

Final Project

Bansara,Gallo,Olivo,Quillo

2023-12-21

Select a website that you want to scrape 300 reviews.

```
library(dplyr)

##
## 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)
library(polite)
library(httr)
library(selectr)

reviews = data.frame()

for(counter in 1:12){
  looplink = paste0("https://www.imdb.com/title/tt1517268/reviews?sort=userRating&dir=desc&ratingFilter=")

  link = looplink

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

  scrapeNodes <- function(selector){
    scrape(session) %>%
      html_nodes(selector) %>%
      html_text(trim = TRUE)
  }

  rating_name <- scrapeNodes("span.display-name-link")

  rating_title <- scrapeNodes("a.title")

  rating_contents <- scrapeNodes("div.content")
}
```

```

rating_date <- scrapeNodes("span.review-date")

rating_ratings <- scrapeNodes("div.ipl-ratings-bar")

reviews <- rbind(reviews, data.frame(name = rating_name,
                                     title = rating_title,
                                     contents = rating_contents,
                                     date = rating_date,
                                     ratings = rating_ratings))
}

```

printing 100 reviews out of 300 because I cant print all 300.

```
reviews[1:100,1:5]
```

```

##              name
## 1          BA_Harrison
## 2      niallmcentegart-93295
## 3      STARONTHETOPOFTHESKY
## 4          kobalplatt
## 5  thedarkknightreturns-96915
## 6          protsenkodv
## 7          yurixander10
## 8          evisadh
## 9          d-s-allen82
## 10         chriswalls-08218
## 11         ericvankeimpema
## 12          nmottel
## 13         pensacolacomputer
## 14         amandakelly-44326
## 15          owenjmiddleton
## 16          kim_shea_berg
## 17          bdewiyah
## 18         craigfish-86029
## 19          li0904426
## 20          eva-rebac
## 21          hamedkhazaei55
## 22         mariekeherwijnen
## 23          steveecho
## 24         taylorelizabethb
## 25          fgarrido-98202
## 26         SarahAppleSauce98
## 27          KrustyKrabnz
## 28          piotr-postowicz
## 29          mayaaaaaaaaaaaa
## 30          bbosma-24403
## 31          jonathanruano
## 32          marismini
## 33          darksun3
## 34          RomeoQQ
## 35          coldinan
## 36         olgakastritsi-81359

```

```

## 37          davelake-79902
## 38    theoneandonlycrazyian
## 39          sparkysparker
## 40          javierfcontreras
## 41          brodyderek
## 42          jannochko26
## 43          jmcygbsn
## 44    alisonacostachinchilla
## 45          tcjames-74057
## 46          princepurple1979
## 47          sevigk
## 48          damian-50231
## 49          hgrill
## 50          amandasluiman
## 51    antoniatejedabarros
## 52          jacobarch
## 53          elihebdon
## 54          Azureus91
## 55          uebermut-1
## 56          rebeccaxjan
## 57          boyd-03573
## 58          enigma49
## 59          liefie2000
## 60          nicri-28979
## 61          beanbaguk
## 62          julian-heinz
## 63    nathanholt-67152
## 64          musajasna
## 65          ryfletch
## 66          softwatcher
## 67          liamjd-36600
## 68          omerus91
## 69    themurphys-11710
## 70    tharunbabu-72513
## 71    bonniepage-40421
## 72          martimusross
## 73          SureCommaNot
## 74    philippe-ripoll
## 75    vincentssheehan
## 76          sandy777
## 77          stusby
## 78          MissSimonetta
## 79    moviesgamesandvideos
## 80          max-850
## 81          ambusched94
## 82          Dello_
## 83    brianjohnson-20043
## 84          nwsquadron
## 85          Joytoyxx
## 86          olelmao
## 87          Criticalstaff
## 88          miriamkheirom
## 89          theamykeown
## 90          gianthaole

```

91 jfgibson73
92 ThereelscoopwithKK
93 otaku-90710
94 Skipfishh
95 jtkwtc
96 ngr-18639
97 asderei
98 abeburnett
99 mfgkqwcx
100 missmaddibishop

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43

Barbie Mo

"The Ma

Elegant in visu

44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97

Disappointing

I

A haphaz
Hollywood

I just w

Hardly any laughter in my theat

The m

Margot Robbie & Ryan Gosling's performances bolster a film with messaging that is heavy-handed and

Funny

I'd probabl

Bas
It v

98
99
100

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50

```

## 51
## 52
## 53
## 54
## 55
## 56
## 57
## 58
## 59
## 60
## 61
## 62
## 63
## 64
## 65
## 66
## 67
## 68
## 69
## 70
## 71
## 72
## 73
## 74
## 75
## 76
## 77
## 78
## 79
## 80
## 81
## 82 I had great expectations for the Barbie movie, mainly because of the writers (Greta Gerwig and N
## 83
## 84
## 85
## 86
## 87
## 88
## 89
## 90
## 91
## 92
## 93
## 94
## 95
## 96
## 97
## 98
## 99
## 100
##
##          date ratings
## 1  19 September 2023    1/10
## 2   13 August 2023     1/10
## 3   9 September 2023    1/10

```

## 4	12 September 2023	1/10
## 5	10 September 2023	1/10
## 6	28 August 2023	1/10
## 7	1 October 2023	1/10
## 8	20 September 2023	1/10
## 9	16 October 2023	1/10
## 10	5 August 2023	1/10
## 11	2 October 2023	1/10
## 12	16 December 2023	1/10
## 13	7 September 2023	1/10
## 14	21 July 2023	1/10
## 15	24 September 2023	1/10
## 16	23 July 2023	1/10
## 17	1 August 2023	1/10
## 18	19 July 2023	1/10
## 19	25 October 2023	1/10
## 20	9 September 2023	1/10
## 21	13 August 2023	1/10
## 22	2 November 2023	1/10
## 23	22 July 2023	1/10
## 24	22 July 2023	1/10
## 25	19 July 2023	1/10
## 26	8 October 2023	2/10
## 27	1 December 2023	2/10
## 28	31 October 2023	2/10
## 29	21 July 2023	2/10
## 30	27 November 2023	2/10
## 31	21 July 2023	2/10
## 32	16 October 2023	2/10
## 33	3 September 2023	2/10
## 34	1 October 2023	2/10
## 35	3 September 2023	2/10
## 36	22 July 2023	2/10
## 37	17 December 2023	2/10
## 38	8 November 2023	2/10
## 39	19 December 2023	2/10
## 40	25 September 2023	2/10
## 41	5 November 2023	2/10
## 42	3 September 2023	2/10
## 43	7 October 2023	2/10
## 44	26 July 2023	2/10
## 45	25 July 2023	2/10
## 46	14 October 2023	2/10
## 47	10 September 2023	2/10
## 48	3 December 2023	2/10
## 49	15 December 2023	2/10
## 50	6 December 2023	2/10
## 51	29 August 2023	3/10
## 52	23 July 2023	3/10
## 53	22 July 2023	3/10
## 54	20 July 2023	3/10
## 55	2 October 2023	3/10
## 56	27 August 2023	3/10
## 57	13 September 2023	3/10


```
## 58      20 July 2023    3/10
## 59      20 July 2023    3/10
## 60     28 October 2023   3/10
## 61      3 November 2023  3/10
## 62      7 November 2023  3/10
## 63      21 July 2023    3/10
## 64      20 July 2023    3/10
## 65      21 July 2023    3/10
## 66      23 July 2023    3/10
## 67     3 September 2023  3/10
## 68      21 July 2023    3/10
## 69      26 July 2023    3/10
## 70      1 August 2023    3/10
## 71    25 September 2023  3/10
## 72      9 October 2023   3/10
## 73      25 July 2023    3/10
## 74     14 August 2023    3/10
## 75      25 July 2023    3/10
## 76    21 November 2023   4/10
## 77     14 August 2023    4/10
## 78      20 July 2023    4/10
## 79     9 September 2023  4/10
## 80     27 August 2023    4/10
## 81    13 September 2023  4/10
## 82      25 July 2023    4/10
## 83      24 July 2023    4/10
## 84      21 July 2023    4/10
## 85      20 July 2023    4/10
## 86      21 July 2023    4/10
## 87    30 September 2023  4/10
## 88      6 September 2023  4/10
## 89     30 August 2023    4/10
## 90      23 July 2023    4/10
## 91    25 September 2023  4/10
## 92     15 August 2023    4/10
## 93      21 July 2023    4/10
## 94    14 November 2023   4/10
## 95      9 September 2023  4/10
## 96    10 September 2023  4/10
## 97      24 July 2023    4/10
## 98      21 July 2023    4/10
## 99      25 July 2023    4/10
## 100     23 July 2023    4/10
```

```
ratingsszz <- reviews$ratings

ratingzznum <- as.factor(ratingsszz)

ratingzznum2 <- as.numeric(ratingzznum)

ratings_mean <- mean(ratingzznum2)

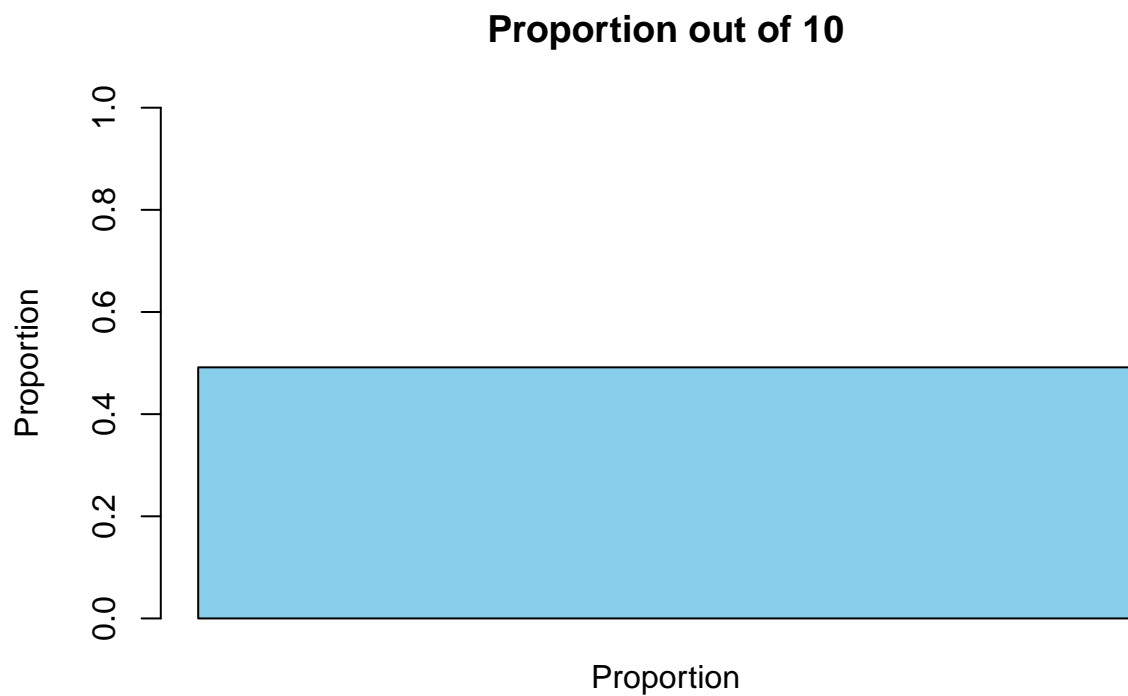
ratings_mean
```

```
## [1] 4.916667
```

```

value <- ratings_mean
total <- 10
proportion <- value / total
barplot(proportion, names.arg = "Proportion", col = "skyblue", main = "Proportion out of 10", ylab = "P

```



#As we observe the mean from our ratings is 4.915557. Over all the 300 reviews we took we got a 4.9 rating