The Anatomy of a Hit*

Modeling Popularity of Billboard Artists using Spotify Audio Features

Luca Carnegie

April 24, 2024

What makes a hit pop song? This paper examines the audio features that characterize mainstream music's biggest hits. By analyzing data from the discographies of the highest-charting artists on the Billboard Hot 100, the research identifies key attributes associated with popular songs. Multivariate regression analysis reveals that higher levels of danceability, explicit lyrics, and loudness are positively related to popularity, while emotional positivity (valence) exhibits a negative relationship. Overall, this work quantifies some attributes underpinning iconic pop successes and can empower professional pop musicians to make more informed creative decisions with their work.

Table of contents

*Code and data are available at: https://github.com/lcarnegie/popularity-modeling. Thank you to Michaela Drouillard for your insights on Spotify and it's metrics. Special hat tip goes to Rajan Maghera and my amazing parents, Cristina and Shane for your constant support and encouragement. I could not have done this without you.

1 Introduction

Throughout history, musicians have aspired to write hit songs and achieve mainstream success in the music industry. However, for major music companies, signing artists without a proven track record is a substantial risk, as most aspirants fail to produce hits. This has led to a concentrated market dominated by a small group of "superstars", such as Taylor Swift, who consistently top the charts (Rosen 1981). Breaking into the music industry, then, presents a unique challenge for aspiring musicians' art: what elements contribute to a pop song's mass appeal?

The field of Hit Song Science, which employs statistical methods to predict song popularity, has gained traction with the increasingly widespread availability of economical computing resources. Researchers (Dodds and Danforth 2010), students (Pham, Kyauk, and Park, n.d.), and likely record companies have attempted to construct models that attempt to explain song popularity, though the latter's efforts remain proprietary, leaving artists without access to this knowledge.

This study diverges from previous approaches by constructing a dataset solely from the discographies of the best-performing artists on the Billboard Hot 100, the preeminent metric of success in the United States music industry. By focusing exclusively on the most successful artists' music, the elements that define enduring hits in the US market can be more easily identified. The analysis employs multivariate regression to estimate a song's Spotify-generated popularity score, using several audio features derived from Spotify's API as predictors.

The regression results reveal statistically significant positive relationships between danceability, explicit lyrics, loudness, and Spotify popularity score, as well as a significant negative relationship between song valence (general positivity) and popularity score, after controlling for other variables.

In the current era of music streaming, competition for stardom is intense. Understanding the key elements driving song success could empower artists to craft hits more effectively and potentially achieve greater independence from record labels (Burke 1997). This data-driven approach could provide a strategic advantage for artists in the pursuit of mainstream success.

The paper is structured as follows: Section 2 provides a detailed overview of the dataset and analyzes each predictor individually; Section 3 outlines the regression model; Section 4 presents and discusses the results; and Section 5 critically examines the findings and their implications.

2 Data Analysis

The Billboard Hot 100 ranks the most popular U.S. songs weekly based on radio plays, sales, and streaming (McCormack 2023). Artists with songs frequently on this chart should have

mass appeal, motivating the analysis of elements that define the popularity of those "hits". Using Billboard also scopes this analysis to solely the US market

Unlike other datasets like The Million Song dataset used by Pham et al., Spotify's web API provides ready-made quantitative audio features (tempo, key, danceability, etc.) as well as calculated popularity score based on recent and total play counts for artists' songs, facilitating analysis of popularity determinants.

Since the popularity score is based on both recency of streams and total number of streams, continually high scores indicate frequent plays long after release. This means this score can be thought of measuring a song's enduring popularity. As the leading U.S. music streamer, Spotify's metrics can reasonably align with Billboard rankings, providing a relatively robust measure of a song's lasting popularity, which is crucial in attempting to infer the elements that make up a "hit".

To analyze the audio feature determinants of popularity, I used R (R Core Team 2023), the tidyverse (Wickham et al. 2019) and related packages. Billboard's "Greatest of All Time Hot 100 Artists" (2015) list identified popular artists. The audio features of the Billboard artists' songs and popularity data were then acquired from Spotify's API via the spotifyr (Thompson et al. 2022) package. Spotify data was current as of 2024, while Billboard rankings were from 2015. The data was cleaned using the dplyr (Wickham, François, et al. 2023) and janitor (Firke 2023) packages and saved using arrow (Richardson et al. 2024). Variables were chosen based on expected impact on popularity, resulting in a dataset with 773 songs.

For this analysis, the popularity score, valence (musical 'positivity'), danceability, mode (major or minor), presence of explicit lyrics, loudness, and song duration were made of interest. A sample of the cleaned dataset is shown in Table 1. This and other tables were created using the knitr (Xie 2023) packages. Visualizations and modelling summaries used ggplot2 (Wickham 2016) and modelsummary (Arel-Bundock 2022).

Table 1: Popularity Scores and Audio Features

Artist Name	Song Name	Popularity	Mode	Valence	Danceability	Explicit	Loudness (dB)	Duration (seconds)
Alicia Keys	Empire State of Mind (Part II) Broken Down	72	1	0.142	0.484	0	-7.784	216.480
Alicia Keys	Fallin'	77	0	0.482	0.652	0	-7.519	210.200
Alicia Keys	If I Ain't Got You	83	1	0.166	0.609	0	-9.129	228.706
Alicia Keys	No One	77	0	0.167	0.644	0	-5.415	253.813
Alicia Keys	Un-thinkable (I'm Ready)	69	1	0.335	0.596	0	-7.892	249.240

The primarily categorical dataset was first analyzed variable-by-variable. Each variable was examined based on its potential impact on popularity scores, measured from 0-100 with 100 being the most popular. For further inference, the popularity score was then modeled using multiple linear regression, with the other variables as predictors.

2.1 Popularity

First, we assess how popularity is distributed across our dataset using a histogram.

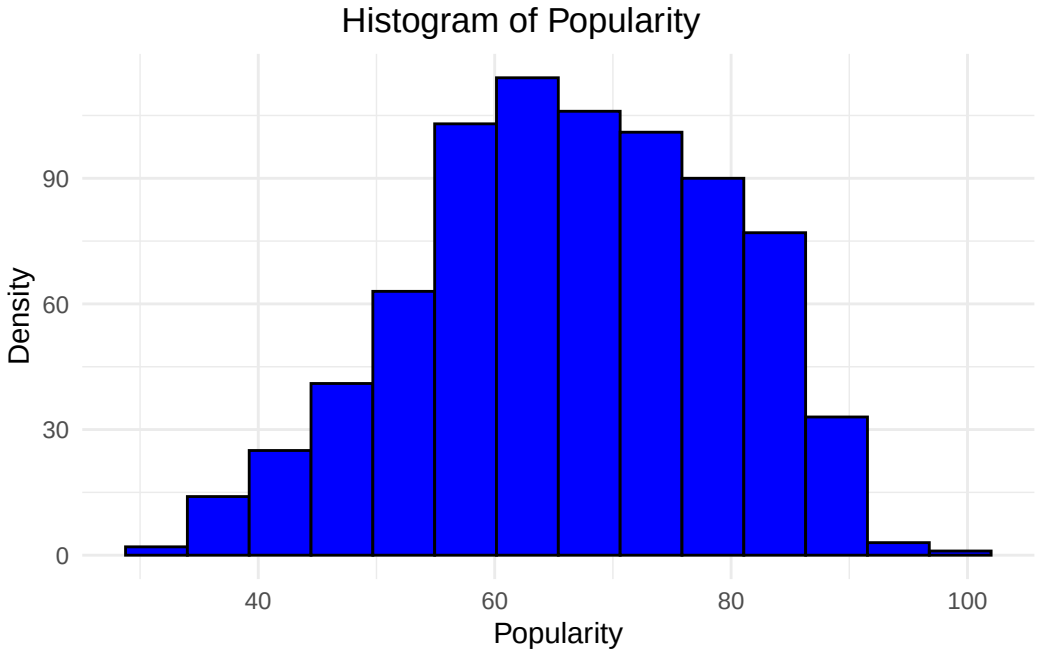


Figure 1: Histogram of Popularity Scores

Against expectations, the histogram in (**popularity-histogram?**) shows no skewed distribution toward higher popularity scores for the most popular artists. Although most songs have scores between 54 and 86, which indicates relatively lasting popularity, the curve looks somewhat bell-shaped. This indicates that there is an even spread of top performing and mediocre artists, even within the subset of the best performing artists in the world.

2.2 Valence

Next, we focus on valence. As one of Spotify's algorithmically generated metrics, valence attempts to measure the musical "positivity" of a song's audio, with scores ranging between

0.0 and 1.0. As Spotify's documentation says: "Tracks with high valence sound more positive (e.g. happy, cheerful, euphoric), while tracks with low valence sound more negative (e.g. sad, depressed, angry)." (n.d.).

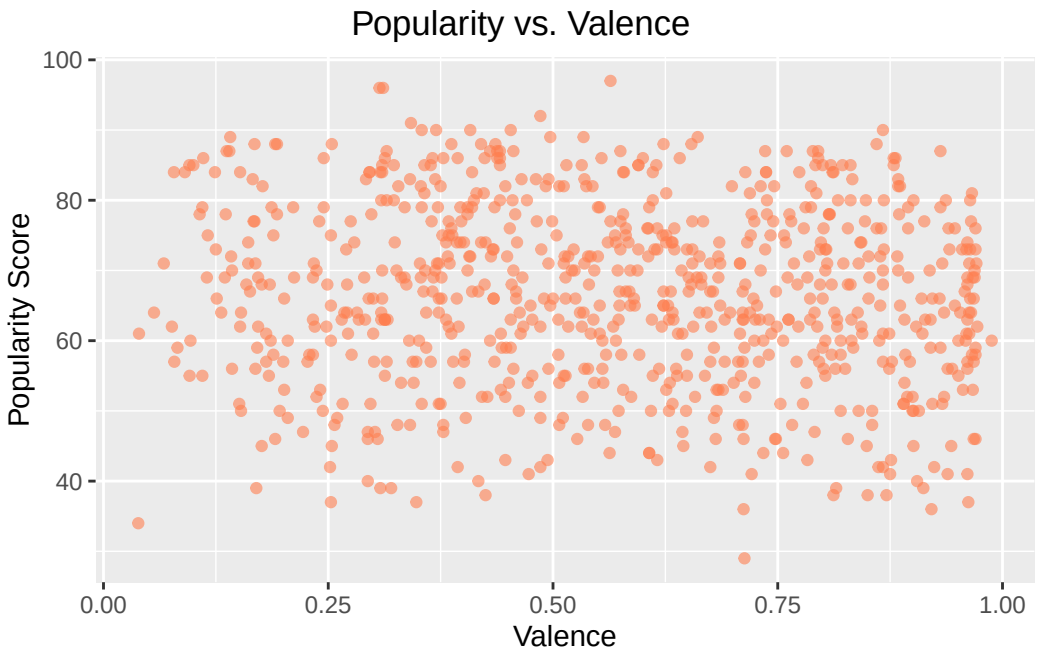


Figure 2: CAPTION

When plotting valence values against popularity scores, (**scatter-valence?**) shows no clear linear trend between valence and popularity, although a clumping of high valence scores near 1.00, suggests potential data issues.

2.3 Danceability

Like valence, danceability is another algorithmically calculated metric by Spotify. From their API documentation, "Danceability describes how suitable a track is for dancing based on a combination of musical elements including tempo, rhythm stability, beat strength, and overall regularity. A value of 0.0 is least danceable and 1.0 is most danceable.".

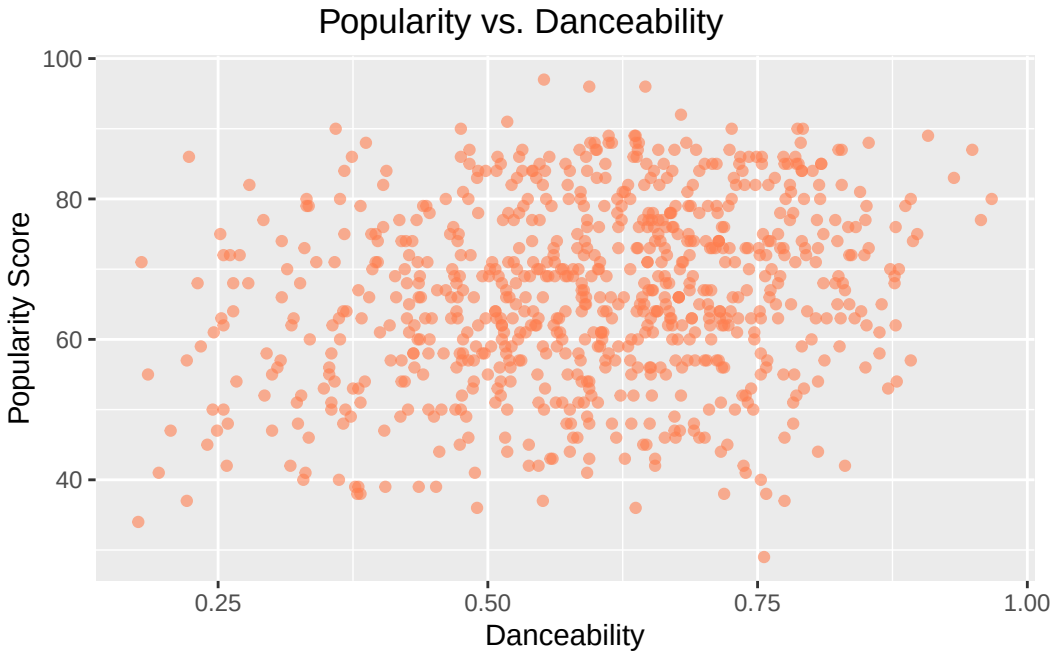


Figure 3: CAPTION

Plotting the danceability scores against popularity in (**scatter-danceability?**) shows a weak association between danceability and popularity score.

2.4 Mode

Musical mode indicates whether a song is in a major ('happier') or minor key ('sadder'). Unlike valence, which is calculated from a variety of physical metrics, musical mode is inherent to musical piece and can be inferred from the arrangement of notes in a song. In the dataset, 1 indicates the song is major while 0 is minor.

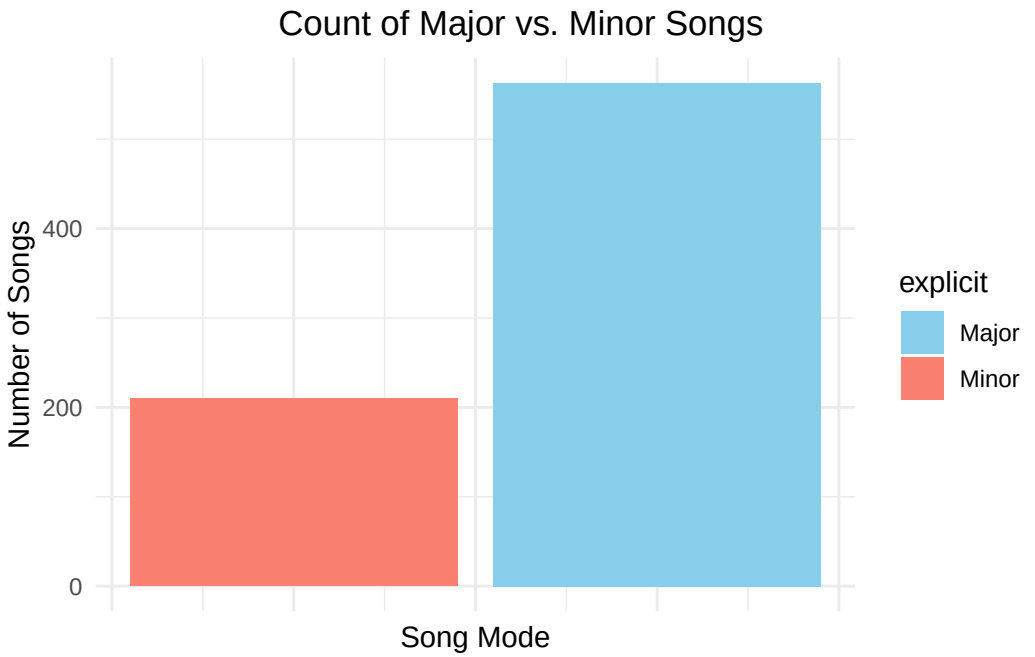


Figure 4: CAPTION

(**maj-min-barchart?**) shows that the most popular songs are in a major key, with more than twice as many major songs than minor songs.

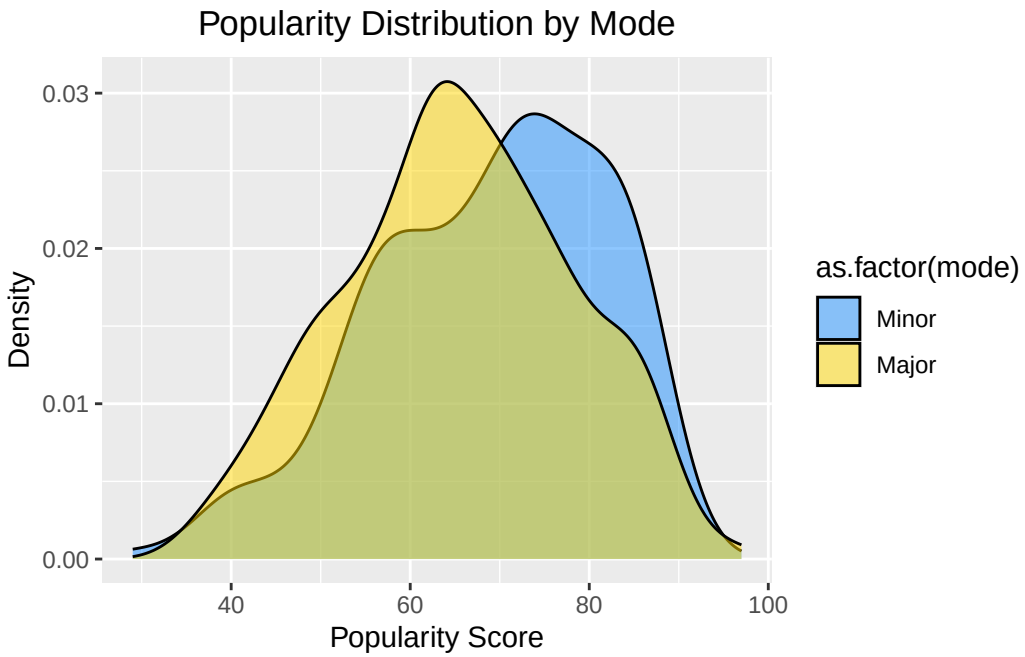


Figure 5: CAPTION

Creating a density plot comparison in (**density-plot?**) shows that although there are more major songs overall, there actually appears to be more songs in minor keys that have higher popularity scores. This means that audiences may tend to have a higher preference for songs in minor keys over major.

2.5 Explicit Status

Spotify records whether a song contains explicit lyrics through reporting by the music publishers.

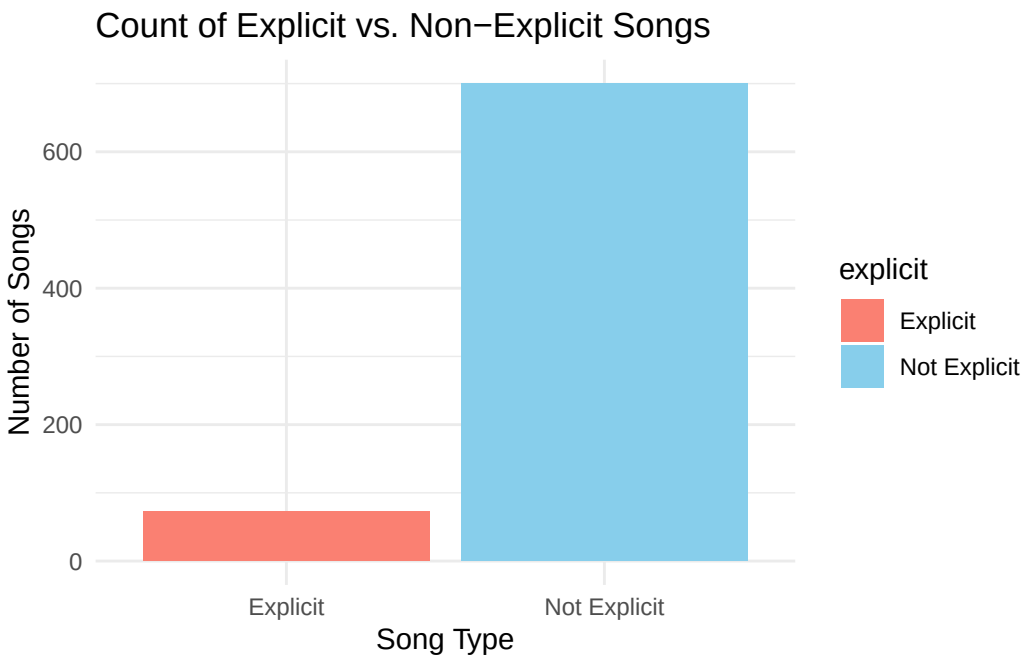


Figure 6: CAPTION

(**explicit-bar-chart?**) shows an imbalanced amount of non-explicit songs among popular artists, with more than seven times that of the number of explicit songs. That could be a artifact of radio stations preferences for songs without explicit language.

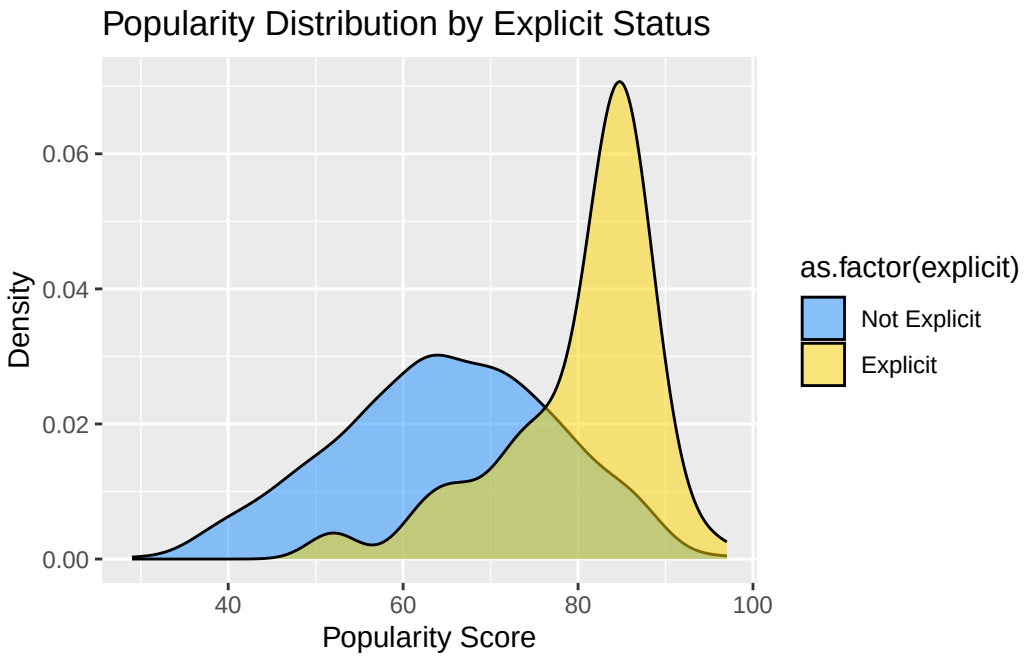


Figure 7: CAPTION

```
List of 1
 $ plot.title:List of 11
  ..$ family       : NULL
  ..$ face         : NULL
  ..$ colour       : NULL
  ..$ size         : NULL
  ..$ hjust        : num 0.5
  ..$ vjust        : NULL
  ..$ angle        : NULL
  ..$ lineheight   : NULL
  ..$ margin       : NULL
  ..$ debug        : NULL
  ..$ inherit.blank: logi FALSE
  ..- attr(*, "class")= chr [1:2] "element_text" "element"
 - attr(*, "class")= chr [1:2] "theme" "gg"
 - attr(*, "complete")= logi FALSE
 - attr(*, "validate")= logi TRUE
```

Comparing the distribution of popularity scores between explicit and not-explicit songs shows a different story when it comes to popularity score on Spotify. Here, we see that explicit songs have the highest density at higher popularity scores, which indicates that explicit songs

tend to be given higher popularity scores on average. This could mean that explicit songs are simply more popular on streaming services like Spotify, where users can choose what music they would like to listen to.

2.6 Loudness

The average loudness of a song, in decibels (dB) is calculated by averaging the height of the waveforms in a particular song. In (**scatter-loudness?**) we plot loudness against popularity to assess any relationship between the two variables.

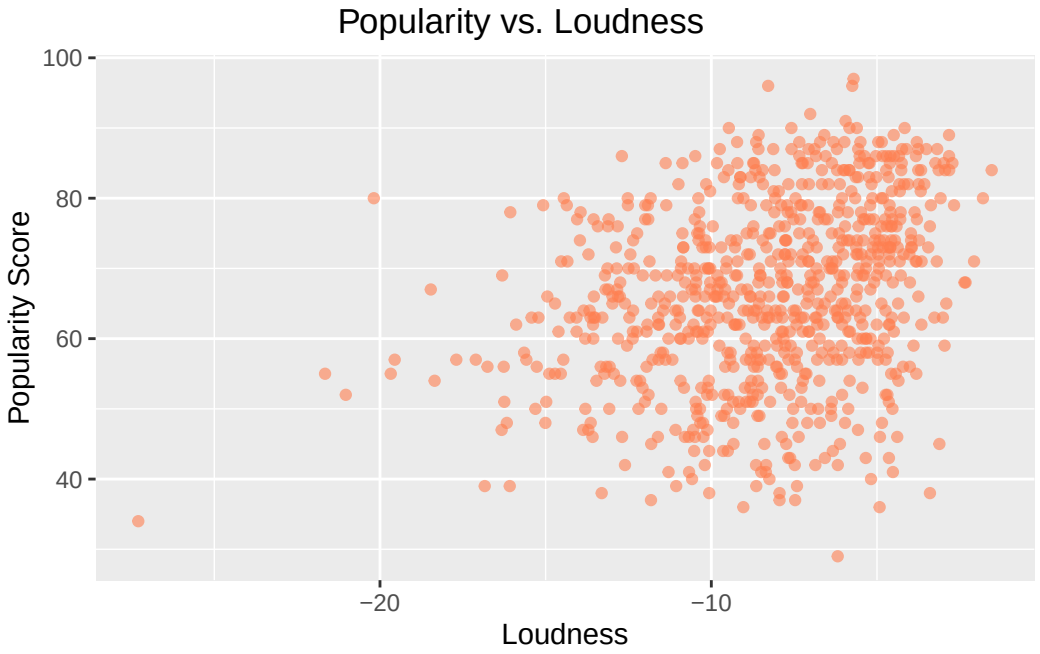


Figure 8: CAPTION

(**scatter-loudness?**) shows no clear relationship between loudness and popularity, likely due to the varied nature of popular music and other confounding factors. That makes sense, since there could be songs that are "known" for being soft and emotional, while others are loud. Given the varied nature of pop music, it is to be expected.

2.7 Duration

Duration's effect on popularity was then investigated.

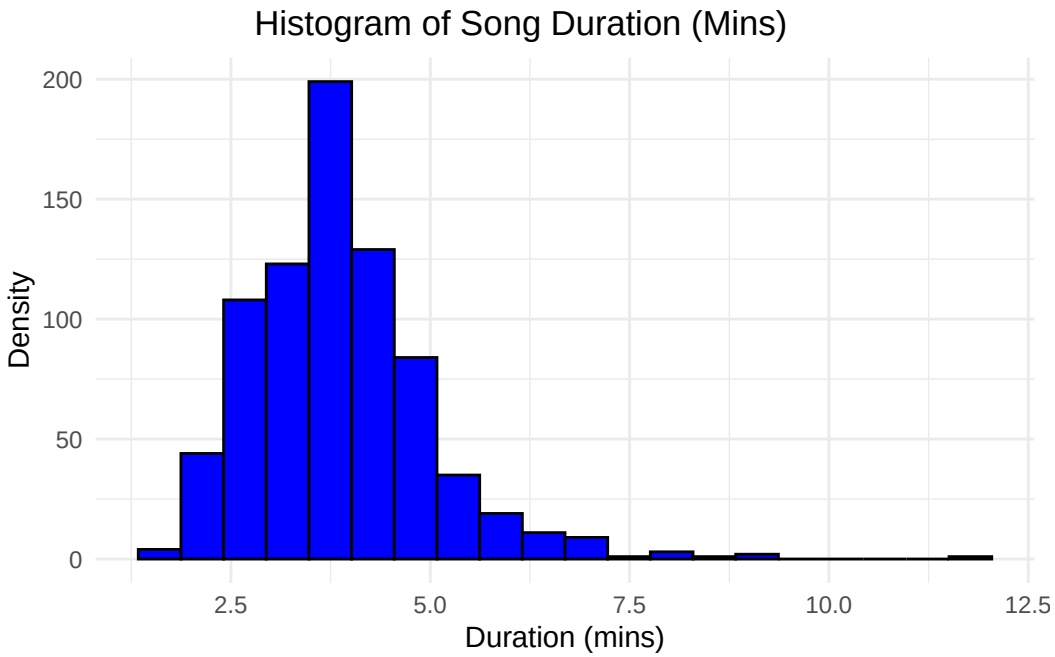


Figure 9: CAPTION

Most popular songs range from 2.5 to 5 minutes based on (**duration-histogram?**), but (**scatter-duration?**) reveals no linear relationship between duration and popularity.

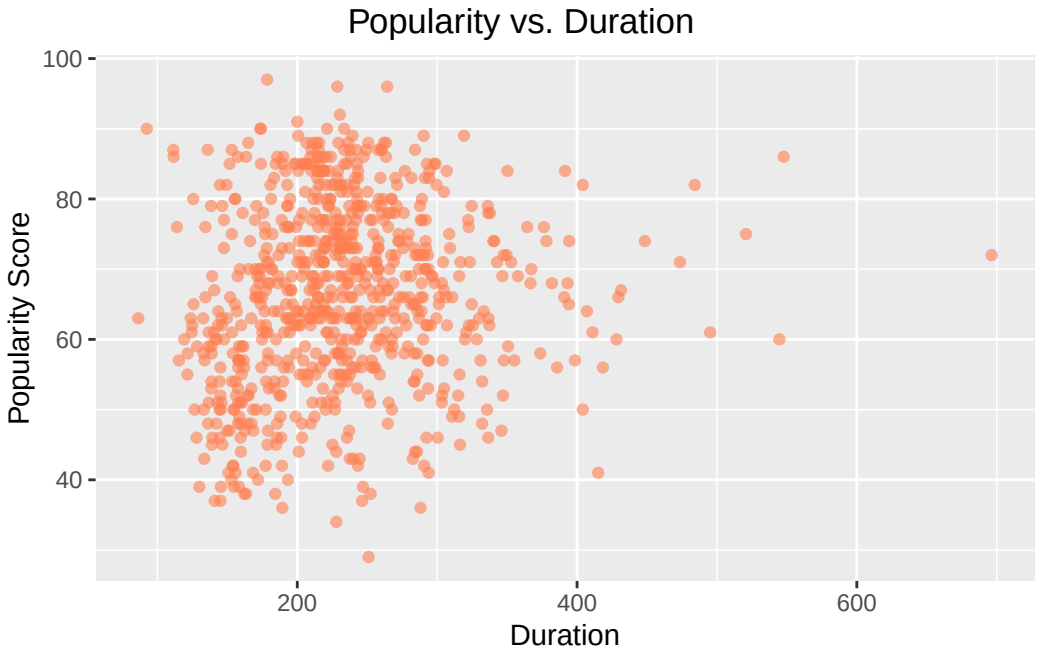


Figure 10: CAPTION

3 Data Modelling

While individual variable analyses provide some surface-level insights into the determinants of the popularity of a song, music's complexity necessitates considering variables together for reasonable and informative inference.

3.1 Model justification

Creating a linear regression model is primarily motivated by the fact that it can account for relationships between the variables and the popularity, as well the relationships between the predictor variables themselves. Running a linear regression model allows us to isolate effects of a particular variable while also controlling for the effects of the other variables at the same time.

Before conducting the analysis, however, several tests were conducted (shown in Section .1) to verify the data fit within the linear regression assumptions of linearity, normally distributed errors, and homoscedasticity of residuals. This ensured that the data was well suited for analysis using a multiple regression.

3.2 Model set-up

We are interested in investigating the relationship between several variables and popularity. These variables are: valence, danceability, mode, explicit lyrics, loudness, and duration. For each of the variables we are investigating, we have a null and an alternative hypothesis.

The null hypothesis (H_0): there is no significant linear relationship between one of the variables and a given Spotify popularity score, holding everything else constant.

The alternative hypothesis (H_1): a significant linear relationship exists between one of the variables and a given Spotify popularity score, holding everything else constant.

For a particular variable, a low p-value for the regression coefficient would provide evidence against the null hypothesis, indicating that the variable has a meaningful effect on Spotify popularity score. On the other hand, if the p-value is high, this would suggest that there is not enough evidence to reject the null hypothesis, and there may be no significant linear relationship between the variable and the Spotify popularity score. Ultimately, the results of the analysis will inform whether the null hypothesis can be rejected or not.

The equation for our linear model can be written as follows:

$S = \beta_0 + \beta_1 V + \beta_2 D + \beta_3 M + \beta_4 E + \beta_5 L + \beta_6 T + \epsilon,$

where,

- S is a Spotify popularity score.
- β_0 is our model constant, better known as S when all the other variables are 0,
- β_1 is effect of valence (measured 0-1) on S,
- β_2 is effect of danceability (measured 0-1) on S,
- β_3 is effect of musical mode (0 or 1) on S,
- β_4 is effect of explicit lyrics (0 or 1) on S,
- β_5 is effect of loudness (measured in decibels) on S,
- β_6 is effect of song duration (measured in seconds) on S, and
- ϵ is the the random error term, which accounts for variation in S that is not explained by the relationship with any of the other variables.

The aim of computing a regression is to estimate a line with the parameters β_0 to β_6 so that the difference (error) between the predicted line and the data points is minimized. By doing this, we get an equation for a line that best fits the data, allowing for the estimation of a given Spotify score for a set of audio features.

In the linear model, the dependent variable S represents the Spotify popularity score measured from 0 to 100.

Table 2: Linear Model of Spotify Popularity Summary

Term	Estimate	Std. Error	Statistic	P-value
(Intercept)	65.85	2.76	23.82	0.00
valence	-10.96	2.04	-5.38	0.00
danceability	18.24	3.27	5.58	0.00
mode	-0.85	0.95	-0.90	0.37
explicit	10.10	1.51	6.70	0.00
loudness	0.82	0.13	6.31	0.00
duration_secs	0.01	0.01	1.66	0.10

- The variable V denotes the valence (emotional positivity) of the song, measured from 0 to 1, where 1 is the most positive. The coefficient β_1 quantifies how a one-unit increase in valence (e.g., from 0.5 to 1.0) affects the popularity score S.
- The variable D represents the danceability of the song, also scaled from 0 to 1, with higher values indicating more danceability. The coefficient β_2 captures the change in S for a one-unit increase in danceability.
- The binary variable M indicates the musical mode (0 for minor, 1 for major), with β_3 reflecting the difference in popularity between major and minor keys.
- The binary variable E indicates explicit lyrics (0 for non-explicit, 1 for explicit), where β_4 estimates the effect of explicit content on popularity.
- The variable L denotes the loudness of the song in decibels, with β_5 quantifying how a one-decibel increase in loudness impacts the popularity score S.
- Finally, T represents the duration of the song in seconds, and β_6 estimates the effect of an additional second on the popularity score.

4 Results

By running our model with the data collected from Spotify, we get the following results:

The intercept (65.85) represents the predicted value of the dependent variable when all the independent variables are set to zero. However, since most of the predictors are likely scaled or binary, the interpretation of the intercept is not very meaningful.

Valence, which measures the emotional positivity of a song on a scale of 0 to 1, has a negative coefficient (-10.96). This suggests that, holding other variables constant, an increase in valence by one unit is associated with a decrease in the predicted Spotify popularity score by approximately 11 points.

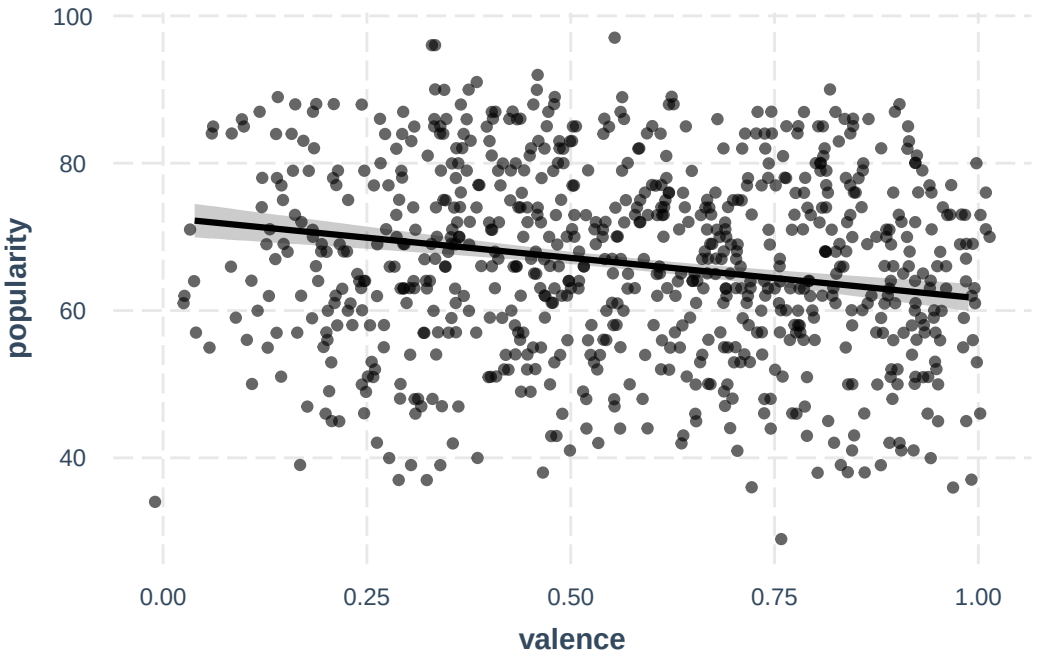
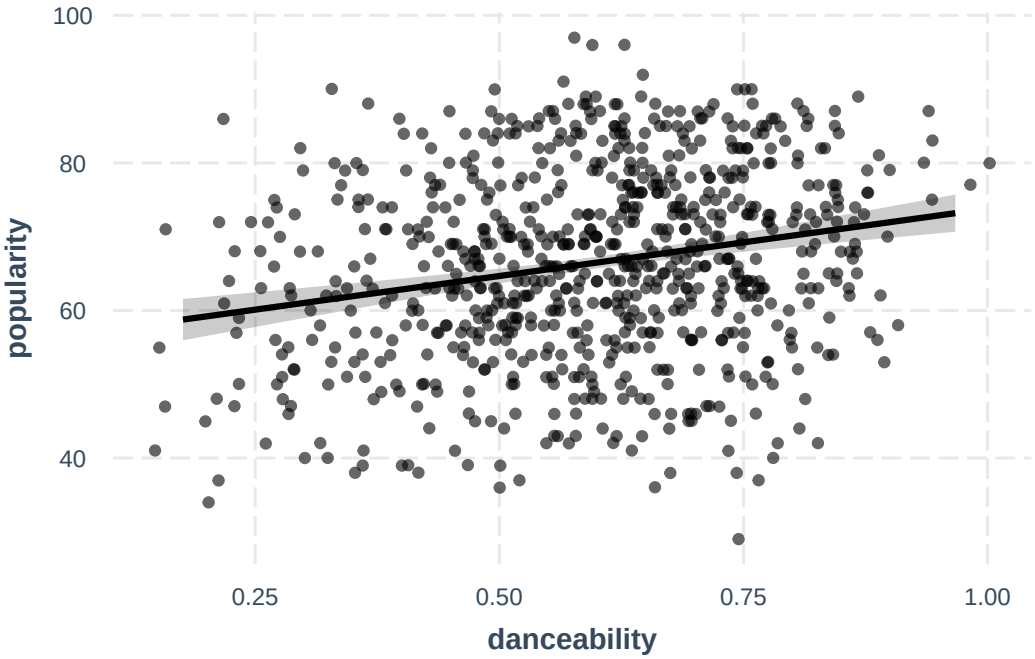
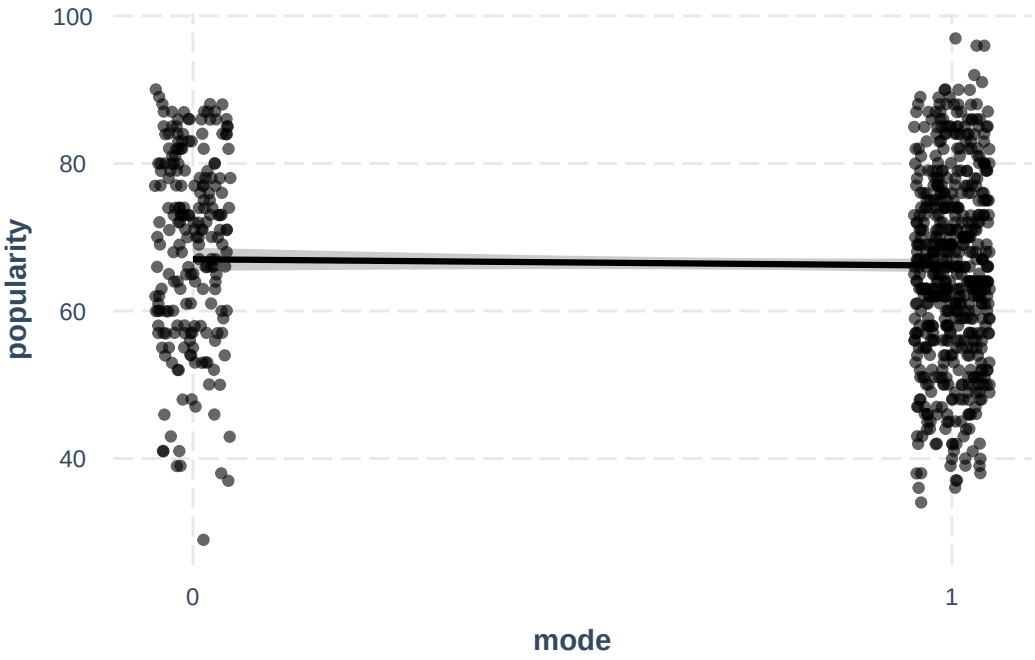


Figure 11: CAPTION

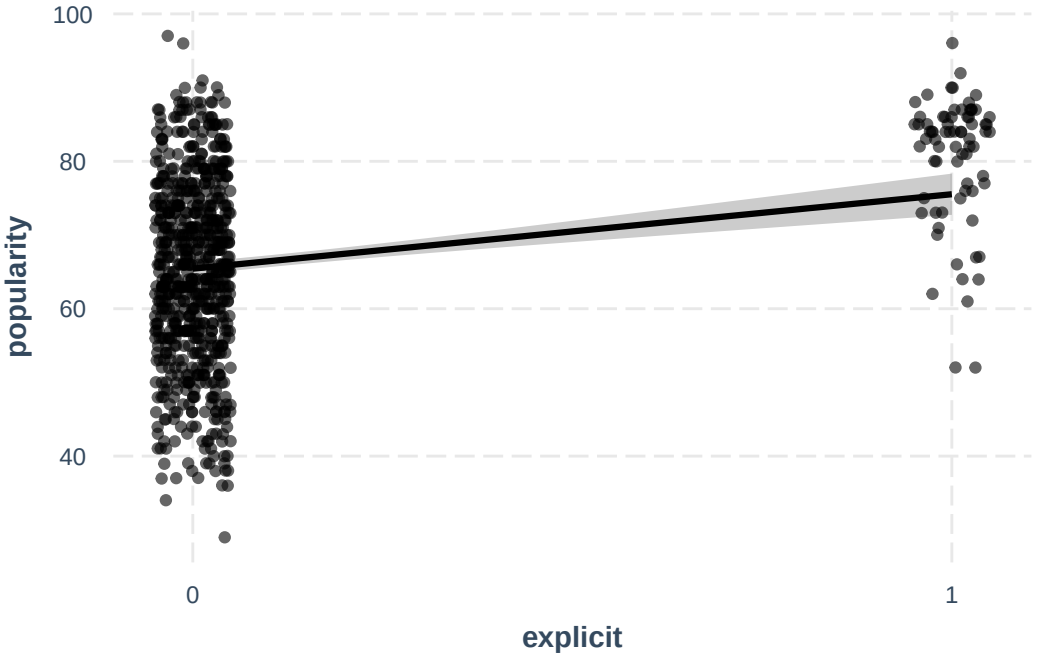
Danceability, also scaled from 0 to 1, has a positive coefficient (18.24). This indicates that, controlling for other factors, a one-unit increase in danceability is associated with an increase in the predicted popularity score by about 18 points.



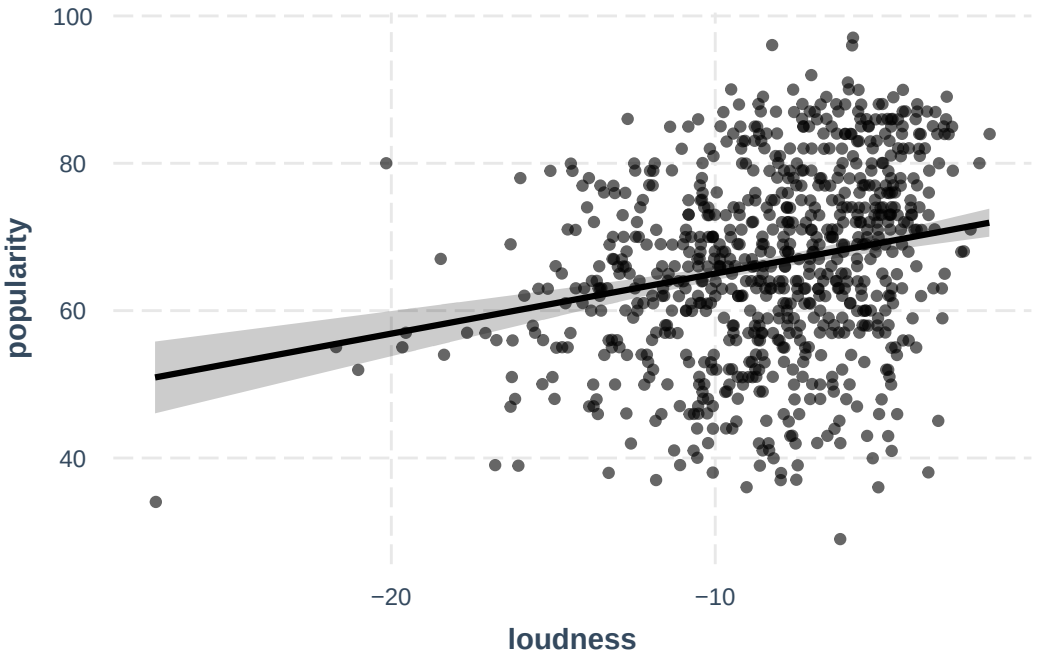
Mode is a binary variable (likely 0 for minor and 1 for major keys), and its coefficient (-0.85) is not statistically significant (p-value = 0.37), suggesting that the musical mode may not have a significant effect on the popularity score after accounting for other predictors.



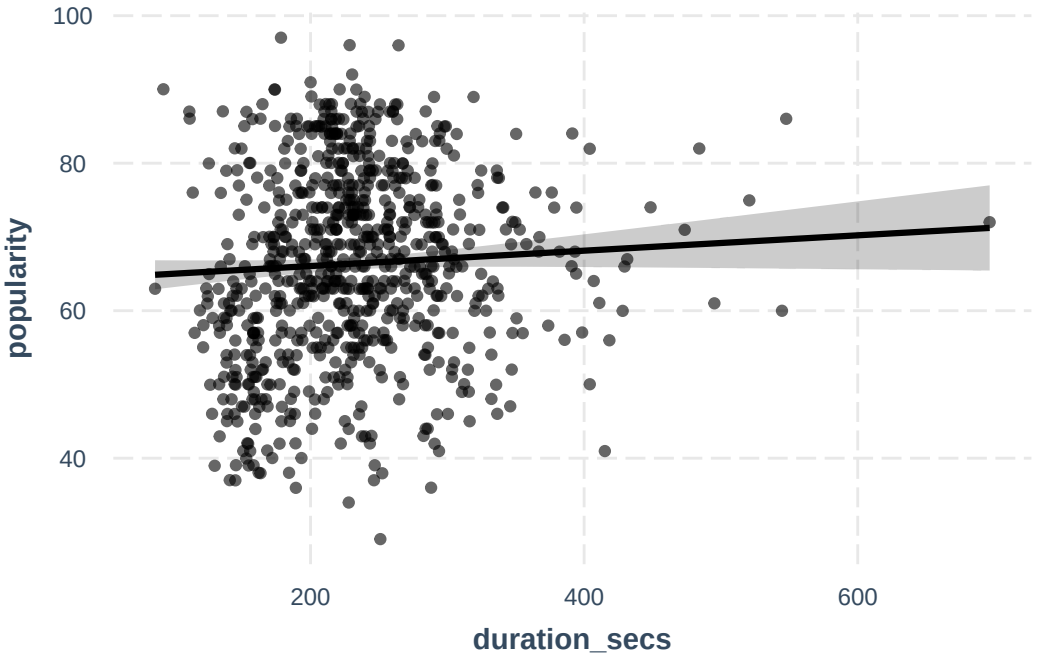
Explicit is another binary variable (possibly indicating the presence of explicit lyrics), with a positive coefficient (10.10). This implies that, holding other variables constant, songs with explicit lyrics are predicted to have a popularity score that is approximately 10 points higher than non-explicit songs.



Loudness, measured in decibels, has a positive coefficient (0.82), indicating that an increase in loudness by one decibel is associated with an increase in the predicted popularity score by 0.82 points, controlling for other factors.



Duration_secs, likely representing the song duration in seconds, has a positive coefficient (0.01) with a p-value of 0.10, which is marginally significant. This suggests that, holding other variables constant, an increase in song duration by one second is associated with a slight increase in the predicted popularity score by 0.01 points.



Taken together, the regression results indicate that danceability, explicit lyrics, loudness, and, to a lesser extent, song duration are associated with higher Spotify popularity scores, while valence (emotional positivity) has a negative association. The musical mode does not appear to have a significant effect on popularity after accounting for other predictors.

5 Discussion

5.1 Considerations of Spotify Metrics

5.2 Impact of Results

5.3 Weaknesses and next steps

Although estimates are relatively precise, increasing the sample size of songs could help resolve this issue and improve precision. Additionally, the Billboard data used ends in 2015, limiting the analysis's relevance to more recent years. Obtaining an updated list from Billboard or another source would enhance the dataset's timeliness. While the sample size is relatively small, this aligns with the focus on top artists.

Moving forward, constructing a more comprehensive dataset encompassing the entire discographies of these artists, with Spotify's assistance, could yield more precise effect estimates.

Repeating this analysis on top artists within specific genres or using different artist rating criteria could uncover genre-specific relationships or highlight the influence of other Spotify API variables not covered in this analysis. Exploring alternative data sources or metrics beyond Spotify's popularity score could also provide additional insights into the determinants of a song's widespread and enduring appeal.

6 References

2015. *Billboard*. https://www.billboard.com/charts/greatest-hot-100-artists/.

———. n.d. Accessed April 15, 2024. https://developer.spotify.com/documentation/web-api/reference/get-an-artist.

Alexander, Rohan. 2023. *Telling Stories with Data*. https://tellingstorieswithdata.com/.

Arel-Bundock, Vincent. 2022. "modelsummary: Data and Model Summaries in R." *Journal of Statistical Software* 103 (1): 1–23. https://doi.org/10.18637/jss.v103.i01.

Burke, Andrew E. 1997. "Small Firm Start-up by Composers in the Recording Industry." *Small Business Economics* 9 (6): 463–71. https://doi.org/10.1023/A:1007968604929.

Dodds, Peter Sheridan, and Christopher M. Danforth. 2010. "Measuring the Happiness of Large-Scale Written Expression: Songs, Blogs, and Presidents." *Journal of Happiness Studies* 11 (4): 441–56. https://doi.org/10.1007/s10902-009-9150-9.

Firke, Sam. 2023. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. https://CRAN.R-project.org/package=janitor.

Gagolewski, Marek. 2022. "stringi: Fast and Portable Character String Processing in R." *Journal of Statistical Software* 103 (2): 1–59. https://doi.org/10.18637/jss.v103.i02.

Gebru, Timnit, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé Iii, and Kate Crawford. 2021. "Datasheets for Datasets." *Communications of the ACM* 64 (12): 86–92.

Kim, Seon Tae, and Joo Hee Oh. 2021. "Music Intelligence: Granular Data and Prediction of Top Ten Hit Songs." *Decision Support Systems* 145 (June): 113535. https://doi.org/10.1016/j.dss.2021.113535.

McCormack, Tara. 2023. https://blog.tunemymusic.com/billboard-top-100-chart-what-is-it-how-is-it-calculated-and-how-to-access-it/.

Pham, James, Edric Kyauk, and Edwin Park. n.d. *Predicting Song Popularity*. https://cs229.stanford.edu/proj2015/140_report.pdf.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Richardson, Neal, Ian Cook, Nic Crane, Dewey Dunnington, Romain François, Jonathan Keane, Dragoș Moldovan-Grünfeld, Jeroen Ooms, Jacob Wujciak-Jens, and Apache Arrow. 2024. *Arrow: Integration to 'Apache' 'Arrow'*. https://CRAN.R-project.org/package=arrow.

Rosen, Sherwin. 1981. "The Economics of Superstars." *The American Economic Review* 71 (5): 845–58. https://www.jstor.org/stable/1803469.

Thompson, Charlie, Daniel Antal, Josiah Parry, Donal Phipps, and Tom Wolff. 2022. *Spotifyr: R Wrapper for the 'Spotify' Web API*. https://CRAN.R-project.org/package=spotifyr.

Wickham, Hadley. 2011. "Testthat: Get Started with Testing." *The R Journal* 3: 5–10. https://journal.r-project.org/archive/2011-1/RJournal_2011-1_Wickham.pdf.

———. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

———. 2022. *Rvest: Easily Harvest (Scrape) Web Pages*. https://CRAN.R-project.org/

package=rvest.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Jennifer Bryan, Malcolm Barrett, and Andy Teucher. 2023. *Usethis: Automate Package and Project Setup*. https://CRAN.R-project.org/package=usethis.

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. https://dplyr.tidyverse.org.

Wickham, Hadley, Jim Hester, and Jennifer Bryan. 2023. *Readr: Read Rectangular Text Data*. https://CRAN.R-project.org/package=readr.

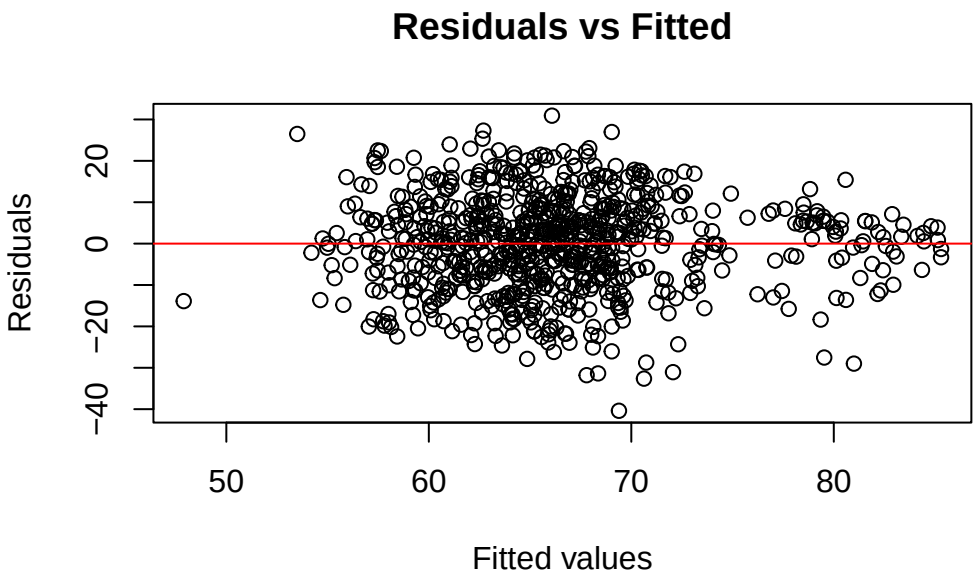
Wickham, Hadley, Jim Hester, and Jeroen Ooms. 2023. *Xml2: Parse XML*. https://CRAN.R-project.org/package=xml2.

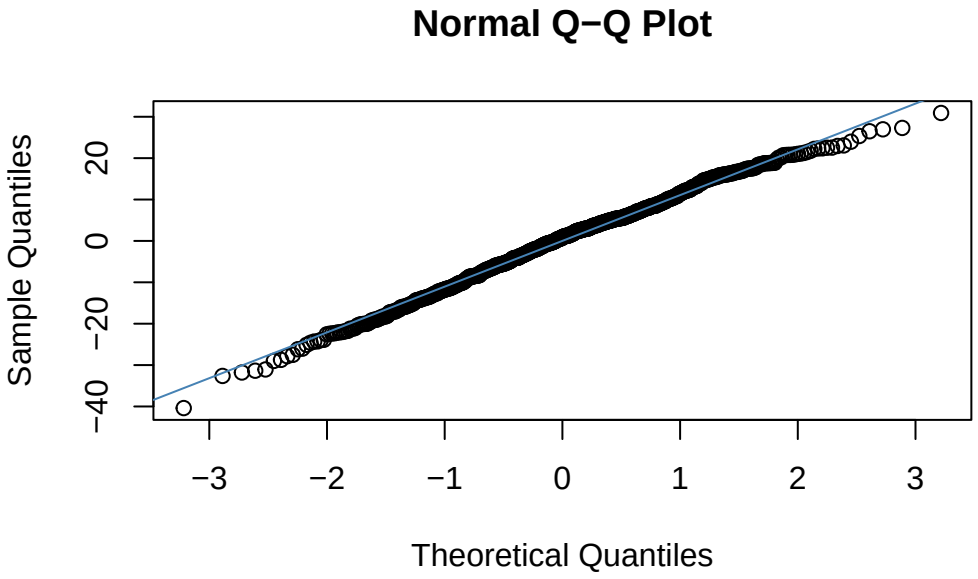
Xie, Yihui. 2023. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. https://yihui.org/knitr/.

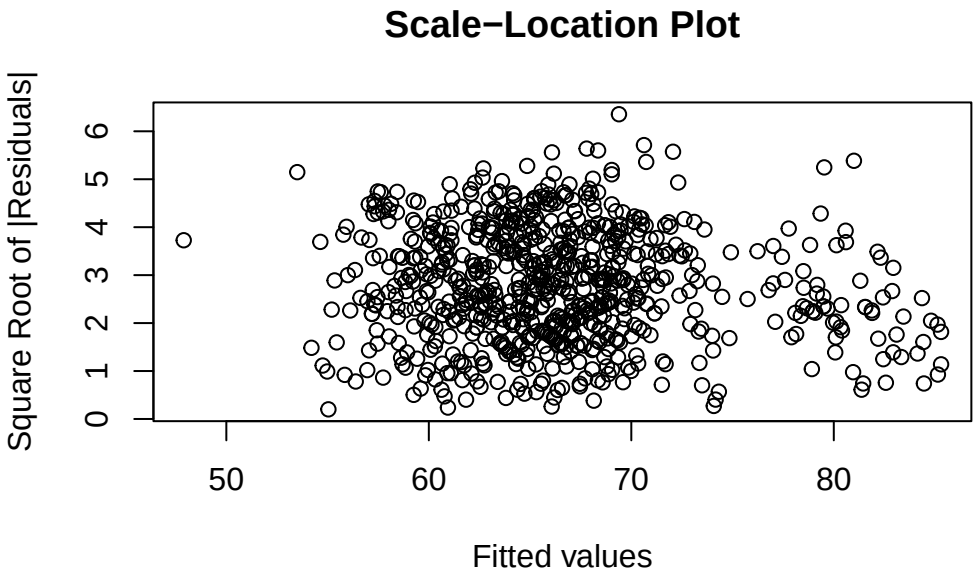
Zhu, Hao. 2021. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. https://CRAN.R-project.org/package=kableExtra.

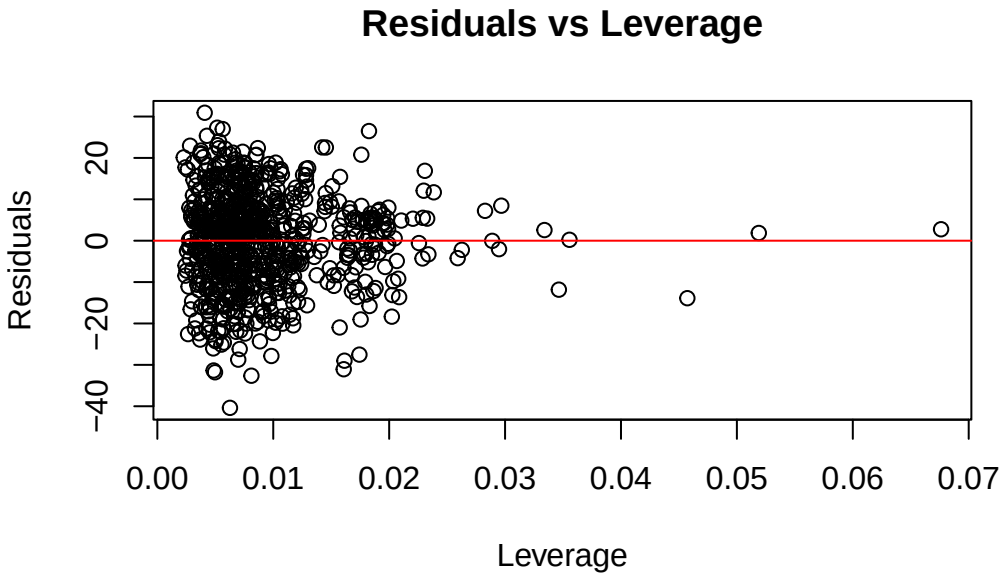
Appendix

.1 Model Testing










```
List of 2
 $ plot.title         :List of 11
  ..$ family       : NULL
```

```
 ..$ face         : NULL
 ..$ colour       : NULL
 ..$ size         : NULL
 ..$ hjust        : num 0.5
 ..$ vjust        : NULL
 ..$ angle        : NULL
 ..$ lineheight   : NULL
 ..$ margin       : NULL
 ..$ debug        : NULL
 ..$ inherit.blank: logi FALSE
 ..- attr(*, "class")= chr [1:2] "element_text" "element"
$ plot.title.position: chr "plot"
- attr(*, "class")= chr [1:2] "theme" "gg"
- attr(*, "complete")= logi FALSE
- attr(*, "validate")= logi TRUE
```