***CS 395 Selected Topics in Computer Science -1***

**Project Details**

* Project team has to be 2 students
* each team must choose public dataset to start working
  + min 200 records
  + min 4 features in each record
* Using Python:
  1. Visualize dataset using different plots
     1. Rug Plot
     2. Scatter Plot
     3. Histogram Plot
     4. Box Plot
  2. Get **mean** and **variance** of dataset
  3. Get different random samples from dataset min 50 samples (10 observation as example)

for each sample calculate

* + - 1. sample mean
      2. sample variance
  1. Get covariance between each pair of features in dataset
  2. Analysis Scatter plots according to covariance matrix
  3. Compare mean of samples means and mean of dataset
  4. Plot samples means vs samples number and get variance of sample mean
  5. Change number of observations (n) of sample and repeat previous step again and note how plot differs with small and large n
* Each team must print
  + Different plots
  + Analysis report about data (include all results of mean, variance…..)