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:

1. Attribute table = 10000
2. Business table = 10000
3. Category table = 10000
4. Checkin table = 10000
5. elite_years table = 10000
6. friend table = 10000
7. hours table = 10000
8. photo table = 10000

9. review table = 100000
10. tip table = 10000
11. user table = 10000

SQL Code to arrive at answer:

```
SELECT COUNT(*) as No_of_ROWS FROM business
```

Copy and Paste the Result Below:

```
+------------+
| No_of_ROWS |
+------------+
|      10000 |
+------------+
```

Note: Similar SQL code like above, was used for finding number of rows for all columns of Yelp dataset.

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.

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

Note: I have mentioned no. of rows and primary/foregin key along side in the result mentioned.

1. Business = 10000 [id - PK]
2. Hours = 1562 [business_id - FK]
3. Category = 2643 [business_id - FK]
4. Attribute = 1115 [business_id - FK]
5. Review = 10000 [id-PK], 8090 [business_id-FK], 9581 [user_id-FK]
6. Checkin = 493 [business_id - FK]
7. Photo = 10000 [id - PK], 6493 [business_id - FK]
8. Tip = 537 [user_id - FK] , 3979 [business_id - FK]
9. User = 10000 [id - PK]

10. Friend = 11 [user_id - FK]
11. Elite_years = 2780 [user_id - FK]

SQL code used to arrive at answer:

```
i. SELECT COUNT(DISTINCT(id)) AS distinct_rows FROM business
ii. SELECT COUNT(DISTINCT(business_id)) AS distinct_rows FROM hours
```

Copy and Paste the Result Below:

```
+---------------+
| distinct_rows |
+---------------+
|         10000 |
+---------------+

+---------------+
| distinct_rows |
+---------------+
|          1562 |
+---------------+
```

Note: Similar SQL code like above, was used for finding distinct rows for all columns.

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

Answer: no

SQL Code to arrive at answer:

```
select *
    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:

1. Table: Review, Column: Stars

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

```
SELECT MIN(stars) as min, MAX(stars) as max, AVG(stars) as avg FROM review
```

1. Table: Business, Column: Stars

```
min: 1.0    max: 5.0    avg: 3.6549
```

```
SELECT MIN(stars) as min, MAX(stars) as max, AVG(stars) as avg FROM business
```

1. Table: Tip, Column: Likes

```
min: 0      max: 2      avg: 0.0144
```

```
SELECT MIN(likes) as min, MAX(likes) as max, AVG(likes) as avg FROM tip
```

1. Table: Checkin, Column: Count

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

```
SELECT MIN(count) as min, MAX(count) as max, AVG(count) as avg FROM checkin
```

1. Table: User, Column: Review_count

```
min: 0      max: 2000   avg: 24.2995
```

```
SELECT MIN(review_count) as min, MAX(review_count) as max, AVG(review_count) as avg FROM user
```

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

SQL code used to arrive at answer:

```
SELECT city,
    SUM(review_count) AS review_count
    FROM business
    GROUP BY city
    ORDER BY review_count DESC
```

Copy and Paste the Result Below:

```
+---------------+-----------------+
| total_reviews | city            |
+---------------+-----------------+
|         82854 | Las Vegas       |
|         34503 | Phoenix         |
|         24113 | Toronto         |
|         20614 | Scottsdale      |
|         12523 | Charlotte       |
|         10871 | Henderson       |
|         10504 | Tempe           |
|          9798 | Pittsburgh      |
|          9448 | Montréal        |
|          8112 | Chandler        |
|          6875 | Mesa            |
|          6380 | Gilbert         |
|          5593 | Cleveland       |
|          5265 | Madison         |
|          4406 | Glendale        |
|          3814 | Mississauga     |
|          2792 | Edinburgh       |
|          2624 | Peoria          |
|          2438 | North Las Vegas |
|          2352 | Markham         |
|          2029 | Champaign       |
|          1849 | Stuttgart       |
|          1520 | Surprise        |
|          1465 | Lakewood        |
|          1155 | Goodyear        |
+---------------+-----------------+
(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",
     COUNT(stars) AS "count"
     FROM
     business
     WHERE city = "Avon"
     GROUP BY stars
```

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

```
+-------------+-------+
| star rating | Count |
+-------------+-------+
|         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 as "star rating",
     COUNT(stars) AS "count"
     FROM
     business
     WHERE city = "Beachwood"
     GROUP BY stars
```

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

```
+-------------+-------+
| star rating | count |
+-------------+-------+
|         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,
    review_count
    FROM
    user
    ORDER BY review_count DESC
    LIMIT 3
```

Copy and Paste the Result Below:

```
+--------+--------------+
| name   | review_count |
+--------+--------------+
| Gerald |         2000 |
| Sara   |         1629 |
| Yuri   |         1339 |
+--------+--------------+
```

8. Does posing more reviews correlate with more fans?

Please explain your findings and interpretation of the results:

Yes. it seems correlated. Users with more reviews approximately have more fans.

SQL code to arrive at answer:

```
SELECT
    range_bin AS fans_bins,
    AVG(review_count) AS avg_no_of_reviews,
    COUNT(*) AS num_user,
    AVG(fans) AS avg_num_of_fans
    FROM (
        SELECT
        CASE
        WHEN fans BETWEEN 0 AND 9 THEN '0 - 9'
        WHEN fans BETWEEN 10 AND 99 THEN '10 - 99'
        ELSE '100-1000' END AS range_bin,
        review_count,
        fans
        FROM user
    ) AS sub_query_result
    GROUP BY sub_query_result.range_bin
```

Copy and paste the result below:

```
+-----------+------------------+----------+-----------------+
| fans_bins | avg_no_of_reviews | num_user | avg_num_of_fans |
+-----------+------------------+----------+-----------------+
| 0 - 9     |     15.0085655315 |     9690 |  0.447265221878 |
| 10 - 99   |     283.326530612 |      294 |   25.5986394558 |
| 100-1000  |             891.5 |       16 |          189.75 |
+-----------+------------------+----------+-----------------+
```

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

Answer: There are more reviews with word "love"

SQL code used to arrive at answer:

```
SELECT reaction_word, count(*) AS word_count
FROM (
      SELECT
        CASE
        WHEN LOWER(text) LIKE '%hate%' THEN 'hate'
        WHEN LOWER(text) LIKE '%love%' THEN 'love'
        ELSE NULL END AS reaction_word
      FROM
      review
)
GROUP BY reaction_word
ORDER BY word_count DESC
```

Copy and Paste the result below:

```
+---------------+------------+
| reaction_word | word_count |
+---------------+------------+
|     None      |    8042     |
|     love      |    1780     |
|     hate      |    178      |
+---------------+-------------+
```

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 |
+-----------+------+
```

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 choose 2 cities and 2 categories. just to get better analysis.

1. Cities: Las Vegas, Phoenix
2. Categories: Restaurants, Shopping

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

SQL Code to arrive at answer:

```
SELECT
    city,
    category.category,
    CASE
    WHEN stars >= 4.0 THEN '4-5'
    WHEN stars >= 2.0 THEN '2-3'
    ELSE 'below 2'
    END AS 'stars_bin',
    COUNT(DISTINCT business.id) AS bus_id_count,
    COUNT(hours) AS bus_open_days_total,
    COUNT(hours)*1.0 / COUNT(DISTINCT business.id) AS bus_open_days_avg
    FROM
    business INNER JOIN hours ON business.id = hours.business_id
    INNER JOIN category ON business.id = category.business_id
    WHERE city IN ('Las Vegas', 'Phoenix') AND category.category IN ('Shopping','Restaurants')
    GROUP BY stars_bin, city, category.category
    ORDER BY city, category.category
```

Copy and paste the result below:

```
+-----------+-------------+-----------+--------------+---------------------+-------------------+
| city      | category    | stars_bin | bus_id_count | bus_open_days_total | bus_open_days_avg |
+-----------+-------------+-----------+--------------+---------------------+-------------------+
| Las Vegas | Restaurants | 2-3       |            1 |                   7 |               7.0 |
| Las Vegas | Restaurants | 4-5       |            2 |                  14 |               7.0 |
| Las Vegas | Shopping    | 2-3       |            2 |                  13 |               6.5 |
| Las Vegas | Shopping    | 4-5       |            2 |                  12 |               6.0 |
| Phoenix   | Restaurants | 2-3       |            3 |                  21 |               7.0 |
| Phoenix   | Restaurants | 4-5       |            2 |                  14 |               7.0 |
| Phoenix   | Shopping    | 2-3       |            1 |                    6 |               6.0 |
+-----------+-------------+-----------+--------------+----------------------+-------------------+
```

Conclusion: When I compare business open days hours average, I can see there is no huge difference in distribution of hours between the groups

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

SQL code to arrive at answer:

```
SELECT
   city,
   category.category,
   CASE
   WHEN stars >= 4.0 THEN '4-5'
   WHEN stars >= 2.0 THEN '2-3'
   ELSE 'Below 2'
   END AS 'stars_bin',
   COUNT(business_id) AS bus_id_count,
   SUM(review_count) AS bus_reviews_total_count,
   SUM(review_count)*1.0/COUNT(DISTINCT business_id) AS bus_reviews_avg_count
   FROM business
   INNER JOIN
   category
   ON business.id = category.business_id
   WHERE city IN ('Las Vegas', 'Phoenix') AND category.category IN ('Shopping', 'Restaurants')
   GROUP BY stars_bin, city, category.category
   ORDER BY city, category.category
```

Copy and paste the result below:

```
+-----------+-------------+-----------+--------------+-------------------------+-----------------------+
| city      | category    | stars_bin | bus_id_count | bus_reviews_total_count | bus_reviews_avg_count |
+-----------+-------------+-----------+--------------+-------------------------+-----------------------+
| Las Vegas | Restaurants | 2-3       |            1 |                     123 |                 123.0 |
| Las Vegas | Restaurants | 4-5       |            3 |                     939 |                 313.0 |
| Las Vegas | Shopping    | 2-3       |            2 |                      17 |                   8.5 |
| Las Vegas | Shopping    | 4-5       |            2 |                      36 |                  18.0 |
| Phoenix    | Restaurants | 2-3       |            3 |                      131 |          43.6666666667 |
| Phoenix    | Restaurants | 4-5       |            3 |                      626 |          208.666666667 |
| Phoenix    | Shopping    | 2-3       |            1 |                       15 |                   15.0 |
| Phoenix    | Shopping    | 4-5       |            1 |                        3 |                    3.0 |
+-----------+-------------+-----------+--------------+------------------------+----------------------+
```

Conclusion:

1. When I compare businesss reviews average count I can see there is clearly a different between them.
2. Businesses with 4-5 stars ratings appromiately have 3-4 times more reviews then business with 2-3 stars ratings.

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

SQL code to arrive at answer:

```
SELECT
    city,
    category.category,
    CASE
    WHEN stars >= 4.0 THEN '4-5'
    WHEN stars >= 2.0 THEN '2-3'
    ELSE 'Below 2'
    END AS 'stars_bin',
    business.latitude,
    business.longitude,
    business.neighborhood ,
    business.address,
    business.postal_code
    FROM business INNER JOIN category ON business.id = category.business_id
    WHERE city IN ('Las Vegas', 'Phoenix') AND category.category IN ('Shopping', 'Restaurants')
    GROUP BY stars_bin, city, category.category
    ORDER BY city, stars_bin
```

Copy and paste the answer below:

```
+-----------+-------------+-----------+----------+-----------+--------------+-------------------------+------
-------+
| city      | category    | stars_bin | latitude | longitude | neighborhood | address                 |
postal_code |
+-----------+-------------+----------+----------+-----------+--------------+-------------------------+------
-------+
| Las Vegas | Restaurants | 2-3      |  36.1003 |   -115.21 |              | 5045 W Tropicana Ave    | 89103
|
| Las Vegas | Shopping    | 2-3      |  36.1007 |  -115.091 | Eastside     | 3808 E Tropicana Ave    | 89121
|
| Las Vegas | Restaurants | 4-5      |  36.1259 |  -115.135 | Eastside     | 3480 S Maryland Pkwy    | 89169
|
| Las Vegas | Shopping    | 4-5      |  36.0964 |  -115.187 |              | 3555 W Reno Ave, Ste F  | 89118
|
| Phoenix   | Restaurants | 2-3      |  33.6536 |  -112.064 |              | 751 E Union Hls Dr      | 85024
|
| Phoenix   | Shopping    | 2-3      |  33.4664 |  -112.018 |              | 2922 E McDowell Rd      | 85008
|
| Phoenix   | Restaurants | 4-5      |  33.5818 |  -112.008 |              | 3375 E Shea Blvd        | 85028
|
| Phoenix   | Shopping    | 4-5      |  33.4944 |  -112.039 |              | 1945 E Indian School Rd | 85016
|
+-----------+-------------+----------+----------+-----------+--------------+-------------------------+------
-------+
```

Conclusions:

1. In Las Vegas, Restaurants and shops with 2-3 stars are mostly in Tropicana Ave area and business with 4-5 star ratings are in other area.
2. In Phoenix, Restaurants and shops with 2-3 stars and businesses with 4-5 stars ratings are in different different areas.

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.

Difference 1: Number of open business across state, city are more compare to closed business and average ratings stars for open and close business is approxmiately same. but average number of reviews are more for open business.

Difference 2: When we consider only business which are reviewed, then Number of review & average ratings stars, both of them in the open-business are higher then closed business.

SQL code used for analysis:

```
SELECT
     is_open,
     COUNT(distinct business.id) as open_bus,
     AVG(stars) as average_stars,
     AVG(review_count) as average_reviews
     FROM
     business
     group by is_open


    copy and paste the answer below:

    +---------+----------+---------------+-----------------+
    | is_open | open_bus | average_stars | average_reviews |
    +---------+----------+---------------+-----------------+
    |       0 |     1520 | 3.52039473684 |   23.1980263158 |
    |       1 |     8480 | 3.67900943396 |   31.7570754717 |
    +---------+----------+---------------+-----------------+
```

```
SELECT is_open,
            count(distinct business.id) num_of_business,
            count(distinct review.id) num_of_review,
            avg(review.stars) avg_stars
     FROM business
     JOIN review ON business.id =  review.business_id
     GROUP BY is_open


    copy and paste the answer below:

    +---------+-----------------+---------------+---------------+
    | is_open | num_of_business | num_of_review |     avg_stars |
    +---------+-----------------+---------------+---------------+
    |       0 |              61 |            71 | 3.64788732394 |
    |       1 |             446 |           565 |  3.7610619469 |
    +---------+-----------------+---------------+---------------+
```

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:

---

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

Analysis: What type of business categories are more opend?

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

Let say I want to open a new business. but I want to understand what kind/category of business are most open. So, I fetched the categories, number of business across the categoris, average review counts and average ratings stars.

From the output, "shopping", "Food", "restaurants" and "bars" are few of the most opend business.

3. Output of your finished dataset:

```
+------------------------+---------+---------------------------+-----------------+---------------+
| category               | is_open | open_business_in_category | avg_review_count |     avg_stars |
+------------------------+---------+---------------------------+-----------------+---------------+
| Restaurants            |       1 |                        53 |   71.1698113208 | 3.45283018868 |
| Shopping               |       1 |                        25 |            37.8 |           4.0 |
| Food                   |       1 |                        20 |            79.4 |         3.725 |
| Restaurants            |       0 |                        18 |   40.6666666667 | 3.47222222222 |
| Health & Medical       |       1 |                        16 |          12.375 |       4.21875 |
| Home Services          |       1 |                        15 |   6.06666666667 | 3.93333333333 |
| Beauty & Spas          |       1 |                        12 |   9.66666666667 | 3.79166666667 |
| Nightlife              |       1 |                        12 |   79.3333333333 |         3.625 |
| Bars                   |       1 |                        11 |   85.9090909091 | 3.63636363636 |
| Active Life            |       1 |                        10 |            13.1 |          4.15 |
| Local Services         |       1 |                        10 |             9.4 |          4.35 |
| Automotive             |       1 |                         9 |             22.0 |            4.5 |
| Nightlife              |       0 |                         8 |           49.875 |           3.25 |
| American (Traditional) |       1 |                         8 |           139.25 |         3.8125 |
| Hotels & Travel        |       1 |                         8 |             46.5 |         3.4375 |
+------------------------+---------+---------------------------+------------------+----------------+
```

4. **SQL Code to arrive at answer:**

```
select
    category.category,
    is_open,
    COUNT(is_open) as "open_business_in_category",
    AVG(review_count) as "avg_review_count",
    AVG(stars) as "avg_stars"
    FROM business inner join category
    on business.id = category.business_id
    group by is_open, category.category
    HAVING count(is_open) >= 8
    order by open_business_in_category desc
```

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,
(useful + funny)*1.0/(useful + funny + cool) AS p_useful_funny,
fans,
CASE
```

```
WHEN (useful + funny)*1.0/(useful + funny + cool) > 0.0 AND
     (useful + funny)*1.0/(useful + funny + cool) <=0.25 THEN 'Low'
WHEN (useful + funny)*1.0/(useful + funny + cool) > 0.25 AND
     (useful + funny)*1.0/(useful + funny + cool) <=0.75 THEN 'Medium'
ELSE 'Strong' END AS Relationship
FROM user
ORDER BY fans DESC
LIMIT 10
```

```
   Copy and Paste the Result Below:

   +-----------+----------------+------+--------------+
   | name      | p_useful_funny | fans | Relationship |
   +-----------+----------------+------+--------------+
   | Amy       | 0.677529011839 |  503 | Medium       |
   | Mimi      | 0.712996389892 |  497 | Medium       |
   | Harald    | 0.666268364881 |  311 | Medium       |
   | Gerald    | 0.569428505853 |  253 | Medium       |
   | Christine | 0.726536295171 |  173 | Medium       |
   | Lisa      | 0.910447761194 |  159 | Strong       |
   | Cat       | 0.617081850534 |  133 | Medium       |
   | William   | 0.666476827792 |  126 | Medium       |
   | Fran      | 0.651356292676 |  124 | Medium       |
   | Lissa     | 0.638859556494 |  120 | Medium       |
   +-----------+----------------+------+--------------+
```

Please explain your findings and interpretation of the results:

On average, there is medium relationship.

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

Analysis: What is the most successfull business category ?

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

Among the business categories, I found the average of stars and the proportion of opening on each category.I have only consider the set of category with more than 10 of business for statistical reason.

From the output, we can see that "Shopping", "Health & Medica", "Home Services","Local Service" and "Beauty & Spas" are successful. They are getting better reviews and higher opening rate.

3. Output of your finished dataset:

```
+------------------------+-----------------+------------------+----------------+
| category               | num_of_business | avg_star_ratings | isopen_average |
+------------------------+-----------------+------------------+----------------+
| Local Services         |              12 |             4.21 |           0.83 |
| Health & Medical       |              17 |             4.09 |           0.94 |
| Home Services          |              16 |              4.0 |           0.94 |
| Shopping               |              30 |             3.98 |           0.83 |
| Beauty & Spas          |              13 |             3.88 |           0.92 |
| American (Traditional) |              11 |             3.82 |           0.73 |
| Food                   |              23 |             3.78 |           0.87 |
| Bars                   |              17 |              3.5 |           0.65 |
| Nightlife              |              20 |             3.48 |            0.6 |
| Restaurants            |              71 |             3.46 |           0.75 |
+------------------------+-----------------+------------------+----------------+
```

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

```sql
SELECT category.category,
       count(business.id) num_of_business,
       round(avg(business.stars),2) avg_star_ratings,
       round(avg(business.is_open),2) isopen_average
FROM (business INNER JOIN category ON business.id = category.business_id)
GROUP BY category.category
HAVING num_of_business > 10
ORDER BY avg_star_ratings DESC, isopen_average DESC
```