# Introduction to Yelp Business User Data



Take a look at the data table above. This is a sample of data that was extracted from Yelp, an application that allows users to leaving ratings and comments about businesses they’ve visited. The data table includes information about the location, type, and ratings of various businesses across the United States and Canada. This information, often referred to as *attributes* or *features* of the data set, is organized into columns (e.g. *business\_id* as the first column*)*. The specific columns are defined in the table below:

|  |  |
| --- | --- |
| Column Name | Description |
| business\_id | a series of letters and numbers assigned to identify each business |
| name | the name of the business |
| address | the physical address of the business |
| city | the city the business is located |
| state | the state or providence where each business is located |
| postal\_code | the postal code (zip code) of the business, in the following formats:  5 numbers – United States  6 letters/numbers – Canada |
| latitude | the latitudinal coordinate of the business |
| longitude | the longitudinal coordinate of the business |
| stars | the average rating of each business, on a scale of 1 to 5 stars |
| review\_count | the number of times the business has been reviewed by a customer |
| is\_open | a binary indicator of whether or not the business is currently open  1 = open  0 = closed |
| categories | a list of the different categories each business could be called |

The table above is a simplified **data dictionary**. A data dictionary is a description of what all of the columns in a data table or a set of data tables contain, called **metadata** (data about data). Metadata can include the column descriptions as well as other information, such as the time that the data was recorded. Data dictionaries can also describe how different data tables relate to one another. A data dictionary is a crucial tool used by data scientists to ensure that data is organized and easy to comprehend.