Lab Exercise 1

Edora Frances Anne V Arcenas BSIT 2C

2024-02-15

SCRAPING FOR NOTEBOOKS

```
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)


notebook_products <- data.frame()

notebooklink1 = "https://www.amazon.co.uk/s?k=notebook&crid=3PE2BSTUHI0D6&qid=1707353799&sprefix=notebo

session1 <- bow(notebooklink1,
           user_agent = "Educational Purpose")

scrapeNodes <- function(selector){
scrape(session1) %>%
  html_nodes(selector) %>%
  html_text(trim = TRUE)
}

scrapedCategory <- "Notebook"

scrapedName <- scrapeNodes("h2.a-size-mini.a-spacing-none.a-color-base.s-line-clamp-4")
scrapedName <- scrapedName[1:45]

scrapedPrice <- scrapeNodes("span.a-offscreen")
scrapedPrice <- scrapedPrice[1:45]

scrapedRatings <- scrapeNodes("span.a-icon-alt")
scrapedRatings <- scrapedRatings[1:45]

scrapedReviews <- scrapeNodes("span.a-size-base.s-underline-text")
```

```
scrapedReviews <- scrapedReviews[1:45]



notebook_products <- rbind(notebook_products, data.frame(category = scrapedCategory,name = scrapedName,
                     price = scrapedPrice,
                     ratings = scrapedRatings,
                     no_of_reviews = scrapedReviews))

###################

notebooklink2 = "https://www.amazon.co.uk/s?k=notebook&page=2&crid=3PE2BSTUHI0D6&qid=1707353804&sprefix=

session2 <- bow(notebooklink2,
           user_agent = "For Educational Purpose")

scrapedCategory2 <- "Notebook"

scrapedName2 <- scrapeNodes("h2.a-size-mini.a-spacing-none.a-color-base.s-line-clamp-4")
scrapedName2 <- scrapedName2[1:45]

scrapedPrice2 <- scrapeNodes("span.a-offscreen")
scrapedPrice2 <- scrapedPrice2[1:45]

scrapedRatings2 <- scrapeNodes("span.a-icon-alt")
scrapedRatings2 <- scrapedRatings2[1:45]

scrapedReviews2 <- scrapeNodes("span.a-size-base.s-underline-text")
scrapedReviews2 <- scrapedReviews2[1:45]



notebook_products <- rbind(notebook_products, data.frame(category = scrapedCategory,name = scrapedName,
                     price = scrapedPrice,
                     ratings = scrapedRatings,
                     no_of_reviews = scrapedReviews))


###################

notebooklink3 = "https://www.amazon.co.uk/s?k=notebook&page=3&crid=3PE2BSTUHI0D6&qid=1707354306&sprefix=

session3 <- bow(notebooklink3,
           user_agent = "For Educational Purpose")

scrapedCategory3 <- "Notebook"

scrapedName3 <- scrapeNodes("h2.a-size-mini.a-spacing-none.a-color-base.s-line-clamp-4")
scrapedName3 <- scrapedName3[1:45]

scrapedPrice3 <- scrapeNodes("span.a-offscreen")
scrapedPrice3 <- scrapedPrice3[1:45]
```

```
scrapedRatings3 <- scrapeNodes("span.a-icon-alt")
scrapedRatings3 <- scrapedRatings3[1:45]

scrapedReviews3 <- scrapeNodes("span.a-size-base.s-underline-text")
scrapedReviews3 <- scrapedReviews3[1:45]



notebook_products <- rbind(notebook_products, data.frame(category = scrapedCategory,name = scrapedName,
                     price = scrapedPrice,
                     ratings = scrapedRatings,
                     no_of_reviews = scrapedReviews))

notebook_products <- notebook_products[1:100,]
#write.csv(notebook_products,"notebook_products.csv")

#View(notebook_products)

##################
```

SCRAPING FOR PENCILS

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


pencil_products <- data.frame()

pencillink1_1_1 = "https://www.amazon.co.uk/s?k=pencil&crid=ROFYFPYBSIDI&qid=1707354893&sprefix=penicl%

session4_4_4 <- bow(pencillink1_1_1,
           user_agent = "Education")

scrapeNodes <- function(selector){
scrape(session4_4_4) %>%
  html_nodes(selector) %>%
  html_text(trim = TRUE)
}

scrapedCategory4 <- "Pencil"

scrapedpencilName <- scrapeNodes("h2.a-size-mini.a-spacing-none.a-color-base.s-line-clamp-4")
scrapedpencilName <- scrapedpencilName[1:45]

scrapedpencilPrice <- scrapeNodes("span.a-offscreen")
scrapedpencilPrice <- scrapedpencilPrice[1:45]

scrapedpencilRatings <- scrapeNodes("span.a-icon-alt")
scrapedpencilRatings <- scrapedpencilRatings[1:45]

scrapedpencilReviews <- scrapeNodes("span.a-size-base.s-underline-text")
scrapedpencilReviews <- scrapedpencilReviews[1:45]
```

```
pencil_products <- rbind(pencil_products, data.frame(category = scrapedCategory4,name = scrapedpencilNa
                    price = scrapedpencilPrice,
                    ratings = scrapedpencilRatings,
                    no_of_reviews = scrapedpencilReviews))

##################

pencillink2_2_2 = "https://www.amazon.co.uk/s?k=pencil&crid=ROFYFPYBSIDI&qid=1707354893&sprefix=penicl%

session5_5_5 <- bow(pencillink2_2_2,
           user_agent = "Educational Purpose")

scrapeNodes <- function(selector){
scrape(session5_5_5) %>%
  html_nodes(selector) %>%
  html_text(trim = TRUE)
}

scrapedCategory5 <- "Pencil"

scrapedpencilName2 <- scrapeNodes("h2.a-size-mini.a-spacing-none.a-color-base.s-line-clamp-4")
scrapedpencilName2 <- scrapedpencilName2[1:45]

scrapedpencilPrice2 <- scrapeNodes("span.a-offscreen")
scrapedpencilPrice2 <- scrapedpencilPrice2[1:45]

scrapedpencilRatings2 <- scrapeNodes("span.a-icon-alt")
scrapedpencilRatings2 <- scrapedpencilRatings2[1:45]

scrapedpencilReviews2 <- scrapeNodes("span.a-size-base.s-underline-text")
scrapedpencilReviews2 <- scrapedpencilReviews2[1:45]



pencil_products <- rbind(pencil_products, data.frame(category = scrapedCategory5,name = scrapedpencilNa
                    price = scrapedpencilPrice2,
                    ratings = scrapedpencilRatings2,
                    no_of_reviews = scrapedpencilReviews2))


##################

pencillink3_3_3 = "https://www.amazon.co.uk/s?k=pencil&crid=ROFYFPYBSIDI&qid=1707354893&sprefix=penicl%

session6_6_6 <- bow(pencillink3_3_3,
           user_agent = "Educational Purpose")

scrapeNodes <- function(selector){
scrape(session6_6_6) %>%
  html_nodes(selector) %>%
  html_text(trim = TRUE)
```

```
}

scrapedCategory6 <- "Pencil"

scrapedpencilName3 <- scrapeNodes("h2.a-size-mini.a-spacing-none.a-color-base.s-line-clamp-4")
scrapedpencilName3 <- scrapedpencilName3[1:45]

scrapedpencilPrice3 <- scrapeNodes("span.a-offscreen")
scrapedpencilPrice3 <- scrapedpencilPrice3[1:45]

scrapedpencilRatings3 <- scrapeNodes("span.a-icon-alt")
scrapedpencilRatings3 <- scrapedpencilRatings3[1:45]

scrapedpencilReviews3 <- scrapeNodes("span.a-size-base.s-underline-text")
scrapedpencilReviews3 <- scrapedpencilReviews3[1:45]



pencil_products <- rbind(pencil_products, data.frame(category = scrapedCategory5,name = scrapedpencilNa
                    price = scrapedpencilPrice2,
                    ratings = scrapedpencilRatings2,
                    no_of_reviews = scrapedpencilReviews2))

pencil_products <- pencil_products[1:100,]
write.csv(pencil_products,"pencil_products.csv")

#View(pencil_products)
```

MERGING TWO PRODUCTS

```
scraped_products <- rbind(notebook_products, pencil_products)
#write.csv(scraped_products, "byCategory.csv")
#View(scraped_products)
```

10 PRODUCTS

product1

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

product1 = data.frame()
scrapedCategory <- "Notebook"


###

  link = "https://www.amazon.co.uk/Blue-Acorn-Spiral-Bound-Sketchbook/product-reviews/B0CH3KXXSN/ref=cm

      sessionProd1 <- bow(link,
                user_agent = "For Educational Purpose")
```

```
      scrapeNodes <- function(selector){
      scrape(sessionProd1) %>%
        html_nodes(selector) %>%
        html_text(trim = TRUE)
      }

      scrapedprodName <- scrapeNodes("h1.a-size-large.a-text-ellipsis")


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


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


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


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


      product1 <- rbind(product1, data.frame(category = scrapedCategory,name = scrapedprodName,
                           reviewer = scrapedReviewer,
                           review = scrapedReview,
                           date = scrapedDate,
                          ratings = scrapedprodRating))

      Sys.sleep(5)
######

      link2 = "https://www.amazon.co.uk/Blue-Acorn-Spiral-Bound-Sketchbook/product-reviews/B0CH3KXXSN/r

  sessionProd2 <- bow(link2,
                user_agent = "For Educational Purpose")

      scrapedReviewer2 <- scrapeNodes("span.a-profile-name")
      scrapedReviewer2 <- scrapedReviewer2[1:10]


      scrapedReview2 <- scrapeNodes("span.a-size-base.review-text.review-text-content")
      scrapedReview2 <- scrapedReview2[1:10]


      scrapedDate2 <- scrapeNodes("span.a-size-base.a-color-secondary.review-date")
      scrapedDate2 <- scrapedDate2[1:10]


      scrapedprodRating2 <- scrapeNodes("span.a-icon-alt")
      scrapedprodRating2 <- scrapedprodRating2[1:10]
```

```
    product1 <- rbind(product1, data.frame(category = scrapedCategory ,name = scrapedprodName,
                        reviewer = scrapedReviewer2,
                        review = scrapedReview2,
                        date = scrapedDate2,
                       ratings = scrapedprodRating2))
    Sys.sleep(5)

    ####

    link3 = "https://www.amazon.co.uk/Blue-Acorn-Spiral-Bound-Sketchbook/product-reviews/B0CH3KXXSN/r

  sessionProd3 <- bow(link3,
              user_agent = "For Educational Purpose")


    scrapedReviewer3 <- scrapeNodes("span.a-profile-name")
    scrapedReviewer3 <- scrapedReviewer3[1:10]


    scrapedReview3 <- scrapeNodes("span.a-size-base.review-text.review-text-content")
    scrapedReview3 <- scrapedReview3[1:10]


    scrapedDate3 <- scrapeNodes("span.a-size-base.a-color-secondary.review-date")
    scrapedDate3 <- scrapedDate3[1:10]


    scrapedprodRating3 <- scrapeNodes("span.a-icon-alt")
    scrapedprodRating3 <- scrapedprodRating3[1:10]


    product1 <- rbind(product1, data.frame(category = scrapedCategory ,name = scrapedprodName,
                        reviewer = scrapedReviewer3,
                        review = scrapedReview3,
                        date = scrapedDate3,
                       ratings = scrapedprodRating3))
    Sys.sleep(5)

#View(product1)
write.csv(product1,"product1.csv")
```

product2

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

product2 = data.frame()
scrapedCategory <- "Notebook"
```

```
###

  link = "https://www.amazon.co.uk/Oxford-Notebook-Subject-College-63756/product-reviews/B08VF9B772/ref=

      sessionProd1 <- bow(link,
                user_agent = "For Educational Purpose")

      scrapeNodes <- function(selector){
      scrape(sessionProd1) %>%
        html_nodes(selector) %>%
        html_text(trim = TRUE)
      }

      scrapedprodName <- scrapeNodes("h1.a-size-large.a-text-ellipsis")


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


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


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


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


      product2 <- rbind(product2, data.frame(category = scrapedCategory,name = scrapedprodName,
                          reviewer = scrapedReviewer,
                          review = scrapedReview,
                          date = scrapedDate,
                         ratings = scrapedprodRating))
      Sys.sleep(5)
######

      link2 = "https://www.amazon.co.uk/Oxford-Notebook-Subject-College-63756/product-reviews/B08VF9B77

  sessionProd2 <- bow(link2,
                user_agent = "For Educational Purpose")

      scrapedReviewer2 <- scrapeNodes("span.a-profile-name")
      scrapedReviewer2 <- scrapedReviewer2[1:10]


      scrapedReview2 <- scrapeNodes("span.a-size-base.review-text.review-text-content")
      scrapedReview2 <- scrapedReview2[1:10]
```

```
      scrapedDate2 <- scrapeNodes("span.a-size-base.a-color-secondary.review-date")
      scrapedDate2 <- scrapedDate2[1:10]


      scrapedprodRating2 <- scrapeNodes("span.a-icon-alt")
      scrapedprodRating2 <- scrapedprodRating2[1:10]


      product2 <- rbind(product2, data.frame(category = scrapedCategory ,name = scrapedprodName,
                          reviewer = scrapedReviewer2,
                          review = scrapedReview2,
                          date = scrapedDate2,
                         ratings = scrapedprodRating2))

      Sys.sleep(5)
      ####

      link3 = "https://www.amazon.co.uk/Oxford-Notebook-Subject-College-63756/product-reviews/B08VF9B771

  sessionProd3 <- bow(link3,
               user_agent = "For Educational Purpose")


      scrapedReviewer3 <- scrapeNodes("span.a-profile-name")
      scrapedReviewer3 <- scrapedReviewer3[1:10]


      scrapedReview3 <- scrapeNodes("span.a-size-base.review-text.review-text-content")
      scrapedReview3 <- scrapedReview3[1:10]


      scrapedDate3 <- scrapeNodes("span.a-size-base.a-color-secondary.review-date")
      scrapedDate3 <- scrapedDate3[1:10]


      scrapedprodRating3 <- scrapeNodes("span.a-icon-alt")
      scrapedprodRating3 <- scrapedprodRating3[1:10]


      product2 <- rbind(product2, data.frame(category = scrapedCategory ,name = scrapedprodName,
                          reviewer = scrapedReviewer3,
                          review = scrapedReview3,
                          date = scrapedDate3,
                         ratings = scrapedprodRating3))
      Sys.sleep(5)
#View(product2)
write.csv(product2,"product2.csv")
```

product3

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

```
library(selectr)

product3 = data.frame()
scrapedCategory <- "Notebook"


###

  link = "https://www.amazon.co.uk/Leitz-Covered-Notebook-Squared-Sheets/product-reviews/B073QYZD5Q/ref=

      sessionProd1 <- bow(link,
               user_agent = "For Educational Purpose")

      scrapeNodes <- function(selector){
      scrape(sessionProd1) %>%
        html_nodes(selector) %>%
        html_text(trim = TRUE)
      }

      scrapedprodName <- scrapeNodes("h1.a-size-large.a-text-ellipsis")


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


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


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


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


      product3 <- rbind(product3, data.frame(category = scrapedCategory,name = scrapedprodName,
                          reviewer = scrapedReviewer,
                          review = scrapedReview,
                          date = scrapedDate,
                         ratings = scrapedprodRating))
      Sys.sleep(5)
######

      link2 = "https://www.amazon.co.uk/Leitz-Covered-Notebook-Squared-Sheets/product-reviews/B073QYZD5

  sessionProd2 <- bow(link2,
               user_agent = "For Educational Purpose")

      scrapedReviewer2 <- scrapeNodes("span.a-profile-name")
      scrapedReviewer2 <- scrapedReviewer2[1:10]
```

```
      scrapedReview2 <- scrapeNodes("span.a-size-base.review-text.review-text-content")
      scrapedReview2 <- scrapedReview2[1:10]


      scrapedDate2 <- scrapeNodes("span.a-size-base.a-color-secondary.review-date")
      scrapedDate2 <- scrapedDate2[1:10]


      scrapedprodRating2 <- scrapeNodes("span.a-icon-alt")
      scrapedprodRating2 <- scrapedprodRating2[1:10]


      product3 <- rbind(product3, data.frame(category = scrapedCategory ,name = scrapedprodName,
                          reviewer = scrapedReviewer2,
                          review = scrapedReview2,
                          date = scrapedDate2,
                         ratings = scrapedprodRating2))

      Sys.sleep(5)
      ####

      link3 = "https://www.amazon.co.uk/Leitz-Covered-Notebook-Squared-Sheets/product-reviews/B073QYZD5

  sessionProd3 <- bow(link3,
                user_agent = "For Educational Purpose")


      scrapedReviewer3 <- scrapeNodes("span.a-profile-name")
      scrapedReviewer3 <- scrapedReviewer3[1:10]


      scrapedReview3 <- scrapeNodes("span.a-size-base.review-text.review-text-content")
      scrapedReview3 <- scrapedReview3[1:10]


      scrapedDate3 <- scrapeNodes("span.a-size-base.a-color-secondary.review-date")
      scrapedDate3 <- scrapedDate3[1:10]


      scrapedprodRating3 <- scrapeNodes("span.a-icon-alt")
      scrapedprodRating3 <- scrapedprodRating3[1:10]


      product3 <- rbind(product3, data.frame(category = scrapedCategory ,name = scrapedprodName,
                          reviewer = scrapedReviewer3,
                          review = scrapedReview3,
                          date = scrapedDate3,
                         ratings = scrapedprodRating3))

      Sys.sleep(5)
#View(product3)
write.csv(product3,"product3.csv")
```

product4

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

product4 = data.frame()
scrapedCategory <- "Notebook"


###

  link = "https://www.amazon.co.uk/Oxford-My-Notes-Wirebound-Perforated/product-reviews/B00BI5N0GW/ref=

      sessionProd1 <- bow(link,
               user_agent = "For Educational Purpose")

      scrapeNodes <- function(selector){
      scrape(sessionProd1) %>%
        html_nodes(selector) %>%
        html_text(trim = TRUE)
      }

      scrapedprodName <- scrapeNodes("h1.a-size-large.a-text-ellipsis")


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


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


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


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


      product4 <- rbind(product4, data.frame(category = scrapedCategory,name = scrapedprodName,
                          reviewer = scrapedReviewer,
                          review = scrapedReview,
                          date = scrapedDate,
                         ratings = scrapedprodRating))

      Sys.sleep(5)
######

      link2 = "https://www.amazon.co.uk/Oxford-My-Notes-Wirebound-Perforated/product-reviews/B00BI5N0GW
```

```
sessionProd2 <- bow(link2,
             user_agent = "For Educational Purpose")

    scrapedReviewer2 <- scrapeNodes("span.a-profile-name")
    scrapedReviewer2 <- scrapedReviewer2[1:10]


    scrapedReview2 <- scrapeNodes("span.a-size-base.review-text.review-text-content")
    scrapedReview2 <- scrapedReview2[1:10]


    scrapedDate2 <- scrapeNodes("span.a-size-base.a-color-secondary.review-date")
    scrapedDate2 <- scrapedDate2[1:10]


    scrapedprodRating2 <- scrapeNodes("span.a-icon-alt")
    scrapedprodRating2 <- scrapedprodRating2[1:10]


    product4 <- rbind(product4, data.frame(category = scrapedCategory ,name = scrapedprodName,
                        reviewer = scrapedReviewer2,
                        review = scrapedReview2,
                        date = scrapedDate2,
                       ratings = scrapedprodRating2))


    Sys.sleep(5)
    ####

    link3 = "https://www.amazon.co.uk/Oxford-My-Notes-Wirebound-Perforated/product-reviews/B00BI5N0GW/

sessionProd3 <- bow(link3,
             user_agent = "For Educational Purpose")


    scrapedReviewer3 <- scrapeNodes("span.a-profile-name")
    scrapedReviewer3 <- scrapedReviewer3[1:10]


    scrapedReview3 <- scrapeNodes("span.a-size-base.review-text.review-text-content")
    scrapedReview3 <- scrapedReview3[1:10]


    scrapedDate3 <- scrapeNodes("span.a-size-base.a-color-secondary.review-date")
    scrapedDate3 <- scrapedDate3[1:10]


    scrapedprodRating3 <- scrapeNodes("span.a-icon-alt")
    scrapedprodRating3 <- scrapedprodRating3[1:10]


    product4 <- rbind(product4, data.frame(category = scrapedCategory ,name = scrapedprodName,
                        reviewer = scrapedReviewer3,
```

```
                          review = scrapedReview3,
                          date = scrapedDate3,
                         ratings = scrapedprodRating3))

      Sys.sleep(5)
#View(product4)
write.csv(product4,"product4.csv")
```

product5

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

product5 = data.frame()
scrapedCategory <- "Notebook"


###

  link = "https://www.amazon.co.uk/Amazon-Basics-Casebound-Notebook-21x13-3x1-6/product-reviews/B01DN8T

      sessionProd1 <- bow(link,
               user_agent = "For Educational Purpose")

      scrapeNodes <- function(selector){
      scrape(sessionProd1) %>%
        html_nodes(selector) %>%
        html_text(trim = TRUE)
      }

      scrapedprodName <- scrapeNodes("h1.a-size-large.a-text-ellipsis")


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


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


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


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


      product5 <- rbind(product5, data.frame(category = scrapedCategory,name = scrapedprodName,
                          reviewer = scrapedReviewer,
```

```
                          review = scrapedReview,
                          date = scrapedDate,
                         ratings = scrapedprodRating))

     Sys.sleep(5)
######

     link2 = "https://www.amazon.co.uk/Amazon-Basics-Casebound-Notebook-21x13-3x1-6/product-reviews/B0

  sessionProd2 <- bow(link2,
              user_agent = "For Educational Purpose")

     scrapedReviewer2 <- scrapeNodes("span.a-profile-name")
     scrapedReviewer2 <- scrapedReviewer2[1:10]


     scrapedReview2 <- scrapeNodes("span.a-size-base.review-text.review-text-content")
     scrapedReview2 <- scrapedReview2[1:10]


     scrapedDate2 <- scrapeNodes("span.a-size-base.a-color-secondary.review-date")
     scrapedDate2 <- scrapedDate2[1:10]


     scrapedprodRating2 <- scrapeNodes("span.a-icon-alt")
     scrapedprodRating2 <- scrapedprodRating2[1:10]


     product5 <- rbind(product5, data.frame(category = scrapedCategory ,name = scrapedprodName,
                          reviewer = scrapedReviewer2,
                          review = scrapedReview2,
                          date = scrapedDate2,
                         ratings = scrapedprodRating2))


     Sys.sleep(5)
     ####

     link3 = "https://www.amazon.co.uk/Amazon-Basics-Casebound-Notebook-21x13-3x1-6/product-reviews/B0

  sessionProd3 <- bow(link3,
              user_agent = "For Educational Purpose")


     scrapedReviewer3 <- scrapeNodes("span.a-profile-name")
     scrapedReviewer3 <- scrapedReviewer3[1:10]


     scrapedReview3 <- scrapeNodes("span.a-size-base.review-text.review-text-content")
     scrapedReview3 <- scrapedReview3[1:10]


     scrapedDate3 <- scrapeNodes("span.a-size-base.a-color-secondary.review-date")
```

```
      scrapedDate3 <- scrapedDate3[1:10]


      scrapedprodRating3 <- scrapeNodes("span.a-icon-alt")
      scrapedprodRating3 <- scrapedprodRating3[1:10]


      product5 <- rbind(product5, data.frame(category = scrapedCategory ,name = scrapedprodName,
                          reviewer = scrapedReviewer3,
                          review = scrapedReview3,
                          date = scrapedDate3,
                         ratings = scrapedprodRating3))
      Sys.sleep(5)

#View(product5)
#write.csv(product5,"product5.csv")
```

product6

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

product6 = data.frame()
scrapedCategory <- "Pencil"


###

  link = "https://www.amazon.co.uk/STAEDTLER-121-2-BK5D-Noris-pencils/product-reviews/B093L3F7GN/ref=cm

      sessionProd1 <- bow(link,
               user_agent = "For Educational Purpose")

      scrapeNodes <- function(selector){
      scrape(sessionProd1) %>%
        html_nodes(selector) %>%
        html_text(trim = TRUE)
      }

      scrapedprodName <- scrapeNodes("h1.a-size-large.a-text-ellipsis")


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


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


      scrapedDate <- scrapeNodes("span.a-size-base.a-color-secondary.review-date")
```

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


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


      product6 <- rbind(product6, data.frame(category = scrapedCategory,name = scrapedprodName,
                          reviewer = scrapedReviewer,
                          review = scrapedReview,
                          date = scrapedDate,
                         ratings = scrapedprodRating))
######

      link2 = "https://www.amazon.co.uk/STAEDTLER-121-2-BK5D-Noris-pencils/product-reviews/B093L3F7GN/r

  sessionProd2 <- bow(link2,
               user_agent = "For Educational Purpose")

      scrapedReviewer2 <- scrapeNodes("span.a-profile-name")
      scrapedReviewer2 <- scrapedReviewer2[1:10]


      scrapedReview2 <- scrapeNodes("span.a-size-base.review-text.review-text-content")
      scrapedReview2 <- scrapedReview2[1:10]


      scrapedDate2 <- scrapeNodes("span.a-size-base.a-color-secondary.review-date")
      scrapedDate2 <- scrapedDate2[1:10]


      scrapedprodRating2 <- scrapeNodes("span.a-icon-alt")
      scrapedprodRating2 <- scrapedprodRating2[1:10]


      product6 <- rbind(product6, data.frame(category = scrapedCategory ,name = scrapedprodName,
                          reviewer = scrapedReviewer2,
                          review = scrapedReview2,
                          date = scrapedDate2,
                         ratings = scrapedprodRating2))


      ####

      link3 = "https://www.amazon.co.uk/STAEDTLER-121-2-BK5D-Noris-pencils/product-reviews/B093L3F7GN/r

  sessionProd3 <- bow(link3,
               user_agent = "For Educational Purpose")


      scrapedReviewer3 <- scrapeNodes("span.a-profile-name")
      scrapedReviewer3 <- scrapedReviewer3[1:10]
```

```
      scrapedReview3 <- scrapeNodes("span.a-size-base.review-text.review-text-content")
      scrapedReview3 <- scrapedReview3[1:10]


      scrapedDate3 <- scrapeNodes("span.a-size-base.a-color-secondary.review-date")
      scrapedDate3 <- scrapedDate3[1:10]


      scrapedprodRating3 <- scrapeNodes("span.a-icon-alt")
      scrapedprodRating3 <- scrapedprodRating3[1:10]


      product6 <- rbind(product6, data.frame(category = scrapedCategory ,name = scrapedprodName,
                          reviewer = scrapedReviewer3,
                          review = scrapedReview3,
                          date = scrapedDate3,
                         ratings = scrapedprodRating3))

      Sys.sleep(5)

#View(product6)
      write.csv(product6,"product6.csv")
```

product7

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

product7 = data.frame()
scrapedCategory <- "Pencil"


###

  link = "https://www.amazon.co.uk/STAEDTLER-121-2-BK5D-Noris-pencils/product-reviews/B093L3F7GN/ref=cm

      sessionProd1 <- bow(link,
               user_agent = "For Educational Purpose")

      scrapeNodes <- function(selector){
      scrape(sessionProd1) %>%
        html_nodes(selector) %>%
        html_text(trim = TRUE)
      }

      scrapedprodName <- scrapeNodes("h1.a-size-large.a-text-ellipsis")


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

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


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


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


      product7 <- rbind(product7, data.frame(category = scrapedCategory,name = scrapedprodName,
                          reviewer = scrapedReviewer,
                          review = scrapedReview,
                          date = scrapedDate,
                         ratings = scrapedprodRating))
      Sys.sleep(5)
######

      link2 = "https://www.amazon.co.uk/STAEDTLER-121-2-BK5D-Noris-pencils/product-reviews/B093L3F7GN/r

  sessionProd2 <- bow(link2,
               user_agent = "For Educational Purpose")

      scrapedReviewer2 <- scrapeNodes("span.a-profile-name")
      scrapedReviewer2 <- scrapedReviewer2[1:10]


      scrapedReview2 <- scrapeNodes("span.a-size-base.review-text.review-text-content")
      scrapedReview2 <- scrapedReview2[1:10]


      scrapedDate2 <- scrapeNodes("span.a-size-base.a-color-secondary.review-date")
      scrapedDate2 <- scrapedDate2[1:10]


      scrapedprodRating2 <- scrapeNodes("span.a-icon-alt")
      scrapedprodRating2 <- scrapedprodRating2[1:10]


      product7 <- rbind(product7, data.frame(category = scrapedCategory ,name = scrapedprodName,
                          reviewer = scrapedReviewer2,
                          review = scrapedReview2,
                          date = scrapedDate2,
                         ratings = scrapedprodRating2))

      Sys.sleep(5)
      ####

      link3 = "https://www.amazon.co.uk/STAEDTLER-121-2-BK5D-Noris-pencils/product-reviews/B093L3F7GN/r

  sessionProd3 <- bow(link3,
```

```
              user_agent = "For Educational Purpose")


      scrapedReviewer3 <- scrapeNodes("span.a-profile-name")
      scrapedReviewer3 <- scrapedReviewer3[1:10]


      scrapedReview3 <- scrapeNodes("span.a-size-base.review-text.review-text-content")
      scrapedReview3 <- scrapedReview3[1:10]


      scrapedDate3 <- scrapeNodes("span.a-size-base.a-color-secondary.review-date")
      scrapedDate3 <- scrapedDate3[1:10]


      scrapedprodRating3 <- scrapeNodes("span.a-icon-alt")
      scrapedprodRating3 <- scrapedprodRating3[1:10]


      product7 <- rbind(product7, data.frame(category = scrapedCategory ,name = scrapedprodName,
                          reviewer = scrapedReviewer3,
                          review = scrapedReview3,
                          date = scrapedDate3,
                         ratings = scrapedprodRating3))
      Sys.sleep(5)

#View(product7)
write.csv(product7,"product7.csv")
```

product8

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

product8 = data.frame()
scrapedCategory <- "Pencil"


###

  link = "https://www.amazon.co.uk/AmazonBasics-Presharpened-Wood-Cased-Pencils/product-reviews/B071JM6

      sessionProd1 <- bow(link,
              user_agent = "For Educational Purpose")

      scrapeNodes <- function(selector){
      scrape(sessionProd1) %>%
        html_nodes(selector) %>%
        html_text(trim = TRUE)
      }
```

```
      scrapedprodName <- scrapeNodes("h1.a-size-large.a-text-ellipsis")


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


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


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


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


      product8 <- rbind(product8, data.frame(category = scrapedCategory,name = scrapedprodName,
                          reviewer = scrapedReviewer,
                          review = scrapedReview,
                          date = scrapedDate,
                         ratings = scrapedprodRating))

      Sys.sleep(5)
######

      link2 = "https://www.amazon.co.uk/AmazonBasics-Presharpened-Wood-Cased-Pencils/product-reviews/B0

  sessionProd2 <- bow(link2,
                user_agent = "For Educational Purpose")

      scrapedReviewer2 <- scrapeNodes("span.a-profile-name")
      scrapedReviewer2 <- scrapedReviewer2[1:10]


      scrapedReview2 <- scrapeNodes("span.a-size-base.review-text.review-text-content")
      scrapedReview2 <- scrapedReview2[1:10]


      scrapedDate2 <- scrapeNodes("span.a-size-base.a-color-secondary.review-date")
      scrapedDate2 <- scrapedDate2[1:10]


      scrapedprodRating2 <- scrapeNodes("span.a-icon-alt")
      scrapedprodRating2 <- scrapedprodRating2[1:10]


      product8 <- rbind(product8, data.frame(category = scrapedCategory ,name = scrapedprodName,
                          reviewer = scrapedReviewer2,
                          review = scrapedReview2,
                          date = scrapedDate2,
```

```
                        ratings = scrapedprodRating2))

    Sys.sleep(5)

    ####

    link3 = "https://www.amazon.co.uk/AmazonBasics-Presharpened-Wood-Cased-Pencils/product-reviews/B0

  sessionProd3 <- bow(link3,
              user_agent = "For Educational Purpose")


    scrapedReviewer3 <- scrapeNodes("span.a-profile-name")
    scrapedReviewer3 <- scrapedReviewer3[1:10]


    scrapedReview3 <- scrapeNodes("span.a-size-base.review-text.review-text-content")
    scrapedReview3 <- scrapedReview3[1:10]


    scrapedDate3 <- scrapeNodes("span.a-size-base.a-color-secondary.review-date")
    scrapedDate3 <- scrapedDate3[1:10]


    scrapedprodRating3 <- scrapeNodes("span.a-icon-alt")
    scrapedprodRating3 <- scrapedprodRating3[1:10]


    product8 <- rbind(product8, data.frame(category = scrapedCategory ,name = scrapedprodName,
                        reviewer = scrapedReviewer3,
                        review = scrapedReview3,
                        date = scrapedDate3,
                       ratings = scrapedprodRating3))
    Sys.sleep(5)

#View(product8)
write.csv(product8,"product8.csv")
```

product9

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

product9 = data.frame()
scrapedCategory <- "Pencil"


###

  link = "https://www.amazon.co.uk/Ecolutions-Evolution-655%C3%82-HB-Pencil-Eraser/product-reviews/B000
```

```
      sessionProd1 <- bow(link,
               user_agent = "For Educational Purpose")

      scrapeNodes <- function(selector){
      scrape(sessionProd1) %>%
        html_nodes(selector) %>%
        html_text(trim = TRUE)
      }

      scrapedprodName <- scrapeNodes("h1.a-size-large.a-text-ellipsis")


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


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


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


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


      product9 <- rbind(product9, data.frame(category = scrapedCategory,name = scrapedprodName,
                          reviewer = scrapedReviewer,
                          review = scrapedReview,
                          date = scrapedDate,
                         ratings = scrapedprodRating))
      Sys.sleep(5)
######

      link2 = "https://www.amazon.co.uk/Ecolutions-Evolution-655%C3%82-HB-Pencil-Eraser/product-reviews,

  sessionProd2 <- bow(link2,
               user_agent = "For Educational Purpose")

      scrapedReviewer2 <- scrapeNodes("span.a-profile-name")
      scrapedReviewer2 <- scrapedReviewer2[1:10]


      scrapedReview2 <- scrapeNodes("span.a-size-base.review-text.review-text-content")
      scrapedReview2 <- scrapedReview2[1:10]


      scrapedDate2 <- scrapeNodes("span.a-size-base.a-color-secondary.review-date")
      scrapedDate2 <- scrapedDate2[1:10]
```

```
      scrapedprodRating2 <- scrapeNodes("span.a-icon-alt")
      scrapedprodRating2 <- scrapedprodRating2[1:10]


      product9 <- rbind(product9, data.frame(category = scrapedCategory ,name = scrapedprodName,
                          reviewer = scrapedReviewer2,
                          review = scrapedReview2,
                          date = scrapedDate2,
                         ratings = scrapedprodRating2))

      Sys.sleep(5)
      ####

      link3 = "https://www.amazon.co.uk/Ecolutions-Evolution-655%C3%82-HB-Pencil-Eraser/product-reviews/

  sessionProd3 <- bow(link3,
              user_agent = "For Educational Purpose")


      scrapedReviewer3 <- scrapeNodes("span.a-profile-name")
      scrapedReviewer3 <- scrapedReviewer3[1:10]


      scrapedReview3 <- scrapeNodes("span.a-size-base.review-text.review-text-content")
      scrapedReview3 <- scrapedReview3[1:10]


      scrapedDate3 <- scrapeNodes("span.a-size-base.a-color-secondary.review-date")
      scrapedDate3 <- scrapedDate3[1:10]


      scrapedprodRating3 <- scrapeNodes("span.a-icon-alt")
      scrapedprodRating3 <- scrapedprodRating3[1:10]


      product9 <- rbind(product9, data.frame(category = scrapedCategory ,name = scrapedprodName,
                          reviewer = scrapedReviewer3,
                          review = scrapedReview3,
                          date = scrapedDate3,
                         ratings = scrapedprodRating3))

      Sys.sleep(5)
#View(product9)
write.csv(product9,"product9.csv")
```

product10

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

product10 = data.frame()
```

```
scrapedCategory <- "Pencil"


###

  link = "https://www.amazon.co.uk/BIC-Evolution-Original-Graphite-Pencils/product-reviews/B003V8BIFE/r

      sessionProd1 <- bow(link,
                user_agent = "For Educational Purpose")

      scrapeNodes <- function(selector){
      scrape(sessionProd1) %>%
        html_nodes(selector) %>%
        html_text(trim = TRUE)
      }

      scrapedprodName <- scrapeNodes("h1.a-size-large.a-text-ellipsis")


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


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


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


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


      product10 <- rbind(product10, data.frame(category = scrapedCategory,name = scrapedprodName,
                           reviewer = scrapedReviewer,
                           review = scrapedReview,
                           date = scrapedDate,
                          ratings = scrapedprodRating))

      Sys.sleep(5)
######

      link2 = "https://www.amazon.co.uk/BIC-Evolution-Original-Graphite-Pencils/product-reviews/B003V8B

  sessionProd2 <- bow(link2,
                user_agent = "For Educational Purpose")

      scrapedReviewer2 <- scrapeNodes("span.a-profile-name")
      scrapedReviewer2 <- scrapedReviewer2[1:10]
```

```
      scrapedReview2 <- scrapeNodes("span.a-size-base.review-text.review-text-content")
      scrapedReview2 <- scrapedReview2[1:10]


      scrapedDate2 <- scrapeNodes("span.a-size-base.a-color-secondary.review-date")
      scrapedDate2 <- scrapedDate2[1:10]


      scrapedprodRating2 <- scrapeNodes("span.a-icon-alt")
      scrapedprodRating2 <- scrapedprodRating2[1:10]


      product10 <- rbind(product10, data.frame(category = scrapedCategory ,name = scrapedprodName,
                          reviewer = scrapedReviewer2,
                          review = scrapedReview2,
                          date = scrapedDate2,
                         ratings = scrapedprodRating2))


      Sys.sleep(5)
      ####

      link3 = "https://www.amazon.co.uk/BIC-Evolution-Original-Graphite-Pencils/product-reviews/B003V8B

  sessionProd3 <- bow(link3,
                user_agent = "For Educational Purpose")


      scrapedReviewer3 <- scrapeNodes("span.a-profile-name")
      scrapedReviewer3 <- scrapedReviewer3[1:10]


      scrapedReview3 <- scrapeNodes("span.a-size-base.review-text.review-text-content")
      scrapedReview3 <- scrapedReview3[1:10]


      scrapedDate3 <- scrapeNodes("span.a-size-base.a-color-secondary.review-date")
      scrapedDate3 <- scrapedDate3[1:10]


      scrapedprodRating3 <- scrapeNodes("span.a-icon-alt")
      scrapedprodRating3 <- scrapedprodRating3[1:10]


      product10 <- rbind(product10, data.frame(category = scrapedCategory ,name = scrapedprodName,
                          reviewer = scrapedReviewer3,
                          review = scrapedReview3,
                          date = scrapedDate3,
                         ratings = scrapedprodRating3))

      Sys.sleep(5)
#View(product10)
write.csv(product10,"product10.csv")
```

```r
all_reviews <- rbind(product1, product2, product3, product4, product5, product6, product7, product8, pr

write.csv(all_reviews,file = "all_reviews.csv")
```

```r
#install.packages("readr")
library(readr)
```

```
##
## Attaching package: 'readr'
```

```
## The following object is masked from 'package:rvest':
##
##     guess_encoding
```

```r
allrevs <- read.csv("all_reviews.csv")
tail(allrevs)
```

```
##       X category
## 295 295   Pencil
## 296 296   Pencil
## 297 297   Pencil
## 298 298   Pencil
## 299 299   Pencil
## 300 300   Pencil
##
## 295 BIC Evolution Original Pencils - Pack of 10 - All-Purpose, Extra- Resistant Erasable Pencils for
## 296 BIC Evolution Original Pencils - Pack of 10 - All-Purpose, Extra- Resistant Erasable Pencils for
## 297 BIC Evolution Original Pencils - Pack of 10 - All-Purpose, Extra- Resistant Erasable Pencils for
## 298 BIC Evolution Original Pencils - Pack of 10 - All-Purpose, Extra- Resistant Erasable Pencils for
## 299 BIC Evolution Original Pencils - Pack of 10 - All-Purpose, Extra- Resistant Erasable Pencils for
## 300 BIC Evolution Original Pencils - Pack of 10 - All-Purpose, Extra- Resistant Erasable Pencils for
##     reviewer
## 295    Dylan
## 296      Kir
## 297     Fred
## 298   PillaF
## 299     Kyle
## 300  kevin w
##
## 295                                                 They don't break, they have a nice look and they
## 296
## 297 The wood used for the pencils are nice and soft and unfortunately so is the graphite, therefore
## 298
## 299
## 300                                      They are good value for money, but I do not like the fact they
##                                                    date            ratings
## 295     Reviewed in the United Kingdom on 15 March 2024 5.0 out of 5 stars
## 296   Reviewed in the United Kingdom on 26 October 2023 5.0 out of 5 stars
## 297 Reviewed in the United Kingdom on 14 September 2023 4.0 out of 5 stars
## 298    Reviewed in the United Kingdom on 9 October 2023 5.0 out of 5 stars
## 299      Reviewed in the United Kingdom on 10 July 2023 5.0 out of 5 stars
## 300    Reviewed in the United Kingdom on 15 August 2023 3.0 out of 5 stars
```