processing of the data:

1. loading it in
2. keeping only one records of song per date

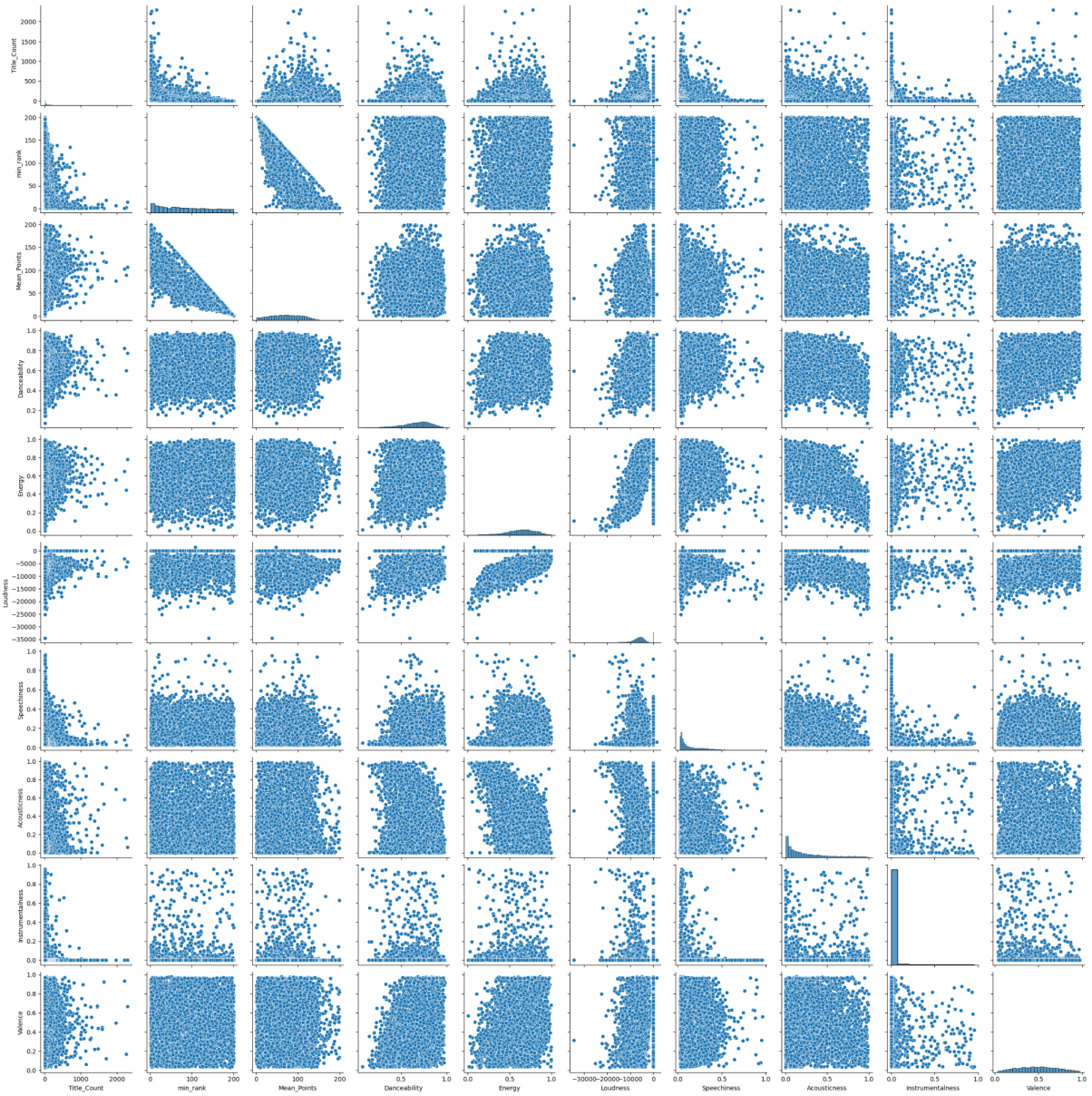
   as our exploration involves looking into the song characteristics and their popularity over time, we do not need separate instances of the same song per date for the different artists
3. adding ‘min_rank’ for the min (top) place a song has achieved in the charts overall
4. adding ‘Title_count’ to the df as the total number a title has occurred in the charts
5. keeping only the first (latest) instance of a song across the whole dataset
6. adding ‘points(Total)_mean’ as the average value of points scored by the song in the charts
7. filtering out the songs which last occurrence in the dataset was in the earliest month

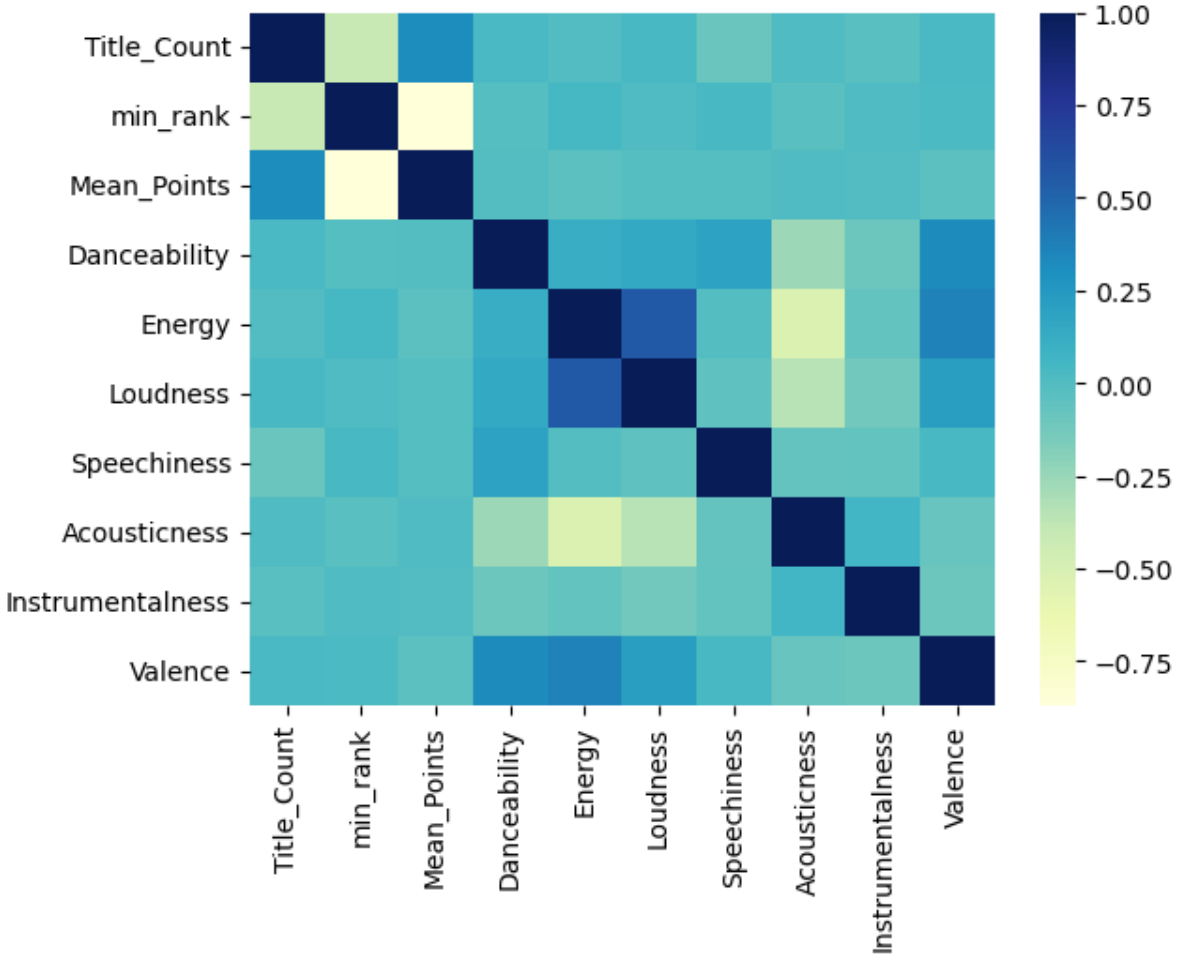
   the songs we could see in there we know that they are quite popular like Ni**as In Paris, Paper planes, stay. purple lamborghini scoring correspondingly 1, 9 and 14 points even though they are one of the top popular songs overall we will filter out for the latest occurrence of a song in the charts to be at least a month after the start of the charts.

   This will ensure that we are not mistakenly saying that a song scored only low points in the charts --> implying that it was not popular when in fact we are getting its end of popularity, coming off the top charts.

   This means that the oldest of the songs in the new dataframe will have a month of past popularity that we can account for.
8. getting

pairplot for correlations for: ['Title',
'Title_Count',
'min_rank',
'Mean_Points',
'Danceability',
'Energy',
'Loudness',
'Speechiness',
'Acousticness',
'Instrumentalness',
'Valence']


Title_Count
min_rank
Mean_Points
Danceability
Energy
Loudness
Speechiness
Acousticness
Instrumentalness
Valence


Title_Count
min_rank
Mean_Points
Danceability
Energy
Loudness
Speechiness
Acousticness
Instrumentalness
Valence
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75