



Web Scrapping: Movie Reviews

Pilsung Kang

School of Industrial Management Engineering

Korea University

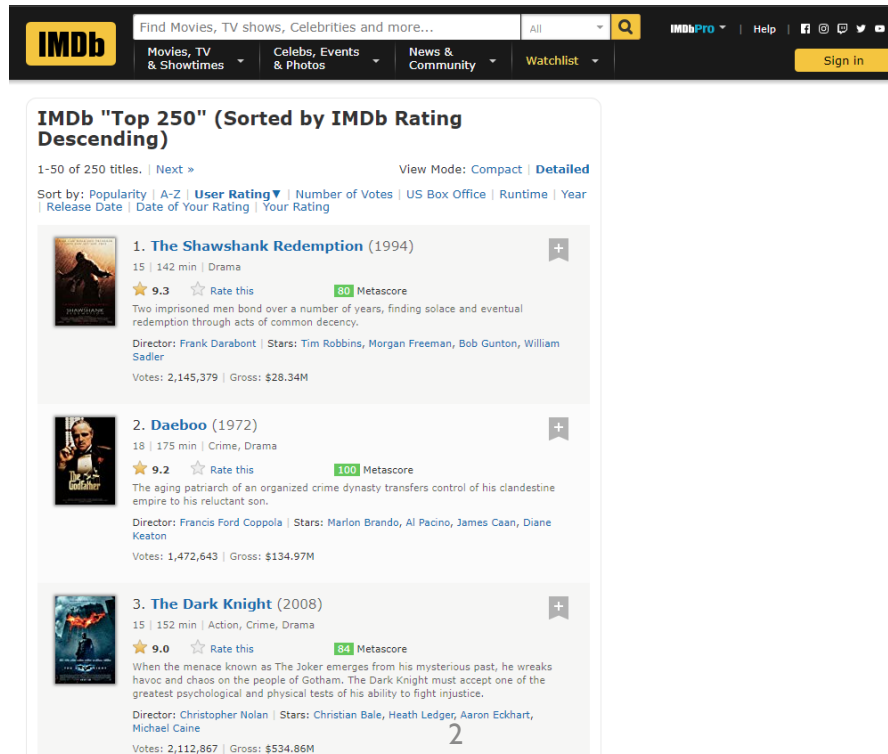
Web Scrapping: IMDB Top 50 Movie Reviews

- Web scraping example 2: Movie review scraping (IMDB Top 50 Movies)

✓ IMDB Top 250:

https://www.imdb.com/search/title/?groups=top_250&sort=user_rating

- ✓ Collect Title, Year, Average Rating, Total Number of Ratings, Summary, Director, Writer, Review Rating, Review Title, Review Text



IMDb "Top 250" (Sorted by IMDb Rating Descending)

1-50 of 250 titles. | Next » View Mode: Compact | Detailed

Sort by: Popularity | A-Z | **User Rating** | Number of Votes | US Box Office | Runtime | Year | Release Date | Date of Your Rating | Your Rating

- 1. The Shawshank Redemption** (1994)
15 | 142 min | Drama
★ 9.3 ☆ Rate this 80 Metascore
Two imprisoned men bond over a number of years, finding solace and eventual redemption through acts of common decency.
Director: Frank Darabont | Stars: Tim Robbins, Morgan Freeman, Bob Gunton, William Sadler
Votes: 2,145,379 | Gross: \$28.34M
- 2. Daeboo** (1972)
18 | 175 min | Crime, Drama
★ 9.2 ☆ Rate this 100 Metascore
The aging patriarch of an organized crime dynasty transfers control of his clandestine empire to his reluctant son.
Director: Francis Ford Coppola | Stars: Marlon Brando, Al Pacino, James Caan, Diane Keaton
Votes: 1,472,643 | Gross: \$134.97M
- 3. The Dark Knight** (2008)
15 | 152 min | Action, Crime, Drama
★ 9.0 ☆ Rate this 84 Metascore
When the menace known as The Joker emerges from his mysterious past, he wreaks havoc and chaos on the people of Gotham. The Dark Knight must accept one of the greatest psychological and physical tests of his ability to fight injustice.
Director: Christopher Nolan | Stars: Christian Bale, Heath Ledger, Aaron Eckhart, Michael Caine
Votes: 2,112,867 | Gross: \$534.86M

Web Scrapping: IMDB Top 50 Movie Reviews

- Step I: Understand the basic structure

- ✓ A total of 250 movies are listed, each page contains 50 movies

- ✓ Each movie has a unique ID

IMDb "Top 250" (Sorted by IMDb Rating Descending)

1-50 of 250 titles. | Next »

View Mode: Compact | Detailed

Sort by: Popularity | A-Z | **User Rating** | Number of Votes | US Box Office | Runtime | Year | Release Date | Date of Your Rating | Your Rating



1. **The Shawshank Redemption** (1994)

15 | 142 min | Drama

★ 9.3 ☆ Rate this 80 Metascore

Two imprisoned men bond over a number of years, finding solace and eventual redemption through acts of common decency.

Director: Frank Darabont | Stars: Tim Robbins, Morgan Freeman, Bob Gunton, William Sadler

Votes: 2,145,379 | Gross: \$28.34M



2. **Daeboo** (1972)

18 | 175 min | Crime, Drama

★ 9.2 ☆ Rate this 100 Metascore

The aging patriarch of an organized crime dynasty transfers control of his clandestine empire to his reluctant son.

Director: Francis Ford Coppola | Stars: Marlon Brando, Al Pacino, James Caan, Diane Keaton

Votes: 1,472,643 | Gross: \$134.97M



3. **The Dark Knight** (2008)

15 | 152 min | Action, Crime, Drama

★ 9.0 ☆ Rate this 84 Metascore

When the menace known as The Joker emerges from his mysterious past, he wreaks havoc and chaos on the people of Gotham. The Dark Knight must accept one of the greatest psychological and physical tests of his ability to fight injustice.

Director: Christopher Nolan | Stars: Christian Bale, Heath Ledger, Aaron Eckhart, Michael Caine

Votes: 2,112,867 | Gross: \$534.86M



4. **The Godfather: Part II** (1974)

18 | 202 min | Crime, Drama

★ 9.0 ☆ Rate this 80 Metascore

The early life and career of Vito Corleone in 1920s New York City is portrayed, while his son, Michael, expands and tightens his grip on the family crime syndicate.

Director: Francis Ford Coppola | Stars: Al Pacino, Robert De Niro, Robert Duvall, Diane Keaton

Votes: 1,024,808 | Gross: \$57.30M



5. **Joker** (2019)

15 | 122 min | Crime, Drama, Thriller

★ 8.9 ☆ Rate this 50 Metascore

In Gotham City, mentally-troubled comedian Arthur Fleck is disregarded and mistreated by society. He then embarks on a downward spiral of revolution and bloody crime. This path brings him face-to-face with his alter-ego: "The Joker".

Director: Todd Phillips | Stars: Joaquin Phoenix, Robert De Niro, Zazie Beetz, Frances Conroy

Votes: 267,033 | Gross: \$96.20M

```
▼<h3 class="list-item-header" style="user-select: auto;">  
  <span class="list-item-index unbold text-primary" style="user-select: auto;">1.</span>  
  <a href="/title/tt0111161/?ref=adv_li_tt" style="user-select: auto;">The Shawshank Redemption</a> == $0  
  <span class="list-item-year text-muted unbold" style="user-select: auto;">(1994)  
</span>  
</h3>
```

```
▼<h3 class="list-item-header" style="user-select: auto;">  
  <span class="list-item-index unbold text-primary" style="user-select: auto;">5.</span>  
  <a href="/title/tt7286456/?ref=adv_li_tt" style="user-select: auto;">Joker</a> == $0  
  <span class="list-item-year text-muted unbold" style="user-select: auto;">(I) (2019)  
</span>  
</h3>
```

Web Scrapping: IMDB Top 50 Movie Reviews

- Step 2: Get the url of each movie

```
url <- 'https://www.imdb.com/search/title/?groups=top_250&sort=user_rating'
...
tmp_list <- read_html(url) %>% html_nodes('h3.lister-item-header') %>%
  html_nodes('a[href^="/title"]') %>% html_attr('href')
```

```
<h3 class="lister-item-header" style="user-select: auto;">
  <span class="lister-item-index unbold text-primary" style="user-select: auto;">1.</span>
  <a href="/title/tt0111161/?ref=adv_li_tt" style="user-select: auto;">The Shawshank Redemption</a> == $0
  <span class="lister-item-year text-muted unbold" style="user-select: auto;">(1994)
</span>
</h3>
```

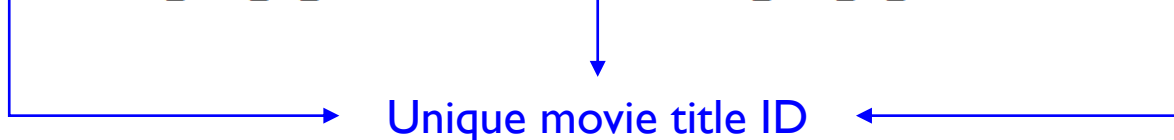
Web Scrapping: IMDB Top 50 Movie Reviews

- Step 2: Get the url of each movie

```
url <- 'https://www.imdb.com/search/title/?groups=top_250&sort=user_rating'
...
tmp_list <- read_html(url) %>% html_nodes('h3.lister-item-header') %>%
  html_nodes('a[href^="/title"]') %>% html_attr('href')
```

> tmp_list

```
[1] "/title/tt0111161/?ref=adv_li_tt" "/title/tt0068646/?ref=adv_li_tt" "/title/tt0468569/?ref=adv_li_tt"
[4] "/title/tt0071562/?ref=adv_li_tt" "/title/tt7286456/?ref=adv_li_tt" "/title/tt0167260/?ref=adv_li_tt"
[7] "/title/tt0110912/?ref=adv_li_tt" "/title/tt0108052/?ref=adv_li_tt" "/title/tt0050083/?ref=adv_li_tt"
[10] "/title/tt1375666/?ref=adv_li_tt" "/title/tt0137523/?ref=adv_li_tt" "/title/tt0120737/?ref=adv_li_tt"
[13] "/title/tt0109830/?ref=adv_li_tt" "/title/tt0060196/?ref=adv_li_tt" "/title/tt3417422/?ref=adv_li_tt"
[16] "/title/tt0167261/?ref=adv_li_tt" "/title/tt0133093/?ref=adv_li_tt" "/title/tt0099685/?ref=adv_li_tt"
[19] "/title/tt0080684/?ref=adv_li_tt" "/title/tt0073486/?ref=adv_li_tt" "/title/tt0056058/?ref=adv_li_tt"
[22] "/title/tt0816692/?ref=adv_li_tt" "/title/tt0317248/?ref=adv_li_tt" "/title/tt0245429/?ref=adv_li_tt"
[25] "/title/tt0120815/?ref=adv_li_tt" "/title/tt0120689/?ref=adv_li_tt" "/title/tt0118799/?ref=adv_li_tt"
[28] "/title/tt0114369/?ref=adv_li_tt" "/title/tt0102926/?ref=adv_li_tt" "/title/tt0076759/?ref=adv_li_tt"
[31] "/title/tt0047478/?ref=adv_li_tt" "/title/tt0038650/?ref=adv_li_tt" "/title/tt6751668/?ref=adv_li_tt"
[34] "/title/tt4154796/?ref=adv_li_tt" "/title/tt4154756/?ref=adv_li_tt" "/title/tt2582802/?ref=adv_li_tt"
[37] "/title/tt1675434/?ref=adv_li_tt" "/title/tt0482571/?ref=adv_li_tt" "/title/tt0407887/?ref=adv_li_tt"
[40] "/title/tt0253474/?ref=adv_li_tt" "/title/tt0172495/?ref=adv_li_tt" "/title/tt0120586/?ref=adv_li_tt"
[43] "/title/tt0114814/?ref=adv_li_tt" "/title/tt0110413/?ref=adv_li_tt" "/title/tt0110357/?ref=adv_li_tt"
[46] "/title/tt0103064/?ref=adv_li_tt" "/title/tt0095765/?ref=adv_li_tt" "/title/tt0095327/?ref=adv_li_tt"
[49] "/title/tt0088763/?ref=adv_li_tt" "/title/tt0064116/?ref=adv_li_tt"
```

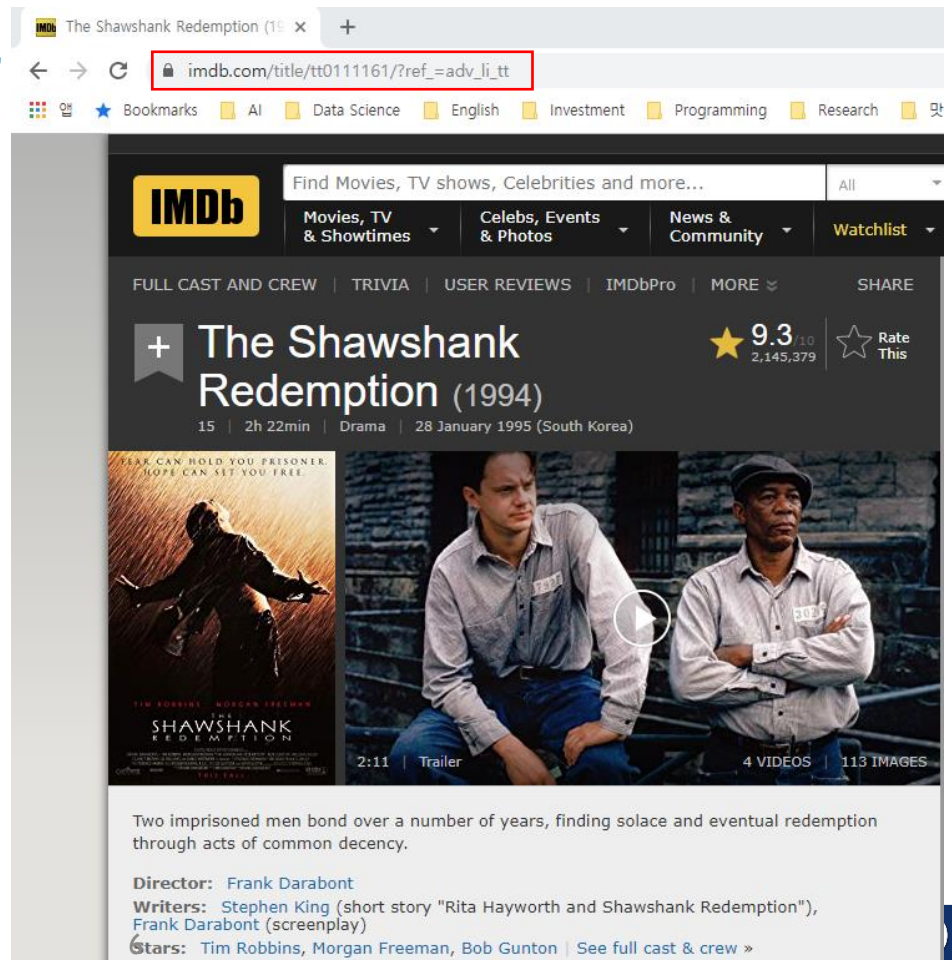


Web Scrapping: IMDB Top 50 Movie Reviews

- Meta Information on Each Movie

```
tmp_url <- paste('http://imdb.com', tmp_list[i], sep="")
tmp_content <- read_html(tmp_url)
```

```
> tmp_url
[1] "http://imdb.com/title/tt0111161/?ref=adv_li_tt"
```



The screenshot shows a web browser displaying the IMDb page for "The Shawshank Redemption (1994)". The browser's address bar shows the URL `imdb.com/title/tt0111161/?ref=adv_li_tt`, which is highlighted with a red rectangle. The IMDb page itself features the movie's title, year, and a star rating of 9.3/10. Below the title, there is a synopsis: "Two imprisoned men bond over a number of years, finding solace and eventual redemption through acts of common decency." The page also lists the director (Frank Darabont), writers (Stephen King, Frank Darabont), and stars (Tim Robbins, Morgan Freeman, Bob Gunton). The IMDb logo is visible in the top left corner of the page.

Web Scrapping: IMDB Top 50 Movie Reviews

- Title and Year

```
# Extract title and year
title_year <- tmp_content %>% html_nodes('div.title_wrapper > h1') %>% html_text %>% str_trim
tmp_title <- substr(title_year, 1, nchar(title_year)-7)
tmp_year <- substr(title_year, nchar(title_year)-4, nchar(title_year)-1)
tmp_year <- as.numeric(tmp_year)
```



The Shawshank Redemption (1994)

15 | 2h 22min | Drama | 28 January 1995 (South Korea)

9.3 10 2,145,379 Rate This

Two imprisoned men bond over a number of years, finding solace and eventual redemption through acts of common decency.

Director: Frank Darabont
Writers: Stephen King (short story "Rita Hayworth and Shawshank Redemption"), Frank Darabont (screenplay)
Stars: Tim Robbins, Morgan Freeman, Bob Gunton | See full cast & crew »

```
▼<div class="title_wrapper" style="user-select: auto;">
▼<h1 class="" style="user-select: auto;"> == $0
  "The Shawshank Redemption&nbsp;"
  ▼<span id="titleYear" style="user-select: auto;">
    "("
    <a href="/year/1994/?ref=tt_ov_inf" style="user-select: auto;">1994</a>
    ")"
  </span>
</h1>
▶<div class="subtext" style="user-select: auto;">...</div>
</div>
```

Web Scrapping: IMDB Top 50 Movie Reviews

- Title and Year

```
# Extract title and year
title_year <- tmp_content %>% html_nodes('div.title_wrapper > h1') %>% html_text %>% str_trim
tmp_title <- substr(title_year, 1, nchar(title_year)-7)
tmp_year <- substr(title_year, nchar(title_year)-4, nchar(title_year)-1)
tmp_year <- as.numeric(tmp_year)
```

```
> title_year <- tmp_content %>% html_nodes('div.title_wrapper > h1') %>% html_text %>% str_trim
> title_year
[1] "The Shawshank Redemption (1994)"
```

```
> tmp_title <- substr(title_year, 1, nchar(title_year)-7)
> tmp_title
[1] "The Shawshank Redemption"
```

```
> tmp_year <- substr(title_year, nchar(title_year)-4, nchar(title_year)-1)
> tmp_year <- as.numeric(tmp_year)
> tmp_year
[1] 1994
```


Web Scrapping: IMDB Top 50 Movie Reviews

- Average Rating

```
# Average rating
```

```
tmp_rating <- tmp_content %>% html_nodes('div.ratingValue > strong > span') %>% html_text  
tmp_rating <- as.numeric(tmp_rating)
```

The Shawshank Redemption (1994)

15 | 2h 22min | Drama | 28 January 1995 (South Korea)

9.3 2,145,379

Rate This

2:11 | Trailer

4 VIDEOS | 113 IMAGES

Two imprisoned men bond over a number of years, finding solace and eventual redemption through acts of common decency.

Director: [Frank Darabont](#)

Writers: [Stephen King](#) (short story "Rita Hayworth and Shawshank Redemption"), [Frank Darabont](#) (screenplay)

Stars: [Tim Robbins](#), [Morgan Freeman](#), [Bob Gunton](#) | [See full cast & crew »](#)

```
▼<div class="ratingValue" style="user-select: auto;">  
▼<strong title="9.3 based on 2,145,379 user ratings" style="user-select: auto;">  
  <span itemprop="ratingValue" style="user-select: auto;">9.3</span> == $0
```

Web Scrapping: IMDB Top 50 Movie Reviews

- Average Rating

```
# Average rating
tmp_rating <- tmp_content %>% html_nodes('div.ratingValue > strong > span') %>% html_text
tmp_rating <- as.numeric(tmp_rating)
```

```
> tmp_content %>% html_nodes('div.ratingValue > strong > span')
{xml_nodeset (1)}
[1] <span itemprop="ratingValue">9.3</span>
```

Web Scrapping: IMDB Top 50 Movie Reviews

- Rating Counts

```
# Rating counts
tmp_count <- tmp_content %>% html_nodes('span.small') %>% html_text
tmp_count <- gsub(",", "", tmp_count)
tmp_count <- as.numeric(tmp_count)
```

The Shawshank Redemption (1994)

15 | 2h 22min | Drama | 28 January 1995 (South Korea)

★ 9.3/10
2,145,379

Rate This

2:11 | Trailer

4 VIDEOS | 113 IMAGES

Two imprisoned men bond over a number of years, finding solace and eventual redemption through acts of common decency.

Director: Frank Darabont

Writers: Stephen King (short story "Rita Hayworth and Shawshank Redemption"), Frank Darabont (screenplay)

Stars: Tim Robbins, Morgan Freeman, Bob Gunton | See full cast & crew »

```
<span class="small" itemprop="ratingCount" style="user-select: auto;">2,145,379</span> == $0
```

Web Scrapping: IMDB Top 50 Movie Reviews

- Rating Counts

```
# Rating counts
tmp_count <- tmp_content %>% html_nodes('span.small') %>% html_text
tmp_count <- gsub(",", "", tmp_count)
tmp_count <- as.numeric(tmp_count)
```

```
> tmp_count <- tmp_content %>% html_nodes('span.small') %>% html_text
> tmp_count
[1] "2,145,379"
> tmp_count <- gsub(",", "", tmp_count)
> tmp_count
[1] "2145379"
> tmp_count <- as.numeric(tmp_count)
> tmp_count
[1] 2145379
```

Web Scrapping: IMDB Top 50 Movie Reviews

- Summary

```
# Summary tmp_summary <- tmp_content %>% html_nodes('div.summary_text') %>%  
  html_text %>% str_trim
```



```
▼<div class="summary_text" style="user-select:  
auto;"> == $0
```

```
"  
                Two imprisoned men bond  
over a number of years, finding solace and  
eventual redemption through acts of common  
decency.  
"
```

```
</div>
```

```
> tmp_summary
```

```
[1] "Two imprisoned men bond over a number of ye  
ars, finding solace and eventual redemption thro  
ugh acts of common decency."
```


Web Scrapping: IMDB Top 50 Movie Reviews

- Director, Writers, and Stars

The Shawshank Redemption (1994)
15 | 2h 22min | Drama | 28 January 1995 (South Korea)
9.3 2,145,379
Rate This

2:11 | Trailer
4 VIDEOS | 113 IMAGES

Two imprisoned men bond over a number of years, finding solace and eventual redemption through acts of common decency.

Director: Frank Darabont

Writers: Stephen King (short story "Rita Hayworth and Shawshank Redemption"), Frank Darabont (screenplay)

Stars: Tim Robbins, Morgan Freeman, Bob Gunton See full cast & crew »

```
▼<div class="credit_summary_item" style="user-select: auto;">
  <h4 class="inline" style="user-select: auto;">
    Director:</h4>
    <a href="/name/nm0001104/?ref=tt_ov_dr" style="
      "user-select: auto;">Frank Darabont</a> == $0
  </div>
▼<div class="credit_summary_item" style="user-select: auto;">
  <h4 class="inline" style="user-select: auto;">
    Writers:</h4>
    <a href="/name/nm0000175/?ref=tt_ov_wr" style="
      "user-select: auto;">Stephen King</a>
    " (short story "Rita Hayworth and Shawshank
    Redemption"), "
    <a href="/name/nm0001104/?ref=tt_ov_wr" style="
      "user-select: auto;">Frank Darabont</a>
    " (screenplay) "
  </div>
▼<div class="credit_summary_item" style="user-select: auto;">
  <h4 class="inline" style="user-select: auto;">
    Stars:</h4>
    <a href="/name/nm0000209/?ref=tt_ov_st_sm"
      style="user-select: auto;">Tim Robbins</a>
    " , "
    <a href="/name/nm0000151/?ref=tt_ov_st_sm"
      style="user-select: auto;">Morgan Freeman</a>
    " , "
    <a href="/name/nm0348409/?ref=tt_ov_st_sm"
      style="user-select: auto;">Bob Gunton</a>
    <span class="ghost" style="user-select: auto;">
    |</span>
    <a href="/fullcredits/?ref=tt_ov_st_sm" style="
      "user-select: auto;">See full cast & crew</a>
    "&nbsp;"
  </div>
</div>
```


Web Scraping: IMDB Top 50 Movie Reviews

- Director, Writers, and Stars

```
tmp_dws <- tmp_content %>% html_nodes('div.credit_summary_item') %>% html_text
tmp_director <- tmp_dws[1] %>% str_trim
tmp_director <- sub("Director:\n", "", tmp_director)

tmp_writer <- tmp_dws[2] %>% str_trim
tmp_writer <- sub("Writers:\n", "", tmp_writer)

tmp_stars <- tmp_dws[3] %>% str_trim
tmp_stars <- strsplit(tmp_stars, "\nSee")[[1]][1]
tmp_stars <- sub("Stars:\n", "", tmp_stars)
tmp_stars <- substr(tmp_stars, 1, nchar(tmp_stars)-1) %>% str_trim
```

```
> tmp_dws
```

```
[1] "\n          Director:\nFrank Darabont      "
```

```
[2] "\n          Writers:\nStephen King (short story \"Rita Hayworth and Shawshank Redempt  
ion\"), Frank Darabont (screenplay)      "
```

```
[3] "\n          Stars:\nTim Robbins, Morgan Freeman, Bob Gunton      |\nSee full ca  
st & crew »\n          "
```

Web Scraping: IMDB Top 50 Movie Reviews

- Director, Writers, and Stars

- ✓ Director

```
> tmp_director <- tmp_dws[1] %>% str_trim
> tmp_director
[1] "Director:\nFrank Darabont"
> tmp_director <- sub("Director:\n", "", tmp_director)
> tmp_director
[1] "Frank Darabont"
```

- ✓ Writers

```
> tmp_writer <- tmp_dws[2] %>% str_trim
> tmp_writer
[1] "Writers:\nStephen King (short story \"Rita Hayworth and Shawshank Redemption\"), Frank Darabont (screenplay)"
> tmp_writer <- sub("Writers:\n", "", tmp_writer)
> tmp_writer
[1] "Stephen King (short story \"Rita Hayworth and Shawshank Redemption\"), Frank Darabont (screenplay)"
```

Web Scraping: IMDB Top 50 Movie Reviews

- Director, Writers, and Stars

- ✓ Stars

```
> tmp_stars <- tmp_dws[3] %>% str_trim
> tmp_stars
[1] "Stars:\nTim Robbins, Morgan Freeman, Bob Gunton          |\nSee full cast & crew
»"
> tmp_stars <- strsplit(tmp_stars, "\nSee")[[1]][1]
> tmp_stars
[1] "Stars:\nTim Robbins, Morgan Freeman, Bob Gunton          |"
> tmp_stars <- sub("Stars:\n", "", tmp_stars)
> tmp_stars
[1] "Tim Robbins, Morgan Freeman, Bob Gunton          |"
> tmp_stars <- substr(tmp_stars, 1, nchar(tmp_stars)-1) %>% str_trim
> tmp_stars
[1] "Tim Robbins, Morgan Freeman, Bob Gunton"
```

Web Scraping: IMDB Top 50 Movie Reviews

- Scrap the First 25 Reviews for Each Movie

```
# Extract the first 25 reviews
title_id <- strsplit(tmp_list[i], "/")[[1]][3]
review_url <- paste("https://www.imdb.com/title/", title_id, "/reviews?ref_=tt_urv", sep="")

tmp_review <- read_html(review_url) %>% html_nodes('div.review-container')
```

```
> title_id
[1] "tt0111161"
> review_url <- paste("https://www.imdb.com/title/", title_id, "/reviews?ref_=tt_urv", sep="")
> review_url
[1] "https://www.imdb.com/title/tt0111161/reviews?ref_=tt_urv"
```


Web Scrapping: IMDB Top 50 Movie Reviews

- (Note): To skip unexpected errors, use tryCatch function
 - ✓ Do the instruction inside the tryCatch
 - ✓ If there is an error, store NULL to the title

```
tryCatch({  
  # Review rating  
  tmp_info <- tmp_review[j] %>% html_nodes('span.rating-other-user-rating > span') %>% html_text  
  tmp_review_rating <- as.numeric(tmp_info[1])  
  
  # Review title  
  tmp_review_title <- tmp_review[j] %>% html_nodes('a.title') %>% html_text  
  tmp_review_title <- tmp_review_title %>% str_trim  
  
  # Review text  
  tmp_review_text <- tmp_review[j] %>% html_nodes('div.text.show-more__control') %>% html_text  
  tmp_review_text <- gsub("\\s+", " ", tmp_review_text)  
  tmp_review_text <- gsub("\"", "", tmp_review_text) %>% str_trim  
  
  # Store the results  
  imdb_top_50[cnt,1] <- tmp_title  
  imdb_top_50[cnt,2] <- tmp_year  
  imdb_top_50[cnt,3] <- tmp_rating  
  imdb_top_50[cnt,4] <- tmp_count  
  imdb_top_50[cnt,5] <- tmp_summary  
  imdb_top_50[cnt,6] <- tmp_director  
  imdb_top_50[cnt,7] <- tmp_writer  
  imdb_top_50[cnt,8] <- tmp_stars  
  imdb_top_50[cnt,9] <- tmp_review_rating  
  imdb_top_50[cnt,10] <- tmp_review_title  
  imdb_top_50[cnt,11] <- tmp_review_text  
  
  cnt <- cnt+1  
}, error = function(e){print("An error occurs, skip the review")})
```


Web Scrapping: IMDB Top 50 Movie Reviews

- Review Rating



The Shawshank Redemption (1994)

User Reviews

[+ Review this title](#)

6,919 Reviews



Hide Spoilers

Filter by Rating:

Show All



Sort by:

Helpfulness



10/10

Tied for the best movie I have ever seen

carflo 26 November 2003

Why do I want to write the 234th comment on The Shawshank Redemption? I am sure - almost everything that could be possibly said about it has been said. But like so many other people who wrote comments, I was and am profoundly moved by this simple and eloquent depiction of hope and friendship and redemption.

The only other movie I have ever seen that effects me as strongly is To Kill a Mockingbird. Both movies leave me feeling cleaner for having watched them.

I didn't intend to see this movie at all: I do not like prison movies and I don't normally watch them. I work at a branch library and one day as I was checking The Shawshank Redemption out to one of our older patrons, she said to me, "Whenever I feel down or

3,007 out of 3,399 found this helpful. Was this review helpful? [Sign in to vote.](#)

[Permalink](#)

```
▼<span class="rating-other-user-rating"
style="user-select: auto;">
  ▶<svg class="ipl-icon ipl-star-icon"
xmlns="http://www.w3.org/2000/svg" fill=
"#000000" height="24" viewBox="0 0 24 24"
width="24" style="user-select: auto;">...
</svg>
  <span style="user-select: auto;">10</span> == $0
  <span class="point-scale" style="user-
select: auto;">/10</span>
</span>
```

Web Scrapping: IMDB Top 50 Movie Reviews

- Review Rating

```
# Review rating
tmp_info <- tmp_review[j] %>% html_nodes('span.rating-other-user-rating > span') %>%
  html_text
tmp_review_rating <- as.numeric(tmp_info[1])
```

```
> tmp_info <- tmp_review[j] %>% html_nodes('span.rating-other-user-rating > span') %>% html_text
> tmp_info
[1] "10"  "/10"
> tmp_review_rating <- as.numeric(tmp_info[1])
> tmp_review_rating
[1] 10
```

Web Scrapping: IMDB Top 50 Movie Reviews

- Review Title



The Shawshank Redemption (1994)

User Reviews

[+ Review this title](#)

6,919 Reviews

☐ Hide Spoilers Filter by Rating: [Show All](#) Sort by: [Helpfulness](#) [↓](#) [↑](#)

★ 10/10

Tied for the best movie I have ever seen

[carlio](#) 26 November 2003

Why do I want to write the 234th comment on The Shawshank Redemption? I am not sure - almost everything that could be possibly said about it has been said. But like so many other people who wrote comments, I was and am profoundly moved by this simple and eloquent depiction of hope and friendship and redemption.

The only other movie I have ever seen that effects me as strongly is To Kill a Mockingbird. Both movies leave me feeling cleaner for having watched them.

I didn't intend to see this movie at all: I do not like prison movies and I don't normally watch them. I work at a branch library and one day as I was checking The Shawshank Redemption out to one of our older patrons, she said to me, "Whenever I feel down or

3,007 out of 3,399 found this helpful. Was this review helpful? [Sign in to vote.](#)

[Permalink](#)

```
<a href="/review/rw0349418/?ref_=tt_urv"
class="title" style="user-select: auto;">
Tied for the best movie I have ever seen
</a> == $0
```

Web Scrapping: IMDB Top 50 Movie Reviews

- Review Title

```
# Review title
tmp_review_title <- tmp_review[j] %>% html_nodes('a.title') %>% html_text
tmp_review_title <- tmp_review_title %>% str_trim
```

```
> tmp_review_title <- tmp_review[j] %>% html_nodes('a.title') %>% html_text
> tmp_review_title
[1] " Tied for the best movie I have ever seen\n"
> tmp_review_title <- tmp_review_title %>% str_trim
> tmp_title
[1] "The Shawshank Redemption"
```

Web Scrapping: IMDB Top 50 Movie Reviews

- Review Text



The Shawshank Redemption (1994)

User Reviews

[+ Review this title](#)

6,919 Reviews

☐ Hide Spoilers Filter by Rating: **Show All** Sort by: **Helpfulness**

★ 10/10

Tied for the best movie I have ever seen
carflo 26 November 2003

Why do I want to write the 234th comment on The Shawshank Redemption? I am not sure - almost everything that could be possibly said about it has been said. But like so many other people who wrote comments, I was and am profoundly moved by this simple and eloquent depiction of hope and friendship and redemption.

The only other movie I have ever seen that effects me as strongly is To Kill a Mockingbird. Both movies leave me feeling cleaner for having watched them.

I didn't intend to see this movie at all: I do not like prison movies and I don't normally watch them. I work at a branch library and one day as I was checking The Shawshank Redemption out to one of our older patrons, she said to me, "Whenever I feel down or depressed, I check out this movie and watch it and it always makes me feel better." At the time, I thought that was very strange. One day there was nothing on TV except things I absolutely would not watch under any circumstance or things that I had seen too many times already. I remembered what she said, so I watched it. I have watched it many many times since then and it gets better with every showing.

3,007 out of 3,399 found this helpful. Was this review helpful? [Sign in to vote.](#)

[Permalink](#)

```
<div class="text show-more__control" style="user-select: auto;"> == $0
```

"Why do I want to write the 234th comment on The Shawshank Redemption? I am not sure - almost everything that could be possibly said about it has been said. But like so many other people who wrote comments, I was and am profoundly moved by this simple and eloquent depiction of hope and friendship and redemption. "

```
<br style="user-select: auto;">
```

```
<br style="user-select: auto;">
```

"The only other movie I have ever seen that effects me as strongly is To Kill a Mockingbird. Both movies leave me feeling cleaner for having watched them."

```
<br style="user-select: auto;">
```

```
<br style="user-select: auto;">
```

"I didn't intend to see this movie at all: I do not like prison movies and I don't normally watch them. I work at a branch library and one day as I was checking The Shawshank Redemption out to one of our older patrons, she said to me, "Whenever I feel down or depressed, I check out this movie and watch it and it always makes me feel better." At the time, I thought that was very strange. One day there was nothing on TV except things I absolutely would not watch under any circumstance or things that I had seen too many times already. I remembered what she said, so I watched it. I have watched it many many times since then and it gets better with every showing."

```
<br style="user-select: auto;">
```

```
<br style="user-select: auto;">
```

"No action, no special effects - just men in prison uniforms talking to each other."

Web Scrapping: IMDB Top 50 Movie Reviews

- Review Text

```
# Review text
tmp_review_text <- tmp_review[j] %>% html_nodes('div.text.show-more__control') %>% html_text
tmp_review_text <- gsub("\\s+", " ", tmp_review_text)
tmp_review_text <- gsub("\"", "", tmp_review_text) %>% str_trim
```

```
> tmp_review_text <- tmp_review[j] %>% html_nodes('div.text.show-more__control') %>% html_text
> tmp_review_text <- gsub("\\s+", " ", tmp_review_text)
> tmp_review_text <- gsub("\"", "", tmp_review_text) %>% str_trim
> tmp_review_text
```

```
[1] "Why do I want to write the 234th comment on The Shawshank Redemption? I am not sure - almost everything that could be possibly said about it has been said. But like so many other people who wrote comments, I was a nd am profoundly moved by this simple and eloquent depiction of hope and friendship and redemption. The only other movie I have ever seen that effects me as strongly is To Kill a Mockingbird. Both movies leave me feeling cleaner for having watched them.I didn't intend to see this movie at all: I do not like prison movies and I don't normally watch them. I work at a branch library and one day as I was checking The Shawshank Redemptio n out to one of our older patrons, she said to me, Whenever I feel down or depressed, I check out this movie and watch it and it always makes me feel better. At the time, I thought that was very strange. One day there was nothing on TV except things I absolutely would not watch under any circumstance or things that I had seen too many times alr... <truncated>
```


Web Scrapping: IMDB Top 50 Movie Reviews

- Store the Results

```
# Store the results
imdb_top_50[cnt,1] <- tmp_title
imdb_top_50[cnt,2] <- tmp_year
imdb_top_50[cnt,3] <- tmp_rating
imdb_top_50[cnt,4] <- tmp_count
imdb_top_50[cnt,5] <- tmp_summary
imdb_top_50[cnt,6] <- tmp_director
imdb_top_50[cnt,7] <- tmp_writer
imdb_top_50[cnt,8] <- tmp_stars
imdb_top_50[cnt,9] <- tmp_review_rating
imdb_top_50[cnt,10] <- tmp_review_title
imdb_top_50[cnt,11] <- tmp_review_text
```

Web Scraping: IMDB Top 50 Movie Reviews

- Post-processing
 - ✓ Assign the column names
 - ✓ Store the result as a Rdata and csv file

```
names(imdb_top_50) <- c("Title", "Year", "Avg.Rating", "RatingCounts", "Summary", "Director",  
                        "Writer", "Stars", "Review.Rating", "Review.Title", "Review.Text")  
  
...  
# Export the result  
save(imdb_top_50, file = "imdb_top_50.RData")  
write.csv(imdb_top_50 , file = "imdb_top_50.csv")
```

