



The Language of Food

How do we leverage user feedback to improve dining experiences?

By Schubert Laforest



Schubert Hervey Laforest

Recent Data Science Immersive Grad, General Assembly
B.A. Linguistics, Political Science & Marketing, Concordia
University

I live at the intersection of data, social science and the internet.

 [schubline](#) 

 [Schubert Hervey
Laforest](#) 

 [schubline](#) 

Some Places I've Worked at





Food is Important to Us

We don't just eat food, we like to talk about it, a lot...

Zumas Tex Mex Grill

Claimed

345 reviews

Write a Review

Add Photo

Share

Save

Tex-Mex, Salad, Sandwiches

Located in Faneuil Hall Marketplace, 7 N Market St, Boston, MA 02109

Get Directions

Aquarium Station and 2 more stations

(617) 367-9114

zumatexmexgrill.com

Message the business

Send to your Phone

It's pretty cool heading downstairs underneath Faneuil Hall and finding a deal in such an otherwise expensive area of Beantown." In 32 reviews

The surprise MVP here was the coconut shrimp as they were perfectly fried and had this crust which retained the very pronounced coconut flavor." In 6 reviews

\$8 Coconut Shrimp

I also got a fried veggie Tex Mex spring roll which was fun and something I've never had before." In 7 reviews

Order Delivery or Takeout

Delivery Takeout

Fee \$3.99

Delivery Min \$0

Arrives by 11:53 am

Delivery Address

reddit r/food

Search Reddit

LOG IN SIGN UP Visit Old Reddit

Posts Quick Search

VIEW SORT HOT

Posted by u/andoh12 Baked Cracker 1 year ago

[MOD POST - PSA **UPDATE**] PLEASE SEE THE FOLLOWING RULE CHANGES FOR POSTING LINKS IN THIS SUB. Discussion

Hello again! We have seen the quality of this sub grow and we have erased a lot of the spam and reposted/stolen content, thanks to you wonderful users! However, we still have some gaps that we need to close, mainly helping to define what [OC] means. So we have added some tags to clarify. Going forward, all link posts must have **only one** of the following tags in the title:

[Homemade] - You made it or someone you know made it. This includes all the "I made, from scratch, had a go at, my try, etc" and those phrases are not needed because this defines that it is made by you or yours. You either made this from ingredients or you assembled this dish at home, from pre-made ingredients. If you made it from scratch or some parts from scratch, you need to be able to discuss what went into the preparation of this dish, including recipes.

43 Comments Share

Posted by u/anibanani 4 hours ago

[Homemade] vietnamese crispy pork belly Image

COMMUNITY DETAILS

r/food

13.6m Subscribers

7.2k Online

Cooking, restaurants, recipes, food network, foodies, talk about it here!

SUBSCRIBE

CREATE POST

ADVERTISEMENT

Image

We Take Pictures of Food,
a Lot ...





And Watch Shows About
It, a Lot...





How Do We Improve Dining Experiences?

For **Customers**

*Think about your **Best** and **Worst** dining
experiences*



How Do We Improve Dining Experiences?

For **Businesses**

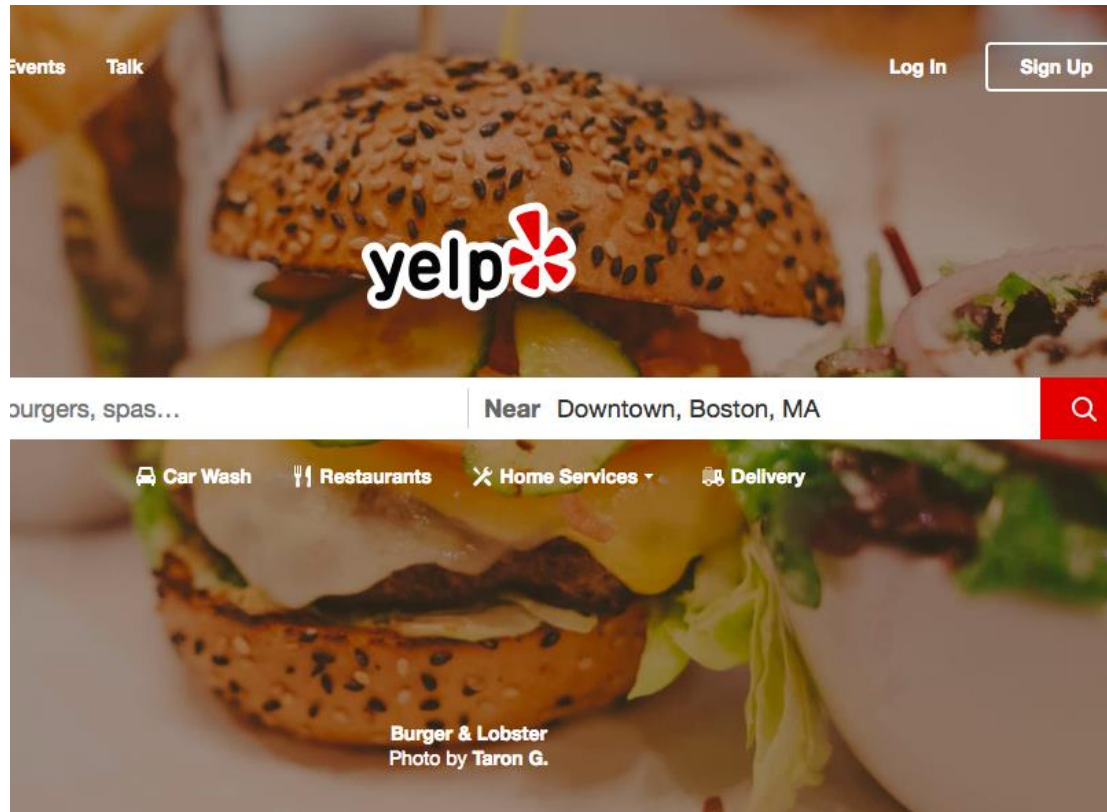


Data!!!

(and a lot of science)

Turning the Business Question into a Data Question

- **Business Question:** How do we leverage **user feedback** to improve dining experiences?
- **Data Question:** How can we use **Recommender Systems** and **Natural Language Processing** to improve dining experiences?
- Replicating a grandmother's intuition and Gordon Ramsay's business acumen.



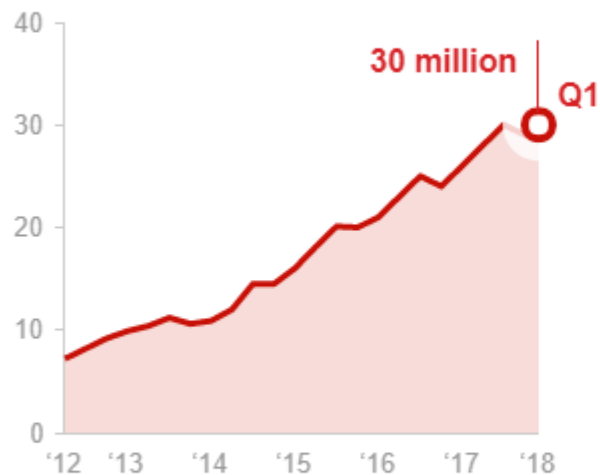
Quick Facts on Yelp Dataset (Kaggle)

- “Small” subset of Yelp Reviews
- **5,200,000** user reviews
- Information on **174,000** businesses
- Spans **11 metropolitan areas**.
- **Stat: A one-star increase in Yelp rating leads to a 5-9 percent increase in revenue**

“ 82% of adults consult online reviews before buying something for the first time, 40% always do ”

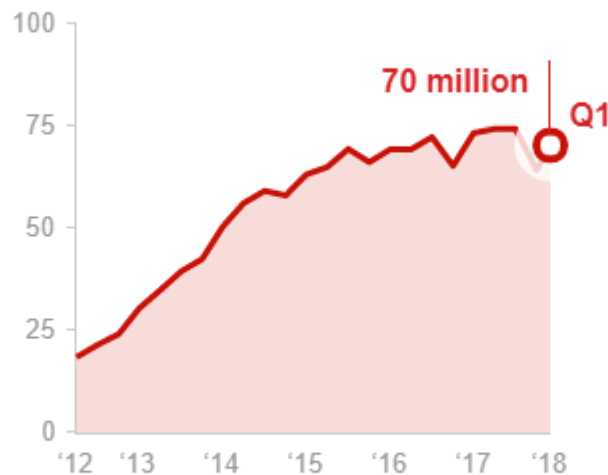
- Fernbach et. al

Average monthly mobile app unique users



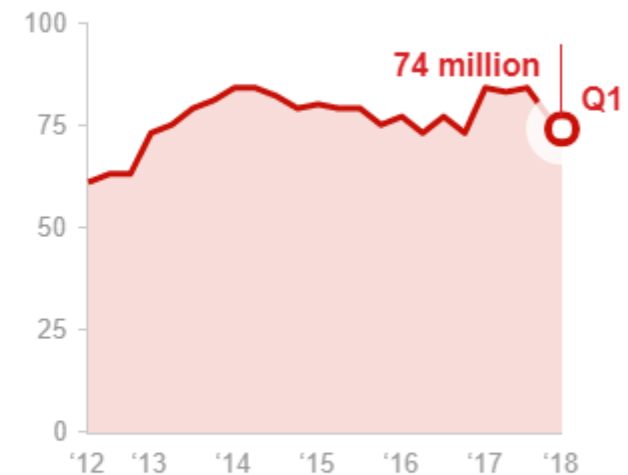
Calculated as the number of unique devices accessing the app on a monthly average basis over a given three-month period, according to internal Yelp logs.

Average monthly mobile web unique visitors



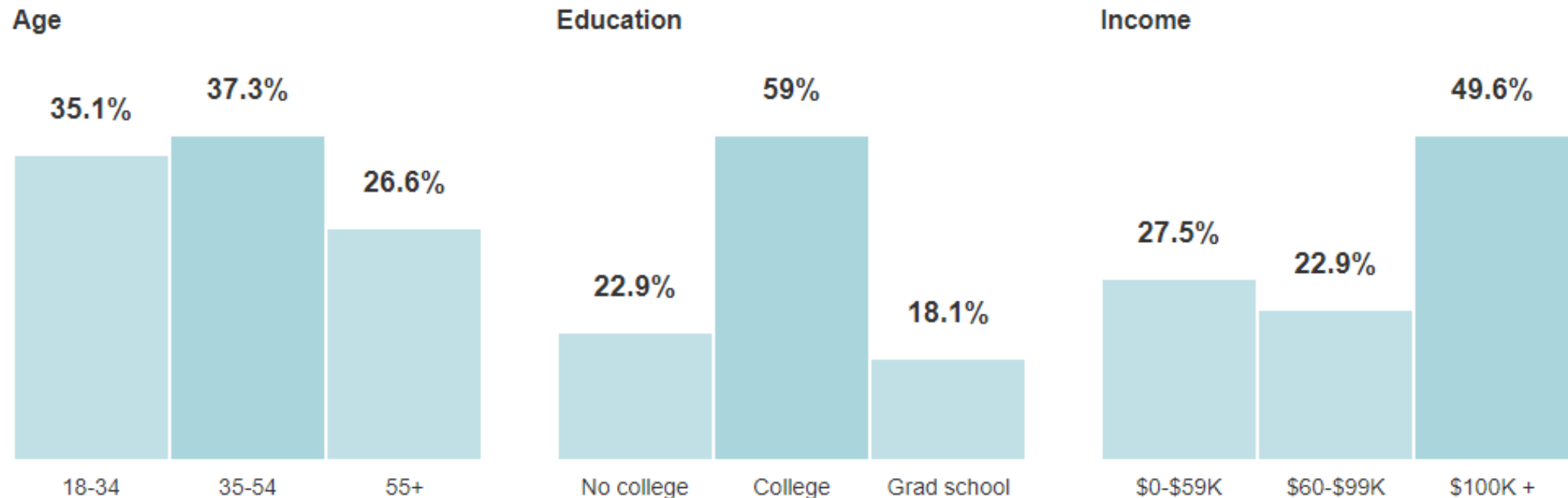
Calculated as the number of "users," as measured by Google Analytics, accessing Yelp via mobile website on a monthly average basis over a given three-month period.

Average monthly desktop unique visitors



Calculated as the number of "users," as measured by Google Analytics, accessing Yelp via desktop computer on an average monthly basis over a given three-month period.

US Yelp User Demographics



Source: comScore. Age and income data via Media Metrix Multi-Platform report as of March 2018.

Education data via Plan Metrix report as of February 2018.

How Do We Make Sense of All This Data and Solve our Business Question?

Hosting Data on AWS

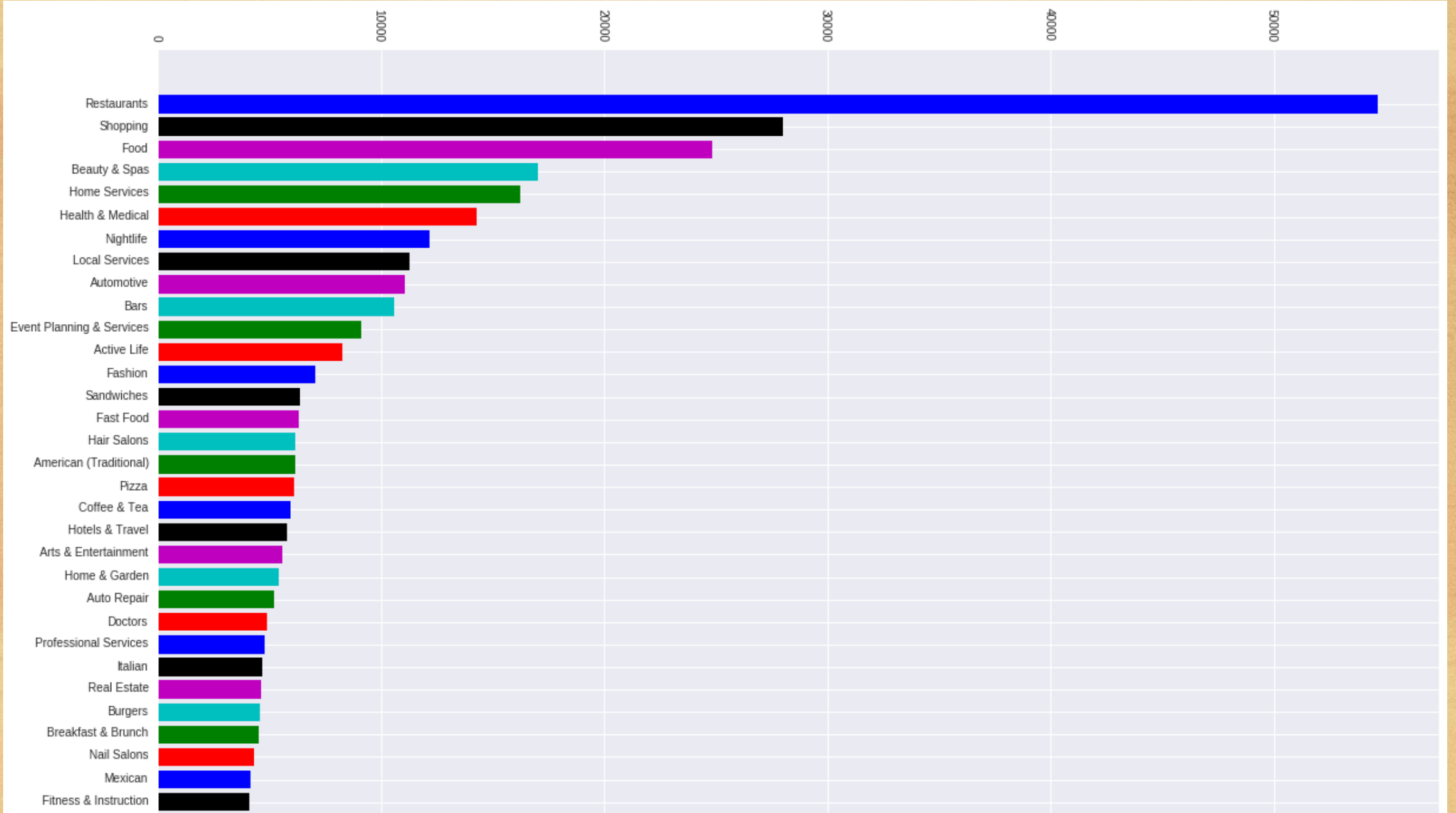
Exploratory Data Analysis

Build a baseline Recommender System

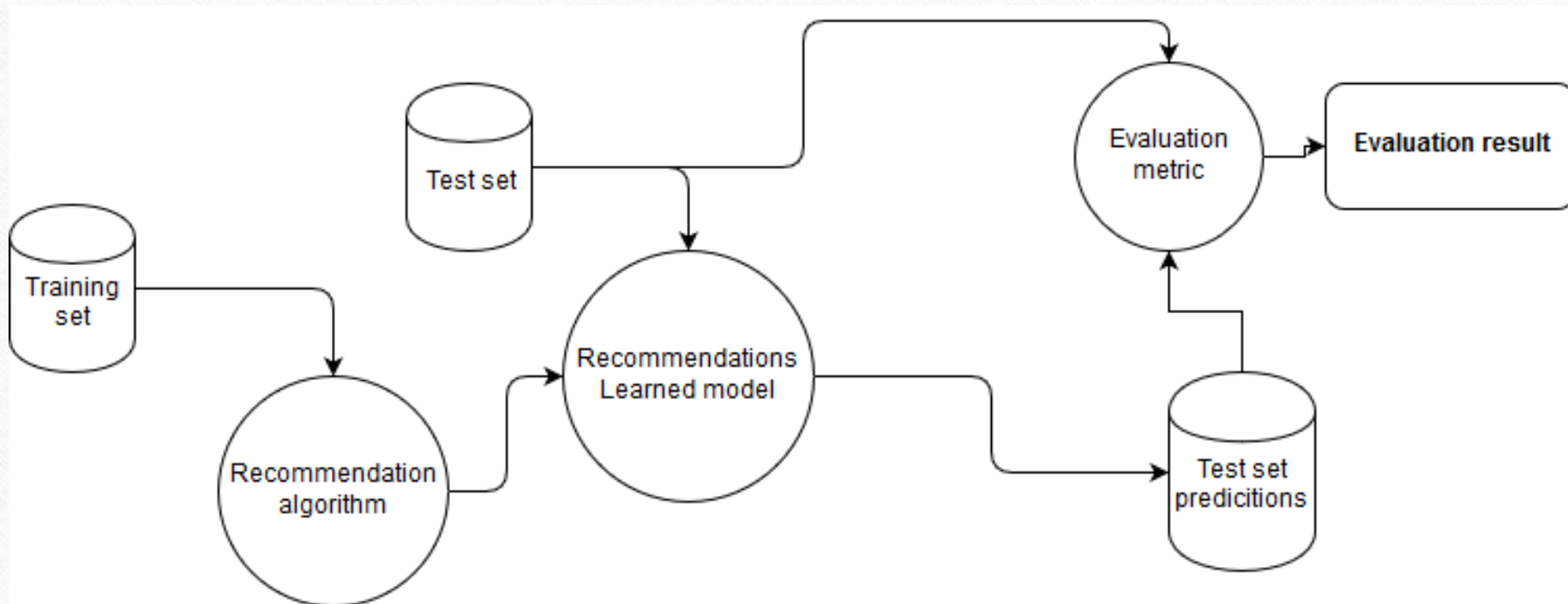
Dive into the language of reviews

Try some creative approaches

Review Count by Category



Process of Building a Recommender System



Recommender Based on Star Ratings

- Built on a subsection of our Data
- **Star-Ratings = Explicit Data**
- Item-Based Recommender: Basically, it suggests similar restaurants





That's Kinda cool!

But what are the possible the shortcomings of this approach?

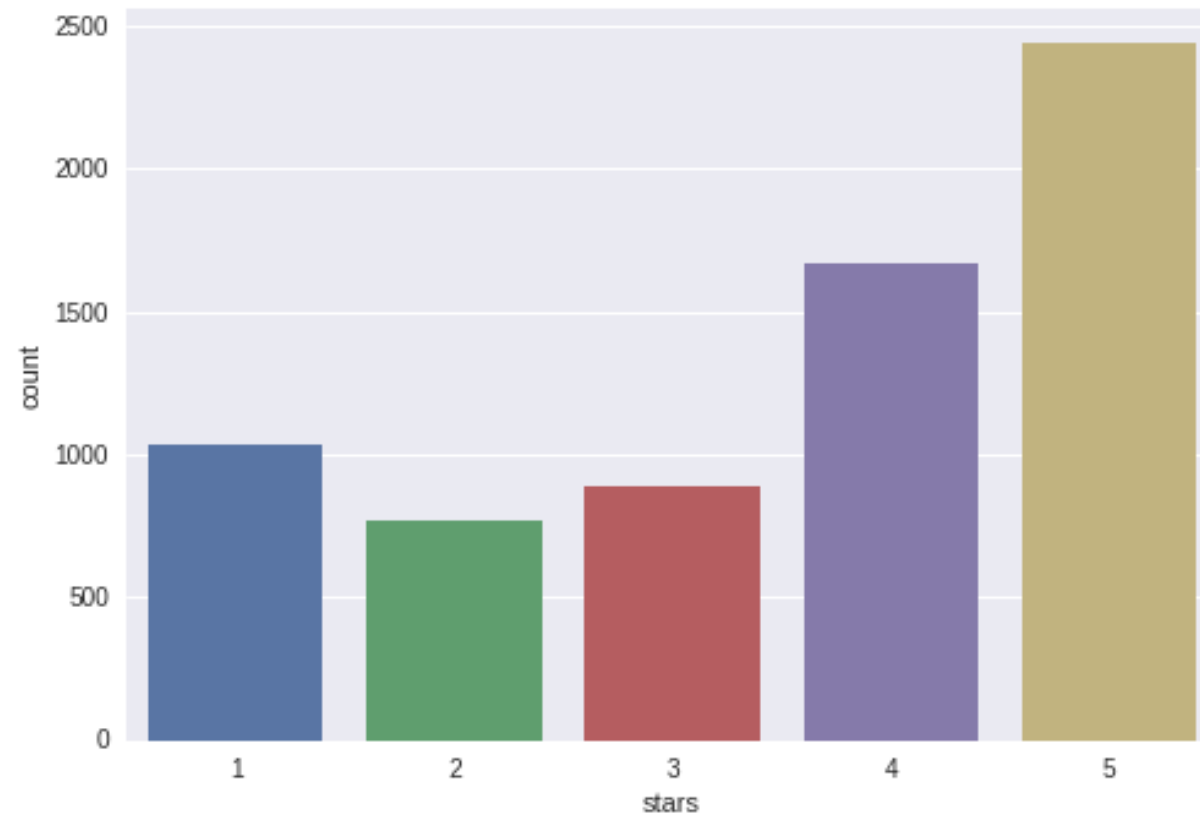
“A model is a simplification of reality”

– Matt Brems

- Every person has their own mental model of how they choose and subsequently evaluate what and where they want to eat
- We all have different palates and preferences.
- Small sample size
- **Heavily skewed data and lopsided databases**

Can we build a better recommender based off of review text?

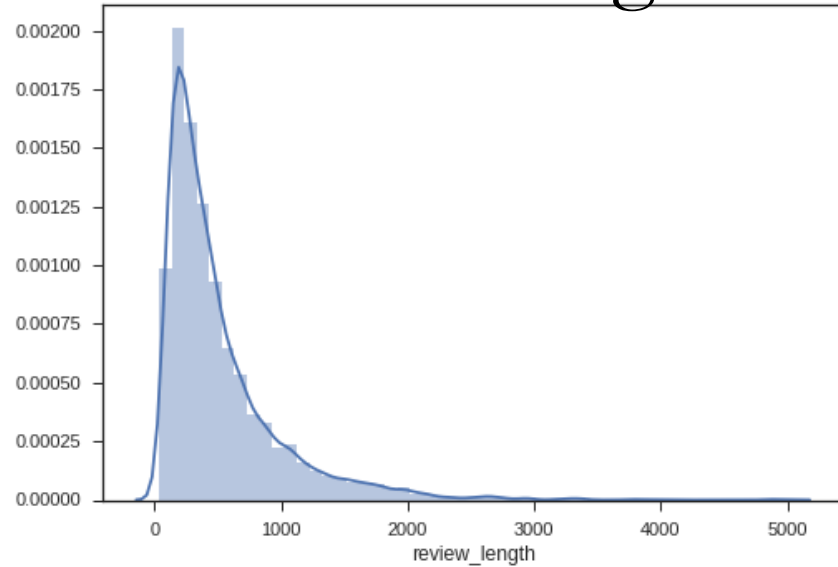
Distribution of Star Score



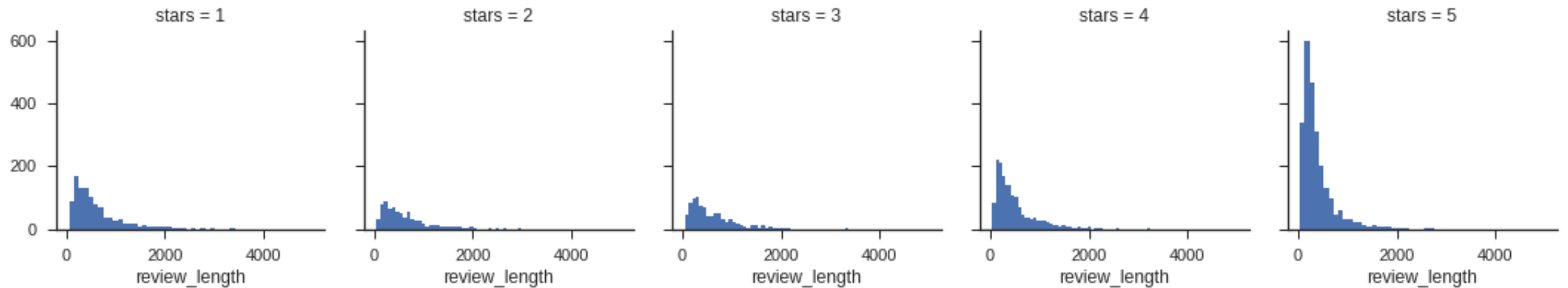
A Closer Look at Review Text

- 'First time **great** food and **service** will be **going back** had the hot dogs more food than 1 person can eat'
- '**Slow service** every time, and **food has been cold twice. Poorly managed** version of this chain. The Rock Hill location is fantastic!'

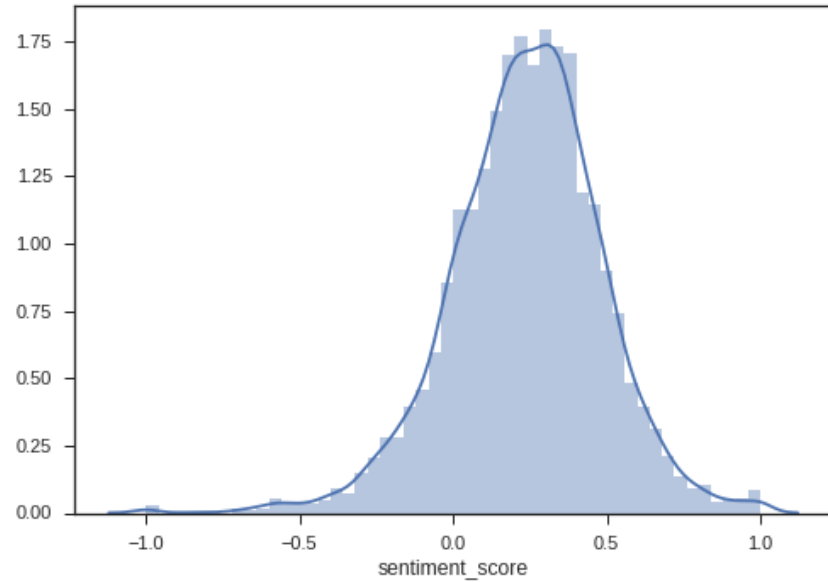
Review Length



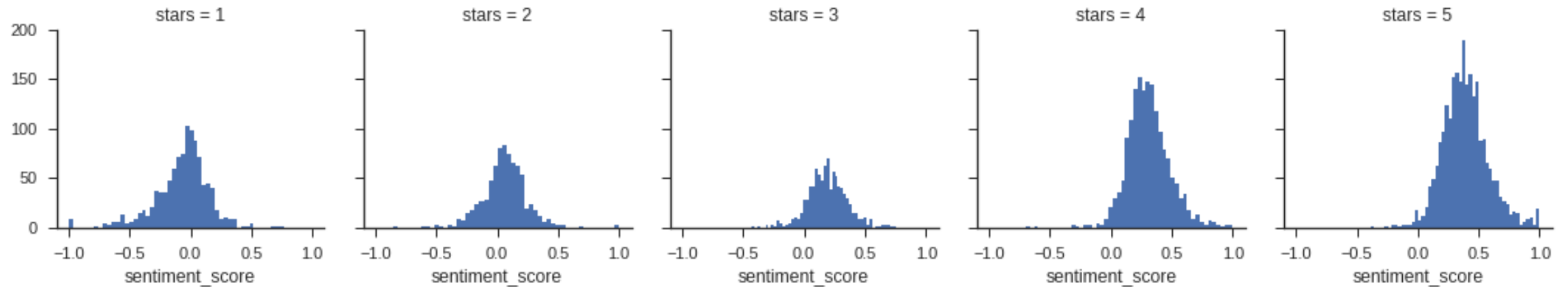
Review Length by Star Rating

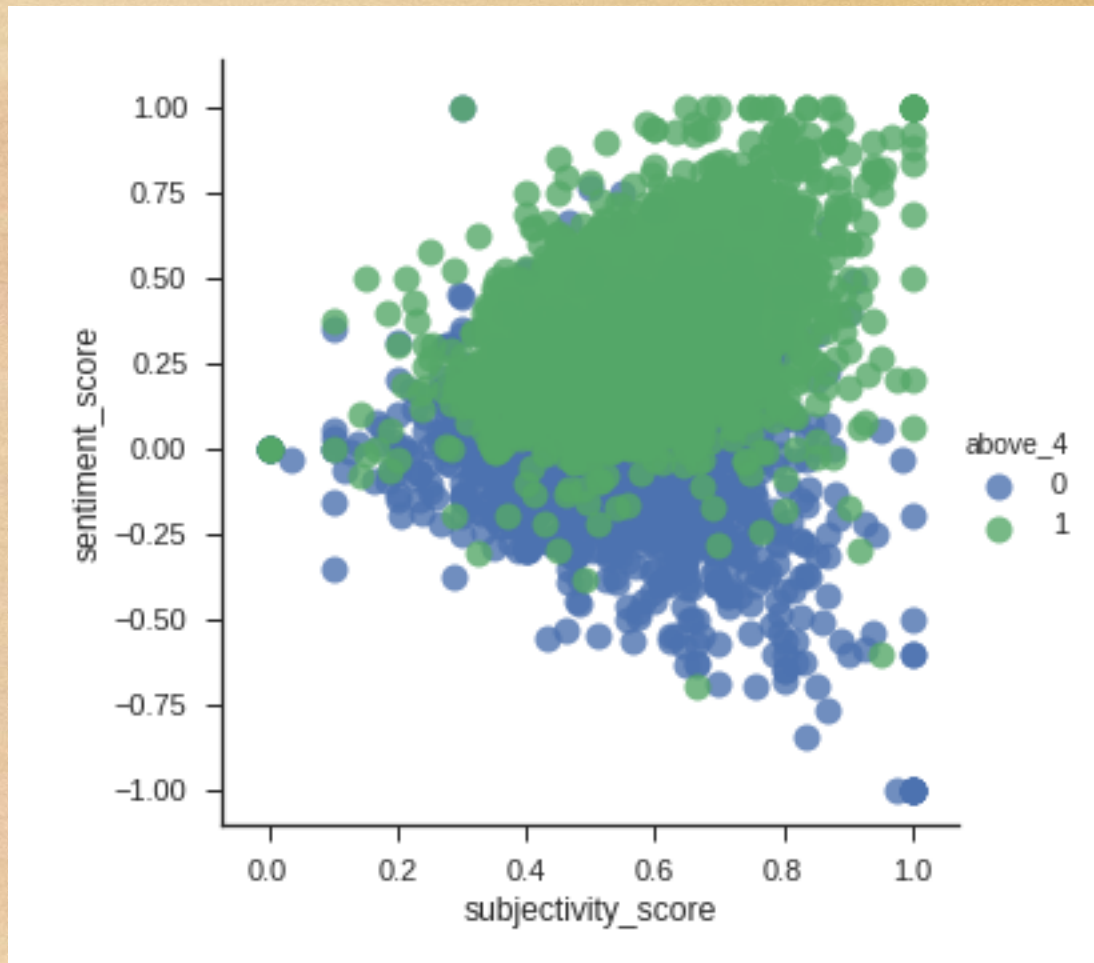
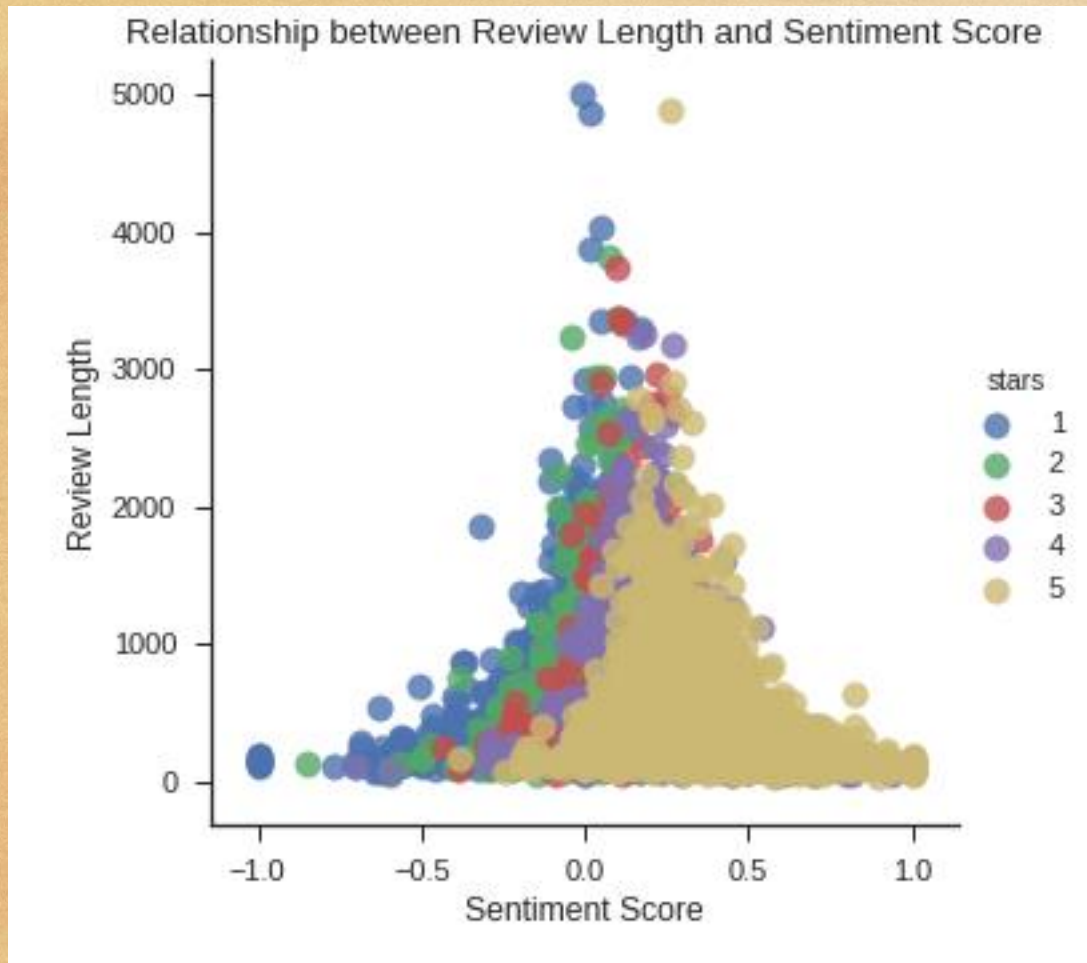


Distribution of Sentiment Score

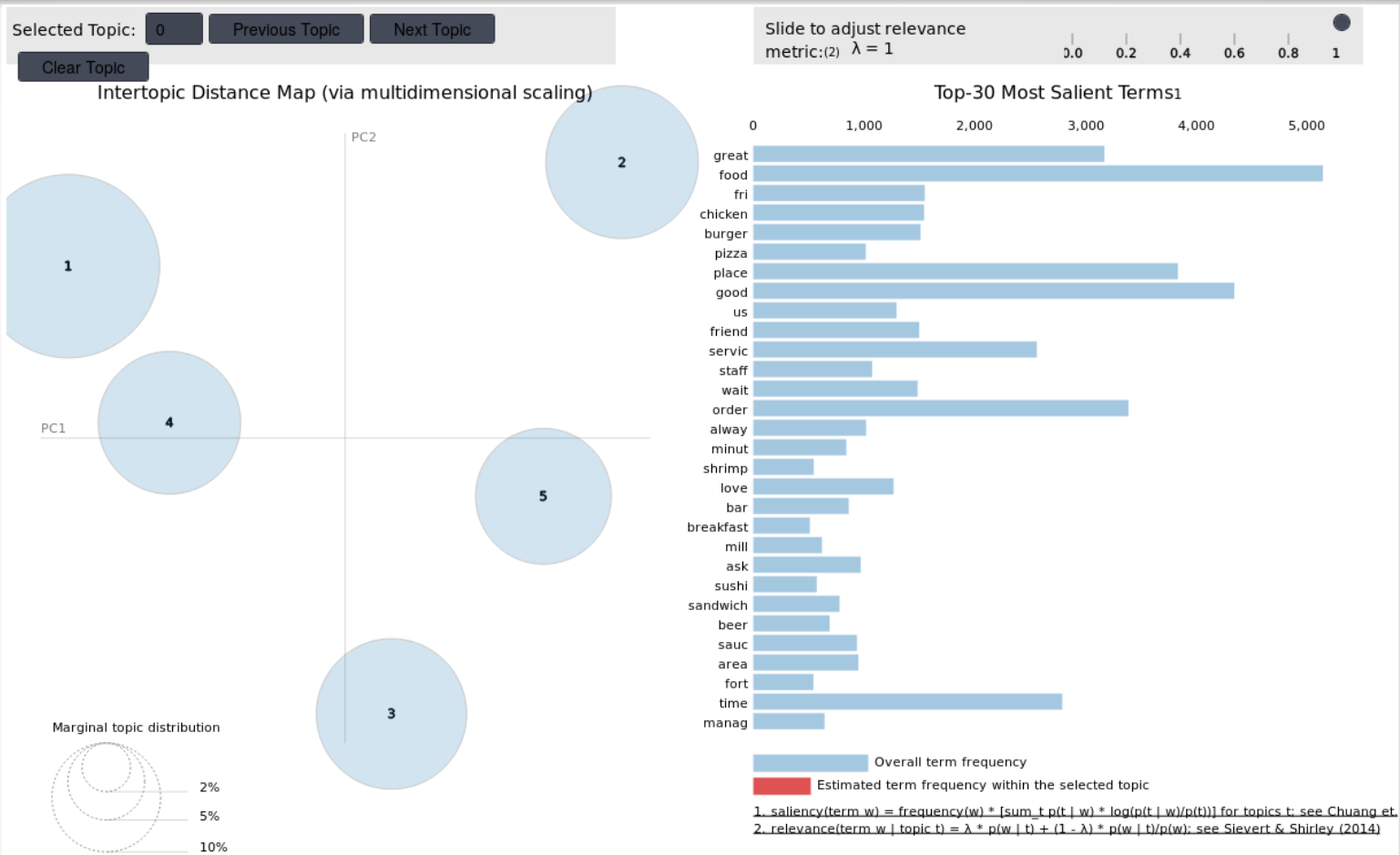


Sentiment by Star Rating

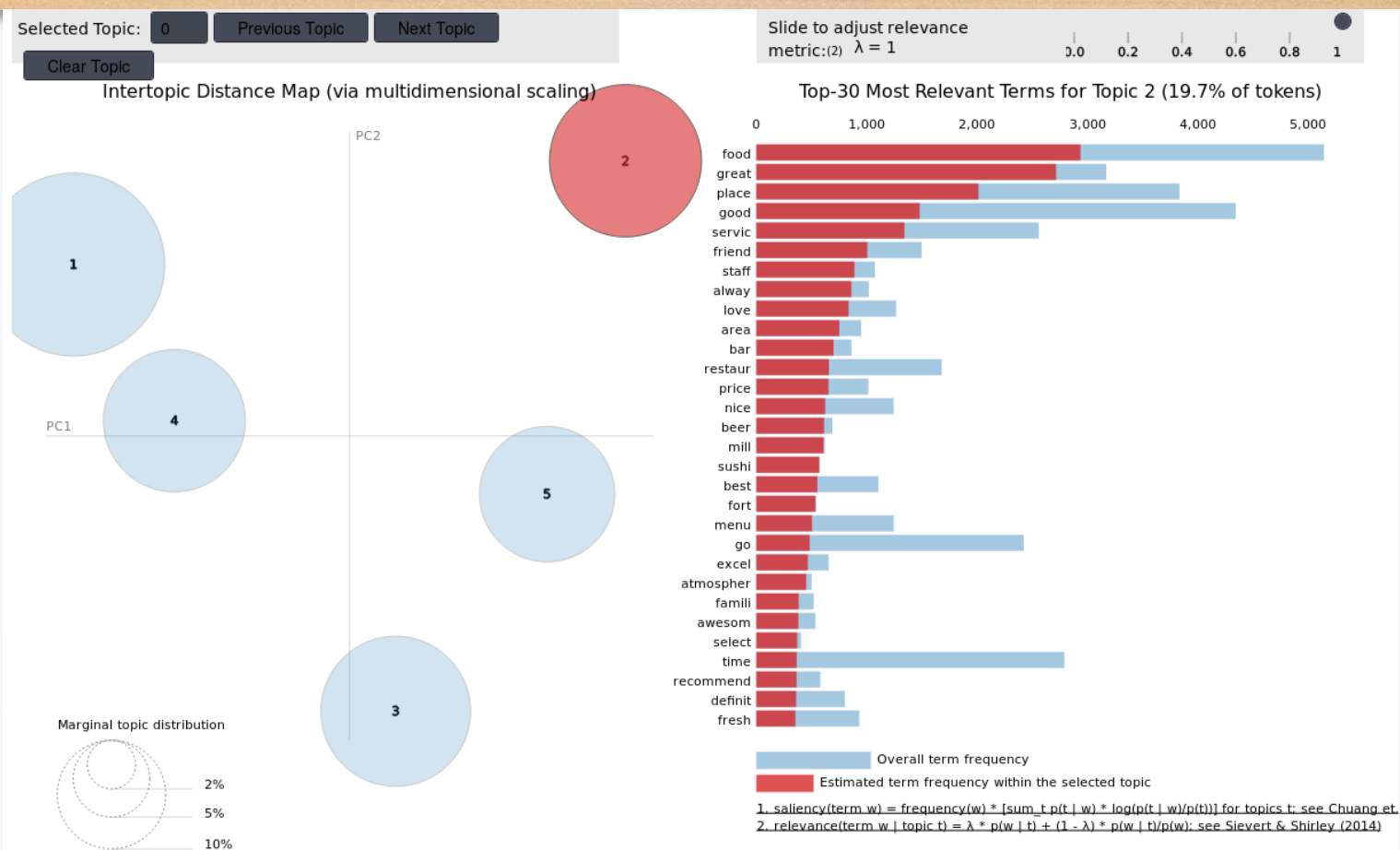




Visualizing Latent Dirichlet Allocation (LDA)



Visualizing Latent Dirichlet Allocation (LDA)



Recurring Themes in Reviews

Five Star Reviews

- “Great Ambiance”
- “Amazing Experience”
- “Nice Servers”
- **Common Words:** “Love, cheap, great, excellent”

One Star Reviews

- “Slow Service”
- “Worst customer service”
- “Absolute Garbage”
- **Common Words:** “Terrible, disappointed, horrible”



Like



Love



Haha



Yay



Wow



Sad



Angry

A Different Kind of Recommender

Using sentiment instead of ratings

So, which is better?

Recommendation based on **Ratings**

or

Recommendations based on **Sentiment**

Preliminary Results

Star-Based Recommender

Scale [1, 5]

- **RMSE:** 1.05
- **MAE:** 1.26



Sentiment-Based Recommender

Scale [-1, 1]

- **RMSE:** 0.784
- **MAE:** 0.748



$$\text{MAE} = \frac{1}{|\hat{R}|} \sum_{\hat{r}_{ui} \in \hat{R}} |r_{ui} - \hat{r}_{ui}|$$

$$\text{RMSE} = \sqrt{\frac{1}{|\hat{R}|} \sum_{\hat{r}_{ui} \in \hat{R}} (r_{ui} - \hat{r}_{ui})^2}$$

Popularity Models

Star Based Ranking

	name	stars
0	"MorningStar Missions Cafe"	5.0
1	"The Flipside Cafe"	5.0
2	"Small Bar Fort Mill"	5.0
3	"Jersey Mike's Subs"	5.0
4	"Pelican's SnoBalls"	5.0
5	"Sweet Dough"	5.0
6	"China Inn"	5.0
7	"Pasquale's Restaurant & Pub"	5.0
8	"Tacos Nayarit #2"	5.0
9	"Papa Murphy's"	5.0

Sentiment Based Ranking

name	sentiment_score
"Golden Bar & Restaurant Equipment"	0.669444
"6500 Tavern"	0.631877
"Takeria Puebla"	0.608333
"El Triunfo Restaurant"	0.599421
"T.I.G. BBQ"	0.589494
"Pop N' Sons"	0.577566
"Tacos Los Parados"	0.569158
"Sleepy Wilson's Barbecue"	0.547558
"Husongs Cantina"	0.542396
"Firelight Barn"	0.536445

Recap

Exploratory Data Analysis

Build a baseline Recommender System

Dive into the language of reviews

Try some creative approaches

Next Steps

- This framework is deployable and can be **A/B tested**. It can be scaled, it can be tuned.
- Evaluating the effectiveness of each Recommender system by pulling live data from **Yelp's API**
- Network Analysis on Reviewers and **Elite** Reviewers
- Technical:
 - **Different Recommenders Based on interaction thresholds** (Popularity → Hybrid Collaborative Filtering)
 - Continue to tune models with more feature engineering
 - More Grid-searching on Single-Value Decomposition (SVD) models for better scores (RMSE, MAE of projected ratings), or HIT - How many recommendations are necessary to use collaborative filtering for new users
 - Identify preferences vs dislikes

Thank You!

“While the individual person is an insoluble puzzle, in the aggregate they becomes a mathematical certainty. You can, for example, never foretell what any one person will do, but you can say with precision what an average number will be up to. Individuals vary, but the percentages remain constant. So says the statistician.” – Sherlock Holmes