# ASSIGNMENT 2 (INDIVIDUAL)

The Subway Reviews Problem

## Instructions

This graded assignment is based on reviews shared by customers on the Yelp platform. The data is released by Yelp for academic use and can be found at [Yelp Open Dataset](https://www.yelp.com/dataset). A subset   
of data is being shared with you for academic use. Please read the [Yelp Dataset Terms of Use](https://s3-media0.fl.yelpcdn.com/assets/srv0/engineering_pages/dc1cabe7cb95/assets/vendor/Dataset_User_Agreement.pdf) before proceeding with the assignment.

The assignment uses names of real companies and individuals as they appear in the dataset. The context laid out, however, is purely fictitious.

Except as indicated, use this document to record all your assignment work and responses to any questions. At a minimum, you will need to turn in a digital copy of this document to your instructor as part of your assignment completion. You may also have additional supporting documents that you will need to submit. Your instructor will provide feedback to help you work through your findings.

**Note:** Though your work will only be seen by those grading the course and will not be used   
or shared outside the course, you should take care to obscure any information you feel might be of a sensitive or confidential nature.

*Complete each assignment part below. Wait to submit the assignment until all parts are complete. Directions to submit your assignment can be found on the assignment page in   
the unit. Information about the grading rubric is available on the assignment page online.   
Do not hesitate to contact your instructor if you have any questions about the assignment.*

# Part One

Problem Setup

Subway CEO John Chidsey received a report from his field officer that customers are unhappy with one of its stores in Milford, CT, where Subway is headquartered. Upon checking the online reviews for this store, he finds that the average rating received by the store is 3.2/5. John is concerned that if a store next to the headquarters is at 3.2/5, stores farther away might be performing even more poorly.

John calls for an urgent meeting of the Head of Customer Service, the Head of Store Operations, the Head of Social Media, and the Chief Data Scientist (your boss). He expresses his concern and urges the team to take measures to improve the average ratings received across all stores in the U.S. to 4.5/5.

His team tells him not to worry by making the following statements:

* Head of Customer Service: “Our ratings are gradually improving, and we will soon reach 4.5/5.”
* Head of Store Operations: “Sandwiches are a tricky business. All sandwich chains suffer from poor customer ratings.”
* Head of Social Media: “The goal of 4.5/5 is unreasonable for national chains like us. Only small, local, and boutique restaurants can achieve such high ratings.”
* Chief Data Scientist: “It is well known that customers make the effort to give a rating only when they are either extremely angry or absolutely delighted with the service. So online ratings are not reliable.”

Your boss, the Chief Data Scientist, gives you two datasets

1. Review Dataset: This data contains details of ratings given by users to restaurants on an online platform. It contains ~400K ratings given to ~2,000 different restaurants along with the date of posting the rating.
2. Restaurant Dataset: This data contains details of restaurants such as name, category, city, and state.

# Part Two

Problem Formulation and Analysis

Answer the following questions.

1. **Part A:** Does the data support the statement made by the Head of Customer Service?

**Part B:** Is this statement valid across all states?

Hint: Construct a plot with year on the x-axis and average rating received on the primary y-axis. Plot a bar chart of number of ratings received on the secondary y- axis.

|  |
| --- |
| Head of Customer Service statement: “Our ratings are gradually improving, and we  will soon reach 4.5/5.”   1. The Data does not align with the statement made by the Head of Customer Service. Subway’s current mean rating of 2.57/5, based on 3,623 reviews,  falls significantly short of the 4.5/5 target claimed by the Head of Customer Service. 2. the statement is not valid across all states. While Subway has 3,623 reviews from 381 locations, the high standard deviation of 1.64 indicates a wide variability in ratings across different locations. This suggests that some locations may perform better, but others receive significantly lower ratings, leading to inconsistent customer satisfaction.   I also created a document to output annual average of subway ratings and saw fluctuations in rating which did not indicate a steady upwards change. |

1. Does the data support the statement made by the Head of Store Operations?

**Part A:** Are sandwiches the only tricky business?

Hint: Identify one or two national competitors of Subway. Create a plot to compare   
the mean and standard deviation of reviews received by Subway with its competitors.

|  |
| --- |
| In the analysis to compare the Head of Operations statement, I saw that Sandwich restaurants have a lower average rating (3.3/5) than non-sandwich restaurants (4.1/5). This supports the notion that sandwich chains face more challenges in achieving high ratings. At 1.2 standard deviation, which is higher than other restaurants for sandwich restaurants suggests greater variability in customer experiences |

1. **Part A:** Does the data support the statement made by the Head of Social Media?

**Part B:** Is it true that average rating decreases as the size of restaurant increases?

Hint: Categorize restaurants with a presence in more than 50 cities as national chains and a presence in only one city as local. Construct a plot to compare ratings received by national chains and local chains.

|  |
| --- |
| 1. National Average Ratings are 3.45/5 with a standard deviation of 1.58. The data supports the claim that national chains face challenges in achieving high ratings. 2. with the given dataset because it only contains data for local chains, defined as restaurants present in a single city. There are no restaurants categorized as national chains |

1. **Part A:** Does the data support the statement made by the Chief Data Scientist?

**Part B:** Is the statement true across all years from 2018 to 2021?

Hint:Create a plot with the ratings on the x-axis and the number of reviews with that rating on the y-axis.

|  |
| --- |
| 1. Most of the ratings are for 1/5 and 5/5, indicating how majority the ratings are connected to strong feelings. 2. The dominance of 1-star and 5-star reviews each year suggests that online ratings are influenced by extreme experiences, making them less reliable for understanding overall customer satisfaction. |

1. Based on the insights generated from Questions 1 through 4, what would be your recommendation to John and the team (at most 100 words)? Your response should include, but may not be limited to the following:

* Is there cause for John to worry? Justify your answer.
* Is the target of 4.5/5 reasonable? If “yes,” justify your answer. If “no,” what do you think would be reasonable and why?
* Support your recommendations by quoting appropriate numbers computed from the data.

|  |
| --- |
| John should be concerned because Subway’s average rating of 2.57 is much lower than its  competitors and far below the target of 4.5/5. Achieving 4.5/5 is very difficult for a national chain like Subway because customers tend to leave reviews only when they are extremely happy or upset.  For example, in 2021, 64% of reviews were either 1-star or 5-star, showing that customer satisfaction  is inconsistent. A more realistic goal would be 3.5/5, which Subway can work toward by focusing on improving service quality and addressing common customer complaints across its locations. |

1. Please upload an additional single .pdf file along with this document with three sections:

* Your Python code appropriately commented
* Additional instructions for running your code, if necessary
* Snapshots of the output of your implementation

1. **Bonus question:** What additional insight on ratings can you provide John?

* Please write a brief comment (at most 50 words).
* Support your argument with up to three data visualizations.

|  |
| --- |
| The Yearly Ratings Distribution shows that most reviews are either very positive or very negative, which means customers are having inconsistent experiences. The Bottom 10 Locations chart helps identify the stores with the lowest ratings, so the team knows exactly where to focus their efforts to improve. Finally, the Monthly Trend of Ratings points out seasonal dips, suggesting there are specific times of the year when Subway might face challenges like staffing or operational issues.    Visual 1:   * Visual 2:   Visual 3: |

*To submit this assignment, please refer to the instructions in the course*.