**Assessment:**

**Effective Data Analysis by Applying Descriptive Statistics**

| **Section B (80 points total)** |
| --- |

**Analyse the Boston Housing Dataset**

The [Boston Housing Dataset](https://drive.google.com/file/d/1prxjaBT-AUsSFBgdCXYlcjGMA76adTvy/view?usp=sharing) is a derived from information collected by the U.S. Census Service concerning housing in the area of [Boston MA](http://www.cs.toronto.edu/~delve/data/boston/bostonDetail.html). The following describes the dataset columns:

* CRIM - per capita crime rate by town
* ZN - proportion of residential land zoned for lots over 25,000 sq.ft.
* INDUS - proportion of non-retail business acres per town.
* CHAS - Charles River dummy variable (1 if tract bounds river; 0 otherwise)
* NOX - nitric oxides concentration (parts per 10 million)
* RM - average number of rooms per dwelling
* AGE - proportion of owner-occupied units built prior to 1940
* DIS - weighted distances to five Boston employment centres
* RAD - index of accessibility to radial highways
* TAX - full-value property-tax rate per $10,000
* PTRATIO - pupil-teacher ratio by town
* B - 1000(Bk - 0.63)^2 where Bk is the proportion of blacks by town
* LSTAT - % lower status of the population
* MEDV - Median value of owner-occupied homes in $1000's

You are required to do following tasks:

1. Load Dataset from Boston Housing Agency into a DataFrame. (*10 points*)
2. For the "Median value of owner-occupied homes" provide a boxplot. (*10 points*)
3. Provide a histogram for the “Charles river variable”. (*10 points*)
4. Provide a boxplot for the MEDV variable vs the AGE variable. (Discretize the age variable into three groups of 35% or less, between 35 and 70% and 70% and over) (*10 points*)
5. Provide a scatter plot to show the relationship between Nitric oxide concentrations (NOX) and the proportion of non-retail business acres per town (INDUS). What can you say about the relationship? (*10 points*)
6. Create a histogram for the pupil to teacher ratio variable (PTRATIO) (*10 points*)
7. Is there a significant difference in median value of houses bounded by the Charles river or not? (CHAS) (T-test for independent samples) (*10 points*)
8. Is there a difference in Median values of houses (MEDV) for each proportion of owner occupied units built prior to 1940 (AGE)? (ANOVA) (*10 points*)