

Titre du projet: Les meilleures chansons hebdomadaires (Top 40 hebdomadaires)

```

1
2 # Importer les packages
3 import requests
4 from bs4 import BeautifulSoup
5 import pandas as pd

```

On crée un ensemble de données des 40 meilleures chansons de chaque semaine d'une année donnée dans ce cas on va procéder ainsi d'abord récupérer le contenu du lien URL

```

1 # collecte du contenu du lien URL
2 url='https://top40weekly.com/'
3 r=requests.get(url)
4 print(r)
5

```

```
<Response [200]>
```

on crée l'objet BeautifulSoup pour pouvoir exploiter des codes HTML et aussi récupérer toutes les balises leurs contenu leurs frères etc le HTML parser pour dire c'est un parser HTML

```

1 doc = BeautifulSoup(r.text, 'html.parser')
2 doc

```

```

    }
  }
}

});
window.addEventListener('LazyLoad::Initialized', function (e) {
  var lazyLoadInstance = e.detail.instance;

  if (window.MutationObserver) {
    var observer = new MutationObserver(function(mutations) {
      var image_count = 0;
      var iframe_count = 0;
      var rocketlazy_count = 0;

      mutations.forEach(function(mutation) {
        for (i = 0; i < mutation.addedNodes.length; i++) {
          if (typeof mutation.addedNodes[i].getElementsByTagName
            return;
          }

          if (typeof mutation.addedNodes[i].getElementsByClassName
            return;
          }

          images = mutation.addedNodes[i].getElementsByTagName('
          is_image = mutation.addedNodes[i].tagName == "IMG";
          iframes = mutation.addedNodes[i].getElementsByTagName(
          is_iframe = mutation.addedNodes[i].tagName == "IFRAME"
          rocket_lazy = mutation.addedNodes[i].getElementsByClas

```

```

        image_count += images.length;
        iframe_count += iframes.length;
        rocketlazy_count += rocket_lazy.length;

        if(is_image){
            image_count += 1;
        }

        if(is_iframe){
            iframe_count += 1;
        }
    }
} );

if(image_count > 0 || iframe_count > 0 || rocketlazy_count > 0)
    lazyLoadInstance.update();
}
} );

var b      = document.getElementsByTagName("body")[0];
var config = { childList: true, subtree: true };

observer.observe(b, config);
    }, false);</script><script async="" data-no-minify="1" src="https://top40w
</script><script>!function(){ "use strict"; var e; e=document, function(){ var t, n; func
</script>

```

```

1 #doc=Soup.find_all(class_="x-text")
2 #print(doc)

```

Dans cette partie on utilise la fonction find pour recuperer le title le texte et strip pour supprimer les espaces

```

1 title=doc.find('header', class_="entry-header").text.strip()
2 title

```

```
'Welcome to the biggest online US Top 40 Archive'
```

Meme procedure pour la date en recuperant d'abord la balise et l'attribut

```

1 date=doc.find('h4', class_="h-custom-headline cs-ta-left h4").text.strip()
2 date

```

```
'This Week's Top 40 - October 30, 2021'
```

le findall va recuperer tout les balises de type p

```

1 p_tag=doc.find_all('p')
2 p_tag
3

```

```
(function() {
```

```

var opts = {
  artist: "",
  song: "",
  detect_artist: true,
  adunit_id: 100003641,
  div_id: "cf_async_" + Math.floor((Math.random() * 999999999))
};
document.write('<div id="'+opts.div_id+'"></div>');var c=function(){cf.sho
})();
</script></p>
<p>1 EASY ON ME by Adele<br/>
2 STAY by the Kid LAROI & Justin Bieber<br/>
3 INDUSTRY BABY by Lil Nas X & Jack Harlow<br/>
4 FANCY LIKE by Walker Hayes<br/>
5 BAD HABITS by Ed Sheeran<br/>
6 WAY 2 SEXY by Drake featuring Future & Young Thug<br/>
7 good 4 u by Olivia Rodrigo<br/>
8 LEVITATING by Dua Lipa<br/>
9 SHIVERS by Ed Sheeran<br/>
10 KISS ME MORE by Doja Cat featuring SZA</p>
<p>11 NEED TO KNOW by Doja Cat<br/>
12 ESSENCE by Wizkid featuring Justin Bieber & Tems<br/>
13 MY UNIVERSE by Coldplay x BTS<br/>
14 HEAT WAVES by Glass Animals<br/>
15 BEGGIN' by Maneskin<br/>
16 YOU RIGHT by Doja Cat & the Weeknd<br/>
17 SAVE YOUR TEARS by the Weeknd & Ariana Grande<br/>
18 KNIFE TALK by Drake featuring 21 Savage & Project Pat<br/>
19 WHO WANT SMOKE??? by Nardo Wick featuring G Herbo, Lil Durk & 21 Savage<br/>
20 BUBBLY by Young Thug with Drake & Travis Scott</p>
<p>21 COLD HEART (PNAU REMIX) by Elton John & Dua Lipa<br/>
22 traitor by Olivia Rodrigo<br/>
23 IF I DIDN'T LOVE YOU - by Jason Aldean & Carrie Underwood<br/>
24 MONTERO (CALL ME BY YOUR NAME) by Lil Nas X<br/>
25 CHASING AFTER YOU by Ryan Hurd with Maren Morris<br/>
26 GIRLS WANT GIRLS by Drake featuring Lil Baby<br/>
27 TAKE MY BREATH by the Weeknd<br/>
28 THAT'S WHAT I WANT by Lil Nas X<br/>
29 PEPAS by Farruko<br/>
30 HAPPIER THAN EVER by Billie Eilish</p>
<p>31 LOVE NWANTITI (AH AH AH) by CKay<br/>
32 MEET ME AT OUR SPOT by THE ANXIETY: WILLOW & Tyler Cole<br/>
33 WOCKESHA by Moneybagg Yo<br/>
34 BUY DIRT by Jordan Davis featuring Luke Bryan<br/>
35 MEMORY I DON'T MESS WITH by Lee Brice<br/>
36 A-O-K by Tai Verdes<br/>
37 MY BOY by Elvie Shane<br/>
38 GHOST by Justin Bieber<br/>
39 PISSSED ME OFF by Lil Durk<br/>
40 YOU SHOULD PROBABLY LEAVE by Chris Stapleton</p>
<p><iframe allow="encrypted-media" allowtransparency="true" data-lazy-src="https:
<p><script async="" src="//pagead2.googlesyndication.com/pagead/js/adsbygoogle.js
<ins class="adsbygoogle" data-ad-client="ca-pub-8091657638982844" data-ad-format=
<script>
  (adsbygoogle = window.adsbygoogle || []).push({});
</script></p>
<p>Join the Top 40 Weekly Newsletter for Fresh Music Content Sent Directly To You

```

ce code recupere le premier paragraphe

```
1 p_tag[3]

<p>1 EASY ON ME by Adele<br/>
2 STAY by the Kid LAROI & Justin Bieber<br/>
3 INDUSTRY BABY by Lil Nas X & Jack Harlow<br/>
4 FANCY LIKE by Walker Hayes<br/>
5 BAD HABITS by Ed Sheeran<br/>
6 WAY 2 SEXY by Drake featuring Future & Young Thug<br/>
7 good 4 u by Olivia Rodrigo<br/>
8 LEVITATING by Dua Lipa<br/>
9 SHIVERS by Ed Sheeran<br/>
10 KISS ME MORE by Doja Cat featuring SZA</p>
```

Dans cette partie on recupere le premier et on l'affiche en texte

```
1 p3=p_tag[3].text.strip()
2 p3 = p3.replace('\n',"")
3 p3

'1 EASY ON ME by Adele2 STAY by the Kid LAROI & Justin Bieber3 INDUSTRY BABY by Lil
Nas X & Jack Harlow4 FANCY LIKE by Walker Hayes5 BAD HABITS by Ed Sheeran6 WAY 2 SEX
Y by Drake featuring Future & Young Thug7 good 4 u by Olivia Rodrigo8 LEVITATING by
Dua Lipa9 SHIVERS by Ed Sheeran10 KISS ME MORE by Doja Cat featuring SZA'
```

```
1 p4=[p3[2:12],p3[23:27],p3[31:60]]
2 p4

['EASY ON ME', 'STAY', 'the Kid LAROI & Justin Bieber']
```

On affiche le second paragraphe

```
1 p_tag[4]

<p>11 NEED TO KNOW by Doja Cat<br/>
12 ESSENCE by Wizkid featuring Justin Bieber & Tems<br/>
13 MY UNIVERSE by Coldplay x BTS<br/>
14 HEAT WAVES by Glass Animals<br/>
15 BEGGIN' by Maneskin<br/>
16 YOU RIGHT by Doja Cat & the Weeknd<br/>
17 SAVE YOUR TEARS by the Weeknd & Ariana Grande<br/>
18 KNIFE TALK by Drake featuring 21 Savage & Project Pat<br/>
19 WHO WANT SMOKE??? by Nardo Wick featuring G Herbo, Lil Durk & 21 Savage<br/>
20 BUBBLY by Young Thug with Drake & Travis Scott</p>
```

On recupere le troisieme paragraphe

```
1 p_tag[5]

<p>21 COLD HEART (PNAU REMIX) by Elton John & Dua Lipa<br/>
22 traitor by Olivia Rodrigo<br/>
23 IF I DIDN'T LOVE YOU - by Jason Aldean & Carrie Underwood<br/>
24 MONTERO (CALL ME BY YOUR NAME) by Lil Nas X<br/>
25 CHASING AFTER YOU by Ryan Hurd with Maren Morris<br/>
26 GIRLS WANT GIRLS by Drake featuring Lil Baby<br/>
```

```

27 TAKE MY BREATH by the Weeknd<br/>
28 THAT'S WHAT I WANT by Lil Nas X<br/>
29 PEPAS by Farruko<br/>
30 HAPPIER THAN EVER by Billie Eilish</p>

```

ceci recupere le dernier paragraphe

```
1 p_tag[6]
```

```

<p>31 LOVE NWANTITI (AH AH AH) by CKay<br/>
32 MEET ME AT OUR SPOT by THE ANXIETY: WILLOW & Tyler Cole<br/>
33 WOCKESHA by Moneybagg Yo<br/>
34 BUY DIRT by Jordan Davis featuring Luke Bryan<br/>
35 MEMORY I DON'T MESS WITH by Lee Brice<br/>
36 A-O-K by Tai Verdes<br/>
37 MY BOY by Elvie Shane<br/>
38 GHOST by Justin Bieber<br/>
39 PISSED ME OFF by Lil Durk<br/>
40 YOU SHOULD PROBABLY LEAVE by Chris Stapleton</p>

```

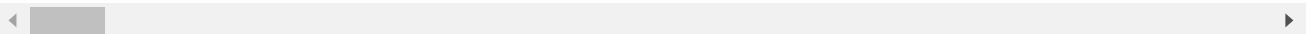
On a recuprerer toutes informations voulues maintenant on va passer à la questios suite : On va capturer pour chacunes des 40 chansons hebdomodaires des informatoins telles le titre de la chanson ,l'artiste et le classement hebdomodaire

```

1 # on utilise ce code pour des iterations
2 if r.ok:
3     print(r.headers)
4

```

```
{'Date': 'Thu, 04 Nov 2021 21:17:08 GMT', 'Content-Type': 'text/html; charset=UTF-8',
```



Dans la partie ci-dessous on va creer une liste d'abord en suite parcourir avec la boucle en utilisant la taille du doc en fin va afficher tout les insformatios dans cette balise

```

1 tops=[]
2 for i in range(len(doc)):
3     top=[]
4     for j in doc.find_all('p'):
5         top.append(j.text)
6     tops.append(top)
7 print(tops)
8
9

```

IOPub data rate exceeded.

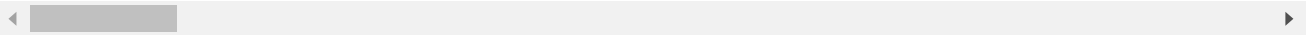
The notebook server will temporarily stop sending output to the client in order to avoid crashing it. To change this limit, set the config variable `--NotebookApp.iopub_data_rate_limit`.

Current values:

NotebookApp.iopub_data_rate_limit=1000000.0 (bytes/sec)
 NotebookApp.rate_limit_window=3.0 (secs)

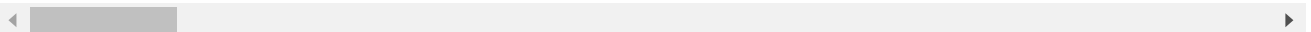
```
1 tops[1]
```

```
["To read Jarrett Nolan's Commentary for this week's chart, click here!",
 '\n/ TFP - Top40Weekly.com /\n(function() {\n\tvar opts = {\n\t\tartist: "",\n\t\ttsco
'\r\n/* TFP - Top40Weekly.com */\r\n(function() {\r\n\tvar opts = {\r\n\t\tartist: '
'1 EASY ON ME by Adele\n2 STAY by the Kid LAROI & Justin Bieber\n3 INDUSTRY BABY by
"11 NEED TO KNOW by Doja Cat\n12 ESSENCE by Wizkid featuring Justin Bieber & Tems\n1
"21 COLD HEART (PNAU REMIX) by Elton John & Dua Lipa\n22 traitor by Olivia Rodrigo\r
"31 LOVE NWANTITI (AH AH AH) by CKay\n32 MEET ME AT OUR SPOT by THE ANXIETY: WILLOW
'',
'\n\n\n      (adsbygoogle = window.adsbygoogle || []).push({});\n',
'Join the Top 40 Weekly Newsletter for Fresh Music Content Sent Directly To Your Int
```



```
1 tops[3]
```

```
["To read Jarrett Nolan's Commentary for this week's chart, click here!",
 '\n/ TFP - Top40Weekly.com /\n(function() {\n\tvar opts = {\n\t\tartist: "",\n\t\ttsco
'\r\n/* TFP - Top40Weekly.com */\r\n(function() {\r\n\tvar opts = {\r\n\t\tartist: '
'1 EASY ON ME by Adele\n2 STAY by the Kid LAROI & Justin Bieber\n3 INDUSTRY BABY by
"11 NEED TO KNOW by Doja Cat\n12 ESSENCE by Wizkid featuring Justin Bieber & Tems\n1
"21 COLD HEART (PNAU REMIX) by Elton John & Dua Lipa\n22 traitor by Olivia Rodrigo\r
"31 LOVE NWANTITI (AH AH AH) by CKay\n32 MEET ME AT OUR SPOT by THE ANXIETY: WILLOW
'',
'\n\n\n      (adsbygoogle = window.adsbygoogle || []).push({});\n',
'Join the Top 40 Weekly Newsletter for Fresh Music Content Sent Directly To Your Int
```



```
1 top[5]
```

```
'21 COLD HEART (PNAU REMIX) by Elton John & Dua Lipa\n22 traitor by Olivia Rodrigo\n
23 IF I DIDN'T LOVE YOU - by Jason Aldean & Carrie Underwood\n24 MONTERO (CALL ME BY
YOUR NAME) by Lil Nas X\n25 CHASING AFTER YOU by Ryan Hurd with Maren Morris\n26 GIR
LS WANT GIRLS by Drake featuring Lil Baby\n27 TAKE MY BREATH by the Weeknd\n28 THA
T'S WHAT I WANT hv Lil Nas X\n29 PFPAS hv Farruko\n30 HAPPTER THAN EVFR hv Billie Fi
```

On utilise quel rappel sur les pages content DANS CETTE partie on va utiliser les pages content pour afficher le contenu de la page . Cette partie c'est pour recuperer quelques informations en utilisant la commande page_content

```
1 page_content=r.text
```

```
1 len(page_content)
```

```
114647
```

```
1 # CE code nous permet de l'ouvrir dans un fichier machine-learning-top40weekly.html
2 with open('machine-learning-top40weekly.html', 'w', encoding='utf-8') as file:
3     file.write(page_content)
```

```
1 # Pour recuperer les 1000 premier caracteres
2 page_content[:1000]
```

```
<!DOCTYPE html>\n<html class="no-js" lang="en-US" prefix="og: https://ogp.me/ns#">  
\n<head>\n<meta charset="UTF-8">\n<meta name="viewport" content="width=device-width,  
initial-scale=1.0">\n<link rel="pingback" href="https://top40weekly.com/xmlrpc.php">  
\n<script>\n(function(w, d) {\n\tw.adthrive = w.adthrive || {};\n\tw.adthrive.cmd =  
w.adthrive.cmd || [];\n\tw.adthrive.plugin = \'adthrive-ads-1.1.2\';\n\tw.adthrive.h  
ost = \'ads.adthrive.com\';\n\tw.\ntvar s = d.createElement(\'script\');\n\tw.ts.async = t  
rue;\n\tw.ts.referrerpolicy=\'no-referrer-when-downgrade\';\n\tw.ts.src = \'https://\' +  
w.adthrive.host + \'/sites/5cadc09cb13d6c3ef8c8bbde/ads.min.js?referrer=\' + w.encod  
eURIComponent(w.location.href) + (window.adthriveCLS && window.adthriveCLS.bucket !=
```

● ×