Data Scientist Role Play

Profiling and Analyzing the Yelp Datasets

Here I have given a 2-part assignment. In the first part, I have asked a series of questions that will help me profile and understand the data just like a data scientist would.

In the second part of the assignment, I have asked to come up with my own inferences and analysis of the data for a particular research question. In order to do that I have prepared the dataset for the analysis.

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 = 10000(id)
ii. Hours = 1562(business_id)
iii. Category= 2643(business_id)
iv. Attribute= 1115(business_id)
v. Review= 10000(id),8090(business_id),9581(user_id)

vi.	Checkin=	493(business_id)
vii.	Photo=	10000(id),6493(business_id)
viii.	Tip=	537(user_id),3979(business_id)
ix.	User=	10000(id)
x.	Friend=	11(user_id)
xi.	Elite_years=	2780(user_id)

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 id, name, review_count, yelping_since, useful, funny, cool,
        fans, average_stars,
           compliment_hot, compliment_more, compliment_profile,
        compliment_cute, compliment_list,
           compliment_note, compliment_plain, compliment_cool,
        compliment_funny, compliment_writer, compliment_photos
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. Find the minimum, maximum, and average 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) from business group by city order by sum(review_count) desc ;

Copy and Paste the Result Below:

```
+-----------------+-------------------+
| city            | sum(review_count) |
+-----------------+-------------------+
| 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 |
+-----------------+-------------------+
```

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) from business where city='Avon' group by stars;

Copy and Paste the Result Below:

```
+-------------+-------+
| 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) from business where
city='Beachwood' group by stars;
```

Copy and Paste the Result Below:

```
+-------------+-------+
| 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 as users,sum(review_count) from user group by fans order by
review_count desc limit 3;
```

Copy and Paste the Result Below:

```
+--------+--------------+
| name   | review_count |
+--------+--------------+
| Gerald |         2000 |
| Sara   |         1834 |
| Yuri   |         1339 |
+--------+--------------+
```

8.Does posing more reviews correlate with more fans?

Please explain your findings and interpretation of the results:

SQL Code:

```
select name, fans,review_count from user order by fans desc
limit 10;
```

Result:

No no. of fans and reviews are not correlated as we can see in the following table below that is our output,for each user there is no relationship between no. of reviews and fans.

```
+-----------+-------------+------+
| name      | review_count | fans |
+-----------+-------------+------+
| Amy       |          609 |  503 |
| Mimi      |          968 |  497 |
| Harald    |         1153 |  311 |
| Gerald    |         2000 |  253 |
| Christine |          930 |  173 |
| Lisa      |          813 |  159 |
| Cat       |          377 |  133 |
| William   |         1215 |  126 |
| Fran      |          862 |  124 |
| Lissa     |          834 |  120 |
+-----------+-------------+------+
```

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

Answer:

SQL code used to arrive at answer:

```
select 'love' Word , count(text) as Count from review where text like
"%love%" union select 'hate' word, count(text) as count from review where
text like "%hate%";
```

```
+------+------------+
| Word |       Count|
+------+------------+
| hate |        232 |
| love |       1780 |
+------+------------+
```

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

SQL code used to arrive at answer:

```
select name as users,sum(fans) from user group by fans 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?

Answer: Yes

SQL code used for analysis:

```
select  stars,count(hours),city, neighborhood, address,hours,
case
when stars in (2.0,2.5,3.0,3.5) then 'lower' else 'upper' end comparison
from business inner join hours on business.id=hours.business_id where
city='Toronto' and stars>1.5 group by stars ;
```

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

Answer: No

SQL code used for analysis:

```
select stars,review_count,city, neighborhood, address,
case
when stars in (2.0,2.5,3.0,3.5) then 'lower' else 'upper' end comparison
from business where city='Avon' and stars>1.5 group by stars ;
```

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

Answer:

So, in our output star rating of customer of the US-based company called **Yelp** is different for different location of the city Toronto(The analysis is done for Toronto city ,any one can do the same for other city).

We can consider the following causes regarding the variation of star rating

- customer's behavior
- Various location
- different quality of day by day business service.

SQL code used for analysis:

```
select  stars,city, neighborhood, address,latitude,longitude,
case
when stars in (2.0,2.5,3.0,3.5) then 'lower' else 'upper' end comparison
from business where city='Toronto' and stars>1.5 group by stars ;
```

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.

Answer:

i. Difference 1:

The data is grouped in such a way that the ones that are open and the ones that are closed, then it can be noticed that how customer's behaviour is changed. In the distribution of star ratings, we can see business services of **Yelp** get highest rating from customer though their services are closed. Though no. of highest ratings such as 4-5 of closed services are less than that of open services

ii. Difference 2:

Average star rating & reviews of those business services **is_opened** is **3.6790** and **31.7570** respectively whereas, the same of **is_closed** is less than is_open that is average star rating and review are **3.5203** and **23.1980** respectively.

iii. Difference 3:

8480 business services are available of **Yelp** and **1520** are closed.

SQL code used for analysis:

For grouping:

```
select is_open,stars,review_count,city, case when is_open='1' then
'Yes' else 'No' end binary_decision from business;
```

i. **Difference 1**

Distribution Of stars_rating

For open service

```
select is_open,stars, count(stars), case when is_open='1' then 'Yes' else
'No' end binary_decision from business where is_open='1' group by stars;
```

For closed service

```
select is_open,stars, count(stars), case when is_open='1' then 'Yes' else
'No' end binary_decision from business where is_open='0' group by stars;
```

ii. **Difference 2**

For open service

```
select is_open, avg(stars),avg(review_count) from business where is_open='1';
```

For closed service

```
select is_open, avg(stars),avg(review_count) from business where is_open='0';
```

iii. **Difference 3**

For open service

```
select is_open,count(is_open) from business where is_open='1' group by is_open;
```

For closed service

```
select is_open,count(is_open) from business where is_open='0' group by is_open;
```

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

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

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

Correlation between star ratings and likes given by the consumer

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

I need two sources of data (tables). First, I join these two tables based on the tables named **users** and **business**. Then I sort them based on rating to see whether there is any correlation between the number of stars and likes.

The reason I chose this analysis and thus, the data sets is that psychologists have shown that how people think about something can completely change even after a few minutes and they think that how people think just after occurrence of an event is a better representative for the quality of that event compared to what they say after thinking about it. Because tip table is related to the occurrence of the event (shopping) and they write a review after hours or even days, comparing these two tables can help us to explore the validity what psychologists claim. As the result shows there is a slight correlation between the number of likes and stars, but this correlation is not strong. So what psychologists claim seems to be fairly valid.

iii. Output of your finished dataset:

SQL Code For Correlation between stars rating and likes given by users

```
select r.stars,t.likes from tip t inner join review r on r.user_id=t.user_id
order by t.likes desc;
```

Output

```
+-------+-------+ | stars | likes | +-------+-------+ | 3 | 2 | | 5 | 2 | | 4 | 1 | | 3 | 1
| | 3 | 1 | | 5 | 1 | | 4 | 1 | | 4 | 1 | | 4 | 1 | | 4 | 1 | | 4 | 1 | | 4 | 1 | | 4 | 1 |
| 4 | 1 | | 4 | 1 | | 4 | 1 | | 4 | 1 | | 4 | 1 | | 4 | 1 | | 4 | 1 | | 4 | 1 | | 4 | 1 | |
4 | 1 | | 5 | 1 | | 5 | 1 | +-------+-------+ (Output limit exceeded, 25 of 1227 total rows
shown)
```

Another example of how thoughts of people change after occurrence of any event like shopping,bike parking etc.

Case studies

There is no relation between stars rating and likes of user in case of bike parking. As how people think about something can completely change even after a few minutes so before the occurrence of the event it seems that that would provide very good services but after the event like after parking bike user’s view was completely changed and this view would be the best representative of that event compared what they said before the occurrence of the event. That’s why there is no correlation between stars which was given before the event is happened and likes which was given after the occurrence of that event.

SQL Code

```
select r.stars,t.likes ,r.useful ,a.name,case when r.useful='1' then 'useful'
else 'not useful' end  binary_decision from review r inner join tip t on
r.business_id=t.business_id inner join attribute a on
a.business_id=r.business_id where a.name like  '%parking%';
```

Output

```
      +-------+-------+--------+----------------+----------------+ | stars | likes |
useful | name | binary_decision | +-------+-------+--------+----------------
+----------------+ | 4 | 0 | 0 | BikeParking | not useful | | 4 | 0 | 0 | BusinessParking |
not useful | | 1 | 0 | 1 | BikeParking | useful | | 1 | 0 | 1 | BikeParking | useful | | 1 |
0 | 1 | BikeParking | useful | | 1 | 0 | 1 | BikeParking | useful | | 1 | 0 | 1 |
BikeParking | useful | | 1 | 0 | 1 | BikeParking | useful | | 1 | 0 | 1 | BikeParking |
useful | | 1 | 0 | 1 | BikeParking | useful | | 1 | 0 | 1 | BikeParking | useful | | 1 | 0 |
1 | BikeParking | useful | | 1 | 0 | 1 | BusinessParking | useful | | 1 | 0 | 1 |
BusinessParking | useful | | 1 | 0 | 1 | BusinessParking | useful | | 1 | 0 | 1 |
BusinessParking | useful | | 1 | 0 | 1 | BusinessParking | useful | | 1 | 0 | 1 |
BusinessParking | useful | | 1 | 0 | 1 | BusinessParking | useful | | 1 | 0 | 1 |
BusinessParking | useful | | 1 | 0 | 1 | BusinessParking | useful | | 1 | 0 | 1 |
BusinessParking | useful | | 5 | 0 | 0 | BikeParking | not useful | | 5 | 0 | 0 |
BikeParking | not useful | | 5 | 0 | 1 | BikeParking | useful | +-------+-------+--------
+----------------+----------------+ (Output limit exceeded, 25 of 38 total rows shown)
```

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

Usefulness, user’s feedback of business for a particular category.How people like that particular category.

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

Here we do inner join operation to the tables review and category in order to analyze how people of US like Korean food.In review table they leaf feedback and help us for our analysis purpose.

Here we can see there is a slight correlation between stars rating and the review what users leaves for us in order to analyze the service quality of this company.If we do a sentiment analysis using this reviews it will be very interesting.Here we can also notice that there is no relation between useful of the reviews and stars rating.

SQL Code and Output

```
      select r.useful,c.category,r.stars,r.date,r.cool,r.funny,r.text from category
      c inner join review r on     c.business_id=r.business_id where
      c.category='Korean';
```

```
+--------+----------+-------+--------------------+------+-------
+-----------------------------------------------------------------------------------------
-----------------------------------------------------------------------------------------
-----------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------+
| useful | category | stars | date | cool | funny | text | +--------+----------+-------
+--------------------+------+-------
+-----------------------------------------------------------------------------------------
-----------------------------------------------------------------------------------------
-----------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------+
| 0 | Korean | 4 | 2016-06-04 00:00:00 | 0 | 0 | I like the food, it was really fresh. The
service was also on point and it was a full house. The beef is really good, any of them. The
fried rice feeds about 3 ppl. I love this place, I come here often. And always bring ppl
here who come to Vegas to visit, they all love it. The price is also very reasonable. For 2
ppl I usually spend about $40 but leave STUFFED. | | 0 | Korean | 4 | 2017-06-27 00:00:00 |
0 | 0 | one of the best korean bbq place we have been. portions are enough, side dishes were
great. place is kinda small and could get really crowded. i little bit expensive but its
worth it. we got free dessert just by checking in.i wish the serves would be more friendly
and attentive. meat has good quality. | | | | | | | | | | | | | | | will definitely come
```

```
back. | | 0 | Korean | 4 | 2015-06-04 00:00:00 | 0 | 0 | Our first Korean BBQ meal in Vegas,
found it in Yelp rated almost five star. The place looks very clean, and the waitress is so
nice! We ordered beef tongue, pork belly and other meat. The meat tasted awesome!!!! | | 0 |
Korean | 5 | 2017-01-16 00:00:00 | 0 | 0 | Atmosphere is amazing! Our server was the best!
Prices are defiantly reasonable compared to everything else and food was to die for ...
Defiantly will be coming back next time I'm in Vegas. | | 0 | Korean | 5 | 2016-04-24
00:00:00 | 0 | 0 | Awsome awsome place to eat highly recomend there banana deserve was
delicious. Loved the BBQ here I'm in love | | 0 | Korean | 5 | 2016-02-17 00:00:00 | 0 | 0 |
Great food, comfortable atmosphere and very friendly staff - kudos to "Back" and "Frank" who
were very friendly and kind! Highly recommend this restaurant! | | 0 | Korean | 4 | 2014-01-
24 00:00:00 | 0 | 0 | My first experience with Korean BBQ. I was a bit intimidated at first.
The place was busy and my rookie flag was flapping wildly. I was able to get through the
ordering process with the help of a super friendly waitress that set me at ease from word
go. Happy hour prices made my experience even more palatable. | | 0 | Korean | 5 | 2015-12-
27 00:00:00 | 0 | 0 | The food was amazing as well as the customer service. The wait was so
worth it. I'd come back everyday if I could! | | 0 | Korean | 5 | 2016-04-04 00:00:00 | 0 |
0 | Great food! Great prices! Great service! One of the beat Korean BBQ places that I've
been to. | +--------+----------+-------+--------------------+------+-------
+------------------------------------------------------------------------------------------
------------------------------------------------------------------------------------------
------------------------------------------------------------------------------------------
-----------------------------------------------------------------------------------------+
```

3)i.Indicate the type of analysis you chose to do

How stars ratings of the coustomers based on services and usefulness of user's feedback are correlated.

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

There is a week correlation between ratings and usefulness.Here we join two tables review and user using inner join in order to analyze the customers' feedback.

SQL Code

```
select r.stars,u.useful,u.review_count from   review r  inner join user u on r.user_id=u.id;
```

Output

```
+-------+--------+--------------+
| stars | useful | review_count |
+-------+--------+--------------+
|     2 |     17 |           71 |
|     5 |      1 |           26 |
|     5 |      0 |            1 |
|     3 |   2654 |          196 |
|     2 |   1402 |          279 |
|     3 |     37 |            8 |
|     1 |      0 |           10 |
|     4 |     38 |          564 |
|     5 |      2 |            8 |
|     4 |      4 |            8 |
|     4 |     72 |          174 |
|     1 |      2 |           13 |
|     5 |      0 |           35 |
|     5 |      4 |          676 |
|     5 |      3 |           96 |
|     4 |      6 |           10 |
|     5 |      0 |           12 |
|     4 |      8 |           83 |
|     5 |   6974 |          198 |
|     4 |    154 |           44 |
|     5 |      5 |           16 |
|     5 |      0 |            7 |
|     4 |      0 |            7 |
|     4 |     21 |          109 |
|     3 |     30 |          235 |
+-------+--------+--------------+
(Output limit exceeded, 25 of 72 total rows shown)
```

4)i.Indicate the type of analysis you chose to do

Here we want to know how the quality of business services for specific category varies with respect to different location of different cities and how many hours they provide services.

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

Here we consider “shopping” category ,anyone can select another category.From our output we can conclude that rating is high for the city **Pittsburgh** and average review is high for the city **Las Vegas** and the customers Services are opened for 10-12 hours(it can be changed for different service provider) except in Mesa where services are closed. Star rating is weekly correlated with average review.

SQL code

Category: Shopping

```
select b.nameas as 'Service Provider'
,b.city
,b.latitude
,b.longitude
,b.address
,b.stars,avg(b.stars)
,avg(b.review_count)
,MAX(CASE
                     WHEN h.hours LIKE "%monday%" THEN
TRIM(h.hours,'%MondayTuesWednesThursFriSatSun|%')
                     END) AS monday_hours,
                     MAX(CASE
                     WHEN h.hours LIKE "%tuesday%" THEN
TRIM(h.hours,'%MondayTuesWednesThursFriSatSun|%')
                     END) AS tuesday_hours,
                     MAX(CASE
                     WHEN h.hours LIKE "%wednesday%" THEN
TRIM(h.hours,'%MondayTuesWednesThursFriSatSun|%')
                     END) AS wednesday_hours,
                     MAX(CASE
                     WHEN h.hours LIKE "%thursday%" THEN
TRIM(h.hours,'%MondayTuesWednesThursFriSatSun|%')
                     END) AS thursday_hours,
                     MAX(CASE
                     WHEN h.hours LIKE "%friday%" THEN
TRIM(h.hours,'%MondayTuesWednesThursFriSatSun|%')
                     END) AS friday_hours,
                     MAX(CASE
                     WHEN h.hours LIKE "%saturday%" THEN
TRIM(h.hours ,'%MondayTuesWednesThursFriSatSun|%')
                     END) AS saturday_hours,
                     MAX(CASE
                     WHEN h.hours LIKE "%sunday%" THEN
TRIM(h.hours,'%MondayTuesWednesThursFriSatSun|%')
                     END) AS sunday_hours,c.category,b.is_open,case when
is_open='1' then 'Open' else 'Close' end Open_or_close from business b inner join
category c on b.id=c.business_id inner join hours h on h.business_id=b.id INNER
JOIN attribute A ON B.id = A.business_id
            where c.category like 'shopping%' group by b.city;
```

Output

```
---------------------------------+--------------+----------+-----------
+-----------------------------+-------+---------------+---------------------+--------------
+---------------+-----------------+----------------+--------------+----------------
+--------------+----------+---------+---------------+ | Service Provider | city | latitude |
longitude | address | stars | avg(b.stars) | avg(b.review_count) | monday_hours |
tuesday_hours | wednesday_hours | thursday_hours | friday_hours | saturday_hours |
sunday_hours | category | is_open | Open_or_close | +----------------------------------
+--------------+----------+-----------+-----------------------------+-------+---------------
+---------------------+--------------+---------------+-----------------+----------------
+--------------+----------------+--------------+------------+-------+---------------+ |
Springmaster Garage Door Service | Chandler | 33.3199 | -111.81 | 1909 E Ray Rd, Ste 9-170 |
5.0 | 4.85714285714 | 5.0 | 9:00-20:00 | 9:00-20:00 | 9:00-20:00 | 9:00-20:00 | 9:00-20:00 |
9:00-20:00 | 5:00-0:00 | Shopping | 1 | Open | | HighLife North Tryon | Charlotte | 35.3167
| -80.7405 | 9605 N Tryon St, Ste C | 4.0 | 3.73076923077 | 6.06593406593 | 9:00-19:00 |
9:00-19:00 | 9:00-19:00 | 9:00-19:00 | 9:00-19:00 | 9:00-17:00 | 12:00-21:00 | Shopping | 1
| Open | | Red Rock Canyon Visitor Center | Las Vegas | 36.1357 | -115.428 | 1000 Scenic
Loop Dr | 4.5 | 4.5 | 32.0 | 8:00-16:30 | 8:00-16:30 | 8:00-16:30 | 8:00-16:30 | 8:00-16:30
| 8:00-16:30 | 8:00-16:30 | Shopping | 1 | Open | | Ghost Armor SS Springs | Mesa | 33.3906
| -111.69 | 6555 E Southern Ave | 2.0 | 2.0 | 3.0 | 10:00-21:00 | 10:00-21:00 | 10:00-21:00
| 10:00-21:00 | 10:00-21:00 | 10:00-21:00 | 11:00-18:00 | Shopping | 0 | Close | | Standard
Restaurant Supply | Phoenix | 33.4664 | -112.018 | 2922 E McDowell Rd | 3.5 | 3.5 | 15.0 |
8:00-18:00 | 8:00-18:00 | 8:00-18:00 | 8:00-18:00 | 8:00-18:00 | 9:00-17:00 | None |
Shopping | 1 | Open | | PRO BIKE+RUN | Pittsburgh | 40.4521 | -80.165 | 3100 Robinson Ln |
5.0 | 5.0 | 8.0 | 10:00-20:00 | 10:00-20:00 | 10:00-20:00 | 10:00-20:00 | 10:00-20:00 |
10:00-18:00 | 12:00-17:00 | Shopping | 1 | Open | | Alterations Express | Strongsville |
41.3141 | -81.8207 | 17240 Royalton Rd | 4.0 | 4.0 | 3.0 | 8:00-19:00 | 8:00-19:00 | 8:00-
19:00 | 8:00-19:00 | 8:00-19:00 | 8:00-18:00 | None | Shopping | 1 | Open | | Bobs Smoke
Shop | Tempe | 33.408 | -111.91 | 1740 E Broadway Rd, Ste 102 | 3.5 | 3.5 | 3.0 | 9:30-22:00
| 9:30-22:00 | 9:30-22:00 | 9:30-22:00 | 9:30-22:00 | 9:30-22:00 | 9:30-22:00 | Shopping | 1
| Open | | Gussied Up | Toronto | 43.6727 | -79.4142 | 1090 Bathurst St | 4.5 | 4.5 | 6.0 |
None | 11:00-19:00 | 11:00-19:00 | 11:00-19:00 | 11:00-19:00 | 11:00-17:00 | 12:00-16:00 |
Shopping | 1 | Open | +----------------------------------+--------------+----------
+-----------+----------------------------+-------+---------------+---------------------
```

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

Objective

Here we want to know whether people like chinese food and how many people,that will also give an effect on star ratings ,more stars rating is also a results of good service of Yelp

Analysis

Here we have to collect all chinese restaurants in different cities .Here we have calculated average stars rating and average of no. of reviews given by customers regarding the services.

In output we can see there is only one chinese restaurant in the corresponding cities except in Edinburgh restaurants is closed and in Las Vegas restaurants got highest average stars and reviews.

Here we can also notice that a moderate positive correlation between average ratings and average reviews in respective cities.

SQL Code

```
SELECT c.category,COUNT(b.name) AS Number_Of_Resturants, AVG(stars),
AVG(review_count),b.city,b.is_open,case when is_open='1' then 'Open' else 'Closed'
end Restuarant_is_open_or_not
FROM business b INNER JOIN category c ON c.business_id = b.id
WHERE c.category like "chin%"
GROUP BY b.city
ORDER BY b.city DESC;
```

Output

```
+----------+---------------------+-----------+-----------------+---------------
+---------+-------------------------+ | category | Number_Of_Resturants | AVG(stars) |
AVG(review_count) | city | is_open | Restuarant_is_open_or_not | +----------
+---------------------+-----------+-----------------+---------------+---------
+-------------------------+ | Chinese | 1 | 1.5 | 4.0 | Toronto | 1 | Open | | Chinese | 1
| 4.0 | 768.0 | Las Vegas | 1 | Open | | Chinese | 1 | 3.5 | 21.0 | Fountain Hills | 1 |
Open | | Chinese | 1 | 3.5 | 3.0 | Edinburgh | 0 | Closed | +----------
+---------------------+-----------+-----------------+---------------+---------
+-------------------------+
```