

PROJECT B, Album Rating**PART II - REPORT**

I chose the Album Rating project as it seemed more interesting. Although web scraping is very new to me, after many trials and errors, I was successfully able to get the wanted review and rating of albums from the popular music site Pitchfork.

The project demands a function `pitchfork` that would take an URL as a parameter and then retrieve the review and the rating. As web scraping is new to me, I figured it would be easier to do a main code first, and later fit into a function. To begin, I used the library (`rvest`) and used `read_html()` function to read the URL given in the project's example. Now I already know the website of any URL that would be tested within my function. So I visited the album link given in the project, right clicked to view page source. I looked for the content I was looking for and found that it's within `<p>` in the class `contents dropcap`. I used `html_nodes()` to retrieve the review and then I used `html_text()` to read the review. To remove all the html, I used `%>%`. Once it worked, I now went to the link again and looked for the class or id that has the rating in it, and figured it was under the class `score`. I repeated the same thing for rating and got the right output. I tested with a few more URLs from pitchfork and when all of it worked, I changed my code to a function and created a list as the project demands and my function `pitchfork` now returns the list with the review and the rating.

Now let's test my function.

#Test1

```
review = pitchfork(' https://pitchfork.com/reviews/albums/davido-a-good-time/')
```

```
review$text
```

```
review$rating
```

Output

```
> review$text
```

```
[1] "For the last several years, Davido has been reshaping the sound of Nigerian pop. As a prolific hitmaker and one of Africa's....."
```

```
.....
```

```
Burna Boy is bringing the sounds of the diaspora back to the motherland, Davido is erecting a monument to his own status as a hometown hero."
```

```
> review$rating
```

```
[1] "8.3"
```

I get the perfect output!

#Test2

```
review = pitchfork(' https://pitchfork.com/reviews/albums/destroyer-have-we-met/')
```

```
review$text
```

```
review$rating
```

Output

```
> review$text
```

```
[1] "Spend enough time listening to Destroyer and the world will start to resemble a Dan Bejar song—when a bon viva....."
```

```
.....
```

```
the audience is as bewildered as us and the artist on stage has never sounded more at home."
```

```
[8] "Listen to our Best New Music playlist on Spotify and Apple Music."
```

```
> review$rating  
[1] "8.5"
```

Here, I get the desired review and rating, but there is an extra line of advertisement text which falls under <p> in the same class. There is an extra coverage of text. I tried several ways to get rid of the extra coverage but nothing seems to work, and this is the best output I get with my function.