Data Scientist Role Play: Profiling and Analyzing the Yelp Dataset Coursera Worksheet

This is a 2-part assignment. In the first part, you are asked a series of questions that will help you profile and understand the data just like a data scientist would. For this first part of the assignment, you will be assessed both on the correctness of your findings, as well as the code you used to arrive at your answer. You will be graded on how easy your code is to read, so remember to use proper formatting and comments where necessary.

In the second part of the assignment, you are asked to come up with your own inferences and analysis of the data for a particular research question you want to answer. You will be required to prepare the dataset for the analysis you choose to do. As with the first part, you will be graded, in part, on how easy your code is to read, so use proper formatting and comments to illustrate and communicate your intent as required.

For both parts of this assignment, use this "worksheet." It provides all the questions you are being asked, and your job will be to transfer your answers and SQL coding where indicated into this worksheet so that your peers can review your work. You should be able to use any Text Editor (Windows Notepad, Apple TextEdit, Notepad ++, Sublime Text, etc.) to copy and paste your answers. If you are going to use Word or some other page layout application, just be careful to make sure your answers and code are lined appropriately. In this case, you may want to save as a PDF to ensure your formatting remains intact for you reviewer.

Part 1: Yelp Dataset Profiling and Understanding

1. Profile the data by finding the total number of records for each of the tables below:
SELECT COUNT(*)
FROM table

i. Attribute table =10000
ii. Business table =10000
iii. Category table =10000
iv. Checkin table =10000
v. elite_years table =10000
vi. friend table = 10000
vii. hours table =10000
viii. photo table = 10000
ix. review table = 10000
x. tip table = 10000
xi. user table =10000

2. Find the total distinct records by either the foreign key or primary key for each table. If two foreign keys are listed in the table, please specify which foreign key.
SELECT COUNT(DISTINCT(key))
FROM table
i. Business =10000
ii. Hours =business_id: 1562
iii. Category =business_id: 2643
iv. Attribute =business_id: 1115
v. Review =id:10000, business_id: 8090
vi. Checkin = business_id: 493
vii. Photo =id: 10000, business_id: 6493
viii. Tip = user_id: 537, business_id: 3979
ix. User = id: 10000
x. Friend = user_id: 11
xi. Elite_years =user_id: 2780

Note: Primary Keys are denoted in the ER-Diagram with a yellow key icon.

3. Are there any columns with null values in the Users table? Indicate "yes," or "no."

    Answer:No

```
        SQL code used to arrive at answer:
SELECT COUNT(*)
            FROM user
            WHERE id IS NULL OR
              name IS NULL OR
              review_count IS NULL OR
              yelping_since IS NULL OR
              useful IS NULL OR
              funny IS NULL OR
              cool IS NULL OR
              fans IS NULL OR
              average_stars IS NULL OR
              compliment_hot IS NULL OR
              compliment_more IS NULL OR
              compliment_profile IS NULL OR
              compliment_cute IS NULL OR
              compliment_list IS NULL OR
              compliment_note IS NULL OR
              compliment_plain IS NULL OR
              compliment_cool IS NULL OR
              compliment_funny IS NULL OR
              compliment_writer IS NULL OR
              compliment_photos IS NULL
```

4. For each table and column listed below, display the smallest (minimum), largest (maximum), and average (mean) value for the following fields:

```
SELECT AVG(column)
FROM table
```
i. Table: Review, Column: Stars

    min:1   max:5   avg:3.7082

    ii. Table: Business, Column: Stars

    min:1   max:5   avg:3.6459

    iii. Table: Tip, Column: Likes

    min:0   max:2   avg:0.0144

    iv. Table: Checkin, Column: Count

    min:1   max:53   avg:1.9414

    v. Table: User, Column: Review_count

    min:0   max:2000   avg:24.2995

5. List the cities with the most reviews in descending order:

```
        SQL code used to arrive at answer:
            SELECT city,
            SUM(review_count) AS reviews
            FROM business
            GROUP BY city
            ORDER BY reviews DESC
```

Copy and Paste the Result Below:

```
+-----------------+---------+
| city            | reviews |
+-----------------+---------+
| Las Vegas       |   82854 |
| Phoenix         |   34503 |
| Toronto         |   24113 |
| Scottsdale      |   20614 |
| Charlotte       |   12523 |
| Henderson       |   10871 |
| Tempe           |   10504 |
| Pittsburgh      |    9798 |
| MontrÃƒÂ©al      |    9448 |
| Chandler        |    8112 |
| Mesa            |    6875 |
| Gilbert         |    6380 |
| Cleveland       |    5593 |
| Madison         |    5265 |
| Glendale        |    4406 |
| Mississauga     |    3814 |
| Edinburgh       |    2792 |
| Peoria          |    2624 |
| North Las Vegas |    2438 |
| Markham         |    2352 |
| Champaign       |    2029 |
| Stuttgart       |    1849 |
| Surprise        |    1520 |
| Lakewood        |    1465 |
| Goodyear        |    1155 |
+-----------------+---------+
```

6. Find the distribution of star ratings to the business in the following cities:

i. Avon

SQL code used to arrive at answer:

```
SELECT stars,
          SUM(review_count) AS count
FROM business
WHERE city == 'Avon'
GROUP BY stars
```

Copy and Paste the Resulting Table Below (2 columns – star rating and count):

```
+-------+-------+
| stars | count |
+-------+-------+
|   1.5 |    10 |
|   2.5 |     6 |
|   3.5 |    88 |
|   4.0 |    21 |
|   4.5 |    31 |
|   5.0 |     3 |
+-------+-------+
```

ii. Beachwood

SQL code used to arrive at answer:

```
SELECT stars,
          SUM(review_count) AS count
FROM business
WHERE city == 'Beachwood'
GROUP BY stars
```

Copy and Paste the Resulting Table Below (2 columns - star rating and count):

```
+-------+-------+
| stars | count |
+-------+-------+
|   2.0 |     8 |
|   2.5 |     3 |
|   3.0 |    11 |
|   3.5 |     6 |
|   4.0 |    69 |
|   4.5 |    17 |
|   5.0 |    23 |
+-------+-------+
```

7. Find the top 3 users based on their total number of reviews:

    SQL code used to arrive at answer:

```
SELECT id,
        name,
        review_count
FROM user
ORDER BY review_count DESC
LIMIT 3
```

    Copy and Paste the Result Below:

```
+------------------------+--------+--------------+
| id                     | name   | review_count |
+------------------------+--------+--------------+
| -G7Zkl1wIWBBmD0KRy_sCw | Gerald |         2000 |
| -3s52C4zL_DHRK0ULG6qtg | Sara   |         1629 |
| -8lbUNlXVSoXqaRRiHiSNg | Yuri   |         1339 |
+------------------------+--------+--------------+
```

8. Does posing more reviews correlate with more fans?

    Yes, but also the amount of time that they have been yelping. The longer they have been yelping and the more reviews they give has a higher fan count.

    Please explain your findings and interpretation of the results:

```
SELECT id,
        name,
        review_count,
        fans,
        yelping_since
FROM user
ORDER BY fans DESC
```

```
+------------------------+-----------+--------------+------+--------------
-------+
| id                     | name      | review_count | fans | yelping_since
|
+------------------------+-----------+--------------+------+--------------
-------+
| -9I98YbNQnLdAmcYfb324Q | Amy       |          609 |  503 | 2007-07-19
00:00:00 |
| -8EnCioUmDygAbsYZmTeRQ | Mimi      |          968 |  497 | 2011-03-30
00:00:00 |
| --2vR0DIsmQ6WfcSzKWigw | Harald    |         1153 |  311 | 2012-11-27
00:00:00 |
| -G7Zkl1wIWBBmD0KRy_sCw | Gerald    |         2000 |  253 | 2012-12-16
00:00:00 |
| -0IiMAZI2SsQ7VmyzJjokQ | Christine |          930 |  173 | 2009-07-08
```

```
00:00:00 |
              | -g3XIcCb2b-BD0QBCcq2Sw | Lisa      |          813 | 159 | 2009-10-05
00:00:00 |
              | -9bbDysuiWeo2VShFJJtcw | Cat       |          377 | 133 | 2009-02-05
00:00:00 |
              | -FZBTkAZEXoP7CYvRV2ZwQ | William   |         1215 | 126 | 2015-02-19
00:00:00 |
              | -9da1xk7zgnnfO1uTVYGkA | Fran      |          862 | 124 | 2012-04-05
00:00:00 |
              | -lh59ko3dxChBSZ9U7LfUw | Lissa     |          834 | 120 | 2007-08-14
00:00:00 |
              | -B-QEUESGWHPE_889WJaeg | Mark      |          861 | 115 | 2009-05-31
00:00:00 |
              | -DmqnhW4Omr3YhmnigaqHg | Tiffany   |          408 | 111 | 2008-10-28
00:00:00 |
              | -cv9PPT7IHux7XUc9dOpkg | bernice   |          255 | 105 | 2007-08-29
00:00:00 |
              | -DFCC64NXgqrxlO8aLU5rg | Roanna    |         1039 | 104 | 2006-03-28
00:00:00 |
              | -IgKkE8JvYNWeGu8ze4P8Q | Angela    |          694 | 101 | 2010-10-01
00:00:00 |
              | -K2Tcgh2EKX6e6HqqIrBIQ | .Hon      |         1246 | 101 | 2006-07-19
00:00:00 |
              | -4viTt9UC44lWCFJwleMNQ | Ben       |          307 |  96 | 2007-03-10
00:00:00 |
              | -3i9bhfvrM3F1wsC9XIB8g | Linda     |          584 |  89 | 2005-08-07
00:00:00 |
              | -kLVfaJytOJY2-QdQoCcNQ | Christina |          842 |  85 | 2012-10-08
00:00:00 |
              | -ePh4Prox7ZXnEBNGKyUEA | Jessica   |          220 |  84 | 2009-01-12
00:00:00 |
              | -4BEUkLvHQntN6qPfKJP2w | Greg      |          408 |  81 | 2008-02-16
00:00:00 |
              | -C-l8EHSLXtZZVfUAUhsPA | Nieves    |          178 |  80 | 2013-07-08
00:00:00 |
              | -dw8f7FLaUmWR7bfJ_Yf0w | Sui       |          754 |  78 | 2009-09-07
00:00:00 |
              | -8lbUNlXVSoXqaRRiHiSNg | Yuri      |         1339 |  76 | 2008-01-03
00:00:00 |
              | -0zEEaDFIjABtPQni0XlHA | Nicole    |          161 |  73 | 2009-04-30
00:00:00 |
              +------------------------+-----------+--------------+------+-------------
-------+
```

9. Are there more reviews with the word "love" or with the word "hate" in them?

　　　　Answer: love has 1780, while hate only has 232   :) 'love prevails'

　　　SQL code used to arrive at answer:

```
            SELECT COUNT(*)
SELECT COUNT(*)
            FROM review
FROM review
            WHERE text LIKE "%love%"
WHERE text LIKE "%hate%"

            = 1780
```


10. Find the top 10 users with the most fans:

　　　SQL code used to arrive at answer:

```
            SELECT id,
                    name,
```

```
              fans
     FROM user
     ORDER BY fans DESC
     LIMIT 10
```

Copy and Paste the Result Below:

```
+------------------------+-----------+------+
| id                     | name      | fans |
+------------------------+-----------+------+
| -9I98YbNQnLdAmcYfb324Q | Amy       |  503 |
| -8EnCioUmDygAbsYZmTeRQ | Mimi      |  497 |
| --2vR0DIsmQ6WfcSzKWigw | Harald    |  311 |
| -G7Zkl1wIWBBmD0KRy_sCw | Gerald    |  253 |
| -0IiMAZI2SsQ7VmyzJjokQ | Christine |  173 |
| -g3XIcCb2b-BD0QBCcq2Sw | Lisa      |  159 |
| -9bbDysuiWeo2VShFJJtcw | Cat       |  133 |
| -FZBTkAZEXoP7CYvRV2ZwQ | William   |  126 |
| -9da1xk7zgnnfO1uTVYGkA | Fran      |  124 |
| -lh59ko3dxChBSZ9U7LfUw | Lissa     |  120 |
+------------------------+-----------+------+
```

Part 2: Inferences and Analysis

    i.   Do the two groups you chose to analyze have a different distribution of hours?

        The 4-5 star group seems to have shorter hours then the 2-3 star group.
        Please note the query returned only three businesses so not a great
        sample size.

    ii.   Do the two groups you chose to analyze have a different number of reviews?

        Yes and no, one of the 4-5 star group has a lot more reviews but then the other
        4-5 star group has close to the same number of reviews as the 2-3 star group

    iii.   Are you able to infer anything from the location data provided between these two
        groups? Explain.

        No, every business is in a different zip-code.

SQL code used for analysis:

```
SELECT B.name,
       B.review_count,
       H.hours,
       postal_code,
       CASE
              WHEN hours LIKE "%monday%" THEN 1
              WHEN hours LIKE "%tuesday%" THEN 2
              WHEN hours LIKE "%wednesday%" THEN 3
              WHEN hours LIKE "%thursday%" THEN 4
              WHEN hours LIKE "%friday%" THEN 5
              WHEN hours LIKE "%saturday%" THEN 6
              WHEN hours LIKE "%sunday%" THEN 7
       END AS ord,
       CASE
              WHEN B.stars BETWEEN 2 AND 3 THEN '2-3 stars'
              WHEN B.stars BETWEEN 4 AND 5 THEN '4-5 stars'
       END AS star_rating
```

```
                    FROM business B INNER JOIN hours H
                    ON B.id = H.business_id
                    INNER JOIN category C
                    ON C.business_id = B.id
                    WHERE (B.city == 'Las Vegas'
                    AND
                    C.category LIKE 'shopping')
                    AND
                    (B.stars BETWEEN 2 AND 3
                    OR
                    B.stars BETWEEN 4 AND 5)
                    GROUP BY stars,ord
                    ORDER BY ord,star_rating ASC
```

2. Group business based on the ones that are open and the ones that are closed. What
differences can you find between the ones that are still open and the ones that are
closed? List at least two differences and the SQL code you used to arrive at your answer.

i.   Difference 1:

               The businesses that are open tend to have more reviews than ones that
               are closed on average.

                       Open: AVG(review_count) = 31.757
                       Closed: AVG(review_count0 = 23.198

        ii.   Difference 2:

               The average star rating is higher for businesses that are open than
               businesses that are closed.

                       Open: AVG(stars) = 3.679
                       Closed: AVG(stars) = 3.520

        SQL code used for analysis:

               SELECT COUNT(DISTINCT(id)),
                       AVG(review_count),
                       SUM(review_count),
                       AVG(stars),
                       is_open
               FROM business
               GROUP BY is_open

3. For this last part of your analysis, you are going to choose the type of analysis you
want to conduct on the Yelp dataset and are going to prepare the data for analysis.

Ideas for analysis include: Parsing out keywords and business attributes for sentiment
analysis, clustering businesses to find commonalities or anomalies between them,
predicting the overall star rating for a business, predicting the number of fans a user
will have, and so on. These are just a few examples to get you started, so feel free to be
creative and come up with your own problem you want to solve. Provide answers, in-line, to
all of the following:

i.   Indicate the type of analysis you chose to do:

               Predicting whether a business will stay open or close. We wish not to
        explicitly
               examine the text of the reviews, but this would be an interesting
        analysis.

ii.   Write 1-2 brief paragraphs on the type of data you will need for your analysis

and why you chose that data:

To better help businesses understand the importance of different factors which

will help their business stay open. Some data that may be important; number of

reviews, star rating of business, hours open, and of course location location. We will gather the latitude and longitude as well as city, state,

postal_code, and address to make processing easier later on. Categories and

attributes will be used to better distinguish between different types of businesses. `is_open` will determine which business is open and which business

have closed (not hours) but permanently.

iii.   Output of your finished dataset:

```
+------------------------+-------------------------------+---------------
---------------+--------------+-------+-------------+---------+-----------+---------------
+-------+-------------+--------------+----------------+----------------+--------------
+----------------+-------------+-----------------------------------------------
------------------------------------------------------------------------------
------------------------------------------------+---------------------------
------------------------------------------------------------------------------
------------------------------------------------------------------------------
------------------------------------------------------------------------------
------------------------+---------+
                | id                     | name                          | address
| city           | state | postal_code | latitude | longitude | review_count | stars |
monday_hours | tuesday_hours | wednesday_hours | thursday_hours | friday_hours |
saturday_hours | sunday_hours | categories
| attributes
| is_open |
                +------------------------+-------------------------------+---------------
---------------+--------------+-------+-------------+---------+-----------+---------------
+-------+-------------+--------------+----------------+----------------+--------------
+----------------+-------------+-----------------------------------------------
------------------------------------------------------------------------------
------------------------------------------------+---------------------------
------------------------------------------------------------------------------
------------------------------------------------------------------------------
------------------------------------------------------------------------------
------------------------+---------+
                | -0DET7VdEQOJVJ_v6klEug | Flaming Kitchen               | 3235 York
Regional Road 7    | Markham        | ON    | L3R 3P9     | 43.8484 | -79.3487 |
25 |   3.0 | 12:00-23:00  | 12:00-23:00   | 12:00-23:00     | 12:00-23:00    | 12:00-23:00
| 12:00-23:00    | 12:00-23:00  | Asian Fusion,Restaurants
|
RestaurantsTableService,GoodForMeal,Alcohol,Caters,HasTV,RestaurantsGoodForGroups,NoiseLev
el,WiFi,RestaurantsAttire,RestaurantsReservations,OutdoorSeating,RestaurantsPriceRange2,Bi
keParking,RestaurantsDelivery,Ambience,RestaurantsTakeOut,GoodForKids,BusinessParking
|      1 |
                | -2HjuT4yjLZ3b5f_abD87Q | Freeman's Car Stereo          | 4821 South
Blvd              | Charlotte      | NC    | 28217       | 35.1727 | -80.8755 |
8 |   3.5 | 9:00-19:00   | 9:00-19:00    | 9:00-19:00      | 9:00-19:00     | 9:00-19:00
| 9:00-17:00     | None         | Electronics,Shopping,Automotive,Car Stereo Installation
| BusinessAcceptsCreditCards,RestaurantsPriceRange2,BusinessParking,WheelchairAccessible
|      1 |
                | -CdstAUdEvci8GeJG8owpQ | Motors & More                 | 2315 Highland
Dr                | Las Vegas      | NV    | 89102       | 36.1465 | -115.167 |        7
|   5.0 | 7:00-17:00   | 7:00-17:00    | 7:00-17:00      | 7:00-17:00     | 7:00-17:00   |
8:00-12:00      | None         | Home Services,Solar Installation,Heating & Air
Conditioning/HVAC
| BusinessAcceptsCreditCards,BusinessAcceptsBitcoin,ByAppointmentOnly
```

```
|       1 |
                        | -K4gAv8_vjx8-2BxkVeRkA | Baby Cakes                     | 4145 Erie St
 | Willoughby      | OH    | 44094       | 41.6399 | -81.4064 |        5 |     3.5 | None
 | 11:00-17:00     | 11:00-17:00   | 11:00-20:00     | 11:00-17:00     | 10:00-17:00     | None
 | Bakeries,Food
 | BusinessAcceptsCreditCards,RestaurantsTakeOut,WheelchairAccessible,RestaurantsDelivery
|       1 |
                        | -PtTGvWsckUL8tTutHr6Ew | Snip-its Rocky River           | 21609 Center
Ridge Rd        | Rocky River   | OH    | 44116       | 41.4595 | -81.8587 |       18
 |   2.5 | 10:00-19:00   | 10:00-19:00   | 10:00-19:00     | 10:00-19:00     | 10:00-19:00     |
9:00-17:30      | 10:00-16:00   | Beauty & Spas,Hair Salons
 |
BusinessAcceptsCreditCards,RestaurantsPriceRange2,GoodForKids,BusinessParking,ByAppointmen
tOnly
|       1 |
                        | -ayZoW_iNDsunYXX_0x1YQ | Standard Restaurant Supply     | 2922 E
McDowell Rd        | Phoenix       | AZ    | 85008       | 33.4664 | -112.018 |
15 |   3.5 | 8:00-18:00    | 8:00-18:00    | 8:00-18:00      | 8:00-18:00      | 8:00-18:00
 | 9:00-17:00     | None          | Shopping,Wholesalers,Restaurant Supplies,Professional
Services,Wholesale Stores
 |
BusinessAcceptsCreditCards,RestaurantsPriceRange2,BusinessParking,BikeParking,WheelchairAc
cessible
|       1 |
                        | -d9qyfNhLMQwVVg_raBKeg | What A Bagel                   | 973 Eglinton
Avenue W       | York          | ON    | M6C 2C4     | 43.6999 | -79.4295 |        8
 |   3.0 | 6:00-15:30    | 6:00-15:30    | 6:00-15:30      | 6:00-15:30      | 6:00-15:30    |
6:00-15:30     | None          | Restaurants,Bagels,Breakfast & Brunch,Food
 | NoiseLevel,RestaurantsAttire,RestaurantsTableService,OutdoorSeating
|       1 |
                        | -hjbcaxaU9yYXY2iI-49sw | Pinnacle Fencing Solutions     |
 | Phoenix        | AZ    | 85060       | 33.4805 | -111.997 |       13 |    4.0 |
8:00-16:00     | 8:00-16:00    | 8:00-16:00    | 8:00-16:00      | 8:00-16:00      | None
 | None          | Home Services,Contractors,Fences & Gates
 | BusinessAcceptsCreditCards,ByAppointmentOnly
|       1 |
                        | -iu4FxdfxN4rU4Fu9BjiFw | Alterations Express            | 17240 Royalton
Rd        | Strongsville  | OH    | 44136       | 41.3141 | -81.8207 |        3 |
4.0 | 8:00-19:00    | 8:00-19:00    | 8:00-19:00      | 8:00-19:00      | 8:00-19:00    |
8:00-18:00     | None          | Shopping,Bridal,Dry Cleaning & Laundry,Local
Services,Sewing & Alterations
 |
BusinessParking,BusinessAcceptsCreditCards,RestaurantsPriceRange2,BusinessAcceptsBitcoin,B
ikeParking,ByAppointmentOnly,WheelchairAccessible
|       1 |
                        | -j4NsiRzSMrMk2N_bGH_SA | Extra Space Storage            | 2880 W Elliot
Rd        | Chandler      | AZ    | 85224       | 33.3496 | -111.892 |        5
 |   4.0 | 8:00-17:30    | 8:00-17:30    | 8:00-17:30      | 8:00-17:30      | 8:00-17:30    |
8:00-17:30     | 10:00-14:00   | Home Services,Self Storage,Movers,Shopping,Local
Services,Home Decor,Home & Garden
 | BusinessAcceptsCreditCards
|       1 |
                        | -uiBBVWI6tMDm2JFbZFrOw | Gussied Up                     | 1090 Bathurst
St        | Toronto       | ON    | M5R 1W5     | 43.6727 | -79.4142 |        6
 |   4.5 | None          | 11:00-19:00   | 11:00-19:00     | 11:00-19:00     | 11:00-19:00   |
11:00-17:00    | 12:00-16:00   | Women's Clothing,Shopping,Fashion
 | BusinessAcceptsCreditCards,RestaurantsPriceRange2,BusinessParking,BikeParking
|       1 |
                        | 0-aPEeNc2zVb5Gp-i7Ckqg | Buddy's Muffler & Exhaust      | 1509 Hickory
Grove Rd        | Gastonia      | NC    | 28056       | 35.2772 | -81.06   |        4
 |   5.0 | 8:30-17:00    | 8:30-17:00    | 8:30-17:00      | 8:30-17:00      | 8:30-17:00    |
9:00-15:00     | None          | Automotive,Auto Repair
 | BusinessAcceptsCreditCards
|       1 |
                        | 01xXe2m_z048W5gcBFpoJA | Five Guys                      | 2641 N 44th
St, Ste 100    | Phoenix       | AZ    | 85008       | 33.478  | -111.986 |
63 |   3.5 | 10:00-22:00   | 10:00-22:00   | 10:00-22:00     | 10:00-22:00     | 10:00-22:00
 | 10:00-22:00    | 10:00-22:00   | American (New),Burgers,Fast Food,Restaurants
 |
```

```
RestaurantsTableService,GoodForMeal,Alcohol,Caters,HasTV,RestaurantsGoodForGroups,NoiseLev
el,WiFi,RestaurantsAttire,RestaurantsReservations,OutdoorSeating,BusinessAcceptsCreditCard
s,RestaurantsPriceRange2,BikeParking,RestaurantsDelivery,Ambience,RestaurantsTakeOut,GoodF
orKids,DriveThru,BusinessParking                              |       1 |
                 | 06I2r8S3tHP_LwGnnkk6Uw | All Storage - Anthem              | 2620 W Horizon
Ridge Pkwy   | Henderson     | NV   | 89052      |    36.0021 |  -115.102 |         3 |
3.5 | 9:00-16:30    | 9:00-16:30    | 9:00-16:30      | 9:00-16:30      | 9:00-16:30    |
9:00-16:30       | None          | Truck Rental,Local Services,Self
Storage,Parking,Automotive
 | BusinessAcceptsCreditCards,BusinessAcceptsBitcoin
 |       1 |
                 | 07h3mGtTovPJE660nX6E-A | Mood                              | 1 Greenside
Place        | Edinburgh     | EDH  | EH1 3AA    |     55.957 |  -3.18502 |
11 |   2.0 | None          | None          | None          | 22:30-3:00      | 22:00-3:00
 | 22:00-3:00    | 22:30-3:00    | Dance Clubs,Nightlife
 |
Alcohol,OutdoorSeating,BusinessAcceptsCreditCards,RestaurantsPriceRange2,AgesAllowed,Music
,Smoking,RestaurantsGoodForGroups,WheelchairAccessible
 |       0 |
                 | 0AJF-USLN6K5T4caooDdjw | Starbucks                         | 4605 E
Chandler Blvd, Ste A | Phoenix      | AZ   | 85048      |    33.3044 |  -111.984 |
52 |   3.0 | 5:00-20:00    | 5:00-20:00    | 5:00-20:00      | 5:00-20:30      | 5:00-20:00
 | 5:00-20:00    | 5:00-20:00    | Coffee & Tea,Food
 |
BusinessParking,Caters,WiFi,OutdoorSeating,BusinessAcceptsCreditCards,RestaurantsPriceRang
e2,BikeParking,RestaurantsTakeOut
 |       1 |
                 | 0B3W6KxkD3o4W4l6cq735w | Big Smoke Burger                  | 260 Yonge
Street       | Toronto       | ON   | M4B 2L9    |    43.6546 |  -79.3805 |
47 |   3.0 | 10:30-21:00   | 10:30-21:00   | 10:30-21:00     | 10:30-21:00     | 10:30-21:00
 | 10:30-21:00   | 11:00-19:00   | Poutineries,Burgers,Restaurants
 |
RestaurantsTableService,GoodForMeal,Alcohol,Caters,HasTV,RestaurantsGoodForGroups,NoiseLev
el,WiFi,RestaurantsAttire,RestaurantsReservations,OutdoorSeating,BusinessAcceptsCreditCard
s,RestaurantsPriceRange2,WheelchairAccessible,BikeParking,RestaurantsDelivery,Ambience,Res
taurantsTakeOut,GoodForKids,DriveThru,BusinessParking |       1 |
                 | 0IySwcfqwJjpHPsYwjpAkg | Subway                            | 2904 Yorkmont
Rd           | Charlotte     | NC   | 28208      |    35.1903 |  -80.9288 |         7
 |   3.5 | 6:00-22:00    | 6:00-22:00    | 6:00-22:00      | 6:00-22:00      | 6:00-22:00    |
10:00-21:00      | None          | Fast Food,Restaurants,Sandwiches
 | Ambience,RestaurantsPriceRange2,GoodForKids
 |       1 |
                 | 0K2rKvqdBmiOAUTebcUohQ | Red Rock Canyon Visitor Center | 1000 Scenic
Loop Dr      | Las Vegas     | NV   | 89161      |    36.1357 |  -115.428 |
32 |   4.5 | 8:00-16:30    | 8:00-16:30    | 8:00-16:30      | 8:00-16:30      | 8:00-16:30
 | 8:00-16:30    | 8:00-16:30    | Education,Visitor Centers,Professional Services,Special
Education,Local Services,Community Service/Non-Profit,Hotels & Travel,Travel Services,Gift
Shops,Shopping,Parks,Hiking,Flowers & Gifts,Active Life |
BusinessAcceptsCreditCards,GoodForKids
 |       1 |
                 | 0Ni7Stqt4RFWDGjOYRi2Bw | Scent From Above Company          | 2501 W Behrend
Dr, Ste 67   | Scottsdale    | AZ   | 85027      |    33.6656 |  -112.111 |        14 |
4.5 | 6:00-16:00    | 6:00-16:00    | 6:00-16:00      | 6:00-16:00      | 6:00-16:00    |
None          | None          | Home Cleaning,Local Services,Professional Services,Carpet
Cleaning,Home Services,Office Cleaning,Window Washing
 | BusinessAcceptsCreditCards,ByAppointmentOnly
 |       1 |
                 | 0WBMEfqXQnEOAIkV-uCW6w | The Charlotte Room                | 19 Charlotte
Street       | Toronto       | ON   | M5V 2H5    |    43.6466 |  -79.3938 |        10
 |   3.5 | 15:00-1:00    | 15:00-1:00    | 15:00-1:00      | 15:00-1:00      | 15:00-2:00    |
18:00-2:00       | None          | Event Planning & Services,Bars,Nightlife,Lounges,Pool
Halls,Venues & Event Spaces
 |
BusinessParking,HasTV,CoatCheck,NoiseLevel,OutdoorSeating,BusinessAcceptsCreditCards,Resta
urantsPriceRange2,Music,WheelchairAccessible,Smoking,Ambience,BestNights,RestaurantsGoodFo
rGroups,HappyHour,GoodForDancing,Alcohol
 |       0 |
                 | 0Y3lHyqRHfWOBuQlS1bM0g | PC Savants                        | 11966 W
Candelaria Ct    | Sun City      | AZ   | 85373      |    33.6901 |  -112.319 |
```

```
11 |    5.0 | 10:00-19:00  | 10:00-19:00  | 10:00-19:00   | 10:00-19:00   | 10:00-19:00
| 11:00-18:00   | 11:00-18:00  | IT Services & Computer Repair,Electronics Repair,Local
Services,Mobile Phone Repair
| BusinessAcceptsCreditCards,BusinessAcceptsBitcoin
|     1 |
                    | 0aKsGxx7XP2TMs_fn_9xVw | Sweet Ruby Jane Confections   | 8975 S Eastern
Ave, Ste 3-B | Las Vegas    | NV   | 89123     |     36.015 | -115.118 |         30 |
4.0 | 10:00-19:00  | 10:00-19:00  | 10:00-19:00   | 10:00-19:00   | 10:00-19:00  |
10:00-19:00   | None       | Food,Chocolatiers & Shops,Bakeries,Specialty Food,Desserts
| BusinessAcceptsCreditCards,RestaurantsPriceRange2,BusinessParking,WheelchairAccessible
|     0 |
                    | 0cxO1Lx2Pi7u6ftWX3Wksg | Oinky's Pork Chop Heaven    | 22483 Emery Rd
| North Randall | OH  | 44128      |    41.4352 | -81.5214 |         3 |   3.0 |
6:00-23:00   | 6:00-23:00   | 6:00-23:00    | 6:00-23:00    | 6:00-23:00    | 6:00-
23:00     | 6:00-23:00   | Soul Food,Restaurants
|
RestaurantsAttire,RestaurantsGoodForGroups,GoodForKids,RestaurantsReservations,Restaurants
TakeOut
|     1 |
                    | 0e-j5VcEn54EZT-FKCUZdw | Sushi Osaka               | 5084 Dundas
Street W     | Toronto     | ON   | M9A 1C2    | 43.6452 | -79.5324 |
8 |   4.5 | 11:00-23:00  | 11:00-23:00  | 11:00-23:00   | 11:00-23:00   | 11:00-23:00
| 11:00-23:00   | 14:00-23:00  | Sushi Bars,Restaurants,Japanese,Korean
| RestaurantsTakeOut,WiFi,RestaurantsGoodForGroups,RestaurantsReservations
|     1 |
                +------------------------+----------------------------+--------------
---------------+--------------+-------+-------------+----------+-----------+-------------
+-------+--------------+--------------+---------------+---------------+-------------
+--------------+--------------+-------------------------------------------------------
-------------------------------------------------------------------------------------
-----------------------------------------------------+---------------------------------
-------------------------------------------------------------------------------------
-------------------------------------------------------------------------------------
-------------------------------------------------------------------------------------
----------------------+--------+
```

iv.   Provide the SQL code you used to create your final dataset:

```
SELECT B.id,
        B.name,
        B.address,
        B.city,
        B.state,
        B.postal_code,
        B.latitude,
        B.longitude,
        B.review_count,
        B.stars,
        MAX(CASE
        WHEN H.hours LIKE "%monday%" THEN
TRIM(H.hours,'%MondayTuesWednesThursFriSatSun|%')
        END) AS monday_hours,
        MAX(CASE
        WHEN H.hours LIKE "%tuesday%" THEN
TRIM(H.hours,'%MondayTuesWednesThursFriSatSun|%')
        END) AS tuesday_hours,
        MAX(CASE
        WHEN H.hours LIKE "%wednesday%" THEN
TRIM(H.hours,'%MondayTuesWednesThursFriSatSun|%')
        END) AS wednesday_hours,
        MAX(CASE
        WHEN H.hours LIKE "%thursday%" THEN
TRIM(H.hours,'%MondayTuesWednesThursFriSatSun|%')
        END) AS thursday_hours,
        MAX(CASE
        WHEN H.hours LIKE "%friday%" THEN
TRIM(H.hours,'%MondayTuesWednesThursFriSatSun|%')
        END) AS friday_hours,
        MAX(CASE
```

```
                              WHEN H.hours LIKE "%saturday%" THEN
 TRIM(H.hours,'%MondayTuesWednesThursFriSatSun|%')
                              END) AS saturday_hours,
                              MAX(CASE
                              WHEN H.hours LIKE "%sunday%" THEN
 TRIM(H.hours,'%MondayTuesWednesThursFriSatSun|%')
                              END) AS sunday_hours,
                              GROUP_CONCAT(DISTINCT(C.category)) AS categories,
                              GROUP_CONCAT(DISTINCT(A.name)) AS attributes,
                              B.is_open
                    FROM business B
                    INNER JOIN hours H
                    ON B.id = H.business_id
                    INNER JOIN category C
                    ON B.id = C.business_id
                    INNER JOIN attribute A
                    ON B.id = A.business_id
                    GROUP BY B.id
```