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. 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 intended 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 recieve over a million streams. Only 240 songs have reached Spotify's "Billions Club" by earning over one billion streams. 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 succesful on Spotify, we will investigate attributes of popular songs and endeavour to reach a consensus on what makes a song more successful than others.

Data

In the following project, we will investigate a dataset of 28,356 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. The data was gathered by spotifyr, an R wrapper for pulling track audio features and other information from Spotify's Web API in bulk. There are 32,833 observations, or songs, in this data set and 23 variables, which means that for each song there are 23 variables to identify it, like song name, artist, tempo, playlist name, and release date. However, because some songs are put on multiple playlists, those 32,833 observations include some duplicate songs. After cleaning the data by removing playlist identifiers and keeping only distinct observations, our dataset has 28,356 observations, and each observation is an individual song. In cleaning the data we had to remove the four variables associated with playlist information (playlist name, playlist genre, playlist subgenre, and playlist ID), so 19 variables remain. Thus, the dataset we will be working with has 28,356 observations and 19 variables. We will be focusing on five variables: song popularity, song danceability, song speechines, song mode, and song duration. 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 higher is better). According to TidyTuesday, "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." Mode describes whether the song is in major or minor key (major is represented by 1 and minor is represented by 0). TidyTuesday describes speechiness as "the presence of spoken words in a track. The more exclusively speech-like the recording (e.g. talk show, audio book, poetry), the closer to 1.0 the attribute value. Values above 0.66 describe tracks that are probably made entirely of spoken words. Values between 0.33 and 0.66 describe tracks that may contain both music and speech, either in sections or layered, including such cases as rap music. Values below 0.33 most likely represent music and other non-speech-like tracks." Song duration refers to the length of the genre in minutes. Popularity, danceability, mode, speechiness, and duration were originally calculated by Spotify's Web API.

Research Question

Our research question is: *do danceability, mode, speechiness, and duration influence the popularity of a song 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 four smaller research questions:

Research Question 1: Do more popular songs have a significantly different danceability than less popular songs? Before analysis, our hypothesis is that more popular songs have a higher danceability than less popular songs.

Research Question 2: Are more popular songs signficantly more likely to be in the major key than less popular songs? Before analysis, our hypothesis is that more popular songs are more likely to be in major key than less popular songs.

Research Question 3: Do more popular songs have a significantly different speechiness than less popular songs? Before analysis, our hypothesis is that more popular songs and less popular songs have the same average speechiness.

Research Question 4: Do more popular songs have a significantly different duration than less popular songs? Before analysis, our hypothesis is that more popular songs have a lower duration than less popular songs.

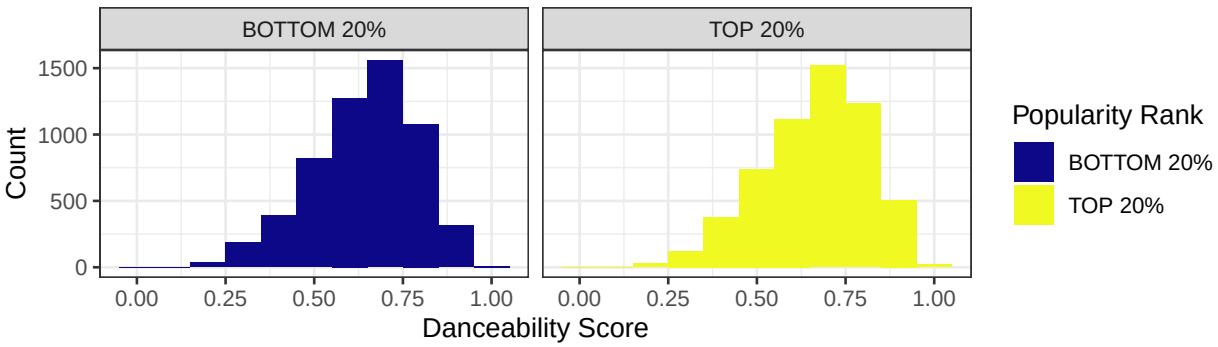
Methodology

When conducting our analysis, our team sought to understand *out of danceability, mode, speechiness, and duration, which factors influence the popularity of a song on Spotify*? To do this, we decided to compare the average attributes of the bottom 20% of songs in terms of popularity with the average attributes of the top 20%. We made this decision because we only wanted to examine the songs that were very unpopular or very popular and we did not want the songs whose popularity was in the middle to decrease the average popularity or attributes of popular songs or increase the average popularity or attributes of unpopular songs. By pulling out the top 20% of songs in terms of popularity and the bottom 20% of songs in terms of popularity, we pulled out the top 5,671 songs and the bottom 5,671 songs, leaving us with 11,342 songs for analysis. In the following analysis, we will be comparing the top 20% of songs with the top 20% of songs in order to collect summary statistics.

Exploratory Data Analysis

Danceability

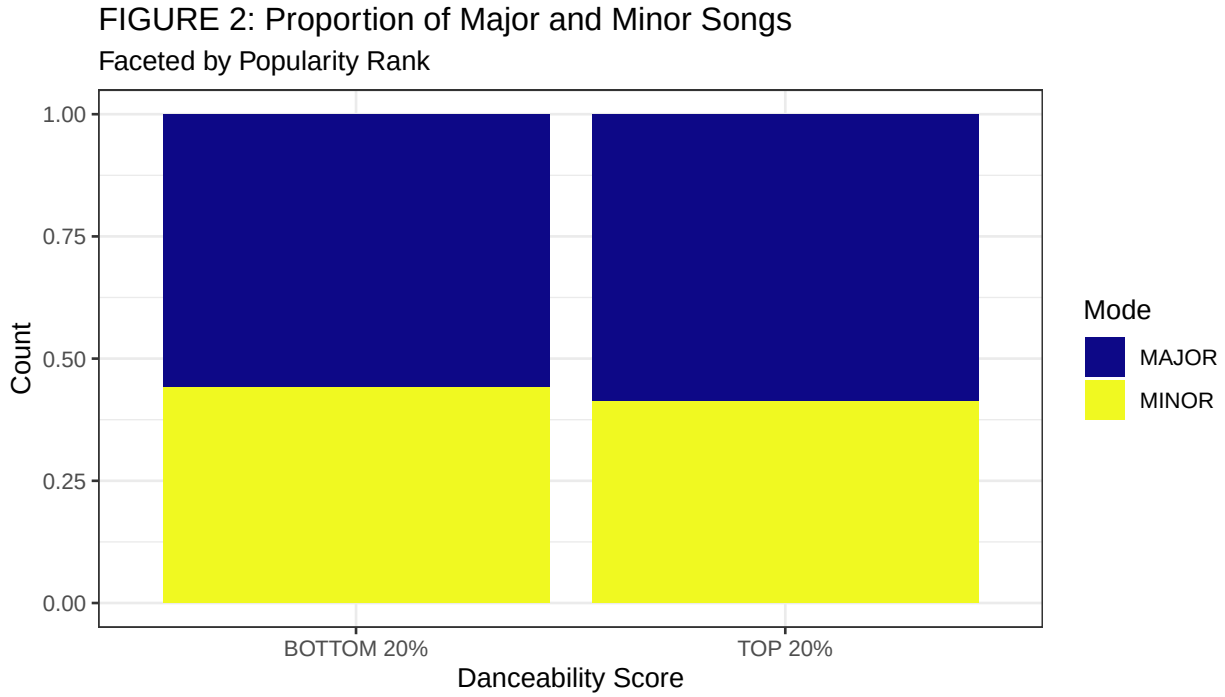
FIGURE 1: Distribution of Danceability Scores



As seen in Figure One, the bottom 20% of songs have a similar distribution of danceability scores as the top 20% of songs. However, with calculations, we determined that the mean danceability of songs in the top 20%

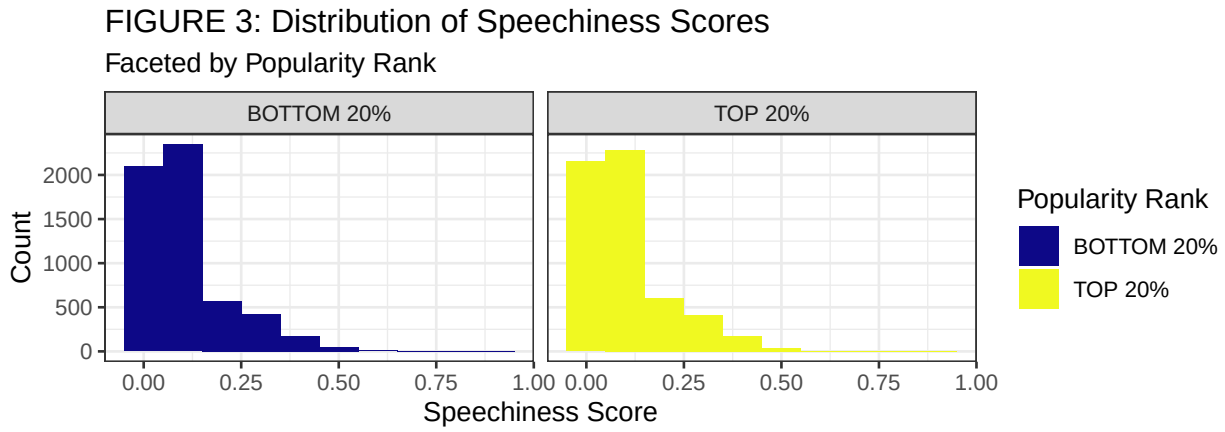
is 0.664 and the mean danceability of songs in the bottom 20% is 0.643, which means the difference in the means is 0.021.

Mode



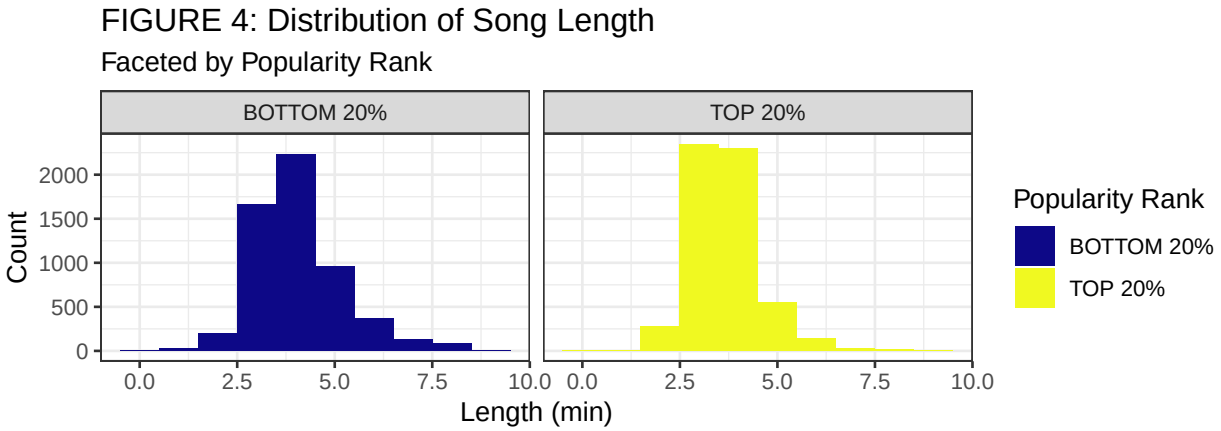
As seen in Figure Two, the bottom 20% of songs have a similar proportion of major and minor songs as the top 20% of songs. However, with calculations, we determined that the proportion of songs in the major key in the top 20% is 58.86% and the proportion of songs in the major key in the bottom 20% is 55.99%, which means the difference in the means is 2.87%.

Speechiness



As seen in Figure Three, the bottom 20% of songs have a similar distribution of speechiness scores as the top 20% of songs. However, with calculations, we determined that the mean speechiness of songs in the top 20% is 0.106 and the mean danceability of songs in the bottom 20% is 0.107, which means the difference in the means is -0.001.

Song Duration



As seen in Figure Four, the bottom 20% of songs have an only slightly different distribution of durations in minutes as the top 20% of songs. However, with calculations, we determined that the mean duration of songs in the top 20% is 3.66 minutes and the mean duration of songs in the bottom 20% is 4.06 minutes, which means the difference in the means is -0.39 minutes.

Analytical Methods

In order to understand which attributes are significantly influential on whether a song is popular or not, we will use Hypothesis Testing. Hypothesis testing will allow us to determine if the data we observed was by chance. It is important to determine if our findings are reflective of the larger population of Spotify songs rather than coincidental.

All of the following p-values will be compared with a 0.05 alpha level and all null distributions calculated with a seed of 1 and 10,000 reps.

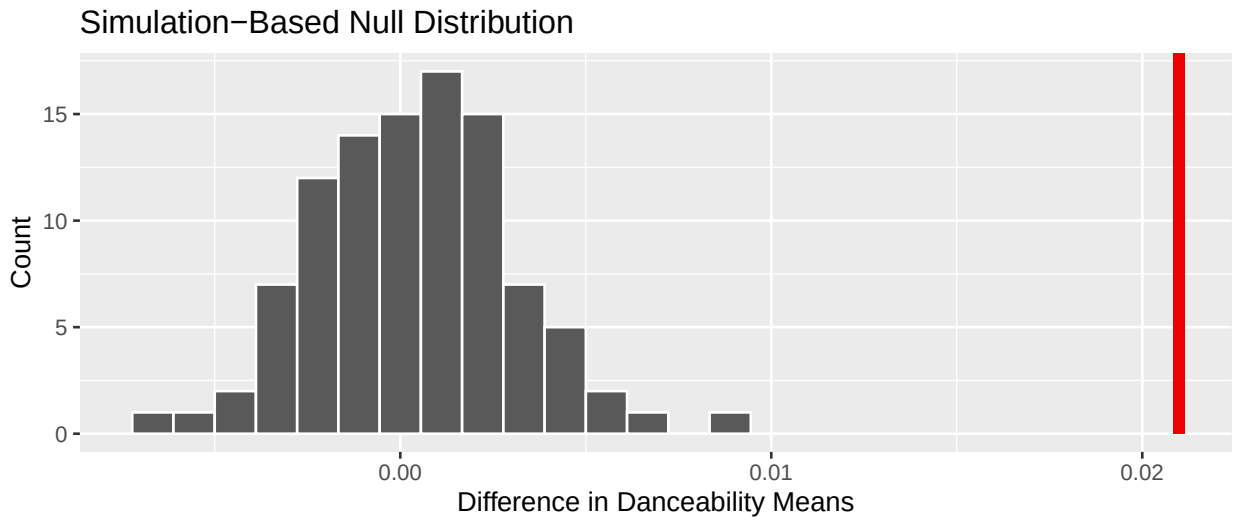
Research Question 1: Exploring danceability trends in Spotify songs.

Null Hypothesis: The mean danceability of more popular songs is not significantly larger than the mean danceability of less popular songs.

Alernative Hypothesis: The mean danceability of more popular songs is significantly larger than the mean danceability of less popular songs.

Let μ_M be the mean danceability of more popular songs and μ_L be the mean danceability of less popular songs.

$$H_0 : \mu_M - \mu_L = 0$$
$$H_a : \mu_M - \mu_L > 0$$



This resulted in a p-value of 0.

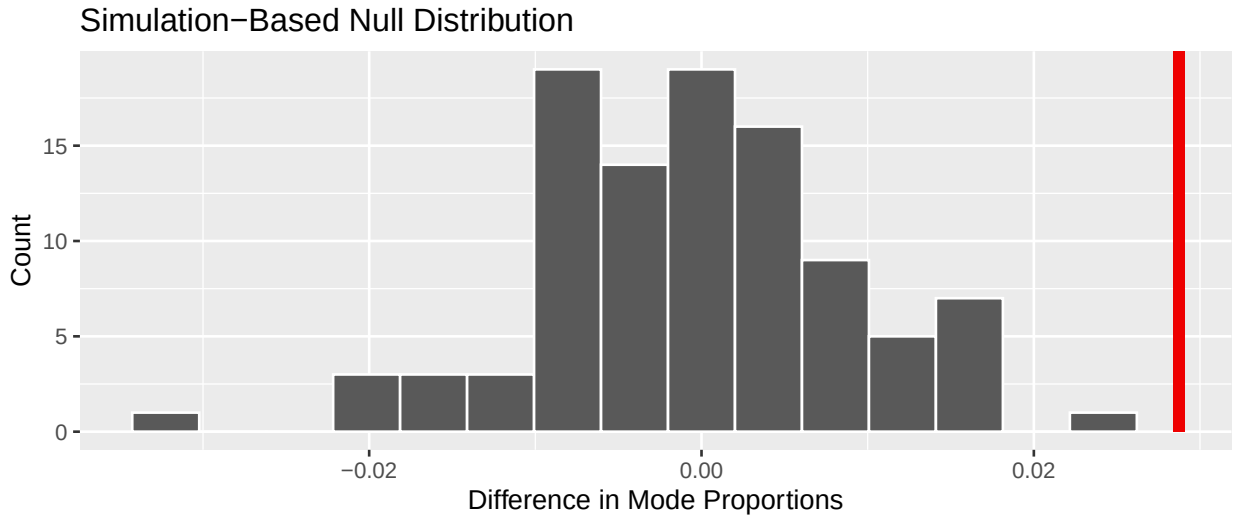
Research Question 2: Exploring mode trends in Spotify songs.

Null Hypothesis: The proportion of major songs for more popular songs is not significantly larger than the proportion of major songs for less popular songs.

Alernative Hypothesis: The proportion of major songs for more popular songs is significantly larger than the proportion of major songs for less popular songs.

Let μ_M be the proportion of major songs for more popular songs and μ_L be the proportion of major songs for less popular songs.

$$H_0 : \mu_M - \mu_L = 0$$
$$H_a : \mu_M - \mu_L > 0$$



This resulted in a p-value of 0.

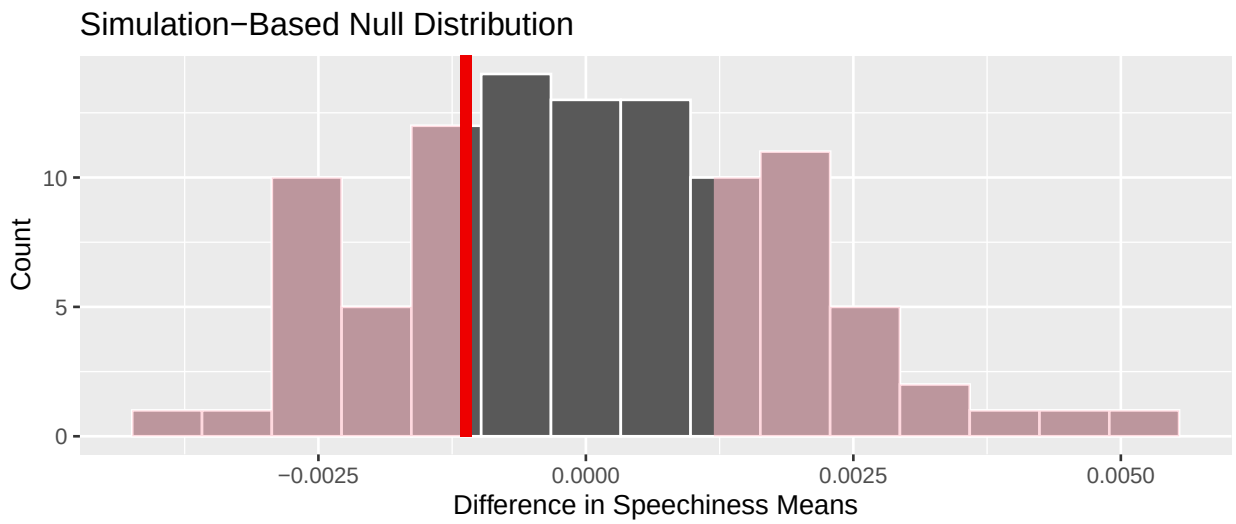
Research Question 3: Exploring speechiness trends in Spotify songs.

Null Hypothesis: The mean speechiness of more popular songs is not significantly different than the mean speechiness of less popular songs.

Alernative Hypothesis: The mean speechiness of more popular songs is significantly different than the mean speechiness of less popular songs.

Let μ_M be the mean speechiness of more popular songs and μ_L be the mean speechiness of less popular songs.

$$H_0 : \mu_M - \mu_L = 0$$
$$H_a : \mu_M - \mu_L \neq 0$$



This resulted in a p-value of 0.54.

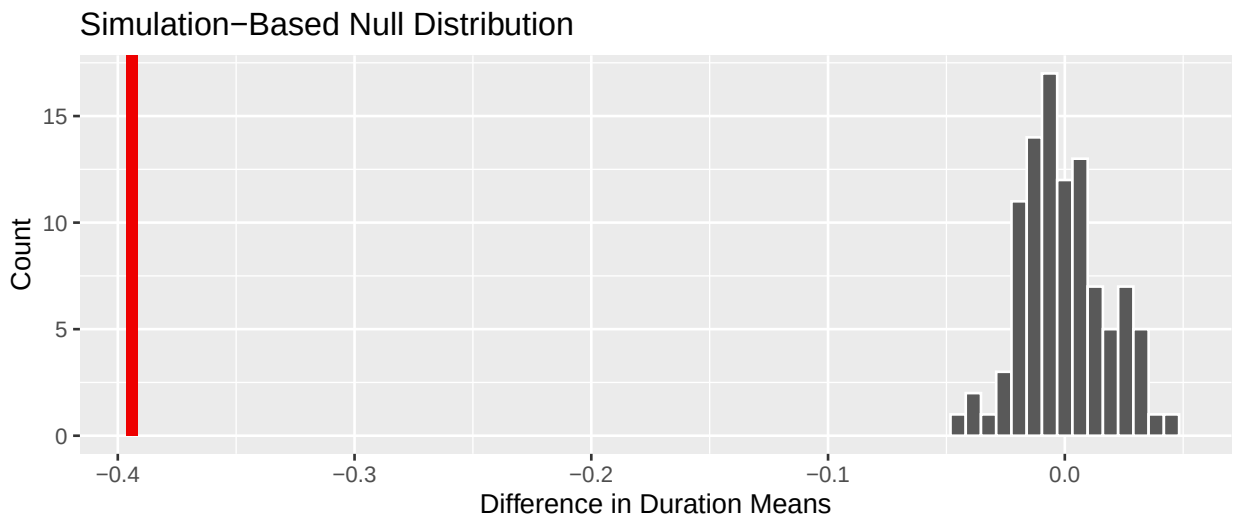
Research Question 4: Exploring duration trends in Spotify songs.

Null Hypothesis: The mean duration of more popular songs is not significantly lower than the mean duration of less popular songs.

Alernative Hypothesis: The mean duration of more popular songs is significantly lower than the mean duration of less popular songs.

Let μ_M be the mean duration of more popular songs and μ_L be the mean duration of less popular songs.

$$H_0 : \mu_M - \mu_L = 0$$
$$H_a : \mu_M - \mu_L < 0$$



This resulted in a p-value of 0.

Results

Research Question 1

The p-value of our hypothesis test was 0, which means that we can reject the null hypothesis. We have enough evidence to suggest that the difference in the danceability means of the most popular songs and the least popular songs is statistically significant, and that more popular songs have a higher danceability than less popular songs. However, the difference in means we calculated was 0.021, meaning that the danceability of more popular songs was only 0.021 higher than the danceability of less popular songs. While that difference may be statistically significant, it such a small difference that it does not give much guidance to artists. It suggests that a slight increase in danceability correlates with higher probability, but the increase is so slight that it would be hard for artists to use this information, as danceability is a hard quality for artists to calculate themselves.

Research Question 2

The p-value of our hypothesis test was 0, which means that we can reject the null hypothesis. We have enough evidence to suggest that popular songs have a larger proportion of songs in a major key than the proportion that less popular songs have. This suggests that more of the popular songs are in the major key. However, the difference in proportion that we calculated was only 2.87%, which suggests that the proportion of major songs for popular songs is only slightly higher than the proportion of major songs for less popular songs. This means that we cannot say that creating a song in a major key correlates with a much greater chance of being popular.

Research Question 3

The p-value of our hypothesis test was 0.54, which is much higher than our significance level of 0.05. This means we do not have evidence to reject the null hypothesis. Thus, we can only say that more popular songs do not have a difference speechiness level than less popular songs, which suggests that increasing or decreasing speechiness would not decrease the likelyhood of a song being popular.

Research Question 4

The p-value our hypothesis test was 0, which means we can reject our null hypothesis. We have enough evidence to suggest that more popular songs are shorter than less popular songs, which may suggest that

creating a shorter song may give it a better chance of being popular.

Discussion

Conclusions

Limitations

Further Research

References

- https://github.com/rfordatascience/tidytuesday/tree/master/data/2020/2020-01-21
- https://newsroom.spotify.com/company-info/#:~:text=Discover%2C%20manage%20and%20share%20over,ad%2Dfree%20music%20listening%20experience.
- https://open.spotify.com/playlist/37i9dQZF1DX7iB3RCnBnN4
- https://loudandclear.byspotify.com/?question=a-million-streams


```
spotify_data_clean %>%
  summarize(mean(speechiness)) %>%
  print()
```

```
## # A tibble: 1 x 1
##   `mean(speechiness)`
##                 <dbl>
## 1               0.108
```

```
spotify_top %>%
  slice(1:1000) %>%
  summarize(mean(speechiness)) %>%
  print()
```

```
## # A tibble: 1 x 1
##   `mean(speechiness)`
##                 <dbl>
## 1               0.115
```