

# Introduction To Data Science

## Midterm Project

**Descriptive Statistics:** It is a study of data analysis to describe, show or summarize data in a meaningful way. It involves the calculation of various measures such as the measure of center tendency and measure of spread. One of the first things we can do with a new data set is to explore it and see what the main characteristics are. To do this you can look at e.g:

- Import the data set(data\_1) as csv and print the data set.
- Find the shape of the data set.
- Show the attributes name of the data set.
- Find the types of data for all attributes.
- Measure of center (mean ,median and mode) for Loan and Interest\_rate attributes.
- Measure of Spread (range and standard Deviation) for Loan and Interest\_rate attributes.
- Find the mode for Credit\_Score attribute.

Now import another data set(data\_2) as csv and print the data set.

- Find the missing value for all attributes.
- Detect the outlier as a missing value.
- Annotate h as 1, m as 2 , and l as 3 from “Type” attribute.
- Recover missing values by the following strategies for Rooms and Price attributes.
  - I. Delete the rows with missing values.
  - II. Recover missing values with the mean value.
  - III. Recover missing values with the median value.
  - IV. Recover missing values with the mode value.

### Remember:

- Make documentation of your project.
- Do not add comments to your code.
- I will open the link in Teams and you need to upload the code as txt file in the given link on or before Tuesday ( 8/11/2022)