Faculty of Computers and Data Science – Alexandria University
Course Name: Data Mining and Analytics
Course Code: 02-24-00206
Semester: Spring 2020/2021

# Final Project

The main goal of this project is to practice and apply clustering algorithms you have learned to real-world tasks.

Your task is to cluster the dataset into an optimal number of clusters, your steps will be as follows:

- Choose your preferred clustering real-world application
- Download any suitable datasets from (Kaggle, UCI repository, etc.)
- Assess and clean your data if needed ( **you can use preprocessed data.)**
- Import the required libraries and perform:
    1. **Agglomerative Hierarchical Clustering Algorithm**, find K clusters, using single linkage strategy, considering Euclidean distance as the distance measure.
        - Visualize your results.
    2. **K-Medoids Clustering Algorithm** find K clusters, use the Manhattan distance.
        - Visualize your results
- Conduct a comparison between the 2 Algorithms results.
- Team Members: only 2 students

Faculty of Computers and Data Science – Alexandria University
Course Name: Data Mining and Analytics
Course Code: 02-24-00206
Semester: Spring 2020/2021

- Please create a **pdf** report called "**Project_Report",** contains:
  - Your dataset link.
  - Dataset description, and your Target from this application.
  - Your plotted graphs for each algorithm.
  - Explain your results and insight by describing your plotted graphs.
  - Comparison between Agglomerative Hierarchical Clustering Algorithm, and K-Medoids Algorithm results.
  - Name, ID, Group and briefly describe the role of each member.

- Submission Details:
  - You should submit your code notebook containing your comments (one drive link)
  - Pdf version of the notebook
  - Project_Report.pdf
  - no plagiarism is allowed
  - Deadline: **10/6/2021 11:59 pm**

**Good Luck**