**In Class Assignment2 (100 points)**

**CS430-01**

**Machine Learning on Cloud**

**Fall 2022**

**Exploratory Data Analysis, train the data using regression techniques, Evaluate the result with Principal Component Analysis (100 points)**

**Goal:** The goal of this assignment is to use Pandas/Matplotlib/Seaborn to explore the dataset, use Sciket-learn libraries do the data preprocessing, split the data and train models with Regression Technique, compare the result with PCA regression.

**The dataset you will have to use :**

<https://www.kaggle.com/code/kingsleyofori/housing-new-dataset/data>

**The target variable is housing price calculation. Detect the best model for housing price calculation out of:**

* 1. **Linear Regression**
  2. **Random Forest Regression**
  3. **PCA with Linear Regression**
  4. **PCA with Random Forest Regression**

**Make sure scale your data before using for PCA**

Ensure your notebook is organized and has proper **Markdown comments on data preprocessing, model performed,** etc. You can assume that after someone see the raw notebook, so it should be clear.

1. **Assignment Submission:** Upload a link to your GitHub repository for the project in the area provided in Moodle by the deadline specified.