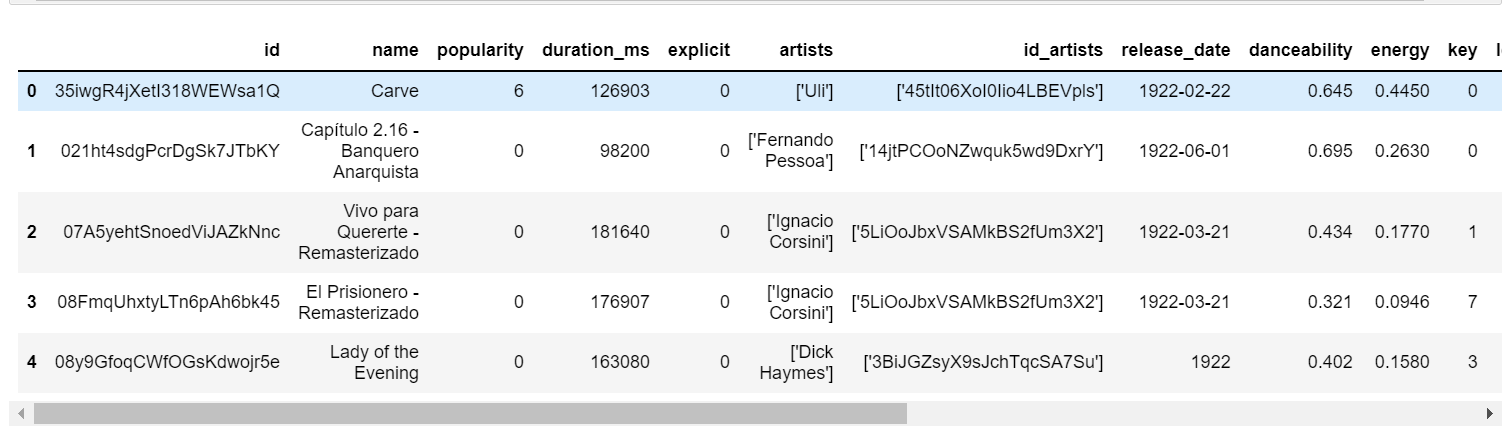
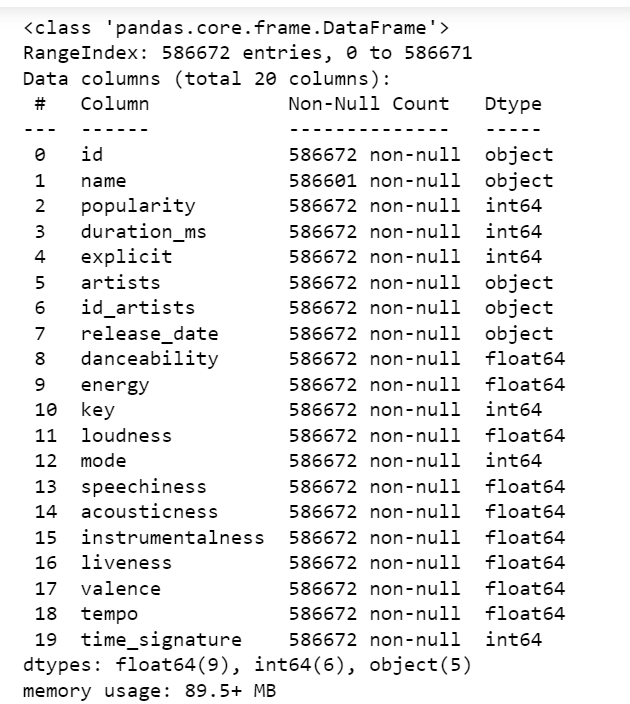
1. checking top 5 rows

code - df\_tracks.head()



1. checking data info

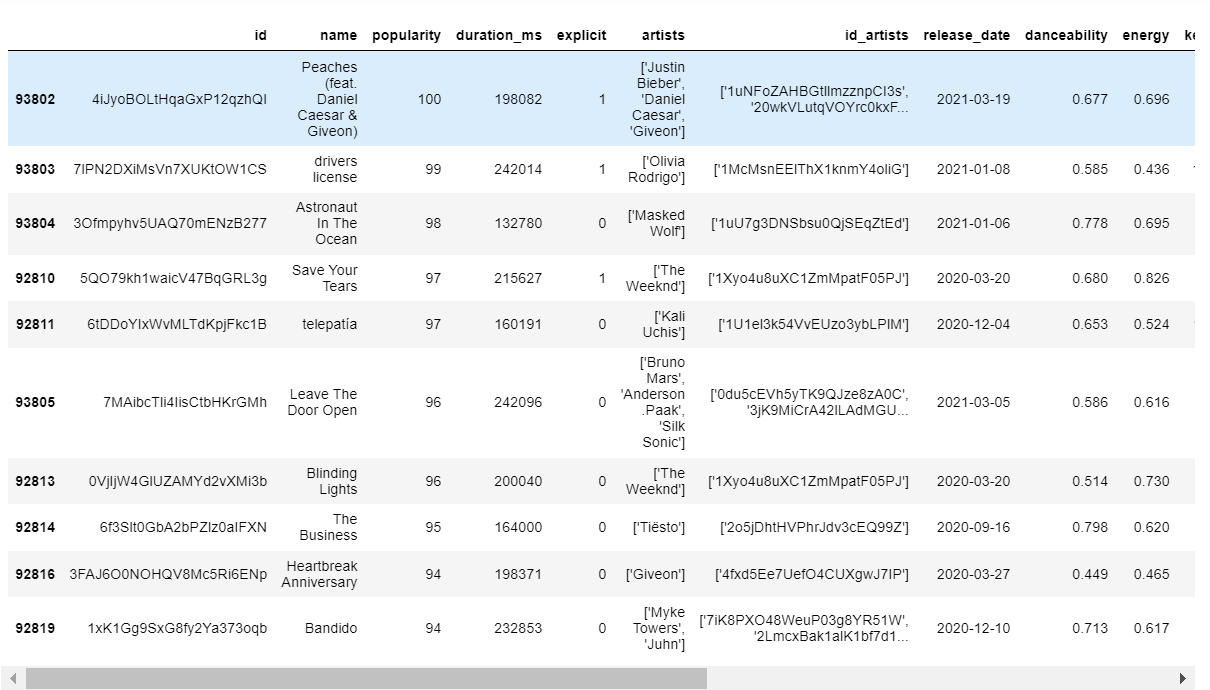
code - df\_tracks.info()



1. Selecting 10 most popular songs present on spotify

Code - most\_popular = df\_tracks.sort\_values(by = 'popularity',ascending=False).head(10)

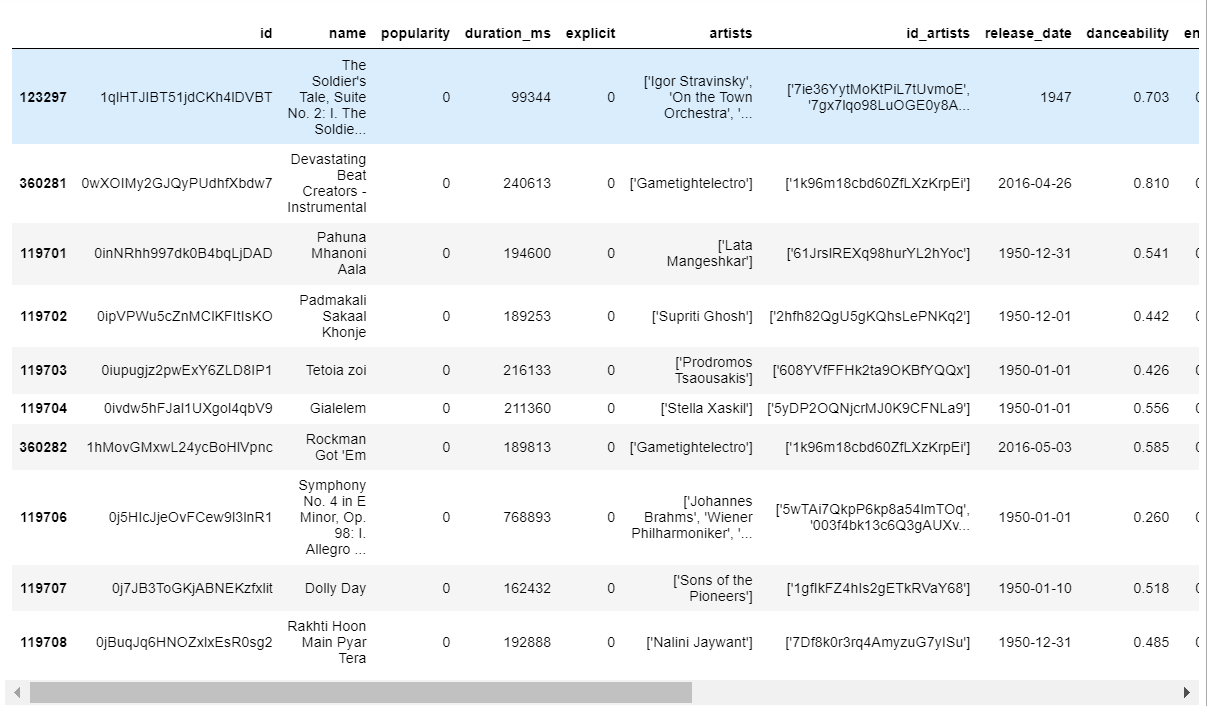
most\_popular



1. Least popular 10 songs present on spotify having popularity < 10

Code - least\_popular = df\_tracks.query('popularity<10', inplace=False).sort\_values('popularity', ascending=True)

least\_popular[:10]

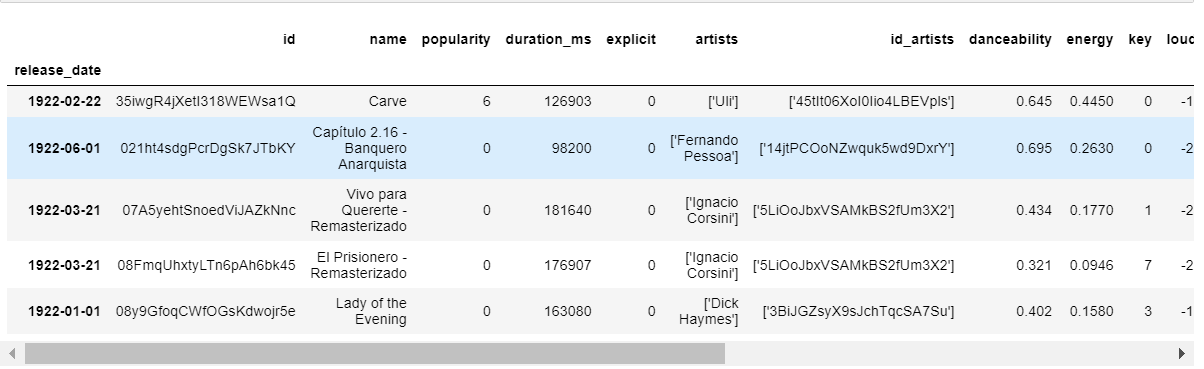


1. setting 'release\_date' column to index and changing it's datatype to datetime

code - df\_tracks\_new = df\_tracks.set\_index('release\_date')

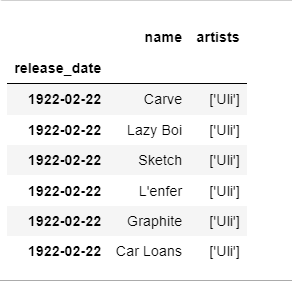
df\_tracks\_new.index = pd.to\_datetime(df\_tracks\_new.index)

df\_tracks\_new.head()



1. selecting the songs released on '1922-02-22'

code - df\_tracks\_new.loc['1922-02-22', ['name','artists']]

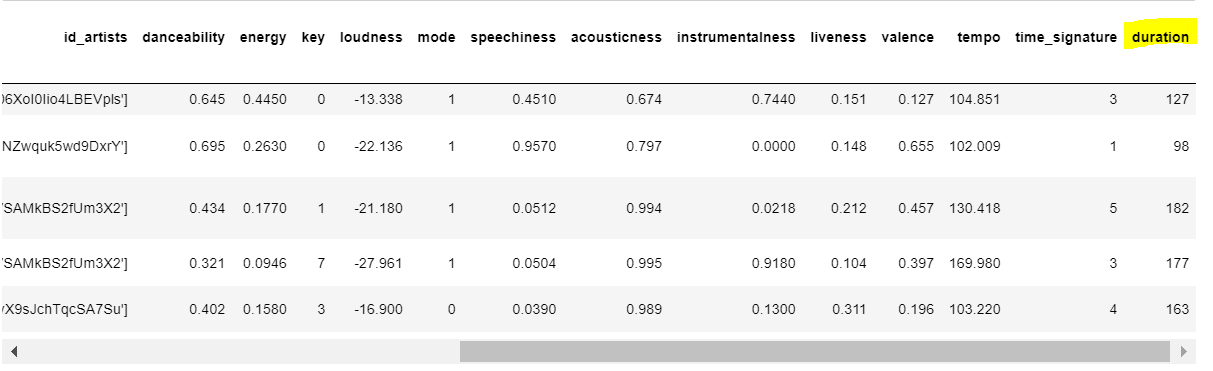


1. create a new column duration for time in seconds from column duration\_ms

code - df\_tracks\_new["duration"] = df\_tracks\_new["duration\_ms"].apply(lambda x: round(x/1000))

df\_tracks\_new.drop('duration\_ms', inplace=True, axis=1)

df\_tracks\_new.head()



1. Taking a sample of 1% of total data and plotting a scatter plot b/w loudness and energy

code - sample\_df = df\_tracks\_new.sample(int(0.01\*len(df\_tracks\_new)))

print(len(sample\_df))

fig = plt.figure(figsize = [10,5])

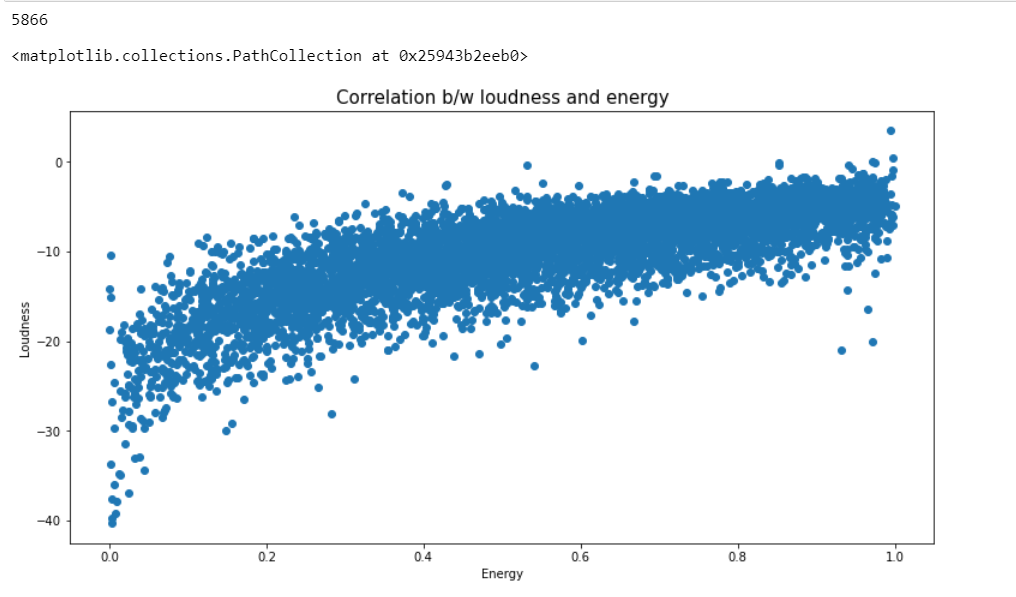
ax = fig.add\_axes([0,0,1,1])

ax.set\_title("Correlation b/w loudness and energy",fontsize=15)

ax.set\_xlabel("Energy")

ax.set\_ylabel("Loudness")

ax.scatter(x = sample\_df['energy'], y = sample\_df['loudness'])



1. plot a histogram for no of songs released per year

code **-** df\_tracks\_new = df\_tracks\_new.reset\_index()

df\_tracks\_new['year'] = df\_tracks\_new['release\_date'].dt.year

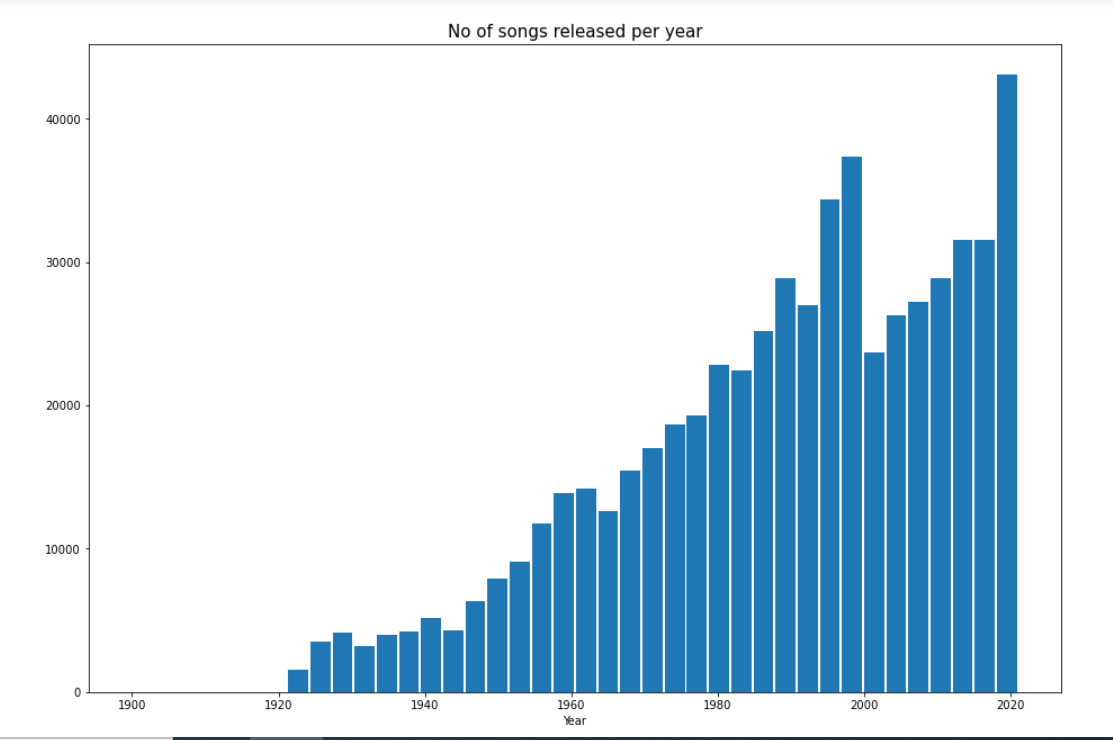
fig = plt.figure(figsize=[12,8])

ax = fig.add\_axes([0,0,1,1])

ax.set\_title("No of songs released per year",fontsize=15)

ax.set\_xlabel("Year")

ax.hist(df\_tracks\_new['year'],bins=40, rwidth=0.9)



1. plot a line chart for average duration of songs over the years

code - dr = df\_tracks\_new.duration

fig = plt.figure(figsize=[10,6])

ax = fig.add\_axes([0,0,1,1])

ax.set\_title("Duration VS year",fontsize=15)

fig = sns.lineplot(x=df\_tracks\_new.year , y=dr)

