



NLP CLASSIFICATION OF REDDIT POSTS

GRATEFUL DEAD VS. PHISH - DAVID BERTSCH



GRATEFUL DEAD

1965 - 1995



PHISH

1983 - PRESENT

Hypotheses

- * Mostly accurate classifications
- * A lot of band-specific language:
 - * Member names
 - * song/album titles
 - * Dates
- * Model would have a hard time with vague text, crossover posts



Gameplan

- * Start simple
- * Tinker/Re-evaluate
 - * Vectorizers
 - * Lemmatization
 - * Model types



Dataset/ Data Cleaning



- * Pushshift API
- * ~5,900 Dead posts
~5,800 Phish posts
- * Past 2 years
- * Post title + post text
- * Removed some non-language related text



Model Iteration

- * Logistic Regression/
Naive Bayes similar
- * Accurate classifications,
but overfit
- * Similar for both
CountVectorizer and
TF-IDF



Model Iteration

- * Random Forests/Support Vector Machines
- * Both increased bias/reduced variance
- * Support Vector Machines more accurate



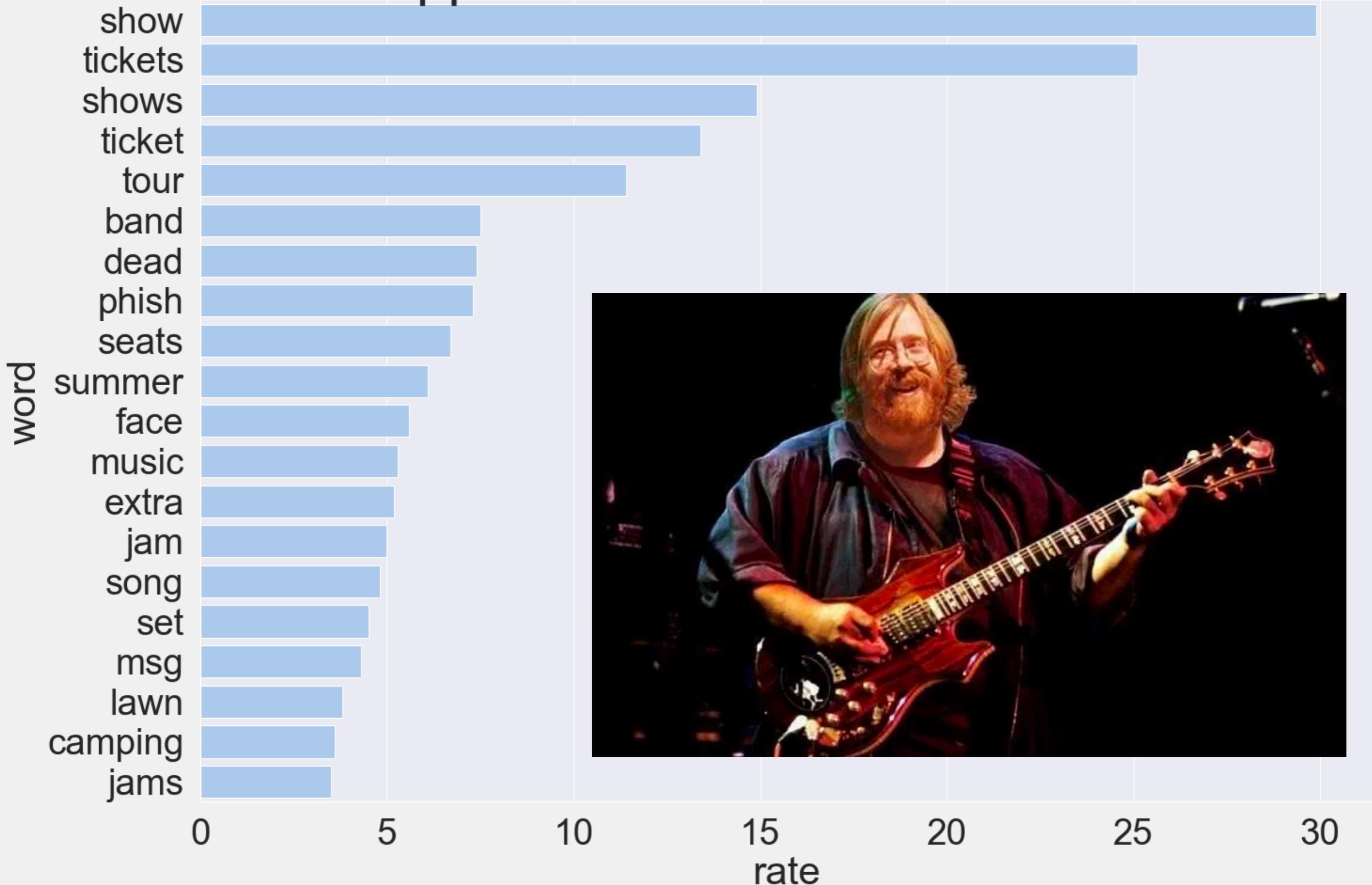
Final Model

- * Lemmatized text
- * TF-IDF Vectorization
- * Support Vector Machines
 - * 2,400 features
 - * Tuned regularization



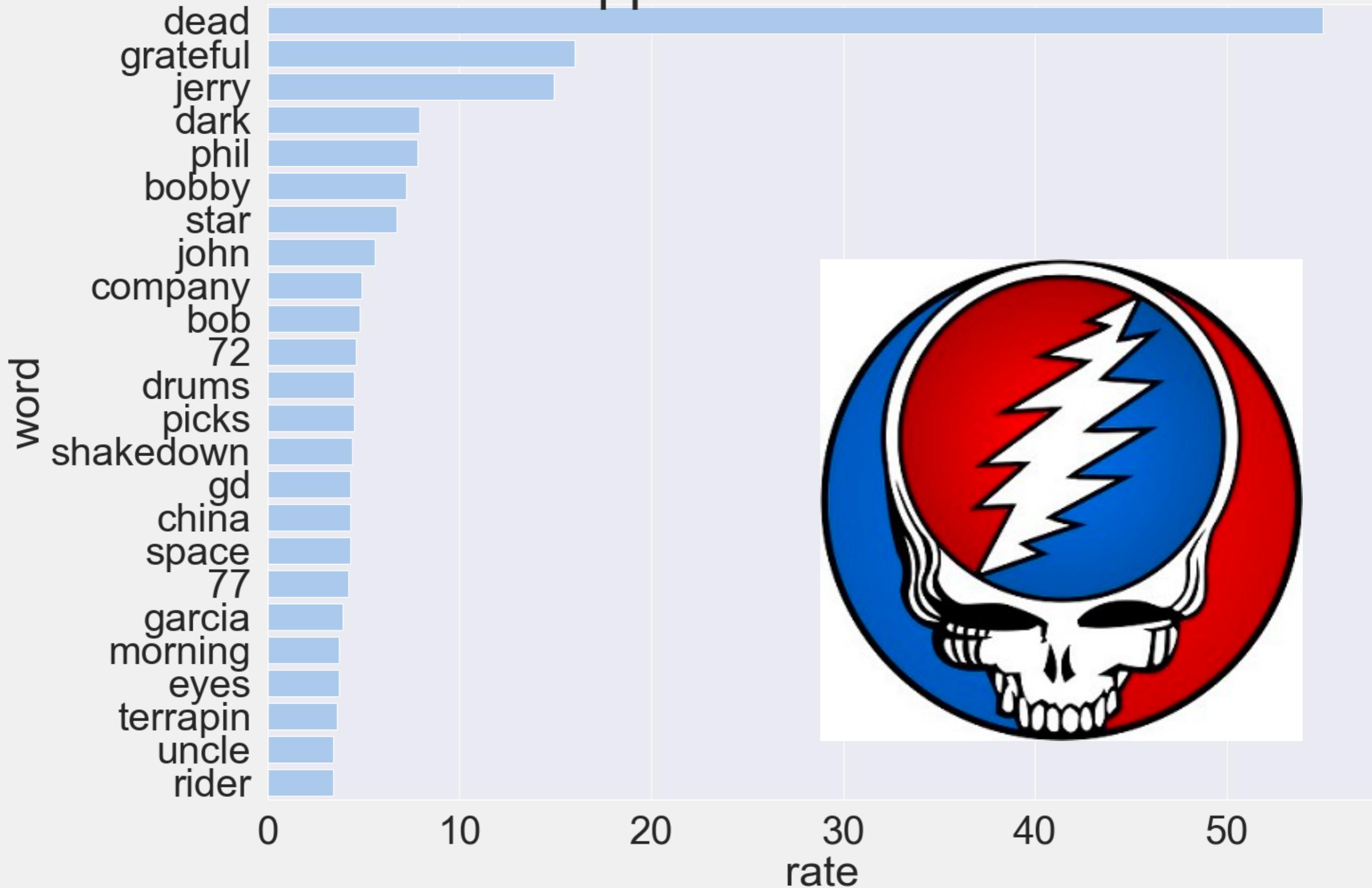
Form Unclear

Word Appearance Rate Per Misclassification



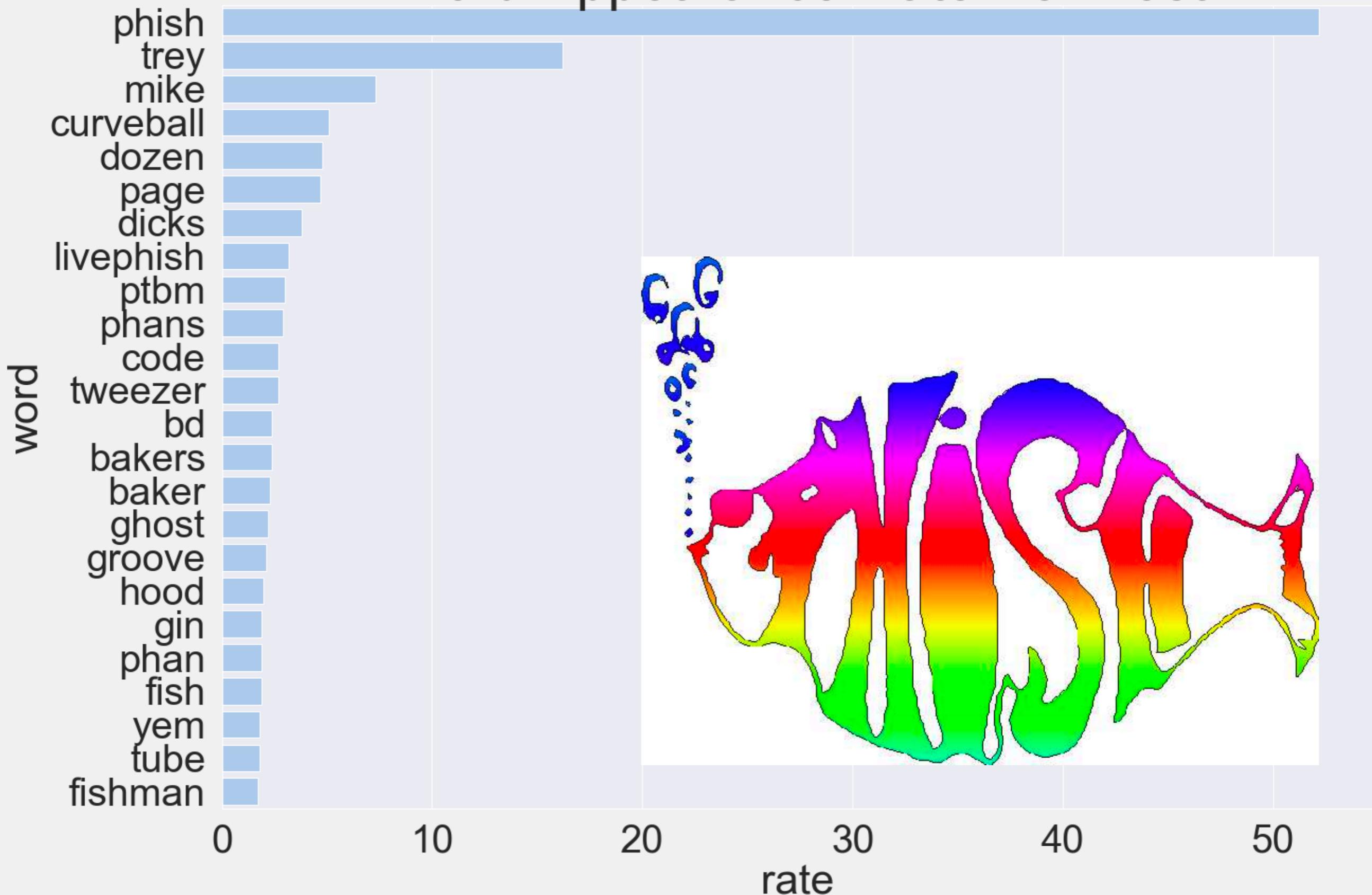
Dead Giveaways

Word Appearance Rate Per Post



Phish Hooks

Word Appearance Rate Per Post



I Need a Miracle

Best place to trade tickets? I'm not looking to scalp or go above face value, just trying to trade my GA for good hard seats. Short wife means I need 1st row of sections.

Anyone get tickets yet? Wondering how much would the price range be so i know what to put aside, thanks!

Philly tickets!? I went to the Wells Fargo box office to avoid online fees. They started selling them around 9:55. The \$50 and \$75 tickets were gone before 10:00!

Blossom pavillion vs. lawn Is it worth the extra 30 or 40 bucks for a back\-\of\-\the\-\pavillion ticket at blossom? Also, do they give you a wristband or something for pavillion access, or do the ushers just visually check your ticket?

Chicago ticket warning... For all you heads looking for cheap Chicago tickets, Don't buy hard tickets from Craigslist. Fake tickets are everywhere on CL right now. Its like The brown acid. Stay AWAY. But if you feel like experimenting buy lottery tickets instead. Be safe be smart

Maybe So, Maybe Not

Favorite Complete Show on YouTube? Thinking about putting on an old show tonight. What would you recommend?

For the nugs webcast are you able to watch on multiple devices at the same time?

Interested in a hoodie?

Favorite Fall Tour Show Recs Needed Making the drive up to Albany tomorrow and need a good show or two for the trip. Any favorite fall tour shows that haven't been officially released before?

Anyone have/know of a driveway/neighborhood/campground that I could park my car at after the show in Phoenix on Sunday?

when the fuck do we get our boxset announcement?

What experiences have you had with cults on the lot? I have been reading about the twelve tribes and am interested in hearing your run ins with them or any other cults. I was thinking of doing a research paper on them do you think they are too dangerous to visit? Thanks!

Sharing in the Groove

As of today, Trey has been alive longer than Jerry. Eternally grateful that we have been able to hear and see them each. Their contributions to music will live on for a long long time.

I posted in the Dead subreddit, but are there any Grateful dead/phish/other sweet bands themed fantasy baseball leagues out there?

Hey guys, I am a huge Grateful Dead fan, but have never listened to Phish. I have seen a lot of crossover in their fan bases, and I was wondering where I should start (certain shows/albums/songs) thank you so much

Can you guys recommend any shows/jams to get me, a huge Deadhead, into Phish?

Deadhead trying fish?

Fall '97 Spring '77 The Dead released the May 1977 box set and Cornell this year. I think its time Phish release the Fall '97 box set. I spent a good amount of time this summer listening to both tours and they both share the good funk in their own way and with ideal energy and tightness. Just wanted to put that out there. I want my goddamn box set.

All Things Reconsidered

- * More data
 - * So model can learn more key terms
 - * Shorter interval between pages scraped
- * Apply subject matter expertise to assign weights to keywords



Questions....



**Been you to have
any spike, man?**