

Do music charts still matter in the streaming age? An analysis predicting Beyoncé's Spotify numbers using a regression model.*

Tam Ly

April 17, 2024

This paper analyzes the number of Spotify streams of a Beyoncé song based on the number of weeks it spent on the *Billboard Hot 100* utilizing data collected from Spotify and Billboard. We find that the number of Spotify streams of a Beyoncé song increases as its duration on the *Billboard Hot 100* increases. Our discovery highlights the enduring relevance of charts in shaping today's music industry.

1 Introduction

Beyoncé Giselle Knowles-Carter is a singer, songwriter, producer, dancer, and one of the most successful and influential artists of our time. Over the course of her career, she has won 32 Grammy Awards, sold beyond 200 million records, and more than 100 of her songs have landed on the *Billboard Hot 100* chart. The *Billboard Hot 100* is a music ranking system that is supposed to showcase 100 of the most popular songs in the United States each week. Listeners can stream these hits on Spotify, the world's most popular audio streaming subscription service.

Do Billboard charts actually reflect which songs are most popular? There is a clear gap in knowing what makes a song famous. In the ever-changing music industry, it is not only about melody and lyrics, but also promotion. In this paper, we focus on music promotion and the importance of Billboard charting. We analyze the number of Spotify streams of a Beyoncé song based on the number of weeks it spent on the *Billboard Hot 100*. The estimand is the popularity of a Beyoncé song and this is considered in the number of Spotify streams.

Using data collected from Spotify and Billboard, we found that the number of Spotify streams of a Beyoncé song increases as its duration on the *Billboard Hot 100* increases.

*Code and data are available at: <https://github.com/atn-ly/beyonce>

The rest of this paper is structured as follows. Section 2 provides the sources, measurement, methodology, and variables of the dataset. Section 3 provides the model set-up and justification. Section 4 provides the results of the model and Section 5 provides a discussion of the results.

2 Data

2.1 Sources

The dataset used in this paper was created using data gathered from Spotify (**spotify?**) and Billboard (**billboard?**). The data was collected by the authors for the purpose of this paper to investigate the number of Spotify streams based on weeks spent on the *Billboard Hot 100* for Beyoncé songs. Song titles, album titles, and the number of Spotify streams were collected from Spotify (**spotify?**), while weeks spent on the *Billboard Hot 100* were collected from Billboard (**billboard?**). There were no similar datasets available that could have been used.

2.2 Measurement

First, we used Spotify (**spotify?**) to collect data on Beyoncé albums. Here, we had to decide which albums to include in our dataset since Beyoncé has 15 different albums in her Spotify discography. We decided to only include her eight studio albums and one soundtrack album and exclude her live albums and compilation albums because these are not as popular. Furthermore, among her eight studio albums, we only looked at the deluxe versions, since these contain the same songs with the same streaming numbers as the standard version along with some additional tracks.

Next, using Spotify (**spotify?**), we collected data on Beyoncé songs from the nine albums we selected and had to decide which types of tracks to include. Since we are interested in songs that charted, we decided to exclude remixes, extended mixes, Spanish versions, interludes, and original demos that were included in her albums because these are not as popular. Note that we also did not include songs where Beyoncé was a guest feature on another artist's song.

Then, we used Spotify (**spotify?**) to collect data on the number of streams for the songs that we selected. Spotify gives one stream to a track if it is played for at least 30 seconds and updates its streaming numbers once a day at approximately 3 PM EST (**spotify?**). We collected this data on April 12, 2024 after 3 PM EST.

Lastly, we collected data on the number of weeks each song spent on the *Billboard Hot 100* from Billboard (**billboard?**). Billboard keeps a chart history for Beyoncé that includes the debut date, peak position, peak date, and weeks spent for each song that charted on the *Billboard Hot 100*. From this list, we only collected the number of weeks spent on the chart for our selected songs. Billboard updates the *Billboard Hot 100* as well as the artist's chart history

every Tuesday (**billboard?**). We collected this data on Friday, April 12, 2024. For songs that did not chart, we recorded them with a 0 in our dataset.

2.3 Methodology

We analyzed the data in R (R Core Team 2023) using the following packages: **arrow** (**arrow?**), **broom.mixed** (**broommixed?**), **ggplot2** (**ggplot2?**), **here** (**here?**), **janitor** (**janitor?**), **knitr** (**knitr?**), **marginaleffects** (**marginaleffects?**), **modelsummary** (**modelsummary?**), **rstanarm** (Goodrich et al. 2022), and **tidyverse** (**tidyverse?**).

2.4 Variables

There are 4 variables in this dataset:

1. **song** which represents the song titles,
2. **album** which represents the album titles,
3. **spotify_streams** which represents the number of Spotify streams, and
4. **wks_on_chart** which represents the number of weeks spent on the *Billboard Hot 100*.

Table 1 shows a sample of the dataset with all the variables and the first 10 out of 140 observations.

Table 1: Sample of the cleaned dataset with the first 10 observations

Song	Album	Spotify Streams	Weeks on Chart
Crazy in Love	Dangerously in Love	1141641440	27
Naughty Girl	Dangerously in Love	197200179	22
Baby Boy	Dangerously in Love	309099095	29
Hip Hop Star	Dangerously in Love	6773025	0
Be With You	Dangerously in Love	14676517	0
Me, Myself and I	Dangerously in Love	167639029	24
Yes	Dangerously in Love	14646805	0
Signs	Dangerously in Love	15041276	0
Speechless	Dangerously in Love	17300089	0
That's How You Like It	Dangerously in Love	13007089	0

Figure 1 summarizes the data with all 140 observations. We see a positive correlation between the number of Spotify streams and weeks spent on the *Billboard Hot 100*.

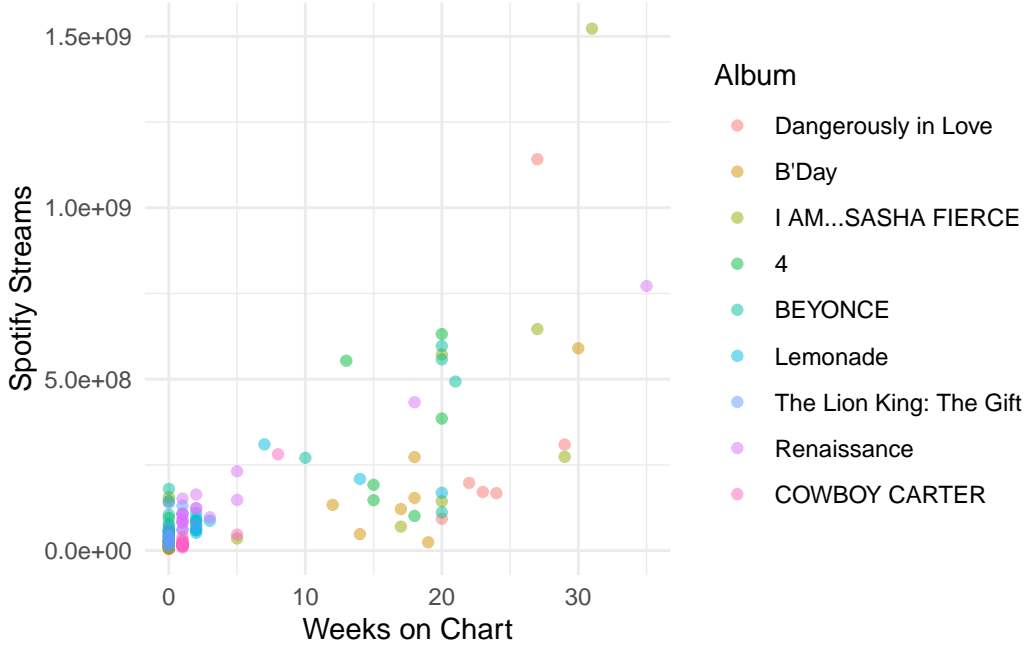


Figure 1: Relationship between the number of Spotify streams and weeks spent on the *Billboard Hot 100* for Beyoncé songs

3 Model

The goal of our modeling strategy is to predict the number of Spotify streams for a Beyoncé song based on the number of weeks it spent on the *Billboard Hot 100*. We used a negative binomial regression model in a Bayesian framework. Negative binomial regression is a type of generalized linear model that is useful for modeling count data.

3.1 Model set-up

The model that we are interested in is:

$$y_i | \mu_i, r \sim \text{NegBinom}(\mu_i, r) \quad (1)$$

$$\log(\mu_i) = \alpha + \beta \times \text{Number of weeks}_i \quad (2)$$

$$\alpha \sim \text{Normal}(18, 2.5) \quad (3)$$

$$\beta \sim \text{Normal}(0, 2.5) \quad (4)$$

$$(5)$$

Where:

- y_i is the outcome variable, representing the number of Spotify streams for song i ,
- μ_i is a parameter for the negative binomial distribution, representing the probability of success in a single trial,
- r is a parameter for the negative binomial distribution, representing the number of successes,
- Number of weeks $_i$ is the predictor variable, representing the number of weeks spent on the *Billboard Hot 100* for song i ,
- α is a parameter, representing the intercept with a specified prior probability distribution that is Normal with a mean of 18 and standard deviation of 2.5, and
- β is a parameter, representing the slope coefficient with a default prior probability distribution that is Normal with a mean of 0 and standard deviation of 2.5.

3.2 Model justification

We expect a positive relationship between the number of Spotify streams and weeks spent on the *Billboard Hot 100* based on the positive correlation in the graph that we observed in [Section 2](#).

Negative binomial regression operates under several assumptions. It assumes linearity between the outcome and predictor variables, independence of observations, and no multicollinearity.

We considered Poisson regression as an alternative model since it is also used for count data. However, one of the restrictions with Poisson regression is that it assumes equal mean and variance. Negative binomial regression relaxes this assumption to allow for over-dispersion. We fitted both and compared them using posterior predictive checks in [Figure 2](#). We see that the negative binomial approach does a better job of fitting the data.

We implemented additional model checking and diagnostic issues. Details and graphs can be found in [Section A](#).

4 Results

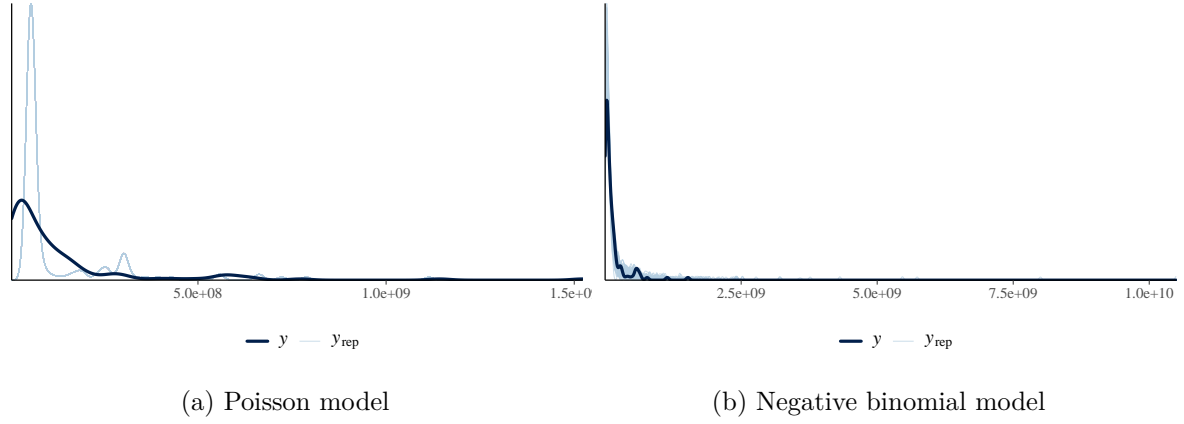


Figure 2: Comparing posterior prediction checks for Poisson and negative binomial models

Table 2: Examining the number of Spotify streams for a Beyoncé song based on the number of weeks it spent on the *Billboard Hot 100*

	Spotify Streams
(Intercept)	17.727 (0.081)
wks_on_chart	0.094 (0.008)
Num.Obs.	140
Log.Lik.	−2687.522
ELPD	−2689.6
ELPD s.e.	14.0
LOOIC	5379.3
LOOIC s.e.	27.9
WAIC	5379.3
RMSE	140 888 217.09

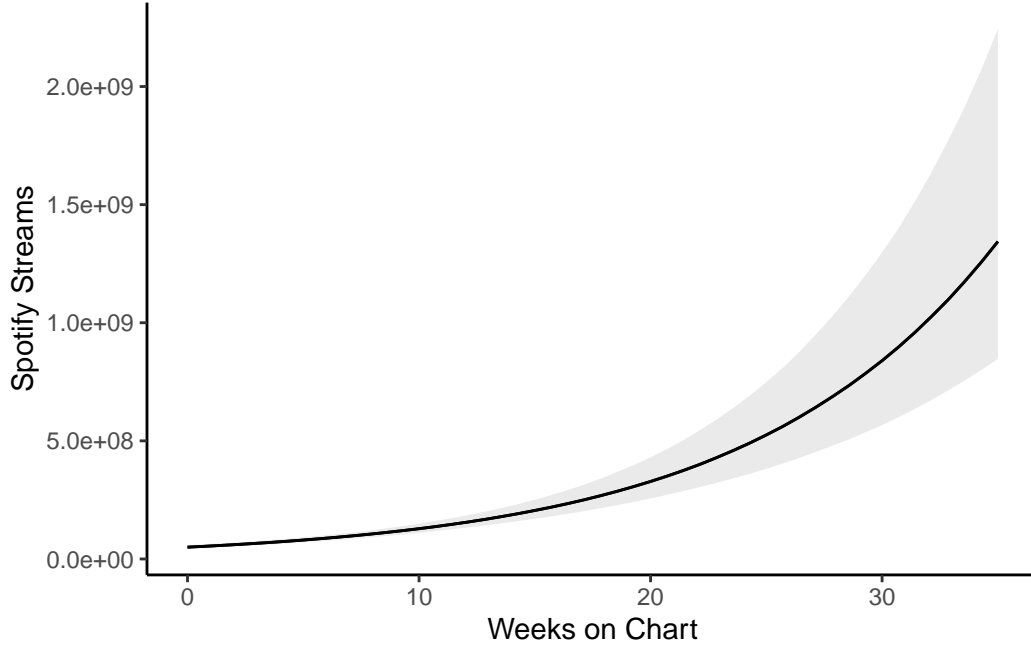


Figure 3: The predicted number of Spotify streams for a Beyoncé song based on the number of weeks it spent on the *Billboard Hot 100*

5 Discussion

5.1 Findings

In this paper, we collected and visualized data from Spotify and Billboard as discussed in Section 2. Then, we used a Bayesian negative binomial regression model as seen in Section 3 to predict the number of Spotify streams for a Beyoncé song based on the number of weeks it spent on the *Billboard Hot 100*. Following our analysis in Section 4, we found that the number of Spotify streams of a Beyoncé song increased as its duration on the *Billboard Hot 100* increased. We will now discuss the historical context behind Billboard and why huge superstars, such as Beyoncé, still care about charting.

5.2 A bit of history

In the 1940s, Billboard started developing music charts to try and determine the most popular song. This was to help inform people in the music industry, including record stores and jukebox operators, about which records to order. The *Billboard Hot 100* was created in 1958 and has undergone complete upheavals since its inception. How exactly does Billboard determine the most popular song? This has always been the subject of much controversy due to its biased

survey methods. Billboard first measured sales by surveying record stores, although it was impossible to know if the store clerks were lying or not. Then, in a book published in the 1970s, co-founder of Casablanca Records, Larry Harris, alleged that industry insiders had heavy influence over the charts. In the 1990s, Billboard began tracking sales based on barcode scans and hip-hop sales seemingly rose dramatically overnight.

Charts are supposed to reflect the industry’s reality and in 2018, it was announced that the *Billboard Hot 100* would take streaming data into account. However, the exact formula behind Billboard’s calculations remains unknown to the public. We only know that one stream does not necessarily equal one sale and that some streams, such as for on demand services and from paid-tier users, count more than others. By the 2010s, the question of what the Billboard charts actually reflect became even more prevalent after rabid fans proved that it was susceptible to coordinated manipulation.

5.3 Why charts still matter

Do artists as big as Beyoncé still care about music charts? Chart success can bring significant music promotion to an artist’s career. Songs that land on the *Billboard Hot 100* also land on Spotify’s biggest playlists, such as Today’s Top Hits which has almost 35 million followers. Cherie Hu, a music industry analyst at the intersection of music and data has deemed playlists “the new album” and “more powerful than radio” (**cherie?**). A placement on Today’s Top Hits can bring up to \$163,000 in additional revenue per track. Moreover, for new artists, high positions in the charts can lead to more ownership and creative control in their musical career. This can help explain the importance of why many of Beyoncé’s longest charting songs come from her earlier albums. For more established artists who have a loyal fanbase (even Beyoncé!), charting is essential for maintaining their brand and curtailing any image of a decline.

5.4 Weaknesses and next steps

Some weaknesses become apparent in our approach in using Spotify streams as a metric for the popularity of a Beyoncé song. First, one of Beyoncé’s albums, *Lemonade*, was added to Spotify three years after its release. This means that the number of Spotify streams for the songs from this album are underrepresented in our dataset. Furthermore, Spotify numbers are also not immune to industry politics that underlie promotional mechanics. For example, all of the major record labels own a sizable share of Spotify and can give some artists preferential treatment on the platform.

In the future, it could be of interest to run the analysis again to give Beyoncé’s latest album released on March 29, 2024, *COWBOY CARTER*, more time to see how long it lasts on the chart. Another possible idea for future work could be to use a different

Appendix

A Model details

A.1 Posterior predictive check

We compare the posterior with the prior. Figure 4a suggests that there is an issue with the default prior we specified for α . Our re-specified model in Figure 4b shows that it does a better job of fitting the data.

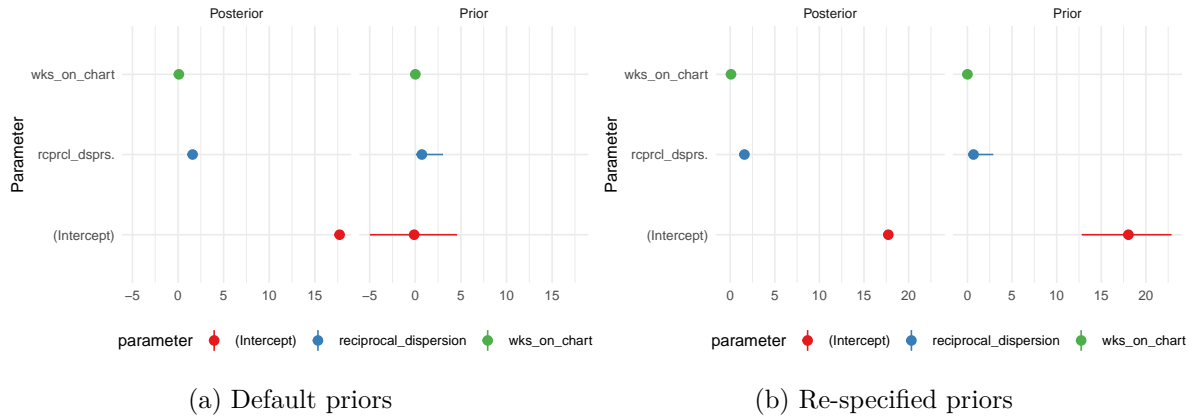


Figure 4: Comparing the posterior with the prior

A.2 Diagnostics

Figure 5a is a trace plot and it shows that there are no horizontal lines that appear to bounce around and have a nice overlap between the chains. Figure 5b is a Rhat plot and it shows that everything is close to 1 and no more than 1.1. This suggests that there are no problems in both.

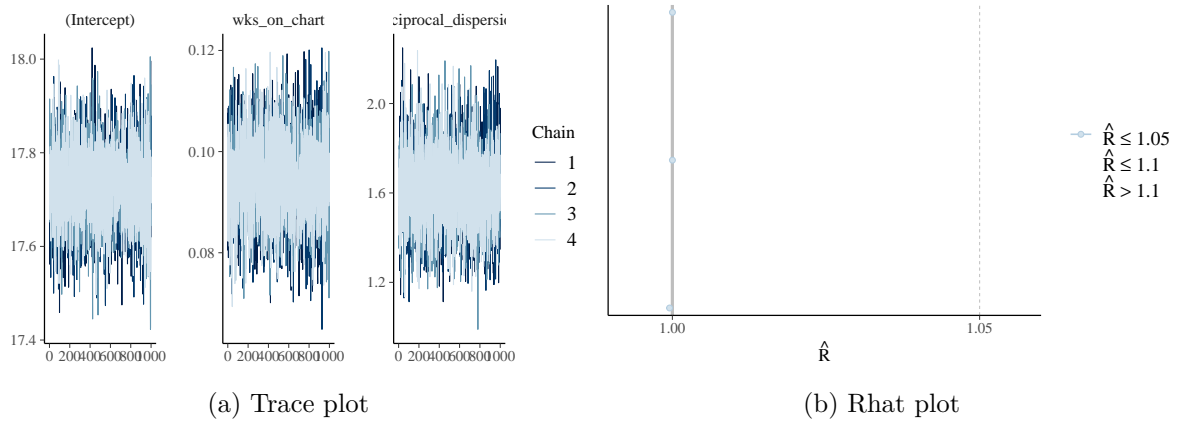


Figure 5: Checking the convergence of the MCMC algorithm

References

- Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. “Rstanarm: Bayesian Applied Regression Modeling via Stan.” <https://mc-stan.org/rstanarm/>.
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.