

Seinfeld: A project about nothing...

By David Sutton

More accurately: The George Costanza fake
occupation prediction engine



Data science problem:

Can we predict which subreddit a post is from the architecture or marine biology subreddit?

George's two favorite fake occupations.



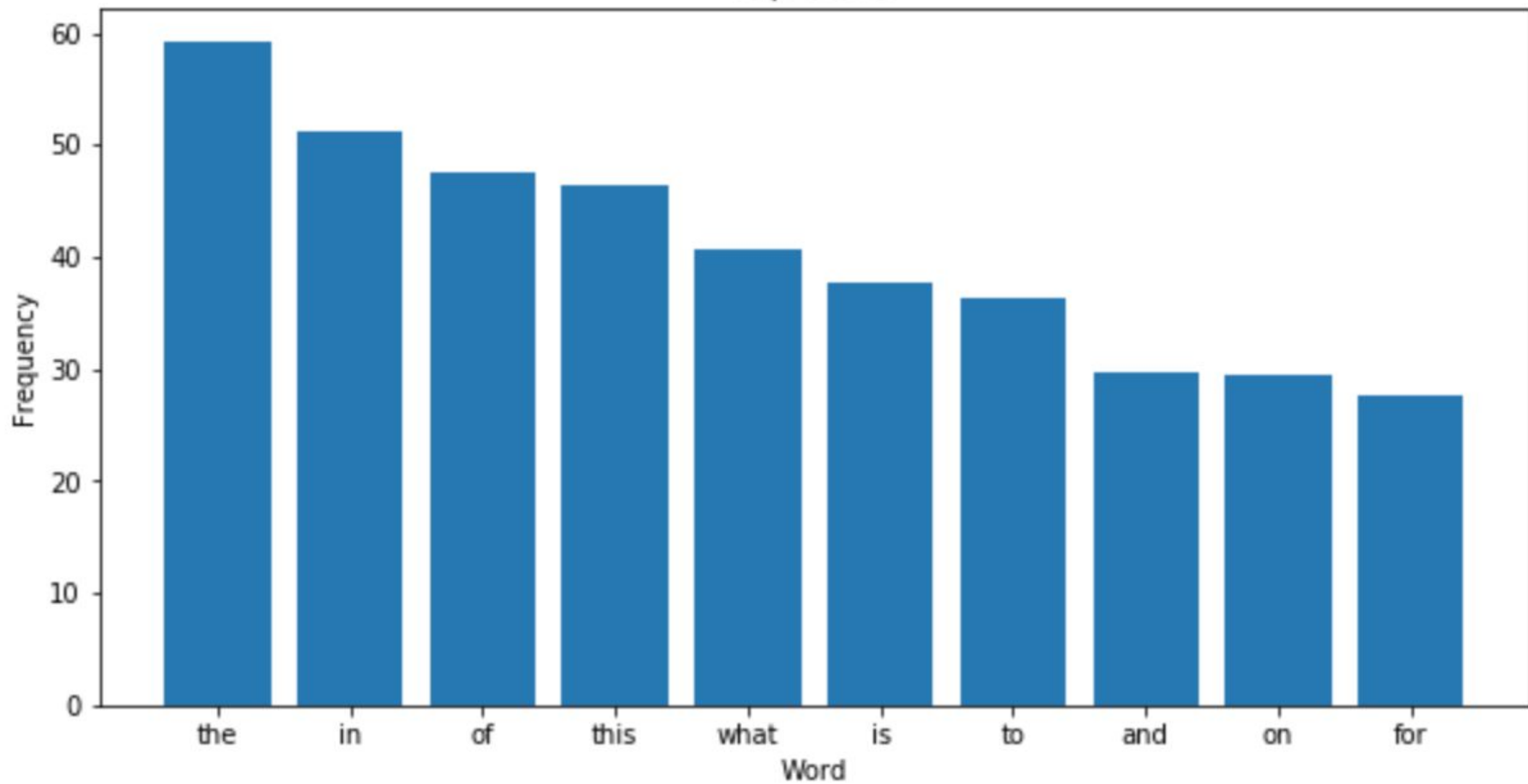
Seinfeld (still) in the news this week:

- National Law Review - [Jerry Seinfeld's Ex-Partner Time Barred in Copyright Dispute Over "Comedians in Cars Getting Coffee"](#)
- PopMatters - [Seinfeld Was Wrong About Keith Hernandez's Spit-Shot, Possibly](#)
- Showbiz Cheat Sheet - ['Seinfeld': Why Susan Ross Had to Be Killed Off on the NBC Hit Comedy](#)
- Screen Rant - [Seinfeld: 10 Best Callbacks Throughout The Series](#)
- Showbiz Cheat Sheet - ['Seinfeld': Jason Alexander on What Makes George Costanza One of the Greatest TV Sitcom Characters of All Time](#)
-

Marine biologist or architect?



Top Words



	coefs	vals
658	practice	1.973410
180	city	2.031105
207	construction	2.133609
403	house	2.263072
234	design	2.455809
64	architect	2.478609
823	style	2.524145
65	architects	2.551391
139	building	2.848882
67	architecture	3.911863

	coefs	vals
514	marine	-3.538680
313	fish	-2.861088
732	sea	-2.350175
748	shark	-2.165473
941	whale	-2.061453
327	found	-1.975851
588	octopus	-1.920220
700	research	-1.910851
585	ocean	-1.816282
269	eel	-1.741440

It's not hard to tell the subreddits apart...



Conclusion:

There is little difficulty distinguishing the two subreddits, for machine or human.



Next steps:

- Troubleshoot coding problems (ran out of time/mojo)
- Work in a multinomial Naive Bayes model (ran out of time)
- Add visualizations for post length and bi grams (ran out of time)
- Share some humorous post titles discovered on this project (ran out of time)
- Make a stronger project (ran out of time)
- Start earlier (no excuses)
- Make a project that better reflects the time and effort spent

Thanks for your time!
And George, you're welcome!



Image Credits:

Photos: [Slide 2](#), [Slide 3](#), [Slide 4](#), [Slide 5](#), [Slide 6](#), [Slide 7](#), [Slide 8](#), [Slide 9](#), [Slide 10](#),
[Slide 11](#)

