Am I The Asshole Prediction

By Robert Malka

(With thanks to Ben Bell & Kenneth Gil-Pasquel!)

Who's the Asshole, Here?

How do people perceive what we're saying and why? Are there any qualities about our tone, word choice, and phrasing that would make someone like us less, think better of us, or carry our message across with further clarity? If we could find answers to this question by sifting through aggregate data, we might be better able to convince a judge/jury that we're not at fault, create a more effective business pitch, or, for the field of self-help, help others be more persuasive.

Shifting away from the narrow perspective of business value-add, we can also examine the limits of machine learning's understanding by means of this dataset – examining where machine learning, and the current state of AI generally, falls (far) short in its assessment of e.g. a person being (or not being) an asshole. Insofar as this can be shown through this project, we validate the value and uniqueness of being human.

We examine the subreddit, Am I The Asshole (AITA), which receives thousands of unique visitors a month. Its goal is to serve as a forum for people to ask the crowd whether or not they, within their self-reported experience, can be understood as an asshole. We investigate features that might serve to better predict whether or not an individual is behaving like an asshole, and experiment with Decision Trees, Random Forests, and Logistic Regressions to predict assholery based on features related to, among other things, the author's post and post title.

The business problem may not have a specific timespan or clear numerical benchmark attached to it, but that does not stop it from being potentially useful to a consulting firm or other entity. We will examine the presence or absence of features relative to the chance of receiving an NTA ("Not the Asshole) or YTA ("You're the Asshole) and, in spite of philosophical limitations, exhibit statistically significant improvements in ML benchmarks.

Huge Optional Aside: *On the Limitations of Our ML Analysis*

You're certainly wondering why I mention something so grandiose as examining the limits of machine learning's analysis within an otherwise-straightforward search for features predicting assholery.

The answer is, simply put, that the features we would *like to have* in examining this dataset, to accurately guess at the crowd's decisions, are not possible through machine learning, and that the features which do have significant predictive power, and which are wonderfully creative and helpful, only reveal the chasm between AI as it is, currently, and the human being, and which show clearly *what problems ought to be worked on with ML and which ought better to be left to humans and their inherent strengths* – problems, for example, like predicting who's an asshole (or, better yet, deciding what an asshole *is*).

Let me explain.

At the heart of this conversation live two famous philosophers: Plato and Heidegger. Plato asserted that there exists a *form* of things, an ideal, which nothing ever truly is, but always aspires to be. For example, a four-legged chair, while coming in many different shapes and sizes (including three- and two-legged chairs), 'borrows' in some way from the form of a chair. That's how we know it's a chair. We adopt this philosophy when we use supervised learning: Every picture of a chair gets labelled a "chair," and gradually some "form," some estimation of a chair, exists in the network. It thus predicts whether or not this or that picture is of a chair (does it have features such that it belongs in this group?), or something else. (Plato applied this philosophy also to The True, The Good, The Beautiful, and so on, but we won't go into that here.)

Heidegger, by contrast, insisted that things do not reach upwards, into idealized forms: They are best understood by means of their *functions*. How do we know whether a chair is a chair? By means of the fact that we *sit on it.* This is why a beanbag chair and a three-legged chair are both chairs: They are still capable of being wielded by a human being for their decided-upon function: Sitting. Heidegger understood objects as being not a "what" but a "what-for." Yet – and this is key – chairs *do not have to be understood as made for sitting.* They could be used as tables, such that humans sit on their knees. They could be sacred emblems representing the great Chair Goddess. They could be building blocks used to build the world's largest blanket fort.

The only way we can understand a chair – *really* understand it – is if we are embodied beings, able to interact with, participate in, and *be inseparable from the World*, as we are. AI is a tool that is part of this world, but it is not able to come to grips with the world because, as of yet, it is not encapsulated within a body/ies able to seamlessly interact with chairs, or beanbags, or assholes.

So, why does this matter?

Because what we most want to know, in this project, is whether person X is an asshole, according to the community – but there are *innumerable cultural grounds* through which to understand this or that person as an asshole (and what the community might *even mean* by designating someone as an asshole or not) – and almost all of them are *inferred* 'features,' not explicitly stated in the dataset. I'll share an example:

I use reading level as one feature to see whether people might be more or less of an asshole, given a certain writing ability. I noticed at least one negative (!) reading level, right off the bat. This means the post was supposedly so confusing the algorithm presumed we wouldn't understand it. Yet the community seemed to have understood: The poster was "Not the Asshole." A cursory review clarified things. The post in full is:

**Title:* I told a goth girl she looked like a clown.*

**Body:* I was four.*

So we see immediately that not only is it not "confusing", but remarkably clear, in context. (When title and body were included, the reading level remained negative, as a slightly-improved -1.4.)

Notice *how much* isn't included in this post, how much we bring to it, to understand why the crowd determined this person was not an asshole. We introduced the archetype most people hold for "a goth girl," and all of the nuanced qualities a goth girl has; the quality of the insult "clown" and how it fits (or doesn't) with being a goth; the fact that this *could have been understood as an insult* or presumed to be so; and how a human being can broadly come to understand and value (or not) the remark as it's made by a four-year-old. There is nothing reducible to "data-crunching" in the interplay of each of these pieces. Even if, for example, we established a rule dictating that remarks made by children should be labelled as “Not The Asshole”, that is a *culturally-contextual understanding*. Is that still the case in China?, Papua New Guinea?, the United States in 1953 or 2031?

Even within the context of this post, a whole story can be developed, perhaps even one in which the author is a less sympathetic character. We can imagine several responses from the goth girl: laughter, maybe, or, if she felt insecure, perhaps she *was,* in some indirect way, insulted; perhaps the child’s tone was unacceptably inappropriate, indicative of a budding future asshole. But the response of the community (a community that, remember, is ever-changing – reliant on who saw and decided to respond to the post) seems to have decided here that a kid's innocence excuses him from what is otherwise a crass remark.

The bottom line: culture, be it American or any other, is not reducible to atomic elements. It is an ecosystem with an interplay of nearly infinite parts, to which we, and not AI, have access. We are part of a culture that understands the significance of being an ‘asshole,’ all the ways in which one can be one, *why* this or that person was called an asshole, whether we agree with that assessment, and so on.

By this project I’m reminded that the words on a page or images on a screen are *referential*, pointing towards things that are in the world. They are not the world itself. It’s therefore impossible to be able to communicate the significance of any event, thing, or experience the data is pointing towards using only the data, since we fill it in with our embodied experience. Word mappings, as one example, capture the relationship between words as they are said, but they do not reveal what those words *embody*. Word mappings will always be a cursory explanation (at best) of the embodied understanding.

Therefore the reader will notice in our analysis that these features fundamentally cover only the shallowest inferences of assholery.

I hope we humans fearlessly assert what is human about us – our ability to make meaning of things, and invent and create those meanings for us to share and celebrate – and not presume AI has that power over us, and that we are helpless before it. It is, after all, stupid, as existentially stupid as can be.

For a full discussion on the limitations of AI, investigate Hubert Dreyfus’ *What AI Can’t Do*.

Further Caveats

Reddit’s AITA subreddit decides whether or not someone is the asshole, given a situation they describe from their own experience, in their own words, based on a majority vote. This majority vote relies on numerous factors, namely:

- The people who are judging are self-selected rather than randomized (they choose to click on the thread).
- People who comment often give nuance in their posts – cultural markers for what is and isn’t appropriate, ask clarifying questions, are sometimes thoughtful and unthoughtful – which isn’t

reflective of the final binary decision given (YTA/NTA). Upon individual inspection of these comments, a reader (or dare I suggest, a majority of readers?) might note a comment in the minority opinion that is overall more persuasive, and more well-thought-out, than the inclinations of the majority judgment.

- The audience making these decisions may be grouped according to time of day, positioning of the post, controversy of the times (related news, fiery memes, something in the culture wars) that leads to one opinion or the other. There is no way, based on the data, to account for these inevitable movements of the national (or international) zeitgeist. We are assessing the smallest snapshot of the particular crowd X's <u>overall</u>, <u>on-balance</u> judgment.

1. Data

My data was received, largely clean, from:

- https://github.com/iterative/aita_dataset

2. Data Cleaning

Fortunately, not much cleaning was required. I changed the "edited" column, signifying whether or not the original post was edited, to "True" or "False" as opposed to epoch time; I changed the epoch time of the post's publication to datetime; and I took out a series of symbols that would have rendered feature analysis inaccurate.

3. EDA

For a full examination, please see the EDA report [link].

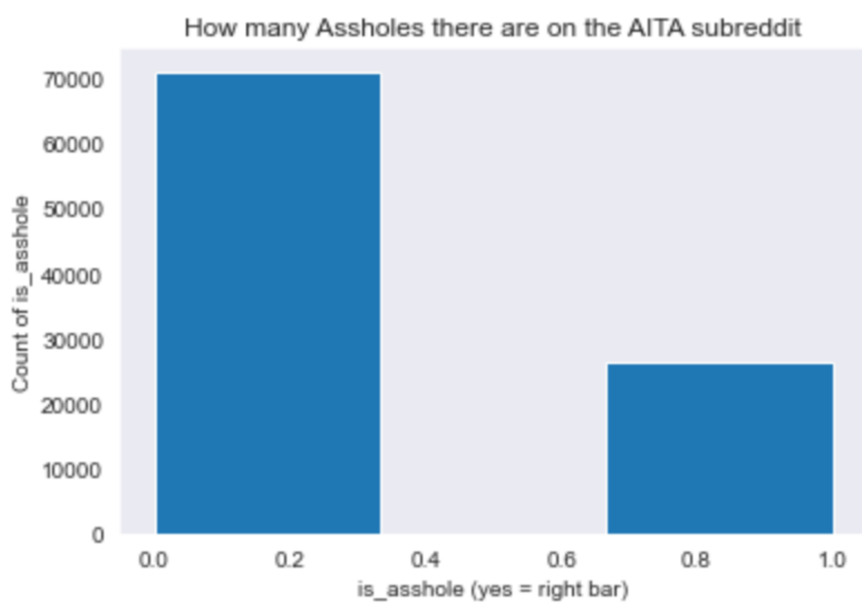
The hypotheses I examine:

- Does the number of I's said in a post, relative to other pronouns, reflect YTA/NTA?
- Does the word count of a post reflect YTA/NTA?
- Does having a question mark (or not) at the end of a post reflect YTA/NTA?
- Does asking "Would I be the asshole [if]" (WIBTA) versus asking "Am I the asshole [for]" (AITA) reflect any differences in YTA/NTA?
- Do we see a difference in likelihood of assholery based on whether or not a subject fits into the following categories?
    - Familial Disagreement
    - Professional Concerns
    - Romantic Relationships
- If we measure the tonal valence of each sentence (highly positive, highly negative), will there be differences in YTA/NTA re: tone?

Some interesting observations:

- There are some (minor) negative correlations, which was thoroughly unexpected. Some of the most remarkable ones are between is_asshole and questionmarklast: *not* having a question mark at the end of your sentence – leaving it as a period, say – correlates slightly with not being an asshole.
- The valence of tonality is not ultimately correlated with being or not being an asshole.
- There is a greater likelihood of being deemed an asshole if the issue is romantic (particularly if it involves sex) than in the other groups.
- The younger the individual being discussed relative to the poster (and the lower in the power structure), the more likely the poster is to be an asshole. We infer that the poster is older and higher in the power structure (e.g. when the subject of children is being discussed, we assume that the writer is older and higher in the familial or broader power structure than the children).
- Conversely, the higher the authority (boss, parent), the less likely the poster is to be an asshole.
- The reading level of the post makes no difference, suggesting the community doesn't judge for the linguistic competency of the poster: The focus is on the moral dimension.
- Use of pronouns (how many times one says "I", for example), along with the post length, have minimal correlation with community judgment. (-0.02 to -0.03 correlation on the heatmap, as seen below.)
- Funnily enough, whether or not a post was edited is a decent indicator of whether or not the poster was an asshole.

Basic facts about the dataset:

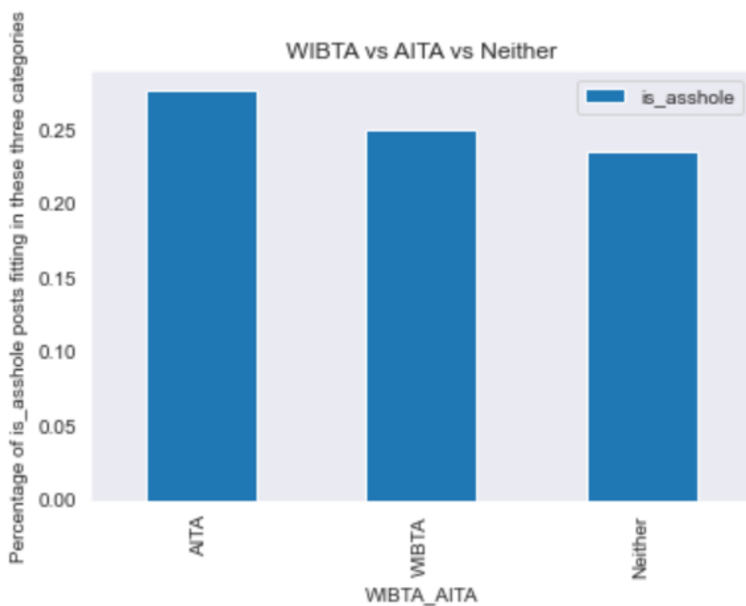


Interesting! So most of the time, people are NOT considered assholes ('0' (the left side) are not deemed assholes) -- it's actually only about 25,000 of 97,000 people -- or a little more than 25% -- of people who are deemed an asshole at all. Not sure if this indicates a generosity in the community, since we have no broad standard. Suffice it to say: More than a quarter of the people who decide to self-reflect on their assholery are confirmed to have behaved like an asshole by the Reddit community.

I also did a heatmap of the initial features, which is displayed below. I opted to examine other features by t-test and frequency relative to the is_asshole global average, which is why the features integral to a number of hypotheses are not included in this heatmap.

	edited	score	num_comments	is_asshole	countI	countHeSheThey	IvsHeSheThey	post_word_count	questionmarklast	bodyreadinglevel
edited	1.000000	0.070000	0.100000	0.090000	0.140000	0.120000	-0.000000	0.150000	-0.380000	0.000000
score	0.070000	1.000000	0.840000	-0.010000	0.050000	0.060000	-0.020000	0.060000	-0.030000	-0.000000
num_comments	0.100000	0.840000	1.000000	0.040000	0.060000	0.060000	-0.010000	0.070000	-0.030000	-0.000000
is_asshole	0.090000	-0.010000	0.040000	1.000000	-0.020000	-0.020000	-0.000000	-0.030000	-0.050000	-0.010000
countI	0.140000	0.050000	0.060000	-0.020000	1.000000	0.640000	0.200000	0.800000	-0.070000	-0.020000
countHeSheThey	0.120000	0.060000	0.060000	-0.020000	0.640000	1.000000	-0.350000	0.900000	-0.060000	0.060000
IvsHeSheThey	-0.000000	-0.020000	-0.010000	-0.000000	0.200000	-0.350000	1.000000	-0.150000	-0.000000	-0.080000
post_word_count	0.150000	0.060000	0.070000	-0.030000	0.800000	0.900000	-0.150000	1.000000	-0.080000	0.070000
questionmarklast	-0.380000	-0.030000	-0.030000	-0.050000	-0.070000	-0.060000	-0.000000	-0.080000	1.000000	-0.010000
bodyreadinglevel	0.000000	-0.000000	-0.000000	-0.010000	-0.020000	0.060000	-0.080000	0.070000	-0.010000	1.000000

One of these variables is WIBTA_AITA, or the question of whether or not someone is more likely to be an asshole using conditional versus present language (“Would I be the asshole” vs “Am I the asshole”). The difference between WIBTA, AITA, and neither is visible in the graph below:



Next, we look at the words that are most likely to predict is_asshole versus the words most likely to predict “not the asshole.”

4. Interesting Words

While there are plenty of words I personally found interesting (the appendix of this report contains a selected list), I'd like to point out just a few of the uniquely interesting ones, from which I allow the reader to write his own story:

First, no words definitively indicated that someone was NTA. The scale was 1 for NTA, and 0 for YTA, on a 100-point scale. The ones I noted were heavily on the NTA side and particularly interesting were *revenge* (0.79), *insecure* (0.70), *included* (0.65), and *unsolicited* (0.65). Apparently, these words suggest contexts within this dataset favorable to the poster. (Note, surprisingly, that *include* is 0.01 – on the opposite side of the scale!)

More neutral, but leaning NTA, we notice *judging* (0.60), *addiction* (0.59), *force* (0.59), *attracted* (0.56). It seems safe to suggest that with judging, there may be a denial ("I'm not judging"), or perhaps a more passive statement ("He's judging me"). Particularly notable is *force* (0.59), which has a very different statistic from *forced* (0.35) and *forcing* (0.39) – which is fascinating.

To be *pc* is pretty neutral (0.51), which is interesting so long as we're not talking about computers, but political correctness; *disagreeing, nsfw, std, lose,* and *cheat* all go right down the middle (0.50), particularly surprising with std, but perhaps more understandable if one is describing whether or not they gave context surrounding their possible std. Finding two random samples, one described calling karma on a friend who had 2 STDs (NTA), and another who failed to properly split up with his FWB (YTA). Makes sense.

Cheated (0.41) and *cheating* (0.32) are both similarly (and strangely) worse than *cheat*, each successively nine points worse than the other.

We slowly enter into firm YTA territory when we see *affair* (0.49), *daddy* (0.46), *shaming* (0.43), *lazy* (0.43), and *sex* (0.39). You probably shouldn't talk to *strangers* (0.42). Even if *god* isn't one of us, He remains as controversial as the *bedroom* and trying to *forgive* (0.38). Whether you *apologize* after an *argument* or not, it ends roughly as badly for you (0.37).

Don't be *racist*… or *vegan* (0.32).

Consent is exactly as fiery as you might expect (0.30), but not as bad as whatever people report happens on or around *Christmas* (0.25). And if you like to talk about demanding *respect*, others being *controlling,* or about who's the *asshole*… well, it's probably you (0.24).

If it deals with a *newborn,* you're (basically) always wrong (0.13). *Abuse* is a *massive* no-no (0.01). *Minor*, be it the noun or the adjective, shafts the storyteller (0.01). Don't even try to be *accommodating* or *noisy* (0.01). And if you have to get to the point where you're *unfriending* or you *block* your friends, and then tell everyone about it… keep it to yourself (0.01).

And, finally, the scale you never knew you needed – the platforms + games where people are most likely to be affiliated with assholery, in descending order:

Tinder 0.42
Reddit 0.38
Instagram 0.36

Minecraft 0.30
Snapchat 0.23
YouTube 0.09
PS4 0.07
Netflix 0.06

You're welcome.

5. Wordcloud

A wordcloud of the titles of each post show the frequency of words outside prepositions, indefinite articles, and so on.

The most common words are nouns – girlfriends, friends, parents. A lot of the verbs are fundamentals of our modern experience: Wanting, telling, asking, letting, refusing. This suggests that many of the conflicts in our modern lives really do come down to clashes of what people want, what people say, what people consent to (or don't). (We recall that there were once worlds in which conflicts centered around beliefs, needs, sacrifices, enemies.) Interestingly, emotions aren't as present (upset, getting mad, and annoyed are tiny). Tragically, cutting is a word that occurs with enough frequency to be present in the wordcloud (it's bigger than "cheating", and only slightly smaller than "making").

6. Modeling

While I played around with accuracy as a metric (able to get as high as 72%), It seems to me that accuracy is the wrong metric for this dataset. Our goal is recall, since it predicts how many true positives there are over the entire set of assholes – and that's exactly what we want to know: How many assholes did we properly predict would be assholes, based on our features?

The model metrics and models I opted for are in the following table:

Model	Recall	Precision	Accuracy	F1
Decision Tree Classifier	0.540	0.555	0.551	0.669
Random Forest Classifier	0.559	0.564	0.561	0.671
Logistic Regression	0.546	0.571	0.565	0.670
Gradient Boosting	0.549	0.559	0.555	0.669

We see that the best model for recall was the Random Forest classifier, at 0.559 - I opted for a max depth of 12. It is second to the Logistic Regression for both Precision and Accuracy, the latter scoring a 0.571 and 0.565. These are mild but notable gains. All models fared only slightly better than chance.

More investigation into the dataset is needed, to continue searching for superior features.

7. Most Helpful Features in Prediction

For the Decision Tree, the most helpful features were *edited* (31.7%), *parents* (25%), *and post_word_count* (8.4%).

For the Random Forest, it was *post_word_count* (8.7%), *bodyreadinglevel* (8.3%), and *loverPostCount* (8.1%).

Logistic Regressions don't have a feature importance method.

For Gradient Boosting, we see it was *edited* (20.1%), *parents* (9.8%), and *relatives* (5.5%).

8. Future Improvements

We can group individual words by their part of speech, which I have left as a separate Dataframe for others to fool around with. Certainly other ways to examine tonality are interesting, as poster considered arrogant might get less sympathy from the community. Being able to examine and weigh comments based on their scores, and the awards they get, might help us put YTA/NTA on a spectrum, allowing us to see it as the more fluid subject that it is. Grouping the titles of posts by more nuanced commonalities than the three headings I created will likely bear more fruit. One could also run handpicked posts through the model and see if the model agrees with them. It's always cool to have a model useful to an individual's moral standards.

9. Acknowledgements

Major thanks to Benjamin Bell, my Springboard mentor, for his vision and help with this project; to Kenneth Gil-Pasquel for his help in cleaning up my work; to the AITA community for this wonderfully-rich data, from which we discovered some tremendous insights.

Appendix: Interesting Words

revenge 0.79	force 0.59	pc 0.51	daddy 0.46
ass 0.78	suggested 0.59	disagreeing 0.50	gfs 0.44
vs 0.75	revealing 0.58	nsfw 0.50	hiding 0.44
insecure 0.70	attracted 0.56	cheat 0.50	shaming 0.43
failing 0.69	dated 0.56	std 0.50	accused 0.43
purpose 0.68	underage 0.55	lose 0.50	faking 0.43
cussing 0.68	bfs 0.54	affair 0.49	scolding 0.43
embarrassed 0.66	yes 0.54	exposing 0.49	replacement 0.43
dumb 0.66	proposal 0.52	fat 0.49	wife 0.43
included 0.65	win 0.52	straight 0.48	criticizing 0.43
unsolicited 0.65	price 0.52	teenage 0.48	lazy 0.43
intentionally 0.64	owed 0.52	casual 0.47	tinder 0.42
failed 0.63	hug 0.52	experience 0.47	daughters 0.42
policy 0.62	valentine 0.52	flirting 0.47	question 0.42
wont 0.62	system 0.52	destroyed 0.47	strangers 0.42
supervisor 0.61	spoiled 0.52	compensation 0.46	worried 0.42
judging 0.60	anonymously 0.52	joint 0.46	age 0.42
please 0.60	bi 0.52	player 0.46	career 0.42
addiction 0.59	nudes 0.51	favorite 0.46	excited 0.42

grandmothers 0.42	anniversary 0.38	pregnant 0.35	workplace 0.32
insulting 0.42	forgive 0.38	transgender 0.35	adopted 0.32
professional 0.41	inappropriate 0.38	dick 0.35	cheating 0.32
babies 0.41	ditching 0.38	gf 0.35	women 0.32
daughter 0.41	sleepover 0.37	race 0.35	racist 0.32
gross 0.41	rejecting 0.37	forced 0.35	vegan 0.32
cheated 0.41	playing 0.37	poop 0.35	sisters 0.32
learn 0.41	girls 0.37	waiter 0.35	breakfast 0.32
gay 0.41	abortion 0.37	beliefs 0.35	addicted 0.32
muslim 0.41	marriage 0.37	romantic 0.35	nude 0.32
entitled 0.41	classmates 0.37	host 0.35	bullying 0.31
showed 0.41	teacher 0.37	bachelor 0.34	exes 0.31
pics 0.41	claims 0.37	sexist 0.34	prom 0.31
kissed 0.40	full 0.37	bridal 0.34	ex 0.31
laughing 0.40	apologize 0.37	bachelorette 0.34	fuck 0.31
girlfriends 0.40	argument 0.37	bestfriend 0.34	slept 0.31
friendships 0.39	doctor 0.36	stepdaughter 0.33	employees 0.31
married 0.39	risk 0.36	assholes 0.33	stepson 0.31
expose 0.39	debt 0.36	naked 0.33	speaking 0.31
joking 0.39	instagram 0.36	son 0.33	waiting 0.30
forcing 0.39	sexuality 0.36	people 0.33	consent 0.30
sex 0.39	pregnancy 0.36	stepmother 0.33	bridesmaid 0.30
girl 0.38	elderly 0.36	professor 0.33	minecraft 0.30
reddit 0.38	worker 0.36	smoking 0.33	medical 0.30
boys 0.38	office 0.36	teen 0.33	younger 0.30
sperm 0.38	brothers 0.35	female 0.33	colleague 0.30
woman 0.38	girlfriend 0.35	date 0.33	therapy 0.30
god 0.38	fiancee 0.35	disability 0.33	artist 0.30
bedroom 0.38	porn 0.35	love 0.32	drama 0.29

gender 0.29	delivery 0.25	bridesmaids 0.24	job 0.22
sibling 0.29	christmas 0.25	spouse 0.24	sales 0.22
child 0.29	push 0.25	ultimatum 0.23	cancer 0.22
bully 0.29	roommates 0.25	puppy 0.23	estranged 0.22
breakup 0.29	stepfather 0.25	brand 0.23	children 0.22
teenager 0.29	fireworks 0.25	animal 0.23	colleagues 0.22
offended 0.27	death 0.25	snapchat 0.23	religious 0.21
falling 0.27	military 0.24	hookup 0.23	fucking 0.21
workers 0.27	respect 0.24	dude 0.23	consulting 0.21
divorced 0.27	lesbian 0.24	bf 0.23	nephews 0.21
boyfriend 0.27	disgusting 0.24	coworker 0.23	employer 0.21
texts 0.27	trusting 0.24	babysitter 0.23	dad 0.21
trans 0.26	controlling 0.24	flatmates 0.23	boss 0.20
kids 0.26	boyfriends 0.24	bff 0.23	father 0.20
strip 0.26	parenting 0.24	easter 0.23	relationships 0.20
uncle 0.26	celebrate 0.24	neighbour 0.23	police 0.20
husband 0.26	religion 0.24	funny 0.23	space 0.20
guests 0.26	partner 0.24	bil 0.23	foster 0.20
roomate 0.26	school 0.24	deaf 0.23	tattoos 0.20
band 0.26	asshole 0.24	cousin 0.23	stepdad 0.20
neighbors 0.26	babysitting 0.24	management 0.23	family 0.20
overweight 0.26	neighbor 0.24	work 0.22	grandfather 0.20
grudge 0.26	roommate 0.24	single 0.22	bosses 0.20
brother 0.26	aunts 0.24	atheist 0.22	divorce 0.20
guy 0.26	autistic 0.24	raise 0.22	marry 0.20
sexual 0.25	aunt 0.24	housemate 0.22	parents 0.20
anxiety 0.25	funeral 0.24	nye 0.22	men 0.20
resulting 0.25	party 0.24	landlord 0.22	depression 0.20
dorm 0.25	cops 0.24	working 0.22	church 0.20

bitch 0.19	mother 0.18	families 0.15	ps4 0.07
company 0.19	mum 0.18	husbands 0.15	netflix 0.06
cousins 0.19	friendship 0.18	babysit 0.14	abuse 0.01
community 0.19	blocked 0.17	grandmother 0.14	noisy 0.01
parent 0.19	mil 0.17	newborn 0.13	carry 0.01
world 0.19	alcoholic 0.17	mothers 0.13	minor 0.01
privacy 0.19	inlaws 0.16	bro 0.13	accommodating 0.01
holding 0.19	grandpa 0.16	fwb 0.13	adopting 0.01
position 0.19	boy 0.16	snitching 0.12	include 0.01
home 0.19	grandma 0.16	relatives 0.12	unfriending 0.01
mom 0.19	nephew 0.15	vasectomy 0.11	block 0.01
snow 0.19	purchase 0.15	stepmom 0.11	
customers 0.19	whose 0.15	doctors 0.10	
handicapped 0.19	boundaries 0.15	baseball 0.09	
fathers 0.18	housemates 0.15	youtube 0.09	
tattoo 0.18	owners 0.15	miscarriage 0.09	