Dataset used: Top 100 Spotify Tracks

Source: https://www.kaggle.com/nadintamer/top-spotify-tracks-of-2018#top2018.csv

Original Data Source: The audio features for each song were extracted using the Spotify Web API and the spotipy Python library. Credit goes to Spotify for calculating the audio feature values.

Programming language of choice: Python

Dataset Description (Source: https://www.kaggle.com/nadintamer/top-spotify-tracks-of-2018#top2018.csv)**:**

Id: *Spotify URI of the song*
Name: *Name of the song*
Artists: *Artist(s) of the song*
Danceability: *Describes how suitable a track is for dancing*
Energy: *Measure from 0.0 to 1.0 and represents a perceptual measure of intensity and activity*
Key: *Integers map to pitches using standard Pitch Class notation. 0 = C, 1 = C♯/D♭, 2 = D etc.*
Loudness: *The overall loudness of a track in decibels (dB)*
Mode: *Indicates the modality (major or minor) of a track. Major = 1 and Minor = 0*
Speechiness: *Detects the presence of spoken words in a track*
Acousticness: *A confidence measure from 0.0 to 1.0 (high) of whether the track is acoustic*
Instrumentalness: *Predicts whether a track contains no vocals*
Liveness: *Detects the presence of an audience in the recording*
Valence: *A measure from 0.0 to 1.0 describing the musical positiveness*
Tempo: *beats per minute (BPM)*
duration_ms: *The duration of the track in milliseconds*
time_signature: *An estimated overall time signature of a track*

Analysis:

- Head of data:

	id	name	artists	danceability	energy	key	loudness	mode	speechiness	acousticness	instrumentalness	liveness	valence	tempo	duration_ms	time_signature
0	6DCZcSspjsKoFjzjrWoCd	God's Plan	Drake	0.754	0.449	7.0	-9.211	1.0	0.1090	0.0332	0.000083	0.552	0.357	77.169	198973.0	4.0
1	3ee8Jmje8o58CHK66QrVC	SAD!	XXXTENTACION	0.740	0.613	8.0	-4.880	1.0	0.1450	0.2580	0.003720	0.123	0.473	75.023	166606.0	4.0
2	0e7ipj03S05BNilyu5bRz	rockstar (feat. 21 Savage)	Post Malone	0.587	0.535	5.0	-6.090	0.0	0.0898	0.1170	0.000066	0.131	0.140	159.847	218147.0	4.0
3	3swc6WTsr7rl9DqQKQA55	Psycho (feat. Ty Dolla $ign)	Post Malone	0.739	0.559	8.0	-8.011	1.0	0.1170	0.5800	0.000000	0.112	0.439	140.124	221440.0	4.0
4	2G7V7zsVDxg1yRsu7Ew9R	In My Feelings	Drake	0.835	0.626	1.0	-5.833	1.0	0.1250	0.0589	0.000060	0.396	0.350	91.030	217925.0	4.0

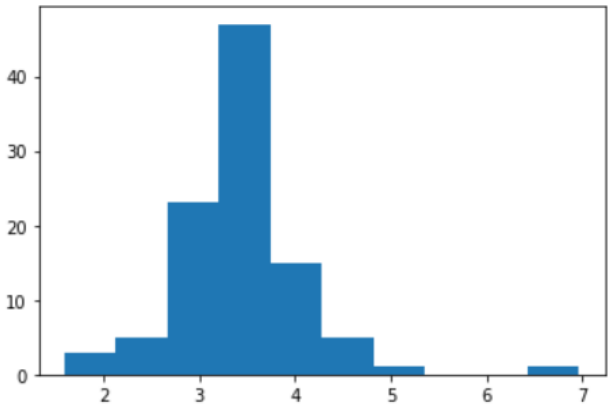
Shape: (100, 16)

- From below info it is observed that, none of the data is missing:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 100 entries, 0 to 99
Data columns (total 16 columns):
id                  100 non-null object
name                100 non-null object
artists             100 non-null object
danceability        100 non-null float64
energy              100 non-null float64
key                 100 non-null float64
loudness            100 non-null float64
mode                100 non-null float64
speechiness         100 non-null float64
acousticness        100 non-null float64
instrumentalness    100 non-null float64
liveness            100 non-null float64
valence             100 non-null float64
tempo               100 non-null float64
duration_ms         100 non-null float64
time_signature      100 non-null float64
dtypes: float64(13), object(3)
memory usage: 12.6+ KB
```

- Added a new column to rank the top 100 songs.

- Histogram plot to observe music duration trend after converting duration in milliseconds to minutes:



Inference:

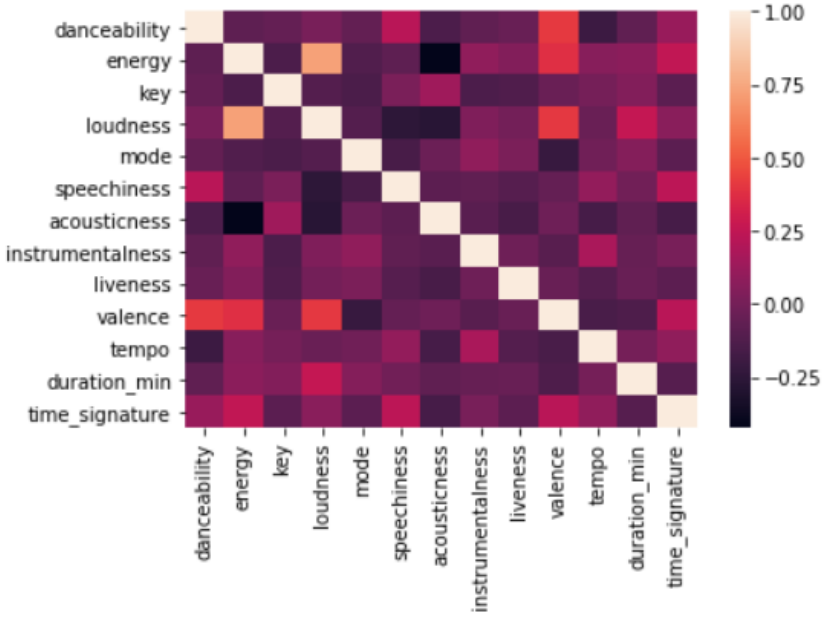
- From this plot, the music duration follows normal distribution, with most of the music having a duration between 3 to 4 minutes.
- There is one track with a duration greater than 6 minutes and is the song with highest duration.
- Identified the track with longest duration: “Te Bot? – Remix” by Nio Garcia with a duration of 6.96 minutes.

- Top 10 music artist with most songs in the top 100 music list are as follows:

  Result:

```
XXXTENTACION       6
Post Malone        6
Drake              4
Ed Sheeran         3
Marshmello         3
Ozuna              2
Imagine Dragons    2
Clean Bandit       2
Camila Cabello     2
The Weeknd         2
Name: artists, dtype: int64
```

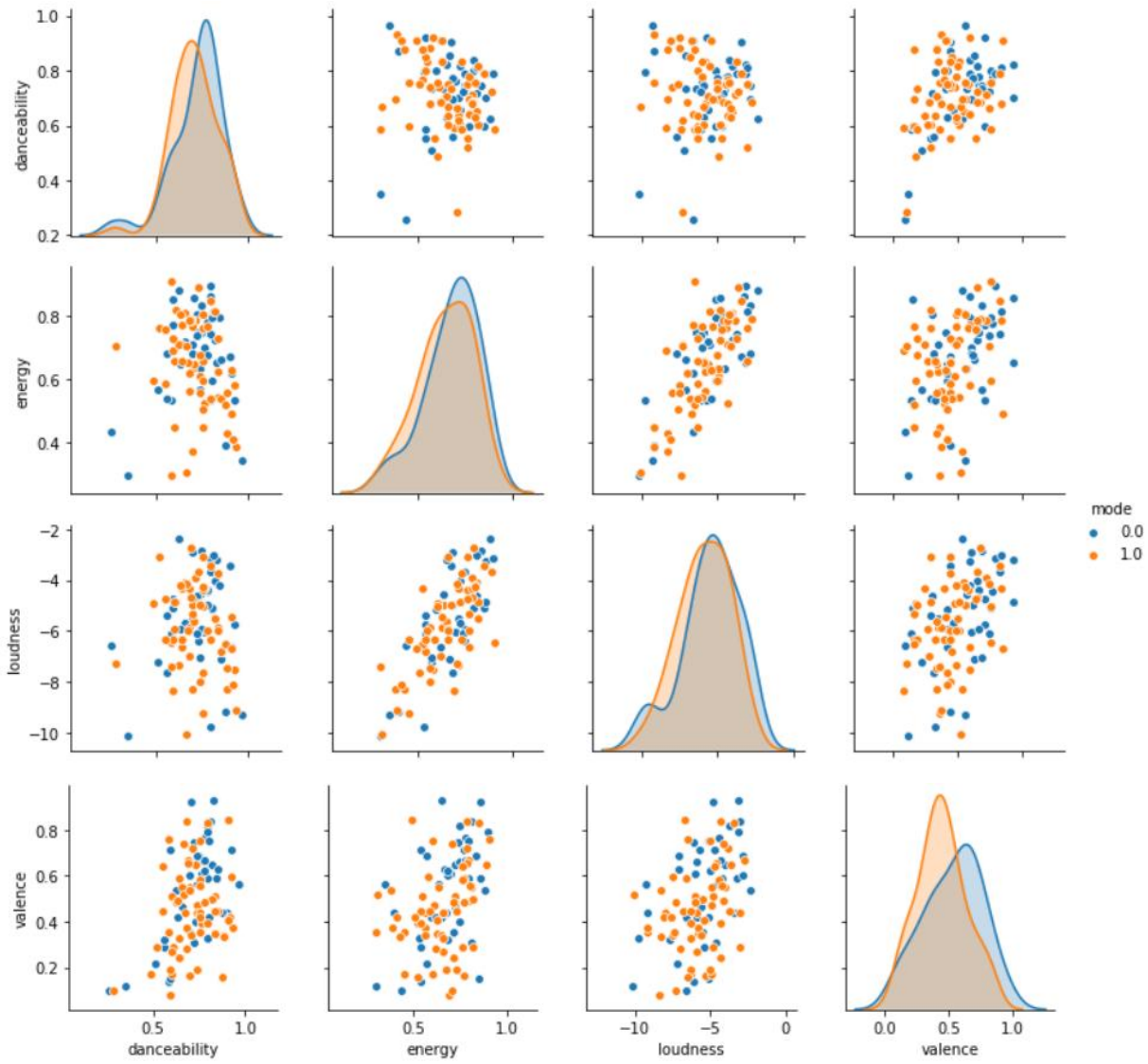
- Heatmap to observe the correlation of all the attributes:



Inference:

- From the heatmap it can be seen that “loudness” and “energy” are highly correlated.
- Valence, dancebility, energy and loudness are also correlated with each other but their correlation is not so strong.

- Pair plot of above four attributes to observe the relationship between each pair:



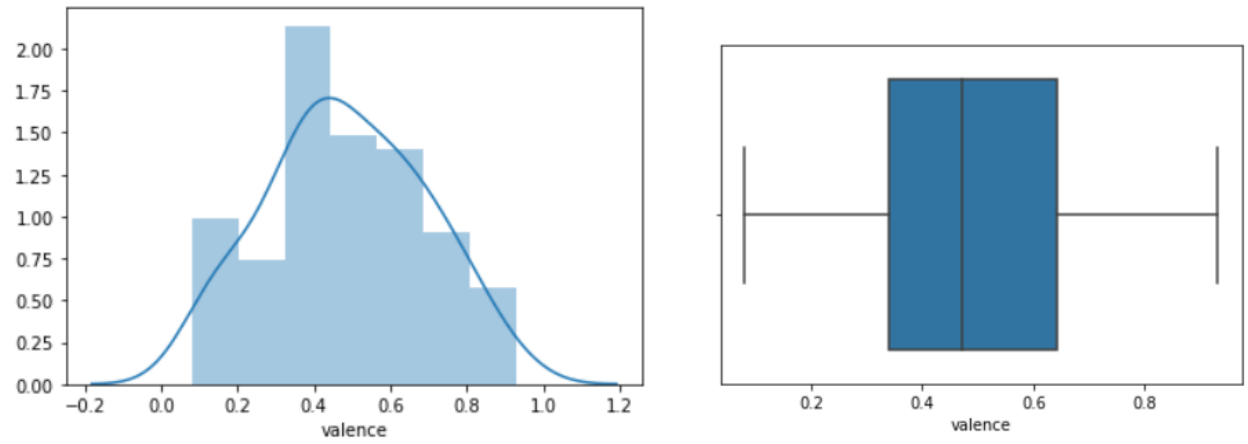
Inference:

- It is observed that energy and loudness are positively correlated.

- Analyzing each of the attributes individually:

  To explore the behavior of each of the attributes and to observe the pattern of those attributes in the top 100 tracks, plotting histograms for all relevant attributes.

Valence:



The valence plot shows normal distribution.

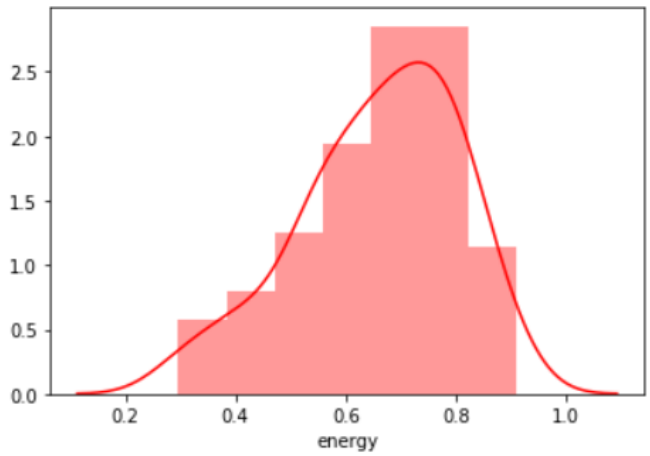
Inference:

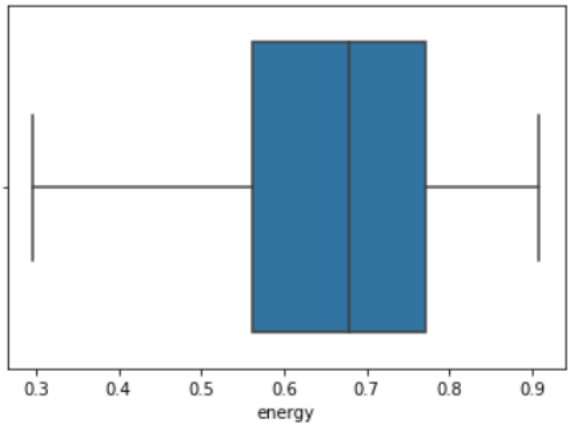
- From the plot, there seems to be equal number of songs with valence more than 0.5 and less than 0.5. To get exact figure checking the number of songs with valence > 0.5 , i.e. positive songs and valence < 0.5 i.e. negative songs.
- There are 45 positive (happier) songs and 55 negative (sad) songs.
- People are likely to listen both positive and negative songs at an equal rate

Top 10 happiest songs:

rank	name	artists	danceability	valence
26	Shape of You	Ed Sheeran	0.825	0.931
47	Finesse (Remix) [feat. Cardi B]	Bruno Mars	0.704	0.926
90	Bella	Wolfine	0.909	0.844
67	D?jala que vuelva (feat. Manuel Turizo)	Piso 21	0.681	0.839
79	Criminal	Natti Natasha	0.814	0.839
42	Dura	Daddy Yankee	0.791	0.828
60	Sin Pijama	Becky G	0.791	0.820
78	1, 2, 3 (feat. Jason Derulo & De La Ghetto)	Sofia Reyes	0.792	0.794
45	Me Niego	Reik	0.777	0.768
13	Nice For What	Drake	0.586	0.757

Energy:

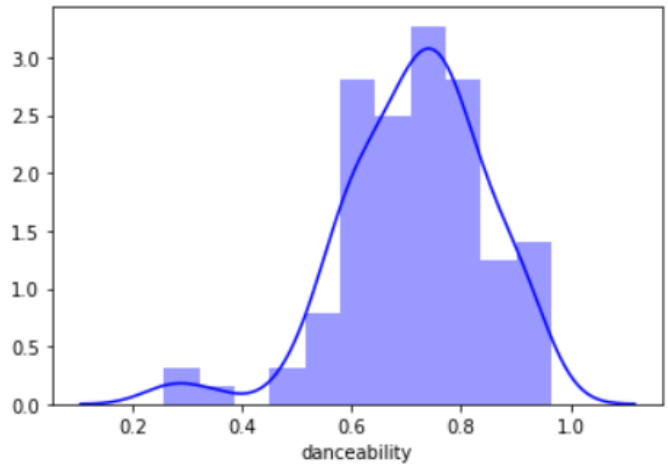


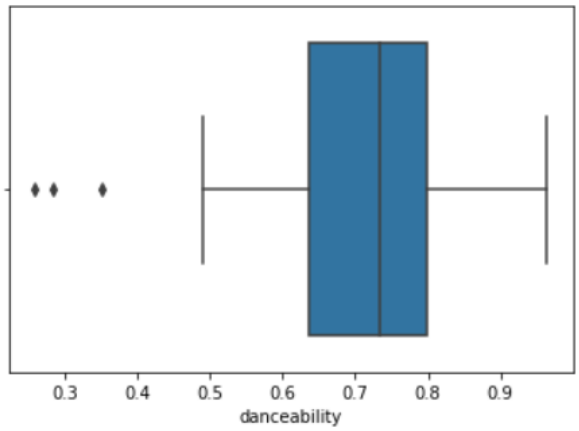


Inference:

- From the plot, there are more tracks with high energy than lower energies.
- People tend to listen to more songs with high energy.
- Most of the songs in top 100 have an energy of greater than 0.5

Danceability:





Inference:

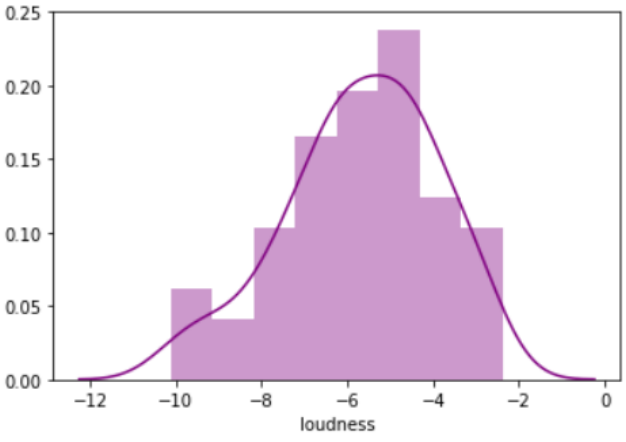
- From the plot, most of the songs in spotify's top 100 have high danceability. Thus, it can be inferred that people tend to like songs with high danceability.
- There are three outliers as per danceability attribute, which have very low danceability.
- To get the number of songs considered danceable,
  Considering, danceability > 0.75 as Very Danceable
  Danceability > 0.5 and < 0.75 as Danceable
  Danceability < 0.5 as Non Danceable

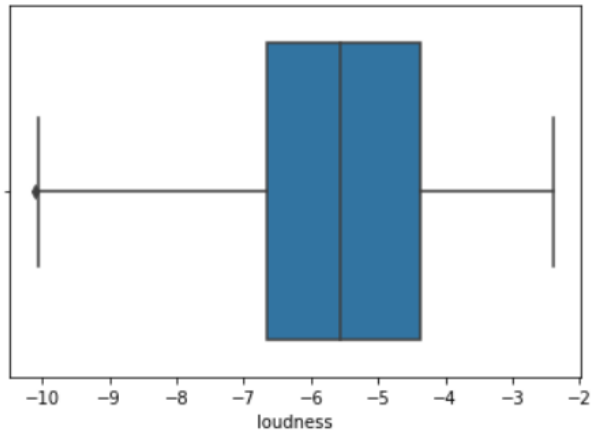
Very Danceable = 43
Danceable = 53
Non Danceable = 4

Top 10 most danceable songs:

rank	name	artists	danceability	valence	tempo
92	Yes Indeed	Lil Baby	0.964	0.562	119.958
56	FEFE (feat. Nicki Minaj & Murda Beatz)	6ix9ine	0.931	0.376	125.978
20	Look Alive (feat. Drake)	BlocBoy JB	0.922	0.595	140.022
19	Moonlight	XXXTENTACION	0.921	0.711	128.009
62	Nonstop	Drake	0.912	0.422	154.983
90	Bella	Wolfine	0.909	0.844	94.016
74	Walk It Talk It	Migos	0.909	0.406	145.905
83	HUMBLE.	Kendrick Lamar	0.908	0.421	150.011
22	Te Bot? - Remix	Nio Garcia	0.903	0.442	96.507
31	Taste (feat. Offset)	Tyga	0.884	0.342	97.994

Loudness:





Inference:

- From the plot, most of the songs in top 100 have loudness ranging from -4dB to -7dB
- There is one loudness outlier as per the boxplot, which has very less loudness.