Do Musicians learn? Evidence from Song Popularity

Dr Preet Deep Singh

09/11/2021

Abstract

Using a dataset of 5,86,672 songs from Spotify, we evaluate whether artists learn from the response their songs get or not. We find that popularity of songs improves with iterations. We plot whether artists have their biggest hits in their first year or subsequent years.

1 Introduction

In a series of papers, I try to examine the following questions . # What does the average album look like: how many sad, what balance # Do albums with similar songs do better # Do albums have multiple hit songs or only 1/2 songs make it ? # Do artists experiment after a hit or do more of the same? Are they scare or overconfident? # Do negative songs do better? # Do winter songs do better? # When does a artist shine to fame: first or next or last? # Is the same song a hit in multiple places? What are unique geographic hits?

In this paper, I look at the most popular hits of artists. I examine whether artists get their biggest hits in the first go and spend the rest of their lives trying to match that success or whether artists enter the market, release music, take feedback and improve in order to produce more popular songs. In case artists are learning from feedback, the popularity of their songs would have an upward trend.

In case artists do not take feedback, their most popular songs would be spread randomly across their careers. For artists who had their most popular song in their first year, the subsequent years would entail a lot of experiments, most of which would fail and eventually they would retire. This means we should see artists with their best hits coming in the first year, dropping off in a couple of years post that. The year/date of the last song by a particular band can be used to check this hypothesis.

To examine learning we regress popularity of the songs on their sequence. We find that sequence has a positive coefficient that is statistically significant while

controlling for other song traits.

Our data spans a total of 586672 songs by 115062 across 102 years. This includes 25864 explicit songs.

2 Data

2.1 Summary

- Total songs
- Total Artists
- Total Albums
- Average Popularity
- Year wise distribution

Table 1: This table shows the summary of the dataset.

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
popularity	586,672	27.570	18.371	0	13	41	100
duration_ms	586,672	230,051.200	126,526.100	3,344	175,093	263,867	5,621,218
explicit	586,672	0.044	0.205	0	0	0	1
danceability	586,672	0.564	0.166	0.000	0.453	0.686	0.991
energy	586,672	0.542	0.252	0.000	0.343	0.748	1.000
key	586,672	5.222	3.519	0	2	8	11
loudness	586,672	−10.206	5.089	−60.000	−12.891	−6.482	5.376
mode	586,672	0.659	0.474	0	0	1	1
speechiness	586,672	0.105	0.180	0.000	0.034	0.076	0.971
acousticness	586,672	0.450	0.349	0.000	0.097	0.785	0.996
instrumentalness	586,672	0.113	0.267	0.000	0.000	0.010	1.000
liveness	586,672	0.214	0.184	0.000	0.098	0.278	1.000
valence	586,672	0.552	0.258	0.000	0.346	0.769	1.000
tempo	586,672	118.465	29.764	0.000	95.600	136.321	246.381
time_signature	586,672	3.873	0.473	0	4	4	5

Figure **??** shows the date wise distribution of songs in our dataset. Figure 1 shows the year wise distribution of songs in our dataset. ##IF this is correct, use the same label command in the previous one.

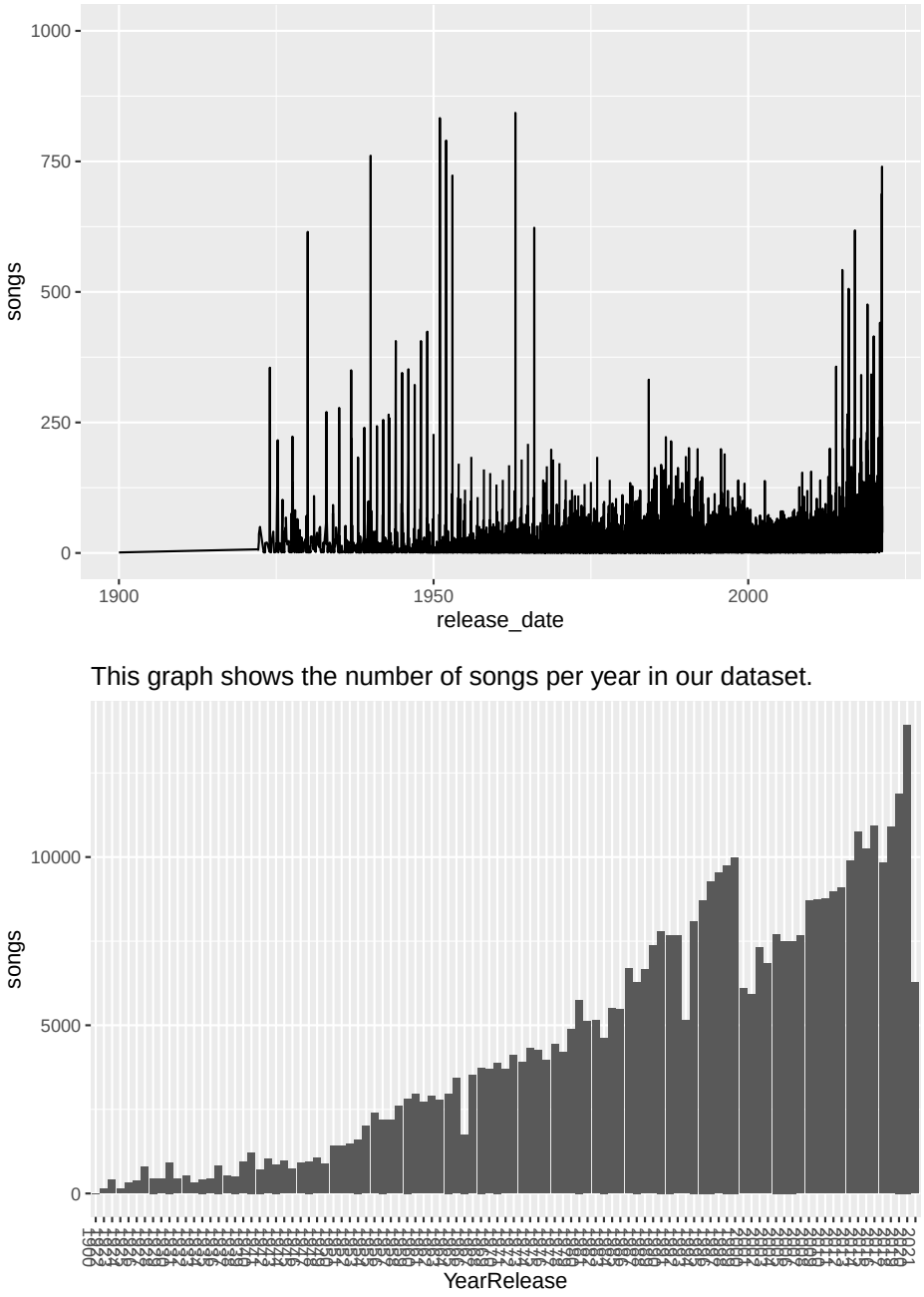

Figure 1: This graph shows the year wise distribution of songs.

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu % Date and time: Sun, Nov 07, 2021 - 23:53:07

Table 2: This table shows the summary of the dates of first song, biggest hit and last song per artist.

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
X1	97,814	58,911.710	32,781.820	1	30,131.2	87,291.8	115,062
Year1	97,814	1,999.853	20.686	1,900	1,991	2,016	2,021
Date1	97,814	11,055.280	7,584.877	−25,567	7,773.2	16,906	18,733
LastYear	97,814	2,002.534	19.832	1,922	1,996	2,017	2,021
GYear	97,814	2,018.000	0.000	2,018	2,018	2,018	2,018
HitYearDiff	97,814	18.147	20.686	−3	2	27	118
LastHitYearDiff	97,814	−15.466	19.832	−96	−22	−1	3

#.....
We look at the year of the first song by the artist and the year of the most popular song by the artist. For example, if the artist released the first song in 1997, and their most popular song was released in 2001, then we note 2001-1997 = 4 as the time till the first hit. In case the most popular song was released in the same year, this number would be 0.

This analysis allows us to answer the following questions # Who got it soonest # How many got it soonest # How many in 1 year # Who got it late # How many after 5 years

2.2 Regression

Sequence of songs Only the most popular song on any date, incase there were multiple songs on the same date Only artists with more than 1 song Get details of the music Regression on Popularity on All the factors Find that sequence is a major contributor along with Explicit

3 Limitations

Limited dataset: It is possible that there are some songs by certain artists that are not part of the half a million songs I use. In some cases, the biggest hit of a few artists might have existed in a particular year and is not accounted for in this study. However, our dataset is not biased in terms of vintage or recency. Therefore, I do not expect this to be systematically lopsided. This data is from Spotify. The popularity of the app has gone up in each of the previous few years. It is possible that songs that were not hits in the first go, in the first geography subsequently became hits. This would not be accounted for in our study.

Table 3: This table shows the results of regression of popularity on sequence and other control variables.

	Dependent variable:	
	popularity.x	
	(1)	(2)
seq	0.010***	0.009***
	(0.001)	(0.001)
duration_ms.x	−0.00000***	0.00000**
	(0.00000)	(0.00000)
danceability	0.177	3.512***
	(0.150)	(0.173)
energy	1.721***	14.882***
	(0.157)	(0.178)
key	−0.033***	0.006
	(0.004)	(0.005)
loudness	0.226***	0.569***
	(0.006)	(0.007)
mode	0.512***	0.136***
	(0.033)	(0.038)
speechiness	−4.433***	−19.589***
	(0.085)	(0.088)
acousticness	0.126*	−9.101***
	(0.074)	(0.082)
instrumentalness	−7.788***	−18.410***
	(0.074)	(0.080)
valence	−0.486***	−11.589***
	(0.096)	(0.106)
tempo	−0.003***	−0.006***
	(0.0005)	(0.001)
time_signature	0.145***	0.145***
	(0.016)	(0.019)
explicit.x	8.430***	13.420***
	(0.115)	(0.132)
Constant	−778.496***	38.088***
	(2.013)	(0.234)
Year	Yes	No
Observations	494,991	494,991
R^2	0.732	0.641
Adjusted R^2	0.732	0.641
Residual Std. Error	9.733 (df = 494975)	11.250 (df = 494976)
F Statistic	89,928.680*** (df = 15; 494975)	63,232.470*** (df = 14; 494976)
Note:		*p<0.1; **p<0.05; ***p<0.01

4 TODO

1. The year differences dont seem to be correct.