

Assessing the Relationship Between Music-Sharing Tweet Popularity and Long-Term Spotify Streaming Popularity

Elizabeth Kiel '23 | Data Science Major Capstone

Research Question

To what extent can the popularity metrics of a tweet containing a music-sharing link predict a song's future popularity?

Background

Musical artists are increasingly attempting to understand and leverage social media to promote their music to wider audiences [3][4]. A better understanding of social media metrics in relation to long-term streaming popularity may improve strategic promotional efforts.

Twitter may be considered the nexus of the online music community, given the immense volume of music-related content on the platform [1]. Spotify is the most popular subscription-based streaming service in the world with over 350 million users [2].

Data

Description

To operationalize music sharing on social media, Professor Mustafaraj and I used the Twitter API to collect ~4 million tweets containing the the Spotify track sharing URL stem ("open.spotify.com") in March 2022. The Twitter API also provided the popularity metrics selected for the final data set:

- **Tweet popularity metrics:** like_count, retweet_count, reply_count, quote_count
- **Tweet author popularity metrics:** # following, # of followers, verified status

To measure long-term popularity, I took a random sample of ~150,000 tracks shared in the tweets in February and March 2023. Popularity scores were collected by passing the unique track or album ID to the Spotify API.

Cleaning

The following variables were created or modified:

- follow_ratio: # following/# followers
- frequency: # of times track/album appeared in sample
- verified: re-coded from Boolean to binary value
- popularity: tracks/albums with a popularity score of 0 were excluded from the datas set to better meet normality assumption of the response variable (Figure I)

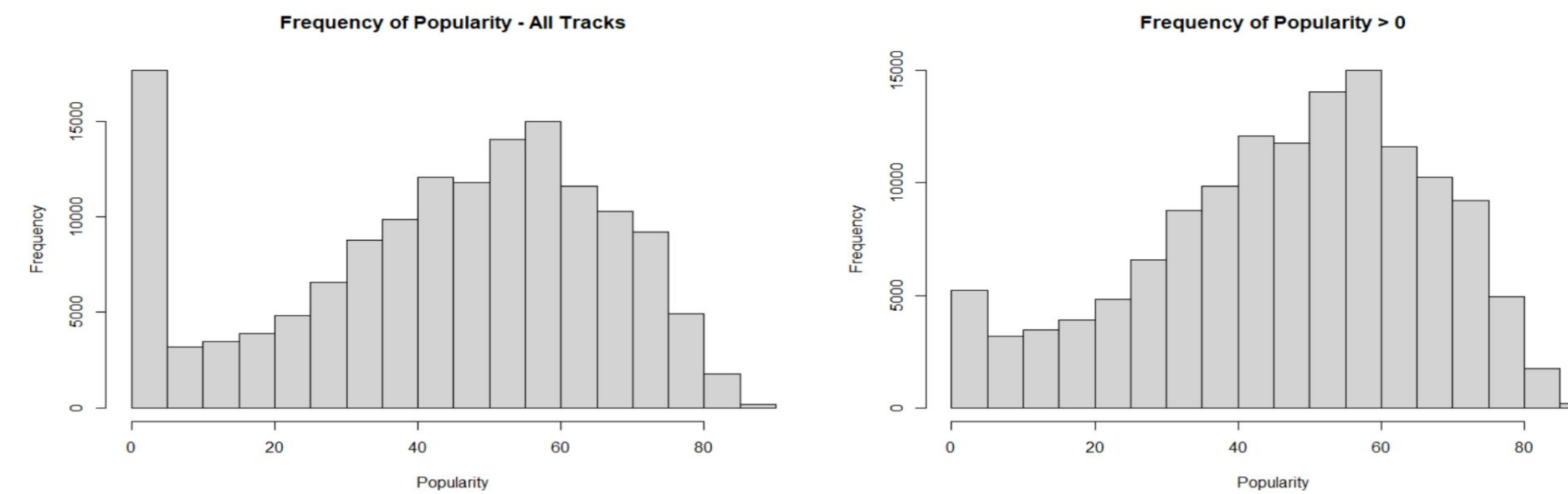


Figure I. Popularity score distribution before (left) and after (right) scores equal to 0 were eliminated

Data Modeling: First Order Model

First using the song data, I ran a multiple linear regression in R to identify the Twitter popularity metrics associated with long-term Spotify popularity score. To select model variables, I performed a stepwise elimination on the full model to minimize AIC.

Using the linear regression diagnostic plots generated by R, I eliminated influential outlier points identified in the regression diagnostic model plots. Utilizing the MASS package in R, I performed a Box-Cox transformation on the response variable, $\lambda = 1.1515$. I refit the data, which yielded a first-order model ($R^2 = 0.03031$, $R^2_{adj} = 0.03025$, $F = 533.8$, $p < 2.2 \cdot 10^{-16}$) of the form:

$$\hat{Popularity}_i = 72.95 + 0.0469 \cdot Frequency_i + 0.1028 \cdot ReplyCount_i - .0318 \cdot RetweetCount_i + .9373 \cdot QuoteCount_i - .0004 \cdot Following_i + .000008 \cdot Followers_i - 19.37 \cdot Verified_i - .3000 \cdot FollowRatio_i$$

	Estimate	SE	t-value	p
Intercept	72.95	0.103	708.215	0*
Frequency	0.04685	0.0007377	63.51	0*
Reply Count	0.1028	0.04137	2.486	0.01294*
Retweet Count	-0.03178	0.01962	-1.619	0.10541
Quote Count	0.9373	0.1787	5.245	0*
Following	-0.0004262	0.00003925	-10.859	0*
Followers	0.000008366	0.000005909	1.416	0.15683
Verified	-19.37	2.736	-7.079	0*
Follow Ratio	-0.02997	0.009877	-3.035	0.00241*

* $p\text{-value} < \alpha = 0.05$

Figure II. Regression coefficient output for first-order model

Data Modeling: Interaction Model

To select model variables for an interaction model, I performed a stepwise process on the first-order model bounded by the second-order model with all interactions to minimize the AIC. This yielded a second-order model that included all 7 variables from the first-order model, as well as 22 interaction terms ($R^2 = 0.03145$, $R^2_{adj} = 0.03129$, $F = 201.7$, $p < 2.2 \cdot 10^{-16}$).

Discussion

- The first-order model describes a positive relationship between popularity of a track and frequency of appearance in the data set, reply count, quote count, and # of followers. It describes a negative relationship between popularity and retweet count, # following, verified status, and follow ratio.
- Like count was excluded as a variable from the first-order model during variable screening
- Overall, the models suggest that the Twitter popularity metrics do not account for much of the variation in Spotify popularity one year after the Tweet

Limitations & Data Ethics

When interpreting these results, it is important to consider that:

- Approximately 8% (n = 12,432) data points were excluded from analysis due to popularity scores < 0

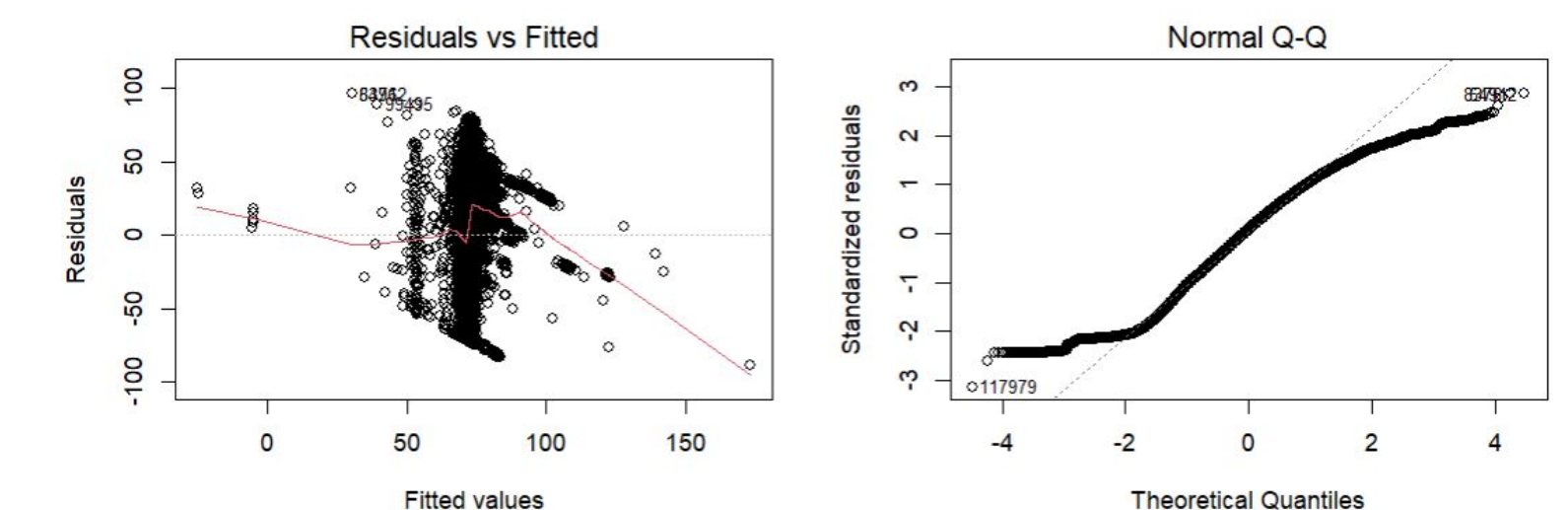


Figure III. Linear regression assumptions for the first-order model were not met

References

- [1] Bengtsson, Linda Ryan, and Jessica Edlom. "Mapping transmedia marketing in the music industry: a methodology." Media and Communication 9.3 (2021): 164-174.
- [2] "Music Streaming App Revenue and Usage Statistics (2023)." Business of Apps, 1 Feb. 2023, <https://www.businessofapps.com/data/music-streaming-market/>.
- [3] Salo, Jari, Mikko Lankinen, and Matti Mäntymäki. "The use of social media for artist marketing: Music industry perspectives and consumer motivations." International Journal on Media Management 15.1 (2013): 23-41.
- [4] Toscher, Benjamin. "Resource Integration, Value Co-Creation, and Service-dominant Logic in Music Marketing: The Case of the TikTok Platform." International Journal of Music Business Research 10.1 (2021): 33-50.