The Analytics of Customer Reviews

*Text analytics **extract meaning** from human language and have the power to offer businesses **insights** into large amounts of data, helping to **drive business decisions**.*

Analysis conducted on **restaurant reviews** provided by **Yelp**, each review being accompanied by a **rating** of 1-5 stars.

> *"Of the several places to get bagels in Oakland, these bagels **are the best**. Really **fresh and tasty**. **However**, I **wasn't crazy** about their bagel and egg sandwiches."*
> ★★★☆☆

Project Goals

1. Assess **customer perception** of restaurants through exploratory data analysis.
2. Use machine learning to **predict** the positive or negative **sentiment** of reviews.
3. Apply the model to a single restaurant – **reveal the key aspects of the service** that drive customer perception.

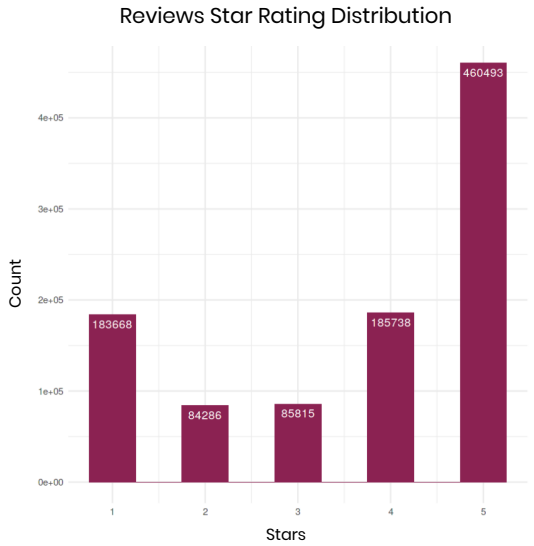
Dataset Statistics

5,261,668 reviews

188,593 businesses

64 reviews on average per restaurant

71% positive reviews (4+ stars)



Customers have the tendency to review a restaurant if they are **highly pleased** or **highly displeased**.

Sentiment Analysis

Sentiment Analysis identifies and categorizes the opinions of the author of the text into positive or negative.

1 Pre-processing

Standardized text, removed unnecessary words, extracted word frequencies.
Obtained **Sentiment** predictor from **Rating***.

2 Machine Learning Model

Classification Tree trained to **categorize** reviews.
It labeled reviews as positive or negative based on the **containing words**.

3 Making Predictions

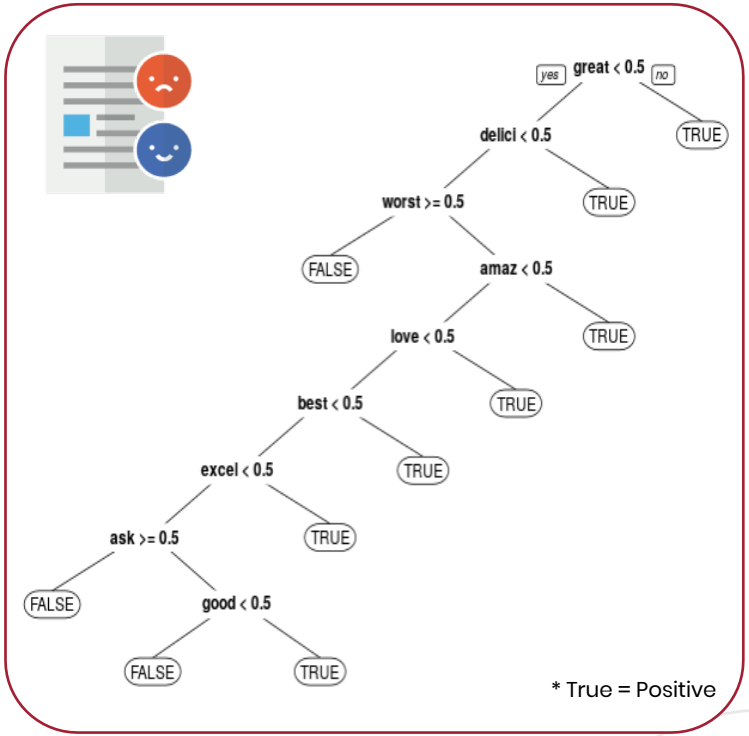
Model **learned sentiment vocabulary** and positive/negative **restaurant aspects**.
Predicted the sentiment of unseen reviews with an **86% accuracy****.

The model was selected due to the **high interpretability** given by the visual output.

This interpretability facilitates **explaining the model** to decision-makers, making it more likely to **drive change**.

*4- and 5-star = Positive; 1- and 2-star = Negative; 3-star reviews excluded
** test-set accuracy over a 71% baseline

Predicting Positive Reviews – Model Output



Model interpretation example: if *great* is present, the review is positive, and if *worst* is present, it is negative, which matches intuition.

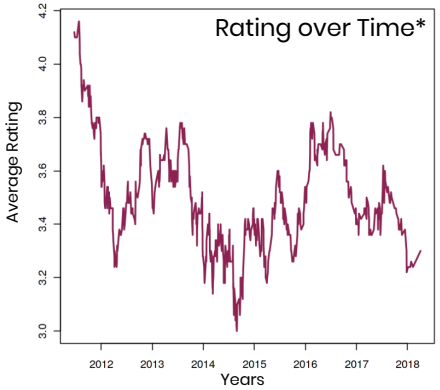
Application on a Single Restaurant – *Wicked Spoon* (Las Vegas Buffet) ★★★☆☆

In the **top 1%** most reviewed businesses (6,500 reviews) despite an **average rating** (3.5 stars).
The goal is to analyze the **discrepancy** between popularity and rating.

The *Wicked Spoon* average rating **dropped significantly in 2014** and has since improved, being currently in a slight **upward trend**.

Application of the trained **classification tree model**:

- Revealing the service aspects of the *Wicked Spoon* seen as **positive and negative** by customers
- Exploring **trends over time**



*50-day moving average

Frequent Review Terms and Sentiment

*Green = Positive, Red = Negative

Removing sentiment indicators (e.g. *great*) led to the identification of the customers' opinion of the **specific business aspects** that drive overall perception.

- The word ***wait*** is mentioned over 1,000 times and was classified as **negative**, almost **400** of mentions being in **2014–15**, matching the period when the sentiment dropped.
- Numerous reviews criticize the **long wait times** in the restaurant.
- This exemplifies an **actionable insight** that the restaurant can choose to **act upon**.

> *"If you plan to eat here expect at least a **two hour wait** regardless of the time or day. You are going to **stand in line to check in** for about an hour and then **wait in a seating line** for about another hour."* ★★☆☆☆