

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

In [11]: import pandas as pd
data = pd.read_csv('C:/Users/STIDHARTH KULAL/OneDrive/Desktop/2nd sem/SpotifyFeatures1.csv')
data.head()
```

Out[11]:

	genre	artist_name	track_name	popularity	acousticness	danceability	duration_ms	energy	instrumentalness	liveness	loudness	speechiness	valence	release_date
0	Movie	Henri Salvador	C'est beau de faire un Show	0	0.611	0.389	99373	0.910	0.000	0.3460	-1.828	0.0525	0.814	7/24/2020
1	Movie	Martin & les fâces	Perdu d'avance (par Gad Elmaleh)	1	0.246	0.990	137373	0.737	0.000	0.1510	-5.599	0.0868	0.816	4/25/2000
2	Movie	Joseph Williams	Don't Let Me Be Lonely Tonight	3	0.952	0.663	152427	0.131	0.000	0.1030	-13.879	0.0362	0.368	7/24/2020
3	Movie	Henri Salvador	Dis-moi Monsieur Gordon Cooper	0	0.703	0.240	152427	0.326	0.000	0.0985	-12.178	0.0395	0.227	7/18/2000
4	Movie	Fabien Natat	Ouverture	4	0.950	0.331	82625	0.225	0.123	0.2020	-21.150	0.0456	0.390	01-01-2000

Out[12]:

```
data.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1528 entries, 0 to 1527
Data columns (total 14 columns):
 #   Column                Non-Null Count  Dtype
--  --
 0   genre                 1528 non-null   object
 1   artist_name          1528 non-null   object
 2   track_name           1528 non-null   object
 3   popularity            1528 non-null   float64
 4   acousticness          1528 non-null   float64
 5   danceability          1528 non-null   float64
 6   duration_ms          1528 non-null   int64
 7   energy                1528 non-null   float64
 8   instrumentalness      1528 non-null   float64
 9   liveness              1528 non-null   float64
10   loudness              1528 non-null   float64
11   speechiness           1528 non-null   float64
12   valence               1528 non-null   float64
13   release_date          1528 non-null   object
dtypes: float64(8), int64(2), object(4)
memory usage: 167.3+ KB
```

Out[13]:

```
#to check the shape of data
data.shape
```

Out[13]:

```
(1528, 14)
```

Out[14]:

```
data.describe()
<bound method NDFrame.describe of
0      genre                 1528 non-null   object
1      artist_name          1528 non-null   object
2      track_name           1528 non-null   object
3      popularity            1528 non-null   float64
4      acousticness          1528 non-null   float64
5      danceability          1528 non-null   float64
6      duration_ms          1528 non-null   int64
7      energy                1528 non-null   float64
8      instrumentalness      1528 non-null   float64
9      liveness              1528 non-null   float64
10     loudness              1528 non-null   float64
11     speechiness           1528 non-null   float64
12     valence               1528 non-null   float64
13     release_date          1528 non-null   object
dtypes: float64(8), int64(2), object(4)
memory usage: 167.3+ KB
1528 rows x 14 columns>
```

Out[15]:

```
#Checking the mean, count, std, min, max and percentage
data.describe()
<bound method NDFrame.describe of
0      genre                 1528 non-null   object
1      artist_name          1528 non-null   object
2      track_name           1528 non-null   object
3      popularity            1528 non-null   float64
4      acousticness          1528 non-null   float64
5      danceability          1528 non-null   float64
6      duration_ms          1528 non-null   int64
7      energy                1528 non-null   float64
8      instrumentalness      1528 non-null   float64
9      liveness              1528 non-null   float64
10     loudness              1528 non-null   float64
11     speechiness           1528 non-null   float64
12     valence               1528 non-null   float64
13     release_date          1528 non-null   object
dtypes: float64(8), int64(2), object(4)
memory usage: 167.3+ KB
1528 rows x 14 columns>
```

Out[16]:

```
#to check missing values
data.isnull().sum()
genre                0
artist_name          0
track_name           0
popularity            0
acousticness          0
danceability          0
duration_ms          0
energy                0
instrumentalness      0
liveness              0
loudness              0
speechiness           0
valence               0
release_date          0
dtype: object
```

Out[17]:

```
#to check the columns of data
data.columns
Index(['genre', 'artist_name', 'track_name', 'popularity', 'acousticness',
       'danceability', 'duration_ms', 'energy', 'instrumentalness', 'liveness',
       'loudness', 'speechiness', 'valence', 'release_date'],
      dtype='object')
```

Out[18]:

```
data.dtypes
genre                object
artist_name          object
track_name           object
popularity            float64
acousticness          float64
danceability          float64
duration_ms          int64
energy                float64
instrumentalness      float64
liveness              float64
loudness              float64
speechiness           float64
valence               float64
release_date          object
dtype: object
```

Out[19]:

```
data.dropna(subset=['valence'], inplace=True)
data.head()
genre                artist_name          track_name          popularity  acousticness  danceability  duration_ms  energy  instrumentalness  liveness  loudness  speechiness  valence  release_date
0  Movie  Henri Salvador  C'est beau de faire un Show      0          0.611      0.389      99373      0.910      0.000      0.3460      -1.828      0.0525      0.814      7/24/2020
1  Movie  Martin & les fâces  Perdu d'avance (par Gad Elmaleh)  1          0.246      0.990      137373      0.737      0.000      0.1510      -5.599      0.0868      0.816      4/25/2000
2  Movie  Joseph Williams  Don't Let Me Be Lonely Tonight    3          0.952      0.663      152427      0.131      0.000      0.1030      -13.879      0.0362      0.368      7/24/2020
3  Movie  Henri Salvador  Dis-moi Monsieur Gordon Cooper    0          0.703      0.240      152427      0.326      0.000      0.0985      -12.178      0.0395      0.227      7/18/2000
4  Movie  Fabien Natat    Ouverture                4          0.950      0.331      82625      0.225      0.123      0.2020      -21.150      0.0456      0.390      01-01-2000
```

Out[20]:

```
#Checking the information of dataset
data.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1528 entries, 0 to 1527
Data columns (total 14 columns):
 #   Column                Non-Null Count  Dtype
--  --
 0   genre                 1528 non-null   object
 1   artist_name          1528 non-null   object
 2   track_name           1528 non-null   object
 3   popularity            1528 non-null   float64
 4   acousticness          1528 non-null   float64
 5   danceability          1528 non-null   float64
 6   duration_ms          1528 non-null   int64
 7   energy                1528 non-null   float64
 8   instrumentalness      1528 non-null   float64
 9   liveness              1528 non-null   float64
10   loudness              1528 non-null   float64
11   speechiness           1528 non-null   float64
12   valence               1528 non-null   float64
13   release_date          1528 non-null   object
dtypes: float64(8), int64(2), object(4)
memory usage: 167.3+ KB
```

Out[21]:

```
#convert the duration from milliseconds to seconds
data['duration_ms'] = data['duration_ms'].apply(lambda x:round(x/1000))
data.drop("duration_ms",inplace=True,axis=1)
data.head()
genre                artist_name          track_name          popularity  acousticness  danceability  duration_ms  energy  instrumentalness  liveness  loudness  speechiness  valence  release_date  duration
0  Movie  Henri Salvador  C'est beau de faire un Show      0          0.611      0.389      99.373      0.910      0.000      0.3460      -1.828      0.0525      0.814      7/24/2020      99
1  Movie  Martin & les fâces  Perdu d'avance (par Gad Elmaleh)  1          0.246      0.990      137.373      0.737      0.000      0.1510      -5.599      0.0868      0.816      4/25/2000      137
2  Movie  Joseph Williams  Don't Let Me Be Lonely Tonight    3          0.952      0.663      152.427      0.131      0.000      0.1030      -13.879      0.0362      0.368      7/24/2020      170
3  Movie  Henri Salvador  Dis-moi Monsieur Gordon Cooper    0          0.703      0.240      152.427      0.326      0.000      0.0985      -12.178      0.0395      0.227      7/18/2000      152
4  Movie  Fabien Natat    Ouverture                4          0.950      0.331      82.625      0.225      0.123      0.2020      -21.150      0.0456      0.390      01-01-2000      83
```

Out[22]:

```
#Converting release_date to index
data.set_index('release_date',inplace=True)
data.index = pd.to_datetime(data.index,infer_datetime_format=True)
data.head()
genre                artist_name          track_name          popularity  acousticness  danceability  energy  instrumentalness  liveness  loudness  speechiness  valence  duration
release_date
2020-07-24  Movie  Henri Salvador  C'est beau de faire un Show      0          0.611      0.389      99.373      0.910      0.000      0.3460      -1.828      0.0525      0.814      99
2020-04-25  Movie  Martin & les fâces  Perdu d'avance (par Gad Elmaleh)  1          0.246      0.990      137.373      0.737      0.000      0.1510      -5.599      0.0868      0.816      137
2020-07-24  Movie  Joseph Williams  Don't Let Me Be Lonely Tonight    3          0.952      0.663      152.427      0.131      0.000      0.1030      -13.879      0.0362      0.368      170
2020-07-18  Movie  Henri Salvador  Dis-moi Monsieur Gordon Cooper    0          0.703      0.240      152.427      0.326      0.000      0.0985      -12.178      0.0395      0.227      152
2020-01-01  Movie  Fabien Natat    Ouverture                4          0.950      0.331      82.625      0.225      0.123      0.2020      -21.150      0.0456      0.390      83
```

Out[23]:

```
# Music Feature correlation heatmap
fig=plt.figure(figsize=(14,6))
sns.heatmap(data.corr(numeric_only=True),cmap='inferno',annot=True)
<Axes: >
```

Out[23]:

Out[24]:

Out[24]:

Out[25]:

Out[25]:

Out[26]:

Out[26]:

Out[27]:

Out[27]:

Out[28]:

Out[29]:

Out[29]:

Out[30]:

Out[30]:

Out[31]:

Out[31]:

Out[31]:

Out[32]:

Out[32]:

Out[33]:

Out[33]: