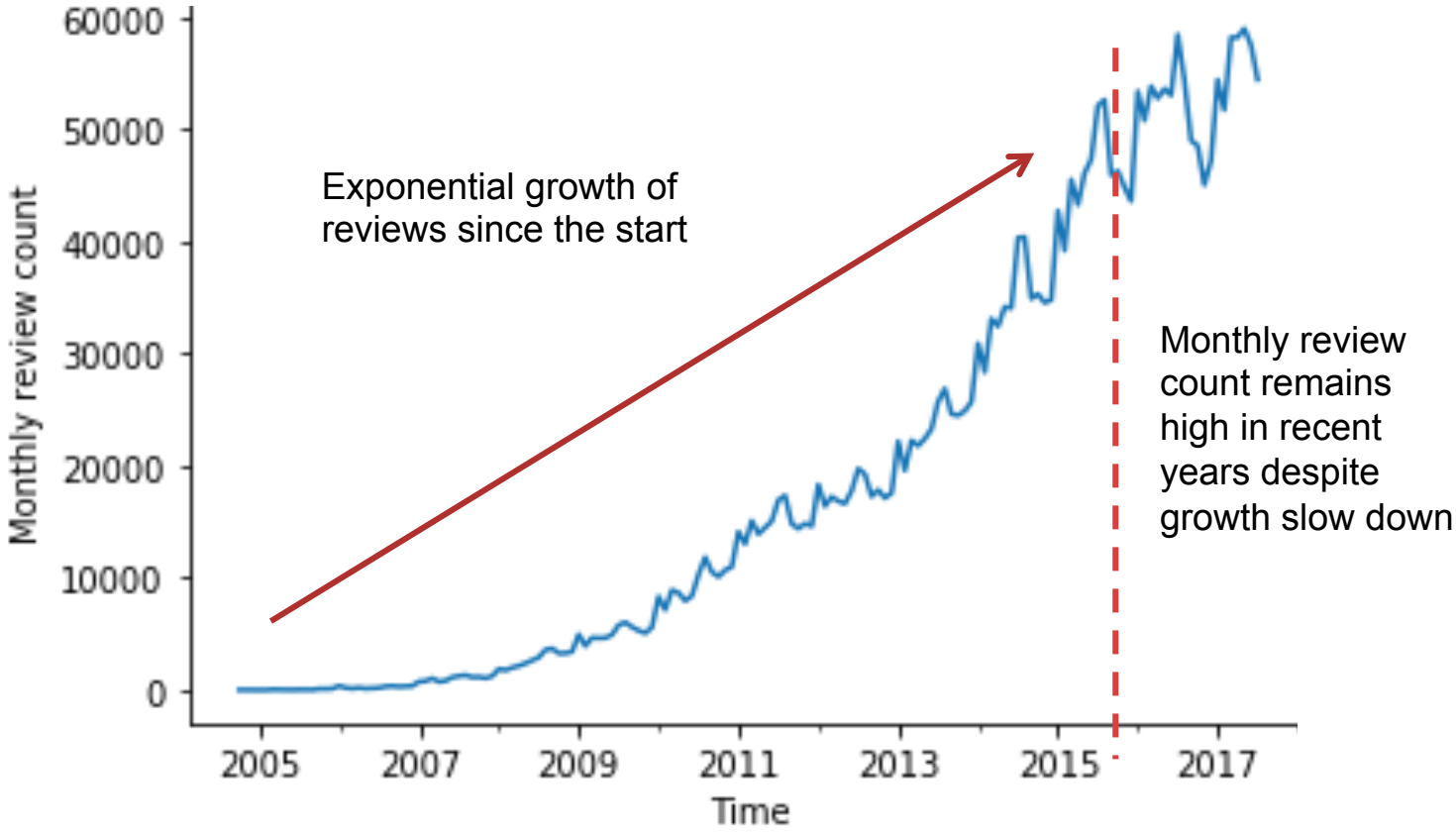
Helping Yelp

SPRINGBOARD - CAPSTONE PROJECT 1

DATA STORYTELLING

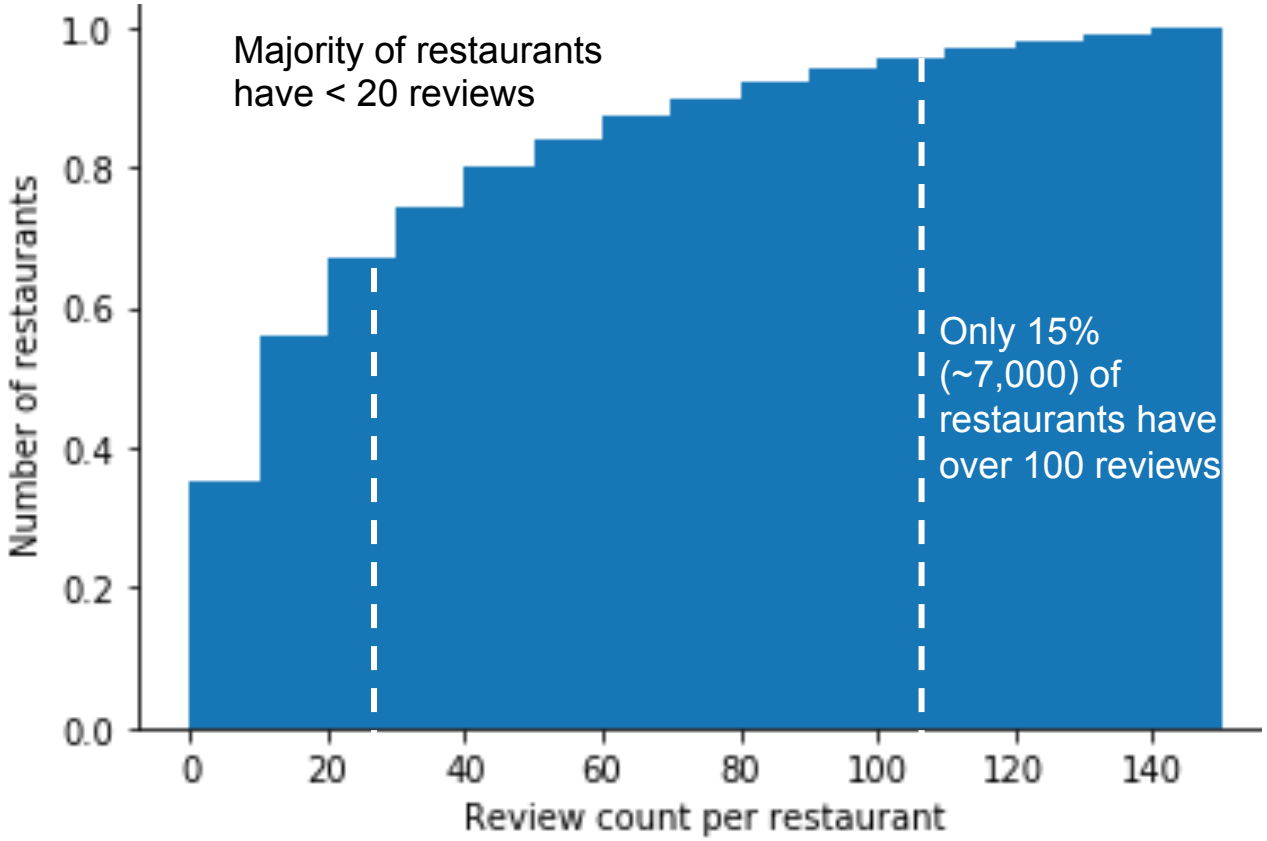
Yelp Reviews continue to be relevant in 2017

MONTHLY RESTAURANT REVIEW COUNTS FOR DATASET



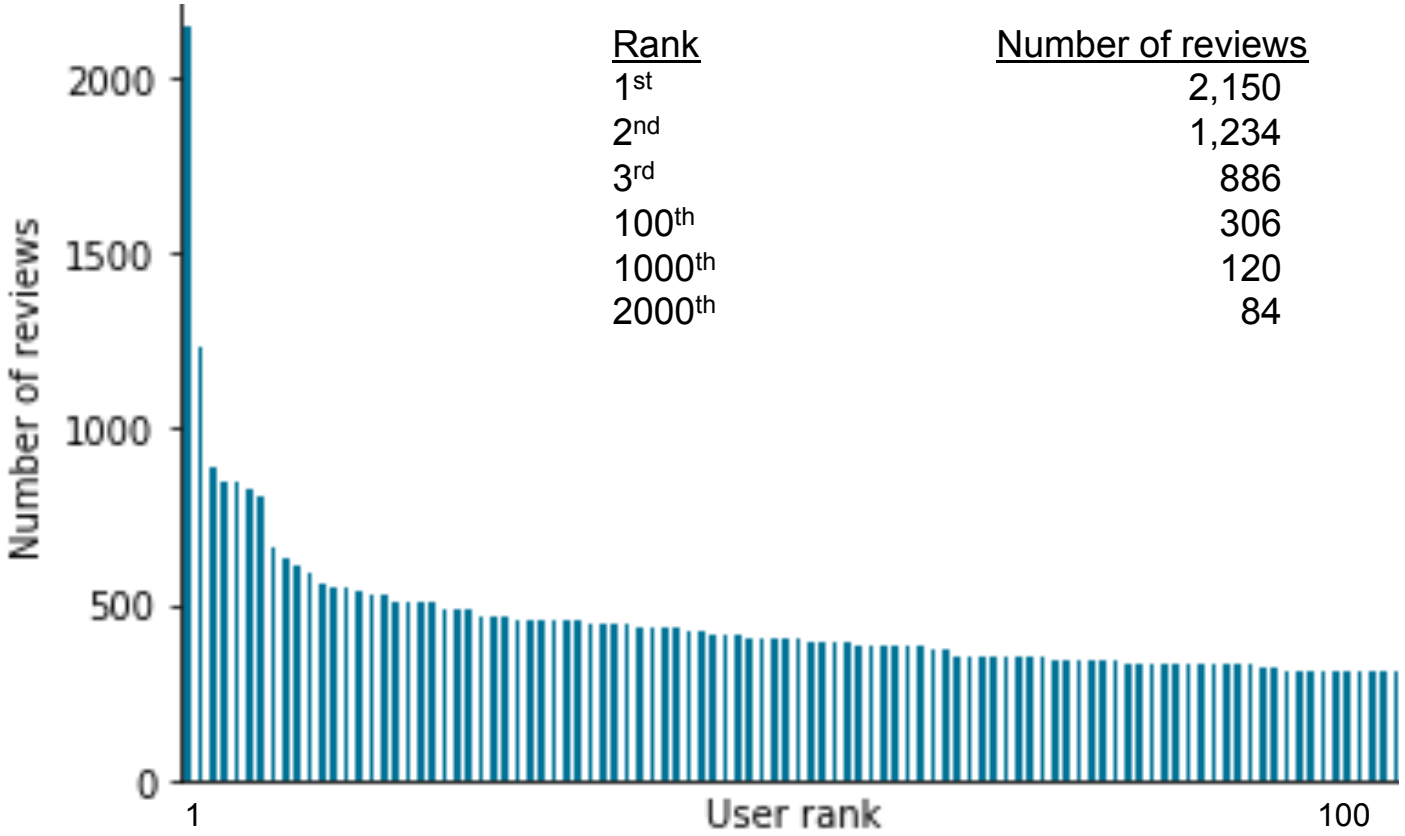
A single review can greatly impact most restaurants

CUMULATIVE DISTRIBUTION OF REVIEW COUNTS



Although the top users are prolific reviewers...

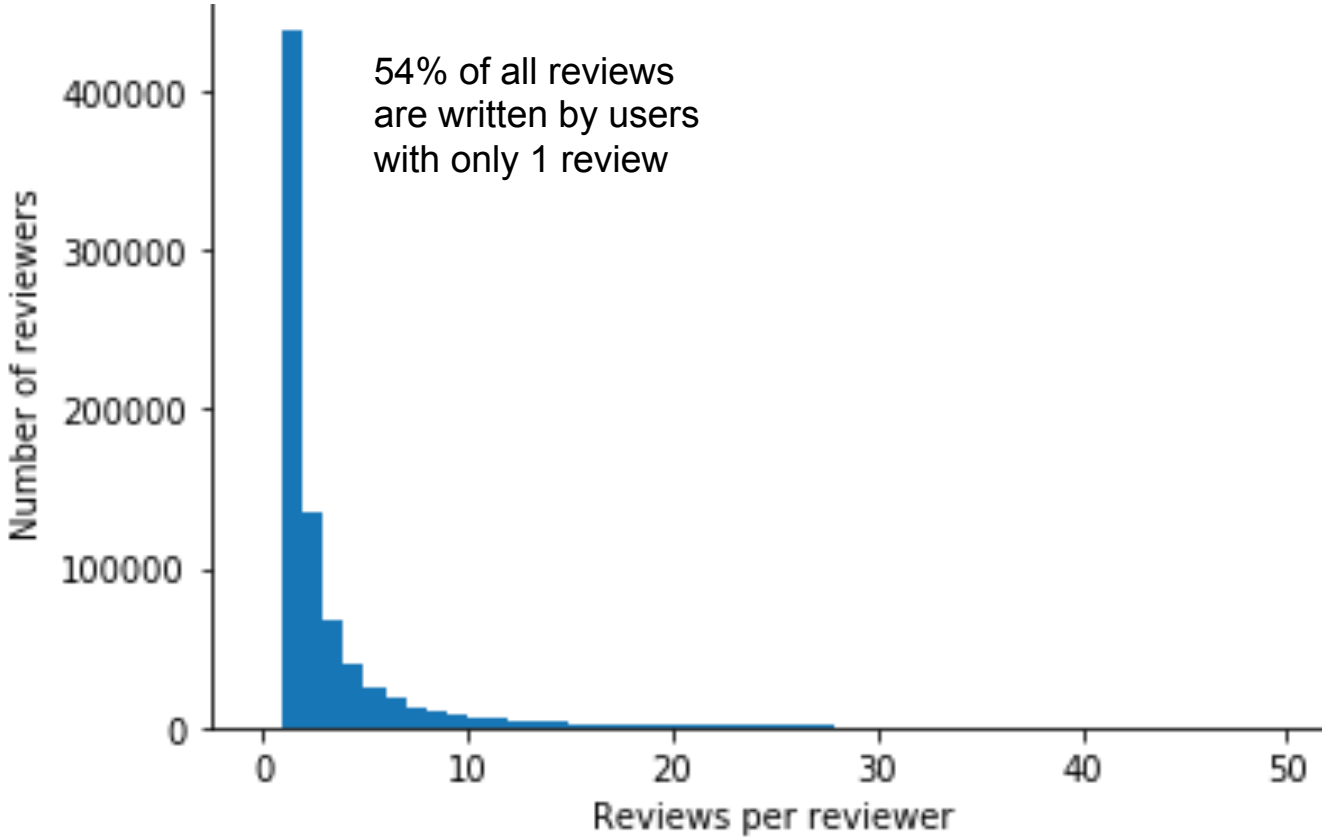
NUMBER OF REVIEWS WRITTEN BY THE TOP REVIEWERS (EACH)



Rank	Number of reviews
1st	2,150
2nd	1,234
3rd	886
100th	306
1000th	120
2000th	84

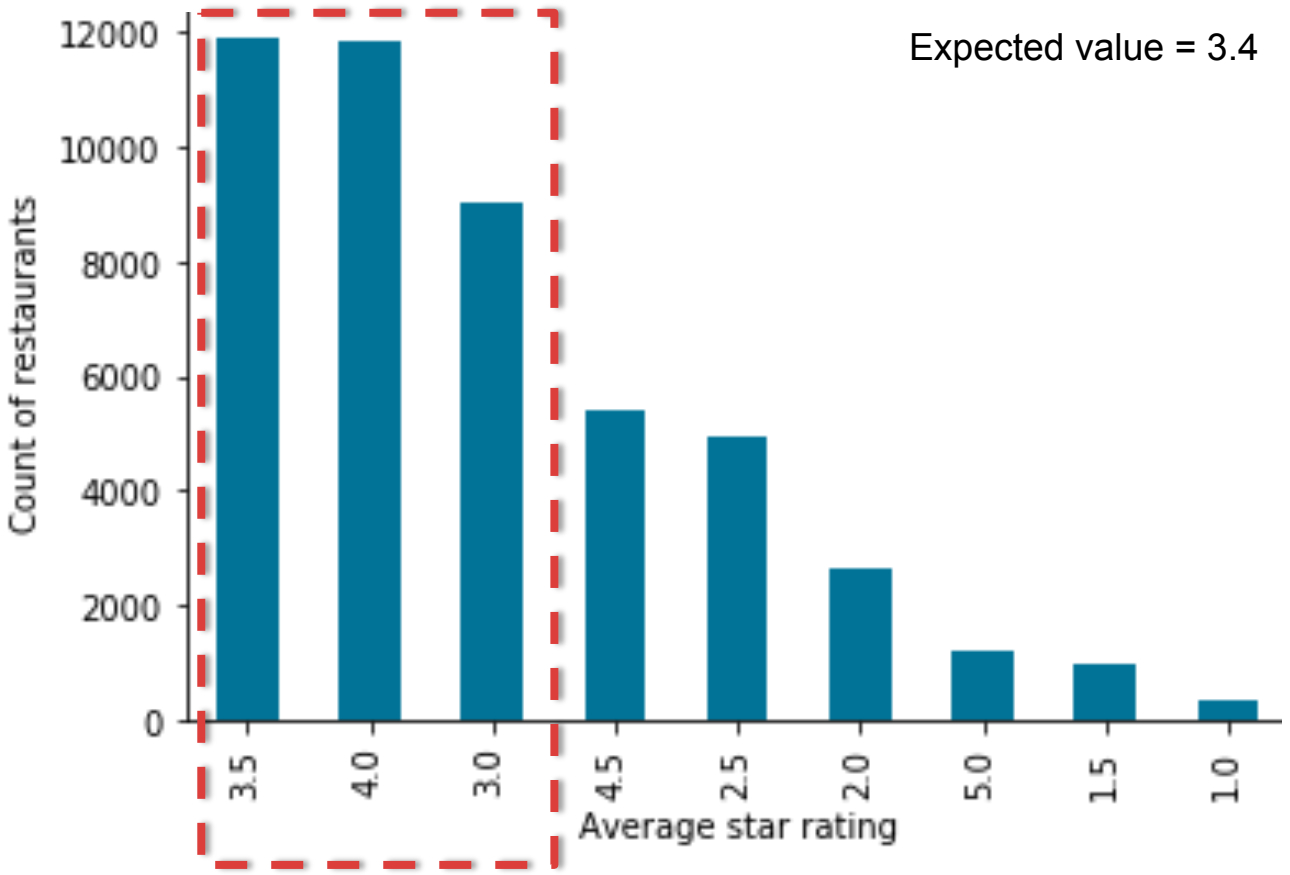
… they only write a small proportion of reviews, most reviews are written by first timers

DISTRIBUTION OF RESTAURANT REVIEWS PER USER



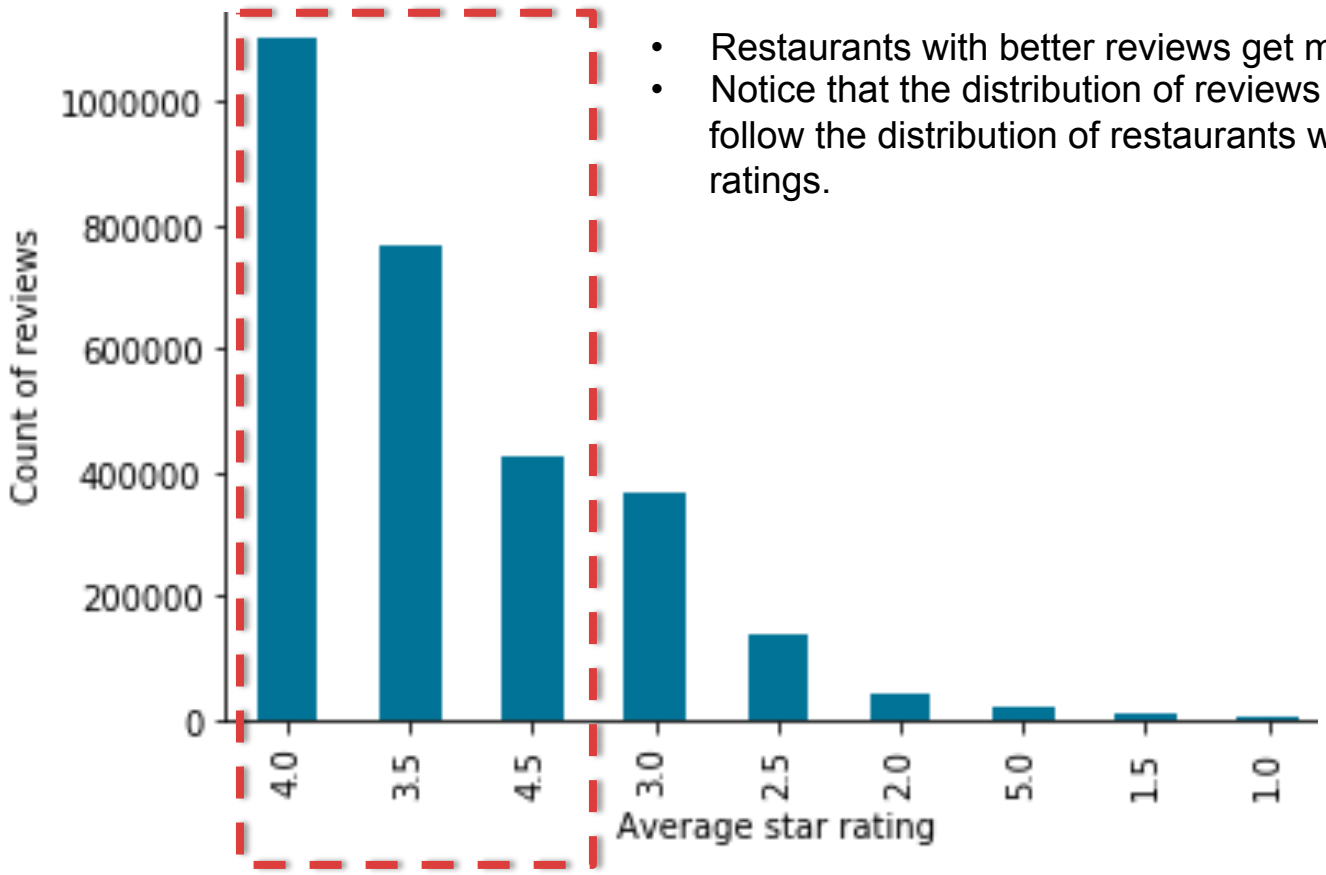
On average, restaurants are rated fair to good (3-4 stars)

DISTRIBUTION OF AVERAGE STAR RATINGS FOR RESTAURANTS



Restaurants with better ratings tend to get more reviews

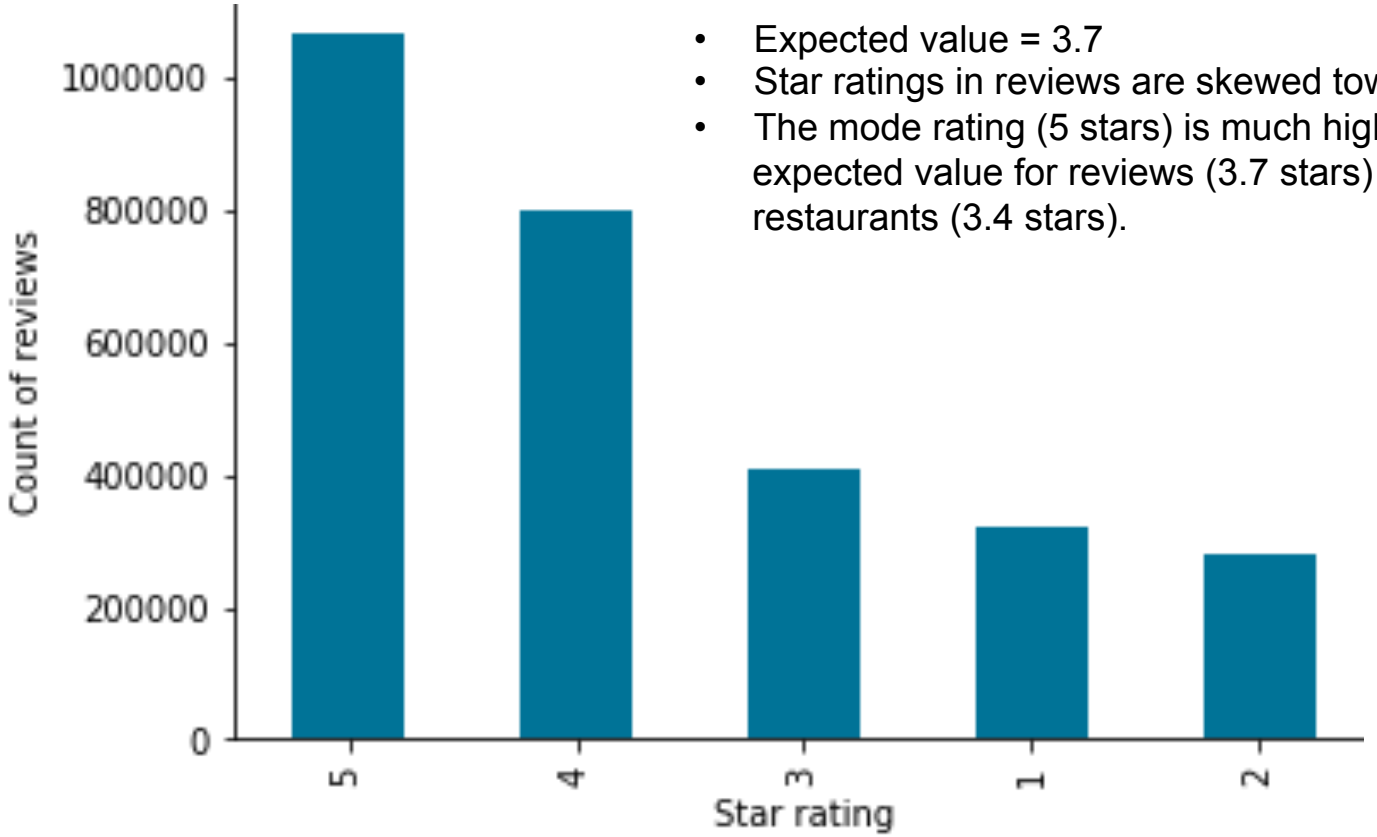
DISTRIBUTION OF REVIEWS BY RESTAURANT'S AVERAGE STAR RATING



- Restaurants with better reviews get more reviews.
- Notice that the distribution of reviews does not follow the distribution of restaurants with those ratings.

There is no obvious ‘representative’ star rating that will be fair to restaurants

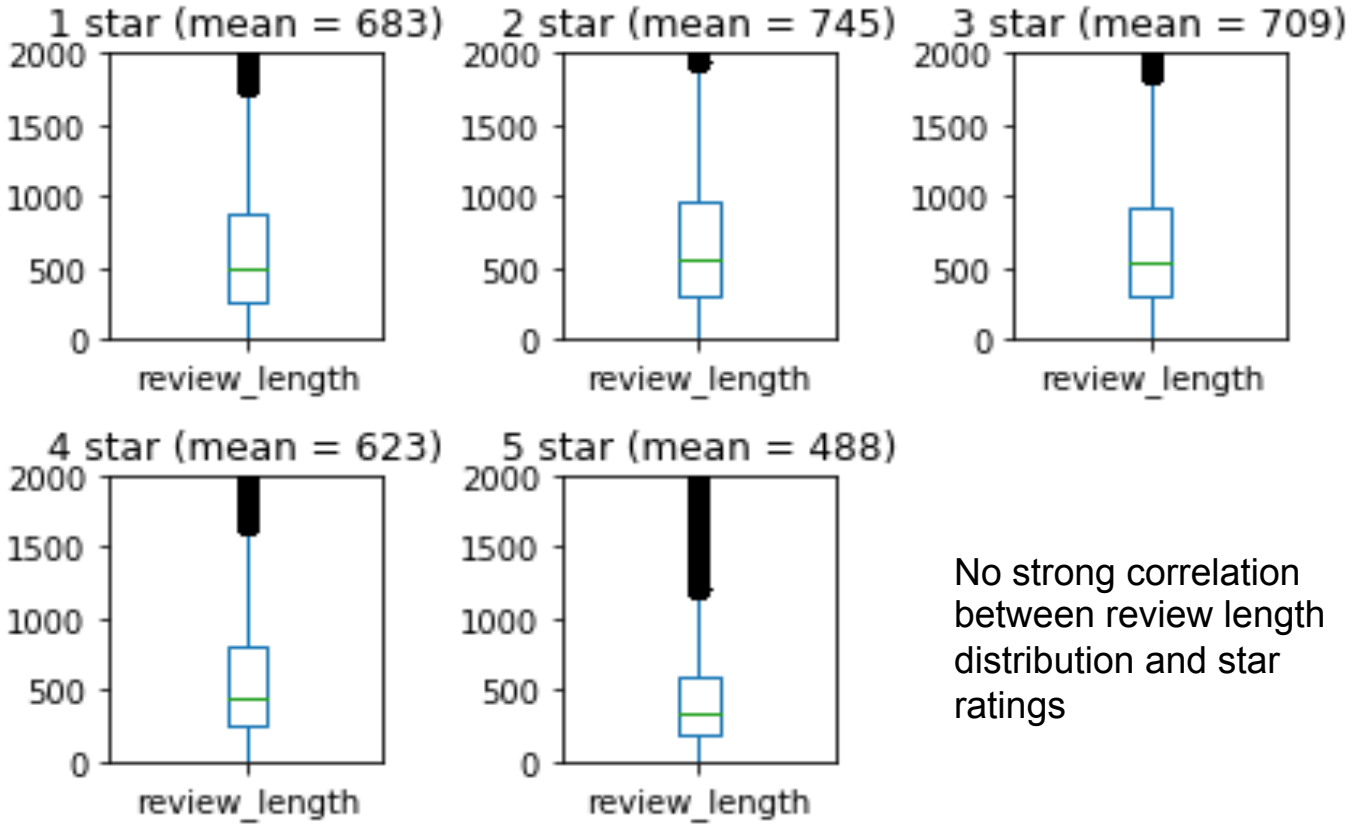
DISTRIBUTION OF STAR RATINGS OF REVIEWS IN THE DATASET



- Expected value = 3.7
- Star ratings in reviews are skewed towards high ratings.
- The mode rating (5 stars) is much higher than the expected value for reviews (3.7 stars) and for restaurants (3.4 stars).

The length of a review appears to be a poor proxy indicator of star ratings

DISTRIBUTION OF NUMBER OF CHARACTERS IN A REVIEW BY STARS



No strong correlation between review length distribution and star ratings

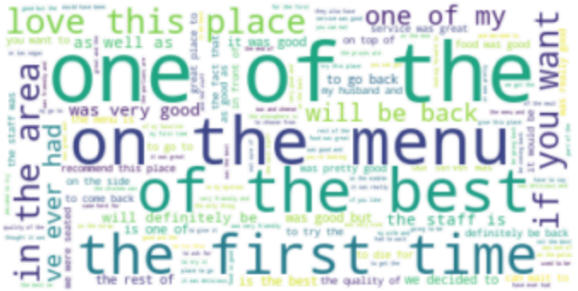
Analysis of the text give might provide a fair ‘representative’ star rating

N-GRAM WORD CLOUDS

Mono-gram

Bi-gram

Tri-gram



Complexity

Basic analysis of top words used, suggest differences based on star ratings

TOP WORDS IN REVIEWS GROUPED BY STAR RATINGS

Word cloud

Rank of word in terms of frequency of use (in text by star rating):

1	time	good	good	good	great
.	like	like	like	great	good
.	one	time	time	like	time
.	back	one	would	time	best
5	get	get	one	one	one
.	would	would	get	really	like
.	go	back	really	get	delicious
.	good	really	great	go	go
	even	go	go	would	back
10	never	table	back	back	love

Basic analysis of top words used, suggest differences based on star ratings

TOP WORDS IN REVIEWS GROUPED BY STAR RATINGS

Word cloud

Rank of word in terms of frequency of use (in text by star rating):

best	153	76	56	31	4
great	60	16	8	2	1
good	8	1	1	1	2
ok	118	50	35	251	540
bad	22	30	48	148	221
never	10	48	110	97	55

Top 10

Next steps

Analyze text and build a model to determine star rating based on review's text