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

df = pd.read_csv('spotify dataset.csv')
df
```

```
                      track_id  \
0      6f807x0ima9a1j3VPbc7VN
1      0r7CVbZTWZgbTCYdfa2P31
2      1z1Hg7Vb0AhHDiEmnDE79l
3      75FpbthrwQmzHlBJLuGdC7
4      1e8PAfcKUYoKkxPhrHqw4x
...                       ...
32828  7bxnKAamR3snQ1VGLuVfC1
32829  5Aevni09Em4575077nkWHz
32830  7ImMqPP3Q1yfUHvsdn7wEo
32831  2m69mhnfQ1Oq6lGtXuYhgX
32832  29zWqhca3zt5NsckZqDf6c

                                                 track_name
track_artist  \
0      I Don't Care (with Justin Bieber) - Loud Luxur...         Ed
Sheeran
1                          Memories - Dillon Francis Remix
Maroon 5
2                          All the Time - Don Diablo Remix      Zara
Larsson
3                       Call You Mine - Keanu Silva Remix  The
Chainsmokers
4                  Someone You Loved - Future Humans Remix     Lewis
Capaldi
...                                                   ...
...
32828                 City Of Lights - Official Radio Edit      Lush &
Simon
32829                  Closer - Sultan & Ned Shepard Remix    Tegan and
Sara
32830                        Sweet Surrender - Radio Edit
Starkillers
32831                       Only For You - Maor Levi Remix
Mat Zo
32832                               Typhoon - Original Mix      Julian
Calor

       track_popularity          track_album_id  \
0                     66  2oCs0DGTsRO98Gh5ZSl2Cx
1                     67  63rPSO264uRjW1X5E6cWv6
2                     70  1HoSmj2eLcsrR0vE9gThr4
```

```
3                    60  1nqYsOef1yKKuGOVchbsk6
4                    69  7m7vv9wlQ4i0LFuJiE2zsQ
...                 ...                     ...
32828                42  2azRoBBWEEEYhqV6sb7JrT
32829                20  6kD6KLxj7s8eCE3ABvAyf5
32830                14  0ltWNSY9JgxoIZO4VzuCa6
32831                15  1fGrOkHnHJcStl14zNx8Jy
32832                27  0X3mUOm6MhxR7PzxG95rAo

                                      track_album_name  \
0      I Don't Care (with Justin Bieber) [Loud Luxury...
1                         Memories (Dillon Francis Remix)
2                         All the Time (Don Diablo Remix)
3                             Call You Mine - The Remixes
4                 Someone You Loved (Future Humans Remix)
...                                                   ...
32828                          City Of Lights (Vocal Mix)
32829                                       Closer Remixed
32830                        Sweet Surrender (Radio Edit)
32831                             Only For You (Remixes)
32832                                      Typhoon/Storm

      track_album_release_date    playlist_name
playlist_id  \
0                   2019-06-14        Pop Remix
37i9dQZF1DXcZDD7cfEKhW
1                   2019-12-13        Pop Remix
37i9dQZF1DXcZDD7cfEKhW
2                   2019-07-05        Pop Remix
37i9dQZF1DXcZDD7cfEKhW
3                   2019-07-19        Pop Remix
37i9dQZF1DXcZDD7cfEKhW
4                   2019-03-05        Pop Remix
37i9dQZF1DXcZDD7cfEKhW
...                        ...               ...                       ..
.
32828               2014-04-28  ♥ EDM LOVE 2020
6jI1gFr6ANFtT8MmTvA2Ux
32829               2013-03-08  ♥ EDM LOVE 2020
6jI1gFr6ANFtT8MmTvA2Ux
32830               2014-04-21  ♥ EDM LOVE 2020
6jI1gFr6ANFtT8MmTvA2Ux
32831               2014-01-01  ♥ EDM LOVE 2020
6jI1gFr6ANFtT8MmTvA2Ux
32832               2014-03-03  ♥ EDM LOVE 2020
6jI1gFr6ANFtT8MmTvA2Ux

      playlist_genre  ... key  loudness  mode  speechiness
acousticness  \
0                 pop  ...   6    -2.634     1       0.0583
```

```
0.102000
1                 pop  ...  11    -4.969     1       0.0373
0.072400
2                 pop  ...   1    -3.432     0       0.0742
0.079400
3                 pop  ...   7    -3.778     1       0.1020
0.028700
4                 pop  ...   1    -4.672     1       0.0359
0.080300
...               ...  ...  ..       ...   ...          ...            .
..
32828             edm  ...   2    -1.814     1       0.0936
0.076600
32829             edm  ...   0    -4.462     1       0.0420
0.001710
32830             edm  ...   6    -4.899     0       0.0481
0.108000
32831             edm  ...   2    -3.361     1       0.1090
0.007920
32832             edm  ...   5    -4.571     0       0.0385
0.000133

       instrumentalness  liveness  valence    tempo  duration_ms
0              0.000000    0.0653   0.5180  122.036       194754
1              0.004210    0.3570   0.6930   99.972       162600
2              0.000023    0.1100   0.6130  124.008       176616
3              0.000009    0.2040   0.2770  121.956       169093
4              0.000000    0.0833   0.7250  123.976       189052
...                 ...       ...      ...      ...          ...
32828          0.000000    0.0668   0.2100  128.170       204375
32829          0.004270    0.3750   0.4000  128.041       353120
32830          0.000001    0.1500   0.4360  127.989       210112
32831          0.127000    0.3430   0.3080  128.008       367432
32832          0.341000    0.7420   0.0894  127.984       337500

[32833 rows x 23 columns]
```

Data Preprocessing

```
df.head()

                 track_id
track_name  \
0  6f807x0ima9a1j3VPbc7VN  I Don't Care (with Justin Bieber) - Loud
Luxur...
1  0r7CVbZTWZgbTCYdfa2P31                      Memories - Dillon Francis
Remix
2  1z1Hg7Vb0AhHDiEmnDE79l                      All the Time - Don Diablo
```

```
Remix
3  75FpbthrwQmzHlBJLuGdC7                  Call You Mine - Keanu Silva
Remix
4  1e8PAfcKUYoKkxPhrHqw4x            Someone You Loved - Future Humans
Remix

       track_artist  track_popularity          track_album_id  \
0        Ed Sheeran                66  2oCs0DGTsRO98Gh5ZSl2Cx
1          Maroon 5                67  63rPSO264uRjW1X5E6cWv6
2      Zara Larsson                70  1HoSmj2eLcsrR0vE9gThr4
3  The Chainsmokers                60  1nqYsOef1yKKuGOVchbsk6
4     Lewis Capaldi                69  7m7vv9wlQ4i0LFuJiE2zsQ

                                    track_album_name
track_album_release_date  \
0  I Don't Care (with Justin Bieber) [Loud Luxury...
2019-06-14
1                     Memories (Dillon Francis Remix)
2019-12-13
2                     All the Time (Don Diablo Remix)
2019-07-05
3                         Call You Mine - The Remixes
2019-07-19
4            Someone You Loved (Future Humans Remix)
2019-03-05

  playlist_name             playlist_id playlist_genre  ... key
loudness  \
0     Pop Remix  37i9dQZF1DXcZDD7cfEKhW            pop  ...   6    -
2.634
1     Pop Remix  37i9dQZF1DXcZDD7cfEKhW            pop  ...  11    -
4.969
2     Pop Remix  37i9dQZF1DXcZDD7cfEKhW            pop  ...   1    -
3.432
3     Pop Remix  37i9dQZF1DXcZDD7cfEKhW            pop  ...   7    -
3.778
4     Pop Remix  37i9dQZF1DXcZDD7cfEKhW            pop  ...   1    -
4.672

   mode  speechiness  acousticness  instrumentalness  liveness
valence  \
0     1       0.0583        0.1020          0.000000    0.0653
0.518
1     1       0.0373        0.0724          0.004210    0.3570
0.693
2     0       0.0742        0.0794          0.000023    0.1100
0.613
3     1       0.1020        0.0287          0.000009    0.2040
0.277
4     1       0.0359        0.0803          0.000000    0.0833
```

```
0.725

     tempo  duration_ms
0  122.036       194754
1   99.972       162600
2  124.008       176616
3  121.956       169093
4  123.976       189052

[5 rows x 23 columns]
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 32833 entries, 0 to 32832
Data columns (total 23 columns):
 #   Column                    Non-Null Count  Dtype  
---  ------                    --------------  -----  
 0   track_id                  32833 non-null  object 
 1   track_name                32828 non-null  object 
 2   track_artist              32828 non-null  object 
 3   track_popularity          32833 non-null  int64  
 4   track_album_id            32833 non-null  object 
 5   track_album_name          32828 non-null  object 
 6   track_album_release_date  32833 non-null  object 
 7   playlist_name             32833 non-null  object 
 8   playlist_id               32833 non-null  object 
 9   playlist_genre            32833 non-null  object 
 10  playlist_subgenre         32833 non-null  object 
 11  danceability              32833 non-null  float64
 12  energy                    32833 non-null  float64
 13  key                       32833 non-null  int64  
 14  loudness                  32833 non-null  float64
 15  mode                      32833 non-null  int64  
 16  speechiness               32833 non-null  float64
 17  acousticness              32833 non-null  float64
 18  instrumentalness          32833 non-null  float64
 19  liveness                  32833 non-null  float64
 20  valence                   32833 non-null  float64
 21  tempo                     32833 non-null  float64
 22  duration_ms               32833 non-null  int64  
dtypes: float64(9), int64(4), object(10)
memory usage: 5.8+ MB
```

```
df.describe()
```

```
       track_popularity  danceability        energy           key  \
count      32833.000000  32833.000000  32833.000000  32833.000000   
mean          42.477081      0.654850      0.698619      5.374471   
std           24.984074      0.145085      0.180910      3.611657   
```

```
min            0.000000      0.000000      0.000175      0.000000
25%           24.000000      0.563000      0.581000      2.000000
50%           45.000000      0.672000      0.721000      6.000000
75%           62.000000      0.761000      0.840000      9.000000
max          100.000000      0.983000      1.000000     11.000000

           loudness          mode   speechiness  acousticness  \
count  32833.000000  32833.000000  32833.000000  32833.000000
mean      -6.719499      0.565711      0.107068      0.175334
std        2.988436      0.495671      0.101314      0.219633
min      -46.448000      0.000000      0.000000      0.000000
25%       -8.171000      0.000000      0.041000      0.015100
50%       -6.166000      1.000000      0.062500      0.080400
75%       -4.645000      1.000000      0.132000      0.255000
max        1.275000      1.000000      0.918000      0.994000

       instrumentalness      liveness       valence         tempo  \
count      32833.000000  32833.000000  32833.000000  32833.000000
mean           0.084747      0.190176      0.510561    120.881132
std            0.224230      0.154317      0.233146     26.903624
min            0.000000      0.000000      0.000000      0.000000
25%            0.000000      0.092700      0.331000     99.960000
50%            0.000016      0.127000      0.512000    121.984000
75%            0.004830      0.248000      0.693000    133.918000
max            0.994000      0.996000      0.991000    239.440000

         duration_ms
count   32833.000000
mean   225799.811622
std     59834.006182
min      4000.000000
25%    187819.000000
50%    216000.000000
75%    253585.000000
max    517810.000000
```

```
df.isnull().sum()
```

```
track_id                    0
track_name                  5
track_artist                5
track_popularity            0
track_album_id              0
track_album_name            5
track_album_release_date    0
playlist_name               0
playlist_id                 0
playlist_genre              0
playlist_subgenre           0
danceability                0
```

```
energy                      0
key                         0
loudness                    0
mode                        0
speechiness                 0
acousticness                0
instrumentalness            0
liveness                    0
valence                     0
tempo                       0
duration_ms                 0
dtype: int64
```

```
df.drop_duplicates()
```

```
                      track_id  \
0      6f807x0ima9a1j3VPbc7VN
1      0r7CVbZTWZgbTCYdfa2P31
2      1z1Hg7Vb0AhHDiEmnDE79l
3      75FpbthrwQmzHlBJLuGdC7
4      1e8PAfcKUYoKkxPhrHqw4x
...                       ...
32828  7bxnKAamR3snQ1VGLuVfC1
32829  5Aevni09Em4575077nkWHz
32830  7ImMqPP3Q1yfUHvsdn7wEo
32831  2m69mhnfQ1Oq6lGtXuYhgX
32832  29zWqhca3zt5NsckZqDf6c

                                                 track_name
track_artist  \
0      I Don't Care (with Justin Bieber) - Loud Luxur...        Ed
Sheeran
1                          Memories - Dillon Francis Remix
Maroon 5
2                          All the Time - Don Diablo Remix      Zara
Larsson
3                       Call You Mine - Keanu Silva Remix  The
Chainsmokers
4                  Someone You Loved - Future Humans Remix     Lewis
Capaldi
...                                                   ...
...
32828                 City Of Lights - Official Radio Edit      Lush &
Simon
32829                  Closer - Sultan & Ned Shepard Remix    Tegan and
Sara
32830                        Sweet Surrender - Radio Edit
Starkillers
32831                       Only For You - Maor Levi Remix
Mat Zo
```

```
32832                              Typhoon - Original Mix      Julian
Calor

       track_popularity          track_album_id  \
0                     66  2oCs0DGTsRO98Gh5ZSl2Cx
1                     67  63rPSO264uRjW1X5E6cWv6
2                     70  1HoSmj2eLcsrR0vE9gThr4
3                     60  1nqYsOef1yKKuGOVchbsk6
4                     69  7m7vv9wlQ4i0LFuJiE2zsQ
...                  ...                     ...
32828                 42  2azRoBBWEEEYhqV6sb7JrT
32829                 20  6kD6KLxj7s8eCE3ABvAyf5
32830                 14  0ltWNSY9JgxoIZO4VzuCa6
32831                 15  1fGrOkHnHJcStl14zNx8Jy
32832                 27  0X3mUOm6MhxR7PzxG95rAo

                                        track_album_name  \
0      I Don't Care (with Justin Bieber) [Loud Luxury...
1                          Memories (Dillon Francis Remix)
2                          All the Time (Don Diablo Remix)
3                              Call You Mine - The Remixes
4                 Someone You Loved (Future Humans Remix)
...                                                   ...
32828                           City Of Lights (Vocal Mix)
32829                                        Closer Remixed
32830                         Sweet Surrender (Radio Edit)
32831                              Only For You (Remixes)
32832                                       Typhoon/Storm

      track_album_release_date    playlist_name
playlist_id  \
0                   2019-06-14        Pop Remix
37i9dQZF1DXcZDD7cfEKhW
1                   2019-12-13        Pop Remix
37i9dQZF1DXcZDD7cfEKhW
2                   2019-07-05        Pop Remix
37i9dQZF1DXcZDD7cfEKhW
3                   2019-07-19        Pop Remix
37i9dQZF1DXcZDD7cfEKhW
4                   2019-03-05        Pop Remix
37i9dQZF1DXcZDD7cfEKhW
...                        ...              ...                       ..
.
32828               2014-04-28  ♥ EDM LOVE 2020
6jI1gFr6ANFtT8MmTvA2Ux
32829               2013-03-08  ♥ EDM LOVE 2020
6jI1gFr6ANFtT8MmTvA2Ux
32830               2014-04-21  ♥ EDM LOVE 2020
6jI1gFr6ANFtT8MmTvA2Ux
32831               2014-01-01  ♥ EDM LOVE 2020
```

```
6jI1gFr6ANFtT8MmTvA2Ux
32832               2014-03-03  ♥ EDM LOVE 2020
6jI1gFr6ANFtT8MmTvA2Ux

      playlist_genre  ... key  loudness  mode  speechiness
acousticness  \
0                pop  ...   6    -2.634     1       0.0583
0.102000
1                pop  ...  11    -4.969     1       0.0373
0.072400
2                pop  ...   1    -3.432     0       0.0742
0.079400
3                pop  ...   7    -3.778     1       0.1020
0.028700
4                pop  ...   1    -4.672     1       0.0359
0.080300
...              ...  ...  ..       ...   ...          ...             .
..
32828            edm  ...   2    -1.814     1       0.0936
0.076600
32829            edm  ...   0    -4.462     1       0.0420
0.001710
32830            edm  ...   6    -4.899     0       0.0481
0.108000
32831            edm  ...   2    -3.361     1       0.1090
0.007920
32832            edm  ...   5    -4.571     0       0.0385
0.000133

       instrumentalness  liveness  valence    tempo  duration_ms
0              0.000000    0.0653   0.5180  122.036       194754
1              0.004210    0.3570   0.6930   99.972       162600
2              0.000023    0.1100   0.6130  124.008       176616
3              0.000009    0.2040   0.2770  121.956       169093
4              0.000000    0.0833   0.7250  123.976       189052
...                 ...       ...      ...      ...          ...
32828          0.000000    0.0668   0.2100  128.170       204375
32829          0.004270    0.3750   0.4000  128.041       353120
32830          0.000001    0.1500   0.4360  127.989       210112
32831          0.127000    0.3430   0.3080  128.008       367432
32832          0.341000    0.7420   0.0894  127.984       337500

[32833 rows x 23 columns]

df.dropna()

                     track_id  \
0      6f807x0ima9a1j3VPbc7VN
1      0r7CVbZTWZgbTCYdfa2P31
2      1z1Hg7Vb0AhHDiEmnDE79l
```

```
3      75FpbthrwQmzHlBJLuGdC7
4      1e8PAfcKUYoKkxPhrHqw4x
...                       ...
32828  7bxnKAamR3snQ1VGLuVfC1
32829  5Aevni09Em4575077nkWHz
32830  7ImMqPP3Q1yfUHvsdn7wEo
32831  2m69mhnfQ1Oq6lGtXuYhgX
32832  29zWqhca3zt5NsckZqDf6c

                                                 track_name
track_artist  \
0      I Don't Care (with Justin Bieber) - Loud Luxur...         Ed
Sheeran
1                         Memories - Dillon Francis Remix
Maroon 5
2                         All the Time - Don Diablo Remix      Zara
Larsson
3                       Call You Mine - Keanu Silva Remix  The
Chainsmokers
4                  Someone You Loved - Future Humans Remix     Lewis
Capaldi
...                                                   ...
...
32828                City Of Lights - Official Radio Edit      Lush &
Simon
32829                 Closer - Sultan & Ned Shepard Remix    Tegan and
Sara
32830                        Sweet Surrender - Radio Edit
Starkillers
32831                       Only For You - Maor Levi Remix
Mat Zo
32832                               Typhoon - Original Mix      Julian
Calor

       track_popularity          track_album_id  \
0                     66  2oCs0DGTsRO98Gh5ZSl2Cx
1                     67  63rPSO264uRjW1X5E6cWv6
2                     70  1HoSmj2eLcsrR0vE9gThr4
3                     60  1nqYsOef1yKKuGOVchbsk6
4                     69  7m7vv9wlQ4i0LFuJiE2zsQ
...                  ...                     ...
32828                 42  2azRoBBWEEEYhqV6sb7JrT
32829                 20  6kD6KLxj7s8eCE3ABvAyf5
32830                 14  0ltWNSY9JgxoIZO4VzuCa6
32831                 15  1fGrOkHnHJcStl14zNx8Jy
32832                 27  0X3mUOm6MhxR7PzxG95rAo

                                          track_album_name  \
0      I Don't Care (with Justin Bieber) [Loud Luxury...
1                         Memories (Dillon Francis Remix)
```

```
2                          All the Time (Don Diablo Remix)
3                              Call You Mine - The Remixes
4                  Someone You Loved (Future Humans Remix)
...                                                      ...
32828                            City Of Lights (Vocal Mix)
32829                                        Closer Remixed
32830                          Sweet Surrender (Radio Edit)
32831                               Only For You (Remixes)
32832                                        Typhoon/Storm

      track_album_release_date    playlist_name
playlist_id  \
0                   2019-06-14        Pop Remix
37i9dQZF1DXcZDD7cfEKhW
1                   2019-12-13        Pop Remix
37i9dQZF1DXcZDD7cfEKhW
2                   2019-07-05        Pop Remix
37i9dQZF1DXcZDD7cfEKhW
3                   2019-07-19        Pop Remix
37i9dQZF1DXcZDD7cfEKhW
4                   2019-03-05        Pop Remix
37i9dQZF1DXcZDD7cfEKhW
...                        ...              ...                      ..
.
32828               2014-04-28  ♥ EDM LOVE 2020
6jI1gFr6ANFtT8MmTvA2Ux
32829               2013-03-08  ♥ EDM LOVE 2020
6jI1gFr6ANFtT8MmTvA2Ux
32830               2014-04-21  ♥ EDM LOVE 2020
6jI1gFr6ANFtT8MmTvA2Ux
32831               2014-01-01  ♥ EDM LOVE 2020
6jI1gFr6ANFtT8MmTvA2Ux
32832               2014-03-03  ♥ EDM LOVE 2020
6jI1gFr6ANFtT8MmTvA2Ux

      playlist_genre  ... key  loudness  mode  speechiness
acousticness  \
0                 pop  ...   6    -2.634     1       0.0583
0.102000
1                 pop  ...  11    -4.969     1       0.0373
0.072400
2                 pop  ...   1    -3.432     0       0.0742
0.079400
3                 pop  ...   7    -3.778     1       0.1020
0.028700
4                 pop  ...   1    -4.672     1       0.0359
0.080300
...               ...  ...  ..       ...   ...          ...           .
..
```

```
32828            edm  ...   2    -1.814     1       0.0936
0.076600
32829            edm  ...   0    -4.462     1       0.0420
0.001710
32830            edm  ...   6    -4.899     0       0.0481
0.108000
32831            edm  ...   2    -3.361     1       0.1090
0.007920
32832            edm  ...   5    -4.571     0       0.0385
0.000133

       instrumentalness  liveness  valence    tempo  duration_ms
0              0.000000    0.0653   0.5180  122.036       194754
1              0.004210    0.3570   0.6930   99.972       162600
2              0.000023    0.1100   0.6130  124.008       176616
3              0.000009    0.2040   0.2770  121.956       169093
4              0.000000    0.0833   0.7250  123.976       189052
...                 ...       ...      ...      ...          ...
32828          0.000000    0.0668   0.2100  128.170       204375
32829          0.004270    0.3750   0.4000  128.041       353120
32830          0.000001    0.1500   0.4360  127.989       210112
32831          0.127000    0.3430   0.3080  128.008       367432
32832          0.341000    0.7420   0.0894  127.984       337500

[32828 rows x 23 columns]

from sklearn.preprocessing import LabelEncoder
label_encoder = LabelEncoder()
categorical_cols = ['track_name',  'playlist_name', 'playlist_genre',
'playlist_subgenre']

label_encoder = LabelEncoder()

for col in categorical_cols:
    df[col + '_encoded'] = label_encoder.fit_transform(df[col])
df.head()

                 track_id
track_name  \
0  6f807x0ima9a1j3VPbc7VN  I Don't Care (with Justin Bieber) - Loud
Luxur...
1  0r7CVbZTWZgbTCYdfa2P31                    Memories - Dillon Francis
Remix
2  1z1Hg7Vb0AhHDiEmnDE79l                    All the Time - Don Diablo
Remix
3  75FpbthrwQmzHlBJLuGdC7                  Call You Mine - Keanu Silva
Remix
4  1e8PAfcKUYoKkxPhrHqw4x            Someone You Loved - Future Humans
Remix
```

```
       track_artist  track_popularity          track_album_id  \
0        Ed Sheeran                66  2oCs0DGTsRO98Gh5ZSl2Cx
1          Maroon 5                67  63rPSO264uRjW1X5E6cWv6
2      Zara Larsson                70  1HoSmj2eLcsrR0vE9gThr4
3  The Chainsmokers                60  1nqYsOef1yKKuGOVchbsk6
4     Lewis Capaldi                69  7m7vv9wlQ4i0LFuJiE2zsQ

                                  track_album_name
track_album_release_date  \
0  I Don't Care (with Justin Bieber) [Loud Luxury...
2019-06-14
1                    Memories (Dillon Francis Remix)
2019-12-13
2                    All the Time (Don Diablo Remix)
2019-07-05
3                        Call You Mine - The Remixes
2019-07-19
4            Someone You Loved (Future Humans Remix)
2019-03-05

  playlist_name             playlist_id playlist_genre  ...
instrumentalness  \
0     Pop Remix  37i9dQZF1DXcZDD7cfEKhW            pop  ...
0.000000
1     Pop Remix  37i9dQZF1DXcZDD7cfEKhW            pop  ...
0.004210
2     Pop Remix  37i9dQZF1DXcZDD7cfEKhW            pop  ...
0.000023
3     Pop Remix  37i9dQZF1DXcZDD7cfEKhW            pop  ...
0.000009
4     Pop Remix  37i9dQZF1DXcZDD7cfEKhW            pop  ...
0.000000

   liveness  valence    tempo  duration_ms  kmeans_cluster  \
0    0.0653    0.518  122.036       194754               1
1    0.3570    0.693   99.972       162600               1
2    0.1100    0.613  124.008       176616               1
3    0.2040    0.277  121.956       169093               0
4    0.0833    0.725  123.976       189052               1

   track_name_encoded  playlist_name_encoded
playlist_genre_encoded  \
0                8898                    292                       2

1               12520                    292                       2

2                 924                    292                       2

3                3020                    292                       2
```

```
4                17910                       292                           2


   playlist_subgenre_encoded
0                          3
1                          3
2                          3
3                          3
4                          3

[5 rows x 28 columns]
```

```
from sklearn.preprocessing import StandardScaler
features = ['danceability', 'energy', 'loudness', 'speechiness',
            'acousticness', 'instrumentalness', 'liveness', 'valence',
'tempo']
X = df[features]
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)
X_scaled_df = pd.DataFrame(X_scaled, columns=features)
for feature in features:
    df[feature + '_scaled'] = X_scaled_df[feature]
print("\nScaled Feature Sample:\n", df[[f + '_scaled' for f in
features]].head())
```

```
Scaled Feature Sample:
    danceability_scaled  energy_scaled  loudness_scaled
speechiness_scaled  \
0             0.642049       1.201614         1.367123           -
0.481362
1             0.490412       0.643317         0.585766           -
0.688642
2             0.138889       1.284529         1.100090           -
0.324422
3             0.435271       1.279002         0.984309           -
0.050024
4            -0.033426       0.742815         0.685151           -
0.702460

   acousticness_scaled  instrumentalness_scaled  liveness_scaled  \
0            -0.333898                -0.377953        -0.809230
1            -0.468670                -0.359177         1.081061
2            -0.436799                -0.377849        -0.519562
3            -0.667642                -0.377911         0.089582
4            -0.432701                -0.377953        -0.692585

   valence_scaled  tempo_scaled
0        0.031908      0.042927
1        0.782522     -0.777198
```

```
2        0.439384      0.116227
3       -1.001795      0.039953
4        0.919777      0.115037
```

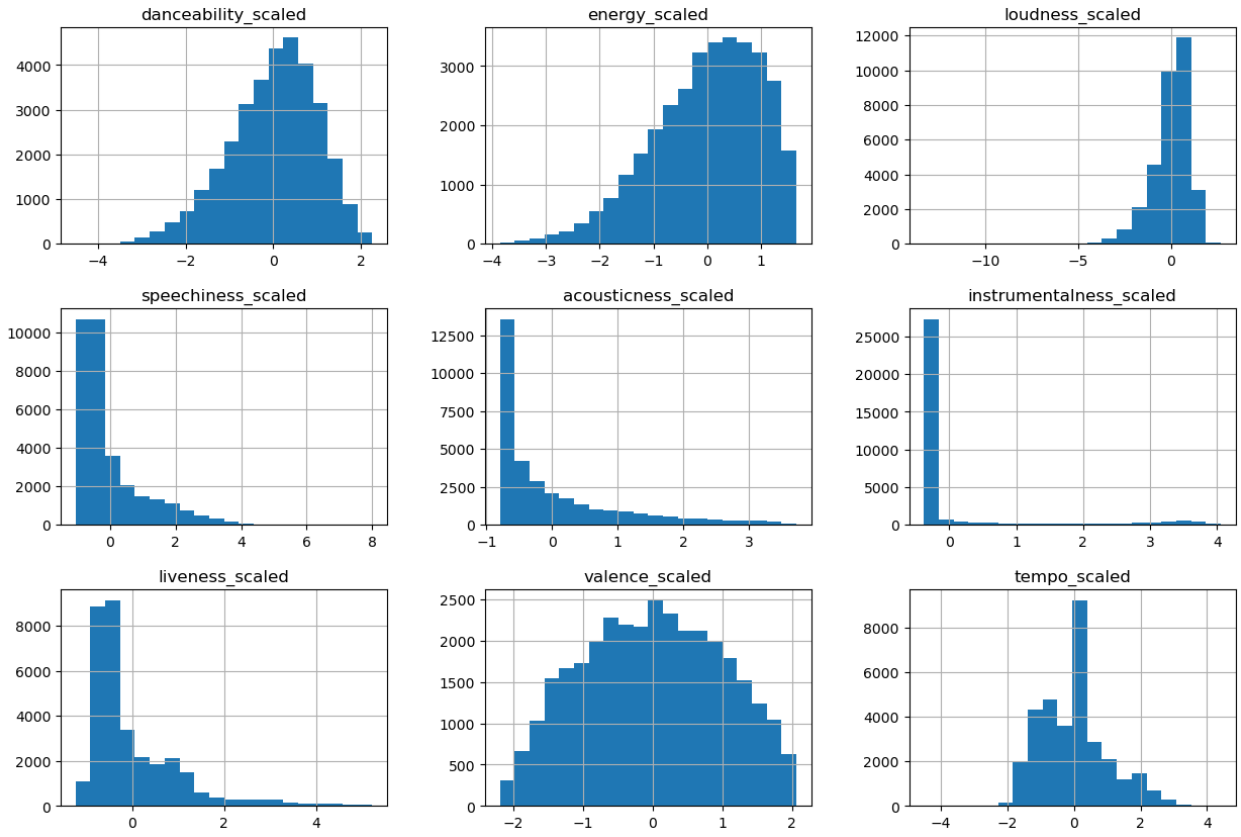
```
from sklearn.model_selection import train_test_split

features_scaled = [feature + '_scaled' for feature in [
    'danceability', 'energy', 'loudness', 'speechiness',
    'acousticness', 'instrumentalness', 'liveness', 'valence', 'tempo'
]]
x = df[features_scaled]
y = df['playlist_genre']
xtrain,xtest,ytrain,ytest= train_test_split(x,y, test_size=0.7)
print("xtrain shape:",xtrain.shape)
print("xtest shape:",xtest.shape)
print("ytrain shape:",ytrain.shape)
print("ytest shape:",ytest.shape)
```

```
xtrain shape: (9849, 9)
xtest shape: (22984, 9)
ytrain shape: (9849,)
ytest shape: (22984,)
```

Data Analysis and Visualizations

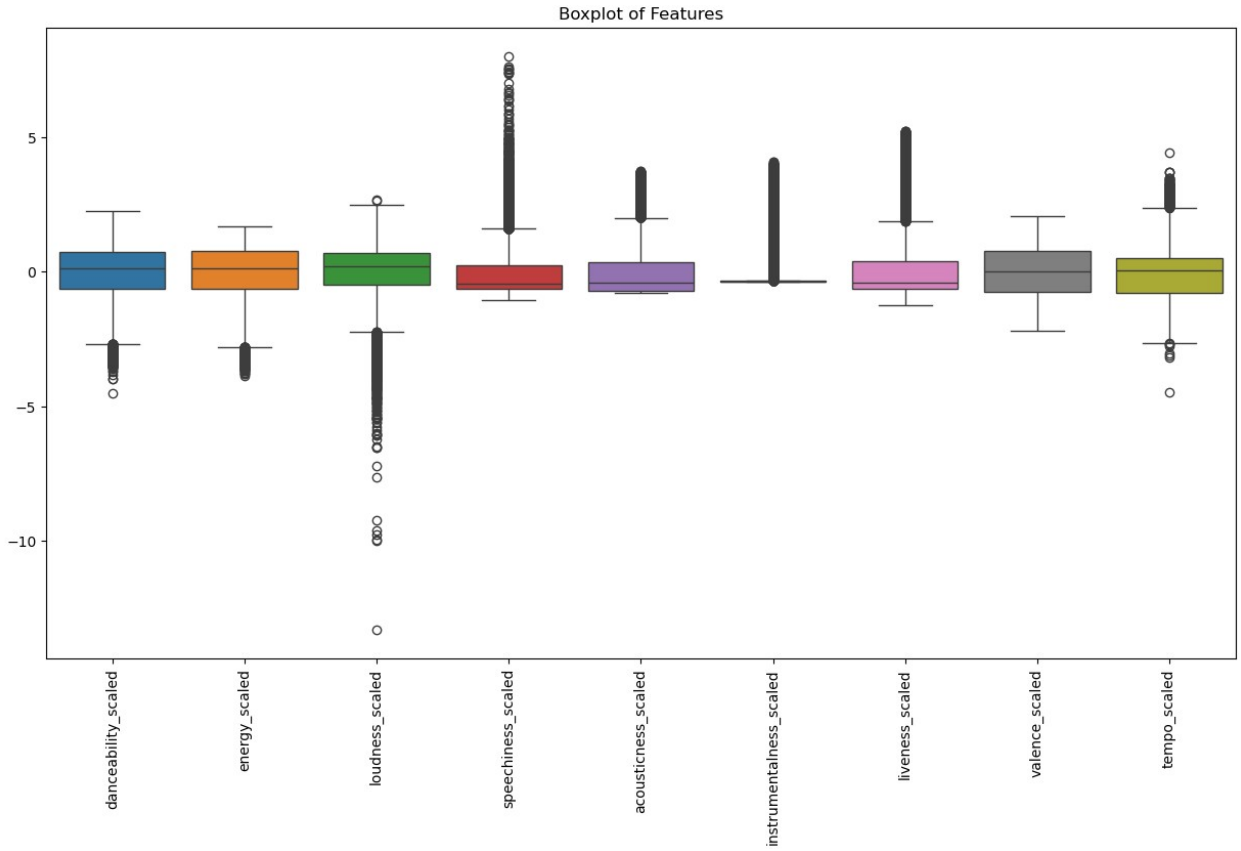
Histogram

```
x.hist(bins=20, figsize=(15,10))
plt.suptitle('Feature Distributions', fontsize=20)
plt.show()
```



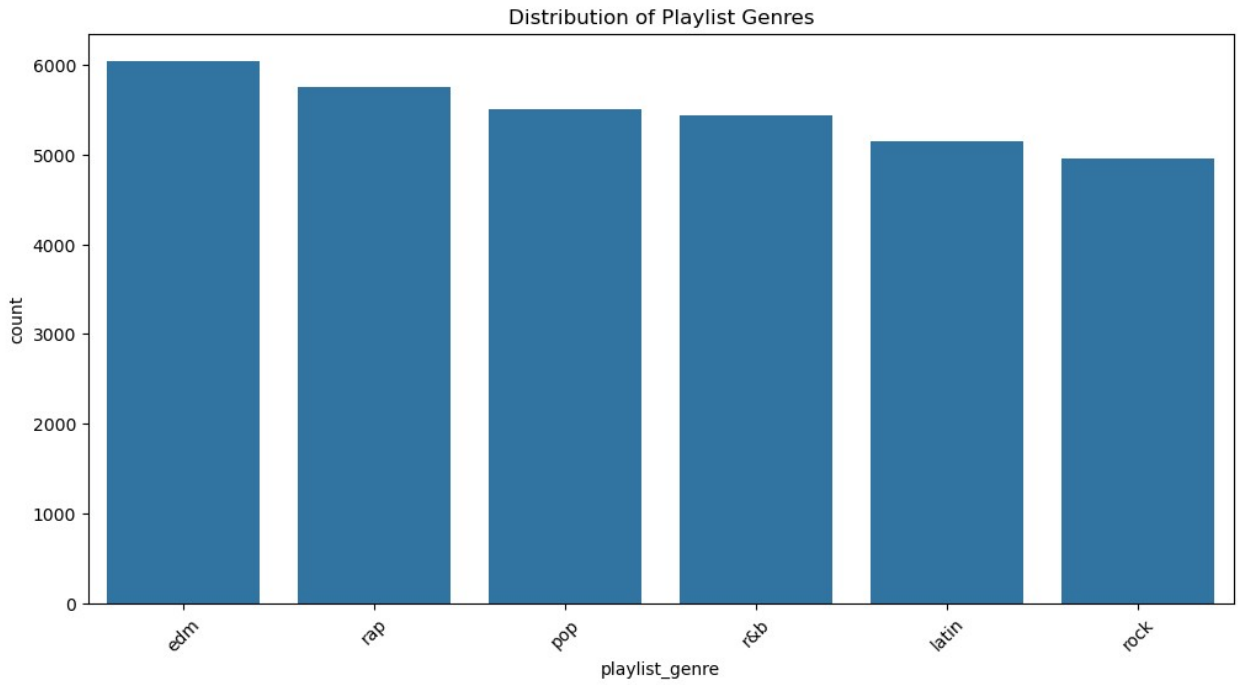
Box plot

```
plt.figure(figsize=(15,8))
sns.boxplot(data=x)
plt.title('Boxplot of Features')
plt.xticks(rotation=90)
plt.show()
```



Countplot for Playlist Genres

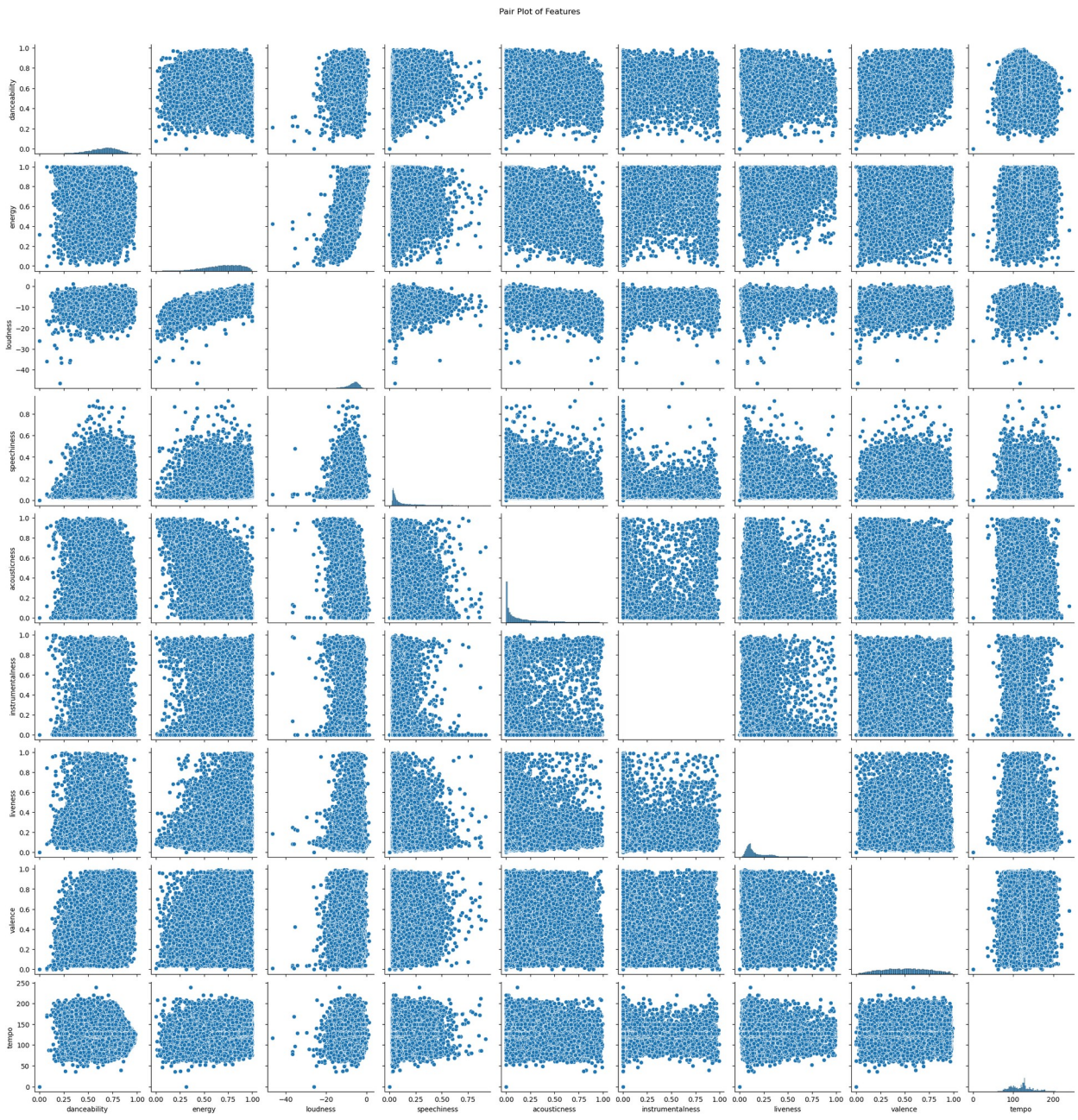
```
plt.figure(figsize=(12,6))
sns.countplot(x='playlist_genre', data=df,
order=df['playlist_genre'].value_counts().index)
plt.title('Distribution of Playlist Genres')
plt.xticks(rotation=45)
plt.show()
```



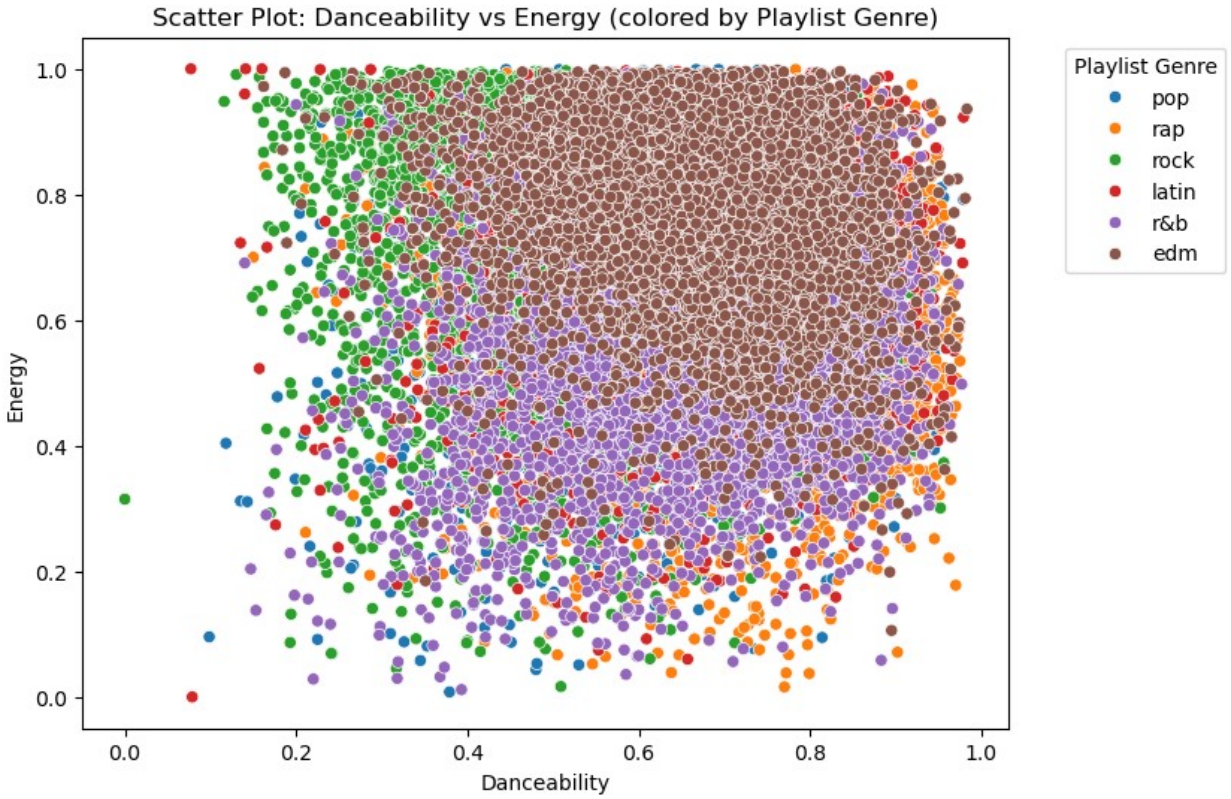
Pair plot

```
plt.figure(figsize=(12,10))
sns.pairplot(df[features])
plt.suptitle("Pair Plot of Features", y=1.02)
plt.show()
```

```
<Figure size 1200x1000 with 0 Axes>
```

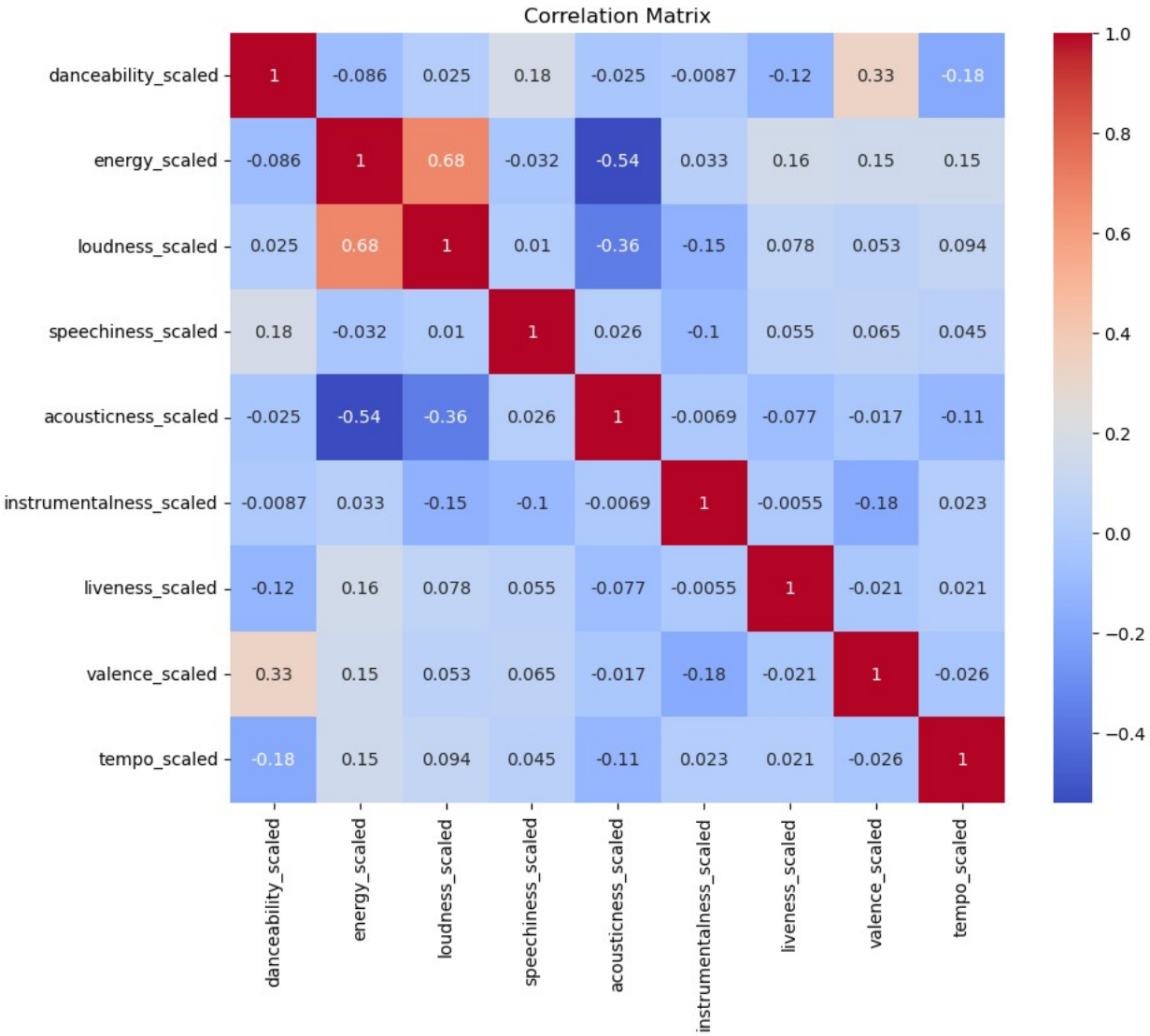



```
plt.figure(figsize=(8,6))
sns.scatterplot(x=df['danceability'], y=df['energy'],
hue=df['playlist_genre'])
plt.title('Scatter Plot: Danceability vs Energy (colored by Playlist
Genre)')
plt.xlabel('Danceability')
plt.ylabel('Energy')
plt.legend(title='Playlist Genre', bbox_to_anchor=(1.05, 1),
loc='upper left')
plt.show()
```



correlation matrix

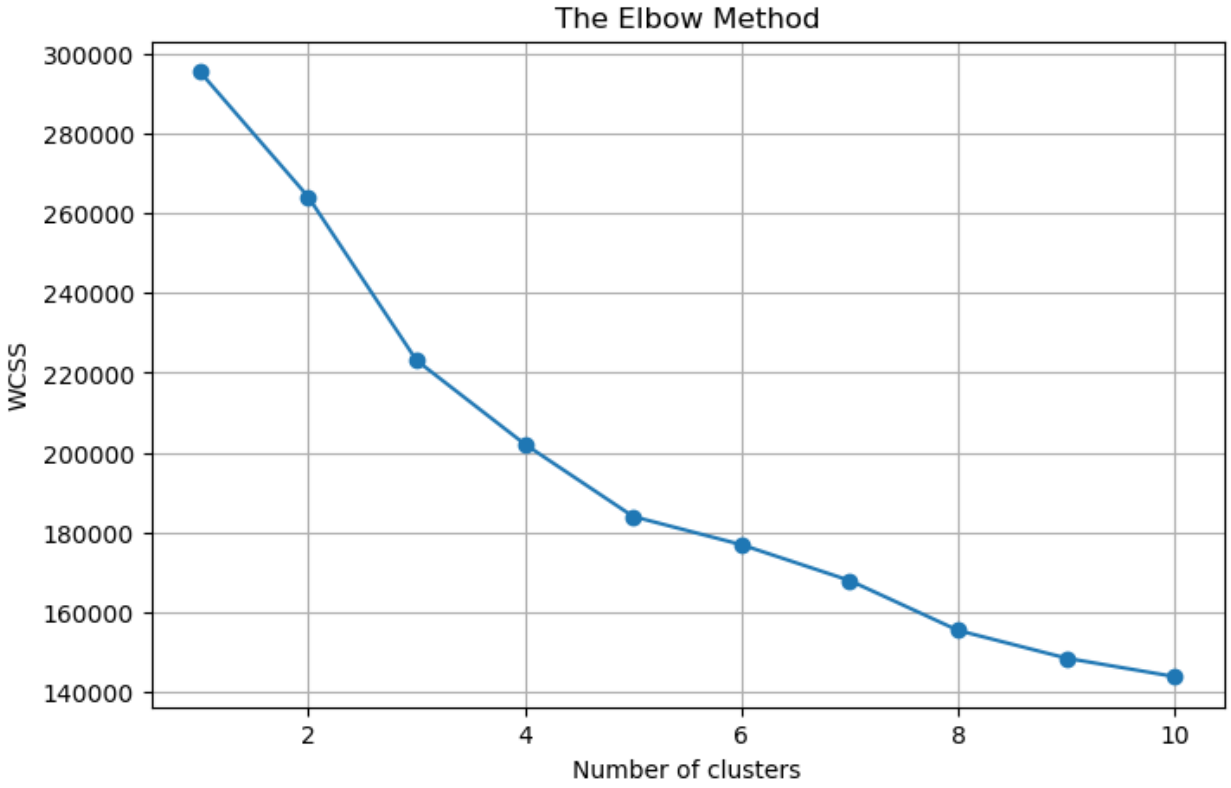
```
corr = x.corr()

plt.figure(figsize=(10,8))
sns.heatmap(corr, annot=True, cmap='coolwarm')
plt.title('Correlation Matrix')
plt.show()
```



Find out and plot different clusters according to different parameters like playlist genres, playlist names

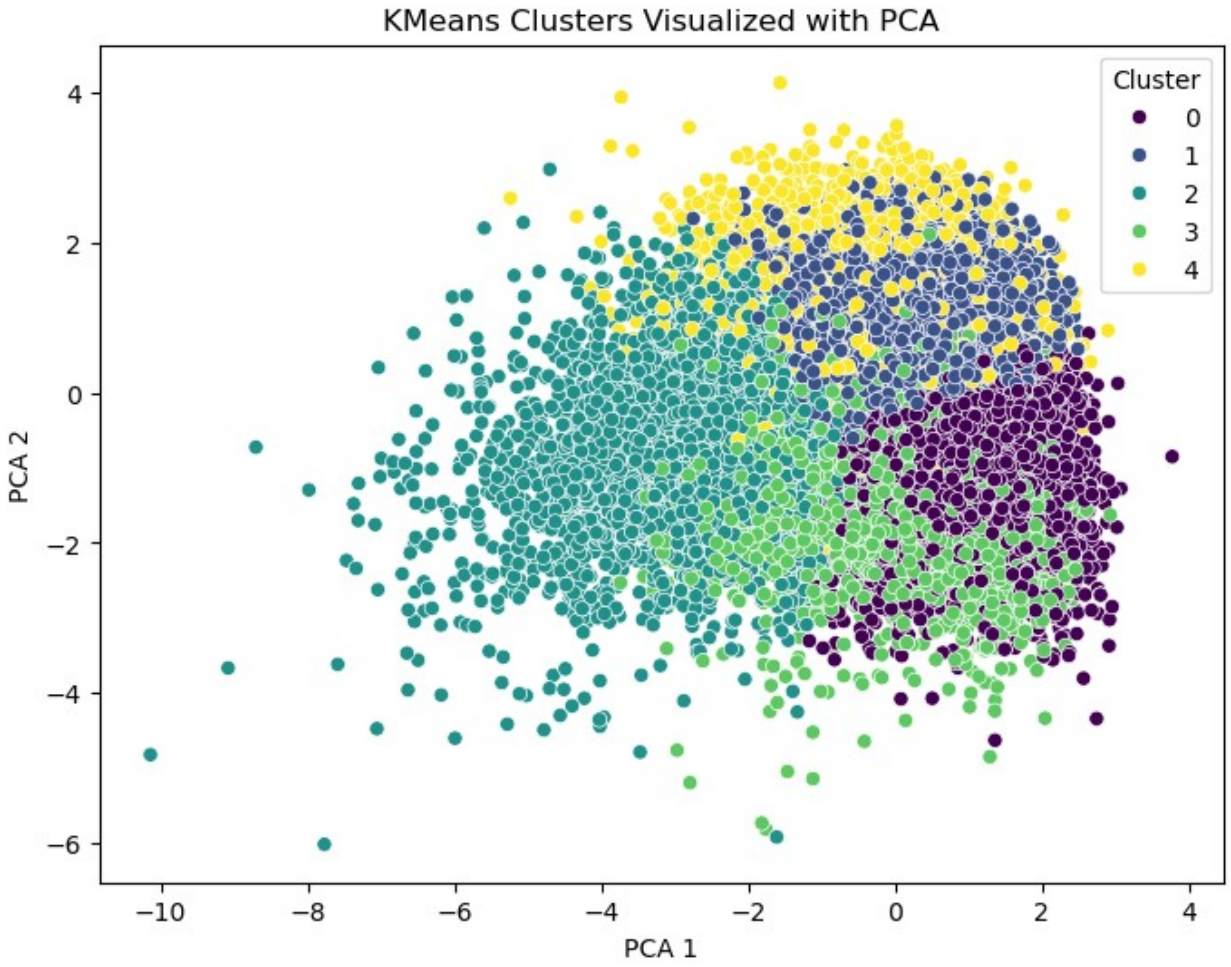
```
 from sklearn.cluster import KMeans
features_scaled = [feature + '_scaled' for feature in [
    'danceability', 'energy', 'loudness', 'speechiness',
    'acousticness', 'instrumentalness', 'liveness', 'valence', 'tempo'
]]

X = df[features_scaled]
wcss = []
```

```
for i in range(1, 11):
    kmeans = KMeans(n_clusters=i, init='k-means++', random_state=42)
    kmeans.fit(X)
    wcss.append(kmeans.inertia_)

plt.figure(figsize=(8,5))
plt.plot(range(1, 11), wcss, marker='o')
plt.title('The Elbow Method')
plt.xlabel('Number of clusters')
plt.ylabel('WCSS')
plt.grid()
plt.show()
```




```
from sklearn.decomposition import PCA

pca = PCA(n_components=2)
X_pca = pca.fit_transform(X)

plt.figure(figsize=(8,6))
sns.scatterplot(x=X_pca[:,0], y=X_pca[:,1], hue=df['kmeans_cluster'],
palette='viridis')
plt.title('KMeans Clusters Visualized with PCA')
plt.xlabel('PCA 1')
plt.ylabel('PCA 2')
```

```
plt.legend(title='Cluster')
plt.show()
```

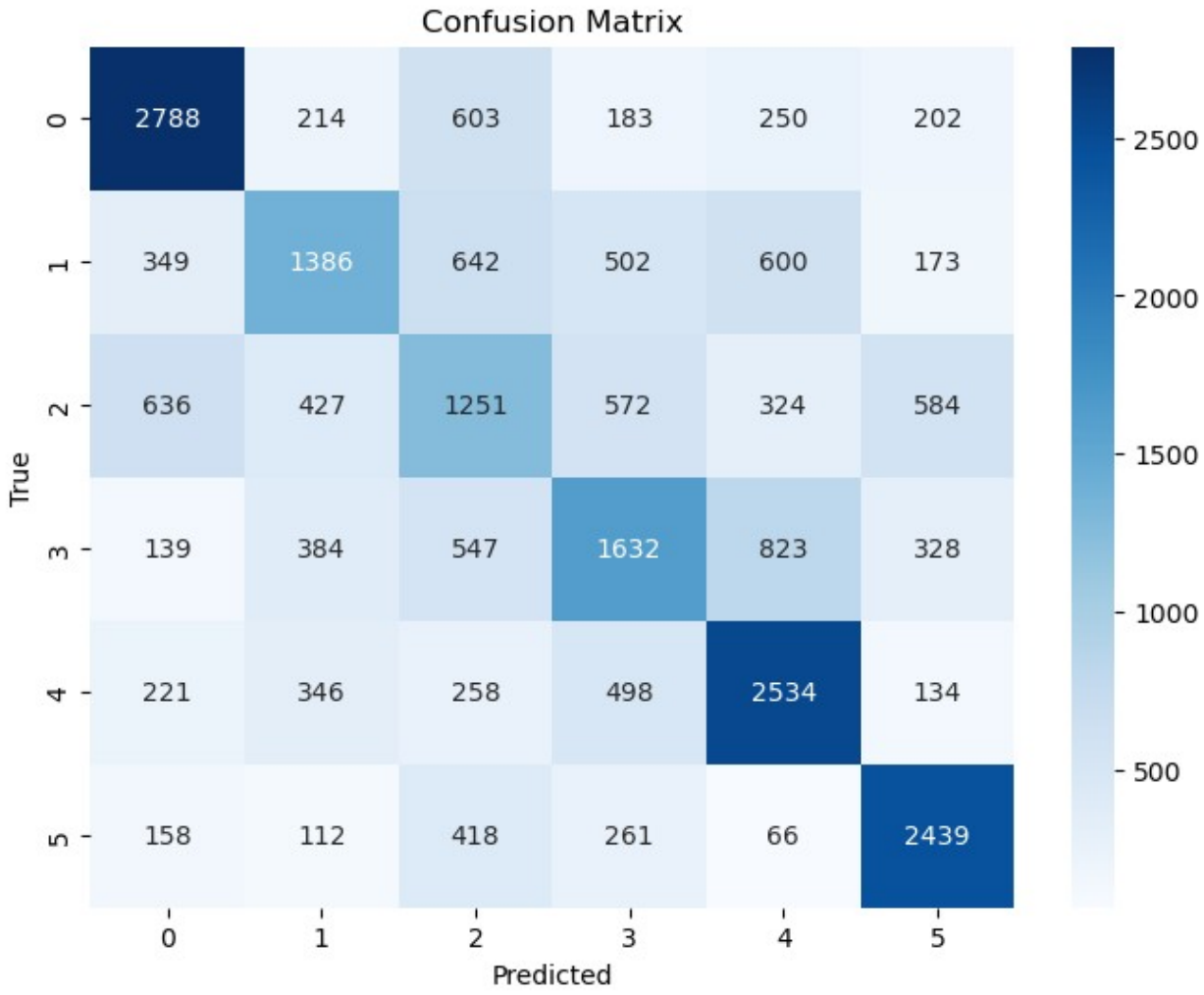


model building

```
from sklearn.ensemble import RandomForestClassifier

# Create Random Forest model
model = RandomForestClassifier(n_estimators=100, random_state=42)

# Train the model
model.fit(xtrain, ytrain)

RandomForestClassifier(random_state=42)

# Predict on test data
y_pred = model.predict(xtest)
```

```python
from sklearn.metrics import accuracy_score, precision_score,
recall_score, f1_score, classification_report, confusion_matrix

# Accuracy
acc = accuracy_score(ytest, y_pred)
print(f"Accuracy: {acc:.2f}")

# Precision, Recall, F1 Score
print("Classification Report:")
print(classification_report(ytest, y_pred))

# Confusion Matrix
import seaborn as sns
import matplotlib.pyplot as plt

plt.figure(figsize=(8,6))
sns.heatmap(confusion_matrix(ytest, y_pred), annot=True, cmap='Blues',
fmt='d')
plt.title('Confusion Matrix')
plt.xlabel('Predicted')
plt.ylabel('True')
plt.show()
```

```
Accuracy: 0.52
Classification Report:
              precision    recall  f1-score   support

         edm       0.65      0.66      0.65      4240
       latin       0.48      0.38      0.43      3652
         pop       0.34      0.33      0.33      3794
         r&b       0.45      0.42      0.44      3853
         rap       0.55      0.63      0.59      3991
        rock       0.63      0.71      0.67      3454

    accuracy                           0.52     22984
   macro avg       0.52      0.52      0.52     22984
weighted avg       0.52      0.52      0.52     22984
```



Final Result

```
from sklearn.cluster import KMeans

# Assume X_scaled is your scaled features (after StandardScaler)
kmeans_model = KMeans(n_clusters=5, random_state=42)
kmeans_model.fit(X_scaled)

# Add cluster labels to your dataframe
df['kmeans_cluster'] = kmeans_model.labels_

print("□ KMeans model trained and clusters assigned!")
```

```
□ KMeans model trained and clusters assigned!
```

```
def recommend_songs(user_preferences, model, scaler, df, features,
top_n=5):
    """
```

```
    Recommend songs based on user preferences.
    """
    import numpy as np

    # Convert user input to array
    user_input = np.array([user_preferences[feature] for feature in
features]).reshape(1, -1)

    # Scale the input
    user_input_scaled = scaler.transform(user_input)

    # Predict the cluster
    cluster_pred = model.predict(user_input_scaled)[0]

    print(f"\n♪ Based on your preferences, you belong to Cluster
{cluster_pred}.\n")

    # Get songs from the same cluster
    recommended_songs = df[df['kmeans_cluster'] == cluster_pred]

    # Randomly pick top_n songs to recommend
    recommendations = recommended_songs.sample(n=top_n,
random_state=42)

    # Check which columns exist
    available_columns = recommendations.columns.tolist()

    # Safe return based on available columns
    columns_to_return = [col for col in ['track_name',
'playlist_genre', 'playlist_name'] if col in available_columns]

    return recommendations[columns_to_return]

user_preferences = {
    'danceability': 0.8,
    'energy': 0.7,
    'loudness': -5.0,
    'speechiness': 0.1,
    'acousticness': 0.2,
    'instrumentalness': 0.0,
    'liveness': 0.1,
    'valence': 0.9,
    'tempo': 120.0
}

recommend_songs(user_preferences, kmeans, scaler, df, features)


♪ Based on your preferences, you belong to Cluster 1.
```

```
C:\Users\user\anaconda3\Lib\site-packages\sklearn\base.py:493:
UserWarning: X does not have valid feature names, but StandardScaler
was fitted with feature names
  warnings.warn(
C:\Users\user\anaconda3\Lib\site-packages\sklearn\base.py:493:
UserWarning: X does not have valid feature names, but KMeans was
fitted with feature names
  warnings.warn(
```

	track_name	playlist_genre	playlist_name
21271	Say My Name	latin	School Dance 2019 (Squeaky Clean)
30098	Photograph - Felix Jaehn Remix	edm	EDM - pop remixes
17570	Mas De Lo Que Te Imaginas (Classic Version)	latin	Latin Pop Classics
6573	Funky Friday	rap	Rap Workout
26279	Aogashima Island!	r&b	Japanese Funk/Soul/NEO/Jazz/Acid