### Mini Project 01 - IMDB Web Scraping

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

```
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"
```

```
# read html
imdb <- read_html(url)
```

```
titles <- imdb %>%
    html_nodes("h3.lister-item-header") %>%
    html_text2()
```

```
titles[1:10]
```

```
'1. The Shawshank Redemption (1994)' · '2. The Godfather (1972)' · '3. The Dark Knight (2008)' ·
'4. The Lord of the Rings: The Return of the King (2003)' · '5. Schindler\'s List (1993)' ·
'6. The Godfather Part II (1974)' · '7. 12 Angry Men (1957)' · '8. Pulp Fiction (1994)' · '9. Inception (2010)' ·
'10. The Lord of the Rings: The Two Towers (2002)'
```

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

```
ratings[1:10]
```

```
9.3 · 9.2 · 9 · 9 · 9 · 9 · 9 · 8.9 · 8.8 · 8.8
```

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

```
num_votes[1:10]
```

```
'Votes: 2,671,065 | Gross: $28.34M | Top 250: #1' · 'Votes: 1,851,156 | Gross: $134.97M | Top 250: #2' ·
'Votes: 2,643,876 | Gross: $534.86M | Top 250: #3' · 'Votes: 1,840,905 | Gross: $377.85M | Top 250: #7' ·
'Votes: 1,352,215 | Gross: $96.90M | Top 250: #6' · 'Votes: 1,267,569 | Gross: $57.30M | Top 250: #4' ·
'Votes: 788,904 | Gross: $4.36M | Top 250: #5' · 'Votes: 2,045,984 | Gross: $107.93M | Top 250: #8' ·
'Votes: 2,343,339 | Gross: $292.58M | Top 250: #14' · 'Votes: 1,662,284 | Gross: $342.55M | Top 250: #13'
```

```
# build a dataset
df <- data.frame(
    title = titles,
    ratings = ratings,
    num_votes = num_votes
)
head(df)
```

A data.frame: 6 × 3

| | title | ratings | num_votes |
|---|---|---|---|
| | <chr> | <dbl> | <chr> |
| 1 | 1. The Shawshank Redemption (1994) | 9.3 | Votes: 2,671,065 \| Gross: $28.34M \| Top 250: #1 |
| 2 | 2. The Godfather (1972) | 9.2 | Votes: 1,851,156 \| Gross: $134.97M \| Top 250: #2 |
| 3 | 3. The Dark Knight (2008) | 9.0 | Votes: 2,643,876 \| Gross: $534.86M \| Top 250: #3 |
| 4 | 4. The Lord of the Rings: The Return of the King (2003) | 9.0 | Votes: 1,840,905 \| Gross: $377.85M \| Top 250: #7 |
| 5 | 5. Schindler's List (1993) | 9.0 | Votes: 1,352,215 \| Gross: $96.90M \| Top 250: #6 |
| 6 | 6. The Godfather Part II (1974) | 9.0 | Votes: 1,267,569 \| Gross: $57.30M \| Top 250: #4 |

### Mini Project 02 - Specphone Phone Database

```
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 <- read_html("https://specphone.com/vivo-X90.html")
```

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

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

```
data.frame(
    att = att,
    value = value
)
```

A data.frame: 32 × 2

| att | value |
|---|---|
| <chr> | <chr> |
| วันเปิดตัว | พฤศจิกายน 2565 |
| วันวางจำหน่าย | ยังไม่วางจำหน่าย |
| ขนาด | 164.10 x 74.40 x 8.50 มม. |
| น้ำหนัก | 196 กรัม |
| วัสดุ | Glass front, glass back or eco leather back |
| SIM | รองรับ 2 ซิมการ์ด (nano sim, nano sim) |
| Technology | HSPA, LTE-A, 5G |
| 2G | 850/900/1800/1900 |
| 3G | 850/900/1900/2100 |
| 4G | 850/900/1900/2100/2600 |
| 5G | 2100/2600/3500/4700 |
| ความเร็ว | HSPA, LTE-A, 5G |
| ประเภท | AMOLED |
| ขนาดหน้าจอ | 6.78 นิ้ว |
| ความละเอียด | 1260 x 2800 pixels |
| ระบบปฏิบัติการ | Android 12 |
| ชิปประมวลผล | MediaTek Dimensity 9200 null 3.05 GHz |
| ชิปกราฟิก | Arm Immortalis-G715 |
| หน่วยความจำ | 8 GB |
| ความจุ | 128 GB |
| Memory Card | ไม่รองรับ |
| กล้องหลัก | ตัวที่ 1: 50 MP, f/1.8, (wide), 1/1.49 ตัวที่ 2: 12 MP, f/2.0, 50mm (telephoto), 1/2.93 ตัวที่ 3: 12 MP, f/2.0, 16mm (ultrawide), 1/2.93 |
| ความละเอียดวีดีโอ | 4K@30/60fps, 1080p@30/60fps, gyro-EIS |
| กล้องหน้า | ตัวที่ 1: 32 MP, f/2.5, 24mm (wide), 1/2.8 |
| Bluetooth | 5.3, A2DP, LE, aptX HD |
| Wi-Fi | 802.11 a/b/g/n/ac/6, dual |
| USB | Type-C |
| GPS | GPS (L1+L5), GLONASS, BDS |
| NFC | รอบรับ |
| ความจุ | 4,810 mAh |
| ประเภท | Non-removable Li-Po Batt |
| Fast Charging | รองรับ (120W) |

```
samsung_url <- read_html("https://specphone.com/brand/Samsung")
```

```
#link to all samsung smart phone
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[1:5]){
    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(attributes = ss_topic,
                      value = ss_detail)
    result <-  bind_rows(result,tmp)
    print("progress...")
}
print(result)
```

```
[1] "progress..."
[1] "progress..."
[1] "progress..."
[1] "progress..."
[1] "progress..."
        attributes
1          วันเปิดตัว
2       วันวางจำหน่าย
3              ขนาด
4            น้ำหนัก
5               วัสดุ
6               SIM
7        Technology
8                2G
9                3G
10               4G
11               5G
12          ความเร็ว
13            ประเภท
14        ขนาดหน้าจอ
```

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

```
   attributes                                    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
6         SIM      รองรับ 2 ซิมการ์ด (nano sim, nano sim)
```

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