

ROLE OF ONLINE COMMUNITIES IN JOB SEARCH EXPERIENCES

Team 2

Truc Kim Ngo

Fiona Mariella Sutanto





INTRODUCTION



Home



Popular



Explore



All

CUSTOM FEEDS



+ Create a custom feed



RECENT



r/careerguidance

WHAT IS REDDIT?

Reddit is a social media platform where users can share content, ask questions, and discuss topics in niche communities called subreddits. Reddit is often described as divided into sub-forums, where people with similar interests can gather to exchange ideas, news, and opinions.

R/CAREERGUIDANCE

These sub-forums are called 'subreddits'. Subreddits are smaller communities within the larger Reddit community. Career Guidance is a subreddit that is described as 'A place to discuss career options, to ask questions and give advice!' It is where people can learn about looking for jobs, voicing concerns regarding their professional life and asking for advice within a community of employees, jobseekers and job hoppers.

PRACTICAL BACKGROUND



Our Intention for this Research:

During the job finding process, a lot of people are under a lot of stress which might affect mental health and we are interested in discovering whether an online community can be one of the factors causing mental distress, such as depression. It can also insight and improvement for content moderation or guidelines that encourage positive and supportive discussions.

Research Question:

1. How do commenters typically respond (positively or negatively) to posts written by individuals sharing their job search experiences?
2. What are the predominant themes in comments on job-seeking posts?

Method

Sentiment Analysis
Topic Modelling



 u/PhysicalAd6081 • 17 hr. ago



I'm getting fired today - they sent me the outlook invite - should I tell them off?

Been working remotely from home at this regional sales position. From the moment I started, this place was a shitshow of disorganization, lack of management and zero onboarding. Embarrassingly so, but I took it in stride. I was poached by this place who sold me a bad bill of goods that I'd have lots of tech support and one-on-ones with partners. They left me high and dry with a micro manager in their first position managing other people. I figured out a month in that I was the 4th person in the role this year. Lots of promises with zero flow through. Wednesday before...

 185 

 101



 Share

TITLE

POST



Optimal_Shirt6637 • 16h ago •

 Top 1% Commenter

I don't agree with doing this for the health of your long term career. You never know who someone from your old company might pop up in the future.

 76   Reply  Award  Share ...



AugustThursday • 11h ago •

I am so sick of hearing this. They are going to talk shiz anyway.

 9   Reply  Award  Share ...



Optimal_Shirt6637 • 11h ago • Edited 11h ago •

 Top 1% Commenter

Huge difference when you give them something to talk shit about. Then instead of just saying you weren't good at your job they're also going to say you're an ass who can't control your emotions in a professional setting. That bars you from finding opportunities more than being bad at the specific set of tasks you were assigned.

POST
EXAMPLE

COMMENTS

PACKAGES USED

FOR EXTRACTING
THE DATA

```
library(SentimentAnalysis)
```

FOR SENTIMENT
ANALYSIS

```
library(RedditExtractor)  
#setwd(dirname(rstudioapi::getActiveDocumentContext()$path))
```

DICTIONARY CHOICE

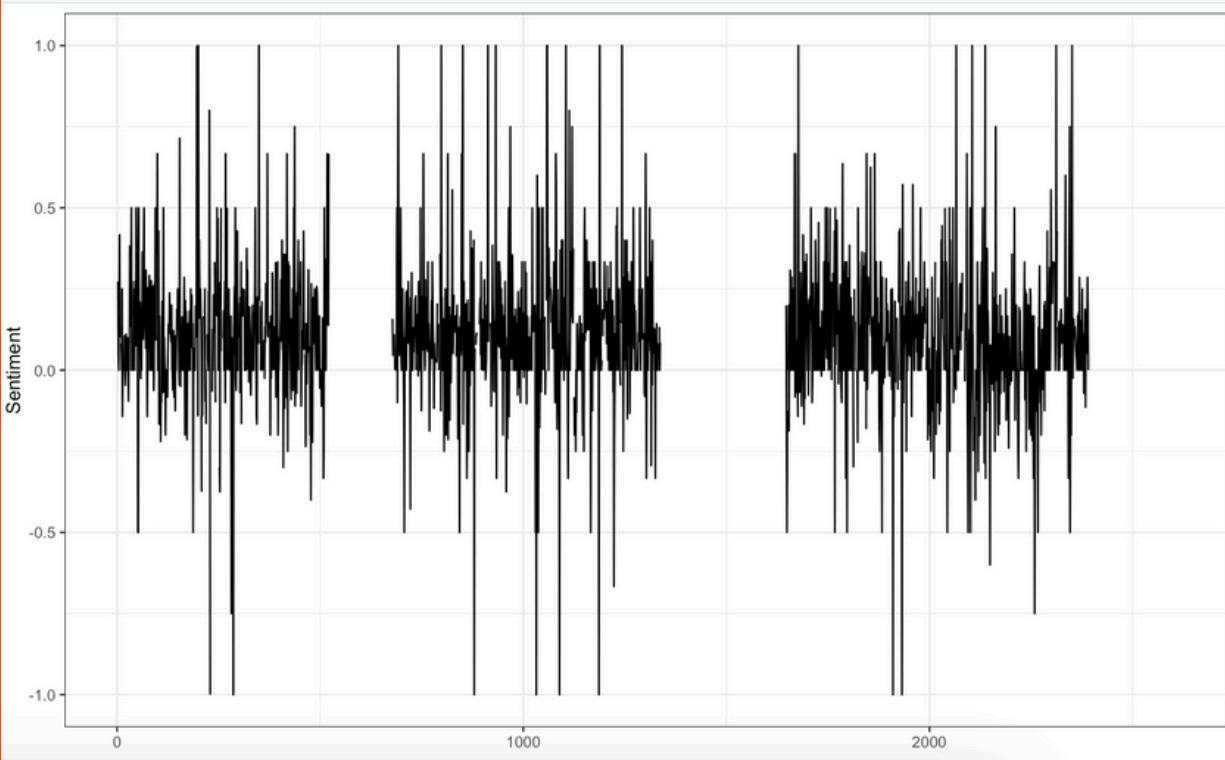
LSD2015 explicitly categorizes words into six types: positive, negative, uncertainty, litigious, constraining, and superfluous, making it more granular than simpler dictionaries that only classify words as positive or negative.

```

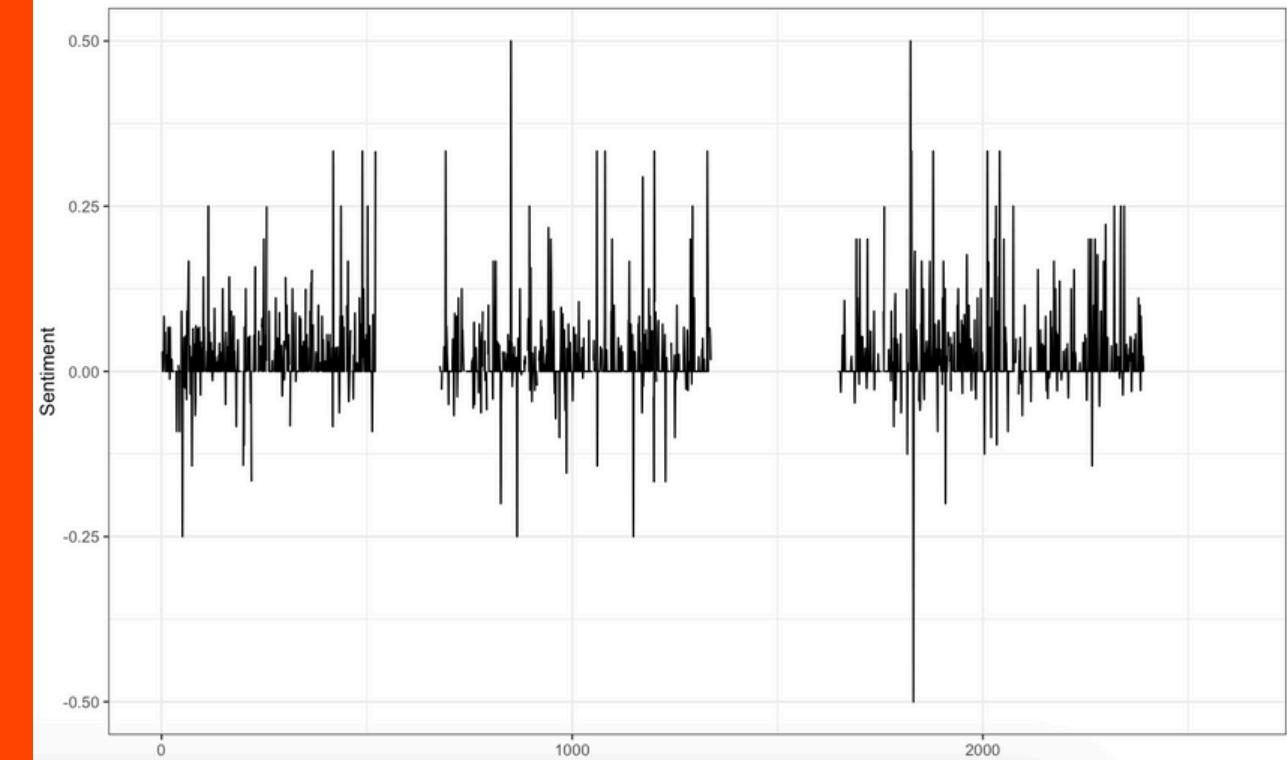
# SENTIMENT ANALYSIS METHOD 1
#####
library(SentimentAnalysis)
sentiment <- analyzeSentiment(merged_df$comment)
# output is sentiment scores based on different dictionary (dictGI, dictHE, dictLM, dictQDAP)
## positive - negative = neutral score
## https://datascience.stackexchange.com/questions/25782/what-do-all-the-column-names-mean-in-the-sentimentanalysis-package

# the difference in scores based on the choice of dictionary
sentimentData <- sentiment$SentimentGI
sentimentData <- sentiment$SentimentHE
sentimentData <- sentiment$SentimentLE
sentimentData <- sentiment$SentimentQDAP
plotSentiment(sentimentData)

```



General Inquirer (Harvard-IV) dictionary



HE dictionary

We also referred to a blog post about the choice of dictionary and apply what is mentioned in the post to our code. The results of sentiment scores using other dictionaries.

DATA COLLECTION METHOD

```
# DATA LOADING
#####
# get the list of the web addresses of the most recent posts
load(file.choose()) # pop-up window, choose objects.RData
CGThreads.df <- data.frame(CGThreads)
CGThreads <- find_thread_urls(subreddit = "careerguidance", sort_by = "new", period = "day") ←
# write.csv(CGThreads, "/Users/User/Desktop/R practice/finalpj_data/redditThreads.csv")

# access each URL and grab information about the comments
CGComments <- get_thread_content(CGThreads$url)
CGComments.df <- data.frame(CGComments$comments)

write.csv(CGComments.df, "/Users/User/Desktop/R practice/finalpj_data/redditComments.csv")

# combine by "url"
require(dplyr)
merged_df <- full_join(CGThreads.df, CGComments.df, by = "url")
colnames(merged_df)[colnames(merged_df) == "comments"] <- "count_comments"
write.csv(merged_df, "/Users/User/Desktop/R practice/finalpj_data/redditMerged.csv")
```

From the Reddit website; sub-reddit: Career Guidance ➤ looking at paragraphs and comments

Using a function called `find_thread_urls` ➤ used for the collection of the **TITLES** and **CONTENT** of posts

Using a function called `get_thread_content` ➤ used for the collecting **COMMENTS**

Period is put as 'day' but can be changed to month or year, depending on the scale of the project.

⌚ It took 30 minutes to collect a day's worth of data!



	date_utc	timestamp.x	title	text	subreddit	count_comments	url
1	2024-11-25	1732572464	I have no idea what to do, hoping to get some advice....	TLDR: I've been unemployed for a year and don't kno...	careerguidance	0	https://www.reddit.com/r/careerguidance/comments...
2	2024-11-25	1732571616	Can you help?	Lost girl SOS Hey 27F have an Associates in IT but I h...	careerguidance	1	https://www.reddit.com/r/careerguidance/comments...
3	2024-11-25	1732571497	How do you deal with rejection in job applications?	I recently got rejected from a job that I really really w...	careerguidance	2	https://www.reddit.com/r/careerguidance/comments...
4	2024-11-25	1732571497	How do you deal with rejection in job applications?	I recently got rejected from a job that I really really w...	careerguidance	2	https://www.reddit.com/r/careerguidance/comments...
5	2024-11-25	1732570841	Which job should I go for?	Hi, just as the title suggests. Any advice for career re...	careerguidance	2	https://www.reddit.com/r/careerguidance/comments...
6	2024-11-25	1732570841	Which job should I go for?	Hi, just as the title suggests. Any advice for career re...	careerguidance	2	https://www.reddit.com/r/careerguidance/comments...
7	2024-11-25	1732570700	How do I quit my job?	Context: I work on a team where all my days off/sick ...	careerguidance	1	https://www.reddit.com/r/careerguidance/comments...
8	2024-11-25	1732570179	What degree/career should I pursue?	Not sure what degree/job to pursue Hi All I am locate...	careerguidance	0	https://www.reddit.com/r/careerguidance/comments...
9	2024-11-25	1732570158	What is the etiquette for quitting?	I am a medical biller for a medical staffing company.	careerguidance	3	https://www.reddit.com/r/careerguidance/comments...
10	2024-11-25	1732570158	What is the etiquette for quitting?	I am a medical biller for a medical staffing company.	careerguidance	3	https://www.reddit.com/r/careerguidance/comments...
11	2024-11-25	1732570158	What is the etiquette for quitting?	I am a medical biller for a medical staffing company.	careerguidance	3	https://www.reddit.com/r/careerguidance/comments...
12	2024-11-25	1732569735	What would you recommend for me?	I wanna add a bit of context so I'll write in bullet poin...	careerguidance	1	https://www.reddit.com/r/careerguidance/comments...
13	2024-11-25	1732569577	Excuses to get off work for 4 days?	Basically, im planning to go to a music festival overse...	careerguidance	13	https://www.reddit.com/r/careerguidance/comments...
14	2024-11-25	1732569577	Excuses to get off work for 4 days?	Basically, im planning to go to a music festival overse...	careerguidance	13	https://www.reddit.com/r/careerguidance/comments...
15	2024-11-25	1732569577	Excuses to get off work for 4 days?	Basically, im planning to go to a music festival overse...	careerguidance	13	https://www.reddit.com/r/careerguidance/comments...
16	2024-11-25	1732569577	Excuses to get off work for 4 days?	Basically, im planning to go to a music festival overse...	careerguidance	13	https://www.reddit.com/r/careerguidance/comments...

POST



SCRAPING RESULTS

COMMENTS →

	author	date	timestamp.y	score	upvotes	downvotes	golds	comment	comment_id
	NA	NA	NA	NA	NA	NA	NA	NA	NA
	ftwmindset	2024-11-25	1732576819	2	2	0	0	Highly recommend going into the event planning ind...	1
	Partytyme2021	2024-11-25	1732572278	1	1	0	0	It s not personal, it s just business. Keep that in min...	1
	ThingsWorkOut	2024-11-25	1732576220	1	1	0	0	Their loss.	2
	thepandapear	2024-11-26	1732651955	2	2	0	0	You might want to look into positions like rehab or w...	1
	Conscious-Quarter423	2024-11-25	1732575085	1	1	0	0	look into high paying allied health careers like perfusi...	2
	Renaissance_Career	2024-11-25	1732571176	1	1	0	0	First of all, it's very considerate of you for thinking ab...	1
	NA	NA	NA	NA	NA	NA	NA	NA	NA
	LittleTatoCakes	2024-11-25	1732570877	3	3	0	0	2 weeks notice after getting the offer. That s complim...	1
	pezziepie85	2024-11-25	1732578958	1	1	0	0	I get the loyalty. I m the same. However you can bet y...	2
	Next-Wishbone1404	2024-11-26	1732628778	1	1	0	0	Give two weeks' notice and stick to it. They will find s...	3
	ftwmindset	2024-11-25	1732576894	1	1	0	0	Have you considered trade school? Electrician, plumbi...	1

This is a combination of the title (posts) dataframe and the comments dataframe, using “URL” as a key to link two tables.

	date_utc	timestamp.x	title	text	comment
1	2024-11-25	1732572464	I have no idea what to do, hoping to get some advice. What would you do?	TLDR: I've been unemployed for a year and don't know what to do. I will answer your question as best as I can. I am a long-time lurker, but this is my first time creating an account and I need some help. Before I start, here's a little backstory: I am with someone who is in the army. It's been a year since I last worked. I learned about Year Up which involves six months of school at a community college followed by a job placement. I interned at Salesforce for 6 months as a UX Project Coordinator Intern. It went well, but I eventually accepted an offer at Accenture as a Project Manager but was affected by a layoff. My problem is this: It's been a year since I've had a job. My girlfriend is getting tired of me being unemployed. I was sold a dream that with Year Up, I wouldn't need a degree. I have come to realize that I do need one to find a good job.	NA
2	2024-11-25	1732571616	Can you help?	Lost girl SOS Hey 27F have an Associates in IT but I haven't worked in IT since 2019 and tbh I've been doing warehouse jobs ever since, but I just got laid off. Now I want to	Highly recommend going into the event planning industry. With your skill sets, you would be a great fit. There are many opportunities in that field.
3	2024-11-25	1732571497	How do you deal with rejection in job applications?	I recently got rejected from a job that I really really wanted, mainly because it wasn't a good fit for me. I have been trying to figure out front office roles but it seems like such a dead end. Most of finding a new job is just a matter of putting yourself out there. You have a little business, try to network with people in your industry.	Its not personal, its just business. Keep that in mind. Many people are experiencing this. Just keep going, keep applying. Try not to give up. You may also want to reach out to people who may already have a job you'd be interested in. Ask them for advice. Most of finding a new job is just a matter of putting yourself out there. You have a little business, try to network with people in your industry.

TASK: CALCULATING SENTIMENT SCORES

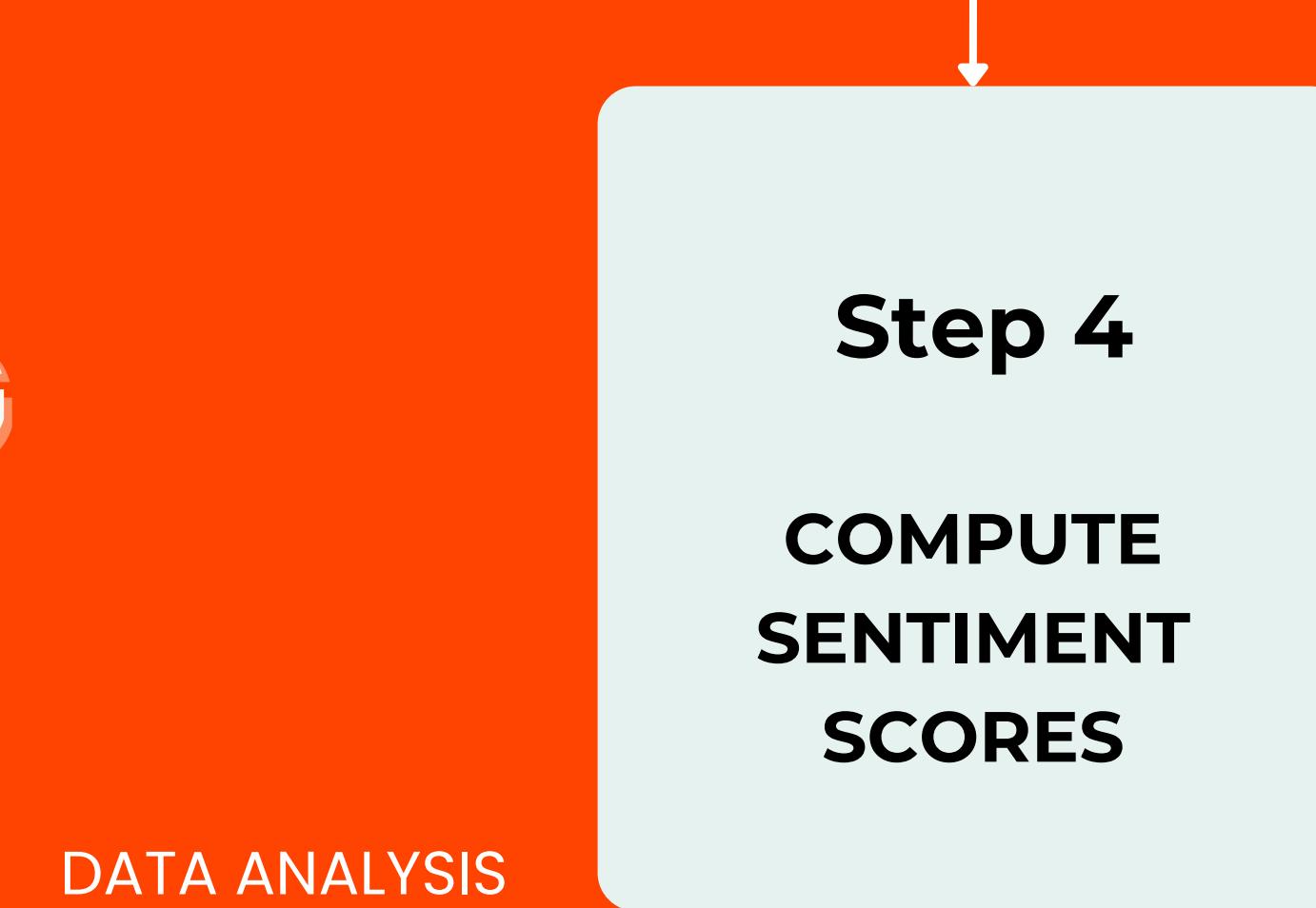


Research Question 1

How do commenters typically respond (positively or negatively) to posts written by individuals sharing their job search experiences?



**TASK:
CALCULATING
SENTIMENT
SCORES**



CHECKING FOR MISSING DATA & STEP 1-3

STEP 0:

```
num_zero_comments <- sum(CGThreads.df$comments == "0") # 351 threads with no comments
num_zero_comments
num_unique_urls <- length(unique(merged_df$url)) # equal number of threads; no prob in merge process!
num_unique_urls
#num_comments <- sum(CGThreads.df$comments) # 2609 comments (we have 8350 comments in total)
#num_comments
```

STEP 1:

```
# basic info about the dictionary that we apply on our analysis: https://quanteda.io/reference/data\_dictionary\_LSD2015.html
data_dictionary_LSD2015
summary(data_dictionary_LSD2015)
merged_df_copy <- merged_df
library(quanteda)
```

STEP 2:

```
additional_stopwords <- c("feel", "like")

cal_sentiment <- function(mydata) {
  # Step 1: Create corpus
  mydata <- cbind(doc_id = seq_len(nrow(mydata)), mydata)
  mydata_corpus <- corpus(mydata, docid_field = "doc_id", text_field = "text")
```

STEP 3:

```
# Step 2: Clean raw data
dfm_mydata_corpus <- mydata_corpus %>%
  tokens(remove_punct = TRUE, remove_symbols = TRUE) %>%
  tokens_wordstem() %>%
  tokens_remove(stopwords("english"), additional_stopwords) %>%
  dfm()

# Step 3: Tokenize and Apply dictionary
toks_dict <- mydata_corpus %>%
  tokens() %>%
  tokens_lookup(dictionary = data_dictionary_LSD2015)
dfm_mydata_LSD2015 <- dfm(toks_dict)
```

STEP 4: CALCULATING SENTIMENT

STEP 4:

```
# Step 4: Calculate sentiment
dict_output <- convert(dfm_mydata_LSD2015, to = "data.frame")
dict_output$sent_score <- log(
  (dict_output[, 3] + dict_output[, 5] + 0.5) /
  (dict_output[, 2] + dict_output[, 4] + 0.5)
)
dict_output <- cbind(dict_output, docvars(mydata_corpus))

# Return the final dictionary output
return(dict_output)
}
```

ADDITIONAL STEP:

**As we have two datasets, we need to repeat step 1 to 4 two times
=> Write a function called “cal_sentiment” and call the function
'two times' so that we do not need to write the codes double times**

```
88 thread_sentiment_df <- cal_sentiment(CGThreads.df)
89 # head(thread_sentiment_score)
90
91 # change the name of "comment" column to "text" so that the function works:
92 colnames(CGComments.df)[colnames(CGComments.df) == "comment"] <- "text"
93
94 comment_sentiment_df <- cal_sentiment(CGComments.df)
95 #head(comment_sentiment_score)
```

Reddit Results

**THREAD'S
SENTIMENT
SCORE**



doc_id	negative	positive	neg_positive	neg_negative	sent_score	date_utc	timestamp	title
1	15	20	0	0	0.27958486	2024-11-25	1732572464	I have no idea what to do, hoping
2	6	8	0	1	0.37948962	2024-11-25	1732571616	Can you help?
3	7	6	0	0	-0.14310084	2024-11-25	1732571497	How do you deal with rejection i
4	0	5	0	0	2.39789527	2024-11-25	1732570841	Which job should I go for?
5	2	2	0	0	0.00000000	2024-11-25	1732570700	How do I quit my job?
6	8	5	0	0	-0.43531807	2024-11-25	1732570179	What degree/career should I pur
7	8	14	0	0	0.53408249	2024-11-25	1732570158	What is the etiquette for quitting
8	5	6	0	0	0.16705408	2024-11-25	1732569735	What would you recommend for
9	10	13	0	0	0.25131443	2024-11-25	1732569577	Excuses to get off work for 4 day
10	2	4	0	0	0.58778666	2024-11-25	1732569391	What would be a good job or con
11	1	3	0	0	0.84729786	2024-11-25	1732568075	Are there any good career interes
12	2	6	1	0	0.61903921	2024-11-25	1732567786	What are your thoughts on Online

doc_id	negative	positive	neg_positive	neg_negative	sent_score	url	author
1	0	6	0	0	2.5649494	https://www.reddit.com/r/careerguidance/comments...	ftwmir
2	1	1	0	0	0.0000000	https://www.reddit.com/r/careerguidance/comments...	Partytin
3	1	0	0	0	-1.0986123	https://www.reddit.com/r/careerguidance/comments...	Things
4	1	9	0	0	1.8458267	https://www.reddit.com/r/careerguidance/comments...	thepar
5	0	2	0	0	1.6094379	https://www.reddit.com/r/careerguidance/comments...	Consci
6	4	8	0	0	0.6359888	https://www.reddit.com/r/careerguidance/comments...	Renais
7	1	5	0	0	1.2992830	https://www.reddit.com/r/careerguidance/comments...	LittleT
8	4	5	0	0	0.2006707	https://www.reddit.com/r/careerguidance/comments...	pezzie
9	0	1	0	0	1.0986123	https://www.reddit.com/r/careerguidance/comments...	Next-1
10	0	1	0	0	1.0986123	https://www.reddit.com/r/careerguidance/comments...	ftwmir
11	2	0	0	0	-1.6094379	https://www.reddit.com/r/careerguidance/comments...	silvern
12	3	2	1	0	-0.5877867	https://www.reddit.com/r/careerguidance/comments...	iheartv
13	0	0	0	0	0.0000000	https://www.reddit.com/r/careerguidance/comments...	[deleted]



**COMMENT'S
SENTIMENT
SCORE**

RESULTS

```
> cor(na.omit(analysis.df[,-1]))  
          sent_score.x sent_score.y  
sent_score.x     1.0000000  0.2158231  
sent_score.y     0.2158231  1.0000000
```

```
> print(cov(analysis.df$sent_score.x, analysis.df$sent_score.y, use="complete"))  
[1] 0.1897255
```

INTERPRETATION



Research Question 1

How do commenters typically respond (positively or negatively) to posts written by individuals sharing their job search experiences?

The correlation coefficient between sent_score.x (posts) and sent_score.y (comments) is 0.2158, which is positive but weak. This indicates that while there is a tendency for the sentiment of comments to align with the sentiment of the posts (positive posts receiving positive comments and negative posts receiving negative comments), the relationship is not very strong. Other factors may influence the sentiment of comments.



TASK: FINDING COMMON THEMES



Research Question 2

What are the predominant themes in comments on job-seeking posts?



TASK: FINDING COMMON THEMES

DATA WRANGLING

DATA ANALYSIS

Step 4
TOPIC MODELLING

```

processed <- textProcessor(CGThreads.df$text, metadata = CGThreads,
                           lowercase = TRUE, #*
                           removestopwords = TRUE, #*
                           removenumbers = TRUE, #*
                           removepunctuation = TRUE, #*
                           stem = TRUE, #*
                           wordLengths = c(3,Inf), #*
                           sparselevel = 1, #*
                           language = "en", #*
                           verbose = TRUE, #*
                           onlycharacter = TRUE, # not def
                           striphtml = FALSE, #*
                           customstopwords = additional_stopwords, #*
                           v1 = FALSE) #*

require(stm)

stm_model <- stm(processed$documents, processed$vocab, K = 5,
                  data = processed$meta, verbose = FALSE,
                  LDAbeta = T,
                  interactions = F,
                  reportevery = 10, |
                  init.type = c("Spectral"), seed = 123)
plot(stm_model)
labelTopics(stm_model)
cloud(stm_model, topic=5)

```

Removing stopwords like “like” or “feel” as they are just filler words when writing comments



Research Question 2

What are the predominant themes in comments on job-seeking posts?

TO ANSWER RS 2:

Run

```

# change number of topics to 10
stm_model12 <- stm(processed$documents, processed$vocab, K = 10,
                     data = processed$meta, verbose = FALSE,
                     LDAbeta = T, ## defaults to TRUE when there are no content covariates
                     interactions = F, ## no content covariates, so no interaction required
                     reportevery = 10, ## displace every 10 iterations
                     init.type = c("Spectral"), seed = 123)
plot(stm_model12)

set.seed(837)
library(quanteda)
library(stm)
require(stm)

model_topic <- function(mydata, num_topics) {
  # Step 1: Create a quanteda corpus
  mydata_corpus <- corpus(mydata, text_field = "text")

  # Step 2: Clean the text data
  my_data_dfm <- mydata_corpus %>%
    tokens(remove_punct = TRUE, remove_symbols = TRUE, remove_numbers = TRUE) %>%
    tokens_wordstem() %>%
    tokens_remove(stopwords("english")) %>%
    dfm()

  # Step 3: Convert quanteda DFM to STM input format
  stm_input <- convert(mydata_dfm, to = "stm")

  # Return the fitted STM model
  return(ctmodel)
}

stmodel <- model_topic(CGThreads.df, 5)

```

S

T

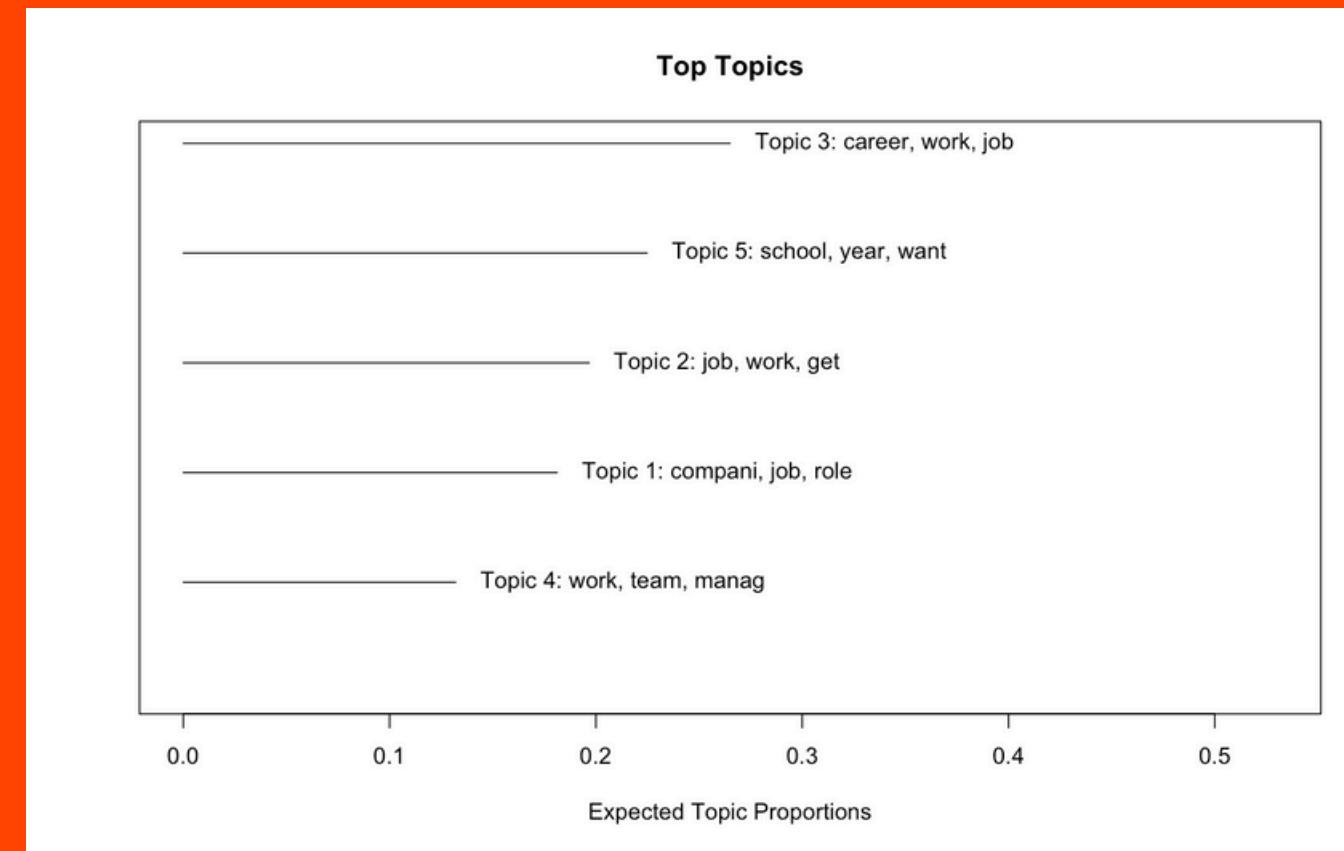
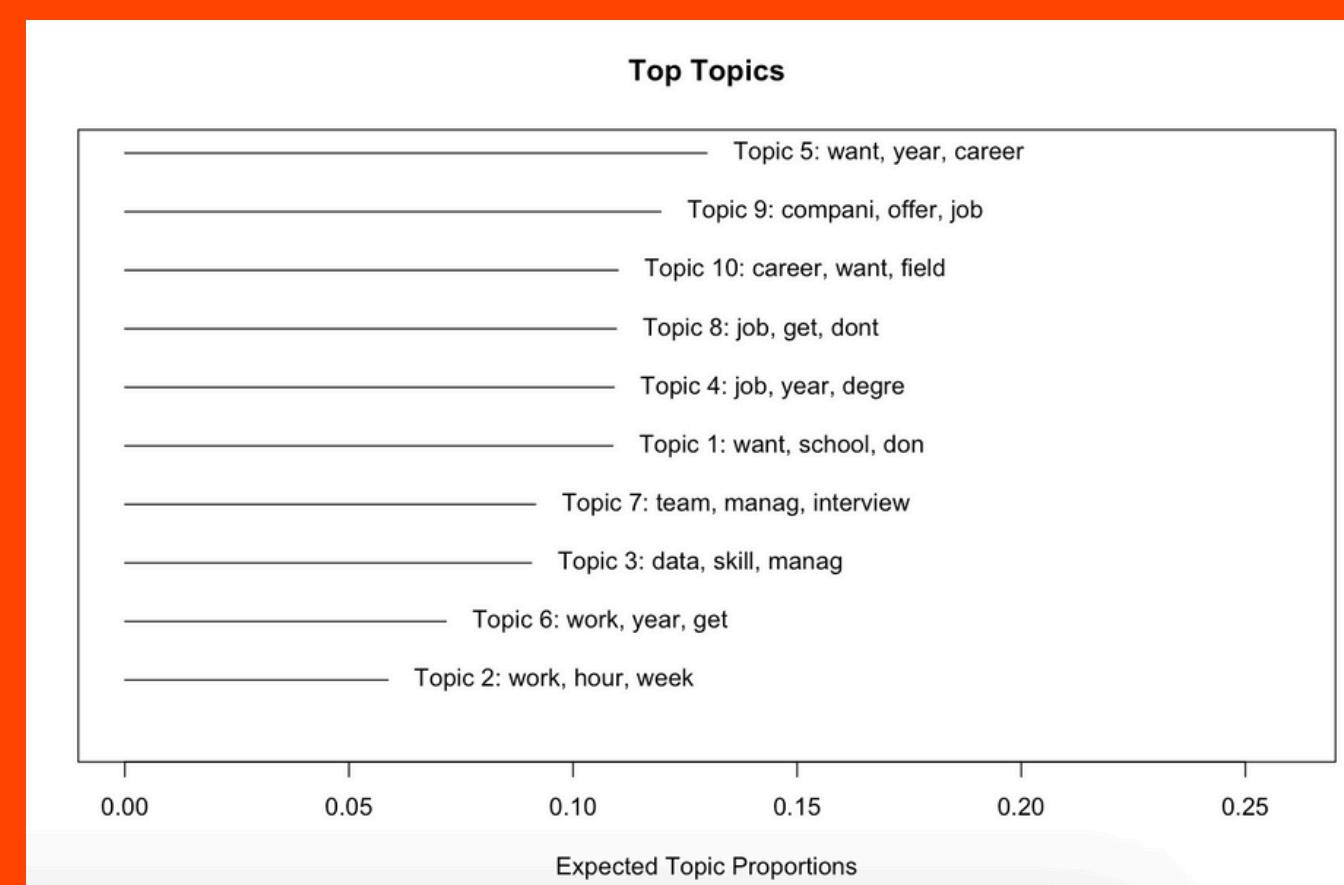
S

D

S

W

R



8

Insights:

Even when we increased the number of topics, we did not find it easier to differentiate between the topics. However, from the results we can infer about the roles of the post writers.

Those who mentioned “work, team, management” might have already succeeded in the interview and are preparing for their corporate life. Those who asked about “school, year, want” are probably students who are considering different career choices.



TOPIC 5: Career aspirations and education

Keywords: "want," "school," "career"

Focuses on career goals, educational backgrounds, and the connection between personal aspirations and professional ambitions.

TOPIC 10: Choosing a career path

Keywords: "career," "want," "field"

Discussions about aligning personal interests with career choices and the steps needed to enter and succeed in a specific industry.

TOPIC 8: Job search strategies

Keywords: "job," "get," "dont"

Advice on effective job search strategies and common pitfalls to avoid and tips on how to successfully navigate the job market and land desired positions.

TOPIC 9: Company and management

Keywords: "compani," "job," "manag"

Focuses on companies, workplace culture, and experiences with management, indicating a theme of sharing or seeking information about potential employers.

TOPIC 4: Job qualifications

Keywords: "job," "year," "degree"

Reflects the necessary qualifications for specific job positions, such as years of experience and degrees required.



TOPIC 1: Career aspirations and education

Keywords: "want," "school," "don"

Focuses on career aspirations and the role of education in achieving personal and professional goals.



TOPIC 7: Teamwork & management skills

Keywords: "team," "manage," "interview"

Highlights the value of teamwork and management skills in job searching & how these skills are evaluated during interviews and their importance in leadership roles.



TOPIC 3:

Data skills & management

Keywords: "data," "skills," "manage"

Covers the growing importance of data analysis and management skills in various job roles & discusses the need for professionals who can interpret data effectively for business decisions.

TOPIC 6:

Gaining work experience

Keywords: "work," "year," "get"

Addresses the importance of gaining work experience and how it impacts job prospects & strategies for acquiring relevant experience and tips for securing employment.

TOPIC 2:

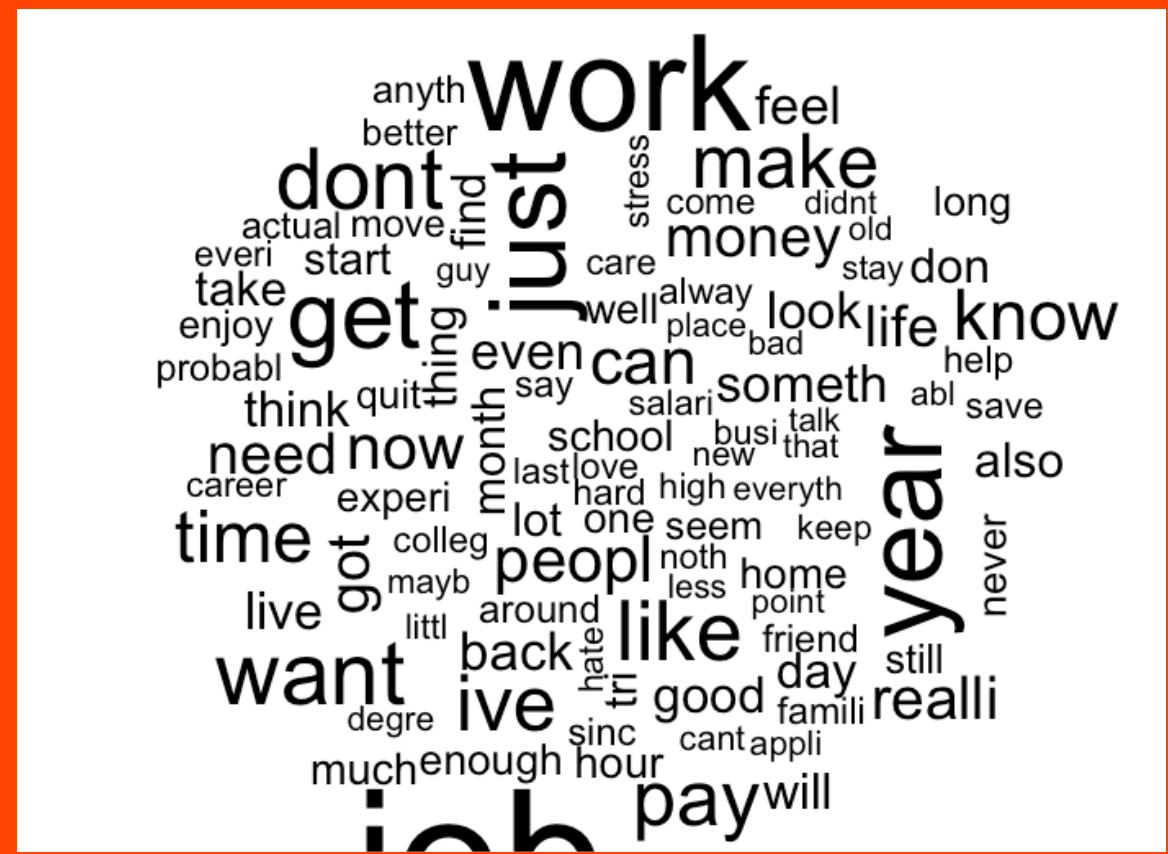
Work hours & scheduling

Keywords: "work," "hour," "week"

Focuses on work-life balance and the distribution of work hours & the impact of work hours on productivity, and employee satisfaction.

Topic 3

Topic 5



Topic 2

time work
feel know
can get
just goal
work
make
use
expect
think
man
often
video
say
ever
call day
hour
stuff
understand
involv
sick
frustrat
friend
speci
told
ive
part
find
question
made
well
good
back
team
situat
improv
tri
sleep
one
focus
peopl
never
anyth
task
mistak
best
see
help
small
manag
even
take
start
other
someth
issu
need
boss
product
new
social
alway
creat
offic
complet
hard
got
cant
colleagu
done
effort
first
job
person
struggl
literide
right
find
ask
reali
good
back
team
improv
tri
sleep
one
focus
peopl
never
anyth
task
mistak
best
see
help
small
manag
even
take
start

Topic 1

success best community like think
thought finish love study pursue get also
advice feel way need current plan parent just
even business lost good graduate program
decide become experitake learn
life can path opportun thank
don sure become educ lot well thing much will abl
passion completfind tri teachtime
math start medic relat someth
comput cours job dontnurs research sinc idea
field amp consid help law
choos high master financ futur
major appli colleg health suggest
scienc want onlin option

WORD CLOUDS

Topic 4

year current
field term transit
project better consid advic path
long cours look manag
can busi growth goal
e degree full
one level everyon
seem sure know process
take secur recent
new want month
engin will help base
like around code next start
advanc now
appreci offer complet salari insight thank
makeget role similar loverequir focus
technolog experi internship much time
job softwar also design interest
etc graduat feel includ entri
data chang

interview

company

job

team

time

week

month

current

now

email

two

said

can

next

anoth

sinc

ago

recent

promot

betteradvic

thing

accept

situat

back

left

worri

due

senior

previous

first

tell

call

make

contract

put

well

appli

pay

just

one

let

come

make

quit

good

year

say

busi

time

receiv

like

leav

also

hour

want

need

hire

however

ask

wait

client

still

take

know

sale

respons

boss

last

end

already

think

person

look

move

got

realli

recruit

offic

perform

give