Lab_exercise#2_octavio

Octavio, Jessa

2024-03-07

```
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)
```

```
  link1 <- "https://www.amazon.com/product-reviews/B0CPXNLBSX/ref=acr_dp_hist_5?ie=UTF8&filterByStar=fi

session <- bow(link1,
               user_agent = "For Educational Purpose")

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

  product_category <- rep("Phone Case", 10)

  product_name <- scrapeNodes("h1.a-size-large.a-text-ellipsis")
  product_name <- rep(product_name, 10)
  product_name <- product_name[1:10]

  verified_reviews <-scrapeNodes("span.a-size-mini.a-color-state.a-text-bold")
  verified_reviews <-verified_reviews[1:10]

  product_reviewer <- scrapeNodes("span.a-profile-name")
```

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

  product_review <- scrapeNodes("span.a-size-base.review-text")
  product_review <- product_review[1:10]

  product_date <- scrapeNodes("span.a-size-base.a-color-secondary.review-date")
  product_date <- product_date[1:10]

  product_rating <- scrapeNodes("span.a-icon-alt")
  product_rating <- product_rating[1:10]

  productreviews1= data.frame()
  productreviews1 <- rbind(productreviews1, data.frame(
                       category = product_category,
                       name = product_name,
                       reviewer = product_reviewer,
                       reviews = product_review,
                       verified = verified_reviews,
                       "date of review" = product_date,
                       ratings = product_rating))
 Sys.sleep(5)
```

```
link2 <- "https://www.amazon.com/product-reviews/B0B94MPFJV/ref=acr_dp_hist_5?ie=UTF8&filterByStar=five_
```

```
  session <- bow(link2,
                 user_agent = "For Educational Purpose")

  scrapeNodes <- function(selector){
    scrape(session) %>%
      html_nodes(selector) %>%
      html_text(trim = TRUE)
  }
  product_category <- rep("Phone Case", 10)

  product_name <- scrapeNodes("h1.a-size-large.a-text-ellipsis")
  product_name <- rep(product_name, 10)
  product_name <- product_name[1:10]

  verified_reviews <-scrapeNodes("span.a-size-mini.a-color-state.a-text-bold")
  verified_reviews <-verified_reviews[1:10]

  product_reviewer <- scrapeNodes("span.a-profile-name")
  product_reviewer <- product_reviewer[1:10]

  product_review <- scrapeNodes("span.a-size-base.review-text")
  product_review <- product_review[1:10]

  product_date <- scrapeNodes("span.a-size-base.a-color-secondary.review-date")
  product_date <- product_date[1:10]

  product_rating <- scrapeNodes("span.a-icon-alt")
  product_rating <- product_rating[1:10]
```

```
  productreviews2= data.frame()
  productreviews2 <- rbind(productreviews2, data.frame(
                       category = product_category,
                       name = product_name,
                       reviewer = product_reviewer,
                       reviews = product_review,
                       verified = verified_reviews,
                       "date of review" = product_date,
                       ratings = product_rating))
 Sys.sleep(5)
```

```
link3 <- "https://www.amazon.com/product-reviews/B0CC1F4V7Q/ref=acr_dp_hist_5?ie=UTF8&filterByStar=five
```

```
  session <- bow(link3,
                 user_agent = "For Educational Purpose")

  scrapeNodes <- function(selector){
    scrape(session) %>%
      html_nodes(selector) %>%
      html_text(trim = TRUE)
  }
  product_category <- rep("Phone Case", 10)

  product_name <- scrapeNodes("h1.a-size-large.a-text-ellipsis")
  product_name <- rep(product_name, 10)
  product_name <- product_name[1:10]

  verified_reviews <-scrapeNodes("span.a-size-mini.a-color-state.a-text-bold")
  verified_reviews <-verified_reviews[1:10]

  product_reviewer <- scrapeNodes("span.a-profile-name")
  product_reviewer <- product_reviewer[1:10]

  product_review <- scrapeNodes("span.a-size-base.review-text")
  product_review <- product_review[1:10]

  product_date <- scrapeNodes("span.a-size-base.a-color-secondary.review-date")
  product_date <- product_date[1:10]

  product_rating <- scrapeNodes("span.a-icon-alt")
  product_rating <- product_rating[1:10]

  productreviews3= data.frame()
  productreviews3 <- rbind(productreviews3, data.frame(
                       category = product_category,
                       name = product_name,
                       reviewer = product_reviewer,
                       reviews = product_review,
                       verified = verified_reviews,
                       "date of review" = product_date,
                       ratings = product_rating))
 Sys.sleep(5)
```

```
link4 <- "https://www.amazon.com/product-reviews/B0CBJZL29J/ref=acr_dp_hist_5?ie=UTF8&filterByStar=five_

  session <- bow(link4,
                 user_agent = "For Educational Purpose")

  scrapeNodes <- function(selector){
    scrape(session) %>%
      html_nodes(selector) %>%
      html_text(trim = TRUE)
  }
  product_category <- rep("Phone Case", 10)

  product_name <- scrapeNodes("h1.a-size-large.a-text-ellipsis")
  product_name <- rep(product_name, 10)
  product_name <- product_name[1:10]

  verified_reviews <-scrapeNodes("span.a-size-mini.a-color-state.a-text-bold")
  verified_reviews <-verified_reviews[1:10]

  product_reviewer <- scrapeNodes("span.a-profile-name")
  product_reviewer <- product_reviewer[1:10]

  product_review <- scrapeNodes("span.a-size-base.review-text.review-text-content")
  product_review <- product_review[1:10]

  product_date <- scrapeNodes("span.a-size-base.a-color-secondary.review-date")
  product_date <- product_date[1:10]

  product_rating <- scrapeNodes("span.a-icon-alt")
  product_rating <- product_rating[1:10]

  productreviews4= data.frame()
  productreviews4 <- rbind(productreviews4, data.frame(
                       category = product_category,
                       name = product_name,
                       reviewer = product_reviewer,
                       reviews = product_review,
                       verified = verified_reviews,
                       "date of review" = product_date,
                       ratings = product_rating))
 Sys.sleep(5)
```

```
link5 <- "https://www.amazon.com/product-reviews/B0CDPGGL3C/ref=acr_dpx_hist_5?ie=UTF8&filterByStar=five

  session <- bow(link5,
                 user_agent = "For Educational Purpose")

  scrapeNodes <- function(selector){
    scrape(session) %>%
      html_nodes(selector) %>%
      html_text(trim = TRUE)
  }
  product_category <- rep("Phone Case", 10)
```

```
  product_name <- scrapeNodes("h1.a-size-large.a-text-ellipsis")
  product_name <- rep(product_name, 10)
  product_name <- product_name[1:10]

  verified_reviews <-scrapeNodes("span.a-size-mini.a-color-state.a-text-bold")
  verified_reviews <-verified_reviews[1:10]

  product_reviewer <- scrapeNodes("span.a-profile-name")
  product_reviewer <- product_reviewer[1:10]

  product_review <- scrapeNodes("span.a-size-base.review-text")
  product_review <- product_review[1:10]

  product_date <- scrapeNodes("span.a-size-base.a-color-secondary.review-date")
  product_date <- product_date[1:10]

  product_rating <- scrapeNodes("span.a-icon-alt")
  product_rating <- product_rating[1:10]

  productreviews5= data.frame()
  productreviews5 <- rbind(productreviews5, data.frame(
                       category = product_category,
                       name = product_name,
                       reviewer = product_reviewer,
                       reviews = product_review,
                       verified = verified_reviews,
                       "date of review" = product_date,
                       ratings = product_rating))
 Sys.sleep(5)

productrev1 <- rbind(productreviews5,productreviews1,productreviews2,productreviews3,productreviews4)
write.csv(productrev1, file = "product1.csv")
```

```
link1 <- "https://www.amazon.com/product-reviews/B0C5LC519H/ref=acr_dpx_hist_5?ie=UTF8&filterByStar=five

  session <- bow(link1,
                 user_agent = "For Educational Purpose")

  scrapeNodes <- function(selector){
    scrape(session) %>%
      html_nodes(selector) %>%
      html_text(trim = TRUE)
  }
  product_category <- rep("Sling Bag", 10)

  product_name <- scrapeNodes("h1.a-size-large.a-text-ellipsis")
  product_name <- rep(product_name, 10)
  product_name <- product_name[1:10]

  verified_reviews <-scrapeNodes("span.a-size-mini.a-color-state.a-text-bold")
  verified_reviews <-verified_reviews[1:10]

  product_reviewer <- scrapeNodes("span.a-profile-name")
```

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

  product_review <- scrapeNodes("span.a-size-base.review-text")
  product_review <- product_review[1:10]

  product_date <- scrapeNodes("span.a-size-base.a-color-secondary.review-date")
  product_date <- product_date[1:10]

  product_rating <- scrapeNodes("span.a-icon-alt")
  product_rating <- product_rating[1:10]

  productreviews1= data.frame()
  productreviews1 <- rbind(productreviews1, data.frame(
                       category = product_category,
                       name = product_name,
                       reviewer = product_reviewer,
                       reviews = product_review,
                       verified = verified_reviews,
                       "date of review" = product_date,
                       ratings = product_rating))
 Sys.sleep(5)
```

```
link2 <- "https://www.amazon.com/product-reviews/B07YD4Q35C/ref=acr_dpx_hist_5?ie=UTF8&filterByStar=five

  session <- bow(link2,
                 user_agent = "For Educational Purpose")

  scrapeNodes <- function(selector){
    scrape(session) %>%
      html_nodes(selector) %>%
      html_text(trim = TRUE)
  }
  product_category <- rep("Sling Bag", 10)

  product_name <- scrapeNodes("h1.a-size-large.a-text-ellipsis")
  product_name <- rep(product_name, 10)
  product_name <- product_name[1:10]

  verified_reviews <-scrapeNodes("span.a-size-mini.a-color-state.a-text-bold")
  verified_reviews <-verified_reviews[1:10]

  product_reviewer <- scrapeNodes("span.a-profile-name")
  product_reviewer <- product_reviewer[1:10]

  product_review <- scrapeNodes("span.a-size-base.review-text")
  product_review <- product_review[1:10]

  product_date <- scrapeNodes("span.a-size-base.a-color-secondary.review-date")
  product_date <- product_date[1:10]

  product_rating <- scrapeNodes("span.a-icon-alt")
  product_rating <- product_rating[1:10]
```

```
  productreviews2= data.frame()
  productreviews2 <- rbind(productreviews2, data.frame(
                       category = product_category,
                       name = product_name,
                       reviewer = product_reviewer,
                       reviews = product_review,
                       verified = verified_reviews,
                       "date of review" = product_date,
                       ratings = product_rating))
 Sys.sleep(5)
```

```
link3 <- "https://www.amazon.com/product-reviews/B07ZKJVJ9H/ref=acr_dpx_hist_5?ie=UTF8&filterByStar=fiv

  session <- bow(link3,
                 user_agent = "For Educational Purpose")

  scrapeNodes <- function(selector){
    scrape(session) %>%
      html_nodes(selector) %>%
      html_text(trim = TRUE)
  }
  product_category <- rep("Sling Bag", 10)

  product_name <- scrapeNodes("h1.a-size-large.a-text-ellipsis")
  product_name <- rep(product_name, 10)
  product_name <- product_name[1:10]

  verified_reviews <-scrapeNodes("span.a-size-mini.a-color-state.a-text-bold")
  verified_reviews <-verified_reviews[1:10]

  product_reviewer <- scrapeNodes("span.a-profile-name")
  product_reviewer <- product_reviewer[1:10]

  product_review <- scrapeNodes("span.a-size-base.review-text")
  product_review <- product_review[1:10]

  product_date <- scrapeNodes("span.a-size-base.a-color-secondary.review-date")
  product_date <- product_date[1:10]

  product_rating <- scrapeNodes("span.a-icon-alt")
  product_rating <- product_rating[1:10]

  productreviews3= data.frame()
  productreviews3 <- rbind(productreviews3, data.frame(
                       category = product_category,
                       name = product_name,
                       reviewer = product_reviewer,
                       reviews = product_review,
                       verified = verified_reviews,
                       "date of review" = product_date,
                       ratings = product_rating))
 Sys.sleep(5)
```

```
link4 <- "https://www.amazon.com/product-reviews/B0BNN6DKLV/ref=acr_dpx_hist_5?ie=UTF8&filterByStar=five

  session <- bow(link4,
                 user_agent = "For Educational Purpose")

  scrapeNodes <- function(selector){
    scrape(session) %>%
      html_nodes(selector) %>%
      html_text(trim = TRUE)
  }
  product_category <- rep("Sling Bag", 10)

  product_name <- scrapeNodes("h1.a-size-large.a-text-ellipsis")
  product_name <- rep(product_name, 10)
  product_name <- product_name[1:10]

  verified_reviews <-scrapeNodes("span.a-size-mini.a-color-state.a-text-bold")
  verified_reviews <-verified_reviews[1:10]

  product_reviewer <- scrapeNodes("span.a-profile-name")
  product_reviewer <- product_reviewer[1:10]

  product_review <- scrapeNodes("span.a-size-base.review-text")
  product_review <- product_review[1:10]

  product_date <- scrapeNodes("span.a-size-base.a-color-secondary.review-date")
  product_date <- product_date[1:10]

  product_rating <- scrapeNodes("span.a-icon-alt")
  product_rating <- product_rating[1:10]

  productreviews4= data.frame()
  productreviews4 <- rbind(productreviews4, data.frame(
                       category = product_category,
                       name = product_name,
                       reviewer = product_reviewer,
                       reviews = product_review,
                       verified = verified_reviews,
                       "date of review" = product_date,
                       ratings = product_rating))
 Sys.sleep(5)
```

```
link5 <- "https://www.amazon.com/product-reviews/B0BNNBH1Q1/ref=acr_dpx_hist_5?ie=UTF8&filterByStar=five

  session <- bow(link5,
                 user_agent = "For Educational Purpose")

  scrapeNodes <- function(selector){
    scrape(session) %>%
      html_nodes(selector) %>%
      html_text(trim = TRUE)
  }
  product_category <- rep("Sling Bag", 10)
```

```
 product_name <- scrapeNodes("h1.a-size-large.a-text-ellipsis")
 product_name <- rep(product_name, 10)
 product_name <- product_name[1:10]

 verified_reviews <-scrapeNodes("span.a-size-mini.a-color-state.a-text-bold")
 verified_reviews <-verified_reviews[1:10]

 product_reviewer <- scrapeNodes("span.a-profile-name")
 product_reviewer <- product_reviewer[1:10]

 product_review <- scrapeNodes("span.a-size-base.review-text")
 product_review <- product_review[1:10]

 product_date <- scrapeNodes("span.a-size-base.a-color-secondary.review-date")
 product_date <- product_date[1:10]

 product_rating <- scrapeNodes("span.a-icon-alt")
 product_rating <- product_rating[1:10]

 productreviews5= data.frame()
 productreviews5 <- rbind(productreviews5, data.frame(
                     category = product_category,
                     name = product_name,
                     reviewer = product_reviewer,
                     reviews = product_review,
                     verified = verified_reviews,
                     "date of review" = product_date,
                     ratings = product_rating))
Sys.sleep(5)
productrev2 <- rbind(productreviews1,productreviews2,productreviews3,productreviews4,productreviews5)
write.csv(productrev2, file = "product2.csv")
```