DS501 Project 1

What motivates reviewer behavior?

MIA BARGER ▪ QUINCY HERSHEY ▪ ALEXANDER MOORE ▪ ETHAN PRIHAR

https://www.yelp.com/dataset

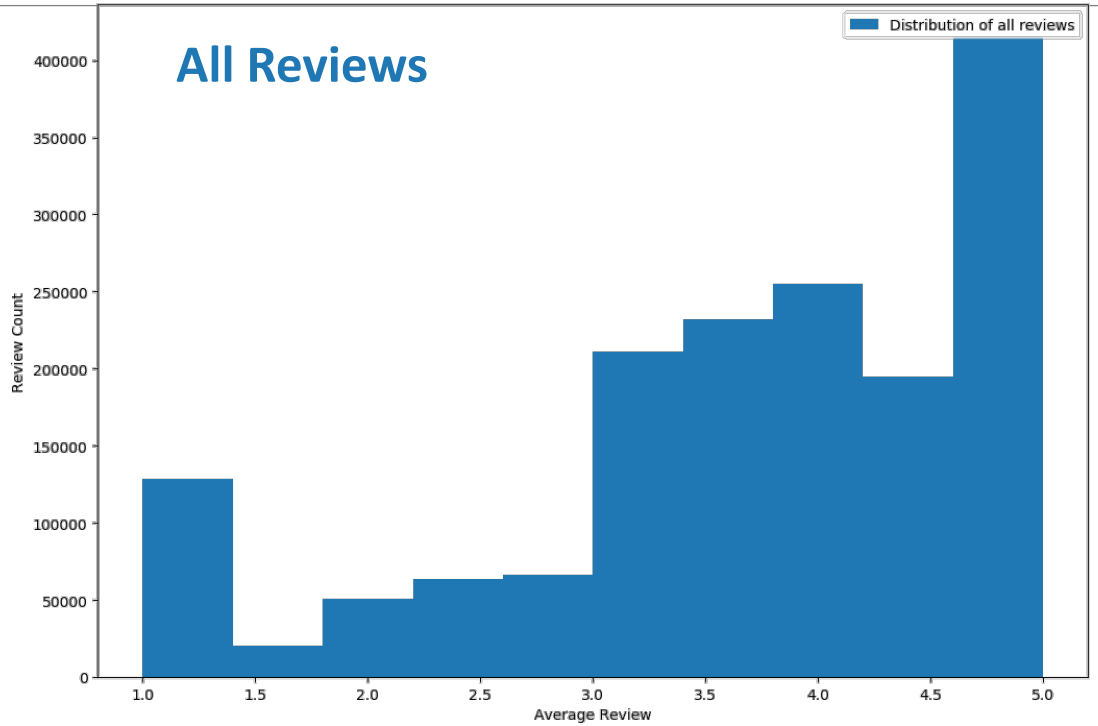
Problem introduction + motivation

Is the Yelp reviewer base homogenous or reflecting distinct user populations and behaviors?

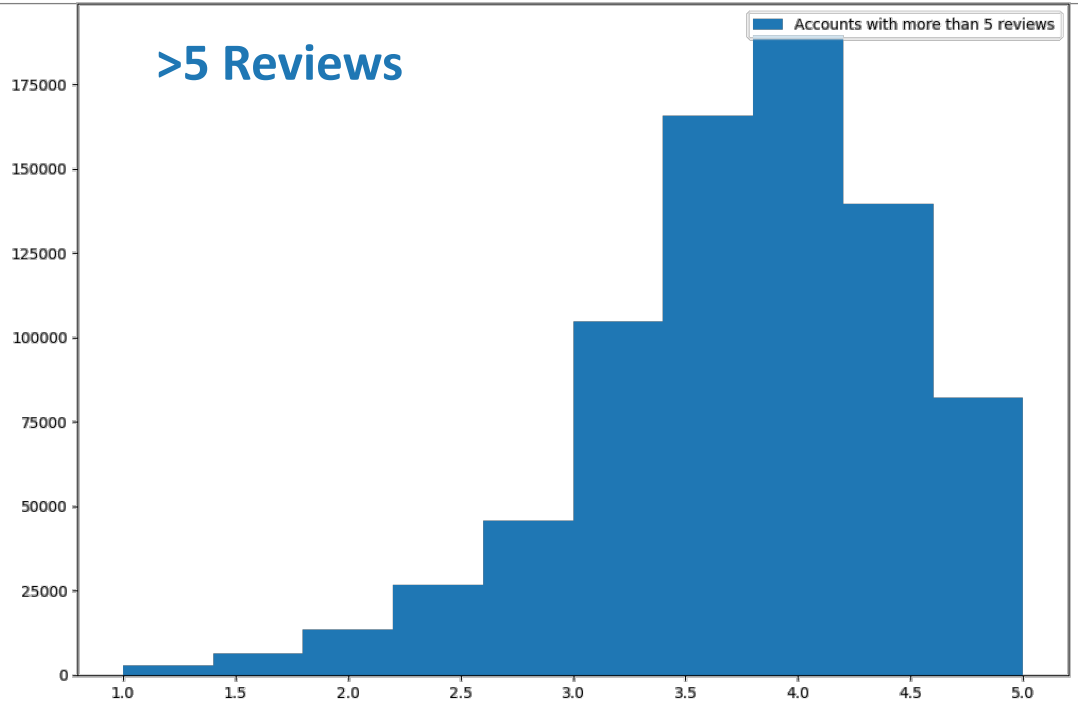
What do distinct subsets to the user base tell us about review bias and account creation?

What factors drive someone to make a Yelp account to leave a single negative review?

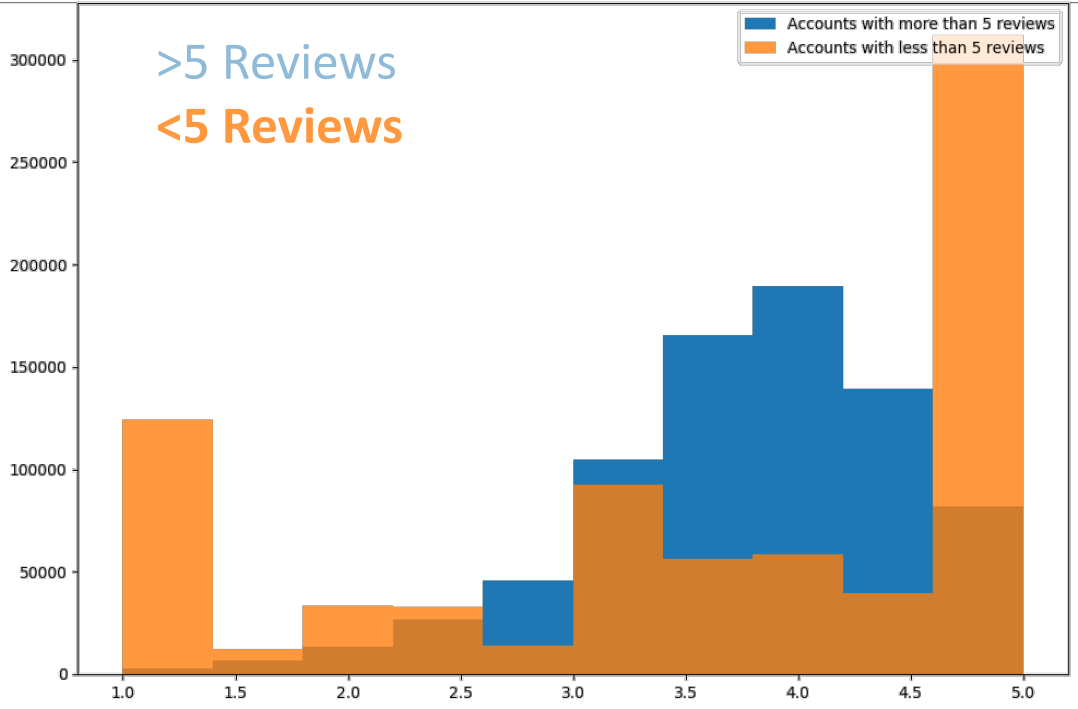
Exploring review distributions



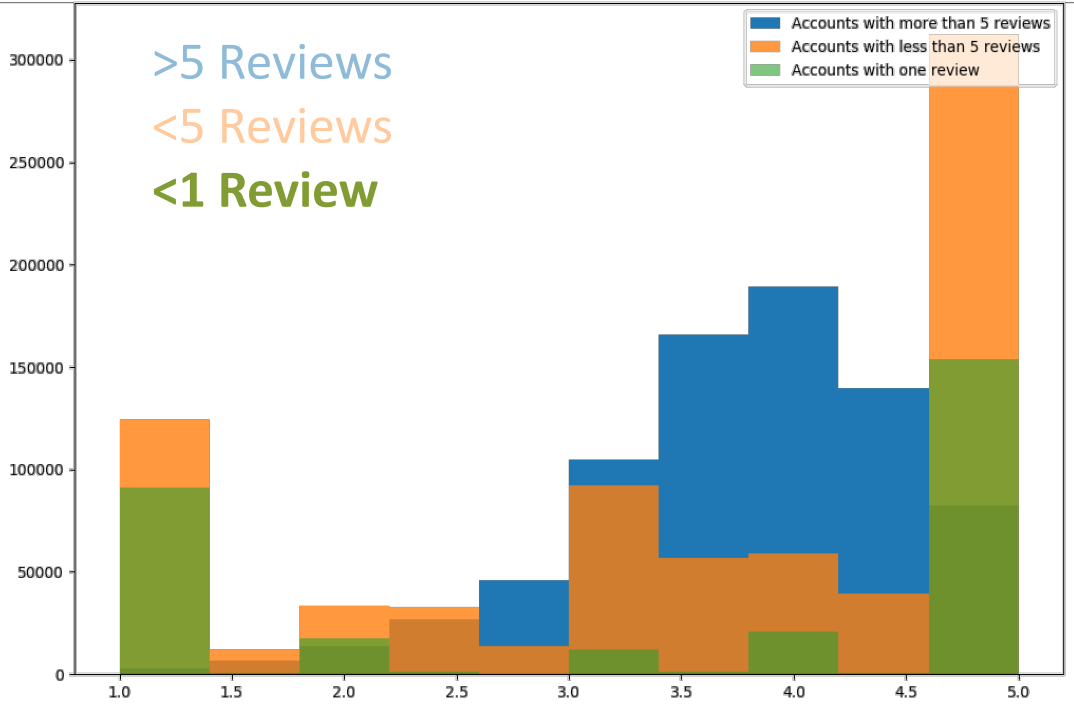
Exploring review distributions



Exploring review distributions



Exploring review distributions



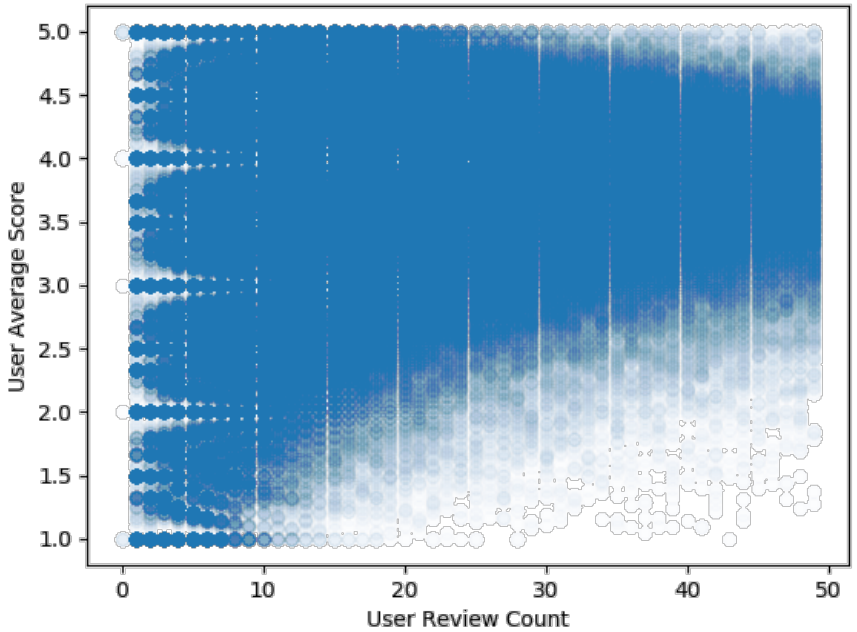
Initial takeaways

Reviews for users >5 reviews ~N(4,?)

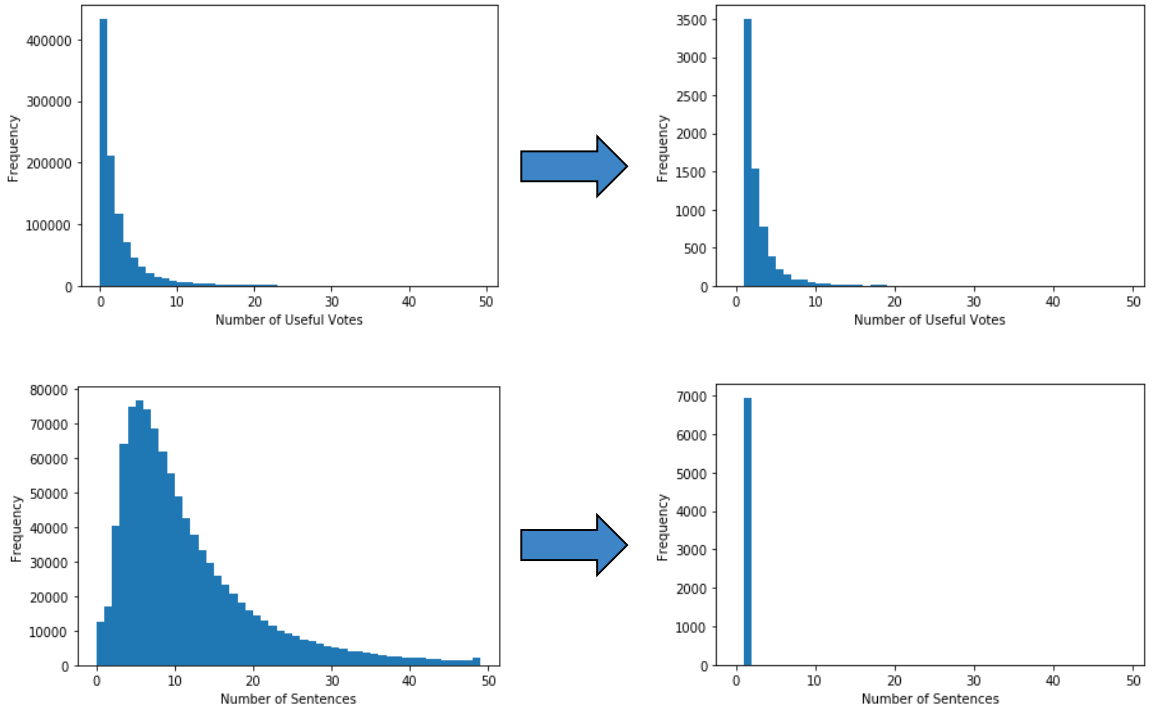
As history declines, distribution → barbell

- Reviews may have silent majority bias
- Experiences drive account creation
- Reviews are driven by tail behavior

What motivates 1 Review, 1 Star users?



Step 1: Identify a relevant subset



Step 2: Extract review features

Embedding: A Neural Network Turns A List of Words into a Vector of Numbers

What the hell happened to this place, the new menu sucks, the size of servings shrunk, and the quality of the food is uber mediocre. Will not come anymore

me and my coworker waited for over an hour for just 2 bowl of phos. and wheb we asking waitress about our orders the waitress just walked away

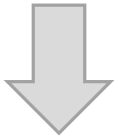
[0.00746889 -0.04002672 0.11342745 ... 0.00595205 0.07626151 -0.00635999]

[0.00746889 -0.01814748 0.04581 ... 0.02196765 0.08440708 -0.00887392]

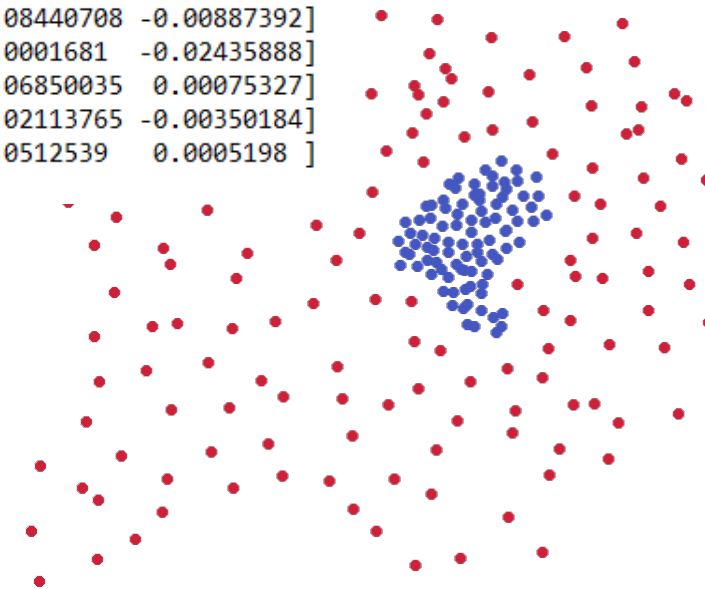
Step 3: Group similar reviews

DBSCAN: A clustering algorithm that clusters based on density

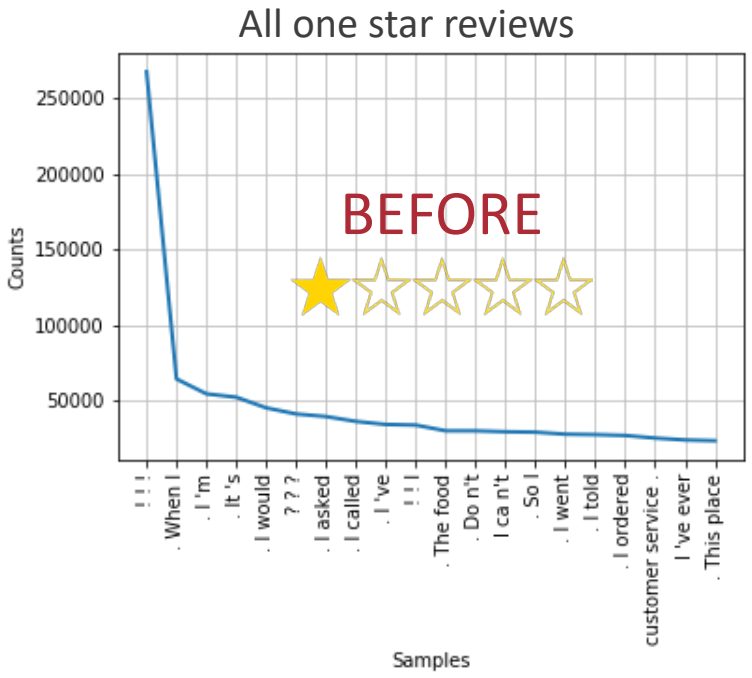
```
[ 0.00746889 -0.04002672  0.11342745 ...  0.00595205  0.07626151 -0.00635999]
[ 0.00746889 -0.01814748  0.04581    ...  0.02196765  0.08440708 -0.00887392]
[ 0.00746889 -0.06945133  0.10283796 ... -0.00449271 -0.0001681  -0.02435888]
[ 0.00746889 -0.00762606  0.08259026 ...  0.03554407  0.06850035  0.00075327]
[ 0.00746889 -0.07472447  0.03286474 ...  0.01323174  0.02113765 -0.00350184]
[ 0.00746889 -0.02781818  0.14583996 ...  0.03921724  0.0512539   0.0005198 ]
```



```
[-1]
[-1]
[ 0]
[ 0]
[-1]
[-1]
```

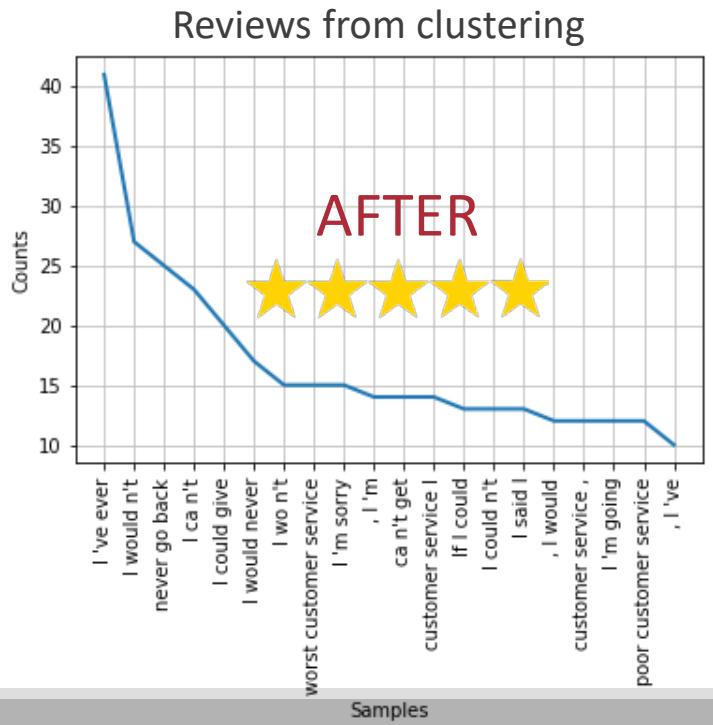


Step 4: Identify Common Phrases



! ! !
. When I
. I ’m
. It’s
. I would
? ? ?
. I asked
. I called
. I ’ve
! ! I
. The food
. Do n’t
I ca n’t
. So I
. I went
. I told
. I ordered
customer service.
I’ve ever
. This place

Step 4: Identify Common Phrases



I ‘ve ever
I would n’t
never go back
I ca n’t
I could give
I would never
I wo n’t
worst customer service
I ‘m sorry
, I ‘m
ca n’t get
customer service I
If I could
I could n’t
I said I
, I would
customer service ,
I ‘m going
poor customer service
, I ‘ve

Conclusion and business implications

Make sure you have good customer service. Tail events drive reviews.

Is Yelp's "Don't Ask for Reviews" policy fighting one bias with another?

Are there better ways to incentivize account creation and cope with free riders?

"Your most unhappy customers are your greatest source of learning"

~Bill Gates