CS 102 IT2C Fermano

Fermano, Aaron John

2024-02-08

```r
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
```

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

```r
url <- "https://www.amazon.co.uk/s?k=watches&crid=10YCJDGWCRU&sprefix=watches+f%2Caps%2C280&ref=nb_sb_n

producdf <- data.frame()

session <- bow(url,
               user_agent = "For Educational Purposes")

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

}

product_name <- scrapeNodes("span.a-size-base-plus.a-color-base.a-text-normal")
product_name <- product_name[1:41]

product_price <- scrapeNodes("span.a-price")
product_price <- product_price[1:41]

product_ratings <- scrapeNodes("i.a-icon.a-icon-star-small.a-star-small-4-5.aok-align-bottom")
product_ratings <- product_ratings[1:41]

total_review <- scrapeNodes("span.a-size-base.s-underline-text")
total_review <- total_review[1:41]
```

```r
product <- rbind(producdf, data.frame(Name = product_name,
                       Price = product_price,
                       Ratings = product_ratings,
                       TotalReview = total_review))

url2<-"https://www.amazon.co.uk/s?k=watches&page=2&crid=1YW0YCJDGWCRU&qid=1707355577&sprefix=watches+f%

producdf2 <- data.frame()

session2 <- bow(url2,
             user_agent = "For Educational Purposes")

 scrapeNodes2 <- function(selector){
    scrape(session2) %>%
      html_nodes(selector) %>%
      html_text(trim = TRUE)

}

product_name2 <- scrapeNodes2("span.a-size-base-plus.a-color-base.a-text-normal")
product_name2 <- product_name2[1:33]

product_price2 <- scrapeNodes2("span.a-price")
product_price2 <- product_price2[1:33]

product_ratings2 <- scrapeNodes2("i.a-icon.a-icon-star-small.a-star-small-4-5.aok-align-bottom")
product_ratings2 <- product_ratings2[1:33]

total_review2 <- scrapeNodes2("span.a-size-base.s-underline-text")
total_review2 <- total_review2[1:33]

product2 <- rbind(producdf2, data.frame(Name = product_name2,
                       Price = product_price2,
                       Ratings = product_ratings2,
                       TotalReview = total_review2))

url3 <- "https://www.amazon.co.uk/s?k=watches&page=3&crid=1YW0YCJDGWCRU&qid=1707458979&sprefix=watches+

productdf3 <- data.frame()

session3 <- bow(url3,
             user_agent = "For Educational Purposes")

 scrapeNodes3 <- function(selector){
    scrape(session3) %>%
      html_nodes(selector) %>%
      html_text(trim = TRUE)

}

product_name3 <- scrapeNodes3("span.a-size-base-plus.a-color-base.a-text-normal")
product_name3 <- product_name3[1:35]
```

```r
product_price3 <- scrapeNodes3("span.a-price")
product_price3 <- product_price3[1:35]

product_ratings3 <- scrapeNodes3("i.a-icon.a-icon-star-small.a-star-small-4-5.aok-align-bottom")
product_ratings3 <- product_ratings3[1:35]

total_review3 <- scrapeNodes3("span.a-size-base.s-underline-text")
total_review3 <- total_review3[1:35]

product3 <- rbind(productdf3, data.frame(Name = product_name3,
                          Price = product_price3,
                          Ratings = product_ratings3,
                          TotalReview = total_review3))
url4 <- "https://www.amazon.co.uk/s?k=headphones&crid=30YCJDGWCRU&sprefix=headp%2Caps%2C430&ref=nb_sb_s

productdf4 <- data.frame()

session4 <- bow(url4,
               user_agent = "For Educational Purposes")

 scrapeNodes4 <- function(selector){
     scrape(session4) %>%
       html_nodes(selector) %>%
       html_text(trim = TRUE)

}


product_name4 <- scrapeNodes4("span.a-size-base-plus.a-color-base.a-text-normal")
product_name4 <- product_name4[1:42]

product_price4 <- scrapeNodes4("span.a-price")
product_price4 <- product_price4[1:42]

product_ratings4 <- scrapeNodes4("i.a-icon.a-icon-star-small.a-star-small-4-5.aok-align-bottom")
product_ratings4 <- product_ratings4[1:42]

total_review4 <- scrapeNodes4("span.a-size-base.s-underline-text")
total_review4 <- total_review4[1:42]

product4 <- rbind(productdf4, data.frame(Name = product_name4,
                          Price = product_price4,
                          Ratings = product_ratings4,
                          TotalReview = total_review4))
url5 <- "https://www.amazon.co.uk/s?k=headphones&page=2&crid=30YCJDGWCRU&qid=1707469052&sprefix=headp%2

productdf5 <- data.frame()

session5 <- bow(url5,
               user_agent = "For Educational Purposes")

 scrapeNodes5 <- function(selector){
```

```r
    scrape(session5) %>%
      html_nodes(selector) %>%
      html_text(trim = TRUE)


}
product_name5 <- scrapeNodes5("span.a-size-base-plus.a-color-base.a-text-normal")
product_name5 <- product_name5[1:42]

product_price5 <- scrapeNodes5("span.a-price")
product_price5 <- product_price5[1:42]

product_ratings5 <- scrapeNodes5("i.a-icon.a-icon-star-small.a-star-small-4-5.aok-align-bottom")
product_ratings5 <- product_ratings5[1:42]

total_review5 <- scrapeNodes5("span.a-size-base.s-underline-text")
total_review5 <- total_review5[1:42]

product5 <- rbind(productdf5, data.frame(Name = product_name5,
                      Price = product_price5,
                      Ratings = product_ratings5,
                      TotalReview = total_review5))
url6 <- "https://www.amazon.co.uk/s?k=headphones&page=3&crid=30YCJDGWCRU&qid=1707471189&sprefix=headp%2


productdf6 <- data.frame()

session6 <- bow(url6,
              user_agent = "For Educational Purposes")


 scrapeNodes6 <- function(selector){
    scrape(session6) %>%
      html_nodes(selector) %>%
      html_text(trim = TRUE)


}
product_name6 <- scrapeNodes6("span.a-size-base-plus.a-color-base.a-text-normal")
product_name6 <- product_name6[1:9]

product_price6 <- scrapeNodes6("span.a-price")
product_price6 <- product_price6[1:9]

product_ratings6 <- scrapeNodes6("i.a-icon.a-icon-star-small.a-star-small-4.aok-align-bottom")
product_ratings6 <- product_ratings6[1:9]

total_review6 <- scrapeNodes6("span.a-size-base.s-underline-text")
total_review6 <- total_review6[1:9]

product6 <- rbind(productdf6, data.frame(Name = product_name6,
                      Price = product_price6,
                      Ratings = product_ratings6,
                      TotalReview = total_review6))
```

```r
url7 <- "https://www.amazon.co.uk/s?k=headphones&page=4&crid=1LIZSZKU5D11R&qid=1707955263&sprefix=headp"

productdf7 <- data.frame()

session7 <- bow(url7,
                user_agent = "For Educational Purposes")


 scrapeNodes7 <- function(selector){
    scrape(session7) %>%
      html_nodes(selector) %>%
      html_text(trim = TRUE)

}

product_name7 <- scrapeNodes7("span.a-size-base-plus.a-color-base.a-text-normal")
product_name7 <- product_name7[1:12]

product_price7 <- scrapeNodes7("span.a-price")
product_price7 <- product_price7[1:12]

product_ratings7 <- scrapeNodes7("i.a-icon.a-icon-star-small.a-star-small-4.aok-align-bottom")
product_ratings7 <- product_ratings7[1:12]

total_review7 <- scrapeNodes7("span.a-size-base.s-underline-text")
total_review7 <- total_review7[1:12]

product7 <- rbind(productdf7, data.frame(Name = product_name7,
                        Price = product_price7,
                        Ratings = product_ratings7,
                        TotalReview = total_review7))

firstprdf <- rbind(product,product2,product3)
secondprdf <- rbind(product4,product5,product6,product7)

url01 <- "https://www.imdb.com/title/tt0111161/reviews?ref_=tt_urv"


 session <- bow(url01,
                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]
```

```r
  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)
```

```r
url02 <- "https://www.imdb.com/title/tt0068646/reviews?ref_=tt_urv"

 session <- bow(url02,
                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()
```

```r
  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)
```

```r
url03 <- "https://www.imdb.com/title/tt0468569/reviews?ref_=tt_urv"

 session <- bow(url03,
                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)
```

```r
url04 <- "https://www.imdb.com/title/tt0071562/reviews?ref_=tt_urv"
```

```r
session <- bow(url04,
               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)
```

```r
url05 <- "https://www.imdb.com/title/tt0050083/reviews?ref_=tt_urv"

session <- bow(url05,
               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)
```

```r
  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
url06 <- "https://www.imdb.com/title/tt0108052/reviews?ref_=tt_urv"

 session <- bow(url06,
                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]
```

```r
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)
```

```r
url07 <- "https://www.imdb.com/title/tt0167260/reviews?ref_=tt_urv"


session <- bow(url07,
               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)
```

```r
url08 <- "https://www.imdb.com/title/tt0110912/reviews?ref_=tt_urv"


 session <- bow(url08,
                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)
```

```r
url09 <- "https://www.imdb.com/title/tt0120737/reviews?ref_=tt_urv"


 session <- bow(url09,
                user_agent = "For Educational Purpose")

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

```r
  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)
```

```r
url10 <- "https://www.imdb.com/title/tt0060196/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")
```

```r
  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))

Sys.sleep(5)
```