

**Technical Task – Stage 2**

The following test is divided into three sections. The order of completion is up to you. Your goal is to complete as many of the given tasks as possible within the deadline. At least one fully completed section is required to be considered. You can find the details of the tasks within each section. At the end, you are expected to prepare a short presentation and provide all source codes for review.

The presentation should include the objective of the task, how it was completed, and the end result (possibly with graphs or images).

The source codes can be shared via GitHub or as a zip file.

Please send the presentation, source codes, and datasets with your name and surname (for zip file) before the deadline for the review.

Best of luck.

**Expected completion duration is 7 days after receiving the e-mail.**

# Section 1:

In this section, you will use the Yelp reviews dataset to answer the questions below. You can access the dataset via [this link](https://eigcom-my.sharepoint.com/:u:/g/personal/elnur_bakhtiiarov_researchlab_digital/EQ_VimwAvKFKoTYo7aCHvgoB3BdP0n2zQMTCX0wTDnP9rQ) (Section 1 data file).

*Please answer the following questions:*

1. How many unique restaurants could be found in this data set? (Hint: Use the [Business\_ID] column for this evaluation.)
2. Which restaurant received the highest number of reviews? What about percentage-wise?
3. Which cities have got at least one 5-star review in Nevada (NV) state?
4. Which city has the highest number of reviews in the Business Category of “Hotels & Travel”? What about percentage-wise?
5. At what day of the week people are more likely to post their reviews?
6. Showcase if there are any trends regarding restaurant performance as time goes by.
7. Based on analyzed data showcase if there are any steps that the restaurant can take to improve their public appeal.
8. Bonus Question – Based on this data set which user had the highest cumulative travel distance? What distance has been covered by him/her?

*Dataset Columns Description:*

'Review\_Date' – the date when the review was posted by the user

'Review\_Text' – text of the review

'User\_ID' – Unique Identification Number of users, who made this post

'Business\_ID' – Unique ID of business. Please note, that this column distinguishes businesses with the same names (for instance chain of restaurants)

'Business\_Name' – Official business name

'Business\_Category' – The category in which this business operates

'City' – City location

'State' – State Location

'Latitude' – X coordinates of business

'Longitude' – Y coordinates of business

'Avg\_Business\_Star\_Rating' – Rating review left by user.

# Section 2:

Part 1:

Create an algorithm that generates an NxN matrix filled with the numbers in a specific pattern starting from the center. This matrix should follow the pattern illustrated in the attached image.

Input: An integer N representing the size of the square matrix (NxN).

Output: A NxN matrix filled with numbers starting from 1 at the center and moving outward in a clockwise manner increasing sequentially, filling the matrix completely.

Example:

For N = 5, the output should look like:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 21 | 22 | 23 | 24 | 25 |
| 20 | 7 | 8 | 9 | 10 |
| 19 | 6 | 1 | 2 | 11 |
| 18 | 5 | 4 | 3 | 12 |
| 17 | 16 | 15 | 14 | 13 |

Part 2:

Calculate and print the sum of the numbers along both diagonals of the result matrix (primary and secondary).

Input: Result matrix from the first part.

Output: Sums of the diagonals separately.

# Section 3:

You are provided with Distributed Acoustic Sensing data (Distributed acoustic sensing - Wikipedia). The fiber cable is sticked to the outside of pipe which goes down to the earth for couple of kilometers. Inside of the pipe is liquid flow. Spatial discretization is 1m (each row in the dataset corresponds for specific depth) and temporal resolution is 1 minute (each column in the well corresponds to a minute).

Your task is to implement some coding with further interpretations/observations of results:

• Brief description of Distributed Acoustic Sensing (DAS) working principle and cases of its applications

• Do data observations: some statistical analysis to get to know your data as well as various types of visualizations and insights from them

• Implement at least one algorithm (coding in Python, MATLAB, etc.) for line detection in the given signal using spatiotemporal data. Implementation of additional algorithm and its comparison with the first one is highly welcomed

• Describe properties of the lines in case they are detected (algorithmically)

Eventually, you need to prepare slides and present your results where you show the workflow of your project in details as well as how algorithms work.

You can access the datasets via [this link](https://eigcom-my.sharepoint.com/:u:/g/personal/elnur_bakhtiiarov_researchlab_digital/EQ_VimwAvKFKoTYo7aCHvgoB3BdP0n2zQMTCX0wTDnP9rQ) (Section 3 data file).

Note: abovementioned things are a minimum program to be done. Feel free to explore beyond!