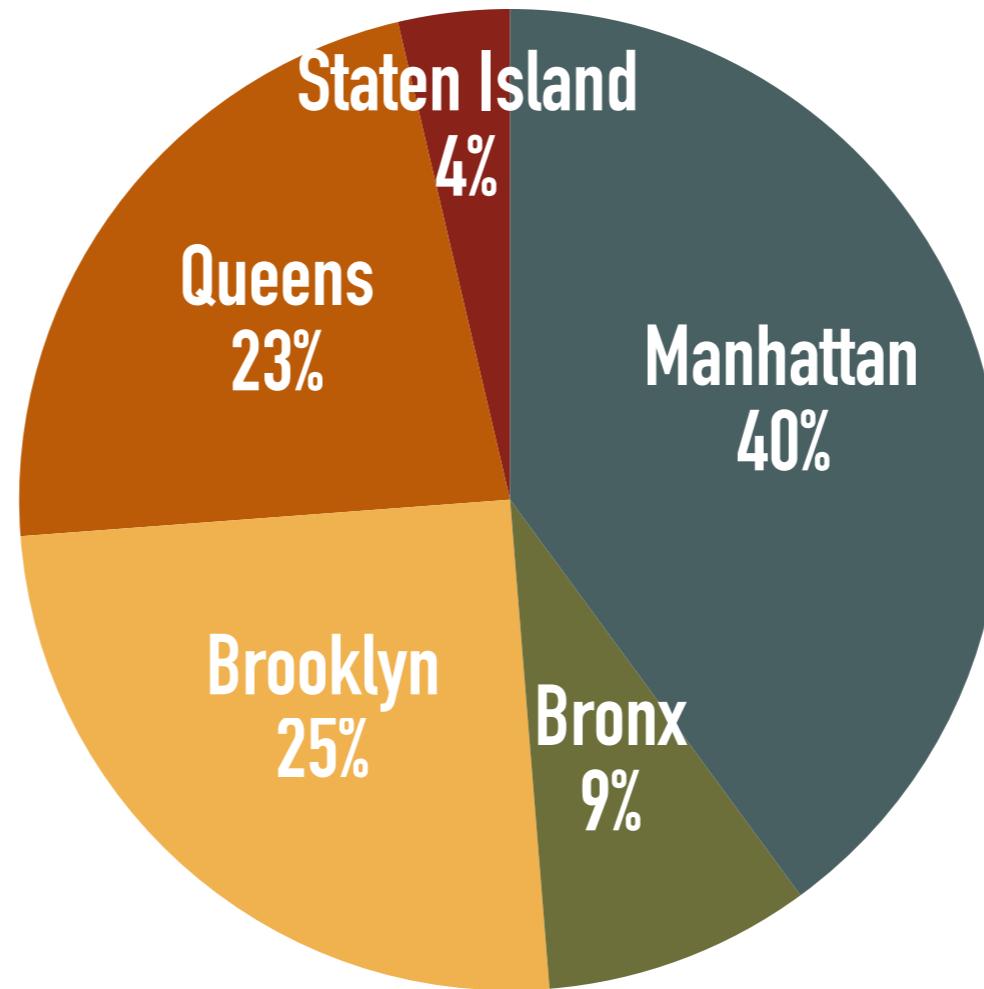




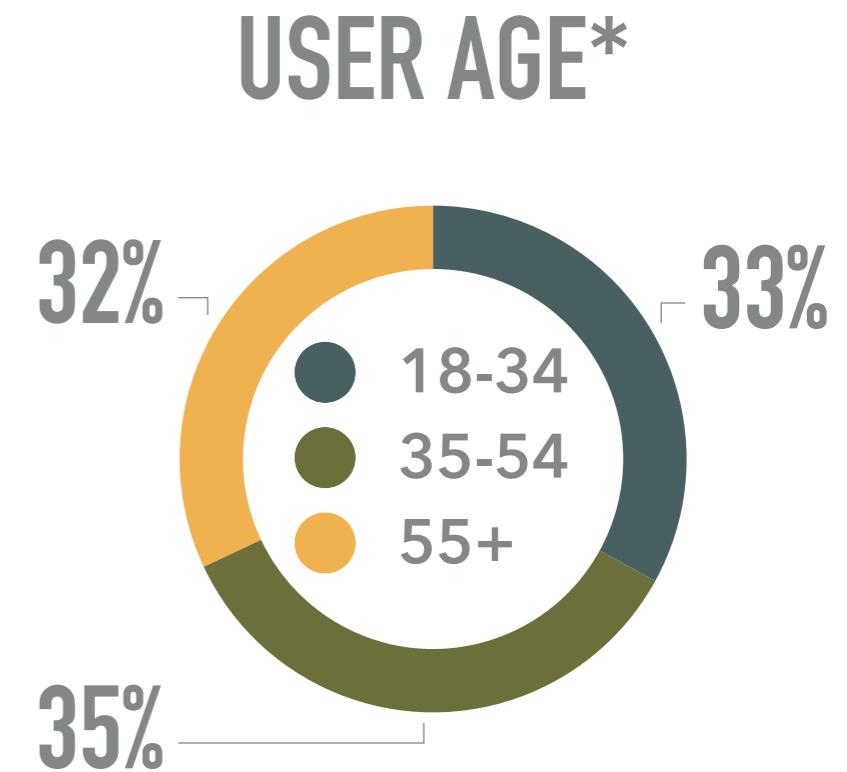
YELP-HELP:

**MANHATTAN RESTAURANT
LANDSCAPE**

- ▶ 27,308 restaurants in NYC*
- ▶ 27,308 restaurants in NYC ➡ 25 years
- ▶ 10,903 restaurants in Manhattan ➡ 10 years



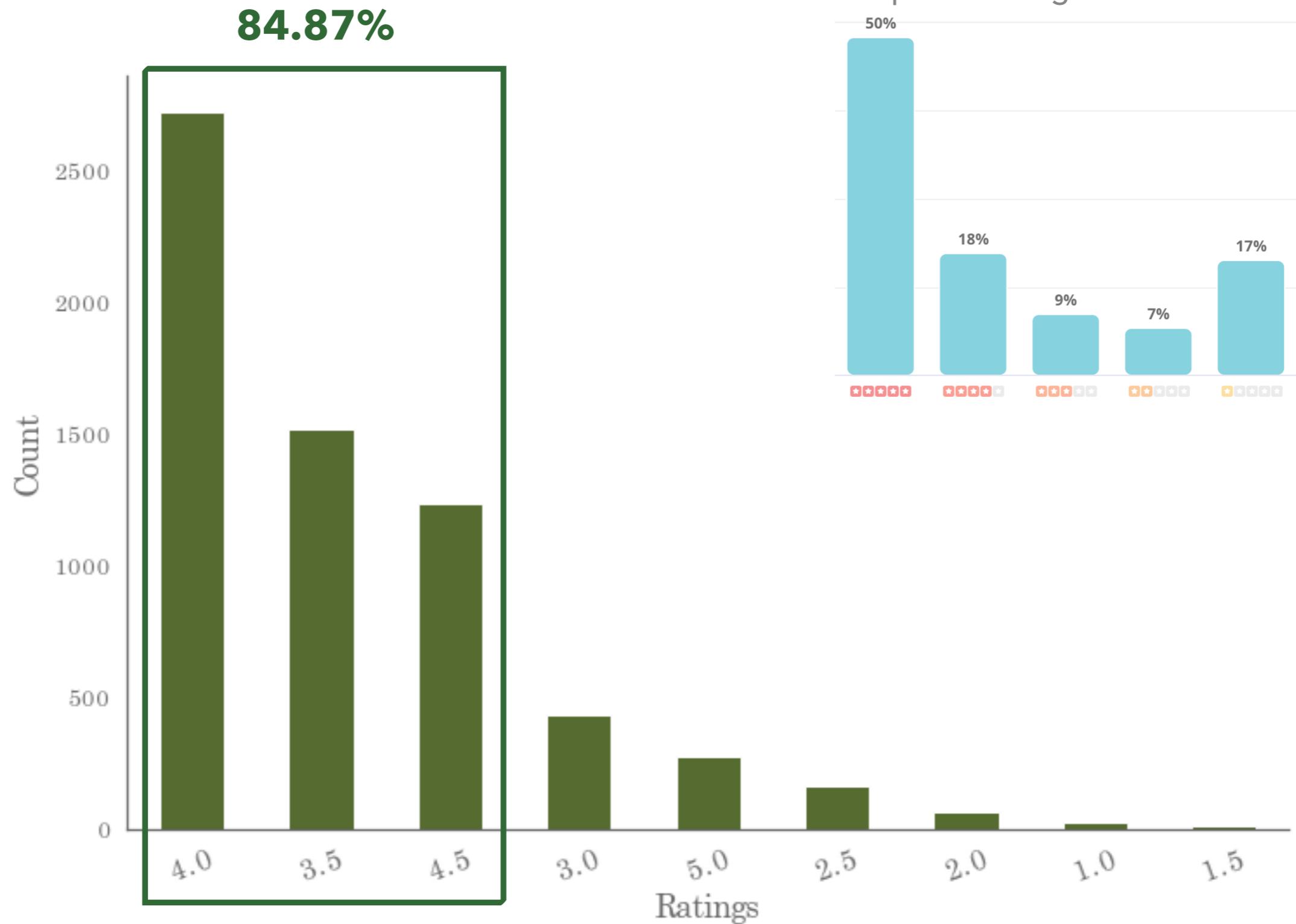
- ▶ 24,233 place to eat in New York, NY registered on Yelp
- ▶ Yelp API
 - ▶ Rating
 - ▶ Claimed
 - ▶ Review count
 - ▶ Price level
 - ▶ Categories (New American, Italian, etc.)
 - ▶ Transactions (pickup, delivery and reservation)
 - ▶ Additionally engineered features
 - ▶ Number of hours open
 - ▶ Neighborhood
 - ▶ Mainstream vs. rare category
 - ▶ Restaurant name length
 - ▶ Messaging feature



*Source: Yelp

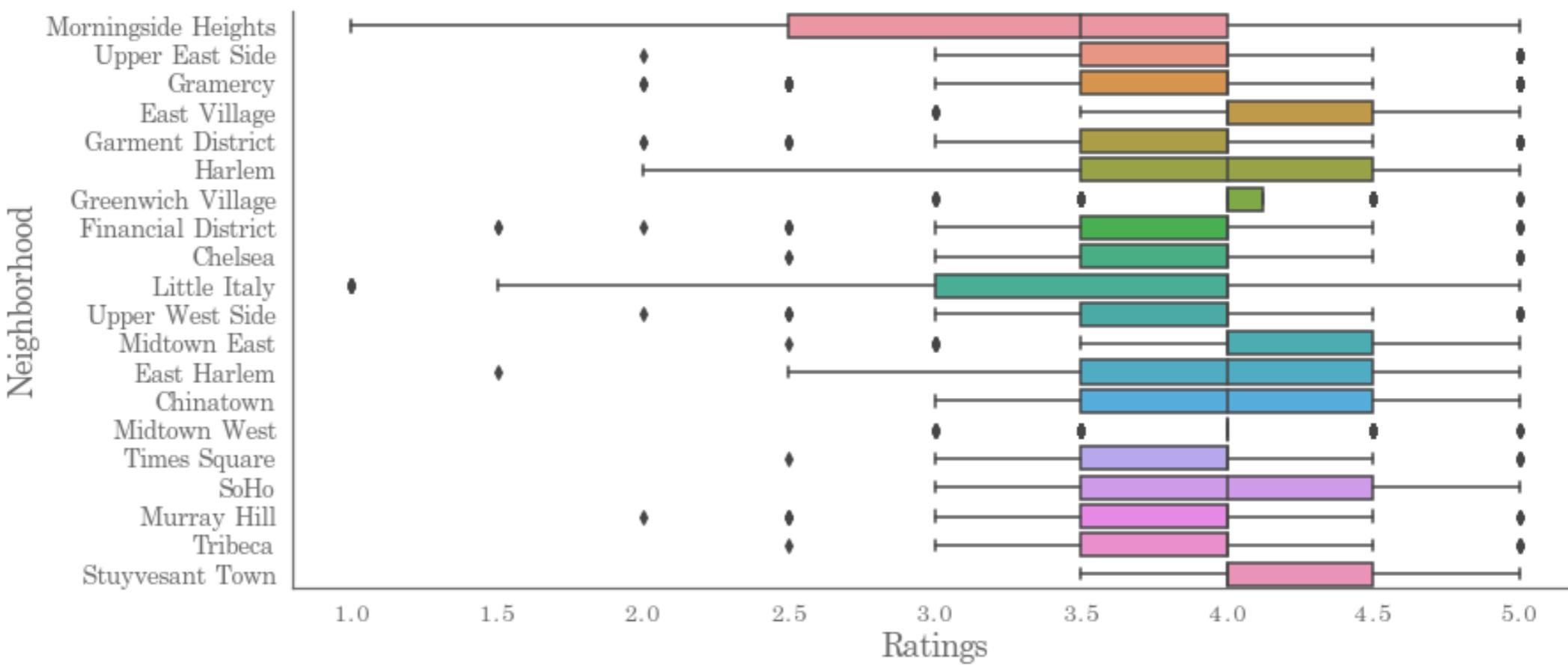
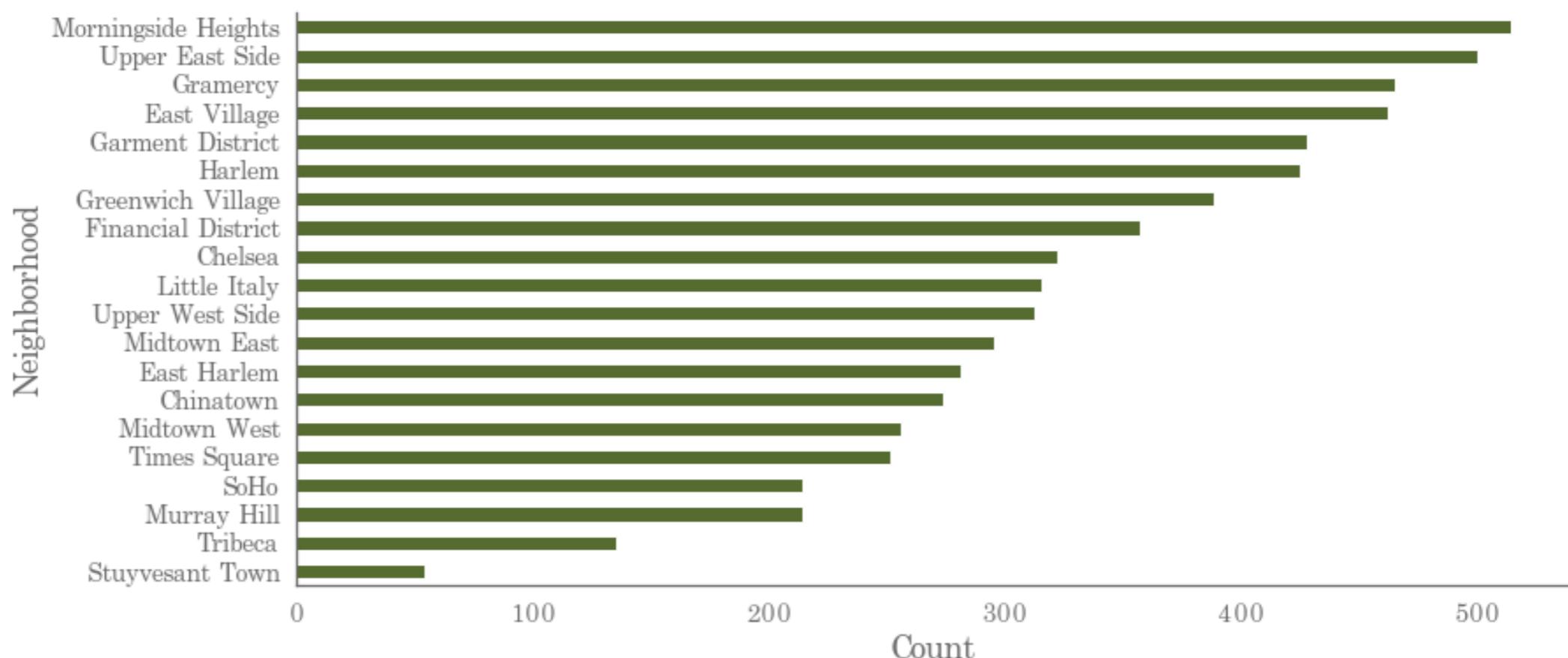
MANHATTAN STAR RATING DISTRIBUTION

4



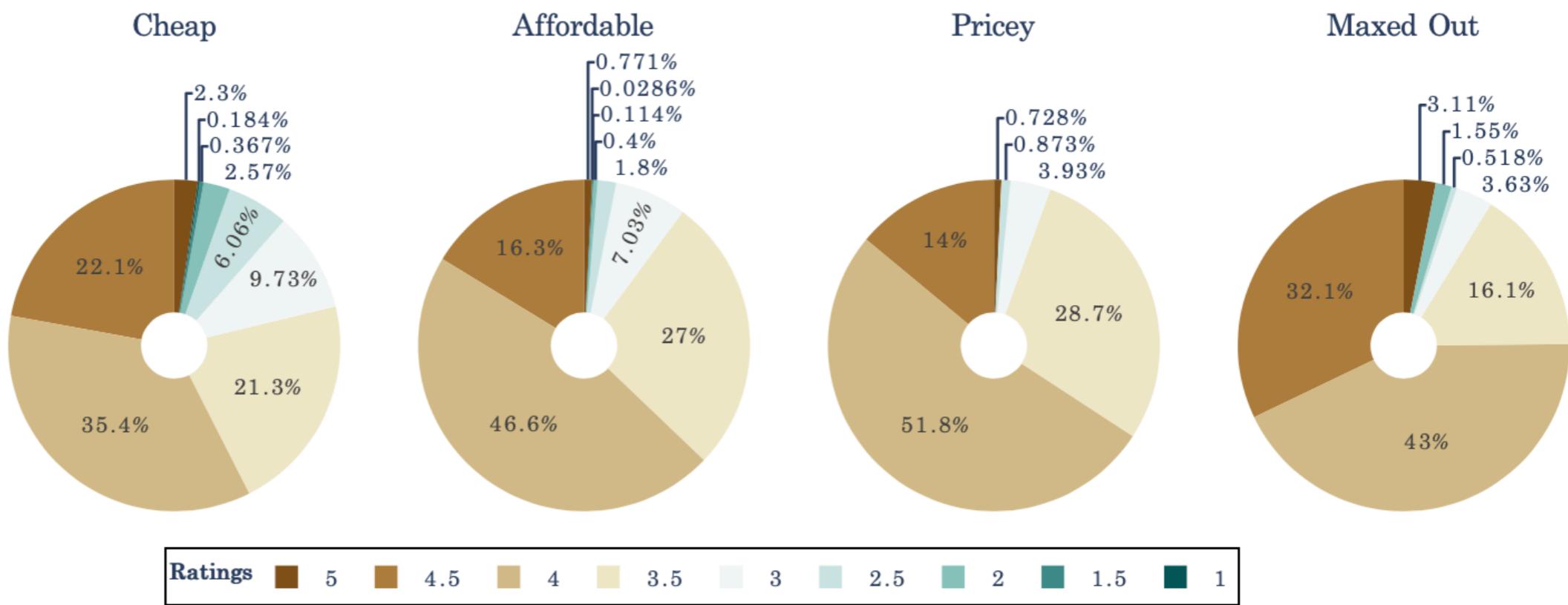
DISTRIBUTION BY NEIGHBORHOOD

5



DISTRIBUTION BY PRICE LEVEL

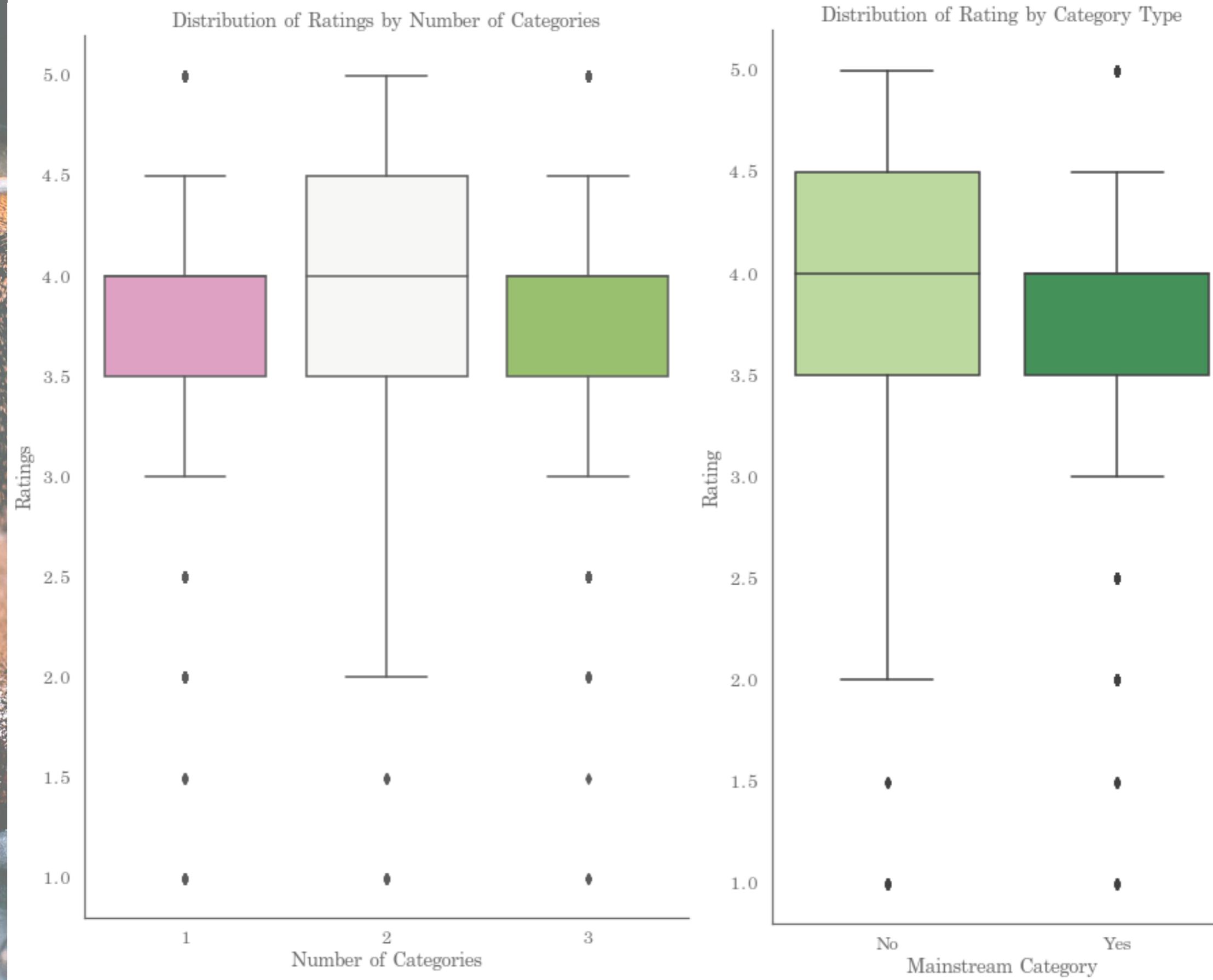
6



Rating	Price Level: Cheap	Price Level: Affordable	Price Level: Pricey	Price Level: Maxed Out
1	2	1	null	null
1.5	4	4	null	null
2	28	14	null	3
2.5	66	63	6	1
3	106	246	27	7
3.5	232	945	197	31
4	385	1631	356	83
4.5	241	570	96	62
5	25	27	5	6

DISTRIBUTION BY CATEGORY

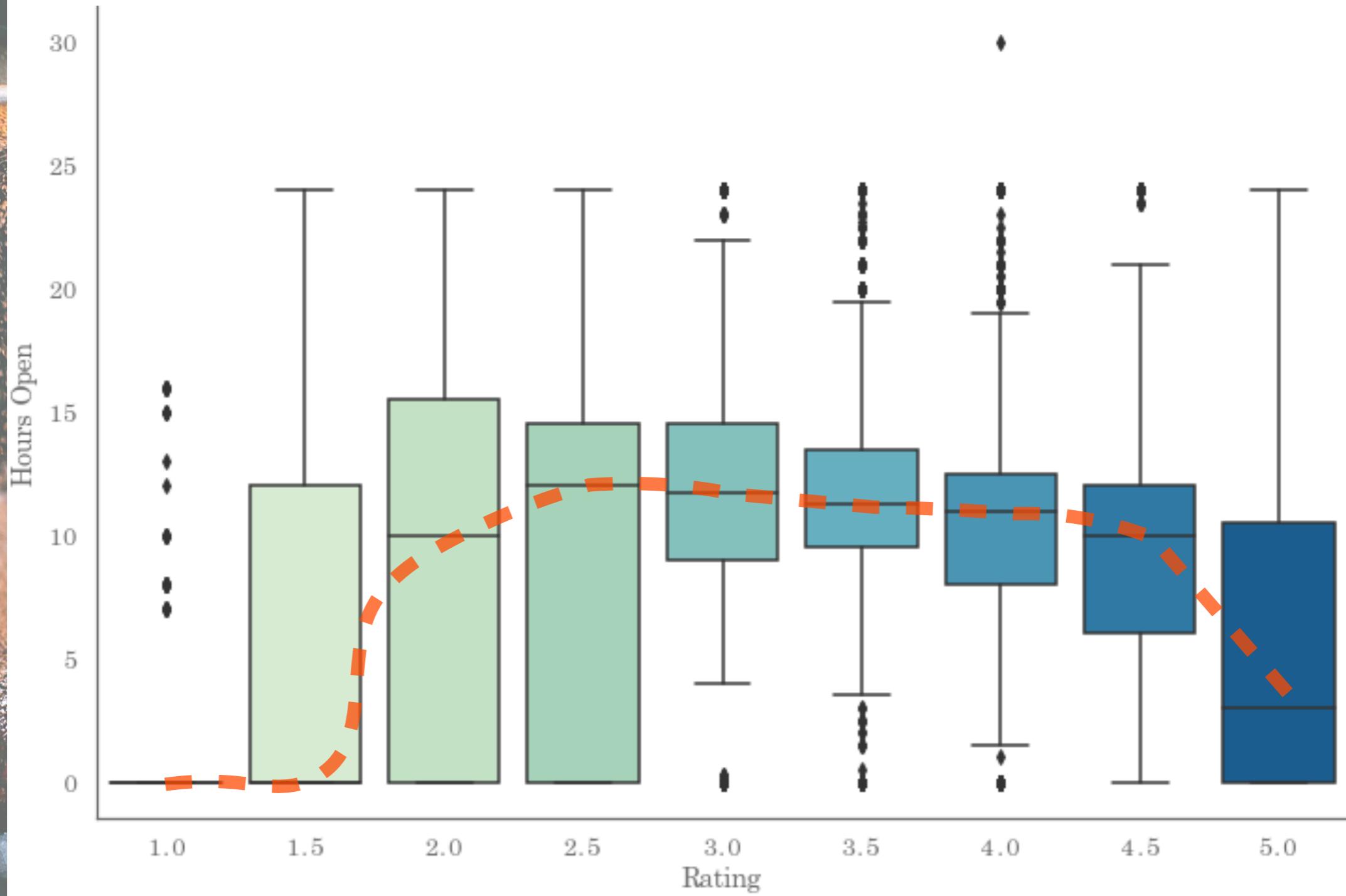
7



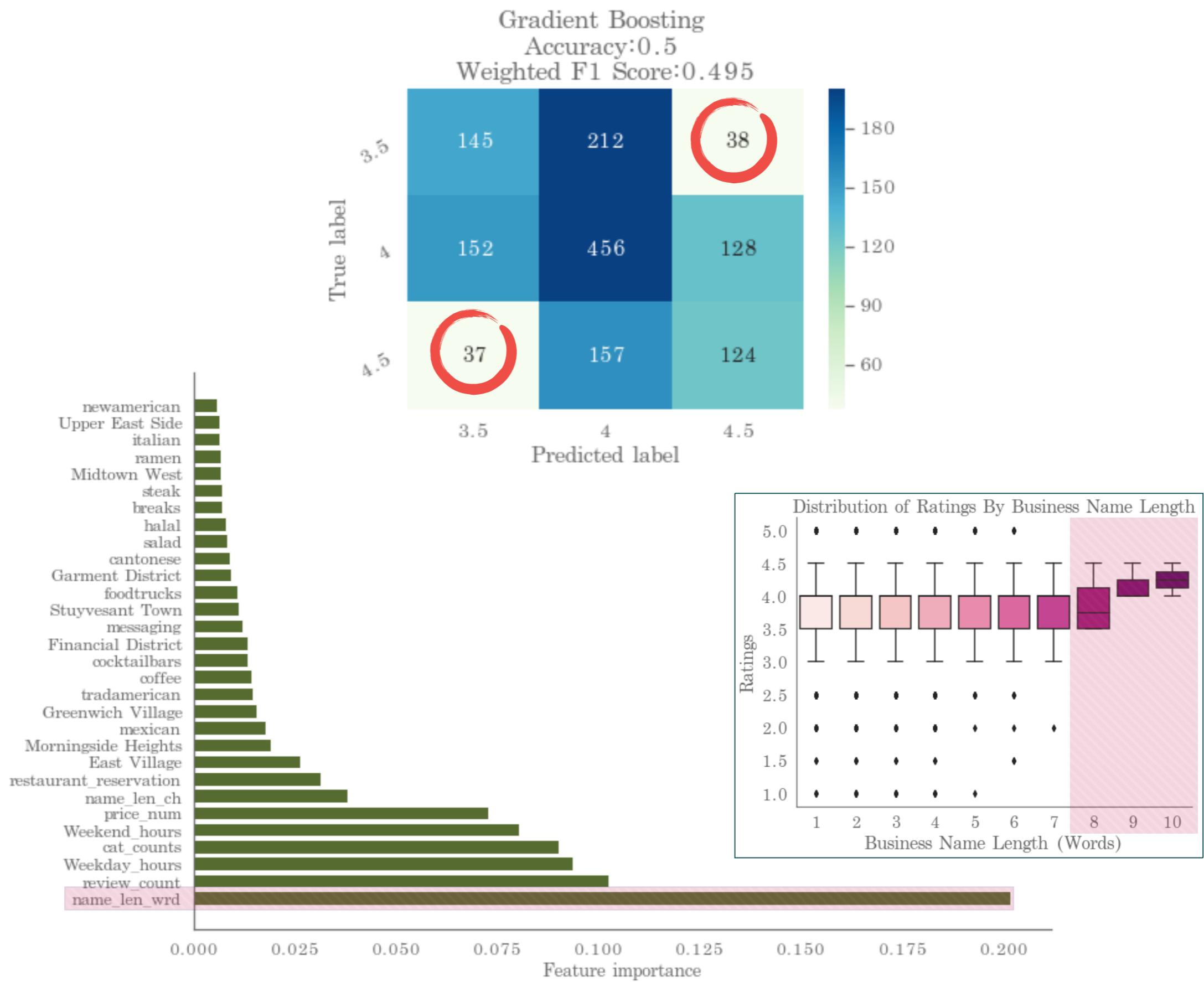


DISTRIBUTION BY OPERATING HOURS

8

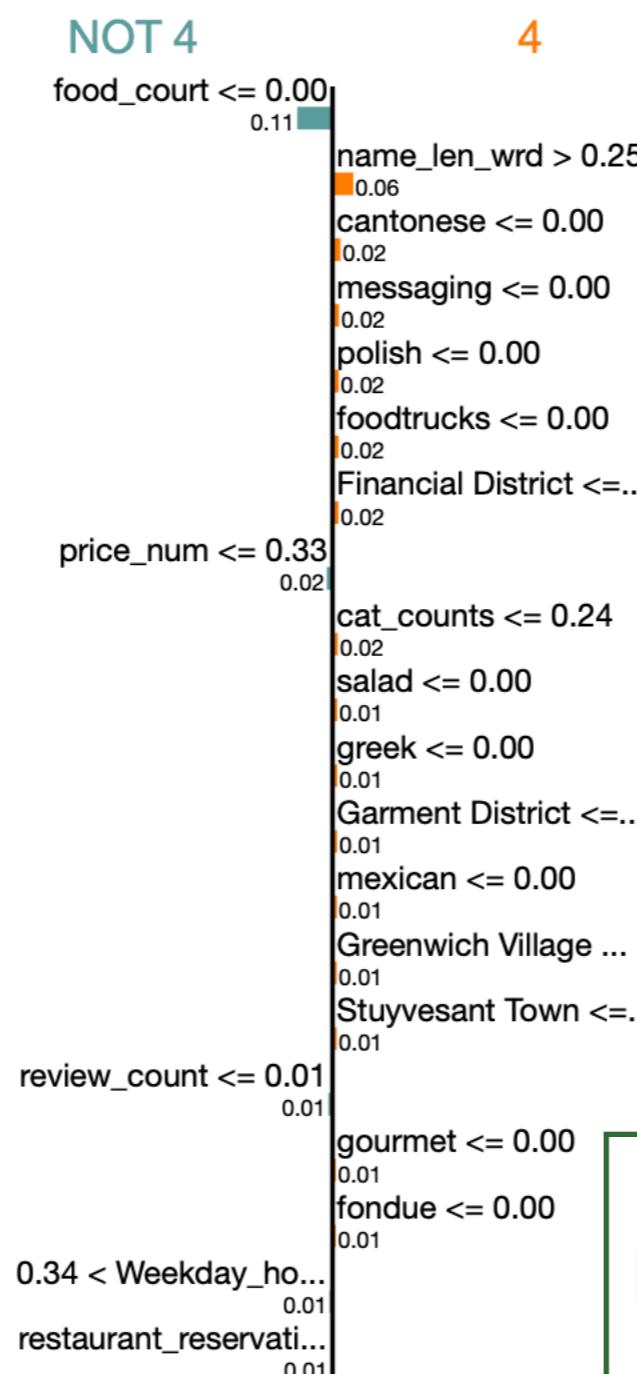
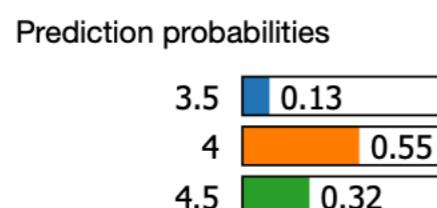


- 
- ▶ Ratings between 3.5 - 4.5 (85% of total)
 - ▶ Feature selection
 - ▶ Top 30 important features from vanilla models (46 total)
 - ▶ Oversampling for class imbalance problem
 - ▶ Run vanilla models to choose optimal algorithm
 - ▶ Refine the model



```
For restaurant # 2577      IL Tavolo di' Palmisano  
Name: name, dtype: object
```

Actual Rating: 4.5
Predicted Rating using Vanilla Gradient Boosting Model: 4.0



Feature	Value
review_count	0.00
name_len_ch	0.46
Weekday_hours	0.43
name_len_wrd	0.38
Weekend_hours	0.52
cat_counts	0.00
price_num	0.33
messaging	0.00
Greenwich Village	0.00

IL Tavolo di' Palmisano

5 reviews

Details

\$\$ • Italian

Edit

Write a Review

Add Photo

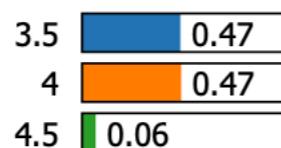
Share

Save

For restaurant # 1494 dell'anima
Name: name, dtype: object

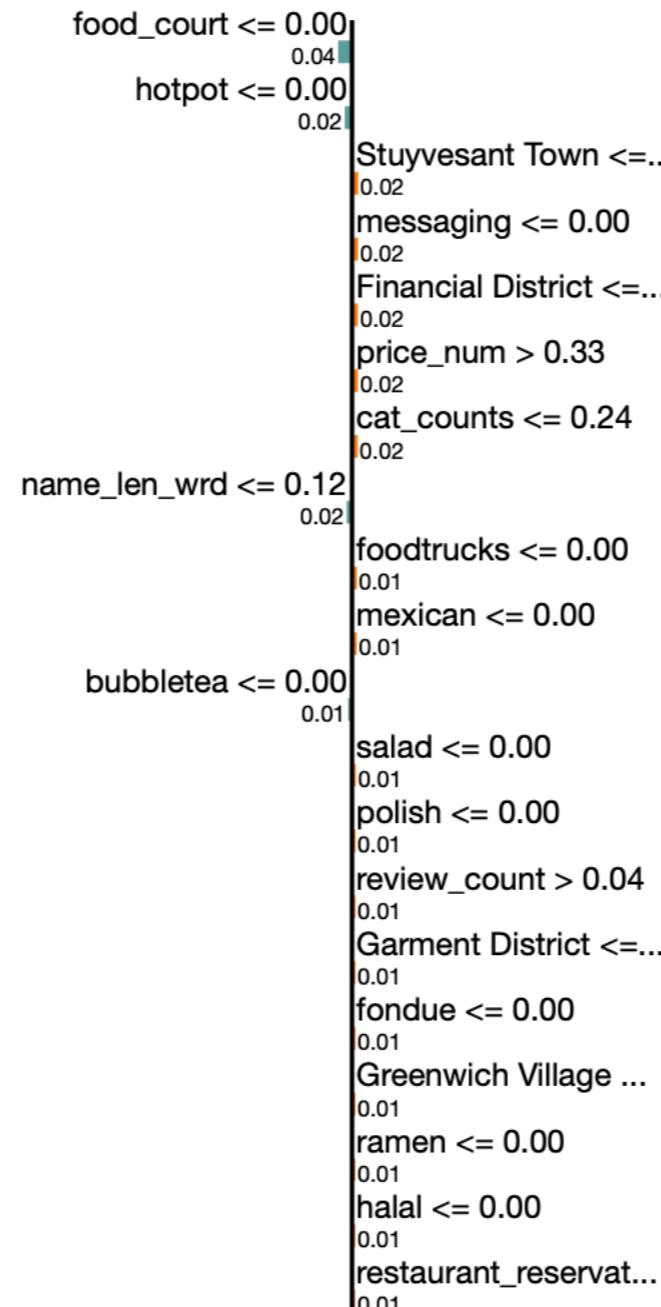
Actual Rating: 4.0
Predicted Rating using Vanilla Gradient Boosting Model: 4.0

Prediction probabilities



NOT 4

4



Feature Value

food_court	0.00
hotpot	0.00
Stuyvesant Town	0.00
messaging	0.00
Financial District	0.00
price_num	0.67
cat_counts	0.00
name_len_wrd	0.00
foodtrucks	0.00

dell'anima Claimed

639 reviews [Details](#)

\$\$\$ · Italian [Edit](#)

Write a Review

Add Photo

Share

Save



► Findings

- ▶ Ratings change over time
- ▶ Target 4-star ratings
- ▶ Restaurant name doesn't necessarily have to be short (or easy)
- ▶ Increase number of reviews
- ▶ New Yorkers are willing to spend money for good food
- ▶ Have two categories (users can modify)

► Next steps

- ▶ More accurate measure of location (neighborhood)
- ▶ More feature engineering
- ▶ Set minimum review count threshold



THANK YOU
FOR YOUR TIME