

```
In [2]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import spotipy
import playlist
import config
import time
import mod1
from sklearn.metrics import pairwise
from sklearn.preprocessing import MinMaxScaler, StandardScaler
from spotipy.oauth2 import SpotifyOAuth
from sklearn.feature_extraction.text import TfidfVectorizer
```

```
In [3]: #create spotipy object to interact with spotify web API
from spotipy.oauth2 import SpotifyClientCredentials
sp = spotipy.Spotify(
    auth_manager=SpotifyClientCredentials
    (client_id=config.SPOTIFY_CLIENT_ID,
     client_secret=config.SPOTIFY_SECRET
    ))
```

```
In [4]: #playlist_link, final_playlist_len = mod1.get_user_playlist()
playlist_link, final_playlist_len = 'https://open.spotify.com/playlist/6ZbjoVzf
playlist1 = playlist.Playlist(playlist_link, sp)
```

```
In [5]: playlist1.normalized_numeric_features
```

Out[5]:

	track_id	0	1	2	3	4	5	
0	2yUbCEiaolfSMLuDo9RMmG	0.619799	0.946043	0.090909	0.794055	0.0	0.223718	0.00
1	1uL78kFWxhz3umTMWJu8n5	0.764706	0.404676	0.545455	0.615488	0.0	0.559196	0.10
2	6j0teRL3WdExzqmy334sh5	0.869440	0.586331	0.090909	0.827914	0.0	0.545543	0.07
3	5VnyRfL844Th0BBBGnhJOL	0.361549	0.541367	0.636364	0.507319	0.0	0.627462	0.55
4	3SEd8nPd8MpGwk6ZZ8tk2j	0.779053	0.836331	1.000000	0.626103	1.0	0.108445	0.00
5	0RyA3o15NOLJYtm9NIDu5c	0.509326	0.852518	0.090909	0.647558	1.0	0.399259	0.26
6	6TrjaOqUwdKQ7KNxJr2s9Y	0.515065	0.438849	0.636364	0.623086	0.0	0.033353	0.17
7	4qKcDkK6siZ7Jp1Jb4m0aL	0.921090	0.500000	0.909091	0.482512	1.0	0.475327	0.00
8	2rqUbIDWJKIMVwh9uJc0Vv	0.800574	0.830935	0.636364	0.616158	1.0	0.494831	0.04
9	0WQcnrsEqrY2e5nexGe3HX	0.889527	0.411871	0.000000	0.703654	1.0	0.093817	0.13
10	0PDwNx6GrHc5sRVy6Hc7GP	0.916786	0.417266	0.454545	0.655827	0.0	0.130096	0.07
11	6dnrppGRbt8fie18Lcd9Dd	0.698709	0.694245	0.909091	0.697285	0.0	0.114882	0.15
12	790MhTXPFrYkrqCKvpg4t1	0.738881	0.764388	0.818182	0.793496	0.0	0.054808	0.30
13	0gHcxtYWQT0HrIGxaxP1KT	0.847920	0.681655	0.545455	0.301710	0.0	0.280281	0.30
14	06Gyf2pFDVmN5K8XAawG96	0.746055	0.395683	0.090909	0.584982	1.0	0.157402	0.00
15	7r6LNJT2LqpLpEyZQJPygt	0.832138	0.464029	0.545455	0.676724	0.0	0.309538	0.40
16	6yHkPtI6UQ7RjtJLBPzbJw	0.193687	0.118705	0.727273	0.197676	1.0	0.093232	0.04
17	5W9QCBJ6Em0XWX333YrNhC	0.593974	0.341727	0.090909	0.361158	1.0	0.676224	0.04
18	78sr3ogs4UzITcCNbXM9cM	0.484935	0.762590	0.272727	0.512795	0.0	0.354398	0.04
19	1DmnEYXa4WfbdhAPwNzgD8	0.665710	0.568345	0.363636	0.298357	0.0	0.724985	1.00
20	4gowy3WT6D1yhMLgRBIf9C	0.727403	0.046763	0.000000	0.203598	1.0	0.711332	0.03
21	0zl3PdjlI320fK0pxBGWNR	0.457676	0.996403	0.636364	0.683093	1.0	0.531890	0.04
22	6urLjX35oXXRYtcH1b3dCh	0.708752	0.406475	0.454545	0.819533	0.0	0.052467	0.09
23	7g7raxdQpiLZT7aOlib4S1	0.843615	0.636691	0.727273	0.777741	1.0	0.169105	0.02
24	4CwumbYOUtgSylJnSrZ7oY	0.581062	0.561151	0.545455	0.427981	1.0	0.024186	0.80
25	0CjHqLvr4EdjBJ0XiKtbCI	1.000000	0.000000	0.000000	0.110850	1.0	1.000000	0.09
26	2tznHmp70DxMyr2XhWLOW0	0.510760	0.670863	0.181818	0.935412	1.0	0.000000	0.11
27	6mUwMth1MQT8NVSMI7IKjx	0.882353	0.399281	0.909091	0.429545	0.0	0.311488	0.13
28	2Th9BGKvfZG8bKQSACitwG	0.799139	0.339928	0.000000	0.449324	1.0	0.639165	0.55
29	4g64deYSvQZ4w9uGpYz2sL	0.152080	0.739209	0.545455	0.694603	0.0	0.691828	0.39
30	5mCPDVBb16L4XQwDdbRUpz	0.758967	0.287770	1.000000	0.048721	1.0	0.025941	0.38
31	7sO5G9EABYOXQKNPNiE9NR	0.860832	0.224820	0.818182	0.394793	1.0	0.350497	0.22
32	7foypmc7KZyU716Yv63BP	0.840746	0.620504	0.090909	0.909375	1.0	0.054613	0.00
33	3ncgNpxLoBQ65ABk4djDyd	0.919656	0.294964	0.090909	0.376578	1.0	0.180807	0.00
34	0gD6lWJhbXHWNVwQziSVvl	0.863702	0.359712	1.000000	0.580512	1.0	0.297835	0.33

	track_id	0	1	2	3	4	5	
35	0AICBILzFCTpUqmAbtzB2z	0.543759	0.559353	0.818182	0.502067	0.0	0.054223	0.00
36	3YeJXuRSNS5FYwOsDu44kD	0.000000	0.931655	0.727273	1.000000	0.0	0.165204	0.24
37	3vWxQ6wOfMr8zH1u7pVFI6	0.634146	0.931655	0.181818	0.854397	1.0	0.284182	0.50
38	2u7mxWSeoqTXndK5e08jMp	0.667145	0.616906	0.545455	0.539278	0.0	0.280281	0.23
39	3ZhTT6yjZwpPph5MIJ53XY	0.589670	0.683453	0.000000	0.500391	1.0	0.134777	0.81
40	1Y500yrabUJY2mIQS3RDt4	0.695839	0.895683	0.818182	0.782657	0.0	0.288083	0.08
41	4YJmZfvlheSziXem8HBWrj	0.793400	0.230216	0.000000	0.246843	1.0	0.327092	0.03
42	0k9RlZtAo4DTQu0ZM6INZ	0.667145	0.305755	0.909091	0.063694	0.0	0.134972	0.02
43	7y6c07pgjZvtHI9kuMVqk1	0.718795	0.751799	0.454545	0.299475	0.0	0.048957	0.14
44	0xl1w2q4VLojeXp4JfazPL	0.893831	0.244604	0.727273	0.000000	1.0	0.723035	0.06
45	7bJ4mu7MHa3rHiNyKjOoSI	0.764706	0.446043	0.545455	0.470220	0.0	0.637215	0.11
46	5Z3GHaZ6ec9bsil5BenrbY	0.744620	0.424460	0.090909	0.610348	1.0	0.030817	0.29
47	3Vo4wlnECJQuz9BIBMOu8i	0.608321	1.000000	0.454545	0.775059	0.0	0.142969	0.02
48	5IZsh9Qf7CbHI9Fcc7Zcsq	0.839311	0.627698	0.727273	0.401721	1.0	0.153501	0.05

```
In [6]: # #tfidf genre lists
# tfidf = TfidfVectorizer()
# tfidf_matrix = tfidf.fit_transform(playlist1.df_playlist.genres.apply(lambda
# genre_df = pd.DataFrame(tfidf_matrix.toarray())
# genre_df.columns = ['genre' + "|" + i for i in tfidf.get_feature_names_out()]
# genre_df.reset_index(drop = True, inplace=True)

# num_features_scaled = playlist1.normalized_numeric_features
# num_features_scaled = pd.concat([num_features_scaled, genre_df], axis=1)
# num_features_scaled.columns
```

```
In [7]: playlist1.tf_idf_transform()
```

```
In [8]: artist_counts, artist_id_counts = playlist1.get_artist_counts2()
avg_audio_values = pd.DataFrame(playlist1.normalized_numeric_features.loc[:,0:]
# avg_audio_values.index
```

```
In [9]: artists_search = mod1.artist_search_results(artist_counts.index, artist_id_cou
recs = playlist1.get_recommendations2(artists_search, avg_audio_values,tf_idf=1
```

```
In [10]: recs
```

Out[10]:

	artist	track	track_id	similarity
0	Drake	Privileged Rappers	7l2nxyx7lkBX5orhkALg0V	0.913764
1	Drake	Spin Bout U	2ZL7WZcjuYKi1KUDtp4kCC	0.908879
2	Drake	Nonstop	0TILq3IA83rQOYtrqBqSct	0.899592
3	Drake	On BS	34tz0eDhGuFErluW3q4mPX	0.899269
4	Drake	Hotline Bling	0wwPcA6wtMf6HUMplRdeP7	0.882253
...	...	...	...	...
110	YBN Nahmir	2 Seater (feat. G-Eazy & Offset)	57IEOoQ3Szy1xbF89JBWR	0.92047
111	YBN Nahmir	Bounce Out With That Remix (feat. Machine Gun ...	5WuU70OzSBj458W7udoBhM	0.892048
112	YBN Nahmir	Spaz (feat. YBN Nahmir)	7d35cilfICdeKTY140KmKT	0.888725
113	YBN Nahmir	Rubbin off the Paint	6V4KHt9xu4TPEnDFoBeacT	0.882871
114	YBN Nahmir	Bows	4Xkk9RBSBgRBAFwPta5FIT	0.881398

115 rows × 4 columns

```
In [11]: # modl.create_new_playlist(recs)
```

## what if i used RMSE as an evaluation metric?

- use the new playlist link to analyze the new playlist
- calculate avg\_audio\_values for the new playlist
- take the error against the old ones

```
In [12]: new_playlist = playlist.Playlist('https://open.spotify.com/playlist/7995dcxuD1k  
new_playlist.tf_idf_transform()  
artist_counts, artist_id_counts = playlist1.get_artist_counts2()  
avg_audio_values2 = pd.DataFrame(new_playlist.normalized_numeric_features.loc[:
```

```
In [13]: # avg_audio_values2
```

```
In [14]: from sklearn.metrics import mean_squared_error, mean_absolute_error
```

```
In [15]: # avg_audio_values.index.values
```

```
In [16]: # avg_audio_values2.index.values
```

```
In [17]: overlap = [column for column in avg_audio_values.index.values if column in avg_  
mean_squared_error(avg_audio_values.loc[overlap, :], avg_audio_values2.loc[over
```

Out[17]: 0.01822299408336292

## use cosine similarity to measure how close the 2 playlists are

```
In [18]: from sklearn.metrics import pairwise
```

```
In [19]: pairwise.cosine_similarity(np.array(avg_audio_values.loc[overlap, :]).reshape(1, 13))
```

Out[19]: array([[0.90773392]])

```
In [33]: def evaluate(summarized_values1, summarized_values2):
    overlap = [column for column in summarized_values1.index.values if column in summarized_values2.index.values]
    # print(overlap)

    mse = mean_squared_error(summarized_values1.loc[overlap, :], summarized_values2.loc[overlap, :])
    print('RMSE', np.sqrt(mse))

    cosine_sim = pairwise.cosine_similarity(np.array(summarized_values1.loc[overlap, :]).reshape(1, 13),
    np.array(summarized_values2.loc[overlap, :]).reshape(1, 13))
    print('cosine_sim', np.sqrt(cosine_sim))
```

```
In [32]: evaluate(avg_audio_values, avg_audio_values2)
```

```
[0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 'genre|alabama_rap', 'genre|atl_hip_hop', 'genre|dance_pop', 'genre|hip_hop', 'genre|melodic_rap', 'genre|philly_rap', 'genre|pop', 'genre|pop_r', 'genre|pop_rap', 'genre|rap', 'genre|slap_house', 'genre|southern_hip_hop', 'genre|trap', 'genre|underground_hip_hop', 'genre|vapor_trap', 'genre|viral_rap']
RMSE 0.13499257047468546
cosine_sim [0.95275071]
```

## Test & Evaluation 1

- note: the playlist link is going to be hardcoded into the cell, as is the final playlist length (25)

```
In [34]: playlist_link = 'https://open.spotify.com/playlist/4uLMmpdcC9DhQo8ptZw10t?si=d0p1'
p1 = playlist.Playlist(playlist_link, mod1.sp)
p1.tf_idf_transform()
artist_counts, artist_id_counts = p1.get_artist_counts2()
summarized_feature_vector = pd.DataFrame(p1.normalized_numeric_features.loc[:, 0:12])
artists_search = mod1.artist_search_results(artist_counts.index, artist_id_counts)
recs = p1.get_recommendations2(artists_search, summarized_feature_vector, tf_idf_weights)
new_p1 = mod1.create_new_playlist(recs)

#create a playlist object with new playlist to get the audio features
new_playlist = playlist.Playlist(new_p1, mod1.sp)
new_playlist.tf_idf_transform()
artist_counts, artist_id_counts = new_playlist.get_artist_counts2()
summarized_feature_vector2 = pd.DataFrame(new_playlist.normalized_numeric_features.loc[:, 0:12])
```

```
#evaluate
evaluate(summarized_feature_vector, summarized_feature_vector2)
```

```
How long would you like your final playlist to be? (please enter an integer):2
5
Choose a name for your new playlist:test_eval_1
Add a description for your new playlist (OPTIONAL):recommendations from: http
s://open.spotify.com/playlist/4ulMmpdcC9DhQo8ptZw10t?si=d076551041774dfe
dict_keys(['collaborative', 'description', 'external_urls', 'href', 'id', 'ima
ges', 'name', 'owner', 'primary_color', 'public', 'snapshot_id', 'tracks', 'ty
pe', 'uri'])
DONE!
Here is the link to your new playlist: https://open.spotify.com/playlist/0vamB
Vo6Xbcs4fQjj2KWDM
RMSE 0.19328275596332697
cosine_sim [0.92191137]
```

## Test & Evaluation 2

```
In [35]: playlist_link = 'https://open.spotify.com/playlist/1zSsAP4PDtRjbD9LDXKAGl?si=3f4f21878b1c4d91'
p2 = playlist.Playlist(playlist_link, mod1.sp)
p2.tf_idf_transform()
artist_counts, artist_id_counts = p2.get_artist_counts2()
summarized_feature_vector = pd.DataFrame(p2.normalized_numeric_features.loc[:,0
artists_search = mod1.artist_search_results(artist_counts.index, artist_id_cou
recs = p2.get_recommendations2(artists_search, summarized_feature_vector, tf_id
new_p2 = mod1.create_new_playlist(recs)

#create a playlist object with new playlist to get the audio features
new_playlist = playlist.Playlist(new_p2, mod1.sp)
new_playlist.tf_idf_transform()
artist_counts, artist_id_counts = new_playlist.get_artist_counts2()
summarized_feature_vector2 = pd.DataFrame(new_playlist.normalized_numeric_featu

#evaluate
evaluate(summarized_feature_vector, summarized_feature_vector2)
```

```
How long would you like your final playlist to be? (please enter an integer):2
5
Choose a name for your new playlist:test_eval_2
Add a description for your new playlist (OPTIONAL):recommendations from: http
s://open.spotify.com/playlist/1zSsAP4PDtRjbD9LDXKAGl?si=3f4f21878b1c4d91
dict_keys(['collaborative', 'description', 'external_urls', 'href', 'id', 'ima
ges', 'name', 'owner', 'primary_color', 'public', 'snapshot_id', 'tracks', 'ty
pe', 'uri'])
DONE!
Here is the link to your new playlist: https://open.spotify.com/playlist/7MrzF
KY8WsoFWioSFnrYov
RMSE 0.17687613481914605
cosine_sim [0.9187298]
```

## Test & Evaluation 3

```
In [36]: playlist_link = 'https://open.spotify.com/playlist/2wiD98vsR7oA7EQuWML9to?si=2c'
```

```

p3 = playlist.Playlist(playlist_link, mod1.sp)
p3.tf_idf_transform()
artist_counts, artist_id_counts = p3.get_artist_counts2()
summarized_feature_vector = pd.DataFrame(p3.normalized_numeric_features.loc[:,0
artists_search = mod1.artist_search_results(artist_counts.index, artist_id_coun
recs = p3.get_recommendations2(artists_search, summarized_feature_vector, tf_idc
new_p3 = mod1.create_new_playlist(recs)

#create a playlist object with new playlist to get the audio features
new_playlist = playlist.Playlist(new_p3, mod1.sp)
new_playlist.tf_idf_transform()
artist_counts, artist_id_counts = new_playlist.get_artist_counts2()
summarized_feature_vector2 = pd.DataFrame(new_playlist.normalized_numeric_featu

#evaluate
evaluate(summarized_feature_vector, summarized_feature_vector2)

```

Input contains NaN.

How long would you like your final playlist to be? (please enter an integer):2  
5

Choose a name for your new playlist:test\_eval\_3

Add a description for your new playlist (OPTIONAL):recommendations from: http  
s://open.spotify.com/playlist/2wiD98vsR7oA7EQuWML9to?si=910cb3df58c2439f  
dict\_keys(['collaborative', 'description', 'external\_urls', 'href', 'id', 'ima  
ges', 'name', 'owner', 'primary\_color', 'public', 'snapshot\_id', 'tracks', 'ty  
pe', 'uri'])

DONE!

Here is the link to your new playlist: <https://open.spotify.com/playlist/6yvdtH7knAWQsCtKC10WSd>

RMSE 0.14233797414401894

cosine\_sim [0.96407933]

## Test & Evaluation 4

```

In [37]: playlist_link = 'https://open.spotify.com/playlist/1QiCuGZfXtVsKGptgU8NVW?si=05
p4 = playlist.Playlist(playlist_link, mod1.sp)
p4.tf_idf_transform()
artist_counts, artist_id_counts = p4.get_artist_counts2()
summarized_feature_vector = pd.DataFrame(p4.normalized_numeric_features.loc[:,0
artists_search = mod1.artist_search_results(artist_counts.index, artist_id_coun
recs = p4.get_recommendations2(artists_search, summarized_feature_vector, tf_idc
new_p4 = mod1.create_new_playlist(recs)

#create a playlist object with new playlist to get the audio features
new_playlist = playlist.Playlist(new_p4, mod1.sp)
new_playlist.tf_idf_transform()
artist_counts, artist_id_counts = new_playlist.get_artist_counts2()
summarized_feature_vector2 = pd.DataFrame(new_playlist.normalized_numeric_featu

#evaluate
evaluate(summarized_feature_vector, summarized_feature_vector2)

```

```

How long would you like your final playlist to be? (please enter an integer):2
5
Choose a name for your new playlist:test_eval_4
Add a description for your new playlist (OPTIONAL):recommendations from:http
s://open.spotify.com/playlist/1QiCuGZfXtVsKGptgU8NVW?si=059be9de7f194b75
dict_keys(['collaborative', 'description', 'external_urls', 'href', 'id', 'ima
ges', 'name', 'owner', 'primary_color', 'public', 'snapshot_id', 'tracks', 'ty
pe', 'uri'])
DONE!
Here is the link to your new playlist: https://open.spotify.com/playlist/3Rrsm
RGpYZFnpuopiWcfa
RMSE 0.1425042566843686
cosine_sim [0.94147226]

```

## Test & Evaluation 5

```

In [38]: playlist_link = 'https://open.spotify.com/playlist/6H2T19cKo9L30lMD7nToHi?si=13
p5 = playlist.Playlist(playlist_link, mod1.sp)
p5.tf_idf_transform()
artist_counts, artist_id_counts = p5.get_artist_counts2()
summarized_feature_vector = pd.DataFrame(p5.normalized_numeric_features.loc[:,0
artists_search = mod1.artist_search_results(artist_counts.index, artist_id_coun
recs = p5.get_recommendations2(artists_search, summarized_feature_vector, tf_ic
new_p5 = mod1.create_new_playlist(recs)

#create a playlist object with new playlist to get the audio features
new_playlist = playlist.Playlist(new_p5, mod1.sp)
new_playlist.tf_idf_transform()
artist_counts, artist_id_counts = new_playlist.get_artist_counts2()
summarized_feature_vector2 = pd.DataFrame(new_playlist.normalized_numeric_featu

#evaluate
evaluate(summarized_feature_vector, summarized_feature_vector2)

```

```

How long would you like your final playlist to be? (please enter an integer):2
5
Choose a name for your new playlist:test_eval_5
Add a description for your new playlist (OPTIONAL):recommendations from: http
s://open.spotify.com/playlist/6H2T19cKo9L30lMD7nToHi?si=2efc02675b5b443e
dict_keys(['collaborative', 'description', 'external_urls', 'href', 'id', 'ima
ges', 'name', 'owner', 'primary_color', 'public', 'snapshot_id', 'tracks', 'ty
pe', 'uri'])
DONE!
Here is the link to your new playlist: https://open.spotify.com/playlist/2ERjk
fGWngaKpsejUOMV5s
RMSE 0.10196259519098216
cosine_sim [0.97755122]

```