IMDB MINI PROJECT #1

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

```
imdb
```

```
{html_document}
<html xmlns:og="http://ogp.me/ns#" xmlns:fb="http://www.facebook.com/2008/fbml"
[1] <head>\n<meta http-equiv="Content-Type" content="text/html; charset=UTF-8 .
[2] <body id="styleguide-v2" class="fixed">\n              <img height="1" widt .
```

```
# movie title
imdb %>%
    html_node("h3.lister-item-header") %>%
    html_text()
```

'\n   1.\n \n   The Shawshank Redemption\n   (1994)\n'

```
# movie title
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. The Godfather Part II (1974)' ·
'6. Schindler\'s List (1993)' · '7. 12 Angry Men (1957)' · '8. Pulp Fiction (1994)' ·
'9. The Lord of the Rings: The Fellowship of the Ring (2001)' · '10. Fight Club (1999)'

```
# number of votes
NOV <- imdb %>%
    html_nodes("p.sort-num_votes-visible") %>%
    html_text2()
NOV[1:10]
```

'Votes: 2,685,310 | Gross: $28.34M | Top 250: #1' · 'Votes: 1,862,105 | Gross: $134.97M | Top 250: #2' ·
'Votes: 2,658,666 | Gross: $534.86M | Top 250: #3' · 'Votes: 1,850,202 | Gross: $377.85M | Top 250: #7' ·
'Votes: 1,274,200 | Gross: $57.30M | Top 250: #4' · 'Votes: 1,358,410 | Gross: $96.90M | Top 250: #6' ·
'Votes: 793,264 | Gross: $4.36M | Top 250: #5' · 'Votes: 2,060,219 | Gross: $107.93M | Top 250: #8' ·
'Votes: 1,879,744 | Gross: $315.54M | Top 250: #9' · 'Votes: 2,130,784 | Gross: $37.03M | Top 250: #12'

```
# build a dataframe
df <- data.frame(
      title = titles,
      num_vote = NOV
)

head(df)
```

A data.frame: 6 × 2

	title	num_vote
	<chr>	<chr>
1	1. The Shawshank Redemption (1994)	Votes: 2,685,310 | Gross: $28.34M | Top 250: #1
2	2. The Godfather (1972)	Votes: 1,862,105 | Gross: $134.97M | Top 250: #2
3	3. The Dark Knight (2008)	Votes: 2,658,666 | Gross: $534.86M | Top 250: #3
4	4. The Lord of the Rings: The Return of the King (2003)	Votes: 1,850,202 | Gross: $377.85M | Top 250: #7
5	5. The Godfather Part II (1974)	Votes: 1,274,200 | Gross: $57.30M | Top 250: #4
6	6. Schindler's List (1993)	Votes: 1,358,410 | Gross: $96.90M | Top 250: #6

MINI PROJECT 02 - Specphone Phone Database

```
library(tidyverse)
library(rvest) # scrape data from internet
url<- "https://specphone.com/Motorola-Moto-X40.html" ## ฝาก url
#read url
url <- read_html(url)
```

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

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

A data.frame: 32 × 2

attribute	value
<chr>	<chr>
วันเปิดตัว	ธันวาคม 2565
วันวางจำหน่าย	ยังไม่วางจำหน่าย
ขนาด	161.20 x 74.00 x 8.60 มม.
น้ำหนัก	199 กรัม
วัสดุ	Glass front, glass back (Gorilla Glass Victus), aluminum frame
SIM	รองรับ 2 ซิมการ์ด (nano sim, nano sim)
Technology	HSPA, LTE-A (CA), 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 (CA), 5G
ประเภท	OLED
ขนาดหน้าจอ	6.70 นิ้ว
ความละเอียด	1080 x 2400 pixels
ระบบปฏิบัติการ	Android 123
ชิปประมวลผล	Qualcomm Snapdragon 8 Gen 2 SM8550 3.2 GHz
ชิปกราฟิก	Adreno 740
หน่วยความจำ	8 GB
ความจุ	128 GB
Memory Card	ไม่รองรับ
กล้องหลัก	ตัวที่ 1: 50 MP, f/1.9, (wide), 1/1.55 ตัวที่ 2: 50 MP, f/2.2, 117˚ (ultrawide), 1/2.76 ตัวที่ 3: 12 MP, f/1.6, (telephoto), 1/2.93
ความละเอียด วีดีโอ	4K@30fps, 1080p@30/60/120/240fps, 720p@960fps, 10-bit HDR10+, gyro-EIS
กล้องหน้า	ตัวที่ 1: 60 MP, f/2.2, (wide), 1/2.8
Bluetooth	5.2, A2DP, LE
Wi-Fi	802.11 a/b/g/n/ac, dual-b
USB	Type-C
GPS	A-GPS, GLONASS, GALILEO,Q
NFC	รอบรับ
ความจุ	4,600 mAh
ประเภท	Non-removable Li-Po Batt
Fast Charging	รองรับ (125W)

```
#Find all mndel on motolora
motorola_url <- read_html("https://specphone.com/brand/Motorola")
model <- motorola_url %>%
    html_nodes("li.mobile-brand-item a") %>%
    html_attr("href")
#add https:
full_links <- paste0("https://specphone.com", model)
full_links
```

'https://specphone.com/Motorola-Moto-X40.html' · 'https://specphone.com/Motorola-Moto-G72.html' ·
'https://specphone.com/Motorola-Moto-G62-5G.html' · 'https://specphone.com/Motorola-Moto-G42.html' ·
'https://specphone.com/Motorola-Moto-E32.html' · 'https://specphone.com/Motorola-Moto-C-3G.html' ·
'https://specphone.com/Motorola-Defy-Mini.html' · 'https://specphone.com/Motorola-Fire-XT.html' ·
'https://specphone.com/Motorola-Moto-E7i-Power.html' · 'https://specphone.com/Motorola-Moto-C-4G.html' ·
'https://specphone.com/Motorola-Moto-E20.html' · 'https://specphone.com/Motorola-Moto-E7.html' ·
'https://specphone.com/Motorola-Moto-M.html' · 'https://specphone.com/Motorola-Moto-G7-Power.html' ·
'https://specphone.com/Motorola-moto-e5-plus.html' · 'https://specphone.com/Motorola-Moto-G5-Plus.html' ·
'https://specphone.com/Motorola-Moto-G31.html' · 'https://specphone.com/Motorola-Moto-G30.html' ·
'https://specphone.com/Motorola-Moto-G.html' · 'https://specphone.com/Motorola-Moto-G6.html' ·
'https://specphone.com/Motorola-One-Action.html' · 'https://specphone.com/Motorola-Moto-G9-Plus.html' ·
'https://specphone.com/Motorola-Moto-G5S.html' · 'https://specphone.com/Motorola-Moto-G51-5G.html' ·
'https://specphone.com/Motorola-RAZR-V.html' · 'https://specphone.com/Motorola-Edge-20-Fusion.html' ·
'https://specphone.com/Motorola-One-Power.html' · 'https://specphone.com/Motorola-RAZR-MAXX.html' ·
'https://specphone.com/Motorola-Moto-X.html' · 'https://specphone.com/Motorola-Atrix.html' ·
'https://specphone.com/Motorola-Moto-Z2-Force.html' · 'https://specphone.com/Motorola-Moto-X-Play.html' ·
'https://specphone.com/Motorola-Edge-30.html' · 'https://specphone.com/Motorola-Moto-One-Fusion-Plus.html' ·
'https://specphone.com/Motorola-Moto-G6-Plus-6GB.html' · 'https://specphone.com/Motorola-RAZR-i.html' ·
'https://specphone.com/Motorola-Droid-RAZR-MAXX-HD.html' ·
'https://specphone.com/Motorola-Edge-30-Pro.html' · 'https://specphone.com/Motorola-Moto-Z.html' ·
'https://specphone.com/Motorola-Razr-5G.html'

```
result <- data.frame()

for (model in full_links[1:10]) {
    moto_topic <- model %>%
        read_html() %>%
        html_nodes("div.topic") %>%
        html_text2()
    moto_value <- model %>%
        read_html() %>%
        html_nodes("div.detail") %>%
        html_text2()

    tmp <- data.frame(attribute = moto_topic,
                          value = moto_value)
    result <- bind_rows(result, tmp)
    print("progress ...")
}
print(result)
```

```
[1] "progress ..."
[1] "progress ..."
[1] "progress ..."
[1] "progress ..."
[1] "progress ..."
[1] "progress ..."
[1] "progress ..."
[1] "progress ..."
[1] "progress ..."
[1] "progress ..."
          attribute
1         วันเปิดตัว
2       วันวางจำหน่าย
3             ขนาด
4           น้ำหนัก
5             วัสดุ
6             SIM
7        Technology
8              2G
9              3G
```

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

	attribute	value
1	วันเปิดตัว	ธันวาคม 2565
2	วันวางจำหน่าย	ยังไม่วางจำหน่าย
3	ขนาด	161.20 x 74.00 x 8.60 มม.
4	น้ำหนัก	199 กรัม

| 5 | วัสดุ Glass front, glass back (Gorilla Glass Victus), aluminum frame |
| 6 | SIM รองรับ 2 ซิมการ์ด (nano sim, nano sim) |

```
#write csv
write_csv(result, "result_mt_phone.csv")
```