ANTWAK ASSIGNMENT:

1. I have given 4 datasets for the analysis
2. I have uploaded the datasets in the form of dataframes and then i have printed first 5 columns of all the dataframes to know the details.
3. Observations:

I:A)1st dataset have 'RowNumber', 'CustomerId', 'Exited' variables.Shape

B)2nd dataset have 'RowNumber', 'CustomerId', 'Surname', 'Geography', 'Gender', 'Age' variables.

C) 3rd dataset have 'RowNumber', 'CustomerId', 'Tenure', 'CreditScore', 'Balance', 'EstimatedSalary' variables.

D) 4th dataset have ‘RowNumber', 'CustomerId', 'NumOfProducts', 'HasChckng',

'IsActiveMember' variables.

II:A)First 3 dataframes have same length=10019, last one have =10021.

B)First 3 dataframes have same unique values of customer id variable =

10000,last one have 10019.

C)All four dataframes have ‘RowNumber', 'CustomerId' variables same.

III:A)After merging total variables are 'RowNumber', 'CustomerId', 'Exited', 'Surname', 'Geography', 'Gender','Age', 'Tenure', 'CreditScore', 'Balance', 'EstimatedSalary','NumOfProducts', 'HasChckng', 'IsActiveMember'

B)Total of 14 variables and length=10002. Missing values(Nan) values are available in 2 variables 'CreditScore'=3, 'Balance'=3617.

C)Missing values are imputed with mean of the respective variable.

D)So after merging 10000 rows × 14 columns with unique values of customer id,

Without unique values 10002 rows x 14 columns.

IV: A) After visualization:

a)Male customer are more compared to female.

b)Customers from western = central geography and east.

c)Users exited are almost equal to 25% of Users stayed.

d)High Percentage of Users in west geography are stayed.