus,min.freq = 10,colors =
brewer.pal(9, "Blues"), random.order = F)
```



```
title1 <- grid::textGrob("Negative Words", gp = grid::gpar(fontsize = 14,
fontface = "bold", col = "blue"))

# Tạo wordCloud cho positive
wordcloudPositive<-wordcloud(positiveCorpus,min.freq = 10,colors =
brewer.pal(9, "Greens"), random.order = F)
```



```
title2 <- grid::textGrob("Positive Words", gp = grid::gpar(fontsize = 14,  
fontface = "bold", col = "green"))  
  
# Tạo wordcloud cho neutral  
wordcloudNeutral<-wordcloud(neutralCorpus,min.freq = 10,colors =  
brewer.pal(9, "Reds"), random.order = F)
```

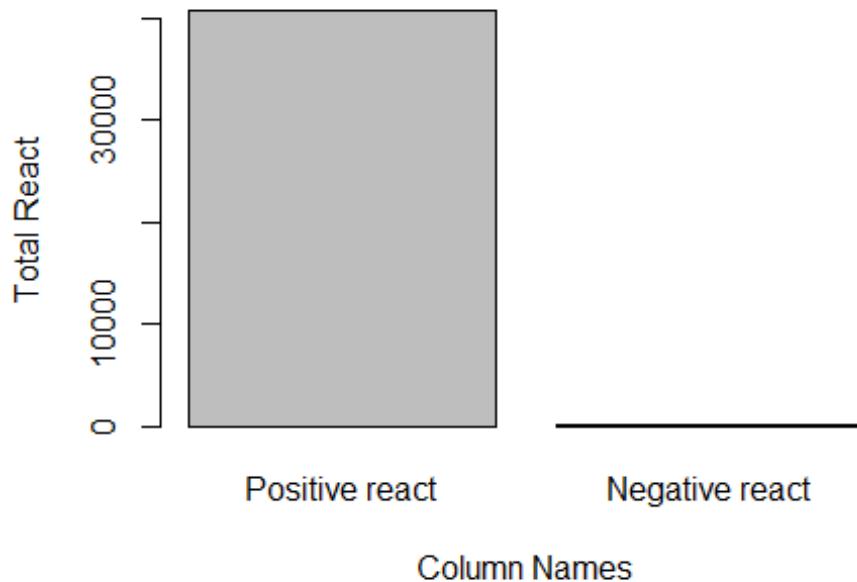


```
title3 <- grid::textGrob("Neutral Words", gp = grid::gpar(fontsize = 14, fontface = "bold", col = "red"))
```

Cả 3 wordcloud của 3 cụm Positive, Negative, Neutral đều cho thấy những từ tương tự nhau như “hàng”, “gà”, “trình”,... Vậy điều gì đã dẫn tới sự khác biệt trong số điểm tương tác của 3 cụm, và đặc điểm nào của nội dung bài đăng đã làm cho cụm Negative có số điểm thấp nhất? Nhóm sẽ bắt đầu bằng việc phân tích số lượng thành phần các loại react mà bài đăng trong cụm này nhận được.

```
#Vẽ đồ thị tính tổng react positive và negative của cụm Negative
sum_positive_react <- negative$like + negative$love + negative$haha +
negative$wow + negative$thankful + negative$pride + negative$care
sum_negative_react <- negative$sad + negative$angry
sum_positive_react <- sum(sum_positive_react)
sum_negative_react <- sum(sum_negative_react)
totalReact <- c(sum_positive_react,sum_negative_react)
#colnames(totalReact) <- c("Positive react", "Negative react")
lbls <- c("Positive react", "Negative react")
barplot(totalReact, names.arg = lbls, main="Bar Chart of React", ylab =
"Total React", xlab = "Column Names")
```

Bar Chart of React



Column Names

Đồ thị trên cho thấy rõ lượng tương tác tiêu cực rất ít có thể nói là không đáng kể so với lượng tương tác tích cực. Từ đây, có thể kết luận rằng các bài đăng nằm trong cụm này có điểm tổng điểm thấp là do có lượng tương tác thấp, không được người dùng đón nhận, quan tâm nhiều. Sau đây, nhóm sẽ phân tích kỹ hơn vào nội dung của các bài đăng trong cụm này.

```
#Chuyển negative thành Corpus
NegativeCorpus <- Corpus(VectorSource(negative$new_message))
# Tạo Term Document matrix
negativeTDM <- TermDocumentMatrix(NegativeCorpus, control =
list(wordLengths=c(1,Inf)))
inspect(negativeTDM)

## <<TermDocumentMatrix (terms: 4401, documents: 1285)>>
## Non-/sparse entries: 42920/5612365
## Sparsity : 99%
## Maximal term length: 93
## Weighting : term frequency (tf)
## Sample :
##          Docs
## Terms    1075 1096 1119 1140 1193 25 547 821 845 957
##   bé      0     0     0     1     0     0     7     0     1     0
##   bóng    1     0     0     0     0     0     2     0     0     0
##   công    1     1     3     0     2     2     1     3     0     1
##   gà      4     0     0     0     0     0     1     0     1     0
##   giải    18    0    12    23    0     0     2     2     0     15
##   hàng    5     0     0     0     0     1     1     0     1     0
##   tặng    3     0     0     0     0     0     7     0     0     0
```

```

## thi 0 2 11 4 0 0 6 6 0 31
## trình 5 0 0 0 1 1 0 0 0 0
## vui 1 0 2 1 0 1 1 3 0 1

# Tìm số lần xuất hiện của từng từ
freqNegative <- rowSums(as.matrix(negativeTDM))
# Tính số từ tương ứng với 2% Lớn nhất
top_2_percent <- length(freqNegative) * 0.02
# Sắp xếp và Lấy 2% từ xuất hiện nhiều nhất
top_words <- head(sort(freqNegative, decreasing = TRUE), n = top_2_percent)
# In ra danh sách các từ xuất hiện nhiều nhất
print(top_words)

## hàng giải vui tặng gà bé trình thi bóng
công 736 623 483 392 391 351 348 336 316
## 308
## dụng 301 301 300 270 270 262 261 259 254
## 253
## mǎn hình thường đồng đơn thành may yêu nhật giá
## 227 246 245 243 239 239 237 237 234 230
## nam viên đội thẻ giao thủ quản nha động địa
## 207 226 225 219 217 214 212 211 211 208
## burger cầu áp tiệc gửi tuần dự tài dịch pepsi
## 189 205 200 200 199 199 199 197 195 194
## nội 189 nội cup toàn đá chúc khoai lý thẻ liên
## hà 189 189 184 184 183 182 181 174 173
## 172 điện nhí hồ năng mua vụ lott voucher hệ nguyễn
## 155 169 164 164 163 160 158 157 156 156
## sơ đẹp quy khuyến học sức nhân bình việt mẹ
## 147 155 154 153 152 152 151 151 149 147
## tập phẩm lưu ảnh mùa hành tuyển combo
## 146 145 142 142 142 141 141 139 138

```

Tiếp theo, nhóm sẽ tìm những từ có khả năng cao xuất hiện cùng với 5 từ có lượt xuất hiện cao nhất để biết được những nội dung phổ biến trong cụm bài đăng này.

```

#Tìm những từ Liên quan
findAssocs(negativeTDM, 'hàng', .5)

## $hàng
##    tặng     giao     dịch      🎁     tích     💑danh 💑member
##    0.55     0.53     0.53     0.52     0.51     0.51     0.50

findAssocs(negativeTDM, 'giải', .5)

## $giải
##    khích     nhì     thường     cầu marketingvn
##    0.71     0.66     0.62     0.56     0.51

findAssocs(negativeTDM, 'gà', .5)

## $gà
##    chiên     pepsi     miếng     khoai     tài
##    0.68     0.67     0.66     0.66     0.61
##    teriyaki     fish     packag     🎉funni     🎉happi
##    0.57     0.57     0.57     0.57     0.57
##    🎉lucki     🎉sinh     thiệp     đùi httpsgooglpcz
##    0.57     0.57     0.56     0.56     0.55
##    nón     nhật     ôm     burger     nugget
##    0.55     0.53     0.53     0.51     0.51
##    gói     khổng     lồ     tiệc     thỏa
##    0.51     0.51     0.51     0.50     0.50
##    gói
##    0.50

findAssocs(negativeTDM, 'tặng', .5)

## $tặng
##    🎁     💑danh 💑member     💕     day     mǎn     may
##    tích
##    0.79     0.79     0.78     0.76     0.71     0.70     0.69
##    0.68
##    1point     đợt     lõ     lũi     kéo     ngẫu memberday
##    hàng
##    0.67     0.66     0.64     0.62     0.60     0.59     0.59
##    0.55
##    nhiên     khủng
##    0.55     0.51

findAssocs(negativeTDM, 'vui', .4)

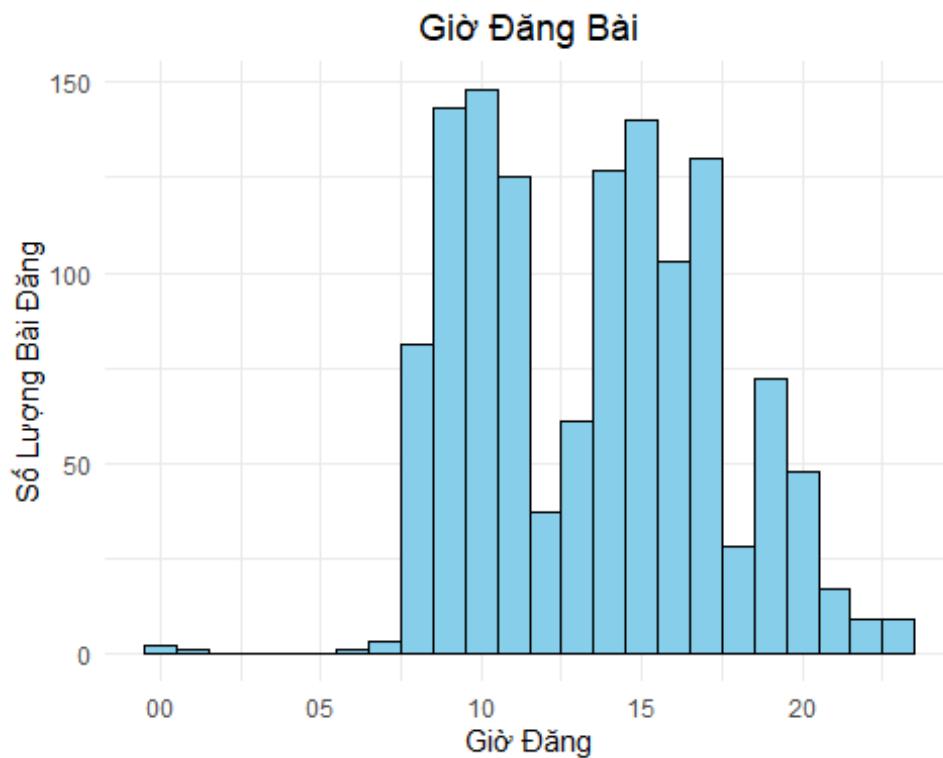
## $vui
##    nhộn     tiệc     sinh     trò     bé     thiệp     nón
##    0.50     0.43     0.43     0.43     0.41     0.41     0.40

```

Từ kết quả trên nhóm có thể ghép ra được những cụm từ cùng chủ đề thường xuyên xuất hiện cùng nhau với tần suất cao và suy đoán được những chủ đề của các bài đăng này: +

Hàng: tặng, giao, dịch , tích, danh,member Chủ đề liên quan: các chương trình khuyến mãi, ưu đãi, và các hoạt động liên quan đến việc tặng quà, giao hàng, giao dịch, tích điểm, đăng ký thành viên, hoặc các thông tin về danh sách thành viên. + Giải: khích, nhì, thưởng, cầu, marketingvn Chủ đề liên quan: công bố kết quả các minigame, các sự kiện thường dành cho khách hàng, thông báo cơ cấu giải thưởng các cuộc thi nhằm marketing cho Lotteria + Gà: chiên, pepsi, miếng, khoai, tây, teriyaki, fish, package, funny, happy, lucky, sinh, thiệp, đùi, nón, nhật, ôm, burger, nugget gói, khổng, lồ, tiệc, thỏa, gói. Chủ đề liên quan: giới thiệu các combo gồm các món bán chạy như gà chiên, gà nugget, burger gà, gà teriyaki, thông báo về các sự kiện, tiệc tùng, sinh nhật, mà mà khách hàng có thể tổ chức tại Lotteria. + Tặng: danh, member, day, mắn, may, tích, lpoint, đợt Chủ đề liên quan: các chương trình khuyến mãi, ưu đãi, và các sự kiện đặc biệt dành riêng cho thành viên của Lotteria vào những ngày đặc biệt hoặc các đợt khuyến mãi lớn như mùa lễ hội Noel,... + Vui: nhộn, tiệc, sinh, trò, bé, thiệp, nón Chủ đề liên quan: thông báo các sự kiện, hoạt động vui chơi, tặng quà khuyến mãi áp dụng dành riêng cho trẻ em tại Lotteria. Dựa vào các cụm từ xuất hiện thường xuyên, chúng ta có được cái nhìn tổng quan về nội dung phổ biến của các bài đăng trong cụm Negative. Chủ yếu là các thông báo về hoạt động mới, sự kiện, giới thiệu các chương trình khuyến mãi như combo, ưu đãi cho thành viên tích điểm và thông báo kết quả của các cuộc thi, minigame. Tuy nhiên, các chủ đề này thường không có điểm đặc biệt đủ để thu hút sự chú ý của người dùng. Trong các bài đăng, thường không có lời kêu gọi hành động (Call-to-action) cũng như phần thưởng kèm theo. Do đó, thông thường những chủ đề này sẽ không nhận được nhiều sự tương tác từ phía người dùng.

```
# Vẽ biểu đồ hiển thị giờ đăng của các bài
negative$Hour <- as.POSIXct(negative$Hour, format = "%H")
ggplot(negative, aes(x = Hour)) +
  geom_histogram(binwidth = 1, fill = "skyblue", color = "black") +
  labs(title = "Giờ Đăng Bài",
       x = "Giờ Đăng",
       y = "Số Lượng Bài Đăng") +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5))
```



Biểu đồ trên cho

thấy các bài đăng trong cụm Negative chủ yếu được đăng tải trong khoảng thời gian từ 8h – 11h và từ 14 – 17h. Điểm chung là cả hai khung giờ này đều nằm trong giờ hành chính, là khoảng thời gian đa số mọi người đều đi học và đi làm, không có thời gian lướt mạng xã hội. Đăng bài truyền thông trong khoảng thời gian này sẽ không có nhiều người xem, do đó, đây rất có thể là nguyên nhân dẫn đến lượt tương tác thấp của các bài đăng nằm trong cụm này. #Kết luận Ban đầu nhóm ghi nhận đây là cụm Negative tức là cụm gồm những bài đăng nghi ngờ có tính chất hoặc chứa từ ngữ có tính chất tiêu cực do điểm tương tác (react_score) của cụm là thấp nhất. Tuy nhiên, qua quá trình phân tích và so sánh các từ phổ biến của cụm với 2 cụm Positive và Neutral không nhận thấy có điểm gì bất thường, cũng như số lượng tương tác tiêu cực từ người dùng chiếm rất ít trong tổng lượng tương tác. Có thể kết luận rằng các bài đăng nằm trong cụm Negative không có tính chất tiêu cực hoặc gây tranh cãi, đây chỉ đơn thuần là những bài có nội dung chủ đề không đủ thu hút người dùng và được đăng vào những khung giờ nằm trong giờ hành chính nên mới có lượt tương tác thấp.

Kết luận

- Insight và Kết Luận: Từ những phân tích trên, nhóm rút ra một số kết luận như sau:
 Giai đoạn từ năm 2012 - 2016, lượng bài viết đăng tải tuy nhiều nhưng lượng tương tác lại thấp, có thể liên hệ nguyên do thực tế là do số lượng người dùng mạng xã hội còn ít và người dùng còn ít quan tâm về các bài viết quảng cáo trên facebook. Giai đoạn từ năm 2016 - 2022, lượng tương tác của fanpage tăng đáng kể, dù số bài đăng không tăng và thậm chí ít hơn giai đoạn trước. Có thể thấy từ năm 2016, fanpage của Lotteria đã phát triển hơn.

Dựa vào phần phân tích 3 cụm positive, negative và neutral, có thể thấy một số đặc điểm từ các bài đăng của Lotteria như: Dùng nhiều từ liên quan tới độ nhận diện thương hiệu: Lotteria, Lotteriavietnam, gà rán, cửa hàng, pepsi, khoai tây, giao hàng,...

Về chủ đề, các chủ đề thường được quan tâm nhiều là Chương trình khuyến mãi và các cuộc thi. Các từ khóa đi kèm là: combo, khuyến mãi, chương trình,... Có thể thấy người dùng quan tâm nhiều tới các phúc lợi từ các chương trình giảm giá, còn về chủ đề các cuộc thi có thể dự đoán nguyên do là những phần thưởng hấp dẫn và những yêu cầu trong cuộc thi thường là bắt buộc người tham gia các bước như like, share, comment nên có tương tác cao hơn. Ngược lại, các chủ đề ít được quan tâm thường là các bài viết thông báo như: thông báo giải thưởng cuộc thi, thông báo sự kiện, các hoạt động cho trẻ em.

Các khung giờ được tương tác cao nhất là từ 10 - 11 giờ sáng, có thể dự đoán do khung giờ này gần tới giờ ăn trưa của học sinh viên, dân văn phòng nên được tương tác cao, là khung giờ phù hợp cho mọi người lựa chọn và đặt đồ ăn. Các khung giờ khuya từ 19 giờ - 21 giờ thường ít được tương tác vì đây là thời gian nghỉ ngơi cuối ngày, và cũng dẫn đến khung giờ hiếm khi đăng bài trên fanpage là từ 23 giờ đến 7 giờ sáng. Còn lại các khung giờ khác thì lượng tương tác ở mức tương đối.

7. Đề Xuất Chiến Lược Cho Tương Lai Dựa vào những kết luận trên, nhóm đưa ra một số đề xuất chiến lược tối ưu hóa về độ phổ biến và tương tác như đa dạng hóa chủ đề bài viết như:
 - Ngoài các bài viết liên quan trực tiếp tới bán hàng như giới thiệu menu, chương trình khuyến mãi,... thì thêm các bài viết nhằm mục đích tăng tương tác với chủ đề bắt trend để dễ tiếp cận và hấp dẫn người dùng trẻ hơn. Tạo ra các loại nội dung đa dạng như video, hình ảnh, bài viết văn bản để phù hợp với sở thích và thói quen tiêu dùng của đối tượng mục tiêu.
 - Lập kế hoạch đăng bài viết vào các khung giờ mà người dùng thường online và hoạt động trên mạng xã hội, như vào giờ trưa và buổi tối sau khi kết thúc công việc: đăng bài về giới thiệu giảm giá, menu, giao hàng,... vào các khung giờ gần bữa ăn (10 - 11 giờ sáng, 16 - 18 giờ tối), các bài viết về giới thiệu sự kiện, cuộc thi vào các giờ nghỉ ngơi và sau bữa ăn (19 - 21 giờ), hạn chế đăng bài vào khung giờ khuya (0 - 7 giờ sáng).
 - Cải thiện các nội dung hình ảnh, video đa dạng tạo ra cảm xúc kích thích tương tác từ cộng đồng. Thúc đẩy sự tham gia của người dùng bằng cách tổ chức các cuộc thi, trò chơi hoặc thăm dò ý kiến để tạo ra môi trường năng động và thú vị.