# SHAI ASSAIGNEMENT

# Youssef Sammha

***First:*** Include libraries (pandas, matblotlib, numby) in order to read, clean, and represent data graphically.

In the paragraph (#Basic Data Exploration), we read the data using the ***read\_csv()*** function, displayed the number of columns and lines using the ***shape()*** function, and displayed the number of null values ​​in each column using the ***isnull().sum()*** functions.

***Second:*** In the cell (#Descriptive Statistics), we calculated the largest and smallest value, the standard deviation, the mean, the median, and the mode for the TotalPay column using the functions ***max()***, ***min()***, ***std()***, ***mean()***, ***mean()***, ***mode()***, respectively, and calculated the range by subtracting the largest value minus the smallest value in the column.

***Third:*** In the cell (#Fill NaN values ​​in the 'TotalPay' column with the mean )

We fill the null values ​​in the TotalPay column with the average using the ***fillna()*** function..

We used the average in order not to add random values ​​that distort the data and to obtain results that are as close to reality as possible. It is a method inspired by reality (when an employee applies to work in a company, he is informed of the average salaries in the company or in the position to which he is applying ).

***Fourth:*** In the cell (#Basic Data Visualization) we drew a bar chart to illustrate the distribution of salary values ​​using the ***histplot()*** function from the ***seaborn*** library..

Then we drew a pie chart to represent the frequency of different values ​​of salaries using functions

***value\_counts().plot(kind='pie', autopct='%1.1f%%')***

***Fifth:*** In the cell (#Grouped Analysis and #Simple Correlation Analysis), first we grouped the salary values ​​according to the department to which the employee belongs using the ***groupby()*** function and calculated the average salaries for each of the departments..

Then we calculated the correlation of the values ​​between the TotalPay and TotalPayBenefits columns using the ***corr()*** function and then we plotted a scatter to display the relationships between the values ​​using the ***scatterplot()*** function..