

For killers we have 137.
The full pull will take for this artist will take 0.38 hours.

Part 2: Pulling Lyrics

Now that we have the links to our lyrics pages, let's go scrape them! Here are the steps for this part.

1. Create an empty folder in our repo called "lyrics".
2. Iterate over the artists in `lyrics_pages`.
3. Create a subfolder in lyrics with the artist's name. For instance, if the artist was Cher you'd have `lyrics/cher/` in your repo.
4. Iterate over the pages.
5. Request the page and extract the lyrics from the returned HTML file using BeautifulSoup.
6. Use the function below, `generate_filename_from_url`, to create a filename based on the lyrics page, then write the lyrics to a text file with that name.

```
In [57]: def generate_filename_from_link(link) :  
      
    if not link :  
        return None  
      
    # drop the http or https and the html  
    name = link.replace("https","").replace("http","")  
    name = link.replace(".html","")  
      
    name = name.replace("/lyrics/", "")  
      
    # Replace useless characters with UNDERSCORE  
    name = name.replace("///","").replace(" ","_").replace("/","_")  
      
    # tack on .txt  
    name = name + ".txt"  
      
    return(name)  
      
In [60]:   
      
    # Make the lyrics folder here. If you'd like to practice your programming, add functionality  
    # that checks to see if the folder exists. If it does, then use shutil.rmtree to remove it and create a new  
      
    if os.path.isdir("lyrics") :  
        shutil.rmtree("lyrics/")  
      
    os.mkdir("lyrics")  
      
In [61]: url_stub = "https://www.azlyrics.com"  
start = time.time()  
total_pages = 0  
      
for artist in lyrics_pages :  
      
    # Use this space to carry out the following steps:  
      
    # 1. Build a subfolder for the artist  
    if not os.path.isdir(f"lyrics/{artist}"):   
        os.mkdir(f"lyrics/{artist}")  
      
    # 2. Iterate over the lyrics pages  
    for artist_lyric_page in lyrics_pages[artist][:20]:  
        print(artist_lyric_page)  
        title = artist_lyric_page.split()  
      
    # 3. Request the lyric page.  
    # Don't forget to add a line like 'time.sleep(5 + 10*random.random())'  
    # to sleep after making the request  
    r = requests.get(artist_lyric_page)  
    time.sleep(5 + 10*random.random())  
      
    # 4. Extract the title and lyrics from the page.  
    soup = BeautifulSoup(r.text, 'html.parser')  
    titles = []  
    for text in soup.find_all('b'):  
        if text and " ".join([ch for ch in text.string.lower() if ch.isalpha()]) in artist_lyric_page:  
            title = text.string.strip("\n")  
            print(f"Extract title: {title}")  
            titles.append(title)  
    lyrics = []  
    for divTag in soup.find_all("div", {"class": "col-xs-12 col-lg-8 text-center"}):  
        for tag in divTag.find_all("div", {"class": ""}):  
            lyric = tag.text.strip("\n")  
            if lyric:  
                print(f"Extract Lyric.")  
                lyrics.append(lyric)  
      
    # 5. Write out the title, two returns ('\n'), and the lyrics. Use 'generate_filename_from_url'  
    # to generate the filename  
    title = title_lyric in zip(titles, lyrics):  
        content = title+"\n"+"Lyric  
        output_file = generate_filename_from_link(artist_lyric_page)  
        with open(f"lyrics/{artist}/{output_file, 'w') as f:  
            f.write(content)  
            print(f"Finished writing the lyric of {title} by {artist}")  
      
    # Remember to pull at least 20 songs per artist. It may be fun to pull all the songs for the artist  
      
https://www.azlyrics.com/lyrics/rush/findingmymay.html  
Extract title: Finding My Way  
Extract lyric: Rush  
Extract lyric:  
Finished writing the lyric of Finding My Way by rush  
https://www.azlyrics.com/lyrics/rush/needsomeLove.html  
Extract title: Need Some Love  
Extract title: Rush  
Extract lyric:  
Finished writing the lyric of Need Some Love by rush  
https://www.azlyrics.com/lyrics/rush/takeafriend.html  
Extract title: Take A Friend  
Extract title: Rush  
Extract lyric:  
Finished writing the lyric of Take A Friend by rush  
https://www.azlyrics.com/lyrics/rush/herereagain.html  
Extract title: Here Again  
Extract title: Rush  
Extract lyric:  
Finished writing the lyric of Here Again by rush  
https://www.azlyrics.com/lyrics/rush/whatyouredoing.html  
Extract title: What You're Doing  
Extract title: Rush  
Extract lyric:  
Finished writing the lyric of What You're Doing by rush  
https://www.azlyrics.com/lyrics/rush/inthemood.html  
Extract title: In The Mood  
Extract title: Rush  
Extract lyric:  
Finished writing the lyric of In The Mood by rush  
https://www.azlyrics.com/lyrics/rush/beforeandafter.html  
Extract title: Before And After  
Extract title: Rush  
Extract lyric:  
Finished writing the lyric of Rush by rush  
https://www.azlyrics.com/lyrics/rush/workingman.html  
Extract title: Working Man  
Extract title: Rush  
Extract lyric:  
Finished writing the lyric of Working Man by rush  
https://www.azlyrics.com/lyrics/rush/anthem.html  
Extract title: Anthem  
Extract lyric:  
Finished writing the lyric of Anthem by rush  
https://www.azlyrics.com/lyrics/rush/bestican.html  
Extract title: Best I Can  
Extract lyric:  
Finished writing the lyric of Best I Can by rush  
https://www.azlyrics.com/lyrics/rush/beneathbetweenandbehind.html  
Extract title: Beneath, Between And Behind  
Extract title: By-Tor And The Snow Dog  
Extract lyric:  
Finished writing the lyric of By-Tor And The Snow Dog by rush  
https://www.azlyrics.com/lyrics/rush/flybynight.html  
Extract title: Fly By Night  
Extract title: Fly By Night  
Extract lyric:  
Finished writing the lyric of Fly By Night by rush  
https://www.azlyrics.com/lyrics/rush/makingmemories.html  
Extract title: Making Memories  
Extract lyric:  
Finished writing the lyric of Making Memories by rush  
https://www.azlyrics.com/lyrics/rush/rivendell.html  
Extract title: Rivendell  
Extract lyric:  
Finished writing the lyric of Rivendell by rush  
https://www.azlyrics.com/lyrics/rush/intheend.html  
Extract title: In The End  
Extract lyric:  
Finished writing the lyric of In The End by rush  
https://www.azlyrics.com/lyrics/rush/bastilleday.html  
Extract title: Bastille Day  
Extract lyric:  
Finished writing the lyric of Bastille Day by rush  
https://www.azlyrics.com/lyrics/rush/ithinkimgoingbald.html  
Extract title: I Think I'm Going Bald  
Extract lyric:  
Finished writing the lyric of I Think I'm Going Bald by rush  
https://www.azlyrics.com/lyrics/rush/lakesidepark.html  
Extract title: Lakeside Park  
Extract lyric:  
Finished writing the lyric of Lakeside Park by rush  
https://www.azlyrics.com/lyrics/rush/thenecromancer.html  
Extract title: The Necromancer  
Extract lyric:  
Finished writing the lyric of The Necromancer by rush  
https://www.azlyrics.com/lyrics/killers/jennywasafriendofmine.html  
Extract title: Jenny Was A Friend Of Mine  
Extract lyric:  
Finished writing the lyric of Jenny Was A Friend Of Mine by killers  
https://www.azlyrics.com/lyrics/killers/mrbrightside39068.html  
Extract title: Mr Brightside  
Extract lyric:  
Finished writing the lyric of Mr Brightside by killers  
https://www.azlyrics.com/lyrics/killers/callitkeyyoumeanit.html  
Extract title: Smile Like You Mean It  
Extract lyric:  
Finished writing the lyric of Smile Like You Mean It by killers  
https://www.azlyrics.com/lyrics/killers/somebodytoldme.html  
Extract title: Somebody Told Me  
Extract lyric:  
Finished writing the lyric of Somebody Told Me by killers  
https://www.azlyrics.com/lyrics/killers/allthesethingsthativedone.html  
Extract title: All These Things That I've Done  
Extract lyric:  
Finished writing the lyric of All These Things That I've Done by killers  
https://www.azlyrics.com/lyrics/killers/andyoureastar.html  
Extract title: Andy, You're A Star  
Extract lyric:  
Finished writing the lyric of Andy, You're A Star by killers  
https://www.azlyrics.com/lyrics/killers/ontop.html  
Extract title: On Top  
Extract lyric:  
Finished writing the lyric of On Top by killers  
https://www.azlyrics.com/lyrics/killers/changeyourmind.html  
Extract title: Change Your Mind  
Extract lyric:  
Finished writing the lyric of Change Your Mind by killers  
https://www.azlyrics.com/lyrics/killers/believesthatilbealright.html  
Extract title: Believe Me Natalie  
Extract lyric:  
Finished writing the lyric of Believe Me Natalie by killers  
https://www.azlyrics.com/lyrics/killers/midnightshow.html  
Extract title: Midnight Show  
Extract lyric:  
Finished writing the lyric of Midnight Show by killers  
https://www.azlyrics.com/lyrics/killers/everythingwillbealright.html  
Extract title: Everything Will Be Alright  
Extract lyric:  
Finished writing the lyric of Everything Will Be Alright by killers  
https://www.azlyrics.com/lyrics/killers/glamorousindierockroll.html  
Extract title: Glamorous Indie Rock & Roll  
Extract lyric:  
Finished writing the lyric of Glamorous Indie Rock & Roll by killers  
https://www.azlyrics.com/lyrics/killers/samstom.html  
Extract title: Sam's Town  
Extract lyric:  
Finished writing the lyric of Sam's Town by killers  
https://www.azlyrics.com/lyrics/killers/enterlude.html  
Extract title: Enterlude  
Extract lyric:  
Finished writing the lyric of Enterlude by killers  
https://www.azlyrics.com/lyrics/killers/whenyouwereyoung.html  
Extract title: When You Were Young  
Extract lyric:  
Finished writing the lyric of When You Were Young by killers  
https://www.azlyrics.com/lyrics/killers/blingconfessionofafaking.html  
Extract title: Bling (Confession Of A King)  
Extract lyric:  
Finished writing the lyric of Bling (Confession Of A King) by killers  
https://www.azlyrics.com/lyrics/killers/forreasonsunknown.html  
Extract title: For Reasons Unknown  
Extract lyric:  
Finished writing the lyric of For Reasons Unknown by killers  
https://www.azlyrics.com/lyrics/killers/readmymind.html  
Extract title: Read My Mind  
Extract lyric:  
Finished writing the lyric of Read My Mind by killers  
https://www.azlyrics.com/lyrics/killers/unclejonny.html  
Extract title: Uncle Jonny  
Extract lyric:  
Finished writing the lyric of Uncle Jonny by killers  
https://www.azlyrics.com/lyrics/killers/bones.html  
Extract title: Bones  
Extract lyric:  
Finished writing the lyric of Bones by killers  
      
In [62]: print(f"Total run time was {round((time.time() - start)/3600,2)} hours.")  
      
Total run time was 0.14 hours.
```

Evaluation

This assignment asks you to pull data from the Twitter API and scrape www.AZlyrics.com. After you have finished the above sections, run all the cells in this notebook. Print this to PDF and submit it, per the instructions.

```
In [63]: # Simple word extractor from Peter Norvig: https://norvig.com/spell-correct.html  
def   
    return re.findall(r'\w+', text.lower())
```

Checking Twitter Data

The output from your `twitter API` pull should be two files per artist, stored in files with formats like `cher_followers.txt` (a list of all follower IDs you pulled) and `cher_followers_data.txt`. These files should be in a folder named `twitter` within the repository directory. This code summarizes the information at a high level to help the instructor evaluate your work.

```
In [64]: twitter_files = os.listdir('twitter')  
twitter_files = [f for f in twitter_files if f != "_DS_Store"]  
artist_handles = list(set([name.split("_")[0] for name in twitter_files]))  
      
print(f"We see two artist handles: {artist_handles[0]} and {artist_handles[1]}.")  
      
We see two artist handles: thekillers and rushtheband.  
      
In [67]: for artist in artist_handles :  
    follower_file = artist + "_followers.txt"  
    follower_data_file = artist + "_follower_data.txt"  
      
    ids = open("twitter/" + follower_file, 'r').readlines()  
      
    print(f"We see {len(ids)-1} in your follower file for {artist}, assuming a header row.")  
    with open("twitter/" + follower_data_file, 'r') as infile :  
      
        # check the headers  
        headers = infile.readline().split("\t")  
        print(f"{'\n' if len(follower_data_file) != 0 else ''}In the follower data file ({follower_data_file}) for {artist}, we have these columns:")  
        print(" " + ".join(headers))  
          
        description_words = []  
        locations = set()  
          
        for idx, line in enumerate(infile.readlines()) :  
            line = line.strip("\n").split("\t")  
              
            try :  
                locations.add(line[3])  
                description_words.extend(words(line[6]))  
            except :  
                pass  
          
    print(f"We have {idx+1} data rows for {artist} in the follower data file.")  
    print(f"{'\n' if len(follower_data_file) != 0 else ''}For {artist} we have {len(locations)} unique locations.")  
    print(f"{'\n' if len(follower_data_file) != 0 else ''}For {artist} we have {len(description_words)} words in the descriptions.")  
    print(f"{'\n' if len(follower_data_file) != 0 else ''}Here are the five most common words:")  
    print(Counter(description_words).most_common(5))  
      
    print("")  
    print("-"*40)  
    print("")  
      
    # Section Complete, however, it doesnt return me any common words. I am unsure how to fix this as the moment  
    # everything appears to be working correctly.  
      
We see 59999 in your follower file for thekillers, assuming a header row.  
In the follower data file (thekillers_follower_data.txt) for thekillers, we have these columns:  
|  
      
We have 539965 data rows for thekillers in the follower data file.  
For thekillers we have 0 unique locations.  
For thekillers we have 0 words in the descriptions.  
Here are the five most common words:  
[]  
-----  
      
We see 59999 in your follower file for rushtheband, assuming a header row.  
In the follower data file (rushtheband_follower_data.txt) for rushtheband, we have these columns:  
|  
      
We have 539980 data rows for rushtheband in the follower data file.  
For rushtheband we have 0 unique locations.  
For rushtheband we have 0 words in the descriptions.  
Here are the five most common words:  
[]  
-----
```

Checking Lyrics

The output from your lyrics scrape should be stored in files located in this path from the directory: `./lyrics/[Artist Name]/[filename From URL]`. This code summarizes the information at a high level to help the instructor evaluate your work.

```
In [68]: artist_folders = os.listdir('lyrics/')  
artist_folders = [f for f in artist_folders if os.path.isdir("lyrics/" + f)]  
      
for artist in artist_folders :  
    artist_files = os.listdir(f"lyrics/" + artist)  
    artist_files = [f for f in artist_files if 'txt' in f or 'csv' in f or 'tsv' in f]  
      
    print(f"{'\n' if len(artist_files) != 0 else ''}For {artist} we have {len(artist_files)} files.")  
    artist_words = []  
      
    for f_name in artist_files :  
        with open("lyrics/" + artist + "/" + f_name, 'a') as infile :  
            artist_words.extend(words(infile.read()))  
      
    print(f"{'\n' if len(artist_files) != 0 else ''}For {artist} we have roughly {len(artist_words)} words, {len(set(artist_words))} are unique.")  
      
For killers we have 20 files.  
For killers we have roughly 5247 words, 771 are unique.  
For rush we have 20 files.  
For rush we have roughly 3562 words, 878 are unique.  
      
In [69]: # End of Homework.
```

```
In [ ] : 
```