

lab_exercise#3

Mel Adry Olivo

2024-02-22

```
#install.packages("dplyr")  
#install.packages("rvest")  
#install.packages("polite")  
#install.packages("httr")
```

```
library(dplyr)
```

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

```
##
```

```
## 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)
```

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

```
library(polite)
```

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

```
library(httr)
```

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

```
reviews <- data.frame()
```

```
link <- "https://www.imdb.com/title/tt0111161/reviews?ref_=tt_urv"
```

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

```
scrapeNodes <- function(selector) {
```

```

scrape(session) %>%
  html_nodes(selector) %>%
  html_text(trim = TRUE)
}

titleMovie <- scrapeNodes("a.subnav_heading")
reviewer <- scrapeNodes("span.display-name-link")[1:10]
scrapedDate <- scrapeNodes("span.review-date")[1:10]
scrapedRating <- scrapeNodes("span.rating-other-user-rating")[1:10]
scrapedReview <- scrapeNodes("div.text.show-more__control")[1:10]
scrapedTitleReview <- scrapeNodes("a.title")[1:10]

reviews <- rbind(reviews, data.frame(
  movie_title = rep(titleMovie, 10),
  reviewer_name = reviewer,
  review_date = scrapedDate,
  review_rating = scrapedRating,
  review_content = scrapedReview,
  review_title = scrapedTitleReview
))

reviews

```

```

##           movie_title      reviewer_name      review_date review_rating
## 1 The Shawshank Redemption hitchcockthelegend      24 July 2010          10/10
## 2 The Shawshank Redemption      Sleepin_Dragon 17 February 2021          10/10
## 3 The Shawshank Redemption              EyeDunno 21 November 2005          10/10
## 4 The Shawshank Redemption      alexkolokotronis 18 February 2008          10/10
## 5 The Shawshank Redemption              kaspen12 10 February 2006          10/10
## 6 The Shawshank Redemption              auuwws   8 October 2020          10/10
## 7 The Shawshank Redemption              Coxer99   25 March 1999          10/10
## 8 The Shawshank Redemption              nowego    15 June 2018          10/10
## 9 The Shawshank Redemption              baumer    13 July 1999          10/10
## 10 The Shawshank Redemption TheLittleSongbird    17 April 2009          10/10
##
## 1
## 2
## 3
## 4
## 5 In its Oscar year, Shawshank Redemption (written and directed by Frank Darabont, after the novell
## 6
## 7
## 8
## 9
## 10
##
##           review_title
## 1           Some birds aren't meant to be caged.
## 2      An incredible movie. One that lives with you.
## 3           Don't Rent Shawshank.
## 4           This is How Movies Should Be Made
## 5      A classic piece of unforgettable film-making.
## 6           the shawshank redemption
## 7      The Shawshank Redemption

```

```
## 8                                Eternal Hope
## 9                                Stephen King's best adapted movie
## 10 Enthralling, fantastic, intriguing, truly remarkable!
```

CREATING CONNECTIONS

```
library(dplyr,dbplyr)
library(RMySQL)
```

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

```
## Loading required package: DBI
```

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

```
#creating connections
connection <- dbConnect(RMySQL::MySQL(),
                        dsn="MariaDB-connection",
                        Server = "localhost",
                        dbname = "olivo_2c",
                        user = "root",
                        password = "")
```

LISTING TABLES AND FIELDS

```
dbListTables(connection)
```

```
## [1] "reviews"
```

```
dbListFields(connection,"reviews")
```

```
## [1] "id"           "movie_title"  "reviewer_name" "review_date"
## [5] "review_rating" "review_content" "review_title"
```

READING DATA FROM DATABASE

```
review_data <- dbGetQuery(connection, "SELECT * FROM olivo_2c.reviews")
glimpse(review_data)
```

```
## Rows: 20
## Columns: 7
## $ id          <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, ~
## $ movie_title <chr> "The Shawshank Redemption", "The Shawshank Redemption",~
## $ reviewer_name <chr> "hitchcockthelegend", "Sleepin_Dragon", "EyeDunno", "al~
## $ review_date  <chr> "24 July 2010", "17 February 2021", "21 November 2005",~
## $ review_rating <chr> "10/10", "10/10", "10/10", "10/10", "10/10", "10/10", "~
## $ review_content <chr> "The Shawshank Redemption is written and directed by Fr~
## $ review_title  <chr> "Some birds aren't meant to be caged.", "An incredible ~
```

INSERTING REVIEWS TO DATABASE

```
### use the gsub function to replace all single quotes in the review_content and review_title columns w
## This will ensure that any single quotes in the text are properly escaped in the SQL query.

for (i in 1:nrow(reviews)) {
  query <- paste0("INSERT INTO reviews (movie_title, reviewer_name, review_date, review_rating, review_content, review_title)
                  VALUES ('", reviews$movie_title[i], "','", reviews$reviewer_name[i], "','", reviews$review_date[i], "','", reviews$review_rating[i], "','", reviews$review_content[i], "','", reviews$review_title[i], "','")")
  # execute query
  query_result <- dbSendQuery(connection, query)

  # clear result after executing query
  dbClearResult(query_result)
}
```

READ AND VIEW INSERTED DATA

```
review_data <- dbGetQuery(connection, "SELECT * FROM olivo_2c.reviews")

# VIEW REVIEWS without primary key "id"
review_data[,c(2:7)]
```

##		movie_title	reviewer_name	review_date	review_rating
## 1	The Shawshank Redemption	hitchcockthelegend	24 July 2010	10/10	
## 2	The Shawshank Redemption	Sleepin_Dragon	17 February 2021	10/10	
## 3	The Shawshank Redemption	EyeDunno	21 November 2005	10/10	
## 4	The Shawshank Redemption	alexkolokotronis	18 February 2008	10/10	
## 5	The Shawshank Redemption	kaspen12	10 February 2006	10/10	
## 6	The Shawshank Redemption	auuwws	8 October 2020	10/10	
## 7	The Shawshank Redemption	Coxer99	25 March 1999	10/10	
## 8	The Shawshank Redemption	nowego	15 June 2018	10/10	
## 9	The Shawshank Redemption	baumer	13 July 1999	10/10	
## 10	The Shawshank Redemption	TheLittleSongbird	17 April 2009	10/10	
## 11	The Shawshank Redemption	hitchcockthelegend	24 July 2010	10/10	
## 12	The Shawshank Redemption	Sleepin_Dragon	17 February 2021	10/10	
## 13	The Shawshank Redemption	EyeDunno	21 November 2005	10/10	
## 14	The Shawshank Redemption	alexkolokotronis	18 February 2008	10/10	
## 15	The Shawshank Redemption	kaspen12	10 February 2006	10/10	
## 16	The Shawshank Redemption	auuwws	8 October 2020	10/10	
## 17	The Shawshank Redemption	Coxer99	25 March 1999	10/10	
## 18	The Shawshank Redemption	nowego	15 June 2018	10/10	
## 19	The Shawshank Redemption	baumer	13 July 1999	10/10	
## 20	The Shawshank Redemption	TheLittleSongbird	17 April 2009	10/10	
## 21	The Shawshank Redemption	hitchcockthelegend	24 July 2010	10/10	
## 22	The Shawshank Redemption	Sleepin_Dragon	17 February 2021	10/10	
## 23	The Shawshank Redemption	EyeDunno	21 November 2005	10/10	
## 24	The Shawshank Redemption	alexkolokotronis	18 February 2008	10/10	
## 25	The Shawshank Redemption	kaspen12	10 February 2006	10/10	

## 26	The Shawshank Redemption	auuwws	8 October 2020	10/10
## 27	The Shawshank Redemption	Coxer99	25 March 1999	10/10
## 28	The Shawshank Redemption	nowego	15 June 2018	10/10
## 29	The Shawshank Redemption	baumer	13 July 1999	10/10
## 30	The Shawshank Redemption	TheLittleSongbird	17 April 2009	10/10

##

1

2

3

4

5 In its Oscar year, Shawshank Redemption (written and directed by Frank Darabont, after the novella

6

7

8

9

10

11

12

13

14

15 In its Oscar year, Shawshank Redemption (written and directed by Frank Darabont, after the novella

16

17

18

19

20

21

22

23

24

25 In its Oscar year, Shawshank Redemption (written and directed by Frank Darabont, after the novella

26

27

28

29

30

##

review_title

1 Some birds aren't meant to be caged.

2 An incredible movie. One that lives with you.

3 Don't Rent Shawshank.

4 This is How Movies Should Be Made

5 A classic piece of unforgettable film-making.

6 the shawshank redemption

7 The Shawshank Redemption

8 Eternal Hope

9 Stephen King's best adapted movie

10 Enthralling, fantastic, intriguing, truly remarkable!

11 Some birds aren't meant to be caged.

12 An incredible movie. One that lives with you.

13 Don't Rent Shawshank.

14 This is How Movies Should Be Made

15 A classic piece of unforgettable film-making.

16 the shawshank redemption

17 The Shawshank Redemption

```
## 18                                Eternal Hope
## 19                        Stephen King's best adapted movie
## 20 Enthralling, fantastic, intriguing, truly remarkable!
## 21                        Some birds aren't meant to be caged.
## 22                An incredible movie. One that lives with you.
## 23                                Don't Rent Shawshank.
## 24                        This is How Movies Should Be Made
## 25                A classic piece of unforgettable film-making.
## 26                                the shawshank redemption
## 27                                The Shawshank Redemption
## 28                                Eternal Hope
## 29                        Stephen King's best adapted movie
## 30 Enthralling, fantastic, intriguing, truly remarkable!
```

```
#View(review_data[,c(2:7)])
```

```
# Export to csv
```

```
write.csv(review_data[,c(2:7)], "review_data.csv")
```

```
# Close the database connection
```

```
dbDisconnect(connection)
```

```
## [1] TRUE
```