**Yelp Database Description for Big Assignment 2024**

The dataset for the big assignment is sourced from the Internet and is based on a real-world business platform, Yelp, which is openly accessible. It includes 16 tables containing information on various aspects of Yelp, such as:

i) attributes of different business brands,

ii) reviews of these businesses, and

iii) the reviewers themselves.

To understand how these tables interrelate, please review each table’s attributes and identify the keys that connect them. This will help you determine how to join the tables effectively. We recommend familiarizing yourself with the dataset before beginning any analysis. If you have experience writing or reading Yelp reviews, this will likely help you better understand the dataset's structure.

While the structure may seem complex at first, it’s not difficult to grasp once you start exploring the data. Keep in mind that real-world database structures are often even more intricate, even without offering any explanation about data structure.

* Please remove unrelated tables in your database before you import the Yelp dataset for your assignment.

Below, the table names are bolded, and attributes within each table are listed inside brackets. This is provided to present the information in the database in a concise way.

**Business** (business\_id, full\_address, active, city, review\_count, business\_name, neighborhoods, latitude, longitude, state, stars)

**Business\_attributes** (business\_id, attribute\_name, attribute\_value)

**Business\_attributes\_ambience** (business\_id, subattribute, value)

**Business\_attributes\_goodfor** (business\_id, subattribute, value)

**Business\_attributes\_hairtypesspecializedin** (business\_id, subattribute, value)

**Business\_attributes\_music** (business\_id, subattribute, value)

**Business\_attributes\_parking** (business\_id, subattribute, value)

**Business\_categories** (business\_id, category)

**Business\_hours** (business\_id, day\_of\_week, opening\_time)

**Checkins** (business\_id, day\_of\_week, times 0-23)

**Compliments** (user\_id, year)

**Elite** (user\_id, year)

**Friends** (user\_id1, user\_id2)

**Reviews** (review\_id, business\_id, user\_id, stars, review\_date, votes\_funny, votes\_useful, votes\_cool, review\_text)

**Tips** (business\_id, user\_id, date\_year, date\_month, date\_day, likes, tip\_text)

**Users** (user\_id, votes\_funny, votes\_useful, votes\_cool, name, average\_stars, review\_count, fans, yelping\_since\_year, yelping\_since\_month)

Description of information each table contains

|  |  |
| --- | --- |
| **Table name (rows of data)** | **Explanation** |
| Business (15 585) | Contains information about the businesses users have reviewed in Yelp. Each business appears only once in this table. |
| Business\_attributes (92 783) | Contains information about attributes (characteristic/ service, etc.) businesses have, as rated or reported by Yelp users. One business may have multiple attributes and not all businesses have attributes. |
| Business\_attributes\_ambience (36 593) | Contains information on the ambience/atmosphere descriptors that users have rated/reported the businesses have. One business may have multiple ambience categories and not all businesses have attributes in ambience. |
| Business\_attributes\_goodfor (28 187) | Contains information on the users’ reviews/reports about what the businesses are good for. Value column indicates if users have reported the goodfor attribute as suitable for a business. One business may have multiple goodfor attributes and not all businesses have goodfor attributes. |
| Business\_attributes\_  Hairtypesspecializedin (320) | Contains information on the specific hair types businesses specialize in. The value column provides more details for the availability of specialization sub attributes. One business may have multiple hair type specialization types and not all businesses offer hair services. |
| Business\_attributes\_music  (3 290) | Contains information about the types of background music businesses play. Value column offers more details for the music sub attribute. One business may have multiple music types and not all businesses play music. |
| Business\_attributes\_parking (46 280) | Contains information about the parking availability options at businesses, where value column offers additional details for the parking sub attribute. One business may have multiple parking types and not all businesses have parking attributes. |
| Business\_categories (43 703) | Contains information about the business categories or business types. One business can belong to multiple categories and not all businesses have been rated/reported to have a category. |
| Business\_hours (47 831) | Provides information on business opening hours on weekly basis. One business may appear on multiple weekdays and have different opening hours on different days of the week. |
| Checkins (80 038) | Contains information on amounts of customer check-ins, categorized based on weekday and time of the day. One business may appear in the table on multiple weekdays and not all businesses are listed in this table. |
| Compliments (98 810) | Contains information about compliments given by users. Compliments are categorized into different types and the number of certain types of compliment user has given is included in the table. One user may have given multiple compliments and in various types. |
| Elite (16 366) | Presents information on the years in which certain users have been recorded as elite by Yelp. One user may have been elite during multiple years. ­­ |
| Friends (303 032) | Shows users who are friends with each other. |
| Reviews (335 022) | Contains information about the reviews users have given for businesses. The review\_id is unique in this table, hence one review appears only once. However, each user may write multiple reviews and each business may receive reviews from multiple users. |
| Tips (113 993) | Provides information about tips (shorter advices/recommendations) users have left for businesses in Yelp. Not all businesses have received tips and not all users have left tip information. Also, one user or one business may be listed multiple times in the table. |
| Users (70 817) | Contains information of all the Yelp users who have given reviews in this dataset. Each user appears only once in this table. |

**Note1:** Some tables contain the key, such as business\_id, in multiple rows.

**Note 2:** The dataset includes data on business reviews from selected cities and regions. The "users" table tracks the review activities of users, but it's important to note that users may have posted reviews for businesses in locations not covered by our dataset. Consequently, a user's total review count may exceed the number of reviews actually included in this dataset.