Profiling and Analyzing Yelp Dataset

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.

Table	Primary Key	Business_id	User_id	Friend_id
Business	10000			
Hours		1562		
Category		2643		
Attribute		1115		
Review	10000	8090	9581	
Checkin		493		
Photo	10000	6493		
Tip		3979	537	
User	10000			

Friend			11	9415
Elite_years			2780	

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
  SUM(CASE WHEN id is null then 1 else 0 end) AS id
  ,SUM(CASE WHEN name is null then 1 else 0 end) AS Name
  ,SUM(CASE WHEN review_count is null then 1 else 0 end) AS Review_count
  ,SUM(CASE WHEN yelping_since is null then 1 else 0 end) AS Yelping_since
  ,SUM(CASE WHEN useful is null then 1 else 0 end) AS Useful
  ,SUM(CASE WHEN funny is null then 1 else 0 end) AS Funny
  ,SUM(CASE WHEN cool is null then 1 else 0 end) AS Cool
  ,SUM(CASE WHEN fans is null then 1 else 0 end) AS Fans
  ,SUM(CASE WHEN average_stars is null then 1 else 0 end) AS Average_stars
  ,SUM(CASE WHEN compliment_hot is null then 1 else 0 end) AS Compliment_hot
  ,SUM(CASE WHEN compliment_more is null then 1 else 0 end) AS Compliment_more
  ,SUM(CASE WHEN compliment_profile is null then 1 else 0 end) AS Compliment_profile
  ,SUM(CASE WHEN compliment_cute is null then 1 else 0 end) AS Compliment_cute
  ,SUM(CASE WHEN compliment_list is null then 1 else 0 end) AS Compliment_list
  ,SUM(CASE WHEN compliment_note is null then 1 else 0 end) AS Compliment_note
  ,SUM(CASE WHEN compliment_plain is null then 1 else 0 end) AS Compliment_plain
  ,SUM(CASE WHEN compliment_cool is null then 1 else 0 end) AS Compliment_cool
  ,SUM(CASE WHEN compliment_funny is null then 1 else 0 end) AS Compliment_funny
  ,SUM(CASE WHEN compliment_writer is null then 1 else 0 end) AS Compliment_writer
  ,SUM(CASE WHEN compliment_photos is null then 1 else 0 end) AS Compliment_photos
FROM User;
```

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

Table	Column	Min	Max	Avg
Review	Stars	1	5	3.7082
Business	Stars	1	5	3.6549

Tip	Likes	0	2	0.0144
Checkin	Count	1	53	1.9414
User	Review_count	0	2000	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 |
| 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

```
SELECT stars AS StarsRating, Count(*) AS Businesses
FROM business
WHERE City = 'Avon'
GROUP BY stars
ORDER BY stars DESC
```

```
+-------------+------------+
| StarsRating | Businesses |
+-------------+------------+
|         5.0 |          1 |
|         4.5 |          1 |
|         4.0 |          2 |
|         3.5 |          3 |
|         2.5 |          2 |
|         1.5 |          1 |
+-------------+------------+
```

ii. Beachwood

```
SELECT stars AS StarsRating, Count(*) AS Businesses
FROM business
WHERE City = 'Beachwood'
GROUP BY stars
ORDER BY stars DESC
```

```
+-------------+------------+
| StarsRating | Businesses |
+-------------+------------+
|         5.0 |          5 |
|         4.5 |          2 |
|         4.0 |          1 |
|         3.5 |          2 |
|         3.0 |          2 |
|         2.5 |          1 |
|         2.0 |          1 |
+-------------+------------+
```

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:

As the table below illustrates, posting more reviews does not necessarily correlate with more fans.

For example, although Gerald has posted the most reviews, he has fewer fans in comparison with Mimi.
Therefore, sorting the users in descending order based on their total number of reviews does not sort the fans in the same order, meaning that there is not a correlation between the total number of reviews and number of fans.

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

```
+-----------+------------------------+--------------+------+
| name      | id                     | review_count | fans |
+-----------+------------------------+--------------+------+
| Gerald    | -G7Zkl1wIWBBmD0KRy_sCw |         2000 |  253 |
| Sara      | -3s52C4zL_DHRK0ULG6qtg |         1629 |   50 |
| Yuri      | -8lbUNlXVSoXqaRRiHiSNg |         1339 |   76 |
| .Hon      | -K2Tcgh2EKX6e6HqqIrBIQ |         1246 |  101 |
| William   | -FZBTkAZEXoP7CYvRV2ZwQ |         1215 |  126 |
| Harald    | --2vR0DIsmQ6WfcSzKWigw |         1153 |  311 |
| eric      | -gokwePdbXjfS0iF7NsUGA |         1116 |   16 |
| Roanna    | -DFCC64NXgqrxlO8aLU5rg |         1039 |  104 |
| Mimi      | -8EnCioUmDygAbsYZmTeRQ |          968 |  497 |
| Christine | -0IiMAZI2SsQ7VmyzJjokQ |          930 |  173 |
| Ed        | -fUARDNuXAfrOn4WLSZLgA |          904 |   38 |
| Nicole    | -hKniZN2OdshWLHYuj21jQ |          864 |   43 |
| Fran      | -9da1xk7zgnnfO1uTVYGkA |          862 |  124 |
| Mark      | -B-QEUESGWHPE_889WJaeg |          861 |  115 |
```

```
| Christina | -kLVfaJytOJY2-QdQoCcNQ |          842 |   85 |
| Dominic   | -kO6984fXByyZm3_6z2JYg |          836 |   37 |
| Lissa     | -lh59ko3dxChBSZ9U7LfUw |          834 |  120 |
| Lisa      | -g3XIcCb2b-BD0QBCcq2Sw |          813 |  159 |
| Alison    | -l9giG8TSDBG1jnUBUXp5w |          775 |   61 |
| Sui       | -dw8f7FLaUmWR7bfJ_Yf0w |          754 |   78 |
| Tim       | -AaBjWJYiQxXkCMDlXfPGw |          702 |   35 |
| L         | -jt1ACMiZljnBFvS6RRvnA |          696 |   10 |
| Angela    | -IgKkE8JvYNWeGu8ze4P8Q |          694 |  101 |
| Crissy    | -hxUwfo3cMnLTv-CAaP69A |          676 |   25 |
| Lyn       | -H6cTbVxeIRYR-atxdielQ |          675 |   45 |
+-----------+------------------------+--------------+------+
(Output limit exceeded, 25 of 10000 total rows shown)
```

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

Answer: Love

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

SQL code used to arrive at answer:

```
SELECT
   SUM(CASE WHEN text like "%love%" then 1 else 0 end) AS Love
   ,SUM(CASE WHEN text like "%hate%" then 1 else 0 end) 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 |
+-----------+------+
```

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?

Yes, they do, the data appears to indicate that the businesses that are open for more hours on average are more likely to have 2-3 stars.

City = Las Vegas OR Category = Restaurants GROUP By stars

```
+-------------+-----------------+-------+------+------------+
| Stars_group | Selection_group | Hours | Days | Businesses |
+-------------+-----------------+-------+------+------------+
| 2-3         | Las Vegas       |  1596 |   27 |          4 |
| 2-3         | Restaurant      |  1874 |  150 |         25 |
| 4-5         | Las Vegas       |  2686 |   49 |          9 |
| 4-5         | Restaurant      |  1155 |  113 |         18 |
+-------------+-----------------+-------+------+------------+
```

Businesses with 2-3 stars appear to have more hours open than the ones with 4-5 but we cannot conclude this for sure because of an irregular distribution on the business population at Las Vegas, so we need to look deeper.

City = Las Vegas

```
+-------+-------+------+------------------------+---------------------------------+
| Stars | Hours | Days | id                     | name                            |
+-------+-------+------+------------------------+---------------------------------+
| 2-3   |   102 |    6 | -9y2L9qSbqukVl8LzEOGdg | Wooly Wonders                   |
| 2-3   |   168 |    7 | 0gWg-kqRLEQbhui8b_v2Xw | Hi Scores - Blue Diamond        |
| 2-3   |    77 |    7 | 1CP8aJa8ILlfM5deroar0Q | Wingstop                        |
| 2-3   |    98 |    7 | 1q44aWEcDN7uRvA2l8xpvQ | Walgreens                       |
| 4-5   |    54 |    6 | -CdstAUdEvci8GeJG8owpQ | Motors & More                   |
| 4-5   |    56 |    7 | 0K2rKvqdBmiOAUTebcUohQ | Red Rock Canyon Visitor Center  |
| 4-5   |   114 |    6 | 0aKsGxx7XP2TMs_fn_9xVw | Sweet Ruby Jane Confections     |
```

4-5	64	7	0sOwP5b_178FkbT_fat8lQ	Vue at Centennial
4-5	40	4	1YdiQM-3oAQMPlFJbI0e0A	Children's Dental Center
4-5	161	7	1ZnVfS-qP19upP_fwOhZsA	Big Wong Restaurant
4-5	49	6	1_iibQxnp0WhQH2m7kXtng	Anthem Pediatrics
4-5	120	7	1aj4TG0eFq6NaPBKk6bK7Q	Jacques Cafe
4-5	45	5	2RhICgMZI6DK-t374VRoow	Desert Medical Equipment

Businesses in Las Vegas with 2-3 stars have a median of 100 hours worked per week while the businesses with 4-5 stars have a median of 56 hours worked per week, but we have too little population with 2-3 stars to conclude something meaningful.

Category = Restaurants

Stars	Hours	Days	id
2-3	100	7	1Ds8V2c7LlwSAA3O-9f4cA
2-3	84	7	1nTMWMa6v-eBKkPYA3gxkQ
2-3	88	7	2LVuwl-eH-8PYikyFmqcTQ
2-3	77	7	-0DET7VdEQOJVJ_v6klEug
2-3	64	7	-OsPCfouYyJ3vjgOKBtzGA
2-3	54	6	-d9qyfNhLMQwVVg_raBKeg
2-3	74	7	0B3W6KxkD3o4W4l6cq735w
2-3	69	7	0CAzhX1w9qGD8iz4F8XZjQ
2-3	119	7	0cxO1Lx2Pi7u6ftWX3Wksg
2-3	70	7	0hBGwOLU2UfiYXkM8wc8Hw
2-3	63	7	1ArRdNrB7RjZ6B3X6JW3eA
2-3	77	7	1CP8aJa8ILlfM5deroar0Q
2-3	81	7	1GaooxqCWHzulI2Ub3CXEw
2-3	80	7	1NyHpXJqSLHnvDCOW0nJDg
2-3	67	7	2JV0xGXsszojof2BuEt_hw
2-3	70	6	2WfY9bow3Mv924gfDB8kqg
2-3	72	6	2yF0qgsSKHdawSRopnXguA
2-3	84	7	01xXe2m_z048W5gcBFpoJA
2-3	91	6	0IySwcfqwJjpHPsYwjpAkg
2-3	27	6	0NDbUCHi9YsRwgG3iZO8Kg
2-3	70	7	0kzPQQL8wVcHlBQzMdRdWQ
2-3	71	6	10Jk5ilimXrfAq8JJ1gISg
2-3	105	7	11bhfBbcFypczdz3N_w6iw
2-3	59	6	1OxSzNUssdRohY5dC-kWVg
2-3	58	7	2z3gnLoBNJPlXswFDESFxQ
4-5	72	7	-3oxnPPPU3YoxO9M1I2idg
4-5	59	7	0kyhbUW6NkpYjJzFBZ64vQ
4-5	73	7	1AxEmgv8Dsr3iU9Aa40jPw
4-5	61	7	1D7U-KEvoQDqWJNiYTNbZg
4-5	91	7	1ZnVfS-qP19upP_fwOhZsA
4-5	63	7	1_y5e1u-o93EKOigXgR3LQ
4-5	60	7	1aj4TG0eFq6NaPBKk6bK7Q
4-5	54	7	27nh-2hNnNkf2dBk9aeKHQ
4-5	70	7	2aiaryk7kgUBhXhVu-9vHg
4-5	40	6	2rcrwnlPd_w5oieGVyDgpw
4-5	77	7	2skQeu3C36VCiB653MIfrw

4-5	24	5	37kk0IW6jL7ZlxZF6k2QBg
4-5	106	7	-n27mJ_jQWGCuIukTvg9Mg
4-5	81	7	0e-j5VcEn54EZT-FKCUZdw
4-5	45	7	1mkDrJRu3VABKy95gxD-Hg
4-5	44	6	1veVZUawy7IhIc5oDpRRQA
4-5	54	6	13eX63udIlRK8NNF0EnwAQ
4-5	81	7	16d3BlncEyCTzb0GxXrBXQ

Meanwhile, we have a larger population of restaurants, so now we can look a little deeper in this data and say that the restaurants with 2-3 stars worked on average 23% more hours per week than the restaurants with 4-5 stars. Restaurants with 2-3 stars have a median of 72 hours worked per week while the restaurants with 4-5 stars have a median of 62 hours worked per week.

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

Yes, in the table below we can appreciate that there is a relationship between more reviews and better stars.

Selection_group	Stars_group	Businesses	Reviews
Las Vegas	2-3	664	34748
Las Vegas	4-5	835	46013
Restaurant	2-3	39	1095
Restaurant	4-5	26	2339
Restaurant - Las Vegas	2-3	1	123
Restaurant - Las Vegas	4-5	3	939

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

The tourism industry influences the market to open more business causing more reviews so businesses in these cities are more likely to have reviews with 4-5 stars.

Note: I don't know for sure if the two groups we are talking about are the stars group or the city and category group but in this case I'm taking in consideration the stars groups.

Stars_group	Reviews	Businesses	city
4-5	46952	838	Las Vegas
2-3	34871	665	Las Vegas
4-5	19848	503	Phoenix
2-3	14061	448	Phoenix
2-3	13992	512	Toronto

```
| 4-5          |   12688 |        325 | Scottsdale |
| 4-5          |    9951 |        439 | Toronto    |
| 2-3          |    7780 |        157 | Scottsdale |
| 4-5          |    7277 |        170 | Montréal   |
| 4-5          |    7023 |        221 | Charlotte  |
| 4-5          |    5769 |        139 | Henderson  |
| 4-5          |    5583 |        139 | Tempe      |
| 2-3          |    5302 |        230 | Charlotte  |
| 4-5          |    5121 |        186 | Pittsburgh |
| 2-3          |    5044 |        129 | Henderson  |
| 2-3          |    4845 |        113 | Tempe      |
| 2-3          |    4652 |        161 | Pittsburgh |
| 4-5          |    4600 |        112 | Gilbert    |
| 4-5          |    4546 |        122 | Chandler   |
| 4-5          |    4022 |        103 | Cleveland  |
+--------------+---------+------------+------------+
```

SQL code used for analysis:

Distribution of hours and days per week

```sql
SELECT DISTINCT
   CASE
       WHEN B.stars >= 2 AND B.stars < 4 THEN '2-3'
       WHEN B.stars >= 4 AND B.stars <= 5 THEN '4-5'
       WHEN B.stars >= 0 AND B.stars < 2 THEN '0-2'
   Else 'Error'
   END AS Stars_group
   ,
   CASE
       WHEN C.Category = 'Restaurants' THEN 'Restaurant'
   Else 'Las Vegas'
   END AS Selection_group
/*
   ,
   TRIM(SUBSTR(SUBSTR(REPLACE(hours,'|',' '), -11),1,5), '-') AS Begin
   ,TRIM(SUBSTR(hours,-5), '-') AS End
   ,CASE
       WHEN
       ABS(TRIM(SUBSTR(SUBSTR(REPLACE(hours,'|',' '), -11),1,5), '-') -
           TRIM(SUBSTR(hours,-5), '-')) = 0
       THEN 24
   Else
       ABS(TRIM(SUBSTR(SUBSTR(REPLACE(hours,'|',' '), -11),1,5), '-') -
           TRIM(SUBSTR(hours,-5), '-'))
```

```
   END
   AS Hours
   ,H.hours
*/
  ,SUM(
   CASE
       WHEN
       ABS(TRIM(SUBSTR(SUBSTR(REPLACE(hours,'|',' '), -11),1,5), '-') -
           TRIM(SUBSTR(hours,-5), '-')) = 0
       THEN 24
   Else
       ABS(TRIM(SUBSTR(SUBSTR(REPLACE(hours,'|',' '), -11),1,5), '-') -
           TRIM(SUBSTR(hours,-5), '-'))
   END
   )
   AS Hours
   ,COUNT(DISTINCT H.hours) AS Days
   ,COUNT(DISTINCT B.id) AS Businesses
--    ,B.id
--    ,B.name
FROM business AS B
INNER JOIN Hours AS H ON B.id = H.business_id
LEFT JOIN Category AS C ON B.id = C.business_id
WHERE
C.Category = 'Restaurants'
OR
City = 'Las Vegas'
GROUP BY Stars_group, Selection_group
ORDER BY Stars_group
```

Distribution of reviews per study group

```
SELECT
 CASE
     WHEN id IN (SELECT business_id FROM Category AS C WHERE Category = 'Restaurants')
AND City = 'Las Vegas' THEN 'Restaurant - Las Vegas'
     WHEN id IN (SELECT business_id FROM Category AS C WHERE Category = 'Restaurants')
THEN 'Restaurant'
     WHEN City = 'Las Vegas' THEN 'Las Vegas'
 Else 'Other'
 END AS Selection_group
 ,
 CASE
     WHEN stars >= 2 AND stars < 4 THEN '2-3'
```

```
     WHEN stars >= 4 AND stars <= 5 THEN '4-5'
     WHEN stars >= 0 AND stars < 2 THEN '0-2'
 Else 'Error'
 END AS Stars_group
 ,COUNT(id) AS Businesses
 ,SUM(review_count) AS Reviews
FROM business
WHERE
Selection_group != 'Other'
AND
Stars_group != '0-2'
GROUP BY Selection_group, Stars_group
```

Top 20 cities with more reviews

```
SELECT
 CASE
     WHEN stars >= 2 AND stars < 4 THEN '2-3'
     WHEN stars >= 4 AND stars <= 5 THEN '4-5'
     WHEN stars >= 0 AND stars < 2 THEN '0-2'
 Else 'Error'
 END AS Stars_group
 ,SUM(review_count) AS Reviews
 ,COUNT(id) AS Businesses
 ,City
FROM business
WHERE
Stars_group != '0-2'
GROUP BY City, Stars_group
ORDER BY Reviews DESC
LIMIT 20
```

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 open businesses have way more reviews than the ones that are closed.

```
+---------+---------+
| is_open | Reviews |
+---------+---------+
|       0 |   35261 |
|       1 |  269300 |
+---------+---------+
```

```
SELECT is_open, SUM(review_count) AS Reviews
FROM business
GROUP BY is_open

SELECT B.is_open, SUM(R.Useful), SUM(R.Funny), SUM(R.Cool)
FROM review AS R
LEFT JOIN business AS B ON R.business_id = B.id
LEFT JOIN user AS U ON R.user_id = U.id
GROUP BY B.is_open
```

```
+---------+---------------+--------------+-------------+
| is_open | SUM(R.Useful) | SUM(R.Funny) | SUM(R.Cool) |
+---------+---------------+--------------+-------------+
|    None |          9525 |         3872 |        4908 |
|       0 |            69 |           15 |          30 |
|       1 |           484 |          152 |         219 |
+---------+---------------+--------------+-------------+
```

ii. Difference 2: The open businesses have more tips than the ones that are closed

```
+---------+-------+------------+
| is_open | Likes | Tips_count |
+---------+-------+------------+
|       0 |     1 |         97 |
|       1 |     9 |        580 |
+---------+-------+------------+
```

```
SELECT B.is_open, SUM(likes) AS Likes, COUNT(*) AS Tips_count
FROM tip AS T
LEFT JOIN business AS B ON T.business_id = B.id
WHERE B.is_open IS NOT NULL
GROUP BY B.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.

i. Indicate the type of analysis you chose to do: Spot a good business opportunity studying the attributes and ratings of the actual businesses.

When there is a good average of reviews the probability of a good amount of clients increase and a good rating for a business could mean that is a service needed.

ii. Write 1-2 brief paragraphs on the type of data you will need for your analysis and why you chose that data: I'm gonna need a dataset with the business attributes, average of stars ratings and count of businesses for each attribute.

name	Stars_bin	Reviews_AVG	Business_count
BusinessAcceptsCreditCards	2-3	34.2580645161	62
RestaurantsPriceRange2	2-3	43.5909090909	44
BusinessParking	2-3	46.3414634146	41
BikeParking	2-3	50.2777777778	36
RestaurantsTakeOut	2-3	57.275862069	29
GoodForKids	2-3	55.75	28
RestaurantsGoodForGroups	2-3	54.8571428571	28
OutdoorSeating	2-3	66.8076923077	26
RestaurantsReservations	2-3	58.8076923077	26
RestaurantsDelivery	2-3	61.92	25
NoiseLevel	2-3	63.125	24
Ambience	2-3	65.652173913	23
HasTV	2-3	65.8260869565	23
RestaurantsAttire	2-3	60.2173913043	23
Alcohol	2-3	68.3181818182	22
GoodForMeal	2-3	62.7727272727	22
RestaurantsTableService	2-3	62.8181818182	22
WiFi	2-3	76.0909090909	22
ByAppointmentOnly	4-5	10.0	21
WheelchairAccessible	2-3	59.85	20
Caters	2-3	77.7368421053	19
AcceptsInsurance	4-5	12.0	8
DriveThru	2-3	41.2857142857	7
DogsAllowed	2-3	86.5	6
BusinessAcceptsBitcoin	4-5	6.4	5

There is a good opportunity of business on WIFI businesses because there is quite quantity of business registers, they have a good average of reviews and they are rated between 4-5 stars

Business Attributes

```
SELECT
  A.name
  ,
  CASE
    WHEN AVG(B.stars) >= 0 AND AVG(B.stars) < 2 THEN '1-2'
    WHEN AVG(B.stars) >= 2 AND AVG(B.stars) < 4 THEN '2-3'
```

```
    WHEN AVG(B.stars) >= 4 AND AVG(B.stars) <= 5 THEN '4-5'
   Else 'Error'
   END AS Stars_bin
   , AVG(review_count) AS Reviews_AVG
   , COUNT(DISTINCT B.id) AS Business_count
FROM business AS B
INNER JOIN Attribute AS A ON A.business_id = B.id
WHERE B.is_open = 1
GROUP BY A.name
ORDER BY Business_count DESC
```