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:

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.

i. Business = Primary Key 10000
ii. Hours = Foreign key Business_id : 1562
iii. Category = Foreign Key Business_id : 2643
iv. Attribute = Foreign Key Business_id : 1115
v. Review = Primary Key 10000 , Foreign Key Business_id: 8090, Foreign Key User_id :9581
vi. Checkin = Foreign Key Business_id: 493
vii. Photo = Primary Key 10000, Foriegn Key Business_id:6493
viii. Tip = Foreign Key User_id: 537, Foreign Key Business_id:3979
ix. User = Primary Key 10000

x. Friend = Foriegn Key User_id: 11
xi. Elite_years = Foriegn Key User_id:2780

Note: Primary Keys are denoted in the ER-Diagram with a yellow key icon.
Foreign keys are denoted as Red Diamonds Thanks for heads up :X

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:

i. Table: Review, Column: Stars

```
min: 1          max: 5          avg:3.7082
```

ii. Table: Business, Column: Stars

```
min: 1          max: 5  avg: 3.6549
```

iii. Table: Tip, Column: Likes

```
min: 0          max: 2          avg: 3.6549
```

```
        iv. Table: Checkin, Column: Count

                min:0           max:1           avg:0.01444


        v. Table: User, Column: Review_count

                min: 1          max:53          avg: 1.9414
```

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

```
        SQL code used to arrive at answer:

SELECT city,  SUM(review_count) AS 'TOTAL_REVIEWS'
        FROM business
        Group BY city
        ORDER BY SUM(review_count) DESC;

        Copy and Paste the Result Below:
         +-----------------+---------------+
| city            | TOTAL_REVIEWS |
+-----------------+---------------+
| Las Vegas       |         82854 |
| Phoenix         |         34503 |
| Toronto         |         24113 |
| Scottsdale      |         20614 |
| Charlotte       |         12523 |
| Henderson       |         10871 |
| Tempe           |         10504 |
| Pittsburgh      |          9798 |
| MontrÃ©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 |
+-----------------+---------------+
(Output limit exceeded, 25 of 362 total rows shown)
```

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 as 'Star_Rating',
SUM(review_count) as 'COUNT'
FROM business
WHERE city == 'Avon'
Group by stars;
```

Copy and Paste the Resulting Table Below (2 columns â€“ star rating and count):

```
+-------------+-------+
| Star_Rating | 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 as 'Star_Rating',
SUM(review_count) as 'COUNT'
FROM business
WHERE city == 'Beachwood'
Group By stars;
```

Copy and Paste the Resulting Table Below (2 columns â€“ star rating and count):

```
+-------------+-------+
| Star_Rating | 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 AS 'ID' , name as 'NAME' , review_count as '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?

        Please explain your findings and interpretation of the results:

By finding the User with the largest quanity of reviews, it was found that Gerald has 2000 reviews but only 253 fans.
The user with the most amount of fans is Amy, she has 503 fans with only 609 reviews.
It seems how long a user has been yelping for is also a factor to the amount of fans they have.
This seems pretty trivial, but to answer this question I am inconclusive about how strong of a correlation there is
between reviewcount and fans, along with yelping_since and fans. I am curiouse what the Correlation Coefficiant would be.

```
+-----------+--------------+------+---------------------+
| name      | review_count | fans | yelping_since       |
+-----------+--------------+------+---------------------+
| Amy       |          609 |  503 | 2007-07-19 00:00:00 |
| Mimi      |          968 |  497 | 2011-03-30 00:00:00 |
| Harald    |         1153 |  311 | 2012-11-27 00:00:00 |
| Gerald    |         2000 |  253 | 2012-12-16 00:00:00 |
| Christine |          930 |  173 | 2009-07-08 00:00:00 |
| Lisa      |          813 |  159 | 2009-10-05 00:00:00 |
| Cat       |          377 |  133 | 2009-02-05 00:00:00 |
| William   |         1215 |  126 | 2015-02-19 00:00:00 |
| Fran      |          862 |  124 | 2012-04-05 00:00:00 |
| Lissa     |          834 |  120 | 2007-08-14 00:00:00 |
| Mark      |          861 |  115 | 2009-05-31 00:00:00 |
| Tiffany   |          408 |  111 | 2008-10-28 00:00:00 |
| bernice   |          255 |  105 | 2007-08-29 00:00:00 |
| Roanna    |         1039 |  104 | 2006-03-28 00:00:00 |
| Angela    |          694 |  101 | 2010-10-01 00:00:00 |
| .Hon      |         1246 |  101 | 2006-07-19 00:00:00 |
| Ben       |          307 |   96 | 2007-03-10 00:00:00 |
| Linda     |          584 |   89 | 2005-08-07 00:00:00 |
| Christina |          842 |   85 | 2012-10-08 00:00:00 |
| Jessica   |          220 |   84 | 2009-01-12 00:00:00 |
| Greg      |          408 |   81 | 2008-02-16 00:00:00 |
| Nieves    |          178 |   80 | 2013-07-08 00:00:00 |
| Sui       |          754 |   78 | 2009-09-07 00:00:00 |
| Yuri      |         1339 |   76 | 2008-01-03 00:00:00 |
| 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: There are more reviews with the word love.

SQL code used to arrive at answer:

```
//Ran two Seperate Queries to get this result, tried union but it combined love/hate
collumns into one collumn//

SELECT COUNT(id) AS 'Number of reviews containing the word hate'
FROM review
WHERE text like '%hate%';
---->232
//A union between these two queries does not work properly if anyone knows why let me
know//

SELECT COUNT(id) AS 'Nubmer of reviews containing the word love'
FROM review
WHERE text like '%love%';
---->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 |
| -1h59ko3dxChBSZ9U7LfUw | Lissa     |  120 |
+------------------------+-----------+------+
```

11. Is there a strong relationship (or correlation) between having a high number of fans and being listed as "useful" or "funny?" Out of the top 10 users with the highest number of fans, what percent are also listed as â€œusefulâ€ or â€œfunnyâ€?

Key:
0% - 25% - Low relationship
26% - 75% - Medium relationship
76% - 100% - Strong relationship

SQL code used to arrive at answer:

```
SELECT name,
fans, useful,funny,review_count,yelping_since
```

```
FROM user
ORDER BY fans DESC

        Copy and Paste the Result Below:
                +-----------+------+--------+--------+--------------+---------------------
+
                | name      | fans | useful |  funny | review_count | yelping_since
|
                +-----------+------+--------+--------+--------------+---------------------
+
                | Amy       |  503 |   3226 |   2554 |          609 | 2007-07-19 00:00:00
|
                | Mimi      |  497 |    257 |    138 |          968 | 2011-03-30 00:00:00
|
                | Harald    |  311 | 122921 | 122419 |         1153 | 2012-11-27 00:00:00
|
                | Gerald    |  253 |  17524 |   2324 |         2000 | 2012-12-16 00:00:00
|
                | Christine |  173 |   4834 |   6646 |          930 | 2009-07-08 00:00:00
|
                | Lisa      |  159 |     48 |     13 |          813 | 2009-10-05 00:00:00
|
                | Cat       |  133 |   1062 |    672 |          377 | 2009-02-05 00:00:00
|
                | William   |  126 |   9363 |   9361 |         1215 | 2015-02-19 00:00:00
|
                | Fran      |  124 |   9851 |   7606 |          862 | 2012-04-05 00:00:00
|
                | Lissa     |  120 |    455 |    150 |          834 | 2007-08-14 00:00:00
|
                | Mark      |  115 |   4008 |    570 |          861 | 2009-05-31 00:00:00
|
                | Tiffany   |  111 |   1366 |    984 |          408 | 2008-10-28 00:00:00
|
                | bernice   |  105 |    120 |    112 |          255 | 2007-08-29 00:00:00
|
                | Roanna    |  104 |   2995 |   1188 |         1039 | 2006-03-28 00:00:00
|
                | Angela    |  101 |    158 |    164 |          694 | 2010-10-01 00:00:00
|
                | .Hon      |  101 |   7850 |   5851 |         1246 | 2006-07-19 00:00:00
|
                | Ben       |   96 |   1180 |   1155 |          307 | 2007-03-10 00:00:00
|
                | Linda     |   89 |   3177 |   2736 |          584 | 2005-08-07 00:00:00
|
                | Christina |   85 |    158 |     34 |          842 | 2012-10-08 00:00:00
|
                | Jessica   |   84 |   2161 |   2091 |          220 | 2009-01-12 00:00:00
|
                | Greg      |   81 |    820 |    753 |          408 | 2008-02-16 00:00:00
|
                | Nieves    |   80 |   1091 |    774 |          178 | 2013-07-08 00:00:00
|
                | Sui       |   78 |      9 |     18 |          754 | 2009-09-07 00:00:00
|
                | Yuri      |   76 |   1166 |    220 |         1339 | 2008-01-03 00:00:00
```

```
|
                | Nicole     |   73 |     13 |     10 |          161 | 2009-04-30 00:00:00
|
                +-----------+------+--------+--------+--------------+---------------------
+
```

        Please explain your findings and interpretation of the results:
I honestly hate this question and this whole assignment; The questions are very vague.
It intuitivally seems like the more people who find a user funny or useful are fans of that yelp reviewer.
Clearly Harald is an outlier to this logic, I dont know what type of mathematics they want us to do here
but its very tedious to calculate Correlation Coeficient or even Variance using SQL Code.
In Conclusion this assignment blows.!

Part 2: Inferences and Analysis

1. Pick one city and category of your choice and group the businesses in that city or category by their overall star rating. Compare the businesses with 2-3 stars to the businesses with 4-5 stars and answer the following questions. Include your code.

//All of the stores in my query were from 3.5-4.5//

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

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

iii. Are you able to infer anything from the location data provided between these two groups? Explain.
No there is only 3 Tobacco Shops in the City of Beachwood, two being in Tempe and one being in Charlotee.
A sample of 3 objects is not enough to infer anything.

SQL code used for analysis:

```
SELECT b.stars , b.city, b.neighborhood, b.is_open,b.review_count ,
c.category,
h.hours
FROM business b
LEFT JOIN category c on b.id=c.business_id
LEFT JOIN hours h on b.id=h.business_id
WHERE category='Tobacco Shops'
GROUP BY STARS;
```

```
+--------+-----------+-----------------+---------+--------------+---------------+----------
------------+
| stars | city      | neighborhood    | is_open | review_count | category      | hours
|
+--------+-----------+-----------------+---------+--------------+---------------+----------
------------+
|   3.5 | Tempe     |                  |       1 |            3 | Tobacco Shops |
Saturday|9:30-22:00  |
|   4.0 | Charlotte | University City |       1 |            5 | Tobacco Shops |
Saturday|12:00-22:00 |
|   4.5 | Tempe     |                  |       0 |           11 | Tobacco Shops | None
|
+-------+-----------+-----------------+---------+--------------+---------------+----------
------------+
```

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 still open have 234039 more reviews than the ones of out business.

ii. Difference 2:
 1427/8480 = .16 that is 16% of the businesses that are open have 5 stars.
 14/138 = .10 that is 10% of the businesses that are closed had 5 stars.
 This shows us that even though your business has 5 stars it still can close down.

SQL code used for analysis:

```
//Difference 1
SELECT sum(review_count) FROM business
WHERE is_open=0;
---> 35261
SELECT sum(review_count) FROM business
WHERE is_open=1;
---> 269300 reviews


//Difference 2
SELECT Count(*) AS 'Number of Closed business with their rating in stars', stars  FROM
business
WHERE is_open=0
Group by stars;

+-------------------------------------------------------+-------+
| Number of Closed business with their rating in stars | stars |
+-------------------------------------------------------+-------+
|                                                    14 |   1.0 |
|                                                    24 |   1.5 |
|                                                    94 |   2.0 |
|                                                   168 |   2.5 |
|                                                   272 |   3.0 |
|                                                   295 |   3.5 |
|                                                   326 |   4.0 |
|                                                   189 |   4.5 |
|                                                   138 |   5.0 |
+-------------------------------------------------------+-------+
                                             SUM: 1514
SELECT Count(*) AS 'Number of Open business with their rating in stars', stars  FROM
business
WHERE is_open=1
Group by stars;

+-----------------------------------------------------+-------+
| Number of Open business with their rating in stars | stars |
+-----------------------------------------------------+-------+
|                                                 142 |   1.0 |
|                                                 182 |   1.5 |
|                                                 472 |   2.0 |
|                                                 722 |   2.5 |
|                                                1124 |   3.0 |
|                                                1483 |   3.5 |
|                                                1679 |   4.0 |
|                                                1249 |   4.5 |
|                                                1427 |   5.0 |
+-----------------------------------------------------+-------+
                                               SUM: 8480
```

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:

I will not forcast anything but will, Help those whose are interested in any automotive business decide when/where to open to be successful.

ii. Write 1-2 brief paragraphs on the type of data you will need for your analysis and why you chose that data:

I will need to utilize data from the business, hours and category tables.
From the business table we will focus on the adress, city, state, and postal code because these are neccesary for postal address.
From the hours table we will find out the hours of operations for each day of the week to give inference on the hours neccessary to stay in business.
We will use the category table to ensure we only query businesses in the Automotive, oil change stations, car wash or auto detailing category.

iii. Output of your finished dataset:

```
 +-------------------------------+----------------------------------------------+-------+-
--------+---------------------+-------------+--------------+-----------------+----------
------+--------------+----------------+--------------+
| name                          | Postal Adress                                 | stars |
is_open | category             | Monday_hours | Tuesday_hours | Wednesday_hours |
Thursday_hours | Friday_hours | Saturday_hours | Sunday_hours |
+-------------------------------+-----------------------------------------------+-------+--
-------+---------------------+--------------+---------------+-----------------+-----------
-----+--------------+----------------+---------------+
| Freeman's Car Stereo          | 4821 South Blvd Charlotte NC 28217            |   3.5 |
1 | Automotive          | 9:00-19:00   |           None |             None |           None
|         None |           None |         None |
| Freeman's Car Stereo          | 4821 South Blvd Charlotte NC 28217            |   3.5 |
1 | Automotive          | None         |     9:00-19:00 |             None |           None
|         None |           None |         None |
| Freeman's Car Stereo          | 4821 South Blvd Charlotte NC 28217            |   3.5 |
1 | Automotive          | None         |           None |             None |           None
|   9:00-19:00 |           None |         None |
| Freeman's Car Stereo          | 4821 South Blvd Charlotte NC 28217            |   3.5 |
1 | Automotive          | None         |           None |       9:00-19:00 |           None
|         None |           None |         None |
| Freeman's Car Stereo          | 4821 South Blvd Charlotte NC 28217            |   3.5 |
1 | Automotive          | None         |           None |             None |     9:00-19:00
|         None |           None |         None |
| Freeman's Car Stereo          | 4821 South Blvd Charlotte NC 28217            |   3.5 |
1 | Automotive          | None         |           None |             None |           None
|         None |     9:00-17:00 |         None |
| Christian Brothers Automotive | 290 E Ocotillo Rd Chandler AZ 85249           |   5.0 |
1 | Automotive          | None         |           None |             None |           None
|   7:00-18:00 |           None |         None |
| Christian Brothers Automotive | 290 E Ocotillo Rd Chandler AZ 85249           |   5.0 |
1 | Automotive          | None         |     7:00-18:00 |             None |           None
|         None |           None |         None |
| Christian Brothers Automotive | 290 E Ocotillo Rd Chandler AZ 85249           |   5.0 |
1 | Automotive          | None         |           None |             None |     7:00-18:00
|         None |           None |         None |
| Christian Brothers Automotive | 290 E Ocotillo Rd Chandler AZ 85249           |   5.0 |
1 | Automotive          | None         |           None |       7:00-18:00 |           None
|         None |           None |         None |
| Christian Brothers Automotive | 290 E Ocotillo Rd Chandler AZ 85249           |   5.0 |
1 | Automotive          | 7:00-18:00   |           None |             None |           None
|         None |           None |         None |
| Christian Brothers Automotive | 290 E Ocotillo Rd Chandler AZ 85249           |   5.0 |
1 | Oil Change Stations | None         |           None |             None |           None
|   7:00-18:00 |           None |         None |
| Christian Brothers Automotive | 290 E Ocotillo Rd Chandler AZ 85249           |   5.0 |
1 | Oil Change Stations | None         |     7:00-18:00 |             None |           None
|         None |           None |         None |
| Christian Brothers Automotive | 290 E Ocotillo Rd Chandler AZ 85249           |   5.0 |
1 | Oil Change Stations | None         |           None |             None |     7:00-18:00
|         None |           None |         None |
| Christian Brothers Automotive | 290 E Ocotillo Rd Chandler AZ 85249           |   5.0 |
1 | Oil Change Stations | None         |           None |       7:00-18:00 |           None
|         None |           None |         None |
| Christian Brothers Automotive | 290 E Ocotillo Rd Chandler AZ 85249           |   5.0 |
1 | Oil Change Stations | 7:00-18:00   |           None |             None |           None
|         None |           None |         None |
| Buddy's Muffler & Exhaust     | 1509 Hickory Grove Rd Gastonia NC 28056       |   5.0 |
1 | Automotive          | 8:30-17:00   |           None |             None |           None
|         None |           None |         None |
| Buddy's Muffler & Exhaust     | 1509 Hickory Grove Rd Gastonia NC 28056       |   5.0 |
1 | Automotive          | None         |     8:30-17:00 |             None |           None
```

```
|         None |           None |         None |
| Buddy's Muffler & Exhaust     | 1509 Hickory Grove Rd Gastonia NC 28056      |   5.0 |
1 | Automotive          | None         |          None |            None |           None
|   8:30-17:00 |           None |         None |
| Buddy's Muffler & Exhaust     | 1509 Hickory Grove Rd Gastonia NC 28056      |   5.0 |
1 | Automotive          | None         |          None |      8:30-17:00 |           None
|         None |           None |         None |
| Buddy's Muffler & Exhaust     | 1509 Hickory Grove Rd Gastonia NC 28056      |   5.0 |
1 | Automotive          | None         |          None |            None |     8:30-17:00
|         None |           None |         None |
| Buddy's Muffler & Exhaust     | 1509 Hickory Grove Rd Gastonia NC 28056      |   5.0 |
1 | Automotive          | None         |          None |            None |           None
|         None |     9:00-15:00 |         None |
| All Storage - Anthem          | 2620 W Horizon Ridge Pkwy Henderson NV 89052 |   3.5 |
1 | Automotive          | 9:00-16:30   |          None |            None |           None
|         None |           None |         None |
| All Storage - Anthem          | 2620 W Horizon Ridge Pkwy Henderson NV 89052 |   3.5 |
1 | Automotive          | None         |    9:00-16:30 |            None |           None
|         None |           None |         None |
| All Storage - Anthem          | 2620 W Horizon Ridge Pkwy Henderson NV 89052 |   3.5 |
1 | Automotive          | None         |          None |            None |           None
|   9:00-16:30 |           None |         None |
+-------------------------------+----------------------------------------------+-------+--
-------+---------------------+--------------+---------------+-----------------+-----------
-----+--------------+----------------+--------------+
(Output limit exceeded, 25 of 66 total rows shown)
```

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

```
SELECT
DISTINCT(b.name),
(b.address||' ' || b.city||' ' ||  b.state||' ' || b.postal_code) AS 'Postal Adress' ,
b.stars , b.is_open,
c.category,
(CASE WHEN h.hours LIKE "%monday%" THEN TRIM(h.hours,'%MondayTuesWednesThursFriSatSun|%')
                           END) AS Monday_hours,
  CASE WHEN h.hours LIKE "%tuesday%" THEN TRIM
(h.hours,'%MondayTuesWednesThursFriSatSun|%')
         END AS Tuesday_hours,
  CASE WHEN h.hours LIKE "%wednesday%" THEN TRIM
(h.hours,'%MondayTuesWednesThursFriSatSun|%')
         END AS Wednesday_hours,
  CASE WHEN h.hours LIKE "%thursday%" THEN TRIM
(h.hours,'%MondayTuesWednesThursFriSatSun|%')
         END AS Thursday_hours,
  CASE WHEN h.hours LIKE "%friday%" THEN TRIM
(h.hours,'%MondayTuesWednesThursFriSatSun|%')
         END AS Friday_hours,
  CASE WHEN h.hours LIKE "%saturday%" THEN TRIM
(h.hours,'%MondayTuesWednesThursFriSatSun|%')
         END AS Saturday_hours,
  CASE WHEN h.hours LIKE "%sunday%" THEN TRIM
(h.hours,'%MondayTuesWednesThursFriSatSun|%')
         END AS Sunday_hours
FROM business b
LEFT JOIN  category c ON b.id=c.business_id
LEFT JOIN hours h ON b.id = h.business_id
WHERE category IN ('Automotive','Oil Change Stations', 'Car Wash', 'Auto Detailing')
AND b.is_open=1
;
```