

# Cx1115 : Programming Quiz : Sample

- Submit a *single* Jupyter Notebook file named **MatID\_Quiz\_solution.ipynb**, where “MatID” is your Matric Number.
- Submit the Notebook file at the same place where you found this problem - in your own Lab Group Course Site.
- Download the data file (in CSV format) posted corresponding to this Lab Quiz. You will need it for the problems.
- You have exactly **60 minutes** to complete the Quiz and submit your solution. Late submission will be penalized.

## Context

The dataset contains information about the Height, Weight, Diameter and Length of Abalones. The goal is to connect the four variables and predict the Length of Abalones using the other variables in the dataset, using **Linear Regression** model.

Plot figures and print outputs as you usually do in a Jupyter Notebook while doing exploratory analysis. If you are supposed to “Comment” something extra, use a Markdown cell in the notebook, similar to what you did for your graded exercises.

## Problems

### Problem 1 : Exploratory Analysis

- Print the overall statistical description of the data and plot standard statistical distributions for each variable. Comment : Which variable has the most “Normal” looking distribution? Which variable has the maximum outliers?
- Print the mutual correlation amongst all the variables using a correlation heatmap. Comment : Which variable has the highest absolute correlation with Length? Is such a strong correlation helpful in predicting Length?

(5 + 5)

### Problem 2 : Uni-Variate Linear Regression

- Partition the dataset randomly into Train and Test datasets - 70% for Train and 30% for Test. On the Train set, fit **three uni-variate** Linear Regression models for Length against each of the three other variables. For each such linear model, print the coefficients of the model you fit, and plot the regression line on a scatterplot.
- Predict Length using each of the three models on both Train and Test datasets. Plot the actual values of Length against the predicted values of Length in each case. Print the Mean Squared Errors and  $R^2$  values in each case.
- Comment : Which of the three uni-variate linear models is the best in terms of predicting Length? Briefly justify.

(9 + 9 + 2)

### Problem 3 : Multi-Variate Linear Regression

- Partition the dataset randomly into Train and Test datasets - 70% for Train and 30% for Test. On the Train set, fit a **single multi-variate** Linear Regression model for Length against all of the three other variables.
- Predict Length using this model on both Train and Test datasets. Print the Mean Squared Error and  $R^2$  values.

(5 + 5)