My Spotify Playlists

Shaniece Gooden
7/18/2025

Outline

- Executive Summary
- Introduction
- Methodology
- Results

Executive Summary

Requested data from Spotify in privacy settings. I created a Python notebook in Kaggle to clean and normalize the JSON file that contained all of my Spotify playlists. I had not had a lot of experience with JSON files before this so there was some trial and error in regards to normalizing the file into a data frame. Once I was able to normalize the file I cleaned up the dataset, and was able to create new playlists for artists that I had in my Favorites playlist that had 5 or more songs.

Introduction

Music is a big part of my life. I listen to music every day during work, doing chores around the house, driving in my car, etc. I use Spotify premium because I enjoy that I can create my own playlist of my favorite songs and artists. I like to have a backup of my playlists because I have lost them in the past and it was very frustrating to create multiple playlists all over again. I wanted an easier way to have a backup on hand so I decided to create a data frame that I can update yearly.

Methodology

JSON and Nested Dictionaries

This was my first time working with a JSON file. It took me a while to figure out how to flatten the data into a data frame. At first I was able to get a data frame with the items column containing all the information that I wanted. I tried splitting the strings. Then I tried to change the data types thinking it would make the string split work, but that also failed. Once I realized that the data was arranged in nested dictionaries I found a solution on the site pandas.pydata.org that allowed me to flatten everything all at once, giving me the data frame that I wanted. After that I was able to clean up the data and column names.

In [2]:

```
#Open and read json file with Spotify Playlists
import json
with open ('/kaggle/input/spotify-playlists/Playlist1.json', 'r') as openfile:
    json.object = json.load(openfile)
#print(json.object)
```

In [3]:

```
#Convert and normalize data from JSON file into dataframe
#Split nested dictionaries in JSON file to get track info, and keep name and lastModifiedDate columns
df=pd.json_normalize(json.object['playlists'], "items", ["name", "lastModifiedDate"])
```

Out[5]:

	addedDate	trackName	artistName	album	playlistName	lastModifiedDate
0	2025-01-20	Boundaries	Durand Bernarr	Wanderlust	Durand Bernarr	2025-01-21
1	2025-01-20	Footprints	Durand Bernarr	Wanderlust	Durand Bernarr	2025-01-21
2	2025-01-20	Lil Bit	Durand Bernarr	Wanderlust	Durand Bernarr	2025-01-21
3	2025-01-20	Leveled	Durand Bernarr	Wanderlust	Durand Bernarr	2025-01-21
4	2025-01-20	Ingredients	Durand Bernarr	Wanderlust	Durand Bernarr	2025-01-21
...	...	...	...	...	...	...
2448	2013-09-30	NaN	NaN	NaN	Wale	2018-09-04
2449	2013-09-30	NaN	NaN	NaN	Wale	2018-09-04
2450	2013-09-30	NaN	NaN	NaN	Wale	2018-09-04
2451	2013-09-30	NaN	NaN	NaN	Wale	2018-09-04
2452	2013-09-30	NaN	NaN	NaN	Wale	2018-09-04

2453 rows × 6 columns

ITunes Playlist

When I first started using Spotify I had a Mac. You can play your iTunes playlist through Spotify, but I didn't do that very often as I could just play the music through iTunes. I've never paid attention to the playlist in Spotify but it could still show up in the data I downloaded, so I wanted to remove any of these playlists/songs from the data because it's not something that I will ever change or update.

In [6]:

```
#List all unique playlists to see if need to remove iTunes from list
pd.unique(df['playlistName'])
```

In [7]:

```
#Drop OneRepublic playlist
One_Rep= df[df['playlistName']=='OneRepublic']
df.drop(One_Rep.index, axis=0, inplace=True)
pd.unique(df['playlistName'])
```

In [8]:

```
#Check Wale playlist to make sure doesn't include iTunes songs
Wale_df=df[df['playlistName']=='Wale']
Wale_df
```

Out[8]:

	addedDate	trackName	artistName	album	playlistName	lastModifiedDate
580	2022-01-10	Poke It Out (feat. J. Cole)	Wale	Folarin II	Wale	2023-01-27
581	2022-01-10	New Balances	Wale	Folarin II	Wale	2023-01-27
582	2022-01-10	Routine (feat. Rick Ross & Meek Mill)	Wale	Wow... That's Crazy	Wale	2023-01-27
583	2022-01-10	On Chill (feat. Jeremih)	Wale	Wow... That's Crazy	Wale	2023-01-27
584	2022-01-10	Sue Me (feat. Kelly Price)	Wale	Wow... That's Crazy	Wale	2023-01-27
585	2022-01-10	Fashion Week (feat. G-Eazy)	Wale	Shine	Wale	2023-01-27
586	2022-01-10	Tiffany Nikes	Wale	Folarin II	Wale	2023-01-27
587	2022-01-10	More Love (feat. Shawn Stockman of Boyz II Men...	Wale	Folarin II	Wale	2023-01-27
588	2022-01-10	Down South (feat. Yella Beezy & Maxo Kream)	Wale	Folarin II	Wale	2023-01-27
589	2022-01-10	BYGPW (Colors)	Wale	The Imperfect Storm	Wale	2023-01-27
590	2023-01-27	So Lonely (feat. Wale)	PJ Morton	Watch The Sun	Wale	2023-01-27

In [9]:

```
#Drop Wale songs using index range
df.drop(index=range(2437,2453), axis=0, inplace=True)
df
```

Out[9]:

	addedDate	trackName	artistName	album	playlistName	lastModifiedDate
0	2025-01-20	Boundaries	Durand Bernarr	Wanderlust	Durand Bernarr	2025-01-21
1	2025-01-20	Footprints	Durand Bernarr	Wanderlust	Durand Bernarr	2025-01-21
2	2025-01-20	Lil Bit	Durand Bernarr	Wanderlust	Durand Bernarr	2025-01-21
3	2025-01-20	Leveled	Durand Bernarr	Wanderlust	Durand Bernarr	2025-01-21
4	2025-01-20	Ingredients	Durand Bernarr	Wanderlust	Durand Bernarr	2025-01-21
...	...	...	...	...	...	...
2421	2013-10-02	Wonderland	Haley Reinhart	Listen Up!	Listen Up!	2022-03-03
2422	2013-10-02	Keep Coming Back	Haley Reinhart	Listen Up!	Listen Up!	2022-03-03
2423	2013-10-02	Hit The Ground Runnin'	Haley Reinhart	Listen Up!	Listen Up!	2022-03-03
2424	2013-10-02	Walking On Heaven	Haley Reinhart	Listen Up!	Listen Up!	2022-03-03
2425	2022-03-03	Talkin' About	Haley Reinhart	Better	Listen Up!	2022-03-03

Making New Playlists

Whenever I have 5 or more songs by an artist in my Favorites playlist, I like to create a separate playlist of just that artist. I do this because it allows me to hear more suggested songs from my favorite artists when I just listen to their music. I will go through my Favorites playlist and find any artists that I have 5 or more songs in the playlist. I will also look for songs that contain artists I already have a playlist for, and add those songs to that playlist removing them from the Favorites.

In [10]:

```
# Find artists with 5 or more songs in Favorites playlist
Fav_play = df[df['playlistName']=='Favorites']
Fav_play5=Fav_play.groupby(['artistName'])[['trackName']].nunique()
Fav_play5=Fav_play5.loc[Fav_play5['trackName']>=5]
Fav_play5
```

Out[10]:

	trackName
artistName	
J. Cole	5
LION BABE	5
My Chemical Romance	5
Princess Nokia	5
Tove Lo	5
Xavier Omär	5

In [11]:

```
#Find indexes of all songs and change playlist name from Favorites to artist name. Some artists had more songs that included other artists; will move those songs to same playlist separately
Fav_play5_artists = ['J. Cole', 'LION BABE', 'My Chemical Romance', 'Princess Nokia', 'Tove Lo']
Fav_play5_ind=df[df['artistName'].isin(Fav_play5_artists)].index
df.loc[Fav_play5_ind]
```

Out[11]:

	addedDate	trackName	artistName	album	playlistName	lastModifiedDate
1611	2013-12-17	Helena	My Chemical Romance	Three Cheers for Sweet Revenge	Favorites	2025-02-20
1612	2013-12-17	I'm Not Okay (I Promise)	My Chemical Romance	Three Cheers for Sweet Revenge	Favorites	2025-02-20
1613	2013-12-17	The Ghost of You	My Chemical Romance	Three Cheers for Sweet Revenge	Favorites	2025-02-20
1714	2014-11-23	She Knows (feat. Amber Coffman & Cults)	J. Cole	Born Sinner	Favorites	2025-02-20
1715	2014-11-23	Can't Get Enough	J. Cole	Cole World: The Sideline Story	Favorites	2025-02-20
1740	2016-06-09	Talking Body	Tove Lo	Queen Of The Clouds	Favorites	2025-02-20

In [14]:

```
#Change playlist names for songs moved to new playlists, and change artist
MCR_ind = [1611, 1612 ,1613, 2179, 2199]
JC_ind = [1714, 1715, 1746, 1784, 1957, 2047, 2241]
Tove_ind = [1740, 1821, 1948, 2076, 2155]
Lion_ind = [1797, 1799, 1838, 1858, 2087]
PN_ind = [1876, 1929, 1960, 1961, 2069, 2116]
XO_ind = [1855, 2095, 2118, 2221, 2279]
df.loc[MCR_ind, 'playlistName'] = 'My Chemical Romance'
df.loc[JC_ind, 'playlistName'] = 'J. Cole'
df.loc[Tove_ind, 'playlistName'] = 'Tove Lo'
df.loc[Lion_ind, 'playlistName'] = 'LION BABE'
df.loc[PN_ind, 'playlistName'] = 'Princess Nokia'
df.loc[XO_ind, 'playlistName'] = 'Xavier Omar'
df.loc[XO_ind, 'artistName'] = 'Xavier Omar'
```

In [15]:

```
#Single songs moved to existing playlists
Sir_songs = ["Fuck Y'all (feat.SiR)"]
Sir_ind = df[df['trackName'].isin(Sir_songs)].index
df.loc[Sir_ind, 'playlistName'] = 'SiR'

Doechii_songs = ['What It Is (Solo Version)']
Doechii_ind = df[df['trackName'].isin(Doechii_songs)].index
df.loc[Doechii_ind, 'playlistName'] = 'Doechii'

JM_songs = ['Do My Thing']
JM_ind = df[df['trackName'].isin(JM_songs)].index
df.loc[JM_ind, 'playlistName'] = 'Janelle Monae'

KL_songs = ['Wow Freestyle (feat. Kendrick Lamar']
KL_ind = df[df['trackName'].isin(KL_songs)].index
df.loc[KL_ind, 'playlistName'] = 'Kendrick Lamar'
```

In [16]:

```
#Artists that had enough songs to create their own playlist
Syd_songs = ['Do U Wrong (feat. Syd)', 'MADE 4 U', 'Take Care of You (feat. Syd)', 'Dollar Bills (feat.
Steve Lacy)', 'Shibuya (feat. Syd)']
Syd_ind = df[df['trackName'].isin(Syd_songs)].index
df.loc[Syd_ind, 'playlistName'] = 'Syd'

NM_songs = ['Monster', 'Anaconda', 'Truffle Butter', 'Rake It Up', 'Anybody (feat. Nicki Minaj)', 'Big
Bank feat. 2 Chainz, Big Sean, Nicki Minaj']
NM_ind = df[df['trackName'].isin(NM_songs)].index
df.loc[NM_ind, 'playlistName'] = 'Nicki Minaj'
```

New Songs

Since this data was downloaded there are 4 playlists that I recently added songs to that may not be in this dataset. Those playlists are: Favorites, Durand Bernarr, SiR, and Tyler the Creator. I created a subset for each of these playlists and found the most recent song added, if there are songs in the playlist that are after the most recent song then I added those songs into the dataset. I did this by creating a separate data frame of the new songs and combining it with the cleaned data frame to get the final result.

In [17]:

```
#Show most recent song that was added to Favorites playl
Fav_play.loc[df['addedDate'].idxmax()]
```

Out[17]:

```
addedDate                2025-02-20
trackName                       WHY
artistName              Sasha Keable
album                           WHY
playlistName              Favorites
lastModifiedDate         2025-02-20
Name: 2285, dtype: object
```

In [18]:

```
#Show most recent song added to SiR playlist
Sir_play = df[df['playlistName']=='SiR']
Sir_play.sort_values(by='addedDate',ascending=False)
```

Out[18]:

	addedDate	trackName	artistName	album	playlistName	lastModifiedDate
60	2025-02-07	Wires in the Way	SiR	Chasing Summer	SiR	2025-02-07
70	2025-02-07	NO EVIL	SiR	HEAVY	SiR	2025-02-07
76	2025-02-07	LIFE IS GOOD (feat. Scribz Riley)	SiR	HEAVY	SiR	2025-02-07
75	2025-02-07	SATISFACTION	SiR	HEAVY	SiR	2025-02-07
74	2025-02-07	ONLY HUMAN	SiR	HEAVY	SiR	2025-02-07
...	...	...	...	...	...	...
36	2025-01-03	The Perfect Remedy	SiR	Seven Sundays	SiR	2025-02-07
37	2025-01-03	Crashing Down	SiR	Seven Sundays	SiR	2025-02-07
38	2025-01-03	The Bullet and the Gun	SiR	Seven Sundays	SiR	2025-02-07
39	2025-01-03	He Deserves Your Love	SiR	Seven Sundays	SiR	2025-02-07
26	2025-01-03	Love You	SiR	Seven Sundays	SiR	2025-02-07

In [21]:

```
#Creat separate dataframe of new songs that were added to playlists after download
new_songs = {'addedDate': ['NaN','NaN','NaN','NaN','NaN','NaN','NaN','NaN',
                           'NaN','NaN','NaN','NaN','NaN','NaN','NaN','NaN',
                           'NaN','NaN','NaN','NaN','NaN','NaN','NaN','NaN'],
             'trackName':['GENEROUS','Flounce','Impact','Jump','No Business','Overqualified','Unspoke
n','Completed','Reaching','PSST!','Here We Are','BLAST!','THAT!','Specialty','Home Alone',
                          'Must Be Nice','Unknown','Fist Bump (feat. Free Nationals)','PIT STOP','Neutra
l','Processing...','ETA?',
                          'AUCTION','Taste'],
             'artistName':['Durand Bernarr','Durand Bernarr, GAWD','Durand Bernarr','Durand Bernarr','Du
rand Bernarr, E-WHIZZ','Durand Bernarr','Durand Bernarr','Durand Bernarr','Durand Bernarr','Durand Bern
arr','Durand Bernarr','Durand Bernarr','Durand Bernarr, T-Pain','Durand Bernarr','Durand Bernarr',
                           'Durand Bernarr','Durand Bernarr','Durand Bernarr, Free Nationals','Durand Ber
narr, DeShawn Jenkins','Durand Bernarr','Durand Bernarr','Durand Bernarr',
                           'Sasha Keable, DESTIN CONRAD','Coco Jones'],
             'album' :['BLOOM','BLOOM','BLOOM','BLOOM','BLOOM','BLOOM','BLOOM','BLOOM','BLOOM','BLOO
M','BLOOM','BLOOM','BLOOM','BLOOM','BLOOM',
```

Out[21]:

	addedDate	trackName	artistName	album	playlistName	lastModifiedDate
0	NaN	GENEROUS	Durand Bernarr	BLOOM	Durand Bernarr	NaN
1	NaN	Flounce	Durand Bernarr, GAWD	BLOOM	Durand Bernarr	NaN
2	NaN	Impact	Durand Bernarr	BLOOM	Durand Bernarr	NaN
3	NaN	Jump	Durand Bernarr	BLOOM	Durand Bernarr	NaN
4	NaN	No Business	Durand Bernarr, E-WHIZZ	BLOOM	Durand Bernarr	NaN
5	NaN	Overqualified	Durand Bernarr	BLOOM	Durand Bernarr	NaN
6	NaN	Unspoken	Durand Bernarr	BLOOM	Durand Bernarr	NaN
7	NaN	Completed	Durand Bernarr	BLOOM	Durand Bernarr	NaN
8	NaN	Reaching	Durand Bernarr	BLOOM	Durand Bernarr	NaN
9	NaN	PSST!	Durand Bernarr	BLOOM	Durand Bernarr	NaN
10	NaN	Here We Are	Durand Bernarr	BLOOM	Durand Bernarr	NaN
11	NaN	BLAST!	Durand Bernarr	BLOOM	Durand Bernarr	NaN
12	NaN	THAT!	Durand Bernarr, T-Pain	BLOOM	Durand Bernarr	NaN
13	NaN	Specialty	Durand Bernarr	BLOOM	Durand Bernarr	NaN
14	NaN	Home Alone	Durand Bernarr	BLOOM	Durand Bernarr	NaN
15	NaN	Must Be Nice	Durand Bernarr	En Route	Durand Bernarr	NaN

Results

In [22]:

```
#Combine dataframes for final result
df=pd.concat([df, new_songs_df], ignore_index=True)
df
```

Out[22]:

	addedDate	trackName	artistName	album	playlistName	lastModifiedDate
0	2025-01-20	Boundaries	Durand Bernarr	Wanderlust	Durand Bernarr	2025-01-21
1	2025-01-20	Footprints	Durand Bernarr	Wanderlust	Durand Bernarr	2025-01-21
2	2025-01-20	Lil Bit	Durand Bernarr	Wanderlust	Durand Bernarr	2025-01-21
3	2025-01-20	Leveled	Durand Bernarr	Wanderlust	Durand Bernarr	2025-01-21
4	2025-01-20	Ingredients	Durand Bernarr	Wanderlust	Durand Bernarr	2025-01-21
...	...	...	...	...	...	...
2445	NaN	Neutral	Durand Bernarr	En Route	Durand Bernarr	NaN
2446	NaN	Processing...	Durand Bernarr	En Route	Durand Bernarr	NaN
2447	NaN	ETA?	Durand Bernarr	En Route	Durand Bernarr	NaN
2448	NaN	AUCTION	Sasha Keable, DESTIN CONRAD	AUCTION	Favorites	NaN
2449	NaN	Taste	Coco Jones	Taste	Favorites	NaN

Final Data Frame

- The final data frame now contains all of my previous playlists and the new ones that I created.
- Next year I can repeat this process using most of the code I have here with a new JSON file.
- Challenges
  - Normalizing the JSON file was definitely a challenge, but with some research and determination I was able to figure out this challenge and in the process learned a new Python skill.
  - In creating the new playlists, I think it would have been easier if I could have searched through the songs based on a certain string of characters given that some of the artists were featured on songs with other artists. This would probably require me to learn more about regex.
  - When trying to determine the most recently added songs to a playlist the first code I used wouldn't work with other playlists. I would like to look into that further to determine why.

Thank you