



YELP A REVIEW ANALYSIS

A BIT OF DOMAIN KNOWLEDGE ON YELP

- ✿ Pros:

- ✿ Democratic Review Process
- ✿ Overall Honest
- ✿ Thousands of very, very, wordy, reviews

- ✿ Cons:

- ✿ No Filter
- ✿ No Guidance
- ✿ No Context

HOW RESTAURANTS INTERSECT WITH YELP

- ✿ Most Restaurants are satisfied with a 4-star rating
- ✿ 3.5 is a motivator for change
- ✿ Any restaurant over 100 reviews cannot maintain 5 stars
- ✿ By 300 reviews your rating will settle, thus 3-star restaurants typically stay 3-stars.
- ✿ Because there's no reliable rubric, individual ratings swing from 1-star or 5-star, with few in between.
- ✿ This leaves most restaurants settling in the 4 zone.
- ✿ Truly amazing restaurants get 4.5s

SOME SAMPLES - TACOS = AWESOME

Taco Bell • Claimed

★★★★★ 32 reviews

\$ Fast Food, Mexican, Tex-Mex

Order Food

\$3.99+ fee • \$0 min • 45-55 mins

Delivery Address

Enter delivery address

Start Order

Review Highlights

- "Went here at 3am and they were so polite and friendly and the **drive thru** was super quick." in 5 reviews
- "Seriously, even with Chick-fil-A and their "my pleasures," this Taco Bell is the best." in 2 reviews
- "They always make you aware of special sales and **promotions.**" in 2 reviews

Los Tacos No.1 • Claimed

★★★★★ 1479 reviews

\$ - Tacos

Review Highlights

- "The chicken tacos were incredible, but I almost regret not getting another **adobada** taco because it was that good." in 131 reviews
- "Excellent tacos right at **Times Square**, awesome flavor and tortillas like in my abuelita's house...just perfect!" in 134 reviews
- "you know it's a gem when the line is always out the door. I get the grilled cactus tacos & grilled **cactus** quesadilla." in 40 reviews

Tacos

3 Michelin Star Restaurants

Per Se • Claimed

★★★★★ 1604 reviews

\$\$\$\$ French

Review Highlights

- "The ambience is modern and warm with a fireplace and a lovely view of **Columbus Circle** and Central Park behind it." in 156 reviews
- "Ok, maybe just me, but seriously if you mention Per Se or **French Laundry** to me I get butterflies and goosebumps." in 208 reviews
- "Inspired by a trip to Baskin Robbins, **Thomas Keller** tops an onion creme fraiche

Le Bernardin • Claimed

★★★★★ 2458 reviews

\$\$\$\$ French, Lounges

Popular Dishes

Lobster Caviar Langoustine

REVIEWS WERE ALSO FLAWED

**Sora K.**

Washington, DC

379 friends

26 reviews

3 photos



8/12/2017

I really like this place. It's not the best Chinese food in NYC, but it's good food. The people that work there are very kind and attentive. I once dropped by at 9:58pm (when they were closing at 10pm), and they stayed open a little longer to take my order and get the food ready. I think they care about their customers and the Chinese food is pretty solid! The prices are cheap and reasonable. I order from this place a lot and they are very reliable.

Useful

Funny

Cool



Not Really Sure What to Do With This

**Ross F.**

NJ, NJ

68 friends

22 reviews

9 photos

Too Much Sentiment



9/7/2014

The entire kitchen and wait staff saw an ice cream truck and ran outside, leaving me alone in the restaurant. 10 minutes later they all came back with ice cream cones.

I still can't believe this actually happened.

Useful 142

Funny 655

Cool 161



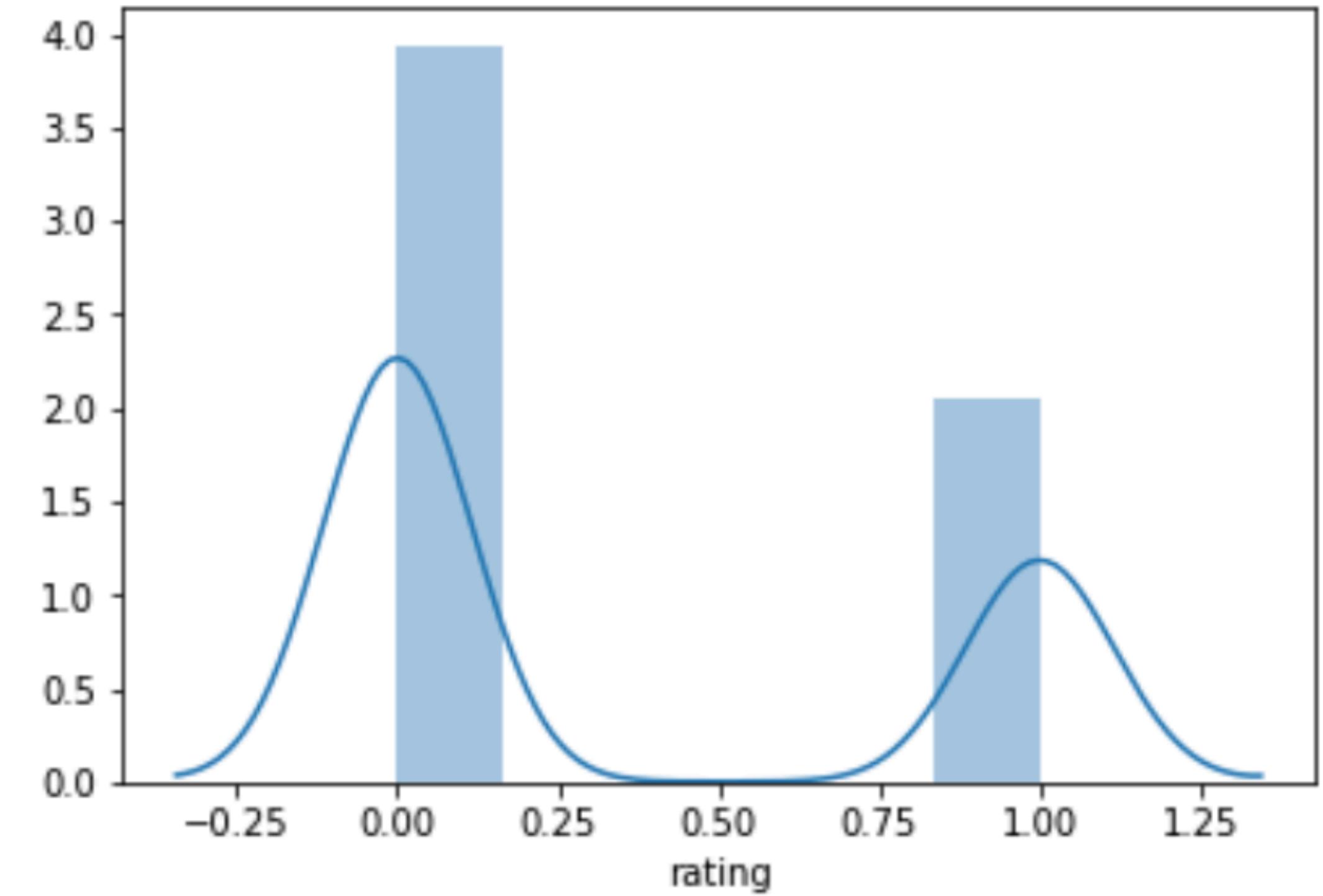
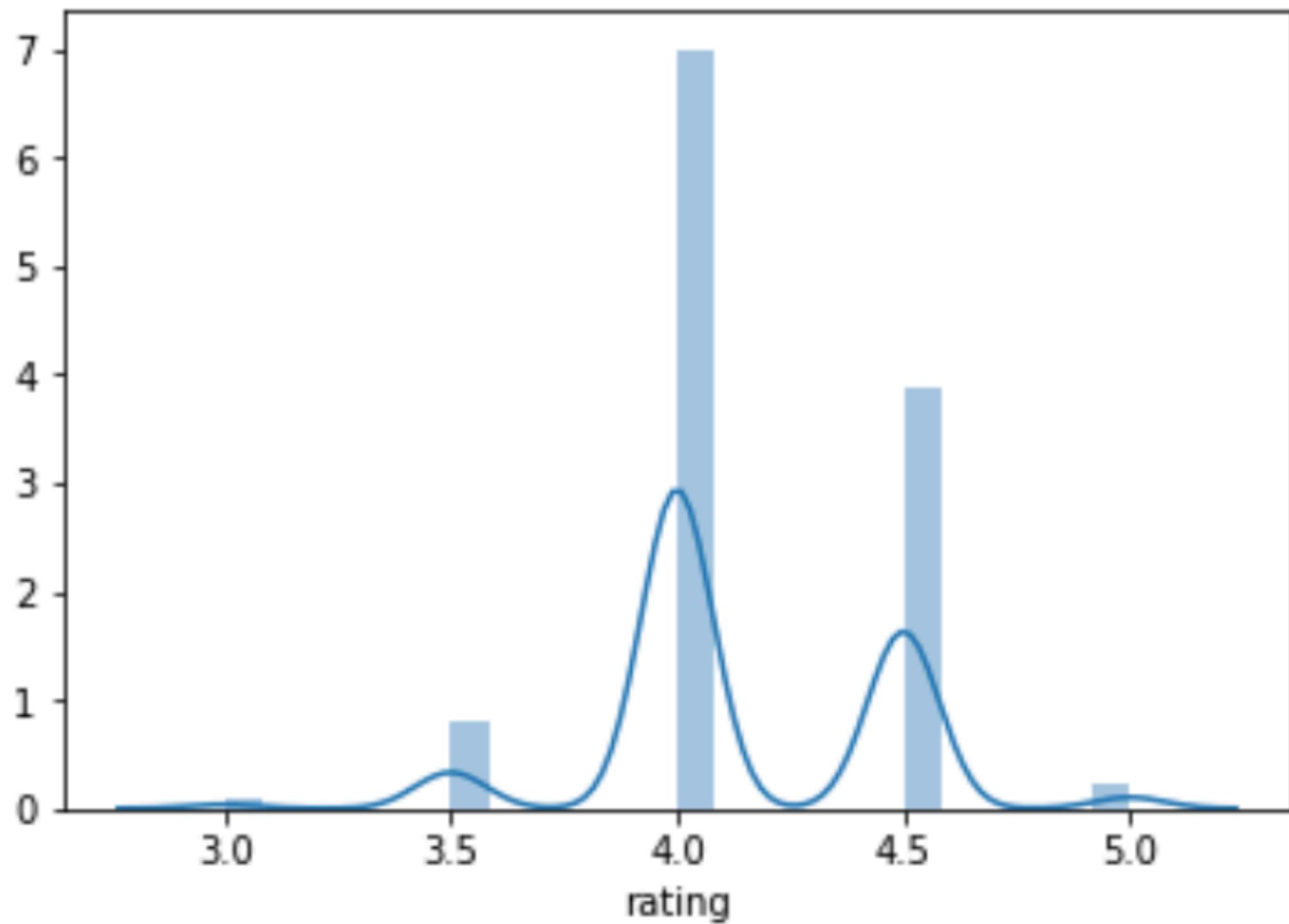


**THAT RESTAURANT: YEAH OR
NAH?**

OUR DATA

- Scraped for Days from Yelp
 - API: 2000 Calls for Restaurants in the NYC Area
 - WebScraping: 20 reviews for each restaurant in the API.
- Post Cleaning:
 - 1625 Restaurants
 - 32,500 Reviews

WHAT DID WE GET FROM THE DATA



- The 4-star rating dominates the majority of our restaurants.
- This leaves with a class imbalance that we resolved by transforming rating into a binary variables
- All restaurants with review 4.5+ are classified 1: Good review (35%)
- The rest under 0: Bad review (65%)

WHAT DID WE DO TO OUR DATA

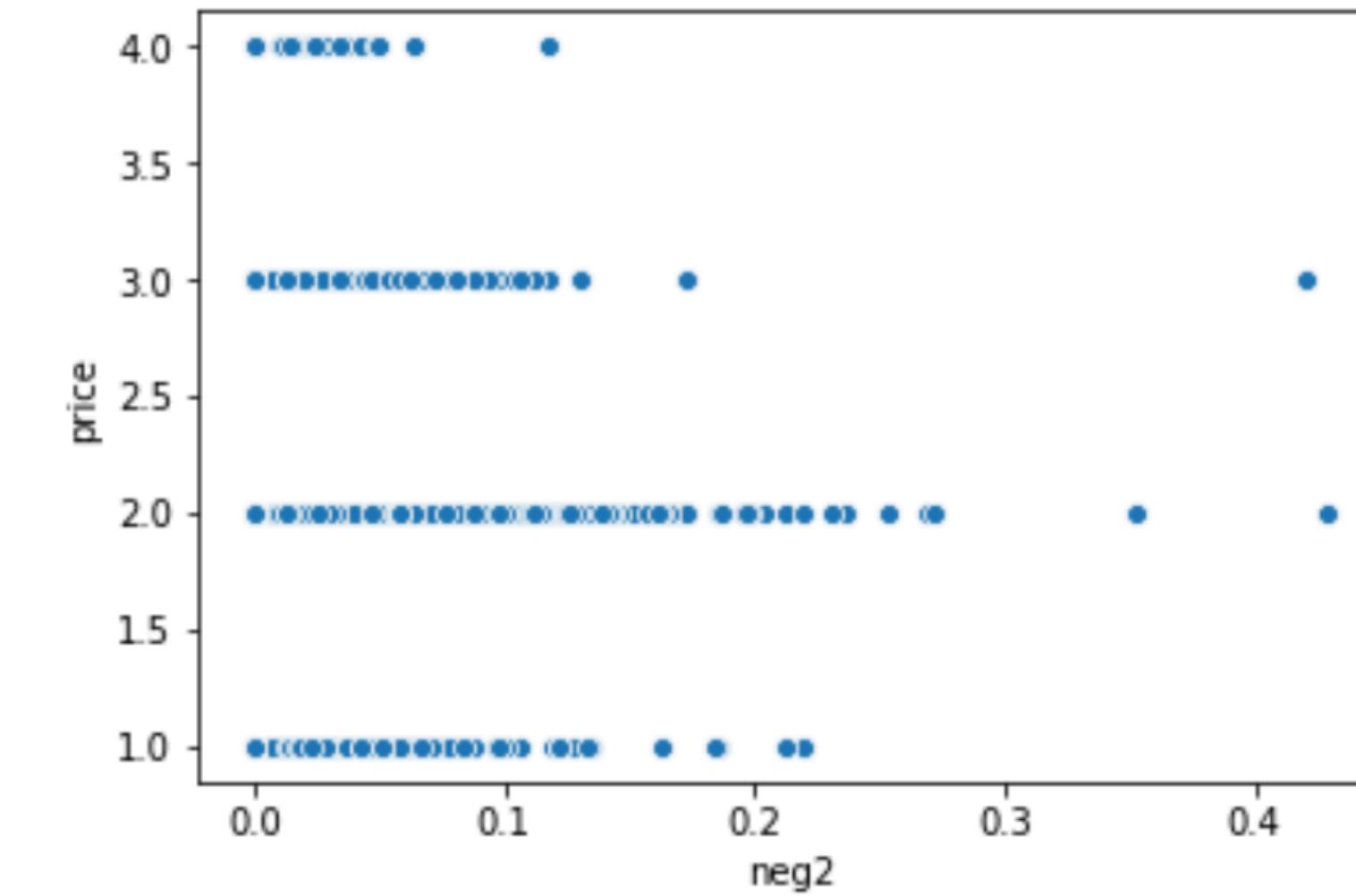
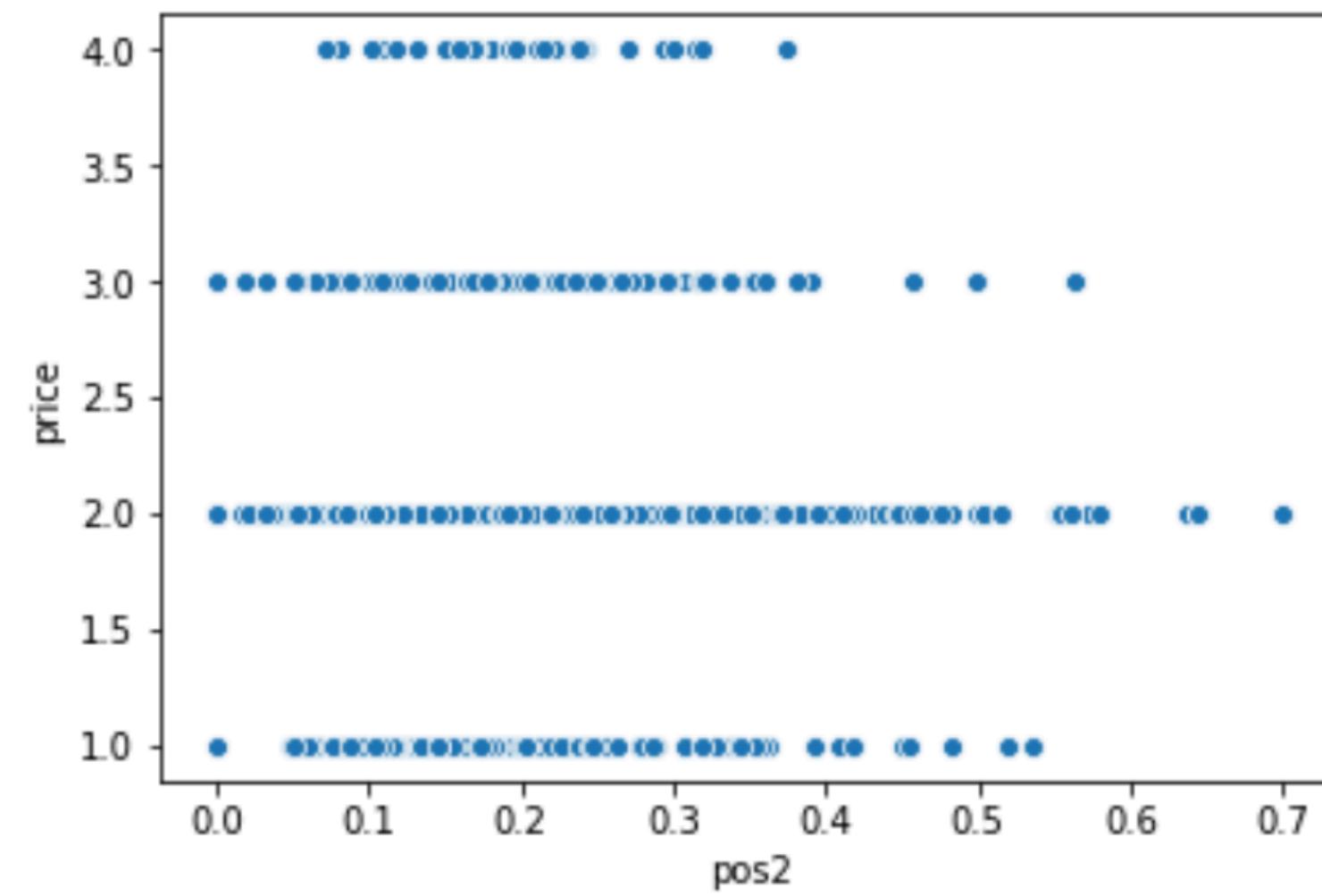
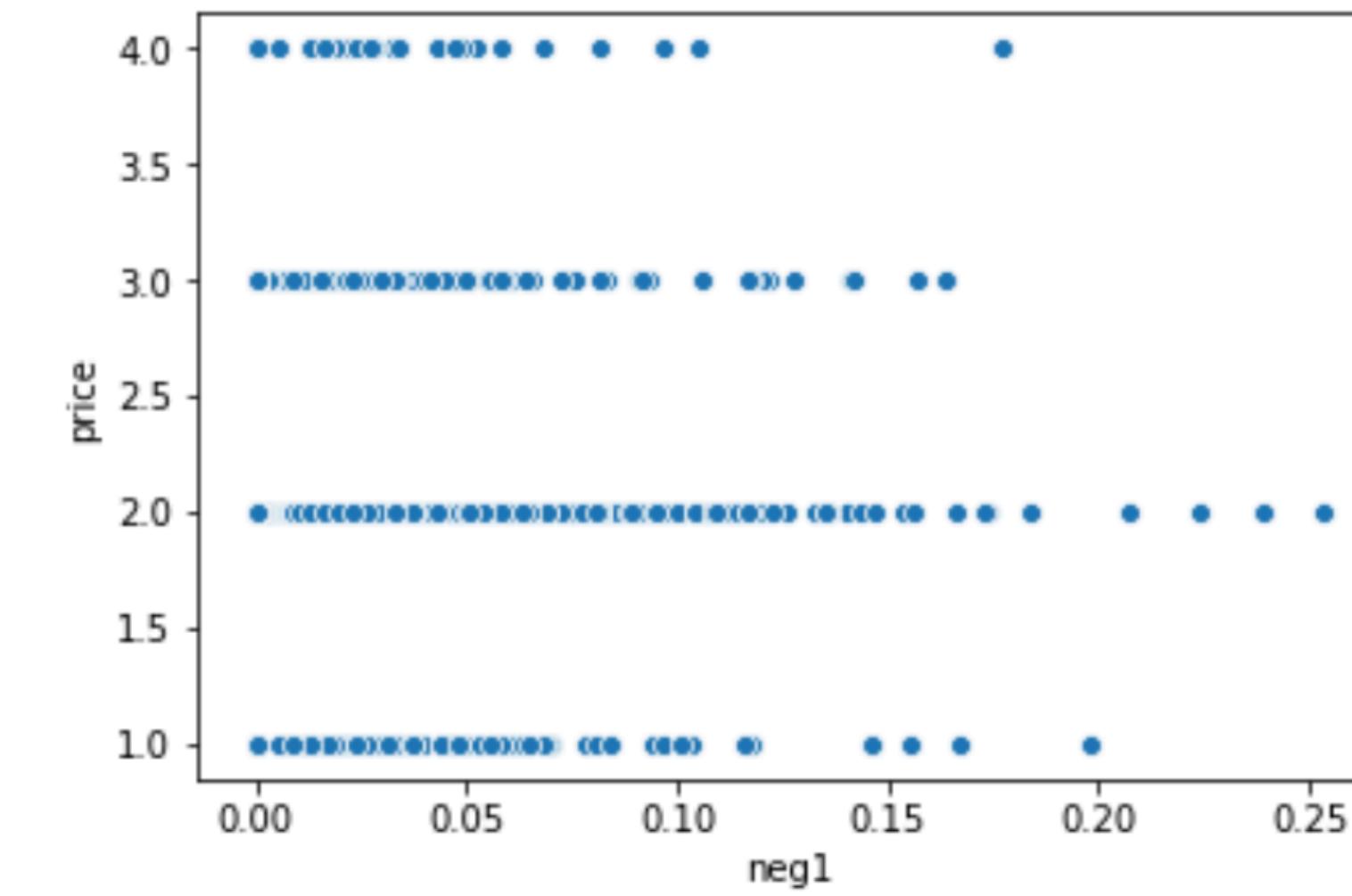
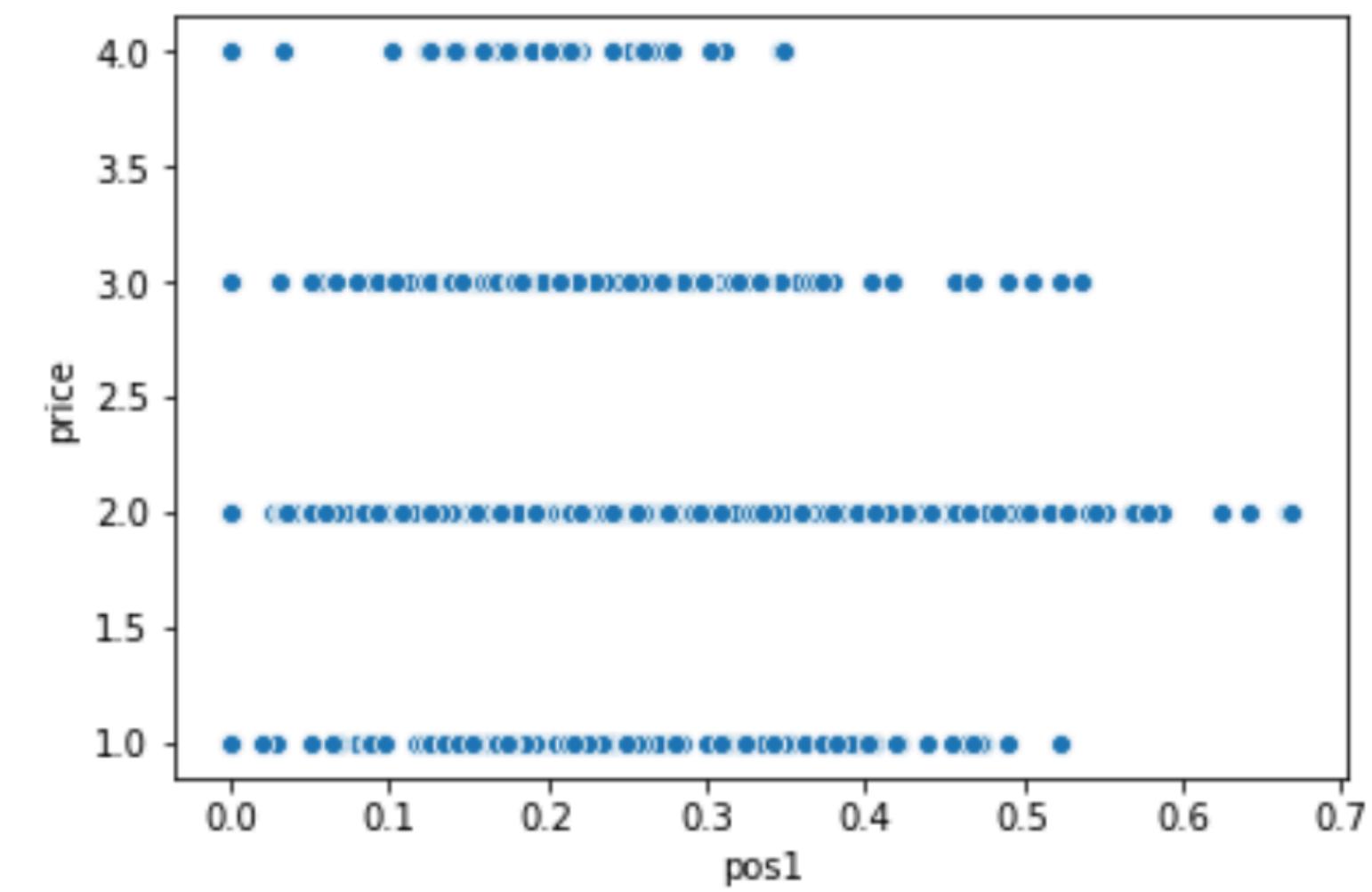


We did not remove any punctuation, any emojis, lowercase any words, or tokenize words.

They would adversely affect the sentiment scores given by VADER

- Drop Null values for ratings
- Remove restaurant with fewer reviews
- Transforming restaurant price for incorporation in models
- Getting sentiment scores from reviews - 97,500 individual scores

HOW THE SENTIMENTS INTERACT WITH PRICE





**LET'S FIND THE BAD
RATINGS**

WHAT ARE THE MODELS USED

- We are looking to predict if a restaurant is Good or Bad based on the sentiments of the reviews
- We don't want to forecast a good restaurant and find out it's bad - **Type I error not acceptable**
- We want to find a model better than the baseline (Dummy: 52.3% accuracy)

MODEL	TRAINING		TESTING	
	PARAMETERS	ACCURACY	F1 SCORE	
NAIVE BAYES	GAUSSIAN/BERNOULLI	58% / 52%	59%/54%	
KNN	N=2 ; TUNING USING GRIDSEARCH	52%	40%	
DECISION TREES	TUNING USING GRIDSEARCH	57%	58%	
BAGGED TREES	N=20;TUNING USING GRIDSEARCH	64%	63%	
RANDOM FORESTS	TUNING - BOOTSTRAP	66%	63%	
XG BOOST	GENETIC SEARCH	65%	65%	
SVM	KERNEL LINEAR; C=6	56%	57%	

WHAT ARE THE TAKE AWAY

- Random Forest and XGBoost predicted our testing sample accurately at 66% and 65% resp.
- Adding the prices of the restaurant didn't make a significant change in the models
- Due to distance between each sentiment of each review - weak learners were better predictors.
- More data needed to increase prediction; topic of reviews worth analyzing separately