+Masters Data in Business
Assignment 2
Spotify - World Music Tracks & Characteristics

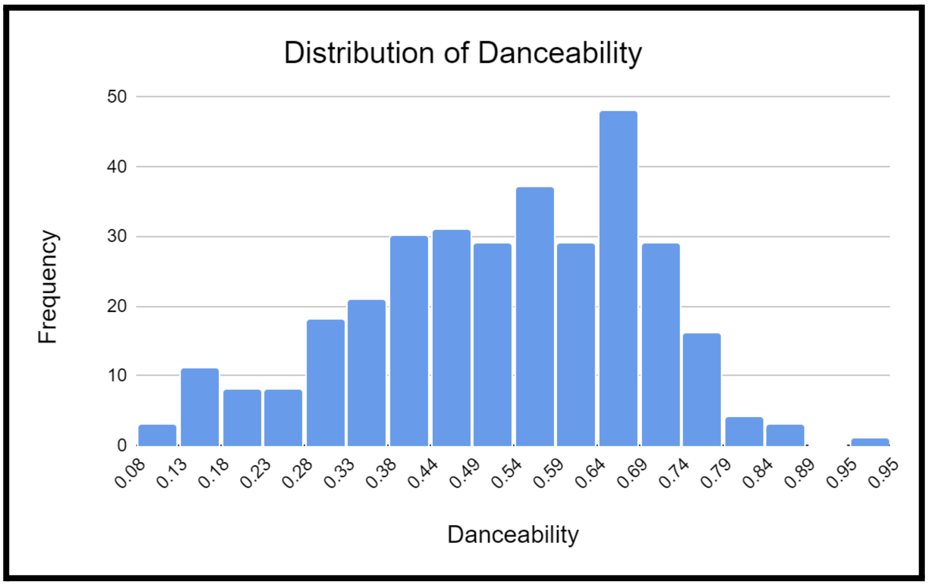
Group 2 : Gabrielle Shamu, Omone Akha, Georgia Sapsani

Task 1 - EDA (Univariate & Bivariate Analyses)

Univariate Analysis (Single Variable)

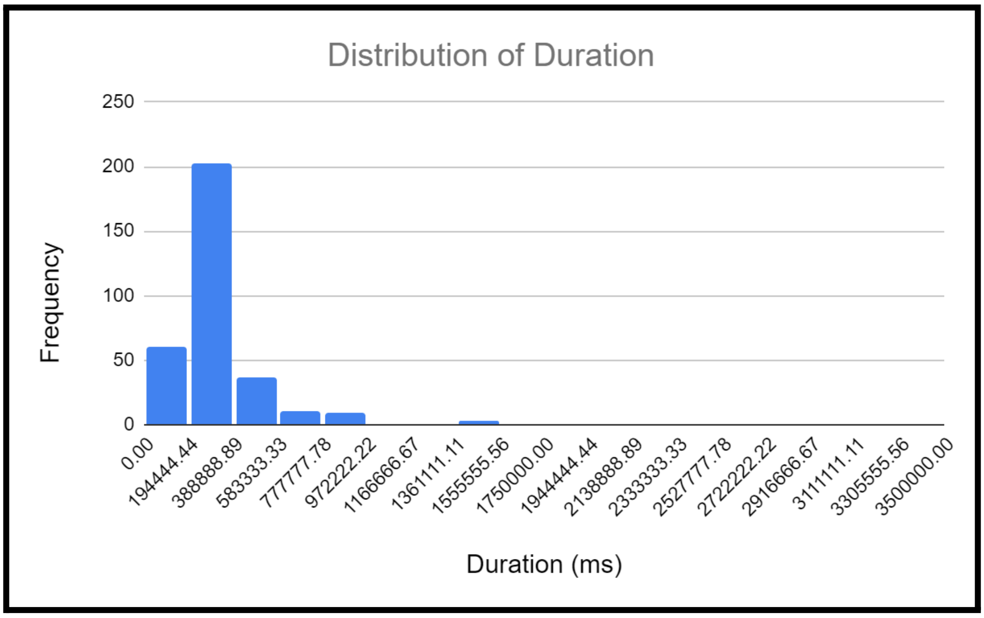
We will conduct a univariate analysis for variables *"Danceability", "Instrumentalness", "Valence"* and *"Duration"*. During this analysis, each variable's mean and range will be conducted, and the distribution will be shown in the visualisations below. The range will show us how the prices are divided between the smallest and largest values. A large range indicates that the data is spread out widely and a small range shows that it is more tightly clustered together.

a. **Danceability (Mean: 0.52, Range:0.8672)**



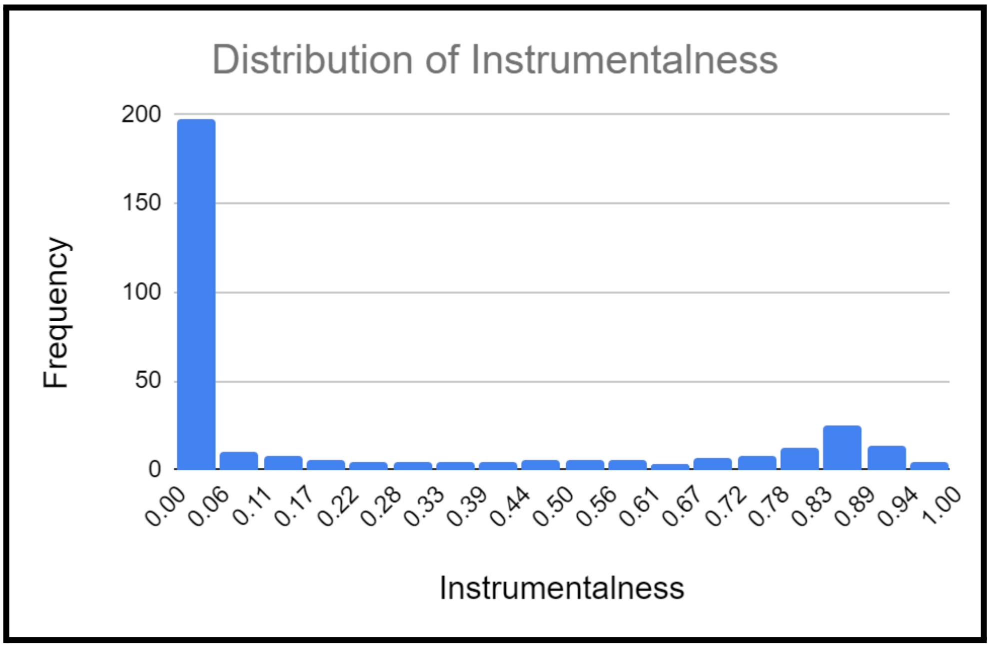
The histogram above shows that the majority of tracks have moderate to high danceability, with a roughly normal distribution between 0.08 and 0.95 (range).

b. **Duration (Mean: 333706.15 ms, Range: 3018770ms)**



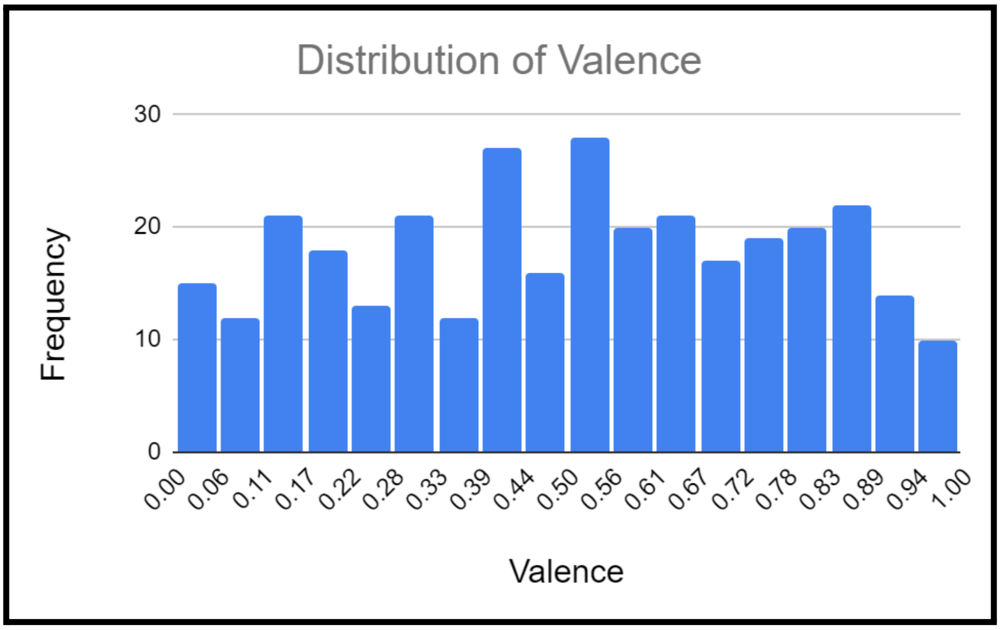
There is a right-skewed distribution for "Duration", which clearly shows that most tracks have a shorter duration.

c. Instrumentallness (Mean: 0.24, Range: 0.967)



The distribution of "Instrumentalness" is skewed heavily towards lower values in a range of 0.00 and 1.00. The histogram shows above that there are a few tracks heavily instrumental.

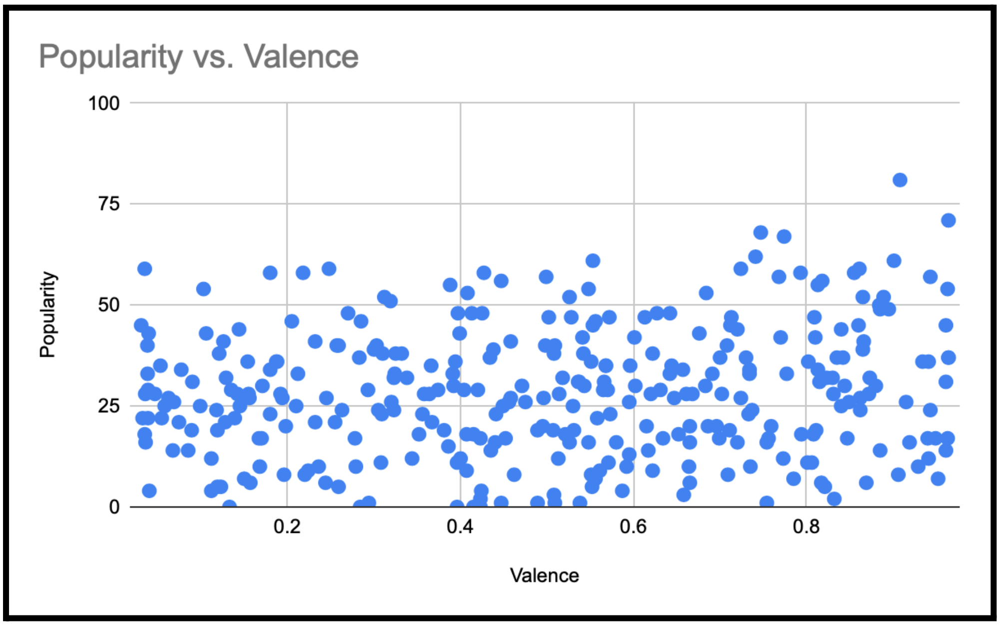
d. Valence (Mean: 0.50, Range: 0.9331)



Last, “Valence” appears to have a symmetrical distribution around the mean value, indicating a balance between positive and negative moods.

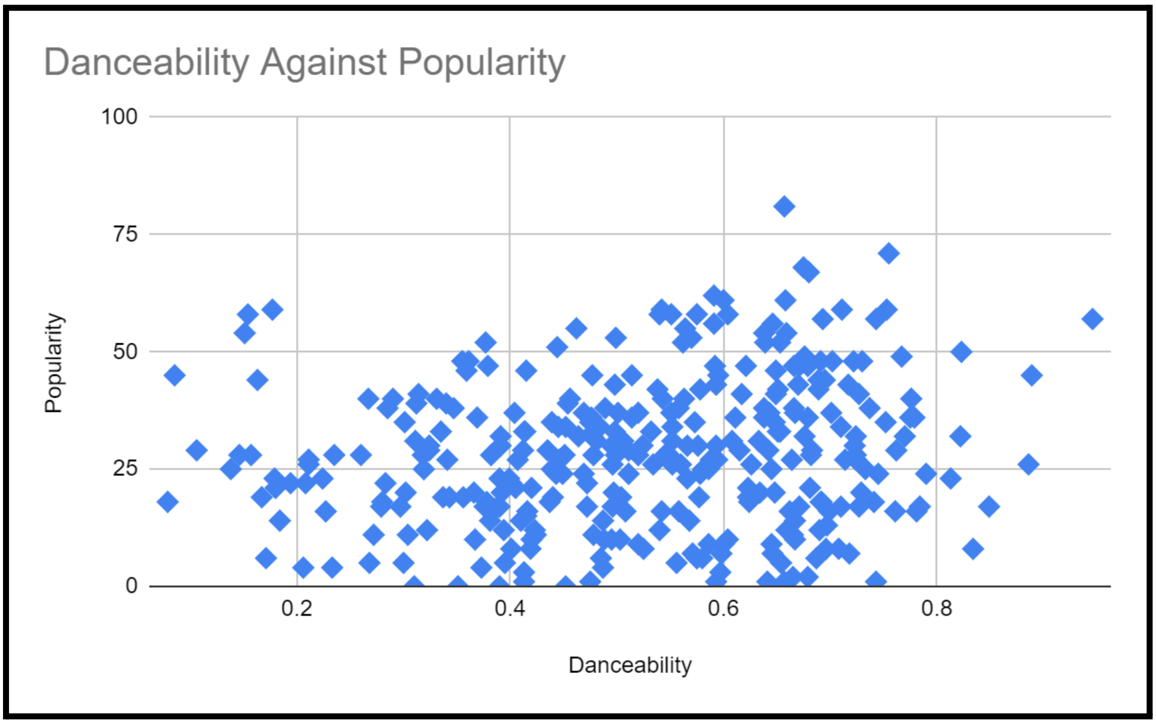
Bivariate Analysis (Two Variables)

a. Popularity vs Valence



The two chosen variables are popularity and valence, looking at the relationship between them and how it has an effect on users' musical preferences. This graph highlights that there is no correlation between the two variables, meaning Valence (how happy/cheerful a song is) does not play a role in the popularity of the music.

b. Popularity vs Danceability



The two chosen variables were danceability and popularity, looking at whether danceability affects the user's choice of music. The graph shows a positive correlation between these two variables. This signifies the importance of danceability to the popularity of music.

Task 2 - Statistical Tests

The following are the statistical tests that were conducted for this dataset.
The level of significance when conducting this hypothesis testing is 0.05.

Duration:

- Null Hypothesis (H_0): There is no significant difference between the hypothesised mean of 300,000 and the actual mean of the duration.
- Alternate Hypothesis (H_1): There is a significant difference between the hypothesised mean of 300,000 and the actual duration mean.
- T-Statistic: 2.274
- P-Value: 0.024
- Conclusion: Since the p-value is less than 0.05, reject the null hypothesis. The mean duration is significantly different from 300,000.

Danceability:

- Null Hypothesis (H_0): There is no significant difference between the hypothesised mean of 0.5 and the actual mean of danceability.
- Alternate Hypothesis (H_1): There is a significant difference between the hypothesised mean of 0.5 and the actual mean of danceability.
- T-Statistic: 2.374
- P-Value: 0.018

- Conclusion: Reject the null hypothesis. The mean danceability is significantly different from 0.5.

Energy:

- Null Hypothesis (H_0): There is no significant difference between the hypothesised mean of 0.5 and the actual mean of energy.
- Alternate Hypothesis (H_1): There is a significant difference between the hypothesised mean of 0.5 and the actual mean of energy.
- T-Statistic: 0.146
- P-Value: 0.884
- Conclusion: Fail to reject the null hypothesis. There is no significant difference between the mean energy and 0.5.

Key:

- Null Hypothesis (H_0): There is no significant difference between the hypothesised mean of 6 and the actual mean of the key.
- Alternate Hypothesis (H_1): There is a significant difference between the hypothesised mean of 6 and the actual mean of the key.
- T-Statistic: -2.835
- P-Value: 0.005
- Conclusion: Reject the null hypothesis. The mean key is significantly different from 6.

Loudness:

- Null Hypothesis (H_0): There is no significant difference between the hypothesised mean of -13 and the actual mean of loudness.
- Alternate Hypothesis (H_1): There is a significant difference between the hypothesised mean of -13 and the actual mean of loudness.
- T-Statistic: 4.207
- P-Value: <0.001
- Conclusion: Reject the null hypothesis. The mean loudness is significantly different from -13.

Speechiness:

- Null Hypothesis (H_0): There is no significant difference between the hypothesised mean of 0.05 and the actual mean of speechiness.
- Alternate Hypothesis (H_1): There is a significant difference between the hypothesised mean of 0.05 and the actual mean of speechiness.
- T-Statistic: 3.110
- P-Value: 0.002
- Conclusion: Reject the null hypothesis. The mean speechiness is significantly different from 0.05.

Acousticness:

- Null Hypothesis (H_0): There is no significant difference between the hypothesised mean of 0.5 and the actual mean of acousticness.
- Alternate Hypothesis (H_1): There is a significant difference between the hypothesised mean of 0.5 and the actual mean of acousticness.
- T-Statistic: 1.060
- P-Value: 0.290
- Conclusion: Fail to reject the null hypothesis. There is no significant difference between the mean acousticness and 0.5.

Instrumentalness:

- Null Hypothesis (H_0): There is no significant difference between the hypothesised mean of 0.3 and the actual mean of instrumentalness.

- Alternate Hypothesis (H_1): There is a significant difference between the hypothesised mean of 0.3 and the actual mean of instrumentalness.
- T-Statistic: -3.133
- P-Value: 0.002
- Conclusion: Reject the null hypothesis. The mean instrumentalness is significantly different from 0.3.

Liveness:
- Null Hypothesis (H_0): There is no significant difference between the hypothesised mean of 0.5 and the actual mean of liveness.
- Alternate Hypothesis (H_1): There is a significant difference between the hypothesised mean of 0.5 and the actual mean of liveness.
- T-Statistic: -24.414
- P-Value: <0.001
- Conclusion: Reject the null hypothesis. The mean liveness is significantly different from 0.5.

Valence:
- Null Hypothesis (H_0): There is no significant difference between the hypothesised mean of 0.5 and the actual mean of valence.
- Alternate Hypothesis (H_1): There is a significant difference between the hypothesised mean of 0.5 and the actual mean of valence.
- T-Statistic: 0.242
- P-Value: 0.809
- Conclusion: Fail to reject the null hypothesis. There is no significant difference between the mean valence and 0.5.

Tempo:
- Null Hypothesis (H_0): There is no significant difference between the hypothesised mean of 120 and the actual mean of tempo.
- Alternate Hypothesis (H_1): There is a significant difference between the hypothesised mean of 120 and the actual mean of tempo.
- T-Statistic: -2.588
- P-Value: 0.010
- Conclusion: Reject the null hypothesis. The mean tempo is significantly different from 120.

Popularity:
- Null Hypothesis (H_0): There is no significant difference between the hypothesised mean of 25 and the actual mean of popularity.
- Alternate Hypothesis (H_1): There is a significant difference between the hypothesised mean of 25 and the actual mean of popularity.
- T-Statistic: 3.792
- P-Value: <0.001
- Conclusion: Reject the null hypothesis. The mean popularity is significantly different from 25.

Based on the results, we can conclude that for most variables (e.g., duration, danceability, loudness, etc.), the mean is significantly different from the hypothesised values. However, for energy, acousticness, and valence, there is no significant difference from the hypothesised means, as we failed to reject the null hypothesis.

For more detail on the specific values used during the statistical test, check the Appendix.

Task 3 - Correlations & Regression Analysis

Correlations

The following *correlations* show the relationship between the variables of our database. Although we didn't identify any strong significant correlations, there is a positive association between loudness and danceability which shows that the louder a song, the more suitable it is for dancing.

Also, there is a negative relationship between acousticness and danceability, which shows that songs that solely or primarily use instruments are less suitable for dancing. There is also a negative relationship between instrumentalness and danceability. This shows that music without any vocals is less ideal for dancing. These two relationships show that the most popular music possibly has vocals and instruments playing simultaneously.
The negative association between liveness and danceability shows that the more the presence of an audience in the recording, the less suitable it is for dancing. It possibly has to do with the clarity of the song as any additional noise in the track makes it less attractive for people to dance.

Moving to popularity, the correlations are not strong but songs that last long (duration), have increased liveness and more spoken words (speechiness) are less popular. On the other hand, songs that are more suitable for dancing and are higher in valence, tend to be more popular. Valence in Spotify[1] describes the musical positiveness conveyed by a track. For example, tracks with high valence sound more positive (happy, cheerful), while tracks with low valence sound more negative (sad, depressed, angry).

Table 1. Correlations

	Correlations
Duration vs Time_Signature	0
Duration vs Danceability	-0.04
Duration vs Energy	-0.01
Loudness vs Danceability	**0.24**
Acousticness vs Danceability	**-0.23**
Instrumentalness vs Danceability	**-0.21**
Liveness vs Danceability	**-0.16**
Tempo vs Danceability	0.01
Popularity vs Duration	**-0.13**
Popularity vs Danceability	**0.16**
Popularity vs Liveness	**-0.14**
Popularity vs Tempo	-0.02
Popularity vs Valence	**0.13**
Popularity vs Instrumentalness	-0.09
Popularity vs Acousticness	0.03

Popularity vs Speechiness	**-0.13**
Popularity vs Mode	-0.05
Popularity vs Energy	-0.01
Popularity vs Key	0.06

Regression Analysis

Based on the variables in our database, we chose *Multiple Linear Regression* for our analysis. Our dependent variable will be "Popularity" and independent "Danceability", "Energy", "Tempo", "Loudness", "Valence", "Acousticness", and "Speechness"

Table 2. Regression Statistics

SUMMARY OUTPUT	
Regression Statistics	
Multiple R	**0.25668898**
R Square	**0.065889232**
Adjusted R Square	0.045327046
Standard Error	15.59459599
Observations	326

Multiple R shows a weak positive correlation between the dependent variable "Popularity" and independent ones. The low value of R Square shows that only 6.59% of the variation in "Popularity" is explained by the independent values in the model, which shows that other factors not included in the model may affect "Popularity".

Table 3. ANOVA (Analysis of Variance)

ANOVA					
	df	SS	MS	F	Significance F
Regression	7	5454.958427	779.2797753	3.204388387	0.002702566
Residual	318	77334.87286	243.1914241		
Total	325	82789.83129			

Table 3 will help us evaluate the overall significance of the regression model. Significance F represents the p-value for the F statistic. The value of 0.00 is less than 0.05 so the model is *statistically significant*, meaning that at least one of the independent values contributes to explaining "Popularity".

Table 4. Coefficient Table

	Coefficients	Standard Error	t Stat	**P-value**	Lower 95%	Upper 95%	Lower 95%	Upper 95%
Intercept	31.14274581	7.998671053	3.893490006	0.000120436	15.4057454	46.87974622	15.4057454	46.87974622
Danceability	**12.30046308**	6.371735878	1.930472844	**0.054437192**	-0.235620658	24.83654682	-0.235620658	24.83654682
Energy (Intensity)	**-10.71071681**	6.886129942	-1.5554044	**0.120844427**	-24.25884612	2.837412497	-24.25884612	2.837412497
Tempo	**-0.00303952**	0.031632957	-0.096087138	**0.923511868**	-0.06527584	0.0591968	-0.06527584	0.0591968
Loudness (dB)	**0.474729661**	0.261739217	1.813750595	**0.070658585**	-0.040229654	0.989688975	-0.040229654	0.989688975
Valence	**6.471961664**	4.401226834	1.470490367	**0.142417737**	-2.18724035	15.13116368	-2.18724035	15.13116368
Acousticness	**1.344195124**	3.654947052	0.367774171	**0.713286253**	-5.84673733	8.535127578	-5.84673733	8.535127578
Speechiness	**-30.13406813**	13.77409495	-2.187734892	**0.02941655**	-57.23393714	-3.034199108	-57.23393714	-3.034199108

"Danceability" appears to have a positive significant impact on "Popularity" (tracks more suitable for dancing the more popular they are). *"Energy"* has a negative but not statistically significant relationship with the dependent value of our model.

Moreover, *"Tempo"* has a very weak and insignificant impact on "Popularity". The variable *"Loudness"* has shown a positive relationship with "Popularity" with borderline significance (a small positive impact). A positive but not significant relationship appears between "Popularity" and variables *"Valence"* and *"Acousticness"* which means that both variables are not significant predictors of "Popularity". Last, *"Speechiness"* which represents the spoken words in a track, appears to have a statistically significant negative effect on "Popularity", the fewer spoken words the more popular is the track.

Task 4 - Potential Recommendations

Problem: Music Apps have a lack of tailoring towards user preferences

How is this shown: Many music apps such as **Apple Music, Spotify,** and **Deezer** offer some level of personalised recommendations through algorithms to help gauge users' musical interests. Music apps rely on algorithms to make personalised recommendations to users. However, these algorithms tend to miss the mark. A user may have a diverse range of genres but the algorithm could skew towards recommending one genre. "For example, if a user is listening to a rock playlist and then they're recommended pop music, it doesn't sit well. For others, the annoyance stems from personalised playlists going off-topic and playing different genres" (Emily Smith, 2022).

Algorithms pertain to having limited contextual awareness. Algorithms revolve around listening patterns as opposed to **context such as location, activity, and mood.** For example, a user may listen to music with high **danceability** while at a party and a user may listen to music with high **energy** while working out. Music apps such as Apple Music and Spotify lack such context within their algorithms and if these apps lack such understanding, the recommendations may have a disconnect from users immediate preferences.

Music apps rely on popularity metrics in their recommendation engines, "music recommendation algorithms tend to favour already popular creators"(gov. uk, 2023). This leads to users being recommended songs that are at the top of the charts, by popular music creators as opposed to their musical preferences. **This brings us back to our problem**, even with systems in place such as algorithms, data sets and machine learning to aid in awareness of individual users' musical tastes, a lack of tailoring towards user preferences within music apps still remains.

Why is this a problem:
This is a big problem for **Stakeholders.** If users are given irrelevant/ repetitive recommendations and fail to cater towards a user's nuanced preferences it will lead to frustration and then disengagement. This could lead to a high turnover of subscription cancellations, due to the lack of proper user personalisation which will directly impact the revenue of these platforms. If users on paid plans feel as though the music apps do not offer proper personalisation or a significantly better experience than free plans within these music apps, they may opt out of paid subscriptions.

There will be a reduction in user loyalty if such neglect of users' preferences continue. Personalisation and recommendation algorithms foster a connection between the user and the platform as the user feels like the music apps "understand them". When this is absent, and there is a sort of laziness within these recommendation algorithms, recommending top songs and artists as opposed to what a user really wants to engage with, reduces user loyalty, meaning a user may feel more comfortable moving to another music app which offers a more personalised experience as opposed to the ones that still fail to accurately do this.

As well as algorithms, music apps use data to help understand users' interests. They collect data on a user's listening habits, likes, skips, etc and song characteristics (the variables mentioned in the dataset i.e. danceability, energy, loudness, etc). Music apps are failing to accurately use this data meaning they are under leveraging **user data**. For example, variables such as **energy, danceability, tempo and valence** are big indicators of the context, mood, and vibe of a track. If these factors were used effectively, they can help music apps predict the type of music/songs users may like. Music apps such as Apple Music and Deezer's failure to enhance and make use of such variables has led to a weakened user experience, as well as a loss of competitive edge when you compare these musical apps to apps that offer more of a personalised experience.

What does the data above represent/show :

- The data above represents the variables that contribute to a user's musical preferences. The dataset above looks at variables such as “Duration”, “Energy”, “Acousticness”, “Tempo, “Instrumentalness”, “Valence”, “Popularity” etc. As mentioned above, music apps rely on **popularity metrics** in their recommendation engines. This is essential to know when it comes to the dataset mentioned. Despite music apps having algorithms that study a user's musical interests, they still tend to recommend users songs, genres, and artists that are popular as opposed to what the user likes.
- Looking into the data set, focusing on the variable of **popularity**, we can see that:

<u>Weak negative correlation:</u>
Popularity vs Duration has a weak negative correlation, less popular songs tend to be longer in duration.
Popularity vs Liveliness, live tracks are less popular this could be due to studio tracks sounding more polished/cleaner
Popularity vs Instrumentalness, instrumental songs appear to be less popular, these could mean audiences prefer vocal tracks
Popularity vs Speechness, songs with more spoken words are less popular,

<u>Weak positive correlation:</u>
Popularity vs Danceability, more danceable songs are slightly more popular suggesting upbeat music appeals to broader audiences
Popularity vs Valence, happier sounding songs are more popular, this could be due to them invoking happier emotions
Popularity vs Key, the key of the song does not significantly influence popularity

<u>Lacks significance:</u>
Popularity vs Energy, energy does not have an impact on popularity

<u>No correlation:</u>

Popularity vs Tempo, tempo has no effect on popularity
Popularity vs Acousticness, acousticness has no effect on popularity
Popularity vs Mode, mode has no effect on popularity
Popularity vs Energy, energy has no effect on popularity

- **Regression statistics:** Multiple R shows a weak positive correlation between the dependent variable “Popularity” and independent ones (Danceability”, “Energy”, “Tempo”, “Loudness”, “Valence”, “Acousticness”, and “Speechness”)

- **Coefficient table:** Danceability" appears to have a positive significant impact on "Popularity" . "Tempo" has a very weak and insignificant impact on "Popularity". "Loudness" has shown a positive relationship with "Popularity"
- **Statistics:** there is strong evidence that the average popularity of songs in this dataset are different from the hypothesised value of 25.

Insights from the data:

- This dataset highlights a disconnect between what is popular and what users may truly want. For example, danceability and high in valence songs are favourable, however songs with instrumentals, longer in duration, or liveliness may appeal to a more niche audience. That shouldn't be disregarded just because these variables **score low when comparing it to popularity**. This implies that music apps may neglect those who prefer longer/less popular music which could lead to dissatisfaction across users.
- Music apps should leverage the data from looking into correlations, in turn bettering tailored recommendations. For example if users prefer music high in danceability and valence, algorithms could prioritise those attributes while providing room for instrumental and/or longer tracks based on user history and patterns. By simply pushing the most popular songs, genres, artists etc, music apps are missing the opportunity to engage with users who have evolving and diverse tastes.
- The correlation data places emphasis on the importance of not just focusing on what is popular, but also having a broader understanding of user preferences. If music apps rely solely on popularity metrics, they run the risk of alienating users who enjoy exploring different genres. By analysing user preferences and **variables beyond popularity**, they can offer a more tailored experience, resonating with individual users which could potentially improve user engagement.

Solutions from the insights:

- The app could use features such as energy or acousticness to suggest songs that fit a users preferences or mood over time. For example, if a user has a tendency of listening to slow tempo, acoustic songs at night and songs high in energy and danceability during the day, the app could tailor night time playlists to include acoustic tracks with a low valence and a daytime playlist to include songs with high levels of energy and valence during the day. Music apps that do not take advantage of such insights are under utilising their data limiting their **potential of user base growth and loyalty.**

- Utilising a broader range of song characteristics (i.e danceability, energy, valence, acousticness) within recommendation algorithms to move beyond the scope of popularity. This would lead to more personalised suggestions, resonating with users.

- Algorithms derived from contextual awareness, having context on a user's location, mood and activity will provide more relevant recommendations. For example if a user is at a gym, they may prefer more high energy music while users who are at home may prefer music low in energy. Morning users may prefer music that is upbeat and high in valence when starting their day, whereas night time users may prefer music that is more relaxed and high in instrumentals. Providing music that immediately fits a user's situation would lead to an increase in engagement.

How do USERS benefit from these solutions:

- **There will be improved personalisation:** analysing the datasets variables such as “energy, valence, danceability” means the algorithms in these apps can understand users musical preferences, offering recommendations that accurately align with their tastes. For example, if a user enjoys high temp, high energy and high liveliness in their music, the app can consistently recommend tracks that have such variables.
- **Better user experience:** with a more nuanced recommendation system, users will not have to worry about irrelevant tracks being recommended to them that do not match their activity/current mood. This inturn minimises frustration and maximises a user's enjoyment.
- **Aids in evolving preferences:** most users' musical interests change over time, if these musical apps stay up to date with analysing user metrics leveraging song characteristics such as “energy”, “speechiness”, “liveness’ etc helps music apps keep up with these changes. This will aid in the users experience, as users will feel as though the app is continuously adapting to their needs and evolving preferences.

How do STAKEHOLDERS benefit from these solutions:

- Stakeholders could use these variables from the above dataset to make more informed decisions about what types of music they promote, as well as possible feature recommendations/improvements. For example, having an awareness of what users prefer high energy and/or high danceability music during the weekends could lead to developing and recommending those users with a playlist tailored for the weekend. It could also lead to partnerships with artists who produce these types of songs. Not only would this boost user engagement, but it reinforces the notion that music apps that take effort in understanding their users and provide a rich personalised experience produce more **new users** as well as existing **loyal users** in comparison to music apps that fail to take advantage of such data. Making an adaptation towards the use of algorithms will bring more **revenue** to stakeholders because of this.

- Having these types of datasets benefits those who play an important role in product development such as Product Managers. They can use the insights from the dataset to aid with their feature prioritisation, prioritising features that focus on a users behavioural patterns. For example, the dataset above shows that songs that are more suitable for dancing and are higher in valence, tend to be more popular. This might lead to stakeholders investing in refining algorithms to properly tailor to user preferences, as well as creating mood-based playlists. Improving their recommendation algorithms and placing systems that help provide a more personal tailoring to user preferences helps grow the user base, bringing more **revenue** for stakeholders.
- Having such a rich dataset like the one provided, stakeholders can use the data to offer personalised advertisements/ sponsorship opportunities. For example, brands/advertisers who are looking to reach users who enjoy high energy, high tempo music could precisely target

their advertisements towards them increasing the effectiveness of such advertisements as well as **revenue** for stakeholders.

- By keeping track of user preferences and analysing how they perform over time (for example, there may be a shift from high energy, high danceability music to low energy, low valence music). Keeping up to date with these variables and the status of them when it comes to users' listening habits could help stakeholders anticipate market trends and make more proactive decisions to adapt to the changing of user interests/demands.

- **Conclusion:** This analysis shows that although popularity is a factor in musical consumption, it is of absolute necessity that music apps dig deeper into user preferences. Leveraging detailed insights from this dataset can enhance their recommendation systems offering a wider range of music that aligns closely to user tastes. This will lead to a more engaging experience, a growth in the user base and increased user loyalty.

Appendix

Table with specific values for the Statistical Tests (Task 2)

	Duration	Danceability	Energy	Key	Loudness
Mean	333706.1503	0.5223806748	0.5020915031	5.438650307	-11.7997546
Standard Deviation	268048.5548	0.1704691616	0.2592828075	3.580405223	5.159427267
Observation	327	327	327	327	327
Hypothesis	300000	0.5	0.5	6	-13
df	326	326	326	326	326
T Statistic	2.273890564	2.374112141	0.145867541	-2.835144405	4.206708621
P Value	0.02362232357	0.01816998843	0.8841161079	0.004866472626	0.00003352877931

	Speechiness	Acousticness	Instrumentalness	Liveness	Valence
Mean	0.06130521472	0.5194457336	0.2402637402	0.2153371166	0.5036242331
Standard Deviation	0.06573369306	0.3317818873	0.3448208242	0.2108457847	0.2703868455
Observation	327	327	327	327	327
Hypothesis	0.05	0.5	0.3	0.5	0.5
df	326	326	326	326	326
T Statistic	3.110030579	1.059852757	-3.132697192	-24.4140482	0.2423842762
P Value	0.00203576225	0.2899961992	0.00188919799	0	0.8086348322

	Tempo	Popularity
Mean	115.9684601	28.34662577
Standard Deviation	28.17410901	15.96051197
Observation	327	327
Hypothesis	120	25
df	326	326
T Statistic	-2.587585123	3.791702095
P Value	0.01009744255	0.0001782485721

References:

1. House of Commons Digital, Culture, Media and Sport Committee, *"Economics of music streaming: Government and Competition and Markets Authority Responses to Committee's Second Report"*, 2021.
2. The Future of Music Streaming Survey, Section 3, Q2, Q3.
3. **The CDEI, *"The impact of recommendation algorithms on the UK's music industry: creators' survey results"*, Section 3, Q1.**
4. brandwatch, *"**https://www.brandwatch.com/blog/pain-points-music-streaming-services/**"*, Emily Smith, 2022.