Yelp Rating Regression Predictor - Feature Descriptions

This project uses A LOT of data, and it can sometimes be confusing when it comes to understanding the different features and what they really are. In this document we will clarify the meaning of each feature in the six different JSON files and hopefully alleviate any confusion.

- yelp_business.json: contains establishment data regarding location and attributes
    - **address:** string, the full address of the business
    - **alcohol?:** integer, 0 or 1 for does not serve alcohol or serves alcohol in some capacity, respectively
    - **attributes:** object, business attributes to values. note: some attribute values might be objects
    - **business_id:** string, 22 character unique string business id
    - **categories:** an array of strings of business categories
    - **city:** string, the city
    - **good_for_kids:** integer, 0 or 1 for not child friendly establishment or child friendly establishment, respectively
    - **has_bike_parking:** integer, 0 or 1 for does not have bike parking or does have bike parking, respectively
    - **has_wifi:** integer, 0 or 1 for does not have wifi or does have wifi, respectively
    - **hours:** an object of key day to value hours, hours are using a 24hr clock
    - **is_open:** integer, 0 or 1 for closed or open, respectively
    - **latitude:** float, latitude
    - **longitude:** float, longitude
    - **name:** string, the business's name
    - **neighborhood:** string, the neighborhood's name
    - **postal_code:** string, the postal code
    - **price_range:** integer, 0, 1, 2, 3, or 4 for price of meal/service with 0 being least expensive and 4 being most expensive
    - **review_count:** integer, number of reviews
    - **stars:** float, star rating, rounded to half-stars
    - **state:** string, 2 character state code, if applicable
    - **take_reservations:** integer, 0 or 1 for does not take reservations or takes reservations, respectively
    - **takes_credit_cards:** integer, 0 or 1 for does not take credit cards or does take credit cards, respectively
- yelp_review.json: contains Yelp review metadata by business
    - **average_review_age:** float, average age of reviews, in days, on business' Yelp page

    - **average_review_length:** float, average length of review, in characters, on business' Yelp page
    - **average_review_sentiment:** float, from -1 to 1, representing the average sentiment of reviews on business' Yelp page, with -1 being most negative, 0 being neutral, and 1 being positive
    - **business_id:** string, 22 character unique string business id
    - **number_cool_votes:** integer, total number of cool votes given to reviews on business' Yelp page
    - **number_funny_votes:** integer, total number of funny votes given to reviews on business' Yelp page
    - **number_useful_votes:** integer, total number of useful votes given to reviews on business' Yelp page
- yelp_user.json: contains user profile metadata by business
    - **average_days_on_yelp:** float, average days on Yelp for all businesses' reviewers
    - **average_number_fans:** float, average number of fans for all businesses' reviewers
    - **average_number_friends:** float, average number of friends for all businesses' reviewers
    - **average_number_years_elite:** float, average number of years with elite status for all businesses' reviewers
    - **average_review_count:** float, average number of total reviews for all businesses' reviewers
    - **business_id:** string, 22 character unique string business id
- yelp_checkin.json: contains online checkin metadata by business
    - **business_id:** string, 22 character unique string business id
    - **time:** nested object of the day of the week with key of the hour (using a 24hr clock) with the count of checkins for that hour (e.g. 14:00 - 14:59)
    - **weekday_checkins:** integer, total number of checkins at business during Mondays, Tuesdays, Wednesdays and Thursdays
    - **weekend_checkins:** integer, total number of checkins at business during Fridays, Saturdays and Sundays
- yelp_tip.json: contains tip metadata by business
    - **average_tip_length:** float, average length of tip, in characters, on business' Yelp page
    - **business_id:** string, 22 character unique string business id
    - **number_tips:** integer, total number of tips left on business' Yelp page
- yelp_photo.json: contains photo metadata by business
    - **average_caption_length:** float, average length of photo caption, in characters, for photos on business' Yelp page
    - **business_id:** string, 22 character unique string business id
    - **number_pics:** integer, total number of pictures posted on business' Yelp page