

Review Dataset: yelp_academic_dataset_review.json

Unit of Observation: Each row in the reviews dataset represents a single user review of a restaurant.

Variables:

- Review_id
 - Type: String (22 characters)
 - Description: A unique identifier assigned to each review.
 - Categorical/Quantitative: Categorical
- User_id
 - Type: String (22 characters)
 - Description: A unique identifier for the user who created the review.
 - Categorical/Quantitative: Categorical
- Business_id
 - Type: String (22 characters)
 - Description: A unique identifier for the business being reviewed.
 - Foreign key linking to Business_id in the business dataset
 - Categorical/Quantitative: Categorical
- Stars
 - Type: Integer
 - Description: The star rating given by the user (e.g., 1-5 stars).
 - Categorical/Quantitative: Quantitative
- Date
 - Type: String (YYYY-MM-DD)
 - Description: The date when the review was posted.
 - Categorical/Quantitative: Categorical
- Text
 - Type: String
 - Description: The full text of the user's review.
 - Categorical/Quantitative: Categorical
- Useful
 - Type: Integer
 - Description: Number of useful votes the review received.
 - Categorical/Quantitative: Quantitative
- Funny
 - Type: Integer
 - Description: Number of funny votes the review received.
 - Categorical/Quantitative: Quantitative
- Cool
 - Type: Integer

- [illegible]

Business Dataset: yelp_academic_dataset_business.json

Unit of Observation: Each row in the business dataset represents a single restaurant.

Variables:

- Business_id
 - Type: String (22 characters)
 - Description: A unique identifier assigned to each business.
 - Primary key linking to Business_id in the reviews dataset
 - Categorical/Quantitative: Categorical
- Name
 - Type: String
 - Description: The name of the business.
 - Categorical/Quantitative: Categorical
- Address
 - Type: String
 - Description: The full address of the business.
 - Categorical/Quantitative: Categorical
- City
 - Type: String
 - Description: The city where the business is located.
 - Categorical/Quantitative: Categorical
- State
 - Type: String
 - Description: The state where the business is located.
 - Categorical/Quantitative: Categorical
- Postal_code
 - Type: String
 - Description: The postal code of the business.
 - Categorical/Quantitative: Categorical
- Latitude
 - Type: Float
 - Description: The latitude of the business.
 - Categorical/Quantitative: Quantitative
- Longitude
 - Type: Float
 - Description: The longitude of the business.
 - Categorical/Quantitative: Quantitative
- Stars
 - Type: Float

- Description: The average star rating for the business (rounded to the nearest half-star).
- Categorical/Quantitative: Quantitative
- Review_count
 - Type: Integer
 - Description: The total number of reviews for the business.
 - Categorical/Quantitative: Quantitative
- Is_open
 - Type: Integer (0 or 1)
 - Description: Indicates whether the business is currently open (1) or closed (0).
 - Categorical/Quantitative: Categorical
- Attributes
 - Type: Object
 - Description: A list of various attributes related to the business.
 - Categorical/Quantitative: Categorical
- Categories
 - Type: Array of Strings
 - Description: A list of relevant categories the business belongs to.
 - Categorical/Quantitative: Categorical
- Hours
 - Type: Object (Key: Day, Value: Hours)
 - Description: The operating hours of the business for each day of the week.
 - Categorical/Quantitative: Categorical

Descriptive Statistics

(1) Total Number of Reviews by Star Rating

