



Top Hits in

SPOTIFY

#



Barrientos Ontiveros Alma Alejandra



Introduction

The data base contains information of some characteristics of the top hits in Spotify.





Variables

artist

Name of the song's artist

song

Song name

danceability

How suitable is the song to be danced?

duration_ms

Song duration

explicit

Does the song contain offensive words?

year

Launch year

popularity


Popularity of the song

key

Key in which the song is composed

energy

Intensity and activity of the song





Variables

loudness

Loudness of the song

mode

Is the key major or
minor?

instrumentalness

How much of the song is instrumental?

acousticness

How acoustic is the song?

liveness

Presence of an audience in the
recording

valence

Positiveness of a track



speechiness

How much of the song
contains voice

tempo

Tempo of the song

genre

Genre of the song

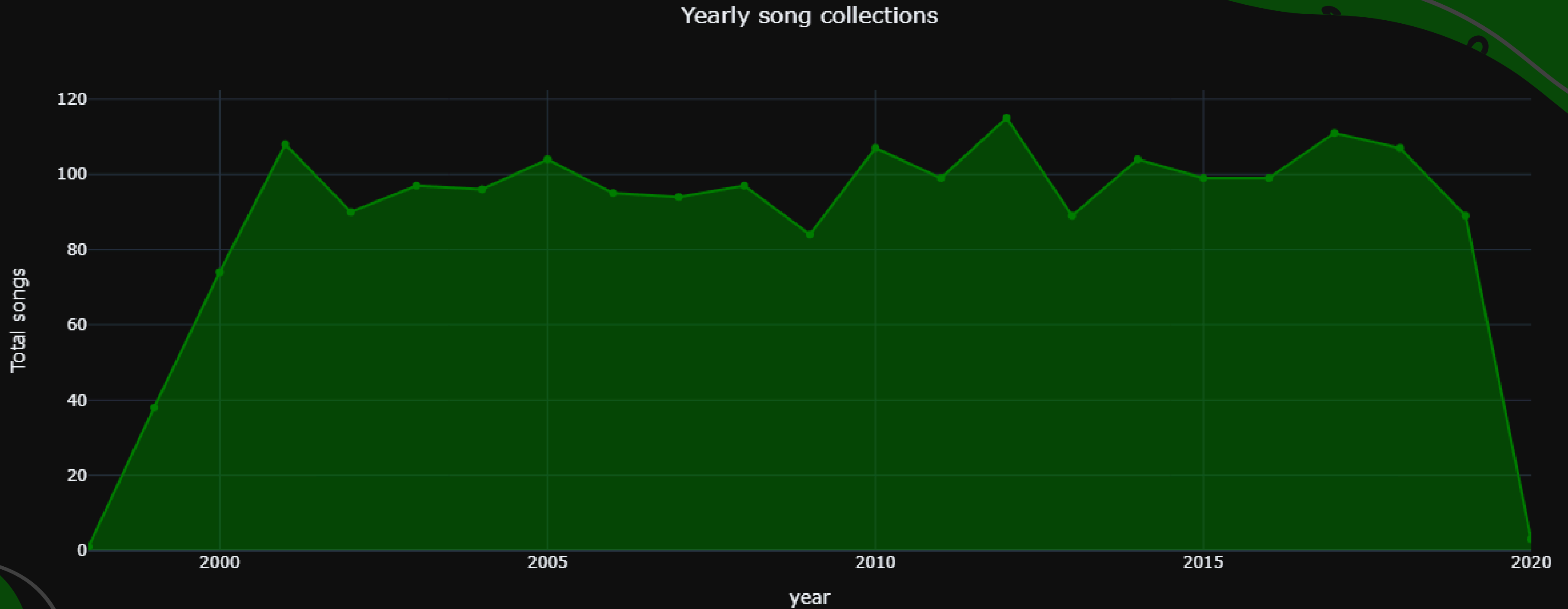


Of what type are the variables?



Three of the variables are objects (artist, song and genre), five are integers (duration_ms, year, popularity, key and mode), nine are floats (danceability, energy, loudness, speechiness, acousticness, instrumentalness, liveness, valence and tempo), and one is a boolean (explicit).





Here we can see how many songs from each year we have. For example

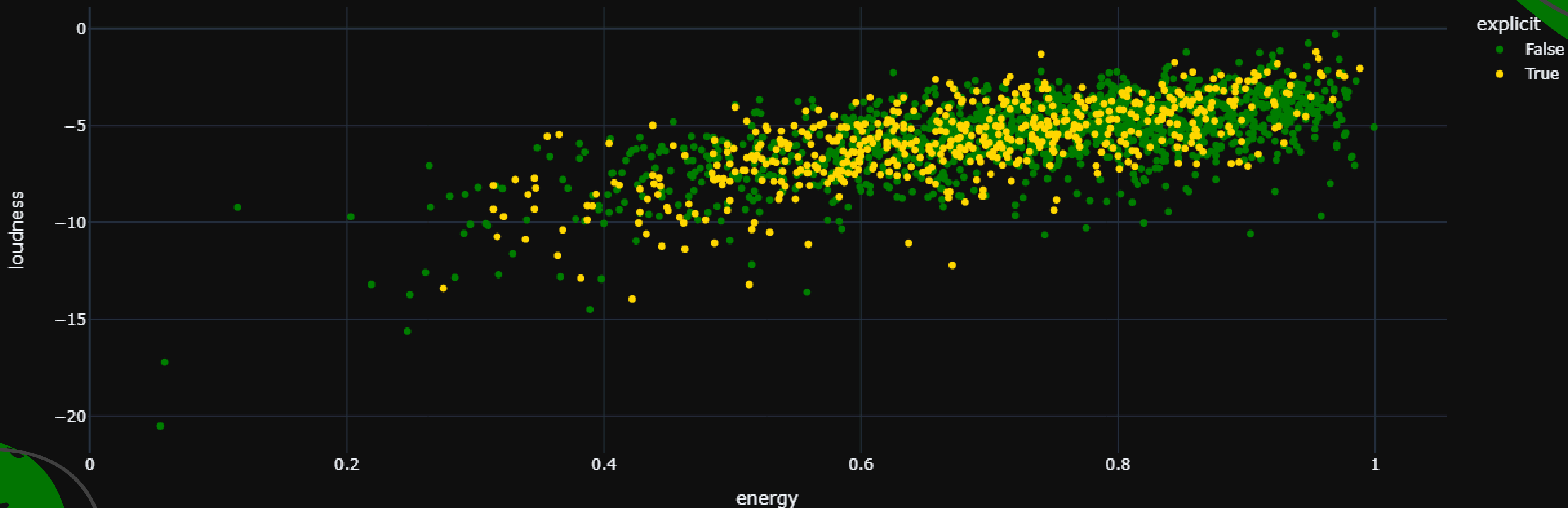


Correlation matrix of columns

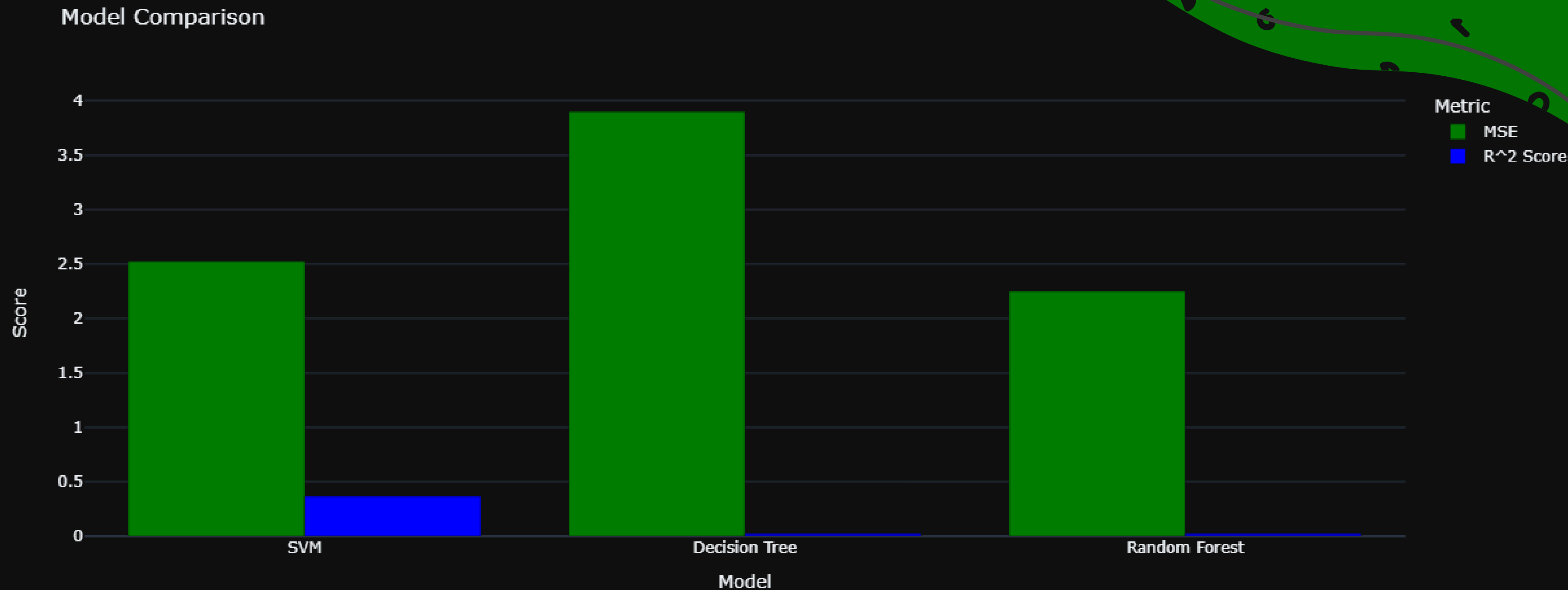
duration_ms	1	-0.3165341	0.05061722	-0.06005683	-0.07876307	-0.002560232	-0.07991182	-0.003847828	0.06699756	0.01092343	-0.00420788	0.02494076	-0.1168698	-0.02860281
year	-0.3165341	1	-0.003825249	0.03353246	-0.1086444	0.007380354	0.01747921	-0.007357771	0.001110708	0.03380855	-0.05026517	-0.0270369	-0.2093645	0.07686651
popularity	0.05061722	-0.003825249	1	-0.00354573	-0.01402125	0.01482272	0.03063232	-0.02135336	0.02116235	0.02461917	-0.04805923	-0.00985576	-0.01614206	0.01428847
danceability	-0.06005683	0.03353246	-0.00354573	1	-0.1040384	0.03273146	-0.03331486	-0.06752847	0.1455897	-0.06542889	0.02320731	-0.1264129	0.4031776	-0.173418
energy	-0.07876307	-0.1086444	-0.01402125	-0.1040384	1	-0.003445685	0.6510155	-0.04065103	-0.05701819	-0.4454685	0.03786112	0.1567607	0.3344744	0.1537191
key	-0.002560232	0.007380354	0.01482272	0.03273146	-0.003445685	1	-0.00747439	-0.1531822	0.007147041	0.002365349	-0.008173222	-0.03307084	0.03697702	-0.001430693
loudness	-0.07991182	0.01747921	0.03063232	-0.03331486	0.6510155	-0.00747439	1	-0.02813256	-0.07638819	-0.3100392	-0.1049246	0.102159	0.2321504	0.0807088
mode	-0.003847828	-0.007357771	-0.02135336	-0.06752847	-0.04065103	-0.1531822	-0.02813256	1	-76.83285μ	0.005744475	-0.03861291	0.02543871	-0.0746809	0.04843437
speechiness	0.06699756	0.001110708	0.02116235	0.1455897	-0.05701819	0.007147041	-0.07638819	-76.83285μ	1	394.1595μ	-0.06295384	0.06117214	0.073605	0.05774747
acousticness	0.01092343	0.03380855	0.02461917	-0.06542889	-0.4454685	0.002365349	-0.3100392	0.005744475	394.1595μ	1	-0.00521442	-0.1100431	-0.1281279	-0.10366
instrumentalness	-0.00420788	-0.05026517	-0.04805923	0.02320731	0.03786112	-0.008173222	-0.1049246	-0.03861291	-0.06295384	-0.00521442	1	-0.03489654	-0.0151915	0.03460754
liveness	0.02494076	-0.0270369	-0.00985576	-0.1264129	0.1567607	-0.03307084	0.102159	0.02543871	0.06117214	-0.1100431	-0.03489654	1	0.01903963	0.02863569
valence	-0.1168698	-0.2093645	-0.01614206	0.4031776	0.3344744	0.03697702	0.2321504	-0.0746809	0.073605	-0.1281279	-0.0151915	0.01903963	1	-0.02507583
tempo	-0.02860281	0.07686651	0.01428847	-0.173418	0.1537191	-0.001430693	0.0807088	0.04843437	0.05774747	-0.10366	0.03460754	0.02863569	-0.02507583	1
	duration_ms	year	popularity	danceability	energy	key	loudness	mode	speechiness	acousticness	instrumentalness	liveness	valence	tempo

Note that the higher correlation is obtained between the variables energy and loudness.

Distribution of Energy VS Loudness



We can appreciate the joint distribution of energy and loudness with a color distinction between explicit and non explicit songs.



As we can see, the lowest MSE is from random forest, where we have our better prediction accuracy

	Model	MSE	R ² Score
0	SVM	2.524004	0.368281
1	Decision Tree	3.900475	0.023771
2	Random Forest	2.248623	0.023771