

## Yelp Reviews Sentiment Analysis - Data Appendix

### *reviews.json*

---

Unit of observation: Each row represents a review associated with each restaurant. Below is a list of all variables in this JSON file:

- **review\_id**: A 22-character unique identifier for the review.
- **user\_id**: A 22-character unique identifier for the user who wrote the review, which maps to *user.json*.
- **business\_id**: A 22-character unique identifier for the business being reviewed, which maps to *business.json*.
- **stars**: The star rating given by the user for the business.
- **date**: The date the review was posted, formatted as YYYY-MM-DD.
- **text**: The full content of the user's review.
- **useful**: The number of times the review was marked as useful by other users.
- **funny**: The number of times the review was marked as funny by other users.
- **cool**: The number of times the review was marked as cool by other users

### *business.json*

---

Unit of observation: Each row represents a restaurant, with several attributes—such as number of stars, the number of reviews associated with each restaurant, and whether or not the restaurant is open or not (for this particular analysis). Below is a list of all variables in this JSON file:

- **business\_id**: A 22-character unique identifier for the business.
- **name**: The name of the business.
- **address**: The full street address of the business.
- **city**: The city where the business is located.
- **state**: A two-character state code representing the business's location.
- **postal code**: The postal or ZIP code of the business.
- **latitude**: The geographic latitude coordinate of the business.
- **longitude**: The geographic longitude coordinate of the business.
- **stars**: The average star rating of the business, rounded to the nearest half-star.
- **review\_count**: The total number of reviews the business has received.
- **is\_open**: A binary value where 1 indicates the business is open and 0 indicates it is closed.

- **attributes:** A collection of key-value pairs describing various features of the business, such as takeout availability and parking options.
- **categories:** A list of strings representing the business categories, such as "Mexican" or "Burgers."
- **hours:** An object mapping each day of the week to the business's operating hours in a 24-hour format.

### ***FINAL\_yelp\_dataframe.csv***

---

Unit of observation: Each row contains the elements associated with each restaurant, in addition to the sentiment score based on the reviews in *reviews.json*.

- **business\_id:** A unique identifier for the business.
- **name:** The name of the business.
- **address:** The full street address where the business is located.
- **city:** The city where the business operates.
- **state:** A two-character state code representing the business's location.
- **postal\_code:** The postal or ZIP code of the business.
- **latitude:** The latitude coordinate of the business location.
- **longitude:** The longitude coordinate of the business location.
- **stars:** The average star rating given by users.
- **review\_count:** The total number of reviews received by the business.
- **is\_open:** A binary indicator where 1 means the business is open, and 0 means it is closed.
- **attributes:** A dictionary containing key-value pairs that describe business features such as parking, payment options, and WiFi availability.
- **categories:** A comma-separated list of categories that describe the type of business.
- **hours:** A dictionary mapping each day of the week to the business's operating hours in a 24-hour format.
- **compound:** The overall sentiment score calculated using sentiment analysis.
- **pos:** The proportion of positive words in the review text.
- **neg:** The proportion of negative words in the review text.
- **neu:** The proportion of neutral words in the review text.
- **useful:** The number of times users marked the review as useful.