Lab_Exercise#1

Octavio, Jessa

2024-02-08

```
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(rvest)
library(polite)
library(httr)
library(selectr)
library(xml2)
```

10 reviews per movie

```
url1 <- "https://www.imdb.com/title/tt0068646/reviews?ref_=tt_urv"


 session <- bow(url1,
                user_agent = "For Educational Purpose")

  scrapeNodes <- function(selector){
    scrape(session) %>%
      html_nodes(selector) %>%
      html_text(trim = TRUE)
  }

  movie_category <- rep("Movie", 10)

  movie_name <- scrapeNodes("a.subnav_heading")
  movie_name <- rep(movie_name, 10)
  movie_name <- movie_name[1:10]

  movie_reviewer <- scrapeNodes("span.display-name-link")
```

```
  movie_reviewer <-movie_reviewer[1:10]

  movie_review <- scrapeNodes("div.text.show-more__control")
  movie_review <- movie_review[1:10]

  movie_date <- scrapeNodes("span.review-date")
  movie_date <- movie_date[1:10]

  movie_rating <- scrapeNodes("span.rating-other-user-rating")
  movie_rating <- movie_rating[1:10]

  moviereviews1= data.frame()

  moviereviews1 <- rbind(moviereviews1, data.frame(
                           category = movie_category,
                           name = movie_name,
                           reviewer = movie_reviewer,
                           reviews = movie_review,
                           "date of review" = movie_date,
                           rating = movie_rating))

Sys.sleep(5)
```

```
#view(movierewies1)
write.csv(moviereviews1, file = "moviereviews1.csv")
```

```
url2 <- "https://www.imdb.com/title/tt0068646/reviews?ref_=tt_urv"

 session <- bow(url2,
                user_agent = "For Educational Purpose")

  scrapeNodes <- function(selector){
    scrape(session) %>%
      html_nodes(selector) %>%
      html_text(trim = TRUE)
  }

  movie_category <- rep("Movie", 10)

  movie_name2 <- scrapeNodes("a.subnav_heading")
  movie_name2 <- rep(movie_name2, 10)
  movie_name2 <- movie_name2[1:10]

  movie_reviewer2 <- scrapeNodes("span.display-name-link")
  movie_reviewer2 <-movie_reviewer2[1:10]

  movie_review2 <- scrapeNodes("div.text.show-more__control")
  movie_review2 <- movie_review2[1:10]

  movie_date2 <- scrapeNodes("span.review-date")
  movie_date2 <- movie_date2[1:10]

  movie_rating2 <- scrapeNodes("span.rating-other-user-rating")
```

```
  movie_rating2 <- movie_rating2[1:10]

  moviereviews2= data.frame()

  moviereviews2 <- rbind(moviereviews2, data.frame(
                           category = movie_category,
                           name = movie_name2,
                           reviewer = movie_reviewer2,
                           reviews = movie_review2,
                           "date of review" = movie_date2,
                           rating = movie_rating2))

Sys.sleep(5)
```

```
#view(movierewies2)
write.csv(moviereviews2, file = "moviereviews2.csv")
```

```
url3 <- "https://www.imdb.com/title/tt0050083/reviews?ref_=tt_urv"

 session <- bow(url3,
                user_agent = "For Educational Purpose")

    scrapeNodes <- function(selector){
      scrape(session) %>%
        html_nodes(selector) %>%
        html_text(trim = TRUE)
    }

  movie_category <- rep("Movie", 10)

  movie_name3 <- scrapeNodes("a.subnav_heading")
  movie_name3 <- rep(movie_name3, 10)
  movie_name3 <- movie_name3[1:10]

  movie_reviewer3 <- scrapeNodes("span.display-name-link")
  movie_reviewer3 <-movie_reviewer3[1:10]

  movie_review3 <- scrapeNodes("div.text.show-more__control")
  movie_review3 <- movie_review3[1:10]

  movie_date3 <- scrapeNodes("span.review-date")
  movie_date3 <- movie_date3[1:10]

  movie_rating3 <- scrapeNodes("span.rating-other-user-rating")
  movie_rating3 <- movie_rating3[1:10]

  moviereviews3= data.frame()

  moviereviews3 <- rbind(moviereviews3, data.frame(
                           category = movie_category,
                           name = movie_name3,
                           reviewer = movie_reviewer3,
                           reviews = movie_review3,
```

```
                          "date of review" = movie_date3,
                          rating = movie_rating3))

Sys.sleep(5)
```

```
#view(movierewies3)
write.csv(moviereviews3, file = "moviereviews3.csv")
```

```
url4 <- "https://www.imdb.com/title/tt0245429/reviews?ref_=tt_urv"

 session <- bow(url4,
                user_agent = "For Educational Purpose")

  scrapeNodes <- function(selector){
    scrape(session) %>%
      html_nodes(selector) %>%
      html_text(trim = TRUE)
  }

  movie_category <- rep("Movie", 10)

  movie_name4 <- scrapeNodes("a.subnav_heading")
  movie_name4 <- rep(movie_name4, 10)

  movie_name4 <- movie_name4[1:10]

  movie_reviewer4 <- scrapeNodes("span.display-name-link")
  movie_reviewer4 <-movie_reviewer4[1:10]

  movie_review4 <- scrapeNodes("div.text.show-more__control")
  movie_review4 <- movie_review4[1:10]

  movie_date4 <- scrapeNodes("span.review-date")
  movie_date4 <- movie_date4[1:10]

  movie_rating4 <- scrapeNodes("span.rating-other-user-rating")
  movie_rating4 <- movie_rating4[1:10]

  moviereviews4= data.frame()

  moviereviews4 <- rbind(moviereviews4, data.frame(
                          category = movie_category,
                          name = movie_name4,
                          reviewer = movie_reviewer4,
                          reviews = movie_review4,
                          "date of review" = movie_date4,
                          rating = movie_rating4))

Sys.sleep(5)
```

```
#view(movierewies4)
write.csv(moviereviews4, file = "moviereviews4.csv")
```

```r
url5 <- "https://www.imdb.com/title/tt0816692/reviews?ref_=tt_urv"

 session <- bow(url5,
                user_agent = "For Educational Purpose")

  scrapeNodes <- function(selector){
    scrape(session) %>%
      html_nodes(selector) %>%
      html_text(trim = TRUE)
  }

  movie_category <- rep("Movie", 10)

  movie_name5 <- scrapeNodes("a.subnav_heading")
  movie_name5 <- rep(movie_name5, 10)
  movie_name5 <- movie_name5[1:10]

  movie_reviewer5 <- scrapeNodes("span.display-name-link")
  movie_reviewer5 <-movie_reviewer5[1:10]

  movie_review5 <- scrapeNodes("div.text.show-more__control")
  movie_review5 <- movie_review5[1:10]

  movie_date5 <- scrapeNodes("span.review-date")
  movie_date5 <- movie_date5[1:10]

  movie_rating5 <- scrapeNodes("span.rating-other-user-rating")
  movie_rating5 <- movie_rating5[1:10]

  moviereviews5= data.frame()

  moviereviews5 <- rbind(moviereviews5, data.frame(
                           category = movie_category,
                           name = movie_name5,
                           reviewer = movie_reviewer5,
                           reviews = movie_review5,
                           "date of review" = movie_date5,
                           rating = movie_rating5))

Sys.sleep(5)
```

```r
#view(movierewies5)
write.csv(moviereviews5, file = "moviereviews5.csv")
```

```r
url6 <- "https://www.imdb.com/title/tt0108052/reviews?ref_=tt_urv"

 session <- bow(url6,
                user_agent = "For Educational Purpose")

  scrapeNodes <- function(selector){
    scrape(session) %>%
      html_nodes(selector) %>%
      html_text(trim = TRUE)
```

```
  }

  movie_category <- rep("Movie", 10)

  movie_name <- scrapeNodes("a.subnav_heading")
  movie_name <- rep(movie_name, 10)
  movie_name <- movie_name[1:10]

  movie_reviewer <- scrapeNodes("span.display-name-link")
  movie_reviewer <-movie_reviewer[1:10]

  movie_review <- scrapeNodes("div.text.show-more__control")
  movie_review <- movie_review[1:10]

  movie_date <- scrapeNodes("span.review-date")
  movie_date <- movie_date[1:10]

  movie_rating <- scrapeNodes("span.rating-other-user-rating")
  movie_rating <- movie_rating[1:10]

  moviereviews6= data.frame()

  moviereviews6 <- rbind(moviereviews6, data.frame(
                          category = movie_category,
                          name = movie_name,
                          reviewer = movie_reviewer,
                          reviews = movie_review,
                          "date of review" = movie_date,
                          rating = movie_rating))

Sys.sleep(5)
```

```
#view(movierewies6)
write.csv(moviereviews6, file = "moviereviews6.csv")
```

```
url7 <- "https://www.imdb.com/title/tt0080684/reviews?ref_=tt_urv"


 session <- bow(url7,
                user_agent = "For Educational Purpose")

  scrapeNodes <- function(selector){
    scrape(session) %>%
      html_nodes(selector) %>%
      html_text(trim = TRUE)
  }

  movie_category <- rep("Movie", 10)

  movie_name <- scrapeNodes("a.subnav_heading")

  movie_name <- rep(movie_name, 10)
  movie_name <- movie_name[1:10]
```

```
  movie_reviewer <- scrapeNodes("span.display-name-link")
  movie_reviewer <-movie_reviewer[1:10]



  movie_review <- scrapeNodes("div.text.show-more__control")
  movie_review <- movie_review[1:10]


  movie_date <- scrapeNodes("span.review-date")

  movie_date <- movie_date[1:10]

  movie_rating <- scrapeNodes("span.rating-other-user-rating")

  movie_rating <- movie_rating[1:10]

  moviereviews7= data.frame()


  moviereviews7 <- rbind(moviereviews7, data.frame(
                          category = movie_category,
                          name = movie_name,
                          reviewer = movie_reviewer,
                          reviews = movie_review,
                          "date of review" = movie_date,
                          rating = movie_rating))

Sys.sleep(5)
```

```
#view(movierewies7)
write.csv(moviereviews7, file = "moviereviews7.csv")
```

```
url8 <- "https://www.imdb.com/title/tt0133093/reviews?ref_=tt_urv"


 session <- bow(url8,
                user_agent = "For Educational Purpose")

  scrapeNodes <- function(selector){
    scrape(session) %>%
      html_nodes(selector) %>%
      html_text(trim = TRUE)
  }

  movie_category <- rep("Movie", 10)

  movie_name <- scrapeNodes("a.subnav_heading")
  movie_name <- rep(movie_name, 10)
  movie_name <- movie_name[1:10]

  movie_reviewer <- scrapeNodes("span.display-name-link")
  movie_reviewer <-movie_reviewer[1:10]
```

```
  movie_review <- scrapeNodes("div.text.show-more__control")
  movie_review <- movie_review[1:10]

  movie_date <- scrapeNodes("span.review-date")
  movie_date <- movie_date[1:10]

  movie_rating <- scrapeNodes("span.rating-other-user-rating")
  movie_rating <- movie_rating[1:10]

  moviereviews8= data.frame()

  moviereviews8 <- rbind(moviereviews8, data.frame(
                          category = movie_category,
                          name = movie_name,
                          reviewer = movie_reviewer,
                          reviews = movie_review,
                          "date of review" = movie_date,
                          rating = movie_rating))

Sys.sleep(5)
```

```
#view(movierewies8)
write.csv(moviereviews8, file = "moviereviews8.csv")
```

```
url9 <- "https://www.imdb.com/title/tt0102926/reviews?ref_=tt_urv"


 session <- bow(url9,
                user_agent = "For Educational Purpose")

  scrapeNodes <- function(selector){
    scrape(session) %>%
      html_nodes(selector) %>%
      html_text(trim = TRUE)
  }

  movie_category <- rep("Movie", 10)

  movie_name <- scrapeNodes("a.subnav_heading")
  movie_name <- rep(movie_name, 10)
  movie_name <- movie_name[1:10]

  movie_reviewer <- scrapeNodes("span.display-name-link")
  movie_reviewer <-movie_reviewer[1:10]

  movie_review <- scrapeNodes("div.text.show-more__control")
  movie_review <- movie_review[1:10]

  movie_date <- scrapeNodes("span.review-date")
  movie_date <- movie_date[1:10]

  movie_rating <- scrapeNodes("span.rating-other-user-rating")
  movie_rating <- movie_rating[1:10]
```

```
  moviereviews9 = data.frame()

  moviereviews9 <- rbind(moviereviews9, data.frame(
                           category = movie_category,
                           name = movie_name,
                           reviewer = movie_reviewer,
                           reviews = movie_review,
                           "date of review" = movie_date,
                           rating = movie_rating))

Sys.sleep(5)
```

```
#view(movierewies9)
write.csv(moviereviews9, file = "moviereviews9.csv")
```

```
url10 <- "https://www.imdb.com/title/tt0088763/reviews?ref_=tt_urv"

 session <- bow(url10,
                user_agent = "For Educational Purpose")

  scrapeNodes <- function(selector){
    scrape(session) %>%
      html_nodes(selector) %>%
      html_text(trim = TRUE)
  }

  movie_category <- rep("Movie", 10)

  movie_name <- scrapeNodes("a.subnav_heading")
  movie_name <- rep(movie_name, 10)
  movie_name <- movie_name[1:10]

  movie_reviewer <- scrapeNodes("span.display-name-link")
  movie_reviewer <-movie_reviewer[1:10]

  movie_review <- scrapeNodes("div.text.show-more__control")
  movie_review <- movie_review[1:10]

  movie_date <- scrapeNodes("span.review-date")
  movie_date <- movie_date[1:10]

  movie_rating <- scrapeNodes("span.rating-other-user-rating")
  movie_rating <- movie_rating[1:10]

  moviereviews10= data.frame()

  moviereviews10 <- rbind(moviereviews10, data.frame(
                           category = movie_category,
                           name = movie_name,
                           reviewer = movie_reviewer,
                           reviews = movie_review,
                           "date of review" = movie_date,
                           rating = movie_rating))
```

```r
Sys.sleep(5)
```

```r
#view(movierewies10)
write.csv(moviereviews10, file = "moviereviews10.csv")
```