

Dataset Description.

The dataset comprises of **three** csv files, namely user.csv, business.csv and review.csv. Note that some of the content, such as id fields are encoded. Note that the files are separated by "^" character.

1. Business.csv file contain basic information about local businesses.

Business.csv file contains the following columns

"business_id", "full_address", "categories"

'business_id': (a unique identifier for the business)

'full_address': (localized address),

'categories': [(localized category names)]

2. Review.csv file contains the star rating given by a user to a business. Use user_id to associate this review with others by the same user. Use business_id to associate this review with others of the same business.

review.csv file contains the following columns

"review_id", "user_id", "business_id", "stars"

'review_id': (a unique identifier for the review)

'user_id': (the identifier of the reviewed business),

'business_id': (the identifier of the authoring user),

'stars': (star rating, integer 1-5), the rating given by the user to a business

3. user.csv file contains aggregate information about a single user across all of Yelp

user.csv file contains the following columns "user_id", "name", "url"

user_id': (unique user identifier),

'name': (first name, last initial, like 'Matt J. '), this column has been made anonymous to preserve privacy

'url': url of the user on yelp