



True Review

A Personalized Restaurant
Recommender

Scope



PROBLEM



EXPLORE THE
DATA



MODELING



RESULTS

Problem

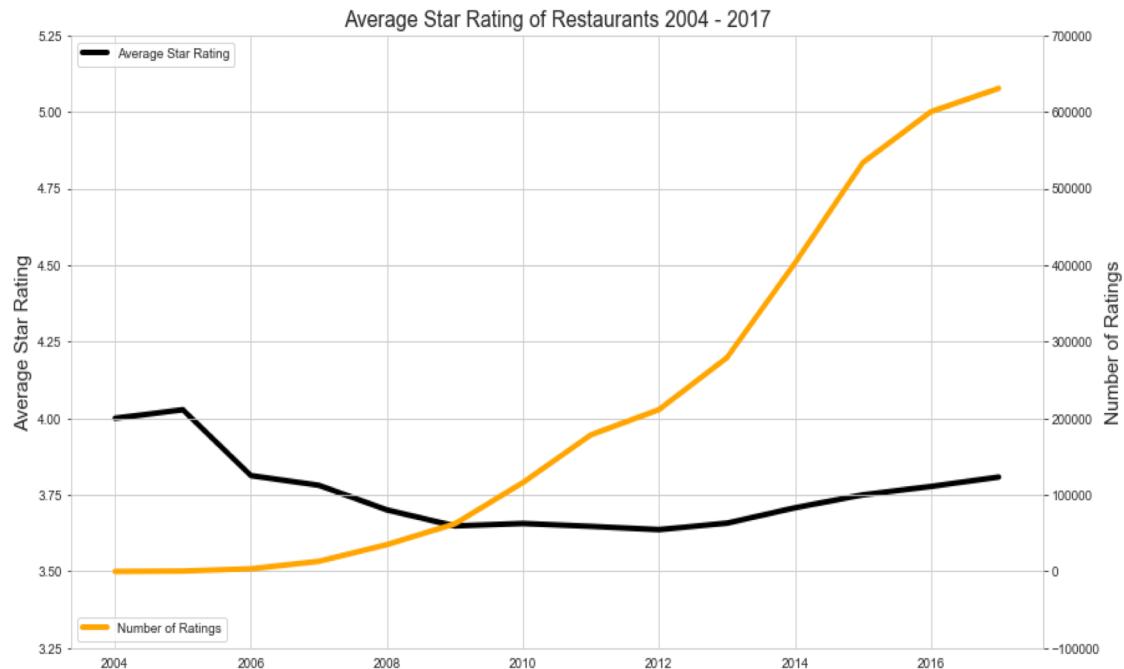


- What are users in Yelp reviews saying?
- Why should I trust Yelp reviewers, I'm different!?
- There are thousands of reviews for many restaurants



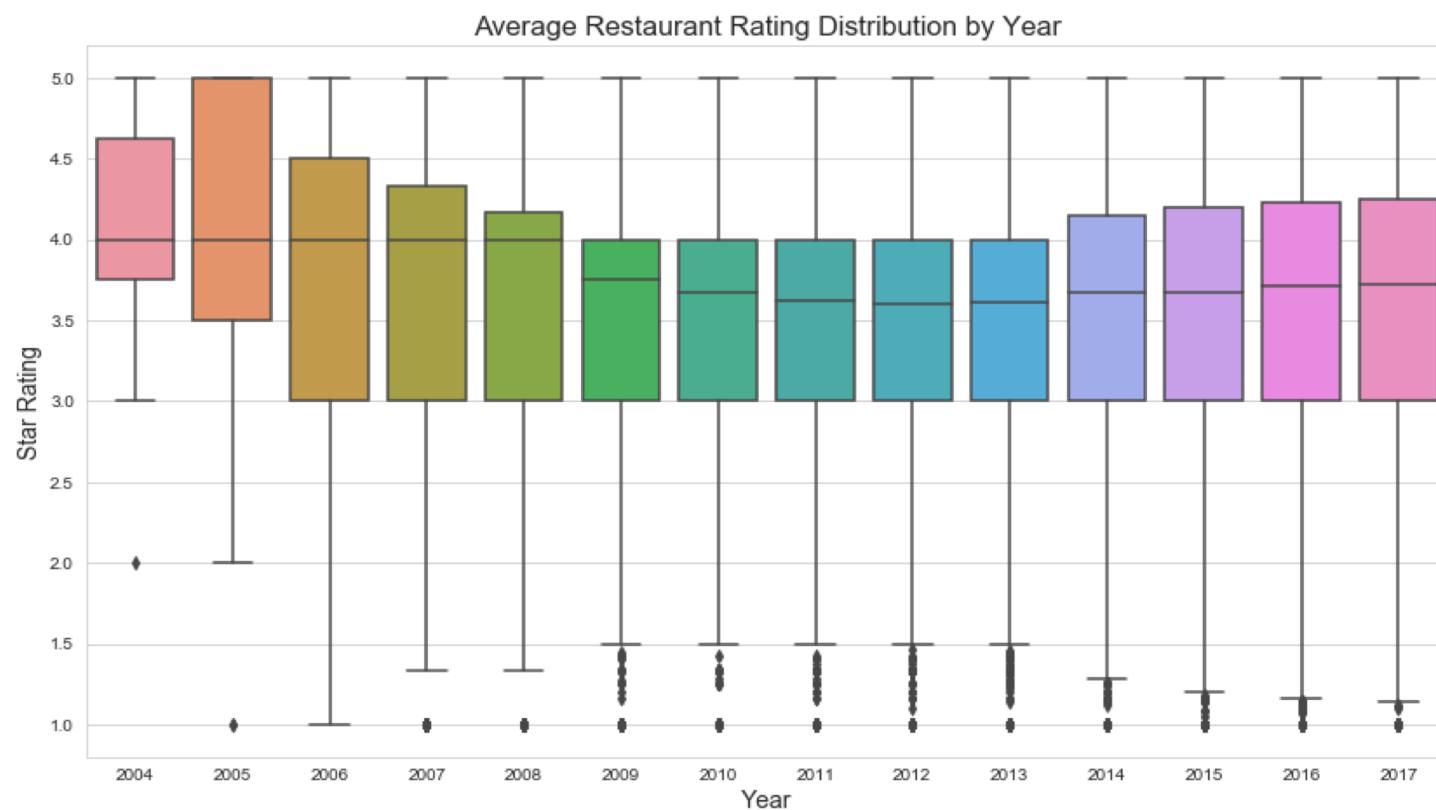
Explore the Data – Reviews Over Time

- Average star rating and number of reviews over the years

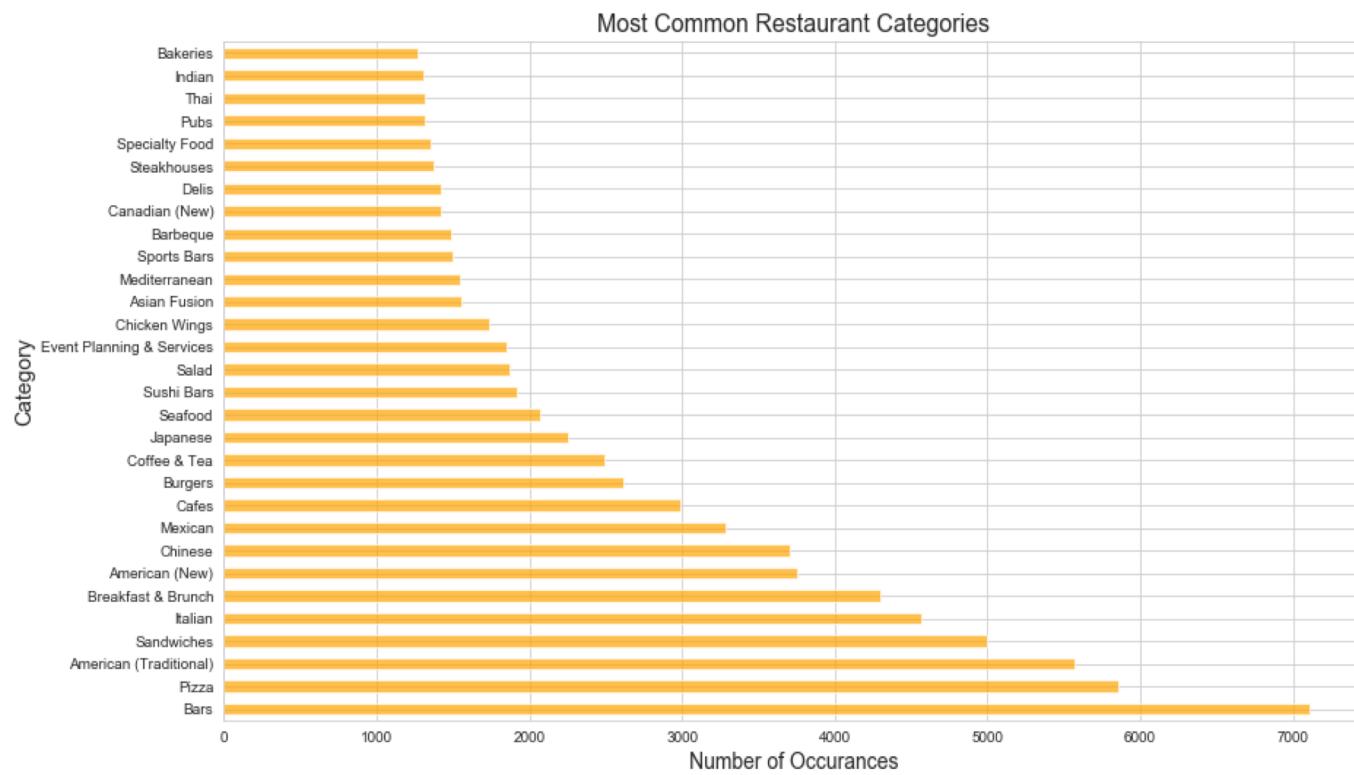




Explore the Data – Distribution of Ratings

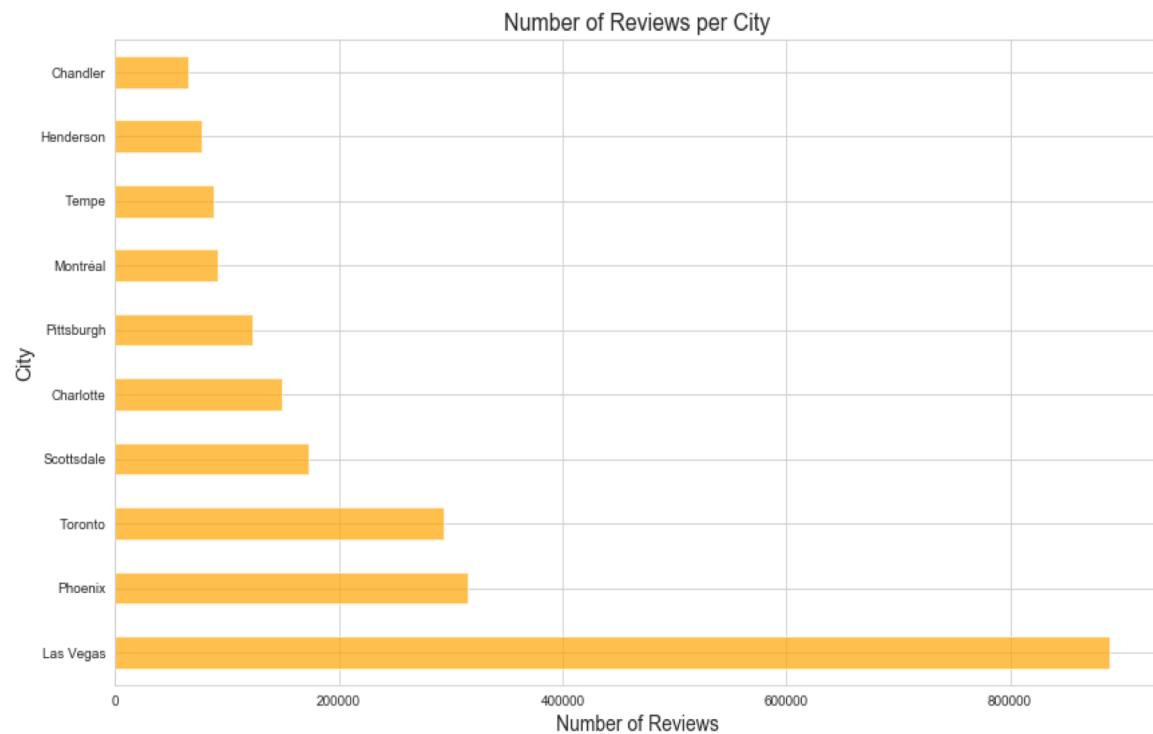


Explore the Data – Common Categories



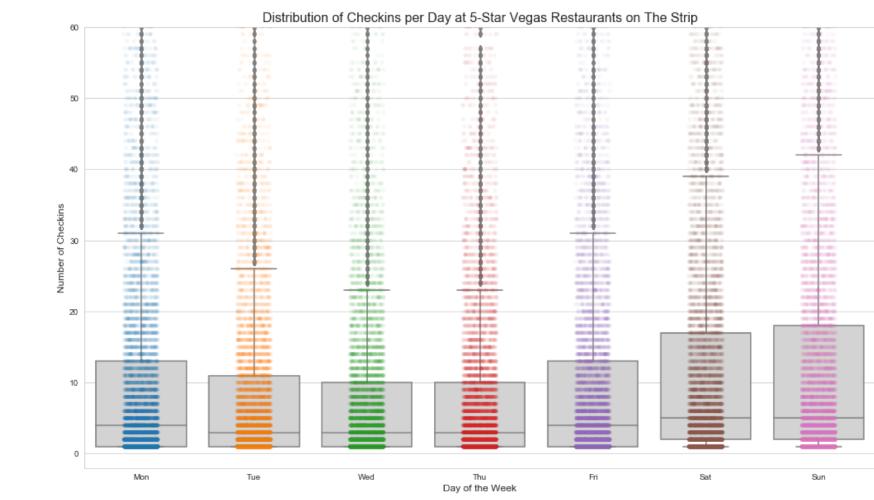
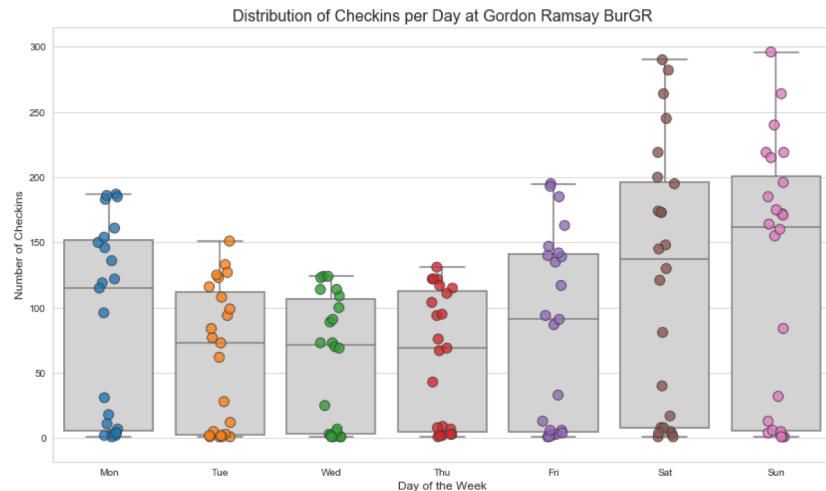


Explore the Data – Number of Reviews



('Henderson', 0.145)
('Chandler', 0.14)
('Tempe', 0.12)
('Scottsdale', 0.076)
('Pittsburgh', 0.049)
('Charlotte', 0.048)
('Phoenix', 0.032)
('Montréal', 0.031)
('Las Vegas', 0.02)
('Toronto', 0.015)

Explore the Data – Check-in Distribution

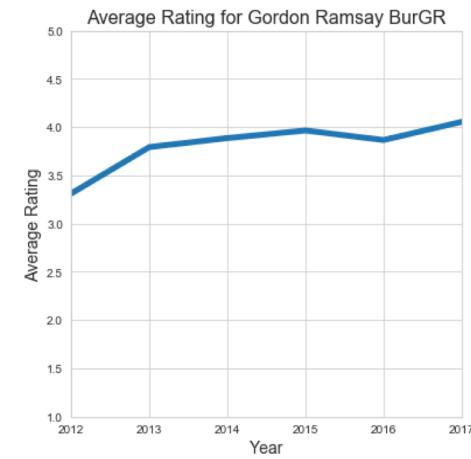
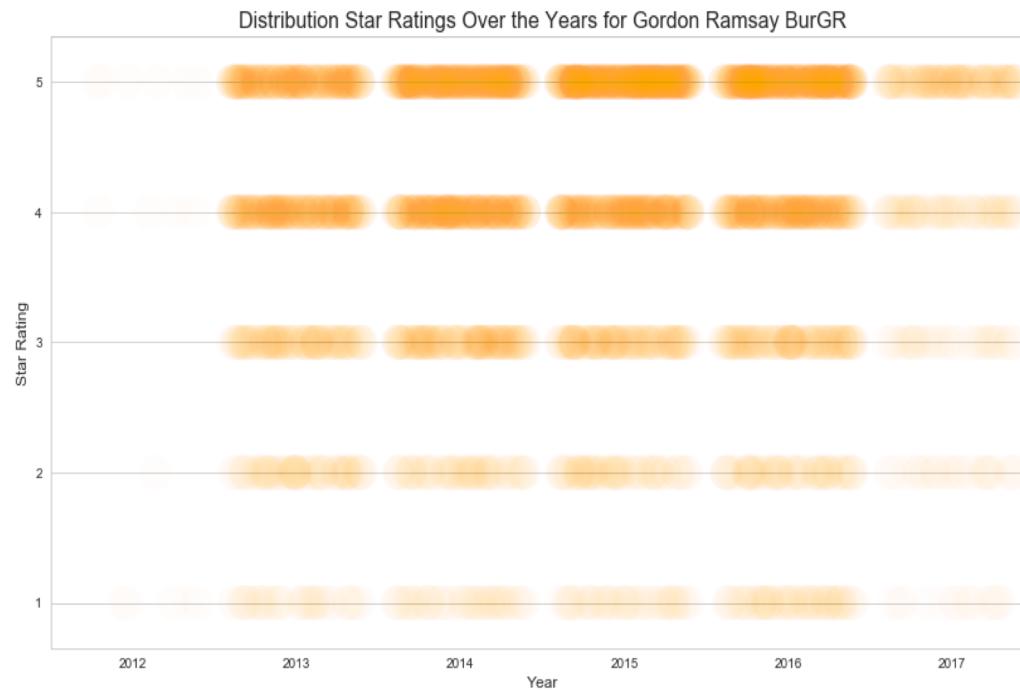


Gordon Ramsay BurGR

vs

Other Vegas Strip Restaurants

Explore the Data – Ratings Over Time

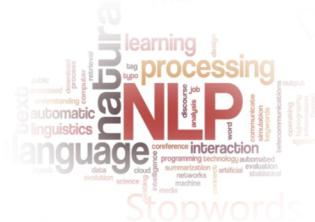


Explore the Data – NLP



Random review for Eddie's House in Scottsdale, AZ:

'I really like this place. I have been numerous amt of times and I keep wanting more. The friendly bartenders, the exciting chef (Eddie). The best part of this place besides the good food and comfort level is their ALL Night Happy Hour. Yes All night \$5 dollar specialty cocktails like an espresso martini or wines of the day. All first courses (apps) are half off too. The apps includes, lambchops (\$19), Tuna tartar with wonton chips (\$9) and so much more. Great s[ot and yearning for more since last night.'



- Remove symbols, characters, etc.
 - Remove small words
 - Form bigrams and trigrams

'numerous amt time keep want friendly bartender exciting chef good_part
comfort level night dollar specialty_cocktail espresso_martini wine day
first_course app half app include lambchop tartar wonton_chip much
great yearning last night'



Explore the Data – LDA

Random review for Eddie's House in Scottsdale, AZ:

'I really like this place. I have been numerous amt of times and I keep wanting more. The friendly bartenders, the exciting chef (Eddie). The best part of this place besides the good food and comfort level is their ALL Night Happy Hour. Yes All night \$5 dollar specialty cocktails like an espresso martini or wines of the day. All first courses (apps) are half off too. The apps includes, lambchops (\$19), Tuna tartar with wonton chips (\$9) and so much more. Great s[ot and yearning for more since last night.'



- LDA Topics Extracted

| | |
|----------------------|------|
| dinner | 0.21 |
| happy hour, drinks | 0.18 |
| cheap, good, service | 0.18 |
| lunch | 0.16 |
| buffet | 0.11 |
| healthy | 0.06 |

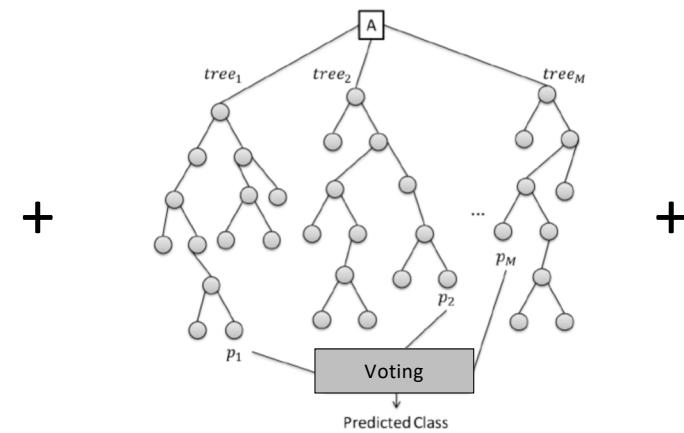
Modeling – Objective



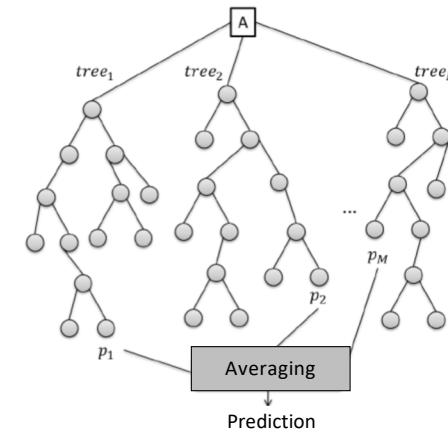
SVD

$$\begin{aligned} M_{m \times n} &= U_{m \times m} \Sigma_{m \times n} V^*_{n \times n} \\ U_{m \times m} U^*_{m \times m} &= I_m \\ V_{n \times n} V^*_{n \times n} &= I_n \end{aligned}$$

Classifier



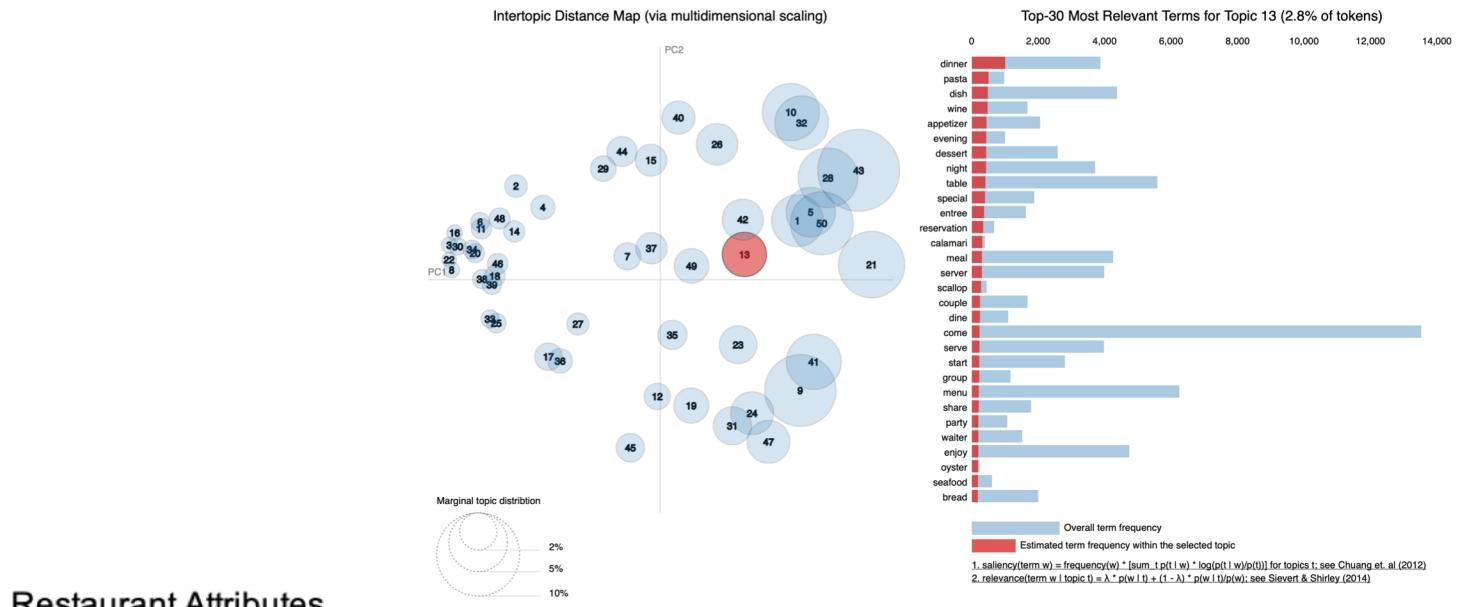
Regressor



Final Rating



Modeling – LDA Topic Modeling



Restaurant Attributes

| | Matrix 1 - Restaurant Categories | | | |
|--------------|----------------------------------|-----|------------|--|
| | Category 1 | ... | Category n | |
| Restaurant 1 | | | | |
| ... | | | | |
| Restaurant n | | | | |

+

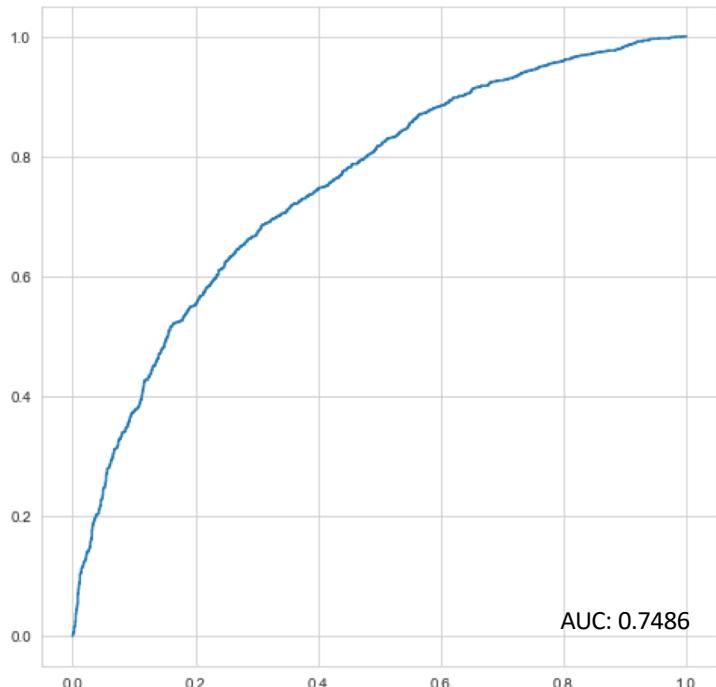
| | Matrix 2 - Restaurant Attributes | | | |
|--------------|----------------------------------|-----|-------------|--|
| | Attribute 1 | ... | Attribute n | |
| Restaurant 1 | | | | |
| ... | | | | |
| Restaurant n | | | | |

+

| | Matrix 3 - User Review Topics Aggregate | | | |
|--------------|---|-----|---------|--|
| | Topic 1 | ... | Topic n | |
| Restaurant 1 | | | | |
| ... | | | | |
| Restaurant n | | | | |



Modeling – RF Classifier and Regressor

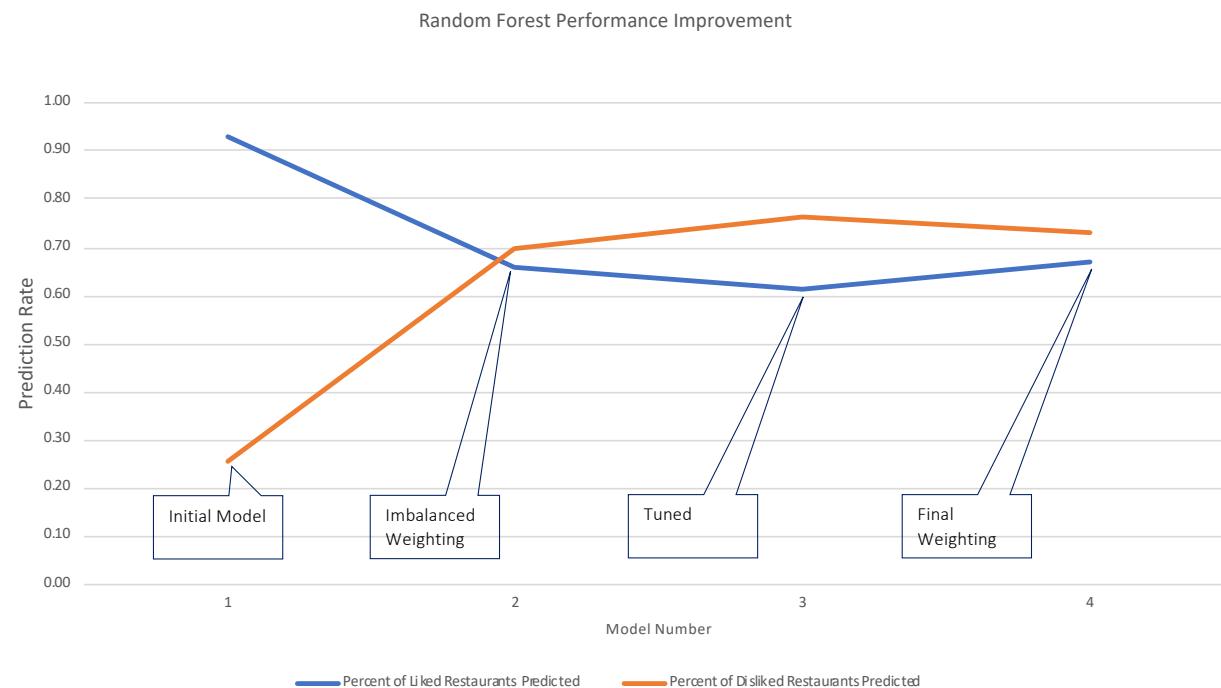


| Random Forest | | Actual | | |
|---------------|---------|---------|-------|-------|
| | | dislike | like | |
| Predicted | dislike | 299 | 870 | 1,169 |
| | like | 179 | 2,364 | 2,543 |
| | | 478 | 3,234 | |

| Random Forest Tuned Weight 1:2.7 | | Actual | | |
|-------------------------------------|---------|---------|-------|-------|
| | | dislike | like | |
| Predicted | dislike | 855 | 1,169 | 1,169 |
| | like | 847 | 2,543 | 2,543 |
| | | 1,702 | 2,010 | |

| Random Forest Regressor | | |
|----------------------------|---------|------|
| RMSE | Initial | 0.94 |
| | Tuned | 0.89 |

Modeling – Improving Random Forest



Results – Combined Predictor



| RMSE Improvement | | RMSE |
|------------------------|---------|------|
| RF Reg. | Initial | 0.94 |
| | Tuned | 0.89 |
| SVD | | 1.01 |
| SVD + RF CLF | | 0.91 |
| SVD + RF CLF + RF Reg. | | 0.74 |

Results – Next Steps



- Add Features – LDA Topic Model on entire dataset
- Add Classifier / Regression models (XGBoost, LightGBM, AdaBoost, etc.)
- Tune model ensemble through GridSearch
- Extract latent features from SVD model
- Create user profile vectors to compare to restaurant profile vectors