Assignment -2

Data visualization

Assignment Date	23 September 2022
Student Name	Gokulakrishnan G
Student Roll Number	2116190701053

Code:

https://colab.research.google.com/drive/1ffC8_O69YQaIRvca5O9phrDEnt38iAKI?usp=sharing#scrollTo=OdlnHPy6WjF6

Download and import dataset

ASSIGNMENT 2 – DATA VISUALISATION

```
[10] import pandas as pd
     import numpy as np
     import seaborn as sns
     import sklearn
     import matplotlib.pyplot as plt
```

```
[2] df = pd.read_csv("/content/drive/MyDrive/Churn_Modelling.csv")
```

```
df.dropna()
```

	RowNumber	CustomerId	Surname	CreditScore	Geography	Gender	Age	Tenure	Balance	NumOfProducts	HasCrCard	IsActiveMember	EstimatedSalary	Exited
0	1	15634602	Hargrave	619	France	Female	42	2	0.00	1	1	1	101348.88	1
1	2	15647311	Hill	608	Spain	Female	41	1	83807.86	1	0	1	112542.58	0
2	3	15619304	Onio	502	France	Female	42	8	159660.80	3	1	0	113931.57	1
3	4	15701354	Boni	699	France	Female	39	1	0.00	2	0	0	93826.63	0
4	5	15737888	Mitchell	850	Spain	Female	43	2	125510.82	1	1	1	79084.10	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
9995	9996	15606229	Obijiaku	771	France	Male	39	5	0.00	2	1	0	96270.64	0
9996	9997	15569892	Johnstone	516	France	Male	35	10	57369.61	1	1	1	101699.77	0
9997	9998	15584532	Liu	709	France	Female	36	7	0.00	1	0	1	42085.58	1
9998	9999	15682355	Sabbatini	772	Germany	Male	42	3	75075.31	2	1	0	92888.52	1
9999	10000	15628319	Walker	792	France	Female	28	4	130142.79	1	1	0	38190.78	0

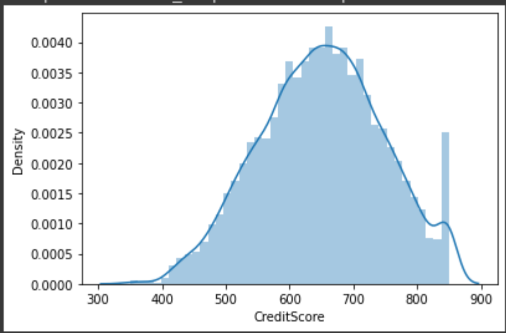
10000 rows × 14 columns

Perform Below Visualizations

UNIVARIATE ANALYSIS

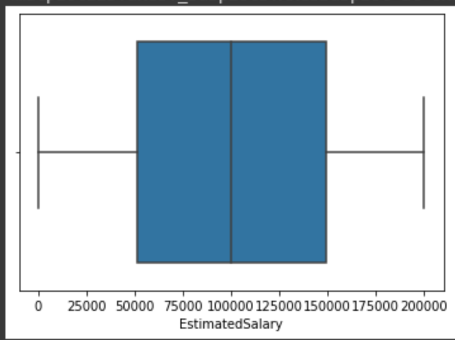
```
sns.distplot(df['CreditScore'])
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2
  warnings.warn(msg, FutureWarning)
<matplotlib.axes._subplots.AxesSubplot at 0x7fb3297cf810>
```



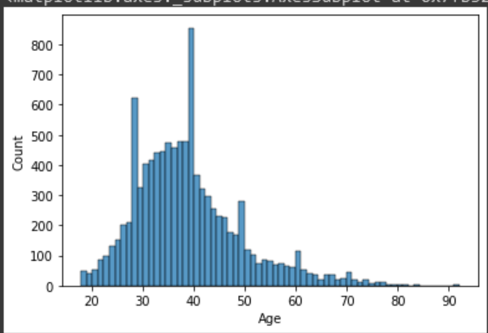

```
sns.boxplot(df['EstimatedSalary'])
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: Fut
  FutureWarning
<matplotlib.axes._subplots.AxesSubplot at 0x7fb328d5f2d0>
```



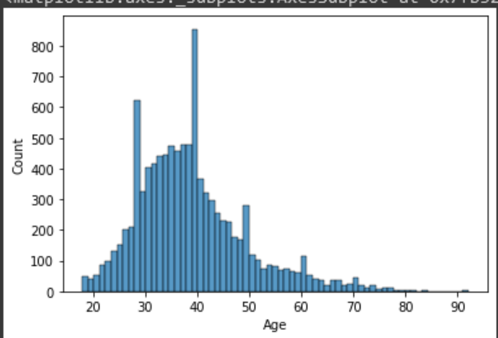

```
sns.histplot(df['Age'])
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fb32905c350>
```




```
sns.histplot(df['Age'])
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fb32905c350>
```

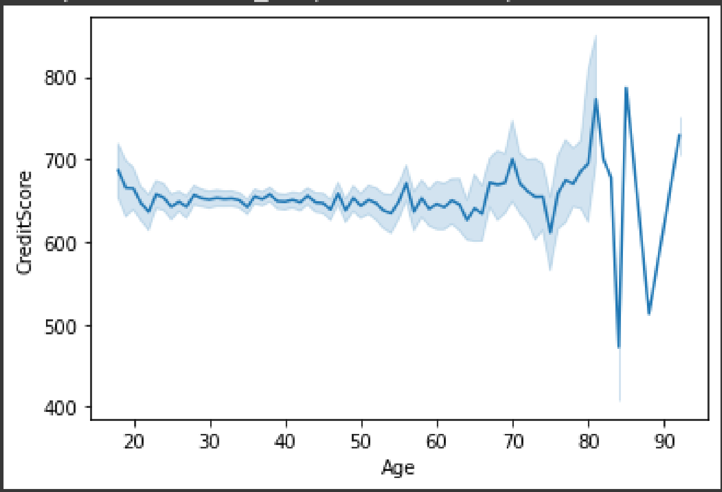


BIVARIATE ANALYSIS

BIVARIATE ANALYSIS

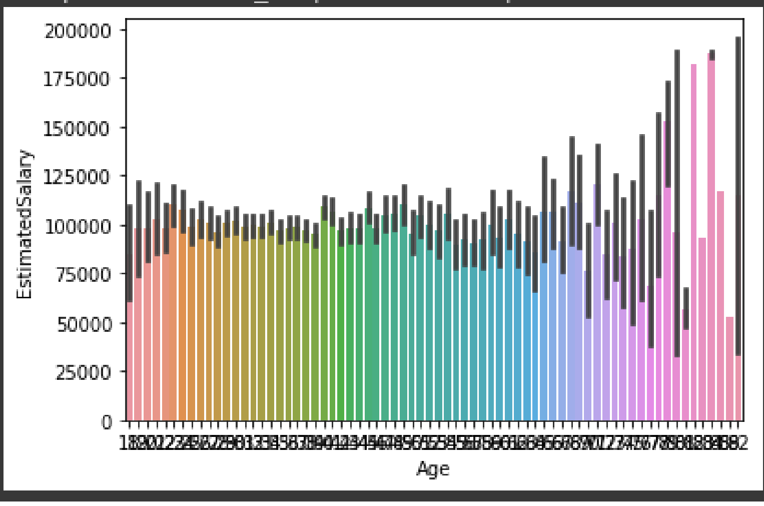
```
sns.lineplot(df['Age'],df['CreditScore'])
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43:
  FutureWarning
<matplotlib.axes._subplots.AxesSubplot at 0x7fb328bf0050>
```




```
[ ] sns.barplot(df['Age'],df['EstimatedSalary'])
```

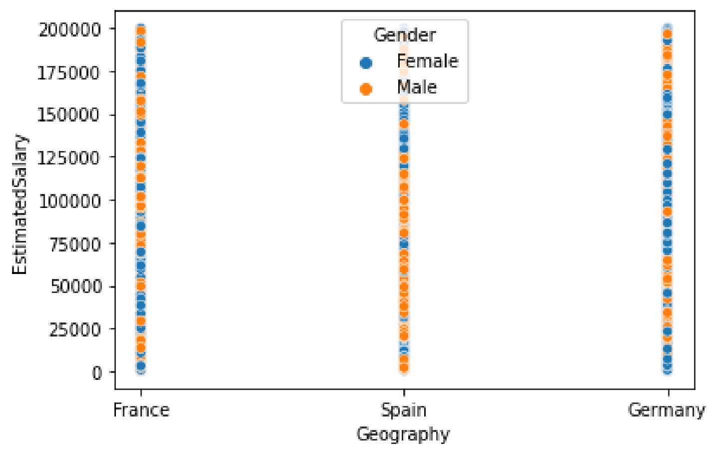
```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43:
  FutureWarning
<matplotlib.axes._subplots.AxesSubplot at 0x7fb328710950>
```



MULTIVARIATE ANALYSIS

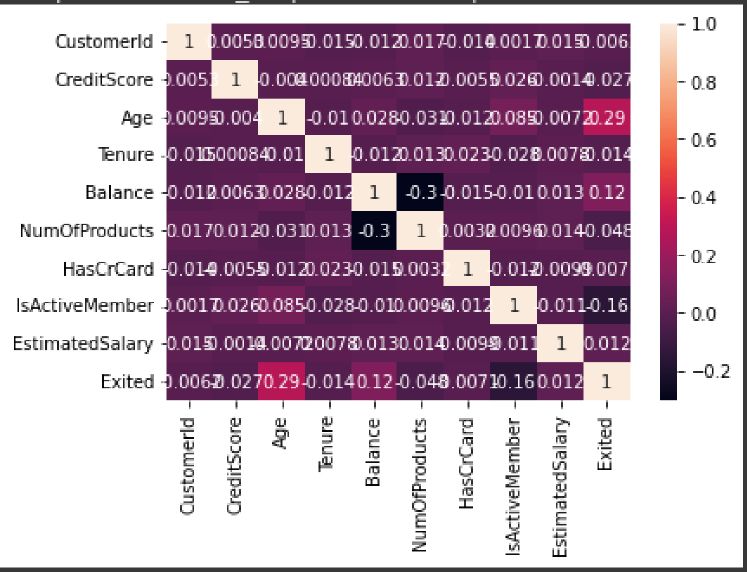
```
sns.scatterplot(df['Geography'],df['EstimatedSalary'],hue = df['Gender'])
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarn
  FutureWarning
<matplotlib.axes._subplots.AxesSubplot at 0x7fb32a7ef950>
```




```
[ ]  sns.heatmap(df.corr(),annot=True)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fb320fd35d0>
```



ENCODING

ENCODING

```
[ ] from sklearn.preprocessing import LabelEncoder, OneHotEncoder
```

```
[ ] label = LabelEncoder()
    oneh = OneHotEncoder()
```

```
df['Geography'] = label.fit_transform(df['Geography'])
df['Gender'] = label.fit_transform(df['Gender'])
df.head()
```

	CustomerId	Surname	CreditScore	Geography	Gender	Age	Tenure	Balance	NumOfProducts	HasCrCard	IsActiveMember	EstimatedSalary	Exited
0	15634602	Hargrave	619	0	0	42	2	0.00	1	1	1	101348.88	1
1	15647311	Hill	608	2	0	41	1	83807.86	1	0	1	112542.58	0
2	15619304	Onio	502	0	0	42	8	159660.80	3	1	0	113931.57	1
3	15701354	Boni	699	0	0	39	1	0.00	2	0	0	93826.63	0
4	15737888	Mitchell	850	2	0	43	2	125510.82	1	1	1	79084.10	0

SPLIT TO X AND Y

SPLIT DATASET

```
x = df.iloc[:,0:13]
x
```

	CustomerId	Surname	CreditScore	Geography	Gender	Age	Tenure	Balance	NumOfProducts	HasCrCard	IsActiveMember	EstimatedSalary	Exited
0	15634602	Hargrave	619	0	0	42	2	0.00	1	1	1	101348.88	1
1	15647311	Hill	608	2	0	41	1	83807.86	1	0	1	112542.58	0
2	15619304	Onio	502	0	0	42	8	159660.80	3	1	0	113931.57	1
3	15701354	Boni	699	0	0	39	1	0.00	2	0	0	93826.63	0
4	15737888	Mitchell	850	2	0	43	2	125510.82	1	1	1	79084.10	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...
9995	15606229	Obijiaku	771	0	1	39	5	0.00	2	1	0	96270.64	0
9996	15569892	Johnstone	516	0	1	35	10	57369.61	1	1	1	101699.77	0
9997	15584532	Liu	709	0	0	36	7	0.00	1	0	1	42085.58	1
9998	15682355	Sabbatini	772	1	1	42	3	75075.31	2	1	0	92888.52	1
9999	15628319	Walker	792	0	0	28	4	130142.79	1	1	0	38190.78	0

10000 rows × 13 columns

```
[ ] y = df['Exited']
    y
```

SCALING

Scaling

```
[ ] from sklearn.preprocessing import StandardScaler
    ss = StandardScaler()
    x_scale = ss.fit_transform(x.iloc[:,4:])
```

```
x_scale
```

```
array([[-1.09598752,  0.29351742, -1.04175968, ...,  0.97024255,
         0.02188649,  1.97716468],
       [-1.09598752,  0.19816383, -1.38753759, ...,  0.97024255,
         0.21653375, -0.50577476],
       [-1.09598752,  0.29351742,  1.03290776, ..., -1.03067011,
         0.2406869 ,  1.97716468],
       ...,
       [-1.09598752, -0.27860412,  0.68712986, ...,  0.97024255,
        -1.00864308,  1.97716468],
       [ 0.91241915,  0.29351742, -0.69598177, ..., -1.03067011,
        -0.12523071,  1.97716468],
       [-1.09598752, -1.04143285, -0.35020386, ..., -1.03067011,
        -1.07636976, -0.50577476]])
```

TRAIN TEST SPLIT

TRAIN TEST SPLIT

```
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test = train_test_split(x,y,test_size=0.2)
x_train
```

	CustomerId	Surname	CreditScore	Geography	Gender	Age	Tenure	Balance	NumOfProducts	HasCrCard	IsActiveMember	EstimatedSalary	Exited
7857	15621457	Chu	850	0	1	27	6	96654.72	2	0	0	152740.16	0
7358	15797767	Ikedinachukwu	600	0	0	49	6	0.00	1	0	1	148087.88	1
1973	15635728	P'an	693	0	1	41	4	0.00	2	0	0	156381.47	0
5450	15580227	Moss	803	0	1	33	6	0.00	2	1	0	115676.61	0
7562	15782089	Mullen	685	0	1	33	6	0.00	1	1	0	58458.26	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...
6304	15653455	Smith	648	0	0	38	2	0.00	2	0	1	9551.49	0
2867	15654211	Milani	559	2	0	27	1	0.00	1	0	1	1050.33	0
8397	15720155	Tao	630	1	1	29	6	131354.39	1	0	1	9324.31	1
121	15580203	Kennedy	674	2	1	39	6	120193.42	1	0	0	100130.95	0
1845	15589076	Henry	737	0	1	36	9	0.00	1	0	1	188670.90	1

8000 rows × 13 columns

```
x_test
```

	CustomerId	Surname	CreditScore	Geography	Gender	Age	Tenure	Balance	NumOfProducts	HasCrCard	IsActiveMember	EstimatedSalary	Exited
1065	15692744	Davison	512	0	1	36	4	152169.12	2	0	0	38629.30	1
4379	15808831	Dale	669	0	1	29	7	0.00	2	1	1	138145.62	0
9998	15682355	Sabbatini	772	1	1	42	3	75075.31	2	1	0	92888.52	1
7781	15759184	Russell	705	0	1	34	7	117715.84	1	1	0	2498.67	0
4402	15720637	Bell	710	1	0	46	10	120530.34	1	1	0	166586.99	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...
4423	15739857	Trentino	785	0	0	40	3	0.00	2	1	1	96832.82	0
1264	15732199	Gether	837	2	1	31	9	104678.62	1	0	1	50972.60	0
146	15705707	Bennelong	635	2	0	29	8	138296.94	2	1	0	141075.51	0
5485	15753837	Young	573	2	1	38	4	0.00	2	1	1	196517.43	0
9364	15780362	Ferrari	607	0	0	49	9	119960.29	2	1	0	103068.22	0

2000 rows × 13 columns