spotify-tracks-dataset-analysis

November 14, 2024

IMPORTING LIBRARIES

[118]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

IMPORTING DATASET

[119]:
```
data = pd.read_csv(r"H:\DA. Python\15.  Spotify Tracks Dataset Analysis\tracks.
↪csv")
```

CHECKING THE DATASET

[120]:
```
data.shape
```

[120]:
```
(586672, 20)
```

[121]:
```
data.head(2)
```

[121]:

	id	name	popularity	duration_ms	explicit	artists	id_artists	release_date	danceability	energy	key	loudness	mode	speechiness	acousticness	instrumentalness	liveness	valence	tempo	time_signature
0	35iwgR4jXetI318WEWsa1Q	Carve	6	126903	0	['Uli']	['45tIt06XoI0Iio4LBEVpls']	22-02-1922	0.645	0.445	0	-13.338	1	0.451	0.674	0.744	0.151	0.127	104.851	3
1	021ht4sdgPcrDgSk7JTbKY	Capítulo 2.16 - Banquero Anarquista	0	98200	0	['Fernando Pessoa']	['14jtPCOoNZwquk5wd9DxrY']	01-06-1922	0.695	0.263	0	-22.136	1	0.957	0.797	0.000	0.148	0.655	102.009	1

[122]:
```
data.tail(2)
```

[122]:

	id	name	popularity	duration_ms	explicit	artists	id_artists	release_date	danceability	energy	key	loudness	mode	speechiness	acousticness	instrumentalness	liveness	valence	tempo	time_signature
586670	45XJsGpFTyzbzeWK8VzR8S	A Day At A Time	58	142003	0	['Gentle Bones', 'Clara Benin']	['4jGPdu95icCKVF31CcFKbS', '5ebPSE9YI5aLeZ1Z2g…	05-03-2021	0.696	0.615	10	-6.212	1	0.0345	0.206	0.000003	0.305	0.438	90.029	4
586671	5Ocn6dZ3BJFPWh4ylwFXtn	Mar de Emociones	38	214360	0	['Afrosound']	['0i4Qda0k4nf7jnNHmSNpYv']	01-07-2015	0.686	0.723	6	-7.067	1	0.0363	0.105	0.000000	0.264	0.975	112.204	4

[123]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 586672 entries, 0 to 586671
Data columns (total 20 columns):
 #   Column            Non-Null Count   Dtype
---  ------            --------------   -----
 0   id                586672 non-null  object
 1   name              586601 non-null  object
 2   popularity        586672 non-null  int64
 3   duration_ms       586672 non-null  int64
 4   explicit          586672 non-null  int64
 5   artists           586672 non-null  object
 6   id_artists        586672 non-null  object
 7   release_date      586672 non-null  object
 8   danceability      586672 non-null  float64
 9   energy            586672 non-null  float64
 10  key               586672 non-null  int64
 11  loudness          586672 non-null  float64
 12  mode              586672 non-null  int64
 13  speechiness       586672 non-null  float64
 14  acousticness      586672 non-null  float64
 15  instrumentalness  586672 non-null  float64
 16  liveness          586672 non-null  float64
 17  valence           586672 non-null  float64
```

```
 18  tempo             586672 non-null  float64
 19  time_signature    586672 non-null  int64
dtypes: float64(9), int64(6), object(5)
memory usage: 89.5+ MB
```

[124]:
```
data.dtypes
```

[124]:
```
id                  object
name                object
popularity           int64
duration_ms          int64
explicit             int64
artists             object
id_artists          object
release_date        object
danceability       float64
energy             float64
key                  int64
loudness           float64
mode                 int64
speechiness        float64
acousticness       float64
instrumentalness   float64
liveness           float64
valence            float64
tempo              float64
time_signature       int64
dtype: object
```

CHECKING FOR NULL VALUES.

[125]:
```
data.isnull().sum()
```

[125]:
```
id                  0
name               71
popularity          0
duration_ms         0
explicit            0
artists             0
id_artists          0
release_date        0
danceability        0
energy              0
key                 0
loudness            0
mode                0
```

```
speechiness          0
acousticness         0
instrumentalness     0
liveness             0
valence              0
tempo                0
time_signature       0
dtype: int64
```

---

10 LEAST POPULAR SONGS

[126]:
```
data.head(2)
```

[126]:
```
                     id                                    name  popularity  \
0  35iwgR4jXetI318WEWsa1Q                                   Carve           6
1  021ht4sdgPcrDgSk7JTbKY  Capítulo 2.16 - Banquero Anarquista           0

   duration_ms  explicit             artists                 id_artists  \
0       126903         0             ['Uli']  ['45tIt06XoI0Iio4LBEVpls']
1        98200         0  ['Fernando Pessoa']  ['14jtPCOoNZwquk5wd9DxrY']

  release_date  danceability  energy  key  loudness  mode  speechiness  \
0   22-02-1922         0.645   0.445    0   -13.338     1        0.451
1   01-06-1922         0.695   0.263    0   -22.136     1        0.957

   acousticness  instrumentalness  liveness  valence    tempo  time_signature
0         0.674             0.744     0.151    0.127  104.851               3
1         0.797             0.000     0.148    0.655  102.009               1
```

[127]:
```
sorted = data.sort_values('popularity' , ascending = True).head(10)
sorted
```

[127]:
```
                     id                                    name  popularity  \
32  1HXdv1z9RlvrcUernyf0MY                The Dear Little Shamrock           0
78  3XxlglpLYkpZIfU1c1C5e4           Pobre Cotorro - Remasterizado           0
77  3ROtnqCtqrMbaDkivZyV5n             Entrerriana - Remasterizado           0
76  3QkuxlHi5RXa2YBr6S52n3   Capítulo 2.9 - Banquero Anarquista           0
75  3FfL0WRdyrEFPlPvmqj2lJ   Capítulo 1.9 - Banquero Anarquista           0
74  3FBXaR512ufrrickKCasRN   Capítulo 1.7 - Banquero Anarquista           0
73  38u3k7zUGTp48G1oA3ky3R                  The Girl That I Marry           0
72  2wWS5DZYc50AzbiXC57vGt  Capítulo 2.14 - Banquero Anarquista           0
71  2vSAIArCi0f9b9Ll612hFK   Capítulo 2.4 - Banquero Anarquista           0
70  2vOhKgXKFpej7mv3mOwoF4  Capítulo 1.21 - Banquero Anarquista           0

    duration_ms  explicit             artists                 id_artists  \
32       191613         0      ['Dick Haymes']  ['3BiJGZsyX9sJchTqcSA7Su']
```

78	159800	0	['Ignacio Corsini']	['5LiOoJbxVSAMkBS2fUm3X2']
77	183720	0	['Ignacio Corsini']	['5LiOoJbxVSAMkBS2fUm3X2']
76	99900	0	['Fernando Pessoa']	['14jtPCOoNZwquk5wd9DxrY']
75	106900	0	['Fernando Pessoa']	['14jtPCOoNZwquk5wd9DxrY']
74	106000	0	['Fernando Pessoa']	['14jtPCOoNZwquk5wd9DxrY']
73	182920	0	['Dick Haymes']	['3BiJGZsyX9sJchTqcSA7Su']
72	95200	0	['Fernando Pessoa']	['14jtPCOoNZwquk5wd9DxrY']
71	97600	0	['Fernando Pessoa']	['14jtPCOoNZwquk5wd9DxrY']
70	102000	0	['Fernando Pessoa']	['14jtPCOoNZwquk5wd9DxrY']

	release_date	danceability	energy	key	loudness	mode	speechiness
32	1922	0.347	0.197	10	-15.652	1	0.0504
78	21-03-1922	0.564	0.053	6	-25.041	0	0.1270
77	21-03-1922	0.431	0.200	5	-23.557	0	0.0322
76	01-06-1922	0.768	0.232	7	-21.678	0	0.9620
75	01-06-1922	0.619	0.173	0	-23.130	1	0.9670
74	01-06-1922	0.735	0.244	10	-22.252	1	0.9560
73	1922	0.185	0.151	7	-17.822	1	0.0391
72	01-06-1922	0.827	0.258	5	-21.896	1	0.9560
71	01-06-1922	0.732	0.176	0	-21.808	1	0.9660
70	01-06-1922	0.625	0.187	10	-24.686	1	0.9630

	acousticness	instrumentalness	liveness	valence	tempo	time_signature
32	0.991	0.000133	0.196	0.246	109.338	3
78	0.996	0.927000	0.117	0.407	140.216	4
77	0.995	0.821000	0.139	0.543	108.650	3
76	0.593	0.000000	0.304	0.591	64.433	4
75	0.730	0.000000	0.145	0.653	181.485	4
74	0.666	0.000000	0.253	0.618	85.079	4
73	0.985	0.798000	0.151	0.135	88.945	3
72	0.658	0.000000	0.316	0.730	41.694	3
71	0.623	0.000000	0.442	0.771	76.708	1
70	0.831	0.000000	0.424	0.576	76.537	3

Descriptive statistics for the numerical variables present in the column.

[128]:
```
data.describe().transpose()
```

[128]:

	count	mean	std	min	25%	\
popularity	586672.0	27.570053	18.370642	0.0	13.0000	
duration_ms	586672.0	230051.167286	126526.087418	3344.0	175093.0000	
explicit	586672.0	0.044086	0.205286	0.0	0.0000	
danceability	586672.0	0.563594	0.166103	0.0	0.4530	
energy	586672.0	0.542036	0.251923	0.0	0.3430	
key	586672.0	5.221603	3.519423	0.0	2.0000	
loudness	586672.0	-10.206067	5.089328	-60.0	-12.8910	
mode	586672.0	0.658797	0.474114	0.0	0.0000	

```
speechiness       586672.0       0.104864       0.179893     0.0       0.0340
acousticness      586672.0       0.449863       0.348837     0.0       0.0969
instrumentalness  586672.0       0.113451       0.266868     0.0       0.0000
liveness          586672.0       0.213935       0.184326     0.0       0.0983
valence           586672.0       0.552292       0.257671     0.0       0.3460
tempo             586672.0     118.464857      29.764108     0.0      95.6000
time_signature    586672.0       3.873382       0.473162     0.0       4.0000

                           50%           75%          max
popularity            27.000000      41.00000      100.000
duration_ms       214893.000000  263867.00000  5621218.000
explicit               0.000000       0.00000        1.000
danceability           0.577000       0.68600        0.991
energy                 0.549000       0.74800        1.000
key                    5.000000       8.00000       11.000
loudness              -9.243000      -6.48200        5.376
mode                   1.000000       1.00000        1.000
speechiness            0.044300       0.07630        0.971
acousticness           0.422000       0.78500        0.996
instrumentalness       0.000024       0.00955        1.000
liveness               0.139000       0.27800        1.000
valence                0.564000       0.76900        1.000
tempo                117.384000     136.32100      246.381
time_signature         4.000000       4.00000        5.000
```

---

10 POPULAR SONGS WHERE POPULARITY IS ABOVE 90

[129]:
```
data.head(2)
```

[129]:
```
                     id                                name  popularity  \
0  35iwgR4jXetI318WEWsa1Q                               Carve           6
1  021ht4sdgPcrDgSk7JTbKY  Capítulo 2.16 - Banquero Anarquista           0

   duration_ms  explicit              artists                 id_artists  \
0       126903         0              ['Uli']  ['45tIt06XoI0Iio4LBEVpls']
1        98200         0  ['Fernando Pessoa']  ['14jtPCOoNZwquk5wd9DxrY']

  release_date  danceability  energy  key  loudness  mode  speechiness  \
0   22-02-1922         0.645   0.445    0   -13.338     1        0.451
1   01-06-1922         0.695   0.263    0   -22.136     1        0.957

   acousticness  instrumentalness  liveness  valence    tempo  time_signature
0         0.674             0.744     0.151    0.127  104.851               3
1         0.797             0.000     0.148    0.655  102.009               1
```

[130]:
```
data[data['popularity'] > 90]
```

[130]:

	id	name	popularity
91866	60ynsPSSKe6O3sfwRnIBRf	Streets	94
91867	6UelLqGlWMcVH1E5c4H7lY	Watermelon Sugar	92
92810	5QO79kh1waicV47BqGRL3g	Save Your Tears	97
92811	6tDDoYIxWvMLTdKpjFkc1B	telepatía	97
92813	0VjIjW4GlUZAMYd2vXMi3b	Blinding Lights	96
92814	6f3Slt0GbA2bPZlz0aIFXN	The Business	95
92816	3FAJ6O0NOHQV8Mc5Ri6ENp	Heartbreak Anniversary	94
92817	27OeeYzk6klgBh83TSvGMA		
92819	1xK1Gg9SxG8fy2Ya373oqb		
92821	2XIc1pqjXV3Cr2BQUGNBck		
92823	3YJJjQPAbDT7mGpX3WtQ9A		
92824	47EiUVwUp4C9fGccaPuUCS		
92826	35mvY5S1H3J2QZyna3TFe0		
92827	3VvA1wSxukMLsvXoXtlwWx		
92829	6ft4hAq6yde8jPZY2i5zLr		
92830	4saklk6nie3yiGePpBwUoc		
92832	45bE4HXI0AwGZXfZtMp8JR		
92833	7vrJn5hDSXRmdXoR30KgF1		
92835	6Im9k8u9iIzKMrmV7BWtlF		
92837	7ytR5pFWmSjzHJIeQkgog4		
92839	6cx06DFPPHchuUAcTxznu9		
92841	1tkg4EHVoqnhR6iFEXb60y		
93802	4iJyoBOLtHqaGxP12qzhQI		
93803	7lPN2DXiMsVn7XUKtOW1CS		
93804	3Ofmpyhv5UAQ70mENzB277		
93805	7MAibcTli4IisCtbHKrGMh		
93806	7Bk0uXKk1uPT0XuQbpFzvs		
93807	4cG7HUWYHBV6R6tHn1gxrl		
93808	5YYW3yRktprLRr47WK219Y		
93809	1XXimziG1uhM0eDNCZCrUl		
93810	5uEYRdEIh9Bo4fpjDd4Na9		
93811	4u4NyuceXP7Uzh7XFJKCr1		
93812	3iw6V4LH7yPj1ESORX9RIN		
93813	2TksvaRivgAEj780DgRB73		
93814	5YaskwnGDZFDRipaqzbwQx		
93815	3aQem4jVGdhtg116TmJnHz		
93816	1diS6nkxMQc3wwC4G1j0bh		

```
92817                                        WITHOUT YOU          94
92819                                            Bandido          94
92821                                  LA NOCHE DE ANOCHE          93
92823                                           Good Days          93
92824                                              DÁKITI          92
92826                                           positions          92
92827                                       Hecha Pa' Mi           92
92829                      Paradise (feat. Dermot Kennedy)          92
92830                                            Dynamite          91
92832                                  you broke me first          91
92833                                            BICHOTA          91
92835                                              34+35          91
92837                         ROCKSTAR (feat. Roddy Ricch)          91
92839                            Head & Heart (feat. MNEK)          91
92841                              What You Know Bout Love          91
93802              Peaches (feat. Daniel Caesar & Giveon)         100
93803                                       drivers license          99
93804                               Astronaut In The Ocean          98
93805                                  Leave The Door Open          96
93806                                                Fiel          94
93807  Friday (feat. Mufasa & Hypeman) - Dopamine Re-…         94
93808                               Ella No Es Tuya - Remix          92
93809                                                  Up          92
93810                                  Goosebumps - Remix          92
93811                                            Hold On          92
93812  Wellerman - Sea Shanty / 220 KID x Billen Ted …         92
93813                                                911          91
93814                                   Your Love (9PM)          91
93815                                       What’s Next          91
93816                                        We're Good          91

       duration_ms  explicit  \
91866       226987         1
91867       174000         0
92810       215627         1
92811       160191         0
92813       200040         0
92814       164000         0
92816       198371         0
92817       161385         1
92819       232853         0
92821       203201         0
92823       279204         1
92824       205090         1
92826       172325         1
92827       186133         0
92829       167903         0
```

```
92830       199054         0
92832       169266         0
92833       178947         1
92835       173711         1
92837       181733         1
92839       166028         0
92841       160000         1
93802       198082         1
93803       242014         1
93804       132780         0
93805       242096         0
93806       261667         0
93807       169153         0
93808       218107         0
93809       156945         1
93810       162803         1
93811       170813         0
93812       116750         0
93813       215891         1
93814       150053         0
93815       178154         1
93816       165507         0
```

```
                                                artists  \
91866                                      ['Doja Cat']
91867                                    ['Harry Styles']
92810                                      ['The Weeknd']
92811                                      ['Kali Uchis']
92813                                      ['The Weeknd']
92814                                          ['Tiësto']
92816                                          ['Giveon']
92817                                   ['The Kid LAROI']
92819                            ['Myke Towers', 'Juhn']
92821                           ['Bad Bunny', 'ROSALÍA']
92823                                             ['SZA']
92824                          ['Bad Bunny', 'Jhay Cortez']
92826                                   ['Ariana Grande']
92827                                            ['Boza']
92829                          ['MEDUZA', 'Dermot Kennedy']
92830                                             ['BTS']
92832                                      ['Tate McRae']
92833                                         ['KAROL G']
92835                                   ['Ariana Grande']
92837                              ['DaBaby', 'Roddy Ricch']
92839                                 ['Joel Corry', 'MNEK']
92841                                           ['Pop Smoke']
93802       ['Justin Bieber', 'Daniel Caesar', 'Giveon']
```

```
93803                                ['Olivia Rodrigo']
93804                                   ['Masked Wolf']
93805    ['Bruno Mars', 'Anderson .Paak', 'Silk Sonic']
93806       ['Los Legendarios', 'Wisin', 'Jhay Cortez']
93807  ['Riton', 'Nightcrawlers', 'Mufasa & Hypeman',…
93808       ['Rochy RD', 'Myke Towers', 'Nicki Nicole']
93809                                        ['Cardi B']
93810                           ['Travis Scott', 'HVME']
93811                                 ['Justin Bieber']
93812         ['Nathan Evans', '220 KID', 'Billen Ted']
93813                                          ['Sech']
93814                           ['ATB', 'Topic', 'A7S']
93815                                       ['Drake']
93816                                    ['Dua Lipa']

                                          id_artists release_date  \
91866                        ['5cj0lLjcoR7YOSnhnX0Po5']   07-11-2019
91867                        ['6KImCVD70vtIoJWnq6nGn3']   13-12-2019
92810                        ['1Xyo4u8uXC1ZmMpatF05PJ']   20-03-2020
92811                        ['1U1el3k54VvEUzo3ybLPlM']   04-12-2020
92813                        ['1Xyo4u8uXC1ZmMpatF05PJ']   20-03-2020
92814                        ['2o5jDhtHVPhrJdv3cEQ99Z']   16-09-2020
92816                        ['4fxd5Ee7UefO4CUXgwJ7IP']   27-03-2020
92817                        ['2tIP7SsRs7vjIcLrU85W8J']   06-11-2020
92819  ['7iK8PXO48WeuP03g8YR51W', '2LmcxBak1alK1bf7d1…   10-12-2020
92821  ['4q3ewBCX7sLwd24euuV69X', '7ltDVBr6mKbRvohxhe…   27-11-2020
92823                        ['7tYKF4w9nC0nq9CsPZTHyP']   25-12-2020
92824  ['4q3ewBCX7sLwd24euuV69X', '0EFisYRi20PTADoJri…   30-10-2020
92826                        ['66CXWjxzNUsdJxJ2JdwvnR']   30-10-2020
92827                        ['2NfSBtmWe7oPw1EmetJVso']   31-07-2020
92829  ['0xRXCcSX89eobfrshSVdyu', '5KNNVgR6LBIABRIomy…   30-10-2020
92830                        ['3NrfpeOtUJi4K4DXYWgMUX']   20-11-2020
92832                        ['45dkTj5sMRSjrmBSBeiHym']   17-04-2020
92833                        ['790FomKkXshlbRYZFtlgla']   23-10-2020
92835                        ['66CXWjxzNUsdJxJ2JdwvnR']   30-10-2020
92837  ['4r63FhuTkUYltbVAg5TQnk', '757aE44tKEUQEqRuT6…   17-04-2020
92839  ['6DgP9otnZw5z6daOntINxp', '7uMh23xWiuR7zsNkuN…   03-07-2020
92841                        ['0eDvMgVFoNV3TpwtrVCoTj']   03-07-2020
93802  ['1uNFoZAHBGtllmzznpCI3s', '20wkVLutqVOYrc0kxF…   19-03-2021
93803                        ['1McMsnEElThX1knmY4oliG']   08-01-2021
93804                        ['1uU7g3DNSbsu0QjSEqZtEd']   06-01-2021
93805  ['0du5cEVh5yTK9QJze8zA0C', '3jK9MiCrA42lLAdMGU…   05-03-2021
93806  ['0n6sKrG0xKAf8xmdqeNGke', '3E6xrwgnVfYCrCs0eP…   04-02-2021
93807  ['7i9j813KFoSBMldGqlh2Z1', '1gALaWbNDnwS2ECV09…   15-01-2021
93808  ['4riOEaOW5hCeqomFDBk0aP', '7iK8PXO48WeuP03g8Y…   03-02-2021
93809                        ['4kYSro6naA4h99UJvo89HB']   05-02-2021
93810  ['0Y5tJX1MQlPlqiwlOH1tJY', '2o08sCWF5yyo2G4DCi…   15-01-2021
```

```
93811                          ['1uNFoZAHBGtllmzznpCI3s']   05-03-2021
93812  ['1PKErrAhYFdfrDymGHRQRo', '4Euia7UzdRshy1DJOS…   21-01-2021
93813                          ['77ziqFxp5gaInVrF2lj4ht']   05-02-2021
93814  ['7jZM5w05mGhw6wTB1okhD9', '0u6GtibW46tFX7koQ6…   15-01-2021
93815                          ['3TVXtAsR1Inumwj472S9r4']   05-03-2021
93816                          ['6M2wZ9GZgrQXHCFfjv46we']   11-02-2021

       danceability  energy  key  loudness  mode  speechiness  acousticness  \
91866         0.749   0.463   11    -8.433     1       0.0828       0.20800
91867         0.548   0.816    0    -4.209     1       0.0465       0.12200
92810         0.680   0.826    0    -5.487     1       0.0309       0.02120
92811         0.653   0.524   11    -9.016     0       0.0502       0.11200
92813         0.514   0.730    1    -5.934     1       0.0598       0.00146
92814         0.798   0.620    8    -7.079     0       0.2320       0.41400
92816         0.449   0.465    0    -8.964     1       0.0791       0.52400
92817         0.662   0.413    0    -7.357     1       0.0299       0.21300
92819         0.713   0.617    8    -4.637     1       0.0887       0.12200
92821         0.856   0.618    7    -4.892     1       0.2860       0.03030
92823         0.436   0.655    1    -8.370     0       0.0583       0.49900
92824         0.731   0.573    4   -10.059     0       0.0544       0.40100
92826         0.737   0.802    0    -4.771     1       0.0878       0.46800
92827         0.725   0.756    4    -5.013     1       0.0572       0.36200
92829         0.632   0.595    8    -7.644     0       0.0401       0.06890
92830         0.746   0.765    6    -4.410     0       0.0993       0.01120
92832         0.667   0.373    4    -9.389     1       0.0500       0.78500
92833         0.863   0.666    1    -4.158     0       0.1520       0.21200
92835         0.830   0.585    0    -6.476     1       0.0940       0.23700
92837         0.746   0.690   11    -7.956     1       0.1640       0.24700
92839         0.734   0.874    8    -3.158     1       0.0662       0.16800
92841         0.709   0.548   10    -8.493     1       0.3530       0.65000
93802         0.677   0.696    0    -6.181     1       0.1190       0.32100
93803         0.585   0.436   10    -8.761     1       0.0601       0.72100
93804         0.778   0.695    4    -6.865     0       0.0913       0.17500
93805         0.586   0.616    5    -7.964     1       0.0324       0.18200
93806         0.849   0.701    6    -4.407     0       0.0600       0.40700
93807         0.824   0.862    2    -3.424     1       0.1260       0.00760
93808         0.585   0.611    6    -4.132     1       0.2880       0.39000
93809         0.868   0.795   11    -6.044     1       0.2690       0.00120
93810         0.841   0.593    1    -7.846     1       0.0379       0.41800
93811         0.658   0.634    1    -5.797     0       0.0413       0.01060
93812         0.722   0.893    0    -3.255     0       0.0475       0.04410
93813         0.795   0.655    7    -3.815     1       0.0422       0.07600
93814         0.669   0.784    7    -5.603     1       0.1120       0.19400
93815         0.781   0.594    0    -6.959     0       0.0485       0.01360
93816         0.722   0.588    6    -5.932     1       0.0544       0.03190

       instrumentalness  liveness  valence    tempo  time_signature
```

91866	0.037100	0.3370	0.1900	90.028	4
91867	0.000000	0.3350	0.5570	95.390	4
92810	0.000012	0.5430	0.6440	118.051	4
92811	0.000000	0.2030	0.5530	83.970	4
92813	0.000095	0.0897	0.3340	171.005	4
92814	0.019200	0.1120	0.2350	120.031	4
92816	0.000001	0.3030	0.5430	89.087	3
92817	0.000000	0.1340	0.4670	93.005	4
92819	0.000000	0.0962	0.6820	168.021	4
92821	0.000000	0.0866	0.3910	81.993	4
92823	0.000008	0.6880	0.4120	121.002	4
92824	0.000052	0.1130	0.1450	109.928	4
92826	0.000000	0.0931	0.6820	144.015	4
92827	0.000685	0.1030	0.8280	100.070	4
92829	0.000000	0.2090	0.4350	124.114	4
92830	0.000000	0.0936	0.7370	114.044	4
92832	0.000000	0.0906	0.0823	124.148	4
92833	0.000493	0.1030	0.8380	163.908	4
92835	0.000000	0.2480	0.4850	109.978	4
92837	0.000000	0.1010	0.4970	89.977	4
92839	0.000011	0.0489	0.9050	122.953	4
92841	0.000002	0.1330	0.5430	83.995	4
93802	0.000000	0.4200	0.4640	90.030	4
93803	0.000013	0.1050	0.1320	143.874	4
93804	0.000000	0.1500	0.4720	149.996	4
93805	0.000000	0.0927	0.7190	148.088	4
93806	0.000000	0.1120	0.5050	98.005	4
93807	0.000132	0.3030	0.8010	122.980	4
93808	0.000000	0.6490	0.9620	97.041	4
93809	0.019300	0.0461	0.8190	166.000	4
93810	0.000000	0.1240	0.8080	124.917	4
93811	0.000000	0.1320	0.2900	139.980	4
93812	0.000937	0.0673	0.4390	119.932	4
93813	0.000002	0.2960	0.7440	93.029	4
93814	0.000006	0.1150	0.5170	125.993	4
93815	0.000000	0.1620	0.0628	129.895	4
93816	0.000000	0.1830	0.5900	134.010	4

[131]:
```
most_popular = data.query('popularity>90', inplace = False).
 ↪sort_values('popularity', ascending = False)
most_popular[:10]
```

[131]:

	id	name	\
93802	4iJyoBOLtHqaGxP12qzhQI	Peaches (feat. Daniel Caesar & Giveon)	
93803	7lPN2DXiMsVn7XUKtOW1CS	drivers license	
93804	3Ofmpyhv5UAQ70mENzB277	Astronaut In The Ocean	
92811	6tDDoYIxWvMLTdKpjFkc1B	telepatía	

```
92810  5QO79kh1waicV47BqGRL3g                        Save Your Tears
92813  0VjIjW4GlUZAMYd2vXMi3b                        Blinding Lights
93805  7MAibcTli4IisCtbHKrGMh                    Leave The Door Open
92814  6f3Slt0GbA2bPZlz0aIFXN                           The Business
91866  60ynsPSSKe6O3sfwRnIBRf                                Streets
93806  7Bk0uXKk1uPT0XuQbpFzvs                                   Fiel

       popularity  duration_ms  explicit  \
93802         100       198082         1
93803          99       242014         1
93804          98       132780         0
92811          97       160191         0
92810          97       215627         1
92813          96       200040         0
93805          96       242096         0
92814          95       164000         0
91866          94       226987         1
93806          94       261667         0

                                            artists  \
93802    ['Justin Bieber', 'Daniel Caesar', 'Giveon']
93803                              ['Olivia Rodrigo']
93804                                 ['Masked Wolf']
92811                                  ['Kali Uchis']
92810                                  ['The Weeknd']
92813                                  ['The Weeknd']
93805  ['Bruno Mars', 'Anderson .Paak', 'Silk Sonic']
92814                                      ['Tiësto']
91866                                    ['Doja Cat']
93806     ['Los Legendarios', 'Wisin', 'Jhay Cortez']

                                          id_artists release_date  \
93802  ['1uNFoZAHBGtllmzznpCI3s', '20wkVLutqVOYrc0kxF…   19-03-2021
93803                          ['1McMsnEElThX1knmY4oliG']   08-01-2021
93804                          ['1uU7g3DNSbsu0QjSEqZtEd']   06-01-2021
92811                          ['1U1el3k54VvEUzo3ybLPlM']   04-12-2020
92810                          ['1Xyo4u8uXC1ZmMpatF05PJ']   20-03-2020
92813                          ['1Xyo4u8uXC1ZmMpatF05PJ']   20-03-2020
93805  ['0du5cEVh5yTK9QJze8zA0C', '3jK9MiCrA42lLAdMGU…   05-03-2021
92814                          ['2o5jDhtHVPhrJdv3cEQ99Z']   16-09-2020
91866                          ['5cj0lLjcoR7YOSnhnX0Po5']   07-11-2019
93806  ['0n6sKrG0xKAf8xmdqeNGke', '3E6xrwgnVfYCrCs0eP…   04-02-2021

       danceability  energy  key  loudness  mode  speechiness  acousticness  \
93802         0.677   0.696    0    -6.181     1       0.1190       0.32100
93803         0.585   0.436   10    -8.761     1       0.0601       0.72100
93804         0.778   0.695    4    -6.865     0       0.0913       0.17500
```

```
92811         0.653   0.524   11    -9.016     0       0.0502       0.11200
92810         0.680   0.826    0    -5.487     1       0.0309       0.02120
92813         0.514   0.730    1    -5.934     1       0.0598       0.00146
93805         0.586   0.616    5    -7.964     1       0.0324       0.18200
92814         0.798   0.620    8    -7.079     0       0.2320       0.41400
91866         0.749   0.463   11    -8.433     1       0.0828       0.20800
93806         0.849   0.701    6    -4.407     0       0.0600       0.40700

       instrumentalness  liveness  valence    tempo  time_signature
93802          0.000000    0.4200    0.464   90.030               4
93803          0.000013    0.1050    0.132  143.874               4
93804          0.000000    0.1500    0.472  149.996               4
92811          0.000000    0.2030    0.553   83.970               4
92810          0.000012    0.5430    0.644  118.051               4
92813          0.000095    0.0897    0.334  171.005               4
93805          0.000000    0.0927    0.719  148.088               4
92814          0.019200    0.1120    0.235  120.031               4
91866          0.037100    0.3370    0.190   90.028               4
93806          0.000000    0.1120    0.505   98.005               4
```

CHECK ARTIST AT THE LOCATION 18.

[132]:
```
data[["artists"]].iloc[18]

#Use data frame + double square bracket and column name
# #using index location method
```

[132]:
```
artists    ['Victor Boucher']
Name: 18, dtype: object
```

CONVERT 'DURATION' FROM MILLISECONDS TO SECONDS

[133]:
```
data.head(2)
```

[133]:
```
                     id                                name  popularity  \
0  35iwgR4jXetI318WEWsa1Q                               Carve           6
1  021ht4sdgPcrDgSk7JTbKY  Capítulo 2.16 - Banquero Anarquista           0

   duration_ms  explicit             artists                 id_artists  \
0       126903         0             ['Uli']  ['45tIt06XoI0Iio4LBEVpls']
1        98200         0  ['Fernando Pessoa']  ['14jtPCOoNZwquk5wd9DxrY']

  release_date  danceability  energy  key  loudness  mode  speechiness  \
0   22-02-1922         0.645   0.445    0   -13.338     1        0.451
1   01-06-1922         0.695   0.263    0   -22.136     1        0.957
```

```
   acousticness  instrumentalness  liveness  valence    tempo  time_signature
0         0.674             0.744     0.151    0.127  104.851               3
1         0.797             0.000     0.148    0.655  102.009               1
```

[134]:
```
data['duration'] = data['duration_ms'].apply(lambda x: round(x/1000))
data.drop('duration_ms',inplace=True,axis=1)

 #Using apply() function
 # x/1000 converts ms to sec
 # round() converts the value to nearest whole number
```

[135]:
```
data.duration.head()
```

[135]:
```
0    127
1     98
2    182
3    177
4    163
Name: duration, dtype: int64
```

SETTING INDEX AS RELEASE DATE

[147]:
```
# Convert the index to datetime, specifying the correct format
data.index = pd.to_datetime(data.index, format='%d-%m-%Y', errors='coerce')

# Display the first 2 rows of the DataFrame
data.head(2)
```

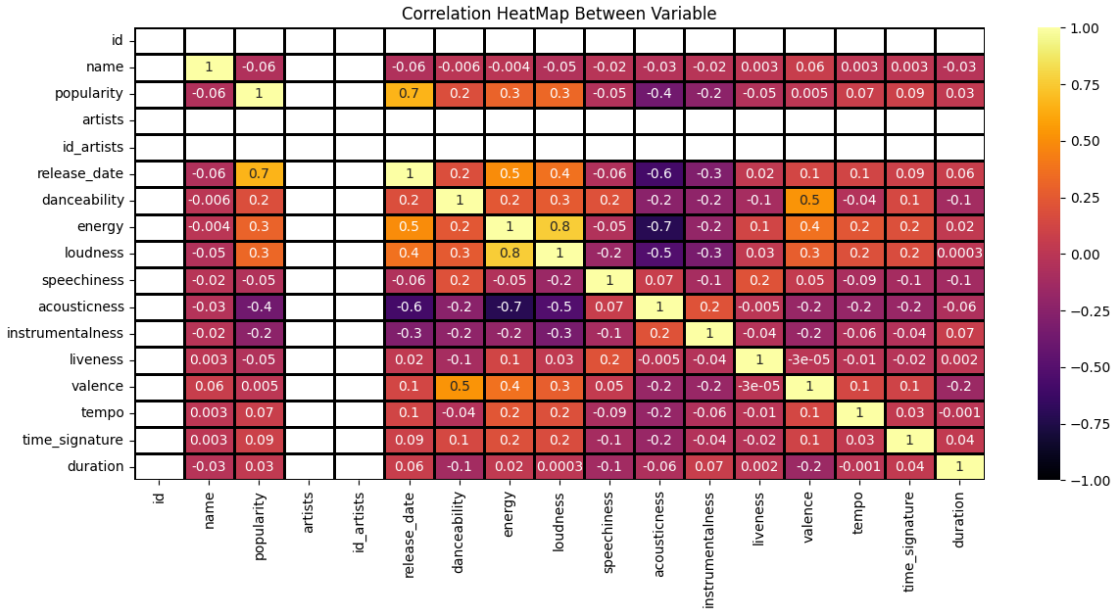
[147]:
```
                                id                                    name  \
release_date
1922-02-22    35iwgR4jXetI318WEWsa1Q                                   Carve
1922-06-01    021ht4sdgPcrDgSk7JTbKY  Capítulo 2.16 - Banquero Anarquista

              popularity  explicit              artists  \
release_date
1922-02-22             6         0              ['Uli']
1922-06-01             0         0  ['Fernando Pessoa']

                           id_artists  danceability  energy  key  loudness  \
release_date
1922-02-22    ['45tIt06XoI0Iio4LBEVpls']         0.645   0.445    0   -13.338
1922-06-01    ['14jtPCOoNZwquk5wd9DxrY']         0.695   0.263    0   -22.136

              mode  speechiness  acousticness  instrumentalness  liveness  \
release_date
```

```
1922-02-22       1         0.451          0.674              0.744      0.151
1922-06-01       1         0.957          0.797              0.000      0.148

              valence    tempo  time_signature  duration
release_date
1922-02-22      0.127  104.851               3       127
1922-06-01      0.655  102.009               1        98
```

CREATING CORRELATION MAP.

[ ]:
```
# unwanted columns are dropped
# "pearson" correlation method is applied
# set figure size to 14,6
# using seaborn heatmap function to create correlation map
# setting annotation = True, this will write the data value in each cell.
# set fmt = .1g (this is a string formating tool to use when adding annotation)
# vmin and vmax are the values to add for color map
# cmap - colormap
```

[ ]:
```
# Drop the specified columns
data_subset = data.drop(["key", "mode", "explicit"], axis=1)

# Convert the entire DataFrame to numeric
data_numeric = data_subset.apply(pd.to_numeric, errors='coerce')

# Calculate correlation
correl = data_numeric.corr(method='pearson')

# Now, proceed with the visualization
plt.figure(figsize=(14, 6))
heatmap = sns.heatmap(correl, annot=True, fmt=".1g", vmin=-1, vmax=1, center=0,␣
 ↪cmap='inferno', linewidths=1, linecolor="Black")
heatmap = plt.title("Correlation HeatMap Between Variable")
```



SAMPLING DATA IN JUST 0.4% OF TOTAL DATA & WILL CREATE 2 REGRESSION PLOT USING THE DATA

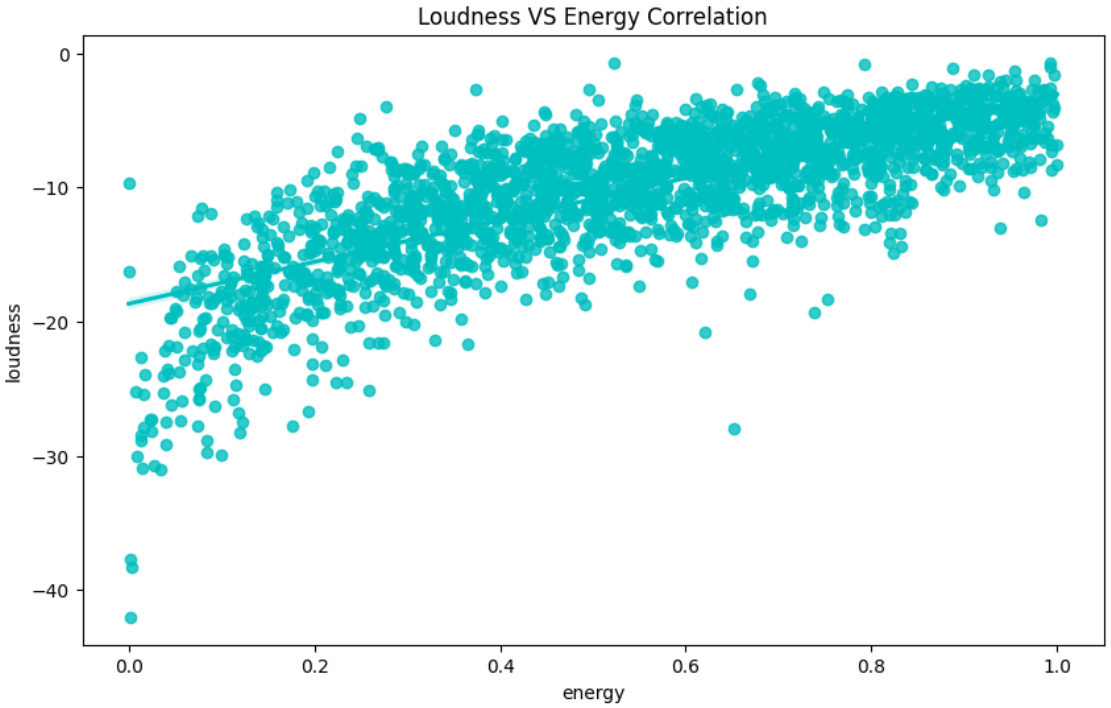
[ ]:
```
sample_data = data.sample(int(0.004*len(data)))

# use int function.
# 0.004 multiplied by len of original data frame
```

[ ]:
```
print(len(sample_data))
```

2346

PLOTTING A REGRESSION PLOT BETWEEN LOUDNESS AND ENERGY

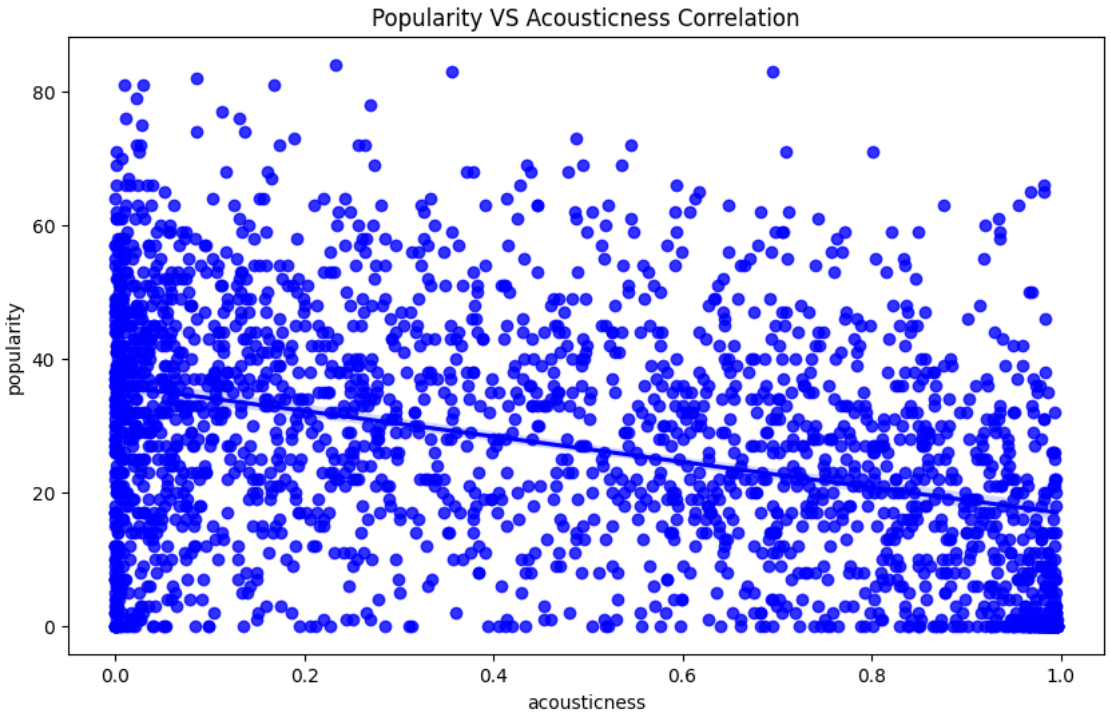
[ ]:
```
plt.figure(figsize=(10,6))
sns.regplot(data=sample_data, y = "loudness", x = "energy", color ="c")
plt.title("Loudness VS Energy Correlation")
plt.show()
```



The regrression line is showing upside - When Energy increases Loudness of the song increases and vise-versa.

PLOTTING A REGRESSION PLOT BETWEEN POPULARITY AND ACCOUSTICNESS.

[ ]:
```
plt.figure(figsize=(10,6))
sns.regplot(data = sample_data, y = 'popularity', x = 'acousticness', color =␣
↪'b')
plt.title('Popularity VS Acousticness Correlation')
plt.show()
```



The regression line is showing downside - If Accousticness of the song increases the Popularity of the song decreases and vice-versa.

CREATE COLUMN YEARS FROM RELEASE DATE COLUMN

[148]:
```
data.head(2)
```

[148]:
```
                                id                                    name  \
release_date
1922-02-22    35iwgR4jXetI318WEWsa1Q                                   Carve
1922-06-01    021ht4sdgPcrDgSk7JTbKY  Capítulo 2.16 - Banquero Anarquista

              popularity  explicit              artists  \
release_date
1922-02-22             6         0              ['Uli']
1922-06-01             0         0  ['Fernando Pessoa']

                           id_artists  danceability  energy  key  loudness  \
release_date
1922-02-22    ['45tIt06XoI0Iio4LBEVpls']         0.645   0.445    0   -13.338
1922-06-01    ['14jtPCOoNZwquk5wd9DxrY']         0.695   0.263    0   -22.136

              mode  speechiness  acousticness  instrumentalness  liveness  \
```

```
release_date
1922-02-22       1         0.451          0.674              0.744      0.151
1922-06-01       1         0.957          0.797              0.000      0.148

               valence    tempo  time_signature  duration
release_date
1922-02-22       0.127  104.851               3       127
1922-06-01       0.655  102.009               1        98
```

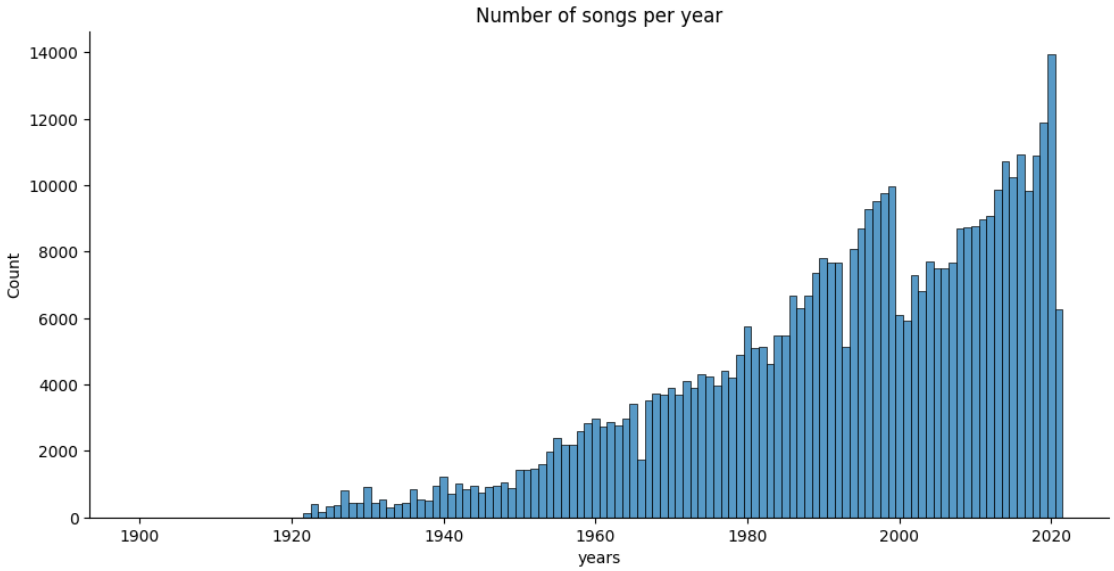
CREATE COLUMN YEARS FROM RELEASE DATE

[153]:
```
data['years'] = data.index.year

years = data['years']

# Create a new column 'years' from the index (release_date)
```

DISTRIBUTION PLOT "NUMBER OF SONGS" SINCE 1922

[154]:
```
sns.displot(years, discrete=True, aspect=2, height=5,kind='hist').
↪set(title="Number of songs per year")
```

[154]: <seaborn.axisgrid.FacetGrid at 0x1f905c30440>



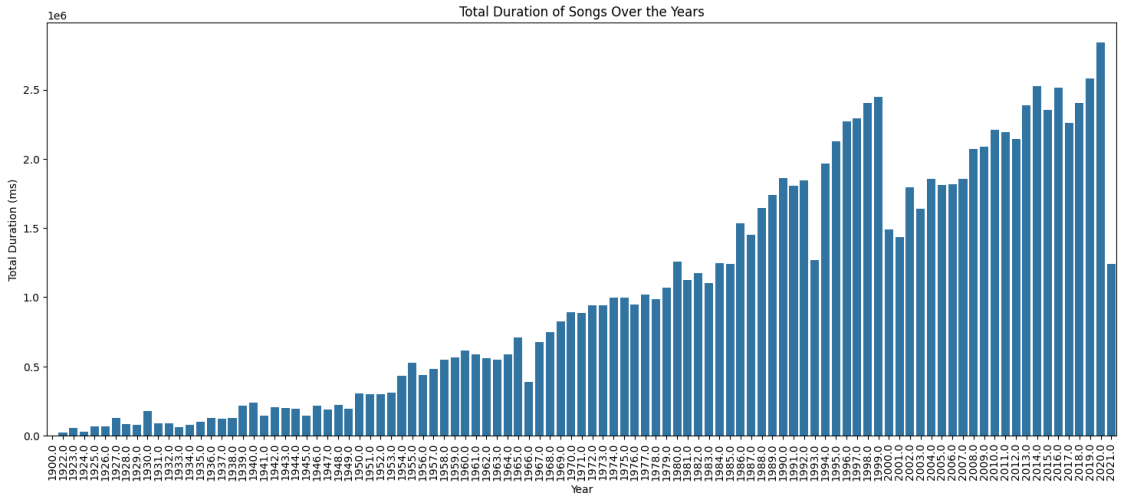
DURATION OF SONGS OVER THE YEARS (BARPLOT)

[ ]:
```
data.head(2)
```

[ ]:
```
                     id                                  name  popularity  \
0  35iwgR4jXetI318WEWsa1Q                                 Carve           6
1  021ht4sdgPcrDgSk7JTbKY  Capítulo 2.16 - Banquero Anarquista           0

   explicit              artists                 id_artists release_date  \
0         0              ['Uli']  ['45tIt06XoI0Iio4LBEVpls']   22-02-1922
1         0  ['Fernando Pessoa']  ['14jtPCOoNZwquk5wd9DxrY']   01-06-1922

   danceability  energy  key  loudness  mode  speechiness  acousticness  \
0         0.645   0.445    0   -13.338     1        0.451         0.674
1         0.695   0.263    0   -22.136     1        0.957         0.797

   instrumentalness  liveness  valence    tempo  time_signature  duration
0             0.744     0.151    0.127  104.851               3       127
1             0.000     0.148    0.655  102.009               1        98
```

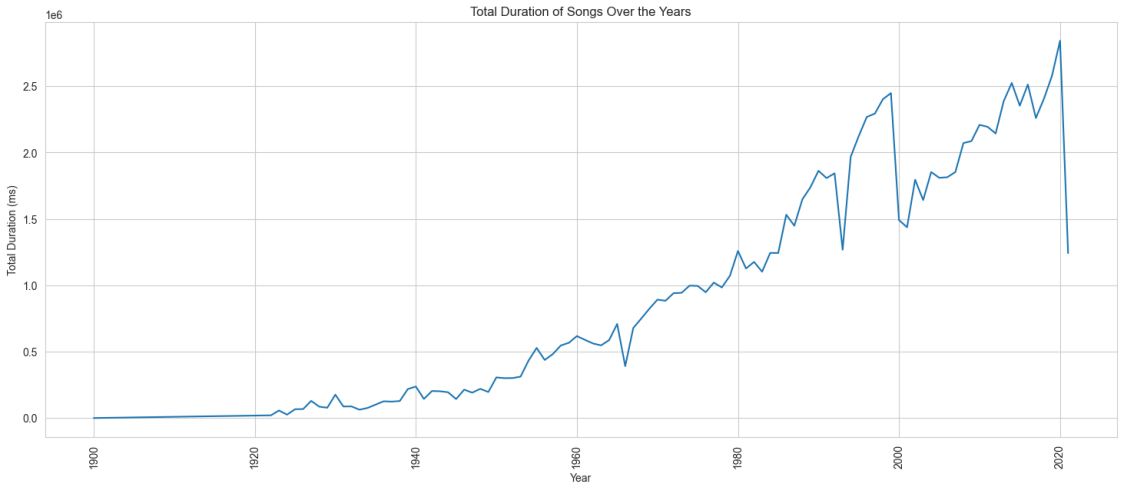
[164]:
```
# Group by 'years' and sum the duration for each year
yearly_total_duration = data.groupby('years')['duration'].sum().reset_index()

# Plot the total duration of songs over the years
fig_dims = (18, 7)  # Dimensions
fig, ax = plt.subplots(figsize=fig_dims)
sns.barplot(x='years', y='duration', data=yearly_total_duration, ax=ax,␣
 ↪errwidth=0)
plt.title("Total Duration of Songs Over the Years")
plt.xlabel("Year")
plt.ylabel("Total Duration (ms)")
plt.xticks(rotation=90)

# Show the plot
plt.show()
```

```
C:\Users\Ebad\AppData\Local\Temp\ipykernel_10724\2684999600.py:7: FutureWarning:

The `errwidth` parameter is deprecated. And will be removed in v0.15.0. Pass
`err_kws={'linewidth': 0}` instead.

  sns.barplot(x='years', y='duration', data=yearly_total_duration, ax=ax,
errwidth=0)
```



DURATION OF SONGS OVER THE YEARS (LINEPLOT)

[171]:
```
# Group by 'years' and calculate the total duration for each year
yearly_total_duration = data.groupby('years')['duration'].sum().reset_index()

# Plot the total duration of songs over the years
fig_dims = (18, 7)  # Dimensions
fig, ax = plt.subplots(figsize=fig_dims)

# Create the line plot with the grouped data
sns.lineplot(x='years', y='duration', data=yearly_total_duration, ax=ax)
plt.title("Total Duration of Songs Over the Years")
plt.xlabel("Year")
plt.ylabel("Total Duration (ms)")
plt.xticks(rotation=90)

# Show the plot
plt.show()
```


Total Duration of Songs Over the Years
1e6
2.5
2.0
1.5
1.0
0.5
0.0
Total Duration (ms)
1900
1920
1940
1960
1980
2000
2020
Year