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 =10,000
ii. Business table =10,000
iii. Category table =10,000
iv. Checkin table = 10,000
v. elite_years table =10,000
vi. friend table = 10,000
vii. hours table =10,000
viii. photo table =10,000
ix. review table = 10,000
x. tip table = 10,000
xi. user table =10,000

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 =10,000
ii. Hours = 1562
iii. Category =2643
iv. Attribute = 1115
v. Review = 8090
vi. Checkin = 493
vii. Photo = 6493
viii. Tip = 3979
ix. User = 10,000
x. Friend = 11
xi. Elite_years = 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:

I have checked every column_name to verify that no record is NULL

```
SELECT *
FROM user
WHERE column_name 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:1.0   max:5.0   avg:3.6549

        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 Total_review
FROM business
GROUP BY city
ORDER BY Total_review DESC;
```

        Copy and Paste the Result Below:

```
+-----------------+--------------+
| city            | Total_review |
+-----------------+--------------+
| Las Vegas       |        82854 |
| Phoenix         |        34503 |
| Toronto         |        24113 |
| Scottsdale      |        20614 |
| Charlotte       |        12523 |
| Henderson       |        10871 |
| Tempe           |        10504 |
| Pittsburgh      |         9798 |
| Montréal        |         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, COUNT(stars)
FROM business
WHERE city="Avon"
GROUP BY stars;
```

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

```
+-------+--------------+
| stars | COUNT(stars) |
+-------+--------------+
|   1.5 |            1 |
|   2.5 |            2 |
|   3.5 |            3 |
|   4.0 |            2 |
|   4.5 |            1 |
|   5.0 |            1 |
+-------+--------------+
```

ii. Beachwood

SQL code used to arrive at answer:

```
SELECT stars, COUNT(stars)
FROM business
WHERE city="Beachwood"
GROUP BY stars;
```

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

```
+-------+--------------+
| stars | COUNT(stars) |
+-------+--------------+
|   2.0 |            1 |
|   2.5 |            1 |
|   3.0 |            2 |
|   3.5 |            2 |
|   4.0 |            1 |
|   4.5 |            2 |
|   5.0 |            5 |
+-------+--------------+
```

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

SQL code used to arrive at answer:

```
SELECT name,COUNT(review_count) AS Total_review
FROM user
GROUP BY name
ORDER BY Total_review DESC
LIMIT 3;
```

Copy and Paste the Result Below:

```
+-------+--------------+
| name  | Total_review |
+-------+--------------+
| John  |          102 |
| David |           90 |
| Chris |           74 |
+-------+--------------+
```

8. Does posing more reviews correlate with more fans?

Please explain your findings and interpretation of the results:

ANS:

NO, posing more reviews have no correlation with more fans. AS seen from below table as reviews decreases there is no pattern increase/decrease in number of fans

```
+----------+--------------+------------+
| name     | Total_review | Total_fans |
+----------+--------------+------------+
| John     |          102 |         46 |
| David    |           90 |         25 |
| Chris    |           74 |         52 |
| Mike     |           74 |        119 |
| Michael  |           72 |         34 |
| Jennifer |           63 |         86 |
| Mark     |           59 |        156 |
| Lisa     |           58 |        207 |
| Melissa  |           58 |        104 |
| Sarah    |           55 |        100 |
| Alex     |           54 |         22 |
| James    |           48 |         86 |
| Jessica  |           45 |        116 |
| Ryan     |           45 |         24 |
| J        |           43 |         13 |
| Michelle |           43 |        133 |
| Andrew   |           41 |        114 |
| Kevin    |           41 |         20 |
| Mary     |           41 |         18 |
| Amanda   |           40 |         26 |
| Ashley   |           40 |         16 |
| Brian    |           40 |         72 |
| Karen    |           40 |        123 |
| Laura    |           39 |         38 |
| Robert   |           39 |          9 |
+----------+--------------+------------+
(Output limit exceeded, 25 of 3454 total rows shown)
```

SQL CODE

```
SELECT name,
COUNT(review_count) AS Total_review,
SUM(fans) AS Total_fans
FROM user
GROUP BY name
ORDER BY Total_review DESC
```

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

Answer: There are more reviews with word “love” than word “hate”

```
+------+------+
| Love | Hate |
+------+------+
| 1780 |  232 |
+------+------+
```

SQL code used to arrive at answer:

```
SELECT DISTINCT ( SELECT COUNT(*)
             FROM review
             WHERE text LIKE '%love%') AS Love,
       (SELECT COUNT(*)
             FROM review
             WHERE text LIKE '%hate%') AS Hate
FROM review ;
```

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

SQL code used to arrive at answer:

```
SELECT name, fans
FROM user
ORDER BY fans DESC
LIMIT 10;
```

Copy and Paste the Result Below:

```
+-----------+------+
| name      | fans |
+-----------+------+
| Amy       |  503 |
| Mimi      |  497 |
| Harald    |  311 |
| Gerald    |  253 |
| Christine |  173 |
| Lisa      |  159 |
| Cat       |  133 |
| William   |  126 |
| Fran      |  124 |
| 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 “useful” or “funny”?

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

Copy and Paste the Result Below:

```
+-----------+------+--------+--------+
| name      | fans | useful |  funny |
+-----------+------+--------+--------+
| Amy       |  503 |   3226 |   2554 |
| Mimi      |  497 |    257 |    138 |
| Harald    |  311 | 122921 | 122419 |
| Gerald    |  253 |  17524 |   2324 |
| Christine |  173 |   4834 |   6646 |
| Lisa      |  159 |     48 |     13 |
| Cat       |  133 |   1062 |    672 |
| William   |  126 |   9363 |   9361 |
| Fran      |  124 |   9851 |   7606 |
| Lissa     |  120 |    455 |    150 |
+-----------+------+--------+--------+
```

Please explain your findings and interpretation of the results:

- There is a low relationship between fans and useful or funny.
- 100% of the top 10 user having highest number of fans are useful as well as funny.

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.

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

For group having stars 2-3 and in city Las vegas, 7 out of 14 have hours 11:00-0:00 and rest have 8:00-22:00

```
+-----------+-----------+----------------------+
| name      | city      | hours                |
+-----------+-----------+----------------------+
| Wingstop  | Las Vegas | Monday|11:00-0:00    |
| Wingstop  | Las Vegas | Tuesday|11:00-0:00   |
| Wingstop  | Las Vegas | Friday|11:00-0:00    |
| Wingstop  | Las Vegas | Wednesday|11:00-0:00 |
| Wingstop  | Las Vegas | Thursday|11:00-0:00  |
| Wingstop  | Las Vegas | Sunday|11:00-0:00    |
| Wingstop  | Las Vegas | Saturday|11:00-0:00  |
| Walgreens | Las Vegas | Monday|8:00-22:00    |
| Walgreens | Las Vegas | Tuesday|8:00-22:00   |
| Walgreens | Las Vegas | Friday|8:00-22:00    |
| Walgreens | Las Vegas | Wednesday|8:00-22:00 |
| Walgreens | Las Vegas | Thursday|8:00-22:00  |
| Walgreens | Las Vegas | Sunday|8:00-22:00    |
| Walgreens | Las Vegas | Saturday|8:00-22:00  |
+-----------+-----------+----------------------+
```

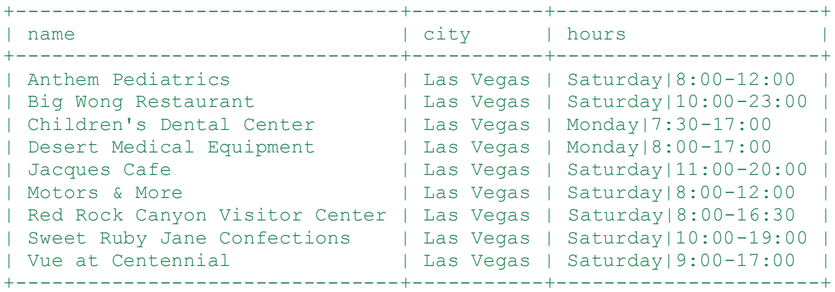
SQL CODE

```
SELECT b.name,b.city,c.hours
FROM business b INNER JOIN hours c
ON b.id=c.business_id
WHERE b.city='Las Vegas'AND
(b.stars BETWEEN 2 AND 3)
```

For the group having 4-5 stars distribution of hours is extremely different from the other group

SQL CODE

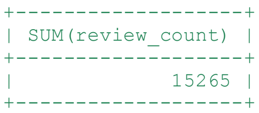
```
SELECT b.name,b.city,c.hours
FROM business b INNER JOIN hours c
ON b.id=c.business_id
WHERE b.city='Las Vegas'AND
(b.stars BETWEEN 4 AND 5)
GROUP BY b.name
```



```
+-------------------------------+-----------+----------------------+
| name                          | city      | hours                |
+-------------------------------+-----------+----------------------+
| Anthem Pediatrics             | Las Vegas | Saturday|8:00-12:00  |
| Big Wong Restaurant           | Las Vegas | Saturday|10:00-23:00 |
| Children's Dental Center      | Las Vegas | Monday|7:30-17:00    |
| Desert Medical Equipment      | Las Vegas | Monday|8:00-17:00    |
| Jacques Cafe                  | Las Vegas | Saturday|11:00-20:00 |
| Motors & More                 | Las Vegas | Saturday|8:00-12:00  |
| Red Rock Canyon Visitor Center | Las Vegas | Saturday|8:00-16:30  |
| Sweet Ruby Jane Confections   | Las Vegas | Saturday|10:00-19:00 |
| Vue at Centennial             | Las Vegas | Saturday|9:00-17:00  |
+-------------------------------+-----------+----------------------+
```

ii. Do the two groups you chose to analyze have a different number of reviews?
Ans: Yes, both groups have different sum of reviews

For 2-3 star group

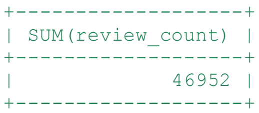


```
+-------------------+
| SUM(review_count) |
+-------------------+
|             15265 |
+-------------------+
```

SQL CODE

```
SELECT SUM(review_count)
FROM business
WHERE city='Las Vegas'AND
(stars BETWEEN 2 AND 3)
```

For 4-5 star group



```
+-------------------+
| SUM(review_count) |
+-------------------+
|             46952 |
+-------------------+
```

SQL CODE

```
SELECT SUM(review_count)
FROM business
WHERE city='Las Vegas'AND
(stars BETWEEN 4 AND 5)
```

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

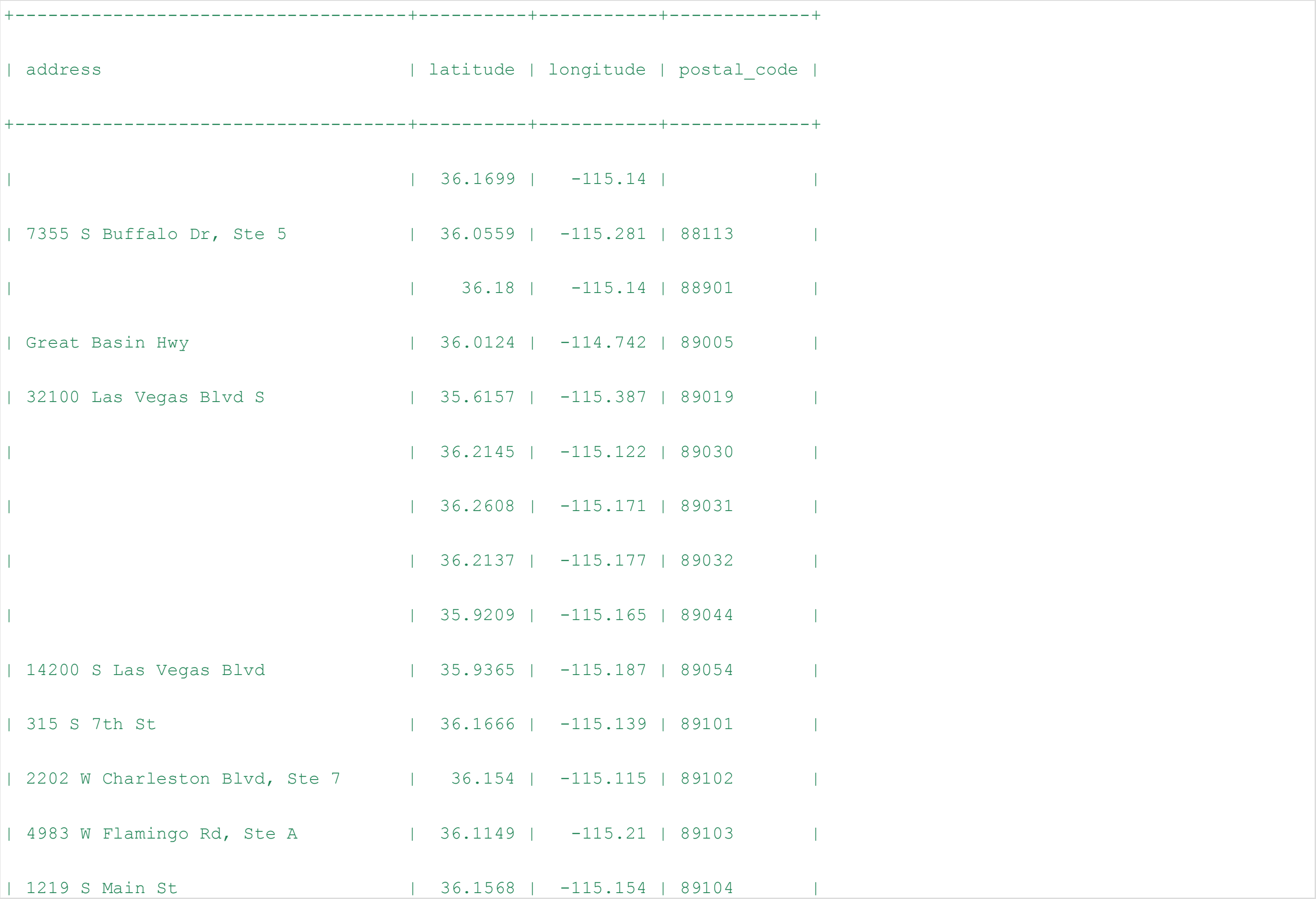
Ans:

- Most of lower stars business are in address of **** las vegas Blvd S and having postal code of 89109
  While for higher rating business most are in postal code 890** of **** las vegas Blvd S.
- Both low rating stars and high stars cities are clustered at around (36.1*,115.1*). So there is possibility of a market area in this region.

```
+---------------------------------------------------------+----------+-----------+-------------+
| address                                                 | latitude | longitude | postal_code |
+---------------------------------------------------------+----------+-----------+-------------+
| 3645 Las Vegas Blvd S                                   |  36.1143 |  -115.171 | 89109       |
| 3355 Las Vegas Blvd S                                   |  36.1221 |  -115.168 | 89162       |
| 8335 Las Vegas Blvd S                                   |   36.038 |  -115.173 | 89123       |
| New York New York Hotel & Casino, 3790 Las Vegas Blvd S |   36.103 |  -115.174 | 89109       |
| 915 S Rainbow Blvd                                      |  36.1611 |  -115.245 | 89145       |
| 6630 N Durango Dr, Ste 180                              |  36.2813 |  -115.287 | 89149       |
| 1109 Western Ave                                        |  36.1584 |  -115.159 | 89102       |
| 3200 S Las Vegas Blvd                                   |  36.1275 |  -115.172 | 89109       |
| 860 East Twain, Ste 102                                 |  36.1193 |  -115.146 | 89169       |
| 3300 S Las Vegas Blvd                                   |  36.1245 |  -115.172 | 89109       |
| 7175 Spring Mountain Rd                                 |  36.1242 |  -115.248 | 89117       |
| 3993 Spring Mountain Rd                                 |  36.1264 |  -115.193 | 89102       |
| 6850 N Durango Dr, Ste 310                              |  36.2858 |  -115.285 | 89149       |
| 410 S Rampart Blvd, Ste 330                             |  36.1672 |  -115.286 | 89145       |
| 3700 W Flamingo Rd                                      |  36.1179 |  -115.187 | 89103       |
| 1251 S Maryland Pkwy                                    |  36.1563 |  -115.137 | 89104       |
| 4255 Spring Mountain Rd                                 |  36.1264 |  -115.198 | 89102       |
| 3000 Paradise Rd                                        |  36.1363 |  -115.151 | 89109       |
| 4055 Palos Verdes St.                                   |  36.1156 |  -115.151 | 89119       |
| 450 Fremont St, Ste 370                                 |  36.1701 |  -115.141 | 89101       |
| 3570 Las Vegas Blvd S                                   |  36.1162 |  -115.175 | 89019       |
| 2075 Festival Plaza Dr                                  |   36.149 |  -115.335 | 89135       |
| 1930 Village Center Cir                                 |  36.1944 |  -115.305 | 89134       |
| 4500 W Tropicana Ave                                    |  36.1027 |  -115.202 | 89103       |
| 3200 Las Vegas Blvd S                                   |   36.127 |  -115.168 | 89109       |
+---------------------------------------------------------+----------+-----------+-------------+
(Output limit exceeded, 25 of 403 total rows shown)
```

SQL CODE

```
SELECT address,latitude,longitude,postal_code
FROM business
WHERE city='Las Vegas'AND
(stars BETWEEN 2 AND 3)
```



```
+------------------------------------+----------+-----------+-------------+
| address                            | latitude | longitude | postal_code |
+------------------------------------+----------+-----------+-------------+
|                                    |  36.1699 |   -115.14 |             |
| 7355 S Buffalo Dr, Ste 5           |  36.0559 |  -115.281 | 88113       |
|                                    |    36.18 |   -115.14 | 88901       |
| Great Basin Hwy                    |  36.0124 |  -114.742 | 89005       |
| 32100 Las Vegas Blvd S             |  35.6157 |  -115.387 | 89019       |
|                                    |  36.2145 |  -115.122 | 89030       |
|                                    |  36.2608 |  -115.171 | 89031       |
|                                    |  36.2137 |  -115.177 | 89032       |
|                                    |  35.9209 |  -115.165 | 89044       |
| 14200 S Las Vegas Blvd             |  35.9365 |  -115.187 | 89054       |
| 315 S 7th St                       |  36.1666 |  -115.139 | 89101       |
| 2202 W Charleston Blvd, Ste 7      |   36.154 |  -115.115 | 89102       |
| 4983 W Flamingo Rd, Ste A          |  36.1149 |   -115.21 | 89103       |
| 1219 S Main St                     |  36.1568 |  -115.154 | 89104       |
| 625 S Grand Central Pkwy, Ste 1254 |  36.1654 |  -115.156 | 89106       |

| 4300 Meadows Ln                    |  36.1725 |  -115.197 | 89107       |

| 5348 Vegas Dr                      |  36.1886 |  -115.214 | 89108       |

| 3765 Las Vegas Blvd S              |   36.105 |  -115.172 | 89109       |

| 3571 Las Vegas Blvd                |  36.1162 |  -115.174 | 89110       |

|                                    |     36.1 |   -115.07 | 89112       |

| 8425 W Windmill Ln                 |    36.04 |  -115.275 | 89113       |

| 4375 Las Vegas Blvd N              |    36.24 |  -115.057 | 89115       |

| 3455 S Durango Dr, Ste 112         |  36.1272 |   -115.28 | 89117       |

| 5447 S Rainbow Blvd, Ste E6        |  36.0896 |  -115.243 | 89118       |

| 6005 S Las Vegas Blvd              |  36.0802 |  -115.171 | 89119       |

+------------------------------------+----------+-----------+-------------+

(Output limit exceeded, 25 of 56 total rows shown)
```

SQL CODE

```
SELECT address,latitude,longitude,postal_code
FROM business
WHERE city='Las Vegas'AND
(stars BETWEEN 4 AND 5)
GROUP BY postal_code
```

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 average star ratings for closed business is 3.52 which is lower than open business 3.67

```
+---------------+
|    AVG(stars) |
+---------------+
| 3.67900943396 |
```

```
 +---------------+

|    AVG(stars) |
+---------------+
| 3.52039473684 |
+---------------+
```

- Similarly average number of review count for closed business is 23.19 and for open business is 31.75. which shows that business with good digital promotion survives more.

ii. Difference 2:
Comparing below observations for top 9 categories with covers 90 % of all categories.

Observations: 1)Food buiness has high survival rate 20/23=87%
2) Medical health and home services has almost 100% survival rate
3)Nightlife and bars has lowest survival business

```
+---------------------------+-------------+
| category                  | categ_count |
+---------------------------+-------------+
| Restaurants               |          71 |
| Shopping                  |          30 |
| Food                      |          23 |
| Nightlife                 |          20 |    All category
| Bars                      |          17 |
| Health & Medical          |          17 |
| Home Services             |          16 |
| Beauty & Spas             |          13 |
| Local Services            |          12 |
```

```
+------------------------+-------------+
| category               | categ_count |
+------------------------+-------------+
| Restaurants            |          53 |
| Shopping               |          25 |
| Food                   |          20 |    For open category
| Health & Medical       |          16 |
| Home Services          |          15 |
| Beauty & Spas          |          12 |
| Nightlife              |          12 |
| Bars                   |          11 |
| Active Life            |          10 |
| Local Services         |          10 |
```

```
+---------------------------+-------------+
| category                  | categ_count |
+---------------------------+-------------+
| Restaurants               |          18 |
| Nightlife                 |           8 |
| Bars                      |           6 |
| Shopping                  |           5 |
| American (New)            |           3 |
| American (Traditional)    |           3 |    For closed category
| Event Planning & Services |           3 |
| Food                      |           3 |
| Desserts                  |           2 |
| Gluten-Free               |           2 |
| Italian                   |           2 |
| Japanese                  |           2 |
| Local Services            |           2 |
```

SQL code used for analysis:

```
SELECT AVG(stars)
FROM business
WHERE is_open=1;

SELECT AVG(stars)
FROM business
WHERE is_open=0;

SELECT category, COUNT(category) AS categ_count
FROM business b INNER JOIN category c
ON b.id=c.business_id
--WHERE is_open=1
GROUP BY category
ORDER BY categ_count DESC;

SELECT category, COUNT(category) AS categ_count
FROM business b INNER JOIN category c
ON b.id=c.business_id
WHERE is_open=1
```

```
GROUP BY category
ORDER BY categ_count DESC;
```

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:
Ans: Predicting the overall star rating of my newly open shopping store

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

Ans: I have opened a new shopping store. I want to know how much customer are happy with my service. This will give me an idea that how much I can expect a return to investment in future with this shopping store.

For this analysis, I am collecting attributes for shopping category.As most of the variables are categorical variable so, using linear regression would be cumbersome here. I would use a random forest algorithm to predict the overall rating for my store.

1)id(a reference attribute)
2)neighbourhood
3)City
4)latitude (1-7) Idependent variable
5)longitude
6)state
7)Review count

8)stars(dependent variable)

iii. Output of your finished dataset:

```
+------------------------+-----------------+--------------+----------+-----------+-------+--------------+-------+
| id                     | neighborhood    | city         | latitude | longitude | state | review_count | stars |
+------------------------+-----------------+--------------+----------+-----------+-------+--------------+-------+
| 25lVJgvthMyvoRz-W6splw |                 | Mesa         |  33.3906 |   -111.69 | AZ    |            3 |   2.0 |
| -iu4FxdfxN4rU4Fu9BjiFw |                 | Strongsville |  41.3141 |  -81.8207 | OH    |            3 |   4.0 |
| 0t2yPpsbObqxB8PRyLRUhg |                 | Pittsburgh   |  40.4521 |   -80.165 | PA    |            8 |   5.0 |
| -ayZoW_iNDsunYXX_0x1YQ |                 | Phoenix      |  33.4664 |  -112.018 | AZ    |           15 |   3.5 |
| 15KgSGyazYR960nTLs5wDQ | University City | Charlotte    |  35.3167 |  -80.7405 | NC    |            5 |   4.0 |
| -uiBBVWI6tMDm2JFbZFrOw | The Annex       | Toronto      |  43.6727 |  -79.4142 | ON    |            6 |   4.5 |
| 1UPbt3BRYU8FmvtEBTXJZQ | South End       | Charlotte    |  35.2026 |   -80.866 | NC    |            6 |   3.5 |
| 0K2rKvqdBmiOAUTebcUohQ |                 | Las Vegas    |  36.1357 |  -115.428 | NV    |           32 |   4.5 |
| -2HjuT4yjLZ3b5f_abD87Q |                 | Charlotte    |  35.1727 |  -80.8755 | NC    |            8 |   3.5 |
| 0V-I5TazN_FeeHg4oiXHDA |                 | Stuttgart    |  48.7782 |    9.1684 | BW    |            3 |   3.5 |
| 0NjV892hH8aymSGo75bpJg |                 | Gilbert      |  33.3353 |   -111.76 | AZ    |            4 |   2.0 |
| 0oDfGJPbqdSigRwIFM-RoQ |                 | Chandler     |  33.3497 |  -111.858 | AZ    |            4 |   5.0 |
| 1FYLWIcM9B9w8F235YKz5w |                 | Tempe        |   33.408 |   -111.91 | AZ    |            3 |   3.5 |
| 1hlwL5E035WQfB7Zb2mLUw |                 | Chandler     |  33.3199 |   -111.81 | AZ    |            7 |   5.0 |
| -j4NsiRzSMrMk2N_bGH_SA |                 | Chandler     |  33.3496 |  -111.892 | AZ    |            5 |   4.0 |
| -Eu04UHRqmGGyvYRDY8-tg | Ohio City       | Cleveland    |  41.4847 |  -81.7031 | OH    |          723 |   4.5 |
| -n27mJ_jQWGCuIukTvg9Mg | High Park       | Toronto      |  43.6553 |  -79.4567 | ON    |           26 |   4.5 |
| -9y2L9qSbqukVl8LzEOGdg | Southeast       | Las Vegas    |  36.0994 |    -115.1 | NV    |           11 |   3.5 |
| 2ZcKa9r9Pci3KZIWuyhP9A | Erindale        | Mississauga  |  43.5489 |  -79.6502 | ON    |           10 |   3.5 |
| 0JoJSub9w_KmONZrDzpFTg |                 | Chandler     |  33.3052 |  -111.903 | AZ    |            3 |   4.0 |
| 1q44aWEcDN7uRvA2l8xpvQ | Eastside        | Las Vegas    |  36.1007 |  -115.091 | NV    |            6 |   2.5 |
| -tKN8LLme5IMC9AjzB9y9Q | Leith           | Edinburgh    |  55.9586 |   -3.1717 | EDH   |            6 |   3.5 |
| 2RhICgMZI6DK-t374VRoow |                 | Las Vegas    |  36.0964 |  -115.187 | NV    |            4 |   5.0 |
| 1KNJI4JT1lT2hsHZM_m28g |                 | Phoenix      |  33.4944 |  -112.039 | AZ    |            3 |   4.5 |
| 1BXyj0B-3hgODg1IFnIDVA | Yorkville       | Toronto      |  43.6693 |  -79.3936 | ON    |            4 |   4.0 |
+------------------------+-----------------+--------------+----------+-----------+-------+--------------+-------+
            (Output limit exceeded, 25 of 30 total rows shown)
```

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

```
SELECT id,neighborhood,city,latitude,longitude,state,review_count,stars
FROM business b INNER JOIN category c
ON b.id=c.business_id
WHERE category='Shopping'
```