The Anatomy of a Hit: Statistically Learning from the Best*

Modeling the most impactful elements of the most successful musicians.

Luca Carnegie

April 18, 2024

We attempt to do some interesting things with models and music - stay tuned!

Table of contents

1 Introduction **3**

2 Data **4**

3 Model **4**

3.1 Model set-up 4

3.1.1 Model justification 4

4 Results **4**

5 Discussion **4**

5.1 First discussion point 4

5.2 Second discussion point 4

5.3 Third discussion point 4

5.4 Weaknesses and next steps 4

Appendix **5**

A Additional data details **5**

B Model details **5**

B.1 Posterior predictive check 5

*Code and data are available at: https://github.com/lcarnegie/valence-modeling.

B.2 Diagnostics 5

References **6**

1 Introduction

Throughout the history of the music industry, musicians successful or not have always pined for one thing - writing a hit and making it big. However, for large music firms, taking on musicians that have yet to generate considerable popularity is an large risk, since they are often gambling on an unknown je ne sais quoi, and often, any given artist is more likely to not produce a hit than produce a hit; this is easily shown through the presence of a few small "superstars" that consistently dominate the music market (Rosen 1981). This poses an interesting question for artists vying to become that superstar: what does make a song a "hit"?

One way to measure how popular a song is through the 'valence' or perceived energy of a song.[investigate the motivation here]. Measuring valence has been done at a national and cross-artistic-work level (Dodds and Danforth 2010), but I investigate solely song valence on an individual basis. Though difficult to exactly predict what individual song could be a hit, it is very possible to wean out insight from previously released music to statistically learn the most important elements of a great song.

Using data from Spotify's API, I construct a linear model of valence or "energy", using the discographies of the ten most popular artists [find a source] of every particular genre in every decade since 1950. I find that [there are interesting things to be found…]. Blah Blah Blah.

In the streaming era, competition for stardom is at an all-time high. Being cognizant of the most impactful elements could encourage artists to focus on what truly works to generate hits. This may allow them to carve their own paths outside of the controlling nature of a record contract (Burke 1997). These insights could be used to give more knowledge and freedom to artists to be able to strike out on their own and carve their own destinies in the world of music.

2 Data

3 Model

3.1 Model set-up

3.1.1 Model justification

4 Results

5 Discussion

5.1 First discussion point

5.2 Second discussion point

5.3 Third discussion point

5.4 Weaknesses and next steps

Weaknesses and next steps should also be included.

Appendix

A Additional data details

B Model details

B.1 Posterior predictive check

In **?@fig-ppcheckandposteriorvsprior-1** we implement a posterior predictive check. This shows...

In **?@fig-ppcheckandposteriorvsprior-2** we compare the posterior with the prior. This shows...

```
# #| eval: true
# #| echo: false
# #| message: false
# #| warning: false
# #| label: fig-ppcheckandposteriorvsprior
# #| layout-ncol: 2
# #| fig-cap: "Examining how the model fits, and is affected by, the data"
# #| fig-subcap: ["Posterior prediction check", "Comparing the posterior with the prior"]
#
# pp_check(first_model) +
#   theme_classic() +
#   theme(legend.position = "bottom")
#
# posterior_vs_prior(first_model) +
#   theme_minimal() +
#   scale_color_brewer(palette = "Set1") +
#   theme(legend.position = "bottom") +
#   coord_flip()
```

B.2 Diagnostics

?@fig-stanareyouokay-1 is a trace plot. It shows... This suggests...

?@fig-stanareyouokay-2 is a hat plot. It shows... This suggests...

References

Alexander, Rohan. 2023. *Telling Stories with Data*. https://tellingstorieswithdata.com/.

Burke, Andrew E. 1997. "Small Firm Start-up by Composers in the Recording Industry." *Small Business Economics* 9 (6): 463–71. https://doi.org/10.1023/A:1007968604929.

Dodds, Peter Sheridan, and Christopher M. Danforth. 2010. "Measuring the Happiness of Large-Scale Written Expression: Songs, Blogs, and Presidents." *Journal of Happiness Studies* 11 (4): 441–56. https://doi.org/10.1007/s10902-009-9150-9.

Friendly, Michael, Chris Dalzell, Martin Monkman, and Dennis Murphy. 2020. *Lahman: Sean "Lahman" Baseball Database*. https://CRAN.R-project.org/package=Lahman.

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. "Rstanarm: Bayesian Applied Regression Modeling via Stan." https://mc-stan.org/rstanarm/.

Kim, Seon Tae, and Joo Hee Oh. 2021. "Music Intelligence: Granular Data and Prediction of Top Ten Hit Songs." *Decision Support Systems* 145 (June): 113535. https://doi.org/10.1016/j.dss.2021.113535.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Rosen, Sherwin. 1981. "The Economics of Superstars." *The American Economic Review* 71 (5): 845–58. https://www.jstor.org/stable/1803469.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.