# Python Assignment – EDA and Visualization

Please create the solutions for these assignments in a jupyter notebook. Name the notebooks as your Name and Emp ID and submit

## Question 1

### Use file F2 to perform tasks 1-6.

1. Find mean and median fare. Also find SD in fare. Using data visualization tools, Check if fare is normal, positively skewed or negatively skewed.
2. Find most common age using python. Divide age into segments and find average fare per segment.
3. Draw box plot of age on categorical variable class.
4. Using hypothesis testing, measure if upper class passengers have a higher survival rate. Mention your null and alternate hypothesis in separate cell in comments.
5. Check interdependence of gender and Pclass on survival to decide whether they should be included in final survival prediction model (You don’t have to build a classification model).
6. Check correlation between variables Fare and Age using visualization tool. Use statistical measures to check if your observation about correlation is correct.

## Question 2

### Use file F1 to perform tasks 7-13.

1. Using Python, find percentage of missing values in each variable. Store the result in a table and display the table sorted on % of missing values descending.
2. Create Gender and Education level box plots to check average income and if outliers exist on basis of income. Write down your comments about outliers in a separate cell.
3. Using python pivot tables, find number in approved and not approved categories on basis of gender
4. Using python pivot tables, find probability of loan getting approved and not approved on basis of credit history
5. Using a suitable visualization tool, compare number of applicants and applicants getting loan on basis of gender and credit history categories. You can use multiple charts of show the data in one chart as per your understanding
6. Create groups on basis of Gender and Property Area. Show the contents of one group. Find and show mean, SD and sum of fares for each group in a table.
7. Fill missing values in loan amount

## Question 3

1. Create 10 csv files in a single folder with same format and headers. Use python to write contents of each file into a single file. Header should be written only once. You are free to choose name of columns and number of rows/columns.