Do Musicians learn? Very slowly

Dr Preet Deep Singh

09/11/2021

Abstract

Using a dataset of 5,86,672 songs from Spotify, we evaluate whether artists are gifted beings who stick to their core creativity or learn and adapt basis their the response their songs get. We find that mean popularity of songs has improved over time but subsequent songs perform worse on averagefor the same artist.

1 Introduction

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 negative coefficient that is statistically significant while controlling for other song traits. This implies that either artists are not able to change as per the needs of the market or there is lack of clarity in what people want. It is believed that artists are, by nature, creative. They create what they feel is good rather than with a commercial mindset. This theory has been questioned and it would be wrong to assume that all artists follow this sort of a pure approach with no regard to fame and fortune. Further, if artists could learn then every artist would continue to produce hits, each better than the last one. In fact, a new artist could simply learn from a database of songs and their popularity. This would mean that every new song would be better than the current best song. This sounds far-fetched.

In a series of papers (Singh [2021b]) , 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?

Some of my other work pertains to perceptions of founders in startups ([Singh, 2021i,d,e,f,j,g]), Olympic medals ([Singh, 2021c,h]), Covid data [Singh, 2020a,b], trading strategy [Singh, 2015], CSR [Singh and Poonawala, 2016], Crowdfunding [Singh and Poonawala, 2021, Singh, 2021a] and some Finance and Director diligence stuff [Singh and Singla, 2016b,a, Singh, 2017] .

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

2 Data

We remove remastered songs from our data for the purpose of analyses.

2.1 Summary

- Total songs
- Average Popularity
- Year wise distribution

Figure 1 shows the date wise distribution of songs in our dataset. Figure 2 shows the year wise distribution of songs in our dataset. ##IF this is correct, use the same label command in the previous one.

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu % Date and time: Mon, Nov 22, 2021 - 23:30:48

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

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
popularity	561,727	27.733	18.364	0	13	41	100
duration_ms	561,727	229,729.200	126,944.300	3,344	175,347	263,427	5,621,218
explicit	561,727	0.046	0.209	0	0	0	1
danceability	561,727	0.566	0.166	0.000	0.456	0.688	0.991
energy	561,727	0.542	0.251	0.000	0.344	0.747	1.000
key	561,727	5.224	3.521	0	2	8	11
loudness	561,727	−10.204	5.098	−60.000	−12.908	−6.464	5.376
mode	561,727	0.657	0.475	0	0	1	1
speechiness	561,727	0.107	0.183	0.000	0.034	0.077	0.971
acousticness	561,727	0.450	0.348	0.000	0.099	0.783	0.996
instrumentalness	561,727	0.111	0.265	0.000	0.000	0.008	1.000
liveness	561,727	0.214	0.184	0.000	0.098	0.278	1.000
valence	561,727	0.552	0.258	0.000	0.346	0.769	1.000
tempo	561,727	118.377	29.794	0.000	95.407	136.286	246.381
time_signature	561,727	3.873	0.475	0	4	4	5

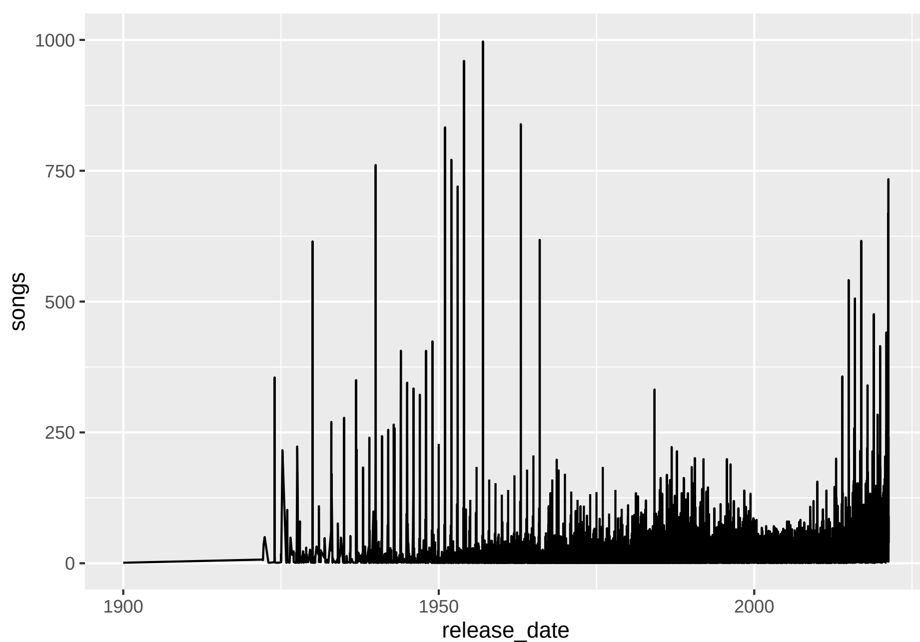

Figure 1: This graph shows the date wise distribution of songs.

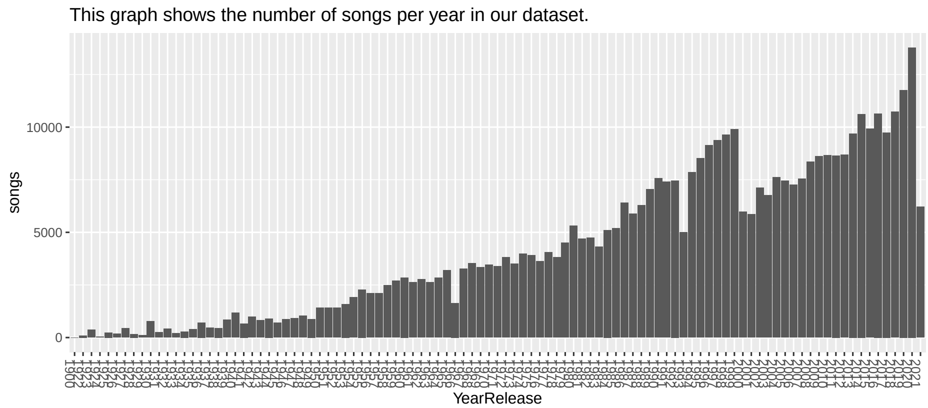

Figure 2: This graph shows the number of songs per year in our dataset

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	86,965	61,652.360	31,079.330	26	34,810	88,158	115,062
Year1	86,965	2,005.602	12.962	1,970	1,997	2,016	2,021
GoldenYear	84,165	2,007.063	12.325	1,970.000	1,999.000	2,017.000	2,021.000
LastYear	86,965	2,007.843	11.948	1,970	2,001	2,017	2,021
GoldenDate2	84,165	2,007.063	12.325	1,970.000	1,999.000	2,017.000	2,021.000
GYear	84,165	2,007.063	12.325	1,970.000	1,999.000	2,017.000	2,021.000
HitYearDiff	84,165	1.110	3.880	0.000	0.000	0.000	47.000
LastHitYearDiff	84,165	0.990	3.620	0.000	0.000	0.000	50.000

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.

Based on the summary of the data, we see that on average the biggest hit of an artist comes about a year after their first song. Surprisingly, the last song of the artist is also almost a year after their biggest hit. This points to the temporary nature of the artist's popularity. These numbers only indicate a typical case, individual cases may vary.

2.2 Visually

I plot popularity of songs over time. In case of learning, I expect to see an upward rising curve. In Figure 3 I plot the popularity of all songs since 1970. Since the plot pertains to songs across time, it is actually plotting collective learning rather than individual learning of each artists. There is a clear upward trend. This could also be because of changing tastes. Music that was produced a decade ago might be closer in feel to that produced recently. We zoom in to this century in Figure 4. We see that there is an overall upward trend since 2013 but it is not the case from 2000 to 2013. If popularity due to learning is a function of Spotify, then we should not be an upward trend between songs from 1970 and 1990.

We look at the average popularity of songs based on their sequence. Figure 5 shows the plot of all songs. Figure 6 shows the plot of all songs except those of artists who have only one song (in the data). This is because a single song would not help us understand whether an artist learnt or not. One of the reasons for only one song could be that the song was so bad that no other record company wanted to invest in that singer again. Some people could have released a song as a 'one-time-activity' rather than a serious professional pursuit. Figure 7 looks at the average popularity of songs by artists that have over 100 songs in the data. These would be people who stayed in the industry for a long time and would have survived owing to popularity [1]. Finally Figure 8 looks at the average popularity of songs by artists that have over 200 songs in the data. There are only two artists who qualify this cut-off.

Spotify launched in 2011 in US [2] whereas the popularity of songs created in those years is not very high.

The general shape of the curve mimics that of a random variable. The hypothesis of learning and continuous improvement ca be rejected.

[1] or a comfortable supply of money.
[2] https://en.wikipedia.org/wiki/Spotify

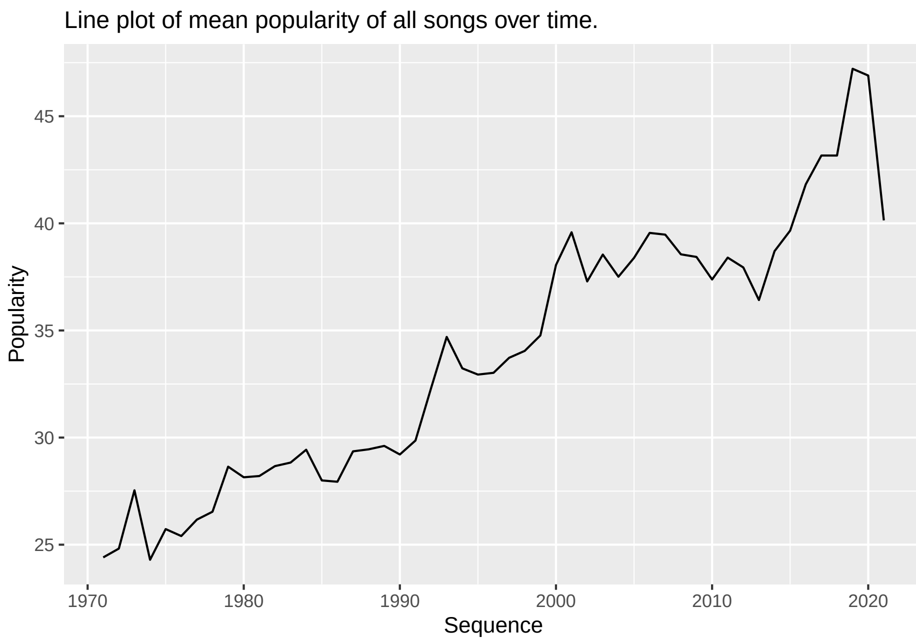

Figure 3: Line plot of mean popularity of all songs

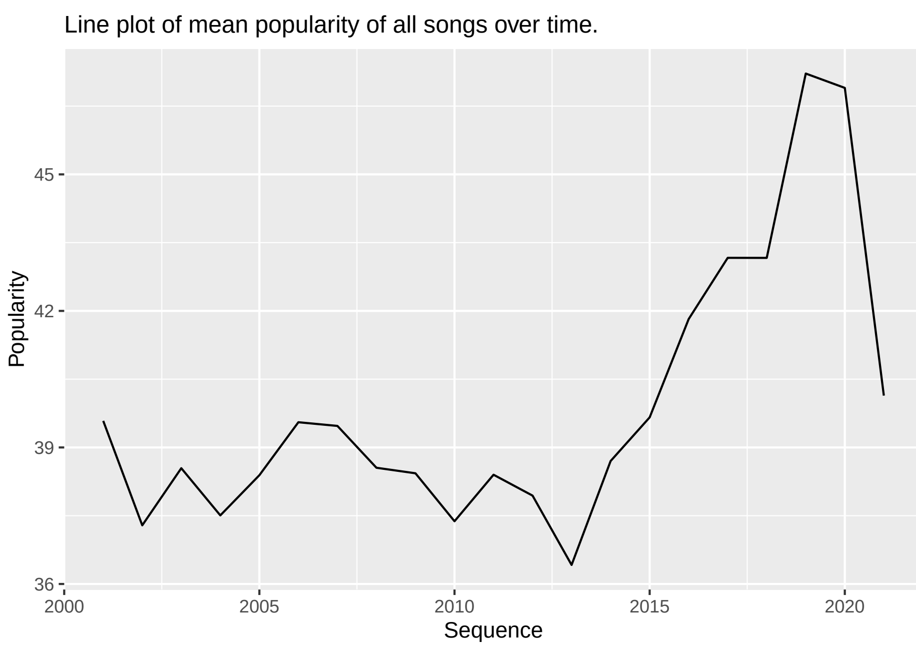

Figure 4: Line plot of mean popularity of all songs

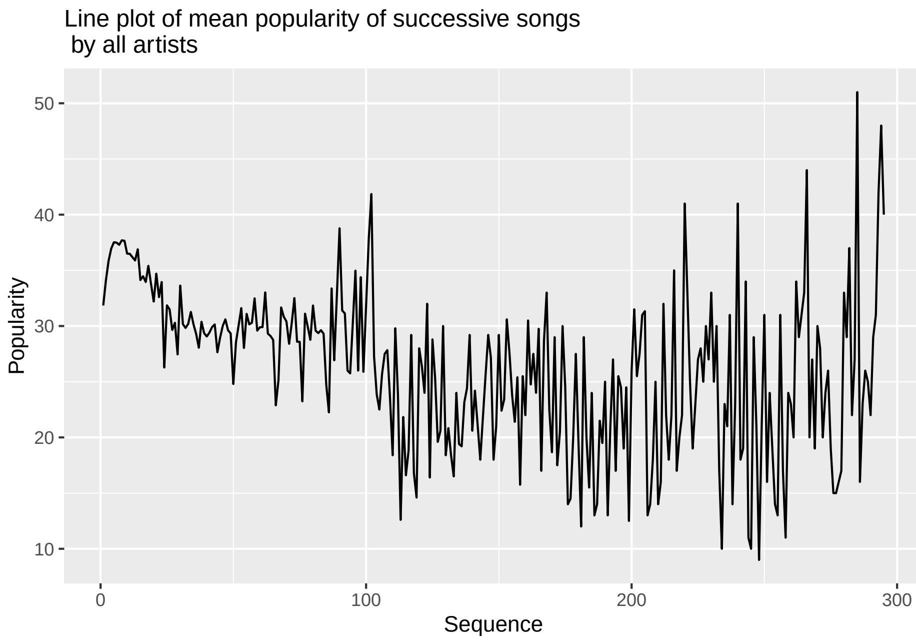

Figure 5: Line plot of mean popularity of successive songs by all artists

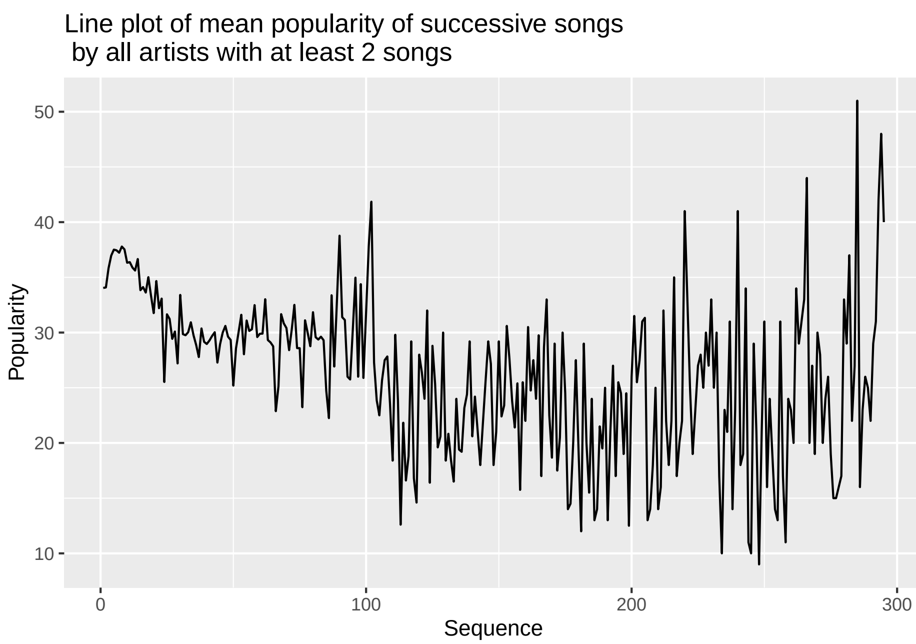

Figure 6: Line plot of mean popularity of successive songs by all artists w/ 1+ songs

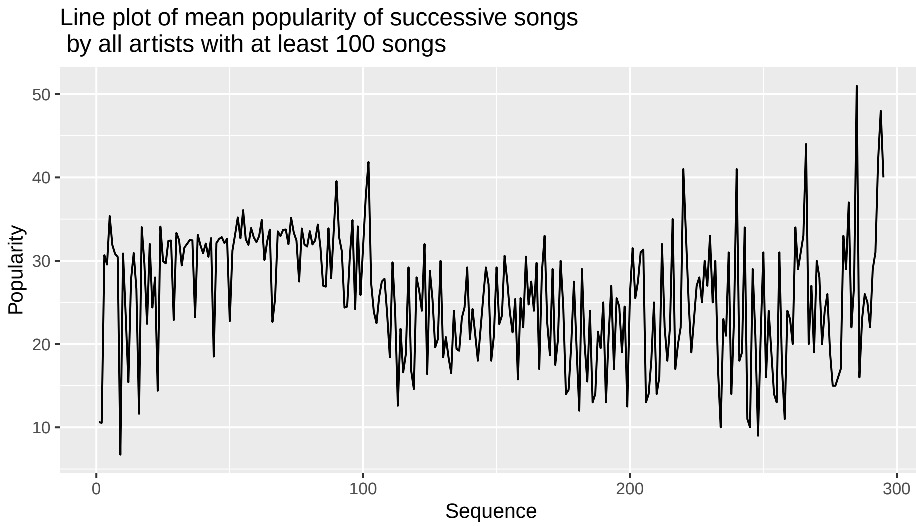

Figure 7: Line plot of mean popularity of successive songs by all artists w/100+ songs

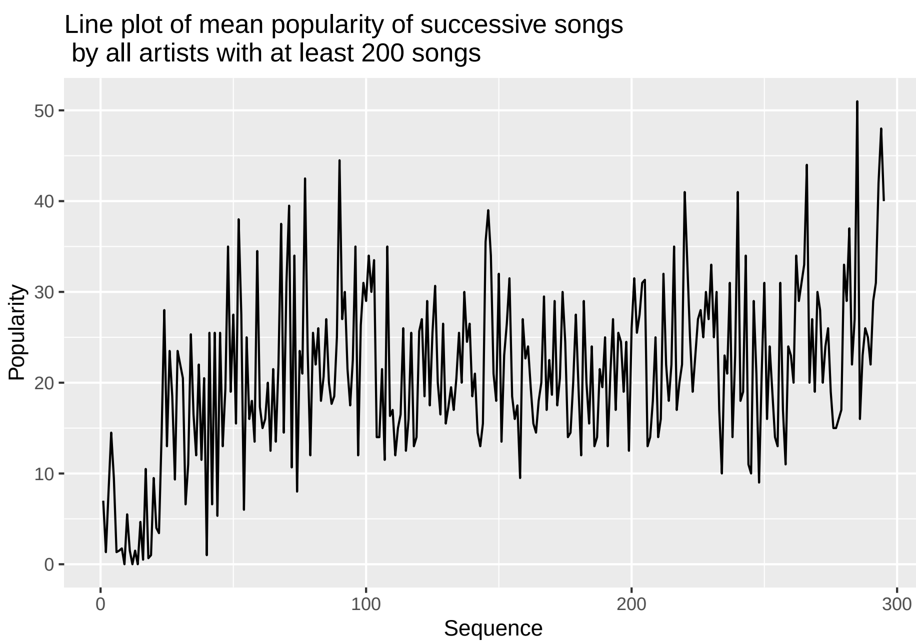

Figure 8: This graph shows the number of songs per year in our dataset

2.3 Regression

I look at the impact of the sequence of song on popularity while controlling for other factors. First, we filter to songs by artists who have at least two songs in our data. This is because we assign a sequence number for each artist. Our data indicates a series of songs, their popularity as well as the sequence in which they were released by the artist. An artist that was more active in the 80s would have their second song in 80s whereas a contemporary artist would have their second song in 2021. For the analysis, we use even time. Therefore, both the second songs are considered at par. This is true for all sequence numbers. The second adjustment we make is that for multiple songs being released on the same day by the same artist. In this case, we only use the value of the most popular song released on the same day. This way we are biasing ourselves towards finding evidence of learning as there would be improvement in taking the maximum instead of the average. We then add details of the songs as provided by Spotify.

Table **??** shows the coefficients on sequence. The first one controls for explicit words in the song and the second one doesn't. However, whether to use explicit words is a matter of choice that contributes to the popluarity of the song. The coefficient is negative in both cases and when controlled for explicit nature, it is statistically significant with a p value of <0.01.

2.4 Drilling Down

In order to examine whether any learning takes place at the start of the career or during the later part, we run part regressions. In Table **??** we show the regression on subsets. The first column contains data for the first 50 songs across artists. The second column shows data for 50-100 songs. The third column includes all data after the first 50 songs and the fourth column only looks at data after 1970. We see that over a longer horizon, and after 1970, the coefficient on sequence is statistically significant. It also shows that the coefficient is negative in both the cases, indicating a negative growth in subsequent production.

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 data is not biased on vintage. 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 the study. While the songs from our study are from almost a century ago, Spotify as an app has only started

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

	Dependent variable:	
	popularity.x	
	(1)	(2)
seq	-0.020^{***}	-0.005
	(0.003)	(0.003)
duration_ms.x	-0.00000^{***}	-0.00000^{***}
	(0.00000)	(0.00000)
danceability	1.644^{***}	19.192^{***}
	(0.359)	(0.378)
energy	-7.196^{***}	-3.021^{***}
	(0.368)	(0.405)
key	-0.012	0.010
	(0.012)	(0.014)
loudness	0.230^{***}	0.741^{***}
	(0.016)	(0.018)
mode	1.033^{***}	0.405^{***}
	(0.091)	(0.100)
speechiness	-2.931^{***}	-9.196^{***}
	(0.281)	(0.303)
acousticness	-3.821^{***}	-11.273^{***}
	(0.196)	(0.211)
instrumentalness	-12.321^{***}	-16.184^{***}
	(0.200)	(0.220)
valence	0.087	-14.306^{***}
	(0.232)	(0.238)
tempo	-0.003^{*}	0.019^{***}
	(0.002)	(0.002)
time_signature	0.267^{***}	0.670^{***}
	(0.102)	(0.112)
explicit.x	7.486^{***}	
	(0.198)	
Constant	-840.374^{***}	43.028^{***}
	(5.557)	(0.660)
Year	Yes	No
Observations	129,541	129,541
R^2	0.351	0.208
Adjusted R^2	0.351	0.208
Residual Std. Error	15.533 (df = 129525)	17.157 (df = 129527)
F Statistic	$4{,}665.166^{***}$ (df = 15; 129525)	$2{,}615.759^{***}$ (df = 13; 129527)
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 4: This table shows the results of regression of popularity on sequence and other control variables

	Dependent variable:			
	popularity.x			
	(1)	(2)	(3)	(4)
seq	0.002	0.005	-0.013^{***}	-0.042^{***}
	(0.007)	(0.015)	(0.004)	(0.003)
duration_ms.x	-0.00000^{***}	0.00000^{***}	0.00001^{***}	-0.00000^{***}
	(0.00000)	(0.00000)	(0.00000)	(0.00000)
danceability	1.663^{***}	10.901^{***}	8.982^{***}	2.725^{***}
	(0.365)	(2.035)	(1.791)	(0.388)
energy	-7.225^{***}	-3.139^{*}	-2.357	-10.382^{***}
	(0.376)	(1.705)	(1.521)	(0.407)
key	-0.012	-0.006	-0.003	-0.013
	(0.013)	(0.053)	(0.048)	(0.013)
loudness	0.225^{***}	-0.118	-0.138^{*}	0.463^{***}
	(0.016)	(0.088)	(0.076)	(0.020)
mode	1.034^{***}	0.363	0.700^{**}	0.955^{***}
	(0.093)	(0.393)	(0.354)	(0.099)
speechiness	-3.754^{***}	5.569^{***}	7.396^{***}	-0.369
	(0.309)	(0.727)	(0.710)	(0.328)
acousticness	-3.921^{***}	-1.881^{**}	-2.431^{***}	-2.924^{***}
	(0.201)	(0.899)	(0.836)	(0.213)
instrumentalness	-12.375^{***}	-8.508^{***}	-7.245^{***}	-13.074^{***}
	(0.203)	(1.484)	(1.307)	(0.254)
valence	0.180	-8.155^{***}	-7.074^{***}	-0.175
	(0.236)	(1.115)	(1.014)	(0.250)
tempo	-0.002	-0.002	-0.003	-0.002
	(0.002)	(0.007)	(0.006)	(0.002)
time_signature	0.211^{**}	0.629^{***}	0.570^{***}	0.073
	(0.107)	(0.225)	(0.213)	(0.122)
explicit.x	7.585^{***}			7.771^{***}
	(0.201)			(0.204)
Constant	-840.631^{***}	-292.962^{***}	-392.815^{***}	-661.565^{***}
	(5.653)	(39.151)	(35.526)	(8.516)
Year	Yes	Yes	Yes	Yes
Observations	126,356	2,611	3,123	113,329
R^2	0.351	0.220	0.272	0.166
Adjusted R^2	0.351	0.216	0.269	0.166
Residual Std. Error	15.642 (df = 126340)	9.423 (df = 2596)	9.300 (df = 3108)	15.851 (df = 113313)
Note:				*p<0.1; **p<0.05; ***p<0.01

recording popularity in the past decade or so. This study assumes that the preference of people has been similar if not same. The study also assumes that the information regarding the popularity of songs was available to the artists. This would not be far-fetched as radio and record sales were always availing on a periodic basis.

References

Dr Preet Deep Singh. Square it off: Trading strategy. *Available at SSRN 3920557*, 2015.

Dr Preet Deep Singh. Emotional crowdfunding blurbs lead to success (of the campaign, at least). *Available at SSRN 3943467*, 2021a.

Dr Preet Deep Singh. How to make a hit song: Invest in a good studio. *Available at SSRN 3939762*, 2021b.

Dr Preet Deep Singh. Olympic medals: Matter of nerves. *Available at SSRN 3901321*, 2021c.

Dr Preet Deep Singh. Perception of founders' education and gender in startups. *Available at SSRN 3916424*, 2021d.

Dr Preet Deep Singh. Perception of founders' education and gender in startups. *Available at SSRN 3916424*, 2021e.

Dr Preet Deep Singh. Perception of founders's age and gender in startups. *Available at SSRN 3917905*, 2021f.

Dr Preet Deep Singh. To start-up or to work in a startup. *Available at SSRN 3932558*, 2021g.

Dr Preet Deep Singh and Sakina Poonawala. Crowdfunding: Word play to make money. *Available at SSRN 3932900*, 2021.

Preet Deep Singh. *Essay 1: Audit committee members' attendance and its impact on earnings management. Essay 2: Exogenously reduced busyness and its impact on earnings management.* PhD thesis, Indian Institute of Management Ahmedabad, 2017.

Preet Deep Singh. Close the big cities, open the small ones: How corona correlates with demographic variables. *Open the Small Ones: How Corona Correlates with Demographic Variables (April 6, 2020)*, 2020a.

Preet Deep Singh. Quantifying the risk of starting a train. *Available at SSRN 3587332*, 2020b.

Preet Deep Singh. Names of olympic athletes: Exploratory study. *Available at SSRN 3895300*, 2021h.

Preet Deep Singh. Perception of gender&founder background in startups. *Available at SSRN 3912640*, 2021i.

Preet Deep Singh. Perception of gender&founder background in startups. *Available at SSRN 3912640*, 2021j.

Preet Deep Singh and Sakina Poonawala. Whether and where to spend mandatory csr? *Available at SSRN 2802866*, 2016.

Preet Deep Singh and Chitra Singla. Executive stock options: Will it work as a good governance mechanism in all scenarios? 2016a.

Preet Deep Singh and Chitra Singla. Impact of independent directors' resignations on firm's governance. 2016b.