

Final Project(Group Activity)

BSIT-2B

2023-12-14

```
library(rvest)
```

```
## Warning: package 'rvest' was built under R version 4.3.2
```

```
library(stringr)
```

```
## Warning: package 'stringr' was built under R version 4.3.2
```

```
library(tm)
```

```
## Warning: package 'tm' was built under R version 4.3.2
```

```
## Loading required package: NLP
```

```
library(wordcloud)
```

```
## Warning: package 'wordcloud' was built under R version 4.3.2
```

```
## Loading required package: RColorBrewer
```

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.3.2
```

```
##
```

```
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
##     filter, lag
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.3.2
```

```
##
```

```
## Attaching package: 'ggplot2'
```

```
## The following object is masked from 'package:NLP':
```

```
##
```

```
##      annotate
```

```
library(syuzhet)
```

```
## Warning: package 'syuzhet' was built under R version 4.3.2
```

```
extract_reviews <- function(url) {  
  page <- read_html(url)  
  
  review_text <- page %>% html_nodes(".text_content") %>%  
    html_text()  
  
  reviews_df <- data.frame(  
    Text_Review = review_text[1:10]  
  )  
  
  return(reviews_df)  
}
```

```
base_url <- "https://www.airlinequality.com/airline-reviews/porter-airlines/page/"
```

```
all_reviews_df <- data.frame()
```

```
for (page_number in 1:37) {  
  url <- paste0(base_url, page_number, "/")  
  
  reviews_page <- extract_reviews(url)  
  
  all_reviews_df <- rbind(all_reviews_df, reviews_page)  
}
```

```
all_reviews_df <- all_reviews_df[1:300,]  
all_reviews_df <- data.frame(Text_Review = all_reviews_df  
View(all_reviews_df)
```

```
write.csv(all_reviews_df, file = "reviewDataset.csv", row.names = FALSE)
```

```
#data cleaning
```

```
cleaned_text <- all_reviews_df$Text_Review
```

```
#clean text using gsub
```

```
cleaned_text <- str_replace_all(cleaned_text, " Trip Verified | Not Verified |Not Verified | Verified
```

```

cleaned_text <- gsub("\\\\", "", cleaned_text)
cleaned_text <- gsub("\\s+", " ", cleaned_text)
cleaned_text <- gsub("[:punct:]", "", cleaned_text)
cleaned_text <- gsub("[:digit:]", "", cleaned_text)
cleaned_text <- str_replace_all(cleaned_text, "[^a-zA-Z0-9]", " ")

# display the cleaned text example...
head(cleaned_text)[1:5]

```

```

## [1] " Porter destroyed my luggage and then did not make it right at all Three large holes two wheels
## [2] " Porter is a great option if you want a flight delay Ive never seen an airline get delayed as m
## [3] " Flight was at pm Loaded the flight told to unload the flight given no communication whatsover
## [4] " Best food Ive ever had on a plane and Im a chef Fantastic service nice new plane impeccably cl
## [5] " I cannot recommend to anyone to fly with Porter We had one flight with them on October a di

```

```

# create a Corpus object
wordCorpus <- Corpus(VectorSource(cleaned_text))

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

```

```

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

```

```

wordCorpus <- tm_map(wordCorpus, removePunctuation)

```

```

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

```

```

wordCorpus <- tm_map(wordCorpus, removeNumbers)

```

```

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

```

```

wordCorpus <- tm_map(wordCorpus, removeWords, stopwords("english"))

```

```

## Warning in tm_map.SimpleCorpus(wordCorpus, removeWords, stopwords("english")):
## transformation drops documents

```

```

wordCorpus <- tm_map(wordCorpus, removeWords, stopwords("SMART"))

```

```

## Warning in tm_map.SimpleCorpus(wordCorpus, removeWords, stopwords("SMART")):
## transformation drops documents

```

```

wordCorpus <- tm_map(wordCorpus, stripWhitespace)

```

```

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

```

```
#additional removal of specific words
```

```
wordCorpus <- tm_map(wordCorpus, removeWords, c("rep", "checkin", "tarmac", "m", "ins", "s", "inperson"
```

```
## Warning in tm_map.SimpleCorpus(wordCorpus, removeWords, c("rep", "checkin", :  
## transformation drops documents
```

```
wordCorpus <- tm_map(wordCorpus, removeWords, c("fo", "dec", "embraer", "ee", "ckal", "c", "st", "tha"
```

```
## Warning in tm_map.SimpleCorpus(wordCorpus, removeWords, c("fo", "dec",  
## "embraer", : transformation drops documents
```

```
wordCorpus <- tm_map(wordCorpus, removeWords, c("ons", "suff", "delayedcancelled", "possibleitsurely",
```

```
## Warning in tm_map.SimpleCorpus(wordCorpus, removeWords, c("ons", "suff", :  
## transformation drops documents
```

```
wordCorpus <- tm_map(wordCorpus, removeWords, c("q", "bombardier", "turboprops", "drinkssnacksalcohol",
```

```
## Warning in tm_map.SimpleCorpus(wordCorpus, removeWords, c("q", "bombardier", :  
## transformation drops documents
```

```
wordCorpus <- tm_map(wordCorpus, removeWords, c("ste", "immigrationcustom", "hyatt", "x", "ottawatoront
```

```
## Warning in tm_map.SimpleCorpus(wordCorpus, removeWords, c("ste",  
## "immigrationcustom", : transformation drops documents
```

```
#some cleaned content
```

```
wordCorpus$content[1:50]
```

```
## [1] " porter destroyed luggage make large holes wheels ripped spoke airport told call person phone  
## [2] " porter great option flight delay ive airline delayed porter air canada dollars extra ticket w  
## [3] " flight pm loaded flight told unload flight communication whatsoever lodging flight told flight  
## [4] " food ive plane im chef fantastic service nice plane impeccably clean pretty comfortable seat  
## [5] " recommend fly porter flight october direct flight layovers lost checked bag checked bag toron  
## [6] " give raving review porter experience pleasant stepping back charming whimsical modern conveni  
## [7] " family member passed unexpectedly flight called cancel return fare money charge seat wont mak  
## [8] " flown porter airlines times friendly staff great flight service food drinks delayed flight tim  
## [9] " airline doesnt electronic boarding pass purchase reserve ticket basic ticket force counter pi  
## [10] " amazing airline time middle seat food choices purchase amazing healthy staff great porter mak  
## [11] " wanted great airline lots room trouble friendly staff toronto halifax return great journeys"  
## [12] " disappointed porter airlines time flying company connecting flight italy arrived montreal pm p  
## [13] " flight short comfortable onboard small bag passengers stored luggage received bag broken hand  
## [14] " smooth flight crew fantastic star food great havent good food plane loved porter "  
## [15] " wonderful flight free internet serve snacks drinks red white wine types juices coffe tea free  
## [16] " complaint groundcrew airport wife assistance wheel chair assistance requested acknowledged t  
## [17] " porter terrible airline originally excited high hopes extremely disappointed flight delayed m  
## [18] " flight cancelled due weather minutes supposed flight booked day porter staff informed circum  
## [19] " worst airline ive experienced cancelled flight hours flight time rebooked night option direct  
## [20] " problems happen plane sitting hour time announce canceling flight problem washroom ridiculous
```

```
## [21] " porter cancelling flights thunder  mondays morning flight cancelled sunday pm flight crew sho
## [22] " air canada timmins toronto fly route frequently airlines porter  porter review relate quality
## [23] " time fly porter airlines heard airline family member confused book flight choose  air canada
## [24] " time flying porter airline bit worried airfare expensive airlines flights porter airline "
## [25] " booking flight airline context family vacation  visit  arrived airport flight couple minutes
## [26] " airline worried destination time flight delayed causing miss connection didnt rebook connecti
## [27] " wonderful staff met staff inspired write review ordinary flight similar flights trouble case p
## [28] " flight united vancouver  canceled due weather rebooked days vancouver catch train frantic sea
## [29] " person wheelchair needing special assistance due unpredictable weather day chose fly home fli
## [30] " flight cancelled due weather conditions rescheduling suggestion didnt work cancelled booking
## [31] " scheduled fly quebec city toronto sunday june pm flight delayed minutes cancelled explanation
## [32] " time flying porter airline business charging bring carry understand budget airlines ryanair p
## [33] " flight  june great service start finish flight attendants accommodating attentive experience
## [34] " experience porter amazing day flight cancel fathers flight join trip anymore realized ended c
## [35] " frequent business traveler experienced poor customer service porter flight boston ottawa cance
## [36] " staff attentive caring checked frequently travelling year made feel extremely cared wonderful
## [37] " cabin staff pleasant pm porter flight back canceled due maintenance issues rebooked pm flight
## [38] " flew vancouver halifax connecting flight toronto april return flight vancouver april booked f
## [39] " time flying porter nightmare landed ottawa flight cancelled pm delayed due mechanical issues v
## [40] " porter airlines quick efficient rectifying issues reimbursed clothing needed purchased baggage
## [41] " airline mess called days prior flight questions baggage allowed wasnt warned expect lose bagg
## [42] " terrible experience porter lost checked bag  trip whistler  inside checked bag worth snowboard
## [43] " originally booked flight paid extra fees ensure make change cancel receive full refund litera
## [44] "  cancelled flight  halifax toronto due coming storm toronto cancelled flight  leaving hours
## [45] " ridiculously unreliable hours flight received notice cancelled ticket transferred flight leav
## [46] " cancelled outbound return flights informed change outbound flight offer flight arrived day re
## [47] " booked tickets trip toronto calgary weeks ago coming march break days departure date receive
## [48] " booked fly calgary toronto march flight cancelled due bad weather toronto email confirmation
## [49] " porter airlines unreliable flight cancelled rebooked flight days original customer service am
## [50] " worst airline company customer service completely professional hated experience recommend pe
```

```
#plotting using wordcloud in reviews
# Generating a word cloud based on the words extracted from airline reviews. The term "flight" emerges
set.seed(1234)
wordcloud(words = wordCorpus$content, min.freq = 1,
          max.words = 200, random.order = FALSE, rot.per = 0.50,
          color = brewer.pal(9, "Set1"))
```

```
## Warning in tm_map.SimpleCorpus(corpus, tm::removePunctuation): transformation
## drops documents
```

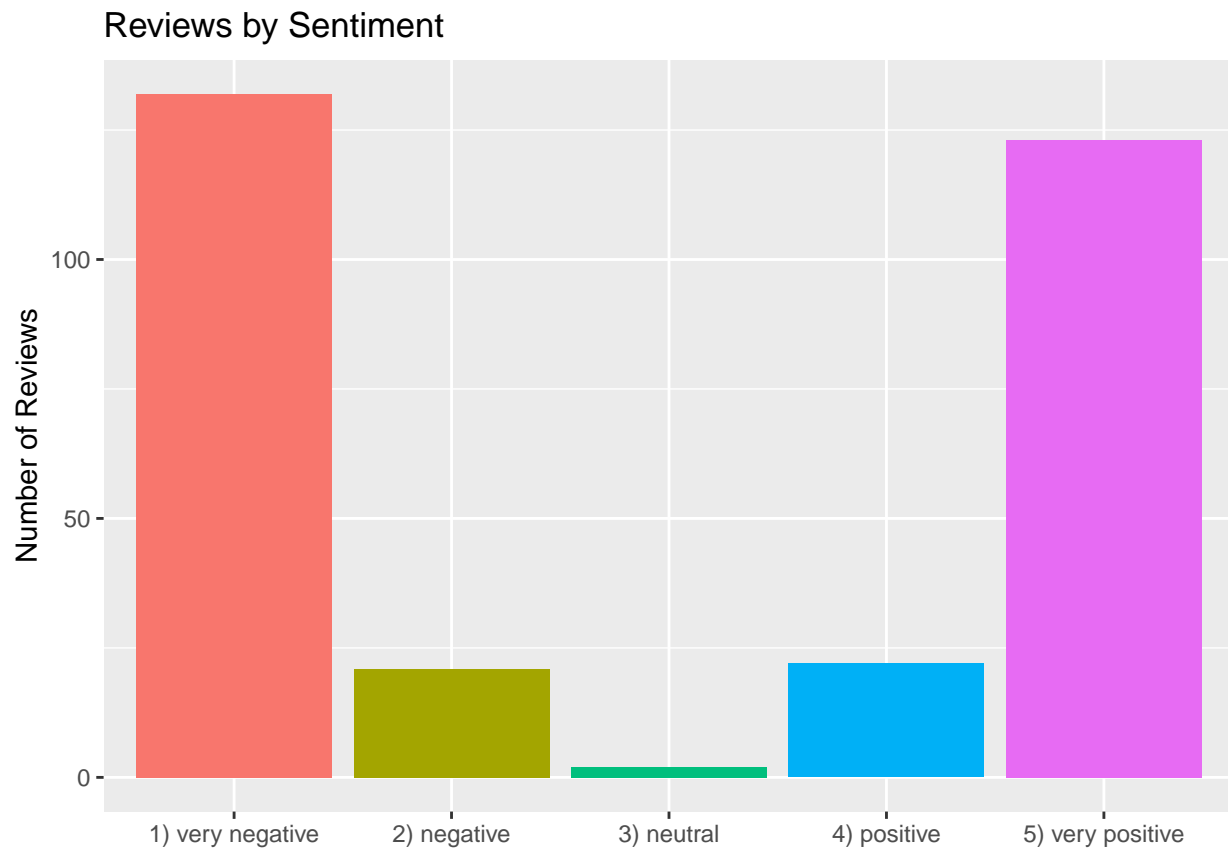
```
## Warning in tm_map.SimpleCorpus(corpus, function(x) tm::removeWords(x,
## tm::stopwords())): transformation drops documents
```



```
reviews$reviewSentiment <- sapply(reviews$reviewSentiment, encodeSentiment)

# Data visualization
#In the bar plot, we observe that the distribution of sentiment scores is skewed towards the extremes.

ggplot(reviews, aes(reviews$reviewSentiment, fill = reviews$reviewSentiment)) +
  geom_bar() +
  theme(legend.position = "none", axis.title.x = element_blank()) +
  ylab("Number of Reviews") +
  ggtitle("Reviews by Sentiment")
```



```
count_sentiment <- reviews %>%
  count(reviews$reviewSentiment)

count_sentiment
```

```
## reviews$reviewSentiment n
## 1 1) very negative 132
## 2 2) negative 21
## 3 3) neutral 2
## 4 4) positive 22
## 5 5) very positive 123
```