

# ACTIVITY2\_SADSAD

BSIT-2B

2024-02-06

```
library(rvest)
```

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

```
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
```

```
scrape_reviews <- function(url) {  
  page <- read_html(url)  
  
  names <- page %>% html_nodes(".display-name-link") %>% html_text()  
  dates <- page %>% html_nodes("span.review-date") %>% html_text()  
  ratings <- page %>% html_nodes("span.rating-other-user-rating") %>% html_text()  
  content_reviews <- page %>% html_nodes("a.title") %>% html_text()  
  reviews <- page %>% html_nodes(".text.show-more__control") %>% html_text()  
  
  scrape_reviews_df = data.frame(  
    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=g4w6ddbmqyzdo6ic4oxwjnbzr3smmarz3u
'https://www.imdb.com/title/tt5776858/reviews/_ajax?&paginationKey=g4w6ddbmqyzdo6ic4oxwjnbqr3t4ycjt3q
'https://www.imdb.com/title/tt5776858/reviews/_ajax?&paginationKey=g4w6ddbmqyzdo6ic4oxwjnjzqpsmscb23u
'https://www.imdb.com/title/tt5776858/reviews/_ajax?&paginationKey=g4w6ddbmqyzdo6ic4oxwjnjvrtrmbr43q
'https://www.imdb.com/title/tt5776858/reviews/_ajax?&paginationKey=g4w6ddbmqyzdo6ic4oxwjnjtrdumsab53q
'https://www.imdb.com/title/tt5776858/reviews/_ajax?&paginationKey=g4w6ddbmqyzdo6ic4oxwjmrqqlrmscj52m
'https://www.imdb.com/title/tt5776858/reviews/_ajax?&paginationKey=g4w6ddbmqyzdo6ic4oxwjmrwqlrmsaj43a
'https://www.imdb.com/title/tt5776858/reviews/_ajax?&paginationKey=g4w6ddbmqyzdo6ic4oxwjmrqupsmoaby2i
'https://www.imdb.com/title/tt5776858/reviews/_ajax?&paginationKey=g4w6ddbmqyzdo6ic4oxwjmrvrdu44az73u
'https://www.imdb.com/title/tt5776858/reviews/_ajax?&paginationKey=g4w6ddbmqyzdo6ic4oxwjmrtrdumsab53q
'https://www.imdb.com/title/tt5776858/reviews/_ajax?&paginationKey=g4w6ddbmqyzdo6ic4oxwjmrqlqm6cby3i
)

all_reviews <- lapply(phantom_thread_urls, scrape_reviews)

final_all_reviews <- do.call(rbind, all_reviews)

View(final_all_reviews)

```