

1.1.2 벅스 크롤링 결과를 엑셀 파일로 저장하기

[1]:

```
1 # 벅스 사이트 접속하기
2 from selenium import webdriver
3 from bs4 import BeautifulSoup
4
5 driver = webdriver.Chrome('C:/Myexam/chromedriver/chromedriver.exe')
6 url = 'https://music.bugs.co.kr/chart'
7 driver.get(url)
8
9 html = driver.page_source
10 soup = BeautifulSoup(html, 'html.parser')
```

C:\Users\Wstudent\AppData\Local\Temp\ipykernel_5512\W3609719275.py:5: DeprecationWarning: executable_path has been deprecated, please pass in a Service object
 driver = webdriver.Chrome('C:/Myexam/chromedriver/chromedriver.exe')

[2]:

```
1 # tr 태그로 곡 정보 찾기 1
2 songs = soup.select('tr')
3 print(len(songs))
```

104

[3]:

```
1 # tr 태그로 곡 정보 찾기 2
2 songs = soup.select('tbody > tr')
3 print(len(songs))
```

103

[4]:

```
1 # tr 태그로 곡 정보 찾기 3
2 songs = soup.select('table > tbody > tr')
3 print(len(songs))
```

103

[5]:

```
1 #tr 태그로 곡 정보 찾기 4
2 songs = soup.select('table.byChart > tbody > tr')
3 print(len(songs))
```

100

[6]:

```
1 # songs 태그 중 첫 번째 태그 출력해보기  
2 print(songs[0])
```

```

<tr albumid="4082425" artistid="20164333" multiartist="N" mvid
="613861" rowtype="track" trackid="6184997">
<input name="_isStream" type="hidden" value="6184997"/>
<input name="_isDown" type="hidden" value="6184997"/>
<td class="check"><input buyminquality="T" disc_id="1" name="che
ck" title="Ditto" type="checkbox" value="6184997"/></td>
<td>
<div class="ranking">
<strong>1</strong>
<p class="change none"><em>0</em><span>변동없음</span></p>
</div>
</td>
<td>
<a class="thumbnail" href="https://music.bugs.co.kr/album/408242
5?wl_ref=list_tr_07_chart" onclick="
">
<span class="mask"></span>

</a>
</td>
<td>
<a class="trackInfo" href="https://music.bugs.co.kr/track/618499
7?wl_ref=list_tr_08_chart" onclick="
">곡정보</a>
</td>
<th scope="row">
<p adult_yn="N" class="title">
<a adultcheckval="1" aria-label="새창" href="javascript:;" oncli
ck="bugs.wiselog.area('list_tr_09_chart');bugs.music.listen('618
4997',true);
" title="Ditto">Ditto</a>
</p>
</th>
<td class="left">
<p class="artist">
<a href="https://music.bugs.co.kr/artist/20164333?wl_ref=list_tr
_10_chart" onclick="
" title="NewJeans">NewJeans</a>
</p>
</td>
<td class="left">
<a class="album" href="https://music.bugs.co.kr/album/4082425?wl
_ref=list_tr_11_chart" onclick="
" title="NewJeans 'OMG'">NewJeans 'OMG'</a>
</td>
<td>
<a adultcheckval="1" aria-label="새창" class="btn play" href="ja
vascript:;" onclick="bugs.wiselog.area('list_tr_12_chart');bugs.
music.listen('6184997',true);
">듣기</a>
</td>

```

```

<td>
<a adultcheckval="1" aria-label="새창" class="btn addPlaylist" href="javascript:;" onclick="bugs.wiselog.area('list_tr_13_chart');bugs.music.listen('6184997');">재생목록에 추가</a>
</td>
<td>
<a class="btn addAlbum" href="javascript:;" onclick="bugs.wiselog.area('list_tr_14_chart');bugs.music.addTrackToMyAlbum(this,'6184997');" track_title="Ditto">내 앨범에 담기</a>
</td>
<td>
<a adultcheckval="1" aria-label="새창" class="btn download flac" href="javascript:;" onclick="bugs.wiselog.area('list_tr_15_chart');bugs.layermenu.down(this,'6184997','true','true','44.1','true','44.1',null);"><em class="flac">flac </em>다운로드</a>
</td>
<td>
<a aria-label="새창" class="btn viewMV" href="javascript:;" onclick="bugs.wiselog.area('list_tr_16_chart');bugs.layermenu.mv(this,6184997,613861,'Y','Y','Y','Y',6184997,'N','ALL','Y');">영상 재생</a>
</td>
<td>
<a artist_disp_nm="NewJeans" artist_id="20164333" class="btnActions" href="javascript:;" layer_type="CHART" layerpositiontarget="td" onclick="bugs.wiselog.area('list_tr_17_chart');bugs.layermenu.trackMoreAction(this,'6184997','0','4082425','bugs.music.listenRadioFromSeed(W'6184997W',W'trackW');','N','Y','_chart');" track_title="Ditto">기타 기능</a>
</td>
</tr>

```

[7]:

```

1 # 한 개의 곡 정보 저장하기
2 song = songs[0]

```

[8]:

```

1 # 벅스 사이트에서 곡 제목 찾기 1
2 title = song.select('a')
3 len(title)

```

[9]:

```
1 # 벅스 사이트에서 곡 제목 찾기 2
2 title = song.select('p > a')
3 len(title)
```

2

[10]:

```
1 # 벅스 사이트에서 곡 제목 찾기 3
2 title = song.select('p.title > a')
3 len(title)
```

1

[11]:

```
1 # 벅스 사이트에서 곡 제목 텍스트 출력하기
2 title = song.select('p.title > a')[0].text
3 title
```

'Ditto'

[12]:

```
1 # 벅스 사이트에서 가수 이름 텍스트 출력하기
2 singer = song.select('p.artist > a')[0].text.strip()
3 singer
```

'NewJeans'

[13]:

```
1 # 벅스 100위 노래 순위 정보 가져오기
2 songs = soup.select('table.byChart > tbody > tr')
3 for song in songs:
4     title = song.select('p.title > a')[0].text
5     singer = song.select('p.artist > a')[0].text
6     print(title, singer, sep = '|')
```

Ditto|NewJeans
OMG|NewJeans
VIBE (feat. Jimin of BTS)|태양
Hype Boy|NewJeans
사건의 지평선|윤하(Younha/ユンナ)
Attention|NewJeans
ANTIFRAGILE|LE SSERAFIM (르세라핌)
Candy|NCT DREAM
After LIKE|IVE (아이브)
NOT SORRY (Feat. pH-1) (Prod. by Slom)|이영지
LOVE DIVE|IVE (아이브)
WHEN I MOVE|카라(Kara)
오르트구름|윤하(Younha/ユンナ)
Nxde|(여자)아이들
Cookie|NewJeans
Shut Down|BLACKPINK
새뽕 (Prod. ZICO) (Feat. 호미들)|지코
I Don't Think That I Like Her|Charlie Puth(찰리 푸스)
Made You Look|Meghan Trainor(메간 트레이너)
FEARLESS|LE SSERAFIM (르세라핌)
Unholy|Sam Smith(샘 스미스)
Rush Hour (Feat. j-hope of BTS)|Crush
Monologue|테이(Tei)
Pink Venom|BLACKPINK
Dreamers [Music from the FIFA World Cup Qatar 2022 Official Soundtrack] (feat. FIFA Sound)|정국
사랑인가 봐|멜로망스(MeloMance)
너의 모든 순간|성시경
TOMBOY|(여자)아이들
자격지심 (Feat. ZICO)|BE'O (비오)
그때 그 순간 그대로 (그그그)|WSG워너비 (가야G)
Say I Love You|우디(Woody)
ELEVEN|IVE (아이브)
STEP|카라(Kara)
그라데이션|10CM
When I Get Old|Christopher(크리스토퍼)
정이라고 하자 (Feat. 10CM)|BIG Naughty (서동현)
That's Hilarious|Charlie Puth(찰리 푸스)
도깨비불 (Illusion)|aespa
STAY|The Kid LAROI
해요 (2022)|#안녕
그댄 행복에 살텐데 (2022)|최유리
내가 아니라도|주호
드라마|아이유(IU)
strawberry moon|아이유(IU)
Dangerously|Charlie Puth(찰리 푸스)
LOVE me|BE'O (비오)
Feel My Rhythm|Red Velvet (레드벨벳)
INVU|태연 (TAEYEON)
Anti-Hero|Taylor Swift(테일러 스위프트)
다정히 내 이름을 부르면|경서예지
딱 10CM만|10CM
Snowman|Sia(시아)

한 페이지가 될 수 있게|DAY6 (데이식스)
Beautiful Liar|몬스타엑스(Monsta X)
모든 날, 모든 순간 (Every day, Every Moment)|폴킴(Paul Kim)
그대를 알고|송하예
나의 X에게|경서
Impurities|LE SSERAFIM (르세라핌)
우린 그렇게 사랑해서|강민경
취중고백|김민석 (멜로망스)
아픈 나를|성시경
마이웨이 (MY WAY) (Prod. R.Tee)|저스디스(JUSTHIS)
주저하는 연인들을 위해|잔나비
Left and Right (Feat. Jung Kook of BTS)|Charlie Puth(찰리 푸스)
보고싶었어|WSG워너비(4FIRE)
사랑은 늘 도망가|임영웅
Hurt|NewJeans
Off My Face|Justin Bieber(저스틴 비버)
봄여름가을겨울 (Still Life)|BIGBANG (빅뱅)
팡파레|다비치
Talk that Talk|TWICE (트와이스)
Birthday|Red Velvet (레드벨벳)
눈이 오잖아(Feat. 헤이즈)|이무진
신호등|이무진
Next Level|aespa
잘가요|주호
FOREVER 1|소녀시대 (GIRLS' GENERATION)
DICE|NMIXX
떠나보낼 준비해 둘걸 그랬어|임한별
I Ain't Worried|OneRepublic(원리퍼블릭)
Cheshire|ITZY (있지)
진심이었던 사람만 바보가 돼|권진아
눈 (EYE) (Feat. BIG Naughty, 저스디스 (JUSTHIS)) (Prod. R.Tee)|
던말릭 (DON MALIK)
Love story|불빨간사춘기
가슴으로 운다|김연지
내 기쁨은 너가 벤틀리를 끄는 거야|김승민
SNEAKERS|ITZY (있지)
미운오리새끼 (Feat. 선우정아, BOBBY) (Prod. R.Tee)|허성현 (Huh)
LAW (Prod. Czaer)|윤미래
WITCH (Feat. 박재범, So!YoON!) (Prod. by Slom)|이영지
GHOST TOWN|Benson Boone
위하여 (We Higher) (Prod. GroovyRoom)|그루비룸 (GroovyRoom)
마지막 사랑|신예영
That That (prod. & feat. SUGA of BTS)|싸이 (PSY)
뽕 (Feat. 팔로알토 (Paloalto), 저스디스 (JUSTHIS)) (Prod. R.Tee)|
던말릭 (DON MALIK)
Santa Tell Me|Ariana Grande(아리아나 그란데)
Celebrity|아이유(IU)
첫 눈|EXO
어른|Sondia
DEJAVU (Feat. 박재범) (Prod. by Slom)|이영지

[14]:

```

1 # 반복문을 이용해 곡과 가수명을 song_data에 저장하기
2 song_data = []
3 rank = 1
4 songs = soup.select('table.byChart > tbody > tr')
5 for song in songs:
6     title = song.select('p.title > a')[0].text
7     singer = song.select('p.artist > a')[0].text
8     song_data.append(['Bugs', rank, title, singer])
9     rank = rank + 1

```

[15]:

```

1 # song_data 리스트를 이용해 데이터프레임 만들기
2 import pandas as pd
3 columns = ['서비스', '순위', '타이틀', '가수']
4 pd_data = pd.DataFrame(song_data, columns = columns)
5 pd_data.info()

```

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 100 entries, 0 to 99

Data columns (total 4 columns):

#	Column	Non-Null Count	Dtype
---	--------	----------------	-------

0	서비스	100 non-null	object
1	순위	100 non-null	int64
2	타이틀	100 non-null	object
3	가수	100 non-null	object

dtypes: int64(1), object(3)

memory usage: 3.2+ KB

[16]:

```

1 # 크롤링 결과를 엑셀 파일로 저장하기
2 pd_data.to_excel('./files/bugs.xlsx', index=False)

```

[17]:

```

1 # 벅스 인기차트를 크롤링한 결과를 엑셀 파일로 저장하기(전체 코드)
2 from selenium import webdriver
3 from bs4 import BeautifulSoup
4 import pandas as pd
5
6 driver = webdriver.Chrome('C:/Myexam/chromedriver/chromedriver.exe')
7 url = 'https://music.bugs.co.kr/chart'
8 driver.get(url)
9
10 html = driver.page_source
11 soup = BeautifulSoup(html, 'html.parser')
12
13 song_data = []
14 rank = 1
15 songs = soup.select('table.byChart > tbody > tr')
16 for song in songs:
17     title = song.select('p.title > a')[0].text
18     singer = song.select('p.artist > a')[0].text
19     song_data.append(['Bugs', rank, title, singer])
20     rank = rank + 1
21
22 columns = ['서비스', '순위', '타이틀', '가수']
23 pd_data = pd.DataFrame(song_data, columns = columns)
24 pd_data.to_excel('./files/bugs.xlsx', index=False)

```

C:\Users\student\AppData\Local\Temp\ipykernel_5512\1205041318.py:6: DeprecationWarning: executable_path has been deprecated, please pass in a Service object
 driver = webdriver.Chrome('C:/Myexam/chromedriver/chromedriver.exe')

[]:

1