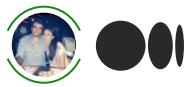


[Open in app](#)

Arda Baysallar

27 Followers [About](#)

Spotify 2019 EDA



Arda Baysallar Just now · 4 min read

My aim on this page is to prepare a spotify 2019 top 50 by countries dashboard that is available in Kaggle.

First I would like to introduce the Kaggle data :

Here are the column names and their explainings :

title = Song's title

artist = Song's artist

top genre = the genre of the track

year = Song's year

beats_per_minute = Beats per minute — The tempo of the song

Energy = Energy of a song — the higher the value, the more energetic song

Danceability = The higher the value, the easier it is the dance to this song

Loudness_dB = Loudness dB higher the value louder the song

Liveness = the higher the value, the more likely the song is a live recording

Valence = The higher the value, the more positive mood for the song.

Length = Duration : Length of the song

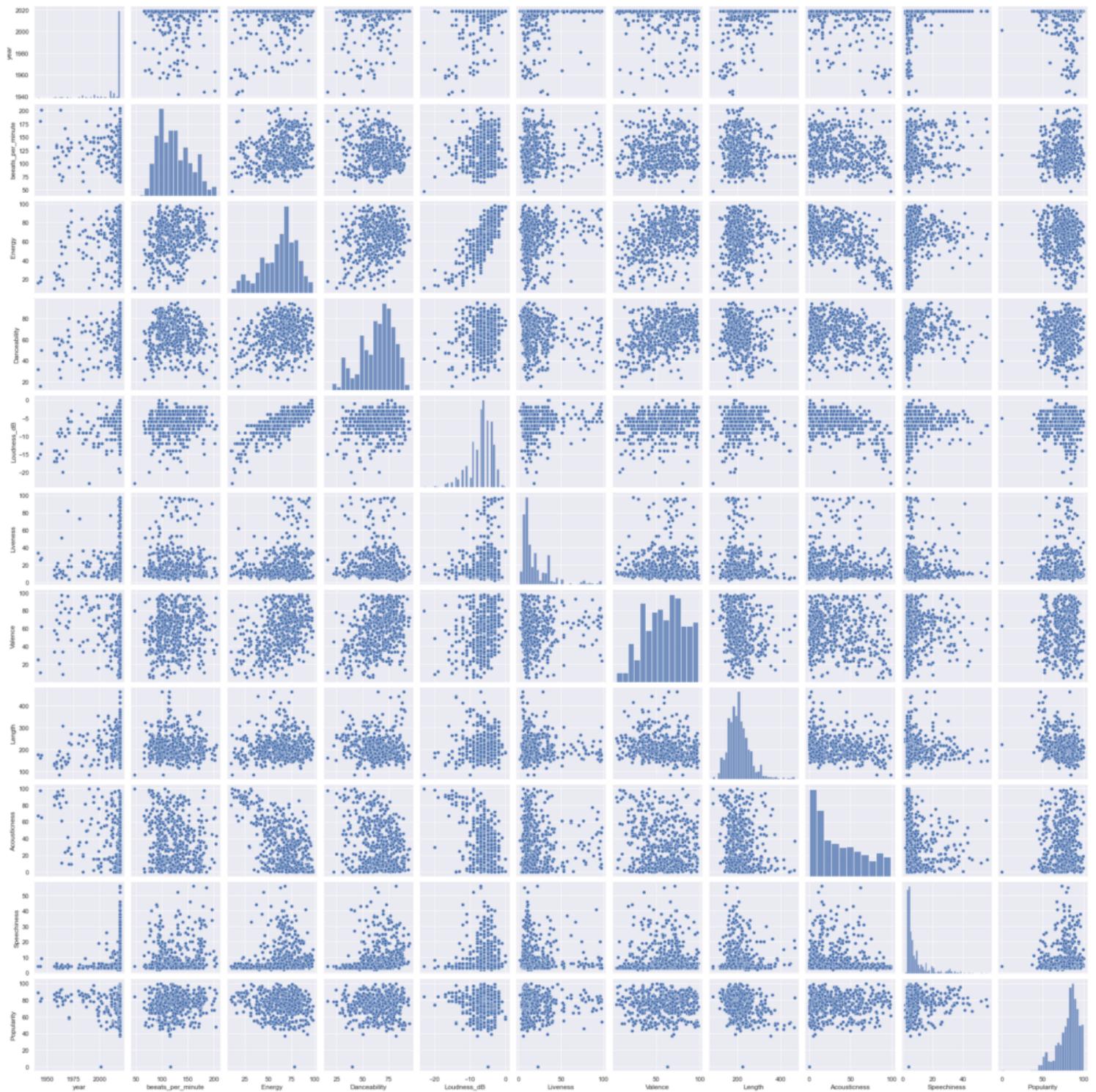
Acousticness = The higher the value the more acoustic the song is

Speechiness = The higher the value the more spoken word the song contains

Popularity = The higher the value the more popular the song is

country = Country where the son was famous

I would like to first start with introducing the data. It is nice and clean data with very few missing rows. It has skewed features here is the pairplot chart we can see speechiness, liveness, loudness_dB and popularity features show skewed data shape :



Pairplot chart

We can see high acousticness songs likely to have a lower energy level and Loud musics

are generally tends to have high energy level. In a way energy level can interpret loudness and accousticness

Moreover popularity tends to be affected negatively from length of the song with a small proportion

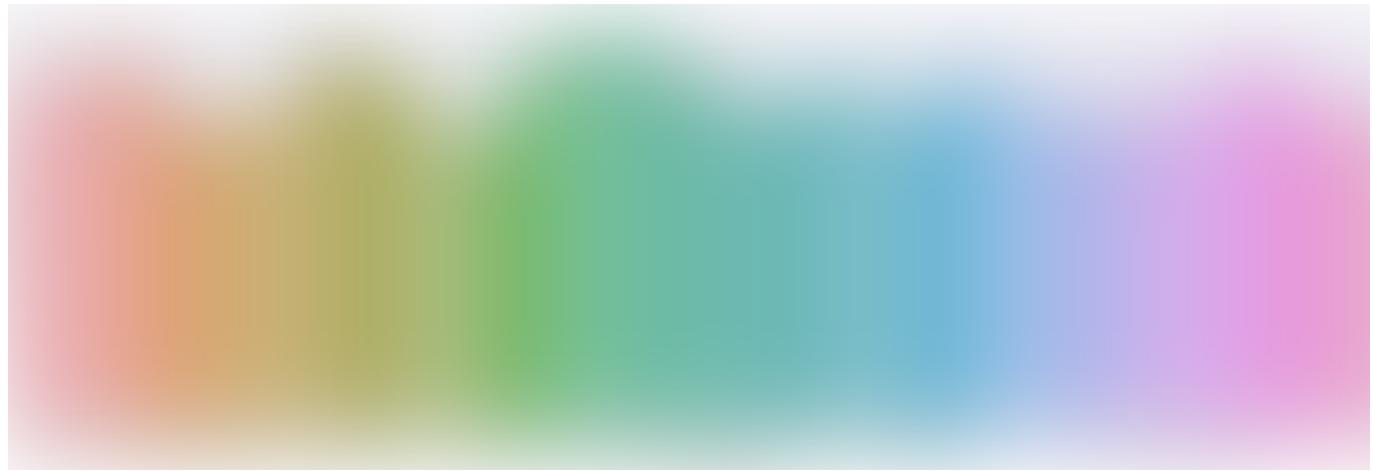


corraletion matrix

HYPOTHESIS :

My first hypothesis is that there is a difference in terms of taste of music between continents. So what I expect is that the countries closer to each other culturally and share same continent would listen similar featured musics

First lets look at danceability and country similarities



We can see that the countries closer to each other have similar danceability value in musics.

Such as :

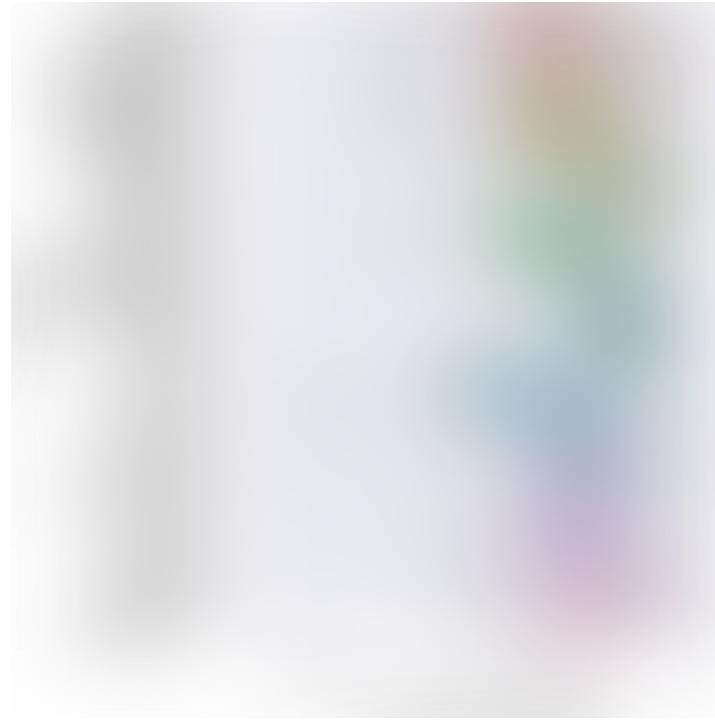
- * Argentina — Brazil — Chile — Bolivia — Columbia
- * Germany — Italy — Belgium (except France and Spain. On the other hand, it is still logical considering that Spain historically affected South America and culturally closer to them rather than location of Europe)

It has same characteristics when we look at the energy level — country attraction



energy level — country

Finally, We can see that loudness change according to countries but locationally close countries tend to have same loudness level in musics



- South America : Argentina, Brazil, Chile, Columbia and Bolivia have similarities closer to -5 dB
- Europe : we can see France — Germany — Italy and Belgium have similarities in their average closer to -8/9dB

Therefore, it is possible to say that cultural similarities + location is great hint to understand potential popularity of the song types in a specific country!

2.

My second hypothesis :

My second hypothesis is that all countries listen their own local musics rather than other music genres :

I use word cloud analysis of python in order to achieve that. It works basically like a word counter. I count the number of words by the genres and number of appearances shows us better view for the genre popularity



All genres in the data set

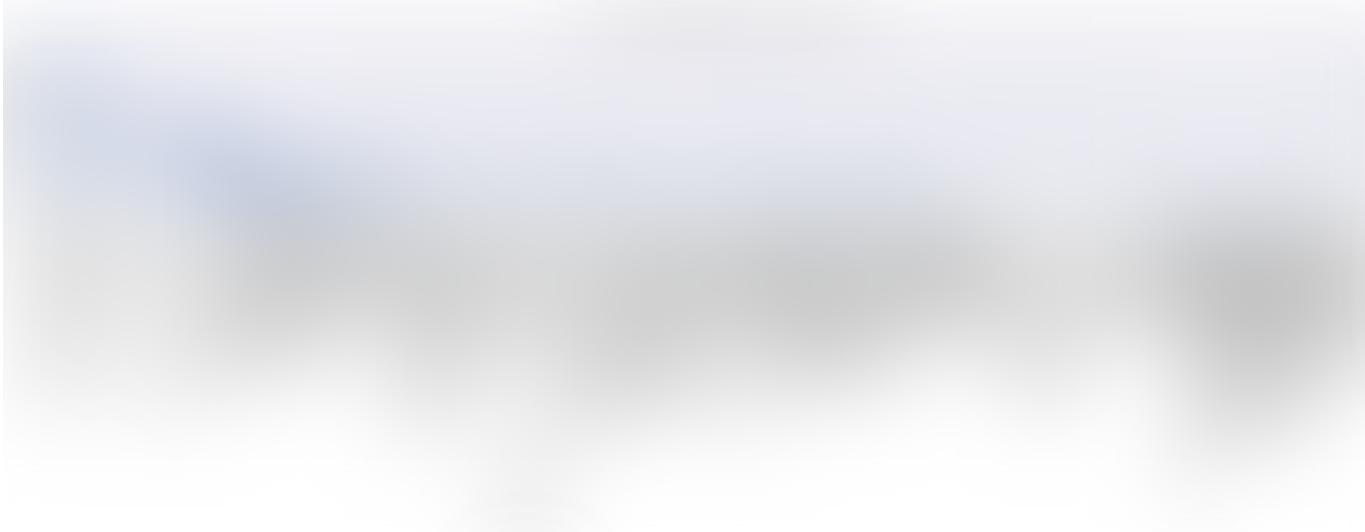
We can see that adult standards are have high density in our data set. Most probably the overall age averages high in our data set countries.



<https://www.worlddata.info/average-age.php>

Such as Japan Germany and Italy are has over 44+ median they are likely to increase the number of ADULT STANDART genre

On the other hand South America have median ages near 25s my other hypothesis is that south america continent will have less adult standards more energetic musics



EUROPE GENRE



SOUTH AMERICA GENRE

To sum up it requires further investigation and bigger data merges in order to check if these hypothesis can be globally generalized rather than spotify data.

[About](#) [Help](#) [Legal](#)[Get the Medium app](#)