Can Yelp Reviews Bridge the Gap Between Consumer Preferences and Industry Priorities?

Business Problem

Directory services apps and review forum apps have changed the way restaurants of different tiers attain brand loyalty and market themselves.

Yelp's historical data on restaurants' reviews, check-ins, and special attributes aid businesses in understanding their impact on customers and what make customers tick.
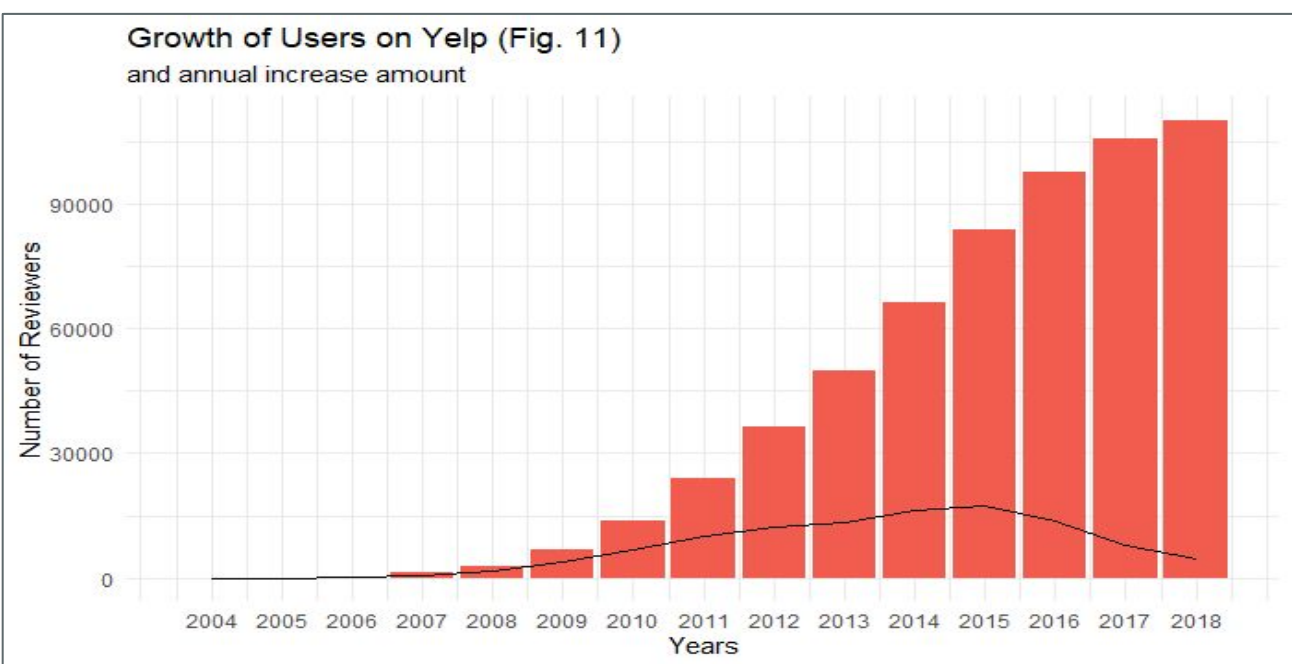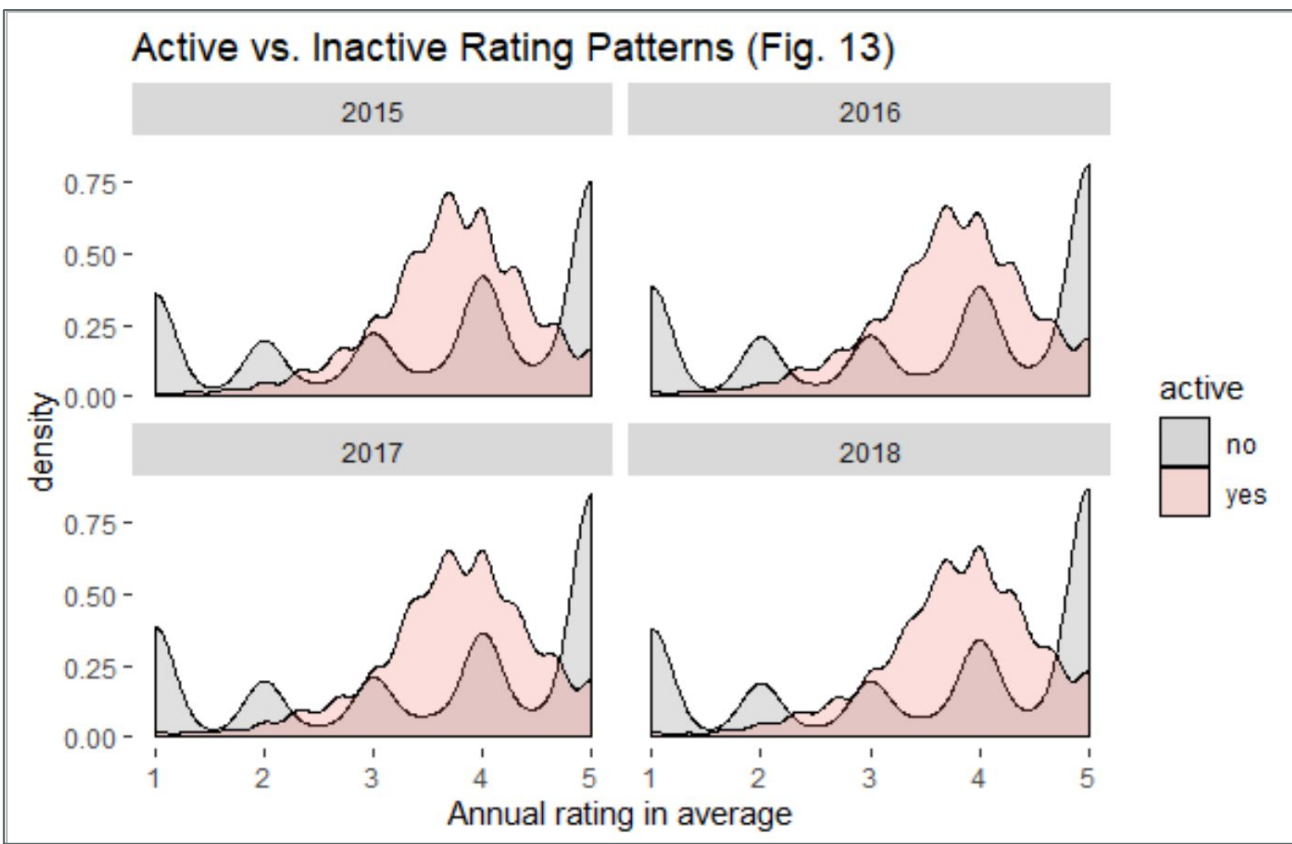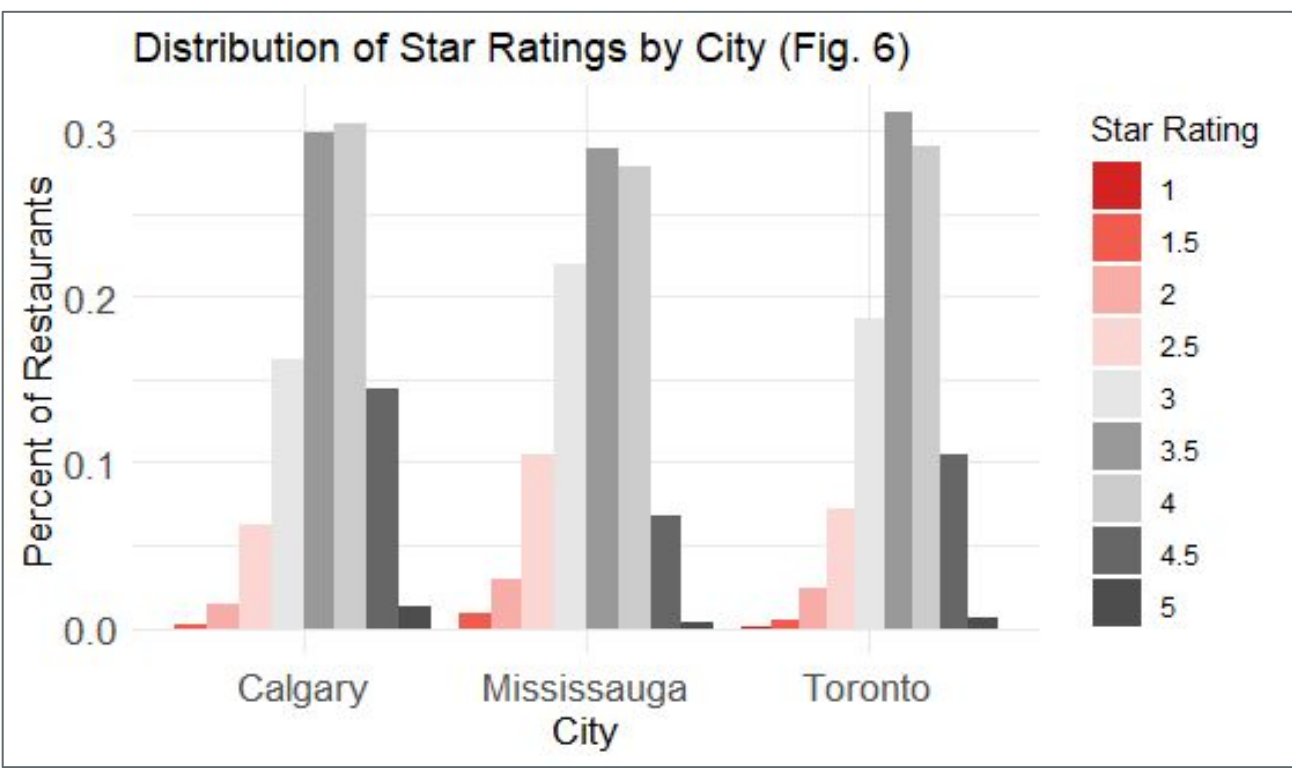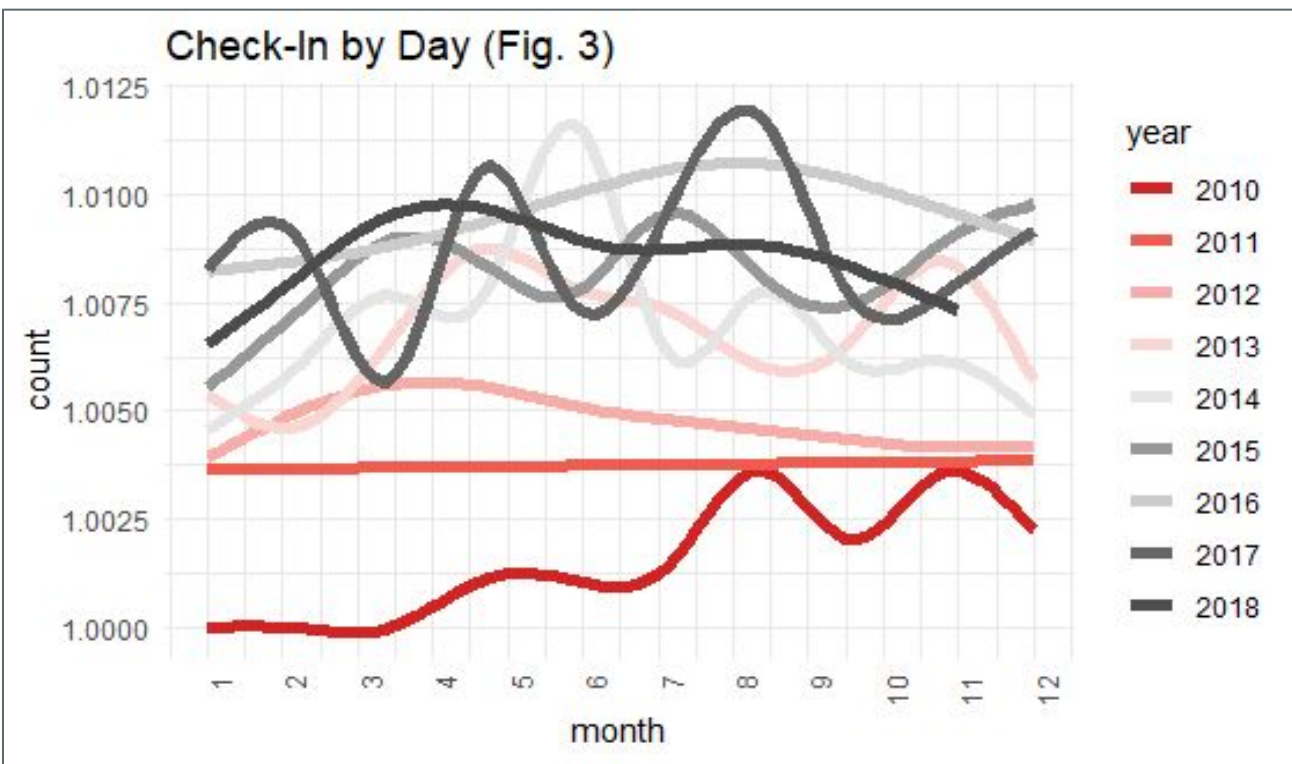
Areas of focus:

- How can a restaurant leverage information on Yelp to increase their chances of attaining their business goals and preventing permanent closure?
- What aspects make customers choose a restaurant over its competitors?

The Dataset

- Found on Kaggle: https://www.kaggle.com/yelp-dataset/yelp-dataset
- Reduced original dataset so keep only:
  - All businesses that were tagged with the phrases "restaurant" or "food" (or similar variations)
  - Eliminated business with category descriptions like "catering".
  - Restaurants with 14+ reviews, (the top 50% of review count).
  - Removed all restaurants not in Toronto, Calgary or Mississauga.
  - Data from 2007 to 2018
  - Final Datasets included ~7,500 restaurants and ~480,000 reviews.

Exploratory Data Analysis

- Contains both opened and closed restaurants; 5,699 (76%) open and 1,831 (24%) closed

- Average restaurant had 111 check-ins

- The average restaurant had 64 reviews.

- Check-ins steadily increased year over year until it peaked between 2015 and 2017. The first year to see a decrease in check-ins from the previous year was 2018 (Fig. 3)

- There appears to be seasonality changes with visits typically higher in the summer than winter (Fig. 3)

- The majority of restaurants in the dataset have a 3.5-star rating, closely followed by 4-star restaurants (Fig. 6)

- The number of new customers increased until 2015 (Fig. 11)

- The distribution of the review length is right-skewed with a mode of around 25 to 50 words.

- Inactive users wrote less, but were more extreme in their ratings while Active users wrote longer reviews and gave more 'mild' ratings to restaurants (Fig. 13)

- Yelpers who gave 1 and 2 star reviews wrote longer reviews that Yelpers who gave 3, 4, and 5 star reviews.

Check-In by Day (Fig. 3)

Distribution of Star Ratings by City (Fig. 6)

Active vs. Inactive Rating Patterns (Fig. 13)

Growth of Users on Yelp (Fig. 11)
and annual increase amount

Methods

Customer Review Topic Modeling

$$$$-restaurants 1-star and 5-star reviews:

- **Topic: Special Occasion** - Yelpers discussed visiting expensive restaurants to celebrate a special occasion, this group often indicated that it was their first time visiting the restaurant and were attracted by special features including tasting menus and the wine list.

- **Topic: Food Preparation** - Yelpers that wrote about this topic used phrases like "perfectly cooked", "great food", and even used specific phrases to describe the foods preparation like "medium rare".

$-restaurants 1-star and 5-star reviews :

- **Topic: Consistency** - These customers use phrases like "can't go wrong", "every time" and "great food". They are regulars at these restaurants and enjoy that they consistently have a good meal, no matter what they choose on the menu.

- **Topic: Friendly Staff** - While these customers do enjoy and mention the food, they use phrases like "super friendly", "staff friendly" and "friendly service" when giving five-star reviews indicating that a welcoming and accommodating staff sets the place apart.

Predicting if a Restaurant Will Go Out of Business

Lasso:

- Running a penalized logistic regression reduced total variables to 15 variables with some impact on the outcome.
- 6 attributes hold most of the weight in predicting whether a restaurant will stay open or not. The presence of a parking lot has the highest positive correlation on keeping a restaurant in business, while the option of only street parking has the strongest negative correlation.
- Interestingly, bike parking also appears to have a negative correlation. Higher star ratings and being kid-friendly have the second highest positive correlation with staying open.

Forward Stepwise:

- The final forward selection model contained 23 variables.
- As expected, many of the variables describing the characteristics of the restaurants were included in the model. Street Parking, Parking Validation, Romantic Ambiance, Full Bar, Bike Parking, Parking Lot, Star Rating, and Trendy Ambiance have the highest correlations to whether the restaurant is open or not.
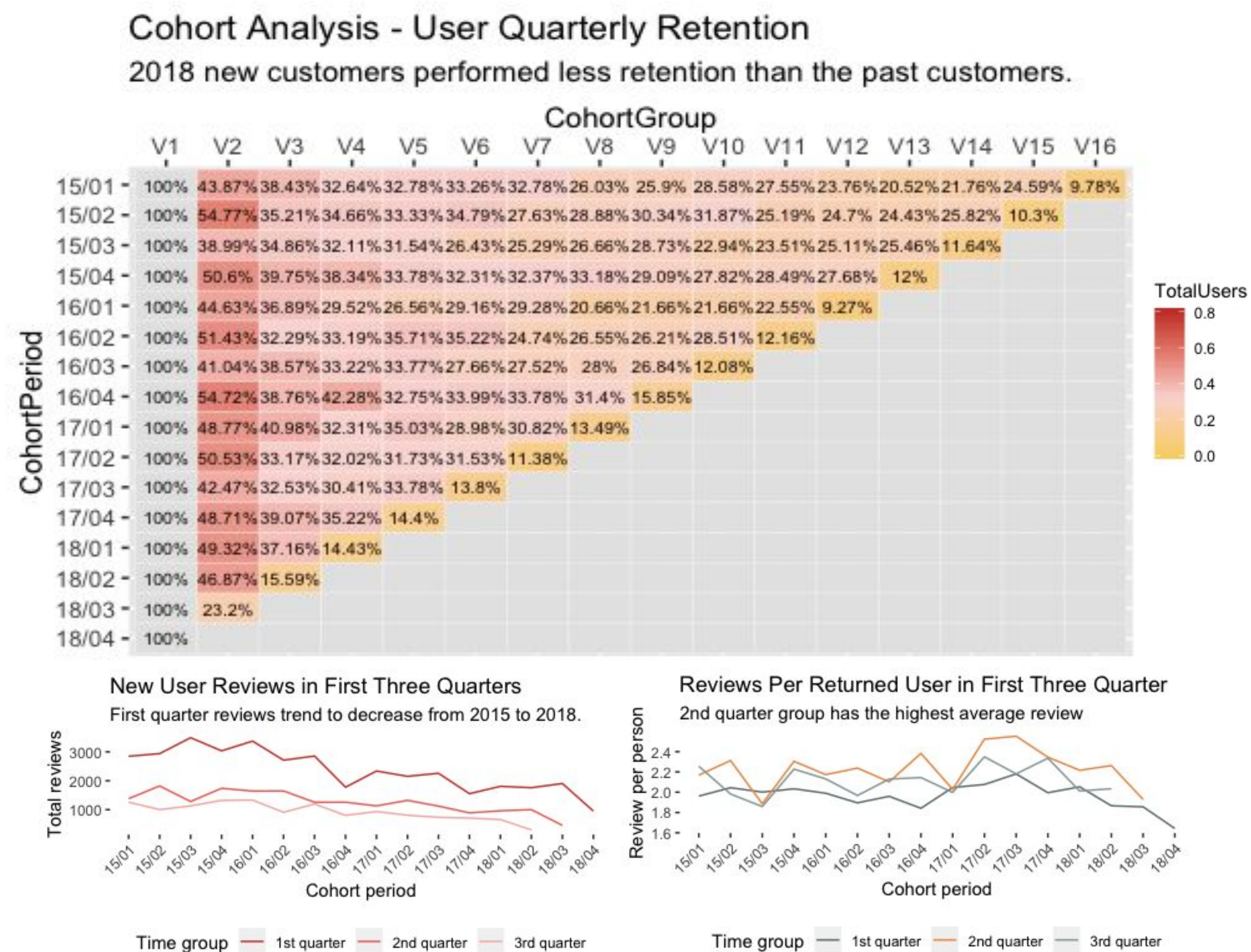
XGBoost

- Three clusters that showed feature importance:
  - **Cluster 1:** Summer was the peak season. Review count and star rating were the most important feature. Because they can reflect the service quality of a restaurant, then the restaurant can make corresponding improvements based on the reviews to reduce the risk of closing.
  - **Cluster 2:** Full bar was very important features. Because alcohol sales have higher gross margins than food, you can improve your bottom line with a full bar.

Feature importance
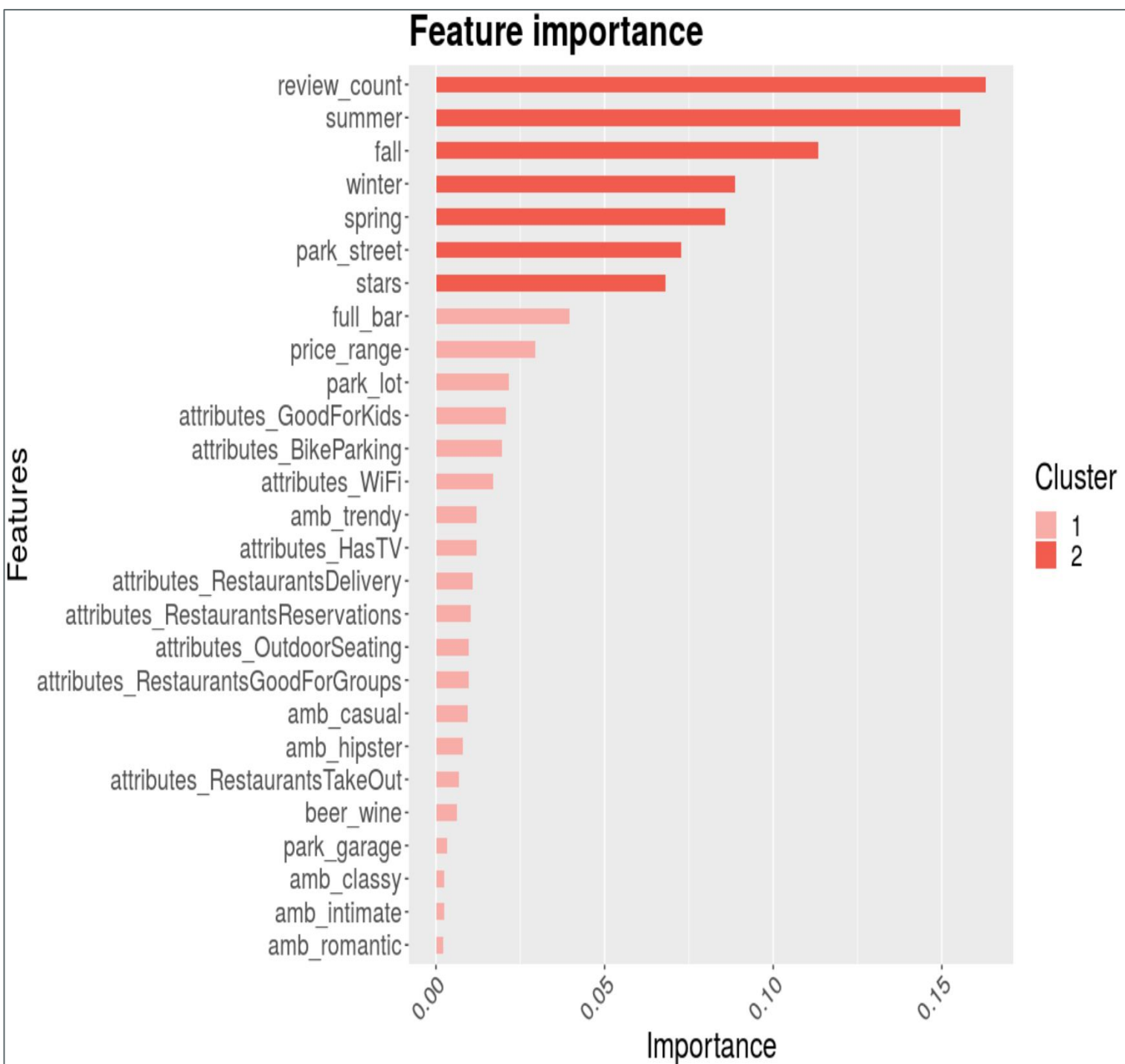
Methods cont.

Cohort Analysis

Inspect user quarterly retention from 2015 to 2018, trending user characteristics in review by timeline.

Cohort Analysis - User Quarterly Retention
2018 new customers performed less retention than the past customers.

New User Reviews in First Three Quarters
First quarter reviews trend to decrease from 2015 to 2018.

Reviews Per Returned User in First Three Quarter
2nd quarter group has the highest average review

User Retention

During 2015 to 2018:
- > 40% users stay on Yelp within six month.
- ~ 40% users stay on Yelp in the first year.
- ~ 30% users stay on Yelp in the second year.
- < 20% users stay on Yelp after second year.

Review posting

- For the number of total review, users posted in the first quarter decreased after 2016.
- The second quarter has the highest review per person,

Recommendation System

Predict customers' favor and engage them contribute to reviews.

- **Model**
  User-Based Collaborative Filtering model

- **Data**
  490 reviewers who rated at least 10 of 670 restaurants in Mississauga,

- **Result**
  Recommend 10 - restaurant list for 117 users.

- **Test Accuracy: 96.70%**
  Setting good rating as 4, 7.9% prediction result are relevant to actual test value..

Conclusions and Recommendations

User activity:

- Reviewers became less active from 2015 to 2018
- The second quarter is the core period to encourage users to post reviews

Customer experience:

- $$$$ - restaurants should focus on creating an experience that is out of the ordinary because many customers that visit are coming to celebrate a special occasion
- $ - restaurants should focus on training and hiring employees that are friendly. As their interactions with customers are shorter, they have less time to make a good impression. They should also be aware that how employees are working behind the counter, when not interacting with customers, can have an impact on the customer's perception of service.

Restaurant attributes:

- Allocating resources to online marketing and online presence (directory service apps) as high review count and star ratings have the strongest significance in predicting that a restaurant will stay in business
- Consider a location with access to a parking lot or look for ways to improve customers' ease of finding parking
- Being accommodating towards kids and encouraging PR teams to drive traffic during the summer

Xinman Liu, Melissa Putur, Jiao Sun, Adil Wahab, Rui Xu