RWorksheet#5B

Barrientos, Delfin, Infiesto

2024-11-20

Amazon Reviews Refrigerator

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

urls <- c(
  "https://www.amazon.com/Silonn-Countertop-Portable-Machine-Self-Cleaning/dp/B0BXXKFJK2",
  "https://www.amazon.com/EUHOMY-Countertop-Machine-Auto-Cleaning-Portable/dp/B0BWHZJHPL",
  "https://www.amazon.com/Silonn-Countertop-Portable-Self-cleaning-Bullet-shaped/dp/B08ZYJ8CRX",
  "https://www.amazon.com/EUHOMY-Countertop-Portable-Machine-Auto-Cleaning/dp/B0D3GV1VQS",
  "https://www.amazon.com/Countertop-Chewable-35-5Lbs-Machine-Self-Cleaning/dp/B0DBHN7WSN",
  "https://www.amazon.com/EUHOMY-Countertop-Removable-Auto-Cleaning-Portable/dp/B0C32SGKMJ",
  "https://www.amazon.com/Upstreman-Adjustable-Thermostat-Refrigerator-Black-BR321/dp/B09RWFZTW",
  "https://www.amazon.com/Silonn-Countertop-One-Click-Operation-Self-Cleaning/dp/B0CD2C8M4D",
  "https://www.amazon.com/ecozy-Portable-Countertop-Self-Cleaning-Standing/dp/B0B498C643",
  "https://www.amazon.com/Countertop-Silonn-Chewable-Self-Cleaning-Function/dp/B0C9SWH3RC"
)
#This is function from GPT, and I understand its function is to make the website think that th
headers <- httr::add_headers(
  `User-Agent` = "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Ge
)
```

```r
scrape_reviews <- function(url) {

  webpage <- httr::GET(url, headers) %>% read_html()


  product_name <- webpage %>%
    html_nodes("#productTitle") %>%
    html_text(trim = TRUE)

  product_reviewers <- webpage %>%
    html_nodes("span.a-profile-name") %>%
    html_text(trim = TRUE)


  product_date <- webpage %>%
    html_nodes(".review-date") %>%
    html_text(trim = TRUE)


  product_individual_rating <- webpage %>%
    html_nodes(".a-icon-alt") %>%
    html_text(trim = TRUE)


  product_text_review <- webpage %>%
    html_nodes(".review-text-content span") %>%
    html_text(trim = TRUE)


  product_title_of_review <- webpage %>%
    html_nodes(".review-title span") %>%
    html_text(trim = TRUE)


  verified_purchase <- webpage %>%
    html_nodes(".review-data span.a-size-mini.a-color-state.a-text-bold") %>%
    html_text(trim = TRUE)


  verified_purchase <- ifelse(verified_purchase == "Verified Purchase", "Yes", "No")


  min_length <- min(
    length(product_reviewers),
    length(product_date),
    length(product_individual_rating),
    length(product_text_review),
```

```r
      length(product_title_of_review),
      length(verified_purchase)
  )


  if (min_length > 0) {
    return(data.frame(
      Product_Name = product_name[1:min_length],
      Reviewers = product_reviewers[1:min_length],
      Date = product_date[1:min_length],
      Individual_Rating = product_individual_rating[1:min_length],
      Text_Review = product_text_review[1:min_length],
      Title_of_Review = product_title_of_review[1:min_length],
      Verified_Purchase = verified_purchase[1:min_length],
      stringsAsFactors = FALSE
    )[1:min(20, nrow(data.frame)), ])
  } else {
    NULL
  }
}


all_reviews <- list()


for (url in urls) {
  product_reviews <- scrape_reviews(url)

  if (!is.null(product_reviews)) {
    all_reviews <- append(all_reviews, list(product_reviews))
  } else {
    cat("Failed to scrape URL:", url, "\n")
  }
}
```

Failed to scrape URL: https://www.amazon.com/EUHOMY-Countertop-Portable-Machine-Auto-Cleanir

```r
final_reviews_in_Refrigerator <- bind_rows(all_reviews)


View(final_reviews_in_Refrigerator)
print(head(final_reviews_in_Refrigerator))
```

```
## 
## 1...1 Silonn Ice Maker Countertop, Portable Ice Machine with Carry Handle, Self-Cleaning Ice
## 2...2
```

```
## 3...3
## 4...4
## 5...5
## 6...6
##                     Reviewers                                                  Date
## 1...1   silonn ice maker Reviewed in the United States on September 11, 2024
## 2...2    Amazon Customer   Reviewed in the United States on October 23, 2024
## 3...3    Amazon Customer    Reviewed in the United States on August 14, 2024
## 4...4        Dirtty Eagle  Reviewed in the United States on November 23, 2024
## 5...5            Channing   Reviewed in the United States on October 25, 2024
## 6...6 kristina m. dieppa    Reviewed in the United States on November 3, 2024
##                                  Individual_Rating
## 1...1                            4.4 out of 5 stars
## 2...2 Previous page of related Sponsored Products
## 3...3     Next page of related Sponsored Products
## 4...4 Previous page of related Sponsored Products
## 5...5     Next page of related Sponsored Products
## 6...6                            4.4 out of 5 stars
##
## 1...1
## 2...2
## 3...3
## 4...4
## 5...5 Surprising yield per fill.  Set and forget, either bin full of ice or out of water, if
## 6...6
##                                  Title_of_Review Verified_Purchase
## 1...1                            5.0 out of 5 stars               Yes
## 2...2                                                             Yes
## 3...3 Compact, Efficient, and Perfect for On-the-Go!               Yes
## 4...4                            5.0 out of 5 stars               Yes
## 5...5                                                             Yes
## 6...6                  Great countertop icemaker               Yes
```

Televisions

```r
library(rvest)
library(httr)
library(dplyr)
library(polite)

urls <- c("https://www.amazon.com/tcl-fire-tv-55-inch-class-s5-4k-smart-tv/dp/B0D4PD799H", "htt
          )

headers <- httr::add_headers(
  `User-Agent` = "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Ge
)
scrape_reviews <- function(url) {
```

```r
webpage <- httr::GET(url, headers) %>% read_html()


product_name <- webpage %>%
  html_nodes("#productTitle") %>%
  html_text(trim = TRUE)

product_reviewers <- webpage %>%
  html_nodes("span.a-profile-name") %>%
  html_text(trim = TRUE)


product_date <- webpage %>%
  html_nodes(".review-date") %>%
  html_text(trim = TRUE)


product_individual_rating <- webpage %>%
  html_nodes(".a-icon-alt") %>%
  html_text(trim = TRUE)


product_text_review <- webpage %>%
  html_nodes(".review-text-content span") %>%
  html_text(trim = TRUE)


product_title_of_review <- webpage %>%
  html_nodes(".review-title span") %>%
  html_text(trim = TRUE)


verified_purchase <- webpage %>%
  html_nodes(".review-data span.a-size-mini.a-color-state.a-text-bold") %>%
  html_text(trim = TRUE)


verified_purchase <- ifelse(verified_purchase == "Verified Purchase", "Yes", "No")


min_length <- min(
  length(product_reviewers),
  length(product_date),
  length(product_individual_rating),
  length(product_text_review),
  length(product_title_of_review),
  length(verified_purchase)
```

```r
    )

  if (min_length > 0) {
   return(data.frame(
      Product_Name = product_name[1:min_length],
      Reviewers = product_reviewers[1:min_length],
      Date = product_date[1:min_length],
      Individual_Rating = product_individual_rating[1:min_length],
      Text_Review = product_text_review[1:min_length],
      Title_of_Review = product_title_of_review[1:min_length],
      Verified_Purchase = verified_purchase[1:min_length],
      stringsAsFactors = FALSE
    )[1:min(20, nrow(data.frame)), ])
  } else {
    NULL
  }
}


all_reviews <- list()


for (url in urls) {
  product_reviews <- scrape_reviews(url)

  if (!is.null(product_reviews)) {
    all_reviews <- append(all_reviews, list(product_reviews))
  } else {
    cat("Failed to scrape URL:", url, "\n")
  }
}


final_reviews_in_TV <- bind_rows(all_reviews)


View(final_reviews_in_TV)
print(head(final_reviews_in_TV))
```

```
##
## 1...1 TCL 55-Inch Class S5 UHD 4K LED Smart TV with Fire TV (55S551F, 2024 Model), Dolby Vi
## 2...2
## 3...3
## 4...4
## 5...5
## 6...6
```

```
##                     Reviewers
## 1...1 Great tv and great price!
## 2...2           Amazon Customer
## 3...3           Amazon Customer
## 4...4                     Kevin
## 5...5                     Kevin
## 6...6               michael b.
##                                                    Date
## 1...1 Reviewed in the United States on November 14, 2024
## 2...2  Reviewed in the United States on October 24, 2024
## 3...3 Reviewed in the United States on November 16, 2024
## 4...4  Reviewed in the United States on November 9, 2024
## 5...5  Reviewed in the United States on November 5, 2024
## 6...6 Reviewed in the United States on November 20, 2024
##                              Individual_Rating
## 1...1                        4.0 out of 5 stars
## 2...2                             Previous page
## 3...3                                 Next page
## 4...4                        4.1 out of 5 stars
## 5...5 Previous page of related Sponsored Products
## 6...6     Next page of related Sponsored Products
##
## 1...1
## 2...2
## 3...3 The Color is fantastic and the picture is pretty great as far as backlight, brightness
## 4...4
## 5...5
## 6...6
##          Title_of_Review Verified_Purchase
## 1...1 5.0 out of 5 stars               Yes
## 2...2                                  Yes
## 3...3     So far so good               Yes
## 4...4 5.0 out of 5 stars               Yes
## 5...5                                  Yes
## 6...6            Great TV               Yes
```

Freezers

```
library(rvest)
library(httr)
library(dplyr)
library(polite)


urls <- c("https://www.amazon.com/Electactic-Adjustable-Thermostat-Removable-Reversible/dp/B0D0
          "https://www.amazon.com/Midea-MRC04M3AWW-White-Freezer-Cubic/dp/B00MVVITWC",
          "https://www.amazon.com/BANGSON-Upright-Convertible-Refrigerator-Apartment/dp/B0DBL4
          "https://www.amazon.com/West-Bend-WB0500CFMW-Temperature-Free-Standing/dp/B0C1PCTG9T"
```

```r
headers <- httr::add_headers(
  `User-Agent` = "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Ge
)
scrape_reviews <- function(url) {

  webpage <- httr::GET(url, headers) %>% read_html()


  product_name <- webpage %>%
    html_nodes("#productTitle") %>%
    html_text(trim = TRUE)

  product_reviewers <- webpage %>%
    html_nodes("span.a-profile-name") %>%
    html_text(trim = TRUE)


  product_date <- webpage %>%
    html_nodes(".review-date") %>%
    html_text(trim = TRUE)


  product_individual_rating <- webpage %>%
    html_nodes(".a-icon-alt") %>%
    html_text(trim = TRUE)


  product_text_review <- webpage %>%
    html_nodes(".review-text-content span") %>%
    html_text(trim = TRUE)


  product_title_of_review <- webpage %>%
    html_nodes(".review-title span") %>%
    html_text(trim = TRUE)


  verified_purchase <- webpage %>%
    html_nodes(".review-data span.a-size-mini.a-color-state.a-text-bold") %>%
    html_text(trim = TRUE)


  verified_purchase <- ifelse(verified_purchase == "Verified Purchase", "Yes", "No")


  min_length <- min(
    length(product_reviewers),
```

```r
      length(product_date),
      length(product_individual_rating),
      length(product_text_review),
      length(product_title_of_review),
      length(verified_purchase)
  )


  if (min_length > 0) {
    return(data.frame(
      Product_Name = product_name[1:min_length],
      Reviewers = product_reviewers[1:min_length],
      Date = product_date[1:min_length],
      Individual_Rating = product_individual_rating[1:min_length],
      Text_Review = product_text_review[1:min_length],
      Title_of_Review = product_title_of_review[1:min_length],
      Verified_Purchase = verified_purchase[1:min_length],
      stringsAsFactors = FALSE
    )[1:min(20, nrow(data.frame)), ])
  } else {
    NULL
  }
}


all_reviews <- list()


for (url in urls) {
  product_reviews <- scrape_reviews(url)

  if (!is.null(product_reviews)) {
    all_reviews <- append(all_reviews, list(product_reviews))
  } else {
    cat("Failed to scrape URL:", url, "\n")
  }
}


final_reviews_in_Freezers <- bind_rows(all_reviews)


View(final_reviews_in_Freezers)
print(head(final_reviews_in_Freezers))


##
## 1...1 3.0 Cu.ft Upright Freezer, Small Mini Freezer with Adjustable Thermostat, Removable S
```

```
## 2...2
## 3...3
## 4...4
## 5...5
## 6...6
##             Reviewers                                          Date
## 1...1 Amazon Customer Reviewed in the United States on February 27, 2024
## 2...2 Amazon Customer Reviewed in the United States on November 19, 2024
## 3...3  Shirley Vatter  Reviewed in the United States on October 21, 2021
## 4...4      Roberto C. Reviewed in the United States on November 20, 2024
## 5...5      David Rowe  Reviewed in the United States on October 24, 2023
## 6...6  Carlene Morris      Reviewed in the United States on May 17, 2023
##                               Individual_Rating
## 1...1                            4.5 out of 5 stars
## 2...2 Previous page of related Sponsored Products
## 3...3     Next page of related Sponsored Products
## 4...4 Previous page of related Sponsored Products
## 5...5     Next page of related Sponsored Products
## 6...6                            4.5 out of 5 stars
##
## 1...1
## 2...2
## 3...3 There is a reason this compact freezer is several pounds lighter than most and less e
## 4...4
## 5...5
## 6...6
##                    Title_of_Review Verified_Purchase
## 1...1         5.0 out of 5 stars                 Yes
## 2...2                                            Yes
## 3...3     Perfect for my needs now               Yes
## 4...4         5.0 out of 5 stars                 Yes
## 5...5                                            Yes
## 6...6 Electactic 3 cf. upright freezer           Yes
```

Make-up

```
library(rvest)
library(httr)
library(dplyr)
library(polite)

urls <- c(
  "https://www.amazon.com/Peripera-Ink-Velvet-Lip-Tint/dp/B08XWHJLT6",
  "https://www.amazon.com/ColorStay-Stocking-Stuffers-Medium-Full-Coverage/dp/B004X8M59E",
  "https://www.amazon.com/Foundation-Coverage-Concealer-Flawless-Control/dp/B07W19Q5CK",
  "https://www.amazon.com/Teenagers-Eyeshadow-Foundation-Concealer-Lipgloss/dp/B0CCS563S6",
  "https://www.amazon.com/l-f-Reviver-Nourishing-High-shine-Cruelty-free/dp/B0CMJZ8G6Y",
```

```r
  "https://www.amazon.com/Liquid-Lipstick-Long-Lasting-Non-Stick-Waterproof/dp/B083QNQQP9","htt
)


headers <- httr::add_headers(
  `User-Agent` = "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Ge
)


scrape_reviews <- function(url) {

  webpage <- httr::GET(url, headers) %>% read_html()

  product_name <- webpage %>%
    html_nodes("#productTitle") %>%
    html_text(trim = TRUE)

  product_reviewers <- webpage %>%
    html_nodes("span.a-profile-name") %>%
    html_text(trim = TRUE)

  product_date <- webpage %>%
    html_nodes(".review-date") %>%
    html_text(trim = TRUE)

  product_individual_rating <- webpage %>%
    html_nodes(".a-icon-alt") %>%
    html_text(trim = TRUE)

  product_text_review <- webpage %>%
    html_nodes(".review-text-content span") %>%
    html_text(trim = TRUE)

  product_title_of_review <- webpage %>%
    html_nodes(".review-title span") %>%
    html_text(trim = TRUE)

  verified_purchase <- webpage %>%
    html_nodes(".review-data span.a-size-mini.a-color-state.a-text-bold") %>%
    html_text(trim = TRUE)

  verified_purchase <- ifelse(verified_purchase == "Verified Purchase", "Yes", "No")

  min_length <- min(
    length(product_reviewers),
    length(product_date),
    length(product_individual_rating),
```

```r
      length(product_text_review),
      length(product_title_of_review),
      length(verified_purchase)
  )

  if (min_length > 0) {
    return(data.frame(
      Product_Name = product_name[1:min_length],
      Reviewers = product_reviewers[1:min_length],
      Date = product_date[1:min_length],
      Individual_Rating = product_individual_rating[1:min_length],
      Text_Review = product_text_review[1:min_length],
      Title_of_Review = product_title_of_review[1:min_length],
      Verified_Purchase = verified_purchase[1:min_length],
      stringsAsFactors = FALSE
    )[1:min(20, nrow(data.frame)), ])
  } else {
    NULL
  }
}


all_reviews <- list()

for (url in urls) {
  product_reviews <- scrape_reviews(url)

  if (!is.null(product_reviews)) {
    all_reviews <- append(all_reviews, list(product_reviews))
  } else {
    cat("Failed to scrape URL:", url, "\n")
  }
}


final_reviews_in_MakeUp <- bind_rows(all_reviews)


View(final_reviews_in_MakeUp)
print(head(final_reviews_in_MakeUp))
```

```
##
## 1...1 Peripera Ink the Velvet Lip Tint - High Pigment Color, Longwear, Weightless, Not Anima
## 2...2
## 3...3
## 4...4
## 5...5
```

```
## 6...6
##                  Reviewers
## 1...1 Peripera Ink Velvet #17
## 2...2        Amazon Customer
## 3...3        Amazon Customer
## 4...4                  Val L
## 5...5                  Val L
## 6...6                   Ange
##                                          Date
## 1...1  Reviewed in the United States on November 4, 2024
## 2...2   Reviewed in the United States on October 3, 2024
## 3...3 Reviewed in the United States on November 14, 2024
## 4...4        Reviewed in the United States on May 8, 2023
## 5...5 Reviewed in the United States on November 21, 2024
## 6...6  Reviewed in the United States on October 10, 2024
##                                  Individual_Rating
## 1...1                           4.1 out of 5 stars
## 2...2 Previous page of related Sponsored Products
## 3...3     Next page of related Sponsored Products
## 4...4 Previous page of related Sponsored Products
## 5...5     Next page of related Sponsored Products
## 6...6                           4.1 out of 5 stars
##
## 1...1
## 2...2
## 3...3
## 4...4
## 5...5
## 6...6 I've been eyeing these lip tints for some time and now that I needed a particular colo
##               Title_of_Review Verified_Purchase
## 1...1   5.0 out of 5 stars               Yes
## 2...2                                     Yes
## 3...3   Gorgeous Lip Cream               Yes
## 4...4   5.0 out of 5 stars               Yes
## 5...5                                     Yes
## 6...6 My Favorite Lip Tint               Yes
```

Phones

```r
library(rvest)
library(httr)
library(dplyr)
library(polite)

urls <- c("https://www.amazon.com/Moto-Play-Unlocked-Camera-Sapphire/dp/B0CP6DDN1H", "https://w

headers <- httr::add_headers(
```

```
    `User-Agent` = "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Ge
)


scrape_reviews <- function(url) {

  webpage <- httr::GET(url, headers) %>% read_html()

  product_name <- webpage %>%
    html_nodes("#productTitle") %>%
    html_text(trim = TRUE)

  product_reviewers <- webpage %>%
    html_nodes("span.a-profile-name") %>%
    html_text(trim = TRUE)

  product_date <- webpage %>%
    html_nodes(".review-date") %>%
    html_text(trim = TRUE)

  product_individual_rating <- webpage %>%
    html_nodes(".a-icon-alt") %>%
    html_text(trim = TRUE)

  product_text_review <- webpage %>%
    html_nodes(".review-text-content span") %>%
    html_text(trim = TRUE)

  product_title_of_review <- webpage %>%
    html_nodes(".review-title span") %>%
    html_text(trim = TRUE)

  verified_purchase <- webpage %>%
    html_nodes(".review-data span.a-size-mini.a-color-state.a-text-bold") %>%
    html_text(trim = TRUE)

  verified_purchase <- ifelse(verified_purchase == "Verified Purchase", "Yes", "No")

  min_length <- min(
    length(product_reviewers),
    length(product_date),
    length(product_individual_rating),
    length(product_text_review),
    length(product_title_of_review),
    length(verified_purchase)
  )
```

```r
  if (min_length > 0) {
    return(data.frame(
      Product_Name = product_name[1:min_length],
      Reviewers = product_reviewers[1:min_length],
      Date = product_date[1:min_length],
      Individual_Rating = product_individual_rating[1:min_length],
      Text_Review = product_text_review[1:min_length],
      Title_of_Review = product_title_of_review[1:min_length],
      Verified_Purchase = verified_purchase[1:min_length],
      stringsAsFactors = FALSE
    )[1:min(20, nrow(data.frame)), ])
  } else {
    NULL
  }
}


all_reviews <- list()

for (url in urls) {
  product_reviews <- scrape_reviews(url)

  if (!is.null(product_reviews)) {
    all_reviews <- append(all_reviews, list(product_reviews))
  } else {
    cat("Failed to scrape URL:", url, "\n")
  }
}


final_reviews_in_CP <- bind_rows(all_reviews)


View(final_reviews_in_CP)
print(head(final_reviews_in_CP))
```

```
##                                                                    Product_Name
## 1...1 Moto G Play | 2024 | Unlocked | Made for US 4/64GB | 50MP Camera | Sapphire Blue
## 2...2                                                                      <NA>
## 3...3                                                                      <NA>
## 4...4                                                                      <NA>
## 5...5                                                                      <NA>
## 6...6                                                                      <NA>
##                     Reviewers
## 1...1 How to switch from iPhone
## 2...2          Amazon Customer
## 3...3          Amazon Customer
```

```
## 4...4               Otto D Woltke
## 5...5                 Bobvinyl
## 6...6                  paula.m
##                                                   Date
## 1...1 Reviewed in the United States on October 30, 2024
## 2...2   Reviewed in the United States on June 15, 2024
## 3...3     Reviewed in the United States on May 8, 2024
## 4...4    Reviewed in the United States on July 14, 2024
## 5...5  Reviewed in the United States on October 7, 2024
## 6...6            Reviewed in Canada on November 13, 2024
##                             Individual_Rating
## 1...1                        4.0 out of 5 stars
## 2...2 Previous page of related Sponsored Products
## 3...3     Next page of related Sponsored Products
## 4...4 Previous page of related Sponsored Products
## 5...5     Next page of related Sponsored Products
## 6...6                          4 out of 5 stars
##
## 1...1
## 2...2
## 3...3 How many apps do you think you have on your phone? 40? 50? how about your phone from
## 4...4
## 5...5
## 6...6
##                             Title_of_Review Verified_Purchase
## 1...1                     5.0 out of 5 stars               Yes
## 2...2                                                      Yes
## 3...3 Great option , to not get hit hard on your wallet.     Yes
## 4...4                     5.0 out of 5 stars               Yes
## 5...5                                                      Yes
## 6...6                 Best bang for your buck               Yes
```