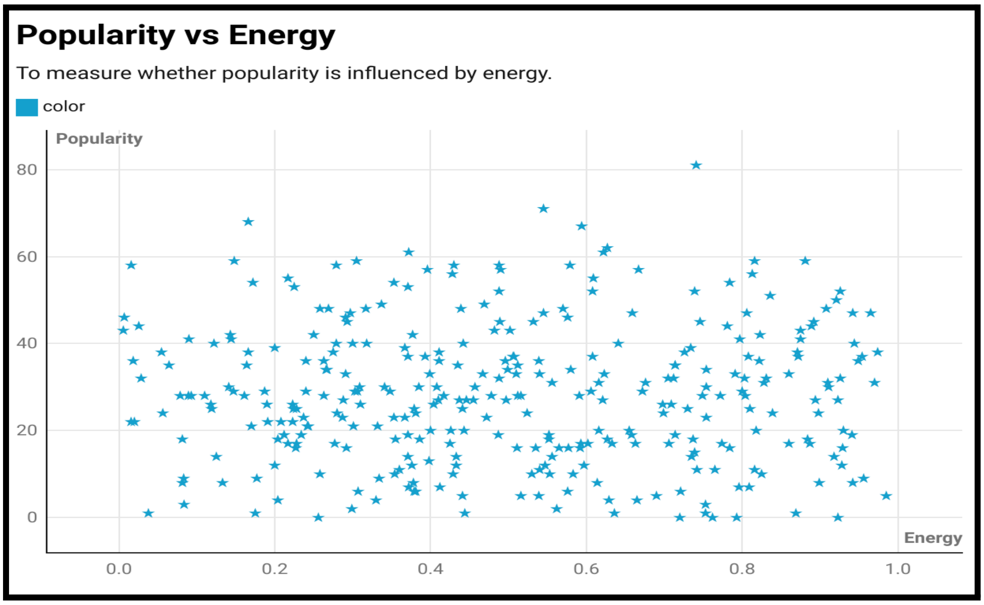
+Masters Data in Business
Assignment 3
Spotify - World Music Tracks & Characteristic Visualisations

Group 2 : Gabrielle Shamu, Omone Akha, Georgia Sapsani

Task 1 - Storytelling

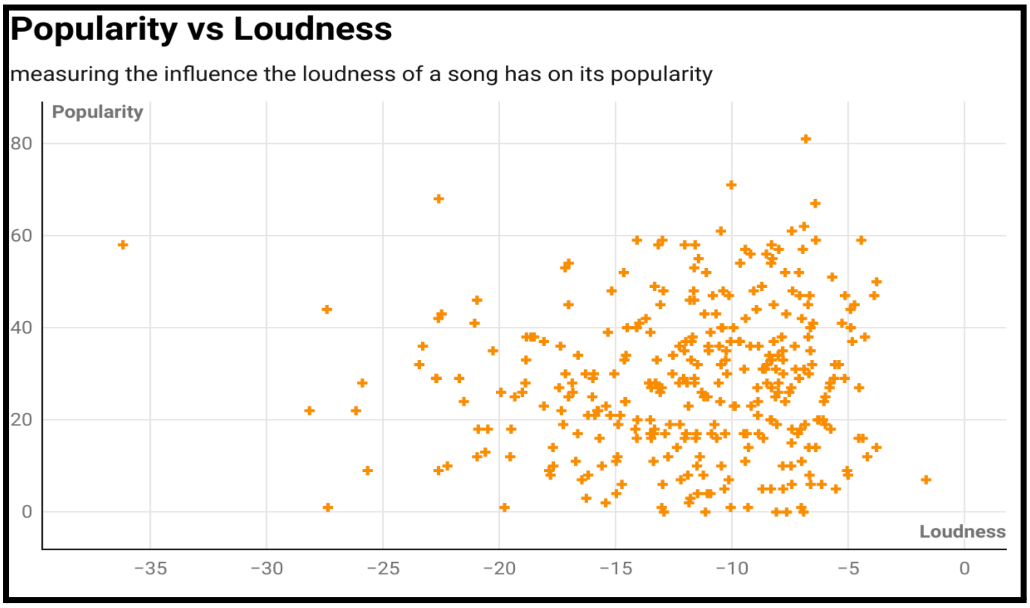
A data narrative was developed based on the key variables that have influenced the popularity of Spotify tracks.

a. Popularity vs Energy (Scatterplot)



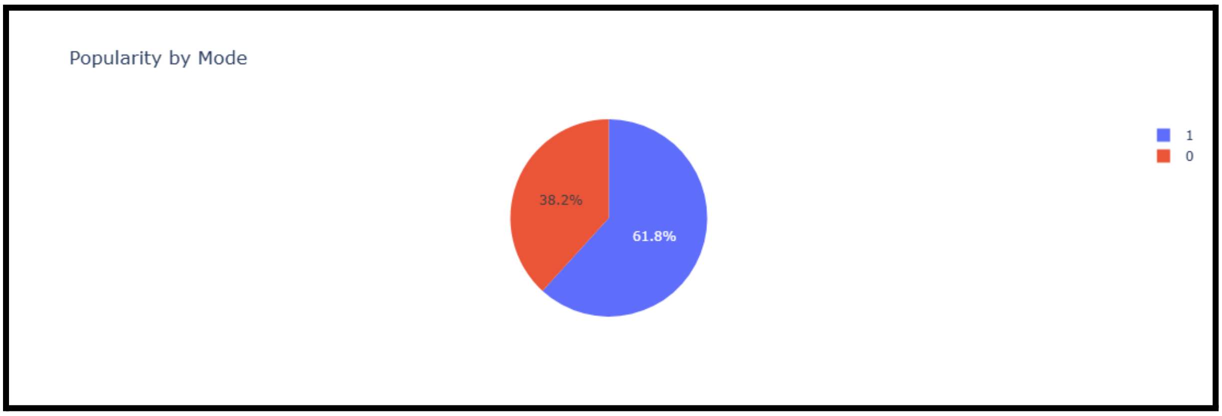
Looking at the scatterplots in this graph, the points appear to be widely spread across the plot with no clear upward or downward trend. This suggests that energy does not have a strong linear relationship with popularity as there doesn't appear to be a significant positive or negative correlation. This implies that popularity is influenced by a range of factors beyond the energy of the music. This spread of data suggests that Energy is not a dominant factor in determining a song's popularity on its own. Even though no strong linear relationship is evident, outliers/ clusters of points could indicate specific trends within certain energy ranges. Taking the time to identify and analyse these outliers (for example the plot on **0.78, 81**) could help uncover niche patterns, such as certain high-energy songs gaining viral popularity or low-energy ballads dominating a particular segment of listeners.

b. Popularity vs Loudness (Scatterplot)



The graph shows a slight trend where, as the loudness of a song increases (moves from -35 decibel to 0 decibel on the x-axis), the popularity (on the y-axis) also appears to slightly increase in some regions. However, this relationship doesn't appear to be strong or linear. There's significant spread and clustering across different ranges of loudness and popularity, indicating a weak/ no clear correlation. There is sparse data from -20 and below, this indicates that very quiet songs tend to be less popular meaning songs with low loudness (below -20) seem to have lower popularity, while most of the popular songs fall between -15 and -5 loudness. This enforces the idea that most popular songs fall within this range of loudness, this clustering may suggest that while loudness may not drive popularity, songs within the -15 and -15 range tend to be likely to achieve higher popularity.

c. Popularity - Mode (Pie Chart)

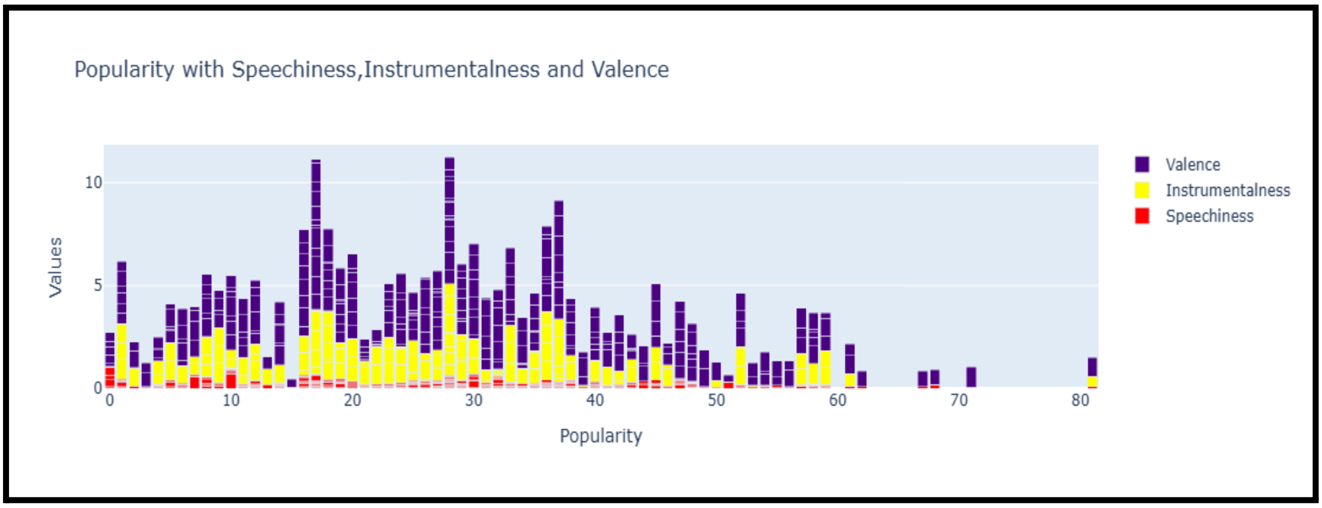


The first chart is a pie chart that breaks down the popularity of tracks by their mode, which refers to whether the song is in the major or minor key. Mode is a musical concept that defines the overall tone and feeling of the track.

- **Mode 1 (blue):** Represents songs in a major key typically associated with happier, more upbeat sounds. A large portion of the pie (61.8%) indicates that most popular songs are in major keys, suggesting that audiences generally prefer uplifting, positive-sounding music.
- **Mode 0 (red):** Represents songs in a minor key, which often evoke feelings of sadness, tension, or introspection. Even though minor-key songs only make up 38.2% of the distribution, they still hold significant weight in popular music, indicating that more melancholic tracks also find substantial success among listeners.

This chart reinforces the idea that while upbeat major-key songs are more popular overall, there is still a strong demand for the emotional depth and complexity offered by minor-key tracks.

d. Popularity - Speechness, Instrumentalness, Valence (Stacked Bar Chart)

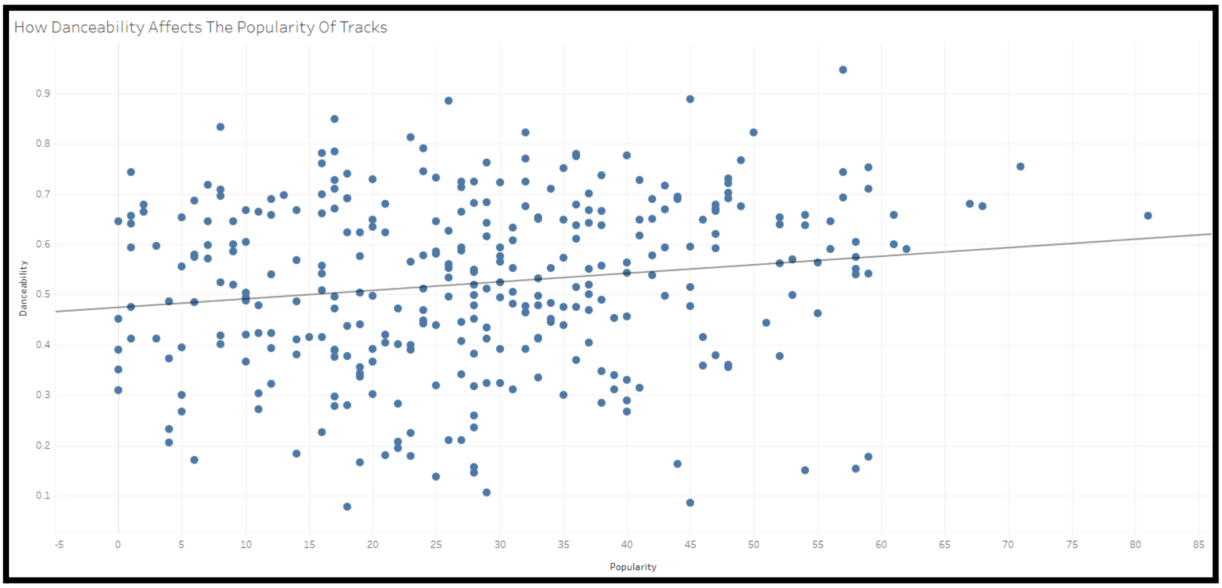


The second chart displays the relationship between a song's popularity (x-axis) and three musical attributes: Speechiness, Instrumentalness, and Valence. These attributes are stacked as bars along the popularity axis, offering insights into how they vary across songs with different popularity levels.

- **Valence (purple):** This feature describes the musical positivity conveyed in a track. Higher valence scores suggest happier, more positive songs, while lower scores indicate more negative or sadder songs. As seen in the chart, valence values as present consistently across different popularity levels, indicating that the emotional tone of a song doesn't follow a strict pattern in terms of how popular a track might become.
- **Instrumentalness (yellow):** This attribute measures the extent to which a track is purely instrumental, without vocals. Songs with higher instrumentalness appear sparsely across the popularity spectrum. However, their prominence increases between the 10 and 40 popularity ranges, which could suggest that moderately popular songs often contain instrumental sections.
- **Speechiness (red):** Speechiness detects the presence of spoken words in a track, like podcasts or vocal-heavy tracks. As highlighted in the chart, tracks with noticeable speech content tend to have very low popularity scores. These appear as small red bands towards the lower end of the popularity scale, suggesting that tracks with more speech are generally less popular.

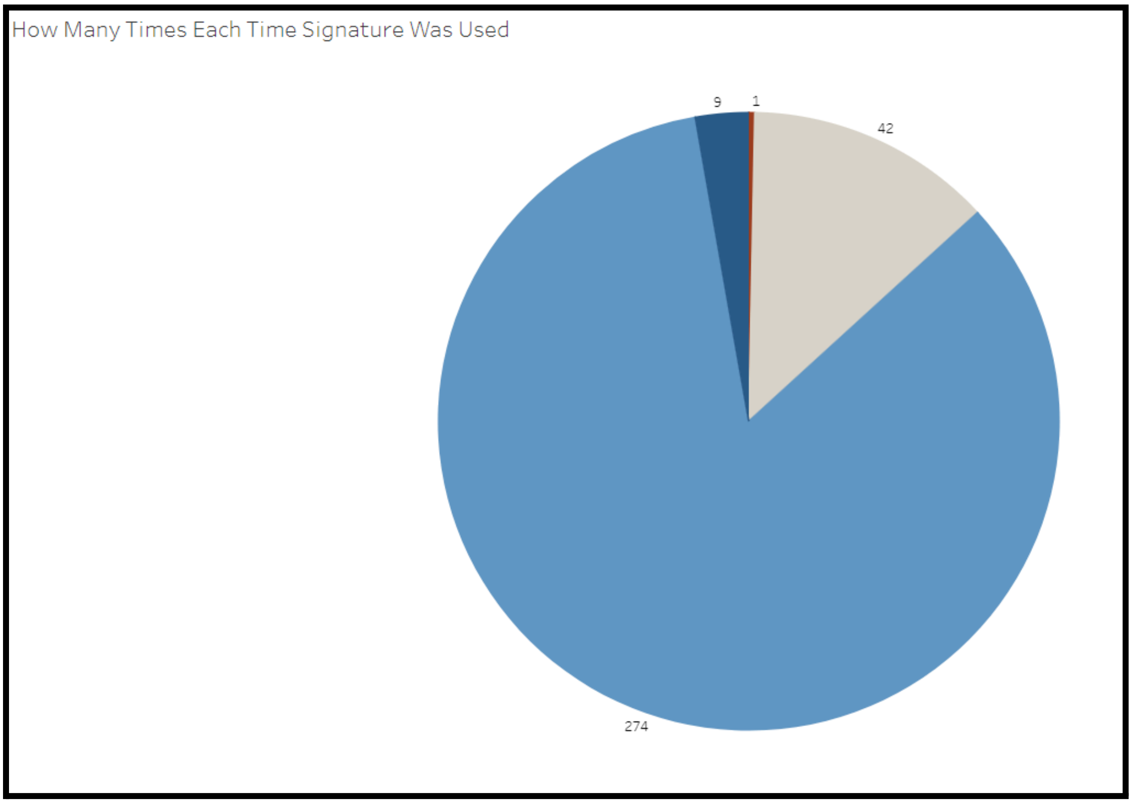
In summary, this chart suggests that while valence (emotional tone) remains relatively constant across varying levels of popularity, purely instrumental tracks or speech-heavy tend to cluster within specific popularity ranges. Instrumental tracks seem more likely to enjoy moderate popularity, whereas speech-heavy tracks tend to skew towards lower popularity.

e. Popularity vs Danceability (Scatterplot Chart)



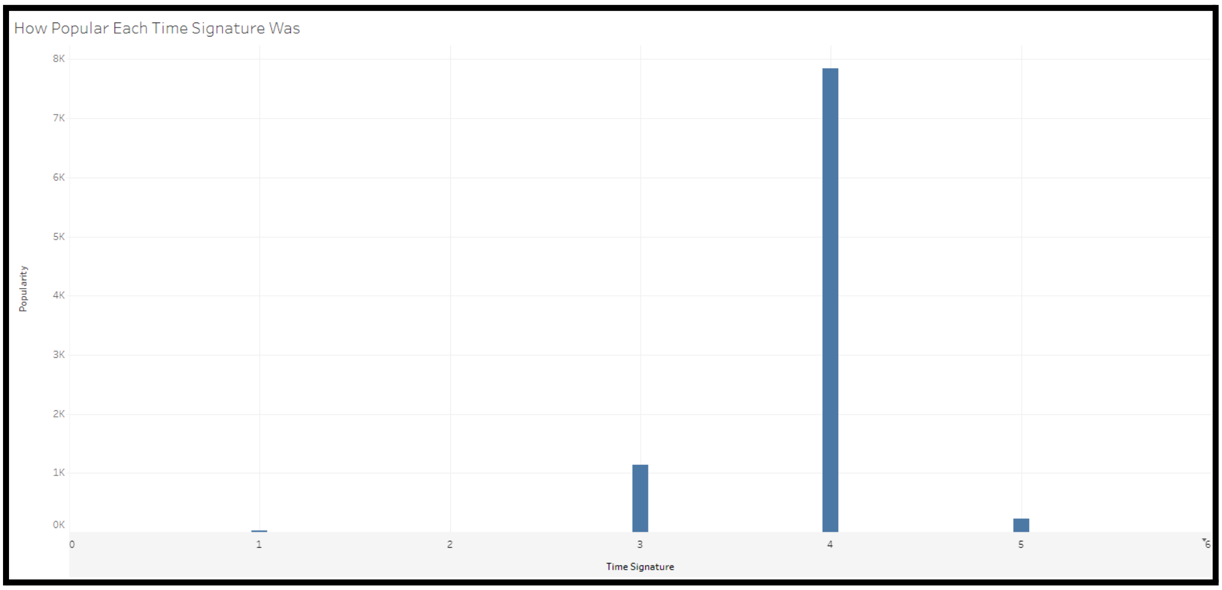
This scatterplot shows a weak positive trend where the danceability (on the y-axis) of the track increases the popularity (on the x-axis). The relationship, however, does not appear to be strong or significant and there is a lot more clustering where the danceability is between 0.4 and 0.7, showing more tracks tend to fall in this range. The tracks that fall lower are more sparsely indicated in this dataset and those that fall on the higher side are barely there but have a very high popularity rating. As a result, despite the positive trend showing that users prefer songs with a high danceability score, there is a chance that the dataset is skewed due to the lack of data. In contrast, the high popularity of these scores may be an indication of what users are looking for in the songs they listen to.

f. Count of Time Signatures within the Dataset (Pie Chart)



The pie chart shows the different time signatures that are used for the tracks within this dataset. The time signature is " The time signature (meter) is a notational convention to specify how many beats are in each bar (or measure)." (Mahajan A., 2023) The pie chart shows tracks that time signature values of 1 (1), 3 (42), 4 (274) and 5 (9). This chart is used to give context to the type of songs within this dataset and to find the most common time signature that is used among the songs found here. This dataset has songs ranging from the 1950s to the 2010s and among all those songs that have 4 beats per bar is the most common.

g. Popularity vs Time Signature (Bar Chart)



This graph shows the sum of popularity for the different time signatures for the tracks within the data collected. The most popular are songs with 4 beats per bar. While this may show that audiences prefer songs with 4 beats per bar, there is a clear skew in this data as most songs within this dataset have 4 beats per bar, with 274 showing this value, compared to the others. In contrast, the songs with the time signature of 3 is 42 and that is the second most used signature within the data.
With this in mind, there must be a reason for this clear divide and maybe because users prefer 4-beat rhythms and songs. Many popular songs use 4 beats per bar are timeless and attract audiences. These include:

- “Let It Be” by The Beatles
- “We Will Rock You” by Queen
- “One Love” by Bob Marley
- “Whole Lotta Love” by Led Zeppelin
- “The Lazy Song” by Bruno Mars
- “Leaving On A Jet Plane” by John Denver
- “Everytime We Touch” by Cascada
- “Smells Like Teen Spirit” by Nirvana
- “Chasing Cars” by Snow Patrol
- “Rolling In The Deep” by Adele
- “Hey Brother” by Avicii
- “Photograph” by Ed Sheeran
- “Love Yourself” by Justin Bieber
- “Firework” by Katy Perry
- “Uptown Funk” by Mark Ronson ft. Bruno Mars
- “Yellow” by Coldplay
- “Can’t Help Falling In Love” by Elvis Presley
- “Blurred Lines” by Robin Thicke ft. T.I., Pharrell

(Henderson, 2023)

All these songs were popular and are still listened to by audiences and used in commercials today. There is a reason for this. The 4 beat rhythm is attractive and catchy. It is a natural melody that is easy to follow and thus easy to listen to. As a result, although this dataset is skewed it does not take away from the fact that the 4-beat time signature is very popular.

Task 2 - Decision Making

As a team, we have chosen to utilize a range of tools and investigate various visualizations through Tableau, Data Wrapper, and Python (Plotly).

Popularity-Mode (Bar chart with Plotly)
Choosing a pie chart to display the distribution of popularity by mode makes sense for several reasons. It shows proportions clearly since we are dealing with two discrete categories (mode 0 and mode 1). Also, with only two categories, a pie chart simplifies the comparison and viewers can instantly see the relative size of each slice. Last, a pie chart is visually engaging and widely recognised. In short, the pie chart is the best fit here because it efficiently communicates the relative proportions of two categories, ensuring the visual remains simple, clear, and easy to interpret.

Popularity - Speechiness, Instrumentalness, Valence (Stacked Bar Chart with Plotly)
Choosing a stacked bar chart to represent popularity alongside speechiness, instrumentalness and valence is an effective choice for several key reasons. Compares multiple variables on the same axis and allows you to display multiple variables together to a single shared variable. Highlights the contribution of each variable making it clear how each feature contributes to the total “composition” of the song’s characteristics for a given popularity range. Also, a stacked bar chart contributes to efficient space usage, showing multiple layers of data without requiring a separate chart for each feature. Furthermore, spot trends and the chart allow for a layered analysis, where you can see how the presence of one attribute might correspond with the presence or absence of another attribute as popularity changes. Moreover, a stacked bar chart focuses on the relative distribution of each attribute and emphasises the combined impact, helping to show how the combination of valence, instrumentalness and speechiness contributes to a song’s popularity rather than isolating them in separate charts.

Popularity and Energy (Scatterplot with Datawrapper)
Datawrapper was used for these visualisations, a scatterplot was chosen as the focus was to look into the relationship and influence Energy and Loudness may when it comes to the popularity of a song. We want to establish whether there is a relationship between popularity and these two variables.

A scattergraph was chosen to measure the relationship between “popularity” and the “energy” of a song. With “popularity” being on the y-axis (dependent variable) and “Energy” being on the x-axis (independent variable). The colour blue was chosen for this visualisation as it is easy on the eyes and caters towards individuals who may live with colour blindness. Stars were chosen for the shape of the plots to add a twist to the graph as opposed to using circles.

Popularity and Loudness (Scatterplot with Datawrapper)

Chosen a scattergraph to measure the relationship between “Popularity” and the loudness of a song, with “popularity” being on the y-axis(dependent variable) and “Loudness” being on the x axis (independent variable). Chosen the colour orange as it is a vibrant colour that is easy to distinguish, and stands out well on a white or light background without overwhelming the eyes.

Popularity vs Danceability (Scatterplot with Tableau)
To show the relationship between these two variables effectively a scatterplot was used. This is because both variables are continuous. This scatterplot shows the correlation between the two and

enables a thorough and clear investigation of the relationship between these two items. A scatter plot makes clear the level of danceability that is most popular and how where the songs within this data set tend to fall when it comes to danceability. Since neither of these values are discrete or time-sequenced then a bar, line or pie chart would be unhelpful and more overwhelming to read, making it harder to draw insights from the data and making outliers harder to recognise.

Count of Time Signatures within the Dataset (Pie Chart with Tableau)
A pie chart was most useful for visualising this data as counting the number of time signatures within the dataset is categorical data. The pie chart effectively visualises the proportional distribution of different time signatures in the dataset, giving an immediate sense of which categories dominate and which are less represented. Thus when breaking down the proportional relationship between the different items effectively this chart is useful. The chart is used to display the frequency distribution of time signatures in the dataset, where each slice of the pie represents a distinct time signature.
Using a pie chart here is most effective as the clear readability of the most dominant and the most rare time signature within the dataset is visible making it easy to draw valuable data from.

Popularity vs Time Signature (Bar Chart with Tableau)
A bar chart was the best choice when comparing the popularity of songs across different time signatures. Time signatures are categorical variables and the bar chat was the most suited chart to show the metric of popularity across the distinct time signatures. The bar chart makes it easier to pick out trends and make comparisons between the different time signatures. A pie chart could have been used here but since this is not a proportional comparison between the popularity of different time signatures it is not the most appropriate chart to use to correlate this relationship, while a scatterplot or line plot would have been even less suitable as there are only 4 different time signatures used within this dataset so any of these graphs would have made the data hard to read.

Task 3 - Conclusions & Recommendations

Problem Formulation
The problem addressed in the analysis is the need for more ability of current music apps, such as Spotify, to provide tailored recommendations based on refined user preferences. While these platforms offer a vast array of musical choices and rely on algorithms to recommend songs, the recommendations often skew towards popular tracks rather than catering to individual user preferences, including attributes like danceability, tempo and instrumentalness. This lack of nuanced personalization creates a suboptimal user experience, as users are exposed to a limited range of tracks, which diminishes the discoverability of new music that aligns with their unique tastes. Solving this issue is essential to enhancing user engagement and loyalty, and increasing subscription retention and revenue for stakehonders.

Data Collection
The data utilised for this analysis was sourced from Spotify through Kaggle and contains information on 300 world music tracks, including key musical characteristics such as danceability, energy, loudness and popularity. These variables provide insights into how different track attributes influence user preferences. Data cleaning was conducted to remove duplicates, missing values and outliers, ensuring a more accurate dataset for analysis. This refined dataset serves as the foundation for understanding the correlations between musical characteristics and popularity, with the ultimate goal of improving music recommendation algorithms.

Data Quality/Limitations
"Data quality is defined as the degree to which data meets a company's expectations of accuracy, validity, completeness, and consistency." With this in mind ensuring the high quality of the dataset "is key to making accurate and informed" decision-making. There are many factors to consider when determining whether the data quality is high or low and many steps to go through to ensure that the quality is as high as possible.
"Completeness is defined as a measure of the percentage of data that is missing within a dataset." When going through the data set during the data cleaning process there were no missing values found and all the necessary data was present thus there was no need to go through the process of imputation and the removal of rows. To make the data more readable the titles or the columns were changed to make it more intuitive for analysis.
Accuracy "of data refers to the level at which the information is reliable and trustworthy" and whether it reflects the real world correctly. When sweeping through the data there were no errors or inaccuracies that were identified. The data validation was done by cross-referencing with the actuarial Spotify website and by comparing with other datasets that had similar data.
The next step taken was assessing the consistency of the data. This is "a measure of the percentage of data that is missing within a dataset." There was only one dataset used as what was needed was a dataset small enough to analyse and explore for the given timeframe as a result no discrepancies would show up.
The next thing checked was the timeliness of the data. "Timeliness measures how up-to-date or antiquated the data is at any given moment." The data for the tracks within this dataset was from 1950 to 2019. Though this does give enough data for the chosen topic, the analysis is for current users, which makes this dataset slightly outdated, and does not reflect the current trend of music tracks and user preferences today.

Relevance was the next data quality step to go through, which is whether the dataset has the correct information to answer the questions for the topic. The data aligns very well with the topic. The data shows different features of each song and how popular that song is. This can help figure out where user preference lies when it comes to song characteristics. The most irrelevant part of the data is the album of the song as this has nothing to do with user preference and insights for this topic can not be drawn from this part of the dataset thus it was excluded from the analysis. The irrelevant data was not removed but was identified and noted down to keep the integrity of the data while maintaining the focus on the task. (Suer, 2023)

An important limitation was the size of the database utilized, which was relatively small. A larger dataset would likely provide more comprehensive and valuable insights into the various tracks and, most importantly, the influence of different variables on the popularity of Spotify tracks.

Data Analysis

Using the dataset that has been cleaned to a high standard and quality the following analysis can be made from them to draw recommendations for the topic at hand.

In regards 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**

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”.

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	-0.0960871 38	**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.
With this along with other analysis from task 1 proper recommendations for user preferences can be deduced.

Recommendations

Overview of our problem- Music apps have a lack of tailoring towards user preferences.

How is this shown: Many music apps like Apple Music, Spotify, and Deezer use algorithms to provide personalised music recommendations based on users' listening habits. However, these algorithms often miss the mark by over emphasising certain genres or relying too heavily on patterns rather than the full range of user preferences. For instance, a user who enjoys multiple genres may find the app recommending mostly one type of music, leading to dissatisfaction. A common issue arises when users, after listening to one genre, are recommended unrelated genres, which feels off-topic and frustrating. “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” (Smith, 2022).

One of the key shortcomings of these algorithms is their lack of contextual awareness. They track listening patterns but fail to account for the user's mood, activity, or environment. For example, a person may listen to high-energy music while working out or prefer more danceable tracks at a party. Yet, apps like Spotify and Apple Music fail to factor in these situational preferences, making recommendations that feel disconnected from users’ immediate needs.

Moreover, these apps tend to rely heavily on popularity metrics. Music recommendation engines often prioritise songs and artists already trending, which results in users being recommended chart-topping tracks over those that truly match their unique tastes (gov.uk, 2023). This over-reliance on popular content creates a disconnect between users' preferences and the music they are shown, despite the vast datasets and machine learning tools available to personalise recommendations.

Why is this a problem: This lack of personalised recommendations is a significant issue for **stakeholders**. When users receive irrelevant or repetitive music suggestions that don't align with their nuanced preferences, it leads to frustration and disengagement. Over time, this can result in higher churn rates and subscription cancellations, as users feel that the service isn’t offering a significantly better experience than the free versions. If paid subscribers feel as though the app fails to deliver proper personalization, they may cancel their subscriptions, moving to competitors who take time in ensuring the user has a personalised experience. Such cancellations will have a direct impact on revenue.

Additionally, user loyalty suffers when music apps fail to tailor recommendations to individual tastes. Personalised experiences help users feel that the app “understands” them, fostering a stronger connection. When apps instead rely on lazy algorithms that promote popular music over tracks suited to a user’s actual preferences, it undermines this connection. As a result, users may be more inclined to switch to a competitor offering a more personalised service.

Beyond just algorithmic limitations, music apps are also under-leveraging the data they collect. These platforms gather valuable information about users’ listening habits, such as song skips, likes, and playlist behaviours, as well as track features like danceability, energy, tempo, and valence. These variables can be strong indicators of a song’s context or mood, yet music apps often fail to fully integrate them into the recommendation system. By neglecting to enhance and apply such data effectively, platforms like Apple Music and Deezer weaken the overall user experience and lose a competitive edge against services that offer deeper, more relevant personalization.

Overview of insights: overall, this dataset reveals a gap between what is popular and what users actually prefer. While songs with high danceability and valence are generally favoured, tracks with longer durations, instrumentals, or lower liveliness may appeal to niche audiences. These preferences shouldn't be overlooked simply because they rank lower in popularity. This suggests that music apps may be neglecting users who enjoy less mainstream or longer tracks, potentially causing dissatisfaction among a wider user base.

Recommendations:

- To improve tailored recommendations, music apps **should use correlation data more effectively**. For instance, if users tend to prefer songs with high danceability and valence, algorithms can prioritise these attributes while still incorporating instrumental or longer tracks based on individual listening patterns. By focusing only on popular songs, genres, or artists, music apps miss opportunities to engage users with diverse and evolving tastes. The correlation data underscores the need to move beyond popularity metrics and adopt a more comprehensive understanding of user preferences. Relying solely on popularity risks alienating users who enjoy exploring different styles of music. By analysing a broader range of user preferences and variables, music apps can offer a more personalised experience, enhancing user satisfaction and engagement.
- **Implement diversity metrics:** stakeholders should consider implementing diversity metrics in the recommendation system to aid in avoiding overloading users with mainstream music, such as weighing songs that better align with a users personal taste profile (such as enjoy music high in danceability and loudness) rather than basing it purely on popularity. This can help offer a more personalised experience for the user. "Tailoring recommendations to users' preferred level of diversity increases overall accuracy metrics" (Burke et al, 2022). Implementing such metrics can help increase user engagement by preventing recommendation fatigue (users growing tired of mainstream, repetitive content being recommended to them). Factoring users' individual tastes provides a richer and more satisfying experience. This approach would foster long term user retention, as users are being properly catered to.
- **Incorporate non-popularity based variables:** analysing the datasets variables such as "Loudness, 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. 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. Utilising a broader range of song characteristics, moving beyond the scope of popularity provides more personalised suggestions, resonating with users.

- **Keep up with evolving preferences:** most users' musical interests change over time, if the musical app stays 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. 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.

- **Playlist tailoring based on metrics, useful for advertising:** 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.

- **Feature prioritisation-** Having these types of datasets benefits stakeholders 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. 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.**

- **Music promotion:**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 provide a rise in **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.

Conclusion: The current focus on recommending popular/trending tracks limits the personalised experience that music apps are supposed to offer, leading to user dissatisfaction. By shifting the recommendation algorithms to incorporate features such as Loudness, mode, speechness, valence, danceability etc music apps can cater to user tastes and preferences.If music apps shift from popularity-driven recommendations to more feature-driven, personalised recommendations then stakeholders can offer a more tailored user experience that matches individual tastes. This will lead to a more engaging experience, a growth in the user base and increased user loyalty.

References

Mahajan A. (2023). Exploring Spotify Dataset. [online] Kaggle.com. Available at: https://www.kaggle.com/code/alankarmahajan/exploring-spotify-dataset [Accessed 27 Sep. 2024]

Henderson, E. (2023). 20 Songs with 4-4 Time Signature - Singersroom.com. [online] Available at: https://singersroom.com/songs-with-4-4-time-signature/. [Accessed 27 Sep. 2024]

Suer, M. (2023). What Is Data Quality and Why Is It Important? [online] Alation. Available at: https://www.alation.com/blog/what-is-data-quality-why-is-it-important/. [Accessed 28 Sep. 2024]

Smith, E.(2022).The biggest pain points consumers have with music streaming services. Userbrandwatch. Available at: *"**https://www.brandwatch.com/blog/pain-points-music-streaming-services/**". [Accessed 29 sep.2024]*

Eskandanian, F. Mobasher, B. Burke, R. (2017). "A Clustering Approach for Personalizing Diversity in Collaborative Recommender Systems"[online] doi.org. 280–284. https://doi.org/10.1145/3079628.3079699 [accessed 29 sep.2024]

House of Commons Digital, Culture, Media and Sport Committee (2021) *"Economics of music streaming: Government and Competition and Markets Authority Responses to Committee's Second Report"*, [accessed 28 sep.2024]