

ASSIGNMENT-4

TEAM ID-PNT2022TMID52462

1. Download the dataset: Dataset

2. Load the dataset.

```
In [3]:import numpy as np import pandas as pd df = pd.read_csv(
"/Churn_Modelling.csv")
```

3. Perform Below Visualizations

- Univariate Analysis

```
In [4]:import seaborn as sns sns.histplot(df.EstimatedSalary,kde=True) Out[4]:
```

- Bi - Variate Analysis

```
In [ ]:import seaborn as sns import matplotlib.pyplot as plt sns.scatterplot(df
.Balance,df.EstimatedSalary) plt.ylim(0,15000)
```

C:\Users\ELCOT\anaconda3\lib\site-packages\seaborn_decorators.py:36:

3.Read dataset and do preprocessing

FutureWarning: Pass the following variables as keyword args: x, y. From version

0.12, the only valid positional argument will be `data`, and passing other

arguments without an explicit keyword will result in an error or

misinterpretation. warnings.warn(Out[]:(0.0, 15000.0)

- Multi - Variate Analysis

4. Perform descriptive statistics on the dataset.

```
In [7]:df=pd.read_csv("/Churn_Modelling.csv") df.describe(include='all')
```

count 10000.00000 1.000000e+04 10000 10000.000000 10000 10000 10000.000000

10000.000000 10000.000000 10000.000000 10000.000000 10000.000000 10000.000000


```
'IsActiveMember','EstimatedSalary','Exited'] import seaborn as sns df.head()
```

Out[11]:

```
RowNumberCustomerIdSurnameCreditScoreGeographyGenderAgeTenureBalance
```

```
NumOfProductsHasCrCardIsActiveMemberEstimatedSalaryExited
```

```
0115634602Hargrave619FranceFemale4220.00111101348.881
```

```
1215647311Hill608SpainFemale41183807.86101112542.580
```

```
2315619304Onio502FranceFemale428159660.80310113931.571
```

```
3415701354Boni699FranceFemale3910.0020093826.630
```

```
4515737888Mitchell850SpainFemale432125510.8211179084.1008. Split the data into
```

```
dependent and independent variables.x=df.iloc[:, :-1].valuesprint(x)
```

```
y=df.iloc[:, -1]._valuesprint(y)[[1 15634602 'Hargrave' ... 1 1 101348.88] [2
```

```
15647311 'Hill' ... 0 1 112542.58] [3 15619304 'Onio' ... 1 0 113931.57] ...
```

```
[9998 15584532 'Liu' ... 0 1 42085.58] [9999 15682355 'Sabbatini' ... 1 0
```

```
92888.52] [10000 15628319 'Walker' ... 1 0 38190.78]][1 0 1 ... 1 1 0]
```

9. Scale the independent variables

```
In [13]:import seaborn as sns df=pd.read_csv("/Churn_Modelling.csv") dff=df[[
```

```
'Balance','Age']] sns.heatmap(dff.corr(), annot=True) sns.set(rc={
```

```
'figure.figsize':(40,40)}) 10. Split the data into training and testingfrom
```

```
scipy.sparse.construct import randomx=df.iloc[:, 1:2].values
```

```
y=df.iloc[:, 2].valuesfrom sklearn.model_selection import train_test_split
```

```
x_train, x_test, y_train,
```

```
y_test=train_test_split(x,y,test_size=0.2,random_state=0)print('Row count of
```

```
x_train table'+ '-' +str(f"{len(x_train):,}")')print('Row count of y_train
```

```
table'+ '-' +str(f"{len(y_train):,}")')print('Row count of x_test
```

```
table'+ '-' +str(f"{len(x_test):,}")')print('Row count of y_test
```

```
table'+ '-' +str(f"{len(y_test):,}")')Row count of x_train table-8,000Row count of
```

```
y_train table-8,000Row count of x_test table-2,000Row count of y_test
```

```
table-2,000
```