Import libraries

In [30]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sb
from warnings import filterwarnings
filterwarnings("ignore")
```

Import Data set

In [31]:
```
df_tracks=pd.read_csv("E:/Download/archive (2)/tracks.csv")
```

In [32]:
```
df_tracks.head()
```

Out[32]:

	id	name	popularity	duration_ms	explicit	artists	
0	35iwgR4jXetI318WEWsa1Q	Carve	6	126903	0	['Uli']	['45tIt06
1	021ht4sdgPcrDgSk7JTbKY	Capítulo 2.16 - Banquero Anarquista	0	98200	0	['Fernando Pessoa']	['14jtPCOoNZ
2	07A5yehtSnoedViJAZkNnc	Vivo para Quererte - Remasterizado	0	181640	0	['Ignacio Corsini']	['5LiOoJbxVS/
3	08FmqUhxtyLTn6pAh6bk45	El Prisionero - Remasterizado	0	176907	0	['Ignacio Corsini']	['5LiOoJbxVS/
4	08y9GfoqCWfOGsKdwojr5e	Lady of the Evening	0	163080	0	['Dick Haymes']	['3BiJGZsyX

Missing Data treatment

In [33]:
```
df_tracks.isna().sum()
```

Out[33]:
```
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 [34]:
```
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 [35]:
```
sorted_df=df_tracks.sort_values("popularity",ascending=True)
```

In [36]:
```
sorted_df.head(10)
```

Out[36]:

	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...	['3WCwCP '7
546222	0yOCz3V5KMm8l1T8EFc60i	恋は水の上で	0	188440	0	['Hibari Misora']	['1m5pM
546221	0y48Hhwe52099UqYjegRCO	私の誕生日	0	173467	0	['Hibari Misora']	['1m5pM
546220	0xCmgtf9ka07hkZg3D6PaV	エル・チョクロ (EL CHOCLO)	0	205280	0	['Hibari Misora']	['1m5pM
546219	0tBXS3VuCPX7KWUFH2nros	恋は不思議なもの	0	185733	0	['Hibari Misora']	['1m5pM
546218	0qrKnQtYDVJhKFAXTHYVS9	ゆうべはどうしたの (WHATSA MALLA U)	0	183427	0	['Hibari Misora']	['1m5pM
546217	0nqsDxOeKSwEzp3AUQAAqS	Screen Director's Playhouse, Music For Million...	0	1767071	0	['Wilms Herbert', 'June Allyson', 'Joseph Kear...	['2rbm8QV '4
546216	0kGEdsxVLYjCdfxM9tbezd	ブルーマンボ	0	162147	0	['Hibari Misora']	['1m5pM
546215	0bc3PUZurUUXrY7yqoOxjq	Screen Director's Playhouse, Trade Winds direc...	0	1776652	0	['Wally Maher', 'Tay Garnett', 'Lurene Tuttle'...	['7hkhJ '
546214	0Wwm0ruSjYMIiWG0nyAI1F	Screen Director's Playhouse, It's A Wonderful ...	0	1767576	0	['Joseph Granby', 'Jimmy Stewart', 'Irene Tedr...	['6GK59I

In [37]:
```
df_tracks.describe().T
```

Out[37]:

	count	mean	std	min	25%	50%	
popularity	586672.0	27.570053	18.370642	0.0	13.0000	27.000000	41
duration_ms	586672.0	230051.167286	126526.087418	3344.0	175093.0000	214893.000000	263867
explicit	586672.0	0.044086	0.205286	0.0	0.0000	0.000000	0
danceability	586672.0	0.563594	0.166103	0.0	0.4530	0.577000	0
energy	586672.0	0.542036	0.251923	0.0	0.3430	0.549000	0
key	586672.0	5.221603	3.519423	0.0	2.0000	5.000000	8
loudness	586672.0	-10.206067	5.089328	-60.0	-12.8910	-9.243000	-6
mode	586672.0	0.658797	0.474114	0.0	0.0000	1.000000	1
speechiness	586672.0	0.104864	0.179893	0.0	0.0340	0.044300	0
acousticness	586672.0	0.449863	0.348837	0.0	0.0969	0.422000	0
instrumentalness	586672.0	0.113451	0.266868	0.0	0.0000	0.000024	0
liveness	586672.0	0.213935	0.184326	0.0	0.0983	0.139000	0
valence	586672.0	0.552292	0.257671	0.0	0.3460	0.564000	0
tempo	586672.0	118.464857	29.764108	0.0	95.6000	117.384000	136
time_signature	586672.0	3.873382	0.473162	0.0	4.0000	4.000000	4

Most popular songs

In [38]:
```
most_popular=df_tracks.query("popularity>90",inplace=False).sort_values('popularity
```

In [39]:
```
most_popular[:10]
```

Out[39]:

	id	name	popularity	duration_ms	explicit	artists	
93802	4iJyoBOLtHqaGxP12qzhQI	Peaches (feat. Daniel Caesar & Giveon)	100	198082	1	['Justin Bieber', 'Daniel Caesar', 'Giveon']	['1uNFo '20
93803	7lPN2DXiMsVn7XUKtOW1CS	drivers license	99	242014	1	['Olivia Rodrigo']	['1McMsı
93804	3Ofmpyhv5UAQ70mENzB277	Astronaut In The Ocean	98	132780	0	['Masked Wolf']	['1uU7g3
92810	5QO79kh1waicV47BqGRL3g	Save Your Tears	97	215627	1	['The Weeknd']	['1Xyo4u8
92811	6tDDoYIxWvMLTdKpjFkc1B	telepatía	97	160191	0	['Kali Uchis']	['1U1el3
92813	0VjIjW4GlUZAMYd2vXMi3b	Blinding Lights	96	200040	0	['The Weeknd']	['1Xyo4u8
93805	7MAibcTli4IisCtbHKrGMh	Leave The Door Open	96	242096	0	['Bruno Mars', 'Anderson .Paak', 'Silk Sonic']	['0du5cE '3jK'
92814	6f3Slt0GbA2bPZlz0aIFXN	The Business	95	164000	0	['Tiësto']	['2o5jDh
91866	60ynsPSSKe6O3sfwRnIBRf	Streets	94	226987	1	['Doja Cat']	['5cj0lL
92816	3FAJ6O0NOHQV8Mc5Ri6ENp	Heartbreak Anniversary	94	198371	0	['Giveon']	['4fxd5E

In [40]:
```
#indexing by release date
df_tracks.set_index("release_date",inplace=True)
df_tracks.index=pd.to_datetime(df_tracks.index)
df_tracks.head()
```

Out[40]:

release_date	id	name	popularity	duration_ms	explicit	artists	
1922-02-22	35iwgR4jXetI318WEWsa1Q	Carve	6	126903	0	['Uli']	
1922-06-01	021ht4sdgPcrDgSk7JTbKY	Capítulo 2.16 - Banquero Anarquista	0	98200	0	['Fernando Pessoa']	['
1922-03-21	07A5yehtSnoedViJAZkNnc	Vivo para Quererte - Remasterizado	0	181640	0	['Ignacio Corsini']	['
1922-03-21	08FmqUhxtyLTn6pAh6bk45	El Prisionero - Remasterizado	0	176907	0	['Ignacio Corsini']	['
1922-01-01	08y9GfoqCWfOGsKdwojr5e	Lady of the Evening	0	163080	0	['Dick Haymes']	

In [41]:
```
df_tracks[["artists"]].iloc[20]
```

Out[41]:
```
artists    ['Jeanne Saint Bonnet']
Name: 1922-01-01 00:00:00, dtype: object
```

In [42]:
```
#convert milisecond in second

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

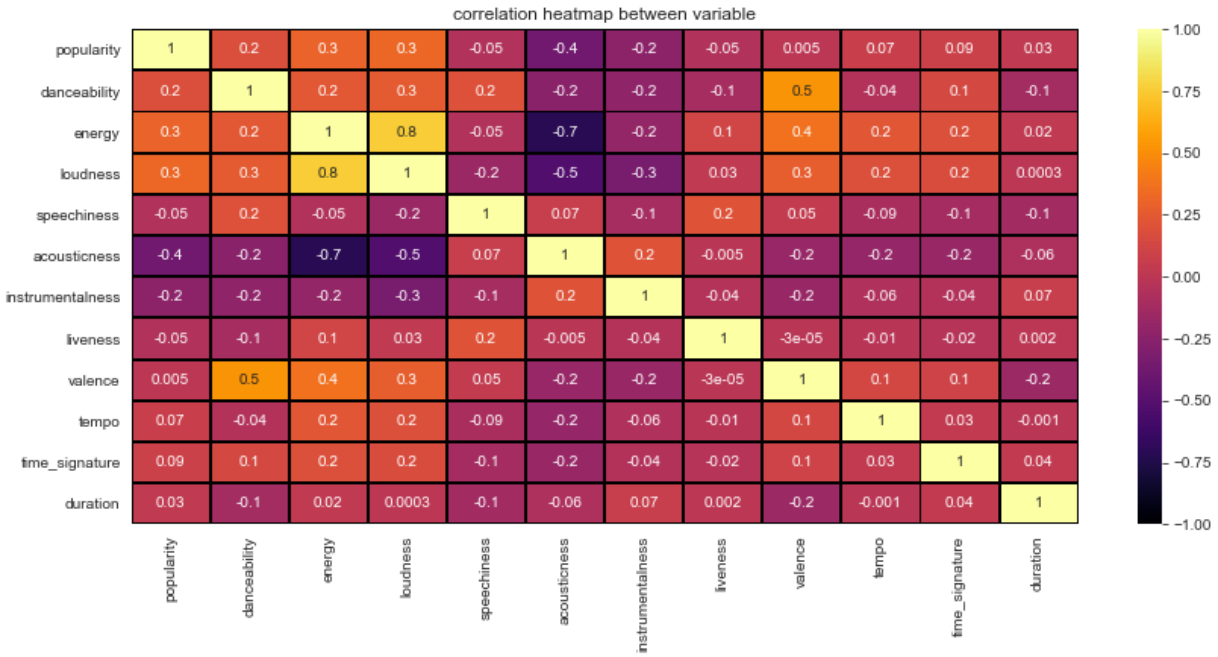
In [43]:
```
df_tracks.duration.head()
```

Out[43]:
```
release_date
1922-02-22    1269
1922-06-01     982
1922-03-21    1816
1922-03-21    1769
1922-01-01    1631
Name: duration, dtype: int64
```

EDA

In [44]:
```
corr_df=df_tracks.drop(["key","mode","explicit"],axis=1).corr(method="pearson")
plt.figure(figsize=(14,6))
heatmap=sb.heatmap(corr_df,annot=True,fmt=".1g",vmin=-1,vmax=1,center=0,cmap="inf
heatmap.set_title("correlation heatmap between variable")
#heatmap.set_xtickslabels(heatmap.get_xticables(),rotation=90)
```

Out[44]:
```
Text(0.5, 1.0, 'correlation heatmap between variable')
```



In [ ]:
```
#highly correlated columns     loudness ~ energy(0.8)
#low correlated colums        energy ~ acouusticness(-0.7)
```

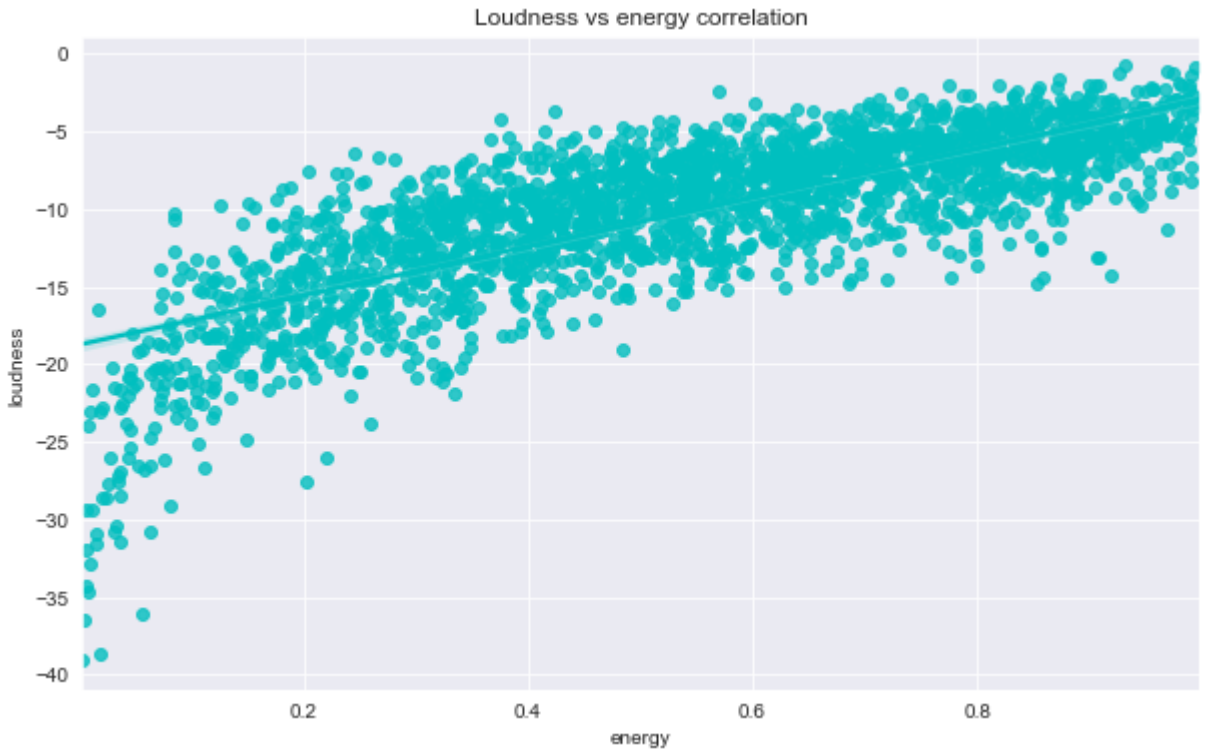
In [45]:
```
#.4% data
sample_df=df_tracks.sample(int(0.004*len(df_tracks)))
print(len(sample_df))
```

```
2346
```

Analysis using Regplot

In [46]:
```
plt.figure(figsize=(10,6))
sb.regplot(data=sample_df,y="loudness",x="energy",color="c").set(title="Loudness
```

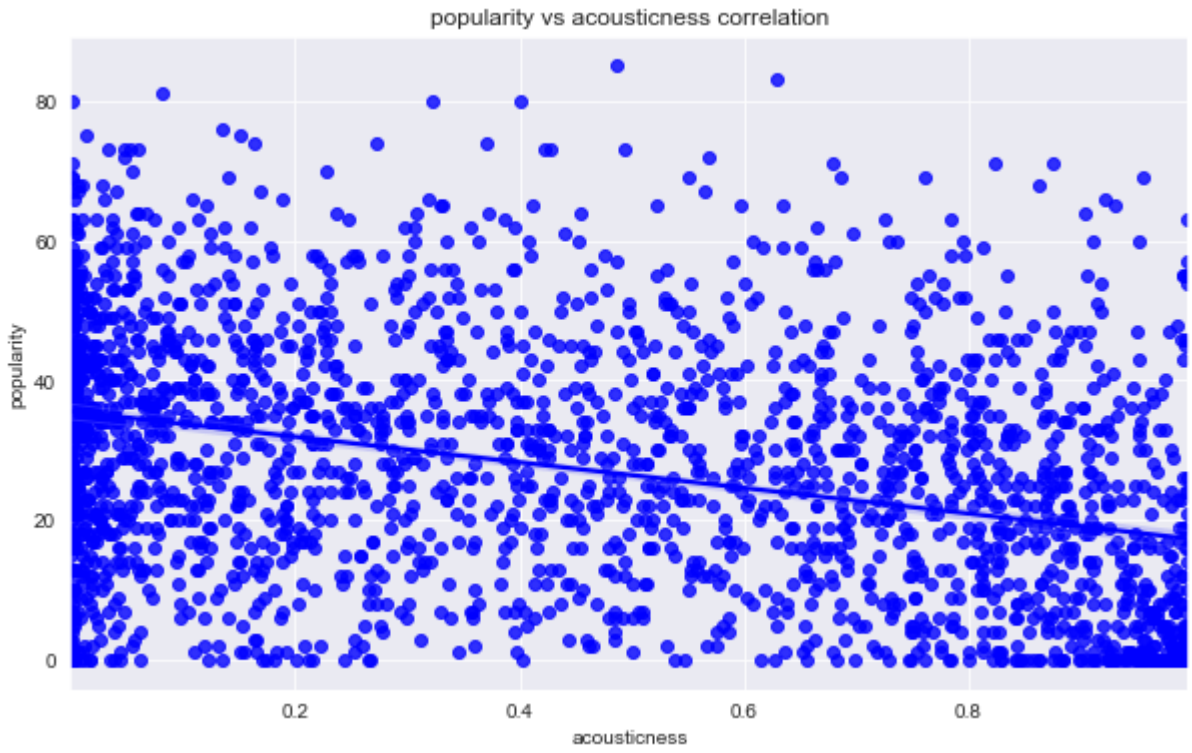
Out[46]:
```
[Text(0.5, 1.0, 'Loudness vs energy correlation')]
```



In [ ]:
```
#insights
#The loudness increases with energy.
```

In [47]:
```
plt.figure(figsize=(10,6))
sb.regplot(data=sample_df,y="popularity",x="acousticness",color="b").set(title="p
```

Out[47]:
```
[Text(0.5, 1.0, 'popularity vs acousticness correlation')]
```



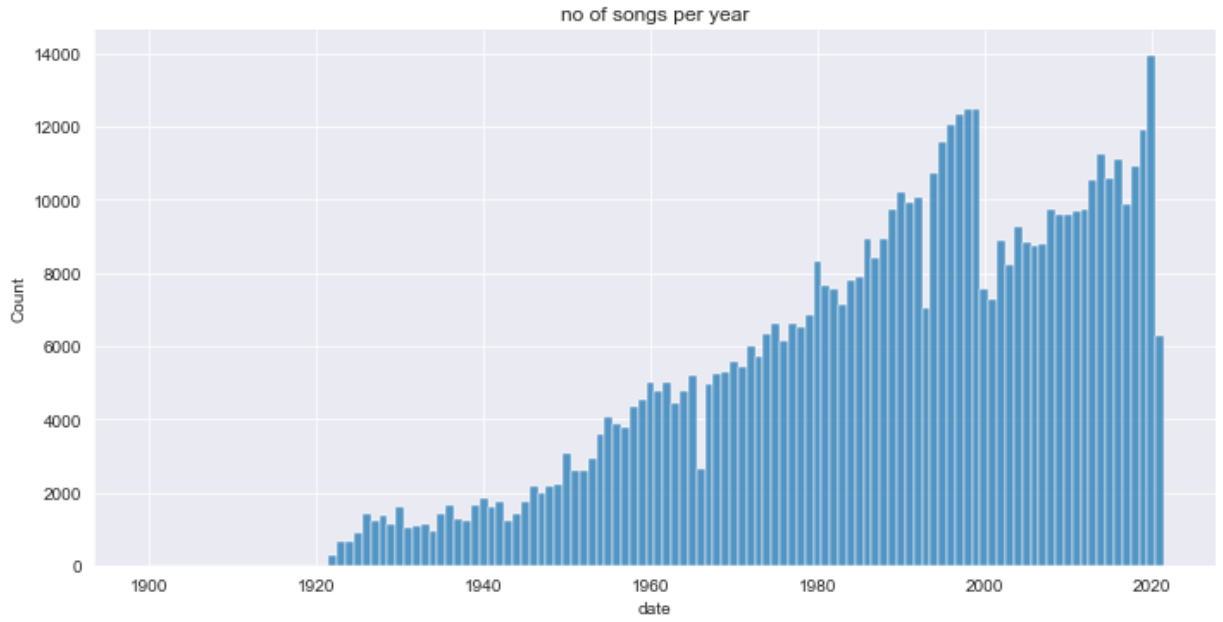
In [ ]:
```
#insights
#acousticness decreases with increasing popularity.
```

Analysis using distplot

In [48]:
```
df_tracks['date']=df_tracks.index.get_level_values('release_date')
df_tracks.dates=pd.to_datetime(df_tracks.date)
years=df_tracks.dates.dt.year
```

In [49]:
```
sb.displot(years,discrete=True,aspect=2,height=5,kind="hist").set(title="no of so
```

Out[49]: <seaborn.axisgrid.FacetGrid at 0x25c10d3eb50>



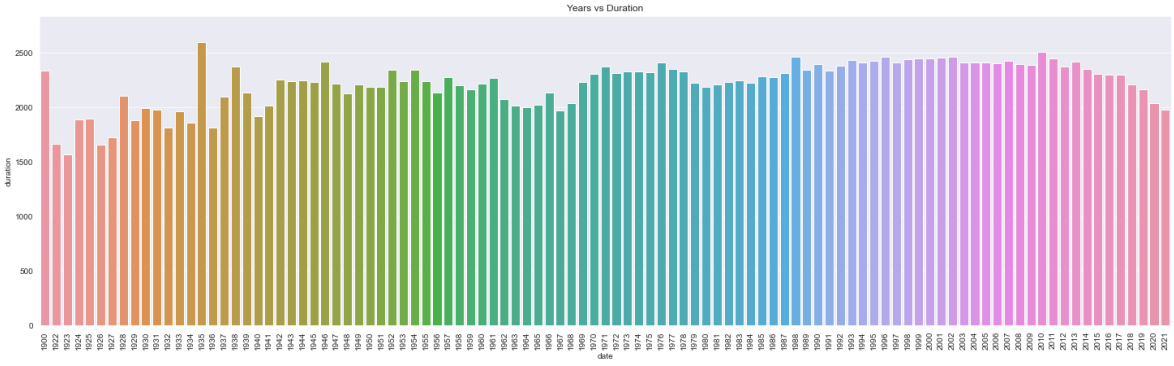
In [ ]:
```
#insights
#The number of songs per year increases over time.
```

Analysis using Barplot

In [50]:
```
total_dr=df_tracks.duration
#insightsfig_dims=(25,7)
fig, ax=plt.subplots(figsize=fig_dims)
fig= sb.barplot(x=years,y=total_dr,ax=ax,errwidth=False).set(title="Years vs Dura
plt.xticks(rotation=90)
```

```
  Text(93, 0, '2014'),
  Text(94, 0, '2015'),
  Text(95, 0, '2016'),
  Text(96, 0, '2017'),
  Text(97, 0, '2018'),
  Text(98, 0, '2019'),

  Text(99, 0, '2020'),
  Text(100, 0, '2021')])
```

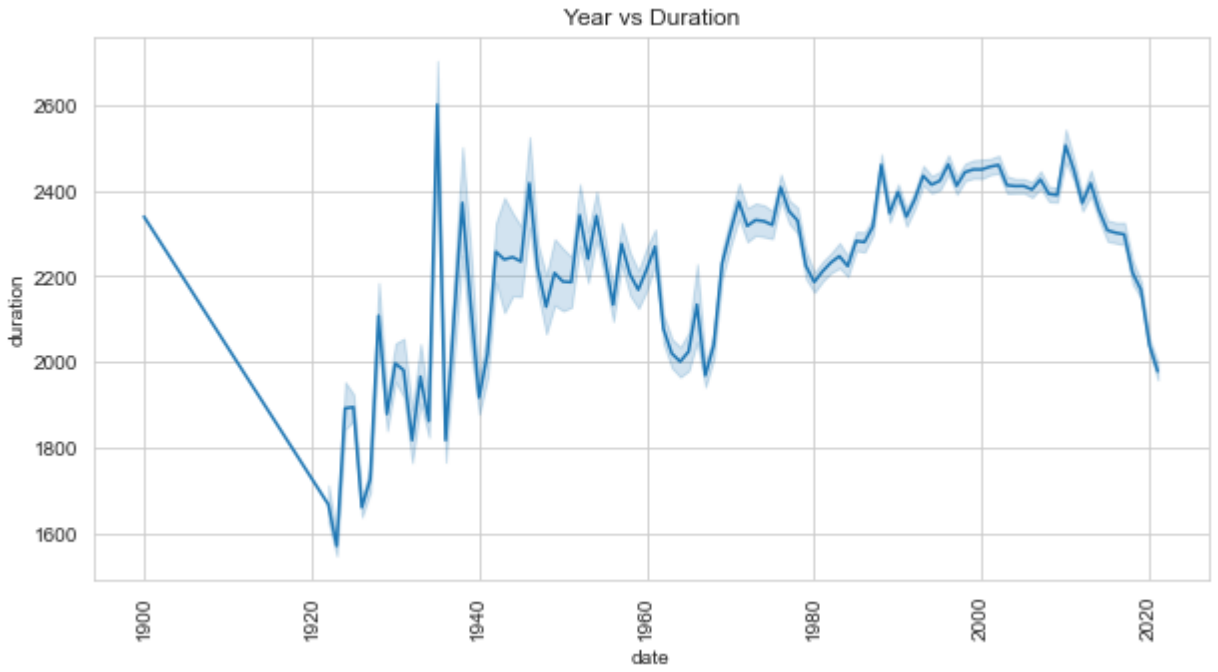


In [ ]:
```
#duration of flactuated songs
#It gradually increased at first, then remained constant for a while before begir
```

Analysis using lineplot

In [51]:
```
total_dr=df_tracks.duration
sb.set_style(style='whitegrid')
fig_dims=(10,5)
fig, ax=plt.subplots(figsize=fig_dims)
fig=sb.lineplot(x=years,y=total_dr,ax=ax).set(title="Year vs Duration")
plt.xticks(rotation=90)
```

Out[51]:
```
(array([1880., 1900., 1920., 1940., 1960., 1980., 2000., 2020., 2040.]),
 [Text(0, 0, ''),
  Text(0, 0, ''),
  Text(0, 0, ''),
  Text(0, 0, ''),
  Text(0, 0, ''),
  Text(0, 0, ''),
  Text(0, 0, ''),
  Text(0, 0, ''),
  Text(0, 0, '')])
```



In [ ]:

In [52]:
```
df_genre=pd.read_csv("E:/Download/archive (1)/SpotifyFeatures.csv")
```

In [53]:
```
df_genre.head()
```

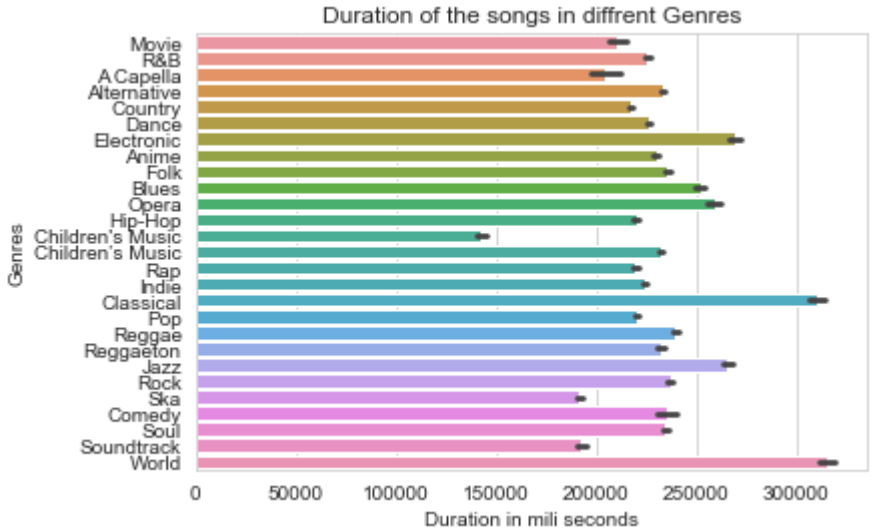
Out[53]:

	genre	artist_name	track_name	track_id	popularity	acousticness	danceability
0	Movie	Henri Salvador	C'est beau de faire un Show	0BRjO6ga9RKCKjfDqeFgWV	0	0.611	0.389
1	Movie	Martin & les fées	Perdu d'avance (par Gad Elmaleh)	0BjC1NfoEOOusryehmNudP	1	0.246	0.590
2	Movie	Joseph Williams	Don't Let Me Be Lonely Tonight	0CoSDzoNIKCRs124s9uTVy	3	0.952	0.663
3	Movie	Henri Salvador	Dis-moi Monsieur Gordon Cooper	0Gc6TVm52BwZD07Ki6tIvf	0	0.703	0.240
4	Movie	Fabien Nataf	Ouverture	0IuslXpMROHdEPvSl1fTQK	4	0.950	0.331

Barplot analysis

In [54]:
```
plt.title("Duration of the songs in diffrent Genres")
sb.color_palette("rocket",as_cmap=True)
sb.barplot(y='genre',x='duration_ms',data=df_genre)
plt.xlabel("Duration in mili seconds")
plt.ylabel("Genres")
```

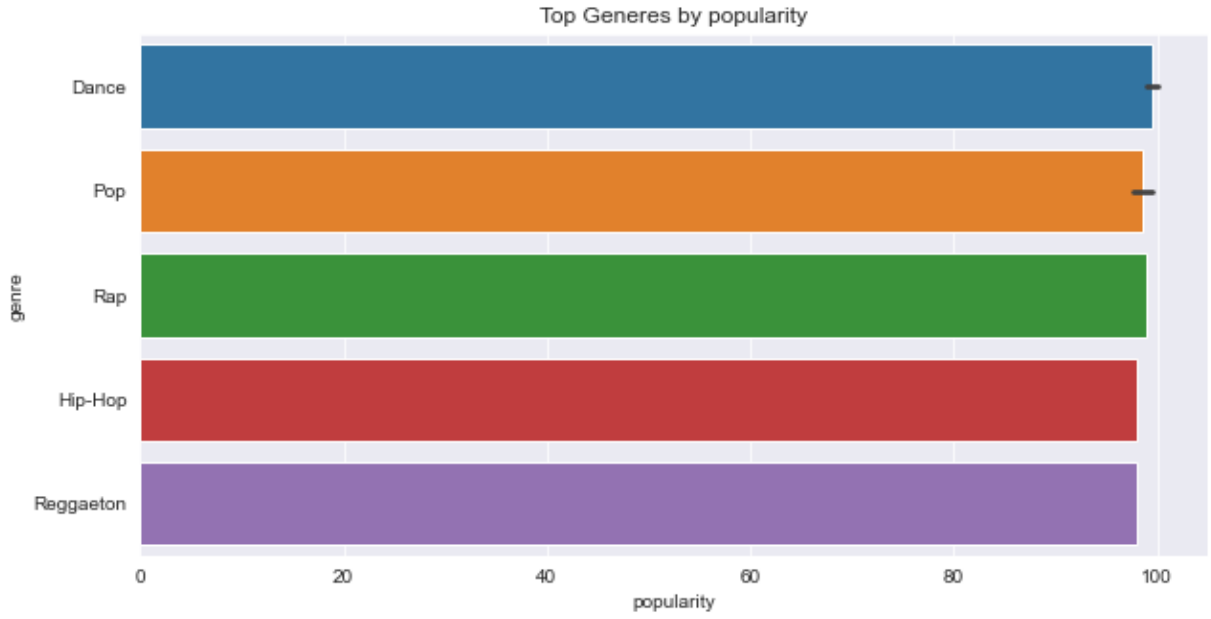
Out[54]:
```
Text(0, 0.5, 'Genres')
```



In [ ]:
```
#insights
#The duration of the songs decreases over the period of time.
```

In [55]:
```
sb.set_style(style="darkgrid")
plt.figure(figsize=(10,5))
famous=df_genre.sort_values("popularity",ascending=False).head(10)
sb.barplot(y='genre',x='popularity',data=famous).set(title="Top Generes by popula
```

Out[55]:
```
[Text(0.5, 1.0, 'Top Generes by popularity')]
```



In [ ]:
```
#insights
#Topest gener is Dance
#Top 5th is Reggaeton
```

In [ ]:

In [ ]: