In [34]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.cluster import KMeans
from sklearn.preprocessing import StandardScaler
from sklearn.decomposition import PCA
```

In [35]:
```
data=pd.read_csv("spotify.csv")
data
```

Out[35]:

	track_id	track_name	track_artist	track_popularity	
0	6f807x0ima9a1j3VPbc7VN	I Don't Care (with Justin Bieber) - Loud Luxur...	Ed Sheeran	66	2oCs0DGTs
1	0r7CVbZTWZgbTCYdfa2P31	Memories - Dillon Francis Remix	Maroon 5	67	63rPSO264
2	1z1Hg7Vb0AhHDiEmnDE79l	All the Time - Don Diablo Remix	Zara Larsson	70	1HoSmj2(
3	75FpbthrwQmzHlBJLuGdC7	Call You Mine - Keanu Silva Remix	The Chainsmokers	60	1nqYsOef1
4	1e8PAfcKUYoKkxPhrHqw4x	Someone You Loved - Future Humans Remix	Lewis Capaldi	69	7m7vv9w
...	...	...	...	...	
32828	7bxnKAamR3snQ1VGLuVfC1	City Of Lights - Official Radio Edit	Lush & Simon	42	2azRoBBWl
32829	5Aevni09Em4575077nkWHz	Closer - Sultan & Ned Shepard Remix	Tegan and Sara	20	6kD6KLxj
32830	7ImMqPP3Q1yfUHvsdn7wEo	Sweet Surrender - Radio Edit	Starkillers	14	0ltWNSY9
32831	2m69mhnfQ1Oq6lGtXuYhgX	Only For You - Maor Levi Remix	Mat Zo	15	1fGrOkH
32832	29zWqhca3zt5NsckZqDf6c	Typhoon - Original Mix	Julian Calor	27	0X3mUOm6M

32833 rows × 23 columns

In [36]:
```
data.describe()
```

Out[36]:

	track_popularity	danceability	energy	key	loudness	m
count	32833.000000	32833.000000	32833.000000	32833.000000	32833.000000	32833.000
mean	42.477081	0.654850	0.698619	5.374471	-6.719499	0.565
std	24.984074	0.145085	0.180910	3.611657	2.988436	0.495
min	0.000000	0.000000	0.000175	0.000000	-46.448000	0.000
25%	24.000000	0.563000	0.581000	2.000000	-8.171000	0.000
50%	45.000000	0.672000	0.721000	6.000000	-6.166000	1.000
75%	62.000000	0.761000	0.840000	9.000000	-4.645000	1.000
max	100.000000	0.983000	1.000000	11.000000	1.275000	1.000

In [37]:
```
data.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
```

In [38]:
```
data.isnull().sum()
```

Out[38]:
```
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
```

In [39]:
```
data=data.dropna()
data.head()
```

Out[39]:

	track_id	track_name	track_artist	track_popularity	track_
0	6f807x0ima9a1j3VPbc7VN	I Don't Care (with Justin Bieber) - Loud Luxur...	Ed Sheeran	66	2oCs0DGTsRO98(
1	0r7CVbZTWZgbTCYdfa2P31	Memories - Dillon Francis Remix	Maroon 5	67	63rPSO264uRjW1
2	1z1Hg7Vb0AhHDiEmnDE79l	All the Time - Don Diablo Remix	Zara Larsson	70	1HoSmj2eLcsrR(
3	75FpbthrwQmzHlBJLuGdC7	Call You Mine - Keanu Silva Remix	The Chainsmokers	60	1nqYsOef1yKKuG
4	1e8PAfcKUYoKkxPhrHqw4x	Someone You Loved - Future Humans Remix	Lewis Capaldi	69	7m7vv9wlQ4i0Ll

5 rows × 23 columns

In [40]:
```
duplicates = data.duplicated(subset=["track_name","track_artist"])
print("Duplicate Rows:")
print(data[duplicates].head())
```

```
Duplicate Rows:
                  track_id  \
193  3BOcWxFUURAomDXRSDFve4
209  4TIkSdsNSfqpuq6ZYvCjAz
232  3sHuIjfAzluc6S9cXoqfqC
272  36orMWv2PgvnzXsd5CJ0yL
294  7rpyHKSH3dkrsKEgv1eNgv

                                          track_name      track_artist
\
193                                  Something Real  Armin van Buuren
209        All You Need To Know (feat. Calle Lehmann)           Gryffin
232                                    Let It Be Me        Steve Aoki
272                          Post Malone (feat. RANI)         Sam Feldt
294  Woke Up Late (feat. Hailee Steinfeld) - Sam Fe...      Drax Project

     track_popularity          track_album_id  \
193                58  5cqwXF2j9LkvFInBFlnQd3
209                68  2IAVHJdaRPFA6MQqXHoG75
232                23  5ocW53VBnOprl6EAMOLGet
272                75  45nsubB5EsRVWWqx0ED1ET
294                56  5VW1WffQj2SqKUhwnNq1xJ

                                    track_album_name  \
193                                          Balance
209                                          Gravity
232                                    Let It Be Me
272         Post Malone (feat. RANI) [Joe Stone Remix]
294  Woke Up Late (feat. Hailee Steinfeld) [Sam Fel...

    track_album_release_date   playlist_name             playlist_id  \
193               25-10-2019      Dance Room  37i9dQZF1DX2ENAPP1Tyed
209               24-10-2019      Dance Room  37i9dQZF1DX2ENAPP1Tyed
232               06-09-2019          Cardio  37i9dQZF1DWSJHnPb1f0X3
272               16-08-2019  Dance Pop Hits  37i9dQZF1DX6pH08wMhkaI
294               02-05-2019  Dance Pop Hits  37i9dQZF1DX6pH08wMhkaI

    playlist_genre  ... key  loudness  mode  speechiness  acousticness
\
193            pop  ...   1    -4.578     1       0.0439       0.04460
209            pop  ...   0    -6.019     1       0.0376       0.00699
232            pop  ...   7    -5.299     1       0.0864       0.07970
272            pop  ...   7    -3.870     1       0.1220       0.07710
294            pop  ...   0    -3.753     1       0.0978       0.07170

     instrumentalness  liveness  valence    tempo  duration_ms
193               0.0    0.0693    0.232  127.922       179531
209               0.0    0.1030    0.219  139.929       238338
232               0.0    0.1060    0.387  114.098       224061
272               0.0    0.1050    0.651  107.356       174444
294               0.0    0.1080    0.499  122.055       196475

[5 rows x 23 columns]
```

In [41]:

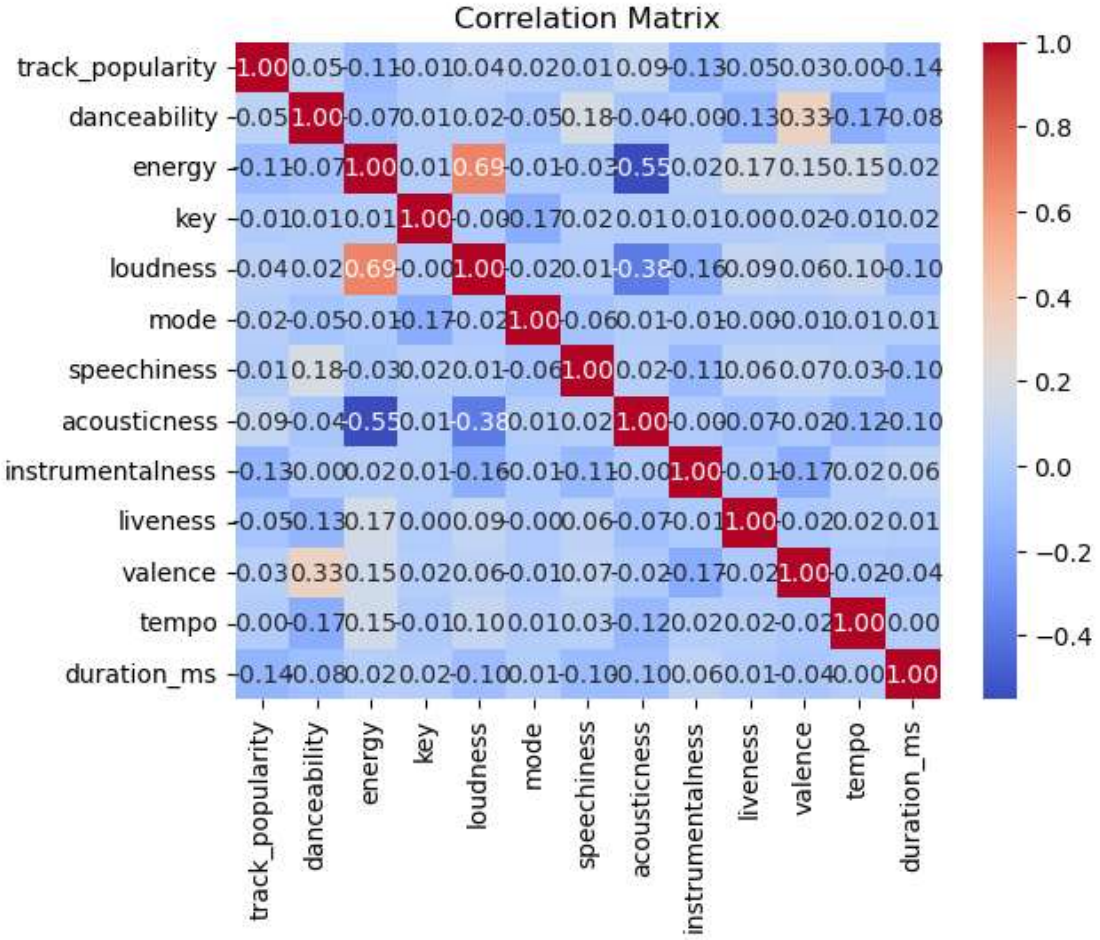
```
# Drop duplicates and keep the first occurrence
data = data.drop_duplicates(subset=["track_name","track_artist"])
print("\nDataFrame after removing duplicates:")
print(data.head())
```

```
DataFrame after removing duplicates:
                track_id                                          track_
name  \
0  6f807x0ima9a1j3VPbc7VN  I Don't Care (with Justin Bieber) - Loud Luxu
r...
1  0r7CVbZTWZgbTCYdfa2P31                    Memories - Dillon Francis R
emix
2  1z1Hg7Vb0AhHDiEmnDE79l                    All the Time - Don Diablo R
emix
3  75FpbthrwQmzHlBJLuGdC7                  Call You Mine - Keanu Silva R
emix
4  1e8PAfcKUYoKkxPhrHqw4x            Someone You Loved - Future Humans R
emix

       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...               14-06
-2019
1                    Memories (Dillon Francis Remix)               13-12
-2019
2                    All the Time (Don Diablo Remix)               05-07
-2019
3                        Call You Mine - The Remixes               19-07
-2019
4            Someone You Loved (Future Humans Remix)               05-03
-2019

  playlist_name            playlist_id playlist_genre  ... key  loudnes
s  \
0     Pop Remix  37i9dQZF1DXcZDD7cfEKhW            pop  ...   6    -2.63
4
1     Pop Remix  37i9dQZF1DXcZDD7cfEKhW            pop  ...  11    -4.96
9
2     Pop Remix  37i9dQZF1DXcZDD7cfEKhW            pop  ...   1    -3.43
2
3     Pop Remix  37i9dQZF1DXcZDD7cfEKhW            pop  ...   7    -3.77
8
4     Pop Remix  37i9dQZF1DXcZDD7cfEKhW            pop  ...   1    -4.67
2

   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]
```

EXPLORATORY DATA ANALYSIS (EDA)

In [42]:
```
#correlation matrix
correlation_matrix = data.corr()
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt=".2f")
plt.title("Correlation Matrix")
plt.show()
```



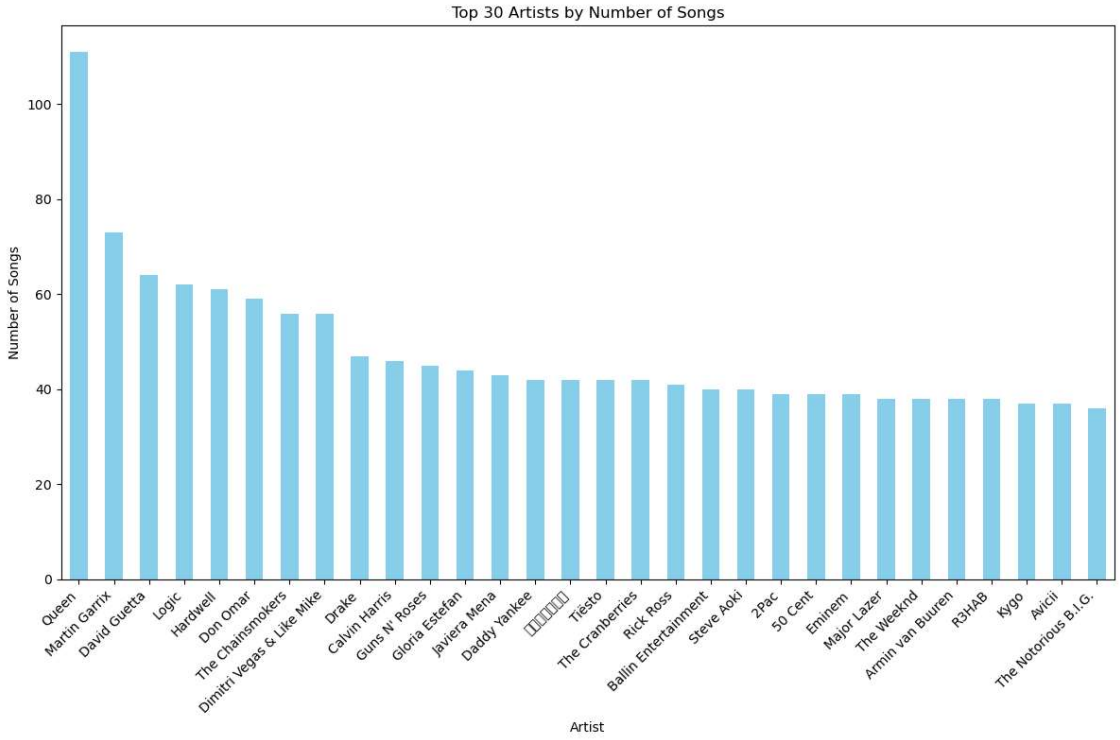
In [43]:
```
artist_counts = data['track_artist'].value_counts()
print(artist_counts.head(30))
```

```
Queen                          111
Martin Garrix                   73
David Guetta                    64
Logic                           62
Hardwell                        61
Don Omar                        59
The Chainsmokers                56
Dimitri Vegas & Like Mike       56
Drake                           47
Calvin Harris                   46
Guns N' Roses                   45
Gloria Estefan                  44
Javiera Mena                    43
Daddy Yankee                    42
オメガトライブ                         42
Tiësto                          42
The Cranberries                 42
Rick Ross                       41
Ballin Entertainment            40
Steve Aoki                      40
2Pac                            39
50 Cent                         39
Eminem                          39
Major Lazer                     38
The Weeknd                      38
Armin van Buuren                38
R3HAB                           38
Kygo                            37
Avicii                          37
The Notorious B.I.G.            36
Name: track_artist, dtype: int64
```

In [44]:

```
top_30_artists = artist_counts.head(30)
#bar graph
plt.figure(figsize=(12, 8))
top_30_artists.plot(kind='bar', color='skyblue')
plt.title('Top 30 Artists by Number of Songs')
plt.xlabel('Artist')
plt.ylabel('Number of Songs')
plt.xticks(rotation=45, ha='right')
plt.tight_layout()
plt.show()
```

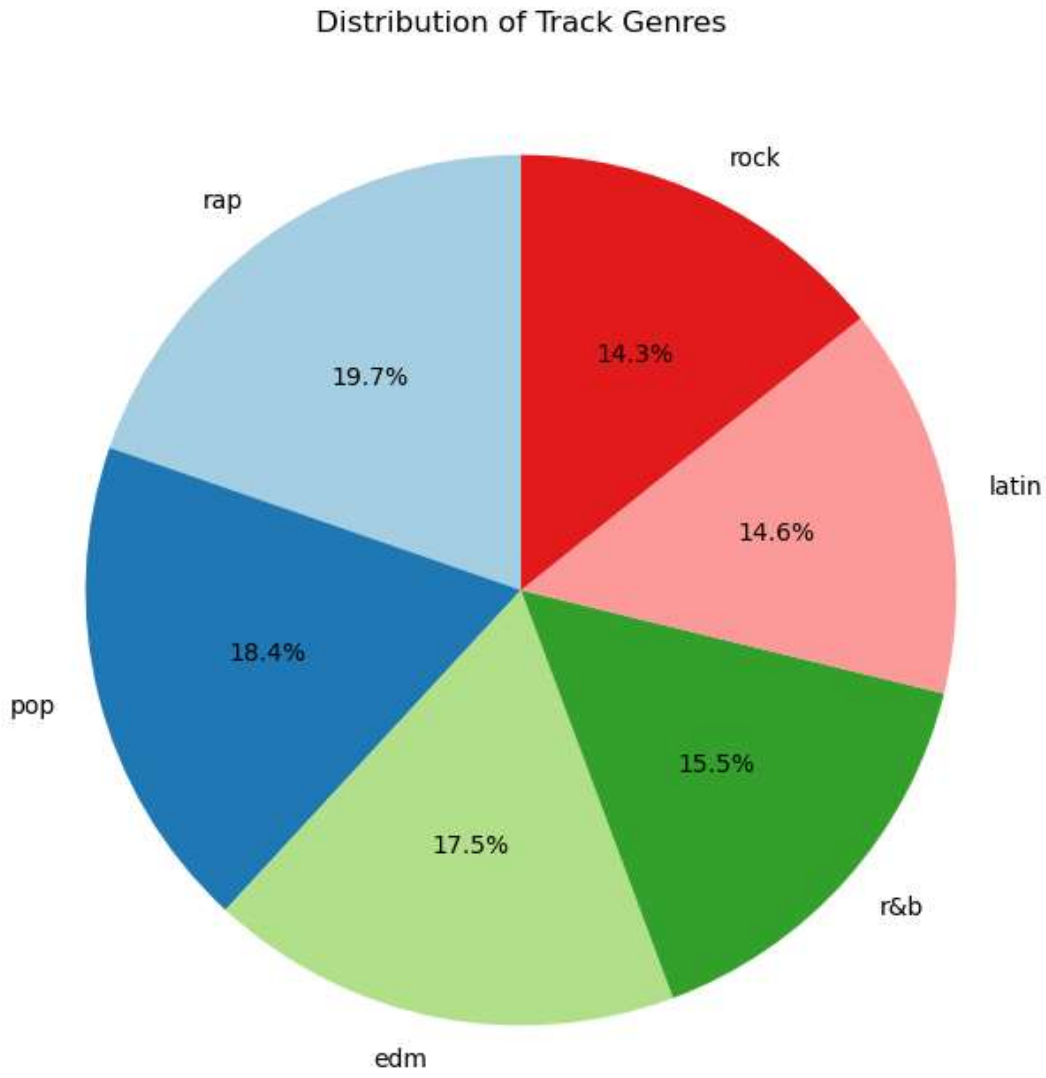
```
C:\Users\HP\AppData\Local\Temp\ipykernel_8836\2882317381.py:9: UserWarni
ng: Glyph 12458 (\N{KATAKANA LETTER O}) missing from current font.
  plt.tight_layout()
C:\Users\HP\AppData\Local\Temp\ipykernel_8836\2882317381.py:9: UserWarni
ng: Glyph 12513 (\N{KATAKANA LETTER ME}) missing from current font.
  plt.tight_layout()
C:\Users\HP\AppData\Local\Temp\ipykernel_8836\2882317381.py:9: UserWarni
ng: Glyph 12460 (\N{KATAKANA LETTER GA}) missing from current font.
  plt.tight_layout()
C:\Users\HP\AppData\Local\Temp\ipykernel_8836\2882317381.py:9: UserWarni
ng: Glyph 12488 (\N{KATAKANA LETTER TO}) missing from current font.
  plt.tight_layout()
C:\Users\HP\AppData\Local\Temp\ipykernel_8836\2882317381.py:9: UserWarni
ng: Glyph 12521 (\N{KATAKANA LETTER RA}) missing from current font.
  plt.tight_layout()
C:\Users\HP\AppData\Local\Temp\ipykernel_8836\2882317381.py:9: UserWarni
ng: Glyph 12452 (\N{KATAKANA LETTER I}) missing from current font.
  plt.tight_layout()
C:\Users\HP\AppData\Local\Temp\ipykernel_8836\2882317381.py:9: UserWarni
ng: Glyph 12502 (\N{KATAKANA LETTER BU}) missing from current font.
  plt.tight_layout()
C:\Users\HP\anaconda3\lib\site-packages\IPython\core\pylabtools.py:151:
UserWarning: Glyph 12458 (\N{KATAKANA LETTER O}) missing from current fo
nt.
  fig.canvas.print_figure(bytes_io, **kw)
C:\Users\HP\anaconda3\lib\site-packages\IPython\core\pylabtools.py:151:
UserWarning: Glyph 12513 (\N{KATAKANA LETTER ME}) missing from current f
ont.
  fig.canvas.print_figure(bytes_io, **kw)
C:\Users\HP\anaconda3\lib\site-packages\IPython\core\pylabtools.py:151:
UserWarning: Glyph 12460 (\N{KATAKANA LETTER GA}) missing from current f
ont.
  fig.canvas.print_figure(bytes_io, **kw)
C:\Users\HP\anaconda3\lib\site-packages\IPython\core\pylabtools.py:151:
UserWarning: Glyph 12488 (\N{KATAKANA LETTER TO}) missing from current f
ont.
  fig.canvas.print_figure(bytes_io, **kw)
C:\Users\HP\anaconda3\lib\site-packages\IPython\core\pylabtools.py:151:
UserWarning: Glyph 12521 (\N{KATAKANA LETTER RA}) missing from current f
ont.
  fig.canvas.print_figure(bytes_io, **kw)
C:\Users\HP\anaconda3\lib\site-packages\IPython\core\pylabtools.py:151:
UserWarning: Glyph 12452 (\N{KATAKANA LETTER I}) missing from current fo
nt.
  fig.canvas.print_figure(bytes_io, **kw)
C:\Users\HP\anaconda3\lib\site-packages\IPython\core\pylabtools.py:151:
UserWarning: Glyph 12502 (\N{KATAKANA LETTER BU}) missing from current f
ont.
  fig.canvas.print_figure(bytes_io, **kw)
```



In [45]:
```
data['playlist_genre'].value_counts()
```

Out[45]:
```
rap      5165
pop      4836
edm      4603
r&b      4067
latin    3819
rock     3739
Name: playlist_genre, dtype: int64
```

In [46]:
```
# Count the number of occurrences for each track genre
genre_counts = data['playlist_genre'].value_counts()

# Plot the pie chart
plt.figure(figsize=(8, 8))
plt.pie(genre_counts, labels=genre_counts.index, autopct='%1.1f%%', starta
plt.title('Distribution of Track Genres')
plt.show()
```

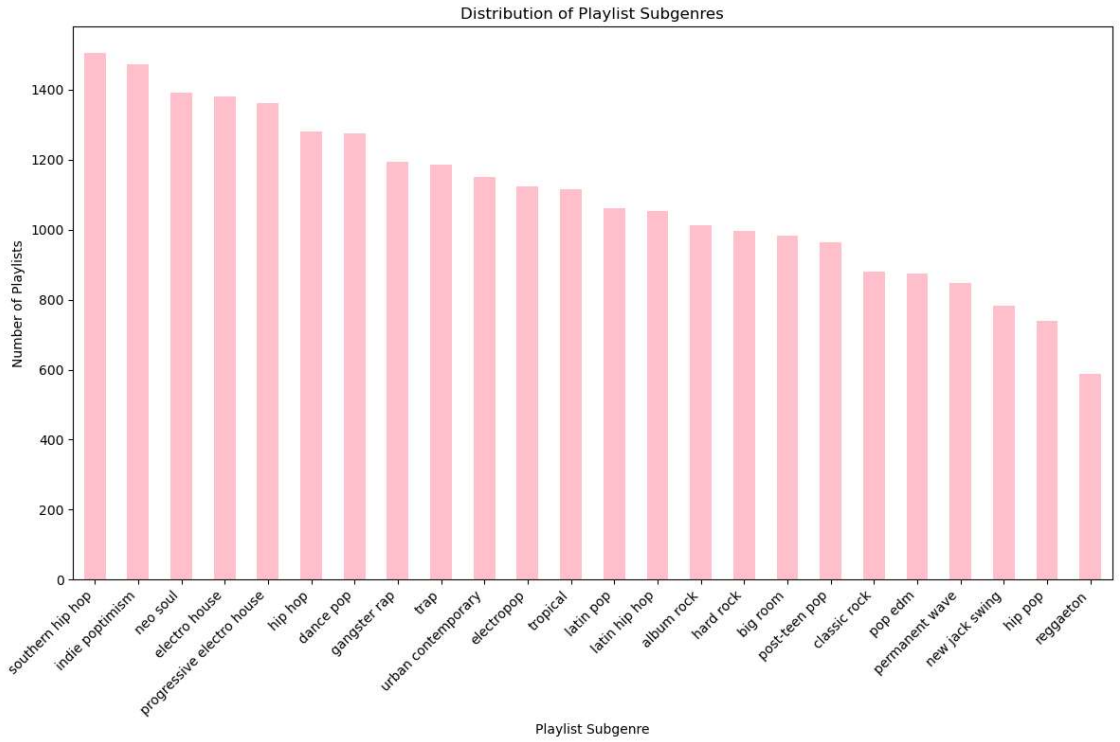


In [47]:
```
data['playlist_subgenre'].value_counts()
```

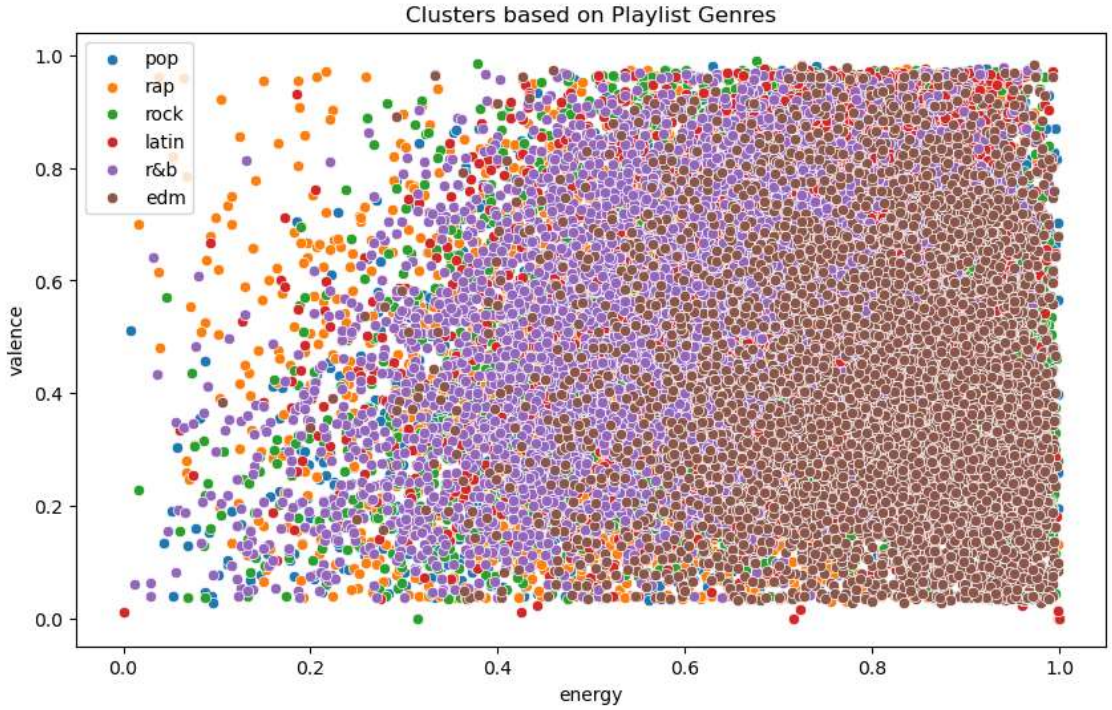
Out[47]:
```
southern hip hop            1505
indie poptimism             1474
neo soul                    1393
electro house               1382
progressive electro house   1363
hip hop                     1281
dance pop                   1276
gangster rap                1194
trap                        1185
urban contemporary          1151
electropop                  1123
tropical                    1117
latin pop                   1061
latin hip hop               1053
album rock                  1013
hard rock                    997
big room                     982
post-teen pop                963
classic rock                 880
pop edm                      876
permanent wave               849
new jack swing               784
hip pop                      739
reggaeton                    588
Name: playlist_subgenre, dtype: int64
```

In [48]:

```
# Count the number of occurrences for each playlist subgenre
subgenre_counts = data['playlist_subgenre'].value_counts()

# Plot the bar graph
plt.figure(figsize=(12, 8))
subgenre_counts.plot(kind='bar', color='pink')
plt.title('Distribution of Playlist Subgenres')
plt.xlabel('Playlist Subgenre')
plt.ylabel('Number of Playlists')
plt.xticks(rotation=45, ha='right')
plt.tight_layout()
plt.show()
```



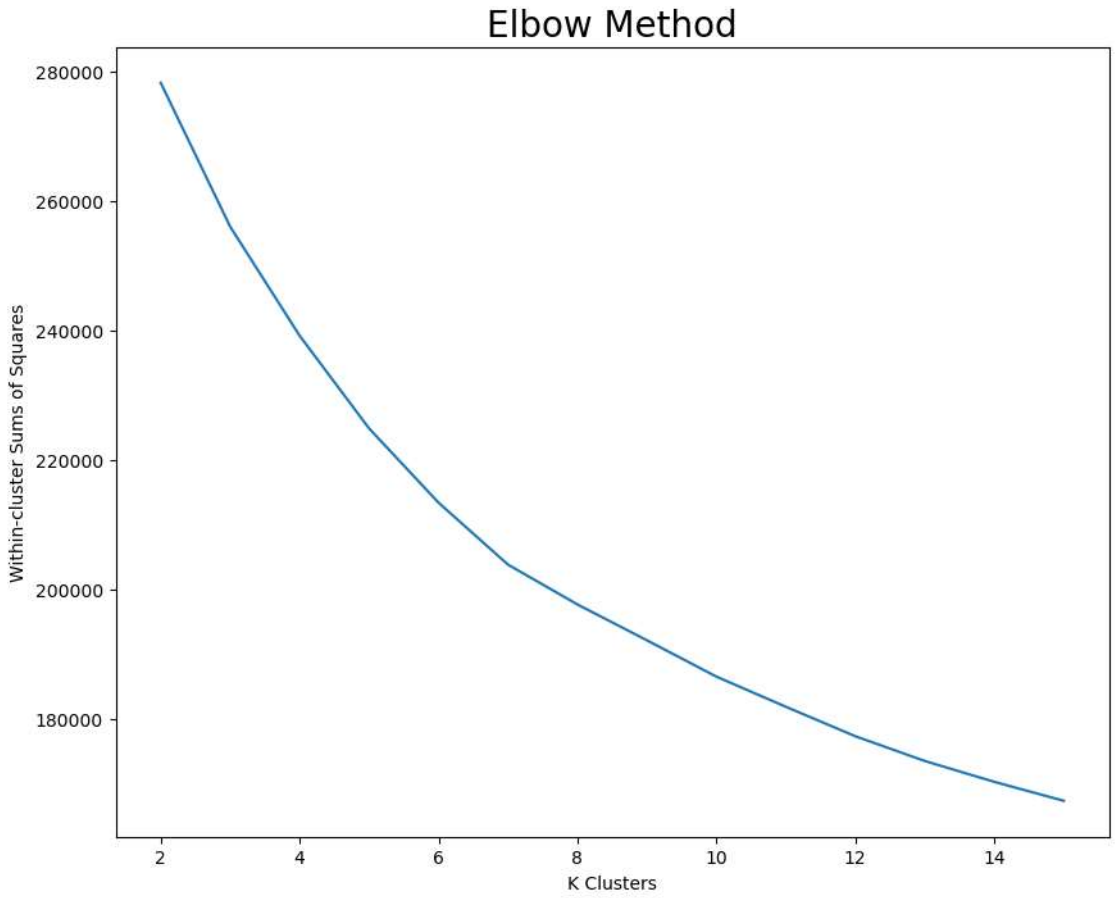
In [49]:
```
# Clustering based on playlist genres
plt.figure(figsize=(10, 6))
playlist_genres = data['playlist_genre'].unique()
for genre in playlist_genres:
    genre_data = data[data['playlist_genre'] == genre]
    sns.scatterplot(x='energy', y='valence', data=genre_data, label=genre)
plt.title('Clusters based on Playlist Genres')
plt.legend()
plt.show()
```



In [ ]:
```
```

In [50]:
```
from sklearn.cluster import KMeans
from sklearn.decomposition import PCA
from sklearn.metrics import silhouette_score
from sklearn.pipeline import Pipeline
from sklearn.preprocessing import StandardScaler
cols = data.iloc[:, 9:].select_dtypes(['uint8', 'int64', 'float64']).colum
```

In [51]:
```
wcss = []
for i in range(2,16):
    km = KMeans(n_clusters=i, init='k-means++', max_iter=500, n_init=10, r
    d = StandardScaler().fit_transform(data[cols])
    km.fit(d)
    wcss.append(km.inertia_)
fig = plt.figure(figsize=(10,8))
ax = sns.lineplot(x=range(2,16), y=wcss)
ax.set_title('Elbow Method')
ax.title.set_size(20)
plt.xlabel('K Clusters')
plt.ylabel('Within-cluster Sums of Squares')
plt.show()
```



In [52]:
```
new_data=data[cols].dropna()
scaler = StandardScaler()
scaled_data = scaler.fit_transform(new_data)
k=4
kmeans = KMeans(n_clusters=k, random_state=1)
cluster_labels = kmeans.fit_predict(scaled_data)
silhouette_avg = silhouette_score(scaled_data, cluster_labels)
print("Silhouette Score:", silhouette_avg)
```

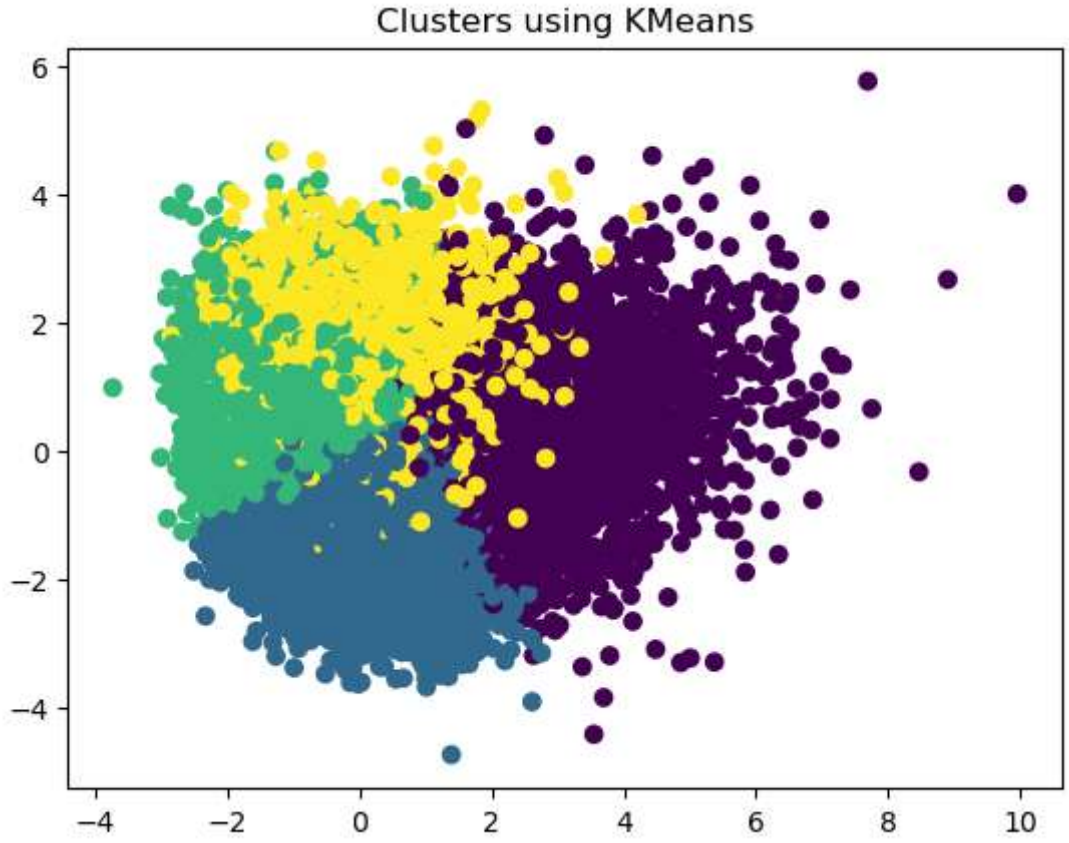
```
Silhouette Score: 0.10559443877113443
```

In [53]:
```
cols
```

Out[53]:
```
Index(['danceability', 'energy', 'key', 'loudness', 'mode', 'speechines
s',
       'acousticness', 'instrumentalness', 'liveness', 'valence', 'temp
o',
       'duration_ms'],
      dtype='object')
```

In [54]:

```
# Build Model (Example using KMeans)
features_for_clustering = ['danceability', 'energy', 'key', 'loudness', 'm
       'speechiness', 'acousticness', 'instrumentalness', 'liveness',
       'valence', 'tempo', 'duration_ms']

X = data[features_for_clustering]

# Standardize the features
scaler = StandardScaler()
X_standardized = scaler.fit_transform(X)

# Use KMeans for clustering
kmeans = KMeans(n_clusters=4, random_state=42)
data['cluster'] = kmeans.fit_predict(X_standardized)

# Plot clusters in 2D using PCA
pca = PCA(n_components=2)
X_pca = pca.fit_transform(X_standardized)
plt.scatter(X_pca[:, 0], X_pca[:, 1], c=data['cluster'], cmap='viridis')
plt.title('Clusters using KMeans')
plt.show()
```



In [55]:
```
top_tracks_cluster_0 = data[data['cluster'] == 0].nlargest(15, 'track_popu
print("Top 15 tracks from Cluster 0:")
print(top_tracks_cluster_0[['track_name', 'track_artist', 'track_popularit
```

```
Top 15 tracks from Cluster 0:
                      track_name   track_artist  track_popularity
711                     Memories       Maroon 5                98
689           everything i wanted  Billie Eilish                97
3155                      Falling  Trevor Daniel                97
344                       bad guy  Billie Eilish                95
718             Someone You Loved  Lewis Capaldi                94
1303          HIGHEST IN THE ROOM   Travis Scott                94
739           Lose You To Love Me   Selena Gomez                93
922                 Dance Monkey    Tones and I                92
5533                      No Idea    Don Toliver                92
742               Before You Go  Lewis Capaldi                90
1319      Trampoline (with ZAYN)          SHAED                90
1782                      7 rings  Ariana Grande                90
5517                     Suicidal      YNW Melly                90
5929   Ballin' (with Roddy Ricch)        Mustard                90
22142       lovely (with Khalid)  Billie Eilish                89
```

In [56]:
```
top_tracks_cluster_1 = data[data['cluster'] == 1].nlargest(15, 'track_popu
print("Top 15 tracks from Cluster 1:")
print(top_tracks_cluster_1[['track_name', 'track_artist', 'track_popularit
```

```
Top 15 tracks from Cluster 1:
                      track_name         track_artist  track_popularity
687                         Tusa              KAROL G                98
5508                     The Box           Roddy Ricch                98
1302             Don't Start Now             Dua Lipa                97
139     RITMO (Bad Boys For Life)  The Black Eyed Peas                96
694                         Yummy        Justin Bieber                95
646                       Ride It               Regard                94
1306             hot girl bummer            blackbear                94
1463      My Oh My (feat. DaBaby)       Camila Cabello                94
702                         China             Anuel AA                93
709                      Señorita         Shawn Mendes                93
5507   Life Is Good (feat. Drake)               Future                93
5514                          BOP               DaBaby                93
11115                        Vete            Bad Bunny                93
1339                     Fantasias       Rauw Alejandro                92
179                  Lose Control               MEDUZA                91
```

In [57]:
```
top_tracks_cluster_2 = data[data['cluster'] == 2].nlargest(15, 'track_popu
print("Top 15 tracks from Cluster 2:")
print(top_tracks_cluster_2[['track_name', 'track_artist', 'track_popularit
```

```
Top 15 tracks from Cluster 2:
                                                track_name    track_artist
\
716                                        Blinding Lights      The Weeknd
730                                              Heartless      The Weeknd
720              Bandit (with YoungBoy Never Broke Again)      Juice WRLD
1314                                    How Do You Sleep?       Sam Smith
1333                                             Callaita       Bad Bunny
1611                                     Watermelon Sugar    Harry Styles
18360                        Goodbyes (feat. Young Thug)     Post Malone
25788                      All I Want for Christmas Is You    Mariah Carey
1327           Good as Hell (feat. Ariana Grande) - Remix           Lizzo
5540   HIGHEST IN THE ROOM (feat. ROSALÍA & Lil Baby)...    Travis Scott
1341                Loco Contigo (feat. J. Balvin & Tyga)        DJ Snake
1344                                             Si Te Vas           Sech
8980                           rockstar (feat. 21 Savage)     Post Malone
10785                                            Robbery      Juice WRLD
18868                                             Quizas  Rich Music LTD

       track_popularity
716                  98
730                  93
720                  92
1314                 91
1333                 90
1611                 90
18360                90
25788                90
1327                 89
5540                 89
1341                 88
1344                 88
8980                 88
10785                88
18868                88
```

In [58]:
```
top_tracks_cluster_3 = data[data['cluster'] == 3].nlargest(15, 'track_popu
print("Top 15 tracks from Cluster 3:")
print(top_tracks_cluster_3[['track_name', 'track_artist', 'track_popularit
```

```
Top 15 tracks from Cluster 3:
                          track_name          track_artist  \
18562  Baila Conmigo (feat. Kelly Ruiz)                 Dayvi
22997                      In My Room           Frank Ocean
634                            Baianá             Bakermat
1110    Feels Like We Only Go Backwards           Tame Impala
1223                     Instant Crush             Daft Punk
5054                 On Melancholy Hill             Gorillaz
14448            Heroes - 2017 Remaster           David Bowie
13622            1979 - Remastered 2012  The Smashing Pumpkins
23031                            A Palé               ROSALÍA
11762                   Born To Be Wild           Steppenwolf
13672                          Reptilia           The Strokes
14051       Blue Monday - 2016 Remaster             New Order
14575             Chamber Of Reflection           Mac DeMarco
4147                              Intro                The xx
10846                   Look at Me Now        Brennan Savage

       track_popularity
18562                83
22997                79
634                  76
1110                 75
1223                 75
5054                 75
14448                75
13622                74
23031                74
11762                73
13672                73
14051                73
14575                73
4147                 72
10846                72
```

In [59]:
```
data['track_popularity'].describe()
```

Out[59]:
```
count    26229.000000
mean        39.508178
std         23.265382
min          0.000000
25%         22.000000
50%         42.000000
75%         58.000000
max         98.000000
Name: track_popularity, dtype: float64
```

In [60]:
```
data['cluster'].value_counts()
```

Out[60]:
```
1    10898
2     8740
0     4285
3     2306
Name: cluster, dtype: int64
```

In [ ]:
```

```

In [61]:

```
from sklearn.metrics.pairwise import cosine_similarity

data['new_index'] = range(1, len(data) + 1)
data.set_index('new_index', inplace=True)

# Create keywords column by combining track_name and track_artist
data['keywords'] = data['track_name'] + ' ' + data['track_artist']

# Extract genres and create a genre matrix
genres = data['playlist_genre'].str.split(',').explode().str.strip()
genre_matrix = pd.get_dummies(genres).groupby(level=0).sum()

# Combine keywords and genre_matrix
feature_matrix = pd.concat([data['keywords'], genre_matrix], axis=1)

# Create user profile based on favorite genres
favorite_genres = ['edm','pop']
user_profile = pd.DataFrame({'genre_matrix': np.isin(genre_matrix.columns,

# Calculate similarity scores
similarity_scores = cosine_similarity(feature_matrix.drop(columns='keyword

# Combine similarity scores with the original data
data_recommendations = pd.concat([data, pd.DataFrame({'similarity': simila

# Get top N recommendations
N = 25
sorted_recommendations = data_recommendations.sort_values(by='similarity',
top_songs = sorted_recommendations.head(N)[["track_name", "track_artist",

# Display top recommendations
print(top_songs)
```

	track_name	track_artist
1	I Don't Care (with Justin Bieber) - Loud Luxur...	Ed Sheeran
23095	The End	SpinRox
23079	Haunted	Dave Mak
23080	Mad Echoes	Swede Dreams
23081	Raise	VITIZE
23082	Gringos	Kevu
23083	Don't Stop	H2HB
23084	Omega	KaZo
23085	Never	Widemode
23086	Mantra	NAEMS
23087	Outta Control	Winning Team
23088	Dominica	NAEMS
23089	No Fear	Latwer
23090	Shotgun	RudeLies
23091	Watch Your Back	FineRefined
23092	Don't Wanna Be Cold	Krimsonn
23093	Take Me	Castion
23078	Back To Life	KRISM
23077	Symphony - Dr Phunk Remix	Sick Individuals
23076	LUNA	Sick Individuals
23067	Up In Smoke	KAAZE
23061	Beast	PURARI
23062	Losing My Religion - Ton Don Remix	Jones & Brock
23063	Black Butterflies	Kill The Buzz
23064	Game Time	Swanky Tunes

	playlist_genre	track_popularity	similarity
1	pop	66.0	0.707107
23095	edm	18.0	0.707107
23079	edm	20.0	0.707107

23080	edm	20.0	0.707107
23081	edm	31.0	0.707107
23082	edm	21.0	0.707107
23083	edm	17.0	0.707107
23084	edm	15.0	0.707107
23085	edm	36.0	0.707107
23086	edm	19.0	0.707107
23087	edm	22.0	0.707107
23088	edm	18.0	0.707107
23089	edm	17.0	0.707107
23090	edm	19.0	0.707107
23091	edm	16.0	0.707107
23092	edm	17.0	0.707107
23093	edm	17.0	0.707107
23078	edm	32.0	0.707107
23077	edm	49.0	0.707107
23076	edm	55.0	0.707107
23067	edm	33.0	0.707107
23061	edm	38.0	0.707107
23062	edm	32.0	0.707107
23063	edm	38.0	0.707107
23064	edm	33.0	0.707107

In [ ]: