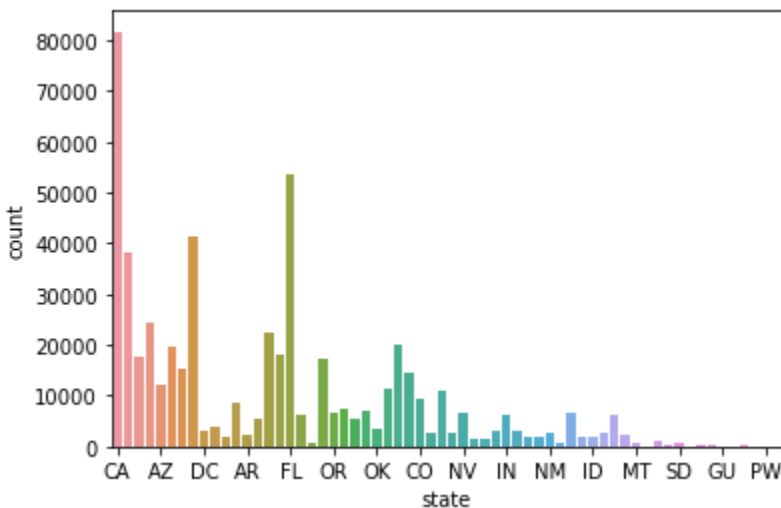


Jason Zalewski

Consumer Disputes

From the data we were given we see we collected almost all of our data from a few of the states. With the biggest players including California and Florida. For building future models or looking for disputes to avoid we should be weary of taking information from states we don't have a good sample size of information on.



I found a simple NLP model that allows sentence relations to be measured with each other. I don't have enough evidence to say the model has capability of finding high correlations between consumer complaints, but I think its potential is worth exploring in the future.

I showed simple unrelated sentences just so it's easy to think about what the models are doing. And the last example is taking two sentences randomly from our dataset.

Taking From Hugging Face A Sentence Transformer Model used to find similarities between sentences

To show relation between the Happy Sentences

Sentences = ["I'm Happy", "I'm full of happiness"]

Relation = 60%

To show relation between a Happy and Mad Sentence

Sentences = ["I'm Happy", "I Don't feel so good today"]

Relation = 0.32%

Will now Attempt to look at the similarities of two sentences randomly in the dataset

consumer_complaint_narrative1 = "It seems like something is really wrong with these prepaid gift cards that makes it so easy for fraudulent transactions while we still have possession of the card."

consumer_complaint_narrative2 = "I asked for proof of a signed contract, I asked for a license to collect in my state, I asked for copies of all information referenced for this debt and still to date, I have not received anything but harassment from this company!"

Relation = 28%

The models will allow us to see relations between sentences. Given enough samples in our data, it can help us find what disputes to avoid.

