



# Restaurant Ratings

Group 10

Aashna Ahuja  
Marwa Eltgani  
Sheng Huang



# Outline

---

- Motivation and Question
- Data Cleaning
- Exploratory Data Analysis
- Sentimental Analysis
- Future Work

# Motivation



- We focus on only restaurants in California because it has the largest number of restaurants in the US.
- We want to help new business owners in California predict their initial rating with features such as: Location(City), Restaurant Type, and Total Hours of Operation per week.



## Question

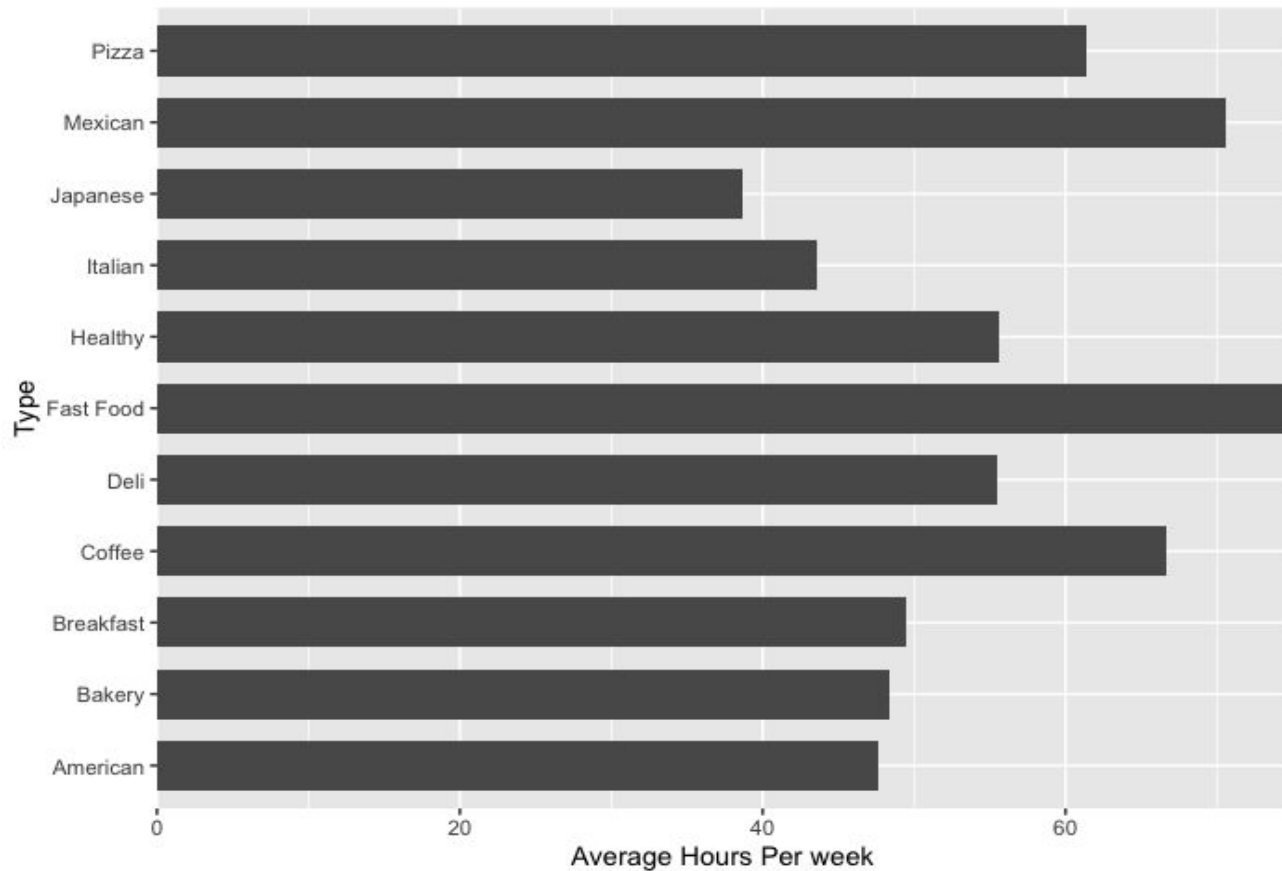
1. What features are associated with high Yelp rating for new business owners?
  - Location
  - Type
  - Hours
  - Sentiment
2. What can these new business owners do to improve their ratings on Yelp?

# Data Cleaning



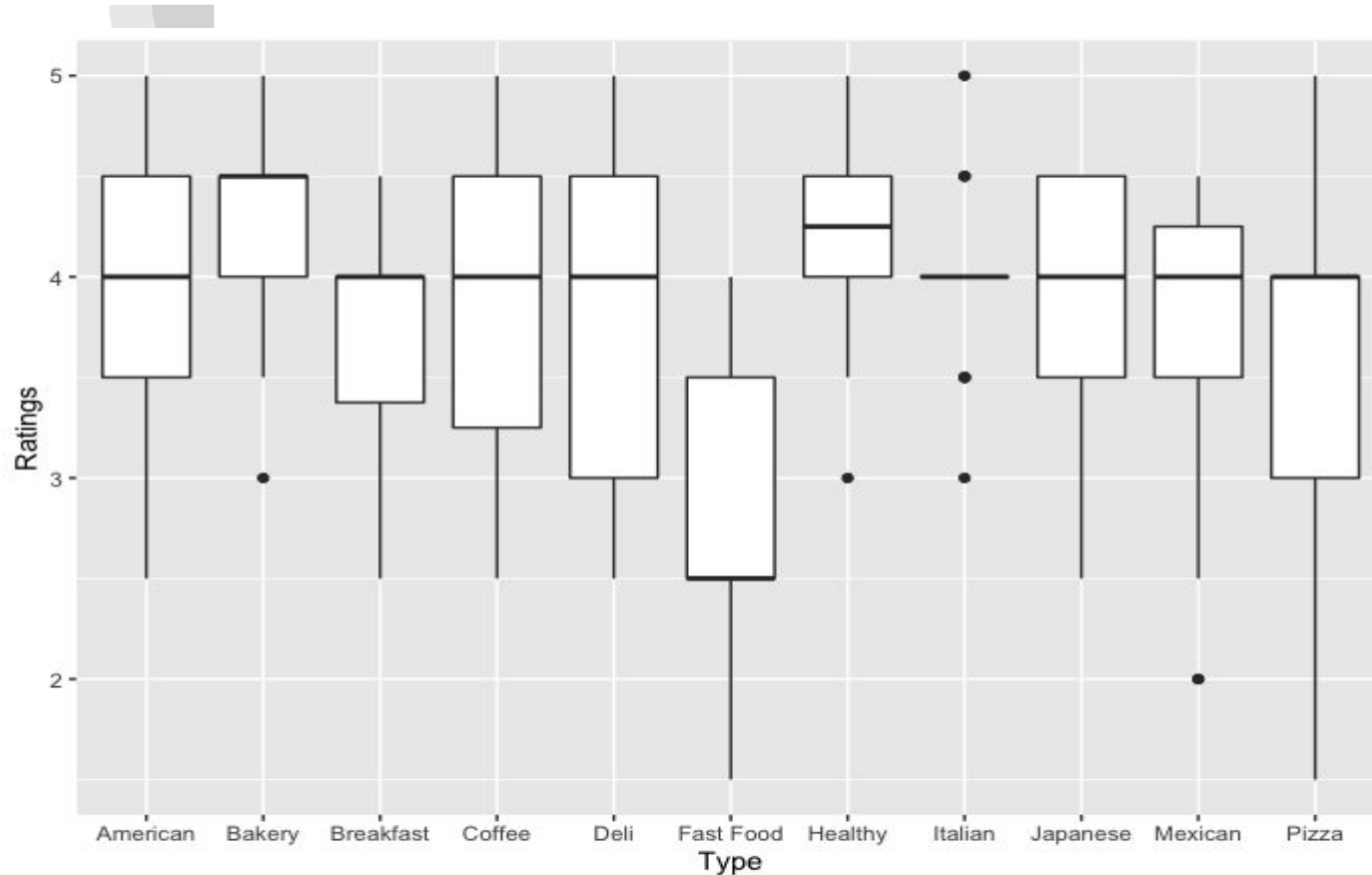
- Filtered the business.json and review.json to restaurants in California
- Dropped closed business
- Classified each restaurant into a type based on their menu or what they consider themselves to be.
- Focused on the most common restaurant types

# Exploratory Data Analysis



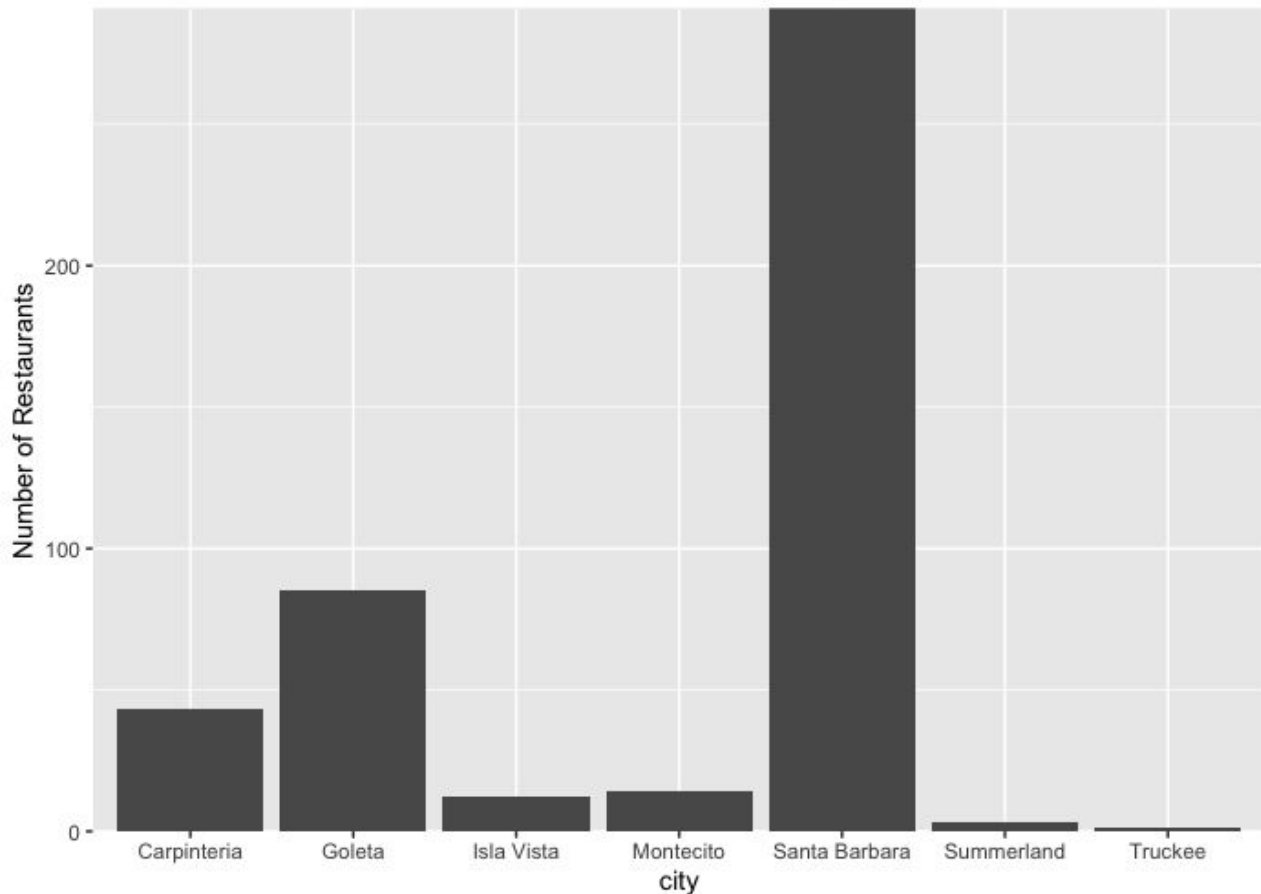
- Fast Food has the highest hours per week.
- Except for Pizza, Coffee & Mexican all other types have less than 60 hours average .

# Exploratory Data Analysis



- Italian food type has a weird pattern
- Some outliers in Bakery and Mexican

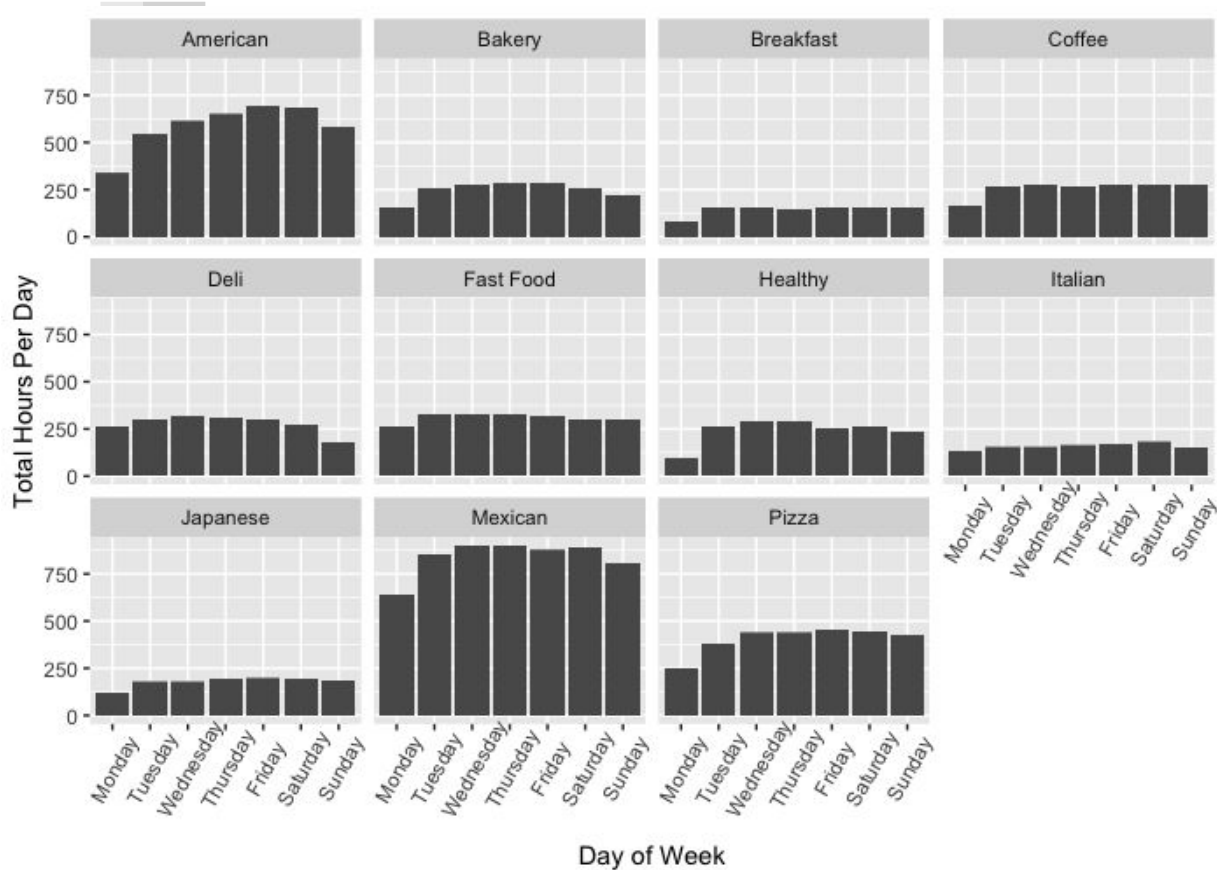
# Exploratory Data Analysis



- Santa Barbara has the most number of restaurants.
- Truckee has the lowest number of restaurants.



# Exploratory Data Analysis

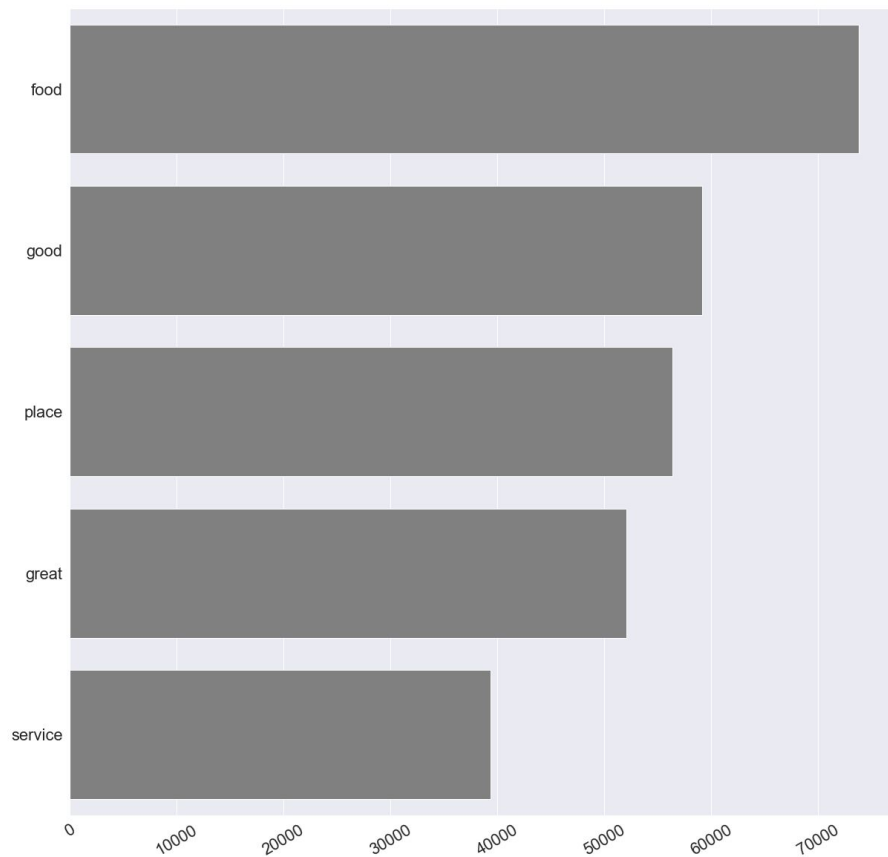


- Monday's total hours is the lowest across all the types of restaurants.

# Sentimental Analysis



# Sentimental Analysis





# Sentiment Score

We used VADER a sentiment analysis tool that is specifically attuned to sentiments expressed in social media.

Business ID	Sentiment Score
Business ID 1	0.765
Business ID 2	-0.235

The sentiment score is in the range of -1 to 1

- ❖ -1 to -0.5 is Negative
- ❖ -0.5 to 0.5 is Neutral
- ❖ 0.5 to 1 is Positive



# Future Work

- Look into standardizing the reviews
- Modeling:
  1. Linear Regression
  2. Random Forest