

## 1.1.3 지니 크롤링 결과를 엑셀 파일로 저장하기

[1]:

```
1 # 지니 사이트에 접속하기
2 from selenium import webdriver
3 from bs4 import BeautifulSoup
4
5 driver = webdriver.Chrome('C:/Myexam/chromedriver/chromedriver.exe')
6 url = 'https://www.genie.co.kr/chart/top200'
7 driver.get(url)
8
9 html = driver.page_source
10 soup = BeautifulSoup(html, 'html.parser')
```

C:\Users\student\AppData\Local\Temp\ipykernel\_12604\264886278.py:5: DeprecationWarning: executable\_path has been deprecated, please pass in a Service object  
driver = webdriver.Chrome('C:/Myexam/chromedriver/chromedriver.exe')

[2]:

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

50

[3]:

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

```

<tr class="list" songid="99570005">
<td class="check"><input class="select-check" title="Ditto" type
="checkbox"/></td>
<td class="number">1

<span class="rank">
<span class="rank"><span class="rank-none"><span class="hide">유
지</span></span></span>
</span>
</td>
<td><a class="cover" href="#" onclick="fnViewAlbumLayer('8332557
7');return false;"><span class="mask"></span></a></td>
<td class="link"><a class="btn-basic btn-info" href="#" onclick
="fnViewSongInfo('99570005');return false;">곡 제목 정보 페이지
</a></td>
<td class="info">
<a class="title ellipsis" href="#" onclick="fnPlaySong('9957000
5','1');return false;" title="재생">

```

Ditto</a>

```

<a class="artist ellipsis" href="#" onclick="fnViewArtist('81490
206');return false;">NewJeans</a>
<div class="toggle-button-box">
<button class="btn artist-etc" onclick="fnRelationArtistList('99
570005');" type="button">외</button>
<ul class="list" id="RelationArtist_99570005"></ul>
</div>
<i class="bar">|</i>
<a class="albumtitle ellipsis" href="#" onclick="fnViewAlbumLaye
r('83325577');return false;">NewJeans 'OMG'</a>
</td>
<td class="btns"><a class="btn-basic btn-listen" href="#" onclik
k="fnPlaySong('99570005','1'); return false;" title="재생">듣
기</a></td>
<td class="btns"><a class="btn-basic btn-add" href="#" onclick
="fnPlaySong('99570005','3'); return false;" title="추가">재생
목록에 추가</a></td>
<td class="btns"><button class="btn-basic btn-album" id="add_my_
album_99570005" onclick="fnAddMyAlbumForm('#add_my_album_9957000
5','99570005',10,10);return false;" songid="99570005" type
="button">플레이리스트에 담기</button></td>
<td class="btns"><a class="btn-basic btn-down" href="#" onclick
="fnDownSong('99570005');return false;" title="다운">다운로드</a>

```

```

</td>
<td class="btns">
<div class="toggle-button-box lyr-mv" id="list-mv_99570005">
<a alt="Ditto" class="btn btn-basic btn-mv" href="#" onclick="fn
PlayMv('99570005','3'); return false;" title="뮤비">뮤직비디오
보기</a>
</div>
</td>
<td class="more">
<div class="toggle-button-box">
<button class="btn btn-basic btn-more" type="button">더보기</but
ton>
<ul class="list">
<li><a class="item" href="#" onclick="shareDo('99570005');return
false;" title="공유">공유하기/음악나누기</a></li>
<li><a class="item" href="#" onclick="fnGiftSong('99570005');ret
urn false;" title="선물">선물하기</a></li>
</ul>
</div>
</td>
</tr>

```

[4]:

```

1 # 한 개의 곡 정보 저장하기
2 song = songs[0]

```

[5]:

```

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

```

1

[6]:

```

1 # 지니 사이트의 곡 제목 출력해보기 1
2 title = song.select('a.title')[0].text
3 title

```

```

'Wn
Wn
Wn
Wn
Ditto'
Wn
Wn
Wn
Wn

```

[7]:

```
1 # 지니 사이트의 곡 제목 출력해보기 2
2 title = song.select('a.title')[0].text.strip()
3 title
```

'Ditto'

[8]:

```
1 # 지니 사이트의 가수명 찾기
2 singer = song.select('a.artist')
3 len(singer)
```

1

[9]:

```
1 # 지니 사이트의 가수명 출력해보기 1
2 singer = song.select('a.artist')[0].text
3 singer
```

'NewJeans'

[10]:

```
1 # 지니 사이트의 가수명 출력해보기 2
2 songs = soup.select('tbody > tr')
3 for song in songs:
4     title = song.select('a.title')[0].text.strip()
5     singer = song.select('a.artist')[0].text
6     print(title, singer, sep = '|')
```

Ditto|NewJeans  
OMG|NewJeans  
Hype boy|NewJeans  
사건의 지평선|윤하 (YOUNHA)  
VIBE (Feat. Jimin of BTS)|태양  
ANTIFRAGILE|LE SSERAFIM (르세라핌)  
After LIKE|IVE (아이브)  
Attention|NewJeans  
NOT SORRY (Feat. pH-1) (Prod. by Slom)|이영지  
LOVE DIVE|IVE (아이브)  
Nxde|(여자)아이들  
새뽕 (Prod. by ZICO) (Feat. 호미들)|지코 (ZICO)  
Candy|NCT DREAM  
사랑은 늘 도망가|임영웅  
Monologue|테이 (Tei)  
Shut Down|BLACKPINK  
우리들의 블루스|임영웅  
Pink Venom|BLACKPINK  
마이웨이 (MY WAY) (Prod. by R.Tee)|저스디스 (JUSTHIS) & R.Tee &  
던말릭 (DON MALIK) & 허성현 (Huh) & KHAN & 맥대디 (Mckdaddy) & L  
OS  
TOMBOY|(여자)아이들  
해요 (2022)|#안녕  
다시 만날 수 있을까|임영웅  
Cookie|NewJeans  
그때 그 순간 그대로 (그그그)|WSG워너비 (가야G)  
Rush Hour (Feat. j-hope of BTS)|Crush  
사랑인가 봐|멜로망스 (MeloMance)  
Say I Love You|우디 (Woody)  
That's Hilarious|Charlie Puth  
이제 나만 믿어요|임영웅  
취중고백|김민석 (멜로망스)  
London Boy|임영웅  
보고싶었어|WSG워너비 (4FIRE)  
무지개|임영웅  
Dreamers (Music from the FIFA World Cup Qatar 2022 Official Soundtrack) (Feat. FIFA Sound)|정국 & 방탄소년단  
ELEVEN|IVE (아이브)  
FEARLESS|LE SSERAFIM (르세라핌)  
Polaroid|임영웅  
다정히 내 이름을 부르면|경서예지 & 전건호  
오르트구름|윤하 (YOUNHA)  
아버지|임영웅  
내가 아니라도|주호  
가슴으로 운다|김연지  
WHEN I MOVE|카라 (KARA)  
I Don't Think That I Like Her|Charlie Puth  
자격지심 (Feat. ZICO)|BE'0 (비오)  
너의 모든 순간|성시경  
잘가요|주호  
A bientot|임영웅  
The Drum|Alan Walker  
그라데이션|10CM

[11]:

```

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

```

[12]:

```

1 # song_data 리스트를 이용해 엑셀 파일로 저장하기
2 import pandas as pd
3
4 columns = ['서비스', '순위', '타이틀', '가수']
5 pd_data = pd.DataFrame(song_data, columns = columns)
6 pd_data.to_excel('./files/genie.xlsx', index=False)

```

[13]:

```

1 # 지니 인기차트를 크롤링한 결과를 엑셀 파일로 저장하기(전체코드)
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://www.genie.co.kr/chart/top200'
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('tbody > tr')
16 for song in songs:
17     title = song.select('a.title')[0].text.strip()
18     singer = song.select('a.artist')[0].text
19     song_data.append(['Genie', 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/genie.xlsx', index=False)

```

C:\Users\student\AppData\Local\Temp\ipykernel\_12604\2563067557.py:6: DeprecationWarning: executable\_path has been deprecated, please pass in a Service object  
 driver = webdriver.Chrome('C:/Myexam/chromedriver/chromedriver.exe')



[ ]:

1	
---	--