CS102_ACTIVITY3_SADSAD

BSIT-2B

2024-02-13

##ACTIVITY 3

```r
library(rvest)
```

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

```r
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
```

```r
library(writexl)
```

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

```r
library(polite)
```

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

```r
#1. Phantom Thread

session <- bow(url = 'https://www.imdb.com/title/tt5776858/reviews?ref_=tt_urv',
               user_agent = "Student's Demo Educational")

session
```

```
## <polite session> https://www.imdb.com/title/tt5776858/reviews?ref_=tt_urv
##      User-agent: Student's Demo Educational
##      robots.txt: 35 rules are defined for 3 bots
##    Crawl delay: 5 sec
##    The path is scrapable for this user-agent

session_scrape <- scrape(session)

scrape_reviews <- function(page_url) {

  reviews_page <- read_html(page_url)
  names <- reviews_page %>%
    html_nodes(".display-name-link") %>%
    html_text()

  dates <- reviews_page %>%
    html_nodes("span.review-date")%>%
    html_text()

  ratings <- reviews_page %>%
    html_nodes("span.rating-other-user-rating") %>%
    html_text()

  content_reviews <- reviews_page %>%
    html_nodes("a.title") %>%
    html_text()

  reviews <- reviews_page %>%
    html_nodes(".text.show-more__control") %>%
    html_text()

  phantom_thread_df = data.frame( Movie_name = c('Phantom Thread'),
                                  User_Name = names[1:25],
                                  Date = dates[1:25],
                                  User_Rating = ratings[1:25],
                                  Content_Review = content_reviews[1:25],
                                  Reviews = reviews[1:25]
  )}

phantom_thread_urls <- c( 'https://www.imdb.com/title/tt5776858/reviews?ref_=tt_urv',
                          'https://www.imdb.com/title/tt5776858/reviews/_ajax?&paginationKey=g4w6ddbmqy;
                          'https://www.imdb.com/title/tt5776858/reviews/_ajax?&paginationKey=g4w6ddbmqy;
                          'https://www.imdb.com/title/tt5776858/reviews/_ajax?&paginationKey=g4w6ddbmqy;
                          'https://www.imdb.com/title/tt5776858/reviews/_ajax?&paginationKey=g4w6ddbmqy;
                          'https://www.imdb.com/title/tt5776858/reviews/_ajax?&paginationKey=g4w6ddbmqy;
                          'https://www.imdb.com/title/tt5776858/reviews/_ajax?&paginationKey=g4w6ddbmqy;
                          'https://www.imdb.com/title/tt5776858/reviews/_ajax?&paginationKey=g4w6ddbmqy;
                          'https://www.imdb.com/title/tt5776858/reviews/_ajax?&paginationKey=g4w6ddbmqy;
                          'https://www.imdb.com/title/tt5776858/reviews/_ajax?&paginationKey=g4w6ddbmqy;
                          'https://www.imdb.com/title/tt5776858/reviews/_ajax?&paginationKey=g4w6ddbmqy;
                          'https://www.imdb.com/title/tt5776858/reviews/_ajax?&paginationKey=g4w6ddbmqy;
)

all_reviews <- lapply(phantom_thread_urls, scrape_reviews)
```

```r
final_all_reviews <- do.call(rbind, all_reviews)

View(final_all_reviews)



#####################################################################################
#2. Alien (1979)




session2<- bow(url = 'https://www.imdb.com/title/tt0078748/reviews/?ref_=tt_ov_rt',
               user_agent = "Student's Demo Educational")

session2
```

```
## <polite session> https://www.imdb.com/title/tt0078748/reviews/?ref_=tt_ov_rt
##      User-agent: Student's Demo Educational
##       robots.txt: 35 rules are defined for 3 bots
##     Crawl delay: 5 sec
##    The path is scrapable for this user-agent
```

```r
session_scrape2 <- scrape(session2)

scrape_reviews2 <- function(page_url) {

  reviews_page <- read_html(page_url)
  names2 <- reviews_page %>%
    html_nodes(".display-name-link") %>%
    html_text()

  dates2 <- reviews_page %>%
    html_nodes("span.review-date")%>%
    html_text()

  ratings2 <- reviews_page %>%
    html_nodes("span.rating-other-user-rating") %>%
    html_text()

  content_reviews2 <- reviews_page %>%
    html_nodes("a.title") %>%
    html_text()

  reviews2 <- reviews_page %>%
    html_nodes(".text.show-more__control") %>%
    html_text()

  AlienMOVIE_df = data.frame(  Movie_name = c('Alien (1979)'),
                               User_Name = names2[1:25],
                               Date = dates2[1:25],
                               User_Rating = ratings2[1:25],
                               Content_Review = content_reviews2[1:25],
                               Reviews = reviews2[1:25]
```

```r
  )}

Alien_MOVIE_urls <- c( 'https://www.imdb.com/title/tt0078748/reviews/?ref_=tt_ov_rt',
                       'https://www.imdb.com/title/tt0078748/reviews/_ajax?&paginationKey=g4w6ddbmqyzdoC
                       'https://www.imdb.com/title/tt0078748/reviews/_ajax?&paginationKey=g4w6ddbmqyzdoC
                       'https://www.imdb.com/title/tt0078748/reviews/_ajax?&paginationKey=g4w6ddbmqyzdoC
                       'https://www.imdb.com/title/tt0078748/reviews/_ajax?&paginationKey=g4w6ddbmqyzdoC
                       'https://www.imdb.com/title/tt0078748/reviews/_ajax?&paginationKey=g4w6ddbmqyzdoC
                       'https://www.imdb.com/title/tt0078748/reviews/_ajax?&paginationKey=g4w6ddbmqyzdoC
                       'https://www.imdb.com/title/tt0078748/reviews/_ajax?&paginationKey=g4w6ddbmqyzdoC
                       'https://www.imdb.com/title/tt0078748/reviews/_ajax?&paginationKey=g4w6ddbmqyzdoC
                       'https://www.imdb.com/title/tt0078748/reviews/_ajax?&paginationKey=g4w6ddbmqyzdoC
                       'https://www.imdb.com/title/tt0078748/reviews/_ajax?&paginationKey=g4w6ddbmqyzdoC
                       'https://www.imdb.com/title/tt0078748/reviews/_ajax?&paginationKey=g4w6ddbmqyzdoC
)

all_reviews2 <- lapply(Alien_MOVIE_urls, scrape_reviews2)

final_all_reviews2 <- do.call(rbind, all_reviews2)

View(final_all_reviews2)


###############################################################################
#3. Salt Burn


session3<- bow(url = 'https://www.imdb.com/title/tt17351924/reviews?ref_=tt_urv',
               user_agent = "Student's Demo Educational")

session3
```

```
## <polite session> https://www.imdb.com/title/tt17351924/reviews?ref_=tt_urv
##      User-agent: Student's Demo Educational
##      robots.txt: 35 rules are defined for 3 bots
##    Crawl delay: 5 sec
##   The path is scrapable for this user-agent
```

```r
session_scrape3 <- scrape(session3)

scrape_reviews3 <- function(page_url) {

  reviews_page <- read_html(page_url)
  names3 <- reviews_page %>%
    html_nodes(".display-name-link") %>%
    html_text()

  dates3 <- reviews_page %>%
    html_nodes("span.review-date")%>%
    html_text()

  ratings3 <- reviews_page %>%
    html_nodes("span.rating-other-user-rating") %>%
```

```r
    html_text()

  content_reviews3 <- reviews_page %>%
    html_nodes("a.title") %>%
    html_text()

  reviews3 <- reviews_page %>%
    html_nodes(".text.show-more__control") %>%
    html_text()

  Saltburn_MOVIE_df = data.frame( Movie_name = c('Salt Burn'),
                                  User_Name = names3[1:25],
                                  Date = dates3[1:25],
                                  User_Rating = ratings3[1:25],
                                  Content_Review = content_reviews3[1:25],
                                  Reviews = reviews3[1:25]
  )}

Saltburn_MOVIE_urls <- c( 'https://www.imdb.com/title/tt17351924/reviews?ref_=tt_urv',
                          'https://www.imdb.com/title/tt17351924/reviews/_ajax?&paginationKey=g4w6ddbmqy
                          'https://www.imdb.com/title/tt17351924/reviews/_ajax?&paginationKey=g4w6ddbmqy
                          'https://www.imdb.com/title/tt17351924/reviews/_ajax?&paginationKey=g4w6ddbmqy
                          'https://www.imdb.com/title/tt17351924/reviews/_ajax?&paginationKey=g4w6ddbmqy
                          'https://www.imdb.com/title/tt17351924/reviews/_ajax?&paginationKey=g4w6ddbmqy
                          'https://www.imdb.com/title/tt17351924/reviews/_ajax?&paginationKey=g4w6ddbmqy
                          'https://www.imdb.com/title/tt17351924/reviews/_ajax?&paginationKey=g4w6ddbmqy
                          'https://www.imdb.com/title/tt17351924/reviews/_ajax?&paginationKey=g4w6ddbmqy
                          'https://www.imdb.com/title/tt17351924/reviews/_ajax?&paginationKey=g4w6ddbmqy
                          'https://www.imdb.com/title/tt17351924/reviews/_ajax?&paginationKey=g4w6ddbmqy
                          'https://www.imdb.com/title/tt17351924/reviews/_ajax?&paginationKey=g4w6ddbmqy
)

all_reviews3 <- lapply(Saltburn_MOVIE_urls, scrape_reviews3)

final_all_reviews3 <- do.call(rbind, all_reviews3)

View(final_all_reviews3)


#######################################################################################################
#4.Puss in Boots: The Last Wish

session4<- bow(url = 'https://www.imdb.com/title/tt3915174/reviews?ref_=tt_urv',
               user_agent = "Student's Demo Educational")

session4


## <polite session> https://www.imdb.com/title/tt3915174/reviews?ref_=tt_urv
##     User-agent: Student's Demo Educational
##     robots.txt: 35 rules are defined for 3 bots
##    Crawl delay: 5 sec
##   The path is scrapable for this user-agent
```

```r
session_scrape4 <- scrape(session4)

scrape_reviews4 <- function(page_url) {

  reviews_page <- read_html(page_url)
  names4 <- reviews_page %>%
    html_nodes(".display-name-link") %>%
    html_text()

  dates4 <- reviews_page %>%
    html_nodes("span.review-date")%>%
    html_text()

  ratings4 <- reviews_page %>%
    html_nodes("span.rating-other-user-rating") %>%
    html_text()

  content_reviews4 <- reviews_page %>%
    html_nodes("a.title") %>%
    html_text()

  reviews4 <- reviews_page %>%
    html_nodes(".text.show-more__control") %>%
    html_text()

  Puss_in_Boots_MOVIE_df = data.frame( Movie_name = c('Puss in Boots: The Last Wish'),
                                       User_Name = names4[1:25],
                                       Date = dates4[1:25],
                                       User_Rating = ratings4[1:25],
                                       Content_Review = content_reviews4[1:25],
                                       Reviews = reviews4[1:25]
  )}

Puss_in_Boots_MOVIE_urls <- c('https://www.imdb.com/title/tt3915174/reviews?ref_=tt_urv',
                              'https://www.imdb.com/title/tt3915174/reviews/_ajax?&paginationKey=g4w6ddk
                              'https://www.imdb.com/title/tt3915174/reviews/_ajax?&paginationKey=g4w6ddk
                              'https://www.imdb.com/title/tt3915174/reviews/_ajax?&paginationKey=g4w6ddk
                              'https://www.imdb.com/title/tt3915174/reviews/_ajax?&paginationKey=g4w6ddk
                              'https://www.imdb.com/title/tt3915174/reviews/_ajax?&paginationKey=g4w6ddk
                              'https://www.imdb.com/title/tt3915174/reviews/_ajax?&paginationKey=g4w6ddk
                              'https://www.imdb.com/title/tt3915174/reviews/_ajax?&paginationKey=g4w6ddk
                              'https://www.imdb.com/title/tt3915174/reviews/_ajax?&paginationKey=g4w6ddk
                              'https://www.imdb.com/title/tt3915174/reviews/_ajax?&paginationKey=g4w6ddk
                              'https://www.imdb.com/title/tt3915174/reviews/_ajax?&paginationKey=g4w6ddk
                              'https://www.imdb.com/title/tt3915174/reviews/_ajax?&paginationKey=g4w6ddk
)

all_reviews4 <- lapply(Puss_in_Boots_MOVIE_urls, scrape_reviews4)

final_all_reviews4 <- do.call(rbind, all_reviews4)

View(final_all_reviews4)
```

```r
##############################################################################
#5.Howl's Moving Castle


session5<- bow(url = 'https://www.imdb.com/title/tt0347149/reviews/?ref_=tt_ov_rt',
               user_agent = "Student's Demo Educational")

session5

## <polite session> https://www.imdb.com/title/tt0347149/reviews/?ref_=tt_ov_rt
##      User-agent: Student's Demo Educational
##       robots.txt: 35 rules are defined for 3 bots
##     Crawl delay: 5 sec
##    The path is scrapable for this user-agent

session_scrape5 <- scrape(session5)

scrape_reviews5 <- function(page_url) {

  reviews_page <- read_html(page_url)
  names5 <- reviews_page %>%
    html_nodes(".display-name-link") %>%
    html_text()

  dates5 <- reviews_page %>%
    html_nodes("span.review-date")%>%
    html_text()

  ratings5 <- reviews_page %>%
    html_nodes("span.rating-other-user-rating") %>%
    html_text()

  content_reviews5 <- reviews_page %>%
    html_nodes("a.title") %>%
    html_text()

  reviews5 <- reviews_page %>%
    html_nodes(".text.show-more__control") %>%
    html_text()

  Howls_Moving_Castle_MOVIE_df = data.frame( Movie_name = c('Howls Moving Castle'),
                                             User_Name = names5[1:25],
                                             Date = dates5[1:25],
                                             User_Rating = ratings5[1:25],
                                             Content_Review = content_reviews5[1:25],
                                             Reviews = reviews5[1:25]
  )}

Howls_Moving_Castle_MOVIE_urls <- c( 'https://www.imdb.com/title/tt0347149/reviews/?ref_=tt_ov_rt',
                                     'https://www.imdb.com/title/tt0347149/reviews/_ajax?&paginationKey=
                                     'https://www.imdb.com/title/tt0347149/reviews/_ajax?&paginationKey=
                                     'https://www.imdb.com/title/tt0347149/reviews/_ajax?&paginationKey=
                                     'https://www.imdb.com/title/tt0347149/reviews/_ajax?&paginationKey=
```

```r
                                                 'https://www.imdb.com/title/tt0347149/reviews/_ajax?&paginationKey=
                                                 'https://www.imdb.com/title/tt0347149/reviews/_ajax?&paginationKey=
                                                 'https://www.imdb.com/title/tt0347149/reviews/_ajax?&paginationKey=
                                                 'https://www.imdb.com/title/tt0347149/reviews/_ajax?&paginationKey=
                                                 'https://www.imdb.com/title/tt0347149/reviews/_ajax?&paginationKey=
                                                 'https://www.imdb.com/title/tt0347149/reviews/_ajax?&paginationKey=
                                                 'https://www.imdb.com/title/tt0347149/reviews/_ajax?&paginationKey=
)

all_reviews5 <- lapply(Howls_Moving_Castle_MOVIE_urls, scrape_reviews5)

final_all_reviews5 <- do.call(rbind, all_reviews5)

View(final_all_reviews5)



###############################################################################################
#6.Black Widow


session6<- bow(url = 'https://www.imdb.com/title/tt3480822/reviews/?ref_=tt_ov_rt',
               user_agent = "Student's Demo Educational")

session6
```

```
## <polite session> https://www.imdb.com/title/tt3480822/reviews/?ref_=tt_ov_rt
##     User-agent: Student's Demo Educational
##      robots.txt: 35 rules are defined for 3 bots
##    Crawl delay: 5 sec
##   The path is scrapable for this user-agent
```

```r
session_scrape6 <- scrape(session6)

scrape_reviews6 <- function(page_url) {

  reviews_page <- read_html(page_url)
  names6 <- reviews_page %>%
    html_nodes(".display-name-link") %>%
    html_text()

  dates6 <- reviews_page %>%
    html_nodes("span.review-date")%>%
    html_text()

  ratings6 <- reviews_page %>%
    html_nodes("span.rating-other-user-rating") %>%
    html_text()

  content_reviews6 <- reviews_page %>%
    html_nodes("a.title") %>%
    html_text()

  reviews6 <- reviews_page %>%
```

```r
    html_nodes(".text.show-more__control") %>%
    html_text()

  Black_Widow_MOVIE_df = data.frame( Movie_name = c('Black Widow'),
                                     User_Name = names6[1:25],
                                     Date = dates6[1:25],
                                     User_Rating = ratings6[1:25],
                                     Content_Review = content_reviews6[1:25],
                                     Reviews = reviews6[1:25]
  )}

Black_Widow_MOVIE_urls <- c( 'https://www.imdb.com/title/tt3480822/reviews/?ref_=tt_ov_rt',
                             'https://www.imdb.com/title/tt3480822/reviews/_ajax?&paginationKey=g4w6ddb
                             'https://www.imdb.com/title/tt3480822/reviews/_ajax?&paginationKey=g4w6ddb
                             'https://www.imdb.com/title/tt3480822/reviews/_ajax?&paginationKey=g4w6ddb
                             'https://www.imdb.com/title/tt3480822/reviews/_ajax?&paginationKey=g4w6ddb
                             'https://www.imdb.com/title/tt3480822/reviews/_ajax?&paginationKey=g4w6ddb
                             'https://www.imdb.com/title/tt3480822/reviews/_ajax?&paginationKey=g4w6ddb
                             'https://www.imdb.com/title/tt3480822/reviews/_ajax?&paginationKey=g4w6ddb
                             'https://www.imdb.com/title/tt3480822/reviews/_ajax?&paginationKey=g4w6ddb
                             'https://www.imdb.com/title/tt3480822/reviews/_ajax?&paginationKey=g4w6ddb
                             'https://www.imdb.com/title/tt3480822/reviews/_ajax?&paginationKey=g4w6ddb
                             'https://www.imdb.com/title/tt3480822/reviews/_ajax?&paginationKey=g4w6ddb
)

all_reviews6 <- lapply(Black_Widow_MOVIE_urls, scrape_reviews6)

final_all_reviews6 <- do.call(rbind, all_reviews6)

View(final_all_reviews6)

###################################################################################
#7. First Blood

session7<- bow(url = 'https://www.imdb.com/title/tt0083944/reviews/?ref_=tt_ov_rt',
               user_agent = "Student's Demo Educational")

session7
```

```
## <polite session> https://www.imdb.com/title/tt0083944/reviews/?ref_=tt_ov_rt
##     User-agent: Student's Demo Educational
##     robots.txt: 35 rules are defined for 3 bots
##    Crawl delay: 5 sec
##   The path is scrapable for this user-agent
```

```r
session_scrape7 <- scrape(session7)

scrape_reviews7 <- function(page_url) {

  reviews_page <- read_html(page_url)
  names7 <- reviews_page %>%
    html_nodes(".display-name-link") %>%
    html_text()
```

```r
  dates7 <- reviews_page %>%
    html_nodes("span.review-date")%>%
    html_text()

  ratings7 <- reviews_page %>%
    html_nodes("span.rating-other-user-rating") %>%
    html_text()

  content_reviews7 <- reviews_page %>%
    html_nodes("a.title") %>%
    html_text()

  reviews7 <- reviews_page %>%
    html_nodes(".text.show-more__control") %>%
    html_text()

  First_Blood_MOVIE_df = data.frame( Movie_name = c('First Blood'),
                                     User_Name = names7[1:25],
                                     Date = dates7[1:25],
                                     User_Rating = ratings7[1:25],
                                     Content_Review = content_reviews7[1:25],
                                     Reviews = reviews7[1:25]
  )}

First_Blood_MOVIE_urls <- c( 'https://www.imdb.com/title/tt0083944/reviews/?ref_=tt_ov_rt',
                             'https://www.imdb.com/title/tt0083944/reviews/_ajax?&paginationKey=g4w6ddb
                             'https://www.imdb.com/title/tt0083944/reviews/_ajax?&paginationKey=g4w6ddb
                             'https://www.imdb.com/title/tt0083944/reviews/_ajax?&paginationKey=g4w6ddb
                             'https://www.imdb.com/title/tt0083944/reviews/_ajax?&paginationKey=g4w6ddb
                             'https://www.imdb.com/title/tt0083944/reviews/_ajax?&paginationKey=g4w6ddb
                             'https://www.imdb.com/title/tt0083944/reviews/_ajax?&paginationKey=g4w6ddb
                             'https://www.imdb.com/title/tt0083944/reviews/_ajax?&paginationKey=g4w6ddb
                             'https://www.imdb.com/title/tt0083944/reviews/_ajax?&paginationKey=g4w6ddb
                             'https://www.imdb.com/title/tt0083944/reviews/_ajax?&paginationKey=g4w6ddb
                             'https://www.imdb.com/title/tt0083944/reviews/_ajax?&paginationKey=g4w6ddb
                             'https://www.imdb.com/title/tt0083944/reviews/_ajax?&paginationKey=g4w6ddb
)

all_reviews7 <- lapply(First_Blood_MOVIE_urls, scrape_reviews7)

final_all_reviews7 <- do.call(rbind, all_reviews7)

View(final_all_reviews7)


###############################################################################
#8. The Meg

session8<- bow(url = 'https://www.imdb.com/title/tt4779682/reviews/?ref_=tt_ov_rt',
               user_agent = "Student's Demo Educational")

session8
```

```
## <polite session> https://www.imdb.com/title/tt4779682/reviews/?ref_=tt_ov_rt
##      User-agent: Student's Demo Educational
##      robots.txt: 35 rules are defined for 3 bots
##    Crawl delay: 5 sec
##   The path is scrapable for this user-agent
```

```r
session_scrape8 <- scrape(session8)

scrape_reviews8 <- function(page_url) {

  reviews_page <- read_html(page_url)
  names8 <- reviews_page %>%
    html_nodes(".display-name-link") %>%
    html_text()

  dates8 <- reviews_page %>%
    html_nodes("span.review-date")%>%
    html_text()

  ratings8 <- reviews_page %>%
    html_nodes("span.rating-other-user-rating") %>%
    html_text()

  content_reviews8 <- reviews_page %>%
    html_nodes("a.title") %>%
    html_text()

  reviews8 <- reviews_page %>%
    html_nodes(".text.show-more__control") %>%
    html_text()

  The_Meg_MOVIE_df = data.frame( Movie_name = c('The Meg'),
                                 User_Name = names8[1:25],
                                 Date = dates8[1:25],
                                 User_Rating = ratings8[1:25],
                                 Content_Review = content_reviews8[1:25],
                                 Reviews = reviews8[1:25]
  )}

The_Meg_MOVIE_urls <- c( 'https://www.imdb.com/title/tt4779682/reviews?ref_=tt_urv',
                         'https://www.imdb.com/title/tt4779682/reviews/_ajax?&paginationKey=g4w6ddbmqyz
                         'https://www.imdb.com/title/tt4779682/reviews/_ajax?&paginationKey=g4w6ddbmqyz
                         'https://www.imdb.com/title/tt4779682/reviews/_ajax?&paginationKey=g4w6ddbmqyz
                         'https://www.imdb.com/title/tt4779682/reviews/_ajax?&paginationKey=g4w6ddbmqyz
                         'https://www.imdb.com/title/tt4779682/reviews/_ajax?&paginationKey=g4w6ddbmqyz
                         'https://www.imdb.com/title/tt4779682/reviews/_ajax?&paginationKey=g4w6ddbmqyz
                         'https://www.imdb.com/title/tt4779682/reviews/_ajax?&paginationKey=g4w6ddbmqyz
                         'https://www.imdb.com/title/tt4779682/reviews/_ajax?&paginationKey=g4w6ddbmqyz
                         'https://www.imdb.com/title/tt4779682/reviews/_ajax?&paginationKey=g4w6ddbmqyz
                         'https://www.imdb.com/title/tt4779682/reviews/_ajax?&paginationKey=g4w6ddbmqyz
                         'https://www.imdb.com/title/tt4779682/reviews/_ajax?&paginationKey=g4w6ddbmqyz
                         'https://www.imdb.com/title/tt4779682/reviews/_ajax?&paginationKey=g4w6ddbmqyz
)

all_reviews8 <- lapply(The_Meg_MOVIE_urls, scrape_reviews8)
```

```r
final_all_reviews8 <- do.call(rbind, all_reviews8)

View(final_all_reviews8)


###############################################################################
#9. Avengers: Infinity War

session9<- bow(url = 'https://www.imdb.com/title/tt4154756/reviews/?ref_=tt_ov_rt',
               user_agent = "Student's Demo Educational")

session9


## <polite session> https://www.imdb.com/title/tt4154756/reviews/?ref_=tt_ov_rt
##      User-agent: Student's Demo Educational
##      robots.txt: 35 rules are defined for 3 bots
##    Crawl delay: 5 sec
##    The path is scrapable for this user-agent

session_scrape9 <- scrape(session9)

scrape_reviews9 <- function(page_url) {

  reviews_page <- read_html(page_url)
  names9 <- reviews_page %>%
    html_nodes(".display-name-link") %>%
    html_text()

  dates9 <- reviews_page %>%
    html_nodes("span.review-date")%>%
    html_text()

  ratings9 <- reviews_page %>%
    html_nodes("span.rating-other-user-rating") %>%
    html_text()

  content_reviews9 <- reviews_page %>%
    html_nodes("a.title") %>%
    html_text()

  reviews9 <- reviews_page %>%
    html_nodes(".text.show-more__control") %>%
    html_text()

  Infinity_War_MOVIE_df = data.frame( Movie_name = c('Avengers: Infinity War'),
                                      User_Name = names9[1:25],
                                      Date = dates9[1:25],
                                      User_Rating = ratings9[1:25],
                                      Content_Review = content_reviews9[1:25],
                                      Reviews = reviews9[1:25]
  )}

Infinity_War_MOVIE_urls <- c( 'https://www.imdb.com/title/tt4154756/reviews/?ref_=tt_ov_rt',
```

```
                                    'https://www.imdb.com/title/tt4154756/reviews/_ajax?&paginationKey=g4w6ddl
                                    'https://www.imdb.com/title/tt4154756/reviews/_ajax?&paginationKey=g4w6ddl
                                    'https://www.imdb.com/title/tt4154756/reviews/_ajax?&paginationKey=g4w6ddl
                                    'https://www.imdb.com/title/tt4154756/reviews/_ajax?&paginationKey=g4w6ddl
                                    'https://www.imdb.com/title/tt4154756/reviews/_ajax?&paginationKey=g4w6ddl
                                    'https://www.imdb.com/title/tt4154756/reviews/_ajax?&paginationKey=g4w6ddl
                                    'https://www.imdb.com/title/tt4154756/reviews/_ajax?&paginationKey=g4w6ddl
                                    'https://www.imdb.com/title/tt4154756/reviews/_ajax?&paginationKey=g4w6ddl
                                    'https://www.imdb.com/title/tt4154756/reviews/_ajax?&paginationKey=g4w6ddl
                                    'https://www.imdb.com/title/tt4154756/reviews/_ajax?&paginationKey=g4w6ddl
                                    'https://www.imdb.com/title/tt4154756/reviews/_ajax?&paginationKey=g4w6ddl
)

all_reviews9 <- lapply(Infinity_War_MOVIE_urls, scrape_reviews9 )

final_all_reviews9  <- do.call(rbind, all_reviews9 )

View(final_all_reviews9)




###############################################################################################
#10. WALL-E

session10<- bow(url = 'https://www.imdb.com/title/tt0910970/reviews/?ref_=ttcrv_ql_2',
                user_agent = "Student's Demo Educational")

session10


## <polite session> https://www.imdb.com/title/tt0910970/reviews/?ref_=ttcrv_ql_2
##      User-agent: Student's Demo Educational
##      robots.txt: 35 rules are defined for 3 bots
##    Crawl delay: 5 sec
##   The path is scrapable for this user-agent

session_scrape10 <- scrape(session10)

scrape_reviews10 <- function(page_url) {

  reviews_page <- read_html(page_url)
  names10 <- reviews_page %>%
    html_nodes(".display-name-link") %>%
    html_text()

  dates10 <- reviews_page %>%
    html_nodes("span.review-date")%>%
    html_text()

  ratings10 <- reviews_page %>%
    html_nodes("span.rating-other-user-rating") %>%
    html_text()
```

```r
  content_reviews10 <- reviews_page %>%
    html_nodes("a.title") %>%
    html_text()

  reviews10 <- reviews_page %>%
    html_nodes(".text.show-more__control") %>%
    html_text()

  WALL_E_MOVIE_df = data.frame( Movie_name = c('WALL-E'),
                                User_Name = names10[1:25],
                                Date = dates10[1:25],
                                User_Rating = ratings10[1:25],
                                Content_Review = content_reviews10[1:25],
                                Reviews = reviews10[1:25]
  )}

WALL_E_MOVIE_urls <- c( 'https://www.imdb.com/title/tt0910970/reviews/?ref_=ttcrv_ql_2',
                        'https://www.imdb.com/title/tt0910970/reviews/_ajax?&paginationKey=g4w6ddbmqyzd
                        'https://www.imdb.com/title/tt0910970/reviews/_ajax?&paginationKey=g4w6ddbmqyzd
                        'https://www.imdb.com/title/tt0910970/reviews/_ajax?&paginationKey=g4w6ddbmqyzd
                        'https://www.imdb.com/title/tt0910970/reviews/_ajax?&paginationKey=g4w6ddbmqyzd
                        'https://www.imdb.com/title/tt0910970/reviews/_ajax?&paginationKey=g4w6ddbmqyzd
                        'https://www.imdb.com/title/tt0910970/reviews/_ajax?&paginationKey=g4w6ddbmqyzd
                        'https://www.imdb.com/title/tt0910970/reviews/_ajax?&paginationKey=g4w6ddbmqyzd
                        'https://www.imdb.com/title/tt0910970/reviews/_ajax?&paginationKey=g4w6ddbmqyzd
                        'https://www.imdb.com/title/tt0910970/reviews/_ajax?&paginationKey=g4w6ddbmqyzd
                        'https://www.imdb.com/title/tt0910970/reviews/_ajax?&paginationKey=g4w6ddbmqyzd
                        'https://www.imdb.com/title/tt0910970/reviews/_ajax?&paginationKey=g4w6ddbmqyzd
)

all_reviews10 <- lapply(WALL_E_MOVIE_urls, scrape_reviews10 )

final_all_reviews10  <- do.call(rbind, all_reviews10)

View(final_all_reviews10)



all_reviews <- rbind(final_all_reviews,
                     final_all_reviews2,
                     final_all_reviews3,
                     final_all_reviews4,
                     final_all_reviews5,
                     final_all_reviews6,
                     final_all_reviews7,
                     final_all_reviews8,
                     final_all_reviews9,
                     final_all_reviews10)

View(all_reviews)
```

```r
write_xlsx(all_reviews, "scrape3000reviews.xlsx")
```