Spotify Data Analysis

In [1]:
```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')
```

In [2]:
```python
df_tracks = pd.read_csv("/kaggle/input/spotify-datasets/tracks.csv")
df_tracks.head()
```

Out[2]:

	id	name	popularity	duration_ms	explicit	artists	
0	35iwgR4jXetl318WEWsa1Q	Carve	6	126903	0	['Uli']	[
1	021ht4sdgPcrDgSk7JTbKY	Capítulo 2.16 - Banquero Anarquista	0	98200	0	['Fernando Pessoa']	['14jtF
2	07A5yehtSnoedViJAZkNnc	Vivo para Quererte - Remasterizado	0	181640	0	['Ignacio Corsini']	['5LiO
3	08FmqUhxtyLTn6pAh6bk45	El Prisionero - Remasterizado	0	176907	0	['Ignacio Corsini']	['5LiO
4	08y9GfoqCWfOGsKdwojr5e	Lady of the Evening	0	163080	0	['Dick Haymes']	['3E

#Finding Null values

In [3]:
```python
pd.isnull(df_tracks).sum()
```

Out[3]:
```
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 [4]:  df_tracks.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
```

```
In [5]: sorted_df = df_tracks.sort_values('popularity',ascending=True).head(10)
        sorted_df
```

Out[5]:

	id	name	popularity	duration_ms	explicit	artists	
546130	181rTRhCcggZPwP2TUcVqm	Newspaper Reports On Abner, 20 February 1935	0	896575	0	['Norris Goff', 'Chester Lauck', 'Carlton Bric...	['3'
546222	0yOCz3V5KMm8l1T8EFc60i	恋は水の上で	0	188440	0	['Hibari Misora']	[
546221	0y48Hhwe52099UqYjegRCO	私の誕生日	0	173467	0	['Hibari Misora']	[
546220	0xCmgtf9ka07hkZg3D6PaV	エル・チョクロ (EL CHOCLO)	0	205280	0	['Hibari Misora']	[
546219	0tBXS3VuCPX7KWUFH2nros	恋は不思議なもの	0	185733	0	['Hibari Misora']	[
546218	0qrKnQtYDVJhKFAXTHYVS9	ゆうべはどうしたの (WHATSA MALLA U)	0	183427	0	['Hibari Misora']	[
546217	0nqsDxOeKSwEzp3AUQAAqS	Screen Director's Playhouse, Music For Million...	0	1767071	0	['Wilms Herbert', 'June Allyson', 'Joseph Kear...	['2r
546216	0kGEdsxVLYjCdfxM9tbezd	ブルーマンボ	0	162147	0	['Hibari Misora']	[
546215	0bc3PUZurUUXrY7yqoOxjq	Screen Director's Playhouse, Trade Winds direc...	0	1776652	0	['Wally Maher', 'Tay Garnett', 'Lurene Tuttle'...	
546214	0Wwm0ruSjYMIiWG0nyAI1F	Screen Director's Playhouse, It's A Wonderful ...	0	1767576	0	['Joseph Granby', 'Jimmy Stewart', 'Irene Tedr...	[

```
In [6]: df_tracks.describe().transpose()
```

Out[6]:

	count	mean	std	min	25%	50%
popularity	586672.0	27.570053	18.370642	0.0	13.0000	27.000000
duration_ms	586672.0	230051.167286	126526.087418	3344.0	175093.0000	214893.000000
explicit	586672.0	0.044086	0.205286	0.0	0.0000	0.000000
danceability	586672.0	0.563594	0.166103	0.0	0.4530	0.577000
energy	586672.0	0.542036	0.251923	0.0	0.3430	0.549000
key	586672.0	5.221603	3.519423	0.0	2.0000	5.000000
loudness	586672.0	-10.206067	5.089328	-60.0	-12.8910	-9.243000
mode	586672.0	0.658797	0.474114	0.0	0.0000	1.000000
speechiness	586672.0	0.104864	0.179893	0.0	0.0340	0.044300
acousticness	586672.0	0.449863	0.348837	0.0	0.0969	0.422000
instrumentalness	586672.0	0.113451	0.266868	0.0	0.0000	0.000024
liveness	586672.0	0.213935	0.184326	0.0	0.0983	0.139000
valence	586672.0	0.552292	0.257671	0.0	0.3460	0.564000
tempo	586672.0	118.464857	29.764108	0.0	95.6000	117.384000
time_signature	586672.0	3.873382	0.473162	0.0	4.0000	4.000000

```
In [7]: most_popular=df_tracks.query('popularity>90',inplace=False).sort_values('popularity',ascendir
        most_popular[:10]
```

Out[7]:

|	id	name	popularity	duration_ms	explicit	artists	
93802	4iJyoBOLtHqaGxP12qzhQI	Peaches (feat. Daniel Caesar & Giveon)	100	198082	1	['Justin Bieber', 'Daniel Caesar', 'Giveon']	[
93803	7lPN2DXiMsVn7XUKtOW1CS	drivers license	99	242014	1	['Olivia Rodrigo']	['
93804	3Ofmpyhv5UAQ70mENzB277	Astronaut In The Ocean	98	132780	0	['Masked Wolf']	['1
92810	5QO79kh1waicV47BqGRL3g	Save Your Tears	97	215627	1	['The Weeknd']	['1
92811	6tDDoYIxWvMLTdKpjFkc1B	telepatía	97	160191	0	['Kali Uchis']	[
92813	0VjIjW4GlUZAMYd2vXMi3b	Blinding Lights	96	200040	0	['The Weeknd']	['1
93805	7MAibcTli4IisCtbHKrGMh	Leave The Door Open	96	242096	0	['Bruno Mars', 'Anderson .Paak', 'Silk Sonic']	['
92814	6f3Slt0GbA2bPZlz0aIFXN	The Business	95	164000	0	['Tiësto']	[
91866	60ynsPSSKe6O3sfwRnIBRf	Streets	94	226987	1	['Doja Cat']	
92816	3FAJ6O0NOHQV8Mc5Ri6ENp	Heartbreak Anniversary	94	198371	0	['Giveon']	[

```
In [8]:  df_tracks.set_index("release_date",inplace=True)
         df_tracks.index=pd.to_datetime(df_tracks.index,format='mixed',dayfirst=True)
         df_tracks.head()
```

Out[8]:

release_date	id	name	popularity	duration_ms	explicit	art
1922-02-22	35iwgR4jXetI318WEWsa1Q	Carve	6	126903	0	['U
1922-06-01	021ht4sdgPcrDgSk7JTbKY	Capítulo 2.16 - Banquero Anarquista	0	98200	0	['Fernar Pess
1922-03-21	07A5yehtSnoedViJAZkNnc	Vivo para Quererte - Remasterizado	0	181640	0	['Igna Cors
1922-03-21	08FmqUhxtyLTn6pAh6bk45	El Prisionero - Remasterizado	0	176907	0	['Igna Cors
1922-01-01	08y9GfoqCWfOGsKdwojr5e	Lady of the Evening	0	163080	0	['D Haym

```
In [9]:  df_tracks['duration']=df_tracks['duration_ms'].apply(lambda x:round(x/1000))
         df_tracks.drop('duration_ms',inplace=True,axis=1)
```

```
In [10]: corr_df=df_tracks.drop(['key','mode','explicit'],axis=1).corr(method='pearson',numeric_only=Tr
         plt.figure(figsize=(14,6))
         heatmap=sns.heatmap(corr_df,annot=True,fmt='.1g',vmin=-1,vmax=1,center=0,cmap='infern
         heatmap.set_title('Correlation Heatmap Between Variable')
         heatmap.set_xticklabels(heatmap.get_xticklabels(),rotation=90);
```
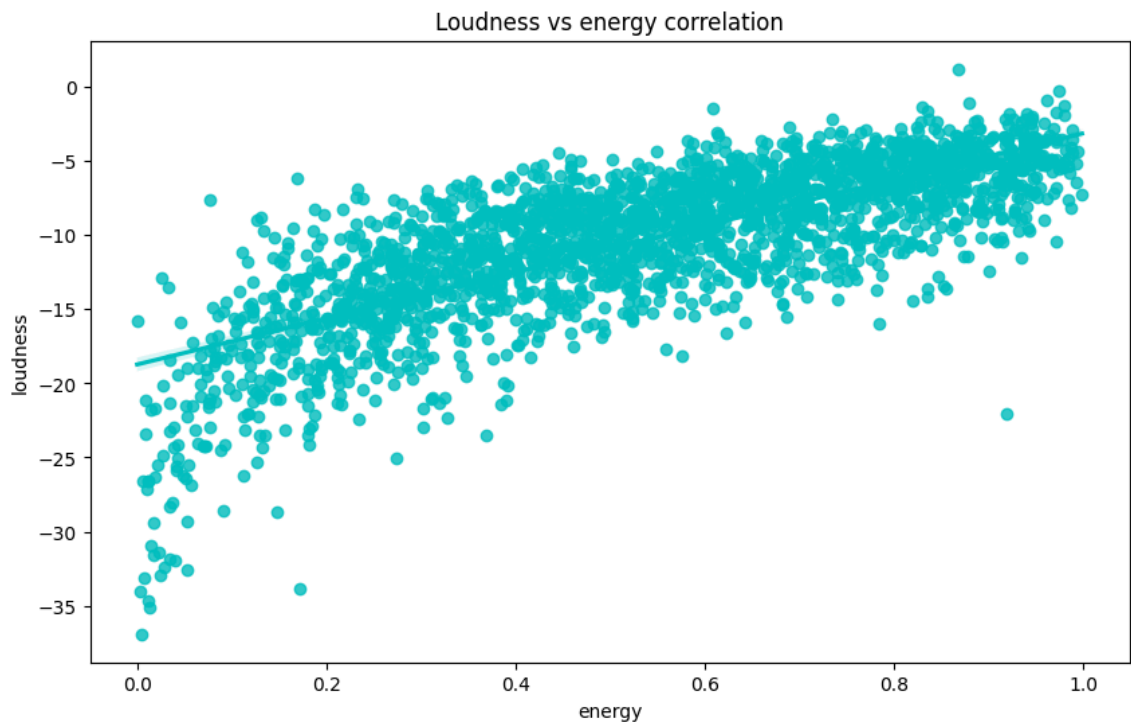

Correlation Heatmap Between Variable
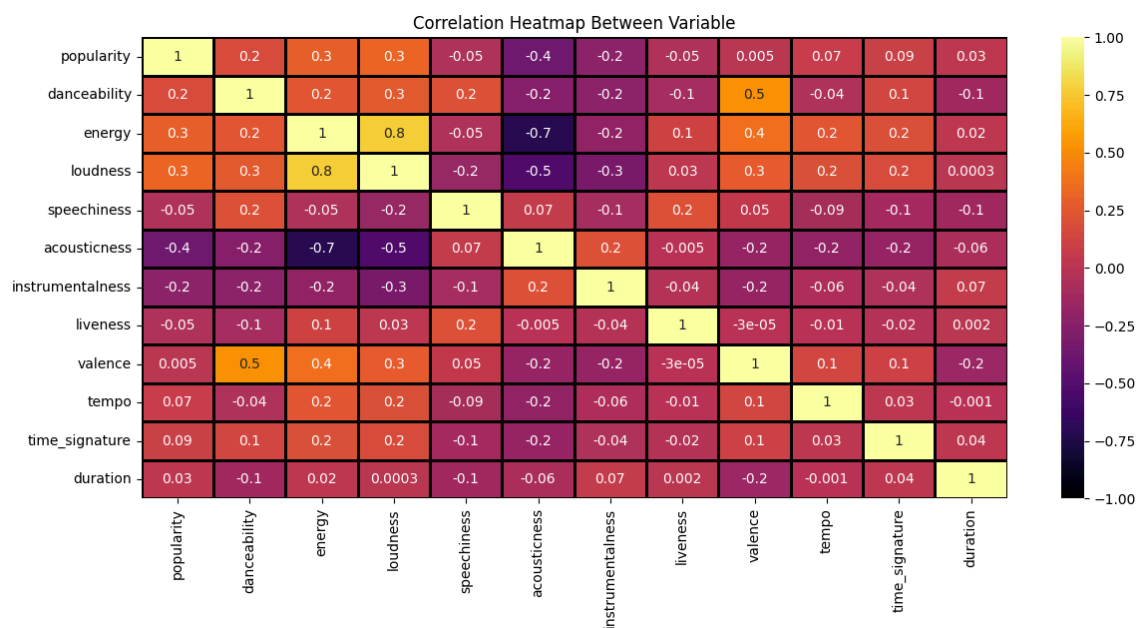
#Sampling data

```
In [11]:  sample_df =df_tracks.sample(int(0.004*len(df_tracks)))
```
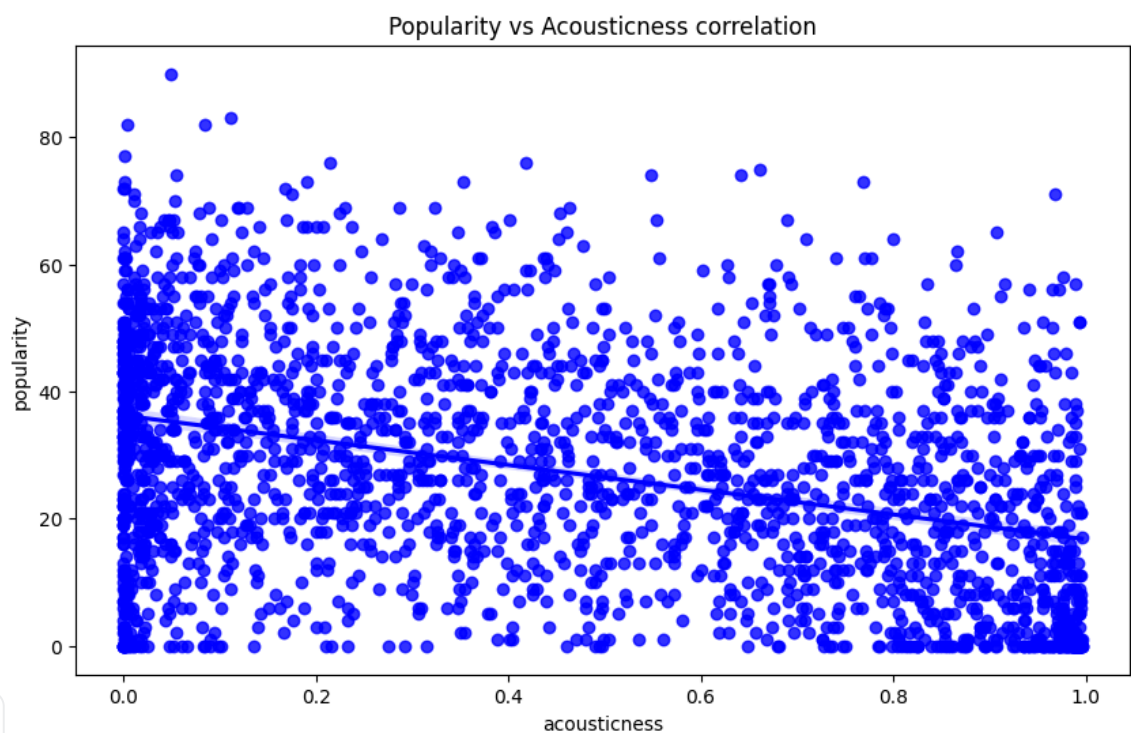
```
In [12]:  print(len(sample_df))
```

2346

```
In [13]:  plt.figure(figsize=(10,6))
          sns.regplot(data=sample_df,y='loudness',x='energy',color='c').set(title="Loudness vs energy c
```


Loudness vs energy correlation

```
In [14]:  plt.figure(figsize=(10,6))
          sns.regplot(data=sample_df,y='popularity',x='acousticness',color='b').set(title="Popularity vs /
```


Popularity vs Acousticness correlation

In [15]:
```python
df_tracks['dates']=df_tracks.index.get_level_values('release_date')
df_tracks.dates=pd.to_datetime(df_tracks.dates)
years=df_tracks.dates.dt.year
```

In [16]:
```python
sns.displot(years,discrete=True,aspect=2,height=5,kind='hist').set(title='Number of songs per
```

Number of songs per year
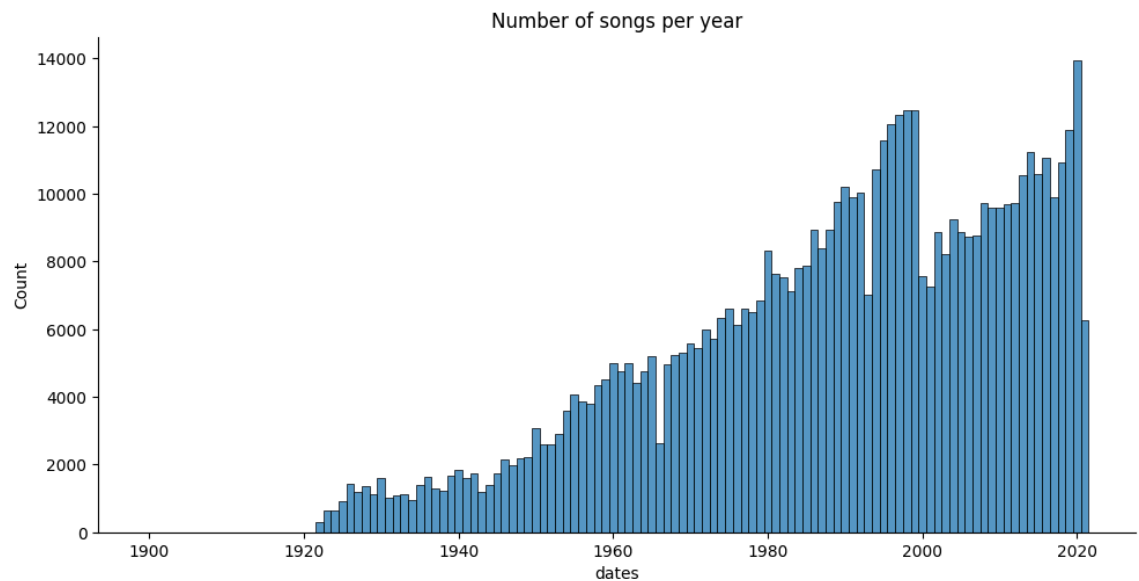
In [17]:
```python
total_dr=df_tracks.duration
fig_dims=(18,7)
fig,ax=plt.subplots(figsize=fig_dims)
fig=sns.barplot(x=years,y=total_dr,ax=ax,errwidth=False).set(title='Year vs Duration')
plt.xticks(rotation=90);
```
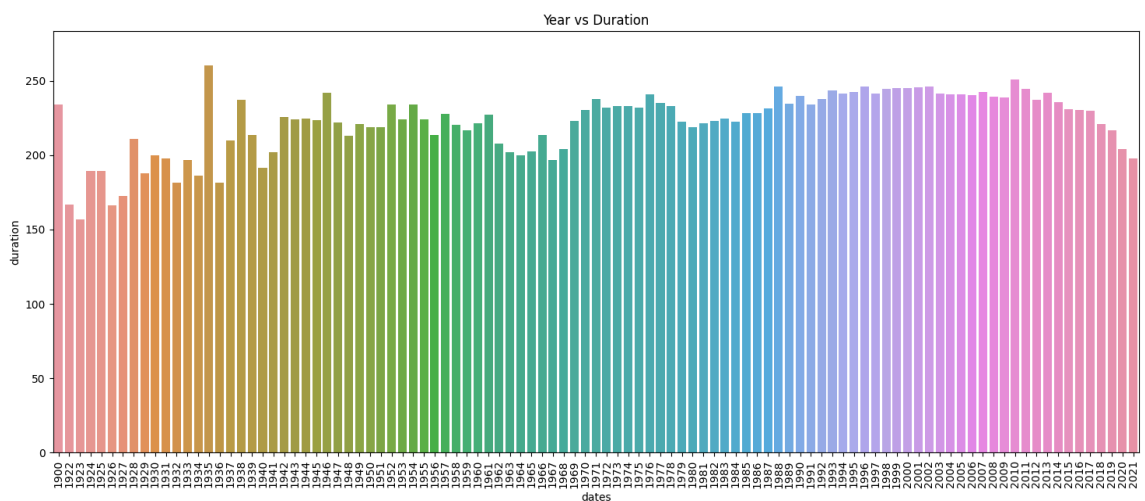
Year vs Duration

In [18]:
```
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);
```
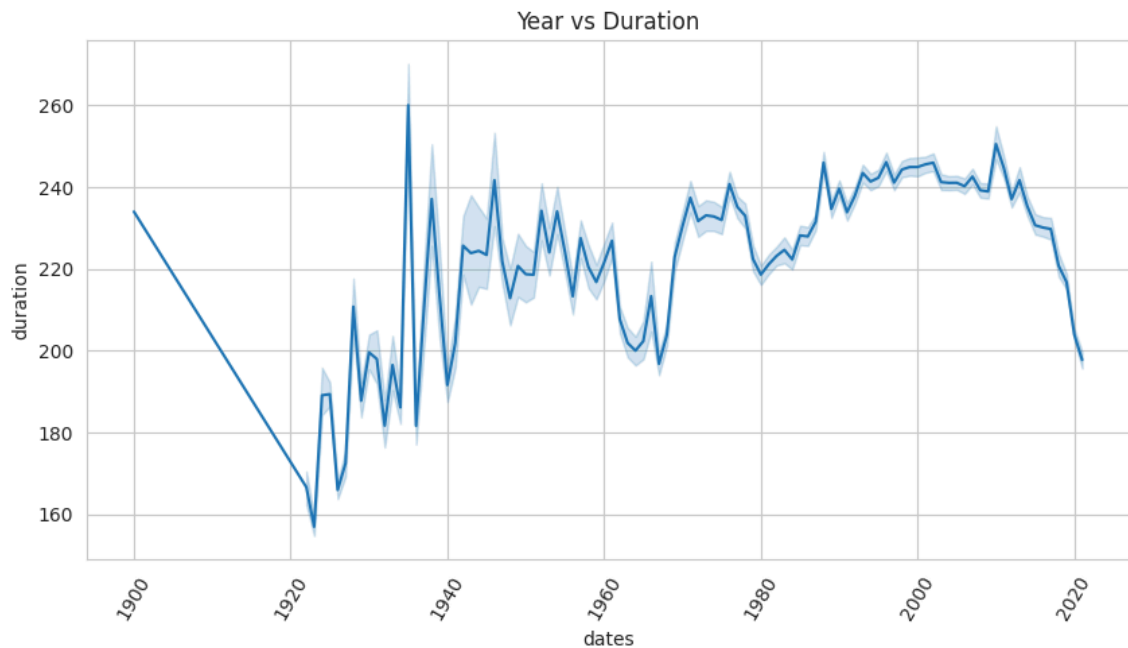

#Analysing genres

In [19]:
```
df_genre=pd.read_csv("/kaggle/input/ultimate-spotify-tracks-db/SpotifyFeatures.csv")
```
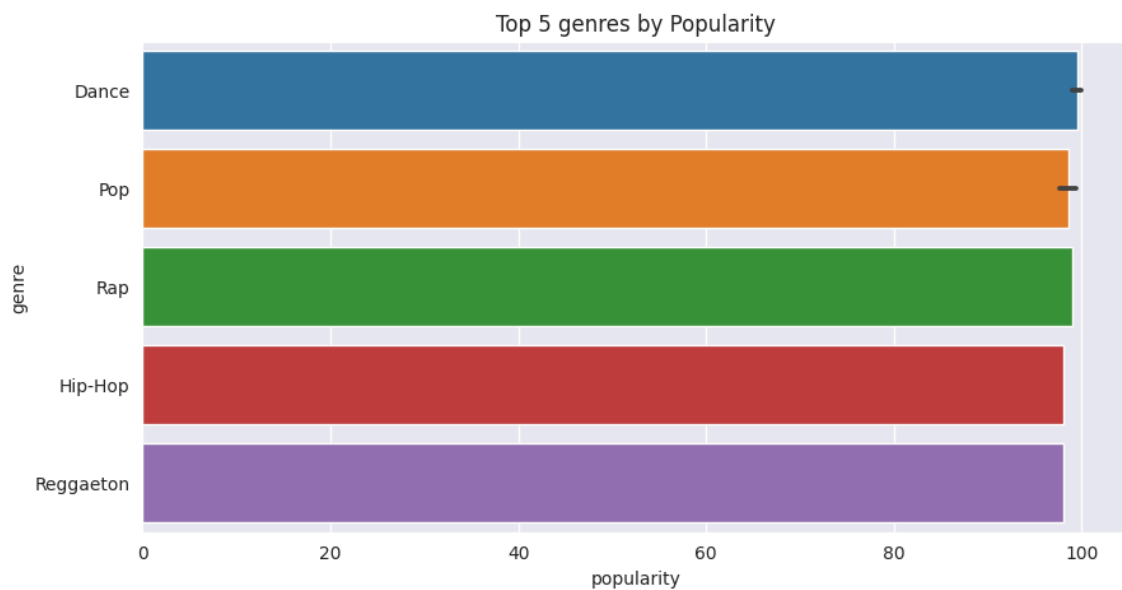
In [20]:
```
df_genre.head()
```

Out[20]:

	genre	artist_name	track_name	track_id	popularity	acousticness	dancea
0	Movie	Henri Salvador	C'est beau de faire un Show	0BRjO6ga9RKCKjfDqeFgWV	0	0.611	
1	Movie	Martin & les fées	Perdu d'avance (par Gad Elmaleh)	0BjC1NfoEOOusryehmNudP	1	0.246	
2	Movie	Joseph Williams	Don't Let Me Be Lonely Tonight	0CoSDzoNIKCRs124s9uTVy	3	0.952	
3	Movie	Henri Salvador	Dis-moi Monsieur Gordon Cooper	0Gc6TVm52BwZD07Ki6tIvf	0	0.703	
4	Movie	Fabien Nataf	Ouverture	0IuslXpMROHdEPvSl1fTQK	4	0.950	

```
In [21]: plt.title("Duration of the songs in different genre")
         sns.color_palette("rocket",as_cmap=True)
         sns.barplot(y='genre',x='duration_ms',data=df_genre)
         plt.xlabel("Duration in milliseconds")
         plt.ylabel('genre');
```
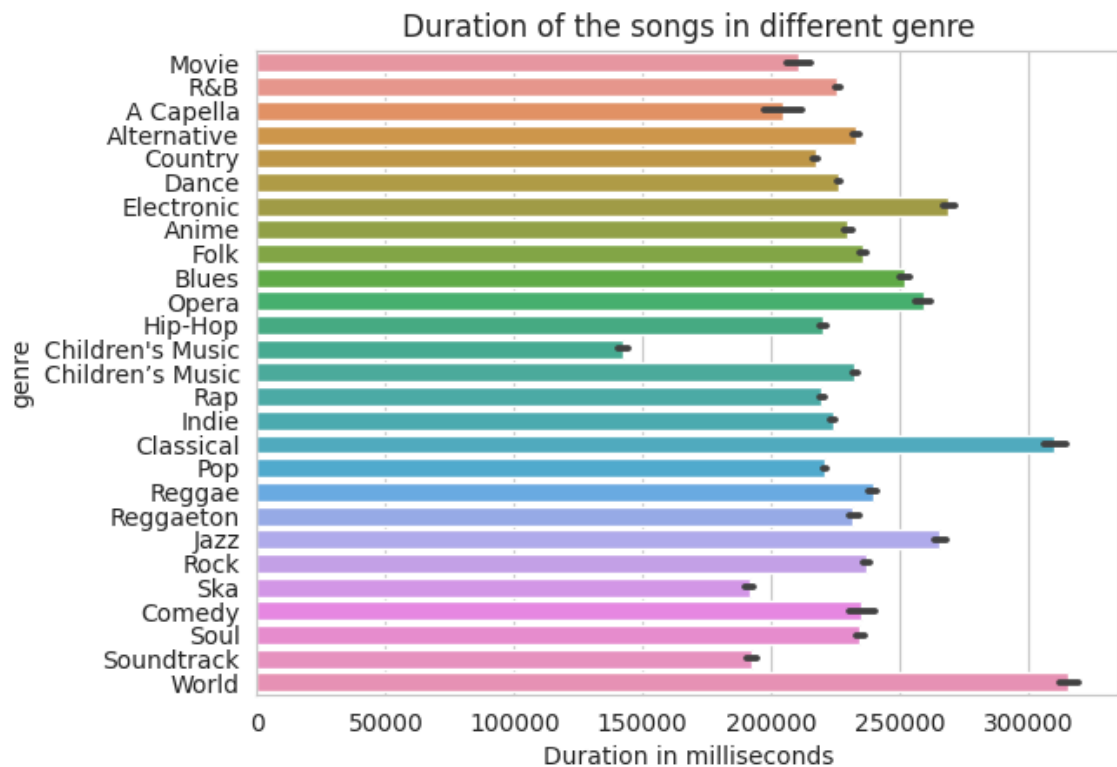


```
In [22]: sns.set_style(style='darkgrid')
         plt.figure(figsize=(10,5))
         famous=df_genre.sort_values('popularity',ascending=False).head(10)
         sns.barplot(y='genre',x='popularity',data=famous).set(title='Top 5 genres by Popularity');
```



```
In [ ]:
```