

Exploring Yelp Reviews Data: Insights and Visualizations using OSEM Process

Rahul Kumar Nalubandhu

Introduction

Conducting a review analysis gives businesses the opportunity to improve customer experience, identify service gaps and gain real time insights amongst many other benefits. This analysis dives into customer reviews entered in Yelp for businesses in different states throughout the United States. In this project I have 3 main questions.

- 1) What is the distribution of the number of reviews for businesses in the dataset, and what can we conclude about the central tendency and spread of the data based on the mean and standard deviation?
- 2) How does the average and total log-scaled review count of businesses vary by state in the Yelp data?
- 3) What is the relationship between the average stars of a restaurant and the number of reviews it has received?
 - Do restaurants with more reviews have a higher average stars or is there no significant relationship?

Problem

The issue with having little to no customer reviews is that this can negatively impact sales. Customer reviews are an important channel to attract customers and increase sales. A benefit in analyzing reviews at the business, city and state level will provide insight into which states and/or businesses have the least customer engagement and allow for proper intervention.

Obtain Data

Yelp is a one-stop platform which enables customers to connect with businesses. More than 80 million people visit this platform in a month to find businesses and service providers. Customers are given the ability to leave reviews and request quotes from local businesses amongst many other things. In return, local business owners are given the ability to communicate with their customers and respond to reviews to build trust with their customers. The customer review data set is acquired directly through Yelp. The data set is 4.04GB (1 point) and split into multiple JSON files (2 points) which contain businesses, reviews, and user data. In addition, the data has punctuation (1 point) and has more than one type of related data (2 points). Based on the point system requirements provided, the yelp data is a 6-point data set.

Scrub Data

Prior to conducting the analysis for the yelp data set, the data is extracted and consolidated into a CSV and JSON files for the following fields: business name, id, city, state, and review count, stars.

OSEMN Report Midterm Project

Visual studio is the primary application utilized to read and extract data in the python programming language for the first two question of interests and for the 3rd question I divided the data into 2 different tables as .csv files and imported them into phpMyAdmin and wrote a SQL Query to join them and generate data to answer the 3rd question. In addition, all null values are deleted from the data set. Null values are deleted to not compromise the integrity of the analysis.

Explore Data

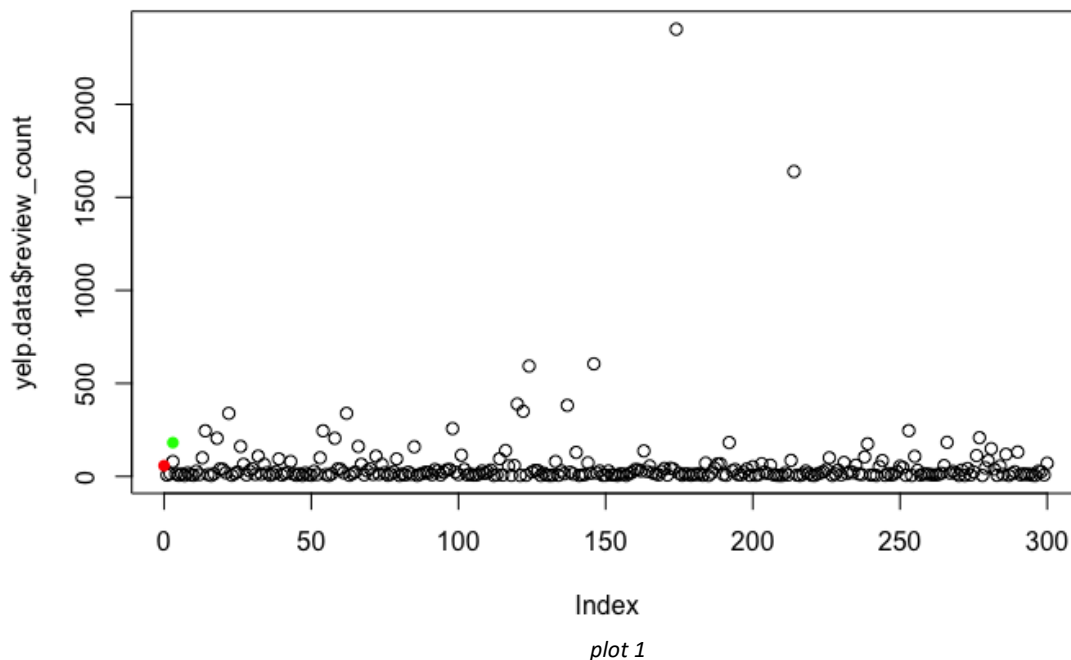
For the 1st two questions, the exploration stage, the csv data is loaded into R-studio. To begin the exploration stage the review count field is plotted.

Analysis and visualizations for 1st question:

Question of interest?

What is the distribution of the number of reviews for businesses in the dataset, and what can we conclude about the central tendency and spread of the data based on the mean and standard deviation?

The generated plot displays the distribution of the number of reviews for the dataset's businesses. Where x-axis is no of business and y axis is review_count of those business. Most of the businesses have fewer than 200 reviews, with a few outliers having more. The median number of reviews, which is around 57.67333, is represented by the red point on the plot. The standard deviation, or green point, is around 181.3557. We may infer from this plot that the majority of the dataset's firms have a relatively low number of reviews, while a few outliers have a high number. We can get a sense of the data's central tendency and distribution from the mean and standard deviation.



OSEMN Report Midterm Project

What is the use of finding mean and standard deviation ?

To visualize the central tendency and distribution of the data, the mean and standard deviation of the review count are computed and plotted. While the standard deviation provides an idea of how much variety there is in the review counts, the mean provides an estimate of the "average" number of reviews for businesses in the dataset. When taken as a whole, these statistics can shed light on the distribution of review counts and the degree of business fluctuation.

Analysis and visualizations for 2nd question:

Question of interest?

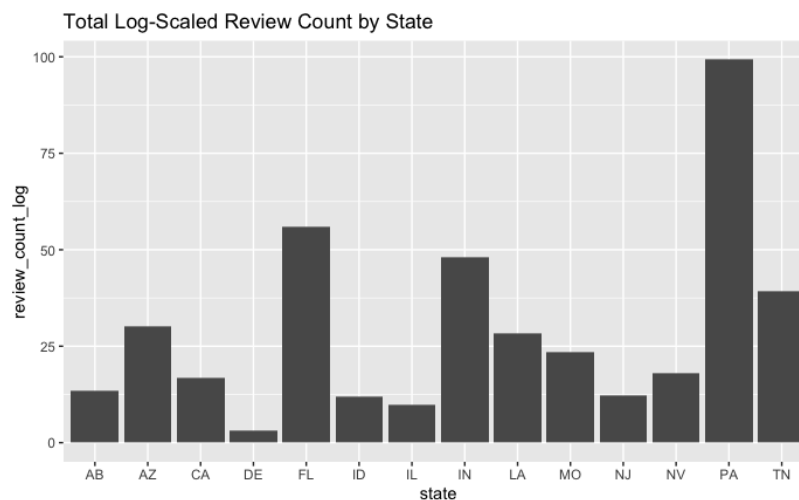
How does the average and total log-scaled review count of businesses vary by state in the Yelp data?

In this two bar charts, one is representing the total log-scaled review count of businesses by state and another representing the average log-scaled review count of businesses by state. The x-axis in both charts shows the states and the y-axis shows the log-scaled review count. By observing these plots, we can conclude that the log-scaled review count of businesses varies by state. Some states have higher total and average log-scaled review count compared to others, indicating a higher concentration of businesses with more reviews in those states.

Why are we doing log-scaling here?

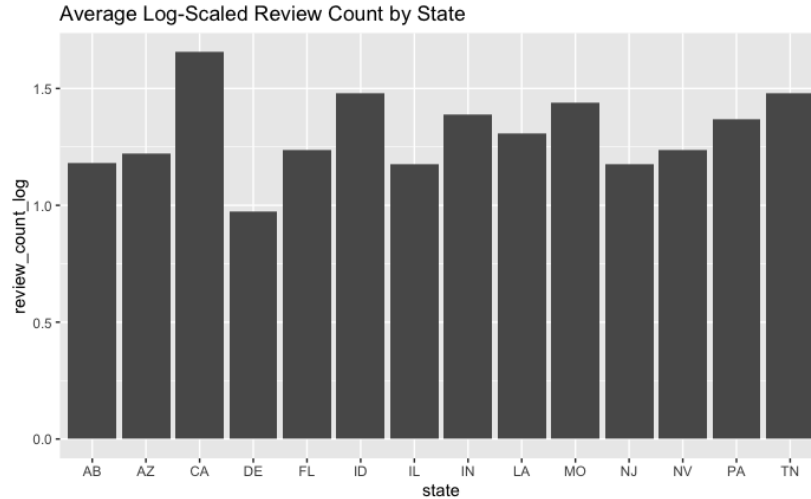
By using the log scaling on the review count data, it can provide a more informative representation of the distribution of review count values, especially if the original data has a skewed distribution.

The log scaling helps to compress the differences in the review count between states with a large difference in their review count, making it easier to see differences in the total review count among states with smaller review counts.



Graph 1

OSEM Report Midterm Project



Graph 2

Analysis and visualizations for 3rd question:

What is the relationship between the average stars of a restaurant and the number of reviews it has received?

- Do restaurants with more reviews have a higher average stars or is there no significant relationship?

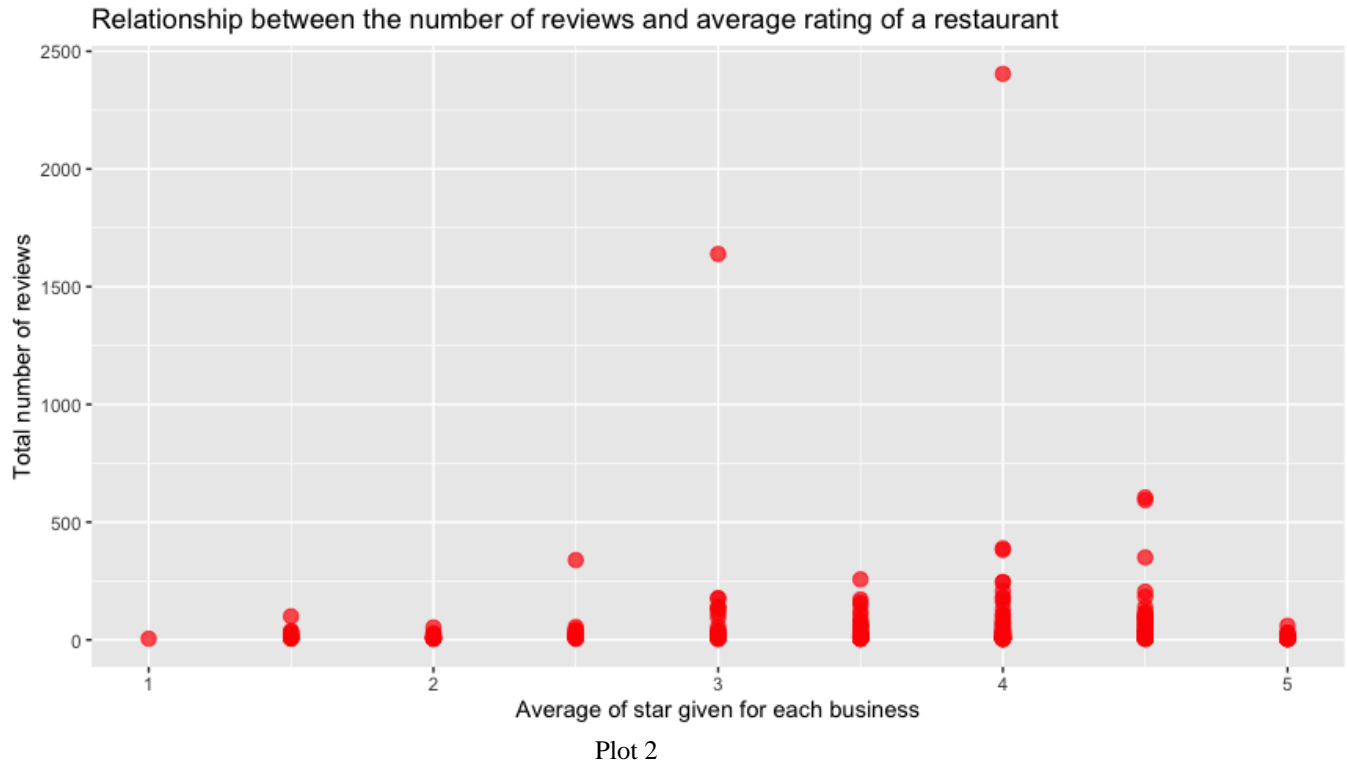
For this question, you will need two data tables :

Restaurant information table - This table should contain information such as the restaurant name, id, state, city. etc.

Review information table - This table should contain information such as the stars, number of reviews, business_id for each restaurant. we will then need to join the two tables using relational database and let's make the business_id as it is the primary key, hence you can access both the number of reviews and the average stars for each restaurant.

From there, you can create a scatter plot (plot2) showing the relationship between the number of reviews and average stars for each restaurant.

OSEMN Report Midterm Project



Hence From the generated plot, you can draw the following conclusions:

1. There doesn't seem to be a clear linear relationship between the number of reviews and the average stars of a restaurant. Some restaurants have many reviews but a relatively low stars, while others have fewer reviews but a higher average stars.
2. There is a large spread of data points, which means that there is a lot of variability in the relationship between the number of reviews and the average stars of restaurants.
3. There is a cluster of points near the 4.0 – 4.5-star mark with a high number of reviews. This suggests that there may be a relationship between high numbers of reviews and a high average stars.

Overall, the graph suggests that while there may be some correlation between the number of reviews and the average stars of a restaurant, it is not a strong or straight forward relationship.

Therefor we cannot still completely answer the question without finding the correlation hence the correlation coefficient can be found in r using the formulae :

```
correlation <- cor(data$ReviewCount, data$AverageStars)
```

```
correlation = 0.03157684
```

But why are we finding the correlation here?

Because the correlation coefficient measures the strength and direction of a linear relationship between two variables. In the context of the relationship between the number of reviews and the average stars of a restaurant, the correlation coefficient can help you determine if there is a strong positive, strong negative, or weak relationship between these two variables.

1. If the correlation coefficient is positive and close to 1, this suggests that there is a strong positive relationship between the number of reviews and the average rating of a restaurant, meaning that as the number of reviews increases, the average rating tends to increase as well.
2. On the other hand, if the correlation coefficient is negative and close to -1, this suggests that there is a strong negative relationship between the number of reviews and the average rating, meaning that as the number of reviews increases, the average rating tends to decrease.
3. If the correlation coefficient is close to 0, this suggests that there is no clear relationship between the number of reviews and the average rating.

Knowing the correlation coefficient can help you make informed conclusions about the relationship between these two variables and can inform future decisions, such as how to allocate resources to improve a restaurant's rating.

so, as we got the correlation coefficient as 0.03157684 which is considered a very weak or close to no linear relationship between two variables. This means that there is not a strong relationship between the number of reviews and the average stars of a restaurant, and that changes in one variable do not consistently predict changes in the other variable.

Model Data

After reviewing the dataset, a decision tree is the most suitable modeling technique for this analysis. A decision tree will contain the different factors that determine if businesses are likely to face less customer reviews due to the state they are in or even the demographics of their target customers. While this is deemed the best approach for this analysis, more data needs to be gathered and analyzed to properly complete this.

Interpret Data

For the 1st analysis, we can interpret that based on the generated plot of the number of reviews for businesses in the Yelp dataset, we can observe that most of the businesses have a relatively low number of reviews, with only a few outliers having significantly higher numbers of reviews. The median number of reviews is approximately 57.67, indicating that half of the businesses have fewer reviews than this value, while the other half has more. The standard deviation, which is around 181.36, implies that there is considerable variation in the review counts of businesses in

OSEMN Report Midterm Project

the dataset. By using the mean and standard deviation, we can visualize the central tendency and distribution of the data, which can help us understand the degree of business fluctuation in the Yelp dataset.

From the second analysis, we can interpret that the log-scaled review count of businesses varies by state and some states have higher total and average log-scaled review counts than others. The use of log scaling on the review count data has allowed for a better representation of the distribution of review count values, which helps in identifying states with high and low review counts. The bar charts have provided a visual representation of the data that is easy to understand and interpret, making it easier to identify patterns in the data. Overall, this analysis has helped us to gain insights into how the review count of businesses varies by state in the Yelp data.

In the 3rd analysis, we can interpret as the correlation coefficient is near to zero which means that it indicates a very weak positive relationship between the two variables, meaning that as the number of reviews a restaurant has received increases, its average stars also slightly increase. However, this relationship is not strong enough to conclude that restaurants with more reviews have a higher average stars. Therefore, it can be concluded that there is no significant relationship between the average stars of a restaurant and the number of reviews it has received.

Process how I loaded data into database:

1. The first step was to extract the relevant data from the original Yelp dataset and convert it into two separate .csv files, containing the required fields.
2. The next step was to import the two .csv files into phpMyAdmin.
3. Within phpMyAdmin, a new table was created with headers for each respective column. During the import of the data, the header row was excluded, and appropriate data types were assigned to each column. The business_id column was also set as the primary key.
4. The above step was repeated for the second .csv file containing business_id and review count data.
5. A SQL query was then written to join the two tables using the primary key.
6. After running the SQL query, the resulting data was exported as a .csv file and loaded into R-studio for visualizations.

OSEMN Report Midterm Project

Appendix

Initial Data:

```
begin.py 1 yelp_academic_dataset_business.json Settings
() yelp_academic_dataset_business.json
1 {"business_id": "Pns214m5f08Kk3d1x6A6A", "name": "Abby Rappoport, LAC, CMQ", "address": "1616 Chapala St, Ste 2", "city": "Santa Barbara", "state": "CA", "postal_code": "93101", "latitude": 34.4266787, "longitude": -119.7111968, "stars": 5.0, "review_count": 122, "is_open": 10, "business_id": "mpf3x-BjTdTfEAsyCzrAYpV", "name": "The UPS Store", "address": "87 Grasso Plaza Shopping Center", "city": "Affton", "state": "MO", "postal_code": "63123", "latitude": 38.551126, "longitude": -90.335695, "stars": 3.0, "review_count": 10, "business_id": "TUPRk1rK1K1_TansVW1NQ", "name": "Target", "address": "5255 E Broadway Blvd", "city": "Tucson", "state": "AZ", "postal_code": "85711", "latitude": 32.223236, "longitude": -110.880452, "stars": 3.5, "review_count": 180, "business_id": "TNTSWMcQd7CvYtjgqe9mV", "name": "St Honore Pastries", "address": "935 Race St", "city": "Philadelphia", "state": "PA", "postal_code": "19107", "latitude": 39.9555052, "longitude": -75.1555641, "stars": 4.0, "review_count": 18, "business_id": "mM6c6_wtDE0EUBK1GX0VfA", "name": "Perklemen Valley Brewery", "address": "101 Walnut St", "city": "Green Lane", "state": "PA", "postal_code": "18054", "latitude": 40.3381827, "longitude": -75.4716585, "stars": 4.5, "review_count": 10, "business_id": "n_BuPQx1shNbnPUS10dU8U", "name": "Famous Footwear", "address": "8522 Eager Road, Dierbergs Brentwood Point", "city": "Brentwood", "state": "MO", "postal_code": "63144", "latitude": 38.627695, "longitude": -90.340465, "stars": 2.0, "review_count": 19, "business_id": "bB00EgKFA10tX9Lf6ZBUZ", "name": "Sonic Drive-In", "address": "615 S Main St", "city": "Ashland City", "state": "TN", "postal_code": "37015", "latitude": 36.269593, "longitude": -87.058943, "stars": 2.0, "review_count": 10, "business_id": "qKRM_ZX51Yqk3btLwA01Q", "name": "Temple Beth-El", "address": "400 Pasadena Ave S", "city": "St. Petersburg", "state": "FL", "postal_code": "33707", "latitude": 27.76659, "longitude": -82.732983, "stars": 3.0, "review_count": 19, "business_id": "U3uofvbfv7f0NHEwAdVAKJJA", "name": "Marshalls", "address": "21785 Village Lakes Sc Dr", "city": "Land O' Lakes", "state": "FL", "postal_code": "33639", "latitude": 28.1904587953, "longitude": -82.4573802199, "stars": 3.5, "review_count": 10, "business_id": "eU5GskpB9bUA71D0MRK", "name": "Vietnamese Food Truck", "address": "", "city": "Tampa Bay", "state": "FL", "postal_code": "33602", "latitude": 27.9552692, "longitude": -82.4563199, "stars": 4.0, "review_count": 10, "business_id": "1l_R0bJpUfHresjy9EGmG", "name": "Denny's", "address": "8901 US 31 S", "city": "Indianapolis", "state": "IN", "postal_code": "46227", "latitude": 39.6371332838, "longitude": -86.127217412, "stars": 2.5, "review_count": 128, "business_id": "jAw5mZm8p003x0J181G", "name": "Adams Dental", "address": "15 N Missouri Ave", "city": "Clearwater", "state": "FL", "postal_code": "33755", "latitude": 27.966235, "longitude": -82.787412, "stars": 5.0, "review_count": 10, "business_id": "00sL1k00h00511_E0m0", "name": "Zio's Italian Market", "address": "2575 E Bay Dr", "city": "Largo", "state": "FL", "postal_code": "33771", "latitude": 27.9161159, "longitude": -82.7604608, "stars": 4.5, "review_count": 100, "business_id": "NUTTq6u9gRbD1180R0A", "name": "Tuna Bay", "address": "203 Race St", "city": "Philadelphia", "state": "PA", "postal_code": "19106", "latitude": 39.953949, "longitude": -75.143262, "stars": 4.0, "review_count": 245, "business_id": "rEby_1lU8y8gghlyb7kA", "name": "Arizona Truck Outfitters", "address": "625 N Stone Ave", "city": "Tucson", "state": "AZ", "postal_code": "85705", "latitude": 32.2208719, "longitude": -110.9723419, "stars": 4.5, "review_count": 14, "business_id": "N0XSSMqRAs0h0b01p0A", "name": "Herb Import Co", "address": "712 Adams St", "city": "New Orleans", "state": "LA", "postal_code": "70118", "latitude": 29.9414679565, "longitude": -90.129952757, "stars": 4.0, "review_count": 5, "business_id": "8uG1S7YVKE2TSn3C0M0K", "name": "Nitty Car Rental", "address": "1241 Airline Dr", "city": "Kenner", "state": "LA", "postal_code": "70062", "latitude": 29.981183, "longitude": -90.2540123, "stars": 3.5, "review_count": 14, "business_id": "R0eac1Q4eh08q7f6TcQ", "name": "BAP", "address": "1224 South St", "city": "Philadelphia", "state": "PA", "postal_code": "19147", "latitude": 39.943223, "longitude": -75.162568, "stars": 4.5, "review_count": 1205, "business_id": "M0DjwqntZKzAae7R0MPEJA", "name": "Rost Coffeehouse and Wine Bar", "address": "10359 104 Street NW", "city": "Edmonton", "state": "AB", "postal_code": "T5J 1B9", "latitude": 53.5460453, "longitude": -113.4991693, "stars": 4.0, "business_id": "qH0d0E1Kx4jYzKyzwPNS3", "name": "Barnes & Noble Booksellers", "address": "8675 River Crossing Blvd", "city": "Indianapolis", "state": "IN", "postal_code": "46240", "latitude": 39.91382, "longitude": -86.105739, "stars": 4.0, "business_id": "kFmV-J2pU6TWN50h0hK1Z", "name": "Hibachi Express", "address": "6625 E 82nd St", "city": "Indianapolis", "state": "IN", "postal_code": "46250", "latitude": 39.4761165, "longitude": -86.8530799, "stars": 4.0, "business_id": "90G5YX1q2G6e2M8A6K1Z", "name": "Romano's Macaroni Grill", "address": "5585 S Virginia St", "city": "Reno", "state": "NV", "postal_code": "89502", "latitude": 39.4761165, "longitude": -119.7893392, "stars": 2.5, "business_id": "41RzR70a5-QaXuvYxEGKG", "name": "Super Dog", "address": "1168 Gallatin Pike S", "city": "Nashville", "state": "TN", "postal_code": "37115", "latitude": 36.248596, "longitude": -86.7199857, "stars": 4.0, "business_id": "P50_C15fA133Hj2VM6Z1Q", "name": "Indian Walk Veterinary Center", "address": "662 Durham Rd", "city": "Newtown", "state": "PA", "postal_code": "18940", "latitude": 40.2734926, "longitude": -74.987496, "stars": 5.0, "business_id": "noByWYNDLQAR9CqkqfDQ", "name": "H&M", "address": "827-833 State St", "city": "Santa Barbara", "state": "CA", "postal_code": "93101", "latitude": 34.4202089255, "longitude": -119.7004597389, "stars": 3.0, "business_id": "tM0wMmFUEXRC9Zduonp7g", "name": "The Green Pheasant", "address": "215 1st Ave S", "city": "Nashville", "state": "TN", "postal_code": "37201", "latitude": 36.1598858, "longitude": -86.7731974, "stars": 4.0, "business_id": "QDN728WofFYKdJhh1S7g", "name": "Bar One", "address": "767 S 9th St", "city": "Philadelphia", "state": "PA", "postal_code": "19147", "latitude": 39.9398245785, "longitude": -75.1574465632, "stars": 4.0, "business_id": "sqSqqLY8N8nZ179c0chZyV", "name": "Domino's Pizza", "address": "3801 Highway 31 W", "city": "White House", "state": "TN", "postal_code": "37188", "latitude": 36.464746722, "longitude": -86.6591868908, "stars": 3.5, "business_id": "fVmn0Z0X0bj2179c0chZyV", "name": "Altitude Trampoline Park - Boise", "address": "1301 N Milwaukee St", "city": "Boise", "state": "ID", "postal_code": "83704", "latitude": 43.616735, "longitude": -116.2853821, "stars": 5.0, "business_id": "Mj0oz24M9N1Be10JkLd_Q", "name": "DeSandro on Main", "address": "4105 Main St", "city": "Philadelphia", "state": "PA", "postal_code": "19127", "latitude": 40.8224662, "longitude": -75.218314, "stars": 3.0, "business_id": "8SshLB4U7e0eUdVtJWnpA", "name": "Danceline", "address": "30 W Lancaster Ave", "city": "Paoli", "state": "PA", "postal_code": "19301", "latitude": 40.0415845065, "longitude": -75.4849526577, "stars": 4.5, "business_id": "KV_Q1o1s180L18du0Gp1YQ", "name": "Ardmore Pizza", "address": "10 Rittenhouse PL", "city": "Ardmore", "state": "PA", "postal_code": "19003", "latitude": 40.0067071, "longitude": -75.289671, "stars": 3.5, "business_id": "W_AW0i11G9eay7ncn67W", "name": "River 127", "address": "100 Iverbille St", "city": "New Orleans", "state": "LA", "postal_code": "70130", "latitude": 29.951359, "longitude": -90.0646715, "stars": 3.0, "business_id": "aPKMqTKf-4bJhYMQxqpQ", "name": "Craft Hall", "address": "901 N Delaware Ave", "city": "Philadelphia", "state": "PA", "postal_code": "19123", "latitude": 39.9628821, "longitude": -75.1356571, "stars": 3.5, "business_id": "ZXV50M9P-Z0K10d9MwWw", "name": "Cheesburger In Paradise", "address": "116 N Pottstown Pike", "city": "Exton", "state": "PA", "postal_code": "19341", "latitude": 40.09092, "longitude": -75.630607, "stars": 2.5, "business_id": "mGf9P0H0eC9Vjg0G0z9g", "name": "T-House", "address": "14735 Concord Pike", "city": "Wilmington", "state": "DE", "postal_code": "19803", "latitude": 39.822176, "longitude": -75.546007, "stars": 2.0, "business_id": "Lc0z0W0L6jW0B0k0JNNj", "name": "Edwardsville Children's Museum", "address": "722 Holyoke Rd", "city": "Edwardsville", "state": "IL", "postal_code": "62025", "latitude": 38.8043945, "longitude": -89.9497332, "stars": 4.5, "business_id": "F50N0M1hN0V9T0B0C9fJA", "name": "Sierra Pro Events", "address": "", "city": "Sparks", "state": "NV", "postal_code": "89431", "latitude": 39.5401545, "longitude": -119.7483949, "stars": 5.0, "business_id": "x1n0q1IpmC0N00d0U8U", "name": "Dianond Hair Company", "address": "5324 W 16th St", "city": "Indianapolis", "state": "IN", "postal_code": "46220", "latitude": 39.7874644177, "longitude": -86.2514209, "stars": 1.5, "business_id": "1jNtN9p77MPPx0fCn0G1", "name": "Tony's Restaurant & 3rd Street Cafe", "address": "312 Plaza St", "city": "Alton", "state": "IL", "postal_code": "62002", "latitude": 38.896563, "longitude": -90.1802032987, "stars": 3.0, "business_id": "Pns214m5f08Kk3d1x6A6A", "name": "Abby Rappoport, LAC, CMQ", "address": "1616 Chapala St, Ste 2", "city": "Santa Barbara", "state": "CA", "postal_code": "93101", "latitude": 34.4266787, "longitude": -119.7111968, "stars": 5.0, "business_id": "mpf3x-BjTdTfEAsyCzrAYpV", "name": "The UPS Store", "address": "87 Grasso Plaza Shopping Center", "city": "Affton", "state": "MO", "postal_code": "63123", "latitude": 38.551126, "longitude": -90.335695, "stars": 3.0, "business_id": "TUPRk1rK1K1_TansVW1NQ", "name": "Target", "address": "5255 E Broadway Blvd", "city": "Tucson", "state": "AZ", "postal_code": "85711", "latitude": 32.223236, "longitude": -110.880452, "stars": 3.5, "business_id": "TNTSWMcQd7CvYtjgqe9mV", "name": "St Honore Pastries", "address": "935 Race St", "city": "Philadelphia", "state": "PA", "postal_code": "19107", "latitude": 39.9555052, "longitude": -75.1555641, "stars": 4.0, "business_id": "mM6c6_wtDE0EUBK1GX0VfA", "name": "Perklemen Valley Brewery", "address": "101 Walnut St", "city": "Green Lane", "state": "PA", "postal_code": "18054", "latitude": 40.3381827, "longitude": -75.4716585, "stars": 4.5, "business_id": "n_BuPQx1shNbnPUS10dU8U", "name": "Famous Footwear", "address": "8522 Eager Road, Dierbergs Brentwood Point", "city": "Brentwood", "state": "MO", "postal_code": "63144", "latitude": 38.627695, "longitude": -90.340465, "stars": 2.0, "business_id": "qKRM_ZX51Yqk3btLwA01Q", "name": "Temple Beth-El", "address": "400 Pasadena Ave S", "city": "St. Petersburg", "state": "FL", "postal_code": "33707", "latitude": 27.76659, "longitude": -82.732983, "stars": 3.0, "business_id": "U3uofvbfv7f0NHEwAdVAKJJA", "name": "Marshalls", "address": "21785 Village Lakes Sc Dr", "city": "Land O' Lakes", "state": "FL", "postal_code": "33639", "latitude": 28.1904587953, "longitude": -82.4573802199, "stars": 3.5, "business_id": "eU5GskpB9bUA71D0MRK", "name": "Vietnamese Food Truck", "address": "", "city": "Tampa Bay", "state": "FL", "postal_code": "33602", "latitude": 27.9552692, "longitude": -82.4563199, "stars": 4.0, "business_id": "1l_R0bJpUfHresjy9EGmG", "name": "Denny's", "address": "8901 US 31 S", "city": "Indianapolis", "state": "IN", "postal_code": "46227", "latitude": 39.6371332838, "longitude": -86.127217412, "stars": 2.5, "business_id": "jAw5mZm8p003x0J181G", "name": "Adams Dental", "address": "15 N Missouri Ave", "city": "Clearwater", "state": "FL", "postal_code": "33755", "latitude": 27.966235, "longitude": -82.787412, "stars": 5.0, "review_count": 10, "business_id": "00sL1k00h00511_E0m0", "name": "Zio's Italian Market", "address": "2575 E Bay Dr", "city": "Largo", "state": "FL", "postal_code": "33771", "latitude": 27.9161159, "longitude": -82.7604608, "stars": 4.5, "review_count": 100, "business_id": "NUTTq6u9gRbD1180R0A", "name": "Tuna Bay", "address": "203 Race St", "city": "Philadelphia", "state": "PA", "postal_code": "19106", "latitude": 39.953949, "longitude": -75.143262, "stars": 4.0, "review_count": 245, "business_id": "rEby_1lU8y8gghlyb7kA", "name": "Arizona Truck Outfitters", "address": "625 N Stone Ave", "city": "Tucson", "state": "AZ", "postal_code": "85705", "latitude": 32.2208719, "longitude": -110.9723419, "stars": 4.5, "review_count": 14, "business_id": "N0XSSMqRAs0h0b01p0A", "name": "Herb Import Co", "address": "712 Adams St", "city": "New Orleans", "state": "LA", "postal_code": "70118", "latitude": 29.9414679565, "longitude": -90.129952757, "stars": 4.0, "review_count": 5, "business_id": "8uG1S7YVKE2TSn3C0M0K", "name": "Nitty Car Rental", "address": "1241 Airline Dr", "city": "Kenner", "state": "LA", "postal_code": "70062", "latitude": 29.981183, "longitude": -90.2540123, "stars": 3.5, "review_count": 14, "business_id": "R0eac1Q4eh08q7f6TcQ", "name": "BAP", "address": "1224 South St", "city": "Philadelphia", "state": "PA", "postal_code": "19147", "latitude": 39.943223, "longitude": -75.162568, "stars": 4.5, "review_count": 1205, "business_id": "M0DjwqntZKzAae7R0MPEJA", "name": "Rost Coffeehouse and Wine Bar", "address": "10359 104 Street NW", "city": "Edmonton", "state": "AB", "postal_code": "T5J 1B9", "latitude": 53.5460453, "longitude": -113.4991693, "stars": 4.0, "review_count": 10, "business_id": "qH0d0E1Kx4jYzKyzwPNS3", "name": "Barnes & Noble Booksellers", "address": "8675 River Crossing Blvd", "city": "Indianapolis", "state": "IN", "postal_code": "46240", "latitude": 39.91382, "longitude": -86.105739, "stars": 4.0, "review_count": 10, "business_id": "kFmV-J2pU6TWN50h0hK1Z", "name": "Hibachi Express", "address": "6625 E 82nd St", "city": "Indianapolis", "state": "IN", "postal_code": "46250", "latitude": 39.4761165, "longitude": -86.8530799, "stars": 4.0, "review_count": 10, "business_id": "90G5YX1q2G6e2M8A6K1Z", "name": "Romano's Macaroni Grill", "address": "5585 S Virginia St", "city": "Reno", "state": "NV", "postal_code": "89502", "latitude": 39.4761165, "longitude": -119.7893392, "stars": 2.5, "review_count": 10, "business_id": "41RzR70a5-QaXuvYxEGKG", "name": "Super Dog", "address": "1168 Gallatin Pike S", "city": "Nashville", "state": "TN", "postal_code": "37115", "latitude": 36.248596, "longitude": -86.7199857, "stars": 4.0, "review_count": 10, "business_id": "P50_C15fA133Hj2VM6Z1Q", "name": "Indian Walk Veterinary Center", "address": "662 Durham Rd", "city": "Newtown", "state": "PA", "postal_code": "18940", "latitude": 40.2734926, "longitude": -74.987496, "stars": 5.0, "review_count": 10, "business_id": "noByWYNDLQAR9CqkqfDQ", "name": "H&M", "address": "827-833 State St", "city": "Santa Barbara", "state": "CA", "postal_code": "93101", "latitude": 34.4202089255, "longitude": -119.7004597389, "stars": 3.0, "review_count": 24, "business_id": "tM0wMmFUEXRC9Zduonp7g", "name": "The Green Pheasant", "address": "215 1st Ave S", "city": "Nashville", "state": "TN", "postal_code": "37201", "latitude": 36.1598858, "longitude": -86.7731974, "stars": 4.0, "review_count": 10, "business_id": "QDN728WofFYKdJhh1S7g", "name": "Bar One", "address": "767 S 9th St", "city": "Philadelphia", "state": "PA", "postal_code": "19147", "latitude": 39.9398245785, "longitude": -75.1574465632, "stars": 4.0, "review_count": 65, "business_id": "sqSqqLY8N8nZ179c0chZyV", "name": "Domino's Pizza", "address": "3801 Highway 31 W", "city": "White House", "state": "TN", "postal_code": "37188", "latitude": 36.464746722, "longitude": -86.6591868908, "stars": 3.5, "review_count": 10, "business_id": "fVmn0Z0X0bj2179c0chZyV", "name": "Altitude Trampoline Park - Boise", "address": "1301 N Milwaukee St", "city": "Boise", "state": "ID", "postal_code": "83704", "latitude": 43.616735, "longitude": -116.2853821, "stars": 5.0, "review_count": 10, "business_id": "Mj0oz24M9N1Be10JkLd_Q", "name": "DeSandro on Main", "address": "4105 Main St", "city": "Philadelphia", "state": "PA", "postal_code": "19127", "latitude": 40.8224662, "longitude": -75.218314, "stars": 3.0, "review_count": 41, "business_id": "8SshLB4U7e0eUdVtJWnpA", "name": "Danceline", "address": "30 W Lancaster Ave", "city": "Paoli", "state": "PA", "postal_code": "19301", "latitude": 40.0415845065, "longitude": -75.4849526577, "stars": 4.5, "review_count": 11, "business_id": "KV_Q1o1s180L18du0Gp1YQ", "name": "Ardmore Pizza", "address": "10 Rittenhouse PL", "city": "Ardmore", "state": "PA", "postal_code": "19003", "latitude": 40.0067071, "longitude": -75.289671, "stars": 3.5, "review_count": 109, "business_id": "W_AW0i11G9eay7ncn67W", "name": "River 127", "address": "100 Iverbille St", "city": "New Orleans", "state": "LA", "postal_code": "70130", "latitude": 29.951359, "longitude": -90.0646715, "stars": 3.0, "review_count": 12, "business_id": "aPKMqTKf-4bJhYMQxqpQ", "name": "Craft Hall", "address": "901 N Delaware Ave", "city": "Philadelphia", "state": "PA", "postal_code": "19123", "latitude": 39.9628821, "longitude": -75.1356571, "stars": 3.5, "review_count": 165, "business_id": "ZXV50M9P-Z0K10d9MwWw", "name": "Cheesburger In Paradise", "address": "116 N Pottstown Pike", "city": "Exton", "state": "PA", "postal_code": "19341", "latitude": 40.09092, "longitude": -75.630607, "stars": 2.5, "review_count": 10, "business_id": "mGf9P0H0eC9Vjg0G0z9g", "name": "T-House", "address": "14735 Concord Pike", "city": "Wilmington", "state": "DE", "postal_code": "19803", "latitude": 39.822176, "longitude": -75.546007, "stars": 2.0, "review_count": 16, "business_id": "Lc0z0W0L6jW0B0k0JNNj", "name": "Edwardsville Children's Museum", "address": "722 Holyoke Rd", "city": "Edwardsville", "state": "IL", "postal_code": "62025", "latitude": 38.8043945, "longitude": -89.9497332, "stars": 4.5, "review_count": 10, "business_id": "F50N0M1hN0V9T0B0C9fJA", "name": "Sierra Pro Events", "address": "", "city": "Sparks", "state": "NV", "postal_code": "89431", "latitude": 39.5401545, "longitude": -119.7483949, "stars": 5.0, "review_count": 7, "business_id": "x1n0q1IpmC0N00d0U8U", "name": "Dianond Hair Company", "address": "5324 W 16th St", "city": "Indianapolis", "state": "IN", "postal_code": "46220", "latitude": 39.7874644177, "longitude": -86.2514209, "stars": 1.5, "review_count": 10, "business_id": "1jNtN9p77MPPx0fCn0G1", "name": "Tony's Restaurant & 3rd Street Cafe", "address": "312 Plaza St", "city": "Alton", "state": "IL", "postal_code": "62002", "latitude": 38.896563, "longitude": -90.1802032987, "stars": 3.0, "review_count": 10, "business_id": "Pns214m5f08Kk3d1x6A6A", "name": "Abby Rappoport, LAC, CMQ", "address": "1616 Chapala St, Ste 2", "city": "Santa Barbara", "state": "CA", "postal_code": "93101", "latitude": 34.4266787, "longitude": -119.7111968, "stars": 5.0, "review_count": 10, "business_id": "mpf3x-BjTdTfEAsyCzrAYpV", "name": "The UPS Store", "address": "87 Grasso Plaza Shopping Center", "city": "Affton", "state": "MO", "postal_code": "63123", "latitude": 38.551126, "longitude": -90.335695, "stars": 3.0, "business_id": "TUPRk1rK1K1_TansVW1NQ", "name": "Target", "address": "5255 E Broadway Blvd", "city": "Tucson", "state": "AZ", "postal_code": "85711", "latitude": 32.223236, "longitude": -110.880452, "stars": 3.5, "business_id": "TNTSWMcQd7CvYtjgqe9mV", "name": "St Honore Pastries", "address": "935 Race St", "city": "Philadelphia", "state": "PA", "postal_code": "19107", "latitude": 39.9555052, "longitude": -75.1555641, "stars": 4.0, "business_id": "mM6c6_wtDE0EUBK1GX0VfA", "name": "Perklemen Valley Brewery", "address": "101 Walnut St", "city": "Green Lane", "state": "PA", "postal_code": "18054", "latitude": 40.3381827, "longitude": -75.4716585, "stars": 4.5, "business_id": "n_BuPQx1shNbnPUS10dU8U", "name": "Famous Footwear", "address": "8522 Eager Road, Dierbergs Brentwood Point", "city": "Brentwood", "state": "MO", "postal_code": "63144", "latitude": 38.627695, "longitude": -90.340465, "stars": 2.0, "business_id": "qKRM_ZX51Yqk3btLwA01Q", "name": "Temple Beth-El", "address": "400 Pasadena Ave S", "city": "St. Petersburg", "state": "FL", "postal_code": "33707", "latitude": 27.76659, "longitude": -82.732983, "stars": 3.0, "business_id": "U3uofvbfv7f0NHEwAdVAKJJA", "name": "Marshalls", "address": "21785 Village Lakes Sc Dr", "city": "Land O' Lakes", "state": "FL", "postal_code": "33639", "latitude": 28.1904587953, "longitude": -82.4573802199, "stars": 3.5, "business_id": "eU5GskpB9bUA71D0MRK", "name": "Vietnamese Food Truck", "address": "", "city": "Tampa Bay", "state": "FL", "postal_code": "33602", "latitude": 27.9552692, "longitude": -82.4563199, "stars": 4.0, "business_id": "1l_R0bJpUfHresjy9EGmG", "name": "Denny's", "address": "8901 US 31 S", "city": "Indianapolis", "state": "IN", "postal_code": "46227", "latitude": 39.6371332838, "longitude": -86.127217412, "stars": 2.5, "business_id": "jAw5mZm8p003x0J181G", "name": "Adams Dental", "address": "15 N Missouri Ave", "city": "Clearwater", "state": "FL", "postal_code": "33755", "latitude": 27.966235, "longitude": -82.787412, "stars": 5.0, "review_count": 10, "business_id": "00sL1k00h00511_E0m0", "name": "Zio's Italian Market", "address": "2575 E Bay Dr", "city": "Largo", "state": "FL", "postal_code": "33771", "latitude": 27.9161159, "longitude": -82.7604608, "stars": 4.5, "review_count": 100, "business_id": "NUTTq6u9gRbD1180R0A", "name": "Tuna Bay", "address": "203 Race St", "city": "Philadelphia", "state": "PA", "postal_code": "19106", "latitude": 39.953949, "longitude": -75.143262, "stars": 4.0, "review_count": 245, "business_id": "rEby_1lU8y8gghlyb7kA", "name": "Arizona Truck Outfitters", "address": "625 N Stone Ave", "city": "Tucson", "state": "AZ", "postal_code": "85705", "latitude": 32.2208719, "longitude": -110.9723419, "stars": 4.5, "review_count": 14, "business_id": "N0XSSMqRAs0h0b01p0A", "name": "Herb Import Co", "address": "712 Adams St", "city": "New Orleans", "state": "LA", "postal_code": "70118", "latitude": 29.9414679565, "longitude": -90.129952757, "stars": 4.0, "review_count": 5, "business_id": "8uG1S7YVKE2TSn3C0M0K", "name": "Nitty Car Rental", "address": "1241 Airline Dr", "city": "Kenner", "state": "LA", "postal_code": "70062", "latitude": 29.981183, "longitude": -90.2540123, "stars": 3.5, "review_count": 14, "business_id": "R0eac1Q4eh08q7f6TcQ", "name": "BAP", "address": "1224 South St", "city": "Philadelphia", "state": "PA", "postal_code": "19147", "latitude": 39.943223, "longitude": -75.162568, "stars": 4.5, "review_count": 1205, "business_id": "M0DjwqntZKzAae7R0MPEJA", "name": "Rost Coffeehouse and Wine Bar", "address": "10359 104 Street NW", "city": "Edmonton", "state": "AB", "postal_code": "T5J 1B9", "latitude": 53.5460453, "longitude": -113.4991693, "stars": 4.0, "review_count": 10, "business_id": "qH0d0E1Kx4jYzKyzwPNS3", "name": "Barnes & Noble Booksellers", "address": "8675 River Crossing Blvd", "city": "Indianapolis", "state": "IN", "postal_code": "46240", "latitude": 39.91382, "longitude": -86.105739, "stars": 4.0, "review_count": 10, "business_id": "kFmV-J2pU6TWN50h0hK1Z", "name": "Hibachi Express", "address": "6625 E 82nd St", "city": "Indianapolis", "state": "IN", "postal_code": "46250", "latitude": 39.4761165, "longitude": -86.8530799, "stars": 4.0, "review_count": 10, "business_id": "90G5YX1q2G6e2M8A6K1Z", "name": "Romano's Macaroni Grill", "address": "5585 S Virginia St", "city": "Reno", "state": "NV", "postal_code": "89502", "latitude": 39.4761165, "longitude": -119.7893392, "stars": 2.5, "review_count": 10, "business_id": "41RzR70a5-QaXuvYxEGKG", "name": "Super Dog", "address": "1168 Gallatin Pike S", "city": "Nashville", "state": "TN", "postal_code": "37115", "latitude": 36.248596, "longitude": -86.7199857, "stars": 4.0, "review_count": 10, "business_id": "P50_C15fA133Hj2VM6Z1Q", "name": "Indian Walk Veterinary Center", "address": "662 Durham Rd", "city": "Newtown", "state": "PA", "postal_code": "18940", "latitude": 40.2734926, "longitude": -74.987496, "stars": 5.0, "review_count": 10, "business_id": "noByWYNDLQAR9CqkqfDQ", "name": "H&M", "address": "827-833 State St", "city": "Santa Barbara", "state": "CA", "postal_code": "93101", "latitude": 34.4202089255, "longitude": -119.7004597389, "stars": 3.0, "review_count": 24, "business_id": "tM0wMmFUEXRC9Zduonp7g", "name": "The Green Pheasant", "address": "215 1st Ave S", "city": "Nashville", "state": "TN", "postal_code": "37201", "latitude": 36.1598858, "longitude": -86.7731974, "stars": 4.0, "review_count": 10, "business_id": "QDN728WofFYKdJhh1S7g", "name": "Bar One", "address": "767 S 9th St", "city": "Philadelphia", "state": "PA", "postal_code": "19147", "latitude": 39.9398245785, "longitude": -75.1574465632, "stars": 4.0, "review_count": 65, "business_id": "sqSqqLY8N8nZ179c0chZyV", "name": "Domino's Pizza", "address": "3801 Highway 31 W", "city": "White House", "state": "TN", "postal_code": "37188", "latitude": 36.464746722, "longitude": -86.6591868908, "stars": 3.5, "review_count": 10, "business_id": "fVmn0Z0X0bj2179c0chZyV", "name": "Altitude Trampoline Park - Boise", "address": "1301 N Milwaukee St", "city": "Boise", "state": "ID", "postal_code": "83704", "latitude": 43.616735, "longitude": -116.2853821, "stars": 5.0, "review_count": 10, "business_id": "Mj0oz24M9N1Be10JkLd_Q", "name": "DeSandro on Main", "address": "4105 Main St", "city": "Philadelphia", "state": "PA", "postal_code": "19127", "latitude": 40.8224662, "longitude": -75.218314, "stars
```


OSEMN Report
Midterm Project

Python code for conversions: for 1st and 2nd question.

```
6  import csv
7  import json
8  import pandas as pd
9  from pathlib import Path
10
11  #read the dataset
12  with open("yelp_academic_dataset_business.json", "r", encoding='utf-8') as infile:
13      data = infile.readlines() #read all lines
14      json_data = []
15      for line in data :
16          line_data = json.loads(line)
17          # handles all Null or empty values. It doest read the whole row
18          if any(val == "None" or val == "" for val in line_data.values()):
19              continue
20          json_data.append(line_data)
21
22  json_dataframe = pd.DataFrame.from_records(json_data[:300])
23  json_dataframe_new = json_dataframe[['name', 'state', 'city', 'review_count']]
24
25  with open("format_json.json", "w") as outfile:
26      outfile.write('[')
27      for i, row in json_dataframe_new.iterrows():
28          json.dump(row.to_dict(), outfile)
29          if i < len(json_dataframe_new) - 1:
30              outfile.write(',')
31              outfile.write('\n')
32      outfile.write(']')
33
34  #convert from json to csv
35  json_dataframe_new.to_csv('format.csv', index = False)
36
37  #convert from csv to json
38  with open("format.csv", "r", encoding='utf-8') as csvFile:
39      csv_read = csv.DictReader(csvFile)
40      conv_json_data = {}
41      for line, rows in enumerate(csv_read, start=1):
42          conv_json_data.update({"Business {:02}".format(line):rows})
43      with open("format.json", "w", encoding='utf-8') as jsonFile:
44          json.dump(conv_json_data, jsonFile, indent=4)
45
```

OSEMN Report Midterm Project

For 3rd question:

```
proj_datawrangling.py 1 x 1st.sql 2ndtable.py 1
proj_datawrangling.py > ...
1
2 #references :
3 # https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#returning-a-view-versus-a-copy
4
5 import csv
6 import json
7 import pandas as pd
8 from pathlib import Path
9
10 #read the dataset
11 with open("yelp_academic_dataset_business.json", "r", encoding='utf-8') as infile:
12     data = infile.readlines() #read all lines
13     json_data = []
14     for line in data :
15         line_data = json.loads(line)
16         if any(val == "None" or val == "" for val in line_data.values()): # handles all Null or empty values. It does read the whole row
17             continue
18         json_data.append(line_data)
19
20 json_dataframe = pd.DataFrame.from_records(json_data[:300])
21 json_dataframe_new = json_dataframe[['name','business_id', 'state', 'city']]
22 json_dataframe_new2 = json_dataframe[['business_id','review_count','stars']]
23
24 with open("businessdetails.json", "w") as outfile:
25     outfile.write('[')
26     for i, row in json_dataframe_new.iterrows():
27         json.dump(row.to_dict(), outfile)
28         if i < len(json_dataframe_new) - 1:
29             outfile.write(',')
30             outfile.write('\n')
31     outfile.write(']')
32
33 with open("businessdetails2.json", "w") as outfile:
34     outfile.write('[')
35     for i, row in json_dataframe_new2.iterrows():
36         json.dump(row.to_dict(), outfile)
37         if i < len(json_dataframe_new2) - 1:
38             outfile.write(',')
39             outfile.write('\n')
40     outfile.write(']')
41
42 #convert from json to csv
43 json_dataframe_new.to_csv('businessdetails.csv',index = False)
44 json_dataframe_new2.to_csv('businessdetails2.csv',index = False)
45
46 #convert from csv to json
47 with open("businessdetails.csv", "r", encoding='utf-8') as csvFile:
48     csv_read = csv.DictReader(csvFile)
49     conv_json_data = {}
50     for line, rows in enumerate(csv_read, start=1):
51         conv_json_data.update({"Business {02}".format(line):rows})
52     with open("businessdetails.json", "w", encoding='utf-8') as jsonFile:
53         json.dump(conv_json_data, jsonFile, indent=4)
54
55 # converting from csv to json for 2table
56 with open("businessdetails2.csv", "r", encoding='utf-8') as csvFile:
57     csv_read = csv.DictReader(csvFile)
58     conv_json_data = {}
59     for line, rows in enumerate(csv_read, start=1):
60         conv_json_data.update({"Business {02}".format(line):rows})
61     with open("businessdetails2.json", "w", encoding='utf-8') as jsonFile:
62         json.dump(conv_json_data, jsonFile, indent=4)
```

OSEMN Report
Midterm Project

CSV Samples:

This is for 1st question:

| | A | B | C | D | E |
|----|--------------------------------|-------|----------------|--------------|---|
| 1 | name | state | city | review_count | |
| 2 | Abby Rappoport, LAC, CMQ | CA | Santa Barbara | 7 | |
| 3 | The UPS Store | MO | Affton | 15 | |
| 4 | Target | AZ | Tucson | 22 | |
| 5 | St Honore Pastries | PA | Philadelphia | 80 | |
| 6 | Perkiomen Valley Brewery | PA | Green Lane | 13 | |
| 7 | Sonic Drive-In | TN | Ashland City | 6 | |
| 8 | Famous Footwear | MO | Brentwood | 13 | |
| 9 | Temple Beth-El | FL | St. Petersburg | 5 | |
| 10 | Tsevi's Pub And Grill | MO | Affton | 19 | |
| 11 | Sonic Drive-In | TN | Nashville | 10 | |
| 12 | Marshalls | FL | Land O' Lakes | 6 | |
| 13 | Denny's | IN | Indianapolis | 28 | |
| 14 | Adams Dental | FL | Clearwater | 10 | |
| 15 | Zio's Italian Market | FL | Largo | 100 | |
| 16 | Tuna Bar | PA | Philadelphia | 245 | |
| 17 | Arizona Truck Outfitters | AZ | Tucson | 10 | |
| 18 | Herb Import Co | LA | New Orleans | 5 | |
| 19 | Nifty Car Rental | LA | Kenner | 14 | |
| 20 | BAP | PA | Philadelphia | 205 | |
| 21 | Roast Coffeehouse and Wine Bar | AB | Edmonton | 40 | |
| 22 | Barnes & Noble Booksellers | IN | Indianapolis | 38 | |
| 23 | Hibachi Express | IN | Indianapolis | 20 | |
| 24 | Romano's Macaroni Grill | NV | Reno | 339 | |
| 25 | Super Dog | TN | Nashville | 6 | |
| 26 | Indian Walk Veterinary Center | PA | Newtown | 15 | |
| 27 | H&M | CA | Santa Barbara | 24 | |
| 28 | The Green Pheasant | TN | Nashville | 161 | |

OSEMN Report Midterm Project

This is for the 2nd question

| | name | state | city | review_count | review_count_log |
|----|----------------------------------|-------|----------------|--------------|------------------|
| 1 | Abby Rappoport, LAC, CMQ | CA | Santa Barbara | 7 | 0.8450980 |
| 2 | The UPS Store | MO | Affton | 15 | 1.1760913 |
| 3 | St Honore Pastries | PA | Philadelphia | 80 | 1.9030900 |
| 4 | Perkiomen Valley Brewery | PA | Green Lane | 13 | 1.1139434 |
| 5 | Sonic Drive-In | TN | Ashland City | 6 | 0.7781513 |
| 6 | Famous Footwear | MO | Brentwood | 13 | 1.1139434 |
| 7 | Temple Beth-El | FL | St. Petersburg | 5 | 0.6989700 |
| 8 | Tsevi's Pub And Grill | MO | Affton | 19 | 1.2787536 |
| 9 | Sonic Drive-In | TN | Nashville | 10 | 1.0000000 |
| 10 | Marshalls | FL | Land O' Lakes | 6 | 0.7781513 |
| 11 | Denny's | IN | Indianapolis | 28 | 1.4471580 |
| 12 | Adams Dental | FL | Clearwater | 10 | 1.0000000 |
| 13 | Zio's Italian Market | FL | Largo | 100 | 2.0000000 |
| 14 | Tuna Bar | PA | Philadelphia | 245 | 2.3891661 |
| 15 | Arizona Truck Outfitters | AZ | Tucson | 10 | 1.0000000 |
| 16 | Herb Import Co | LA | New Orleans | 5 | 0.6989700 |
| 17 | Nifty Car Rental | LA | Kenner | 14 | 1.1461280 |
| 18 | BAP | PA | Philadelphia | 205 | 2.3117539 |
| 19 | Roast Coffeehouse and Wine Bar | AB | Edmonton | 40 | 1.6020600 |
| 20 | Barnes & Noble Booksellers | IN | Indianapolis | 38 | 1.5797836 |
| 21 | Hibachi Express | IN | Indianapolis | 20 | 1.3010300 |
| 22 | Romano's Macaroni Grill | NV | Reno | 339 | 2.5301997 |
| 23 | Super Dog | TN | Nashville | 6 | 0.7781513 |
| 24 | Indian Walk Veterinary Center | PA | Newtown | 15 | 1.1760913 |
| 25 | H&M | CA | Santa Barbara | 24 | 1.3802112 |
| 26 | The Green Pheasant | TN | Nashville | 161 | 2.2068259 |
| 27 | Bar One | PA | Philadelphia | 65 | 1.8129134 |
| 28 | Domino's Pizza | TN | White House | 8 | 0.9030900 |
| 29 | Altitude Trampoline Park - Boise | ID | Boise | 30 | 1.4771213 |
| 30 | DeSandro on Main | PA | Philadelphia | 41 | 1.6127839 |

OSEMN Report
Midterm Project

This is for 3rd question

| A | B | C | D | E | F |
|------------------------|--------------|-------|---|---|---|
| business_id | review_count | stars | | | |
| Pns2l4eNsfO8kk83dixA6A | 7 | 5 | | | |
| mpf3x-BjTdTEA3yCZrAYPw | 15 | 3 | | | |
| tUFrWirKiKi_TAnsVWINQQ | 22 | 3.5 | | | |
| MTSW4McQd7CbVtyjqoe9mw | 80 | 4 | | | |
| mWMc6_wTdE0EUBKIGXDVF | 13 | 4.5 | | | |
| CF33F8-E6oudUQ46HnavjQ | 6 | 2 | | | |
| n_0UpQx1hsNbnPUSlodU8w | 13 | 2.5 | | | |
| qkRM_2X51Yqyk3btlwAQlg | 5 | 3.5 | | | |
| k0hlBqXX-Bt0vf1op7Jr1w | 19 | 3 | | | |
| bBDDEgkFA1Otx9Lfe7BZUQ | 10 | 1.5 | | | |
| UJsufbvfyfONHeWdvAHKjA | 6 | 3.5 | | | |
| il_Ro8jwPIHresjw9EGmBg | 28 | 2.5 | | | |
| jaxMSolnw8Poo3XeMJt8lQ | 10 | 5 | | | |
| 0bPLkL0QhhPO5kt1_EXmNQ | 100 | 4.5 | | | |
| MUTTqe8uqyMdB1186RmNeA | 245 | 4 | | | |
| rBmpy_Y1UbBx8ggHlyb7hA | 10 | 4.5 | | | |
| MOXSSHqrASOnhgbWDJlpQA | 5 | 4 | | | |
| 8wGISYjYkE2tSqn3cDMu8A | 14 | 3.5 | | | |
| ROeacJQwBeh05Rqg7F6TCg | 205 | 4.5 | | | |
| WKMJwqnfZKsAae75RMP6jA | 40 | 4 | | | |
| qhDdDel3K4jy2KyzwFN53w | 38 | 4 | | | |
| kfNv-JZpuN6TVNSO6hHdkw | 20 | 4 | | | |
| 9OG5YkX1g2GReZM0AskizA | 339 | 2.5 | | | |
| 4iRzR7OaS-QaSxuvYxEGKA | 6 | 4 | | | |
| PSO_C1Sfa13JHjzVNW6ziQ | 15 | 5 | | | |
| noByYnTDLQAra9ccqxdfDw | 24 | 3 | | | |
| tMkwHmWFUEXrC9ZduonpTg | 161 | 4 | | | |
| QdN72BWoyFypdGJhhI5r7g | 65 | 4 | | | |
| sqSqLy0sN8n2IZrAbzidQ | 8 | 3.5 | | | |
| fvWn8oXXwbj2l79cochZyw | 30 | 5 | | | |
| Mjboz24M9NIBeiOJKLEd_Q | 41 | 3 | | | |
| 8sshLb4UU7emeUDvtJWnpA | 11 | 4.5 | | | |
| kV_Q1oqis8Qli8dUoGpTyQ | 109 | 3.5 | | | |
| w_AMNol1iG9eay7ncmc67w | 12 | 3 | | | |
| aPNXGTDkf-4bihvMBQxqpQ | 65 | 3.5 | | | |

OSEMN Report Midterm Project

JSON Samples :

```
{
  "Business 01": {
    "name": "Abby Rappoport, LAC, CMQ",
    "business_id": "Pns2l4eNsf08kk83dixA6A",
    "state": "CA",
    "city": "Santa Barbara"
  },
  "Business 02": {
    "name": "The UPS Store",
    "business_id": "mpf3x-BjTdTEA3yCZrAYPw",
    "state": "MO",
    "city": "Affton"
  },
  "Business 03": {
    "name": "Target",
    "business_id": "tUfRWirKiKi_TAnsVWINQQ",
    "state": "AZ",
    "city": "Tucson"
  },
  "Business 04": {
    "name": "St Honore Pastries",
    "business_id": "MTSW4McQd7CbVtyjqoe9mw",
    "state": "PA",
    "city": "Philadelphia"
  },
  "Business 05": {
    "name": "Perkiomen Valley Brewery",
    "business_id": "mWMc6_wTdE0EUBKIGXDvFA",
    "state": "PA",
    "city": "Green Lane"
  },
  "Business 06": {
    "name": "Sonic Drive-In",
    "business_id": "CF33F8-E6oudUQ46HnavjQ",
    "state": "TN",
    "city": "Ashland City"
  },
  "Business 07": {
    "name": "Famous Footwear",
    "business_id": "n_0UpQx1hsNbnPUSlodU8w",
    "state": "MO",
    "city": "Brentwood"
  },
  "Business 08": {
    "name": "Temple Beth-El",
    "business_id": "qkRM_2X51Yqxk3btlwAQIg",
    "state": "FL",
    "city": "St. Petersburg"
  },
  "Business 09": {
    "name": "Tsevi's Pub And Grill",
    "business_id": "k0hlBqXX-Bt0vflop7Jr1w",
    "state": "MO",
    "city": "Affton"
  },
  "Business 10": {
    "name": "Sonic Drive-In",
    "business_id": "bBDEgkFA10tx9Lfe7BZUQ",
    "state": "TN",
    "city": "Nashville"
  },
  "Business 11": {
    "name": "Marshalls",
    "business_id": "UJsufbvfyfONHeWdvAHKjA",
    "state": "FL",
    "city": "Land O' Lakes"
  },
  "Business 12": {
```

OSEMN Report Midterm Project

```
proj_datawrangling.py 1  {} businessdetails2.json x 1st.sql 2ndtable.py 1
businessdetails2.json > {} Business 04
1  {
2    "Business 01": {
3      "business_id": "Pns2l4eNsf08kk83dixA6A",
4      "review_count": "7",
5      "stars": "5.0"
6    },
7    "Business 02": {
8      "business_id": "mpf3x-BjTdTEA3yCZrAYPw",
9      "review_count": "15",
10     "stars": "3.0"
11   },
12   "Business 03": {
13     "business_id": "tUFrWirKiKi_TAnsVWINQQ",
14     "review_count": "22",
15     "stars": "3.5"
16   },
17   "Business 04": {
18     "business_id": "MTSW4McQd7CbVtyjqoe9mw",
19     "review_count": "80",
20     "stars": "4.0"
21   },
22   "Business 05": {
23     "business_id": "mWMc6_wTdE0EUBKIGXDvfA",
24     "review_count": "13",
25     "stars": "4.5"
26   },
27   "Business 06": {
28     "business_id": "CF33F8-E6oudUQ46HnavjQ",
29     "review_count": "6",
30     "stars": "2.0"
31   },
32   "Business 07": {
33     "business_id": "n_0UpQx1hsNbnPUSlodU8w",
34     "review_count": "13",
35     "stars": "2.5"
36   },
37   "Business 08": {
38     "business_id": "qkRM_2X51Yqk3btlwAQIg",
39     "review_count": "5",
40     "stars": "3.5"
41   },
42   "Business 09": {
43     "business_id": "k0hlBqXX-Bt0vf1op7Jr1w",
44     "review_count": "19",
45     "stars": "3.0"
46   },
47   "Business 10": {
48     "business_id": "bBDEgkFA10tx9Lfe7BZUQ",
49     "review_count": "10",
50     "stars": "1.5"
51   },
52   "Business 11": {
53     "business_id": "UJsufbvfyfONHeWdvAHKjA",
54     "review_count": "6",
55     "stars": "3.5"
56   },
57   "Business 12": {
58     "business_id": "il_Ro8jwPLHresjw9EGmBg",
59     "review_count": "28",
60     "stars": "2.5"
61   },
62   "Business 13": {
63     "business_id": "jaxMSoInw8Poo3XeMJt8lQ",
64     "review_count": "10",
65     "stars": "5.0"
66   },
67   "Business 14": {
68     "business_id": "0bPl.kl.00hbp05kt1.FXmN0".
```


OSEMN Report Midterm Project

SQL Query:

```
1st.sql
1
2 -- What is the relationship between the average rating of a restaurant and the number of reviews it has received?
3
4 -- Do restaurants with more reviews have a higher average rating or is there no significant relationship?
5
6 SELECT D.business_id, D.name, D.state, D.city, SUM(R.review_count) AS ReviewCount, AVG(R.stars) AS AverageStars
7 FROM business_details AS D
8 JOIN business_review AS R
9 ON D.business_id = R.business_id
10 GROUP BY D.business_id, D.name, D.state, D.city;
11
12
```

phpMyAdmin: below are the 2 tables loaded into phpMyAdmin

The screenshot shows the phpMyAdmin interface. On the left is a sidebar with a tree view of the database structure, including tables like 'actor', 'address', 'business_details', 'business_review', etc. The main area displays the 'business_details' table. At the top, there are tabs for 'Browse', 'Structure', 'SQL', 'Search', 'Insert', 'Export', 'Import', 'Operations', 'Tracking', and 'Triggers'. Below these is a table with columns: 'name', 'business_id', 'state', and 'city'. Each row represents a restaurant entry with its name, unique business ID, state, and city. The table contains 20 rows of data.

| | name | business_id | state | city |
|--------------------------|---------------------------------------|-------------------------|-------|----------------|
| <input type="checkbox"/> | Pathmark | -4dYswJy7SPcbcERvitmlg | PA | Philadelphia |
| <input type="checkbox"/> | B2 Bluefin | -7GDqSUaXrpC8QI7nDBxWA | PA | Bala Cynwyd |
| <input type="checkbox"/> | David Thomas Trailways | -aeZuatjCDMV1X4gCTz9Ug | PA | Philadelphia |
| <input type="checkbox"/> | Gilley's Antique Mall | -ZXycCyDjj1QimnJyGtYtg | IN | Plainfield |
| <input type="checkbox"/> | Elegance African Hair Braiding | 07JLsUlnZr5OSiYH2XVi4g | PA | Philadelphia |
| <input type="checkbox"/> | Zio's Italian Market | 0bPLkLOQhhPO5kt1_EXmNQ | FL | Largo |
| <input type="checkbox"/> | Charlie's Market | 0qNpTGToqPwOLi2hAdx4Xw | FL | Tampa |
| <input type="checkbox"/> | Whole Foods Market | 0vCsslIn-TYnDeOSNT4rlfg | IN | Carmel |
| <input type="checkbox"/> | The Pepper Pott | 1dSKEitDDgIkaApe6UNMSA | TN | Nashville |
| <input type="checkbox"/> | Rita's Italian Ice | 1GnJDrpp-jyDRCptavD9RQ | PA | Aston |
| <input type="checkbox"/> | Ricardo's Italian Cafe | 1MeIwdbTnZOBFCkOrgaxuw | MO | Saint Louis |
| <input type="checkbox"/> | Enjoy The Mountain | 25UwwOC0wvF9CZ_3B6vWtA | CA | Santa Barbara |
| <input type="checkbox"/> | AIM Mail Center | 2Bi3ef7c0CrLxHD_-yy_tw | AZ | Oro Valley |
| <input type="checkbox"/> | ReAnimator Coffee | 2dlQX5sP9X6Dlm1MmNOISw | PA | Philadelphia |
| <input type="checkbox"/> | Mighty Dustless | 2n9HHBxG7yAyAUwXXa49aw | PA | Southampton |
| <input type="checkbox"/> | Total Access Urgent Care | 2oxrOO3c9_mQmqM9kwEm9Q | MO | Saint Louis |
| <input type="checkbox"/> | Cheeseburger In Paradise | 2xVsWBNFwZOxlOdd9Mwnww | PA | Exton |
| <input type="checkbox"/> | Rippy's Central Pike Collision Repair | 35dqrE4z9qWCtCn749oorg | TN | Hermitage |
| <input type="checkbox"/> | Golden Chopstick Chinese Restaurant | 3BJxm-HnvzdwD1zjmSbmyQ | PA | Philadelphia |
| <input type="checkbox"/> | The Handlebar Boise | 3S9GYtMOYhbZi4MMPHOHT0Q | ID | Boise |
| <input type="checkbox"/> | Walmart Supercenter | 4dVrWc4Nhuw9apfk9BWEcA | FL | Tarpon Springs |
| <input type="checkbox"/> | Eat Mah Taco @ Pal's Lounge | 4IcB3QyMEA85UTWFKh9O9A | LA | New Orleans |
| <input type="checkbox"/> | Super Dog | 4IRzR7OaS-QaSxuvYxEGKA | TN | Nashville |

OSEMN Report Midterm Project

phpMyAdmin

Recent Favorites

New

cs512_nalubarr

- New
- actor
- address
- bsg_cert
- bsg_cert_people
- bsg_people
- bsg_planets
- bsg_ship_assignment
- bsg_ship_class
- bsg_ship_instance
- business_details
- business_review
- category
- city
- country
- customer
- demo_table
- film
- film_actor
- film_category
- film_text
- inventory
- language
- payment
- rental
- staff
- store
- information_schema

Server: classmysql.engr.oregonstate.edu » Database: cs512_nalubarr » Table: business_review

Browse Structure SQL Search Insert Export Import Operations

Extra options

| | | | | business_id | review_count | stars |
|--------------------------|------|------|--------|------------------------|--------------|-------|
| <input type="checkbox"/> | Edit | Copy | Delete | -4dYswJy7SPcbcERvitmIg | 34 | 2.5 |
| <input type="checkbox"/> | Edit | Copy | Delete | -7GDqSUaXrpC8QI7nDBxWA | 92 | 4 |
| <input type="checkbox"/> | Edit | Copy | Delete | -aeZuatjCDMV1X4gCTz9Ug | 6 | 4.5 |
| <input type="checkbox"/> | Edit | Copy | Delete | -ZXYcCyDjj1QimnJyGtYtg | 24 | 4.5 |
| <input type="checkbox"/> | Edit | Copy | Delete | 07JLsUlnZr5OSiYH2XVi4g | 8 | 2 |
| <input type="checkbox"/> | Edit | Copy | Delete | 0bPLkL0QhhPO5kt1_EXmNQ | 100 | 4.5 |
| <input type="checkbox"/> | Edit | Copy | Delete | 0qNpTGTCqPwOLi2hADx4Xw | 9 | 3 |
| <input type="checkbox"/> | Edit | Copy | Delete | 0vCsslN-TYnDeOSNT4rlfg | 44 | 3.5 |
| <input type="checkbox"/> | Edit | Copy | Delete | 1dSKEitDDgIkaApe6UNMSA | 18 | 5 |
| <input type="checkbox"/> | Edit | Copy | Delete | 1GnJDrpp-jyDRCptavD9RQ | 7 | 4 |
| <input type="checkbox"/> | Edit | Copy | Delete | 1MelwdbTnZOBFCOrgaxuw | 80 | 3.5 |
| <input type="checkbox"/> | Edit | Copy | Delete | 25Uww0C0wvF9CZ_3B6vWtA | 48 | 4.5 |
| <input type="checkbox"/> | Edit | Copy | Delete | 2Bi3ef7c0CrLxHD_-yy_tw | 13 | 3 |
| <input type="checkbox"/> | Edit | Copy | Delete | 2dIQX5sP9X6DIm1MmNOISw | 105 | 4 |
| <input type="checkbox"/> | Edit | Copy | Delete | 2n9HHBxG7yAyAUwXXa49aw | 5 | 4 |
| <input type="checkbox"/> | Edit | Copy | Delete | 2oxrOO3c9_mQmqM9kwEm9Q | 6 | 3.5 |
| <input type="checkbox"/> | Edit | Copy | Delete | 2xVsWBNFwZOxIOdd9Mwnww | 20 | 2.5 |
| <input type="checkbox"/> | Edit | Copy | Delete | 35dqrE4z9qWCtCn749oorg | 5 | 2.5 |
| <input type="checkbox"/> | Edit | Copy | Delete | 3BJxm-HnvzdwD1zjmSbmyQ | 137 | 3 |
| <input type="checkbox"/> | Edit | Copy | Delete | 3S9GYtMOYhbZI4MMPOHT0Q | 34 | 4.5 |
| <input type="checkbox"/> | Edit | Copy | Delete | 4dVrWc4Nhuw9apfk9BWEcA | 29 | 2 |
| <input type="checkbox"/> | Edit | Copy | Delete | 4lcB3QyMEA85UTWFKh9O9A | 8 | 4.5 |

Console

OSEMN Report Midterm Project

SQL query output:

This is the result_report generated by the query.

phpMyAdmin

Recent Favorites

New

cs512_nalubanr

New

actor

address

bsg_cert

bsg_cert_people

bsg_people

bsg_planets

bsg_ship_assignment

bsg_ship_class

bsg_ship_instance

business_details

business_review

category

city

country

customer

demo_table

film

film_actor

film_category

film_text

inventory

language

payment

rental

staff

store

information_schema

Server: classmysql.engr.oregonstate.edu » Database: cs512_nalubanr » Table: business_details

Browse Structure SQL Search Insert Export Import Operations Tracking Triggers

Showing rows 0 - 199 (300 total, Query took 0.0001 seconds.)

SELECT D.business_id, D.name, D.state, D.city, SUM(R.review_count) AS ReviewCount, AVG(R.stars) AS AverageStars FROM busin
D.business_id, D.name, D.state, D.city;

Profiling [Edit inline] [Edit] [Explain SQL] [Create PHP code] [Refresh]

1 > >> | Show all | Number of rows: 25 Filter rows: Search this table

Extra options

| business_id | name | state | city | ReviewCount | AverageStars |
|-------------------------|---------------------------------------|-------|---------------|-------------|--------------|
| -4dYswJy7SPcbcERvitmlg | Pathmark | PA | Philadelphia | 34 | 2.5 |
| -7GDqSUaXrpC8Ql7nDBxWA | B2 Bluefin | PA | Bala Cynwyd | 92 | 4 |
| -aeZuatCDMV1X4gCTz9Ug | David Thomas Trailways | PA | Philadelphia | 6 | 4.5 |
| -ZYxCyDij1QimnJyGtYtg | Gilley's Antique Mall | IN | Plainfield | 24 | 4.5 |
| 07JLsUlInZr5OSiYH2XVi4g | Elegance African Hair Braiding | PA | Philadelphia | 8 | 2 |
| 0bPLkL0QhhPO5kt1_EXmNQ | Zio's Italian Market | FL | Largo | 100 | 4.5 |
| 0qNpTGtqPwOLi2hADx4Xw | Charlie's Market | FL | Tampa | 9 | 3 |
| 0vCsslN-TYnDeOSNT4rlfg | Whole Foods Market | IN | Carmel | 44 | 3.5 |
| 1dSKEitDDglkaApe6UNMSA | The Pepper Pott | TN | Nashville | 18 | 5 |
| 1GnJDrpp-jyDRCptavD9RQ | Rita's Italian Ice | PA | Aston | 7 | 4 |
| 1MelwdbTnZOBFCOrgaxuw | Ricardo's Italian Cafe | MO | Saint Louis | 80 | 3.5 |
| 25Uww0C0wvF9CZ_3B6vWtA | Enjoy The Mountain | CA | Santa Barbara | 48 | 4.5 |
| 2Bi3ef7c0CrLxHD_yy_tw | AIM Mail Center | AZ | Oro Valley | 13 | 3 |
| 2dIQX5sP9X6Dlm1MmNOISw | ReAnimator Coffee | PA | Philadelphia | 105 | 4 |
| 2n9HHBxG7yAyAUwXXa49aw | Mighty Dustless | PA | Southampton | 5 | 4 |
| 2oxrOO3c9_mQmqM9kwEm9Q | Total Access Urgent Care | MO | Saint Louis | 6 | 3.5 |
| 2xVsWBNFwZOxIOdd9Mwnnw | Cheeseburger In Paradise | PA | Exton | 20 | 2.5 |
| 35dqrE4z9qWCiCn749oorg | Rippy's Central Pike Collision Repair | TN | Hermitage | 5 | 2.5 |
| 3BJxm-HnvzdwD1zjmSbmyQ | Golden Chopstick Chinese Restaurant | PA | Philadelphia | 137 | 3 |
| 3S9GYtMOYhbZi4MMPOHT0Q | The Handlebar Boise | ID | Boise | 34 | 4.5 |

Console

SELECT * FROM business_details

OSEMN Report
Midterm Project

This is the result_report exported as .csv.

| business_id | name | state | city | ReviewCount | AverageStars |
|------------------------|---------------------------------------|-------|----------------|-------------|--------------|
| -4dYswJy7SPcbcERvitmlg | Pathmark | PA | Philadelphia | 34 | 2.5 |
| -7GDqSUaXrpC8QI7nDBxWA | B2 Bluefin | PA | Bala Cynwyd | 92 | 4 |
| -aeZuatjCDMV1X4gCTz9Ug | David Thomas Trailways | PA | Philadelphia | 6 | 4.5 |
| -ZXYcCyDjj1QimnJyGtYtg | Gilley's Antique Mall | IN | Plainfield | 24 | 4.5 |
| 07JLsUlnZr5OSiYH2XVi4g | Elegance African Hair Braiding | PA | Philadelphia | 8 | 2 |
| 0bPLkL0QhhPO5kt1_EXmNQ | Zio's Italian Market | FL | Largo | 100 | 4.5 |
| 0qNpTGTcqPwOLi2hADx4Xw | Charlie's Market | FL | Tampa | 9 | 3 |
| 0vCsslN-TYnDeOSNT4rlfg | Whole Foods Market | IN | Carmel | 44 | 3.5 |
| 1dSKEitDDglkaApe6UNMSA | The Pepper Pott | TN | Nashville | 18 | 5 |
| 1GnJDrpp-jyDRCptavD9RQ | Rita's Italian Ice | PA | Aston | 7 | 4 |
| 1MelwdbTnZOBFCORgaxuw | Ricardo's Italian Cafe | MO | Saint Louis | 80 | 3.5 |
| 25Uww0C0wvF9CZ_3B6vWtA | Enjoy The Mountain | CA | Santa Barbara | 48 | 4.5 |
| 2Bi3ef7c0CrLxHD_-yy_tw | AIM Mail Center | AZ | Oro Valley | 13 | 3 |
| 2dlQX5sP9X6Dlm1MmNOISw | ReAnimator Coffee | PA | Philadelphia | 105 | 4 |
| 2n9HHBxG7yAyAUwXXa49aw | Mighty Dustless | PA | Southampton | 5 | 4 |
| 2oxrOO3c9_mQmqM9kwEm9Q | Total Access Urgent Care | MO | Saint Louis | 6 | 3.5 |
| 2xVsWBNFwZOxIOdd9Mwnww | Cheeseburger In Paradise | PA | Exton | 20 | 2.5 |
| 35dqrE4z9qWctCn749oorg | Rippy's Central Pike Collision Repair | TN | Hermitage | 5 | 2.5 |
| 3BJxm-HnvzdwD1zjmSbmyQ | Golden Chopstick Chinese Restaurant | PA | Philadelphia | 137 | 3 |
| 3S9GYtMOYhbZI4MMPOHT0Q | The Handlebar Boise | ID | Boise | 34 | 4.5 |
| 4dVrWc4Nhuw9apfk9BWEcA | Walmart Supercenter | FL | Tarpon Springs | 29 | 2 |
| 4lcB3QyMEA85UTWFKh9O9A | Eat Mah Taco @ Pal's Lounge | LA | New Orleans | 8 | 4.5 |
| 4iRzR7OaS-QaSxuvYxEGKA | Super Dog | TN | Nashville | 6 | 4 |
| 59jkCUm4yHciKL1KW_1HJA | Roma Pizza & Italian Eatery | PA | Warrington | 25 | 2.5 |
| 5BmQX4UVJY19mMtafMg7JA | Breadland Organic Whole Grain Bakery | AB | Edmonton | 23 | 4 |
| 5BYQP7i7ckUlgSzX4Yhnmw | A-1 Expert Cleaning Services | PA | Levittown | 5 | 5 |
| 6BF3ealzHtoNHSxKTFnTyg | Mister B's Cigar Box | IN | Avon | 13 | 3.5 |
| 6bFx8j2KqPfmT0XEUS1RMg | Center City Emergency Dentist | PA | Philadelphia | 120 | 3.5 |
| 6czqMScx8MRDirv7fxX7pw | Frankie's Fruit Bar | NV | Reno | 34 | 4.5 |
| 6l_DA3uqOox50rBvjrQ_-g | Aloft Philadelphia Airport | PA | Philadelphia | 175 | 3 |
| 7cICRnNhdLv2Wi9i76Miw | Rier Brewery and Tap Room | IN | Indianapolis | 139 | 4.5 |

OSEMN Report Midterm Project

R-Studio code : visualizations for 1st two questions.

```
datawranglingplots.Rmd x yelp.data x
Go to file/function Addins
Source Visual
8 knitr::opts_chunk$set(echo = TRUE)
9
10 library(ggplot2)
11 library(magrittr)
12 library(dplyr)
13 library(tidyverse)
14 library(data.table)
15 ^
16
17
18 ^{r}
19 # load data and calculate the number of businesses in each group
20 yelp.data <- read.csv("format.csv", head = T)
21
22 #print data - only prints 10 first and 10 last rows
23 head(yelp.data, n=20)
24 tail(yelp.data, n=20)
25
26 #plot for review count
27 #a couple outliers can be see in the plot
28
29 #add mean and sd points
30 mean_count <- mean(yelp.data$review_count)
31 sd_count <- sd(yelp.data$review_count)
32 plot(yelp.data$review_count)
33 mean_count
34 sd_count
35 points(x=0, y = mean_count, col = "red", pch = 16)
36 points(x=3, y = sd_count, col = "green", pch = 16)
37 ^
38
39 ^{r}
40 #total review count by state barchart
41 # Log scaling
42 yelp.data$review_count_log <- log10(yelp.data$review_count)
43 # Total log-scaled review count by state bar chart
44 sum.plot <- ggplot(yelp.data, aes(x = state, y = review_count_log)) + stat_summary(fun = sum, geom = "bar")
45 print(sum.plot + ggtitle("Total Log-Scaled Review Count by State"))
46 ^
47
48 ^{r}
49 #average review count by state barchart
50 # Log scaling
51 yelp.data$review_count_log <- log10(yelp.data$review_count)
52 # Average log-scaled review count by state bar chart
53 avg.plot <- ggplot(yelp.data, aes(x = state, y = review_count_log)) + stat_summary(fun.y = mean, geom = "bar")
54 print(avg.plot + ggtitle("Average Log-Scaled Review Count by State"))
55 avg.plot <- ggplot(yelp.data, aes(x = state, y = review_count)) + stat_summary(fun.y = mean, geom = "bar")
56 print(avg.plot + ggtitle("Average Review Count by State"))
57 ^
58
59
60 ^{r}
61 #Count of business name in each state
62 yelp.data.table <- yelp.data %>% group_by(state, name) %>%
63   summarise(total_count=n(), .groups = 'drop') %>%
64   as.data.frame()
65 yelp.data.table
66 ^
67 ^{r}
68 yelp.data.table
36:51 [x] Chunk 2 [x]
Console
```

OSEMN Report Midterm Project

This is the r-studio code for 3rd question analysis.



This report has some data mixed from group project and the 3rd question is the different one and the visualizations developed for 3rd question uses 2 .csv files loaded into phpMyAdmin, and a query written in SQL is executed to relate the 2 tables and the data generated from the query is exported as .csv and loaded into r-studio for generating graph.

References

- e-satisfaction. (n.d.). *7 reasons why customer reviews are important*. Retrieved January 27, 2023, from <https://www.e-satisfaction.com/7-reasons-why-customer-reviews-are-important/#:~:text=Analyzing%20reviews%20left%20by%20your,what%20your%20customers%20truly%20want>
- Convert CSV to JSON using python with new headers*. Stack Overflow. (1969, July 1). Retrieved January 27, 2023, from <https://stackoverflow.com/questions/73089933/convert-csv-to-json-using-python-with-new-headers>
- Rana, D. S. (2022, September 12). *Learn Review Analysis: Why, how, data sources with Free Trial*. Learn Review Analysis | Why, How, Data Sources with Free Trial. Retrieved January 29, 2023, from <https://www.repustate.com/blog/review-analysis/#:~:text=Review%20analysis%20allows%20you%20to,new%20sales%20opportunities%2C%20and%20more.>
- Indexing and selecting data - pandas 1.5.3 documentation*. (n.d.). Retrieved January 29, 2023, from https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy.
- Yelp for Business. (2023, January 23). *Plan, start, grow, and advertise your small business*. Yelp for Business. Retrieved January 29, 2023, from <https://business.yelp.com/>
- PH717 module 9 - correlation and regression*. The Correlation Coefficient (r). (n.d.). Retrieved February 12, 2023, from <https://sphweb.bumc.bu.edu/otlt/MPH-Modules/PH717-QuantCore/PH717-Module9-Correlation-Regression/PH717-Module9-Correlation-Regression4.html>
- GeeksforGeeks. (2021, July 21). *Plot mean and standard deviation using GGPlot2 in R*. GeeksforGeeks. Retrieved February 13, 2023, from <https://www.geeksforgeeks.org/plot-mean-and-standard-deviation-using-ggplot2-in-r/>