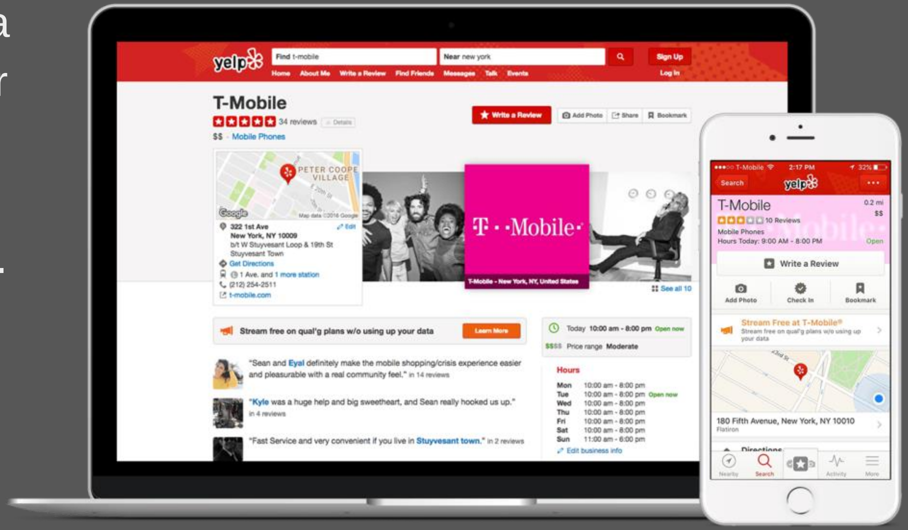
Help Phoenix Restaurants on Yelp Attract More Customers

Section 2 - Group 11 Final Presentation

Chaoying Bao, Ruoyuan Guo, Peisheng Li, Youping Wang

Background

- Nowadays, people prefer to use Yelp to find a good restaurant and leave comments for their views on restaurants they have been to.
- However, the number of restaurants on Yelp in Phoenix has decreased from 2014 to 2018.
- It comes to be worthwhile to analyze customers’ reviews since reviews indicate customers’ real thoughts on restaurants which can help merchants to improve and innovate their restaurants.

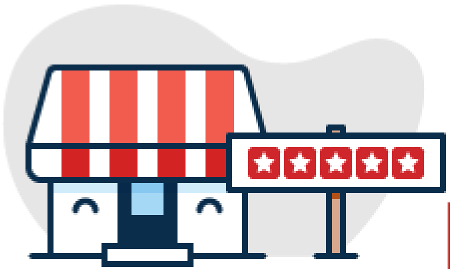


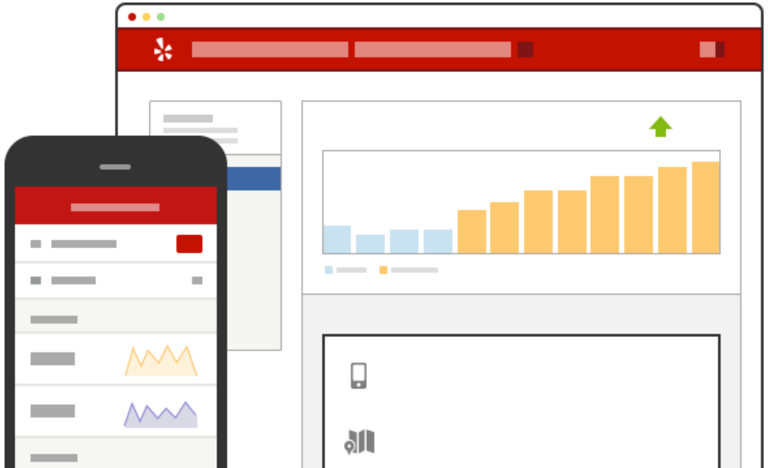
Problem Statement

- Explore the factors leading to customers’ satisfaction and dissatisfaction.
- Build a recommender system to provide more related information for customers so that restaurants can attract more target customers.

Significance

- Help users find their favorite restaurants.
- Help restaurant to find their advantages and disadvantages so that they can utilize and improve them to increase their sales.
- Increase the usage of Yelp in Phoenix.





Data Collection

Yelp Dataset: https://www.yelp.com/dataset/download

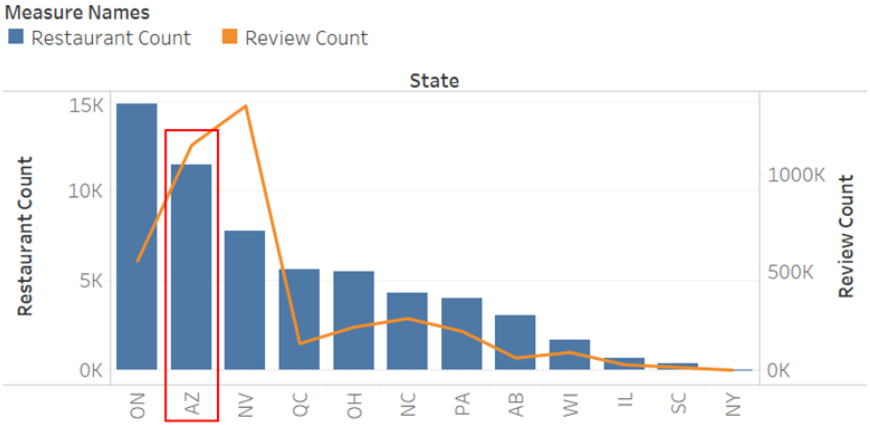
Data Description

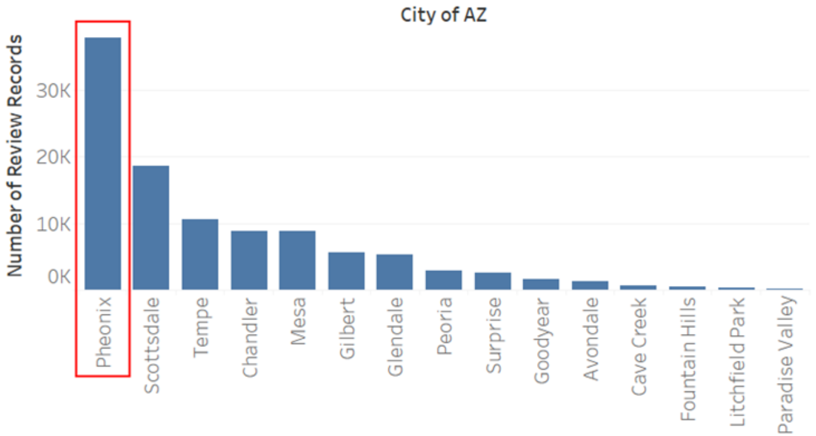
Name	Size	Type	Description
Business	131MB	JSON	Contains business data including location data, attributes, and categories.
Review	4.97GB	JSON	Contains full review text data including the user_id that wrote the review and the business_id the review is written for.
User	2.31GB	JSON	User data including the user's friend mapping and all the metadata associated with the user.

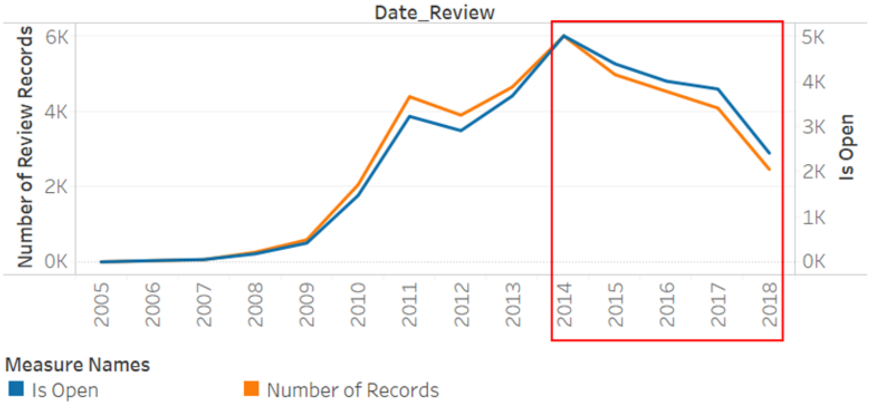
Methodology - Overall

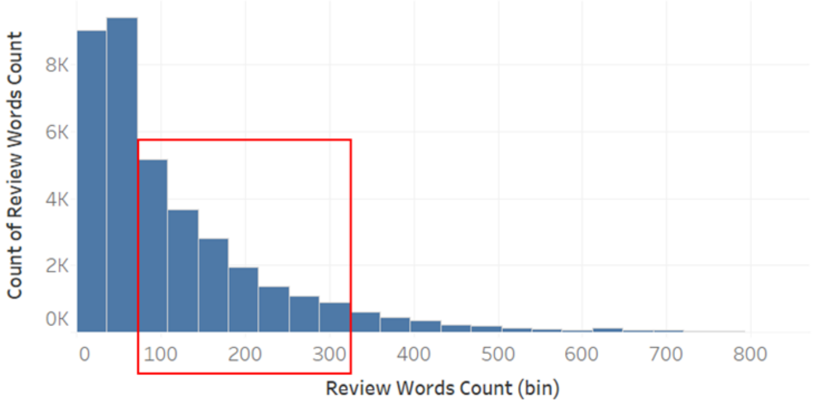
Data Collection → Descriptive Stats & Data Preprocessing → Variables Selection → Model Building → Evaluation

Descriptive Stats

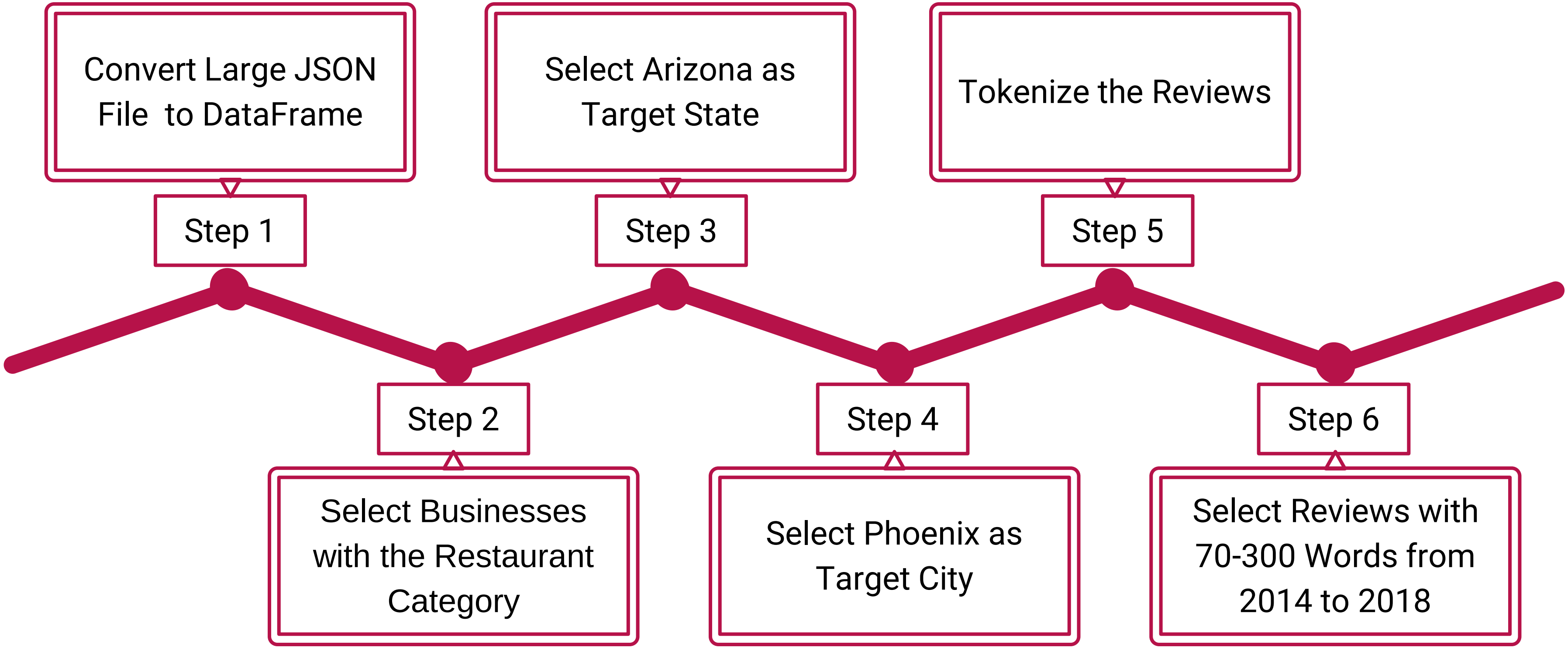








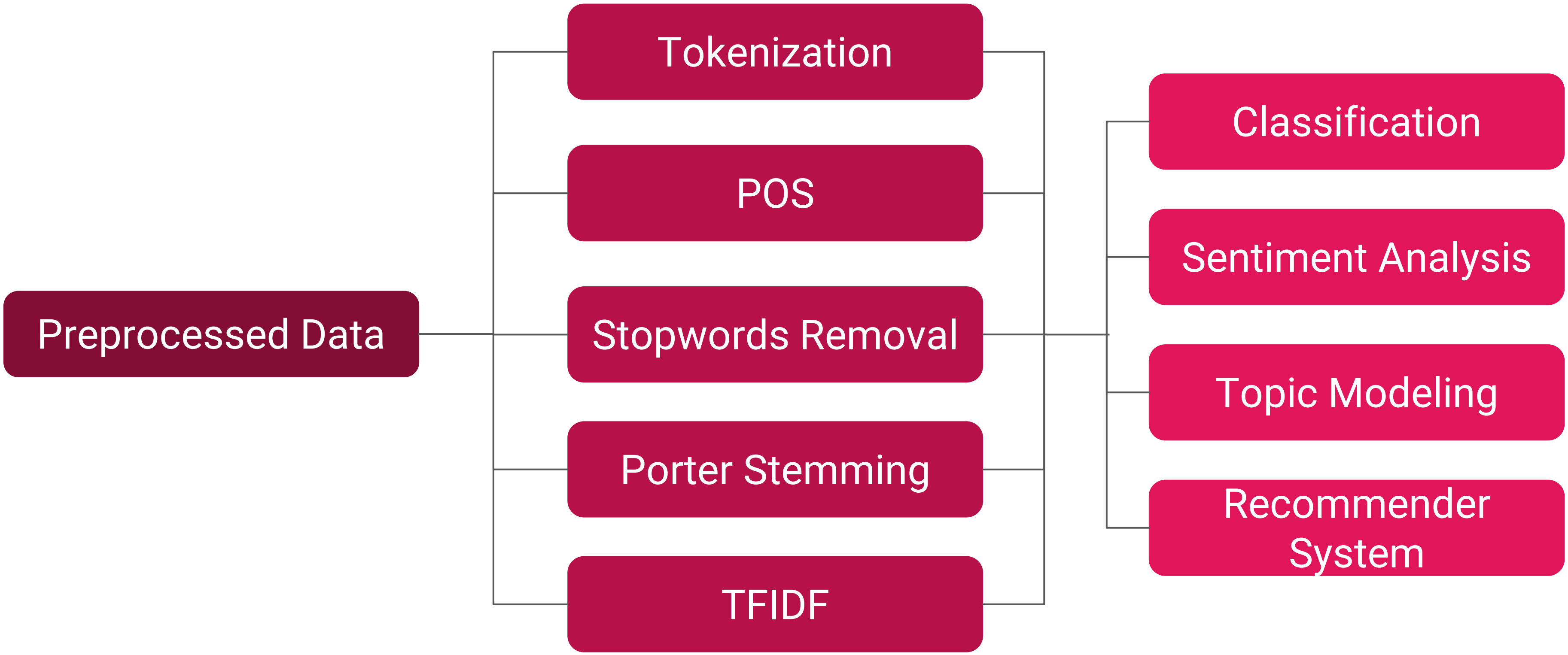
Data Preprocessing



Variables Selection

Variable Type	Variable Name	Definition	Data Type	Example
Independent Variable	useful	Number of useful votes sent by the user	Integer	0
	funny	Number of funny votes sent by the user	Integer	0
	cool	Number of cool votes sent by the user	Integer	0
	postal_code	Postal code of a restaurant	String	94107
	review_count	Number of reviews of a restaurant	Integer	1198
	review_words_counts	Number of words in a review	Integer	68
	negative	Negative score of a review calculated by NLTK Vader	Float	0.022
	neutral	Neutral score of a review calculated by NLTK Vader	Float	0.856
	positive	Positive score of a review calculated by NLTK Vader	Float	0.121
	compound	Compound score of a review calculated by NLTK Vader	Float	0.9664
	“0” - “9”	Weight of 10 topics that were modelled based on POS noun tags in a review	Float	0
Dependent Variable	rating_group	Two groups of business ratings	Integer	0

Model Building



Rating Prediction

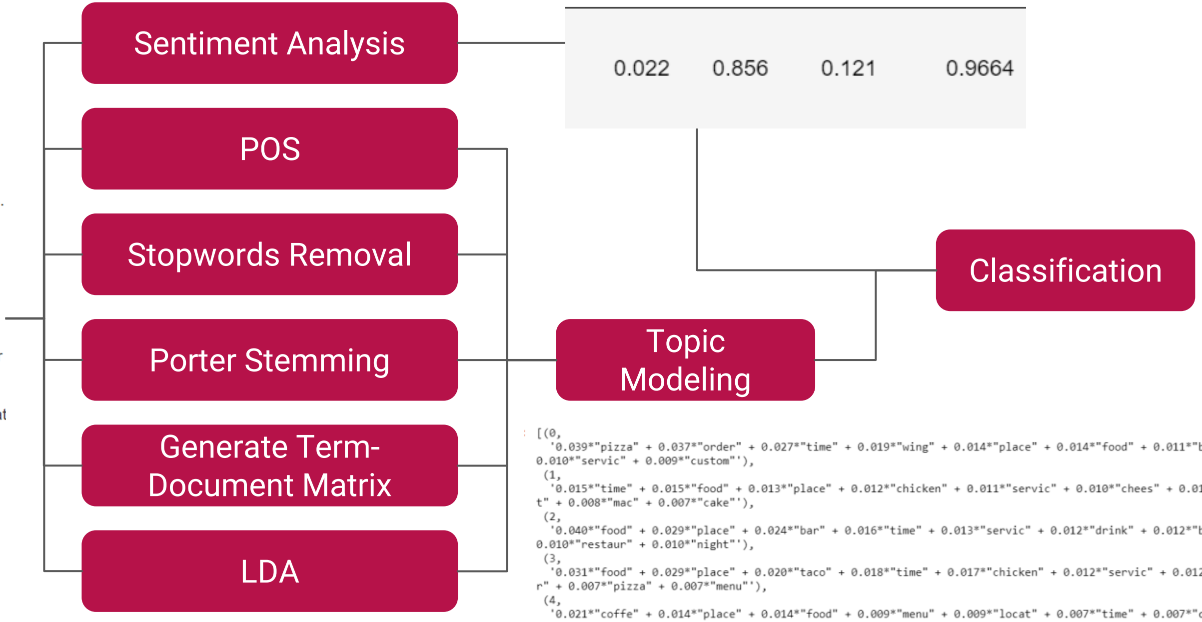
★★★★★ 7/16/2015

2 photos 19 check-ins

A must go destination in Phoenix.
Chinese - Mexican - Jamaican Fusion Fast Casual (or as I like to call it - Confusion!)
It's a Phoenix original. It's been on Diner's Dives and Drive Ins.

If it's your first time, go up to the counter (to the left of the cash registers) and ask for the orientation - they'll walk you through the ordering process, and give you samples of several of the dishes. After you do that, grab an order form and pick a 2 item combo and choose 7 9 Black and White. That's all your really need to know. :-) That's Jade Red Chicken, Jerk Chicken, Black Beans and White Rice. Jade Red Chicken is sort of their take on sweet and sour, but with a spicy twist. Jerk Chicken is Jamaican influenced (and can be quite spicy some days). The Black Beans are amazing - stir them into your rice (easier to eat both with your chopsticks when you do this). And, did I mention, you get a snickerdoodle with your order? Yep! Best snickerdoodles around, and one comes with every dish!

If you haven't been to Chino's - it's time - make a run for the border/great wall/beach!
(and make sure to read the non-PC T-shirts while you wait - funny!)



negative	neutral	positive	compound
0.022	0.856	0.121	0.9664

```
[(0,
  '0.039*"pizza" + 0.037*"order" + 0.027*"time" + 0.019*"wing" + 0.014*"place" + 0.014*"food" + 0.011*"burrito" + 0.010*"fri" + 0.010*"servic" + 0.009*"custom"'),
 (1,
  '0.015*"time" + 0.015*"food" + 0.013*"place" + 0.012*"chicken" + 0.011*"servic" + 0.010*"chees" + 0.010*"cream" + 0.009*"donut" + 0.008*"mac" + 0.007*"cake"'),
 (2,
  '0.040*"food" + 0.029*"place" + 0.024*"bar" + 0.016*"time" + 0.013*"servic" + 0.012*"drink" + 0.012*"beer" + 0.011*"pizza" + 0.010*"restaur" + 0.010*"night"'),
 (3,
  '0.031*"food" + 0.029*"place" + 0.020*"taco" + 0.018*"time" + 0.017*"chicken" + 0.012*"servic" + 0.012*"salad" + 0.009*"order" + 0.007*"pizza" + 0.007*"menu"'),
 (4,
  '0.021*"coffe" + 0.014*"place" + 0.014*"food" + 0.009*"menu" + 0.009*"locat" + 0.007*"time" + 0.007*"chicken" + 0.006*"ice" + 0.006*"day" + 0.005*"servic"'),
 (5,
  '0.031*"place" + 0.027*"burger" + 0.021*"food" + 0.021*"time" + 0.012*"restaur" + 0.010*"menu" + 0.010*"fri" + 0.009*"sauc" + 0.009*"servic" + 0.008*"flavor"'),
 (6,
  '0.037*"place" + 0.026*"food" + 0.025*"servic" + 0.020*"time" + 0.010*"meal" + 0.010*"staff" + 0.008*"locat" + 0.008*"breakfast" + 0.008*"restaur" + 0.007*"sushi"'),
 (7,
  '0.029*"food" + 0.018*"order" + 0.018*"servic" + 0.018*"place" + 0.014*"time" + 0.012*"drink" + 0.010*"hour" + 0.009*"restaur" + 0.009*"server" + 0.008*"beer"'),
 (8,
  '0.032*"food" + 0.020*"place" + 0.013*"time" + 0.012*"sandwich" + 0.012*"servic" + 0.011*"restaur" + 0.009*"bread" + 0.008*"price" + 0.007*"lunch" + 0.007*"thing"'),
 (9,
  '0.037*"place" + 0.020*"food" + 0.015*"servic" + 0.014*"time" + 0.011*"breakfast" + 0.011*"coffe" + 0.010*"lunch" + 0.010*"sandwich" + 0.009*"egg" + 0.008*"menu"')]
```

Content-based Recommender System

Approach 1- Recommending based on the similarity of item contents[3]

Step 1: Restaurant names and reviews_words were grouped by the restaurants' names and put into a new DataFrame.

title	review_words
1000 Degrees Neapolitan Pizzeria	[absolut, best, pizza, arizona, use, huge, fan...
1130 The Restaurant	[total, disappoint, way, restaur, go, lettuc, ...
2601 on Central	[kind, gener, vascil, go, give, benefit, new, ...
32 Shea	[restaur, sold, reserv, judgement, month, see,...
3on Smith Cafe	[new, neighborhood, hot, spot, sure, awesom, p...
40th Street Cafe	[love, biscuit, gravi, side, bacon, fast, hot,...

Step 2: Use cosine similarity function to analyze similarities between restaurants

```
array([[1.        , 0.08746385, 0.07888007, ..., 0.10516693, 0.52821066,
        0.12681706],
       [0.08746385, 1.        , 0.19087136, ..., 0.03331894, 0.27904066,
        0.2236817 ],
       [0.07888007, 0.19087136, 1.        , ..., 0.09929231, 0.21925814,
        0.22616243],
       ...,
       [0.10516693, 0.03331894, 0.09929231, ..., 1.        , 0.12304447,
        0.11809198],
       [0.52821066, 0.27904066, 0.21925814, ..., 0.12304447, 1.        ,
        0.29922899],
       [0.12681706, 0.2236817 , 0.22616243, ..., 0.11809198, 0.29922899,
        1.        ]])
```

Step 3: Build the function that takes names as input, and returns the top 5 similar restaurants

```
indices = pd.Series(df_3.index)
recommended=[]

def recommendations(title, cosine_sim = cosine_sim):


    idx = indices[indices == title].index[0]

    score_series = pd.Series(cosine_sim[idx]).sort_values(ascending = False)

    top_5_indexes = list(score_series.iloc[1:6].index)
    for i in top_5_indexes:
        recommended.append(list(df_3.index)[i])

    return recommended
```

```
recommendations('Pita Bistro')
```

```
['Chipotle Mexican Grill',
 'Chino Bandido',
 'Pita Jungle',
 'Cowboy Ciao',
 'Valle Luna']
```

Content-based Recommender System

Approach 2 - Recommending based on the prediction of user's rating [4]

Step 1: Create item profile with TFIDF scores (2993 restaurants * 6744 keywords)

	pita	kabab	falafel	baklava	eastern	matt
--g-a85VwrdZJNf0R95GcQ	0.624557	0.560562	0.378748	0.276890	0.274918	0.0
0KpoeCt1E-SQsUBwtjLAEw	0.344505	0.000000	0.626751	0.000000	0.000000	0.0
0N2y8rNxbet6p4UIBWTOrw	0.266268	0.000000	0.000000	0.196745	0.000000	0.0
1E1Qp9HWSZmqruir3sTeKw	0.378842	0.000000	0.306320	0.000000	0.000000	0.0
3x45Q9c5G6VBicedNKrXxQ	0.674633	0.000000	0.000000	0.000000	0.000000	0.0

Step 2: Create user profile with users' rating (2993 restaurants * 14510 users)

user_id / business_id	--2HUmLkcNHZp0xw6AMBPg	--41c9TI0C9OGewIR7Qyzg	--4q8EyqThydQm-eKZpS-A
--g-a85VwrdZJNf0R95GcQ	0.0	0.0	0.0
-050d_Xlor1NpCuWkblVaQ	0.0	0.0	0.0
-0WegMt6Cy966qIDKhu6jA	0.0	0.0	0.0
-0alra_B6iALlfqAriBSYA	0.0	0.0	0.0
-0tgMGI7D9B10YjSN2ujLA	0.0	0.0	0.0

Step 3: Generate user's preference profile (14510 users * 6744 keywords)

user_id	pita	kabab	falafel	baklava	eastern	matt	breakfast	wait	hash	brown
--2HUmLkcNHZp0xw6AMBPg	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
--41c9TI0C9OGewIR7Qyzg	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
--4q8EyqThydQm-eKZpS-A	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
--4rAAfZnEIAKJE80aliYg	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
--CluK7sUpaNzalLAIHJKA	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Step 4: Compute the predicted rating (2993 restaurants*14510 users)

user_id / business_id	--2HUmLkcNHZp0xw6AMBPg	--41c9TI0C9OGewIR7Qyzg	--4q8EyqThydQm-eKZpS-A
--g-a85VwrdZJNf0R95GcQ	0.0	0.0	0.0
-050d_Xlor1NpCuWkblVaQ	0.0	0.0	0.0
-0WegMt6Cy966qIDKhu6jA	0.0	0.0	0.0
-0alra_B6iALlfqAriBSYA	0.0	0.0	0.0
-0tgMGI7D9B10YjSN2ujLA	0.0	0.0	0.0

Results - Rating Prediction

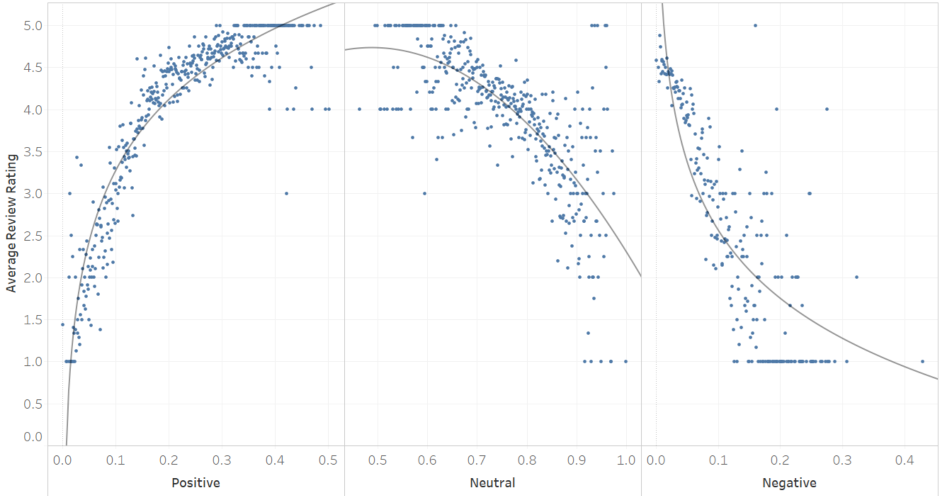
Results for output field RatingGroup

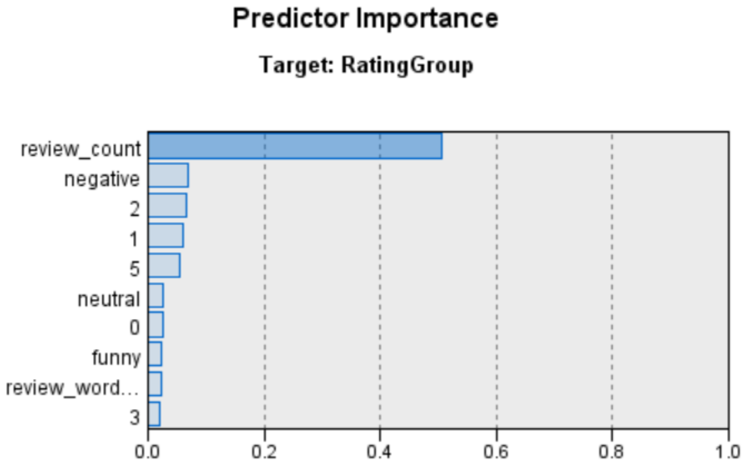
Comparing $C-RatingGroup with RatingGroup

'Partition'	1_Training		2_Testing	
Correct	6,337	90.93%	1,459	80.39%
Wrong	632	9.07%	356	19.61%
Total	6,969		1,815	

Coincidence Matrix for $C-RatingGroup (rows show actuals)

'Partition' = 1_Training	0	1
0	3,214	270
1	362	3,123
'Partition' = 2_Testing	0	1
0	173	104
1	252	1,286





Results - Recommender System (Approach 1)

```
recommendations('Pita Bistro')
```

```
['Chipotle Mexican Grill',
 'Chino Bandido',
 'Pita Jungle',
 'Cowboy Ciao',
 'Valle Luna']
```

Pita bistro

Recommend

Chipotle Mexican Grill

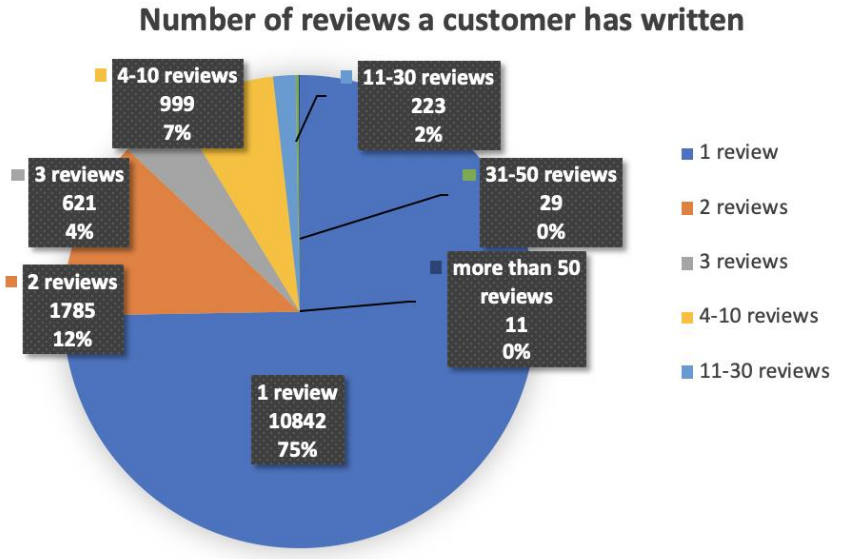
Pita Jungle

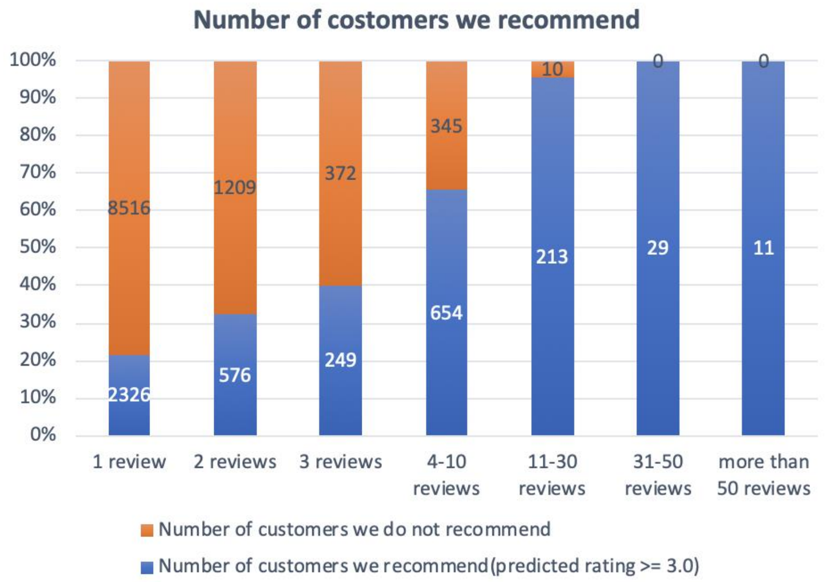
Valle Luna

Chino Bandido

Cowboy Ciao

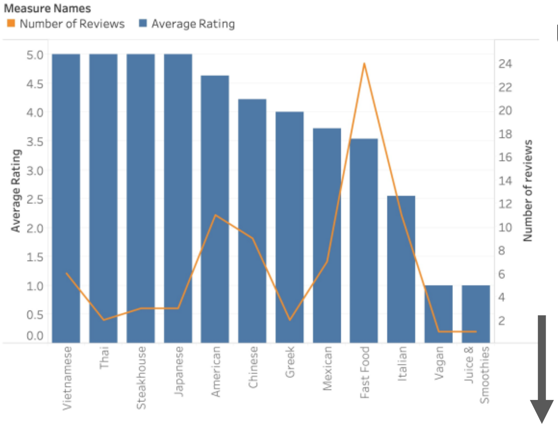
Results - Recommender System (Approach 2)



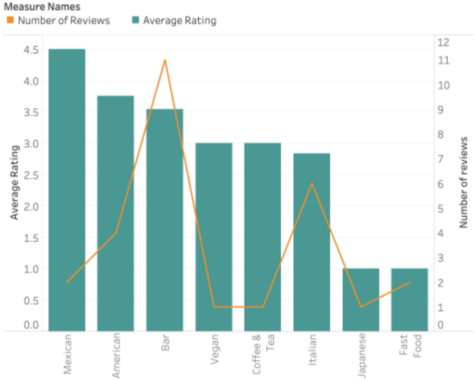


- There are 14510 customers in our dataset before selecting reviews with 70-300 words and special years, but 10842 customers (75%) only have one review.
- As the user provides more inputs, the engine becomes more and more effective.
- For users who created less than 10 reviews in our user profile, our recommender system with the first approach recommends better.

Results - Recommender System (Approach 2)



User 1: Taylor
- 80 records in user profile
- Often writes reviews for Fast Food and American restaurants
- Loves Vietnamese, Thai, Steakhouse, Japanese and American foods.



User 2: Cris
- 28 records in user profile
- Often writes reviews for Bar and Italian restaurants
- Loves Mexican and American food, enjoys bar as well

Restaurants we recommend

Restaurant Name	Restaurant Categories		
Speedy Street Taco	Mexican	Food Stands	
Lone Star Steakhouse	Steakhouses	Seafood	American (Traditional)
Dairy Queen	American (New)		
Pho Noodles	Vietnamese	Noodles	
Rita's Mexican Food & Mariscos	Seafood	Mexican	

Restaurants we recommend

Restaurant Name	Restaurant Categories							
Tarbell's	American (New)	Italian	Desserts	Nightlife	Bars			
Dick's Hideaway	New Mexican Cuisine	Breakfast & Brunch	American (Traditional)	Tex-Mex	Nightlife	Mexican	Bars	
Buffalo Wild Wings	American (New)	Chicken Wings	American (Traditional)	Sports Bars	Nightlife	Bars		
The Capital Grille	Seafood	American (Traditional)	Wine Bars	Nightlife	Steakhouses	Bars		
Rosie McCaffrey's Irish Pub & Restaurant	Irish Pub	Nightlife	Pubs	Irish		Bars		
16th Street Sports Bar & Grill	Nightlife	American (Traditional)	Sports Bars	Wine & Spirits	Food	Burgers	Beer	Bars

Conclusions

For Restaurants

- Customers' sentiment in reviews is significantly related to the actual rating they give the restaurants, so restaurants shouldn't ignore customers' reviews.
- Based on the model, the number of reviews a restaurant receives, the negative score in a review, and topic 2, 1, and 5 are important predictors for the restaurant rating.

For Customers

- For new users or users who have limited records, the recommender system (approach 1) could be used to recommend restaurants based on the similarity of each restaurant's reviews.
- For other users with enough records, the recommender system (approach 2) should be used, since it takes user's preference into consideration.

Recommendation

For Restaurants

- Topic 2, 1, and 5 all mention food, time, place, and service. Based on our model, the more related to these three topics, the more likely a restaurant to get low rating. Therefore, those restaurants who get low rating should pay more attention to these aspects and improve them.

For Customers

Since the number of reviews is the most important predictor for the restaurant rating, customers are recommended to write reviews about their opinions, which will contribute to not only the restaurant's improvement, but also the customers' future experience.

Project timeline

2/29 - 3/4

Project Proposal

- Choose an application domain
- Find a publicly available dataset
- Identify a NLP problem
- Apply appropriate techniques to solve this problem

3/4 - 3/11

Data Preprocessing

- Data cleaning
- Data transformation
- Data reduction

3/11 - 4/1

Model building

- Writing code to build models
- Compare the performance of multiple techniques

4/1 - 4/15

Evaluation

- Evaluate the results of different restaurants
- Gain business insights

4/15 - 4/22

Final Report

- Write final report
- Prepare for final presentation

References

[1] https://github.com/karantyagi/Restaurant-Recommendations-with-Yelp
[2] https://monkeylearn.com/text-mining/
[3]https://towardsdatascience.com/how-to-build-from-scratch-a-content-based-movie-recommender-with-natural-language-processing-25ad400eb243
[4]https://www.analyticsvidhya.com/blog/2015/08/beginners-guide-learn-content-based-recommender-systems/

Thanks!

Chaoying Bao, Ruoyuan Guo, Peisheng Li, Youping Wang