

# Mini Project 01 - IMDB web scraping

```
library(tidyverse)
library(rvest) # scrape data from internet
```

Warning message in system("timedatectl", intern = TRUE):

"running command 'timedatectl' had status 1"

Warning message:

"Failed to locate timezone database"

— Attaching packages — tidyverse 1.3.1

```
✓ ggplot2 3.3.5    ✓ purrr  0.3.4
✓ tibble  3.1.5    ✓ dplyr  1.0.7
✓ tidyr   1.1.4    ✓ stringr 1.4.0
✓ readr   2.0.2    ✓ forcats 0.5.1
```

— Conflicts — tidyverse\_conflicts()

```
✗ dplyr::filter() masks stats::filter()
✗ purrr::flatten() masks jsonlite::flatten()
✗ dplyr::lag()     masks stats::lag()
```

Attaching package: 'rvest'

```
url <- "https://www.imdb.com/search/title/?groups=top_100&sort=user_rating,desc"
```

```
print(url)
```

```
[1] "https://www.imdb.com/search/title/?groups=top_100&sort=user_rating,desc"
```

```
imdb <- read_html(url)
```

```
## movie title
titles <- imdb %>%
  html_nodes("h3.lister-item-header") %>%
  html_text2() %>%
  head(10)
```

```
ratings <- imdb %>%
  html_nodes("div.ratings-imdb-rating") %>%
  html_text2() %>%
  as.numeric()
```

```
num_votes <- imdb %>%
  html_nodes("p.sort-num_votes-visible") %>%
  html_text2()
```

```
df <- data.frame(
  title = titles,
  rating = ratings,
  num_vote = num_votes
)
df
```

A data.frame: 50 × 3

title	rating	num_vote
<chr>	<dbl>	<chr>
1. The Shawshank Redemption (1994)	9.3	Votes: 2,679,543   Gross: \$28.34M   Top 250: #1
2. The Godfather (1972)	9.2	Votes: 1,857,665   Gross: \$134.97M   Top 250: #2
3. The Dark Knight (2008)	9.0	Votes: 2,652,882   Gross: \$534.86M   Top 250: #3
4. The Lord of the Rings: The Return of the King (2003)	9.0	Votes: 1,846,579   Gross: \$377.85M   Top 250: #7
5. The Godfather Part II (1974)	9.0	Votes: 1,271,596   Gross: \$57.30M   Top 250: #4
6. Schindler's List (1993)	9.0	Votes: 1,356,010   Gross: \$96.90M   Top 250: #6
7. 12 Angry Men (1957)	9.0	Votes: 791,580   Gross: \$4.36M   Top 250: #5
8. Pulp Fiction (1994)	8.9	Votes: 2,054,948   Gross: \$107.93M   Top 250: #8
9. The Lord of the Rings: The Fellowship of the Ring (2001)	8.8	Votes: 1,875,924   Gross: \$315.54M   Top 250: #9
10. Fight Club (1999)	8.8	Votes: 2,125,092   Gross: \$37.03M   Top 250: #12
1. The Shawshank Redemption (1994)	8.8	Votes: 2,352,434   Gross: \$292.58M   Top 250: #14
2. The Godfather (1972)	8.8	Votes: 2,078,972   Gross: \$330.25M   Top 250: #11
3. The Dark Knight (2008)	8.8	Votes: 1,667,453   Gross: \$342.55M   Top 250: #13
4. The Lord of the Rings: The Return of the King (2003)	8.8	Votes: 762,819   Gross: \$6.10M   Top 250: #10
5. The Godfather Part II (1974)	8.7	Votes: 1,162,303   Gross: \$46.84M   Top 250: #17
6. Schindler's List (1993)	8.7	Votes: 1,913,532   Gross: \$171.48M   Top 250: #16
7. 12 Angry Men (1957)	8.7	Votes: 1,009,049   Gross: \$112.00M   Top 250: #18
8. Pulp Fiction (1994)	8.7	Votes: 1,293,239   Gross: \$290.48M   Top 250: #15
9. The Lord of the Rings: The Fellowship of the Ring (2001)	8.6	Votes: 463,234   Top 250: #21
10. Fight Club (1999)	8.6	Votes: 1,830,763   Gross: \$188.02M   Top 250: #26
1. The Shawshank Redemption (1994)	8.6	Votes: 1,653,028   Gross: \$100.13M   Top 250: #19
2. The Godfather (1972)	8.6	Votes: 1,302,705   Gross: \$136.80M   Top 250: #27
3. The Dark Knight (2008)	8.6	Votes: 1,365,832   Gross: \$322.74M   Top 250: #28
4. The Lord of the Rings: The Return of the King (2003)	8.6	Votes: 1,433,492   Gross: \$130.74M   Top 250: #22
5. The Godfather Part II (1974)	8.6	Votes: 1,100,490   Gross: \$204.84M   Top 250: #29
6. Schindler's List (1993)	8.6	Votes: 1,392,421   Gross: \$216.54M   Top 250: #24
7. 12 Angry Men (1957)	8.6	Votes: 758,124   Gross: \$7.56M   Top 250: #23
8. Pulp Fiction (1994)	8.6	Votes: 764,910   Gross: \$10.06M   Top 250: #31
9. The Lord of the Rings: The Fellowship of the Ring (2001)	8.6	Votes: 696,364   Gross: \$57.60M   Top 250: #25
10. Fight Club (1999)	8.6	Votes: 347,193   Gross: \$0.27M   Top 250: #20
1. The Shawshank Redemption (1994)	8.6	Votes: 57,000   Top 250: #44

1. The Shawshank Redemption (1994)	8.6	Votes: 57,998   Top 250: #44
2. The Godfather (1972)	8.5	Votes: 863,089   Gross: \$13.09M   Top 250: #42
3. The Dark Knight (2008)	8.5	Votes: 1,501,417   Gross: \$187.71M   Top 250: #37
4. The Lord of the Rings: The Return of the King (2003)	8.5	Votes: 804,234   Gross: \$53.37M   Top 250: #34
5. The Godfather Part II (1974)	8.5	Votes: 1,206,669   Gross: \$210.61M   Top 250: #30
6. Schindler's List (1993)	8.5	Votes: 1,162,175   Gross: \$19.50M   Top 250: #35
7. 12 Angry Men (1957)	8.5	Votes: 884,400   Gross: \$78.90M   Top 250: #51
8. Pulp Fiction (1994)	8.5	Votes: 1,326,463   Gross: \$132.38M   Top 250: #39
9. The Lord of the Rings: The Fellowship of the Ring (2001)	8.5	Votes: 1,334,334   Gross: \$53.09M   Top 250: #41
10. Fight Club (1999)	8.5	Votes: 1,123,943   Gross: \$6.72M   Top 250: #38
1. The Shawshank Redemption (1994)	8.5	Votes: 669,123   Gross: \$83.47M   Top 250: #53
2. The Godfather (1972)	8.5	Votes: 493,414   Gross: \$36.76M   Top 250: #49
3. The Dark Knight (2008)	8.5	Votes: 1,086,720   Gross: \$23.34M   Top 250: #40
4. The Lord of the Rings: The Return of the King (2003)	8.5	Votes: 1,059,405   Gross: \$422.78M   Top 250: #36
5. The Godfather Part II (1974)	8.5	Votes: 331,181   Gross: \$5.32M   Top 250: #48
6. Schindler's List (1993)	8.5	Votes: 860,032   Gross: \$13.18M   Top 250: #46
7. 12 Angry Men (1957)	8.5	Votes: 833,553   Gross: \$32.57M   Top 250: #33
8. Pulp Fiction (1994)	8.5	Votes: 573,480   Gross: \$1.02M   Top 250: #43
9. The Lord of the Rings: The Fellowship of the Ring (2001)	8.5	Votes: 673,604   Gross: \$32.00M   Top 250: #32
10. Fight Club (1999)	8.5	Votes: 278,970   Top 250: #45

## Mini Project 02 - Specphone Phone Database

```
library(tidyverse)
library(rvest)
```

```
url <- read_html("https://specphone.com/Samsung-Galaxy-A04.html")
```

```
att <- url %>%
  html_nodes("div.topic") %>%
  html_text2()

value <- url %>%
  html_nodes("div.detail") %>%
  html_text2()
```

```
df_samsungA04 <- data.frame(attribute = att, value = value)
df_samsungA04
```

A data.frame: 31 × 2

attribute	value
<chr>	<chr>
วันเปิดตัว	ตุลาคม 2565
วันวางจำหน่าย	ยังไม่วางจำหน่าย
ขนาด	164.40 x 76.30 x 9.10 มม.
น้ำหนัก	192 กรัม
วัสดุ	Glass front, plastic back, plastic frame
SIM	รองรับ 2 ซิมการ์ด (nano sim, nano sim)
Technology	HSPA 42.2/5.76 Mbps, LTE-A
2G	850/900/1800/1900
3G	850/900/1900/2100
4G	850/900/1900/2100/2600
5G	-
ความเร็ว	HSPA 42.2/5.76 Mbps, LTE-A
ประเภท	PLS LCD
ขนาดหน้าจอ	6.50 นิ้ว
ความละเอียด	720 x 1600 pixels
ระบบปฏิบัติการ	Android 12
ชิปประมวลผล	Spreadtrum Unisoc SC9863A 1.6 GHz
ชิปกราฟิก	PowerVR GE8322
หน่วยความจำ	3 GB
ความจุ	32 GB
Memory Card	microSD (1)
กล้องหลัก	ตัวที่ 1: 50 MP, f/1.8, (wide), AF ตัวที่ 2: 2 MP, f/2.4, (depth)
ความละเอียดวิดีโอ	1080p@30fps
กล้องหน้า	ตัวที่ 1: 5 MP, f/2.2
Bluetooth	5.0, A2DP, LE
Wi-Fi	802.11 a/b/g/n/ac, dual-b
USB	Type-C
GPS	GLONASS, GALILEO, BDS
NFC	ไม่รองรับ
ความจุ	5,000 mAh
ประเภท	Non-removable Li-Po Batt

```
## links to all samsung smartphone
samsung_url <- read_html("https://specphone.com/brand/Samsung")
```

```
links <- samsung_url %>%  
  html_nodes("li.mobile-brand-item a") %>%  
  html_attr("href")
```

```
full_links <- paste0("https://specphone.com", links)  
full_links
```

'https://specphone.com/Samsung-Galaxy-M13.html' · 'https://specphone.com/Samsung-Galaxy-A23.html' ·  
'https://specphone.com/Samsung-Galaxy-A13.html' · 'https://specphone.com/Samsung-Galaxy-M32-5G.html' ·  
'https://specphone.com/Samsung-Galaxy-A12-Nacho.html' ·  
'https://specphone.com/Samsung-Galaxy-Pocket-Neo.html' ·  
'https://specphone.com/Samsung-Galaxy-Young.html' · 'https://specphone.com/Samsung-Galaxy-J1-Mini.html' ·  
'https://specphone.com/Samsung-Galaxy-A01-Core-1-16GB.html' ·  
'https://specphone.com/Samsung-Galaxy-V-PLUS.html' · 'https://specphone.com/Samsung-Galaxy-Young-2.html' ·  
'https://specphone.com/Samsung-Galaxy-M02.html' · 'https://specphone.com/Samsung-Galaxy-A11.html' ·  
'https://specphone.com/Samsung-Galaxy-J2-Pro-2018.html' ·  
'https://specphone.com/Samsung-Galaxy-A12-2021.html' ·  
'https://specphone.com/Samsung-Galaxy-A21s-3-32GB.html' · 'https://specphone.com/Samsung-Galaxy-J5.html' ·  
'https://specphone.com/Samsung-Galaxy-J4.html' · 'https://specphone.com/Samsung-Galaxy-Core-2-Duos.html' ·  
'https://specphone.com/Samsung-Galaxy-Ace-Plus.html' · 'https://specphone.com/Samsung-Galaxy-A20.html' ·  
'https://specphone.com/Samsung-Galaxy-Chat.html' · 'https://specphone.com/Samsung-Galaxy-Gio.html' ·  
'https://specphone.com/Samsung-Galaxy-Tab-A7-Lite-LTE.html' ·  
'https://specphone.com/Samsung-Galaxy-Tab-A-10.5WIFI.html' ·  
'https://specphone.com/Samsung-Galaxy-Alpha.html' · 'https://specphone.com/Samsung-Galaxy-S3-Slim.html' ·  
'https://specphone.com/Samsung-Galaxy-S4-zoom.html' ·  
'https://specphone.com/Samsung-Galaxy-Xcover-2.html' ·  
'https://specphone.com/Samsung-Galaxy-Tab-8.9-3G-16GB.html' ·  
'https://specphone.com/Samsung-Galaxy-Tab-A8-LTE-2021.html' ·  
'https://specphone.com/Samsung-Galaxy-A8-2018.html' ·  
'https://specphone.com/Samsung-Galaxy-Tab4-8.0-wifi.html' ·  
'https://specphone.com/Samsung-Galaxy-M33-5G.html' · 'https://specphone.com/Samsung-Galaxy-A50.html' ·  
'https://specphone.com/Samsung-Galaxy-E7.html' · 'https://specphone.com/Samsung-Galaxy-S6.html' ·  
'https://specphone.com/Samsung-Galaxy-S20-FE.html' ·  
'https://specphone.com/Samsung-Galaxy-Tab-S4-WIFI.html' · 'https://specphone.com/Samsung-Galaxy-S7.html' ·  
'https://specphone.com/Samsung-Galaxy-Note-5-Exynos.html' ·  
'https://specphone.com/Samsung-Galaxy-TabPRO-12.2-LTE.html' ·  
'https://specphone.com/Samsung-Galaxy-S4-Active.html' ·  
'https://specphone.com/Samsung-Galaxy-Tab-Active-3.html' ·  
'https://specphone.com/Samsung-Galaxy-Tab-S3-9.7.html' ·  
'https://specphone.com/Samsung-Galaxy-S6-edge.html' ·  
'https://specphone.com/Samsung-Galaxy-Note-4-Exynos.html' ·  
'https://specphone.com/Samsung-Galaxy-Round.html' ·  
'https://specphone.com/Samsung-Galaxy-Note-20-Ultra-5G.html' · 'https://specphone.com/Samsung-ATIV-Q.html' ·  
'https://specphone.com/Samsung-ATIV-Smart-PC-PRO.html' ·  
'https://specphone.com/Samsung-Galaxy-S22-Ultra12-128GB.html' ·  
'https://specphone.com/Samsung-Galaxy-Z-Flip-5G.html' · 'https://specphone.com/Samsung-Galaxy-Z-Flip.html' ·  
'https://specphone.com/Samsung-Galaxy-Tab-S8-Ultra-5G.html' ·  
'https://specphone.com/Samsung-Galaxy-S21-Ultra-16-512GB.html' ·  
'https://specphone.com/Samsung-Galaxy-S10-Plus-Ram-12GB.html' ·  
'https://specphone.com/Samsung-Galaxy-Z-Fold-3.html' · 'https://specphone.com/Samsung-Galaxy-Z-Fold4.html' ·  
'https://specphone.com/Samsung-Galaxy-Z-Fold-2-5G.html'



```
result <- data.frame()

for (link in full_links[1:10]){
  ss_topic <- link %>%
    read_html() %>%
    html_nodes("div.topic") %>%
    html_text2()

  ss_detail <- link %>%
    read_html() %>%
    html_nodes("div.detail") %>%
    html_text2()

  tmp <- data.frame(attribute = ss_topic, value = ss_detail)
  result <- bind_rows(result, tmp)
}
```

```
result
write_csv(result, "result_ss_phone.csv")
```

attribute	value
<chr>	<chr>
วันเปิดตัว	มิถุนายน 2565
วันวางจำหน่าย	ยังไม่วางจำหน่าย
ขนาด	165.40 x 76.90 x 8.40 มม.
น้ำหนัก	192 กรัม
วัสดุ	Glass front, plastic back, plastic frame
SIM	รองรับ 2 ซิมการ์ด (nano sim, nano sim)
Technology	HSPA 42.2/5.76 Mbps, LTE-A
2G	850/900/1800/1900
3G	850/900/1900/2100
4G	850/900/1900/2100
5G	-
ความเร็ว	HSPA 42.2/5.76 Mbps, LTE-A
ประเภท	PLS LCD
ขนาดหน้าจอ	6.60 นิ้ว
ความละเอียด	1080 x 2408 pixels
ระบบปฏิบัติการ	Android 12
ชิปประมวลผล	Samsung Exynos 850 S5E3830 2 GHz
ชิปกราฟิก	Mali-G52 MP1
หน่วยความจำ	4 GB
ความจุ	64 GB
Memory Card	microSD (1)
กล้องหลัก	ตัวที่ 1: 50 MP, f/1.8, (wide), PDAF ตัวที่ 2: 5 MP, f/2.2, 123° (ultrawide) ตัวที่ 3: 2 MP, f/2.4, (depth)
ความละเอียดวิดีโอ	1080p@30fps
กล้องหน้า	ตัวที่ 1: 8 MP, f/2.2, (wide)
Bluetooth	5.0, A2DP, LE
Wi-Fi	802.11 a/b/g/n/ac, dual-b
USB	Type-C
GPS	A-GPS, GLONASS, GALILEO,
NFC	รองรับ
ความจุ	5,000 mAh
:	:
ขนาด	121.40 x 62.90 x 10.70 มม.
น้ำหนัก	123 กรัม
วัสดุ	Plastic
SIM	รองรับ 2 ซิมการ์ด (micro sim, micro sim)

Technology	EDGE, HSPA
2G	850/900/1800/1900
3G	900/2100
4G	-
5G	-
ความเร็ว	EDGE, HSPA
ประเภท	TFT LCD
ขนาดหน้าจอ	4.00 นิ้ว
ความละเอียด	800 x 480 pixels
ฟีเจอร์เพิ่มเติม	สามารถให้ PC ใช้อินเทอร์เน็ตของสมาร์ทโฟนได้ด้วย Internet Tethering สามารถใช้งาน Mobile Hotspot โดยแชร์อินเทอร์เน็ตบนสมาร์ทโฟนให้กับเครื่องอื่นๆแบบไร้สายได้ สามารถใช้งานเป็น USB Mass Storage ได้เพื่อโอนหรือย้ายข้อมูลบน PC ใช้อินเตอร์เฟซ TouchWiz UX มาพร้อมกับ Android 4.4 รองรับ 2 ซิม
ระบบปฏิบัติการ	Android 4.4
ชิปประมวลผล	Dual Core 0 MHz
ชิปกราฟิก	VideoCore IV
หน่วยความจำ	512 MB
ความจุ	4 GB
Memory Card	microSD (64)
กล้องหลัก	ตัวที่ 1: 3 MP
ความละเอียดวิดีโอ	480
กล้องหน้า	ตัวที่ 1: MP
Bluetooth	4
Wi-Fi	802.11b/g/n
USB	micro USB

```
print(head(result), 3)
```

	attribute	value
1	วันเปิดตัว	มิถุนายน 2565
2	วันวางจำหน่าย	ยังไม่วางจำหน่าย
3	ขนาด	165.40 x 76.90 x 8.40 มม.
4	น้ำหนัก	192 กรัม
5	วัสดุ	Glass front, plastic back, plastic frame

