Activity 2(Movie Review)

Jasper Conlu

2024-02-27

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:

```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)
library(stringr)
library(ggplot2)
library(readr)
```

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

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

```r
polite::use_manners(save_as = 'polite_scrape.R')
```

```
## v Setting active project to '/cloud/project'
```

```r
session <- bow(url = 'https://www.imdb.com/title/tt7286456/reviews/?ref_=tt_ov_rt', user_agent = "Educa

session
```

```
## <polite session> https://www.imdb.com/title/tt7286456/reviews/?ref_=tt_ov_rt
##     User-agent: Educational
##     robots.txt: 35 rules are defined for 3 bots
##     Crawl delay: 5 sec
```

```
##   The path is scrapable for this user-agent
session_scrape <- scrape(session)

J_reviews <- function(page_url) {

  page <- read_html(page_url)
  names <- page %>% html_nodes(".display-name-link") %>% html_text()
  dates <- page %>% html_nodes("span.review-date") %>% html_text()
  ratings <- page %>% html_nodes("span.rating-other-user-rating") %>% html_text()
  content_reviews <- page %>% html_nodes("a.title") %>% html_text()
  reviews <- page %>% html_nodes(".text.show-more__control") %>% html_text()

  J_df= data.frame(
    Movie_Name=c('Joker'),
    Reviewer_Name = names[1:25],
    Content_Review = content_reviews[1:25],
    Date = dates[1:25],
    User_Rating = ratings[1:25],
    Reviews = reviews[1:25]
  )
}

Joker_urls<-c('https://www.imdb.com/title/tt7286456/reviews/?ref_=tt_ov_rt',
              'https://www.imdb.com/title/tt3915174/reviews/_ajax?&paginationKey=g4xojermtizcsyab7gthhn
              'https://www.imdb.com/title/tt3915174/reviews/_ajax?&paginationKey=g4u6dermtizcsyql7svxtnl
              'https://www.imdb.com/title/tt3915174/reviews/_ajax?&paginationKey=g4v6jermtizcsyqe72th7nz
              'https://www.imdb.com/title/tt3915174/reviews/_ajax?&paginationKey=g4s6rermtizcsyqg7kxxhml
              'https://www.imdb.com/title/tt3915174/reviews/_ajax?&paginationKey=g4w6hcbsqyxdo6ih7svxfmz
              'https://www.imdb.com/title/tt3915174/reviews/_ajax?&paginationKey=g4w6lbjsqyxdo6ih7suxvmz
              'https://www.imdb.com/title/tt17351924/reviews/_ajax?&paginationKey=g4w6ncbsqyxdo6ih7wth5r
              'https://www.imdb.com/title/tt17351924/reviews/_ajax?&paginationKey=g4w6rbjsqyxdo6ih7wuhvr
              'https://www.imdb.com/title/tt17351924/reviews/_ajax?&paginationKey=g4xohcbsqyxdo6ih7wvxhr
              'https://www.imdb.com/title/tt17351924/reviews/_ajax?&paginationKey=g4xolbjsqyxdo6ih7wwxho
              'https://www.imdb.com/title/tt17351924/reviews/_ajax?&paginationKey=g4xoncbsqyxdo6ih7wxxxr
)

Joker_reviews <- lapply(Joker_urls, J_reviews)
Joker_Review <- do.call(rbind, Joker_reviews)
```