In [3]:
```
import numpy as np
import pandas as pd
import seaborn as sns
from matplotlib import pyplot as plt
```

In [4]:
```
spotify = pd.read_csv("C:\\Users\\SUBHAJIT\\Desktop\\My Document\\Study Document\\Projects\\Spotify Data Analysis\\tracks.csv")
```

In [5]:
```
spotify
```

Out[5]:

	id	name	popularity	duration_ms	explicit	artists	id_artists	release_date	danceability	energy	key	loudness	mode	speechiness	a
0	35iwgR4jXetI318WEWsa1Q	Carve	6	126903	0	['Uli']	['45tlt06XoI0Iio4LBEVpls']	1922-02-22	0.645	0.4450	0	-13.338	1	0.4510	
1	021ht4sdgPcrDgSk7JTbKY	Capítulo 2.16 - Banquero Anarquista	0	98200	0	['Fernando Pessoa']	['14jtPCOoNZwquk5wd9DxrY']	1922-06-01	0.695	0.2630	0	-22.136	1	0.9570	
2	07A5yehtSnoedViJAZkNnc	Vivo para Quererte - Remasterizado	0	181640	0	['Ignacio Corsini']	['5LiOoJbxVSAMkBS2fUm3X2']	1922-03-21	0.434	0.1770	1	-21.180	1	0.0512	
3	08FmqUhxtyLTn6pAh6bk45	El Prisionero - Remasterizado	0	176907	0	['Ignacio Corsini']	['5LiOoJbxVSAMkBS2fUm3X2']	1922-03-21	0.321	0.0946	7	-27.961	1	0.0504	
4	08y9GfoqCWfOGsKdwojr5e	Lady of the Evening	0	163080	0	['Dick Haymes']	['3BiJGZsyX9sJchTqcSA7Su']	1922	0.402	0.1580	3	-16.900	0	0.0390	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
586667	5rgu12WBIHQtvej2MdHSH0	云与海	50	258267	0	['阿YueYue']	['1QLBXKM5GCpyQQSVMNZqrZ']	2020-09-26	0.560	0.5180	0	-7.471	0	0.0292	
586668	0NuWgxEp51CutD2pJoF4OM	blind	72	153293	0	['ROLE MODEL']	['1dy5WNgIKQU6ezkpZs4y8z']	2020-10-21	0.765	0.6630	0	-5.223	1	0.0652	
586669	27Y1N4Q4U3EfDU5Ubw8ws2	What They'll Say About Us	70	187601	0	['FINNEAS']	['37M5pPGs6V1fchFJSgCguX']	2020-09-02	0.535	0.3140	7	-12.823	0	0.0408	
586670	45XJsGpFTyzbzeWK8VzR8S	A Day At A Time	58	142003	0	['Gentle Bones', 'Clara Benin']	['4jGPdu95icCKVF31CcFKbS', '5ebPSE9YI5aLeZ1Z2g...	2021-03-05	0.696	0.6150	10	-6.212	1	0.0345	
586671	5Ocn6dZ3BJFPWh4ylwFXtn	Mar de Emociones	38	214360	0	['Afrosound']	['0i4Qda0k4nf7jnNHmSNpYv']	2015-07-01	0.686	0.7230	6	-7.067	1	0.0363	

586672 rows × 20 columns

In [6]:
```
spotify.head()
```

Out[6]:

	id	name	popularity	duration_ms	explicit	artists	id_artists	release_date	danceability	energy	key	loudness	mode	speechiness	acousticness	in
0	35iwgR4jXetI318WEWsa1Q	Carve	6	126903	0	['Uli']	['45tlt06XoI0Iio4LBEVpls']	1922-02-22	0.645	0.4450	0	-13.338	1	0.4510	0.674	
1	021ht4sdgPcrDgSk7JTbKY	Capítulo 2.16 - Banquero Anarquista	0	98200	0	['Fernando Pessoa']	['14jtPCOoNZwquk5wd9DxrY']	1922-06-01	0.695	0.2630	0	-22.136	1	0.9570	0.797	
2	07A5yehtSnoedViJAZkNnc	Vivo para Quererte - Remasterizado	0	181640	0	['Ignacio Corsini']	['5LiOoJbxVSAMkBS2fUm3X2']	1922-03-21	0.434	0.1770	1	-21.180	1	0.0512	0.994	
3	08FmqUhxtyLTn6pAh6bk45	El Prisionero - Remasterizado	0	176907	0	['Ignacio Corsini']	['5LiOoJbxVSAMkBS2fUm3X2']	1922-03-21	0.321	0.0946	7	-27.961	1	0.0504	0.995	
4	08y9GfoqCWfOGsKdwojr5e	Lady of the Evening	0	163080	0	['Dick Haymes']	['3BiJGZsyX9sJchTqcSA7Su']	1922	0.402	0.1580	3	-16.900	0	0.0390	0.989	

In [7]:
```
spotify.tail()
```

Out[7]:

	id	name	popularity	duration_ms	explicit	artists	id_artists	release_date	danceability	energy	key	loudness	mode	speechiness	acoustic
586667	5rgu12WBIHQtvej2MdHSH0	云与海	50	258267	0	['阿YueYue']	['1QLBXKM5GCpyQQSVMNZqrZ']	2020-09-26	0.560	0.518	0	-7.471	0	0.0292	
586668	0NuWgxEp51CutD2pJoF4OM	blind	72	153293	0	['ROLE MODEL']	['1dy5WNgIKQU6ezkpZs4y8z']	2020-10-21	0.765	0.663	0	-5.223	1	0.0652	
586669	27Y1N4Q4U3EfDU5Ubw8ws2	What They'll Say About Us	70	187601	0	['FINNEAS']	['37M5pPGs6V1fchFJSgCguX']	2020-09-02	0.535	0.314	7	-12.823	0	0.0408	
586670	45XJsGpFTyzbzeWK8VzR8S	A Day At A Time	58	142003	0	['Gentle Bones', 'Clara Benin']	['4jGPdu95icCKVF31CcFKbS', '5ebPSE9YI5aLeZ1Z2g...	2021-03-05	0.696	0.615	10	-6.212	1	0.0345	
586671	5Ocn6dZ3BJFPWh4ylwFXtn	Mar de Emociones	38	214360	0	['Afrosound']	['0i4Qda0k4nf7jnNHmSNpYv']	2015-07-01	0.686	0.723	6	-7.067	1	0.0363	

In [8]:
```
pd.isnull(spotify)
```

Out[8]:

	id	name	popularity	duration_ms	explicit	artists	id_artists	release_date	danceability	energy	key	loudness	mode	speechiness	acousticness	instrumentalness	liveness	valence	tempo	t
0	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	
1	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	
2	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	
3	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	
4	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
586667	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	
586668	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	
586669	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	

586670	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
586671	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False

586672 rows × 20 columns

In [9]:
```
pd.isnull(spotify).sum()
```

Out[9]:
```
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
```

In [11]:
```
spotify.describe()
```

Out[11]:

	popularity	duration_ms	explicit	danceability	energy	key	loudness	mode	speechiness	acousticness	instrumentalness	liveness	valenc
count	586672.000000	5.866720e+05	586672.000000	586672.000000	586672.000000	586672.000000	586672.000000	586672.000000	586672.000000	586672.000000	586672.000000	586672.000000	586672.00000
mean	27.570053	2.300512e+05	0.044086	0.563594	0.542036	5.221603	-10.206067	0.658797	0.104864	0.449863	0.113451	0.213935	0.55229
std	18.370642	1.265261e+05	0.205286	0.166103	0.251923	3.519423	5.089328	0.474114	0.179893	0.348837	0.266868	0.184326	0.25767
min	0.000000	3.344000e+03	0.000000	0.000000	0.000000	0.000000	-60.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000
25%	13.000000	1.750930e+05	0.000000	0.453000	0.343000	2.000000	-12.891000	0.000000	0.034000	0.096900	0.000000	0.098300	0.34600
50%	27.000000	2.148930e+05	0.000000	0.577000	0.549000	5.000000	-9.243000	1.000000	0.044300	0.422000	0.000024	0.139000	0.56400
75%	41.000000	2.638670e+05	0.000000	0.686000	0.748000	8.000000	-6.482000	1.000000	0.076300	0.785000	0.009550	0.278000	0.76900
max	100.000000	5.621218e+06	1.000000	0.991000	1.000000	11.000000	5.376000	1.000000	0.971000	0.996000	1.000000	1.000000	1.00000

In [12]:
```
spotify.columns
```

Out[12]:
```
Index(['id', 'name', 'popularity', 'duration_ms', 'explicit', 'artists',
       'id_artists', 'release_date', 'danceability', 'energy', 'key',
       'loudness', 'mode', 'speechiness', 'acousticness', 'instrumentalness',
       'liveness', 'valence', 'tempo', 'time_signature'],
      dtype='object')
```

In [14]:
```
sort = spotify.sort_values('popularity', ascending =True)
sort
```

Out[14]:

	id	name	popularity	duration_ms	explicit	artists	id_artists	release_date	danceability	energy	key	loudness	mode	speechiness	acousticnes
546130	181rTRhCcggZPwP2TUcVqm	Newspaper Reports On Abner, 20 February 1935	0	896575	0	['Norris Goff', 'Chester Lauck', 'Carlton Bric...	['3WCwCPDMpGzrt0Qz6quumy', '7vk8UqABg0Sga78GI3...	1935-02-20	0.595	0.262	8	-17.746	1	0.9320	0.993
546222	0yOCz3V5KMm8I1T8EFc60i	恋は水の上で	0	188440	0	['Hibari Misora']	['1m5pMY5blqJwdxJ7vqQtuN']	1949	0.418	0.388	0	-8.580	1	0.0358	0.925
546221	0y48Hhwe52099UqYjegRCO	私の誕生日	0	173467	0	['Hibari Misora']	['1m5pMY5blqJwdxJ7vqQtuN']	1949	0.642	0.178	5	-11.700	1	0.0501	0.993
546220	0xCmgtf9ka07hkZg3D6PaV	エル・チョクロ (EL CHOCLO)	0	205280	0	['Hibari Misora']	['1m5pMY5blqJwdxJ7vqQtuN']	1949	0.695	0.467	0	-12.236	0	0.0422	0.827
546219	0tBXS3VuCPX7KWUFH2nros	恋は不思議なもの	0	185733	0	['Hibari Misora']	['1m5pMY5blqJwdxJ7vqQtuN']	1949	0.389	0.388	2	-8.221	1	0.0351	0.869
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
92811	6tDDoYIxWvMLTdKpjFkc1B	telepatía	97	160191	0	['Kali Uchis']	['1U1el3k54VvEUzo3ybLPlM']	2020-12-04	0.653	0.524	11	-9.016	0	0.0502	0.112
92810	5QO79kh1waicV47BqGRL3g	Save Your Tears	97	215627	1	['The Weeknd']	['1Xyo4u8uXC1ZmMpatF05PJ']	2020-03-20	0.680	0.826	0	-5.487	1	0.0309	0.021
93804	3Ofmpyhv5UAQ70mENzB277	Astronaut In The Ocean	98	132780	0	['Masked Wolf']	['1uU7g3DNSbsu0QjSEqZtEd']	2021-01-06	0.778	0.695	4	-6.865	0	0.0913	0.175
93803	7IPN2DXiMsVn7XUKtOW1CS	drivers license	99	242014	1	['Olivia Rodrigo']	['1McMsnEElThX1knmY4oliG']	2021-01-08	0.585	0.436	10	-8.761	1	0.0601	0.721
93802	4iJyoBOLtHqaGxP12qzhQI	Peaches (feat. Daniel Caesar & Giveon)	100	198082	1	['Justin Bieber', 'Daniel Caesar', 'Giveon']	['1uNFoZAHBGtllmzznpCI3s', '20wkVLutqVOYrc0kxF...	2021-03-19	0.677	0.696	0	-6.181	1	0.1190	0.321

586672 rows × 20 columns

In [15]:
```
sort = spotify.sort_values('popularity', ascending = True).head(10)
sort
```

Out[15]:

	id	name	popularity	duration_ms	explicit	artists	id_artists	release_date	danceability	energy	key	loudness	mode	speechiness	acousticness
546130	181rTRhCcggZPwP2TUcVqm	Newspaper Reports On Abner, 20 February 1935	0	896575	0	['Norris Goff', 'Chester Lauck', 'Carlton Bric...	['3WCwCPDMpGzrt0Qz6quumy', '7vk8UqABg0Sga78GI3...	1935-02-20	0.595	0.262	8	-17.746	1	0.9320	0.993
546222	0yOCz3V5KMm8I1T8EFc60i	恋は水の上で	0	188440	0	['Hibari Misora']	['1m5pMY5blqJwdxJ7vqQtuN']	1949	0.418	0.388	0	-8.580	1	0.0358	0.925
546221	0y48Hhwe52099UqYjegRCO	私の誕生日	0	173467	0	['Hibari Misora']	['1m5pMY5blqJwdxJ7vqQtuN']	1949	0.642	0.178	5	-11.700	1	0.0501	0.993
546220	0xCmgtf9ka07hkZg3D6PaV	エル・チョクロ (EL CHOCLO)	0	205280	0	['Hibari Misora']	['1m5pMY5blqJwdxJ7vqQtuN']	1949	0.695	0.467	0	-12.236	0	0.0422	0.827
546219	0tBXS3VuCPX7KWUFH2nros	恋は不思議なもの	0	185733	0	['Hibari Misora']	['1m5pMY5blqJwdxJ7vqQtuN']	1949	0.389	0.388	2	-8.221	1	0.0351	0.869
546218	0qrKnQtYDVJhKFAXTHYVS9	ゆうべはどうしたの (WHATSAMALLA U)	0	183427	0	['Hibari Misora']	['1m5pMY5blqJwdxJ7vqQtuN']	1949	0.631	0.249	5	-11.883	1	0.0355	0.951
546217	0nqsDxOeKSwEzp3AUQAAqS	Screen Director's Playhouse, Music For Million...	0	1767071	0	['Wilms Herbert', 'June Allyson', 'Joseph Kear...	['2rbm8QWvmnVwxFo84EVM1h', '4yW5adMgyIfHFzaL9i...	1949-04-10	0.533	0.317	7	-13.047	1	0.9180	0.682
546216	0kGEdsxVLYjCdfxM9tbezd	ブルーマンボ	0	162147	0	['Hibari Misora']	['1m5pMY5blqJwdxJ7vqQtuN']	1949	0.529	0.546	0	-6.462	0	0.0418	0.784
546215	0bc3PUZurUUXrY7yqoOxjq	Screen Director's Playhouse, Trade Winds direc...	0	1776652	0	['Wally Maher', 'Tay Garnett', 'Lurene Tuttle'...	['7hkhJTTI3VnUGVWUt8SJXT', '3kYeelpRCgJz4fQYDv...	1949-05-19	0.599	0.321	0	-15.428	0	0.9330	0.808
546214	0Wwm0ruSjYMIiWG0nyAI1F	Screen Director's Playhouse, It's A Wonderful ...	0	1767576	0	['Joseph Granby', 'Jimmy Stewart', 'Irene Tedr...	['6GK59BC4LJzqR0OpHAX2S3', '58BzBaExrnrx898sby...	1949-05-08	0.645	0.341	8	-12.177	1	0.8670	0.690

In [16]:
```
most = spotify.query('popularity>90', inplace = False).sort_values('popularity', ascending = False)
most[:10]
```

Out[16]:

	id	name	popularity	duration_ms	explicit	artists	id_artists	release_date	danceability	energy	key	loudness	mode	speechiness	acousticness
93802	4iJyoBOLtHqaGxP12qzhQI	Peaches (feat. Daniel Caesar & Giveon)	100	198082	1	['Justin Bieber', 'Daniel Caesar', 'Giveon']	['1uNFoZAHBGtllmzznpCI3s', '20wkVLutqVOYrc0kxF...	2021-03-19	0.677	0.696	0	-6.181	1	0.1190	0.32100
93803	7lPN2DXiMsVn7XUKtOW1CS	drivers license	99	242014	1	['Olivia Rodrigo']	['1McMsnEElThX1knmY4oliG']	2021-01-08	0.585	0.436	10	-8.761	1	0.0601	0.72100
93804	3Ofmpyhv5UAQ70mENzB277	Astronaut In The Ocean	98	132780	0	['Masked Wolf']	['1uU7g3DNSbsu0QjSEqZtEd']	2021-01-06	0.778	0.695	4	-6.865	0	0.0913	0.17500
92810	5QO79kh1waicV47BqGRL3g	Save Your Tears	97	215627	1	['The Weeknd']	['1Xyo4u8uXC1ZmMpatF05PJ']	2020-03-20	0.680	0.826	0	-5.487	1	0.0309	0.02120
92811	6tDDoYlxWvMLTdKpjFkc1B	telepatía	97	160191	0	['Kali Uchis']	['1U1el3k54VvEUzo3ybLPlM']	2020-12-04	0.653	0.524	11	-9.016	0	0.0502	0.11200
92813	0VjIjW4GlUZAMYd2vXMi3b	Blinding Lights	96	200040	0	['The Weeknd']	['1Xyo4u8uXC1ZmMpatF05PJ']	2020-03-20	0.514	0.730	1	-5.934	1	0.0598	0.00146
93805	7MAibcTIi4lisCtbHKrGMh	Leave The Door Open	96	242096	0	['Bruno Mars', 'Anderson .Paak', 'Silk Sonic']	['0du5cEVh5yTK9QJze8zA0C', '3jK9MiCrA42ILAdMGU...	2021-03-05	0.586	0.616	5	-7.964	1	0.0324	0.18200
92814	6f3Slt0GbA2bPZlz0aIFXN	The Business	95	164000	0	['Tiësto']	['2o5jDhtHVPhrJdv3cEQ99Z']	2020-09-16	0.798	0.620	8	-7.079	0	0.2320	0.41400
91866	60ynsPSSKe6O3sfwRnIBRf	Streets	94	226987	1	['Doja Cat']	['5cj0lLjcoR7YOSnhnX0Po5']	2019-11-07	0.749	0.463	11	-8.433	1	0.0828	0.20800
92816	3FAJ6O0NOHQV8Mc5Ri6ENp	Heartbreak Anniversary	94	198371	0	['Giveon']	['4fxd5Ee7UefO4CUXgwJ7IP']	2020-03-27	0.449	0.465	0	-8.964	1	0.0791	0.52400

In [19]:
```
spotify.columns
```

Out[19]:
```
Index(['id', 'name', 'popularity', 'duration_ms', 'explicit', 'artists',
       'id_artists', 'danceability', 'energy', 'key', 'loudness', 'mode',
       'speechiness', 'acousticness', 'instrumentalness', 'liveness',
       'valence', 'tempo', 'time_signature'],
      dtype='object')
```

In [20]:
```
spotify
```

Out[20]:

	id	name	popularity	duration_ms	explicit	artists	id_artists	danceability	energy	key	loudness	mode	speechiness	acousticness
release_date														

1922-02-22	35iwgR4jXetI318WEWsa1Q	Carve	6	126903	0	['Uli']	['45tIt06XoI0Iio4LBEVpls']	0.645	0.4450	0	-13.338	1	0.4510	0.674
1922-06-01	021ht4sdgPcrDgSk7JTbKY	Capítulo 2.16 - Banquero Anarquista	0	98200	0	['Fernando Pessoa']	['14jtPCOoNZwquk5wd9DxrY']	0.695	0.2630	0	-22.136	1	0.9570	0.797
1922-03-21	07A5yehtSnoedViJAZkNnc	Vivo para Quererte - Remasterizado	0	181640	0	['Ignacio Corsini']	['5LiOoJbxVSAMkBS2fUm3X2']	0.434	0.1770	1	-21.180	1	0.0512	0.994
1922-03-21	08FmqUhxtyLTn6pAh6bk45	El Prisionero - Remasterizado	0	176907	0	['Ignacio Corsini']	['5LiOoJbxVSAMkBS2fUm3X2']	0.321	0.0946	7	-27.961	1	0.0504	0.995
1922	08y9GfoqCWfOGsKdwojr5e	Lady of the Evening	0	163080	0	['Dick Haymes']	['3BiJGZsyX9sJchTqcSA7Su']	0.402	0.1580	3	-16.900	0	0.0390	0.989
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
2020-09-26	5rgu12WBIHQtvej2MdHSH0	云与海	50	258267	0	['阿YueYue']	['1QLBXKM5GCpyQQSVMNZqrZ']	0.560	0.5180	0	-7.471	0	0.0292	0.785
2020-10-21	0NuWgxEp51CutD2pJoF4OM	blind	72	153293	0	['ROLE MODEL']	['1dy5WNgIKQU6ezkpZs4y8z']	0.765	0.6630	0	-5.223	1	0.0652	0.141
2020-09-02	27Y1N4Q4U3EfDU5Ubw8ws2	What They'll Say About Us	70	187601	0	['FINNEAS']	['37M5pPGs6V1fchFJSgCguX']	0.535	0.3140	7	-12.823	0	0.0408	0.895
2021-03-05	45XJsGpFTyzbzeWK8VzR8S	A Day At A Time	58	142003	0	['Gentle Bones', 'Clara Benin']	['4jGPdu95icCKVF31CcFKbS', '5ebPSE9YI5aLeZ1Z2g...	0.696	0.6150	10	-6.212	1	0.0345	0.206
2015-07-01	5Ocn6dZ3BJFPWh4ylwFXtn	Mar de Emociones	38	214360	0	['Afrosound']	['0i4Qda0k4nf7jnNHmSNpYv']	0.686	0.7230	6	-7.067	1	0.0363	0.105

586672 rows × 19 columns

In [21]:
```
spotify.columns
```

Out[21]:
```
Index(['id', 'name', 'popularity', 'duration_ms', 'explicit', 'artists',
       'id_artists', 'danceability', 'energy', 'key', 'loudness', 'mode',
       'speechiness', 'acousticness', 'instrumentalness', 'liveness',
       'valence', 'tempo', 'time_signature'],
      dtype='object')
```

In [22]:
```
spotify.describe().transpose()
```

Out[22]:

	count	mean	std	min	25%	50%	75%	max
popularity	586672.0	27.570053	18.370642	0.0	13.0000	27.000000	41.00000	100.000
duration_ms	586672.0	230051.167286	126526.087418	3344.0	175093.0000	214893.000000	263867.00000	5621218.000
explicit	586672.0	0.044086	0.205286	0.0	0.0000	0.000000	0.00000	1.000
danceability	586672.0	0.563594	0.166103	0.0	0.4530	0.577000	0.68600	0.991
energy	586672.0	0.542036	0.251923	0.0	0.3430	0.549000	0.74800	1.000
key	586672.0	5.221603	3.519423	0.0	2.0000	5.000000	8.00000	11.000
loudness	586672.0	-10.206067	5.089328	-60.0	-12.8910	-9.243000	-6.48200	5.376
mode	586672.0	0.658797	0.474114	0.0	0.0000	1.000000	1.00000	1.000
speechiness	586672.0	0.104864	0.179893	0.0	0.0340	0.044300	0.07630	0.971
acousticness	586672.0	0.449863	0.348837	0.0	0.0969	0.422000	0.78500	0.996
instrumentalness	586672.0	0.113451	0.266868	0.0	0.0000	0.000024	0.00955	1.000
liveness	586672.0	0.213935	0.184326	0.0	0.0983	0.139000	0.27800	1.000
valence	586672.0	0.552292	0.257671	0.0	0.3460	0.564000	0.76900	1.000
tempo	586672.0	118.464857	29.764108	0.0	95.6000	117.384000	136.32100	246.381
time_signature	586672.0	3.873382	0.473162	0.0	4.0000	4.000000	4.00000	5.000

In [24]:
```
spotify[["artists"]].iloc[18]
```

Out[24]:
```
artists    ['Victor Boucher']
Name: 1922, dtype: object
```

In [25]:
```
spotify["duration"] = spotify["duration_ms"].apply(lambda x: round(x / 100))
spotify.drop("duration_ms", inplace =True, axis =1)
```

In [26]:
```
spotify.duration.head()
```

Out[26]:
```
release_date
1922-02-22    1269
1922-06-01     982
1922-03-21    1816
1922-03-21    1769
1922          1631
Name: duration, dtype: int64
```

In [33]:
```
print(spotify.info())
print(spotify.head())
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 586672 entries, 1922-02-22 to 2015-07-01
Data columns (total 19 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   explicit          586672 non-null  int64
 4   artists           586672 non-null  object
 5   id_artists        586672 non-null  object
 6   danceability      586672 non-null  float64
 7   energy            586672 non-null  float64
 8   key               586672 non-null  int64
 9   loudness          586672 non-null  float64
```

```
 10  mode              586672 non-null  int64
 11  speechiness       586672 non-null  float64
 12  acousticness      586672 non-null  float64
 13  instrumentalness  586672 non-null  float64
 14  liveness          586672 non-null  float64
 15  valence           586672 non-null  float64
 16  tempo             586672 non-null  float64
 17  time_signature    586672 non-null  int64
 18  duration          586672 non-null  int64
dtypes: float64(9), int64(6), object(4)
memory usage: 89.5+ MB
None
```

release_date	id	name	popularity	explicit	artists	id_artists	danceability	energy	key	loudness	mode	speechiness	acousticness	instrumentalness	liveness	valence	tempo	time_signature	duration
1922-02-22	35iwgR4jXetI318WEWsa1Q	Carve	6	0	['Uli']	['45tIt06XoI0Iio4LBEVpls']	0.645	0.4450	0	-13.338	1	0.4510	0.674	0.7440	0.151	0.127	104.851	3	1269
1922-06-01	021ht4sdgPcrDgSk7JTbKY	Capítulo 2.16 - Banquero Anarquista	0	0	['Fernando Pessoa']	['14jtPCOoNZwquk5wd9DxrY']	0.695	0.2630	0	-22.136	1	0.9570	0.797	0.0000	0.148	0.655	102.009	1	982
1922-03-21	07A5yehtSnoedViJAZkNnc	Vivo para Quererte - Remasterizado	0	0	['Ignacio Corsini']	['5LiOoJbxVSAMkBS2fUm3X2']	0.434	0.1770	1	-21.180	1	0.0512	0.994	0.0218	0.212	0.457	130.418	5	1816
1922-03-21	08FmqUhxtyLTn6pAh6bk45	El Prisionero - Remasterizado	0	0	['Ignacio Corsini']	['5LiOoJbxVSAMkBS2fUm3X2']	0.321	0.0946	7	-27.961	1	0.0504	0.995	0.9180	0.104	0.397	169.980	3	1769
1922	08y9GfoqCWfOGsKdwojr5e	Lady of the Evening	0	0	['Dick Haymes']	['3BiJGZsyX9sJchTqcSA7Su']	0.402	0.1580	3	-16.900	0	0.0390	0.989	0.1300	0.311	0.196	103.220	4	1631

In [46]:
```
spotify.columns
```

Out[46]:
```
RangeIndex(start=0, stop=5, step=1)
```

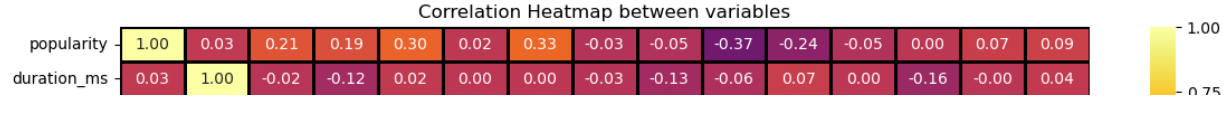
In [47]:
```
spotify = pd.read_csv("C:\\Users\\SUBHAJIT\\Desktop\\My Document\\Study Document\\Projects\\Spotify Data Analysis\\tracks.csv")
```

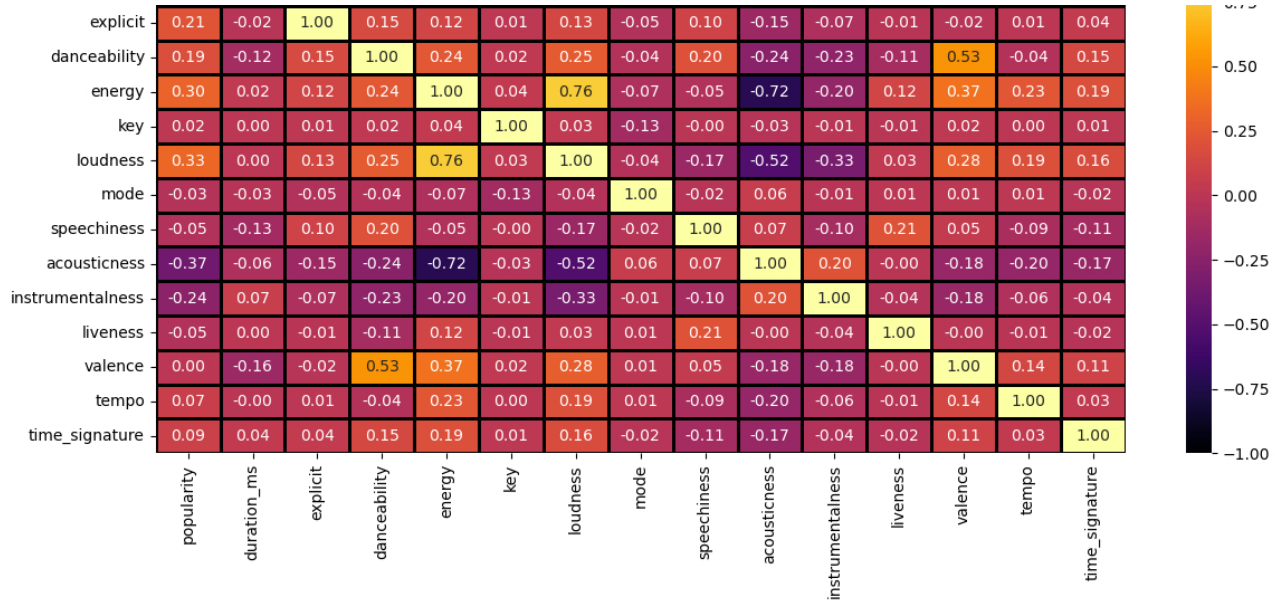
In [48]:
```
spotify.columns
```

Out[48]:
```
Index(['id', 'name', 'popularity', 'duration_ms', 'explicit', 'artists',
       'id_artists', 'release_date', 'danceability', 'energy', 'key',
       'loudness', 'mode', 'speechiness', 'acousticness', 'instrumentalness',
       'liveness', 'valence', 'tempo', 'time_signature'],
      dtype='object')
```

In [52]:
```
file_path = r'C:\Users\SUBHAJIT\Desktop\My Document\Study Document\Projects\Spotify Data Analysis\tracks.csv'
spotify = pd.read_csv(file_path)
print(spotify.info())
numeric_columns = spotify.select_dtypes(include=np.number).columns
corr_df = spotify[numeric_columns].corr(method='pearson')
plt.figure(figsize=(14, 6))
heatmap = sns.heatmap(data=corr_df, annot=True, fmt='.2f', vmin=-1, vmax=1, center=0, cmap='inferno', linewidths=1, linecolor='black')
heatmap.set_title('Correlation Heatmap between variables')
heatmap.set_xticklabels(heatmap.get_xticklabels(), rotation=90)
plt.show()
```

```
<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
None
```





In [53]:
```
spotify.columns
```

Out[53]:
```
Index(['id', 'name', 'popularity', 'duration_ms', 'explicit', 'artists',
       'id_artists', 'release_date', 'danceability', 'energy', 'key',
       'loudness', 'mode', 'speechiness', 'acousticness', 'instrumentalness',
       'liveness', 'valence', 'tempo', 'time_signature'],
      dtype='object')
```

In [54]:
```
spotify["duration"] = spotify["duration_ms"].apply(lambda x: round(x / 100))
spotify.drop("duration_ms", inplace =True, axis =1)
```

In [55]:
```
spotify.duration.head()
```

Out[55]:
```
0    1269
1     982
2    1816
3    1769
4    1631
Name: duration, dtype: int64
```

In [56]:
```
print(spotify.info())
print(spotify.head())
```

```
<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   explicit          586672 non-null  int64
 4   artists           586672 non-null  object
 5   id_artists        586672 non-null  object
 6   release_date      586672 non-null  object
 7   danceability      586672 non-null  float64
 8   energy            586672 non-null  float64
 9   key               586672 non-null  int64
 10  loudness          586672 non-null  float64
 11  mode              586672 non-null  int64
 12  speechiness       586672 non-null  float64
 13  acousticness      586672 non-null  float64
 14  instrumentalness  586672 non-null  float64
 15  liveness          586672 non-null  float64
 16  valence           586672 non-null  float64
 17  tempo             586672 non-null  float64
 18  time_signature    586672 non-null  int64
 19  duration          586672 non-null  int64
dtypes: float64(9), int64(6), object(5)
memory usage: 89.5+ MB
None
                     id                              name  popularity  \
0  35iwgR4jXetI318WEWsa1Q                             Carve           6
1  021ht4sdgPcrDgSk7JTbKY  Capítulo 2.16 - Banquero Anarquista           0
2  07A5yehtSnoedViJAZkNnc    Vivo para Quererte - Remasterizado           0
3  08FmqUhxtyLTn6pAh6bk45         El Prisionero - Remasterizado           0
4  08y9GfoqCWfOGsKdwojr5e                 Lady of the Evening           0

   explicit              artists                 id_artists release_date  \
0         0             ['Uli']  ['45tIt06XoI0Iio4LBEVpls']   1922-02-22
1         0  ['Fernando Pessoa']  ['14jtPCOoNZwquk5wd9DxrY']   1922-06-01
2         0  ['Ignacio Corsini']  ['5LiOoJbxVSAMkBS2fUm3X2']   1922-03-21
3         0  ['Ignacio Corsini']  ['5LiOoJbxVSAMkBS2fUm3X2']   1922-03-21
4         0      ['Dick Haymes']  ['3BiJGZsyX9sJchTqcSA7Su']         1922

   danceability  energy  key  loudness  mode  speechiness  acousticness  \
0         0.645  0.4450    0   -13.338     1       0.4510         0.674
1         0.695  0.2630    0   -22.136     1       0.9570         0.797
2         0.434  0.1770    1   -21.180     1       0.0512         0.994
3         0.321  0.0946    7   -27.961     1       0.0504         0.995
4         0.402  0.1580    3   -16.900     0       0.0390         0.989

   instrumentalness  liveness  valence    tempo  time_signature  duration
0            0.7440     0.151    0.127  104.851               3      1269
1            0.0000     0.148    0.655  102.009               1       982
2            0.0218     0.212    0.457  130.418               5      1816
```

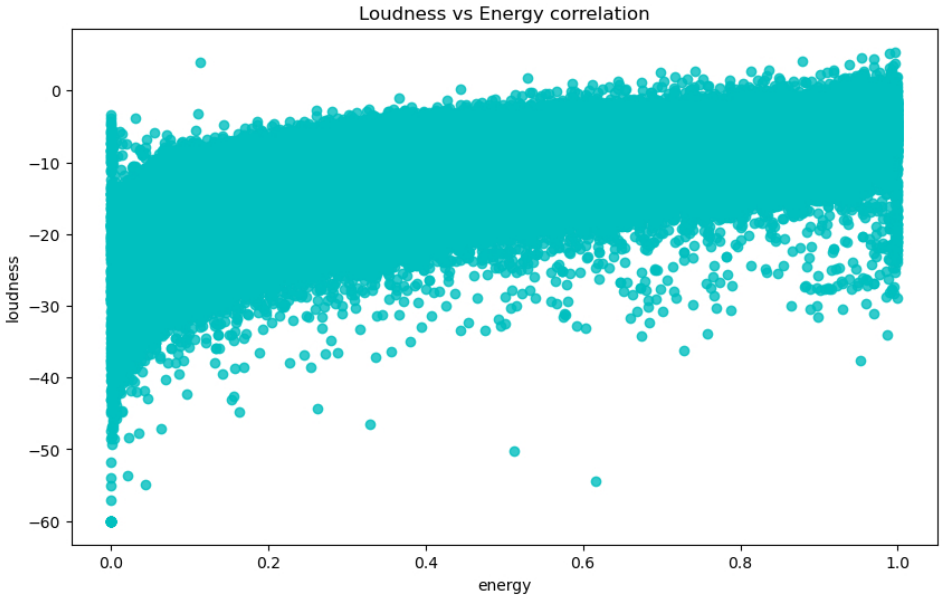
```
5          0.9180    0.164    0.597  169.960          5      1769
4          0.1300    0.311    0.196  103.220          4      1631
```

In [57]:
```
sample_df = spotify.sample(int(0.004 * len(spotify)))
```

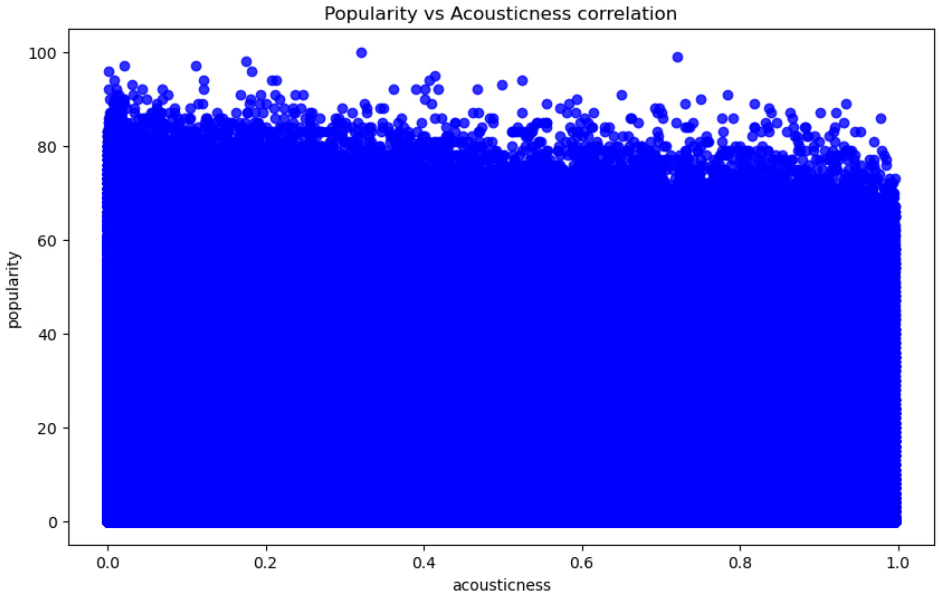
In [58]:
```
print(len(sample_df))
```

```
2346
```

In [61]:
```
file_path = r'C:\Users\SUBHAJIT\Desktop\My Document\Study Document\Projects\Spotify Data Analysis\tracks.csv'
data = pd.read_csv(file_path)
plt.figure(figsize=(10, 6))
sns.regplot(data=data, y="loudness", x="energy", color="c").set(title="Loudness vs Energy correlation")
plt.show()
```



In [66]:
```
file_path = r'C:\Users\SUBHAJIT\Desktop\My Document\Study Document\Projects\Spotify Data Analysis\tracks.csv'
data = pd.read_csv(file_path)
plt.figure(figsize=(10, 6))
sns.regplot(data=data, y="popularity", x="acousticness", color="b").set(title="Popularity vs Acousticness correlation")
plt.show()
```



In [75]:
```
#spotify['dates'] = spotify.index.get_level_values('release_date')
#spotify.dates = pd.to_datetime(spotify.dates)
#years = spotify.dates.dt.year
print(spotify['release_date'].unique())
```

```
['1922-02-22' '1922-06-01' '1922-03-21' ... '1991-05' '1996-10-06'
 '2009-11-28']
```

In [76]:
```
spotify['dates'] = pd.to_datetime(spotify['release_date'], errors='coerce')
```

In [78]:
```
spotify['dates'] = pd.to_datetime(spotify['release_date'], format='%Y-%m-%d', errors='coerce')
```

In [79]:
```
spotify['dates']
```

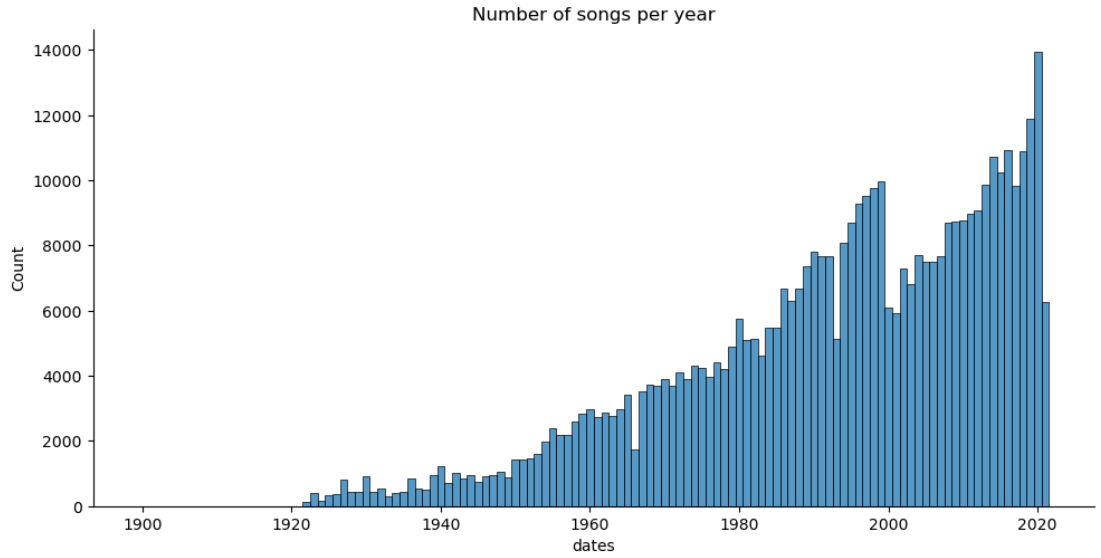
Out[79]:
```
0        1922-02-22
1        1922-06-01
2        1922-03-21
3        1922-03-21
4               NaT
            ...
586667   2020-09-26
```

```
586668   2020-10-21
586669   2020-09-02
586670   2021-03-05
586671   2015-07-01
Name: dates, Length: 586672, dtype: datetime64[ns]
```

In [83]:
```
years = spotify['dates'].dt.year
sns.displot(years, discrete =True, aspect =2, height =5, kind ="hist").set(title= "Number of songs per year")
```

```
D:\anaconda\anaconda3\Lib\site-packages\seaborn\axisgrid.py:118: UserWarning: The figure layout has changed to tight
  self._figure.tight_layout(*args, **kwargs)
```

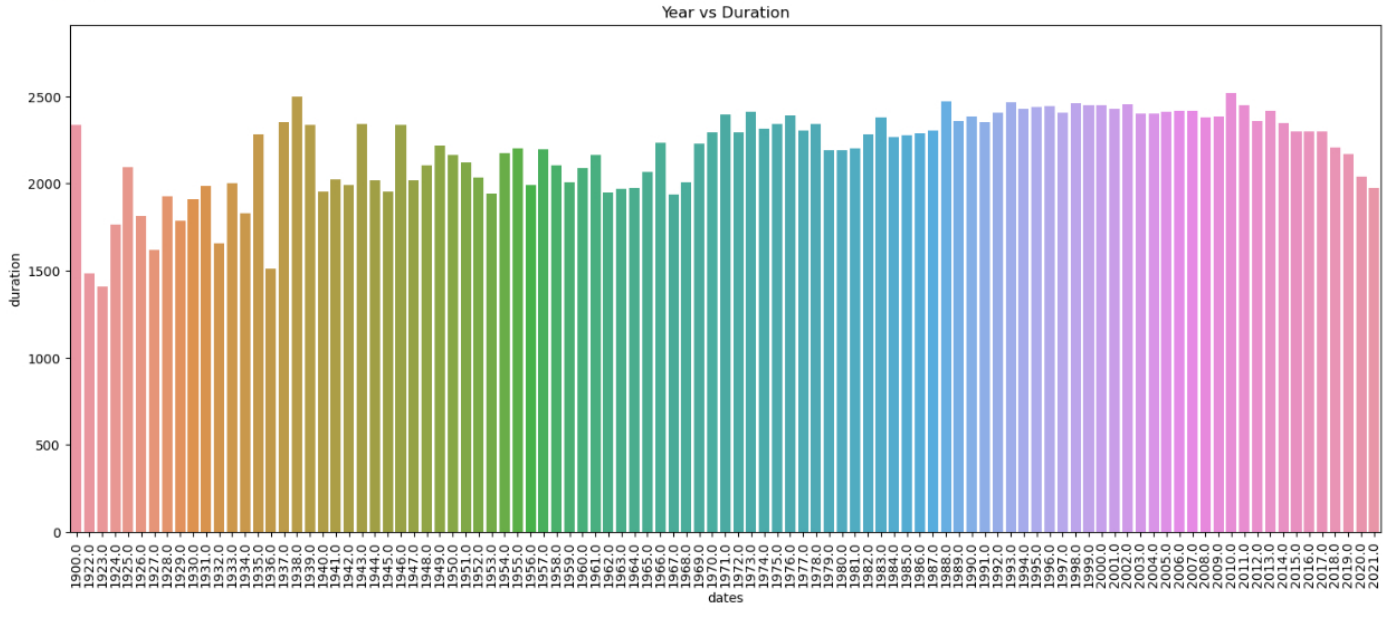
Out[83]:
```
<seaborn.axisgrid.FacetGrid at 0x183d28f8050>
```



In [84]:
```
total_dr = spotify.duration
fig_dims =(18, 7)
fig,  ax = plt.subplots(figsize = fig_dims)
fig = sns.barplot(x = years, y = total_dr, ax = ax, errwidth = False).set(title = "Year vs Duration")
plt.xticks(rotation =90)
```

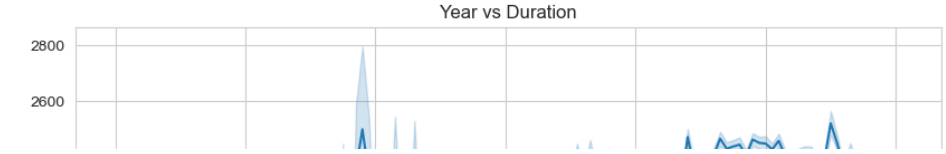
Out[84]:
```
(array([  0,   1,   2,   3,   4,   5,   6,   7,   8,   9,  10,  11,  12,
         13,  14,  15,  16,  17,  18,  19,  20,  21,  22,  23,  24,  25,
         26,  27,  28,  29,  30,  31,  32,  33,  34,  35,  36,  37,  38,
         39,  40,  41,  42,  43,  44,  45,  46,  47,  48,  49,  50,  51,
         52,  53,  54,  55,  56,  57,  58,  59,  60,  61,  62,  63,  64,
         65,  66,  67,  68,  69,  70,  71,  72,  73,  74,  75,  76,  77,
         78,  79,  80,  81,  82,  83,  84,  85,  86,  87,  88,  89,  90,
         91,  92,  93,  94,  95,  96,  97,  98,  99, 100]),
 [Text(0, 0, '1900.0'),
  Text(1, 0, '1922.0'),
  Text(2, 0, '1923.0'),
  Text(3, 0, '1924.0'),
  Text(4, 0, '1925.0'),
  Text(5, 0, '1926.0'),
  Text(6, 0, '1927.0'),
  Text(7, 0, '1928.0'),
  Text(8, 0, '1929.0'),
  Text(9, 0, '1930.0'),
  Text(10, 0, '1931.0'),
  Text(11, 0, '1932.0'),
  Text(12, 0, '1933.0'),
  Text(13, 0, '1934.0'),
  Text(14, 0, '1935.0'),
  Text(15, 0, '1936.0'),
  Text(16, 0, '1937.0'),
  Text(17, 0, '1938.0'),
  Text(18, 0, '1939.0'),
  Text(19, 0, '1940.0'),
  Text(20, 0, '1941.0'),
  Text(21, 0, '1942.0'),
  Text(22, 0, '1943.0'),
  Text(23, 0, '1944.0'),
  Text(24, 0, '1945.0'),
  Text(25, 0, '1946.0'),
  Text(26, 0, '1947.0'),
  Text(27, 0, '1948.0'),
  Text(28, 0, '1949.0'),
  Text(29, 0, '1950.0'),
  Text(30, 0, '1951.0'),
  Text(31, 0, '1952.0'),
  Text(32, 0, '1953.0'),
  Text(33, 0, '1954.0'),
  Text(34, 0, '1955.0'),
  Text(35, 0, '1956.0'),
  Text(36, 0, '1957.0'),
  Text(37, 0, '1958.0'),
  Text(38, 0, '1959.0'),
  Text(39, 0, '1960.0'),
  Text(40, 0, '1961.0'),
  Text(41, 0, '1962.0'),
  Text(42, 0, '1963.0'),
  Text(43, 0, '1964.0'),
  Text(44, 0, '1965.0'),
  Text(45, 0, '1966.0'),
  Text(46, 0, '1967.0'),
  Text(47, 0, '1968.0'),
  Text(48, 0, '1969.0'),
  Text(49, 0, '1970.0'),
  Text(50, 0, '1971.0'),
```

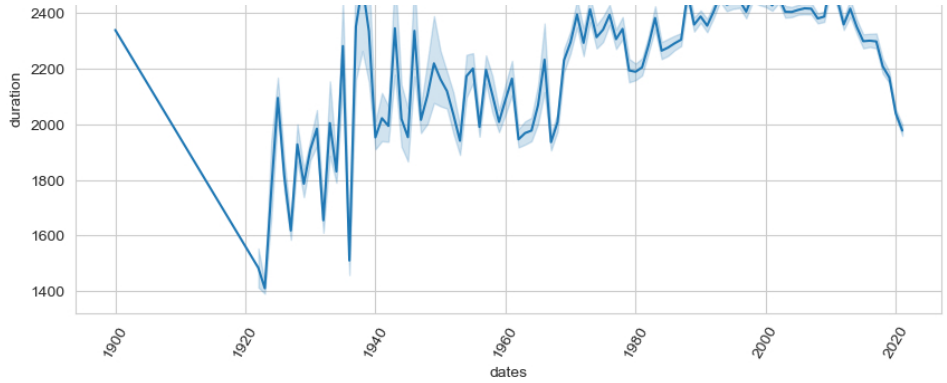
```
Text(51, 0, '1972.0'),
Text(52, 0, '1973.0'),
Text(53, 0, '1974.0'),
Text(54, 0, '1975.0'),
Text(55, 0, '1976.0'),
Text(56, 0, '1977.0'),
Text(57, 0, '1978.0'),
Text(58, 0, '1979.0'),
Text(59, 0, '1980.0'),
Text(60, 0, '1981.0'),
Text(61, 0, '1982.0'),
Text(62, 0, '1983.0'),
Text(63, 0, '1984.0'),
Text(64, 0, '1985.0'),
Text(65, 0, '1986.0'),
Text(66, 0, '1987.0'),
Text(67, 0, '1988.0'),
Text(68, 0, '1989.0'),
Text(69, 0, '1990.0'),
Text(70, 0, '1991.0'),
Text(71, 0, '1992.0'),
Text(72, 0, '1993.0'),
Text(73, 0, '1994.0'),
Text(74, 0, '1995.0'),
Text(75, 0, '1996.0'),
Text(76, 0, '1997.0'),
Text(77, 0, '1998.0'),
Text(78, 0, '1999.0'),
Text(79, 0, '2000.0'),
Text(80, 0, '2001.0'),
Text(81, 0, '2002.0'),
Text(82, 0, '2003.0'),
Text(83, 0, '2004.0'),
Text(84, 0, '2005.0'),
Text(85, 0, '2006.0'),
Text(86, 0, '2007.0'),
Text(87, 0, '2008.0'),
Text(88, 0, '2009.0'),
Text(89, 0, '2010.0'),
Text(90, 0, '2011.0'),
Text(91, 0, '2012.0'),
Text(92, 0, '2013.0'),
Text(93, 0, '2014.0'),
Text(94, 0, '2015.0'),
Text(95, 0, '2016.0'),
Text(96, 0, '2017.0'),
Text(97, 0, '2018.0'),
Text(98, 0, '2019.0'),
Text(99, 0, '2020.0'),
Text(100, 0, '2021.0')])
```



In [86]:
```
total_dr = spotify.duration
sns.set_style(style="whitegrid")
fig_dims =(10, 5)
fig,  ax = plt.subplots(figsize = fig_dims)
fig = sns.lineplot(x = years, y = total_dr, ax = ax).set(title = "Year vs Duration")
plt.xticks(rotation =60)
```

Out[86]:
```
(array([1880., 1900., 1920., 1940., 1960., 1980., 2000., 2020., 2040.]),
 [Text(1880.0, 0, '1880'),
  Text(1900.0, 0, '1900'),
  Text(1920.0, 0, '1920'),
  Text(1940.0, 0, '1940'),
  Text(1960.0, 0, '1960'),
  Text(1980.0, 0, '1980'),
  Text(2000.0, 0, '2000'),
  Text(2020.0, 0, '2020'),
  Text(2040.0, 0, '2040')])
```





In [87]:
```
spotify1 = pd.read_csv("C:\\Users\\SUBHAJIT\\Desktop\\My Document\\Study Document\\Projects\\Spotify Data Analysis\\SpotifyFeatures.csv")
```

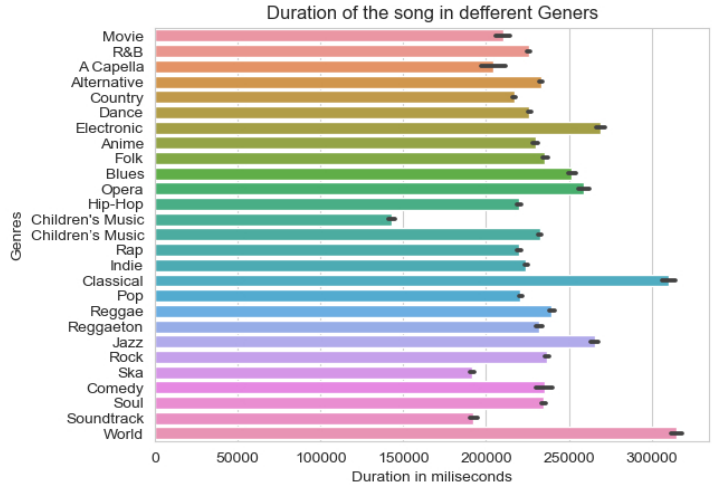
In [88]:
```
spotify1.head()
```

Out[88]:

	genre	artist_name	track_name	track_id	popularity	acousticness	danceability	duration_ms	energy	instrumentalness	key	liveness	loudness	mode	speechiness	tempo	time_signa
0	Movie	Henri Salvador	C'est beau de faire un Show	0BRjO6ga9RKCKjfDqeFgWV	0	0.611	0.389	99373	0.910	0.000	C#	0.3460	-1.828	Major	0.0525	166.969	
1	Movie	Martin & les fées	Perdu d'avance (par Gad Elmaleh)	0BjC1NfoEOOusryehmNudP	1	0.246	0.590	137373	0.737	0.000	F#	0.1510	-5.559	Minor	0.0868	174.003	
2	Movie	Joseph Williams	Don't Let Me Be Lonely Tonight	0CoSDzoNIKCRs124s9uTVy	3	0.952	0.663	170267	0.131	0.000	C	0.1030	-13.879	Minor	0.0362	99.488	
3	Movie	Henri Salvador	Dis-moi Monsieur Gordon Cooper	0Gc6TVm52BwZD07Ki6tIvf	0	0.703	0.240	152427	0.326	0.000	C#	0.0985	-12.178	Major	0.0395	171.758	
4	Movie	Fabien Nataf	Ouverture	0IusIXpMROHdEPvSl1fTQK	4	0.950	0.331	82625	0.225	0.123	F	0.2020	-21.150	Major	0.0456	140.576	

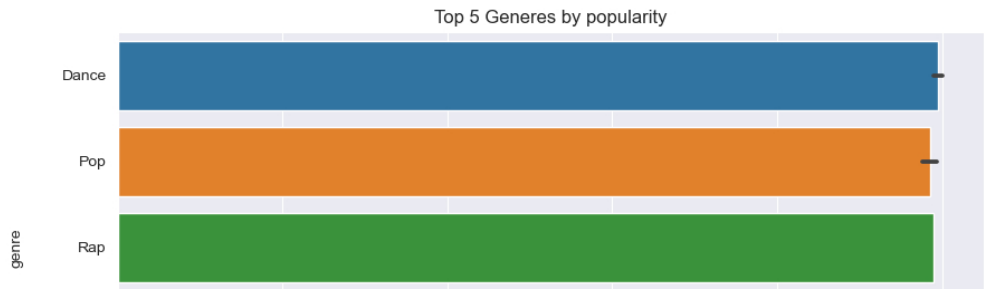
In [89]:
```
plt.title("Duration of the song in defferent Geners")
sns.color_palette('rocket', as_cmap = True)
sns.barplot(y = 'genre', x = 'duration_ms', data = spotify1)
plt.xlabel("Duration in miliseconds")
plt.ylabel("Genres")
```

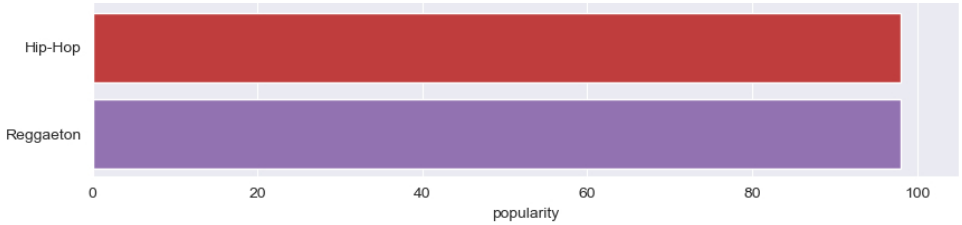
Out[89]:
```
Text(0, 0.5, 'Genres')
```



In [90]:
```
sns.set_style(style ="darkgrid")
plt.figure(figsize =(10, 5))
famous = spotify1.sort_values("popularity", ascending =False).head(10)
sns.barplot(y ='genre', x ='popularity', data =famous).set(title = "Top 5 Generes by popularity")
```

Out[90]:
```
[Text(0.5, 1.0, 'Top 5 Generes by popularity')]
```





In [91]:
```
spotify1.columns
```

Out[91]:
```
Index(['genre', 'artist_name', 'track_name', 'track_id', 'popularity',
       'acousticness', 'danceability', 'duration_ms', 'energy',
       'instrumentalness', 'key', 'liveness', 'loudness', 'mode',
       'speechiness', 'tempo', 'time_signature', 'valence'],
      dtype='object')
```

In [ ]: