

```
from google.colab import drive
drive.mount('/content/gdrive')
```

Mounted at /content/gdrive

```
from google.colab import files
uploaded = files.upload()
```

<IPython.core.display.HTML object>

Saving Churn\_Modelling.csv to Churn\_Modelling.csv

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

```
import io
```

```
df = pd.read_csv(io.BytesIO(uploaded['Churn_Modelling.csv']))
print(df)
```

|       | RowNumber | CustomerId | Surname   | CreditScore | Geography | Gender |
|-------|-----------|------------|-----------|-------------|-----------|--------|
| Age \ |           |            |           |             |           |        |
| 0     | 1         | 15634602   | Hargrave  | 619         | France    | Female |
| 42    |           |            |           |             |           |        |
| 1     | 2         | 15647311   | Hill      | 608         | Spain     | Female |
| 41    |           |            |           |             |           |        |
| 2     | 3         | 15619304   | Onio      | 502         | France    | Female |
| 42    |           |            |           |             |           |        |
| 3     | 4         | 15701354   | Boni      | 699         | France    | Female |
| 39    |           |            |           |             |           |        |
| 4     | 5         | 15737888   | Mitchell  | 850         | Spain     | Female |
| 43    |           |            |           |             |           |        |
| ...   | ...       | ...        | ...       | ...         | ...       | ...    |
| ...   |           |            |           |             |           |        |
| 9995  | 9996      | 15606229   | Obijiaku  | 771         | France    | Male   |
| 39    |           |            |           |             |           |        |
| 9996  | 9997      | 15569892   | Johnstone | 516         | France    | Male   |
| 35    |           |            |           |             |           |        |
| 9997  | 9998      | 15584532   | Liu       | 709         | France    | Female |
| 36    |           |            |           |             |           |        |
| 9998  | 9999      | 15682355   | Sabbatini | 772         | Germany   | Male   |
| 42    |           |            |           |             |           |        |
| 9999  | 10000     | 15628319   | Walker    | 792         | France    | Female |
| 28    |           |            |           |             |           |        |

|   | Tenure | Balance   | NumOfProducts | HasCrCard | IsActiveMember | \ |
|---|--------|-----------|---------------|-----------|----------------|---|
| 0 | 2      | 0.00      | 1             | 1         |                | 1 |
| 1 | 1      | 83807.86  | 1             | 0         |                | 1 |
| 2 | 8      | 159660.80 | 3             | 1         |                | 0 |

|      |     |           |     |     |     |
|------|-----|-----------|-----|-----|-----|
| 3    | 1   | 0.00      | 2   | 0   | 0   |
| 4    | 2   | 125510.82 | 1   | 1   | 1   |
| ...  | ... | ...       | ... | ... | ... |
| 9995 | 5   | 0.00      | 2   | 1   | 0   |
| 9996 | 10  | 57369.61  | 1   | 1   | 1   |
| 9997 | 7   | 0.00      | 1   | 0   | 1   |
| 9998 | 3   | 75075.31  | 2   | 1   | 0   |
| 9999 | 4   | 130142.79 | 1   | 1   | 0   |

|      | EstimatedSalary | Exited |
|------|-----------------|--------|
| 0    | 101348.88       | 1      |
| 1    | 112542.58       | 0      |
| 2    | 113931.57       | 1      |
| 3    | 93826.63        | 0      |
| 4    | 79084.10        | 0      |
| ...  | ...             | ...    |
| 9995 | 96270.64        | 0      |
| 9996 | 101699.77       | 0      |
| 9997 | 42085.58        | 1      |
| 9998 | 92888.52        | 1      |
| 9999 | 38190.78        | 0      |

[10000 rows x 14 columns]

df.head()

|   | RowNumber | CustomerId | Surname  | CreditScore | Geography | Gender | Age |
|---|-----------|------------|----------|-------------|-----------|--------|-----|
| 0 | 1         | 15634602   | Hargrave | 619         | France    | Female | 42  |
| 1 | 2         | 15647311   | Hill     | 608         | Spain     | Female | 41  |
| 2 | 3         | 15619304   | Onio     | 502         | France    | Female | 42  |
| 3 | 4         | 15701354   | Boni     | 699         | France    | Female | 39  |
| 4 | 5         | 15737888   | Mitchell | 850         | Spain     | Female | 43  |

|   | Tenure | Balance   | NumOfProducts | HasCrCard | IsActiveMember | \ |
|---|--------|-----------|---------------|-----------|----------------|---|
| 0 | 2      | 0.00      | 1             | 1         | 1              |   |
| 1 | 1      | 83807.86  | 1             | 0         | 1              |   |
| 2 | 8      | 159660.80 | 3             | 1         | 0              |   |
| 3 | 1      | 0.00      | 2             | 0         | 0              |   |
| 4 | 2      | 125510.82 | 1             | 1         | 1              |   |

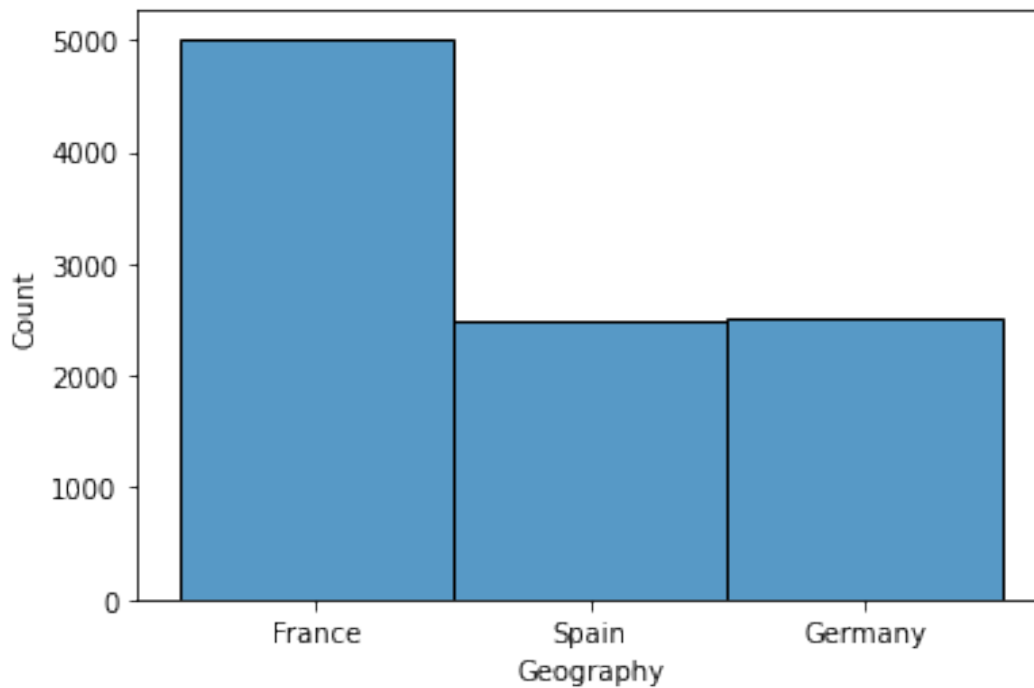
|   | EstimatedSalary | Exited |
|---|-----------------|--------|
| 0 | 101348.88       | 1      |
| 1 | 112542.58       | 0      |
| 2 | 113931.57       | 1      |

```
3          93826.63      0
4          79084.10      0
```

3 a)Univariate Analysis

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

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f021875f6d0>
```

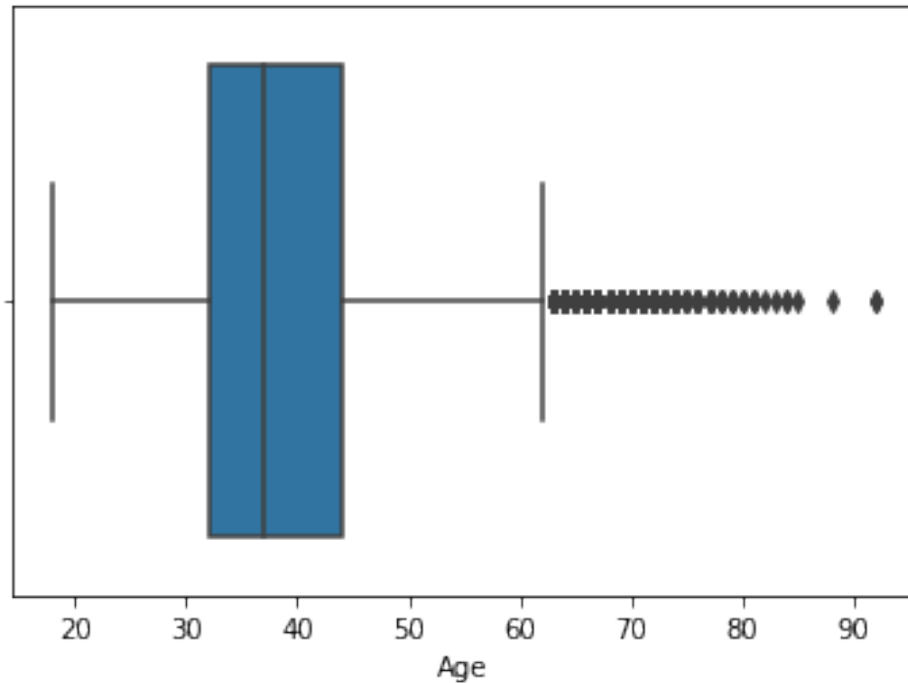


```
sns.boxplot(df['Age'])
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43:
FutureWarning: Pass the following variable as a keyword arg: x. 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.
```

```
FutureWarning
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f021819fd90>
```



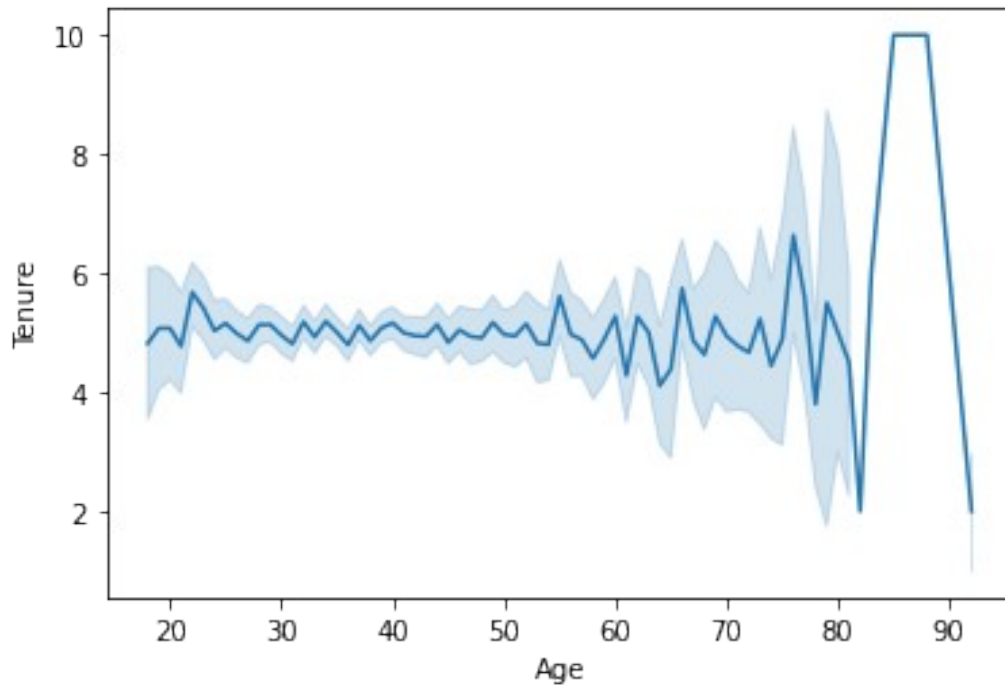
### 3 b) Bi - Variate Analysis

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

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43:  
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.
```

```
FutureWarning
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f0218179250>
```

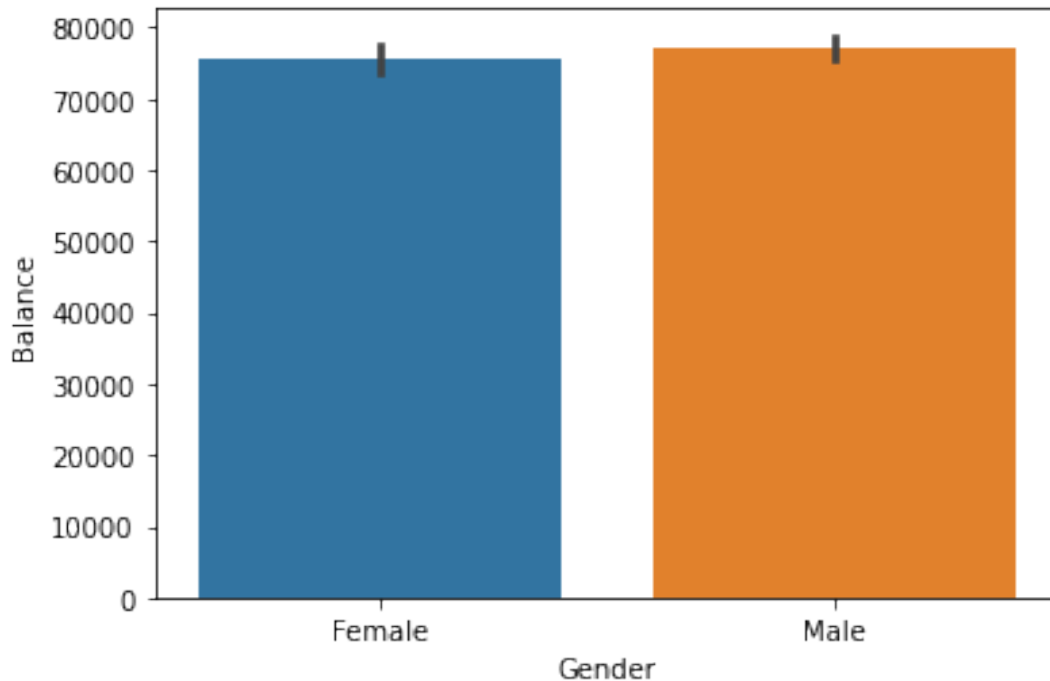


```
sns.barplot(df['Gender'], df['Balance'])
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43:  
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.
```

```
FutureWarning
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f0217dcca90>
```



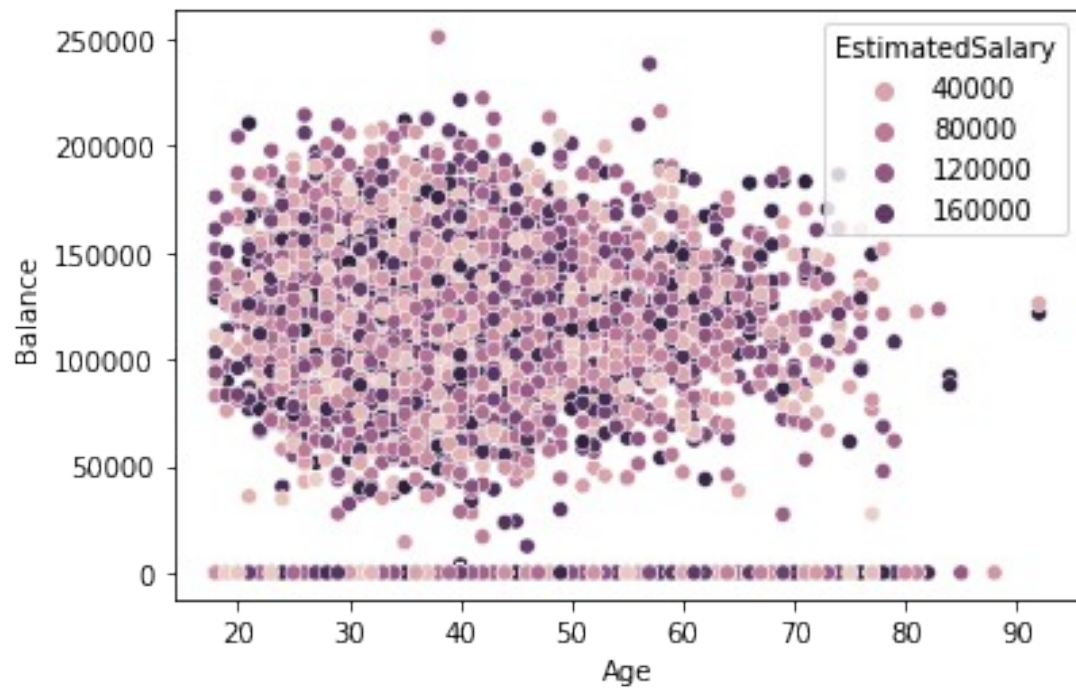
### 3 c) Multi - Variate Analysis

```
sns.scatterplot(df['Age'], df['Balance'], hue = df['EstimatedSalary'])
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43:  
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.
```

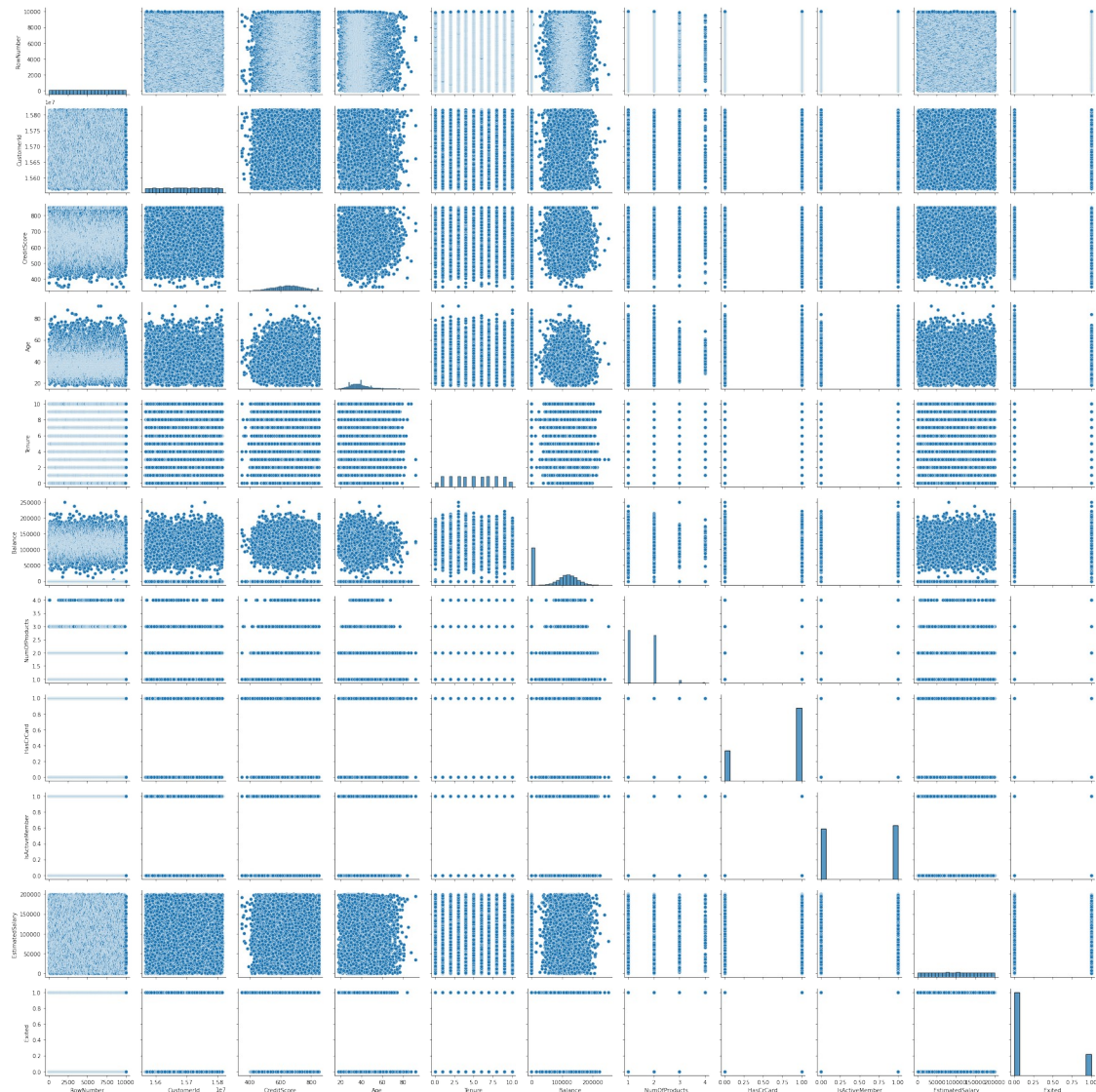
```
FutureWarning
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f0217d2bcd0>
```



```
sns.pairplot(df)
```

```
<seaborn.axisgrid.PairGrid at 0x7f0217d1c110>
```



#### 4) Descriptive Statistics

```
df.mean()
```

```
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:1:
FutureWarning: Dropping of nuisance columns in DataFrame reductions
(with 'numeric_only=None') is deprecated; in a future version this
will raise TypeError. Select only valid columns before calling the
reduction.
```

```
"""Entry point for launching an IPython kernel.
```

```
RowNumber      5.000500e+03
CustomerId     1.569094e+07
CreditScore    6.505288e+02
Age            3.892180e+01
Tenure         5.012800e+00
Balance        7.648589e+04
```



```
NumOfProducts      1.530200e+00
HasCrCard           7.055000e-01
IsActiveMember      5.151000e-01
EstimatedSalary     1.000902e+05
Exited              2.037000e-01
dtype: float64
```

```
df.median()
```

```
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:1:
FutureWarning: Dropping of nuisance columns in DataFrame reductions
(with 'numeric_only=None') is deprecated; in a future version this
will raise TypeError.  Select only valid columns before calling the
reduction.
```

```
"""Entry point for launching an IPython kernel.
```

```
RowNumber          5.000500e+03
CustomerId          1.569074e+07
CreditScore        6.520000e+02
Age                 3.700000e+01
Tenure              5.000000e+00
Balance            9.719854e+04
NumOfProducts      1.000000e+00
HasCrCard           1.000000e+00
IsActiveMember      1.000000e+00
EstimatedSalary     1.001939e+05
Exited              0.000000e+00
dtype: float64
```

```
df.mode()
```

|       | RowNumber | CustomerId | Surname | CreditScore | Geography | Gender |
|-------|-----------|------------|---------|-------------|-----------|--------|
| Age \ |           |            |         |             |           |        |
| 0     | 1         | 15565701   | Smith   | 850.0       | France    | Male   |
| 37.0  |           |            |         |             |           |        |
| 1     | 2         | 15565706   | NaN     | NaN         | NaN       | NaN    |
| NaN   |           |            |         |             |           |        |
| 2     | 3         | 15565714   | NaN     | NaN         | NaN       | NaN    |
| NaN   |           |            |         |             |           |        |
| 3     | 4         | 15565779   | NaN     | NaN         | NaN       | NaN    |
| NaN   |           |            |         |             |           |        |
| 4     | 5         | 15565796   | NaN     | NaN         | NaN       | NaN    |
| NaN   |           |            |         |             |           |        |
| ...   | ...       | ...        | ...     | ...         | ...       | ...    |
| .     |           |            |         |             |           | ..     |
| 9995  | 9996      | 15815628   | NaN     | NaN         | NaN       | NaN    |
| NaN   |           |            |         |             |           |        |
| 9996  | 9997      | 15815645   | NaN     | NaN         | NaN       | NaN    |
| NaN   |           |            |         |             |           |        |
| 9997  | 9998      | 15815656   | NaN     | NaN         | NaN       | NaN    |
| NaN   |           |            |         |             |           |        |

|      |       |          |     |     |     |     |
|------|-------|----------|-----|-----|-----|-----|
| 9998 | 9999  | 15815660 | NaN | NaN | NaN | NaN |
| NaN  |       |          |     |     |     |     |
| 9999 | 10000 | 15815690 | NaN | NaN | NaN | NaN |
| NaN  |       |          |     |     |     |     |

|      | Tenure | Balance | NumOfProducts | HasCrCard | IsActiveMember | \ |
|------|--------|---------|---------------|-----------|----------------|---|
| 0    | 2.0    | 0.0     | 1.0           | 1.0       | 1.0            |   |
| 1    | NaN    | NaN     | NaN           | NaN       | NaN            |   |
| 2    | NaN    | NaN     | NaN           | NaN       | NaN            |   |
| 3    | NaN    | NaN     | NaN           | NaN       | NaN            |   |
| 4    | NaN    | NaN     | NaN           | NaN       | NaN            |   |
| ...  | ...    | ...     | ...           | ...       | ...            |   |
| 9995 | NaN    | NaN     | NaN           | NaN       | NaN            |   |
| 9996 | NaN    | NaN     | NaN           | NaN       | NaN            |   |
| 9997 | NaN    | NaN     | NaN           | NaN       | NaN            |   |
| 9998 | NaN    | NaN     | NaN           | NaN       | NaN            |   |
| 9999 | NaN    | NaN     | NaN           | NaN       | NaN            |   |

|      | EstimatedSalary | Exited |
|------|-----------------|--------|
| 0    | 24924.92        | 0.0    |
| 1    | NaN             | NaN    |
| 2    | NaN             | NaN    |
| 3    | NaN             | NaN    |
| 4    | NaN             | NaN    |
| ...  | ...             | ...    |
| 9995 | NaN             | NaN    |
| 9996 | NaN             | NaN    |
| 9997 | NaN             | NaN    |
| 9998 | NaN             | NaN    |
| 9999 | NaN             | NaN    |

[10000 rows x 14 columns]

5)Handle the Missing values

df.isnull().any()

|                 |       |
|-----------------|-------|
| RowNumber       | False |
| CustomerId      | False |
| Surname         | False |
| CreditScore     | False |
| Geography       | False |
| Gender          | False |
| Age             | False |
| Tenure          | False |
| Balance         | False |
| NumOfProducts   | False |
| HasCrCard       | False |
| IsActiveMember  | False |
| EstimatedSalary | False |

```
Exited          False
dtype: bool
```

6)Find the outliers and replace the outliers

```
df["Tenure"] = np.where(df["Tenure"] >10, np.median(df["Tenure"]))
df["Tenure"]
```

```
0      2
1      1
2      8
3      1
4      2
..
9995   5
9996  10
9997   7
9998   3
9999   4
```

```
Name: Tenure, Length: 10000, dtype: object
```

7)Check for Categorical columns and perform encoding.

```
from sklearn.preprocessing import LabelEncoder, OneHotEncoder
le = LabelEncoder()
oneh = OneHotEncoder()
df['Geography'] = le.fit_transform(df['Geography'])
df.head()
```

|       | RowNumber | CustomerId | Surname  | CreditScore | Geography | Gender |
|-------|-----------|------------|----------|-------------|-----------|--------|
| Age \ |           |            |          |             |           |        |
| 0     | 1         | 15634602   | Hargrave | 619         | 0         | Female |
| 42    |           |            |          |             |           |        |
| 1     | 2         | 15647311   | Hill     | 608         | 2         | Female |
| 41    |           |            |          |             |           |        |
| 2     | 3         | 15619304   | Onio     | 502         | 0         | Female |
| 42    |           |            |          |             |           |        |
| 3     | 4         | 15701354   | Boni     | 699         | 0         | Female |
| 39    |           |            |          |             |           |        |
| 4     | 5         | 15737888   | Mitchell | 850         | 2         | Female |
| 43    |           |            |          |             |           |        |

|   | Tenure | Balance   | NumOfProducts | HasCrCard | IsActiveMember | \ |
|---|--------|-----------|---------------|-----------|----------------|---|
| 0 | 2      | 0.00      | 1             | 1         |                | 1 |
| 1 | 1      | 83807.86  | 1             | 0         |                | 1 |
| 2 | 8      | 159660.80 | 3             | 1         |                | 0 |
| 3 | 1      | 0.00      | 2             | 0         |                | 0 |
| 4 | 2      | 125510.82 | 1             | 1         |                | 1 |

```
EstimatedSalary  Exited
```

```
0      101348.88      1
1      112542.58      0
2      113931.57      1
3       93826.63      0
4       79084.10      0
```

```
df['Balance'] = le.fit_transform(df['Balance'])
df['NumOfProducts'] = le.fit_transform(df['NumOfProducts'])
```

```
df.head()
```

|       | RowNumber | CustomerId | Surname  | CreditScore | Geography | Gender |
|-------|-----------|------------|----------|-------------|-----------|--------|
| Age \ |           |            |          |             |           |        |
| 0     | 1         | 15634602   | Hargrave | 619         | 0         | Female |
| 42    |           |            |          |             |           |        |
| 1     | 2         | 15647311   | Hill     | 608         | 2         | Female |
| 41    |           |            |          |             |           |        |
| 2     | 3         | 15619304   | Onio     | 502         | 0         | Female |
| 42    |           |            |          |             |           |        |
| 3     | 4         | 15701354   | Boni     | 699         | 0         | Female |
| 39    |           |            |          |             |           |        |
| 4     | 5         | 15737888   | Mitchell | 850         | 2         | Female |
| 43    |           |            |          |             |           |        |

|                   | Tenure | Balance | NumOfProducts | HasCrCard | IsActiveMember |
|-------------------|--------|---------|---------------|-----------|----------------|
| EstimatedSalary \ |        |         |               |           |                |
| 0                 | 2      | 0       | 0             | 1         | 1              |
| 101348.88         |        |         |               |           |                |
| 1                 | 1      | 743     | 0             | 0         | 1              |
| 112542.58         |        |         |               |           |                |
| 2                 | 8      | 5793    | 2             | 1         | 0              |
| 113931.57         |        |         |               |           |                |
| 3                 | 1      | 0       | 1             | 0         | 0              |
| 93826.63          |        |         |               |           |                |
| 4                 | 2      | 3696    | 0             | 1         | 1              |
| 79084.10          |        |         |               |           |                |

|   | Exited |
|---|--------|
| 0 | 1      |
| 1 | 0      |
| 2 | 1      |
| 3 | 0      |
| 4 | 0      |

8) Split the data into dependent and independent variables.

```
#Independent
```

```
x = df.iloc[:, 0:6]
```

```
x
```

|      | RowNumber | CustomerId | Surname   | CreditScore | Geography | Gender |
|------|-----------|------------|-----------|-------------|-----------|--------|
| 0    | 1         | 15634602   | Hargrave  | 619         | 0         | Female |
| 1    | 2         | 15647311   | Hill      | 608         | 2         | Female |
| 2    | 3         | 15619304   | Onio      | 502         | 0         | Female |
| 3    | 4         | 15701354   | Boni      | 699         | 0         | Female |
| 4    | 5         | 15737888   | Mitchell  | 850         | 2         | Female |
| ...  | ...       | ...        | ...       | ...         | ...       | ...    |
| 9995 | 9996      | 15606229   | Obijiaku  | 771         | 0         | Male   |
| 9996 | 9997      | 15569892   | Johnstone | 516         | 0         | Male   |
| 9997 | 9998      | 15584532   | Liu       | 709         | 0         | Female |
| 9998 | 9999      | 15682355   | Sabbatini | 772         | 1         | Male   |
| 9999 | 10000     | 15628319   | Walker    | 792         | 0         | Female |

[10000 rows x 6 columns]

*#Dependent*

Y=df.iloc[:, -1].values

Y

array([1, 0, 1, ..., 1, 1, 0])

9)Scale the independent variables

```
import pandas as pd
from sklearn.preprocessing import MinMaxScaler
scaler=MinMaxScaler()
df[["RowNumber"]]=scaler.fit_transform(df[["RowNumber"]])
print(df)
```

|       | RowNumber | CustomerId | Surname   | CreditScore | Geography | Gender |
|-------|-----------|------------|-----------|-------------|-----------|--------|
| Age \ |           |            |           |             |           |        |
| 0     | 0.0000    | 15634602   | Hargrave  | 619         | 0         | Female |
| 42    |           |            |           |             |           |        |
| 1     | 0.0001    | 15647311   | Hill      | 608         | 2         | Female |
| 41    |           |            |           |             |           |        |
| 2     | 0.0002    | 15619304   | Onio      | 502         | 0         | Female |
| 42    |           |            |           |             |           |        |
| 3     | 0.0003    | 15701354   | Boni      | 699         | 0         | Female |
| 39    |           |            |           |             |           |        |
| 4     | 0.0004    | 15737888   | Mitchell  | 850         | 2         | Female |
| 43    |           |            |           |             |           |        |
| ...   | ...       | ...        | ...       | ...         | ...       | ...    |
| ...   |           |            |           |             |           |        |
| 9995  | 0.9996    | 15606229   | Obijiaku  | 771         | 0         | Male   |
| 39    |           |            |           |             |           |        |
| 9996  | 0.9997    | 15569892   | Johnstone | 516         | 0         | Male   |
| 35    |           |            |           |             |           |        |
| 9997  | 0.9998    | 15584532   | Liu       | 709         | 0         | Female |
| 36    |           |            |           |             |           |        |
| 9998  | 0.9999    | 15682355   | Sabbatini | 772         | 1         | Male   |

```

42
9999      1.0000      15628319      Walker      792      0  Female
28

```

```

      Tenure  Balance  NumOfProducts  HasCrCard  IsActiveMember  \
0          2         0              0          1              1
1          1        743              0          0              1
2          8       5793              2          1              0
3          1         0              1          0              0
4          2       3696              0          1              1
...      ...      ...      ...      ...      ...
9995        5         0              1          1              0
9996       10        124              0          1              1
9997        7         0              0          0              1
9998        3        427              1          1              0
9999        4       4112              0          1              0

```

```

      EstimatedSalary  Exited
0          101348.88        1
1          112542.58        0
2          113931.57        1
3           93826.63        0
4           79084.10        0
...      ...      ...
9995          96270.64        0
9996         101699.77        0
9997          42085.58        1
9998          92888.52        1
9999          38190.78        0

```

```
[10000 rows x 14 columns]
```

10) Split the data into training and testing

```

from sklearn.model_selection import train_test_split
train_size=0.7
X=df.drop(columns=['Tenure']).copy()
Y=df['Tenure']
X_train, X_rem, Y_train, Y_rem=train_test_split(X,Y,train_size=0.8)
test_size=0.3
X_valid,X_test,Y_valid,Y_test=train_test_split(X_rem,Y_rem,test_size=0
.5)
print(X_train.shape),print(Y_train.shape)
print(X_valid.shape),print(Y_valid.shape)
print(X_test.shape),print(Y_test.shape)

(8000, 13)
(8000,)
(1000, 13)
(1000,)

```

(1000, 13)

(1000,)

(None, None)