



Yelp Coffee Shop Analysis

Andrew Field

Contents

Questions to Answer, Hypotheses, Approach

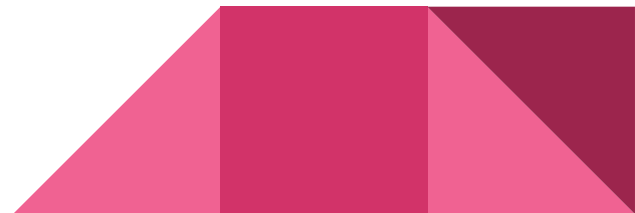
Discuss Technical Challenges

Entity Relationship Diagram

Initial Findings

Deeper Analysis

Hypotheses Results



Questions to Answer

1. How do business hours relate to review scores?

Gain insight into best practice regarding store hours

2. How do amenities impact customer experience?

Does Wi-Fi, TV, outdoor seating translate into happier customers?

3. What are the most important factors in determining customer satisfaction?

Understand customer perspective

Improve customer satisfaction



Questions to Answer (cont.)

4. How does customer volume fluctuate throughout the year?

Optimize timing of promotions to animate customers

Set expectations for business



Hypotheses

1. More off-days will correlate with lower ratings. More 24-hour days will correlate with higher ratings.

The more available a business is, the happier the customers will be

2. Businesses with Wi-Fi, TV, outdoor seating will have better ratings

Amenities will help to improve customer experience

3. Service and price will be key factors in customer satisfaction and dissatisfaction

These factors are commonly highly prioritized among customers



Hypotheses (cont.)

4. Coffee businesses will have more customers in the summer and less in the winter

Customers will go out more when the weather is warm



Approach

1. Focus on coffee shop reviews

Reviews of other business will be less relevant

2. Analyze correlation between features

What features are most correlated with high ratings, low ratings?

3. Look into frequently used words

What is important to customers

Connections between certain words and ratings



Technical Challenges

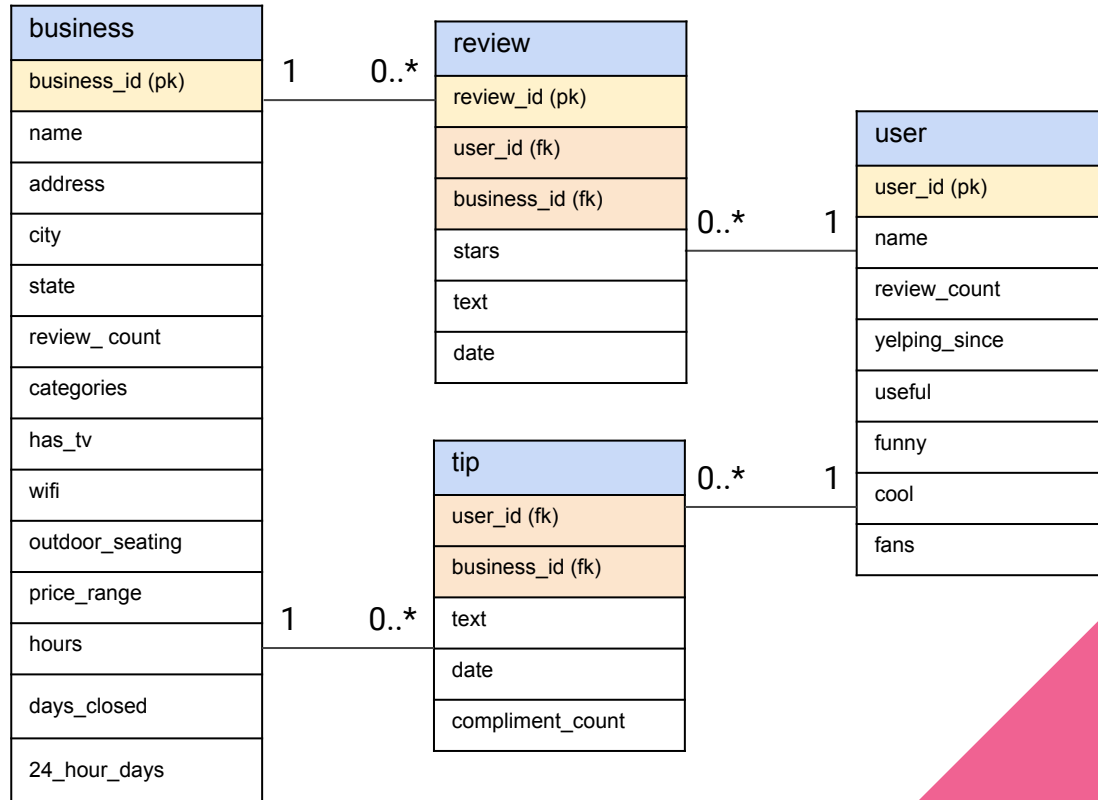
Yelp review data was voluminous. File transfer and data processing was a challenge with limited resources.

Format of some data in Yelp tables was complex. Column values included dictionaries of free formatted values; needed to commit time to data wrangling before beginning analysis.

Data was inconsistent and sometimes incomplete, requiring workarounds and defensive code.



Entity Relationship Diagram (ERD)



Initial Findings

Initial look into the data shows evidence my hypotheses about business hours were not correct.

Businesses with no off-days had worse ratings. Businesses with 1-6 off-days all have similar ratings.

Generally, the more 24-hour days a business has, the worse their ratings are.

Data needs to be investigated further to determine correlation between other statistics.

days_closed	stars
0	3.329774
1	4.209273
2	4.173697
3	4.183673
4	4.128571
5	4.309524
6	4.233333

24_hour_days	stars
0	3.584182
1	3.975403
2	1.725806
3	1.875000
4	2.250000
5	2.500000
6	2.500000
7	2.594194

Deeper Analysis - Pearson Correlation

Overview of Pearson correlations between features.

Strongest correlations:

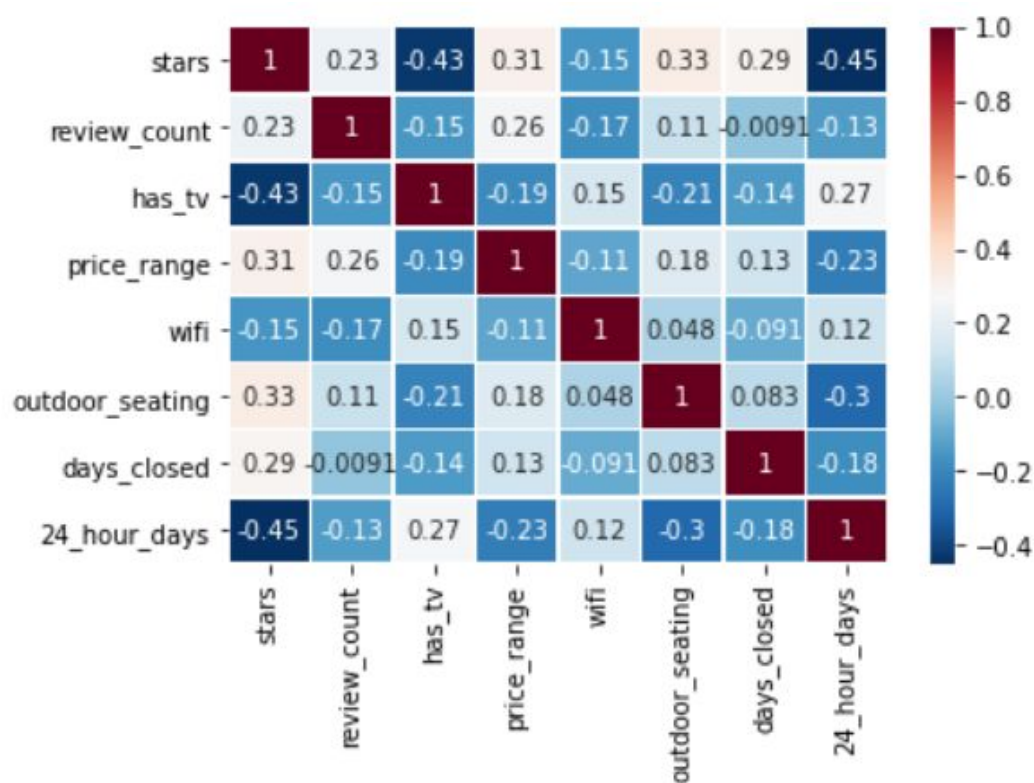
24 hour days, stars

Has TV, stars

Outdoor seating, stars

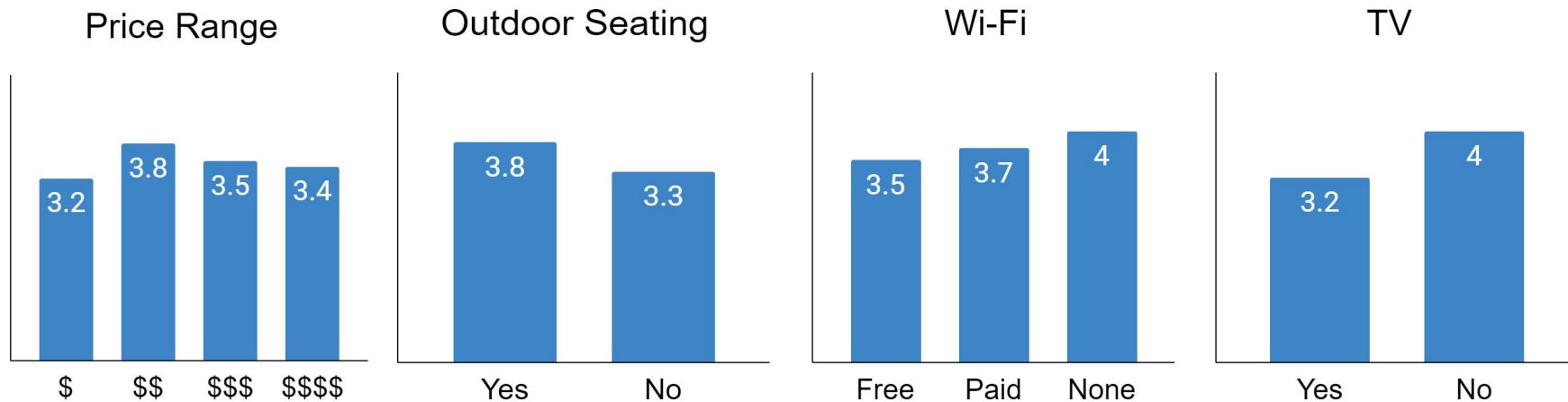
Price range, stars

24 hour days, outdoor seating

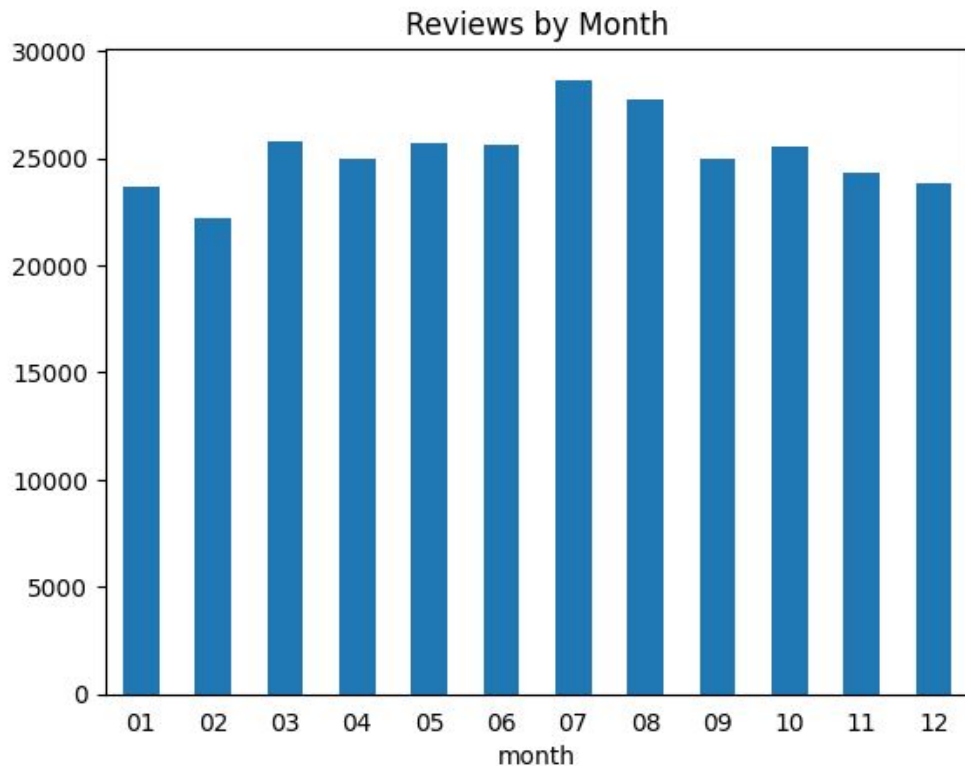


Deeper Analysis - Features and Amenities

Graphing relationships between some key features and business' ratings



Deeper Analysis - Time



Data taken from reviews
between 1-1-2007 and
12-31-2018

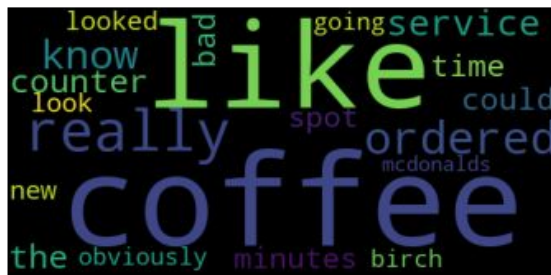
Deeper Analysis - Text

Looking into most commonly used review words

1 star reviews



2 star reviews



3 star reviews



4 star reviews



5 star reviews



Deeper Analysis - Text (cont.)

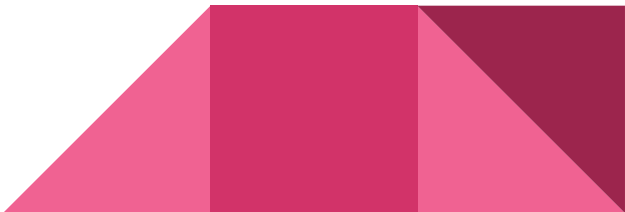
Analyzing most commonly used review words

Low Ratings: told, asked, said, minutes, order

Dissatisfied customers' frustration is mostly associated with quality and speed of service. Coffee is in the most common 1-2 words for all reviews except those with 1 star. Extreme customer dissatisfaction does not seem to be related to the product itself.

High Ratings: food (breakfast, donuts)

Quality food is a key way to improve customer satisfaction. Coffee businesses with great ratings provide more than just coffee.



Results of Hypotheses

1. More off-days will correlate with lower ratings. More 24-hour days will correlate with higher ratings.


This was incorrect. Businesses with no off-days had worse ratings and, generally, 24-hour days correlated with lower ratings.

2. Businesses with Wi-Fi, TV, outdoor seating will have better ratings

Partially correct. Outdoor seating correlated with better ratings. However, Wi-Fi and TV correlated with worse ratings.

3. Service and price will be key factors in customer satisfaction and dissatisfaction

Based on findings from the text analysis, service is an important factor. Price was more nuanced. It was a common word only in 1 star reviews. In analysis of price ranges, mid-range priced coffee businesses were rated the best on average.



Results of Hypotheses (cont.)

4. Coffee businesses will have more customers in the summer and less in the winter

Correct. Coffee businesses' most busy months were July and August. Least busy months were December, January and February.

