Spotify Study Playlists

Presented by Christy Sato

Data Description

Data Sources:

- Spotify API
  - Tracks, artists, and features of songs (danceability, energy, etc) in a playlist
  - 2,316 observations (songs) and 15 columns
  - 20 study playlists, 9 other
- MongoDB
  - #studyplaylists, #studysongs, #studymusic
  - 74 tweets

```
data.dtypes
```

```
playlist_name        object
date_added           object
track_name           object
first_artist         object
danceability        float64
energy              float64
key                   int64
loudness            float64
speechiness         float64
acousticness        float64
instrumentalness    float64
liveness            float64
valence             float64
tempo               float64
duration_ms           int64
dtype: object
```

Analysis Question Description

What features of songs make it a good study song in a study playlist?

What are the significant differences from a study playlist and other genre playlists?

What is the frequency of words in a collection of tweets that includes #studyplaylists?

Study Playlists

```
studydf.describe()
```

	danceability	energy	key	loudness	speechiness	acousticness	instrumentalness	liveness	valence	tempo	duration_ms
count	1527.000000	1527.000000	1527.000000	1527.000000	1527.000000	1527.000000	1527.000000	1527.000000	1527.000000	1527.000000	1527.000000
mean	0.501934	0.290835	5.179437	-16.211210	0.069125	0.687455	0.787400	0.132331	0.256014	110.972791	213624.620825
std	0.209249	0.234266	3.474758	6.617235	0.082457	0.348716	0.253052	0.099262	0.221909	32.740743	103197.438970
min	0.000000	0.000877	0.000000	-39.952000	0.000000	0.000002	0.000000	0.032800	0.000000	0.000000	51337.000000
25%	0.335000	0.099150	2.000000	-20.763500	0.036500	0.437000	0.802000	0.095500	0.073400	82.752000	142375.000000
50%	0.518000	0.232000	5.000000	-14.860000	0.043600	0.860000	0.886000	0.108000	0.186000	109.982000	192008.000000
75%	0.673000	0.440000	8.000000	-11.133000	0.059750	0.967000	0.925000	0.123000	0.363000	130.039000	255730.000000
max	0.943000	0.984000	11.000000	-2.921000	0.915000	0.996000	0.997000	0.946000	0.979000	218.418000	806926.000000

```
studydf.groupby(['playlist_name']).mean()
```

playlist_name	danceability	energy	key	loudness	speechiness	acousticness	instrumentalness	liveness	valence	tempo	duration_ms
Beats to think to	0.729800	0.694850	5.125000	-10.170925	0.050643	0.041506	0.863687	0.125013	0.231197	123.477350	435681.975000
Chill Lofi Study Beats	0.690497	0.265963	5.242775	-13.377867	0.171994	0.682682	0.740774	0.140253	0.448846	103.946607	117114.404624
Deep Concentration	0.738920	0.717720	5.200000	-10.747380	0.057296	0.059395	0.862320	0.115742	0.250742	124.732060	390145.500000
Focus Flow	0.682353	0.567426	5.397059	-8.928853	0.131265	0.269527	0.702453	0.154309	0.670850	112.222588	133045.735294
Intense Studying	0.333285	0.174012	5.416667	-19.582972	0.042118	0.892000	0.798151	0.124219	0.150753	111.743264	212287.472222
Just Focus	0.395367	0.095944	5.304762	-22.329495	0.048947	0.958419	0.920895	0.120482	0.202031	102.824619	176705.485714
Music for Concentration	0.399091	0.074478	4.988889	-22.439033	0.052339	0.956422	0.913089	0.113276	0.194336	105.953111	154919.555556
Peaceful Guitar	0.636547	0.145848	5.368421	-15.796316	0.072255	0.931558	0.804126	0.111098	0.251235	115.013505	168114.642105
Perfect Concentration	0.330600	0.047783	5.466667	-28.628689	0.048153	0.953111	0.800552	0.130316	0.172798	101.401889	194076.666667
Productive Morning	0.406752	0.375152	5.283582	-13.738000	0.035473	0.583559	0.843896	0.125560	0.137767	117.636269	280966.507463
Quiet Hours	0.490336	0.289170	5.253012	-17.018964	0.101017	0.776986	0.679661	0.195565	0.233302	109.658470	172813.530120
Reading Chill Out	0.295443	0.140614	4.606557	-21.184090	0.039793	0.858311	0.904254	0.123250	0.105879	99.743443	209773.918033
Reading Soundtrack	0.215105	0.115809	5.206897	-21.498448	0.040719	0.869793	0.795338	0.104229	0.065002	95.297862	203277.844828
Soft Focus	0.406205	0.175945	5.136364	-17.343045	0.040916	0.881507	0.717321	0.156457	0.253523	107.855727	259993.750000
Study Zone	0.629480	0.527700	5.480000	-7.821520	0.079356	0.374512	0.015160	0.168178	0.366444	113.600280	204051.240000
Workday Lounge	0.519313	0.306515	4.822917	-15.421427	0.041533	0.570328	0.820729	0.144903	0.188425	115.907010	231769.020833

Key takeaways:

- Playlists have different types of songs which affect the features
- Best study playlists depends on the user's preferences

Comparison to Other Genre Playlists

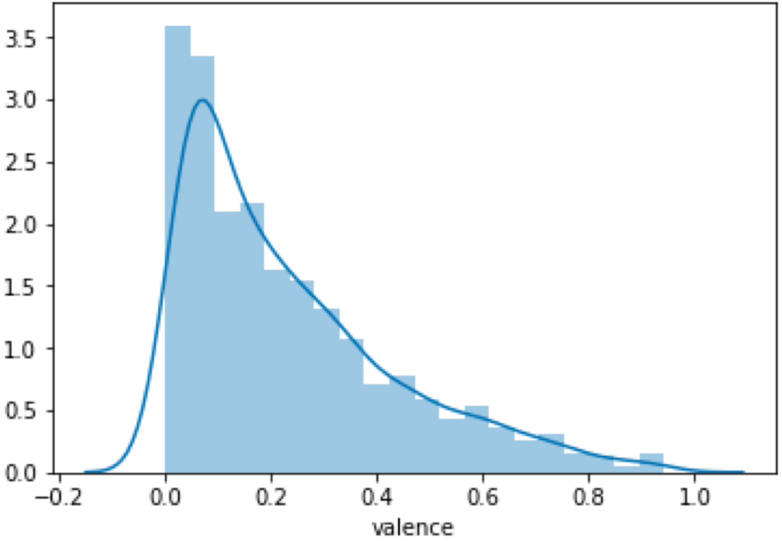
```
myplaylists.groupby(['playlist_name']).mean()
```

playlist_name	danceability	energy	key	loudness	speechiness	acousticness	instrumentalness	liveness	valence	tempo	duration_ms
Channel X	0.698860	0.593260	4.760000	-6.719760	0.138066	0.231788	0.000238	0.165296	0.456538	126.746540	207381.880000
Discover Weekly	0.636467	0.497667	4.933333	-8.614367	0.139517	0.365571	0.026400	0.149433	0.424187	120.862333	245481.000000
Hip-Hop Drive	0.731200	0.696190	5.440000	-6.165670	0.234807	0.099292	0.000275	0.225833	0.633284	118.821220	254855.780000
Mega Hit Mix	0.651480	0.631000	4.906667	-6.099800	0.097444	0.267512	0.008144	0.151508	0.437219	117.073653	206178.946667
Oldies	0.665025	0.631475	5.323232	-6.174889	0.118691	0.163340	0.000232	0.162580	0.550968	114.982793	242955.611111
TGIF	0.666557	0.689500	4.685714	-5.479671	0.105664	0.176040	0.005357	0.162774	0.490314	118.919886	191719.085714
Tastebreakers	0.592420	0.598860	5.080000	-7.572260	0.118448	0.283884	0.027732	0.174540	0.471680	118.319340	211553.920000
Todays Top Hits	0.710160	0.647580	6.080000	-5.750980	0.132612	0.198060	0.011192	0.167352	0.511860	114.461620	187614.160000

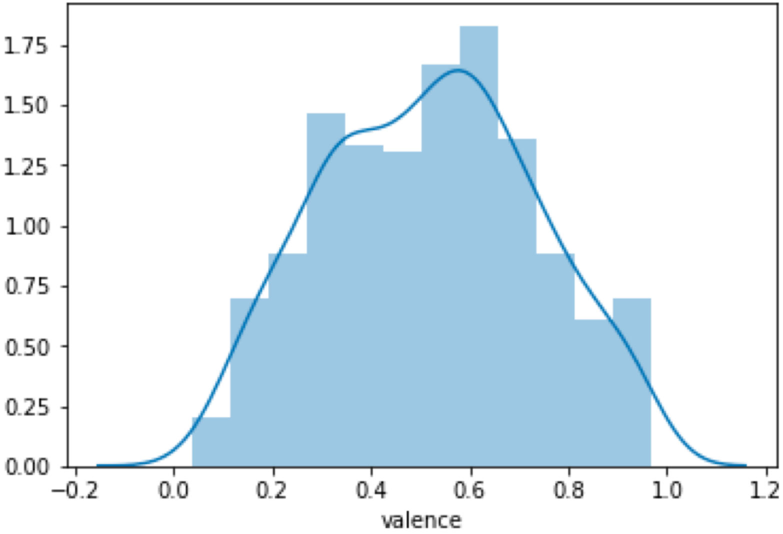
Study Playlists

My Playlists

```python
import seaborn as sns
```




```python
sns.distplot(studydf["valence"])
```

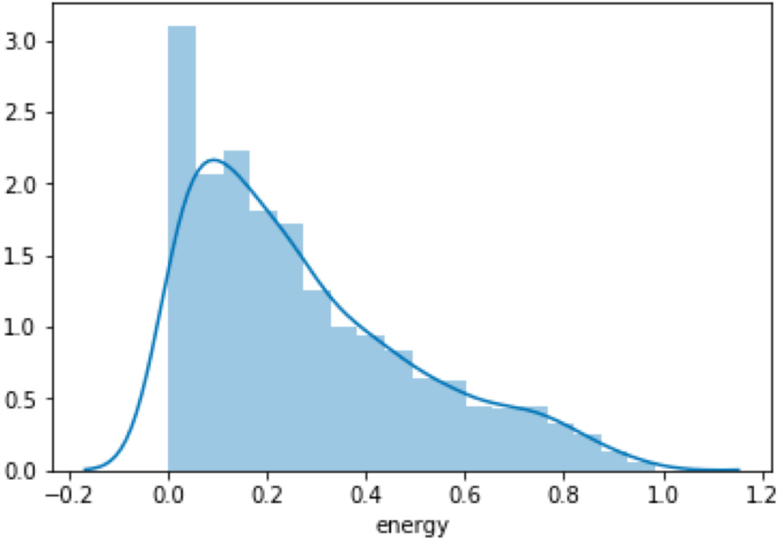



```python
sns.distplot(myplaylists['valence'])
```

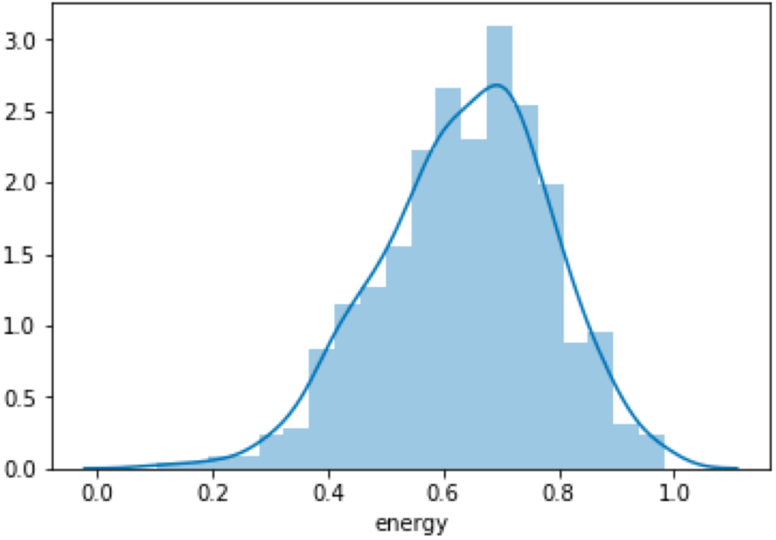
Valence = how positive(happy) or negative songs are

Study Playlists

My Playlists




```
sns.distplot(studydf['energy'])
```




```
sns.distplot(myplaylists['energy'])
```

Study Playlist Tweets

```
#create a data frame with specific information from tweets
tweetdata['id'] = [tweet['id'] for tweet in tweetlist]
tweetdata['user'] = [tweet['user']['screen_name'] for tweet in tweetlist]
tweetdata['text'] = [tweet['text'] for tweet in tweetlist]
tweetdata['lang'] = [tweet['lang'] for tweet in tweetlist]
tweetdata['place'] = [tweet['place'] for tweet in tweetlist]
tweetdata['retweet_count'] = [tweet['retweet_count'] for tweet in tweetlist]
tweetdata['favorite_count'] = [tweet['favorite_count'] for tweet in tweetlist]
```

```
tweetdata
```

	id	user	text	lang	place	retweet_count	favorite_count
0	1115922024562483206	StudentLifes98	#StudyPlaylist which somehow consists of rando...	en	None	0	0
1	1117914008495874049	DubStep_Edm_	RT @fluorescenceboi: #music #electronic #edm #...	und	None	4	0
2	1117909454186524675	ProtestMusica	RT @fluorescenceboi: #music #electronic #edm #...	und	None	4	0
3	1117900321638862849	BotLofi	RT @fluorescenceboi: #music #electronic #edm #...	und	None	4	0
4	1117887297406431234	fluorescenceboi	#music #electronic #edm #electro #remix #songr...	und	None	4	2
5	1117827461906038785	bgmchannelbgm	Deep Sleeping Music 👇 \nhttps://t.co/aVU3FTFvxr\...	en	None	0	0
6	1117712525540450305	HellaChuffed06	Crunch time work ahead of me this week. Time t...	en	None	0	2

Most Frequent Words

1. Import nltk and SentimentIntensityAnalyzer

1. Tokenize all the tweets
    a. 1975 tokens

1. Lower case all the tokens

1. Remove all the stopwords

1. Remove all unnecessary symbols

```
tweetFD = nltk.FreqDist(token_list)
top_words = tweetFD.most_common(30)
for word, freq in top_words:
    print(word, freq)
```

```
https 63
music 34
rt 25
electronic 16
remix 16
song 15
tasterdaysontour 15
edm 14
studymusic 14
electro 13
songremix 12
dubstep 12
bassdrop 12
songwriter 11
new 10
now 10
study_music_cam 9
got 9
fluorescenceboi 8
f4f 8
focus 8
study 8
deep 7
nowplaying 7
day 7
lofi 7
follow4follow 6
ambient 6
ta 6
track 6
```

Questions?