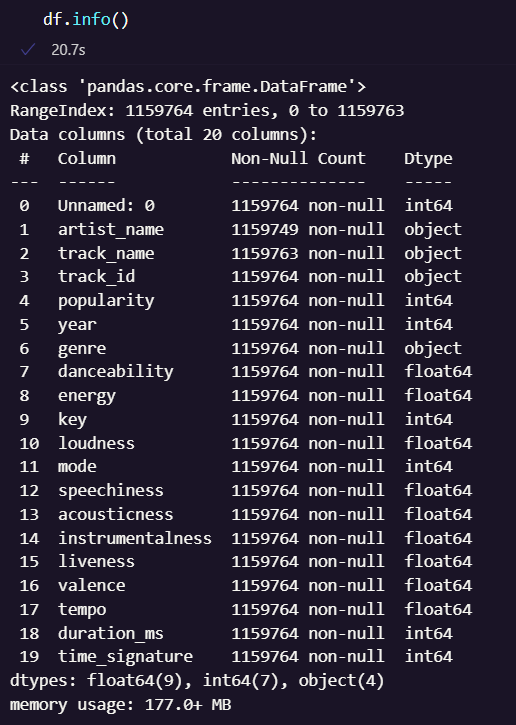
MDSC-102 Final Lab

# Dataset:

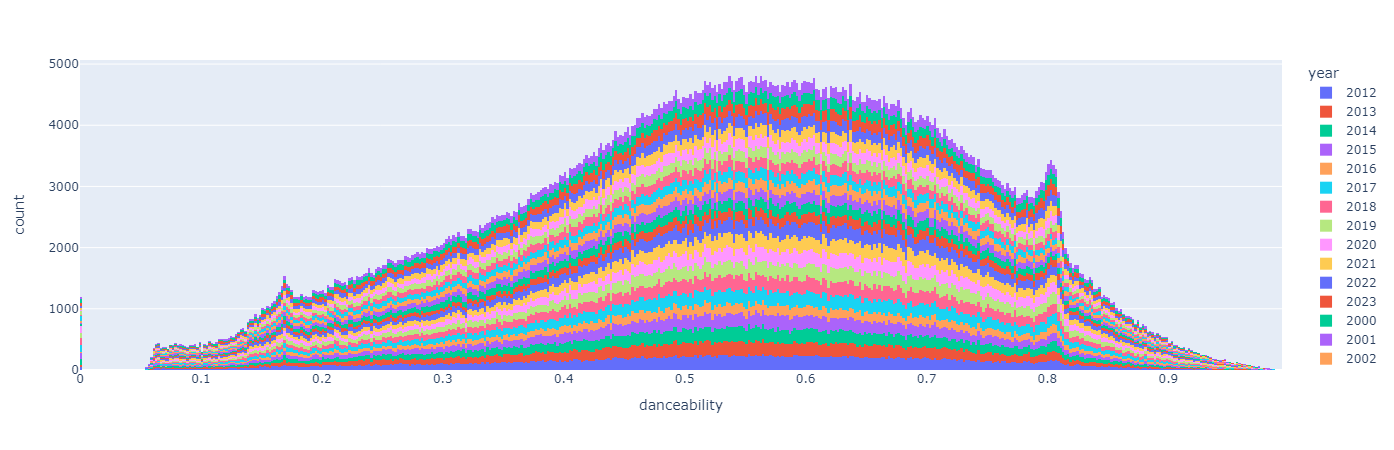
Spotify Dataset containing details of the 1 million+ songs on the platform.

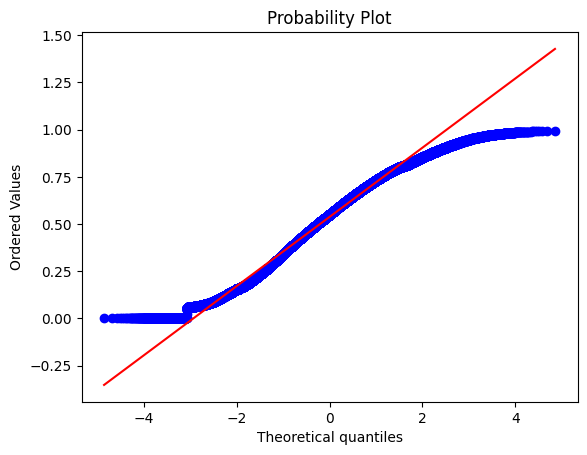
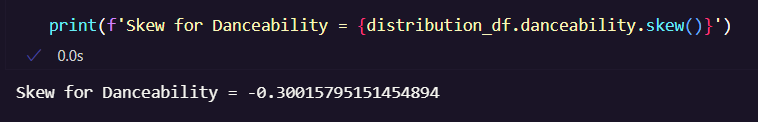


Each record has:

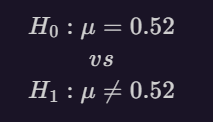
* artist name
* track name
* unique track id
* popularity metric ranging from 0 to 100
* year when the song was released
* genre of the music
* sound metrics such as:
  + danceability
  + energy
  + key
  + loudness
  + mode
  + speechiness
  + acousticness
  + instrumentalness
  + liveness
  + valence
  + tempo
  + duration
  + time signature.

Looking at the continuous features, we can start by analysing the Danceability feature:

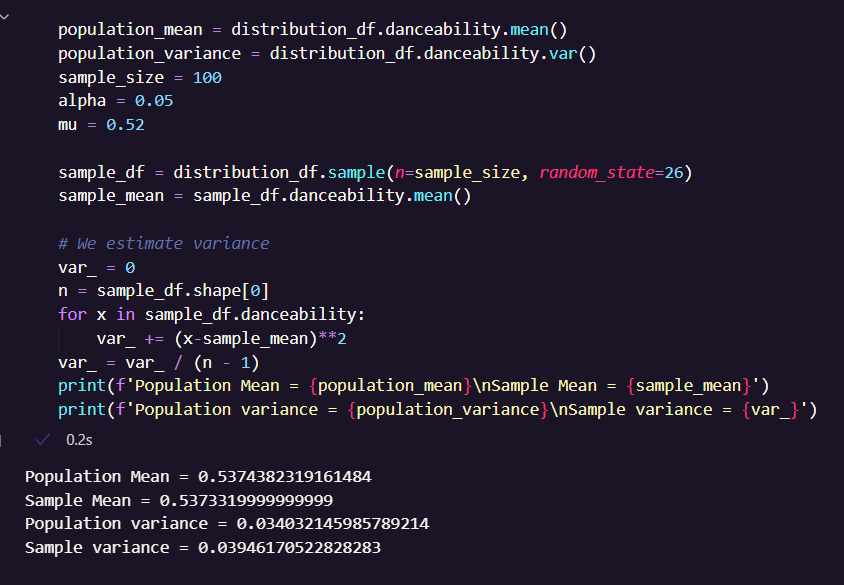


The data is mostly normal, but there is some skewness. It is small enough for us to avoid applying the boxcox method.

Let us assume the following hypothesis for the Danceability distribution:



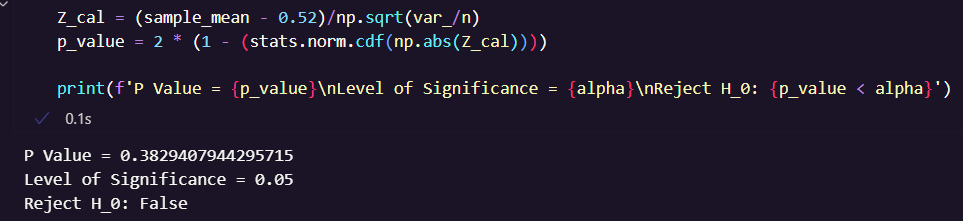
Here, Variance is unknown.



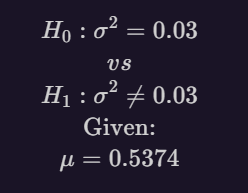
We see that the sample mean is relatively close to the population mean.

But our null hypothesis assumes that the mean is equal to 0.52

Let us calculate the Zcal value and the associated p-value while deciding to either reject the null hypothesis or not with level of significance 0.05.

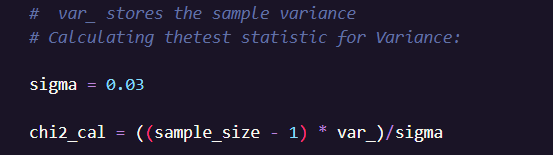


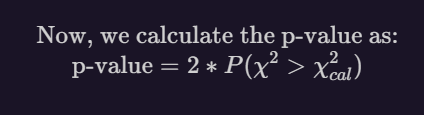
Similarly, we can test the variance:

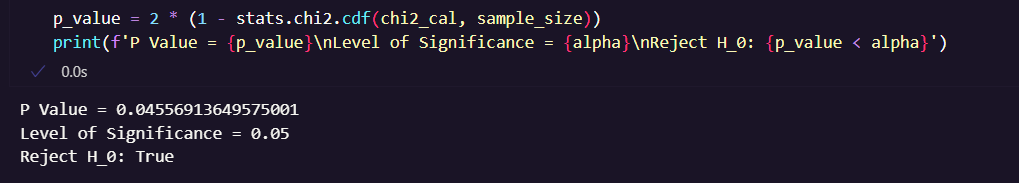


The null hypothesis proposes that the variance is equal to 0.03, but the population variance was 0.034 and sample variance was 0.039

We calculate the test statistic χ2cal

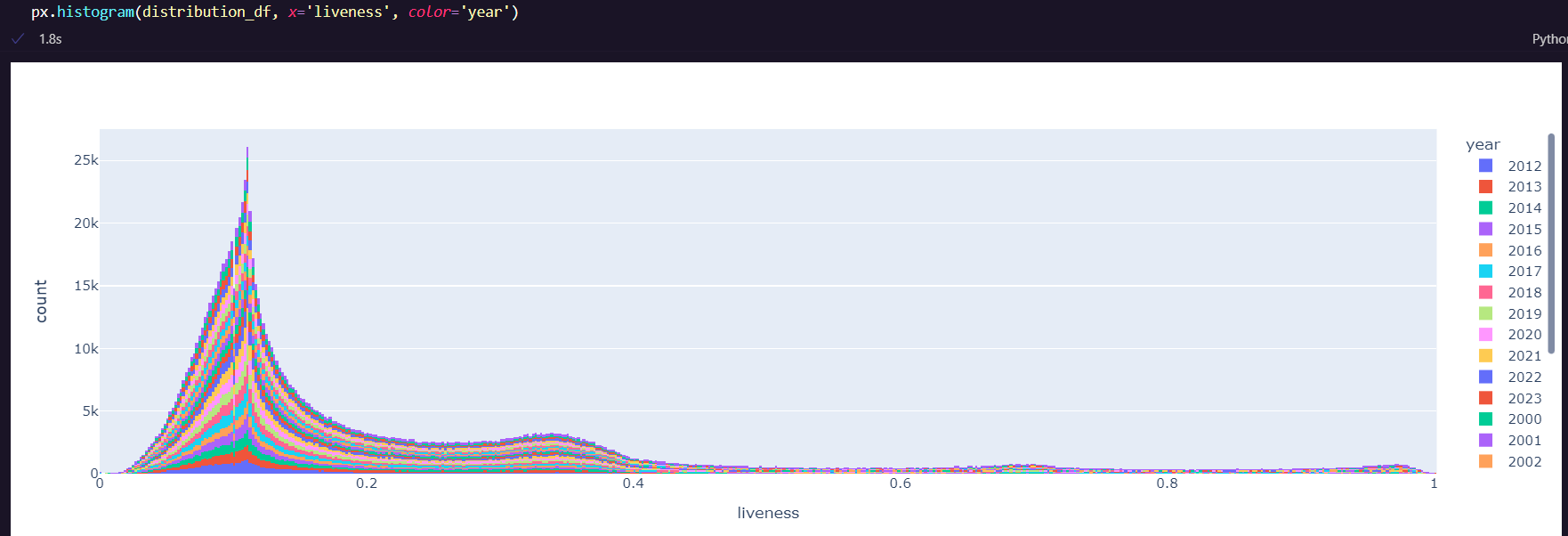


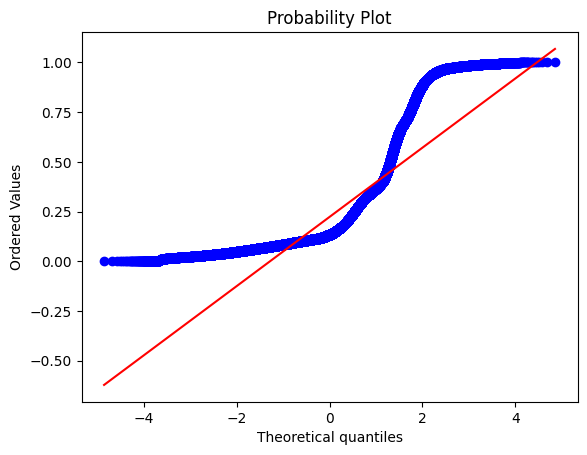


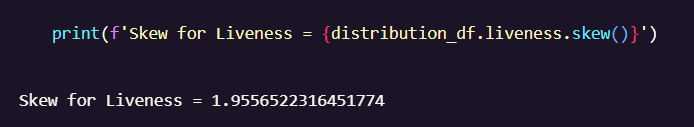


Therefore, we reject the null hypothesis.

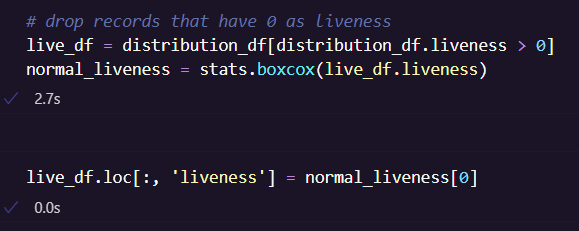
In a similar manner, if we look at the Liveness feature:

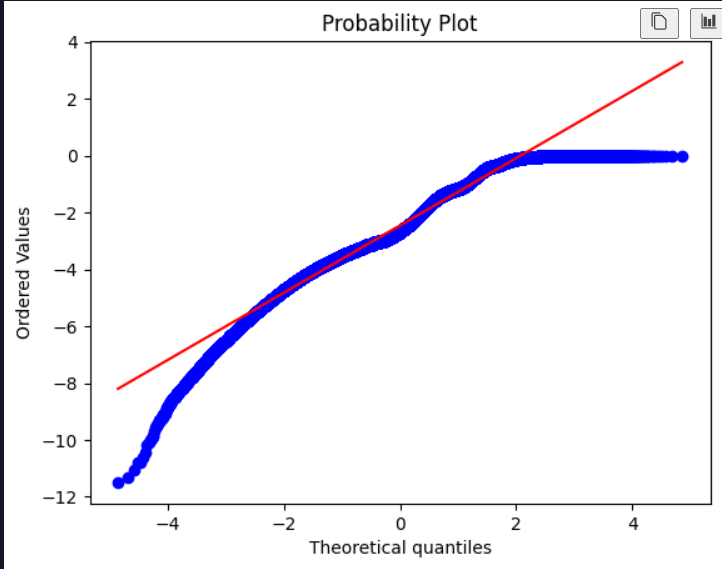
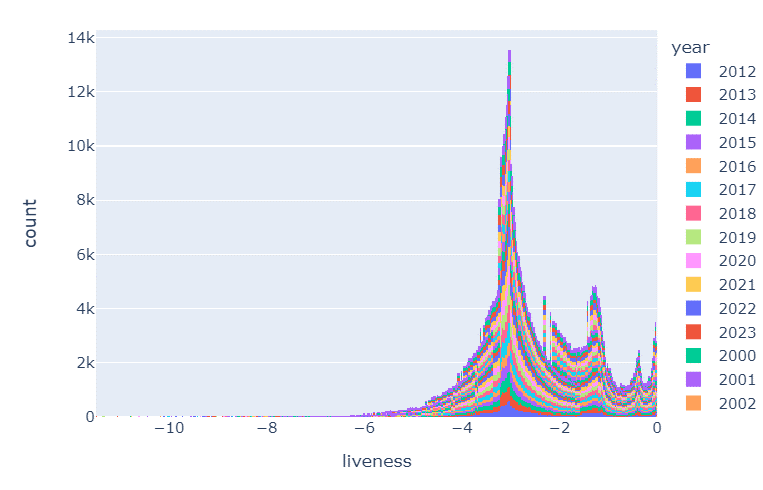


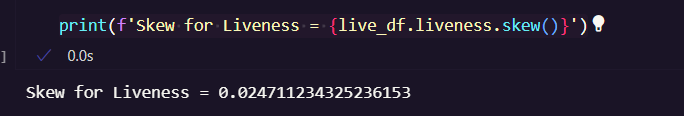
The data is pretty skewed.



We can use the boxcox function to make this data normal.

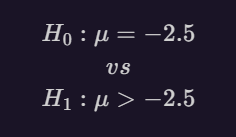




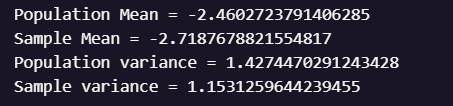


The new skewness of the data is down to 0.0247, although the plot does not look very normal.

Let us test the following hypothesis:

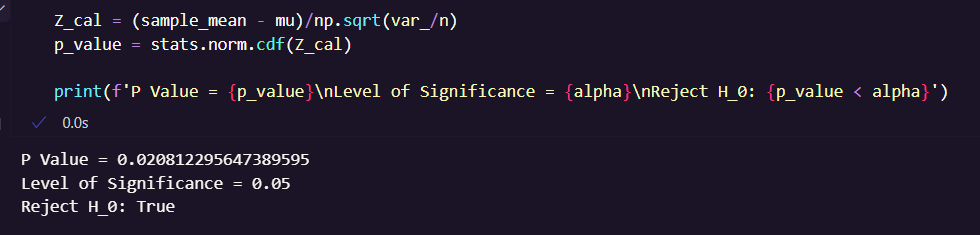


Studying the population, and sample of 100 records, we find that:

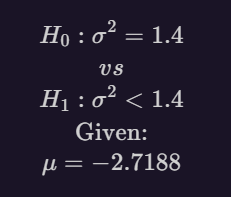


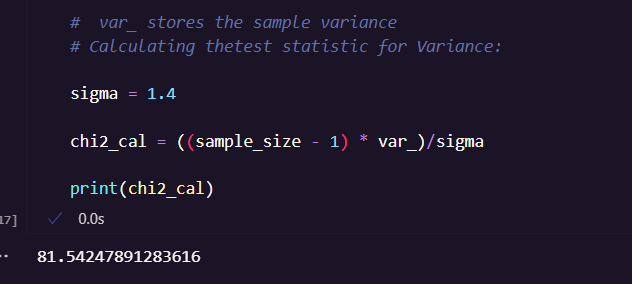
We calculate the Zcal test statistic, and calculate the left sided p-value.

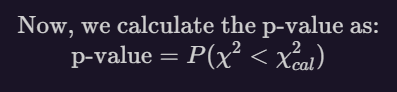
We see that the p-value is less than the level of significance. Therefore, we reject the null hypothesis.

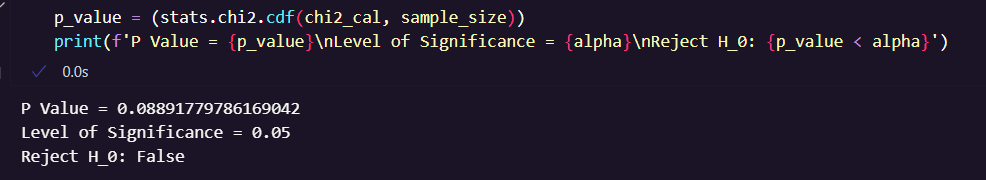


Now, let us test the variance of this data:



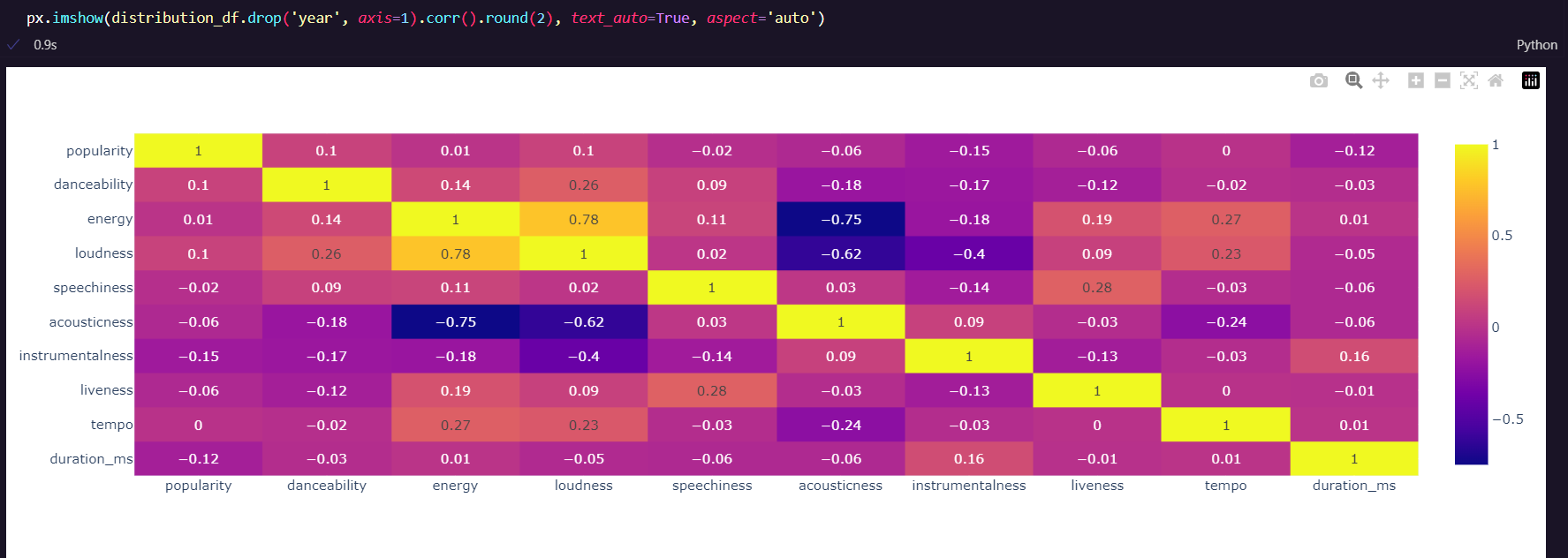




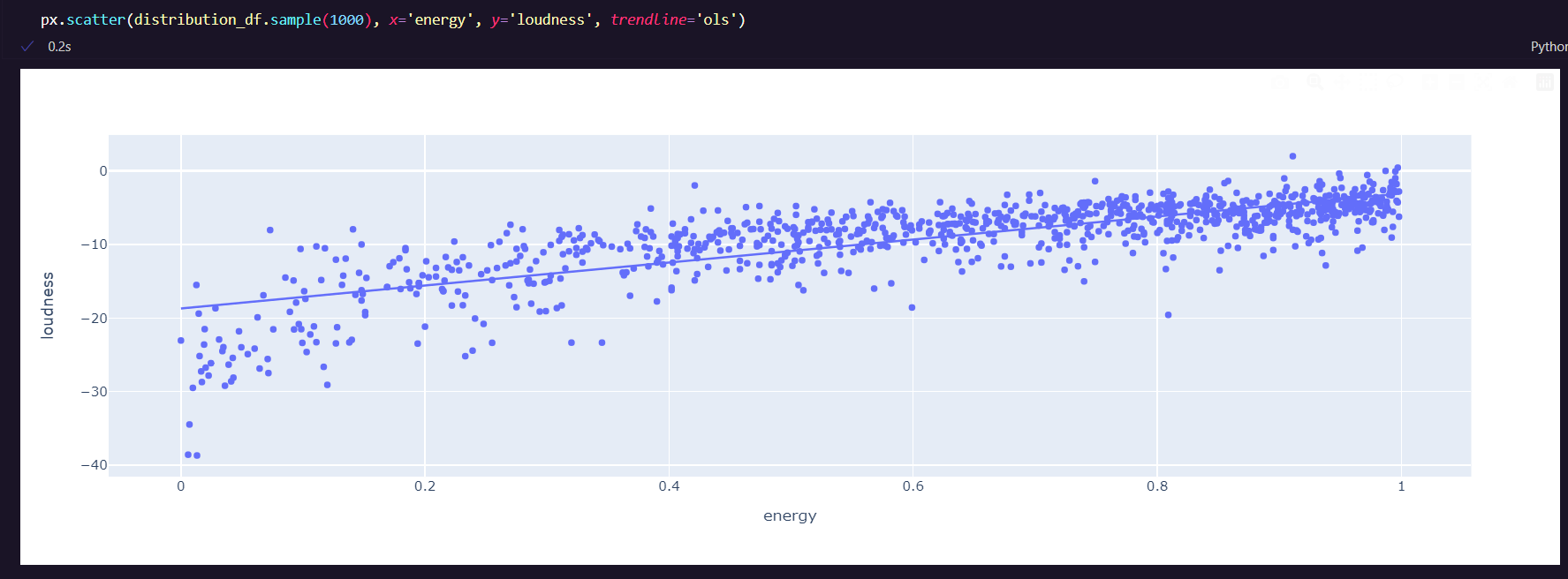


Therefore, we do not reject the null hypothesis.

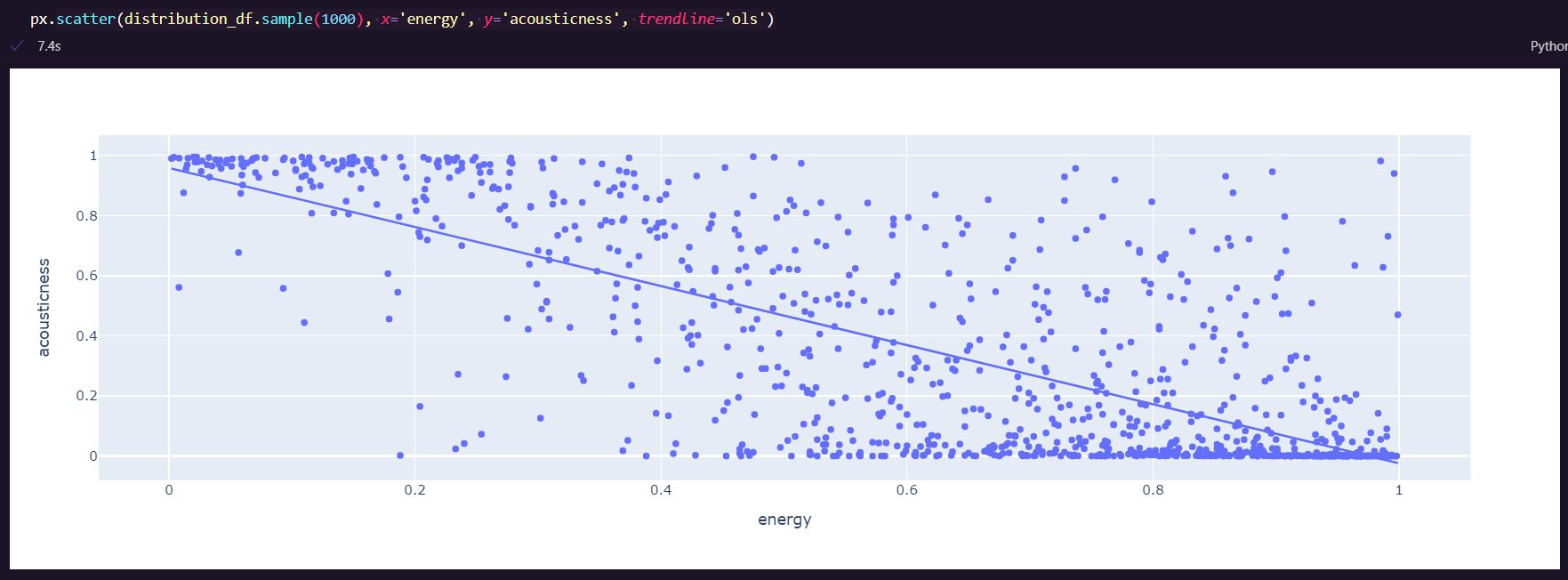
# Analysing correlation between music metrics.



Here, we see a very negative correlation between energy and acousticness, and a positive correlation between energy and loudness. On plotting these data points, here’s what we find:

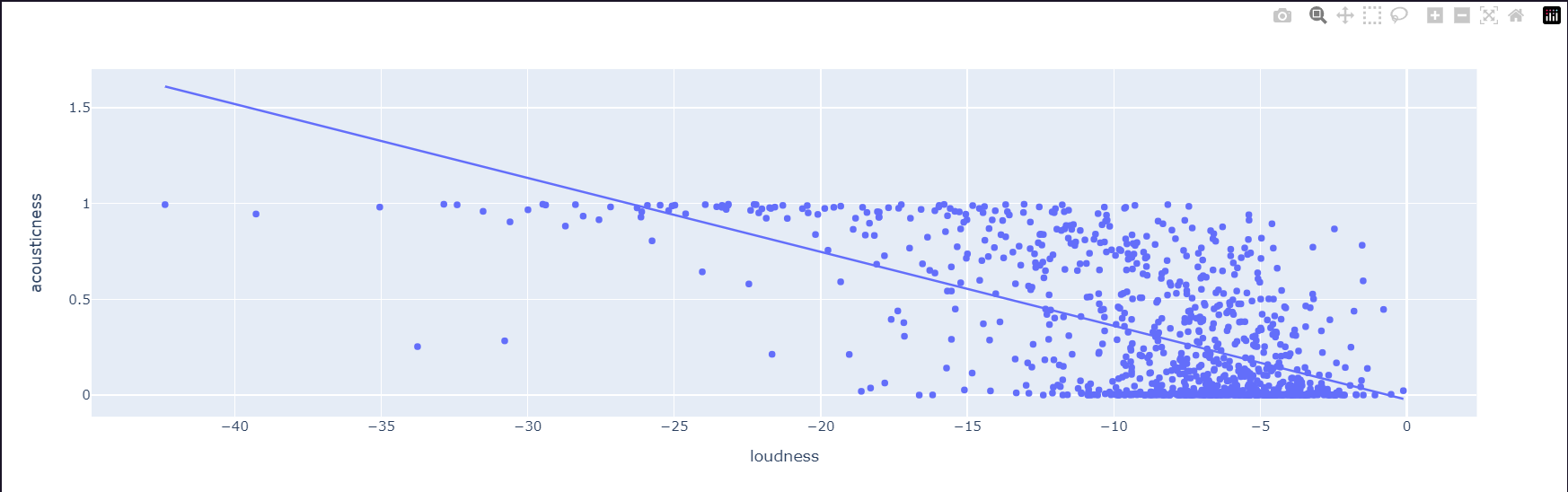


Energy and loudness are indeed positively correlated.

****

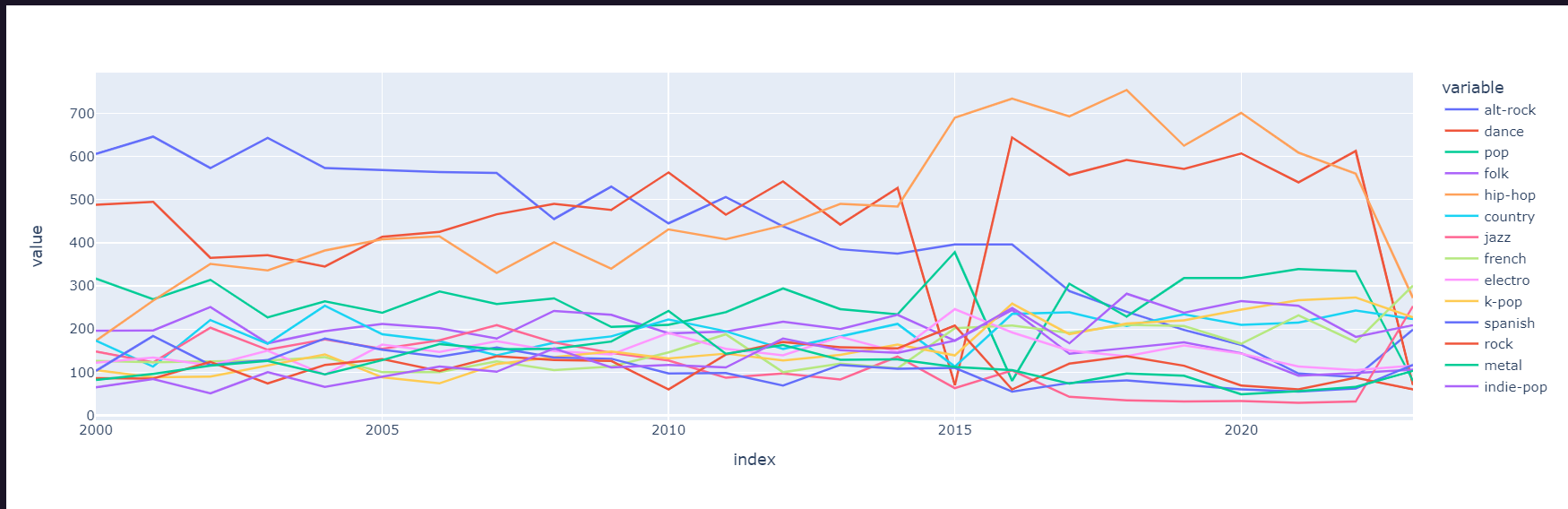
Whereas acousticness and energy are negatively correlated.

Since energy and loudness are positively correlated, and energy and acousticness are negatively correlated, this correlation carries on to show that loudness and acousticness are also negatively correlated.



# Analysing Genre trends

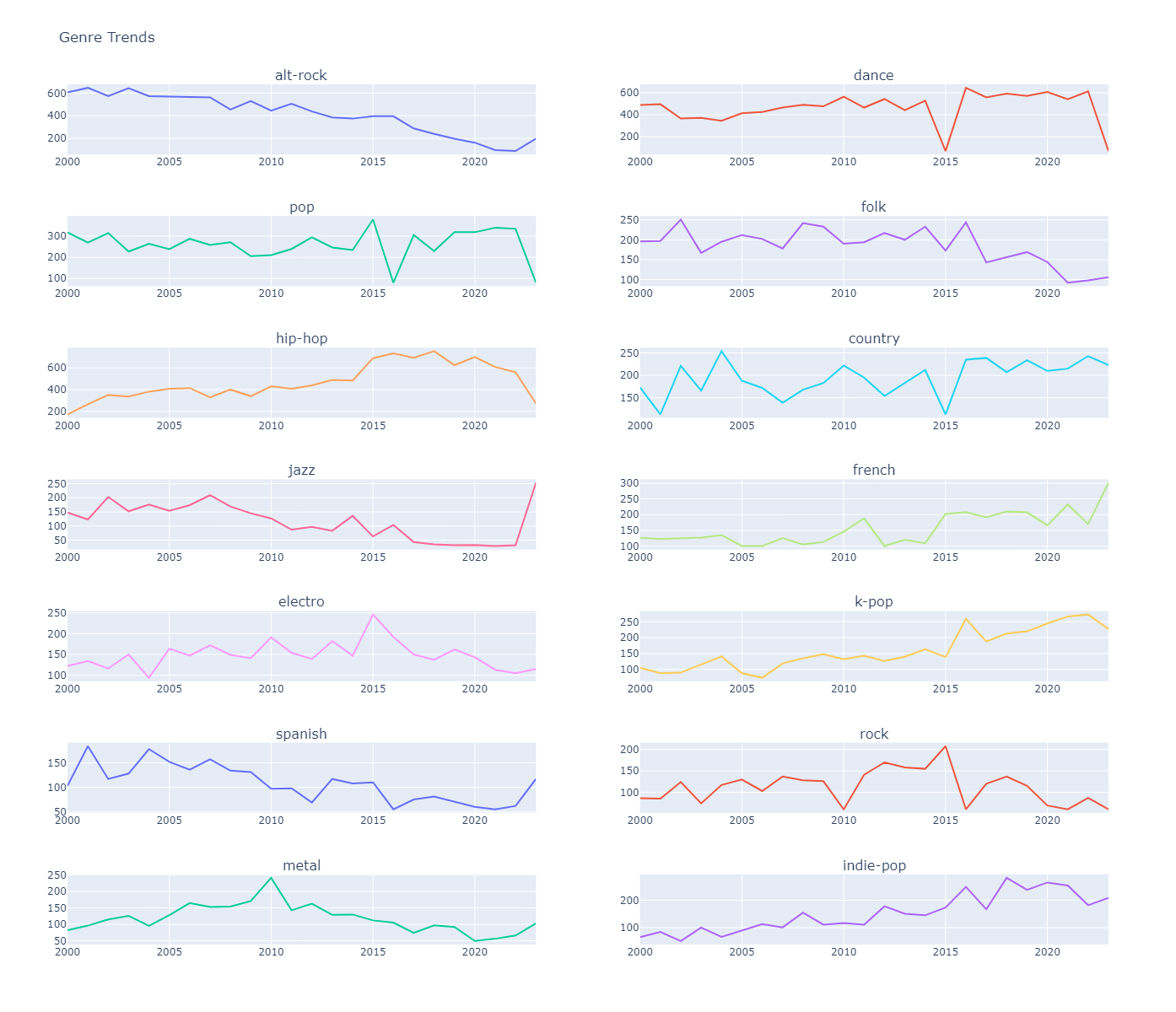
We study the 5000 most popular songs, and the most prominent genres from them, and plot their frequency over the years.



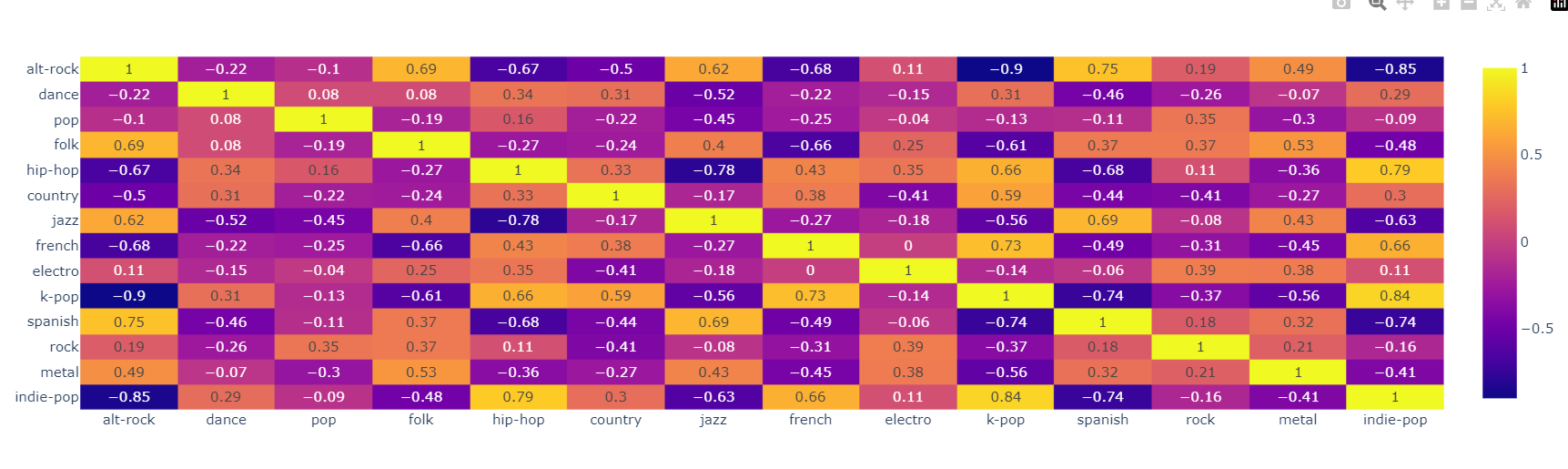
This looks a bit messy, so on visualising the trend of each genre:

We see that alt-rock has steadily declined in popularity over the years, while dance has remained considerably stable, along with a high rise in the popularity of hip-hop.

Genres like French, k-pop and indie-pop have also been steadily rising in popularity.



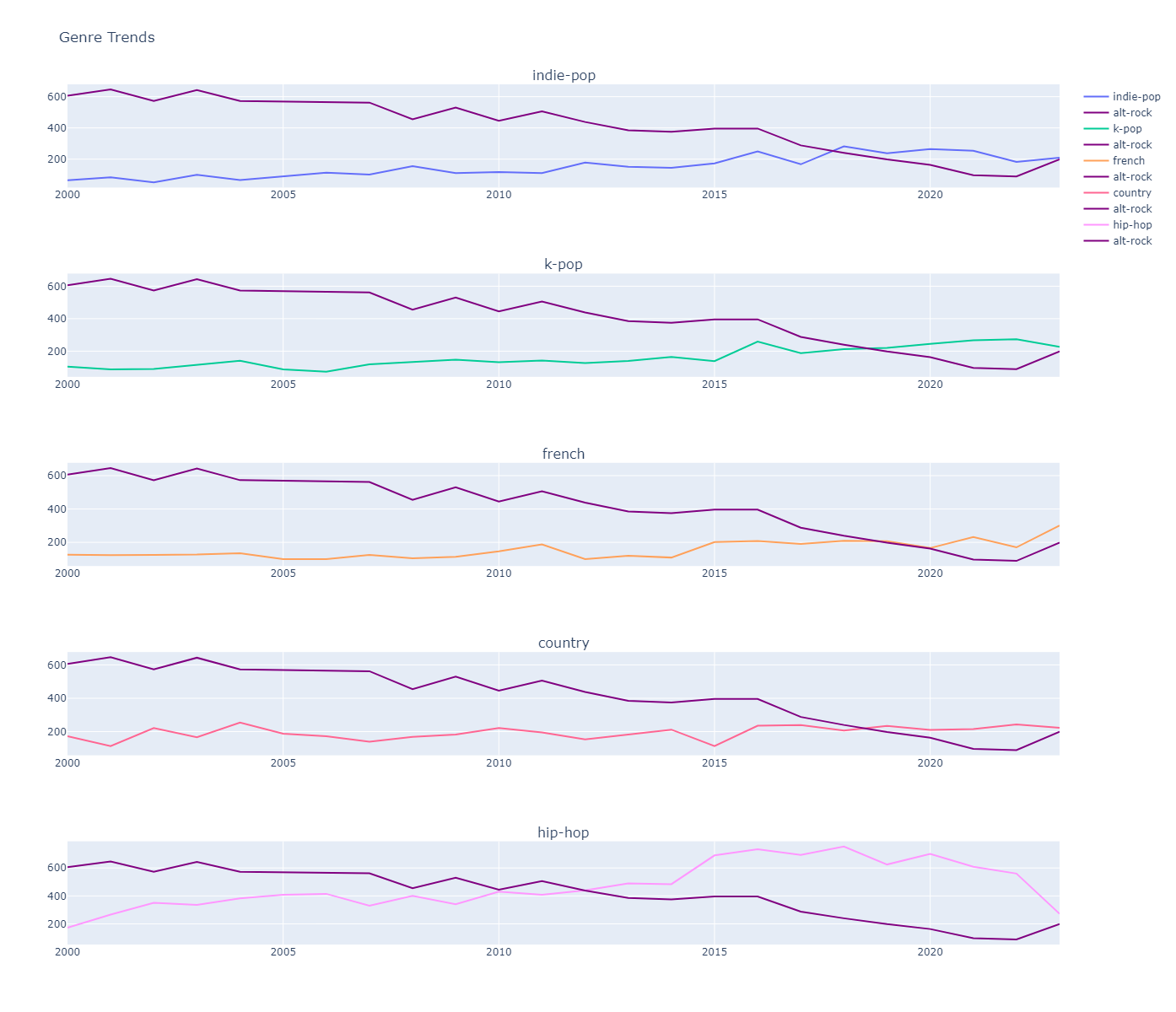
This plot shows us the frequency of the appearance of these genres in the 5000 most popular songs of each year.



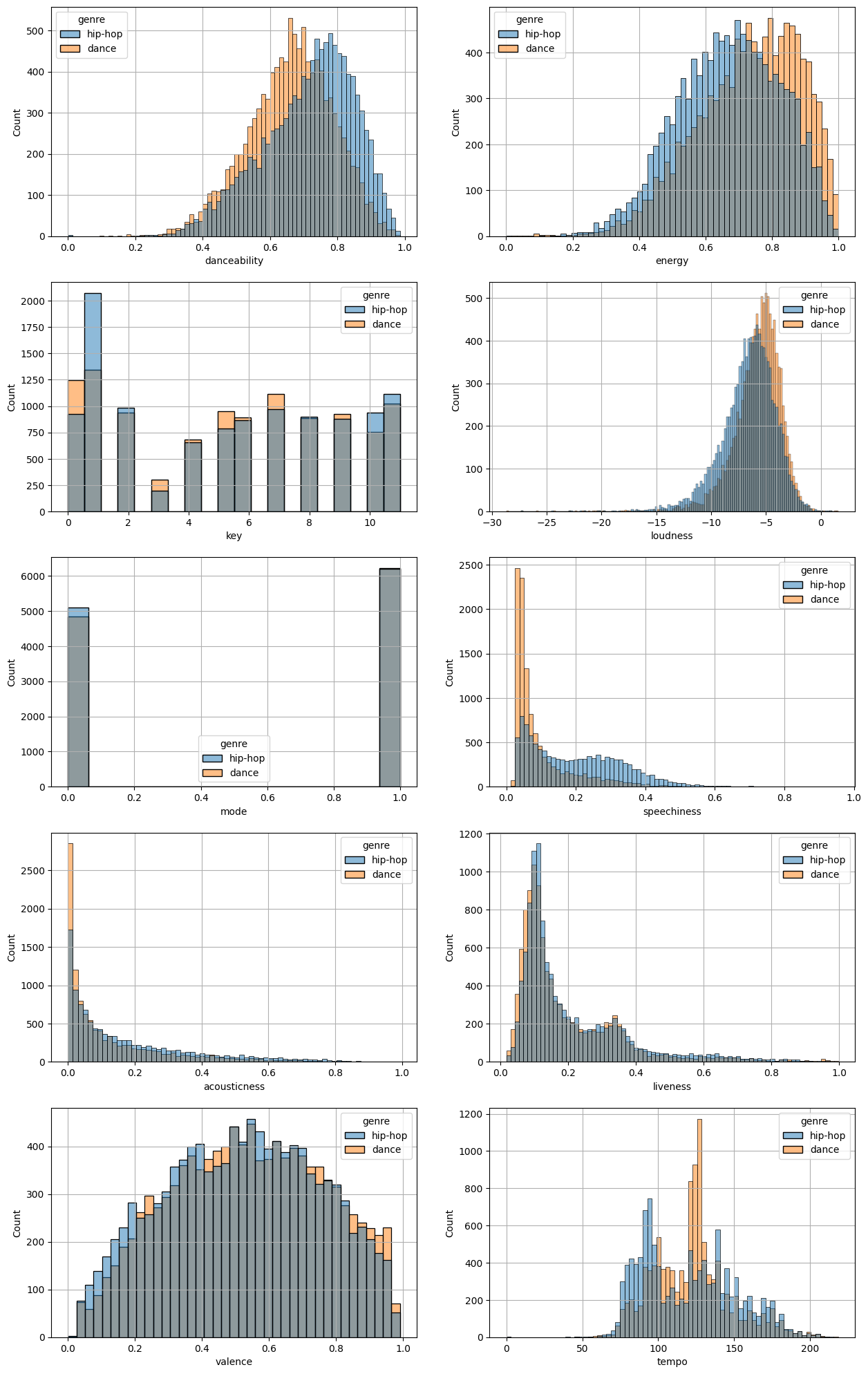
This heatmap shows us the correlation between the frequencies of the popular genres.

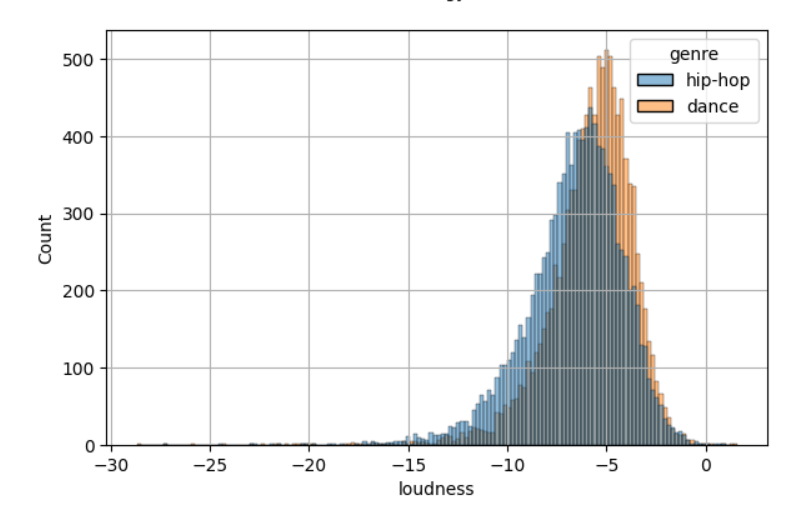
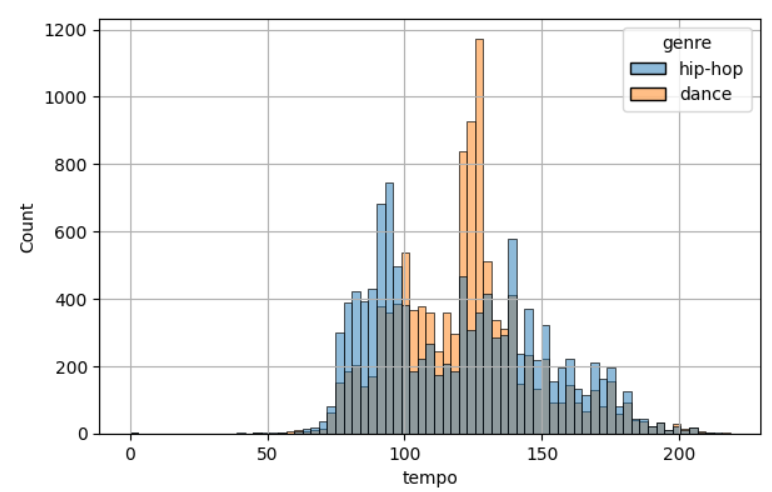
* On further analysing the decline of alt-rock, it is strongly negatively correlated to the following genres:
  + Indie-Pop
  + K-pop
  + French
  + Country
  + Hip-Hop

If we try to plot the decline of alt-rock against the steadily rising genres:

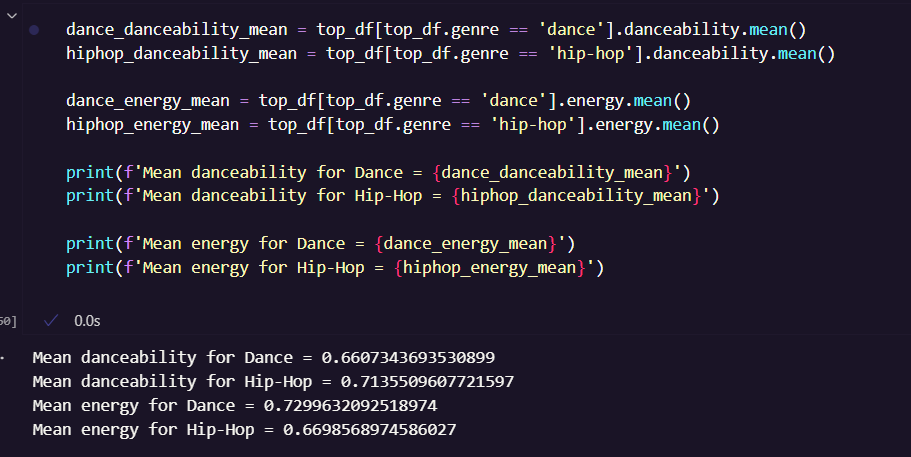


# Studying Dance and Hip-Hop and why they have dominated the market

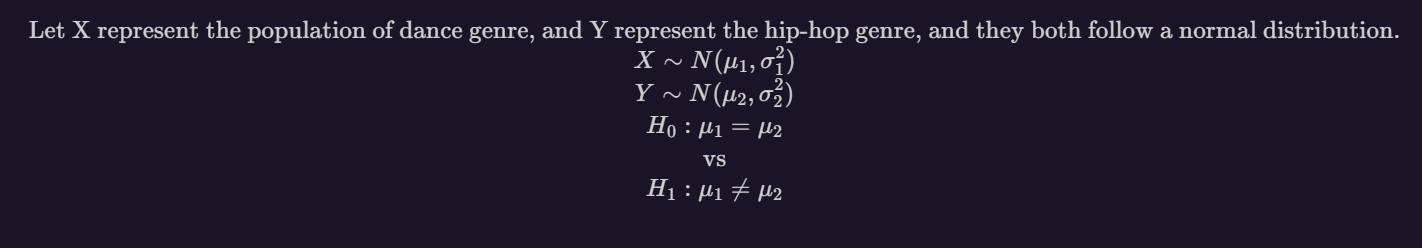


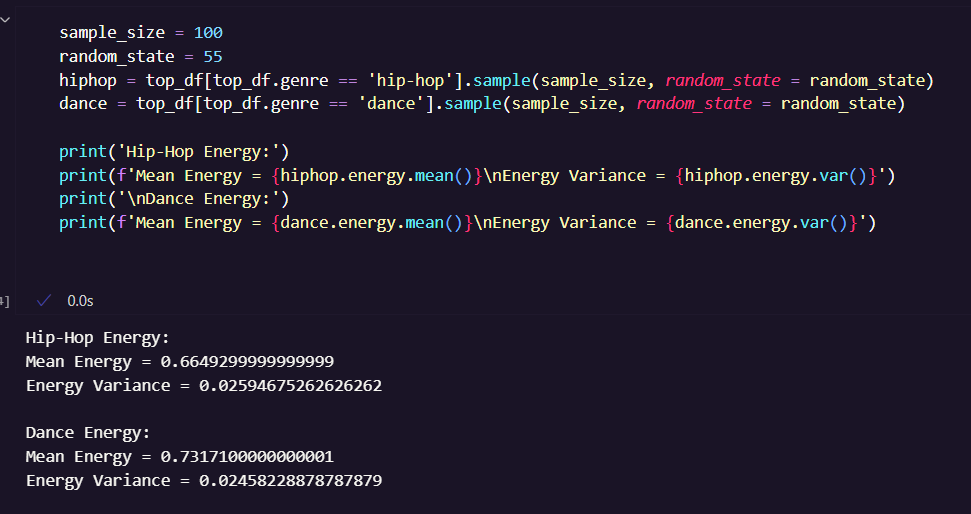
 

We see that hip-hop and dance have similar features, only that surprisingly enough, hip-hop has a higher danceability mean than the dance genre itself, but dance has higher enery mean.

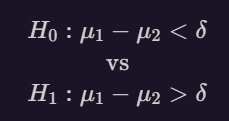


However, let us test the hypothesis that the mean energy for dance is equal to the energy for hip-hop.

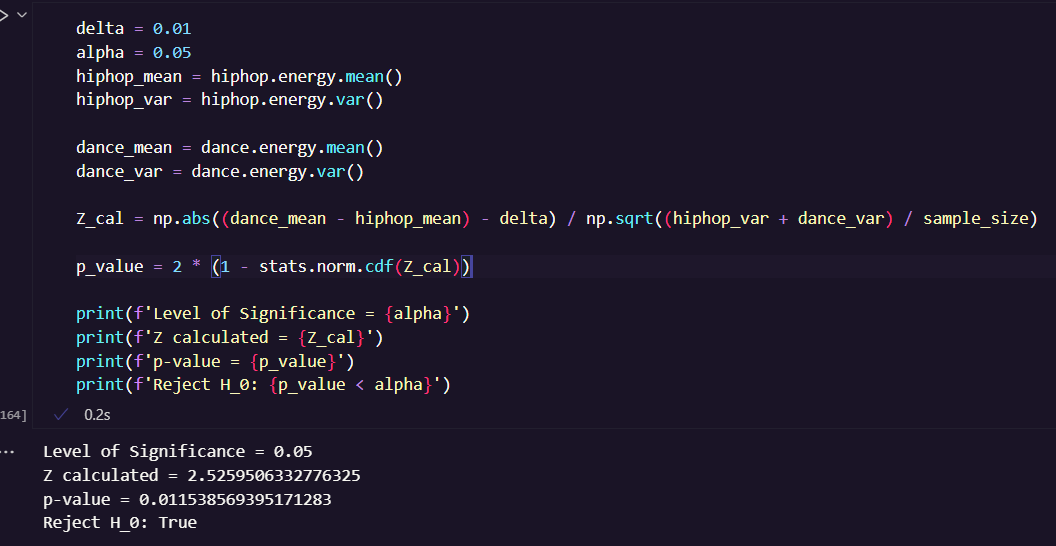




Equivalently, we can write our hypothesis as:



Where we take our delta as 0.01



Therefore, we reject the null hypothesis.

The mean energy of dance genre is not equal to the mean energy of hip-hop, as is shown by the plot above.