Investigating Popularity of Songs on Spotify

Dancing Queens: Karianna Klassen, Madeline Waterfield, Mark Yukelis, Shrikrishna Sriram

Introduction and Data

Background and Significance

Spotify, a Swedish company founded in 2006 by Daniel Ek and Martin Lorentzon, is self-described as "the world's most popular audio streaming subscription service." Across 183 markets, Spotify has 406 million users, 180 million subscribers, 82 million tracks, and 3.6 million podcasts (Spotify 2022a). Unfortunately for artists, the large number of users does not change the fact that only a small percentage of the 82 million tracks reach over 1 million streams. According to Spotify's Loud&Clear website that intends to increase transparency for artists, earning 1 million streams would put a song in the top 719,000 tracks, which means that only 0.87% of songs receive over a million streams (Spotify 2022c). Only 240 songs have reached Spotify's "Billions Club" by earning over one billion streams (Spotify 2022b). For small artists, "going viral" or earning a significant number of streams is difficult and unlikely. In an effort to help artists understand how to create a song that will be successful on Spotify, we will investigate attributes of popular songs and attempt to reach a consensus on what makes a song more successful than others.

Data

In the following project, we will investigate a dataset of 30,385 Spotify songs in order to determine whether songs with certain characteristics are more likely to be popular than others. The dataset spotify_songs.csv is from a TidyTuesday launch on January 21, 2020 (Tom Mock and Wolff 2020). The data was gathered by spotifyr, an R wrapper for pulling track audio features and other information from Spotify's Web API in bulk. The original dataset has 32,833 observations, or songs, and 23 variables, and was created by Kaylin Pavlik. However, because some songs are put on multiple playlists, those 32,833 observations include some duplicate songs. After cleaning the data by removing three playlist identifiers (subgenre, id, and name) and keeping only distinct observations, our dataset has 30,385 observations, and each observation is a song. Thus, the dataset we will be working with has 30,385 observations and 20 variables. We will be focusing on six variables: popularity, genre, length, tempo, loudness, and acousticness. Song popularity refers to the popularity of a song (or number of streams) relative to other songs in the dataset, ranked from 0-100 (where a higher value denotes a more popular song). Genre refers to the genre of the playlist the song is located on, of which there are six: Pop, Rap, R&B, Latin, EDM, and Rock. Song duration in the original dataset refers to song length in milliseconds, but we mutated a new variable to measure song length in minutes. Song tempo refers to average number of beats per minutes (BPM) in a track, and represents the speed or pace of the song. Song loudness refers to the average loudness of a song in decibells (dB), ranging from 0 to -60 dB (with 0 being the loudest). Song acousticness is a measure of how confident the Spotify API is that a song is acoustic, meaning it lacks electronic modification. Acousticness is measured from 0 to 1.0, with 1.0 being 100% confident that a song is acoustic. The higher the acousticness score, the less electronic modificaiton it contains. Popularity, genre, length, tempo, loudness, and acousticness were originally determined by Spotify's Web API.

Research Question

Our research question is: **What characteristics make a song on Spotify more likely to be popular than other songs on Spotify?** We are endeavouring to find out what qualities make a song popular in order to be able to tell artists what kind of song has the most likelihood of becoming popular. We can break up our research question into two smaller research questions:

Research Question 1: Do some genres on Spotify tend to have more popular songs than others? Before our exploration, we hypothesize that the pop genre will have the most popular songs, while all other genres will not be significantly different than the mean.

Research Question 2: Do length, tempo, loudness and acoustiness tend affect popularity on Spotify, and in what way? Before our exploration, we hypothesize that they all are significant, with length having an overall negative effect while tempo, loudness and acoustiness have an overall positive effect.

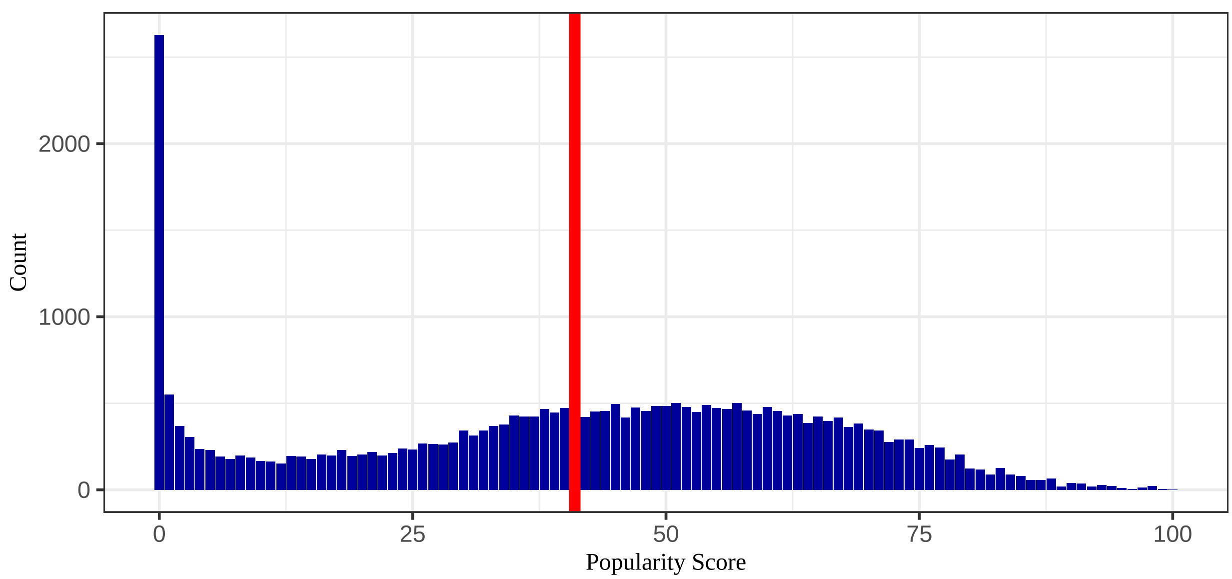
Methodology for Exploratory Data Analysis.

When conducting our analysis, our team sought to understand what qualities make a song more likely to be popular. To do this, we decided to compare the mean attributes of the top 10% of songs in terms of popularity and the mean attributes of the bottom 90% of songs in terms of popularity, hoping that differences (or continuities) in summary statistics would give us insight into how more popular songs are different from less popular songs.

Exploratory Data Analysis

Genre

Figure 1: Number of Tracks per Popularity Score



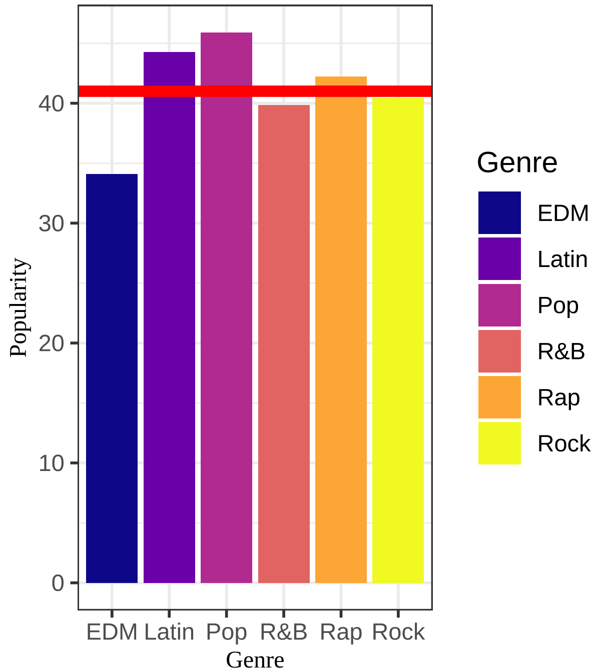
*Red line represents average popularity

As seen in Figure 1, the average popularity score of all songs in the dataset is 41.004.

Table 1: Genre and Popularity Summary Statistics

Genre	Count	% of Total	Avg. Popularity	Diff. from Overall Avg.
Pop	5132	16.89	45.91	4.90
Latin	4641	15.27	44.26	3.25
Rap	5486	18.05	42.21	1.21
Rock	4451	14.65	40.43	-0.57
R&B	5138	16.91	39.84	-1.16
EDM	5537	18.22	34.07	-6.93

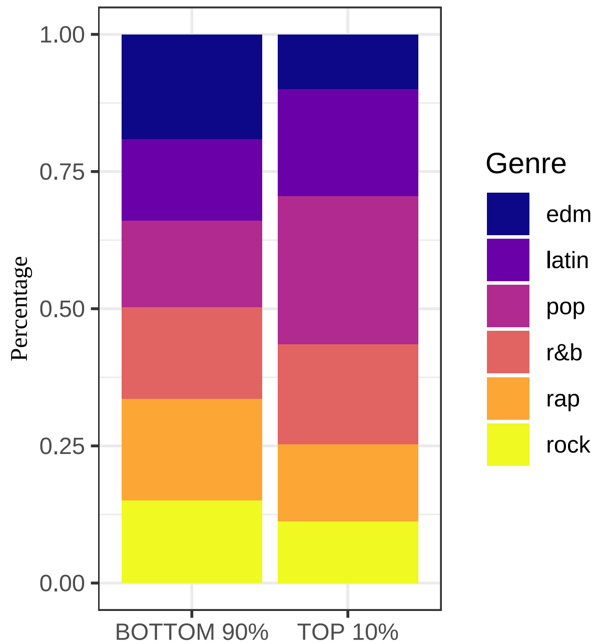
Figure 2: Average Popularity of Each Genre



*Red Line Shows Overall Average Popularity

Figure 3: Proportions of Genre

For the Bottom 90% vs. the Top 10%



As seen in Table 1 and Figure 2, the genres Pop, Latin, and Rap all have higher average popularity scores than the overall average popularity, while R&B and EDM have lower average popularity scores. Figure three shows that the Pop genre has the largest percentage of songs in the top 10%, and that the proportion of Pop music in the top 10% is larger than the proportion of Pop music in the bottom 90%, which further points to Pop music being the most popular.

Length, Tempo, Loudness, & Acousticness

Table 2: Summary Statistics for Length, Tempo, Loudness, & Acousticness; Comparison of Top 10% and Bottom 90%

Percentile	Avg. Length (min)	Avg. Tempo (BPM)	Avg. Loudness (dB)	Avg. Acousticness
BOTTOM 90%	3.78	120.91	-6.84	0.17
TOP 10%	3.59	120.67	-6.10	0.20

It is interesting that there is not a lot of difference between our averages: There is a 0.19 min difference in length between the bottom 90% and the top 10%, a 0.24 BPM difference in tempo, a -0.74 dB difference in loudness, and a -0.03 point difference in acousticness. These small differences imply that length, tempo, loudness, and acousticness, may not have a significant impact on popularity, but that requires further investigation.

Figure 4: Density Plot of the Tempo of top 10% vs bottom 90% of songs

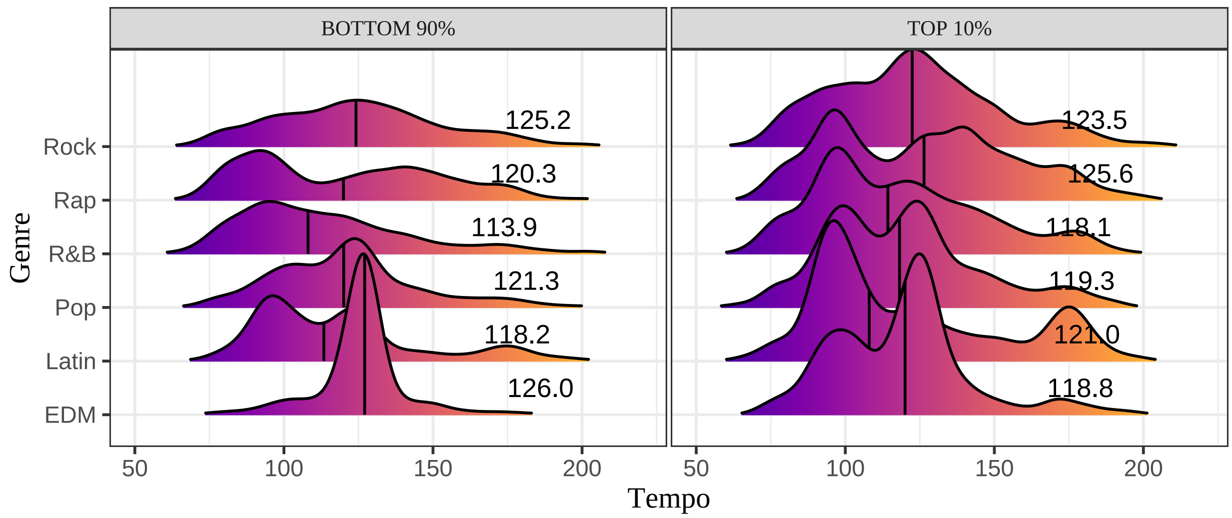


Figure 4 shows that there are only small differences in the average tempo of the top 10% and the bottom 90% of songs when separated by Genre. It is interesting to note that the distribution of tempo is somewhat different. This requires further investigation (Wilke 2017).

Figure 5: Density Plot of Length of top 10% vs bottom 90% of songs

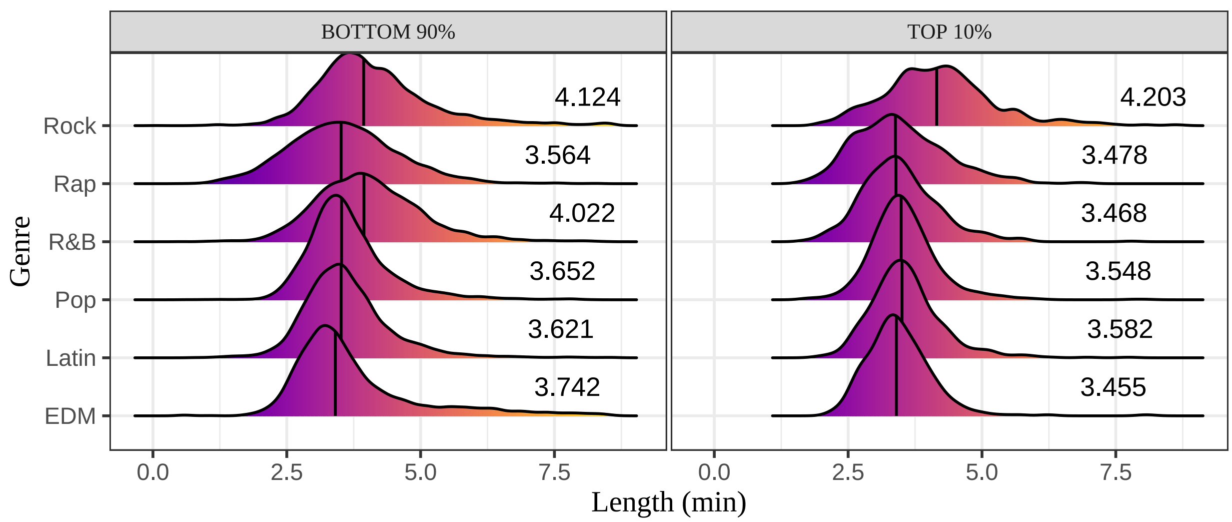


Figure 5 shows for all genres except Rock, the average length for the bottom 90% of songs is higher than the average length for the top 10% of songs, suggesting that slightly shorter songs may do better in those genres. For Rock, the average length for the bottom 90% of songs is 0.079 minutes less than the average length for the top 10%, suggesting maybe that longer songs may do slightly better in the Rock genre. This requires further investigation (Wilke 2017).

Figure 6: Density Plot of Loudness of top 10% vs bottom 90% of songs

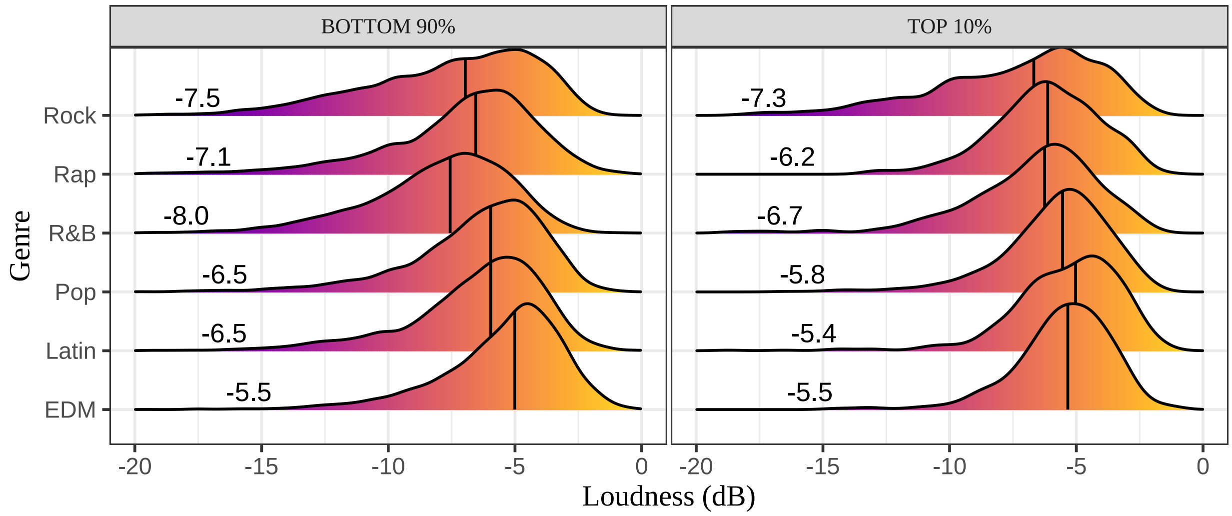


Figure 6 shows that for all genres except EDM, the average loudness of the bottom 90% of songs is lower (thus quieter) than the average loudness of the top 10% of songs, suggesting that louder songs may be popular for those genres. For EDM, the average loudness is the same for both percentiles (-5.5 BPM), suggesting that loudness may not contribute to popularity for that genre. This requires further investigation (Wilke 2017).

Figure 7: Density Plot of Acousticness of top 10% vs bottom 90% of songs

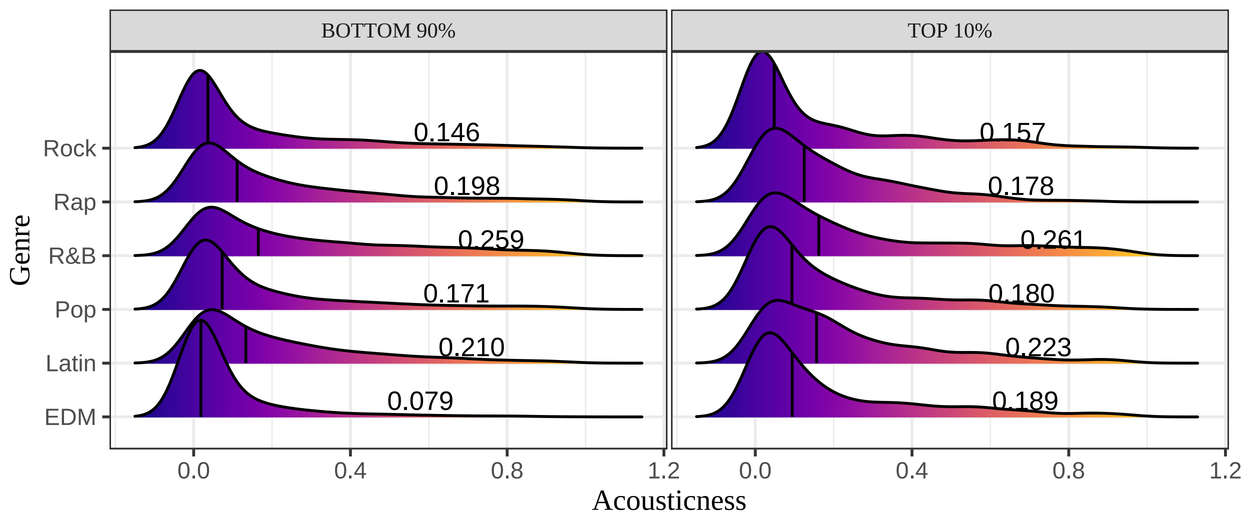


Figure 7 shows that for all genres except rap, average acousticness for the bottom 90% of songs is lower than the average acousticness for the top 10% of songs. This may suggest that more acousticness may correlate to a more popular song for those genres (Wilke 2017).

Methodology for Results:

Given our exploratory data analysis, we decided the best way to answer our research questions was through a combination of simulated null distributions and multiple regression models. We simulated null distributions to see if certain genres overall were significantly more popular than others; we used multiple regression models to see if certain factors are significant in predicting popularity and in what way.

After the exploratory analysis we conducted, we decided to examine exclusively the Pop and Rock genres. We chose to examine the Pop genre because we found that the average popularity of the Pop genre was the largest, suggesting that perhaps Pop songs are more popular on average than any other genre, and we chose to examine the Rock genre because its average popularity deviated the least from the overall average popularity. The overall average popularity is 41.004.

Another reason we selected these genres was that we saw that they tend to behave differently (see figures 4 through 7). To further delve into this and answer the second part of our research question, we generated multiple regression models for these two genres and also for the entire dataset(Fox 2021). This allowed us to see if these two genres behaved differently from the dataset overall and each other. The multiple regression models include the variables length, tempo, acousticness, and loudness. We chose those variables because they are independent from other variables. Other variables in the dataset, like danceability, depend on things like tempo and rhythm stability. However, length, tempo, acousticness, and loudness are not dependent on any other variables.

Results

Research Question 1: Do some genres tend have more popular songs than others?

Simulated Null Distribution for Pop Genre:

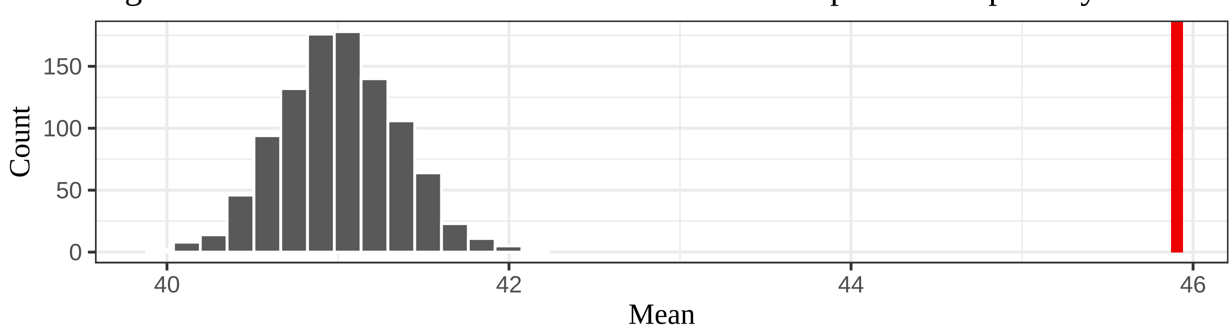
H_0: There is no significant difference between the mean popularity of the data set as a whole and the mean popularity of songs in the pop genre.

H_A: The mean popularity of the pop genre is significantly greater than the overall mean popularity of the data set.

$$H_0 : \mu_{pop} = \mu_{overall}$$

$$H_A : \mu_{pop} > \mu_{overall}$$

Figure 8: Simulation-Based Null Distribution for Pop Genre Popularity



p-value
0

Given $\alpha = 0.05$ and a p value of 0, we reject our null hypothesis. In context, we conclude that the mean popularity of a song in the pop genre is significantly greater than the overall mean of the data set. Overall, this answer part of our research question in that we know pop songs tend to be more popular than others.

Simulated Null Distribution for Rock:

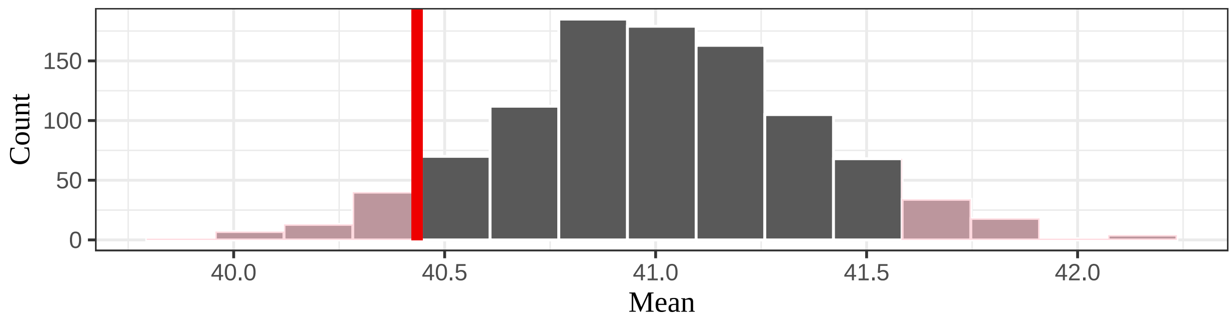
H_0: There is no significant difference between the mean popularity of the data set as a whole and the mean popularity of songs in the rock genre.

H_A: There is a significant difference between the mean popularity of the data set as a whole and the mean popularity of songs in the rock genre.

$$H_0 : \mu_{rock} = \mu_{overall}$$

$$H_A : \mu_{rock} \neq \mu_{overall}$$

Figure 9: Simulation-Based Null Distribution for Rock Genre Popularity



p-value
0.114

Given $\alpha = 0.05$ and a p value of 0.114, we fail to reject our null hypothesis. For the rock genre, we can conclude that its mean is not significantly different than the overall mean of the data set. While pop songs tend to be more popular, rock songs are average when it comes to their popularity.

Research Question 2: Do length, tempo, loudness and acoustiness tend affect popularity, and in what way?

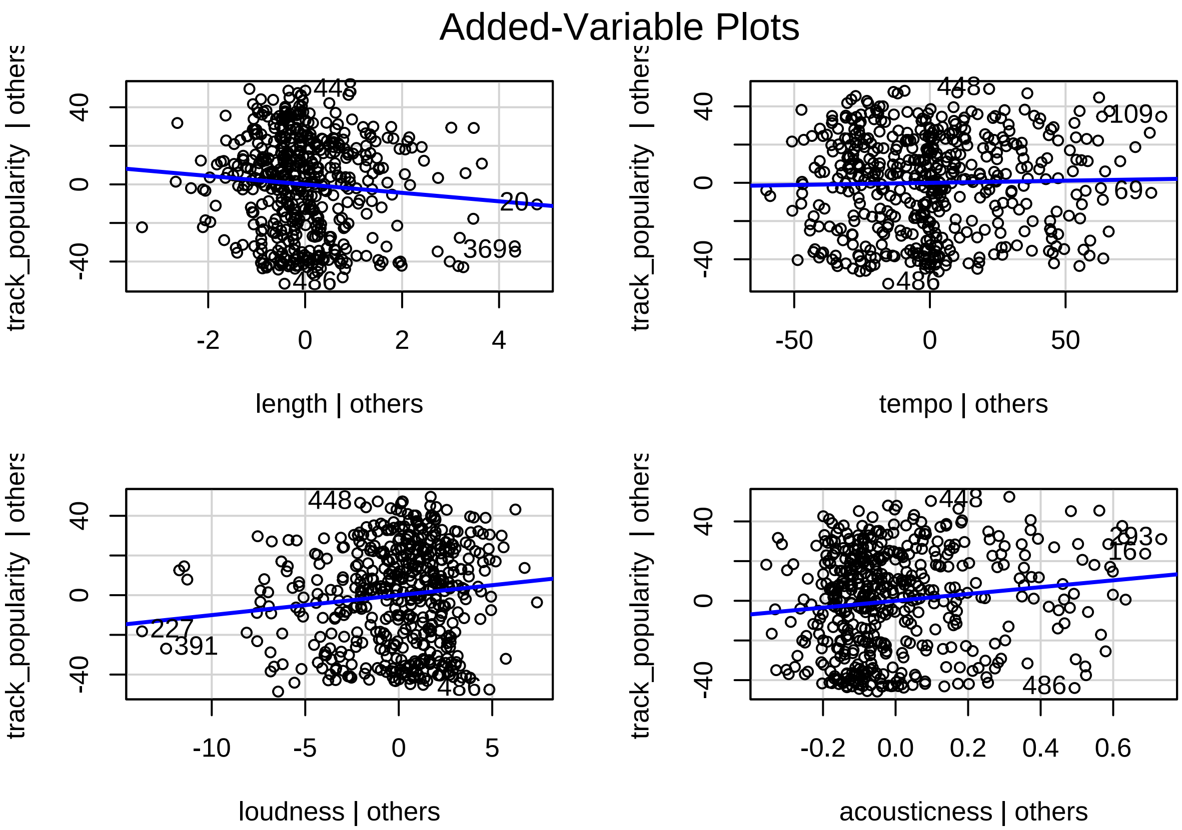
Overall Multiple Regression Model

term	estimate	std.error	statistic	p.value
(Intercept)	54.3060195	0.8848123	61.3757507	0.0000000
length	-3.0992066	0.1393814	-22.2354432	0.0000000
tempo	0.0039545	0.0051615	0.7661538	0.4435907
loudness	0.6217895	0.0499428	12.4500269	0.0000000
acousticness	11.7662402	0.6760409	17.4046270	0.0000000

Given a $\alpha = 0.05$ and p-values of 0 for length, loudness, and acousticness, we can reject the null that β is zero and say that these factors are significant predictors of popularity. It is interesting to note that length negatively affects popularity, while loudness and acousticness positively affect popularity. Lastly, given a p-value of 0.4435907 for tempo, we fail to reject the null that β is zero and conclude that tempo does not have a significant affect on popularity. Figure 10 visualizes each variable.

Our fitted model is: $\hat{y} = 54.3 - 3.1x_1 + 0.6x_2 + 11.7x_3$ where x_1 is length, x_2 is loudness, and x_3 is acousticness.

Figure 10: Visualizing Linear Regression of Length, Tempo, Loudness, and Acousticness for All Genres



Note 2: The notation A / Others depicts that all variables that are not on the X or Y axis are held constant.

Pop Multiple Regression Model

term	estimate	std.error	statistic	p.value
(Intercept)	70.2614936	2.4120698	29.129129	0.0000000
length	-3.8454177	0.4504469	-8.536895	0.0000000
tempo	-0.0340889	0.0136832	-2.491286	0.0127595
loudness	1.2116132	0.1446872	8.374020	0.0000000
acousticness	8.5529279	1.7168332	4.981805	0.0000007

Given a $\alpha = 0.05$ and the p-values of above, we can reject the null that β is zero and say all of our factors are significant predictors of popularity. It is interesting to note that tempo for pop has a negative affect on popularity, whereas it was not a significant predictor for the data set overall. It is also interesting that acoustiness does not affect popularity as much as it did in our overal model, whereas length, tempo and loudness all have stronger affects than in our overall model.

Our fitted model is: $\hat{y} = 70.26 - 3.8x_1 - 0.03x_2 + 1.2x_3 + 8.6x_4$ where x_1 is length, x_2 is tempo, x_3 is loudness, and x_4 is acousticness.

Rock Multiple Regression Model

term	estimate	std.error	statistic	p.value
(Intercept)	45.0786226	2.3605649	19.0965403	0.0000000
length	-0.8131631	0.3406902	-2.3868112	0.0170368
tempo	-0.0010234	0.0131465	-0.0778468	0.9379534
loudness	0.1866091	0.1240823	1.5039144	0.1326744
acousticness	1.6997955	1.9482296	0.8724821	0.3829925

Given a $\alpha = 0.05$ and the p-values of above, we can reject the null that β is zero and say length is significant predictor of popularity. On the other hand, we fail to reject the null that β is zero for tempo, loudness and acousticness and conclude that they are not significant predictors of popularity. This is notably different than what we saw with the pop genre and with our model for the entire dataset.

Our fitted model is: $\hat{y} = 45.1 - 0.8x_1$ where x_1 is length.

Discussion

Conclusions:

We looked at data from Spotify to examine what makes a song popular. We found that across genres, not only is popularity different, but also how certain factors affect popularity.

For our first research question, we asked *Do some genres tend have more popular songs than others on Spotify?*, and chose to focus specifically on the Rock and Pop genres after our exploratory data analysis. Looking at average popularity scores for Pop songs in comparison to average scores for all songs, we found that music in the pop genre tends to be significantly more popular that other genres. For music in the rock genre, we see average popularity. This supports part of the hypothesis we stated in the beginning: Pop songs are more popular than other genres, and the Rock genre has average popularity. However, we hypothesized that the popularity of the other four genres, Rap, R&B, Latin, and Edm, would not be significantly different than the average popularity, but we did not completed simulation-based hypothesis testing for those genres, so we cannot make a judgement on them.

For our second research question, we asked *do length, tempo, loudness and acoustiness tend affect popularity, and in what way?*. For the data set as a whole, we found insufficient evidence to demonstrate tempo to be a significant predictor of popularity. However, loudness, acousticness, and length were all found to be significant predictors of popularity. We also saw pop behaving slightly different from the data set as a whole, while rock behaved much more different. For pop, we saw all our factors being significant when it comes to predicting popularity. For rock, we saw only length being a significant predictor. It is interesting though that all factors that are significant in several models had the effected popularity the same way (either positively or negatively) in all of those models. Shorter length, higher tempo, smaller loudness (loudness is a negative variable, so a positive β means a negative relationship), and higher acousticness increase popularity when those factors are significant.

Looking back at the goal of our project, we wanted make some recommendations to new or small artists who want to make more popular songs. Based on our findings, our recommendation to them would be to make Pop songs that are shorter, slightly quieter, and with higher acousticness (lower electronic modification).

Limitations:

One limitation of our findings is that while we attempted to remove all duplicate songs from our dataset by removing three playlist identifiers, we had to keep one playlist identifier, genre, in order to examine the genre of the songs. This means that there may be some songs that factored into the data of two genres, because many songs are considered to be part of two genres (for example, Pop-Rock). While we removed as many duplicates as we could, the possible presence of duplicates in our data means that our findings may not be accurate.

Another limitation is that it is unclear whether Kaylin Pavlik gathered the data randomly from public playlists, private playlists, or both. The data being from only public or private playlists could mean that it is not representative of the entire population.

Table 8: R^2 Values Across Regression Models

Overall	Pop	Rock
0.0321	0.036	0.002

A further limitation is in the method of analysis we used in answering our second research question. For our multiple regression model, R^2 is very low across all models. This indicates that we do not model the variability of our data well. If we were looking to better model the data rather than analyze what factors are significant, it would probably be better to test other types of regression. For example, one would assume that song loudness is not a purely negative relationship, as a song that is excessively quiet would likely be listened to less.

Further Research:

We only looked at two genres in our evaluation of the impact of genre on popularity, so it would be interesting to see how other genres influence song popularity to expand on our findings on genre. If we redid our project, we would attempt to examine all of the genres. We could also look into how different factors work in interaction with one another. For instance, it maybe true that people expect certain tempos of certain genres and songs that fit with said expectations are more popular. We looked at factors independently but they definitely could potentially act in interaction to impact popularity.

It would also potentially be a good next step to look to other music streaming services like Pandora or Amazon Music, as well as more outdated modes of music listening like album and cd sales, to see how those results compare to those gleaned from Spotify data. Further, it might be interesting to examine what factors impact the type of music a person likes such as age, gender, location, etc. To find data on this, it might be possible to find something from Spotify or other streaming platforms, though data sharing rules might limit our ability to get this data straight from the source. One way to collect this data would be to randomly sample Spotify users and have them share basic information on themselves as well as the information from the year in review Spotify sends out every year.

If we were to start over, as mentioned in the limitations section, we might look to testing other types of regression to better account for how these factors operate in the real world. We also might look into smaller fractions of popularity, i.e. the top 1% of songs, and see if there is different behavior. This would give us interesting insight on the difference between songs that are just popular and the songs that are true hits.

References

Fox, John. 2021. *Car: Companion to Applied Regression.* John Fox, CRAN. https://cran.r-project.org/web/packages/car/index.html.

Spotify. 2022a. *About Spotify.* Spotify. https://newsroom.spotify.com/company-info.

———. 2022b. *Billion's Club Playlist.* Spotify. https://open.spotify.com/playlist/37i9dQZF1DX7iB3RCnBnN4.

———. 2022c. *Loud &Amp; Clear by Spotify.* Spotify. https://loudandclear.byspotify.com/?question=a-million-streams.

Tom Mock, Josiah Parry, Charlie Thompson, and Tom Wolff. 2020. *Spotify Songs: A Database of the Musical Characteristics of Songs.* Tidy Tuesday. https://github.com/rfordatascience/tidytuesday/tree/master/data/2020/2020-01-21.

Wilke, Claus. 2017. *Ggridges: A CRAN Package for the Construction of Ridgeline Plots.* Wilke Lab, CRAN. https://wilkelab.org/ggridges/.