# CMS COLLEGE OF ENGINEERING AND TECNOLOGY

# Appachigoundanpathy, Kummitipathy , Coimbatore, Tamil Nadu 641032

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING.

WEB PHISHING DETECTION (ASSIGNMENT 2)

DATE : 30-09-2022

PROBLEM: PERFORM TASKS ACCORDINGLY

NAME :SARANYA R

**OUTPUT**:

# **SCREENSHOTS:**

#### 1.Download the Dataset

## 2.Load the Dataset

In [1]: import numpy as np import pandas as pd import seaborn as sns import matplotlib.pyplot as plt import sklearn

In [2]: data = pd.read\_csv(r"C:\Users\chand\Downloads\IBM-Project\Assignments\Ass-2\Churn\_Modelling.csv")

## 3.Perform below Visualizations

**Univariate Analysis** 

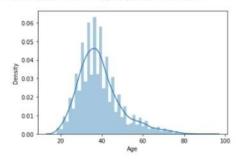
## Univariate Analysis

#### In [3]: sns.distplot(data['Age'])

C:\Users\chand\anaconda3\lib\site-packages\seaborn\distributions.py:2619: FutureWarning: `distplot` is a deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with similar fle xibility) or `histplot` (an axes-level function for histograms).

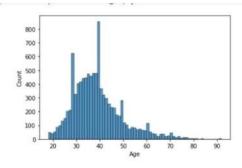
warnings.warn(msg, FutureWarning)

Out[3]: <AxesSubplot:xlabel='Age', ylabel='Density'>



In [4]: sns.histplot(data['Age'])

Out[4]: <AxesSubplot:xlabel='Age', ylabel='Count'>

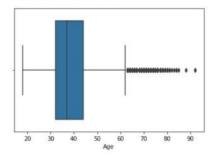


#### In [5]: sns.boxplot(data['Age'])

C:\Users\chand\anaconda3\lib\site-packages\seaborn\\_decorators.py:36: FutureWarning: Pass the following variable as a keyword a rg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit ke yword will result in an error or misinterpretation.

warnings.warn(

Out[5]: <AxesSubplot:xlabel='Age'>



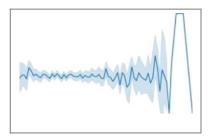
## Bi - Variate Analysis

In [6]: sns. I JnepJH (data[ 'Age '], dat a [ 'renure '])

:c • \users\charid\aria \"onda3t1ib\site-packagés\seaborn\\_decoi ators.py: 3s: FutureMa'rnTng: Pass. the..l-ollowing variable's as kéyukiñd args: x., y.. Crom ver'sion e:.\*z, 'the' on'ly valid'posi'tional argument will be 'data', and passing bther argdments Without añ..éxpli'cit keyword mill. result in an error or m'isiriterpretation.

warnings.warn(

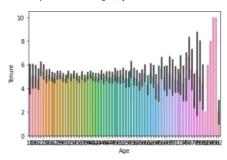
 $\label{eq:out_approx} \mbox{Out} \ [6] : \quad \mbox{``Axessubplot: xlabel = 'Age', ylabel= 'Tenure')}$ 



in [7]: sns.barplof (data ['age'],data ['Tenure'])

C:\users\chand\anaconda3\1ib\site-packages\seaborn'\: decorators.py:36: FutureHaenIng: Pass .the following 'variables as .keywor.d ar gs: x; y.- Froic version '\(\frac{1}{2}\): 13j' 'the \(\text{amy val4d positional argument -wi.II be "data"; and passing other argument s without: an exp11 \(\text{wit; ke rd uS11 result . in...an error or misinterpretation.}\)
aarnihgs.'warn(

Out[7]: «AxessubpJot xlabel•'Age, ylabe1•'tenure'>



In [B]: sns.scat terplot (data['Age'j, data['IeR ure'j)

C',\usérs\chdnd\#ñaconda3\lib\site-packages\seaborn\"deforators.'#y:3'6: "FutureWzrnifg: Pâss the fdll0wing variables as keyword âr gs: x, y Fr'a ve'rsioA 8, 12, 'the "only valid positional arguiïént. uill be; /"data\*, and passing: other argun 'nts.' without an exg'licit keyuord'-will 'rgsult in' an error ar qisIntei-pretat'ion:
uarnIrigs. Varn(

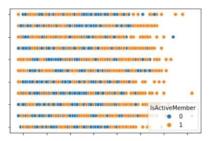
Out B): <a href="AxesSubplot">AxesSubplot</a> • xlabel- 'Age', y1abel- 'Tenure ">

# Multi - Variate Analysis

 $\label{local_prop} $$ In [9]: sns.sc atterp1ot(data['Age'], data[*Tenure'], hue=data['Is4ct iveMeober']) $$$ 

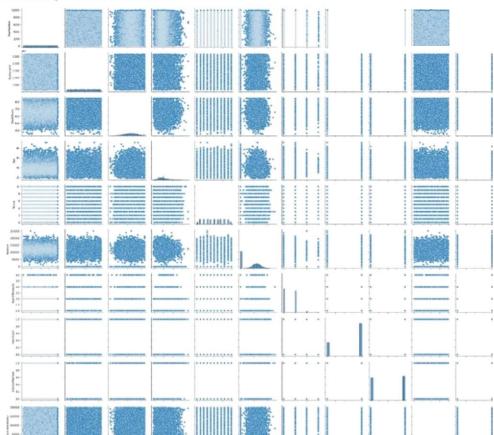
.C,\Usens\chand\anaconOaJ\lih\site-packageslseaborn\ decorators.py:36::'FutureNarning:.Pags thp following variables as keyword ar gs: "xx, y-. rro ti versiori e.a2;: 'th'e o'n1'y valid. positional .'argwient: all I be -data" an'd passing:-other argii ents without. an...explicit keyvoe0 "will result In an erroe. or uis"interpretat1on. namings .mam (

Out [9]: < AxesSubplot: xlabel='nge', ylabeJ='Tenure'>









## 4. Descriptive Statistics

#### In [ LL] : data .mean( )

c: \Users\chand\App0ata\Local\Temp\1pykernel\_79s8\5319e3386.py:1• Future taming: Or opplng" crf nuisance columns th DataFraae reductions (with 'nueer1c\_on1ya one'-) '1s degrecated; In a future version this "will raise TypeError. Select only valid columns before calling the reduction.

data. aeant)

Out[It] RowNumber 5.000s00e+e3 1.569B94e+67 Custonerld CreditScore 6.565288e+02 Age 3.B921s0e+61 Tenure Balance 5.81280<del>0a+00</del> 7.6a85s9e+0a NuxofProducts 1.5302<del>00 TOO</del> HasCrCard 7.655800e-61 isAct ivef4enber 5. IS 1084c-01 EstinatedSa1ary 1.000902e+05 2.637a00e-61

Exited dtype f1oat64

#### In [T2]: data .median()

c:.\Users\chand\Ag@ata\i-rra1\Tel\1py?erne1\_7968\4184645713.py:1: Futureuarning: Dropping of nuisance co1uiens in DataFraaie reductions (n1th 'nuaer'tc\_on1y-done') 1s deprecated; in a future .version this will .ra1se TypeError. se1ect only va11d co1uons before callings the reduction.

Out [i2]: Ro 8unber s.e8eseee•e3 custoneria 1.56987ae+a7 creditscore 6.szeeeee+e2 Age Tenure 3.7eaa68e+01 9.719854a+B4 Balance NunofPnoduct s HastrCard 1.000000r•80 zsActivenember 1. 001939e+05 Estinatedsalary Exited

#### In [13] : data.code ()

0	1	1SS6S70\	Smith	850.0	France	Male 37.0	2.0	0.0	1.D	J.0	1.0	24gZ4.'
1	2	1SS6fi708	NaN	NaN	NaN	NaN NaN	NaN	NaN	NaN	NaN	NaN	Nz
2	3	15S857i4	NaN	NaN	NaN	NaP NaN	NaN	NaN	NaN	NaN	NaP	N*
3	4	1S56577B	NaN	NaN	NaN	NaN NaM	Nat	NaN	NaN	NaN	NaN	N*
4	5	1S565796	NaN	NaN	NaN	NaN NaN	NaN	NaN	NaN	NaN	htañt	N*
eess	9090	15815828	NaN	NaN	NaN	NaN NaN	NaN	NaN	NaN	NaN	NaN	Nz
9997	9888	1581li65tl	NaN	NaN	NaN	NaN NaN	Na/'/	NaN	NaN	NaN	NaN	Na
eses	9999	15815660	NaN	NaN	NaN	NaN NaN	NaN	NaN	NaN	NaN	NaN	Na
9g99	10000	15815690	NaN	NaN	NaN	NaU NaN	NaN	NaN	NaN	NaN	UaN	Na

## 5. Missing Values

## i« [i4): data.isnull().any()

Out [Id	False
---------	---

# There are no missing values

# 6. Handling Outliners

I n [ \s ] : data .quant ile ( [e . i j )

" , '	R	Re v inch+r	Cussomsda	Crean been Age 7sn	ur+ BHan	cc t4uw C	OfProdum HssCfisza	lsAnuvtht	cmb*r Esdrnaoa	Ssls y bit+a	
	0.1	1000.9	15591167.1	024.0 21.0	4.0	0.0	10	0.0	0.0	20273.58	0.0

En [Z7]: data.quantiLe([0. 1, 0. 5])

Out[17]:		RowNumber	Customerld	CreditScore	Age	Tenure	Balance	NumOfProducts	HasCrCard	IsActiveMember	<b>EstimatedSalary</b>	Exited
	0.1	1000.9	15591167.1	521.0	27.0	1.0	0.00	1.0	0.0	0.0	20273.5B0	0.0
	0.5	5000.S	15690T38.0	652.0	37.0	5.0	97198.54	10	.0	1.0	100193.915	0.0

 $\ln$  [  $i\,B$  ) : data .quantile ( [8. i, 0. 9 ] )

utl18]:		RowNumbet	Custon'ierld	CraaitSoora	Age	Tenure	Balance NumD	Balance NumDfRrodoms HasCrCard IsAedvgMember Estimated Salary Exited									
	0.1	1000 8	155B1167.1	524.0	27.0	4.0	0 000	10	0.0	00	20273580	0.0					
	0.9	9000.1	15790630.7	778.0	53.0	9.0	14g244.7gZ	10	1.0	1.0	179674.704	1.0					

# 7.Perform Encoding

In [i9]: £•roa sk! earn ii»fiior\ preproc essing

in [2B]: Ie = preprotessing.tabe1fincoaer()

I n [ 21] : oneb = preproc es st rig. oneuotEncoder( )

En [ 22 ] : data[' Age'] - te .£ it\_t ran s I arm dad a I' Age'])

In [ 2 7 ] : data. head( )

Out[23]:	RowNumber	Customerld	Surname	CreditScore	Geography	Gender	Age	Tenure	Balance	NumOfProducts	HasCrCard	IsActiveMember	Estimated Sala
	0 1	15634602	Hargrave	619	France	Female	24	Z	0 00	1	1	1	101348.88
	1 2	15647311	Hill	608	Spain	Female	23	1	83807.86	1	0	1	112542.58
	<b>z</b> 3	1M1g04	anio	502	France	Femae	24	8	1598B0.80	3	1	0	113931.57
	4 4	5701354	8oni	B99	France	Female	21	1	0 00	2	0	0	9382663
	4 S	15737888	Michell	BPO	Span	Fsmale	25	2	12551082		1	1	79084 10

# 8. Split into Dependent and Independent variables (X and Y)

In [24]: x • Oata.iloc[:,B:12]

In (25]: x

Out[25]:		RowNumber	Customerld	Surname	CreditScore	Geography	Gender	Age	Tenure	Balance	NumOfProducts	HasCrCard	IsActiveMember
	0		15034602	Hargrave	819	France	Femak	24	2	0.00	1	1	1
	1	2	1S647311	Hill	608	Spaip	Femak	23	1	83807.88	1	0	1
	2	3	15819304	O«ia	502	France	Female	24	8	15IXI80.80	3	1	0
	3	<	is70ia5*	Bob	BBC	Fmme	Ferna/e	21	1	0.00	2	0	0
	4	5	15737888	Mitchell	850	Spain	Femae	25	2	12551082	1	1	1
	9995	9996	15808229	i	771	France	Mde	21	5	0.00	2	1	0
	9996	9897	15580802	Johnstone	518	France	Mâle	17	10	57369.81	1	1	1
	9997	9998	15584532	Liu	709	France	Female	18	7	0.00	1	0	1
	9998	98B8	15882355	Sabbaoni	772	Gotnan'r	Mae	24	3	7507531	2	1	0
	9999	UDIO	15028319	WBke/	792	France	remak	10	4	130142.79	1	1	0

10Ql0 rows x 12 columns

In [26]: y - data['Balance']

# 9. Scale Independent values

## 10. Split the data into train and test

```
In [12]: from sklearn.model_selection import train_test_split x_train, x_test, y_train, y_test = train_test_split(x_scaled, y, test_size = 0.3, random_state = 0)
In [13]: x_train
Out[13]: array([[ 0.92889885],
                    [ 1.39655257],
[-0.4532777 ],
                    [-0.60119484],
                    [ 1.67853045],
[-0.78548505]])
In [15]: x_train.shape
Out[15]: (7000, 1)
In [16]: y_train
Out[16]: 7681
                    146193.60
                   0.00
160979.68
            9031
            3691
            202
                          0.00
                    143262.04
            5625
                    120074.97
114440.24
           9225
           4859
            3264
                    161274.05
           2732 108076.33
           Name: Balance, Length: 7000, dtype: float64
```

\*\*\*\*\*\*\*THANKING

YOU\*