Main research topic:

What are the factors affecting Song Popularity: The Analysis of Solo Performances, Collaborations, and Audience engagements? How does the presence of one or more highly popular songs in an album affect the popularity and listener engagement of other tracks within the same album?

Question 1: Solo vs Collab

Background:

Is there a relationship between a song's popularity and whether the track was a solo or a collaboration? We hypothesise that a collaborated song gains more attention, which may result in the success of a song and yield a better popularity score and performance on some of the musical attributes such as danceability and energy.

Data analysis steps :

We addressed the question of whether a song's popularity correlates with whether it is performed solo or in collaboration using an external dataset of 114,000 samples from Spotify, which included columns such as track popularity, artists, and musical features. We classified tracks as purely solo or with collaboration and explored the popularity scores. As the first observation, the dataset has several repeated instances owing to associating more than one genre with the same track. To do this, we removed duplicates and for tracks with more than 1 entry, we merged all genres into a new column.

We also created a binary variable collaboration to indicate if a track is a collaboration or a single artist track. Then we set a threshold for classifying tracks as popular. Tracks with a popularity score of 50 or above were considered popular. This allowed us to compare the average popularity of solo tracks vs collaborative ones. A popularity score of 50 was selected as the benchmark for a popular song, as indicated in the graph (Appendix C-1). This value aligns with the 75th percentile, suggesting relatively few tracks surpass a popularity score of 50.

As part of the cleaning, we determined there were some outliers based on the popularity column – using a boxplot – and investigated whether or not they affected overall trends. We also checked no null values in essential columns including artists, track_id, and popularity. We computed summary statistics for features such as danceability, energy, and tempo for both solo and collaborative tracks to see if certain features correlate to popularity. We then performed a t-test to compare the mean popularity score of solo with collaborative tracks, confirming the significance of the differences. Lastly, we generated bar plots, scatterplots, and boxplots to summarize the findings and show the relationship between track popularity with whether it was a solo track or a collaboration. These visualizations focused on the trends and enabled us to gain a holistic understanding of the dataset. Solo tracks have an average popularity score of 74.00, while collaborative tracks have an average popularity score of 86.88. That said, we observed potential biases in the dataset, such as the underrepresentation of niche tracks, which should be considered when interpreting the results.

Main Findings :

After cleaning and analyzing the data, we compared the average popularity of songs and found that collaboration tracks were generally more popular than solo tracks . Additionally, we created a graph comparing the duration_ms of solo and collaboration songs, which revealed that collaborated songs tend to have a longer duration on average."

As seen in appendix (A-1) the graph indicates the popularity scores of two categories of music tracks: solo tracks and collaboration tracks. The x-axis is the figure of speech, popularly measured by a score of 0-100. On the y-axis is the frequency within each popularity range.

Some of the key observations identified within this graph are that the distribution is highly right-skewed values, implying that there are many tracks with low popularity scores compared to the few with high scores. The number of jams per day reaching the peak is about 20-25 in the case of solo tracks and 40-45 in case of collaboration tracks. The given circles for the solo tracks are spread larger

and reach higher popularity scores than the collaboration tracks. These density distributions partially overlay, this suggests that some particular records are similarly popular: microphone solo tracks and collaborative records.

The graph implies that although the frequencies of the solo and collaboration tracks look like they follow the same curve, the popularity scores for the collaboration track appears to have a slightly higher mean.

The graph given in appendix (A-2) is a heatmap labelled as 'Comparison of Musical Features for Solo and Collaboration Tracks.'

On the horizontal axis, we have the track type starting from 'Solo' at the left end and 'Collaboration' at the right end. On the y-axis, there are several factors which relate to music; duration, the extent to which the content is explicit, its danceability, energy, key, loudness, mode, how speech like it sounds, its acoustics, instrumentality, liveness, valence, tempo and time signatures.

The demarcation of shades within each cell is anchored on the mean value of the applicable attribute for the specified track kind. A code with a black background and numeration in dark blue means that the variable is close to zero, while a code in red means that its value is high. A mixed value is grayed out while intermediate values are in shades of purple.

If the same attribute has different levels of color intensity within solo and collaboration tracks then we can compare the average value of each. For instance, while looking at collaboration tracks, duration and explicitness are relatively higher and danceability as shown by the darker shade of red in those cells. In contrast, solo tracks are more spread to the right and up, which means that they have bigger acousticness and instrumentalness scores, illustrated by darker blue color.

There are attributes, for example, 'key' and 'time signature,' which have little difference between the solo and collaboration tracks as the shades in their cells suggest. This means that there is not much difference between these attributes of the two track types.

Altogether, the presented heatmap offers an exhaustive and simple approach of comparing the musical attribute differences between the analyzed solo and collaboration tracks. It enables fast identification of trends and patterns to know the characteristics of each track type.

To further understand the distribution of energy and danceability among solo and collaborated songs. The graph found in Appendix (A-3), breaks down the relationship between the following attributes.

In Terms of danceability of songs, both solo tracks and collaboration tracks look very similar, and the density function has a maximum danceability value of 0.6. This we can deduce when comparing the collaboration tracks to other tracks where most bars are slightly higher in danceability.

The energy distribution is also identical for solo and collaboration tracks with maximum at about energy 0.5. The distribution of energy is also similar between solo and collaboration tracks, with a peak around an energy value of 0.5. However, collaboration tracks have a more pronounced peak and a slightly wider spread compared to solo tracks, suggesting that collaboration tracks tend to have a greater variation in energy levels.

Comparing danceability as well as energy distributions, there is some similarity in these aspects between solo tracks and collaboration tracks. However, collaboration tracks are found to be just a little more danceable and vary in terms of energy level than the solo tracks.ity value of 0.6. Collaboration tracks tend to have slightly higher danceability values, as evidenced by the taller bars in the higher danceability range.

The graph titled "Average Popularity: "The Mean of Blog Frequency" is a horizontal bar chart showing the means of blog frequency between solo and collaboration tracks.

The horizontal axis is the type of track: solo or collaboration; the vertical axis is the average popularity score. The orange bar shows a mean number of listens per solo track, the blue bar illustrates a mean number of listens per collaboration track.

In order to understand average popularity among solo and collaborated songs, the graph seen in Appendix (A-4) was created to further understand the relationship.

Collaboration tracks are more popular with heights of the blue bar being notably higher than those of the orange bar. These results raises the possibility that a collaboration between artists could increase demand, although the reasons may be the combined fan base, which on its own is not necessarily higher than individual popularity, or the appeal of such projects collective.Nevertheless,

these results suggest that average popularity is the most widely recognized analytica format, however, in this context, it is noteworthy that the outcomes do not reflect other parameters that may affect popularity like the type of a musical, the time of release, and promotion campaigns.collaboration), and the y-axis represents the average popularity score. The orange bar represents the average popularity of solo tracks, while the blue bar represents the average popularity of collaboration tracks.

This information suggests that collaborations between artists may contribute to increased popularity, possibly due to the combined fan bases or the unique appeal of collaborative efforts. However, it is important to note that this graph only shows average popularity and does not account for other factors that may influence popularity, such as genre, release date, or marketing efforts.

Drawbacks of the Analysis Performed:

As the dataset is large, processing time and power become a concern that affects the feasibility of certain analyses. Same song was part of many other genres making it tough to relate which category it should be counted in. As well as single songs by the artist's were present in some albums affecting their rating just by the presence of a particular song.

Question 2 : Popularity comparison

Background:

We aim to explore how the presence of one or more highly popular songs in an album impacts the popularity and listener engagement of other tracks within the same album. The data reveals that adding a highly popular song to an album significantly enhances the album's overall popularity. This effect may stem from increased exposure, as listeners drawn to the popular track are more likely to explore other songs within the album. Additionally, the presence of a standout hit can elevate the perceived value of the album, encouraging more engagement.

Data analysis:

The tracks are categorized based on their popularity based on a threshold score of 50. The analysis then examines album popularity. The mean popularity of tracks within each album is calculated and stored in a dataset. Albums are categorized as "popular albums" if they contain at least one highly popular song, and "unpopular albums" if they do not. Finally, our analysis examines the distribution of popular songs within albums. The total number of tracks and the count of popular tracks are computed for each album, and the percentage of popular songs is calculated.

The albums shown in Appendix B-1 splits into two different categories between albums with no popular songs and albums with at least one popular song. The y-axis represents the number of songs in the album or the frequency of songs within the albums. The results of the graph show that the albums with at least one popular song are more popular than albums without any popular songs. Most of the songs with no popular songs range from slightly popular (popularity scores 1–10) to somewhat popular (popularity scores 30–50), while most albums with at least one popular song mostly range from moderately popular (popularity scores 30–50) to highly popular (popularity scores 60–80). Overall, this graph reveals that albums containing at least one highly popular song were more likely to have higher average popularity.

In the second graph (Appendix B-2), dives more into the popular albums. We decided to remove all the popular songs from within a popular album and find the average popularity score within these albums. This will give insight on how the other songs perform within a popular album. The graph revealed that the average popularity is closer to 50 and is higher than albums that don't have popular songs all together. This suggests that popular songs in an album have an affect on the popularity of other songs within an album.

Lasty, the third graph derived from the dataset (Appendix B-3) represents the percentage of popular songs within an album. Data suggest that almost all the songs (100%) within a popular album have a popularity score of more than 50. This translates to the fact that there is a strong relationship between the impact a popular song has on a particular album and its tracks.

Drawbacks of the Analysis Performed:

The analysis did not account for external factors like marketing efforts, artist reputation, or social media trends, which can significantly influence an album's popularity. The absence of listener demographic information further limits the ability to generalize findings to diverse groups. The focus on quantitative metrics, such as popularity, ignored qualitative aspects of music like lyrics or emotional resonance, which also influence engagement.

Word Cited:

What Is the Spotify Popularity Index? | Two Story Melody

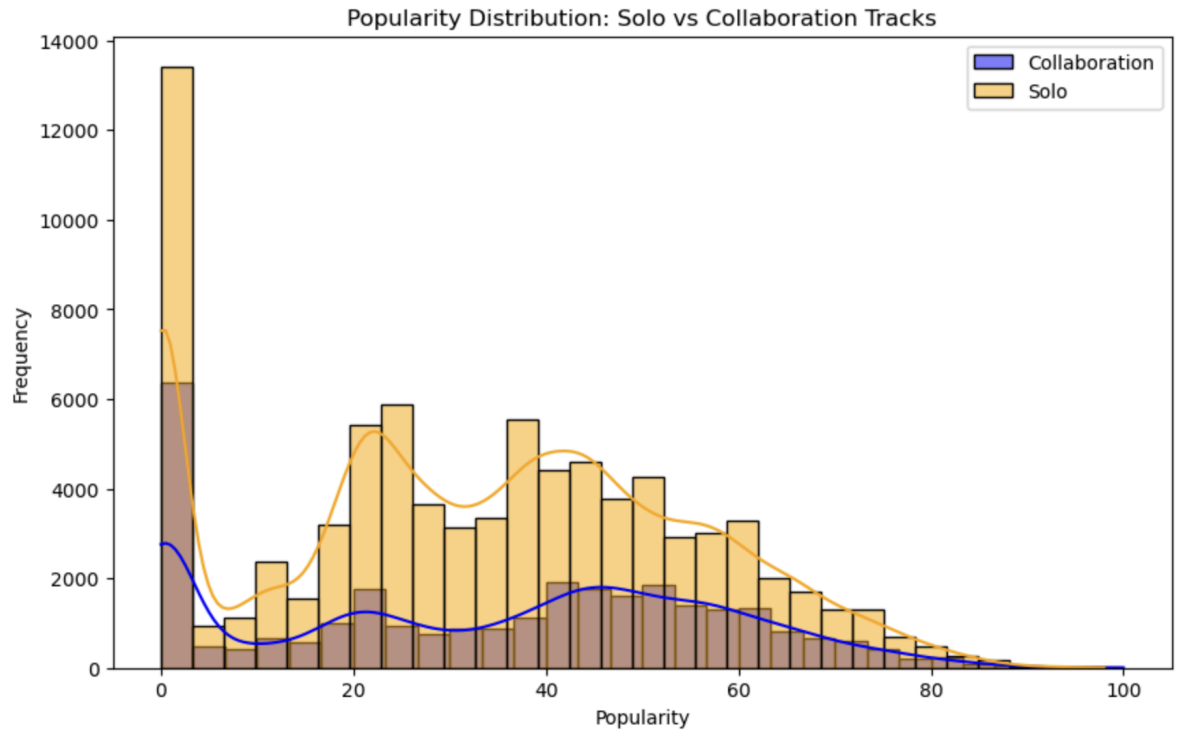
https://help.chartmetric.com/en/articles/1560578-what-is-spotify-popularity-index

https://musicat.musicstax.com/spotify-popularity-index

https://www.cbc.ca/radio/undertheinfluence/marketing-hit-songs-1.2947741

Appendix:

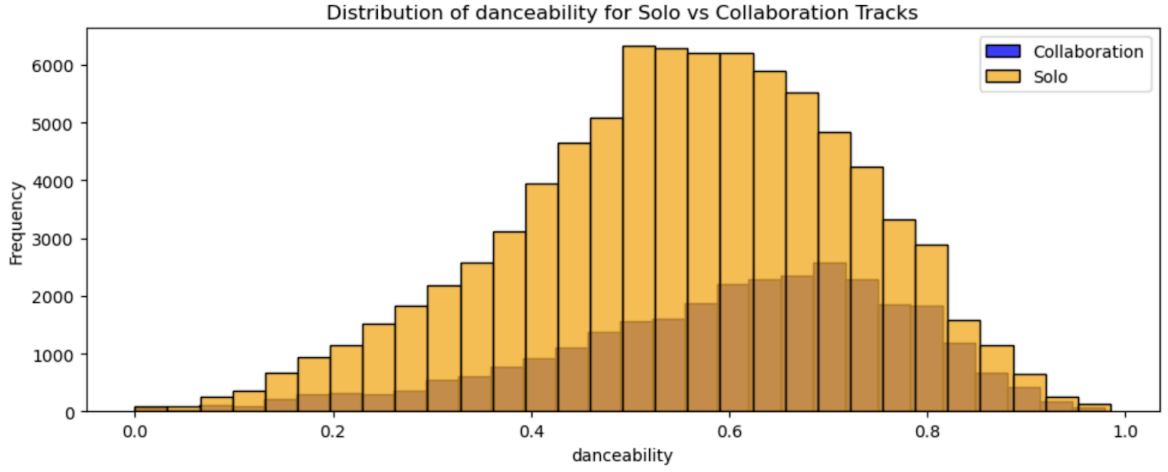
A-1

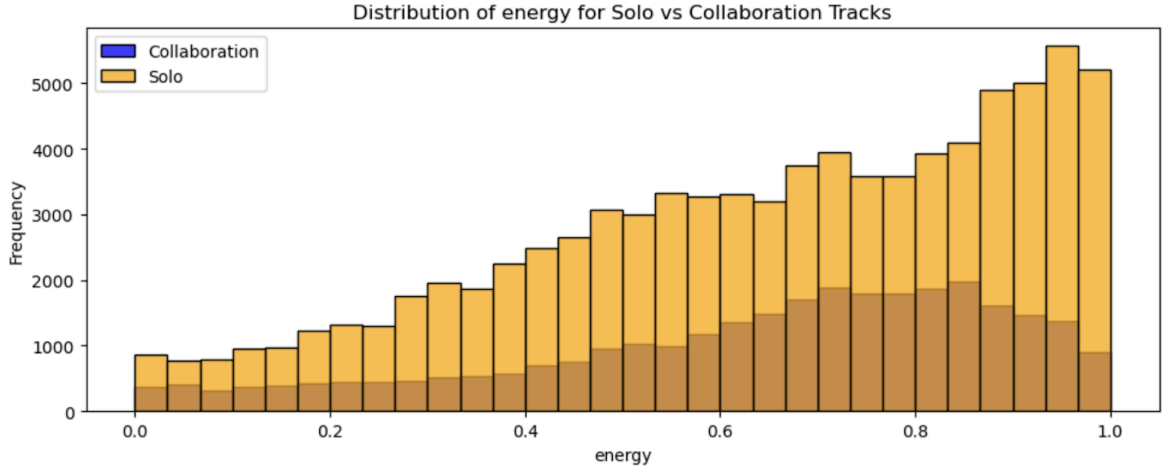


A-2

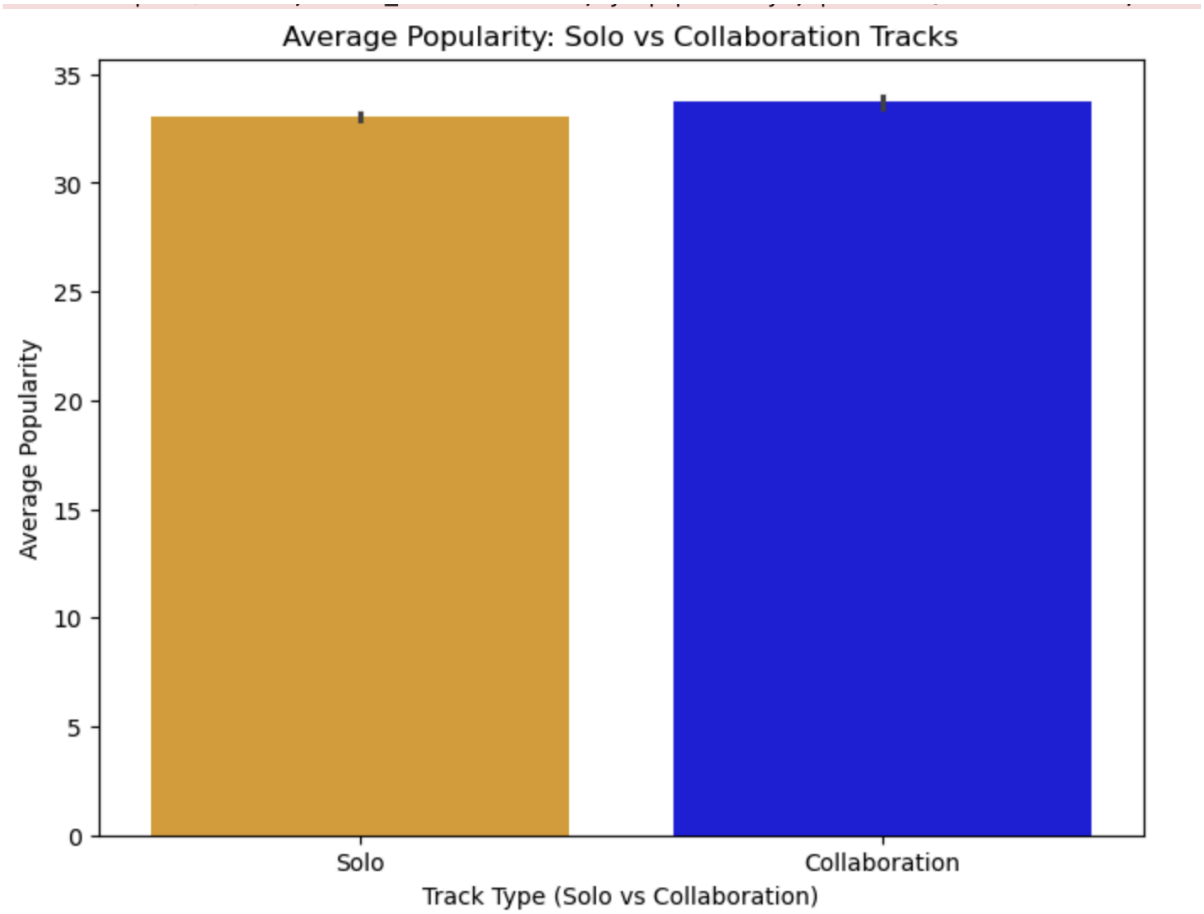

Heatmap of Musical Attributes for Solo vs Collaboration Tracks
duration_ms
explicit
danceability
energy
key
loudness
mode
speechiness
acousticness
instrumentalness
liveness
valence
tempo
time_signature
227295.01
230085.34
0.08
0.11
0.55
0.60
0.64
0.64
5.28
5.39
-8.34
-8.04
0.66
0.58
0.08
0.09
0.32
0.30
0.16
0.14
0.22
0.20
0.48
0.46
122.40
121.44
3.90
3.92
Musical Attributes
Solo
Collaboration
Track Type
Mean Value
200000
150000
100000
50000
0

A-3


Distribution of danceability for Solo vs Collaboration Tracks
Collaboration
Solo
Frequency
danceability
0
1000
2000
3000
4000
5000
6000
0.0
0.2
0.4
0.6
0.8
1.0


Distribution of energy for Solo vs Collaboration Tracks
Collaboration
Solo
Frequency
energy
0
1000
2000
3000
4000
5000
0.0
0.2
0.4
0.6
0.8
1.0

A-4

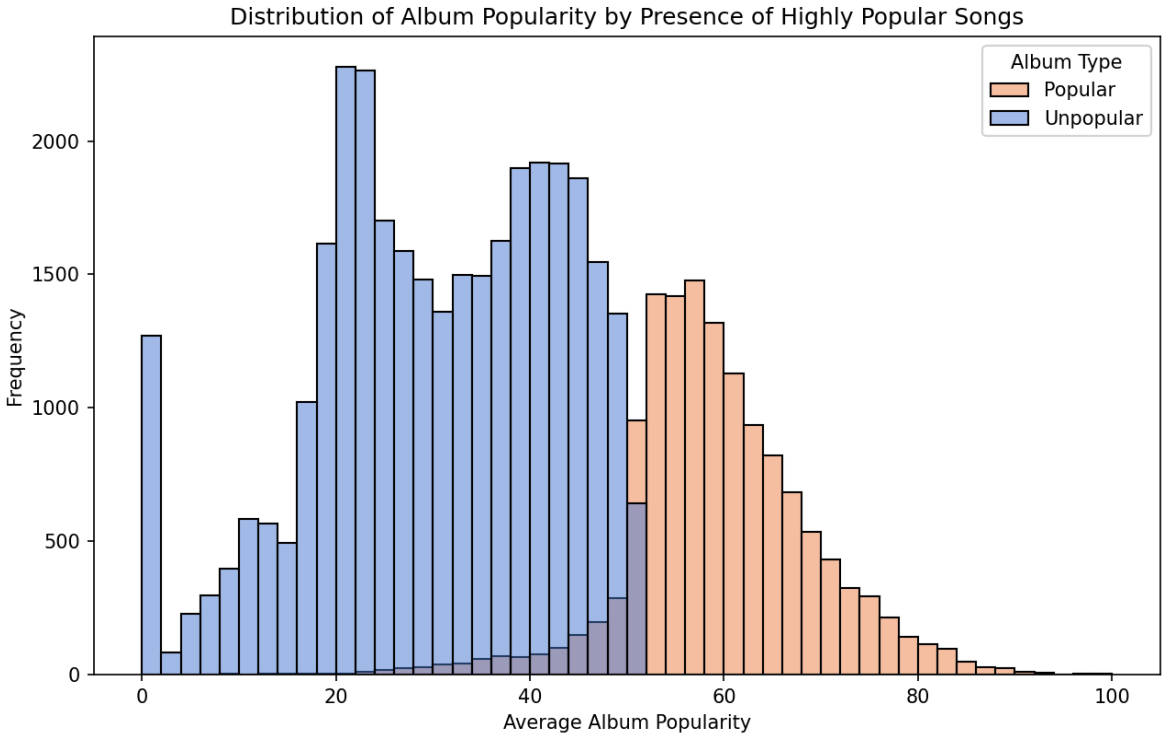


A-5

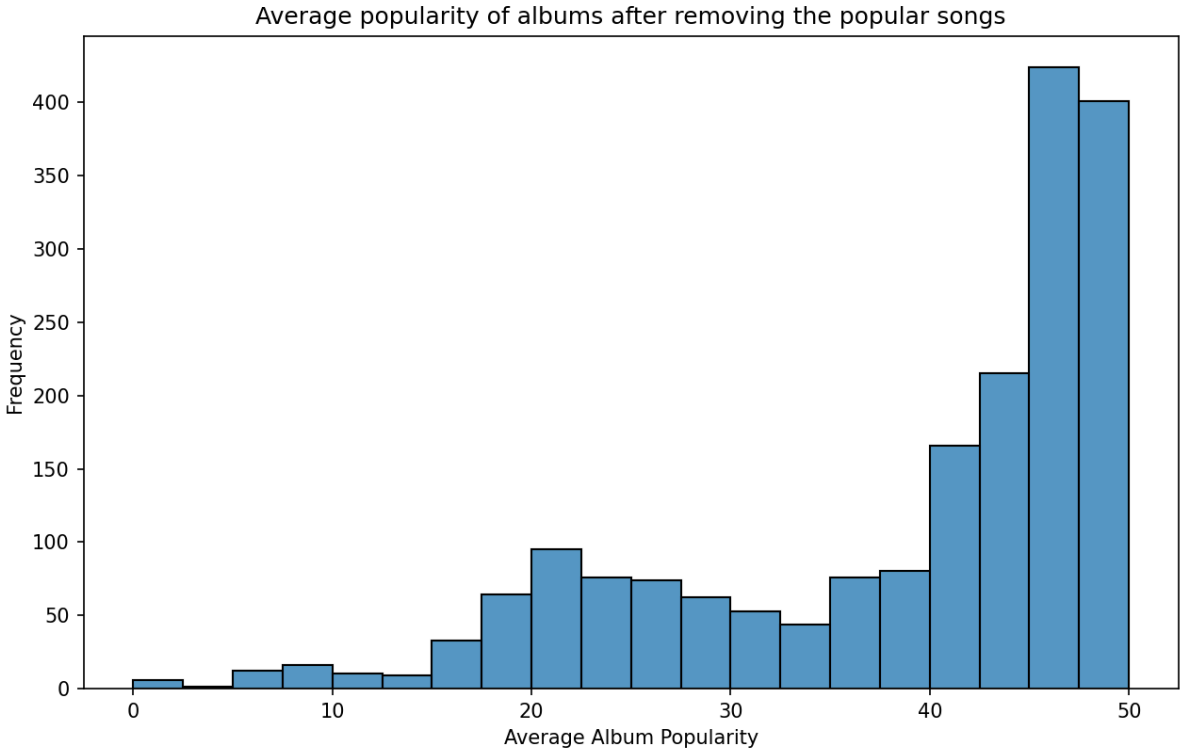
Rank	artists	track_name	popularity
1	Sam Smith;Kim Petras	Unholy (feat. Kim Petras)	100
2	Bizarrap;Quevedo	Quevedo: Bzrp Music Sessions, Vol. 52	99
3	David Guetta;Bebe Rexha	I'm Good (Blue)	98
4	Manuel Turizo	La Bachata	98
5	Bad Bunny;Chencho Corleone	Me Porto Bonito	97
6	Bad Bunny	Tití Me Preguntó	97
7	Bad Bunny	Efecto	96
8	Chris Brown	Under The Influence	96
9	OneRepublic	I Ain't Worried	96
10	Harry Styles	As It Was	95

Rank	track_genre	danceability	energy	popularity
1	pop-film	0.597146	0.6045619	59.283
2	k-pop	0.6478788788788788	0.6757467467467467	56.952952952952955
3	chill	0.664346	0.42672289999999996	53.651
4	sad	0.692378	0.4624701	52.379
5	grunge	0.457062	0.80329	49.594
6	indian	0.592273	0.567121	49.539
7	anime	0.5374514	0.6741083	48.772
8	emo	0.599321	0.6699669202999999	48.128
9	sertanejo	0.591647	0.710391	47.866
10	pop	0.630441	0.606437	47.576
11	progressive-house	0.6239349999999999	0.813359	46.615
12	piano	0.4550983	0.32010256400000003	45.273
13	mandopop	0.546532	0.4984339	45.025
14	deep-house	0.710448	0.741855	44.808
15	brazil	0.562948	0.620721	44.67
16	electronic	0.652945	0.6947523	44.325
17	pagode	0.577723	0.7121230000000001	44.298
18	ambient	0.3678668	0.23716178999999998	44.191
19	british	0.5012758	0.50712697	43.802
20	metal	0.46428800000000003	0.840273	43.705
21	metalcore	0.4238	0.914485	43.477
22	psych-rock	0.502554	0.5615028	42.778
23	hard-rock	0.48225	0.7950389999999999	42.775
24	acoustic	0.549593	0.43536810000000004	42.483
25	dubstep	0.518087	0.758969	42.317

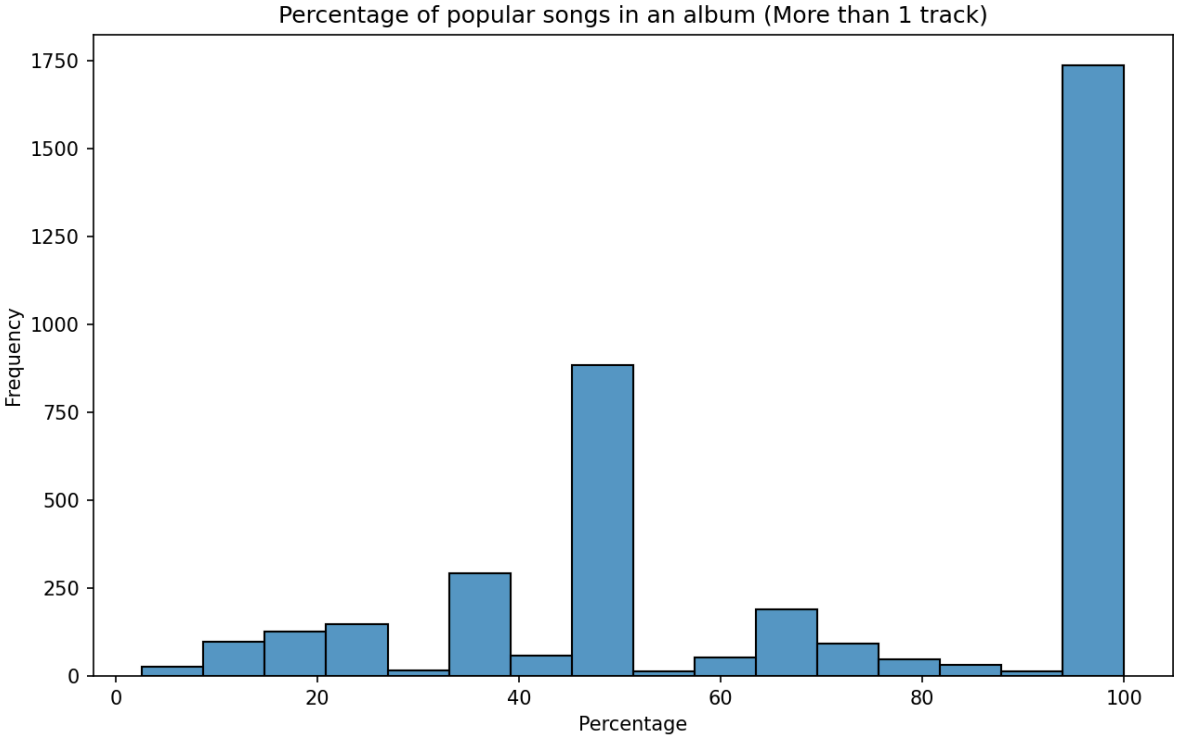
B-1



B-2


Average popularity of albums after removing the popular songs
Frequency
Average Album Popularity
0
50
100
150
200
250
300
350
400
0
10
20
30
40
50

B-3


Percentage of popular songs in an album (More than 1 track)
Frequency
Percentage
0
250
500
750
1000
1250
1500
1750
0
20
40
60
80
100

C-1