RWorksheet#5

2023-12-01

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
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(polite)
library(xml2)
library(magrittr)
library(rvest)
library(httr)
```

```
polite::use_manners(save_as = "polite_scrape.R")
```

```
## v Setting active project to 'F:/RScraping'
```

```
url_m1 <- 'https://www.imdb.com/title/tt0436992/reviews?spoiler=hide&sort=curated&dir=desc&ratingFilter
session_m1 <- bow(url_m1,
                  user_agent = "Educational")

session_m1
```

```
## <polite session> https://www.imdb.com/title/tt0436992/reviews?spoiler=hide&sort=curated&dir=desc&rat
##     User-agent: Educational
##     robots.txt: 34 rules are defined for 2 bots
##    Crawl delay: 5 sec
##   The path is scrapable for this user-agent
```

```r
reviewerName_m1 <- character(0)
dateReviewed_m1 <- character(0)
userRating_m1 <- character(0)
titleReview_m1 <- character(0)
textReview_m1 <- character(0)

movies_m1 <- scrape(session_m1) %>%
  html_elements('div.lister-item')


reviewerName_m1 <- scrape(session_m1) %>%
  html_nodes('span.display-name-link') %>%
  html_text()



dateReviewed_m1 <- scrape(session_m1) %>%
  html_nodes('span.review-date') %>%
  html_text()


userRating_m1 <- movies_m1 %>%
  html_element(".rating-other-user-rating") %>%
  html_text()


titleReview_m1 <- scrape(session_m1) %>%
  html_nodes('a.title') %>%
  html_text()

textReview_m1 <- scrape(session_m1) %>%
  html_nodes('div.text.show-more__control') %>%
  html_text()

drWhoDF_m1 <- data.frame(userRating_m1, dateReviewed_m1, reviewerName_m1, titleReview_m1, textReview_m1
colnames(drWhoDF_m1) <- c("User Rating", "Date Reviewed", "Reviewer Name", "Title Review", "Text Review
View(drWhoDF_m1)
```