The purpose of the Mini Project is to reinforce skills that have been covered in modules 4,5,6.

Problem 1: Build a regression model to predict "median\_house\_value". Try out the regression algorithms we used in class

Data - https://www.kaggle.com/datasets/camnugent/california-housing-prices

Problem 2: Build a classification model to predict stroke (= 1 if a person had a stroke else 0). Try out the classification algorithms we used in class

Data - https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset

Problem 3: Build a clustering model to segment customers

Data - https://www.kaggle.com/datasets/imakash3011/customer-personality-analysis

The following items should be part of your submission:

1. Git Version Control:

- Identify a suitable folder structure to store your code, data, models, outputs

- Initialize a Git repository for the project to track code changes

- Commit code changes related to data preprocessing, analysis, and database integration.

2. Environment Management:

- Use environments to manage project dependencies and ensure a consistent environment for all contributors.

- Use a requirements.txt file to document and share the required dependencies (should be part of the git file).

3. Perform the necessary data preprocessing and EDA

4. Modelling and results. Try to save the model in pickle format after training it and reload the model when you predict on the test dataset. (For this assignment: include your model pickle file in the git repository)

5. Conclusion and Reporting: Prepare a report in any presentable format to summarize the eda, model, and support results with visualization and stats.

Output expected:

1) Github link to a repository that contains the following:

i) code

ii) info to reproduce the environment

iii)All your code files

iv) final report can be of any presentable format - html, ppt, word, pdf (not jupyter notebooks with code)

v) Readme file - Readme file should contain a summary of the project and the steps to reproduce your code

2) 8 minute presentation of the results(not more than 8 minutes)

Presentation to be held on August 28th