

# Analyzing Factors Contributing to the Popularity of Songs by The Strokes

By analyzing explicitness in lyrics and track length duration

Divya Gupta

November 21, 2024

## Abstract

This paper aims to examine and understand the formula behind the popularity of certain songs by The Strokes. This is done by examining the relationship between track popularity, lyrical explicitness, and track length duration using Spotify's data available through its API. This analysis reveals that explicitness is not a strong indicator for song popularity, illustrated by a 28% presence of explicit lyrics in The Strokes' top 50 most popular tracks. On the other hand, there is a positive correlation between the track length and track popularity. These findings are further illustrated via visual graphs.

## 1 Introduction

The Strokes are a critically acclaimed rock band that were formed in New York in 1998. Since the 90s, they have steadily gained popularity, with thought-invoking lyrics centering politics and philosophy - explored through globally-acclaimed songs like "Ize of the World", "New York City Cops" and "Last Nite", throughout the 2000s. Spotify provides valuable metrics regarding their global popularity as well as specific information such as specific track popularities. This paper aims to examine the "formula" behind a popular The Strokes song, and examines various variables such as lyrical explicitness and track length to understand the recipe behind a popular song by the band. Inherently, the paper answers two research questions, 1) *"Does lyrical explicitness impact track popularity for The Strokes?"* and 2) *"Is track duration a good predictor for track popularity for The Strokes?"*

Spotify's Web API was used to analyze the top 50 most popular tracks by The Strokes. Further data was extrapolated regarding lyrical explicitness across their top 50 songs and its potential influence on track popularity. Since The Strokes are known for their brave lyricism as they combat politically-charged and emotional topics in their songs, lyrical explicitness can be a relevant predictor for track popularity among their audience. Further, we examined track length and its relation to track popularity. In a time of reduced attention spans and an expansive access to music through Spotify or Apple Music, there is a race to make songs popular by tapping into the formula of shorter but catchier tunes [5]. It is worth examining whether this formula stands true and works for The Strokes. This paper is thus structured as follows: Section 2 explores the data set, measurement and methodology, Section 3 visually displays the results, and Section 4 discusses implications of our results and concludes the paper.

## 2 Data

The dataset was procured using Spotify's Web API in tandem with the "spotifyr" package in R. This gave us access to variables such as:

Name: The name of the song

Popularity: A score (between 0-100) calculated via user engagement metrics on Spotify, illustrating the popularity of a certain track. Here, 100 is the most popular track while 0 shows low popularity.

Explicit: A binary variable (TRUE or FALSE) indicating whether the song is explicit.

Duration\_ms: The length of the track in milliseconds.

The data was organized to extract The Strokes' top 50 most popular tracks on Spotify. Then, they were analyzed for the distribution between explicit and non-explicit tracks as well as the distribution of popularity across tracks across various track lengths. These analyses are further explored in the Results section. All scripts are available on GitHub.

### 2.1 Measurement

There were three variables being contested during this analysis - popularity, explicitness, and duration.

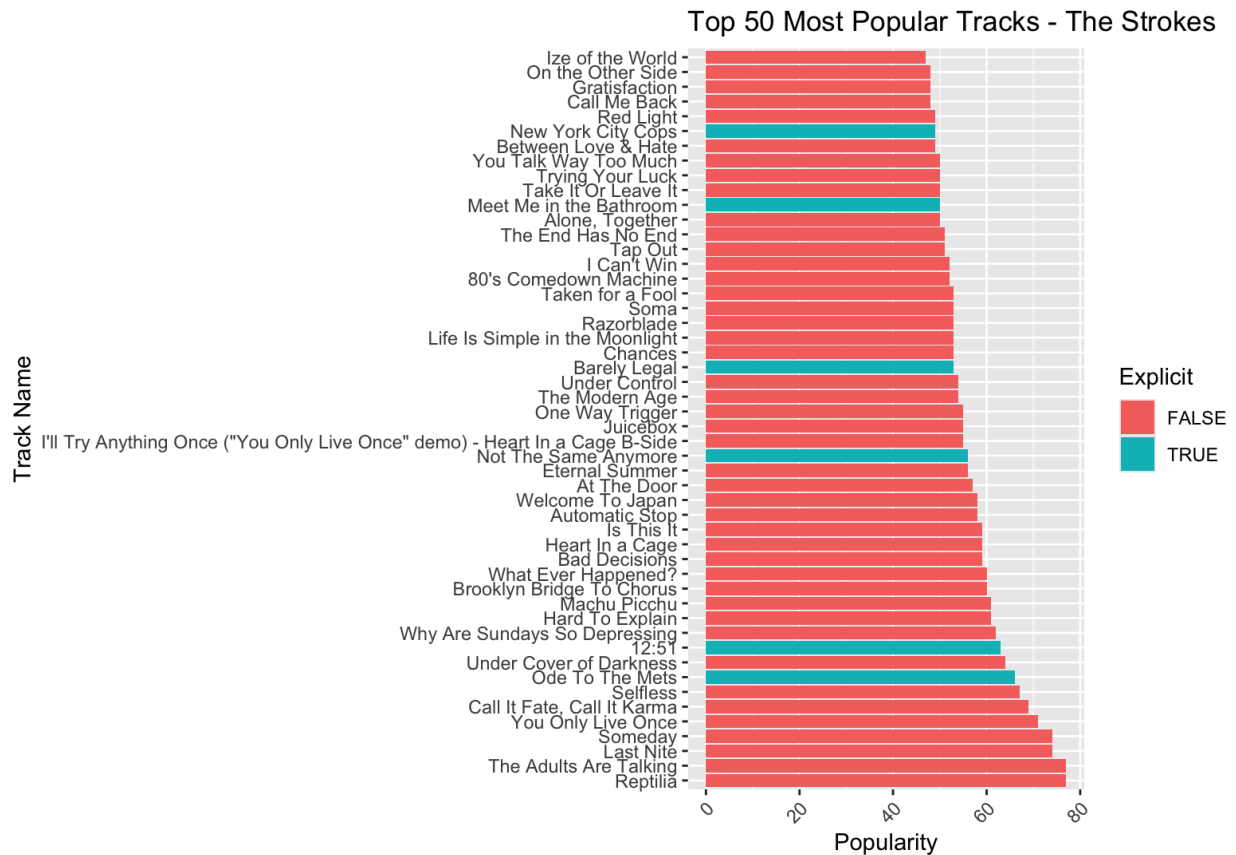
Popularity is defined by Spotify using track play counts (where a play counts if the track is played for at least one minute). This metric can be used to assess user interest in certain tracks.

Explicitness is tracked through the Spotify variable "explicit" which labels a track TRUE if it contains strong language, mature content, or inappropriate words.

Duration of a track is measured by Spotify in milliseconds. For this analysis, it was converted into seconds for simplicity as all The Strokes songs are at least one minute in length.

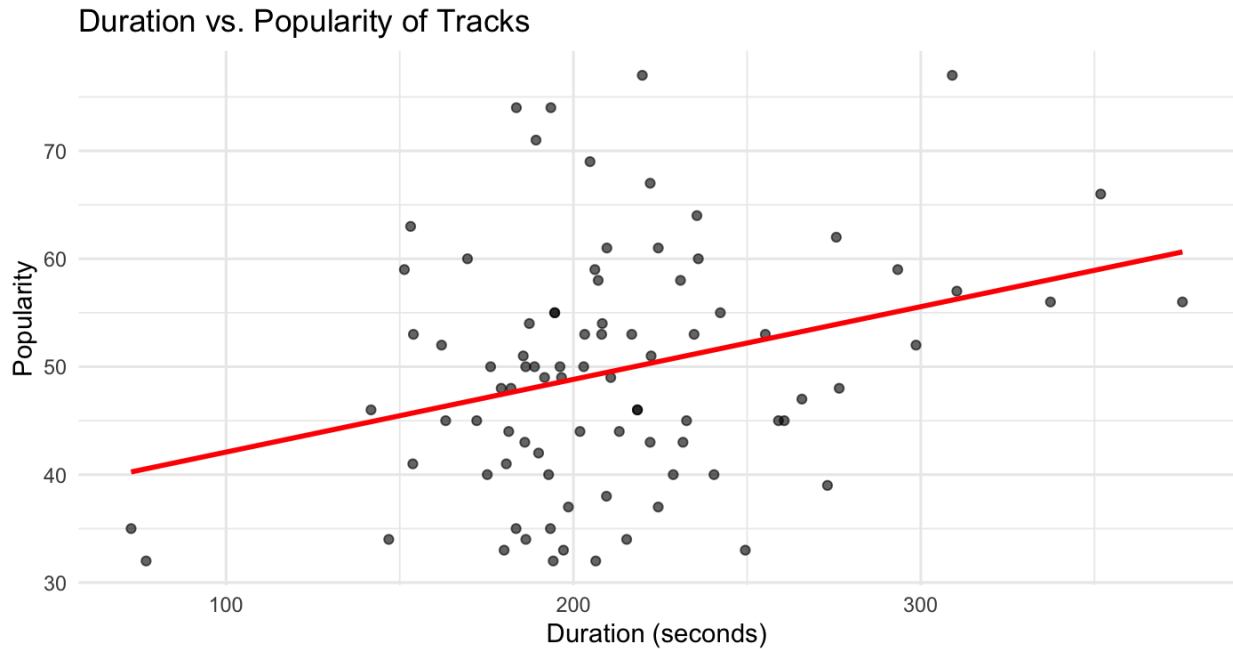
## 3 Results

In the top 50 most popular tracks, it was found that 28% of them are explicit. Figure 1 illustrates the distribution of tracks by popularity along with their status as explicit or non-explicit. It can be seen that there is no strong correlation indicating that explicitness is a predictor for song popularity. Extremely popular songs such as "Reptilia" have no explicit lyrics, whereas other songs such as "12:51" and "Ode to the Mets" contain explicit lyrics and still chart very highly, indicating their popularity.



*Figure 1: Distribution of Popularity Across Explicit and Non-Explicit Tracks*

Figure 2 shows a scatterplot of the track duration versus track popularity to determine if the length of the track has an influence on how popular it is, ie, how much the users engage with it. Here, we can see that there is a slight positive correlation between longer tracks and track popularity, illustrating that longer tracks are overall favored by listeners of The Strokes. The red line highlights the best fit relationship between these two metrics and shows an upward tick as the length of the track increases.



*Figure 2: Duration vs. Popularity of Tracks*

## 4 Discussion

This analysis indicates that explicitness is not necessarily a predictor for song popularity as The Strokes' top 50 songs were evenly distributed between explicit and non-explicit songs. This suggests that there is something else in the lyricism apart from the explicitness that drives The Strokes' musical success. Secondly, contrary to current trends that aim to make shorter songs to boost popularity ratings and retain user engagement, The Strokes' top 50 songs have a slight skewness for longer songs. Further, their longer songs are outperforming shorter songs - illustrating an overall value in track content (such as musical organization and lyricism) that retains a listener's interest for longer periods of time.

These findings suggest that the strength of The Strokes' music lies not in their ability to write explicit lyrics, but in their overall messaging and relatability. A user's interest in longer track can be indicative of their interest in the content and messaging of the band's music.

However, while these results are useful, it is of note that they are extracted from only one data source and cannot account for the totality of The Strokes' fanbase. A greater analysis needs to be conducted, pooling various data sources such as Apple Music, Youtube, online forums, and vinyl and record sales to understand the complete nature of user interest in The Strokes' tracks.

## References

1. Thompson, Charlie, et al. 2022. \*spotifyr: R Wrapper for the Spotify Web API\*. <https://CRAN.R-project.org/package=spotifyr>.
2. Wickham, Hadley, et al. 2019. "Welcome to the Tidyverse." \*Journal of Open Source Software\* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.
3. R Core Team. 2023. \*R: A Language and Environment for Statistical Computing\*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
4. Spotify. 2023. "Spotify Developer Documentation." <https://developer.spotify.com/documentation/>.
5. Giraldo, 2020. "Influence of platforms in a digitalized world: a focus on the music industry." Università Ca' Foscari Venezia. <http://hdl.handle.net/10579/18084>