spotify-songs-genre-segmentation

November 5, 2024

1 Perform data pre-processing operations.

[91]:
```
import pandas as pd
import numpy as np
```

[92]:
```
df = pd.read_csv("spotify dataset.csv")
df
```

[92]:

```
                      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]
```

[93]:
```
df.describe()
```

[93]:
```
       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
```

[94]:
```
df.head()
```

[94]:
```
                 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]
```

[95]:
```
df.tail()
```

[95]:
```
                    track_id                            track_name  \
32828  7bxnKAamR3snQ1VGLuVfC1  City Of Lights - Official Radio Edit
32829  5Aevni09Em4575077nkWHz   Closer - Sultan & Ned Shepard Remix
32830  7ImMqPP3Q1yfUHvsdn7wEo          Sweet Surrender - Radio Edit
32831  2m69mhnfQ1Oq6lGtXuYhgX        Only For You - Maor Levi Remix
32832  29zWqhca3zt5NsckZqDf6c                Typhoon - Original Mix

        track_artist  track_popularity          track_album_id  \
32828    Lush & Simon                42  2azRoBBWEEEYhqV6sb7JrT
32829  Tegan and Sara                20  6kD6KLxj7s8eCE3ABvAyf5
32830     Starkillers                14  0ltWNSY9JgxoIZO4VzuCa6
32831          Mat Zo                15  1fGrOkHnHJcStl14zNx8Jy
32832    Julian Calor                27  0X3mUOm6MhxR7PzxG95rAo

                  track_album_name track_album_release_date    playlist_name  \
32828    City Of Lights (Vocal Mix)               2014-04-28    EDM LOVE 2020
32829                Closer Remixed               2013-03-08    EDM LOVE 2020
32830  Sweet Surrender (Radio Edit)               2014-04-21    EDM LOVE 2020
32831        Only For You (Remixes)               2014-01-01    EDM LOVE 2020
32832                 Typhoon/Storm               2014-03-03    EDM LOVE 2020

                playlist_id playlist_genre  … key  loudness  mode  \
32828  6jI1gFr6ANFtT8MmTvA2Ux            edm  …   2    -1.814     1
32829  6jI1gFr6ANFtT8MmTvA2Ux            edm  …   0    -4.462     1
32830  6jI1gFr6ANFtT8MmTvA2Ux            edm  …   6    -4.899     0
32831  6jI1gFr6ANFtT8MmTvA2Ux            edm  …   2    -3.361     1
32832  6jI1gFr6ANFtT8MmTvA2Ux            edm  …   5    -4.571     0

       speechiness  acousticness  instrumentalness  liveness  valence  \
32828       0.0936      0.076600          0.000000    0.0668   0.2100
32829       0.0420      0.001710          0.004270    0.3750   0.4000
32830       0.0481      0.108000          0.000001    0.1500   0.4360
32831       0.1090      0.007920          0.127000    0.3430   0.3080
32832       0.0385      0.000133          0.341000    0.7420   0.0894
```

```
          tempo  duration_ms
32828  128.170       204375
32829  128.041       353120
32830  127.989       210112
32831  128.008       367432
32832  127.984       337500

[5 rows x 23 columns]
```

[97]:
```
df.isnull().values.any()
```

[97]:
```
True
```

[98]:
```
df.isnull().mean()
```

[98]:
```
track_id                    0.000000
track_name                  0.000152
track_artist                0.000152
track_popularity            0.000000
track_album_id              0.000000
track_album_name            0.000152
track_album_release_date    0.000000
playlist_name               0.000000
playlist_id                 0.000000
playlist_genre              0.000000
playlist_subgenre           0.000000
danceability                0.000000
energy                      0.000000
key                         0.000000
loudness                    0.000000
mode                        0.000000
speechiness                 0.000000
acousticness                0.000000
instrumentalness            0.000000
liveness                    0.000000
valence                     0.000000
tempo                       0.000000
duration_ms                 0.000000
dtype: float64
```

2 Label Encoding on text/Categorical data

[46]:
```
from sklearn import preprocessing
label_encoder = preprocessing.LabelEncoder()
df['track_name']= label_encoder.fit_transform(df['track_name'])

df['track_name'].unique()


df['track_artist']= label_encoder.fit_transform(df['track_artist'])

df['track_artist'].unique()


df['track_album_name']= label_encoder.fit_transform(df['track_album_name'])

df['track_album_name'].unique()

df['track_id']= label_encoder.fit_transform(df['track_id'])

df['track_id'].unique()


df['track_album_id']= label_encoder.fit_transform(df['track_album_id'])

df['track_album_id'].unique()


df['track_album_release_date']= label_encoder.
 ↪fit_transform(df['track_album_release_date'])

df['track_album_release_date'].unique()


df['playlist_name']= label_encoder.fit_transform(df['playlist_name'])

df['playlist_name'].unique()



df['playlist_id']= label_encoder.fit_transform(df['playlist_id'])

df['playlist_id'].unique()


df['playlist_genre']= label_encoder.fit_transform(df['playlist_genre'])
```

```
df['playlist_genre'].unique()
```

[46]:
```
array([2, 4, 5, 1, 3, 0])
```

[47]:
```
df
```

[47]:

	track_id	track_name	track_artist	track_popularity	track_album_id	\
0	24150	8898	2782	66	8225	
1	3061	12520	6084	67	17650	
2	7219	924	10416	70	3798	
3	25699	3020	9215	60	5293	
4	5987	17910	5402	69	21936	
…	…	…	…	…	…	
32828	26856	3567	5725	42	7586	
32829	18774	3642	9102	20	19610	
32830	26465	18844	8746	14	2263	
32831	10083	14439	6140	15	4914	
32832	7864	20779	4707	27	1558	

	track_album_name	track_album_release_date	playlist_name	playlist_id	\
0	7614	4315	292	235	
1	10410	4492	292	235	
2	985	4335	292	235	
3	2798	4348	292	235	
4	14843	4220	292	235	
…	…	…	…	…	
32828	3240	2830	443	420	
32829	3317	2580	443	420	
32830	15576	2825	443	420	
32831	11962	2760	443	420	
32832	17683	2794	443	420	

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

	instrumentalness	liveness	valence	tempo	duration_ms
0	0.000000	0.0653	0.5180	122.036	194754

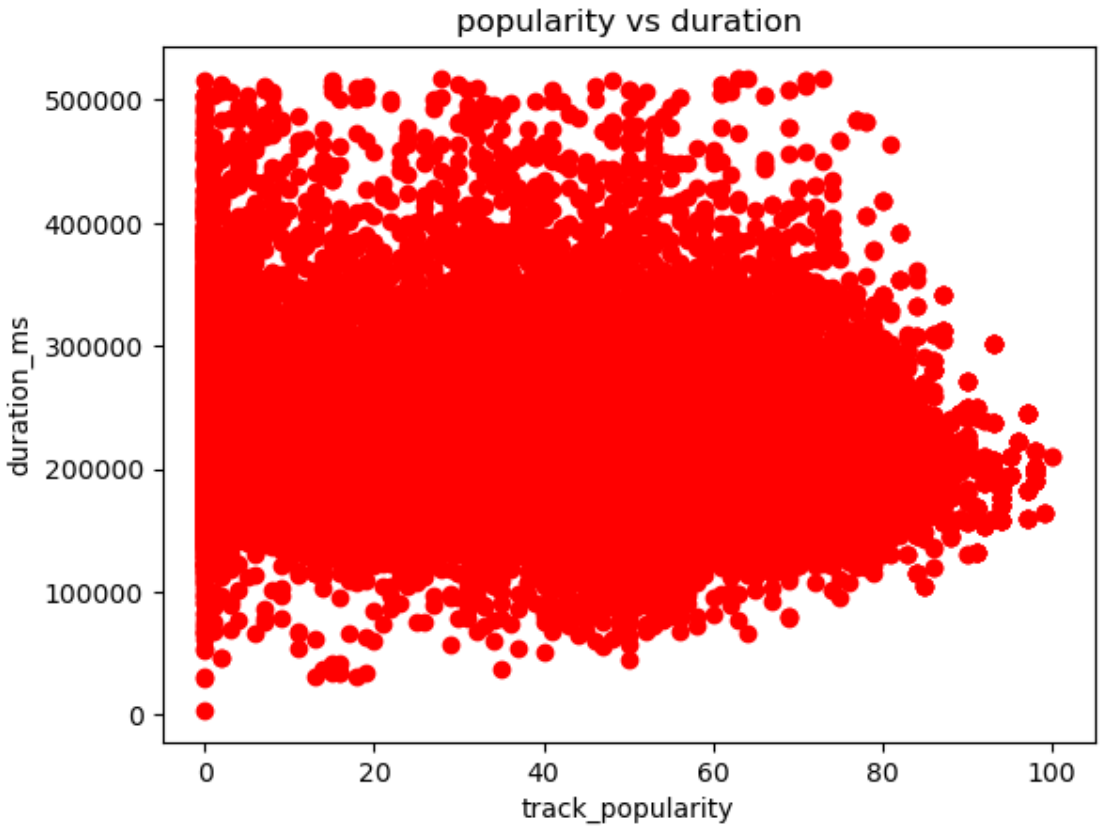
```
1              0.004210    0.3570   0.6930   99.972        162600
2              0.000023    0.1100   0.6130  124.008        176616
3              0.000009    0.2040   0.2770  121.956        169093
4              0.000000    0.0833   0.7250  123.976        189052
…                 …          …        …        …              …
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]
```

[48]:
```
df.isnull().mean()
```

[48]:
```
track_id                  0.0
track_name                0.0
track_artist              0.0
track_popularity          0.0
track_album_id            0.0
track_album_name          0.0
track_album_release_date  0.0
playlist_name             0.0
playlist_id               0.0
playlist_genre            0.0
playlist_subgenre         0.0
danceability              0.0
energy                    0.0
key                       0.0
loudness                  0.0
mode                      0.0
speechiness               0.0
acousticness              0.0
instrumentalness          0.0
liveness                  0.0
valence                   0.0
tempo                     0.0
duration_ms               0.0
dtype: float64
```

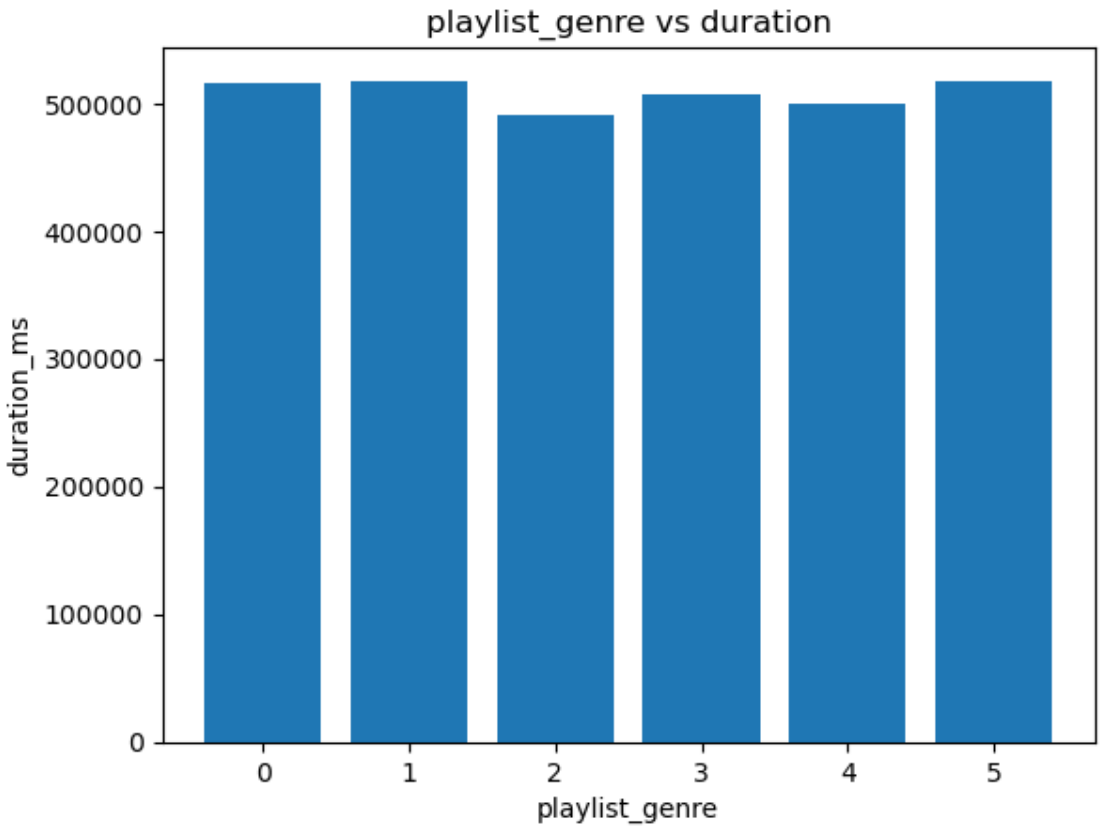
3 data analysis and visualizations draw all the possible plots to provide essential informations and to derive some meaningful insights.

[49]:
```
import matplotlib.pyplot as plt
x=df["track_popularity"]
y=df["duration_ms"]
plt.scatter(x, y, c='r')
plt.xlabel("track_popularity")
plt.ylabel("duration_ms")
plt.title("popularity vs duration")
plt.show()
```

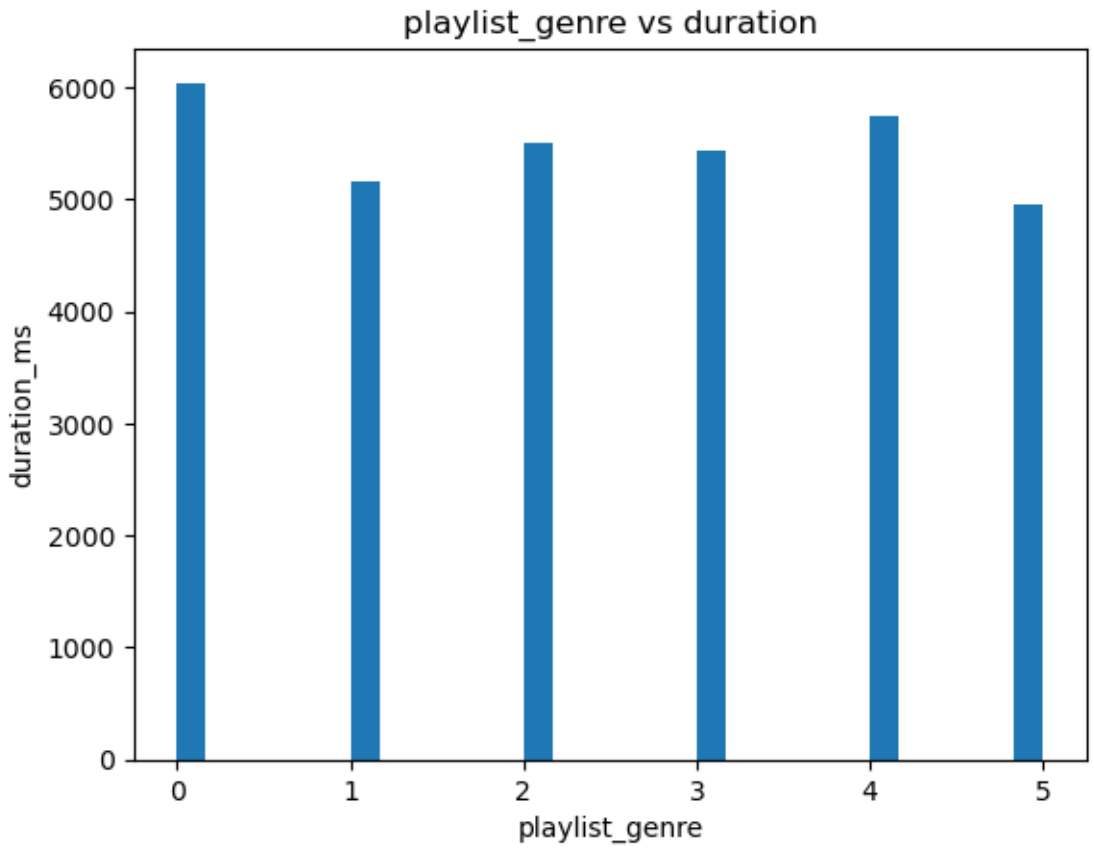


[12]:
```
import matplotlib.pyplot as plt
x=df["playlist_genre"]
y=df["duration_ms"]
plt.bar(x, y)
plt.xlabel("playlist_genre")
plt.ylabel("duration_ms")
plt.title("playlist_genre vs duration")
```

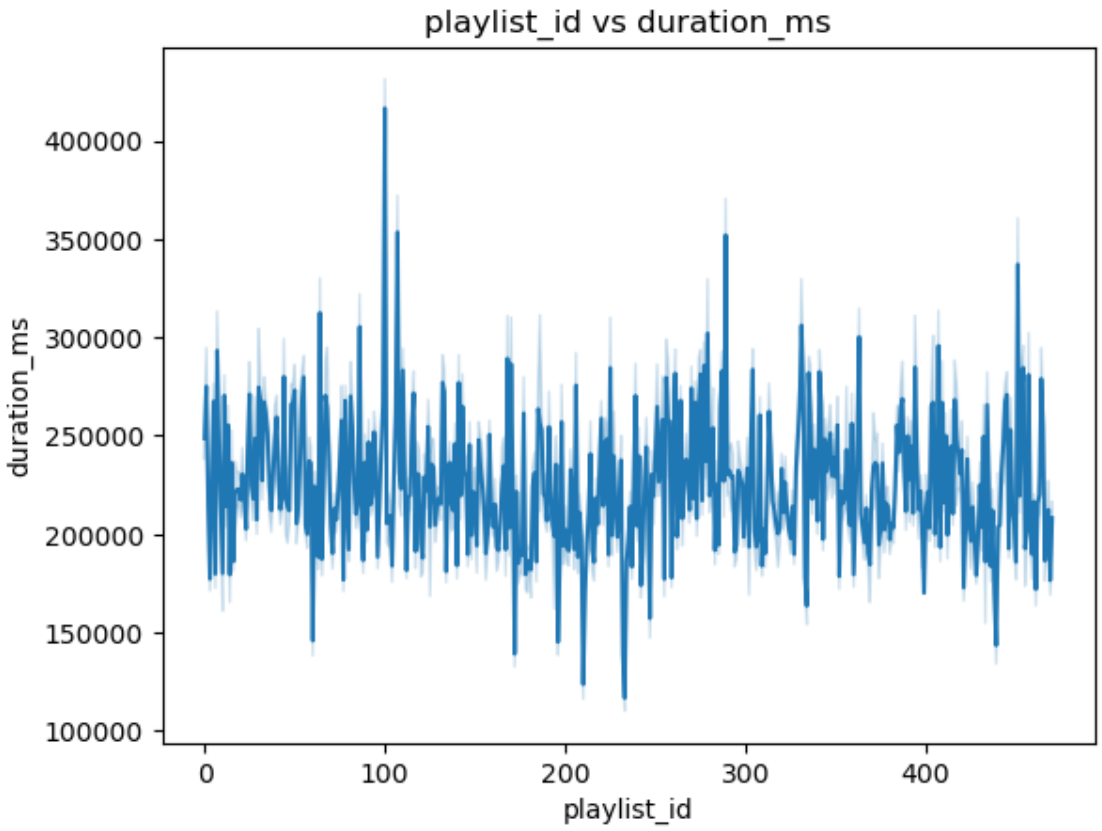
```
plt.show()
```



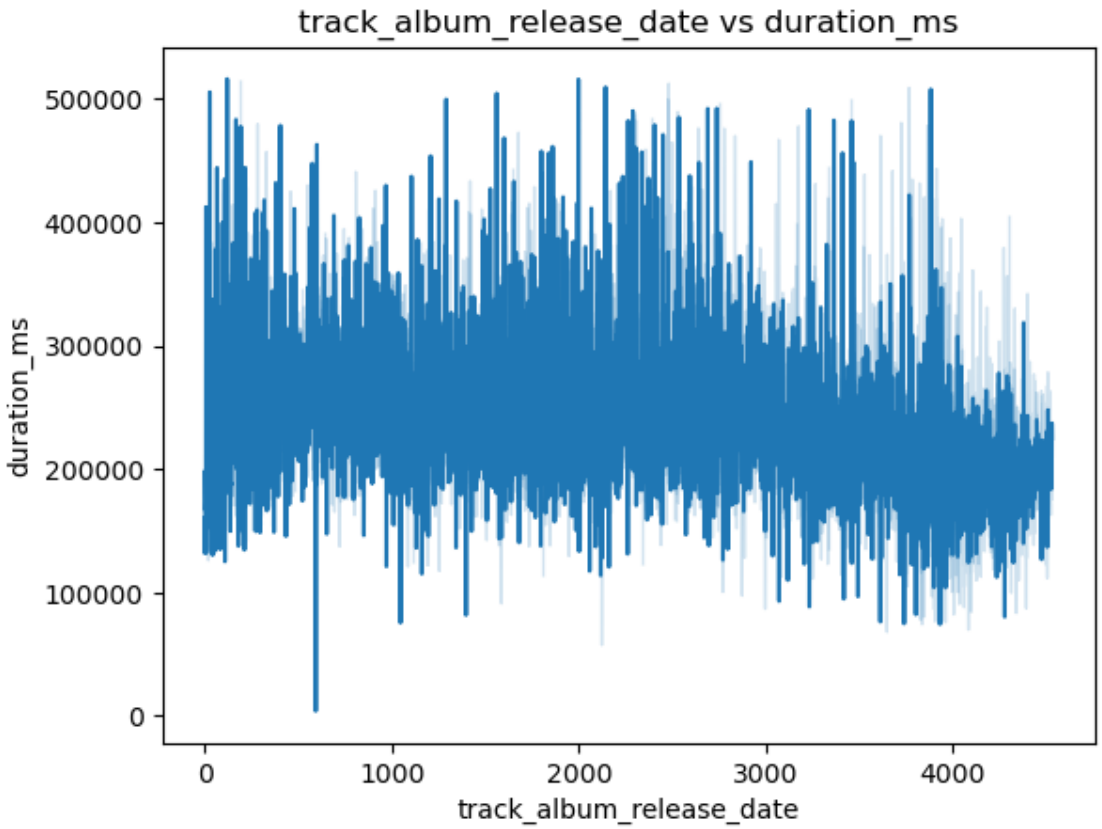
[13]:
```
import matplotlib.pyplot as plt
x=df["playlist_genre"]
y=df["duration_ms"]
plt.hist(x, bins=30)
plt.xlabel("playlist_genre")
plt.ylabel("duration_ms")
plt.title("playlist_genre vs duration")
plt.show()
```



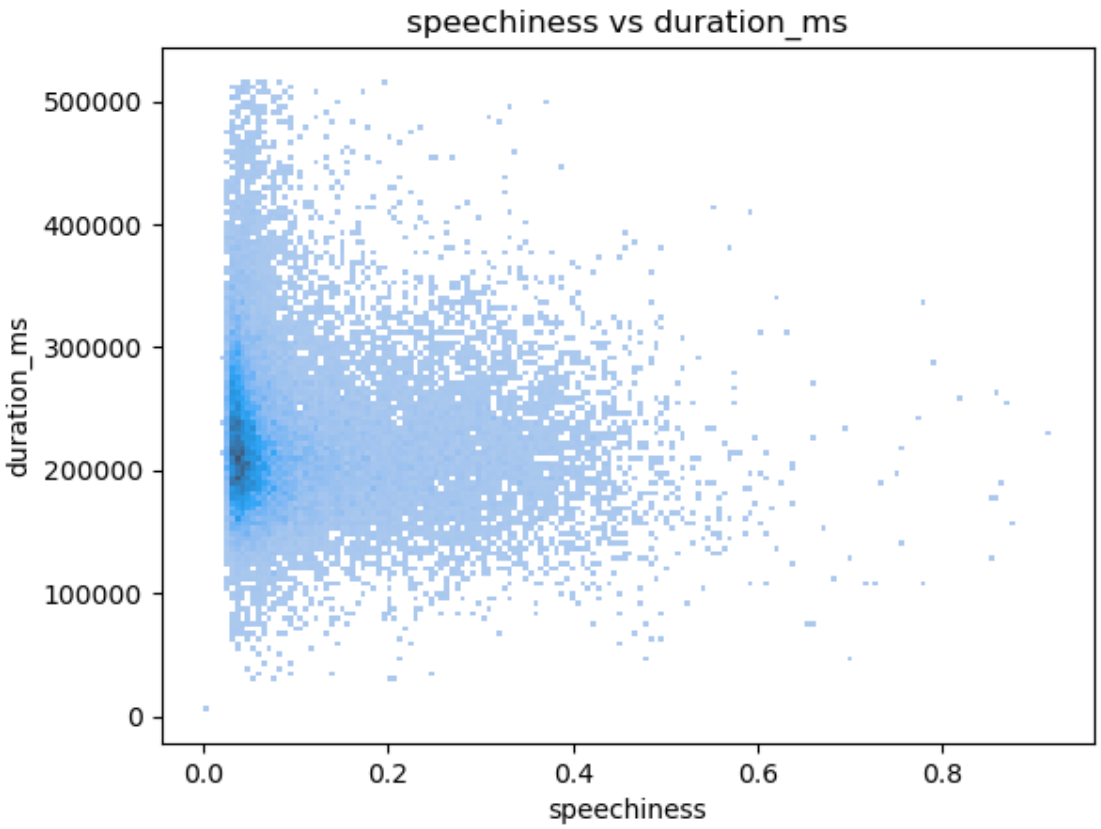
[15]:
```
import seaborn as sns
sns.lineplot(x="playlist_id",y="duration_ms",data = df)
plt.xlabel('playlist_id')
plt.ylabel('duration_ms')
plt.title("playlist_id vs duration_ms")
plt.show()
```



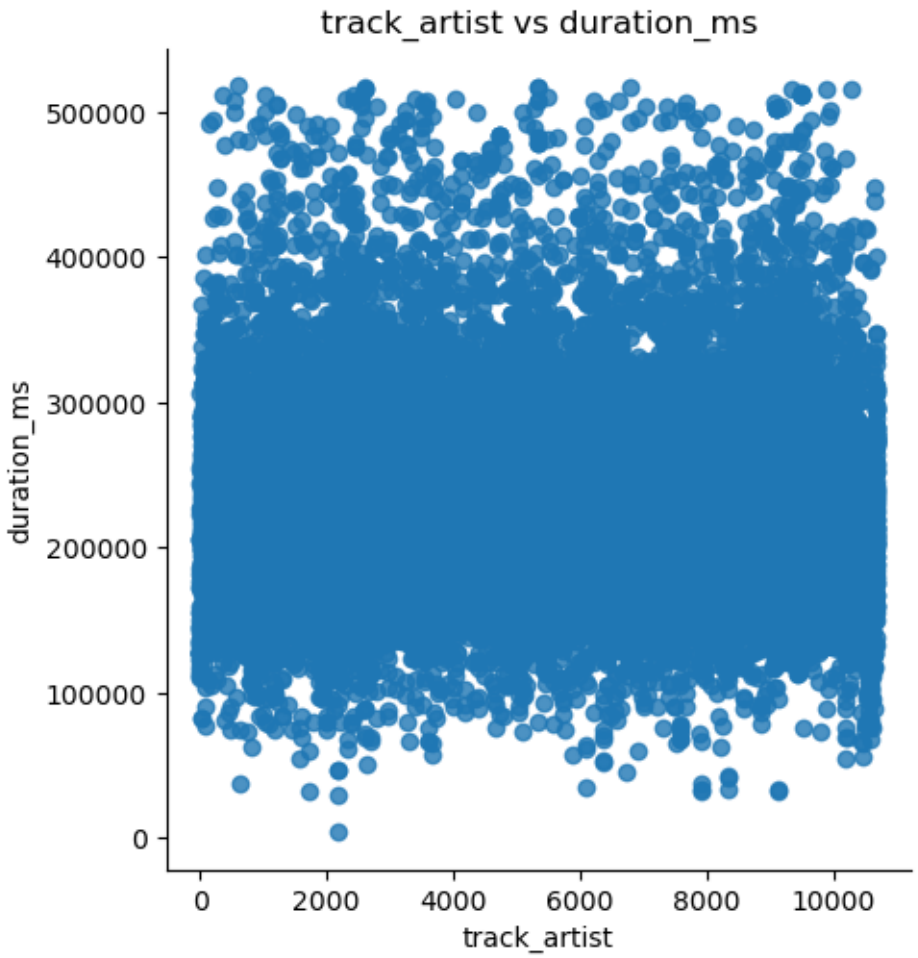
[16]:
```
import seaborn as sns
sns.lineplot(x="track_album_release_date",y="duration_ms",data = df)
plt.xlabel('track_album_release_date')
plt.ylabel('duration_ms')
plt.title("track_album_release_date vs duration_ms")
plt.show()
```



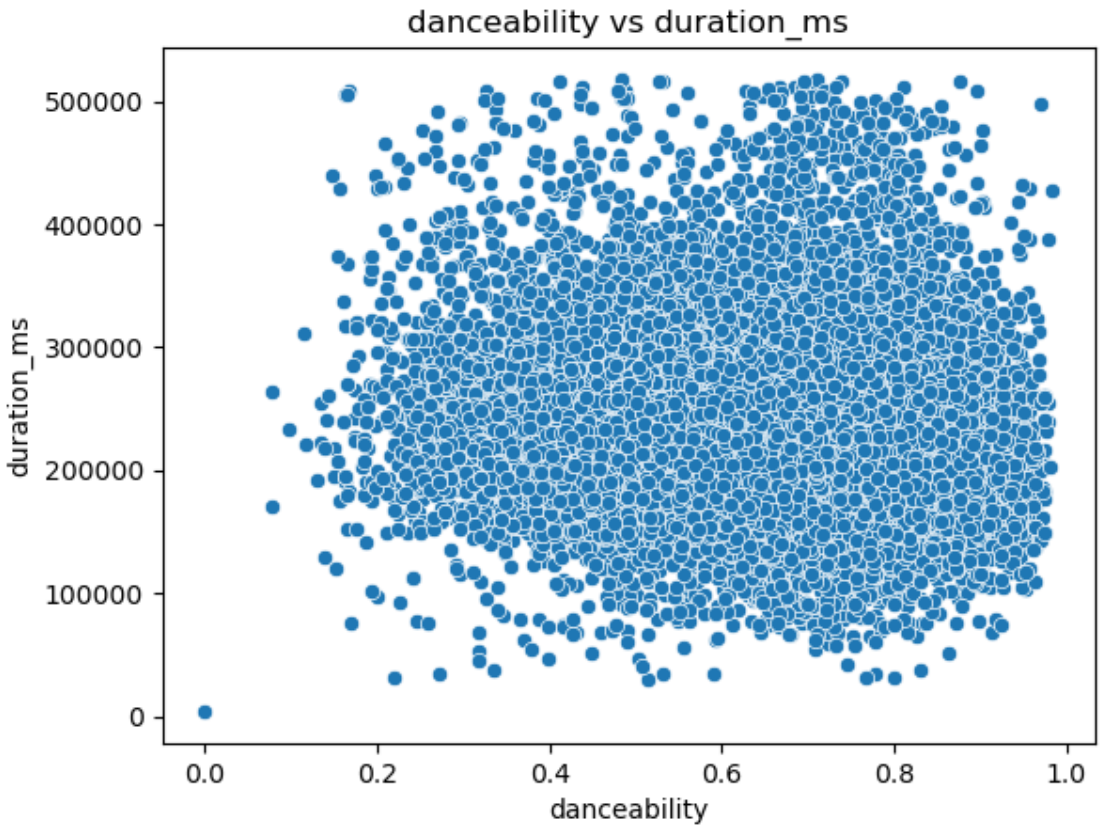
[22]:
```
import seaborn as sns
sns.histplot(x="speechiness",y="duration_ms",data = df)
plt.xlabel('speechiness')
plt.ylabel('duration_ms')
plt.title("speechiness vs duration_ms")
plt.show()
```



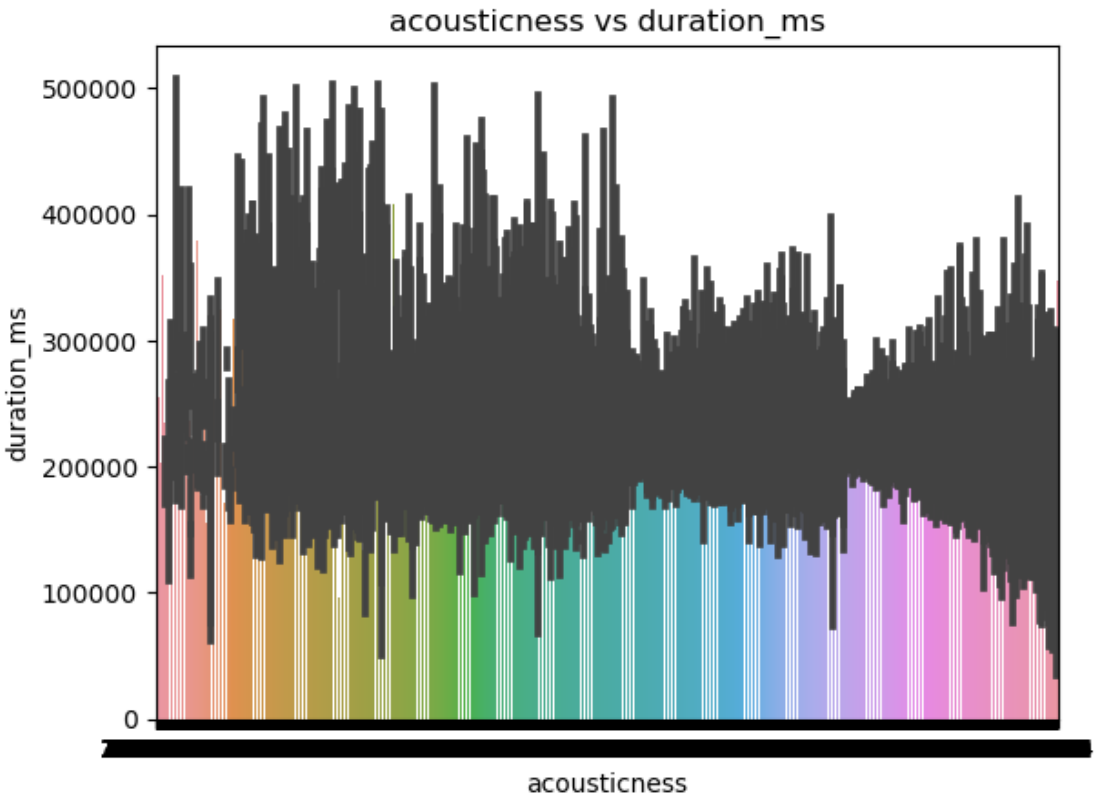
[23]:
```
import seaborn as sns
sns.lmplot(x="track_artist",y="duration_ms",data = df)
plt.xlabel('track_artist')
plt.ylabel('duration_ms')
plt.title("track_artist vs duration_ms")
plt.show()
```



[24]:
```
import seaborn as sns
sns.scatterplot(x="danceability",y="duration_ms",data = df)
plt.xlabel('danceability')
plt.ylabel('duration_ms')
plt.title("danceability vs duration_ms")
plt.show()
```



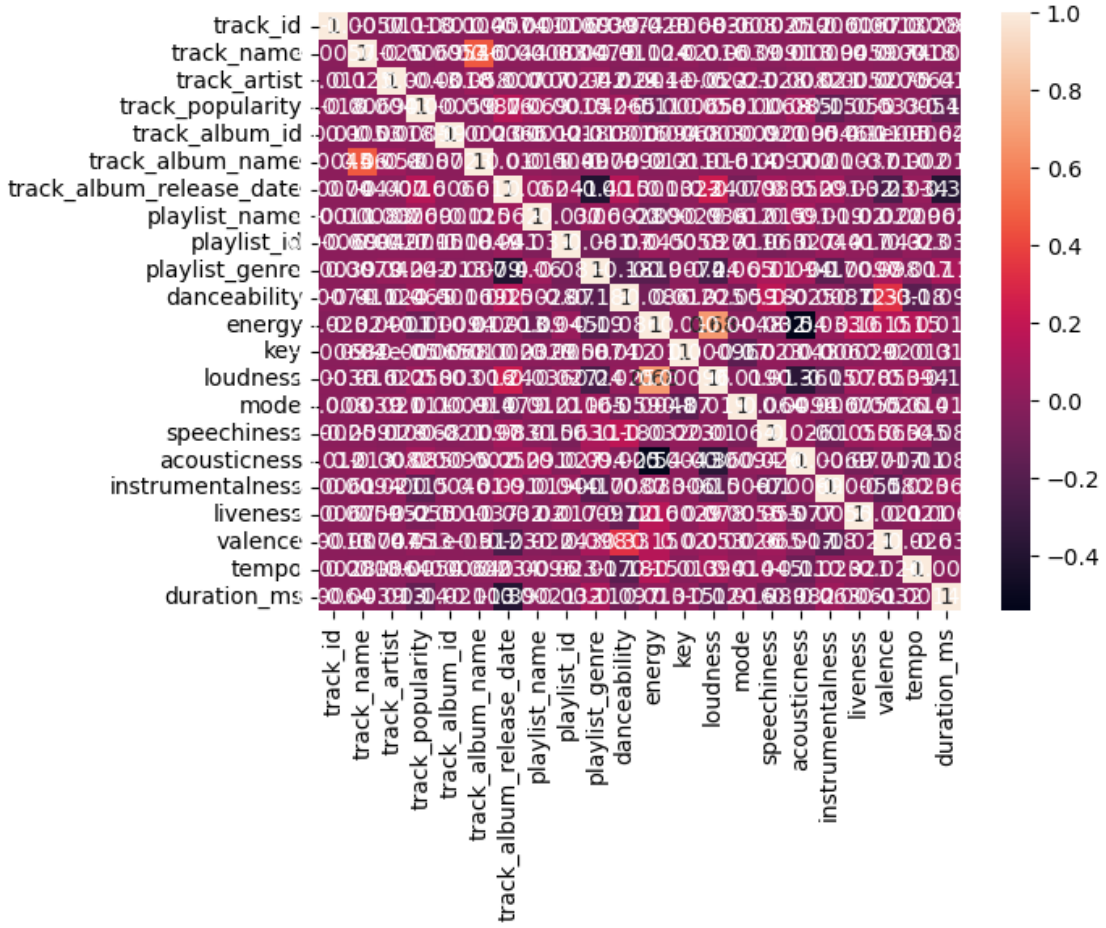
[25]:
```
import seaborn as sns
sns.barplot(x="acousticness",y="duration_ms",data = df)
plt.xlabel('acousticness')
plt.ylabel('duration_ms')
plt.title("acousticness vs duration_ms")
plt.show()
```



4 correlation matrix of features according to the datasets.

[50]:
```
import pandas as pd
import seaborn as sns

corr_matrix = df.corr()
sns.heatmap(corr_matrix, annot=True)
```

[50]: <AxesSubplot:>



[54]:
```
df
```

[54]:

	track_id	track_name	track_artist	track_popularity	track_album_id	\
0	24150	8898	2782	66	8225	
1	3061	12520	6084	67	17650	
2	7219	924	10416	70	3798	
3	25699	3020	9215	60	5293	
4	5987	17910	5402	69	21936	
…	…	…	…	…	…	
32828	26856	3567	5725	42	7586	
32829	18774	3642	9102	20	19610	
32830	26465	18844	8746	14	2263	
32831	10083	14439	6140	15	4914	
32832	7864	20779	4707	27	1558	

	track_album_name	track_album_release_date	playlist_name	playlist_id	\
0	7614	4315	292	235	
1	10410	4492	292	235	

```
2                985                        4335            292            235
3               2798                        4348            292            235
4              14843                        4220            292            235
…                …                           …              …              …
32828           3240                        2830            443            420
32829           3317                        2580            443            420
32830          15576                        2825            443            420
32831          11962                        2760            443            420
32832          17683                        2794            443            420

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

5 K-MEANS CLUSTERING

[67]:
```
from sklearn.cluster import KMeans
```

[69]:
```
km=KMeans(n_clusters=5)
km.
  ↪fit(df[["track_id","track_name","track_artist","track_popularity","track_album_id","track_a
```

[69]:
```
KMeans(n_clusters=5)
```

[70]:
```
km.cluster_centers_
```

[70]:
```
array([[ 1.42482100e+04,  1.17467784e+04,  5.28273025e+03,
         4.51691131e+01,  1.12719536e+04,  9.84536196e+03,
         3.35369785e+03,  2.23745081e+02,  2.37454013e+02,
         2.17410993e+00,  5.36937851e+00, -6.14435251e+00,
         5.62617114e-01,  1.02879208e-01,  1.69702072e-01,
         5.48390250e-02,  1.89602951e-01,  5.16497744e-01,
         1.21237090e+02,  2.02992766e+05],
       [ 1.48570498e+04,  1.17353564e+04,  5.29455859e+03,
         2.92724609e+01,  1.07538809e+04,  9.88431738e+03,
         2.19600879e+03,  2.45007812e+02,  2.15260742e+02,
         2.34277344e+00,  5.61914062e+00, -8.41162891e+00,
         5.87890625e-01,  8.14125977e-02,  1.34467660e-01,
         2.93981202e-01,  1.90673437e-01,  4.30458789e-01,
         1.22401687e+02,  4.16082950e+05],
       [ 1.40822678e+04,  1.16474995e+04,  5.41733058e+03,
         4.50459547e+01,  1.13681104e+04,  9.80891780e+03,
         3.75160227e+03,  2.24315696e+02,  2.37237055e+02,
         2.26650485e+00,  5.32815534e+00, -6.83892023e+00,
         5.50000000e-01,  1.27787508e-01,  2.06002904e-01,
         1.28926606e-01,  1.90105599e-01,  5.02570684e-01,
         1.21370595e+02,  1.55649497e+05],
       [ 1.41615277e+04,  1.15761735e+04,  5.18329255e+03,
         3.59390735e+01,  1.11801609e+04,  9.53898263e+03,
         2.02418907e+03,  2.14733384e+02,  2.32675982e+02,
         2.93856999e+00,  5.43479355e+00, -7.55409542e+00,
         5.74269889e-01,  1.01774673e-01,  1.60529617e-01,
         1.09516954e-01,  1.97011999e-01,  5.02592170e-01,
         1.20274979e+02,  3.09521367e+05],
       [ 1.42112540e+04,  1.16755223e+04,  5.29219211e+03,
         4.12492937e+01,  1.12121242e+04,  9.72849022e+03,
         2.55550525e+03,  2.22911289e+02,  2.28407956e+02,
         2.74347384e+00,  5.35879760e+00, -6.89812928e+00,
         5.74754210e-01,  1.04005729e-01,  1.73440078e-01,
         6.18511468e-02,  1.87929321e-01,  5.20394691e-01,
         1.20120208e+02,  2.48203431e+05]])
```

[71]:
```
df["cluster_group"]=km.labels_
```

[72]:
```
df
```

[72]:

	track_id	track_name	track_artist	track_popularity	track_album_id	\
0	24150	8898	2782	66	8225	
1	3061	12520	6084	67	17650	

```
2          7219         924         10416                70            3798
3         25699        3020          9215                60            5293
4          5987       17910          5402                69           21936
…             …           …             …                 …               …
32828     26856        3567          5725                42            7586
32829     18774        3642          9102                20           19610
32830     26465       18844          8746                14            2263
32831     10083       14439          6140                15            4914
32832      7864       20779          4707                27            1558

       track_album_name  track_album_release_date  playlist_name  playlist_id  \
0                  7614                      4315            292          235
1                 10410                      4492            292          235
2                   985                      4335            292          235
3                  2798                      4348            292          235
4                 14843                      4220            292          235
…                     …                         …              …            …
32828              3240                      2830            443          420
32829              3317                      2580            443          420
32830             15576                      2825            443          420
32831             11962                      2760            443          420
32832             17683                      2794            443          420

       playlist_genre  … loudness  mode  speechiness  acousticness  \
0                   2  …   -2.634     1       0.0583      0.102000
1                   2  …   -4.969     1       0.0373      0.072400
2                   2  …   -3.432     0       0.0742      0.079400
3                   2  …   -3.778     1       0.1020      0.028700
4                   2  …   -4.672     1       0.0359      0.080300
…                 …  …       …   …           …             …
32828               0  …   -1.814     1       0.0936      0.076600
32829               0  …   -4.462     1       0.0420      0.001710
32830               0  …   -4.899     0       0.0481      0.108000
32831               0  …   -3.361     1       0.1090      0.007920
32832               0  …   -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
```

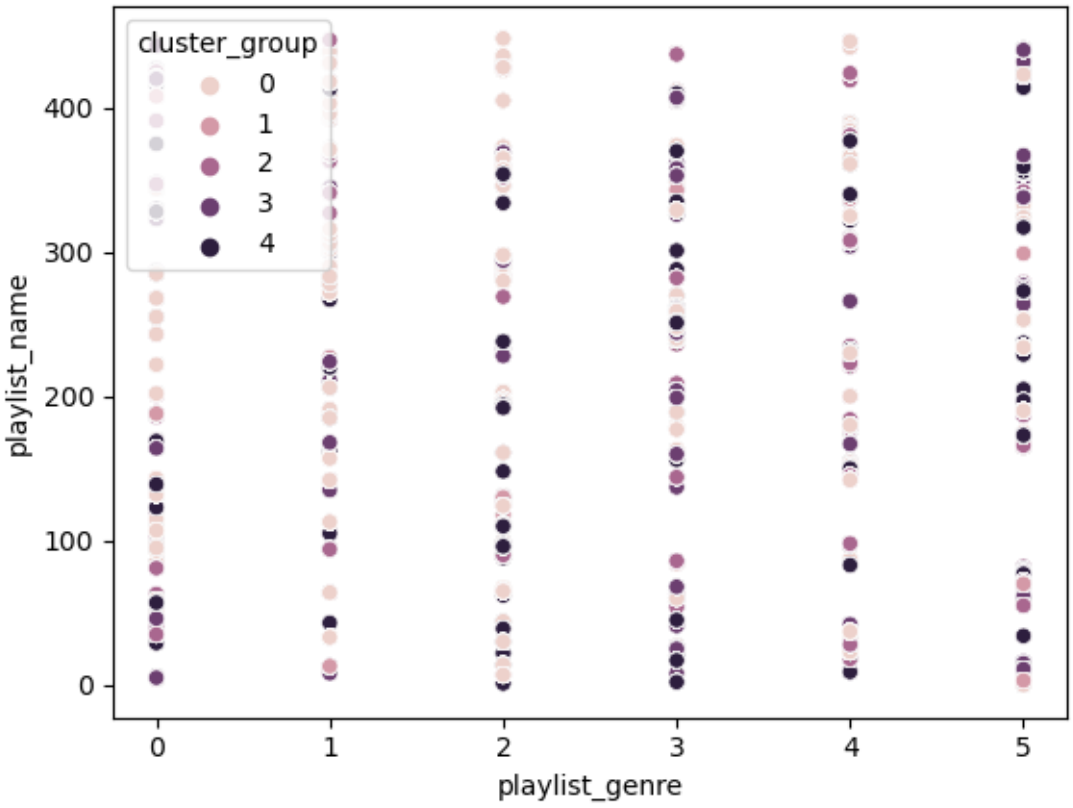
```
32832          0.341000    0.7420    0.0894  127.984       337500
```

```
       cluster_group
0                  0
1                  2
2                  2
3                  2
4                  0
…                  …
32828              0
32829              3
32830              0
32831              1
32832              3

[32833 rows x 24 columns]
```

6 plot different clusters according to different parameters like playlist genres , playlist names.

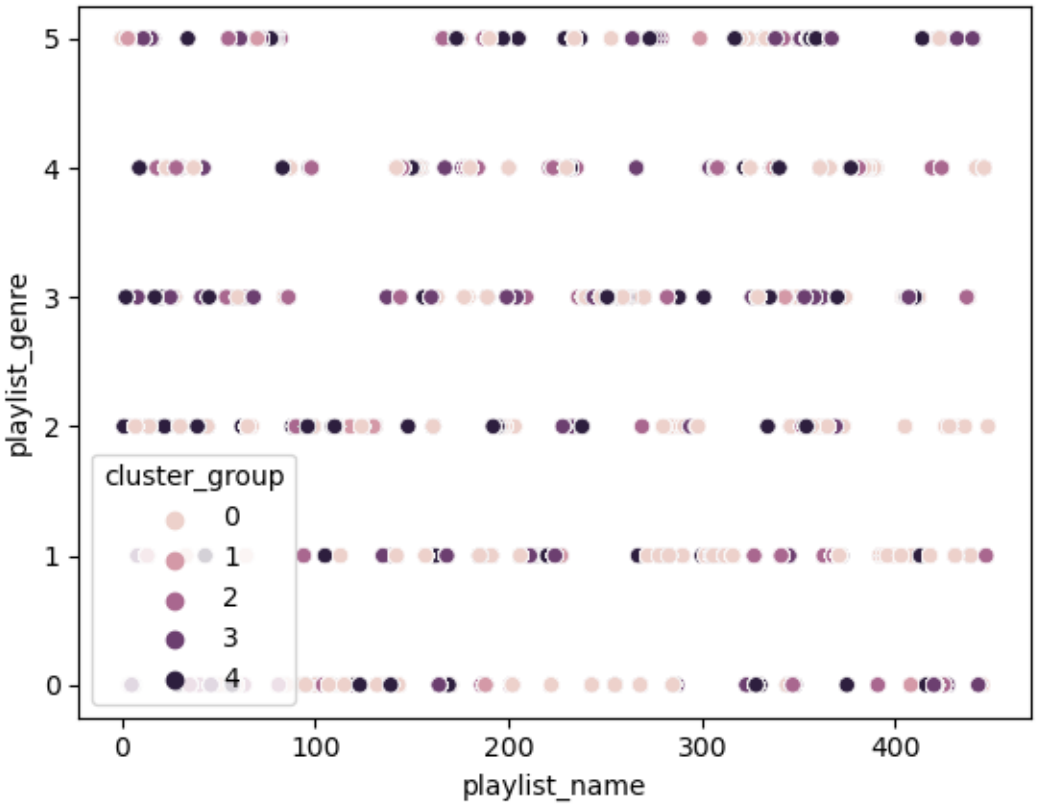
[87]:
```
sns.
↪scatterplot(x="playlist_genre",y="playlist_name",data=df,hue="cluster_group")
```

[87]:
```
<AxesSubplot:xlabel='playlist_genre', ylabel='playlist_name'>
```



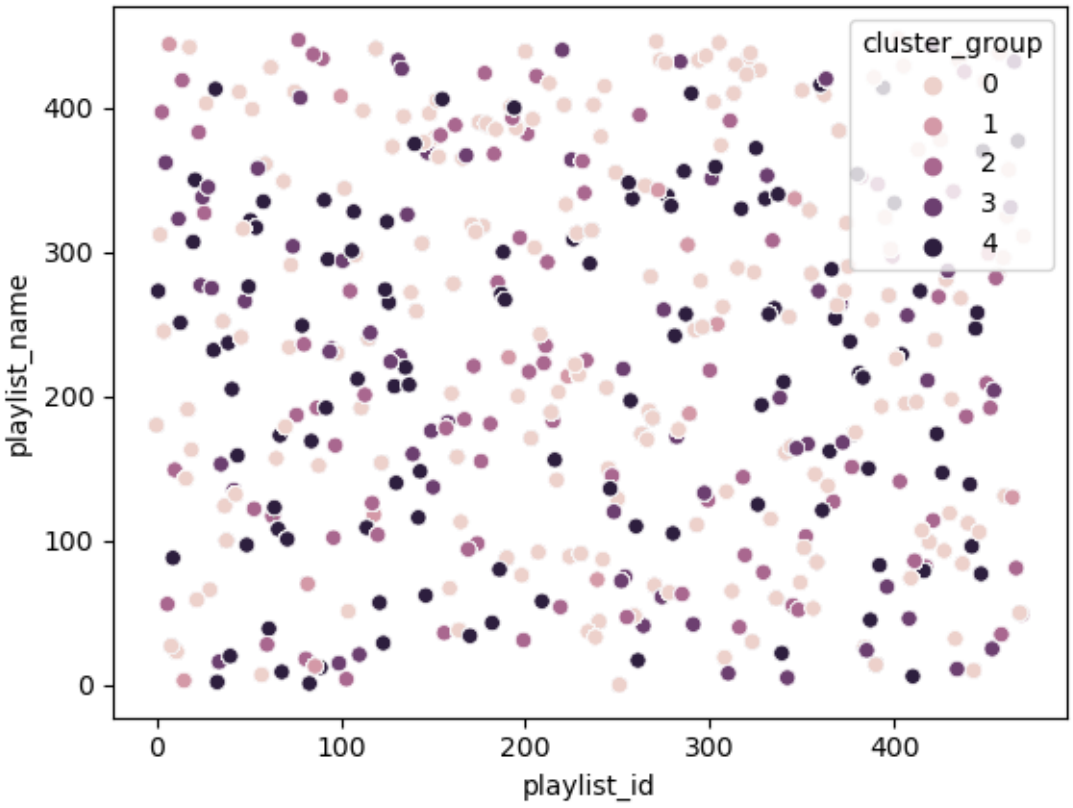
[76]:
```
sns.
↪scatterplot(x="playlist_name",y="playlist_genre",data=df,hue="cluster_group")
```

[76]: <AxesSubplot:xlabel='playlist_name', ylabel='playlist_genre'>



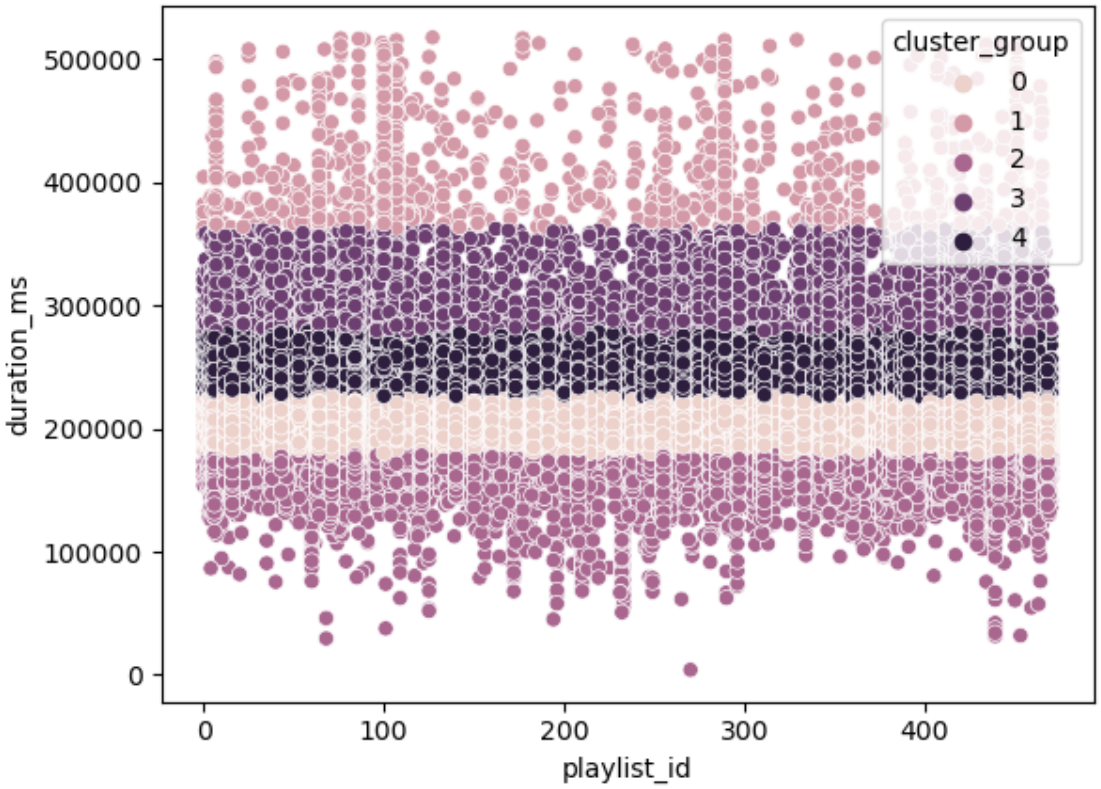
[77]:
```
sns.scatterplot(x="playlist_id",y="playlist_name",data=df,hue="cluster_group")
```

[77]:
```
<AxesSubplot:xlabel='playlist_id', ylabel='playlist_name'>
```



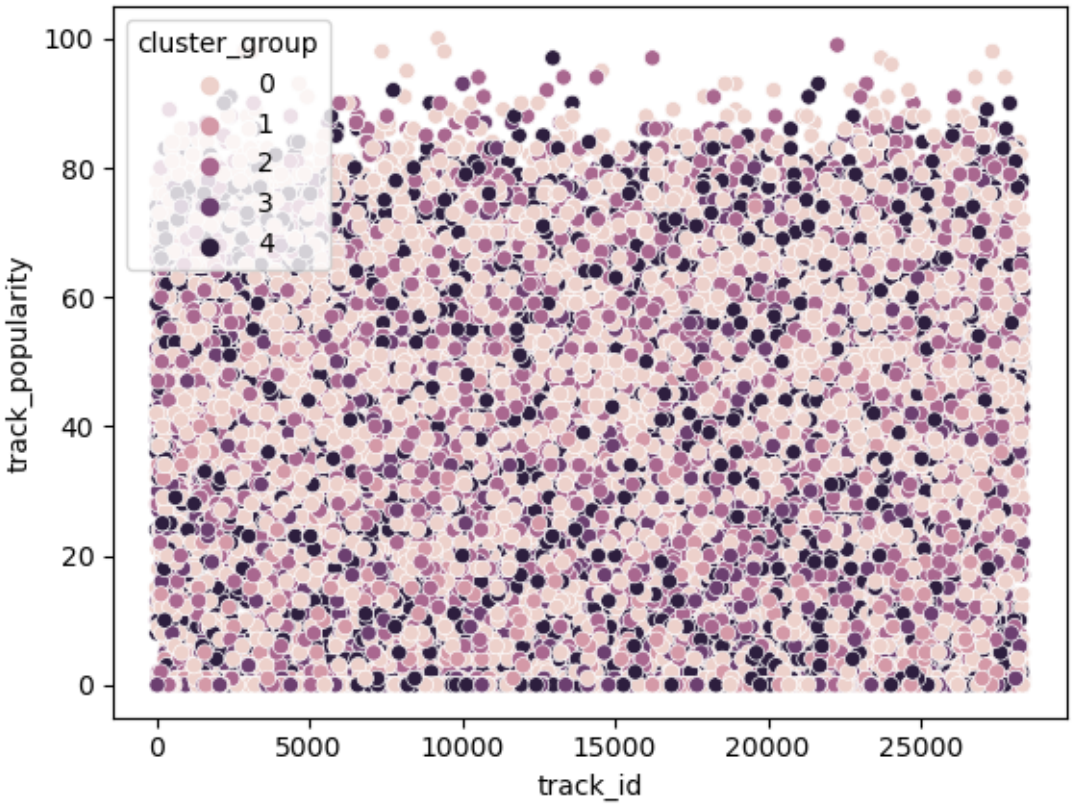
[86]:
```
sns.scatterplot(x="playlist_id",y="duration_ms",data=df,hue="cluster_group")
```

[86]: <AxesSubplot:xlabel='playlist_id', ylabel='duration_ms'>



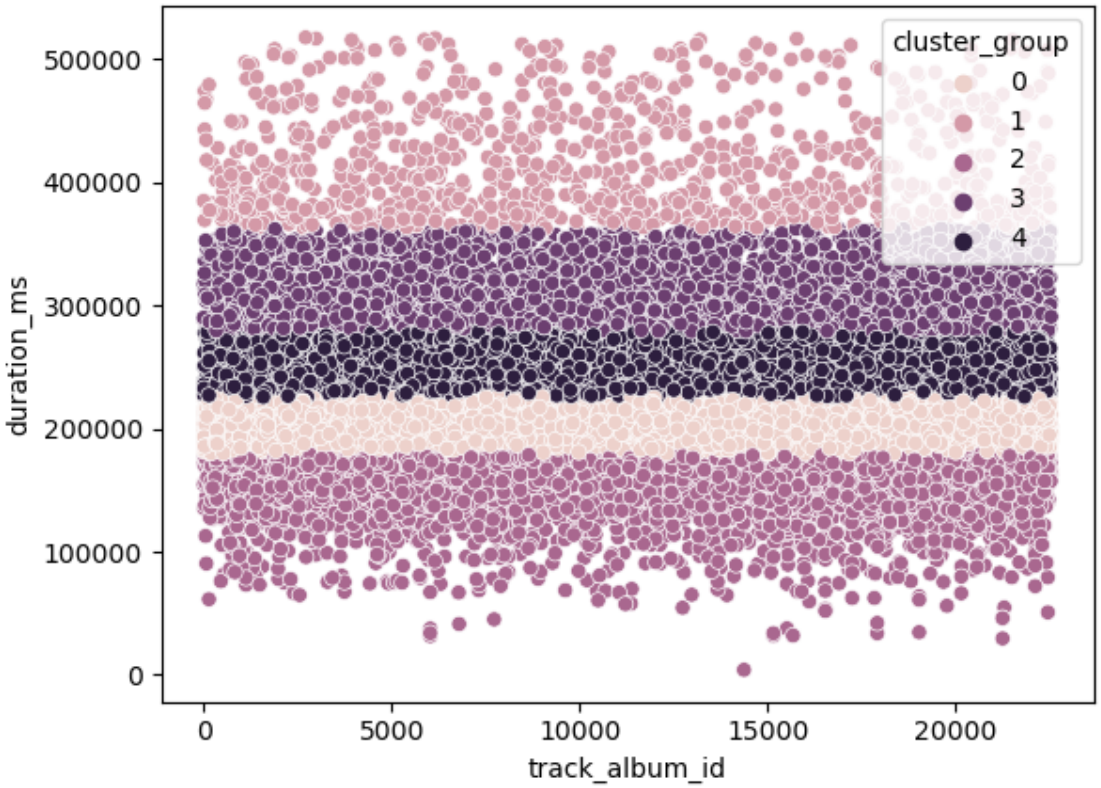
[81]:
```
sns.scatterplot(x="track_id",y="track_popularity",data=df,hue="cluster_group")
```

[81]:
```
<AxesSubplot:xlabel='track_id', ylabel='track_popularity'>
```



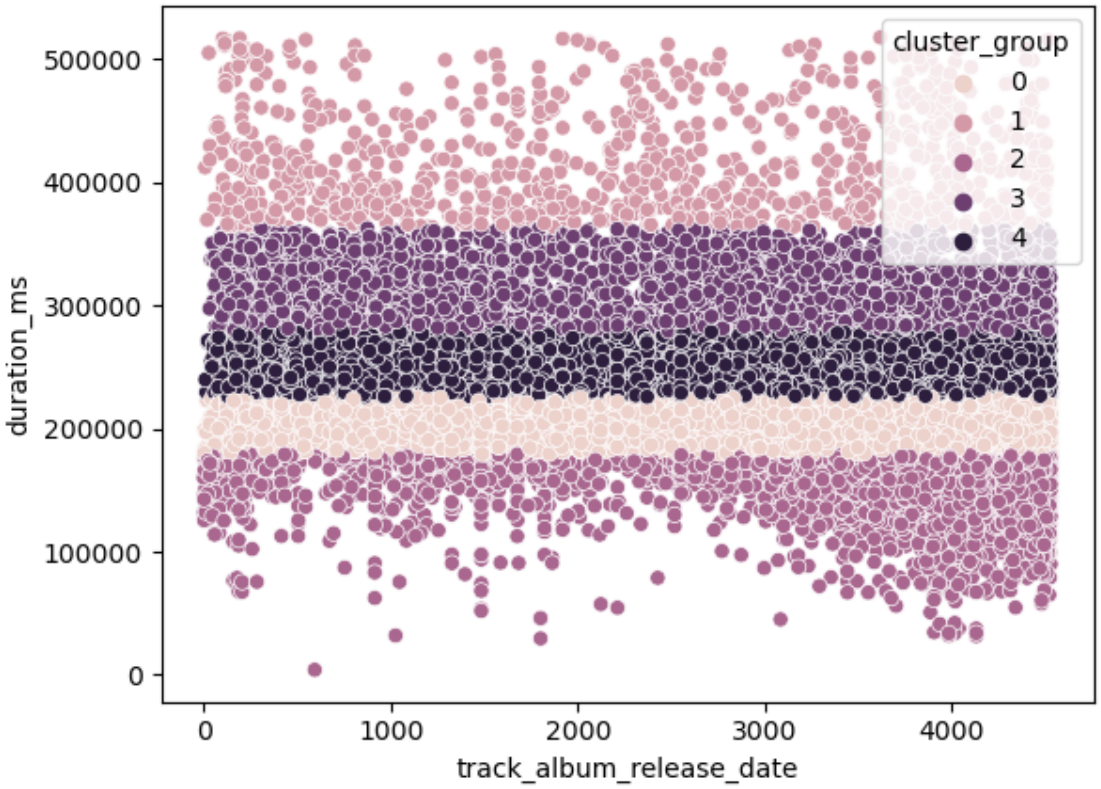
[85]:
```
sns.scatterplot(x="track_album_id",y="duration_ms",data=df,hue="cluster_group")
```

[85]:
```
<AxesSubplot:xlabel='track_album_id', ylabel='duration_ms'>
```



[89]:
```
sns.
↪scatterplot(x="track_album_release_date",y="duration_ms",data=df,hue="cluster_group")
```

[89]: <AxesSubplot:xlabel='track_album_release_date', ylabel='duration_ms'>


cluster_group
0
1
2
3
4
duration_ms
500000
400000
300000
200000
100000
0
0
1000
2000
3000
4000
track_album_release_date

[ ]: