# Relational Schema

Business(
    Business_Id: varchar(25) [PK],
    name: varchar(30),
    address: varchar(65),
    city: varchar(20),
    state: varchar(15),
    postal_code: int [FK to Location.postal_code],
    latitude: float(10),
    longitude: float(10),
    stars: float(10),
    review_count: int,
    attributes: varchar(1000),
    categories: varchar(1000),
    monday_hours: varchar(9),
    tuesday_hours: varchar(9),
    wednesday_hours: varchar(9),
    thursday_hours: varchar(9),
    friday_hours: varchar(9),
    saturday_hours: varchar(9),
    sunday_hours: varchar(9)
)

The Business table has information about the businesses in the Yelp dataset. The following are descriptions of its attributes:

- business_Id: Unique identifier for each of the businesses in the dataset
- name: The name of the business
- address: The address of the business
- city: The city in which the business is located
- state: The state in which the business is located
- postal_code: The postal code in which the business is located
- latitude: The latitude in which the business is located
- longitude: The latitude in which the business is located
- stars: The average number of stars the business has received, across all reviews in the dataset
- review_count: The number of reviews the business has received
- attributes: The attributes of the business (e.g., “Food”)
- categories: The categories of the business (e.g., “BusinessAcceptsCreditCards”)
- monday_hours: The hours during which the business is open on Monday
- tuesday_hours: The hours during which the business is open on Tuesday
- wednesday_hours: The hours during which the business is open on Wednesday
- thursday_hours:  The hours during which the business is open on Thursday

- friday_hours:  The hours during which the business is open on Friday
- saturday_hours:  The hours during which the business is open on Saturday
- sunday_hours:  The hours during which the business is open on Sunday

Weather_By_Zip(
    postal_code: int [PK],
    month: varchar(9) [PK],
    avg_temp: float(20),
    high_temp: float(20),
    low_temp: float(20),
    total_rain_inches float(10)
)

The Weather_By_Zip table has information about the average weather conditions throughout the year for different postal codes.

- postal_code: The postal code corresponding to the weather conditions
- month: The month of the weather conditions
- avg_temp: The average temperature for this postal code in this month
- high_temp: The high temperature for this postal code in this month
- low_temp: The low temperature for this postal code in this month
- total_rain_inches: The total amount of rain, in inches, this postal code received

Reviews(
    review_id: varchar(25) [PK],
    user_id: varchar(25) [FK to Users.user_id],
    business_id: varchar(100) [FK to Business.Business_Id],
    stars: int,
    useful: int,
    funny: int,
    cool: int,
    text: varchar(6000),
    date: varchar(24)
)

The Reviews table has information about all of the reviews in the Yelp dataset.

- review_id: The ID corresponding to the review in the Yelp dataset
- user_id: The ID corresponding to the user in the Yelp dataset
- business_id: The ID of the business which the review concerns
- stars: The number of stars which the user gave the business
- useful: The number of users who have indicated the review as useful
- funny: The number of users who have indicated the review as funny
- cool: The number of users who have indicated the review as cool
- text: The text of the review
- date: The date on which the review was published to Yelp

Tips(
    tip_id: int [PK], user_id: varchar(25) [FK to Users.user_id],
    business_id: varchar(30),
    text: varchar(3000),
    date: varchar(24),
    compliment_count: int
)

The Tips table has recommendations from users to other users.

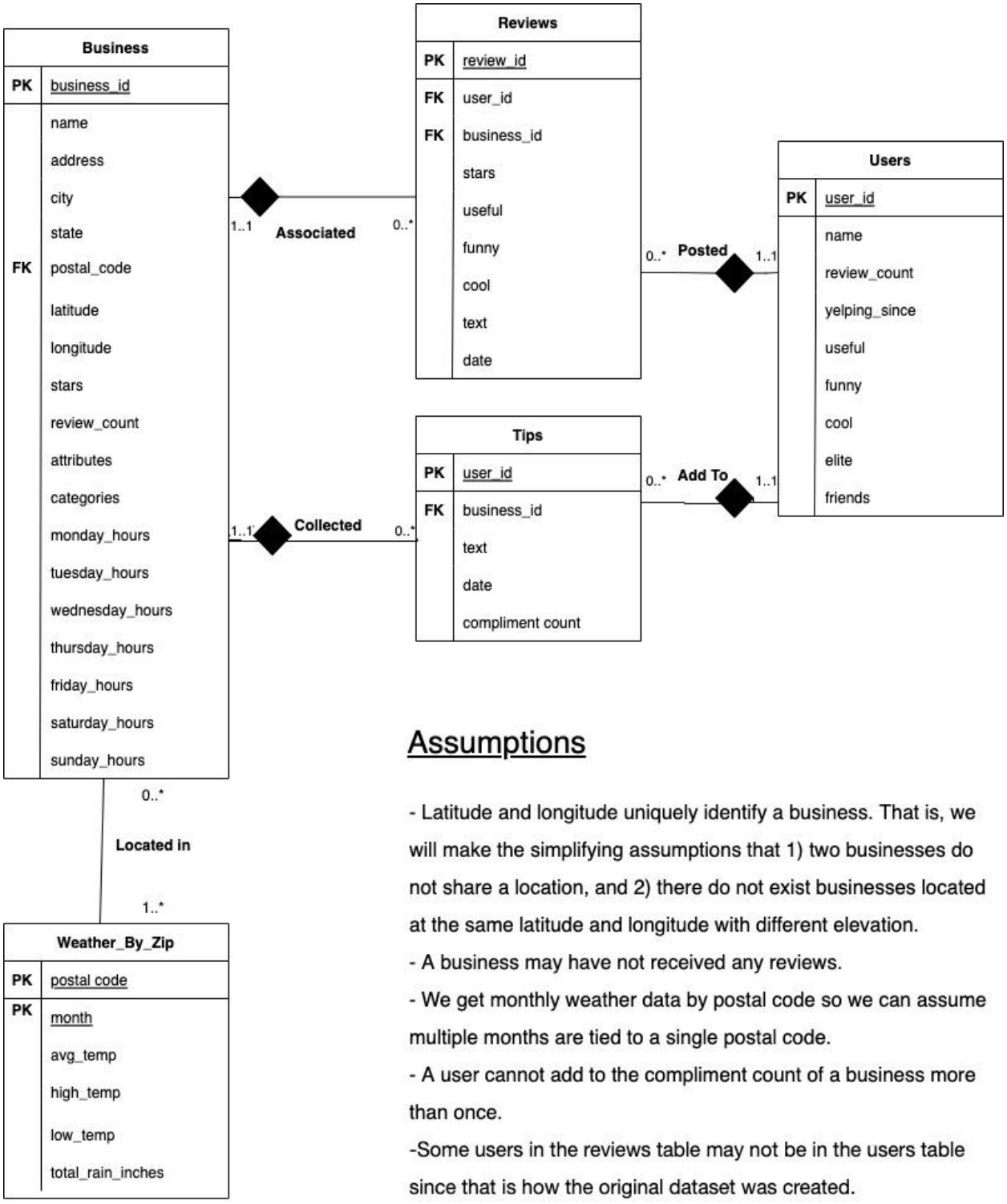
- tip_id: The ID corresponding to the tip the user has given
- business_id: The ID corresponding to the business the user gave a tip for
- text: The text of the tip
- date: The date on which the tip was given
- compliment_count: The number of people who found the tip helpful

Users(
    user_id: varchar(25) [PK],
    name: varchar(30),
    review_count: int,
    yelping_since: varchar(24),
    useful: int,
    funny: int,
    cool: int,
    elite: int,
    friends: int
)

The Users table has information about Yelp users in the dataset.

- user_id: The ID corresponding to the user
- name: The user's username
- review_count: The number of reviews which the user has given
- yelping_since: The date on which the user began using Yelp
- useful: The number of users who have marked this user's review as useful
- funny: The number of users who have marked this user's review as funny
- cool: The number of users who have marked this user's review as cool
- elite: The number of users who have marked this user's review as elite
- friends: The number of users who are friends with this user on Yelp

# ER/UML Diagram




## Assumptions

- Latitude and longitude uniquely identify a business. That is, we will make the simplifying assumptions that 1) two businesses do not share a location, and 2) there do not exist businesses located at the same latitude and longitude with different elevation.
- A business may have not received any reviews.
- We get monthly weather data by postal code so we can assume multiple months are tied to a single postal code.
- A user cannot add to the compliment count of a business more than once.
-Some users in the reviews table may not be in the users table since that is how the original dataset was created.