**Executive Summary**

Using Twitter API, our primary objective is to determine how significant the Louisville shooting was to LinkedIn's sudden trending status. By incorporating the data retrieved into R, we'll break down how that data is applied to each analytical checkpoint and discuss our findings to provide accurate results. For the data and analysis part, we focused on English reviews and filtered the data to 7,752 tweets. And then we conducted a behavioral analysis to understand the distribution of posts, and the favorite distribution to see what types of posts get the most favors. We used topic analysis to find out whether the influence of shootings has decreased, and people focused more on the data leak on LinkedIn and tried to let the platform solve the issue. Overall, the sentiment analysis told us the overall sentiment for LinkedIn is positive but there is space for improvement.

**1. Introduction to the business problem/issue/campaign/event/brand/product and main research questions**

A mass shooting incident occurred in Nashville, Tennessee, leading to several casualties. The event quickly gained attention, and people turned to social media platforms to share their thoughts and feelings about the tragedy.

According to some Twitter users, the shooter had a LinkedIn profile who also revealed that they had altered their pronouns on the social media site following the shooting. We were curious about the talks around the shooting incident and the mention of LinkedIn on Twitter as a result.

The analysis' main goal was to learn more about how much and how differently the shooting incident was discussed on LinkedIn. We gathered information from tweets regarding LinkedIn and the shooting incident to do this. Additionally, we were interested in learning how individuals used LinkedIn and what features were discussed on Twitter.

We analyzed the gathered data and sought to answer several questions:

1. We wanted to know how people use LinkedIn.
2. To understand what people discussed about LinkedIn on Twitter.
3. We tried to determine the extent to which the shooting incident was talked about on LinkedIn and in what light.

In a nutshell, we wanted to obtain a more comprehensive understanding of how people use the platform and what topics they debate on Twitter by collecting tweets about LinkedIn.

**2. Data & Methodology**

The process of gathering data required a Twitter account. Using Melvin's account, we used Twitter API to accumulate 10,000 tweets that had "LinkedIn". After further analysis, we were able to retrieve a total of 9,954 tweets out of 10,000 within the timeframe of April 10-18th. After gathering our results, we saved it as a RDS file and created a new script for the raw data to begin the methodology process.

In terms of the methodology aspect, we utilized behavioral analysis, topic model, sentiment analysis, and world cloud. Let's go into detail with each analysis process.

## 3. Analysis & Findings

- **Behavioral analysis**
    - We look at the distribution of different posts by days and hours to see when did LinkedIn become trending.
    - We look at favorite counts and find the most heated post was a post talking about the pronounced issues and gender inequity problems behind it.
    - Then looking at retweets, quotes, and private and self-replies of the data to see the composition of LinkedIn's posts.
- **Topic model**
    - We first look at the shooting keywords and then find out there were 11 posts mentioning the shooting, so we consider it as a trivial influence.
    - Based on word clouds, we find 'petition' as an initiative word from the user side, and then using the word, we find the campaign "save the bee" mentioned the data leak and the irresponsible behavior that LinkedIn just deleted users' accounts.
    - The first topic is user experience, the second topic is the England banking recruitment issue, and the third one is "Save the bee".
- **Sentiment analysis**
    - We did the sentimental analysis based on the data and found out that the mean sentimental score of the data is 0.14 which is greater than 0, indicating a slightly positive tone overall. However, some of the posts and comments also criticized LinkedIn for allowing him to change his pronouns after the shooting; these posts' sentiment scores are below 0. The majority of the posts and comments expressed sympathy, condolences, and support for the victims and their families, as well as gratitude for the first responders and law enforcement officers.

## 4. Recommendations

As for all replies, there are 829 posts that were private replies and self replies which can show that LinkedIn still needs to improve on the activeness of its account to increase user engagement. LinkedIn can increase user engagement and activity by developing more user-friendly features, such as discussion forums, commenting tools, and social media integrations. Additionally, LinkedIn should encourage users to create and share content by recognizing and rewarding high-quality posts, articles, and remarks. In addition, LinkedIn can leverage its extensive user network to create online events and webinars that bring together like-minded individuals for networking and information sharing.

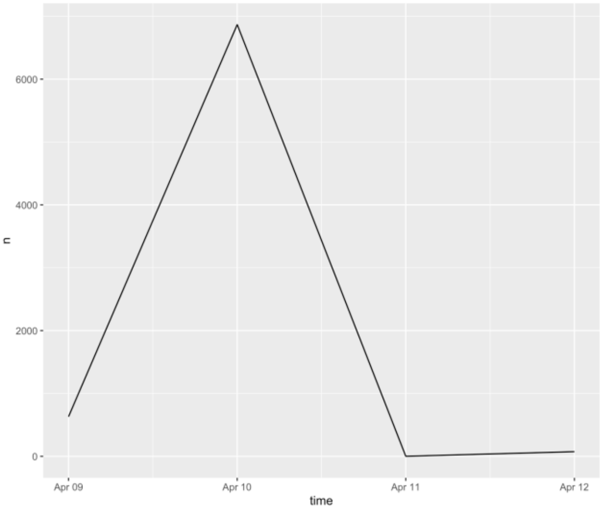
LinkedIn can establish a comprehensive diversity and inclusion policy that addresses gender equity issues and promotes a respectful and inclusive environment for all users to increase awareness of gender equity. LinkedIn should also provide its employees and users with training and resources to promote gender equality and sensitivity. In addition, LinkedIn can collaborate with organizations committed to gender equality to promote diversity and inclusion on the platform.

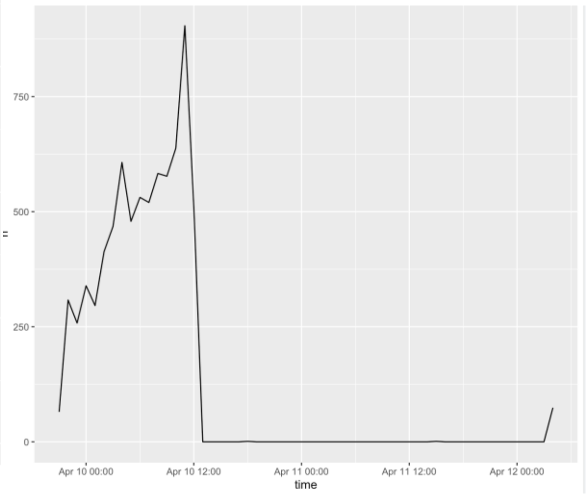
LinkedIn should prioritize data security and user privacy by investing in the most advanced security measures and performing routine security audits. LinkedIn should also give users greater control over their data and make it simple for them to delete their accounts or eliminate their personal information if they so choose. LinkedIn can also be transparent regarding its data security policies and practices and provide explicit explanations of how user data is collected, stored, and utilized. In addition, LinkedIn can educate users on best practices for keeping their accounts and data secure, as well as offer resources and support to users who experience data breaches or security issues. By implementing these suggestions, LinkedIn can enhance its reputation and provide its users with a superior experience.

**APPENDIX**

**Social Media Research Analysis and Findings**

> ts_plot(ENGREVIEWs,by="days")

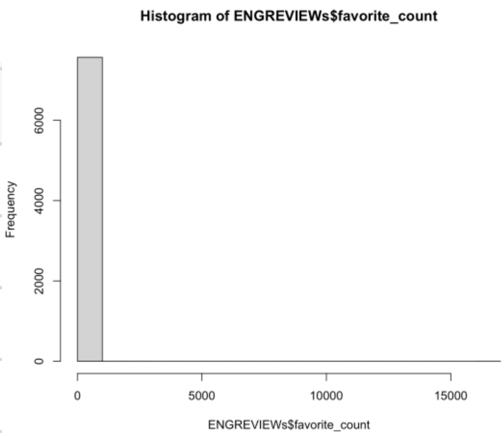



We can tell from the graph that on April 10th, the number of Linkedin Twitter posts went to a peak value. And it is the reason why we want to explore the reason behind it.
> ts_plot(ENGREVIEWs,by="hours")
And to be more precise, we can see from 10 am to 12 pm that day, the discussion went to a peak. So, we can search what is the reason behind the trend.

> tsdata=ts_data(ENGREVIEWs,by="days")

| | time | n |
|---|---|---|
| 1 | 2023-04-09 | 631 |
| 2 | 2023-04-10 | 6866 |
| 3 | 2023-04-11 | 1 |
| 4 | 2023-04-12 | 74 |




> summary(ENGREVIEWs$favorite_count)

| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max | NA |
|---|---|---|---|---|---|---|
| 0.00 | 0.00 | 0.00 | 3.982 | 0.00 | 16979 | 1 |

On average, Linkdin's Twitter posts did not show high favourite numbers compared to its followers' number (1.7M Followers). While during the day, the maximum number of favourites of one Linkdln post is 16979.

And then we tried to find the post with the highest number of favourites. To do so,

> which (ENGREVIEWs$favorite_count == 16979)

> ENGREVIEWs$full_text[c(1)]

[1] "The woman behind the BudLight Dylan Mulvaney campaign uses She/Her pronouns on LinkedIn. \n\nIn an unsurfaced video prior to the brand's collaboration with Dylan, the brands VP of Marketing, talked about "having a campaign that's truly inclusive and appeals to women". \n\nSo she… https://t.co/Xan0s2LBb5 https://t.co/u46aUnT6o8"


> quote<-na.omit(ENGREVIEWs$is_quote_status)

```
> #Analysis for quote tweet
> summary(quote)
   Mode   FALSE    TRUE 
logical    7466     106 
> sum(quote)/(nrow(temp)-1)
[1] 0.01399894
```

As shown by the result, among 7572 posts, there are 106 (1.4%) posts added their own comments when retweeting the messages from other accounts.

```
> temp$is_retweet <- grepl('^RT', temp$full_text)
> summary(temp$is_retweet)
   Mode   FALSE    TRUE 
logical    4916    2657 
> sum(temp$is_retweet)/nrow(temp)
[1] 0.3508517
```

We can see there are 2657 posts (35.09%) among 7573 posts retweeted which is a relatively high interaction rate.

```
> # Analysis of self replies
> temp$t1 <- temp$in_reply_to_screen_name
> temp$t1[is.na(temp$in_reply_to_screen_name)] = ""
> temp$is_self_reply<-temp$t1==temp$in_reply_to_screen_name
> summary(temp$is_self_reply)
   Mode    TRUE    NA's
logical     829    6744
> SELFREPLY=temp[temp$is_self_reply,]
> # Analysis on "private replies" replies that do not show in followers' news feeds
> temp$is_private=temp$is_reply==1 & temp$is_self_reply==0
> summary(temp$is_private)
   Mode   FALSE
logical    7573
```

```
> # Analysis for all replies (private replies and self replies)
> temp$is_reply<-!(is.na(temp$in_reply_to_screen_name))
> summary(temp$is_reply)
   Mode   FALSE    TRUE
logical    6744     829
> sum(temp$is_reply)/nrow(temp)
[1] 0.1094678
```

As for all replies, there are 829 posts that were private replies and self replies which can show that LinkedIn still needs to improve on the activeness of its account to increase user engagement. There is 829 are self replies while no private replies in the data we collected.

```
> summary(RT$favorite_count)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
      0       0       0       0       0       0 
> summary(QT$favorite_count)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's 
  0.000   0.000   0.000   4.113   1.000 217.000       1 
> summary(SELFREPLY$favorite_count)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's 
  0.000   0.000   0.000   2.527   1.000 266.000    6744 
> summary(OWNNOTREPLY$favorite_count)
    Min.    1st Qu.     Median       Mean    3rd Qu.       Max.       NA's 
   0.000      0.000      0.000      6.835      1.000  16979.000          1 
```

Among those posts, LinkedIn's "own" messages, which means not retweets, not quotes, not replies get the highest favorite counts.

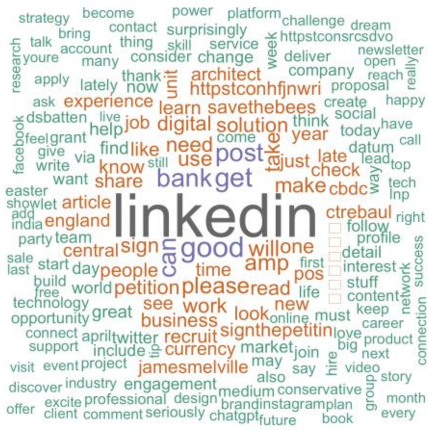
Now we tried to more focus on the relationship between shooting and LinkedIn.

```
> #Analysis on the topic shooting
> words <- c("shooting","shooter","Louisville")
> temp1<-temp[rowSums(sapply(words, grepl, temp$full_text)) > 0, , drop = FALSE]
> temp1$full_text
 [1] "RT @Islam_pfr: 🔻Officials used \"she\" to refer to the suspect of the Nashville shooting, but according to a social media post and a Linkedi
n…"
 [2] "RT @Islam_pfr: 🔻Officials used \"she\" to refer to the suspect of the Nashville shooting, but according to a social media post and a Linkedi
n…"
 [3] "RT @Islam_pfr: 🔻Officials used \"she\" to refer to the suspect of the Nashville shooting, but according to a social media post and a Linkedi
n…"
 [4] "RT @Islam_pfr: 🔻Officials used \"she\" to refer to the suspect of the Nashville shooting, but according to a social media post and a Linkedi
n…"
 [5] "RT @Islam_pfr: 🔻Officials used \"she\" to refer to the suspect of the Nashville shooting, but according to a social media post and a Linkedi
n…"
 [6] "RT @Islam_pfr: 🔻Officials used \"she\" to refer to the suspect of the Nashville shooting, but according to a social media post and a Linkedi
n…"
 [7] "RT @Islam_pfr: 🔻Officials used \"she\" to refer to the suspect of the Nashville shooting, but according to a social media post and a Linkedi
n…"
 [8] "🔻Officials used \"she\" to refer to the suspect of the Nashville shooting, but according to a social media post and a Linkedin profile, he was
male\n✍️6 people have been murdered and they're worried about the shooter's gender🤐\n#DeclineOfAHegemone https://t.co/aGIxaQbSo4"
 [9] "RT @DC_Draino: *Unconfirmed* reports identify the Nashville shooter as “Audrey Hale”, a biological female that identifies as “He/Him” on th…"
[10] "How to stop mass shootings.  #men #wealth #opportunity https://t.co/BtqZW4WCNg"
[11] "It's always lovely shooting spring images. You get the feel for warmer weather, even if you are lighting it in the studio.\nHere's some random
selects from some of the editorial work I've done this season for spring issues.\nIncluding M&amp;S, Olive mag, GoodFo…https://t.co/4senFgGwnl"
```

We have three keywords which are “shooting、shooter 、 Louisville ” and then we use words, grepl, and sapply to find the posts that had those keywords.  Then we figured out that there are 72.73% of posts that were retweeted.

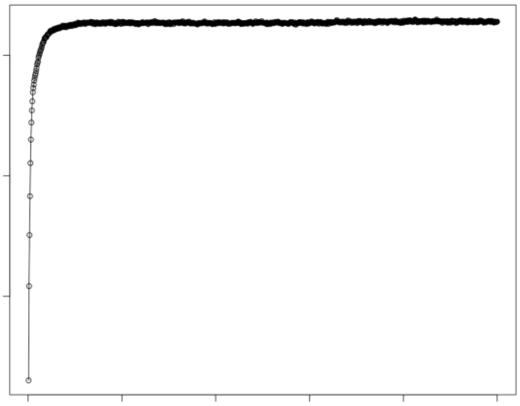
```
> temp1$is_retweet <- grepl('^RT', temp1$full_text)
> summary(temp1$is_retweet)
   Mode   FALSE    TRUE 
logical       3       8 
> sum(temp1$is_retweet)/nrow(temp1)
[1] 0.7272727
```


Then we want to analyze the data with more society topics since the shooting topic is less discussed among those posts.




Then we grabbed the “petition” in those posts.

One of the petitions we found on LinkedIn is called save the bees. However, the campaign itself focus on saving the bees. While the author mentioned in the first place that LinkedIn closed his account for no reason after the disclosure of a file with 6.5 million user passwords(TREBAUL, 2023). He mentioned the mismanagement and limited speech freedom that LinkedIn should improve on.



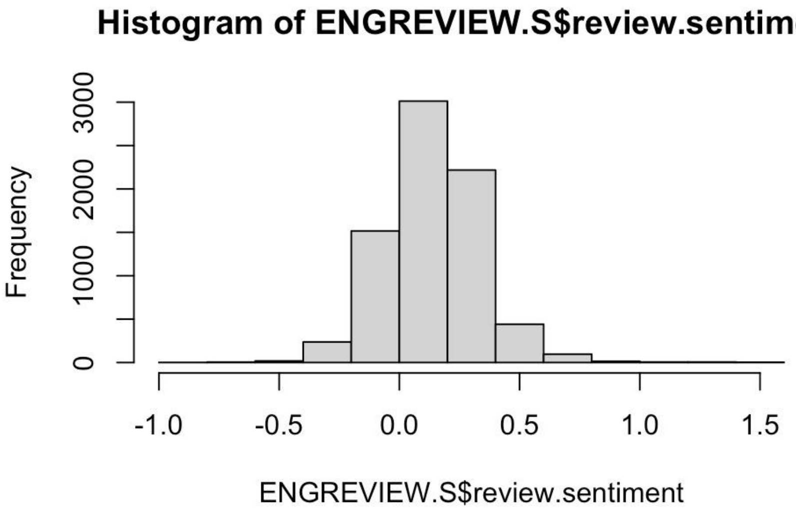
```
> TOPIC = top.topic.words(result$topics, 10, by.score=TRUE)
> TOPIC
      [,1]       [,2]         [,3]              
 [1,] "linkedin" "bank"       "sign"            
 [2,] "post"     "solution"   "petition"        
 [3,] "just"     "digital"    "savethebees"     
 [4,] "can"      "recruit"    "signthepetitin"  
 [5,] "take"     "pos"        "💯🐝🙏"            
 [6,] "know"     "architect"  "httpstconhfjnwri"
 [7,] "make"     "central"    "ctrebaul"        
 [8,] "good"     "currency"   "please"          
 [9,] "think"    "unit"       "join"            
[10,] "get"      "england"    "project"         
```

We later ran the topic analysis among those posts, the first topic should be the user experience using LinkedIn, the second one should be a bank issue, and the third one is the #SavetheBees petition.

```
> topicproportion[1:10,]
            [,1]       [,2]       [,3]
 [1,] 0.76923077 0.03846154 0.19230769
 [2,] 1.00000000 0.00000000 0.00000000
 [3,] 0.75000000 0.00000000 0.25000000
 [4,] 1.00000000 0.00000000 0.00000000
 [5,] 0.00000000 0.09090909 0.90909091
 [6,] 1.00000000 0.00000000 0.00000000
 [7,] 0.04545455 0.68181818 0.27272727
 [8,] 1.00000000 0.00000000 0.00000000
 [9,] 0.81818182 0.09090909 0.09090909
[10,] 0.91666667 0.00000000 0.08333333
> colMeans(topicproportion)
[1] 0.4680354 0.2299593 0.3020053
```

After that we ran sentimental analysis on the data, and the results are shown below,

```
> summary(ENGREVIEW.S$review.sentiment)
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
-0.87212  0.01928  0.12995  0.14205  0.25000  1.57098
```




From this sentimental analysis, we can get that the overall sentiment is positive, indicating a slightly positive tone overall. However, some of the posts and comments were also criticized and these posts' sentiment scores are below 0.

**5. References**

TREBAUL , P. (2023) LET'S #SAVETHEBEES TOGETHER LINKEDIN ! #BEEBAN, Change.org. Available at: https://www.change.org/p/linkedin-let-s-save-the-bees?utm_source=share_petition&utm_medium=custom_url&recruited_by_id=456c8970-feca-11e8-95c0-e383d2242133.

6. Appendix

1) R Script for data analysis

```
# Using Module 2 Behavioral Data to analyze the LinkedIn data

library(rtweet)

library(data.table)

library(ggplot2)

library(textcat)

library(tm)

library(textstem)

library(RColorBrewer)

library(wordcloud)

library(lda)

library(topicmodels)

library(textcat)

# read the data object

LinkedIn = readRDS("testTWLinkedIn.RDS")

#Divide text by different languages

language=textcat(LinkedIn$text)

sort(table(language),decreasing=T)

# Select only English reviews

ENGREVIEWs=LinkedIn[language=="english",]

# Trend Analysis: find out the number of tweets over time

# Times Series Analysis

ts_plot(ENGREVIEWs,by="days")
```

```
ts_plot(ENGREVIEWs,by="hours")

tsdata=ts_data(ENGREVIEWs,by="days")

#Are these social media messages effective?

# use summary statistics

summary(ENGREVIEWs$favorite_count)

# use histogram

hist(ENGREVIEWs$favorite_count,nclass=20)

which (ENGREVIEWs$favorite_count == 16979)

ENGREVIEWs$full_text[c(1)]

# There are different types of social media messages

# Original tweets

# Retweets: broadcast messages from other accounts without changing the content

# Quote Tweets: Add own comments when retweeting the messages from other accounts

# Private replies to someone: do not show up in the followers' feeds

# Self replies to 3M itself: show up in the followers' feeds

temp=ENGREVIEWs

quote<-na.omit(ENGREVIEWs$is_quote_status)

#Analysis for quote tweet

summary(quote)

sum(quote)/(nrow(temp)-1)

#Analysis for retweet

temp$is_retweet <- grepl('^RT', temp$full_text)

summary(temp$is_retweet)

sum(temp$is_retweet)/nrow(temp)

RT=temp[temp$is_retweet,]

# Analysis for all replies (private replies and self replies)

temp$is_reply<-!(is.na(temp$in_reply_to_screen_name))

summary(temp$is_reply)

sum(temp$is_reply)/nrow(temp)
```

```
# Analysis of self replies

temp$t1 <- temp$in_reply_to_screen_name

temp$t1[is.na(temp$in_reply_to_screen_name)] = ""

temp$is_self_reply<-temp$t1==temp$in_reply_to_screen_name

summary(temp$is_self_reply)

SELFREPLY=temp[temp$is_self_reply,]

# Analysis on "private replies" replies that do not show in followers' news feeds

temp$is_private=temp$is_reply==1 & temp$is_self_reply==0

summary(temp$is_private)

# Analysis on LinkedIn's "own" messages, not retweets, not quote, not replies

OWNNOTREPLY=temp[temp$is_retweet==0 & temp$is_quote_status==0 & temp$is_reply==0,]

QT=temp[temp$is_quote_status,]

# Which type of messages is more liked?

summary(RT$favorite_count)

summary(QT$favorite_count)

summary(SELFREPLY$favorite_count)

summary(OWNNOTREPLY$favorite_count)

#Analysis on the topic shooting

words <- c("shooting","shooter","Louisville")

temp1<-temp[rowSums(sapply(words, grepl, temp$full_text)) > 0, , drop = FALSE]

temp1$full_text

# search the full_text column and if the text starts with RT, it is retweet.

temp1$is_retweet <- grepl('^RT', temp1$full_text)

summary(temp1$is_retweet)

sum(temp1$is_retweet)/nrow(temp1)

#real data topic analysis further

tp=ENGREVIEWs$text

docs = VCorpus(VectorSource(tp))

# Preprocess the data:
```

```
# Transform to lower case

docs <-tm_map(docs,content_transformer(tolower))

#remove punctuation

docs = tm_map(docs, removePunctuation)

#Strip digits

docs = tm_map(docs, removeNumbers)

#remove stopwords

docs = tm_map(docs, removeWords, stopwords("english"))

#remove whitespace

docs = tm_map(docs, stripWhitespace)

# lemmatize the words

# e.g. ran and run will both be converted to run

docs = tm_map(docs, content_transformer(lemmatize_strings))

#convert data into document term matrix format

DTM = DocumentTermMatrix(docs)

# Generate wordcloud for all the documents

# Calculate the frequency of each word

DTMMATRIX=as.matrix(DTM)

WD = sort(colSums(DTMMATRIX),decreasing=TRUE)

# Transform the data to fit with the wordcloud package

WCLOUD=data.table(words=names(WD),freq=WD)

par(mar=c(1,1,1,1))

wordcloud(words = WCLOUD$words, freq = WCLOUD$freq,

        scale=c(5,1),

        min.freq = 2,

        max.words=200,

        random.order=FALSE,

        rot.per=0.1,

        colors=brewer.pal(8, "Dark2"))
```

```
words2 <- c("petition")
temp2<-temp[rowSums(sapply(words2, grepl, temp$full_text)) > 0, , drop = FALSE]
temp2$full_text
# Apply topic models to identify common themes
# Transform the data to fit with the topic model package
input = dtm2ldaformat(DTM)
# set the random seed so results are replicable
set.seed(12345)
# select model parameters to be 3 topics,  select parameters
K=3
N=1000
result = lda.collapsed.gibbs.sampler(
  input$documents,
  K,                # The number of topics.
  input$vocab,
  N, # The number of iteration
  alpha=1/K,          # The Dirichlet hyper parameter for topic proportion
  eta=0.1,            # The Dirichlet hyper parameter for topic multinomial
  compute.log.likelihood=TRUE)
plot(result$log.likelihoods[1,],type="o")
# Get the top words in each cluster.
# Top words are the characteristics of the relevant topic.
#different topics may have same keywords
TOPIC = top.topic.words(result$topics, 10, by.score=TRUE)
TOPIC
# Count number of topic keywords for each sentence
T1=t(result$document_sums)
# Calculate the topic assignment for each sentence
topicproportion=T1/rowSums(T1)
```

```
topicproportion[1:10,]

colMeans(topicproportion)

#sentiment score

library(sentimentr)

temp3<-ENGREVIEWs$text

MYTEXT=get_sentences(temp3)

SENTENCE.S = sentiment(MYTEXT)

SENTENCE.S = as.data.table(SENTENCE.S)

REVIEW.S=SENTENCE.S[,

list(

 review.sentiment = mean(sentiment)

),

),

 by=element_id]

ENGREVIEW.S = cbind(ENGREVIEWs, REVIEW.S)

summary(ENGREVIEW.S$review.sentiment)

hist(ENGREVIEW.S$review.sentiment)
```