**Set 1**

**CA2:** BYOD Practical

**Course Code:** INTM577

**Allocation Date:** 06-03-2023

**Submission Date:** 06-03-2023

**Reg No: 12202342 Roll No : A08**

**Max. Marks: 50**

**IMPORTANT GUIDELINES:**

1. Each question is of 10 marks.
2. Copied cases directly get zero marks.
3. Maximum time to upload the file is 6th February 2023 2:00pm. After the deadline zero marks will be awarded.

Q1. YouTube has gained an audience of billions of users including educators and scholars. While the academic literature provides some evidence that YouTube has been studied and written about, little is known about priorities for YouTube research. A study represents trend analysis and content analysis method to obtain data on research topics, issues category, research settings and sampling, research design, research method and data analysis on articles published regarding YouTube in selected journals.

Analyse the sentiment of Audience for youtube link: “[Bill Gates on Covid Vaccine, Climate Change, Future of Electric Vehicles - YouTube](https://www.youtube.com/watch?v=j_mVkbaodyo)”. Analyse the data using R and answer the following questions.

Source code:

---

title: "Budget Case Study"

output:

flexdashboard::flex\_dashboard:

orientation: columns

vertical\_layout: fill

---

```{r setup, include=FALSE}

library(flexdashboard)

library(tm)

library(syuzhet)

library(ggplot2)

library(wordcloud)

library(tuber) # youtube API

library(magrittr) # Pipes %>%, %T>% and equals(), extract().

library(tidyverse) # all tidyverse packages

library(purrr) # package for iterating/extracting data

```

Column {data-width=450}

-----------------------------------------------------------------------

### Chart A

```{r}

client\_id <- "375771870105-dlhu23ucu0r42qc029v0a9jhkjabaqml.apps.googleusercontent.com"

client\_secret <- "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"

# use the youtube oauth

yt\_oauth(app\_id = client\_id,

app\_secret = client\_secret,token= '')

data1= get\_all\_comments(video\_id="j\_mVkbaodyo")

b<-write.csv(data1,file="T5.csv",row.names=FALSE)

a<-read.csv("D:/Class files/Module 3/INTM577(BA2)/T5.csv")

##### tm ####

library(tm)

corpus = iconv(a$textDisplay, "latin1", "UTF-8")

corpus<- Corpus(VectorSource(corpus))

toSpace <- content\_transformer(function (x , pattern ) gsub(pattern, " ", x))

docs=corpus

docs <- tm\_map(docs, toSpace, "/")

docs <- tm\_map(docs, toSpace, "@")

docs <- tm\_map(docs, toSpace, "\\|")

corpus=docs

corpus<- tm\_map(corpus,tolower)

corpus<-tm\_map(corpus,removePunctuation)# remove puntuations like , .

corpus<- tm\_map(corpus,removeNumbers)

cleanset<-tm\_map(corpus,removeWords,stopwords('english'))# remove common words

removeURL<- function(x)gsub('http[[:alnum:]]=','',x)

cleanset<-tm\_map(cleanset,content\_transformer(removeURL))

x=cleanset

tdm<-TermDocumentMatrix(cleanset)

tdm # display information

```

Column {.tabset}

-----------------------------------------------------------------------

### Sentimental analysis

```{r}

tdm<-as.matrix(tdm)

v=sort(rowSums(tdm))

library(wordcloud)

w<-data.frame(names(v),v)

colnames(w)<-c('word','freq')

set.seed(1234)

wordcloud(words=w$word,freq=w$freq)

write.csv(tdm,"tdm.csv")

#### sentimental analysis #####

library(syuzhet)

data=read.csv("tdm.csv")

mysentiment\_tech<-get\_nrc\_sentiment(data$X)

```

### Visualization

```{r}

#calculationg total score for each sentiment

Sentimentscores\_tech<-data.frame(colSums(mysentiment\_tech[,]))

names(Sentimentscores\_tech)<-"Score"

Sentimentscores\_tech<-cbind("sentiment"=rownames(Sentimentscores\_tech),Sentimentscores\_tech)

rownames(Sentimentscores\_tech)<-NULL

Sentimentscores\_tech\_r=Sentimentscores\_tech

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

library(ggplot2)

ggplot(data=Sentimentscores\_tech,aes(x=sentiment,y=Score))+

geom\_bar(aes(fill=sentiment),stat = "identity")+

theme(legend.position="none")+

xlab("Sentiments")+ylab("scores")+ggtitle("R Program")

```

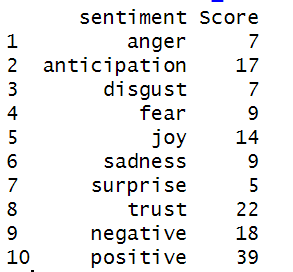
### High frequency

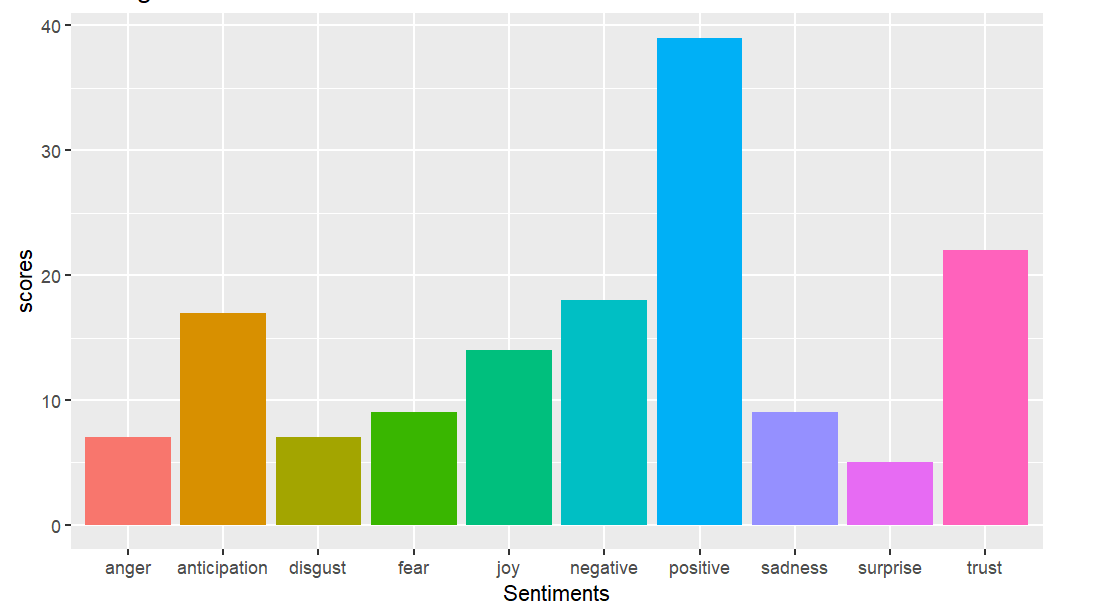
```{R}

tail(w,n=5)

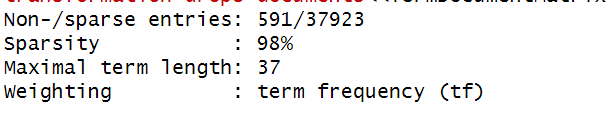
```

1. Display the sentimental analysis.

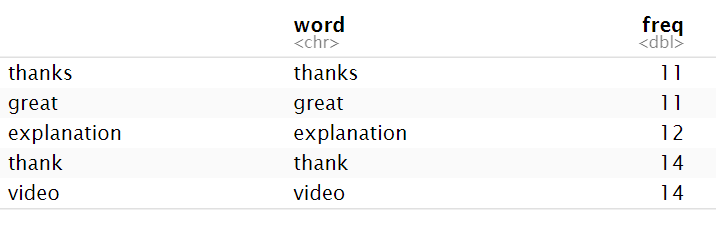


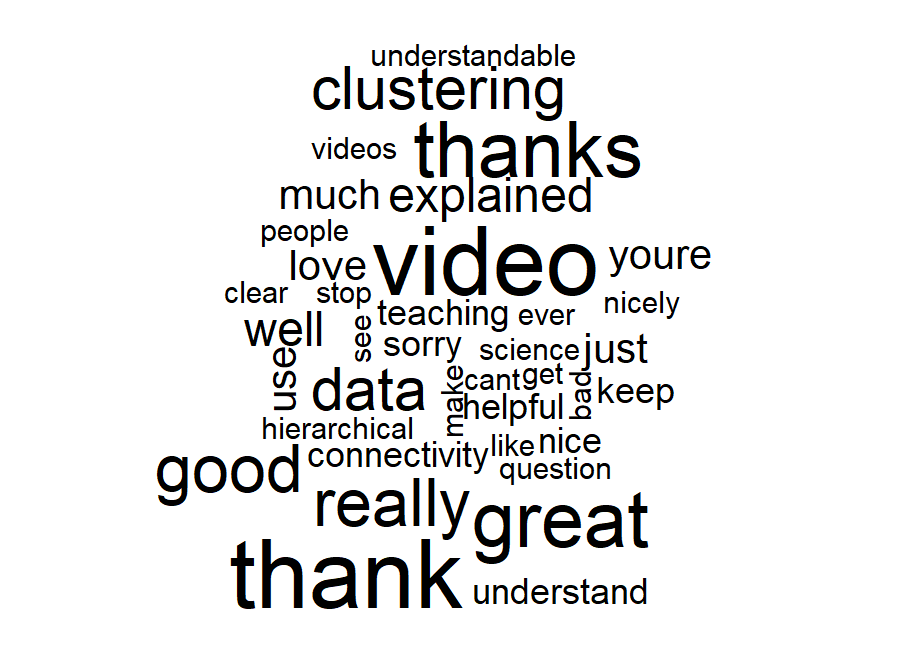


1. Perform the pre-processing and analyse the data.

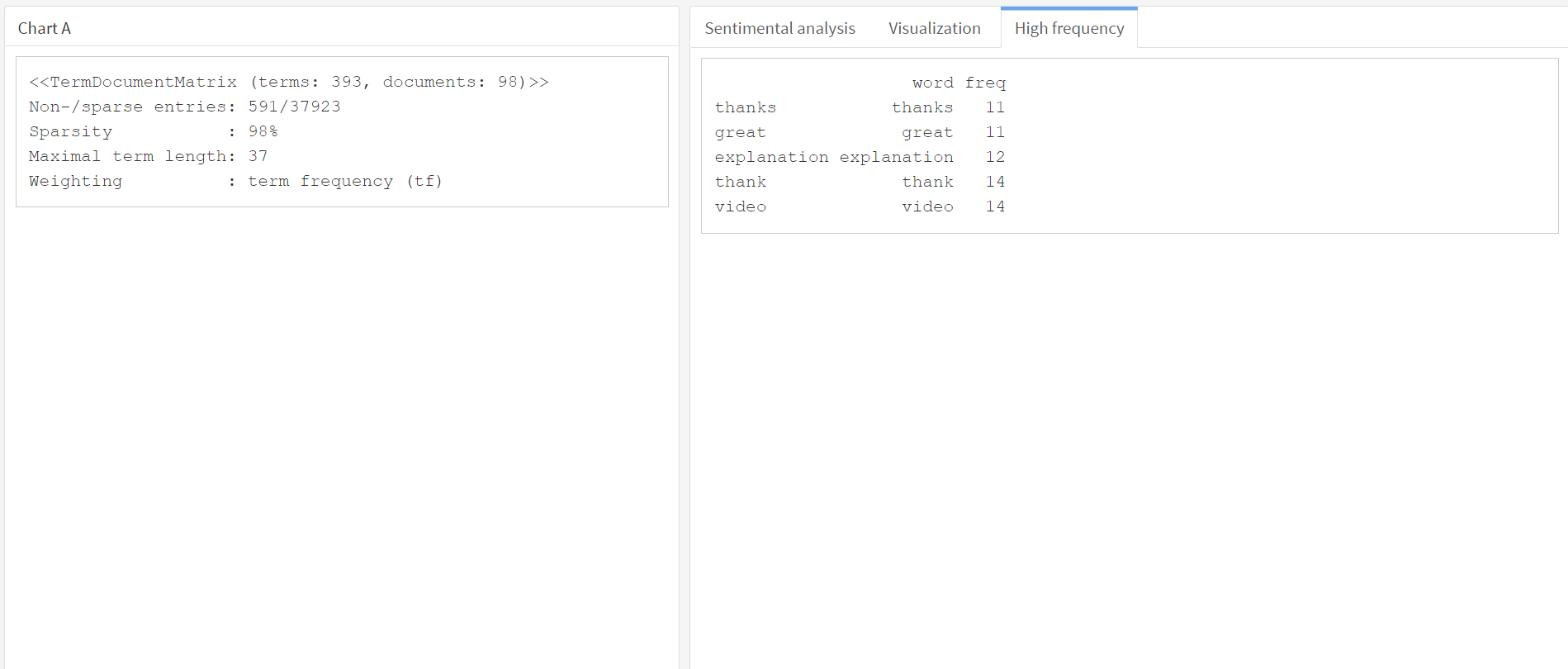


1. Which keywords used more frequently (using WORDCLOUD ).





1. Create a dashboard representing the above scenarios.



5. Analyse the data and mention 3-4 points to conclude the complete analysis.

1. Based on the analysis, There was the positive response as compared to negative.

2. More trust is available in the video.

3. The client id and secret is must from youtube developer account.

4. Based on analysis, the video contain only few comments.

