

Group_7_Scraping_Songs

April 16, 2021

1 Group 7 - Song Scraper (Billboard and Genius)

```
[1]: import requests

[2]: res = requests.get("https://www.billboard.com/articles/news/hot-100-turns-60/
↳8468142/hot-100-all-time-biggest-hits-songs-list")

[3]: res.raise_for_status()

[4]: song_list_page = res.content

[5]: from bs4 import BeautifulSoup

[6]: song_list_soup = BeautifulSoup(song_list_page)

[9]: song_divs = song_list_soup.find_all("div", {"class": "article__html"})

[21]: import re

[25]: pattern = re.compile(r"^\d*[\.,]")

[45]: def sanitize_name(name):
    sanitized_name = ""
    for c in name:
        if c.isalpha():
            sanitized_name += c
        elif c == " ":
            sanitized_name += '-'
    return sanitized_name

[46]: for cont in song_divs:
    song_items = cont.find_all("p")
    for item in song_items:
        items = item.stripped_strings
        try:
            current_index = next(items)
        except StopIteration:
```

```

        continue

    if pattern.match(current_index):
        song_title = sanitize_name(next(items))
        song_year = next(items)
        try:
            singer = sanitize_name(next(items))
        except StopIteration:
            continue

        print(f"https://genius.com/{singer}-{song_title}-lyrics")
        res = requests.get(f"https://genius.com/
↪{singer}-{song_title}-lyrics")
        res.raise_for_status()

```

<https://genius.com/Chubby-Checker-The-Twist-lyrics>
<https://genius.com/Santana-Smooth-lyrics>
<https://genius.com/Bobby-Darin-Mack-the-Knife-lyrics>
<https://genius.com/Mark-Ronson-Uptown-Funk-lyrics>
<https://genius.com/Leann-Rimes-How-Do-I-Live-lyrics>
<https://genius.com/LMFAO-Party-Rock-Anthem-lyrics>
<https://genius.com/The-Black-Eyed-Peas-I-Gotta-Feeling-lyrics>
<https://genius.com/Los-Del-Rio-Macarena-Bayside-Boys-Mix-lyrics>
<https://genius.com/Ed-Sheeran-Shape-of-You-lyrics>
<https://genius.com/Debby-Boone-You-Light-Up-My-Life-lyrics>
<https://genius.com/The-Beatles-Hey-Jude-lyrics>
<https://genius.com/The-Chainsmokers-Feat-Halsey-Closer-lyrics>

↪ -----

HTTPError Traceback (most recent call↪
 ↪last)

```

<ipython-input-46-0d1dde6af401> in <module>
    18         print(f"https://genius.com/{singer}-{song_title}-lyrics")
    19         res = requests.get(f"https://genius.com/
↪{singer}-{song_title}-lyrics")
    ---> 20         res.raise_for_status()

```

```

/usr/local/lib/python3.8/site-packages/requests/models.py in ↪
↪raise_for_status(self)
    939
    940         if http_error_msg:
--> 941             raise HTTPError(http_error_msg, response=self)

```

```
942
```

```
943     def close(self):
```

```
        HTTPError: 404 Client Error: Not Found for url: https://genius.com/  
↳The-Chainsmokers-Feat-Halsey-Closer-lyrics
```

```
[ ]:
```