ADS 509 Module 1: APIs and Web Scraping This notebook has three parts. In the first part you will pull data from the Twitter API. In the second, you will scrape lyrics from AZLyrics.com. In the last part, you'll run code that verifies the completeness of your data pull. For this assignment you have chosen two musical artists who have at least 100,000 Twitter followers and 20 songs with lyrics on AZLyrics.com. In this part of the assignment we pull the some of the user information for the followers of your artist and store them in text files. **General Assignment Instructions** These instructions are included in every assignment, to remind you of the coding standards for the class. Feel free to delete this cell after reading it. One sign of mature code is conforming to a style guide. We recommend the Google Python Style Guide. If you use a different style guide, please include a cell with a link. Your code should be relatively easy-to-read, sensibly commented, and clean. Writing code is a messy process, so please be sure to edit your final submission. Remove any cells that are not needed or parts of cells that contain unnecessary code. Remove inessential import statements and make sure that all such statements are moved into the designated cell. Make use of non-code cells for written commentary. These cells should be grammatical and clearly written. In some of these cells you will have questions to answer. The questions will be marked by a "Q:" and will have a corresponding "A:" spot for you. Make sure to answer every question marked with a Q: for full credit. Twitter API Pull In [23]: # pip install tweepy In [24]: # for the twitter section import tweepy import os import datetime import re from pprint import pprint # for the lyrics scrape section import requests import time from bs4 import BeautifulSoup from collections import defaultdict, Counter In [59]: import seaborn as sns import pandas as pd import numpy as np import os from scipy import stats import matplotlib as mpl import matplotlib.pyplot as plt import random import requests from bs4 import BeautifulSoup import shutil We need bring in our API keys. Since API keys should be kept secret, we'll keep them in a file called api_keys.py . This file should be stored in the directory where you store this notebook. The example file is provided for you on Blackboard. The example has API keys that are not functional, so you'll need to get Twitter credentials and replace the placeholder keys. from api keys import api key, api key secret, access token, access token secret In [27]: auth = tweepy.AppAuthHandler(api_key, api_key_secret) api = tweepy.API(auth, wait_on_rate_limit=True) print ('API Host', api.host) #print ('API Version', api.api_root) API Host api.twitter.com # dir(api) Testing the API The Twitter APIs are quite rich. Let's play around with some of the features before we dive into this section of the assignment. For our testing, it's convenient to have a small data set to play with. We will seed the code with the handle of John Chandler, one of the instructors in this course. His handle is @37chandler. Feel free to use a different handle if you would like to look at someone else's data. We will write code to explore a few aspects of the API: 1. Pull all the follower IDs for @katymck. 2. Explore the user object, which gives us information about Twitter users. 3. Pull some user objects for the followers. 4. Pull the last few tweets by @katymck. # I had to change api.followers_ids to api.get_followers_ids since it returned the error that it had no foll handle = "37chandler" followers = [] for page in tweepy.Cursor(api.get_follower_ids, # This is how we will get around the issue of not being able to grab all ids at on # Once the rate limit is hit, we will be that we must wait 15 mins (900 secs) wait_on_rate_limit=True, wait_on_rate_limit_notify=True, compression=True, screen_name=handle).pages(): # The page variable comes back as a list, so we have to use .extend rather than .append followers.extend(page) print(f"Here are the first five follower ids for {handle} out of the {len(followers)} total.") followers[:5] # Section Complete Here are the first five follower ids for 37chandler out of the 189 total. Out[29]: [257285645, 1469785454576820225, 1181131341687066624, 257686741, 2306579816] We have the follower IDs, which are unique numbers identifying the user, but we'd like to get some more information on these users. Twitter allows us to pull "fully hydrated user objects", which is a fancy way of saying "all the information about the user". Let's look at user object for our starting handle. user = api.get_user(user_id=8813602) print(user. json) # Section Complete # I needed to do user id = because it no longer displays the users, it returns an error that it expects 1 bu {'id': 8813602, 'id str': '8813602', 'name': 'Stacey Burns', 'screen name': 'WentRogue', 'location': 'Minnea polis Witch District', 'profile_location': None, 'description': 'Once inspired the NYT Science Times to "try harder." And now I tweet about my cats. Cis; she/her. I used to be someone's big sister. I am still not O K.', 'url': 'https://t.co/CgQ2gGvTks', 'entities': {'url': {'urls': [{'url': 'https://t.co/CgQ2gGvTks', 'exp anded url': 'https://unrestrictmn.org/petition-reproductive-freedom/?ms=SBtwitter', 'display_url': 'unrestri ctmn.org/petition-repro...', 'indices': [0, 23]}]}, 'description': {'urls': []}}, 'protected': False, 'followe rs count': 10554, 'friends count': 4453, 'listed count': 278, 'created at': 'Tue Sep 11 16:25:31 +0000 200 7', 'favourites count': 141964, 'utc offset': None, 'time zone': None, 'geo enabled': False, 'verified': Tru e, 'statuses count': 184673, 'lang': None, 'status': {'created_at': 'Mon May 16 21:42:28 +0000 2022', 'id': 1526317181767667717, 'id str': '1526317181767667717', 'text': '@chris steller What a photo!!!', 'truncated': False, 'entities': {'hashtags': [], 'symbols': [], 'user_mentions': [{'screen_name': 'chris_steller', 'nam e': 'Chris Steller', 'id': 14812037, 'id_str': '14812037', 'indices': [0, 14]}], 'urls': []}, 'source': 'Twitter for iPhone', 'in_reply_to_status_id': 1 526316858340782080, 'in_reply_to_status_id_str': '1526316858340782080', 'in_reply_to_user_id': 14812037, 'in reply to user id str': '14812037', 'in reply to screen name': 'chris steller', 'geo': None, 'coordinates': None, 'place': None, 'contributors': None, 'is_quote_status': False, 'retweet_count': 0, 'favorite_count': 2, 'favorited': False, 'retweeted': False, 'lang': 'en'}, 'contributors enabled': False, 'is translator': Fa lse, 'is_translation_enabled': False, 'profile_background_color': 'E9E8EB', 'profile_background_image_url': 'http://abs.twimg.com/images/themes/theme5/bg.gif', 'profile background image url https': 'https://abs.twim g.com/images/themes/theme5/bg.gif', 'profile_background_tile': False, 'profile_image_url': 'http://pbs.twim g.com/profile_images/1511695479217786886/BkdOZPwV_normal.jpg', 'profile_image_url_https': 'https://pbs.twim g.com/profile_images/1511695479217786886/BkdOZPwV_normal.jpg', 'profile_banner_url': 'https://pbs.twimg.com/ profile banners/8813602/1618767093', 'profile link color': 'FF0000', 'profile sidebar border color': '00000 0', 'profile sidebar fill color': 'E9E8EB', 'profile text color': 'OC3E53', 'profile use background image': True, 'has_extended_profile': True, 'default_profile': False, 'default_profile_image': False, 'following': N one, 'follow_request_sent': None, 'notifications': None, 'translator_type': 'none', 'withheld_in_countries': print(f"{len(user. json.keys())} individual fields are being returned in the json portion of the user objection 45 individual fields are being returned in the _json portion of the user object Now a few questions for you about the user object. Q: How many fields are being returned in the _json portion of the user object? A: There are 45 fields being returned, they can be seen above for the user I chose from the list of followers. Q: Are any of the fields within the user object non-scaler? TK correct term A: I am not sure what TK correct term means, but there appears to be non-scaler fields that are not integers or other scaler fields. This can be seen in the above listing. Q: How many friends, followers, favorites, and statuses does this user have? A: Friends: 4434, Followers: 10523, Favorites: 141585, Statuses: 184339 **This section seems to keep changing when I run each day. We can map the follower IDs onto screen names by accessing the screen_name key within the user object. Modify the code below to also print out how many people the follower is following and how many followers they have. In [32]: # I had to add in user id= to be able to run properly, update from tweepy ids to lookup = followers[:10] for user obj in api.lookup users(user id=ids to lookup) : print(f"{handle} is followed by {user obj.screen name}") # Add code here to print out friends and followers of `handle` print(f"{user obj.screen name} is following {user_obj._json['friends_count']} people") print(f"{user obj.screen name} has {user obj.json['followers count']} followers\n") # Section Complete 37chandler is followed by HicSvntDraconez HicSvntDraconez is following 1546 people HicSvntDraconez has 33 followers 37chandler is followed by JohnOCo70713197 JohnOCo70713197 is following 8 people JohnOCo70713197 has 1 followers 37chandler is followed by CodeGradeCom CodeGradeCom is following 2711 people CodeGradeCom has 388 followers 37chandler is followed by cleverhoods cleverhoods is following 2774 people cleverhoods has 3371 followers 37chandler is followed by PaulNaish78 PaulNaish78 is following 19266 people PaulNaish78 has 19524 followers 37chandler is followed by mplsFietser mplsFietser is following 2651 people mplsFietser has 2783 followers 37chandler is followed by echallstrom echallstrom is following 457 people echallstrom has 304 followers 37chandler is followed by byler_t117 byler t117 is following 440 people byler_t117 has 48 followers 37chandler is followed by Community Owner Community Owner is following 47 people Community_Owner has 31 followers 37chandler is followed by DeepakC64237257 DeepakC64237257 is following 574 people DeepakC64237257 has 31 followers Although you won't need it for this assignment, individual tweets (called "statuses" in the API) can be a rich source of text-based data. To illustrate the concepts, let's look at the last few tweets for this user. You are encouraged to explore the status object and marvel in the richness of the data that is available. tweet_count = 0 for status in tweepy.Cursor(api.user_timeline, screen_name=handle).items(): tweet_count += 1 print(f"The tweet was tweeted at {status.created_at}.") print(f"The original tweet has been retweeted {status.retweet_count} times.") clean_status = status.text clean_status = clean_status.replace("\n"," ") print(f"{clean status}") print("\n"*2) if tweet_count > 10 : break #Section Complete The tweet was tweeted at 2022-05-13 23:46:46+00:00. The original tweet has been retweeted 5 times. RT @johnhollinger: @NateSilver538 Atlanta still leads the nation in "further West than you think" The tweet was tweeted at 2022-05-13 12:03:18+00:00. The original tweet has been retweeted 1478 times. RT @tomscocca: It was helpful to talk with @pareene about the experience of turning away from one of the mos t comfortable default beliefs o... The tweet was tweeted at 2022-05-12 14:41:18+00:00. The original tweet has been retweeted 187 times. RT @JamesTateHill: I Actually Thought You Had Dropped the Class: A Memoir of Your Final Grade The tweet was tweeted at 2022-05-12 12:19:43+00:00. The original tweet has been retweeted 0 times. @BluehairCoffee @WedgeLIVE I try to always take a pic of it. https://t.co/7i4nIgBKFM The tweet was tweeted at 2022-05-12 03:18:56+00:00. The original tweet has been retweeted 0 times. @LiberalwKnives @WedgeLIVE @BluehairCoffee It seemed deep enough to total some cars at 8:45. Also, 🦓, neigh The tweet was tweeted at 2022-05-12 03:16:18+00:00. The original tweet has been retweeted 0 times. ${\tt @WedgeLIVE~2200~block~of~Garfield.~aka,~the~former~Lake~Blaisdell.~https://t.co/SLmOzjjWUw}$ The tweet was tweeted at 2022-05-11 00:47:28+00:00. The original tweet has been retweeted 412 times. RT @ThePlumLineGS: Terrific @Milbank piece vividly highlighting the long trail of lying, deception, norm-shr edding, and all around bad-acti ... The tweet was tweeted at 2022-05-08 03:20:25+00:00. The original tweet has been retweeted 0 times. @WedgeLIVE Had such a glorious cross today. First time in 10 years. Excited for our new weapons in the @TheW arOnCars The tweet was tweeted at 2022-05-05 20:23:42+00:00. The original tweet has been retweeted 120 times. RT @tomtomorrow: I don't know how to explain to you that you should care that we are two years away from fas cist theocracy. The tweet was tweeted at 2022-05-03 18:34:51+00:00. The original tweet has been retweeted 0 times. This whole **Z** https://t.co/YHVewCyhJy The tweet was tweeted at 2022-05-03 12:02:15+00:00. The original tweet has been retweeted 1334 times. RT @QasimRashid: Alito claims Roe V Wade "was wrongly decided because Constitution makes no specific mention of abortion rights." Constitu... **Pulling Follower Information** In this next section of the assignment, we will pull information about the followers of your two artists. We must first get the follower IDs, then we will be able to "hydrate" the IDs, pulling the user objects for them. Once we have those user objects we will extract some fields that we can use in future analyses. The Twitter API only allows users to make 15 requests per 15 minutes when pulling followers. Each request allows you to gather 5000 follower ids. Tweepy will grab the 15 requests quickly then wait 15 minutes, rather than slowly pull the requests over the time period. Before we start grabbing follower IDs, let's first just check how long it would take to pull all of the followers. To do this we use the followers_count item from the user object. In [34]: # I needed to update the screen name= in order to run, update to tweepy # I'm also putting the handles in a list to iterate through below handles = ['rushtheband','thekillers'] # This will iterate through each Twitter handle that we're collecting from for screen name in handles: # Tells Tweepy we want information on the handle we're collecting from # The next line specifies which information we want, which in this case is the number of followers user = api.get user(screen name=screen name) followers_count = user.followers_count # Let's see roughly how long it will take to grab all the follower IDs. print(f''' @{screen name} has {followers count} followers. That will take roughly $\{followers count/(5000*15*4):.2f\}$ hours to pull the followers. **'''**) # Section Complete @rushtheband has 302858 followers. That will take roughly 1.01 hours to pull the followers. @thekillers has 4781598 followers. That will take roughly 15.94 hours to pull the followers. As we pull data for each artist we will write their data to a folder called "twitter", so we will make that folder if needed. # Make the "twitter" folder here. If you'd like to practice your programming, add functionality # that checks to see if the folder exists. If it does, then "unlink" it. Then create a new one. if not os.path.isdir("twitter") : #shutil.rmtree("twitter/") os.mkdir("twitter") # Section Complete In this following cells, use the api.followers_ids (and the tweepy.Cursor functionality) to pull some of the followers for your two artists. As you pull the data, write the follower ids to a file called [artist name]_followers.txt in the "twitter" folder. For instance, for Cher I would create a file named cher_followers.txt. As you pull the data, also store it in an object like a list or a data frame. num followers to pull = 60*1000 # feel free to use this to limit the number of followers you pull. In [37]: # Modify the below code stub to pull the follower IDs and write them to a file. # Grabs the time when we start making requests to the API start time = datetime.datetime.now() for idx, handle in enumerate(handles): output file = "./twitter/"+handle + " followers.txt" user = api.get user(screen name=handle) followers_count = user.followers_count # Pull and store the follower IDs followers = [] for page in tweepy.Cursor(api.get follower ids, wait on rate limit=True, wait on rate limit notify=True, compression=True, count = int(num_followers_to_pull/15), screen_name=handle).pages(): followers.extend(page) # If you've pulled num_followers_to_pull, feel free to break out paged twitter API response if len(followers) == num_followers_to_pull: break print(f"Pulled {len(followers)} followers of {handle}") # Write the IDs to the output file in the `twitter` folder. with open(output_file,'w') as f: for follower in followers: f.write(str(follower) + "\n") print(f"Wrote {len(followers)} followers of {handle} to file: {output file}") # Sleep before the next pull if idx != len(handles)-1: time.sleep(60*15) # Let's see how long it took to grab all follower IDs end_time = datetime.datetime.now() print(end_time - start_time) # Section Complete Rate limit reached. Sleeping for: 546 Pulled 60000 followers of rushtheband Wrote 60000 followers of rushtheband to file: ./twitter/rushtheband_followers.txt Pulled 60000 followers of thekillers Wrote 60000 followers of thekillers to file: ./twitter/thekillers_followers.txt 0:24:16.508331 Now that you have your follower ids, gather some information that we can use in future assignments on them. Using the lookup_users function, pull the user objects for your followers. These requests are limited to 900 per 15 minutes, but you can request 100 users at a time. At 90,000 users per 15 minutes, the rate limiter on pulls might be bandwidth rather than API limits. Extract the following fields from the user object: screen_name name id location followers_count friends_count description These can all be accessed via these names in the object. Store the fields with one user per row in a tab-delimited text file with the name [artist name]_follower_data.txt . For instance, for Cher I would create a file named cher_follower_data.txt . handles = ['rushtheband','thekillers'] # in this cell, do the following In [42]: # 1. Set up a data frame or dictionary to hold the user information # 2. Use the `lookup_users` api function to pull sets of 100 users at a time # 3. Store the listed fields in your data frame or dictionary. # 4. Write the user information in tab-delimited form to the follower data text file. import json for handle in handles: input_file = "./twitter/"+handle + "_followers.txt" followers_id = [int(x) for x in open(input_file,'r').read().split('\n') if x] users_count = len(followers_id) #1 Set up Dataframe: my_list_of_dicts = [] try: print(f"Getting user information for followers of {handle}") # 2. Use the `lookup_users` api function to pull sets of 100 users at a time for i in range(int(users_count / 100) + 1): each_json_tweet = api.lookup_users(user_id=followers_id[i*100:min((i+1)*100, users_count)]) # 3. Store the listed fields in your data frame or dictionary. fields = ['screen_name', 'name', 'id', 'location', 'followers_count', 'friends_count', 'description'] user_data = [{key: x._json[key] for key in fields} for x in each_json_tweet] print(user_data) my_list_of_dicts.extend(user_data) print('getting users batch:', i) except tweepy.errors.TweepyException as e: print('Something went wrong, quitting...', e) time.sleep(60*15) output_file = "./twitter/"+handle +"_follower_data.txt" # 4. Write the user information in tab-delimited form to the follower data text file. with open(output_file, 'w') as file: file.write(json.dumps(my_list_of_dicts, indent=4)) # Section Complete Getting user information for followers of rushtheband Something went wrong, quitting... 404 Not Found 17 - No user matches for specified terms. Getting user information for followers of thekillers Something went wrong, quitting... 404 Not Found 17 - No user matches for specified terms. One note: the user's description can have tabs or returns in it, so make sure to clean those out of the description before writing them to the file. Here's an example of how you might do this. tricky_description = """ In [43]: Home by Warsan Shire no one leaves home unless home is the mouth of a shark. you only run for the border when you see the whole city running as well. # This won't work in a tab-delimited text file. clean_description = re.sub(r"\s+"," ",tricky_description) clean_description Out[43]: ' Home by Warsan Shire no one leaves home unless home is the mouth of a shark. you only run for the border w hen you see the whole city running as well. ' Lyrics Scrape This section asks you to pull data from the Twitter API and scrape www.AZLyrics.com. In the notebooks where you do that work you are asked to store the data in specific ways. artists = {'rush':"https://www.azlyrics.com/r/rush.html", 'killers':"https://www.azlyrics.com/k/killers.html"} # we'll use this dictionary to hold both the artist name and the link on AZlyrics # Section Complete A Note on Rate Limiting The lyrics site, www.azlyrics.com, does not have an explicit maximum on number of requests in any one time, but in our testing it appears that too many requests in too short a time will cause the site to stop returning lyrics pages. (Entertainingly, the page that gets returned seems to only have the song title to a Tom Jones song.) Whenever you call requests.get to retrieve a page, put a time.sleep(5 + 10*random.random()) on the next line. This will help you not to get blocked. If you do get blocked, which you can identify if the returned pages are not correct, just request a lyrics page through your browser. You'll be asked to perform a CAPTCHA and then your requests should start working again. Part 1: Finding Links to Songs Lyrics That general artist page has a list of all songs for that artist with links to the individual song pages. Q: Take a look at the robots.txt page on www.azlyrics.com. (You can read more about these pages here.) Is the scraping we are about to do allowed or disallowed by this page? How do you know? A: yes it appears that the scraping we are doing is allowed since I have the outputs of the song title and lyrics. # Let's set up a dictionary of lists to hold our links lyrics pages = defaultdict(list) for artist, artist page in artists.items(): # request the page and sleep r = requests.get(artist page) time.sleep(5 + 10*random.random()) # now extract the links to lyrics pages from this page # store the links `lyrics pages` where the key is the artist and the # value is a list of links. soup = BeautifulSoup(r.text, 'html.parser') urls = [] for link in soup.find all('a'): url = link.get('href') if url and (artist in url) and ('/lyrics/' in url): full url = 'https://www.azlyrics.com'+link.get('href') lyrics pages[artist].append(full url) # Section Complete In [54]: lyrics pages Out[54]: defaultdict(list, {'rush': ['https://www.azlyrics.com/lyrics/rush/findingmyway.html', 'https://www.azlyrics.com/lyrics/rush/needsomelove.html', 'https://www.azlyrics.com/lyrics/rush/takeafriend.html', 'https://www.azlyrics.com/lyrics/rush/hereagain.html', 'https://www.azlyrics.com/lyrics/rush/whatyouredoing.html', 'https://www.azlyrics.com/lyrics/rush/inthemood.html', 'https://www.azlyrics.com/lyrics/rush/beforeandafter.html', 'https://www.azlyrics.com/lyrics/rush/workingman.html', 'https://www.azlyrics.com/lyrics/rush/anthem.html', 'https://www.azlyrics.com/lyrics/rush/bestican.html', 'https://www.azlyrics.com/lyrics/rush/beneathbetweenandbehind.html', 'https://www.azlyrics.com/lyrics/rush/bytorandthesnowdog.html', 'https://www.azlyrics.com/lyrics/rush/flybynight.html', 'https://www.azlyrics.com/lyrics/rush/makingmemories.html', 'https://www.azlyrics.com/lyrics/rush/rivendell.html', 'https://www.azlyrics.com/lyrics/rush/intheend.html', 'https://www.azlyrics.com/lyrics/rush/bastilleday.html', 'https://www.azlyrics.com/lyrics/rush/ithinkimgoingbald.html', 'https://www.azlyrics.com/lyrics/rush/lakesidepark.html', 'https://www.azlyrics.com/lyrics/rush/thenecromancer.html', 'https://www.azlyrics.com/lyrics/rush/thefountainoflamneth.html', 'https://www.azlyrics.com/lyrics/rush/2112.html', 'https://www.azlyrics.com/lyrics/rush/apassagetobangkok.html', 'https://www.azlyrics.com/lyrics/rush/thetwilightzone.html', 'https://www.azlyrics.com/lyrics/rush/lessons.html', 'https://www.azlyrics.com/lyrics/rush/tears.html', 'https://www.azlyrics.com/lyrics/rush/somethingfornothing.html', 'https://www.azlyrics.com/lyrics/rush/afarewelltokings.html', 'https://www.azlyrics.com/lyrics/rush/xanadu.html', 'https://www.azlyrics.com/lyrics/rush/closertotheheart.html', 'https://www.azlyrics.com/lyrics/rush/cinderellaman.html', 'https://www.azlyrics.com/lyrics/rush/madrigal.html', 'https://www.azlyrics.com/lyrics/rush/cygnusx1bookonethevoyage.html', 'https://www.azlyrics.com/lyrics/rush/cygnusx1bookiihemispheres.html', 'https://www.azlyrics.com/lyrics/rush/circumstances.html', 'https://www.azlyrics.com/lyrics/rush/thetrees.html', 'https://www.azlyrics.com/lyrics/rush/thespiritofradio.html', 'https://www.azlyrics.com/lyrics/rush/freewill.html', 'https://www.azlyrics.com/lyrics/rush/jacobsladder.html', 'https://www.azlyrics.com/lyrics/rush/entrenous.html', 'https://www.azlyrics.com/lyrics/rush/differentstrings.html', 'https://www.azlyrics.com/lyrics/rush/naturalscience.html', 'https://www.azlyrics.com/lyrics/rush/tomsawyer.html', 'https://www.azlyrics.com/lyrics/rush/redbarchetta.html', 'https://www.azlyrics.com/lyrics/rush/limelight.html', 'https://www.azlyrics.com/lyrics/rush/thecameraeye.html', 'https://www.azlyrics.com/lyrics/rush/witchhunt.html', 'https://www.azlyrics.com/lyrics/rush/vitalsigns.html' 'https://www.azlyrics.com/lyrics/rush/subdivisions.html', 'https://www.azlyrics.com/lyrics/rush/theanalogkid.html', 'https://www.azlyrics.com/lyrics/rush/chemistry.html', 'https://www.azlyrics.com/lyrics/rush/digitalman.html', 'https://www.azlyrics.com/lyrics/rush/theweapon.html', 'https://www.azlyrics.com/lyrics/rush/newworldman.html', 'https://www.azlyrics.com/lyrics/rush/losingit.html', 'https://www.azlyrics.com/lyrics/rush/countdown.html', 'https://www.azlyrics.com/lyrics/rush/distantearlywarning.html', 'https://www.azlyrics.com/lyrics/rush/afterimage.html', 'https://www.azlyrics.com/lyrics/rush/redsectora.html', 'https://www.azlyrics.com/lyrics/rush/theenemywithin.html', 'https://www.azlyrics.com/lyrics/rush/thebodyelectric.html', 'https://www.azlyrics.com/lyrics/rush/kidgloves.html', 'https://www.azlyrics.com/lyrics/rush/redlenses.html', 'https://www.azlyrics.com/lyrics/rush/betweenthewheels.html', 'https://www.azlyrics.com/lyrics/rush/thebigmoney.html', 'https://www.azlyrics.com/lyrics/rush/granddesigns.html', 'https://www.azlyrics.com/lyrics/rush/manhattanproject.html', 'https://www.azlyrics.com/lyrics/rush/marathon.html', 'https://www.azlyrics.com/lyrics/rush/territories.html', 'https://www.azlyrics.com/lyrics/rush/middletowndreams.html', 'https://www.azlyrics.com/lyrics/rush/emotiondetector.html', 'https://www.azlyrics.com/lyrics/rush/mysticrhythms.html', 'https://www.azlyrics.com/lyrics/rush/forceten.html', 'https://www.azlyrics.com/lyrics/rush/timestandstill.html', 'https://www.azlyrics.com/lyrics/rush/opensecrets.html', 'https://www.azlyrics.com/lyrics/rush/secondnature.html', 'https://www.azlyrics.com/lyrics/rush/primemover.html', 'https://www.azlyrics.com/lyrics/rush/lockandkey.html', 'https://www.azlyrics.com/lyrics/rush/mission.html', 'https://www.azlyrics.com/lyrics/rush/turnthepage.html', 'https://www.azlyrics.com/lyrics/rush/taishan.html', 'https://www.azlyrics.com/lyrics/rush/highwater.html', 'https://www.azlyrics.com/lyrics/rush/showdonttell.html', 'https://www.azlyrics.com/lyrics/rush/chainlightning.html', 'https://www.azlyrics.com/lyrics/rush/thepass.html' 'https://www.azlyrics.com/lyrics/rush/warpaint.html', 'https://www.azlyrics.com/lyrics/rush/scars.html', 'https://www.azlyrics.com/lyrics/rush/presto.html', 'https://www.azlyrics.com/lyrics/rush/superconductor.html', 'https://www.azlyrics.com/lyrics/rush/anagramformongo.html', 'https://www.azlyrics.com/lyrics/rush/redtide.html', 'https://www.azlyrics.com/lyrics/rush/handoverfist.html', 'https://www.azlyrics.com/lyrics/rush/availablelight.html', 'https://www.azlyrics.com/lyrics/rush/dreamline.html', 'https://www.azlyrics.com/lyrics/rush/bravado.html', 'https://www.azlyrics.com/lyrics/rush/rollthebones.html', 'https://www.azlyrics.com/lyrics/rush/faceup.html', 'https://www.azlyrics.com/lyrics/rush/thebigwheel.html', 'https://www.azlyrics.com/lyrics/rush/heresy.html', 'https://www.azlyrics.com/lyrics/rush/ghostofachance.html', 'https://www.azlyrics.com/lyrics/rush/neurotica.html', 'https://www.azlyrics.com/lyrics/rush/youbetyourlife.html', 'https://www.azlyrics.com/lyrics/rush/animate.html', 'https://www.azlyrics.com/lyrics/rush/stickitout.html', 'https://www.azlyrics.com/lyrics/rush/cuttothechase.html', 'https://www.azlyrics.com/lyrics/rush/nobodyshero.html', 'https://www.azlyrics.com/lyrics/rush/betweensunandmoon.html', 'https://www.azlyrics.com/lyrics/rush/alienshore.html', 'https://www.azlyrics.com/lyrics/rush/thespeedoflove.html', 'https://www.azlyrics.com/lyrics/rush/doubleagent.html', 'https://www.azlyrics.com/lyrics/rush/coldfire.html', 'https://www.azlyrics.com/lyrics/rush/everydayglory.html', 'https://www.azlyrics.com/lyrics/rush/testforecho.html', 'https://www.azlyrics.com/lyrics/rush/driven.html', 'https://www.azlyrics.com/lyrics/rush/halftheworld.html', 'https://www.azlyrics.com/lyrics/rush/thecolorofright.html', 'https://www.azlyrics.com/lyrics/rush/timeandmotion.html', 'https://www.azlyrics.com/lyrics/rush/totem.html', 'https://www.azlyrics.com/lyrics/rush/dogyears.html' 'https://www.azlyrics.com/lyrics/rush/virtuality.html', 'https://www.azlyrics.com/lyrics/rush/resist.html', 'https://www.azlyrics.com/lyrics/rush/carveawaythestone.html', 'https://www.azlyrics.com/lyrics/rush/onelittlevictory.html', 'https://www.azlyrics.com/lyrics/rush/ceilingunlimited.html', 'https://www.azlyrics.com/lyrics/rush/ghostrider.html', 'https://www.azlyrics.com/lyrics/rush/peaceablekingdom.html', 'https://www.azlyrics.com/lyrics/rush/thestarslookdown.html', 'https://www.azlyrics.com/lyrics/rush/howitis.html', 'https://www.azlyrics.com/lyrics/rush/vaportrail.html', 'https://www.azlyrics.com/lyrics/rush/secrettouch.html', 'https://www.azlyrics.com/lyrics/rush/earthshine.html', 'https://www.azlyrics.com/lyrics/rush/sweetmiracle.html', 'https://www.azlyrics.com/lyrics/rush/nocturne.html', 'https://www.azlyrics.com/lyrics/rush/freeze.html', 'https://www.azlyrics.com/lyrics/rush/outofthecradle.html', 'https://www.azlyrics.com/lyrics/rush/summertimeblues.html' 'https://www.azlyrics.com/lyrics/rush/heartfullofsoul.html', 'https://www.azlyrics.com/lyrics/rush/theseeker.html', 'https://www.azlyrics.com/lyrics/rush/forwhatitsworth.html', 'https://www.azlyrics.com/lyrics/rush/shapesofthings.html', 'https://www.azlyrics.com/lyrics/rush/mrsoul.html', 'https://www.azlyrics.com/lyrics/rush/crossroads.html', 'https://www.azlyrics.com/lyrics/rush/sevenandsevenis.html', 'https://www.azlyrics.com/lyrics/rush/farcry.html', 'https://www.azlyrics.com/lyrics/rush/armorandsword.html', 'https://www.azlyrics.com/lyrics/rush/workinthemangels.html', 'https://www.azlyrics.com/lyrics/rush/thelargerbowlapantoum.html', 'https://www.azlyrics.com/lyrics/rush/spindrift.html', 'https://www.azlyrics.com/lyrics/rush/thewaythewindblows.html', 'https://www.azlyrics.com/lyrics/rush/faithless.html', 'https://www.azlyrics.com/lyrics/rush/bravestface.html' 'https://www.azlyrics.com/lyrics/rush/goodnewsfirst.html', 'https://www.azlyrics.com/lyrics/rush/malignantnarcissism.html', 'https://www.azlyrics.com/lyrics/rush/weholdon.html', 'https://www.azlyrics.com/lyrics/rush/caravan.html', 'https://www.azlyrics.com/lyrics/rush/bu2b.html', 'https://www.azlyrics.com/lyrics/rush/clockworkangels.html', 'https://www.azlyrics.com/lyrics/rush/theanarchist.html', 'https://www.azlyrics.com/lyrics/rush/carnies.html', 'https://www.azlyrics.com/lyrics/rush/haloeffect.html', 'https://www.azlyrics.com/lyrics/rush/sevencitiesofgold.html', 'https://www.azlyrics.com/lyrics/rush/thewreckers.html', 'https://www.azlyrics.com/lyrics/rush/headlongflight.html', 'https://www.azlyrics.com/lyrics/rush/bu2b2.html', 'https://www.azlyrics.com/lyrics/rush/wishthemwell.html', 'https://www.azlyrics.com/lyrics/rush/thegarden.html', 'https://www.azlyrics.com/lyrics/rush/badboy.html', 'https://www.azlyrics.com/lyrics/rush/fancydancer.html', 'https://www.azlyrics.com/lyrics/rush/gardenroad.html' 'https://www.azlyrics.com/lyrics/rush/notfadeaway.html', 'https://www.azlyrics.com/lyrics/rush/youcantfightit.html'], 'killers': ['https://www.azlyrics.com/lyrics/killers/jennywasafriendofmine.html', 'https://www.azlyrics.com/lyrics/killers/mrbrightside39068.html', 'https://www.azlyrics.com/lyrics/killers/smilelikeyoumeanit.html', 'https://www.azlyrics.com/lyrics/killers/somebodytoldme.html', 'https://www.azlyrics.com/lyrics/killers/allthesethingsthativedone.html', 'https://www.azlyrics.com/lyrics/killers/andyyoureastar.html', 'https://www.azlyrics.com/lyrics/killers/ontop.html', 'https://www.azlyrics.com/lyrics/killers/changeyourmind.html', 'https://www.azlyrics.com/lyrics/killers/believemenatalie.html', 'https://www.azlyrics.com/lyrics/killers/midnightshow.html', 'https://www.azlyrics.com/lyrics/killers/everythingwillbealright.html', 'https://www.azlyrics.com/lyrics/killers/glamorousindierockroll.html', 'https://www.azlyrics.com/lyrics/killers/samstown.html', 'https://www.azlyrics.com/lyrics/killers/enterlude.html', 'https://www.azlyrics.com/lyrics/killers/whenyouwereyoung.html', 'https://www.azlyrics.com/lyrics/killers/blingconfessionofaking.html', 'https://www.azlyrics.com/lyrics/killers/forreasonsunknown.html', 'https://www.azlyrics.com/lyrics/killers/readmymind.html', 'https://www.azlyrics.com/lyrics/killers/unclejonny.html', 'https://www.azlyrics.com/lyrics/killers/bones.html', 'https://www.azlyrics.com/lyrics/killers/mylist.html' 'https://www.azlyrics.com/lyrics/killers/thisriveriswild.html', 'https://www.azlyrics.com/lyrics/killers/whydoikeepcounting.html', 'https://www.azlyrics.com/lyrics/killers/exitlude.html', 'https://www.azlyrics.com/lyrics/killers/wherethewhiteboysdance.html', 'https://www.azlyrics.com/lyrics/killers/alltheprettyfaces.html', 'https://www.azlyrics.com/lyrics/killers/daddyseyes.html', 'https://www.azlyrics.com/lyrics/killers/tranquilize.html', 'https://www.azlyrics.com/lyrics/killers/shadowplay.html', 'https://www.azlyrics.com/lyrics/killers/alltheprettyfaces.html', 'https://www.azlyrics.com/lyrics/killers/leavethebourbonontheshelf.html', 'https://www.azlyrics.com/lyrics/killers/sweettalk.html', 'https://www.azlyrics.com/lyrics/killers/underthegun.html', 'https://www.azlyrics.com/lyrics/killers/wherethewhiteboysdance.html', 'https://www.azlyrics.com/lyrics/killers/showyouhow.html', 'https://www.azlyrics.com/lyrics/killers/moveaway.html', 'https://www.azlyrics.com/lyrics/killers/glamorousindierockroll.html', 'https://www.azlyrics.com/lyrics/killers/wholetyougo.html', 'https://www.azlyrics.com/lyrics/killers/theballadofmichaelvalentine.html', 'https://www.azlyrics.com/lyrics/killers/rubydonttakeyourlovetotown.html', 'https://www.azlyrics.com/lyrics/killers/daddyseyes.html', 'https://www.azlyrics.com/lyrics/killers/samstownabbeyroadversion.html', 'https://www.azlyrics.com/lyrics/killers/romeoandjuliet.html', 'https://www.azlyrics.com/lyrics/killers/mrbrightsidejacquelucontsthinwhitedukemix.html', 'https://www.azlyrics.com/lyrics/killers/losingtouch.html', 'https://www.azlyrics.com/lyrics/killers/human.html', 'https://www.azlyrics.com/lyrics/killers/spaceman.html', 'https://www.azlyrics.com/lyrics/killers/joyride.html', 'https://www.azlyrics.com/lyrics/killers/adustlandfairytale.html', 'https://www.azlyrics.com/lyrics/killers/thisisyourlife.html', 'https://www.azlyrics.com/lyrics/killers/icantstay.html', 'https://www.azlyrics.com/lyrics/killers/neontiger.html', 'https://www.azlyrics.com/lyrics/killers/theworldwelivein.html', 'https://www.azlyrics.com/lyrics/killers/goodnighttravelwell.html', 'https://www.azlyrics.com/lyrics/killers/forgetaboutwhatisaid.html', 'https://www.azlyrics.com/lyrics/killers/tidalwave.html', 'https://www.azlyrics.com/lyrics/killers/acripplingblow.html', 'https://www.azlyrics.com/lyrics/killers/agreatbigsled.html', 'https://www.azlyrics.com/lyrics/killers/dontshootmesanta.html', 'https://www.azlyrics.com/lyrics/killers/josephbetteryouthanme.html', 'https://www.azlyrics.com/lyrics/killers/happybirthdayguadalupe.html', 'https://www.azlyrics.com/lyrics/killers/boots.html', 'https://www.azlyrics.com/lyrics/killers/thecowboyschristmasball.html', 'https://www.azlyrics.com/lyrics/killers/fleshandbone.html', 'https://www.azlyrics.com/lyrics/killers/runaways.html', 'https://www.azlyrics.com/lyrics/killers/thewayitwas.html', 'https://www.azlyrics.com/lyrics/killers/herewithme.html', 'https://www.azlyrics.com/lyrics/killers/amatteroftime.html', 'https://www.azlyrics.com/lyrics/killers/deadlinesandcommitments.html', 'https://www.azlyrics.com/lyrics/killers/missatomicbomb.html', 'https://www.azlyrics.com/lyrics/killers/therisingtide.html', 'https://www.azlyrics.com/lyrics/killers/heartofagirl.html', 'https://www.azlyrics.com/lyrics/killers/fromhereonout.html', 'https://www.azlyrics.com/lyrics/killers/bestill.html', 'https://www.azlyrics.com/lyrics/killers/battleborn.html', 'https://www.azlyrics.com/lyrics/killers/carrymehome.html', 'https://www.azlyrics.com/lyrics/killers/prizefighter.html', 'https://www.azlyrics.com/lyrics/killers/wonderfulwonderful.html', 'https://www.azlyrics.com/lyrics/killers/theman.html', 'https://www.azlyrics.com/lyrics/killers/rut.html', 'https://www.azlyrics.com/lyrics/killers/lifetocome.html', 'https://www.azlyrics.com/lyrics/killers/runforcover.html', 'https://www.azlyrics.com/lyrics/killers/tysonvsdouglas.html', 'https://www.azlyrics.com/lyrics/killers/somekindoflove.html', 'https://www.azlyrics.com/lyrics/killers/outofmymind.html', 'https://www.azlyrics.com/lyrics/killers/thecalling.html', 'https://www.azlyrics.com/lyrics/killers/haveallthesongsbeenwritten.html', 'https://www.azlyrics.com/lyrics/killers/moneyonstraight.html', 'https://www.azlyrics.com/lyrics/killers/myownsoulswarning.html', 'https://www.azlyrics.com/lyrics/killers/blowback.html', 'https://www.azlyrics.com/lyrics/killers/dyingbreed.html', 'https://www.azlyrics.com/lyrics/killers/caution.html', 'https://www.azlyrics.com/lyrics/killers/lightningfields.html', 'https://www.azlyrics.com/lyrics/killers/fireinbone.html', 'https://www.azlyrics.com/lyrics/killers/runningtowardsaplace.html', 'https://www.azlyrics.com/lyrics/killers/mygod.html', 'https://www.azlyrics.com/lyrics/killers/whenthedreamsrundry.html', 'https://www.azlyrics.com/lyrics/killers/implodingthemirage.html', 'https://www.azlyrics.com/lyrics/killers/cestlavie.html', 'https://www.azlyrics.com/lyrics/killers/cautionwasatchstyle.html', 'https://www.azlyrics.com/lyrics/killers/westhills.html', 'https://www.azlyrics.com/lyrics/killers/quiettown.html', 'https://www.azlyrics.com/lyrics/killers/terriblething.html', 'https://www.azlyrics.com/lyrics/killers/cody.html', 'https://www.azlyrics.com/lyrics/killers/sleepwalker.html', 'https://www.azlyrics.com/lyrics/killers/runawayhorses.html', 'https://www.azlyrics.com/lyrics/killers/inthecaroutside.html', 'https://www.azlyrics.com/lyrics/killers/inanotherlife.html', 'https://www.azlyrics.com/lyrics/killers/desperatethings.html', 'https://www.azlyrics.com/lyrics/killers/pressuremachine.html', 'https://www.azlyrics.com/lyrics/killers/thegettingby.html', 'https://www.azlyrics.com/lyrics/killers/thegettingbyii.html', 'https://www.azlyrics.com/lyrics/killers/thegettingbyiii.html', 'https://www.azlyrics.com/lyrics/killers/thegettingbyiv.html', 'https://www.azlyrics.com/lyrics/killers/thegettingbyv.html', 'https://www.azlyrics.com/lyrics/killers/runawayhorsesii.html', 'https://www.azlyrics.com/lyrics/killers/westhillsii.html', 'https://www.azlyrics.com/lyrics/killers/westhillsiii.html', 'https://www.azlyrics.com/lyrics/killers/awhitedemonlovesong.html', 'https://www.azlyrics.com/lyrics/killers/christmasinla.html', 'https://www.azlyrics.com/lyrics/killers/desperate.html', 'https://www.azlyrics.com/lyrics/killers/dirtsledding.html', 'https://www.azlyrics.com/lyrics/killers/dustland.html', 'https://www.azlyrics.com/lyrics/killers/fourwinds.html', 'https://www.azlyrics.com/lyrics/killers/gettrashed.html', 'https://www.azlyrics.com/lyrics/killers/ifeelitinmybones.html', 'https://www.azlyrics.com/lyrics/killers/illbehomeforchristmas.html', 'https://www.azlyrics.com/lyrics/killers/joelthelumpofcoal.html', 'https://www.azlyrics.com/lyrics/killers/justanothergirl.html', 'https://www.azlyrics.com/lyrics/killers/landofthefree.html', 'https://www.azlyrics.com/lyrics/killers/monalisasandmadhatters.html', 'https://www.azlyrics.com/lyrics/killers/peaceofmind.html', 'https://www.azlyrics.com/lyrics/killers/replaceable.html' 'https://www.azlyrics.com/lyrics/killers/shotatthenight.html', 'https://www.azlyrics.com/lyrics/killers/spaceshipadventure.html', 'https://www.azlyrics.com/lyrics/killers/whereisshesoftsurrender.html', 'https://www.azlyrics.com/lyrics/killers/whydontyoufindoutforyourself.html']}) Let's make sure we have enough lyrics pages to scrape. In [55]: for artist, lp in lyrics pages.items() : assert(len(set(lp)) > 20) # Let's see how long it's going to take to pull these lyrics # if we're waiting `5 + 10*random.random()` seconds for artist, links in lyrics pages.items() : print(f"For {artist} we have {len(links)}.") print(f"The full pull will take for this artist will take {round(len(links)*10/3600,2)} hours.") For rush we have 171. The full pull will take for this artist will take 0.47 hours.

In [57]:	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.
	<pre>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. def generate_filename_from_link(link) : if not link : return None</pre>
	<pre>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 chareacters with UNDERSCORE name = name.replace("://","").replace(".","_") # tack on .txt name = name + ".txt"</pre>
In [60]:	<pre>name = name + ".txt" return(name)</pre>
In [61]:	<pre>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}/"):</pre>
	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 lyrics 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.
	<pre>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("\"") 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": ""}):</pre>
	<pre>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. for title, lyric in zip(set(titles),lyrics): content = title+"\n"+"\n"+lyric output_file = generate_filename_from_link(artist_lyric_page) with open(f'lyrics/{artist}/'+output_file,'w') as f: f.write(content)</pre>
	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/findingmyway.html Extract title: Finding My Way Extract title: 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/hereagain.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 lyric. Finished writing the lyric of Beneath, Between And Behind by rush https://www.azlyrics.com/lyrics/rush/bytorandthesnowdog.html 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 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/smilelikeyoumeanit.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/andyyoureastar.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/believemenatalie.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/samstown.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/blingconfessionofaking.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.
In [62]:	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 print(f"Total run time was {round((time.time() - start)/3600,2)} hours.")
	Total run time was {round((time.time() - start)/3600,2)} 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]:	·
	follower IDs you pulled) and <pre>cher_followers_data.txt</pre> . These files should be in a folder named <pre>twitter</pre> within the repository directory. This code summarizes the information at a high level to help the instructor evaluate your work. 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]:	<pre>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</pre>
	<pre># check the headers headers = infile.readline().split("\t") print(f"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")</pre>
	<pre>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.")</pre>
	<pre>print(f"For {artist} we have {len(locations)} unique locations.") print(f"For {artist} we have {len(description_words)} words in the descriptions.") print("Here are the five most common words:") print(Counter(description_words).most_common(5)) print("") print("") print("")</pre>
	# Section Complete, however, it doesn't return me any common words. I am unsure how to fix this as the momen # 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:
	Here are the five most common words: []
In [68]:	<pre>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("lyrics/" + artist) artist_files = [f for f in artist_files if 'txt' in f or 'csv' in f or 'tsv' in f] print(f"For {artist} we have {len(artist_files)} files.") artist_words = []</pre>
	<pre>for f_name in artist_files : with open("lyrics/" + artist + "/" + f_name) as infile : artist_words.extend(words(infile.read())) print(f"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.</pre>
In []:	