**Assignment DAY 09**

dataset.head()

To see the first 5 records in the data sheet.

dataset.head(10) prints first 10 records

dataset.tail() prints the last 5 records

dataset.tail(10) prints the last 10 records

dataset.columns col in the sheet

we can identify the no of columns in the data.

dataset.info() complete description (information about the sheet)

dataset[MonthlyIncome].mean() to find the AVG

By this we can get the average salary for the employees to take the decision.

dataset[['MonthlyIncome','EmployeeCount','YearsAtCompany']].mean()

dataset[['MonthlyIncome','EmployeeCount','YearsAtCompany']].median()

we can get the MIDDLE VALUES .

dataset[['MonthlyIncome','EmployeeCount','YearsAtCompany']].mode()

to check the MOST FRQUENTLY ACCOURED.

dataset[['MonthlyIncome','EmployeeCount','YearsAtCompany']].var()

HOW MUCH VARIATION IN THE DATA.

dataset[['MonthlyIncome','EmployeeCount','YearsAtCompany']].std()

By this we can analyze the Consistency of the data.

dataset[['MonthlyIncome','EmployeeCount','YearsAtCompany']].sum()

dataset[['MonthlyIncome','EmployeeCount','YearsAtCompany']].skew()

to get the Symmetric data.

dataset[['MonthlyIncome','EmployeeCount','YearsAtCompany']].kurt()

to identify the Peakness of the data.

plt.boxplot(dataset.MonthlyIncome) it will give the graph

we can identify the Outlier.

plt.scatter(dataset.MonthlyIncome,dataset.YearsAtCompany)

by using the scatter we can identify the Outlier in dataset.

plt.hist(dataset.MonthlyIncome)

By using hist method we can get the skewness in the form of pictorial.

So by the diagram we can see the skewness Is left side, so It means it is a **positive skewness**.