

# CS102 Activity2

2024-02-06

```
library(dplyr)
```

```
##  
## 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(polite)  
library(xml2)  
library(magrittr)  
library(rvest)  
library(httr)
```

```
Keys <- c('g4w6ddbmqyzdo6ic4oxwjnr3smwajy3ykdd6hpaxb7yv35pjt6udc2oe4vxnb4dy3ik6g6vacybfohautcdhjuwL  
'g4w6ddbmqyzdo6ic4oxwjnzxrps44bzt2mnt777natepeud5pjt6uds6oa4vlnrnb4dwpdyddjbidtthzjvknhdvuacda  
'g4w6ddbmqyzdo6ic4oxwjnzsrpsmybzz3motv6hfahep2tttfjrk4d2yoqyfjmbods2dhi6gej2z6drga52tx4yji  
'g4w6ddbmqyzdo6ic4oxwjnbzrhtmqqby3ektr6hmb3d74wt5pjt6udc7om2vjmjib4d3sbiuib73l2uugyiqamq4w7v  
'g4w6ddbmqyzdo6ic4oxwjnbqlr4ubjy3ektd4poadb76tttfjrk4dcyouyvxpxzoakcnhar44ks4qjs7yrcpyxmb3ja  
'g4w6ddbmqyzdo6ic4oxwjnbtrlu4ucj634ndd6xeb7b7gttfjrk4dssoa4vvnrjoav7lvtwyifymnhf33vpv465m4lq  
'g4w6ddbmqyzdo6ic4oxwjnjyrtq4qab63qktx77ialapuv35pjt6ucs4ouyfvnrdb4dsn5kzynt3hwwniviy5dpiuepy  
'g4w6ddbmqyzdo6ic4oxwjnjwqpu4qbr53int56hmb7d6mqbdf4q26bk5peyvxjchswwsblrza5z6dhvoubu5gs7jym',  
'g4w6ddbmqyzdo6ic4oxwjnjxrhu42bz63imdv6ppapc7gu35pjt6udk2oizf7prob4dyfngiexvmcltazeibjono6uvh  
'g4w6ddbmqyzdo6ic4oxwjnjurxr42bz23eotd6hla7a7qvdgoq366ss5oyyflmrjdr4k4iy5dixztbdig3djhuzqdazd  
'g4w6ddbmqyzdo6ic4oxwjnjurlsmwbr43intz7xmape7gulhoq366ss3oezv5nzdj4k5kyby542dlm332rbqvxyzrur  
'g4w6ddbmqyzdo6ic4oxwjnjvr3qm2bb33ent56pka7b74v3foq366ss3oezv5mjidb4k5kjwnydlc3r3nzm5kw7sx3ag
```

```
titles <- character(0)  
names <- character(0)  
dates <-character(0)  
ratings <- character(0)  
reviews <- character(0)
```

```
for(i in 1:12){  
  url <- paste('https://www.imdb.com/title/tt0167260/reviews/_ajax?&paginationKey=', Keys[i], sep = "  
  
  session <- bow(url,  
                  user_agent = "Educational")  
  
  session  
  #Get the Parent Node  
  parent <- scrape(session) %>%  
    html_elements('div.review-container')
```

```

reviewTitle <- parent %>%
  html_node('.title') %>%
  html_text()

reviewerName <- parent %>%
  html_node('.display-name-link') %>%
  html_text()

dateUploaded <- parent %>%
  html_node('.review-date') %>%
  html_text()

userRating <- parent %>%
  html_node('.ipl-ratings-bar') %>%
  html_text()

reviewContent <- parent %>%
  html_node('.content') %>%
  html_text()

titles <- c(titles,reviewTitle)
names <- c(names, reviewerName)
dates <- c(dates, dateUploaded)
ratings <- c(ratings, userRating)
reviews <- c(reviews, reviewContent)

}

df <- data.frame(titles,names,dates,ratings,reviews)

```