

```
In [1]: import pandas as pd
```

```
In [2]: #Comencem convertint el csv en un dataframe per treballar amb pandas

# csv to df of only 10.000 first songs
msd_df = pd.read_csv('./datasets/msd_reduced.csv', index_col=0)

#During the csv making process, I was merging different dfs into one df and df index info was lost. Let's reset
msd_df.reset_index(drop=True,inplace=True)

print ("Number of Rows: %d"%len(msd_df))
print ("Number of Columns: %d"%len(msd_df.columns))
pd.set_option('display.max_columns', None)
display(msd_df)
```

Number of Rows: 926096
Number of Columns: 20

	danceability	duration	end_of_fade_in	energy	key	loudness	mode	start_of_fade_out	tempo	time_signature	artist_fa
0	0.0	252.05506	2.049	0.0	10	-4.829	0	236.635	87.002	4	0
1	0.0	156.55138	0.258	0.0	9	-10.555	1	148.660	150.778	1	0
2	0.0	138.97098	0.000	0.0	7	-2.060	1	138.971	177.768	4	C
3	0.0	145.05751	0.000	0.0	7	-4.654	1	138.687	87.433	4	C
4	0.0	514.29832	0.000	0.0	5	-7.806	0	506.717	140.035	4	0
...
926091	0.0	185.33832	0.000	0.0	1	-5.968	0	170.974	120.009	4	0
926092	0.0	217.44281	0.000	0.0	7	-8.339	0	217.443	92.159	7	C
926093	0.0	244.16608	3.048	0.0	5	-12.646	1	240.089	156.132	3	C
926094	0.0	553.03791	0.223	0.0	11	-8.218	0	548.989	137.089	4	C
926095	0.0	295.07873	0.000	0.0	6	-5.801	1	292.182	164.037	4	C

926096 rows x 20 columns

```
In [3]: # Cerquem els artistes de la llista de l'Esther, amb un hottnesss alt
artist_list = ["Johnny Cash", "The Rolling Stones", "Bruce Springsteen", "Ray Charles", "Muse", "Céline Dion",
artists_song_list = msd_df.loc[msd_df['artist_name'].isin(artist_list)]
artists_song_list = artists_song_list.drop_duplicates(subset=['title', 'artist_name'], keep='first')
# display
print ("Number of Rows: %d"%len(artists_song_list))
print ("Number of Columns: %d"%len(artists_song_list.columns))
pd.set_option('display.max_columns', None)
display(artists_song_list)
```

Number of Rows: 3107
Number of Columns: 20

	danceability	duration	end_of_fade_in	energy	key	loudness	mode	start_of_fade_out	tempo	time_signature	artist_fa
178	0.0	223.94730	0.316	0.0	7	-7.417	1	216.160	119.301	4	C
904	0.0	247.61424	2.798	0.0	2	-7.744	1	233.616	120.583	4	C
1001	0.0	110.52363	0.096	0.0	4	-16.179	1	101.994	101.793	4	C
1151	0.0	317.90975	2.862	0.0	1	-7.714	1	303.386	90.024	4	C
1161	0.0	324.88444	0.159	0.0	11	-9.383	0	316.674	129.272	4	C
...
924484	0.0	217.44281	0.183	0.0	7	-7.515	0	214.332	87.010	4	C
925238	0.0	93.36118	0.177	0.0	9	-16.960	0	90.720	69.131	4	C
925446	0.0	675.91791	0.000	0.0	10	-8.901	0	671.057	92.820	3	C
925702	0.0	147.93098	0.171	0.0	9	-5.114	1	140.463	86.709	3	C
925809	0.0	246.49098	0.334	0.0	0	-14.073	1	234.672	70.055	3	C

3107 rows x 20 columns

```
In [4]: print(len(artists_song_list['artist_name'].unique()))

31
```

```
In [5]: # export to csv file
artists_song_list.to_csv("artists_hotness.csv")
```