

An Analysis of The Beatles' Music

Janessa Tran

April 27, 2017

1 Introduction

The Beatles were an English rock band who rose to fame during the 20th century. They are known today for their contribution to music, film, literature, art, fashion, and have an ongoing impact on generations of popular culture. Their music expresses powerful ideas of love, peace, and imagination, and have earned them many accolades and awards over the course of their music career. During the 60s, the message of love and peace was arguably most prominent as the band vocalized an anti-war stance on the Vietnam War. Their lyrics became increasingly opinionated and optimistic, which some attribute to their meeting with Nobel prize winner Bertrand Russell, who was very active in speaking out about the war through his writings [6]. During this tumultuous time, it is understandable how music about love could be appealing. Regardless of the times, however, love has always been a prevalent subject for music by the Beatles. According to The Guardian's DataBlog in "Beatles Lyrics and the Words They Used Most", the top verb "love" is used 613 times throughout their music [7]. The Beatles often sang about love, but are these the songs that most appealed to their audience? The Rolling Stone's biography about The Beatles claims "the Beatles proved that rock and roll could embrace a limitless variety of harmonies, structures, and sounds" [8]. Moreover, the biography emphasizes the Beatles' influence on setting higher standards for musicians, making it a norm rather than exceptional to write your own material. Clearly, the Beatles' music impacted the industry in many ways and while they no longer create music, generations of people continue to enjoy their songs. What exactly makes The Beatles' music popular across generations? In this paper, songs by The Beatles are analyzed to better understand the factors that make these once popular songs also popular today. This will be accomplished through identifying and analyzing the relationships between the popularity of a song with the following: simplicity of the song lyrics, the length of the song, the track number of the song on the album, whether or not it mentions love, and whether or not the song was released as a single.

2 Data Description

The data is collected using the Spotify Web API [2]. Available in sixty markets around the world and available at no cost, Spotify is one of the most popular music streaming services in use, making it an appealing source for data. Moreover, Spotify has a wide library of music, including all songs released by the Beatles rendering convenience in gathering data for the analysis. The analysis incorporates data collected on songs from fifteen albums by the Beatles and focuses on three markets in particular: Mexico, United States, and Canada.

We continue with a discussion of how the explanatory variables are quantified and the motivations for choosing these variables. To begin, the simplicity of song lyrics and duration of the song is measured in light of the nature of popular music today. Many pop songs have the characteristics of being catchy, short, simple, and melodic. In order to capture the simplicity of each Beatles song, we gathered data on the unique word count of the lyrics using a Python script. The script reads in the lyrics of each song, creates a dictionary of all unique words in the song, and outputs the count of words in the dictionary. The data on the shortness of the song is simply gathered by obtaining the duration of each song in milliseconds. As mentioned previously, the Beatles often wrote about love, therefore this variable is included to explain the variation in song popularity. In the January 1981 issue of Playboy, Lennon says "in the early days, we didn't care about lyrics as long as the song had some vague theme - she loves you, he loves him, they all love each other" [5]. In order to measure a song as being a love song objectively, we use a dummy variable which has the value of "1" if the song mentions love in the lyrics, and "0" if the song does not. This data is also gathered using a Python script which reads through lyrics of each song, searching for the word "love" at least once. Next, the track number of the song is gathered as a variable of explaining popularity. This variable considers that artists may put their better songs earlier in the album, to showcase their best work. Therefore, songs with lower track numbers may be popular just by nature of the songs being better. Another reason this variable is included is because most people listen to albums sequentially, so earlier songs get more exposure and are more likely to be relatively popular. Our last

explanatory variable, the dummy variable *Single*, has a value of "1" if the song was once released as a single, or "0" if not. Singles are typically released prior to the album, as a means to provide a preview of an album and promote it, or to keep the musicians relevant in public. According to award-winning singer-songwriter Shantell Ogden in "Musicians: The Truth About Releasing a Single", singles cost thousands of dollars in advertising and promotions to get out in the market [4]. As the investment in releasing a single is so high, examining the correlation between singles and popularity seems worthwhile for this model. Lastly, the popularity of the song was gathered using Spotify's metric of popularity. The metric is an integer with a value between 0 and 100, with 100 being the most popular. The popularity, according to the Spotify Developer documentation is "calculated by algorithm and is based, in the most part, on the total number of plays the track has had and how recent those plays are" [1]. This means that songs that are frequently played now will have a higher popularity score than a song that was frequently played in the past. The summary statistics and scatterplots of the data can be found at the end of the paper in Figure 3 and Figure 2 respectively.

3 Empirical Results

To estimate the degree to which the aforementioned explanatory variables correlate with the popularity of a Beatles song, we estimate a function with the form

$$Popularity = B_0 + B_1TrackNumber + B_2Duration + B_3UniqueWordCount + B_4MentionsLove + B_5ReleasedSingle$$

where the Popularity and the explanatory variables are defined as listed in the table below.

Variable	Description
Track Number	Track number on album (first track = 1).
Duration	Duration of song in milliseconds.
Unique Word Count	Number of unique words in song lyrics, to measure simplicity.
Mentions Love	Dummy variable equal to 1 if lyrics include word "love", 0 if not.
Single	Dummy variable equal to 1 if released as single, 0 if not.

Table 1: Description for Explanatory Variables in Empirical Analysis.

The coefficients are interpreted with respect to their impact on popularity. The estimated coefficient B_1 should be interpreted as the estimated marginal effect of a one unit increase in track number on the song's popularity. The estimated coefficient B_2 should be interpreted as a one millisecond increase in a song's duration on the song's popularity. The estimated coefficient B_3 should be interpreted as the estimated marginal effect of a one word increase in count of unique words on a song's popularity. The estimated coefficient B_4 should be interpreted as the estimated effect on the intercept of the song lyrics including "love" on popularity. Lastly, the estimated coefficient B_5 should be interpreted as the estimated effect on the intercept of a song previously released as a single.

$$Popularity = 45.9373 - 0.1989TrackNumber + 2.4784E - 05Duration + 0.0183UniqueWordCount - 0.8828MentionsLove + 4.6629ReleasedSingle$$

The results of the regression indicate that all other independent variables held constant, increasing Track Number decreases Popularity by 0.1989, increasing Duration by one millisecond increases Popularity by $2.47E - 5$, increasing the count of Unique Words increases Popularity by 0.0183, a song that mentions "love" decreases Popularity by 0.883, and a song once released as a single increases Popularity by 4.6629. Of all the explanatory variables, Track Number, Duration, and Single are statistically significant estimates at the 5% significance level, while the others are statistically insignificant. Among our predictions regarding the relationships of our independent variables to our dependent variable, two results align with our previous notions: a higher Track Number is negatively correlated with Popularity and a song released as a single is positively correlated with Popularity. This regression suggests, however, that the typical characteristics that make pop music popular (simplicity and brevity among other qualities), may not apply for The Beatles songs observed.

Conducting a hypothesis test on the overall significance of the regression function, We ask the question: All else equal, does the regression model explain a significant portion of the observed variation in The Beatles songs popularity on Spotify? We state our null hypothesis and alternative hypothesis: all of the true unobserved slope parameters equal zero or at least one of the true unobserved slope parameters is statistically different from zero. Because our p-value, $1.68337E - 08$ is less than 0.05, we reject the null hypothesis and conclude that at

least one of the independent variables Track Number, Duration, Unique Word Count, Mentions Love, or Single is statistically related to Popularity.

Another point of interest is analyzing the subset of songs produced before and after the abrupt death of Brian Epstein in 1967, the founder and manager of the Beatles, to see if there are structural differences in the model. The years following Epstein's death, the Beatles were personally involved in financial and legal conflicts which may have impacted their music-making [3]. A Chow Test was conducted to answer the question "Is there a structural difference between the popularity of songs released under Epstein's management (prior to Epstein's death in 1967) and songs not released under his management?" Because our test statistic, an F-statistic of value 0.148791957, is less than our critical value 2.286, we fail to reject the null hypothesis at the 5% significance level and conclude that there is no structural difference between songs released under Epstein's management and not under his management.

4 Conclusion

In this paper, we estimate popularity of songs for 327 Beatles songs available on Spotify. Our pooled sample regression results suggest that Track Number, Duration, and songs released previously as singles have a statistically significant relationship with Popularity, while Unique Word Count and songs that mention love do not. While our regression results for song duration and simplicity contradict our understanding of what makes Pop Music popular among the general public, we did not measure melody of songs and this may have a more apparent relationship to song popularity. Although a few of our independent variables are not statistically related to our dependent variable, the overall regression model was statistically significant at the five-percent level and we have a clearer understanding of what may explain the variation in popularity of songs by The Beatles.

References

- [1] Spotify API Endpoint Reference, 2016.
- [2] Spotify Web API, 2016.
- [3] M. Beschloss. The Fifth Man: Brian Epstein and the Beatles. 2015.
- [4] S. O. Dave Cool. Musicians: The Truth About Releasing a Single. 2015.
- [5] M. Evans. *The Beatles: Paperback Writer 40 Years of Classic Writing*. Plexus Publishing Limited, 2012.
- [6] T. Garner. Vietnam War Songs: What Happened When the Beatles met bertran russell. 2016.
- [7] S. Rogers. Beatles Lyrics and Words They Used Most. 2016.
- [8] R. Stone. The Beatles Bio. 2001.

Figure 1: Results of regressing Popularity on Track Number, Duration, Unique Word Count, Mentions Love, and Singles

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.418165802
R Square	0.174862638
Adjusted R Square	0.157002522
Standard Error	6.402176671
Observations	237

ANOVA					
	df	SS	MS	F	Significance F
Regression	5	2006.494907	401.2989814	9.790677566	1.68337E-08
Residual	231	9468.197076	40.98786613		
Total	236	11474.69198			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	45.93734641	1.552337442	29.59237158	1.5183E-80	42.87879666	48.99589616	42.87879666	48.99589616
Track Number	-0.198919834	0.081144954	-2.451413484	0.014971735	-0.358798654	-0.039041014	-0.358798654	-0.039041014
Duration (ms)	2.47842E-05	8.09967E-06	3.059909422	0.00247567	8.82558E-06	4.07429E-05	8.82558E-06	4.07429E-05
Unique Word Count	0.018325348	0.014423593	1.270512048	0.205180397	-0.010093264	0.046743961	-0.010093264	0.046743961
Mentions Love	-0.88280076	0.871038254	-1.013504006	0.311879864	-2.598995812	0.833394292	-2.598995812	0.833394292
Released Single	4.662986111	0.950039427	4.908202732	1.73681E-06	2.791136102	6.534836119	2.791136102	6.534836119

Figure 2: Scatterplots of the independent variable against the dependent variable (Popularity)

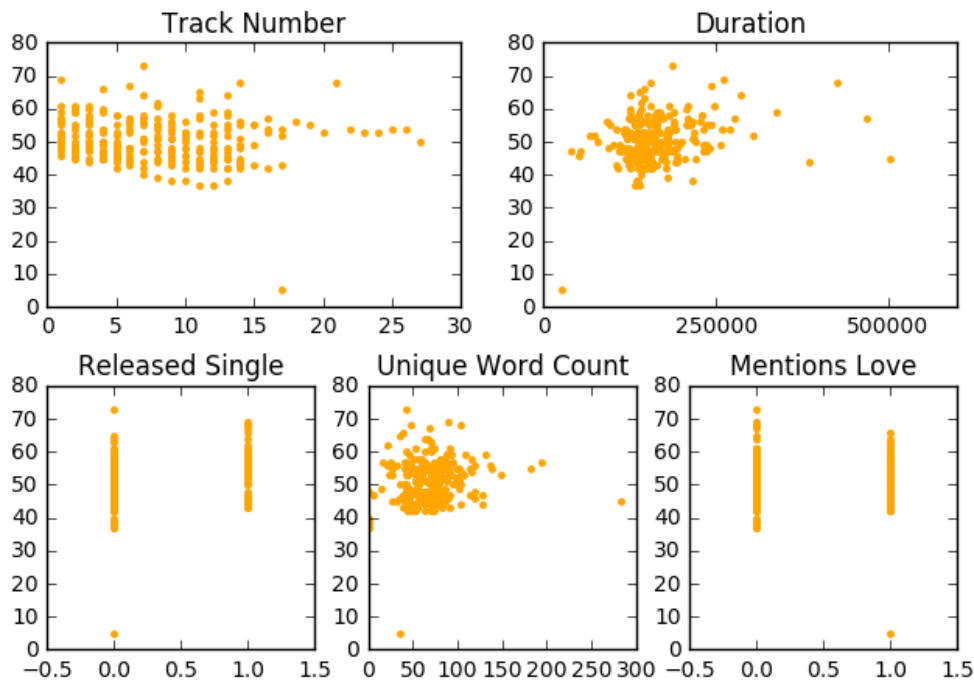


Figure 3: Summary statistics of variables in regression

	count	mean	std	min	25%	50%	75%	max
Popularity	237.0	50.548523	6.972917	5.0	46.0	50.0	55.0	73.0
Track Number	237.0	8.333333	5.186630	1.0	4.0	8.0	12.0	27.0
Duration (ms)	237.0	164345.991561	57035.153436	25986.0	134720.0	154160.0	180893.0	502013.0
Unique Word Count	237.0	68.118143	31.946903	1.0	51.0	67.0	82.0	282.0
Mentions Love	237.0	0.375527	0.485284	0.0	0.0	0.0	1.0	1.0
Released Single	237.0	0.274262	0.447086	0.0	0.0	0.0	1.0	1.0