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". 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

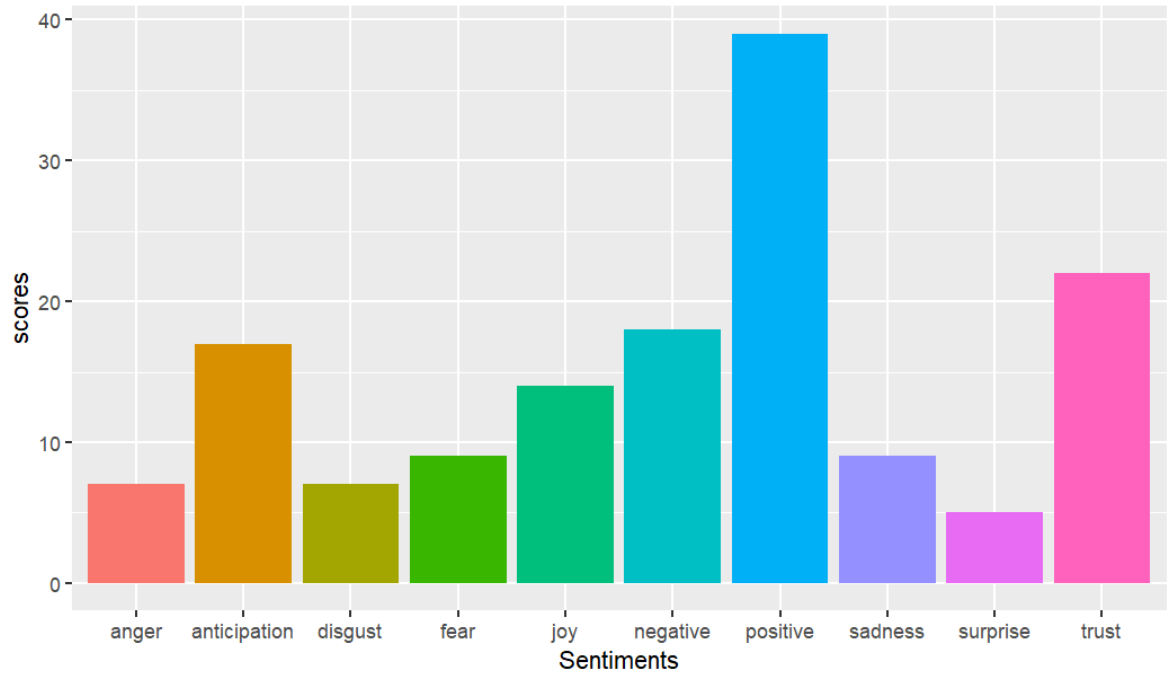
```{R}

tail(w,n=5)

```

1. Display the sentimental analysis.

```
      sentiment Score
1         anger     7
2  anticipation    17
3       disgust     7
4          fear     9
5           joy    14
6       sadness     9
7      surprise     5
8         trust    22
9      negative    18
10     positive    39
```

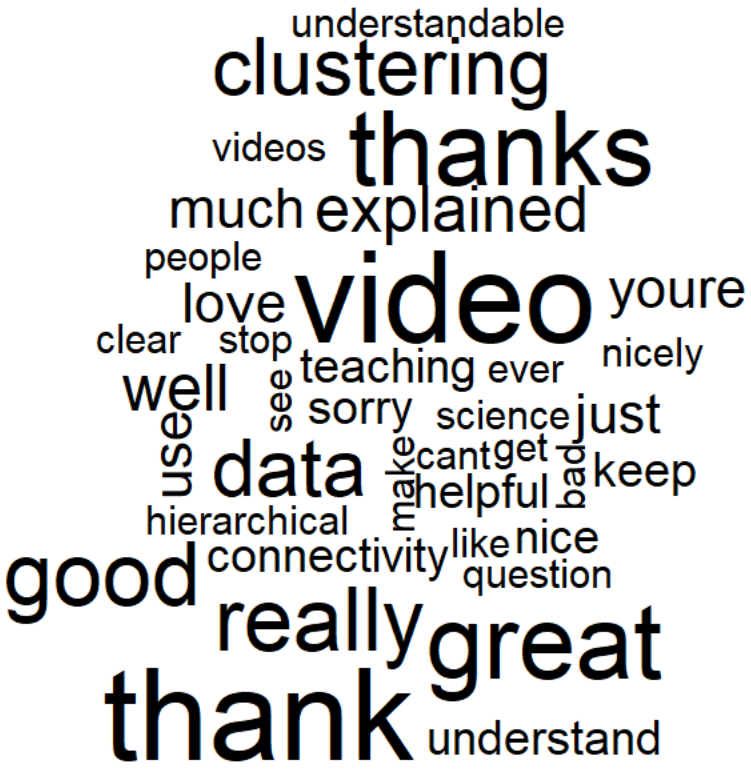


2. Perform the pre-processing and analyse the data.
```

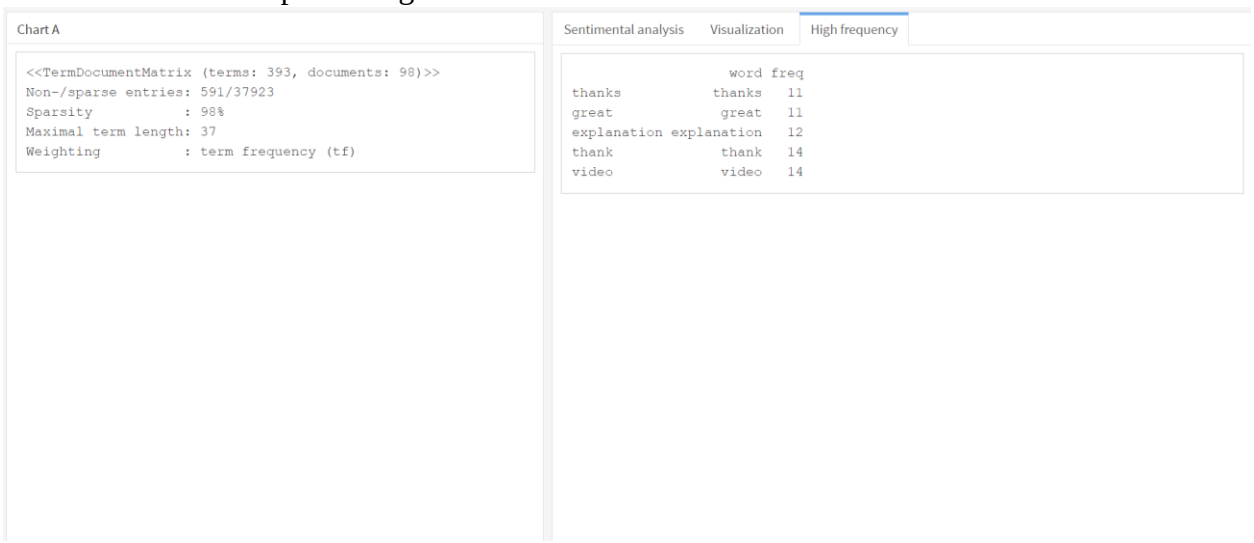
```
Non-/sparse entries: 591/37923
Sparsity           : 98%
Maximal term length: 37
Weighting          : term frequency (tf)
```

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

	word <chr>	freq <dbl>
thanks	thanks	11
great	great	11
explanation	explanation	12
thank	thank	14
video	video	14



4. Create a dashboard representing the above scenarios.



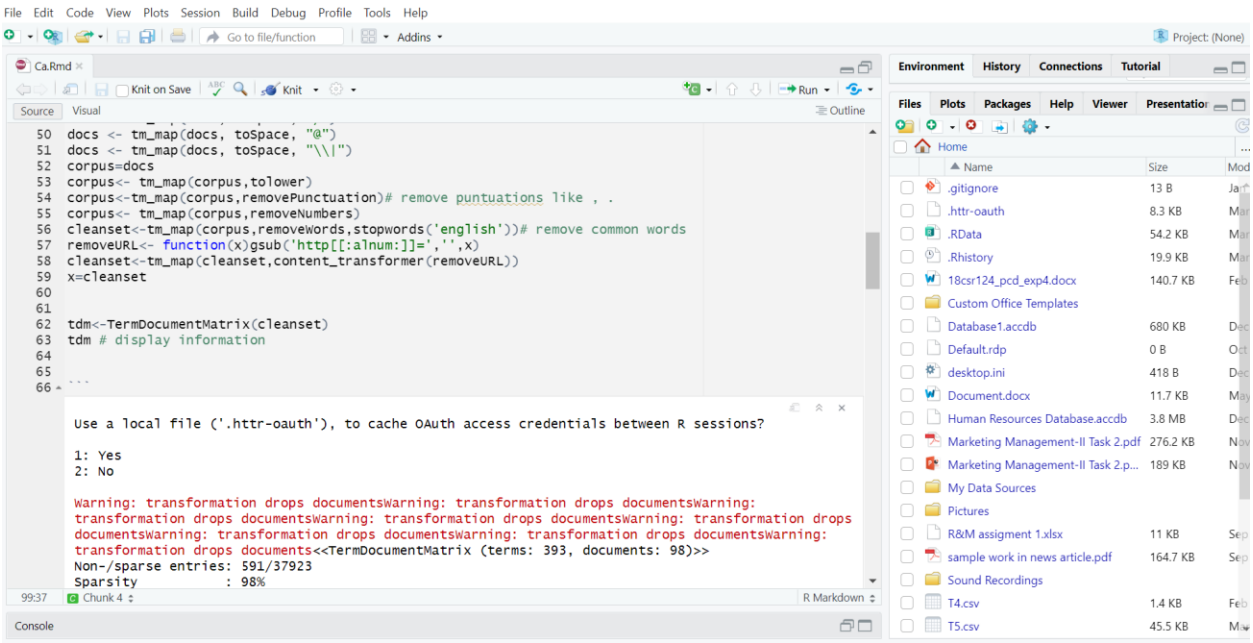
5. Analyse the data and mention 3-4 points to conclude the complete analysis.

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.


File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
Project: (None)
Ca.Rmd
Knit on Save
Knit
Run
Source
Visual
Outline
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
```
Use a local file ('.httr-oauth'), to cache OAuth access credentials between R sessions?
1: Yes
2: No
Warning: transformation drops documentsWarning: transformation drops documentsWarning: transformation drops documentsWarning: transformation drops documentsWarning: transformation drops documentsWarning: transformation drops documentsWarning: transformation drops documentsWarning: transformation drops documents<<TermDocumentMatrix (terms: 393, documents: 98)>>
Non-/sparse entries: 591/37923
Sparsity : 98%
99:37
Chunk 4
R Markdown
Console
Environment
History
Connections
Tutorial
Files
Plots
Packages
Help
Viewer
Presentation
Home
Name
Size
.gitignore 13 B
.httr-oauth 8.3 KB
.RData 54.2 KB
.Rhistory 19.9 KB
18csr124_pcd_exp4.docx 140.7 KB
Custom Office Templates
Database1.accdb 680 KB
Default.rdp 0 B
desktop.ini 418 B
Document.docx 11.7 KB
Human Resources Database.accdb 3.8 MB
Marketing Management-II Task 2.pdf 276.2 KB
Marketing Management-II Task 2.p... 189 KB
My Data Sources
Pictures
R&M assigment 1.xlsx 11 KB
sample work in news article.pdf 164.7 KB
Sound Recordings
T4.csv 1.4 KB
T5.csv 45.5 KB
```