

Group 9

Andrey Sumadic, Aeron Jesse Edombingo, Ariane Soldevilla

2023-12-20

```
#install.packages("rvest")
library(rvest)
```

```
## Warning: package 'rvest' was built under R version 4.3.2
```

```
url <- "https://www.amazon.com/dp/BOCHWRXH8B/ref=va_live_carousel?pf_rd_r=V3MNET3SVF8MNH31WGD1&pf_rd_p="
url
```

```
## [1] "https://www.amazon.com/dp/BOCHWRXH8B/ref=va_live_carousel?pf_rd_r=V3MNET3SVF8MNH31WGD1&pf_rd_p="
```

```
scrape <- read_html(url)
```

```
scrape_amazon_reviews <- function(url, max_reviews = 300) {
  scrape <- read_html(url)

  username <- scrape %>%
    html_nodes(".a-profile-name") %>%
    html_text()

  ratings <- scrape %>%
    html_nodes(".review-rating") %>%
    html_text()

  reviews <- scrape %>%
    html_nodes(".review-text-content span") %>%
    html_text()

  # Find the minimum length among the vectors
  min_length <- min(length(username), length(ratings), length(reviews))

  # Create a data frame only with complete rows
  data <- data.frame(
    Username = username[1:min_length],
    Ratings = ratings[1:min_length],
    Reviews = reviews[1:min_length]
  )

  return(data)
}
```

```
reviews_data <- scrape_amazon_reviews(url, max_reviews = 300)
reviews_data
```

```
##                               Username           Ratings
## 1 AirPods Pro (2nd Gen) with MagSafe charging case (USB-C) 5.0 out of 5 stars
## 2                               Amazon Customer 5.0 out of 5 stars
## 3                               HJeffK 5.0 out of 5 stars
## 4                               HJeffK 5.0 out of 5 stars
## 5                               Lexi 5.0 out of 5 stars
## 6                               Harmanjeetsingh 5.0 out of 5 stars
## 7                               Simon 4.0 out of 5 stars
##
## 1 I am a rare duck in the tech space as I really don't have a preference between Android and iOS. I l
## 2
## 3
## 4
## 5
## 6
## 7
```