# Music Taste Across Decades:

## Spotify Audio Feature Analysis of 33,354 Musical Pieces

Christopher El Khouri

Johns Hopkins University

605.662.83

# Abstract

This paper performs a statistical and visual analysis of audio feature data of 33,354 musical tracks, whose audio features have been measured by Spotify. Hypotheses testing was performed testing the relationships between genres, popularity, audio features, and time. Popularity is related to genre and time. Instrumentalness, danceability, valence, and energy are correlated with popularity. Genres have characteristic audio features that change with time. Herd music preference trends towards low instrumentalness (≈0) and highly danceable tracks (>0.5). Furthermore, herd energy preference forms an Inverted-U shape indicating a sweet spot between energy measures of 0.4 and 0.9. Popularity measures increase with valence up until 0.5, further increases in valence do not result in higher popularity probability.


# Introduction

Before the 20th century, the recording of music was not yet possible, if somebody wanted to enjoy some music they would have had to be able to sing, play an instrument, or attend a live musical performance. The 20th and 21st century saw the development of many different mediums of music from the radio to LPs to cassettes and CDs to MP3s and finally to where we are today in the digital age of music streaming platforms, giving us the utmost freedom in playing whatever we want whenever we want to. Moreover, the type of music has changed over the years, the 20th century saw the birth of a vast variety of popular genres that we enjoy today with their popularities changing as the years flew by.

Spotify is one of the most popular music streaming services in the United States with 44.2 million monthly users as of September 2019 (Verto Analytics, 2019). In addition to being a digital music streaming platform, Spotify provides the service of providing a user with analytical audio features of a selected track. In this paper, we are going to analyze the audio features measured by Spotify and determine their correlations with the different genres, the different decades, and mass popularity. Finding correlations between audio features, time, and popularity can provide various insights into consumer music trends and their changes with time. This can lead to implications regarding the marketing of music from a platform, record label, and a musician's perspective.

# Background

Research published by the South East European University applied principles of data mining and audio visualization on Spotify audio data (Trpkovska, Kajtazi, Bexheti, & Kadriu, 2019). Trpkovska, Kajtazi, Bexheti, & Kadriu utilized 18,598 tracks from Spotify's Top Songs 2017 and analyzed whether there are any correlative patterns among them. Correlations were found between the variables of the audio features and concluded that "energy" and "loudness" are highly positively correlated; valence is positively correlated with danceability, energy, and loudness (Trpkovska, Kajtazi, Bexheti, & Kadriu, 2019). Loudness and speechiness; tempo and danceability; tempo and valence are all negatively correlated with each other. Furthermore; a density plot of energy, valence, and Danceability was produced showing that high danceability and energy were prominent in hit tracks, whereas around 0.50 valence was prominent (Trpkovska, Kajtazi, Bexheti, & Kadriu, 2019). Due to the study only considering popular tracks within a single year (2017-2018), the limited size and span of the data considered limits the significance of the implications and insights. Nevertheless, it puts us on a more refined track for this research topic.

Moving on to a study of relatively higher significance; North, Krause, Sheridan, & Ritchie published a paper that analyzed 143,353 tracks that have enjoyed any degree of commercial success in the United Kingdom since the beginning of the 20th century (North, Krause, Sheridan, & Ritchie, 2017). The paper tested 3 hypotheses regarding the relationship between energy, typicality, and popularity (North, Krause, Sheridan, & Ritchie, 2017). One of the hypotheses that were tested was that there should be an inverted-U relationship between energy and hit popularity, indicating that moderately arousing music is like the most, the result of the test was that the relationship was U-shaped rather than inverted-U (North, Krause, Sheridan, & Ritchie, 2017).

The suggestion of an inverted-U relationship refers to Daniel Berlyne's theory (Berlyne, 1971). A literature review published in 2017 assesses the relevance and significance of Berlyne's theory on music preference (Chmiel & Schubert, 2017). Berlyne argued that an organism is aroused by various aspects of the external environment and that this arousal drives the approach and aversion systems of the organism (Chmiel & Schubert, 2017). Berlyne proposed that an intermediate level of arousal is preferred and that preference is an index of this level of arousal (Chmiel & Schubert, 2017). This proposed inverted-U function is shown to the pattern outlined in Figure 1, also known as the Wundt curve (Berlyne, 1971) Chmiel and Schubert investigated the inverted-U model of preference for music over the results of 57 studies (Chmiel & Schubert, 2017). The variables tested mainly consisted of familiarity and complexity, as a result, 50 of the 57 studies (88%) were compatible with the inverted U-model. This contradicts the results produced by North, Krause, Sheridan, & Ritchie when using energy as a variable and popularity of music within the United Kindom.

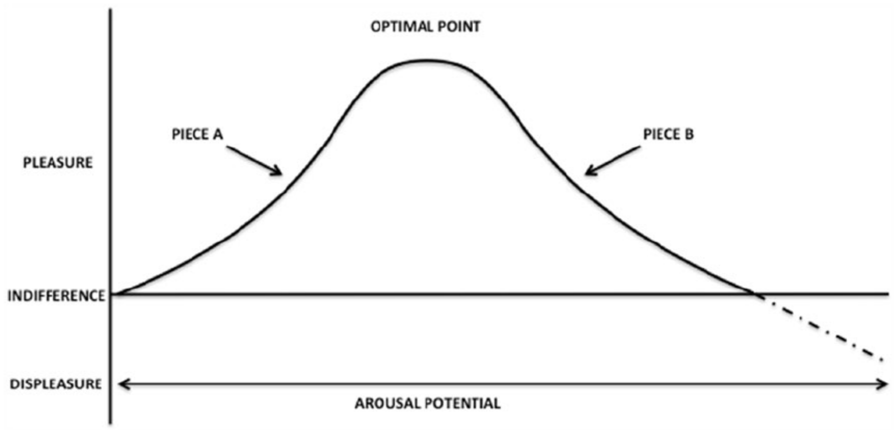


*Figure 1: Wundt Curve*

One year later; North, Krause, Sheridan, & Ritchie conducted another study but considered music that was popular within the United States (North, Krause, Sheridan, & Ritchie, 2018). The study considered 204,506 pieces of music that have achieved any degree of commercial success in the United States since the beginning of the 20th century (North, Krause, Sheridan, & Ritchie, 2018). The study tested a total of 5 hypotheses: there should be an inverted-U relationship between energy and measures of popularity; there should be a positive relationship between the typicality of a piece and its popularity measure; there should be an association between both energy and tempo and mood scores; popularity scores should be associated with mood scores; the mood scores should differ between genres (North, Krause, Sheridan, & Ritchie, 2018). These different mood scores correlate to the Spotify audio features valence and energy. The results of the hypotheses showed that energy has an inverted-U relationship with hit popularity; there was no relationship between typicality and popularity; energy and tempo were associated with mood scores towards higher arousal; popularity was associated with mood scores; and mood scores differed between genres (North, Krause, Sheridan, & Ritchie, 2018).

# Approach

The dataset that will be used is The Spotify Hit Predictor Dataset (1960-2019) (Ansari, 2020) found on Kaggle; it contains 41,106 tracks split between 6 decades along with their audio features measured by Spotify. Ansari added a target column where the value '1' denotes a popular track (hit) and '0' denotes an unpopular track (flop), the detailed criteria of a hit and flop tracks can be found on Appendix 1: Dictionary. The dataset is split into 6 different files, separating the tracks by decade: 1960s; 1970s; 1980s; 1990s; 2000s; and 2010s where half of the tracks are 'hits' and half are 'flops'.

Using Ansari's data, the research objective of analyzing the audio features measured by Spotify and determining their correlations with the different genres, the different decades, and on mass popularity will be performed by testing the following hypotheses:

**Hypothesis 1:** There is a statistically significant relationship between the genre and decade of a track and its popularity

**Hypothesis 2:** There is a statistically significant relationship between the audio features of a track and the popularity.

**Hypothesis 3:** There is a statistically significant relationship between the audio features of a track, the genre, and the decade.

A user friendly dashboard for data exploration can be found on Appendix 12: Dashboard.

# Data Preprocessing

## Feature Generation

The dataset provided (Ansari, 2020) did not contain the genre of the track, they had to be extracted from Spotify's database. Spotify stores genre as an artist property where each artist could have multiple genres. Furthermore, the genres stored on Spotify's API include sub-genres, i.e. including "Canadian pop" and "pop" as two different genres. Using Spotipy, a python library used to access Spotify's API, the first 4 genres of each artist was extracted (Lamere, 2014). Subsequently, each sub-genre was categorized into the main genre accordingly. Appendix 3: Genre Extraction Code contains the Python codes that were developed for this process.

The main genres defined in this project are:

- Pop
- R&B
- Latin
- Caribbean
- Blues
- Jazz
- Hip hop
- Rock
- Electronic
- Folk
- Classical
- Flamenco
- Metal
- Country
- Avant-garde
- Comedy
- Easy listening
- Other

The choice of main genres is heavily based on AllMusic's genre list (AllMusic, 2020).

In addition to the 4 sub-genre and main-genre columns; the columns 'Decade', 'key_name', 'mode_name', 'key_full', and 'hit_flop' were added to further aid our analysis. The full dictionary of this data can be found in Appendix 1: Dictionary. Appendix 2: Genres lists the main genres with their corresponding sub-genres.

## Outlier Removal

Figure 2 shows a chart of the percentage of the total number of tracks vs genre. 8.7% of the tracks of the data across all decades were denoted as 'Other' as in not classifiable to any of the main genres defined above. The genres Flamenco, Avant-garde, and Comedy constitute a sum of 1.1% of the tracks.

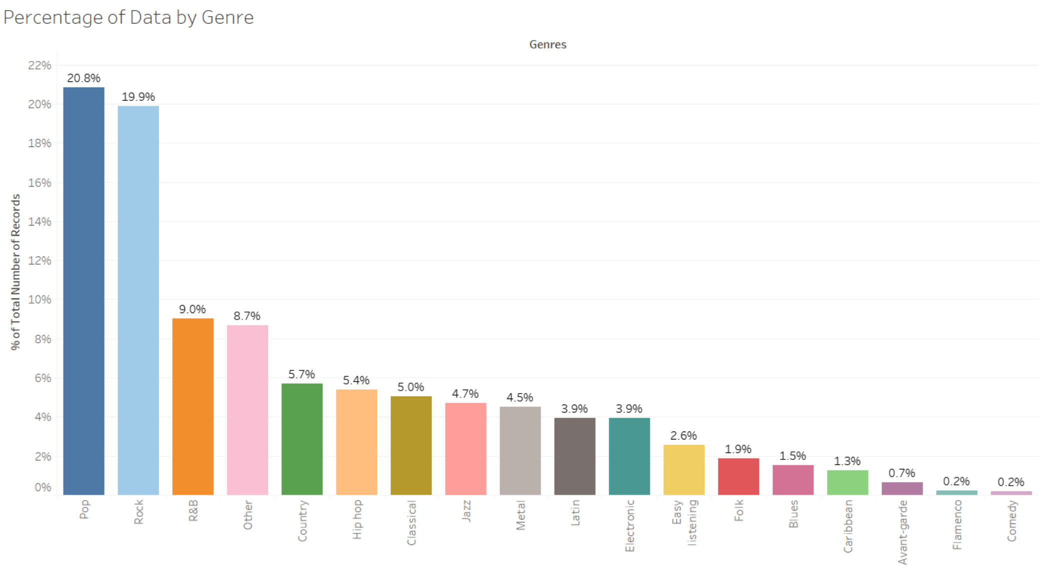


*Figure 2: Distribution of genres pre-processing*

Therefore, the genres Other, Flamenco, Avant-garde, and Comedy were excluded from the analysis resulting in a total of 9.8% of the tracks removed, removing the genre outliers, and a total of 37,095 tracks remaining.

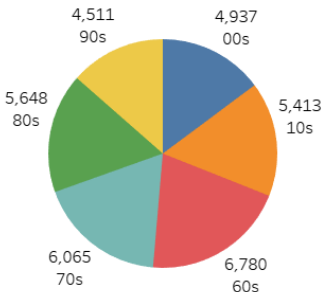


*Figure 3: Distribution of Decades*

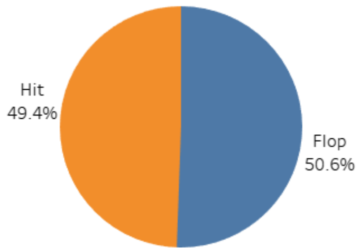


*Figure 4: Distribution of Hits and Flops*

The next step in outlier removal is the removal of audio feature outliers, this was done by separating the data by genre, by decade, and by target. Approximately 10% of the data was removed and resulted in a total of 33,354 remaining tracks. The python script used for this process can be found in Appendix 4: Outlier Removal. Figure 3 shows the distribution of data amongst the decades, and Figure 4 the distribution of hits and flops. The number of hits and flops were kept as close as possible to minimize bias.

# Analysis

## Hypothesis 1: Genre Popularity and Decade

Analyzing all 33,354 tracks; the top 5 genres by order of popularity are Pop, Rock, R&B, Country, and Hip Hop. Figure 5 shows a chart of the percentage of popular tracks (target=1) vs the top 5 genres by decade with some clear changes between each decade.

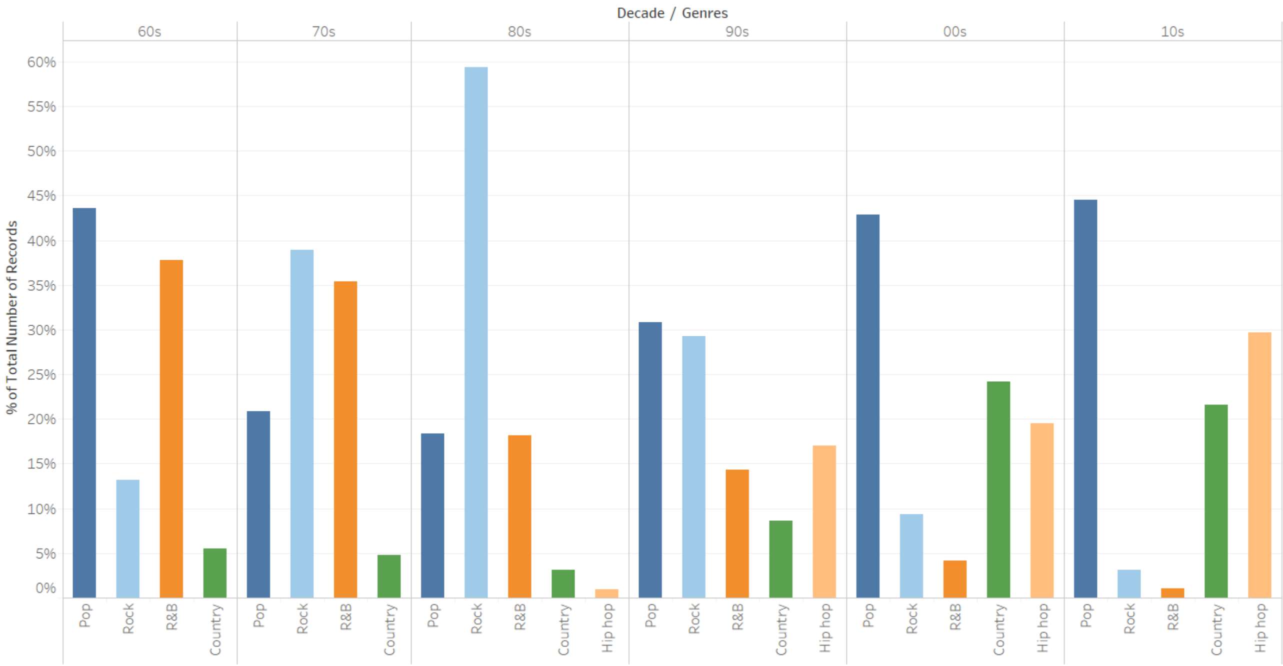


*Figure 5: Popular Genres by Decade*

In the 60s, Pop and R&B were the top genres, Rock managed to rise above Pop in the 70s with R&B still competing. The 80s show the peak of Rock's popularity and the birth of popular Hip Hop, the 90s show Rock's popularity diminished by approx. 50% with Pop having the highest popularity and Hip Hop rising. The 2000s and 2010s show Pop reclaiming its throne whit Country and Hip Hop side by side for the 2nd and 3rd place. The visual analysis of Genre Popularity and Decade can be found in Appendix 5: Hypothesis 1 Visualizations.

## Hypothesis 2: Audio Features and Popularity

Performing a correlation analysis among the audio features and target of each track on Python confirms the results found in the previous research by the South East European University (Trpkovska, Kajtazi, Bexheti, & Kadriu, 2019) seem to hold here.  Energy and Loudness are highly positively correlated with a value of 0.77; Valence (how cheerful a track is) is positively correlated with danceability, energy and loudness with values of 0.55,0.33, and 0.27 respectively. Furthermore, energy is extremely negatively correlated with acousticness with a value of -0.73; loudness is negatively correlated with acousticness with a value of -0.58; sections are very positively correlated with the duration with a value of 0.86. The code behind this process can be found in Appendix 6: Correlation Heat Map.

Removing the variables that resulted in high inter-correlation (>0.7) to minimize variable bias results in the following correlation map. Loudness, acousticness, and sections were removed due to their high correlations with other audio features. Looking at the correlations with target; Instrumentalness, Danceability, valence, and energy have the highest absolute correlations sorted by the order.

Figures 7-9 show the plots of the number of tracks vs the feature. Separating the tracks into hits and flops produces a probability distribution type of chart. Hits are represented by the color orange and flops are represented by the color blue.

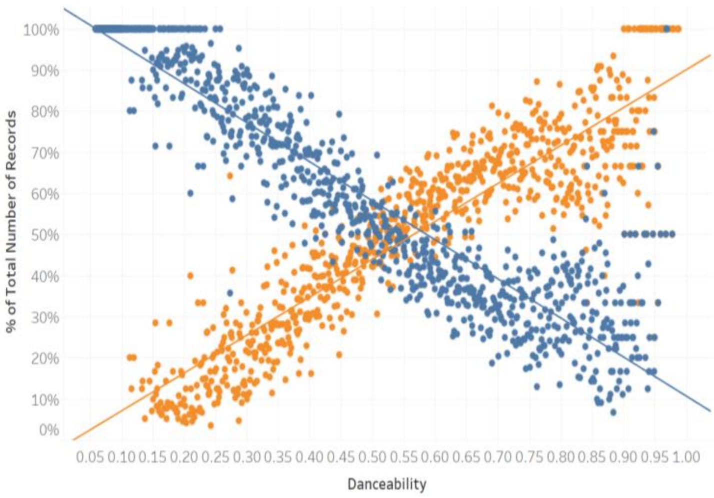


*Figure 7: Danceability and Popularity*

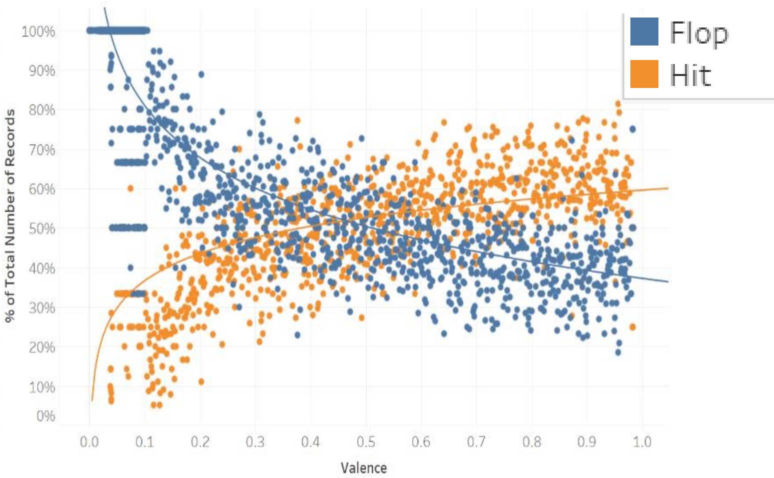


*Figure 8: Valence and Popularity*

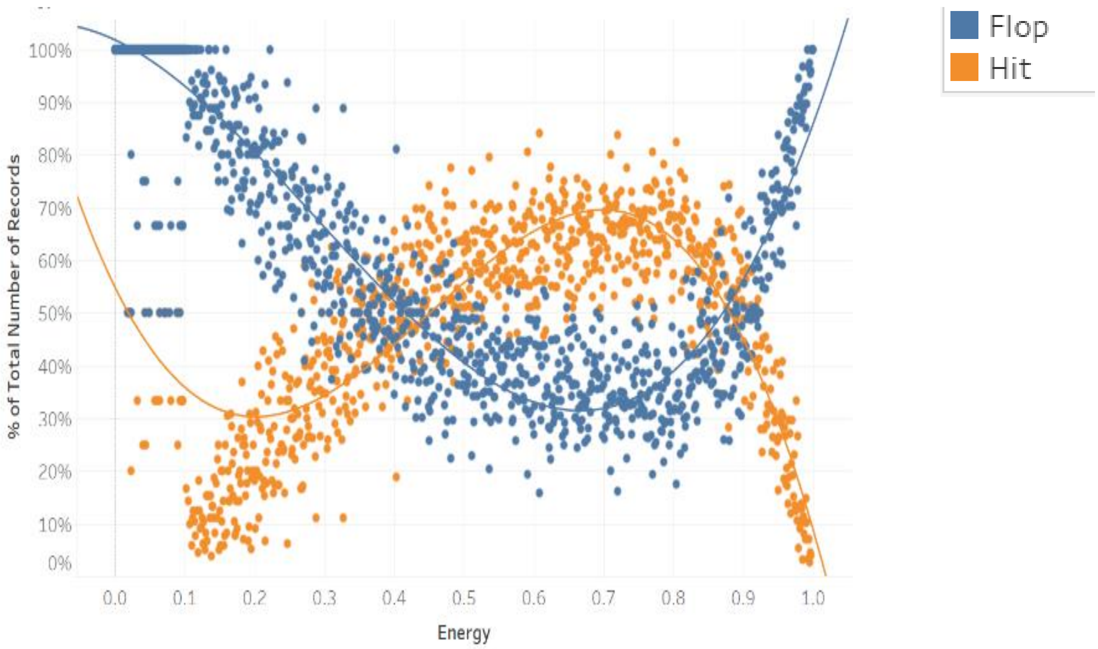


*Figure 9: Energy and Popularity*

Figure 7 is a plot of the danceability, where the probability distribution appears to be highly linear with tracks having higher danceability having a higher hit probability. Figure 8 is a plot of the valence, where tracks having a valence measure of less than 0.5 have a higher chance of being flops, however, a valence measure of 0.5 or greater only manages to keep the probability of a track being a hit track at around 50%. The probability distribution of valence seems to be leaning towards a logarithmic model. Figure 9 is a plot of the energy measure, where the popular tracks are clearly showing a curve very similar to the Inverted-U Wundt Curve shown in Figure 1 with the tracks having a higher probability of popularity between energy measures of around 0.4 and 0.9. North, Krause, Sheridan, & Ritchie found a similar result in their research of popularity and energy within the United States (North, Krause, Sheridan, & Ritchie, 2018). The visual analysis of all the audio features can be found in Appendix 7: Hypothesis 2 Visualizations.

**Hypothesis 3: Audio Features, Genre, Decade**

Figure 10 shows the change in medians of energy, danceability, and valence for the top 5 genres; Pop, Rock, R&B, Country, and Hip Hop. Danceability is distinguishable especially between Hip Hop and Rock; R&B has a much higher valence than the other 4 genres, and Rock tops the energy chart.

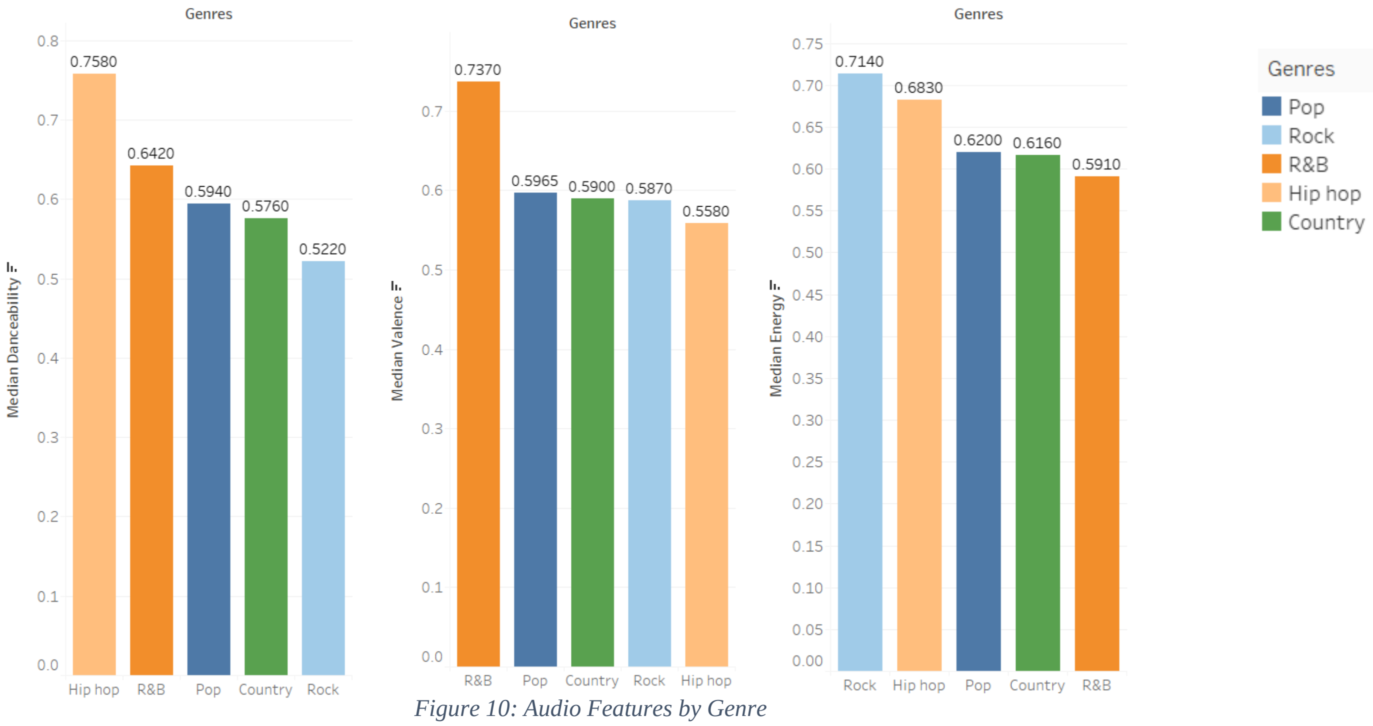


*Figure 10: Audio Features by Genre (Top Genres)*

Figure 11 shows the change in the medians of; Danceability, valence, and energy with time. Danceability has increased by approx. 20%, valence is shown to have decreased by approx. 25%, whereas energy has increased by approx. 70%. Figure 12 shows the change of the medians but with the main focus on Rock and Hip Hop indicating that the genre audio features change with time. The full visual analysis of Hypothesis 3 can be found in Appendix 8:Hypothesis 3 Visualizations.

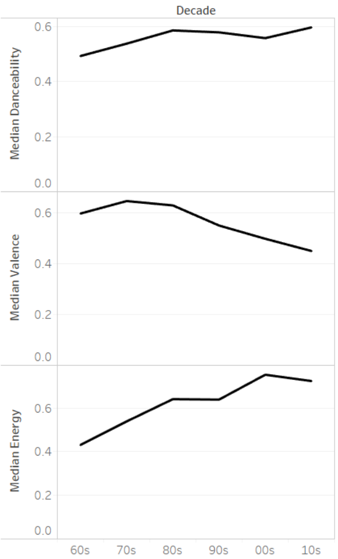


*Figure 11: Change of Audio Features with Decade (All tracks)*

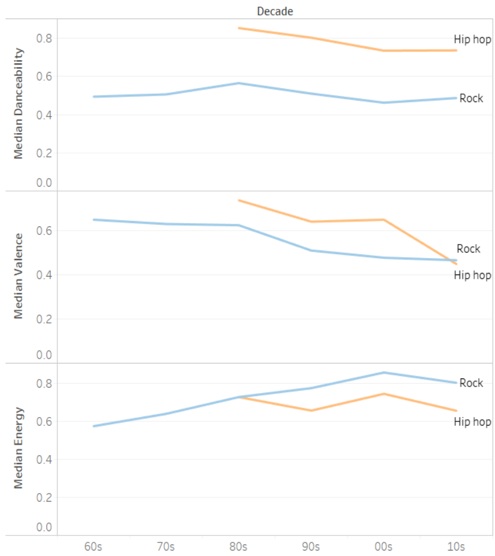


*Figure 12: Change of Audio Features with Decade (Rock and Hip Hop)*

# Results

## Hypothesis 1: Genre Popularity and Decade

*Table 1: Hypothesis 1 Results*

| Hypothesis 1:Chi-Squared Test of Independence | | | | | | |
|---|---|---|---|---|---|---|
| Categorical Variable 1 | Categorical Variable 2 | Observation Variable | df | Critical Value | $X^2$ | p-value |
| Genre | Decade | Number of Hits | 65 | 105.99 | 9916.17 | **<0.001** |

As shown in Table 1, a chi-squared test of independence was performed on Python testing our hypothesis on all 14 genres. P-value <0.001 was achieved therefore rejecting the null hypothesis and confirming that there is a statistical relationship between the popularity of a genre and the decade. The Python code of the test can be found in Appendix 9: Hypothesis 1 Test.

## Hypothesis 2: Audio Features and Popularity

As shown in Table 2, a logistic regression model was performed on Python testing the relationship of the independent variables: Danceability, energy, Instrumentalness, and valence on the dependent variable target. The overall p-value of the test was <0.001 with all the independent variables except valence having a p-value of <0.001 and valence having a p-value of 0.001. We can confirm a statistical relationship between the above mentioned independent variables and popularity. The code behind the test can be found in Appendix 10: Hypothesis 2 Test.

*Table 2: Hypothesis 2 Results*

| **Hypothesis 2:Logistic Regression** | | | |
|---|---|---|---|
| **Dep. Variable:** | target | **Df Residuals:** | 33349 |
| **Model:** | Logit | **Df Model:** | 4 |
| **Method:** | MLE | **Pseudo R-squ.:** | 0.2453 |
| **converged:** | TRUE | **Log-Likelihood:** | -17446 |
| **Covariance Type:** | nonrobust | **LL-Null:** | -23117 |
| **No. Observations:** | 33354 | **LLR p-value:** | **<0.001** |

| **Ind. Variables** | **coef** | **std err** | **z** | **P>\|z\|** | **[0.025** | **0.975]** |
|---|---|---|---|---|---|---|
| Danceability | 3.9003 | 0.096 | 40.657 | **<0.001** | 3.712 | 4.088 |
| energy | 0.6304 | 0.058 | 10.876 | **<0.001** | 0.517 | 0.744 |
| Instrumentalness | -7.0114 | 0.182 | -38.485 | **<0.001** | -7.368 | -6.654 |
| valence | -0.1981 | 0.062 | -3.209 | **0.001** | -0.319 | -0.077 |
| intercept | -1.9799 | 0.057 | -34.587 | **<0.001** | -2.092 | -1.868 |

## Hypothesis 3: Audio Features, Genre, Decade

As shown in Table 3, a Two Way ANOVA test was performed on each of Danceability, energy, Instrumentalness, and valence on Genre and Decade. All tests achieved a p-value of <0.001 confirming a statistical relationship between each of the mentioned audio features on genre, on the decade, and on both genre and decade. The code behind the test can be found in Appendix 11: Hypothesis 3 Test.

*Table 3: Hypothesis 3 Test Results*

| Hypothesis 3:Two-Way ANOVA | | | | | | |
|---|---|---|---|---|---|---|
| No | Observation Variable | Test Variables | sum_sq | df | F | p-value |
| 1 | Danceability | Genre | 335.3268 | 13 | 1080.019 | **<0.001** |
| | | Decade | 27.07262 | 5 | 226.708 | **<0.001** |
| | | Genre: Decade | 32.80516 | 65 | 21.13174 | **<0.001** |
| | | Residual | 794.5958 | 33270 | | |
| 2 | Energy | Genre | 488.3581 | 13 | 999.2658 | **<0.001** |
| | | Decade | 101.6127 | 5 | 540.5849 | **<0.001** |
| | | Genre: Decade | 72.44554 | 65 | 29.64724 | **<0.001** |
| | | Residual | 1250.739 | 33270 | | |
| 3 | Instrumentalness | Genre | 815.6999 | 13 | 1054.43 | **<0.001** |
| | | Decade | 15.74211 | 5 | 52.90829 | **<0.001** |
| | | Genre: Decade | 98.9025 | 65 | 25.56964 | **<0.001** |
| | | Residual | 1979.803 | 33270 | | |
| 4 | Valence | Genre | 417.7572 | 13 | 587.067 | **<0.001** |
| | | Decade | 51.91061 | 5 | 189.6676 | **<0.001** |
| | | Genre: Decade | 42.29069 | 65 | 11.88608 | **<0.001** |
| | | Residual | 1821.15 | 33270 | | |

# Conclusion

Given that all our tests yielded significant results, we have a better understanding of how music preference has changed over time swaying from one genre to another. Several analytical questions can be answered using this data and results and some of the questions will be discussed below.

**Question 1.** Which of the track audio features are correlated with popularity?

Based on the correlation table produced in Appendix 6; Instrumentalness, Danceability, Valence, and Energy are the audio features with the most significant relationships with popularity.

**Question 2.** How have the audio features of popular tracks (hits) changed with time?

Figure 13 below shows how the median of the audio features of hit tracks are changing with time. Instrumentalness and Valence are following a downward trend, whereas Danceability and Energy are following an upward trend.

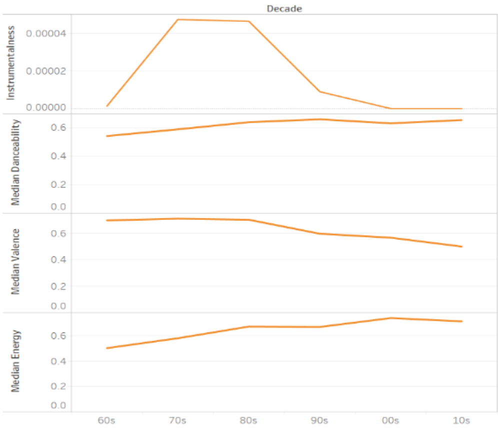


*Figure 13: Change of Audio Features with Decade (Hit Tracks)*

**Question 3.** What forecasts can be made on how genres and herd music preferences have changed with time?

Based on this data and the analysis that has been performed throughout this paper, it seems that Instrumentalness will remain near 0, Danceability will continue to increase, valence will continue to decrease, energy may remain somewhat the same. However, we can see that herd popularity seems to be heading towards more danceable, low valence, non-instrumental, and energetic music. Converting the audio features into the top genres analyzed in the paper; Pop and Hip Hop are good candidates for the most popular genres of the 2020s. Appendix 8-P shows how the top genres audio features have changed with time, Pop and Hip Hop are moving with the trends of the audio features. However, for us to more accurately validate how music genres have changed, further analysis is required focusing on the remaining genres.

# References

AllMusic. (2020, May 08). *Music Genres*. Retrieved from AllMusic: https://www.allmusic.com/genres

Ansari, F. (2020, April 26). *The Spotify Hit Predictor Dataset (1960-2019)*. Retrieved from Kaggle: https://www.kaggle.com/theoverman/the-spotify-hit-predictor-dataset

Berlyne, D. E. (1971). *Aesthetics and Psychobiology*. New York: Appleton-Century-Crofts.

Chmiel, A., & Schubert, E. (2017). Back to the inverted-U for music preference: A review of the literature. *Psychology of Music*, 886-909.

Lamere, P. (2014). *Welcome to Spotipy!* Retrieved from spotipy 2.0 documentation: https://spotipy.readthedocs.io/en/2.11.2/

North, A. C., Krause, A. E., Sheridan, L. P., & Ritchie, D. (2017). Energy, Typicality,and Music Sales: A Computerized Analysis of 143,353 Pieces. *Empirical Studies of the Arts*, 214-229.

North, A. C., Krause, A. E., Sheridan, L. P., & Ritchie, D. (2018). Popularity, Mood, Energy, and Typicality in Music: A Computerized Analysis of 204,506 Pieces. *American Psychological Association*, 89-109.

Spotify AB. (2020). *Home*. Retrieved from Spotify for Developers: https://developer.spotify.com/

Trpkovska, M. A., Kajtazi, A., Bexheti, L. A., & Kadriu, A. (2019). Applying Data Mining and Data Visualization within the Scope of Audio Data Using Spotify. *International Conference on Information Systems Post-Implementation & Change Management*, 197-204.

Verto Analytics. (2019). *Verto Statista Report September 2019.* Verto Analytics.