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

```
import warnings
warnings.filterwarnings('ignore')
```

```
df=pd.read_csv('/content/Churn_Modelling (1).csv')
```

```
df.head()
```

	RowNumber	CustomerId	Surname	CreditScore	Geography	Gender	Age	Tenure	Ba
0	1	15634602	Hargrave	619	France	Female	42	2	
1	2	15647311	Hill	608	Spain	Female	41	1	83
2	3	15619304	Onio	502	France	Female	42	8	159
3	4	15701354	Boni	699	France	Female	39	1	
4	5	15737888	Mitchell	850	Spain	Female	43	2	125

```
df.describe()
```

	RowNumber	CustomerId	CreditScore	Age	Tenure	Bala
count	10000.00000	1.000000e+04	10000.000000	10000.000000	10000.000000	10000.000
mean	5000.50000	1.569094e+07	650.528800	38.921800	5.012800	76485.889
std	2886.89568	7.193619e+04	96.653299	10.487806	2.892174	62397.405
min	1.00000	1.556570e+07	350.000000	18.000000	0.000000	0.000
25%	2500.75000	1.562853e+07	584.000000	32.000000	3.000000	0.000
50%	5000.50000	1.569074e+07	652.000000	37.000000	5.000000	97198.540
75%	7500.25000	1.575323e+07	718.000000	44.000000	7.000000	127644.240
max	10000.00000	1.581569e+07	850.000000	92.000000	10.000000	250898.090

```
df.info()
```

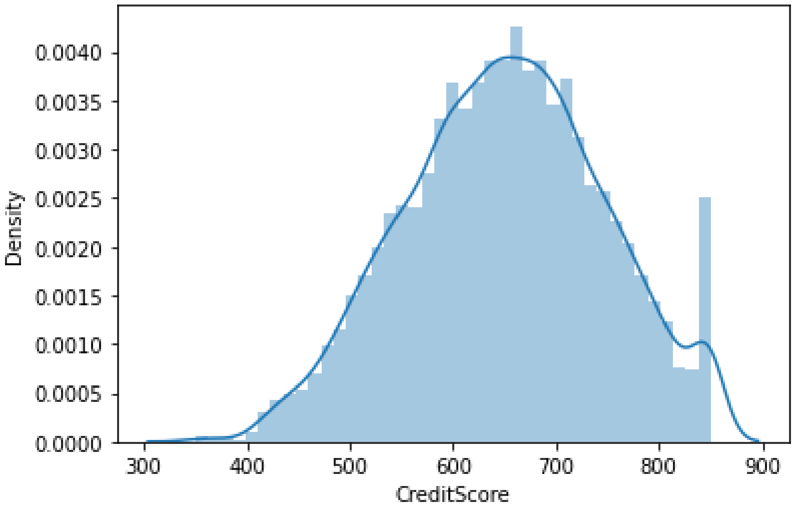
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10000 entries, 0 to 9999
Data columns (total 14 columns):
 #   Column           Non-Null Count  Dtype  
---  ------           --------------  -----  
 0   RowNumber        10000 non-null  int64  
 1   CustomerId       10000 non-null  int64  
 2   Surname          10000 non-null  object 
 3   CreditScore      10000 non-null  int64  
 4   Geography        10000 non-null  object 
 5   Gender           10000 non-null  object 
 6   Age              10000 non-null  int64  
 7   Tenure           10000 non-null  int64  
 8   Balance          10000 non-null  float64
 9   NumOfProducts    10000 non-null  int64  
 10  HasCrCard        10000 non-null  int64  
 11  IsActiveMember   10000 non-null  int64  
 12  EstimatedSalary  10000 non-null  float64
 13  Exited           10000 non-null  int64  
dtypes: float64(2), int64(9), object(3)
memory usage: 1.1+ MB
```

```
df.head(2)
```

	RowNumber	CustomerId	Surname	CreditScore	Geography	Gender	Age	Tenure	Bal
0	1	15634602	Hargrave	619	France	Female	42	2	
1	2	15647311	Hill	608	Spain	Female	41	1	838

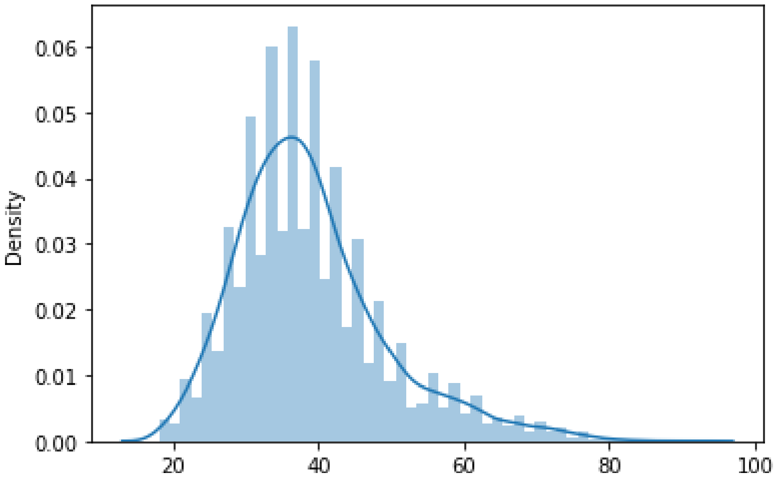
```
sns.distplot(df.CreditScore)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fedf5f6a250>
```



```
sns.distplot(df.Age)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fedf5df3210>
```



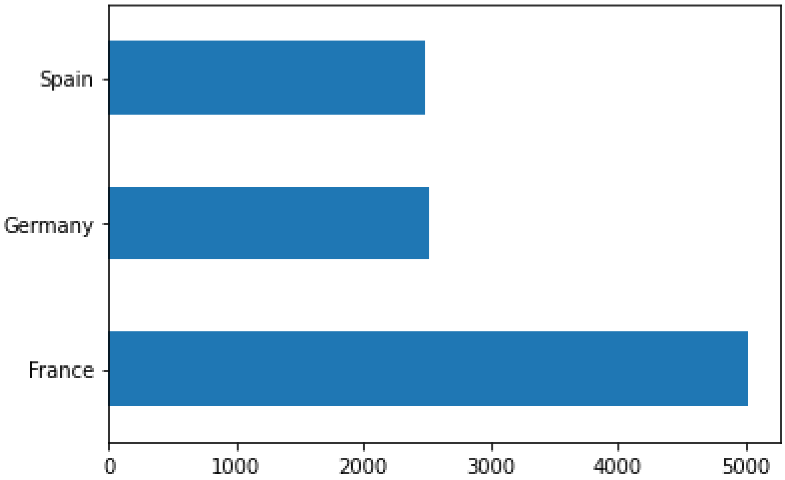
```
ind='barh')df.Gender.value_counts().plot(k
```

```
  File "<ipython-input-17-7c8b3896840f>", line 1
    ind='barh')df.Gender.value_counts().plot(k
              ^
SyntaxError: invalid syntax
```

SEARCH STACK OVERFLOW

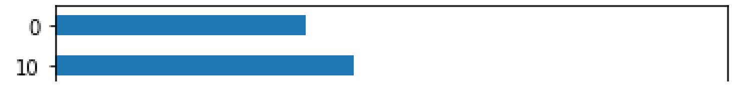
```
df.Geography.value_counts().plot(kind='barh')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fedf5813d10>
```



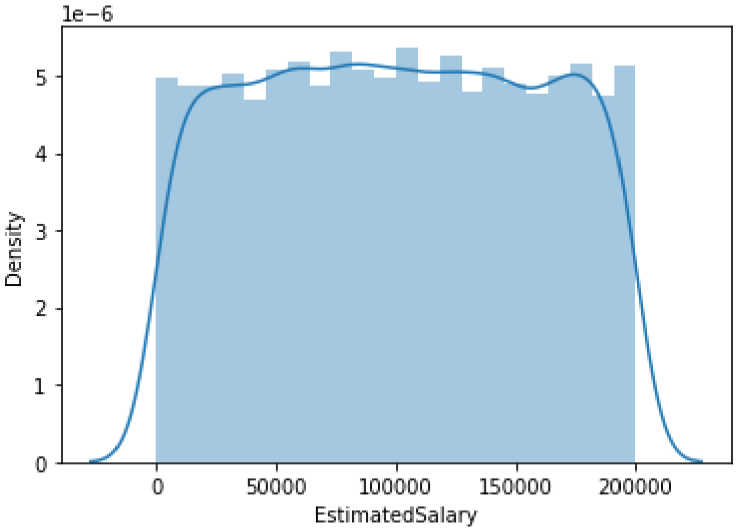
```
df.Tenure.value_counts().plot(kind='barh')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fedf57c7bd0>
```



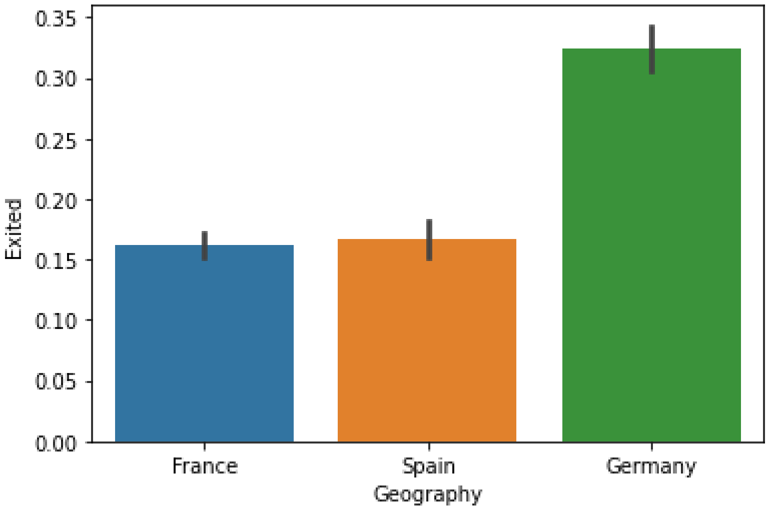
```
sns.distplot(df.EstimatedSalary)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fedf5738a90>
```



```
sns.barplot(df.Geography, df.Exited)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fedf5611950>
```

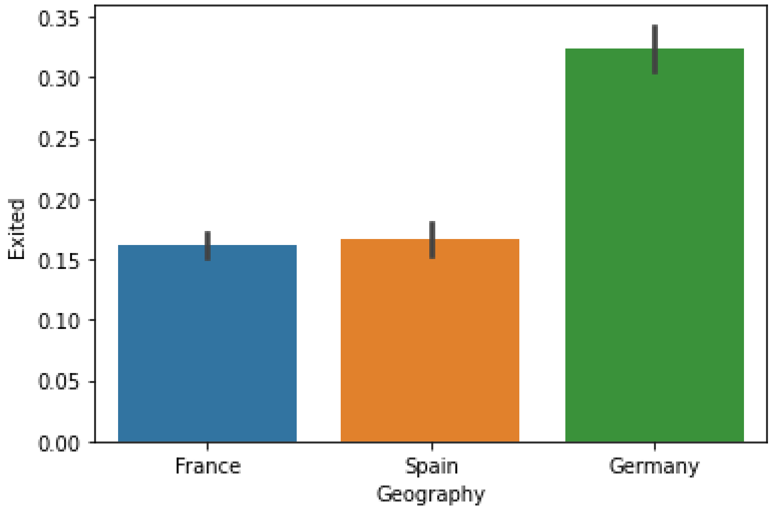


```
df.head(2)
```

	RowNumber	CustomerId	Surname	CreditScore	Geography	Gender	Age	Tenure	Bal
0	1	15634602	Hargrave	619	France	Female	42	2	
1	2	15647311	Hill	608	Spain	Female	41	1	838

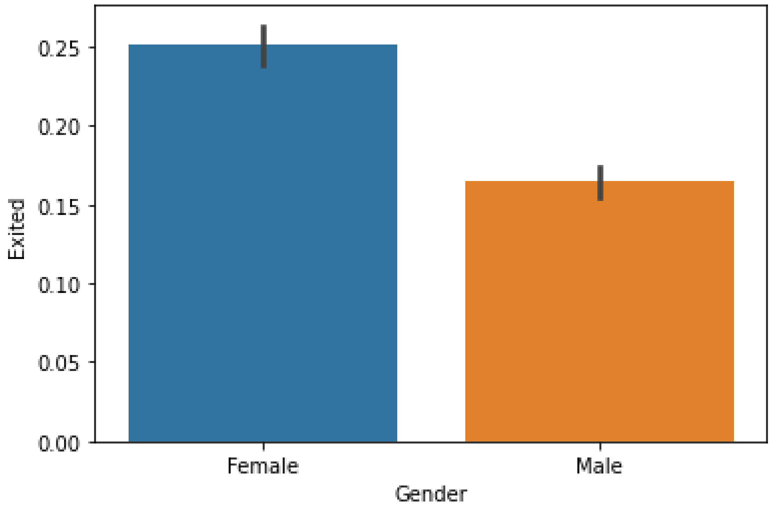
```
sns.barplot(x='Geography',y='Exited',data=df)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fedf55fe050>
```



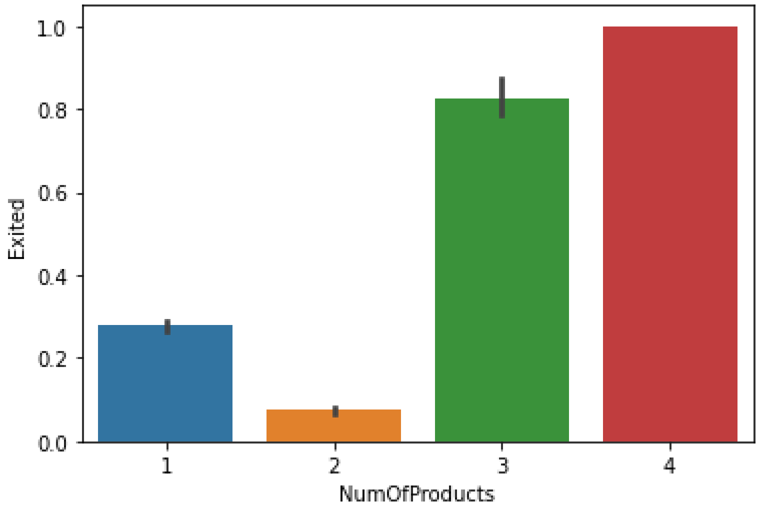
```
sns.barplot(x='Gender',y='Exited',data=df)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fedf55756d0>
```



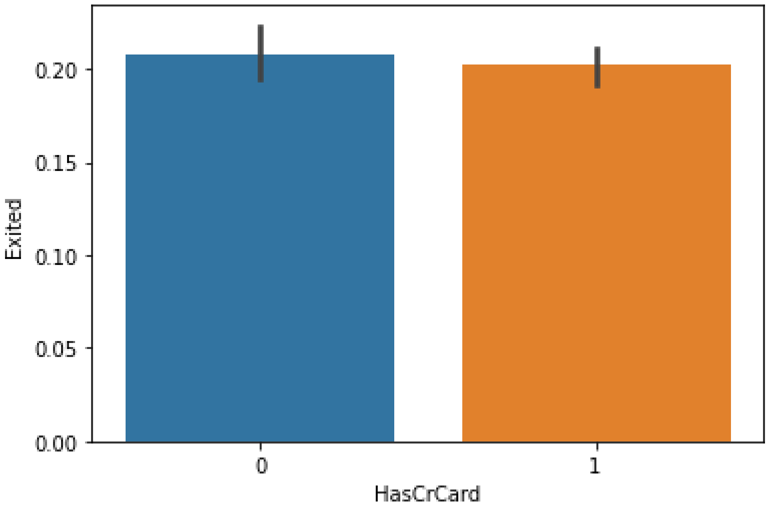
```
sns.barplot(x='NumOfProducts',y='Exited',data=df)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fedf54d3b90>
```



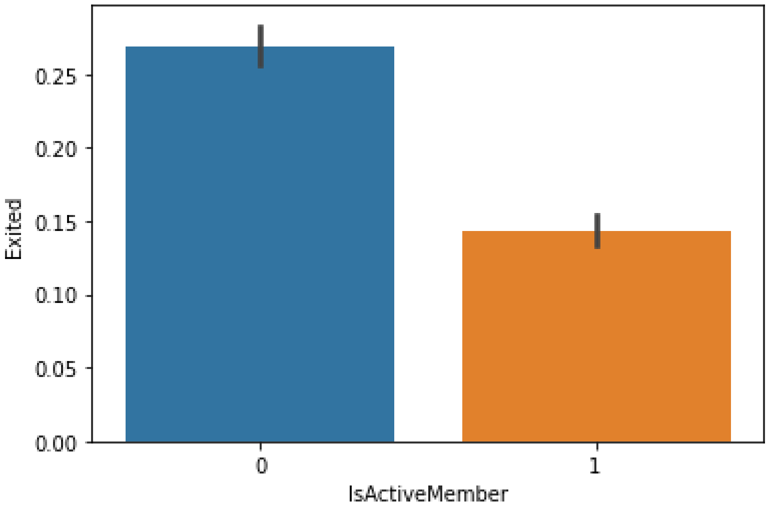
```
sns.barplot(x='HasCrCard',y='Exited',data=df)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fedf54cfd90>
```



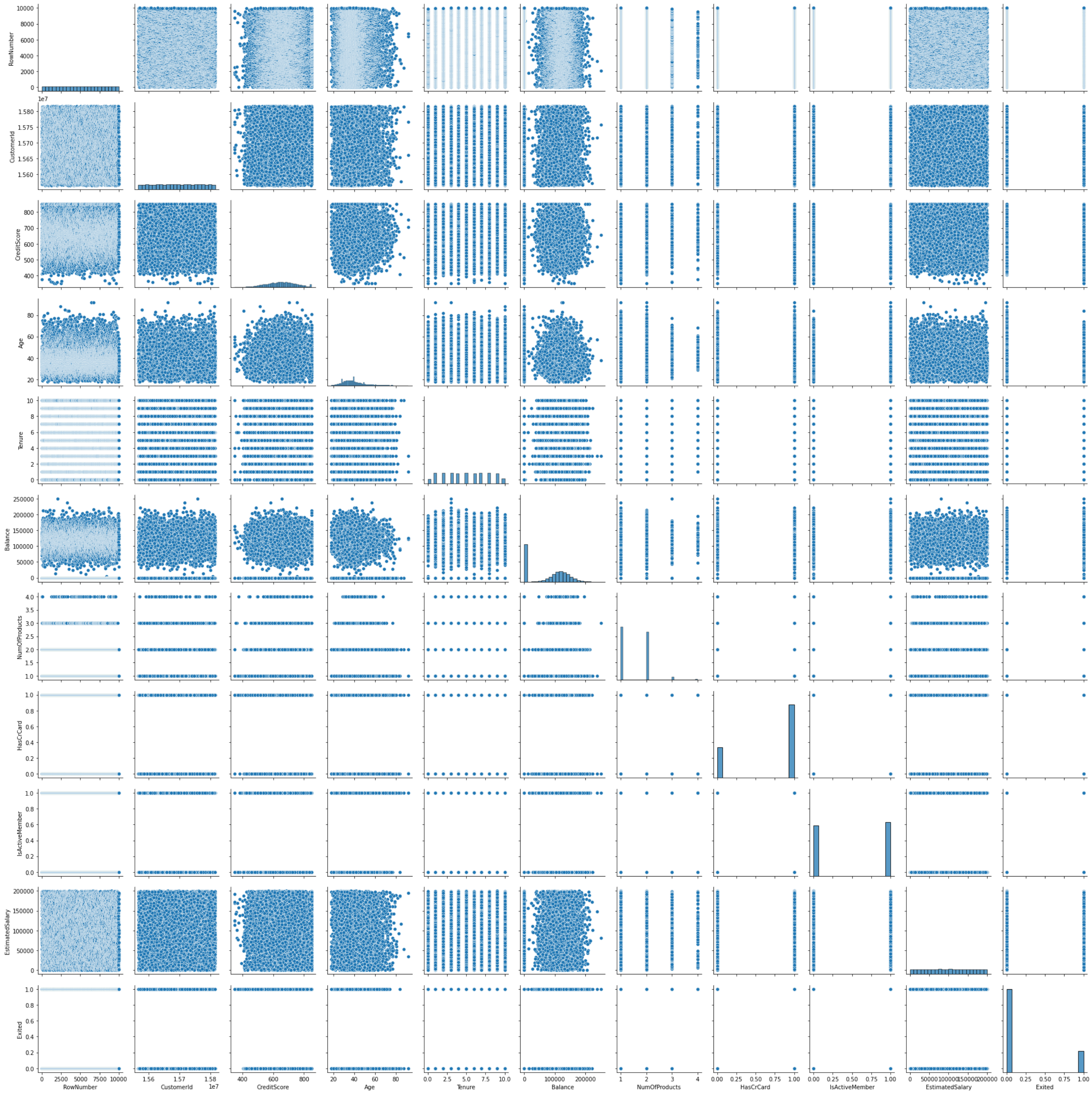
```
sns.barplot(x='IsActiveMember',y='Exited',data=df)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fedf5542350>
```



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

```
<seaborn.axisgrid.PairGrid at 0x7fedf56e7c10>
```



```
plt.figure(figsize=(8,5))
sns.heatmap(df.corr(),annot=True)V
```

```
  File "<ipython-input-32-a07fc315aa27>", line 2
    sns.heatmap(df.corr(),annot=True)V
                                     ^
SyntaxError: invalid syntax
```

SEARCH STACK OVERFLOW

```
df.Exited.value_counts()
```

```
0    7963
1    2037
Name: Exited, dtype: int64
```

```
df.isnull().sum()
```

```
RowNumber          0
CustomerId         0
Surname            0
CreditScore        0
Geography          0
Gender             0
Age                0
Tenure             0
Balance            0
NumOfProducts      0
HasCrCard          0
IsActiveMember     0
EstimatedSalary    0
Exited             0
dtype: int64
```

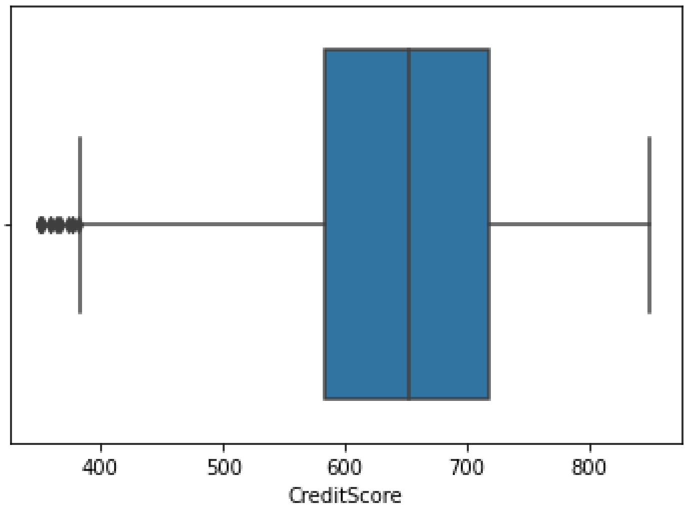
```
#No missing values
```

```
df.head(2)
```

	RowNumber	CustomerId	Surname	CreditScore	Geography	Gender	Age	Tenure	Bal
0	1	15634602	Hargrave	619	France	Female	42	2	
1	2	15647311	Hill	608	Spain	Female	41	1	838

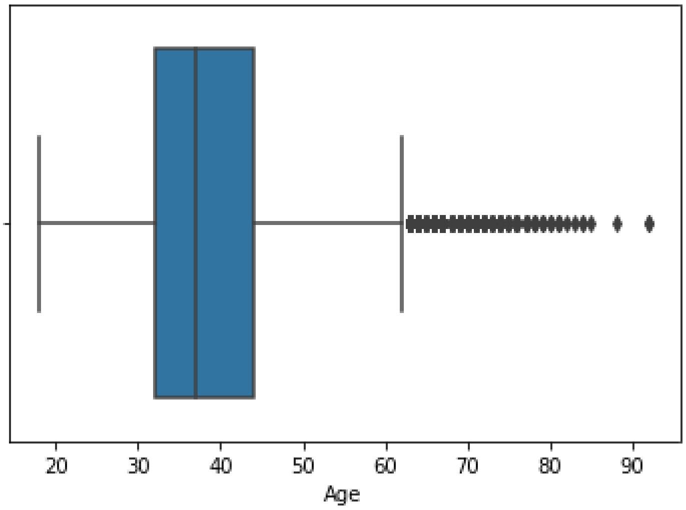
```
sns.boxplot(df.CreditScore)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fedf5824150>
```



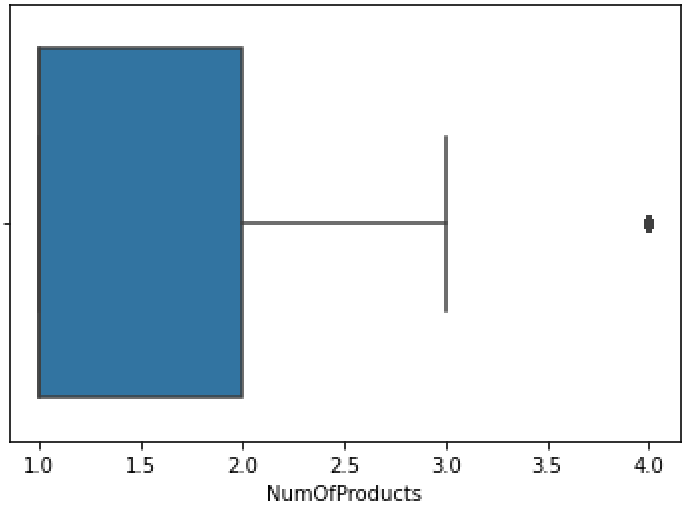
```
sns.boxplot(df.Age)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fedf06da790>
```



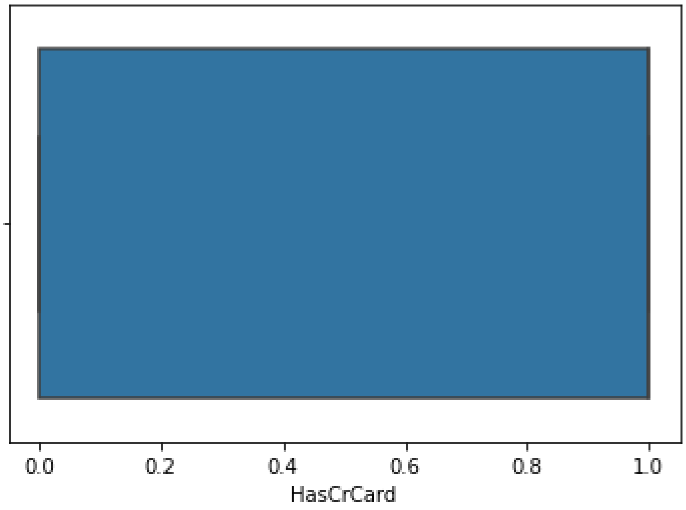
```
sns.boxplot(df.NumOfProducts)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fedef89c250>
```



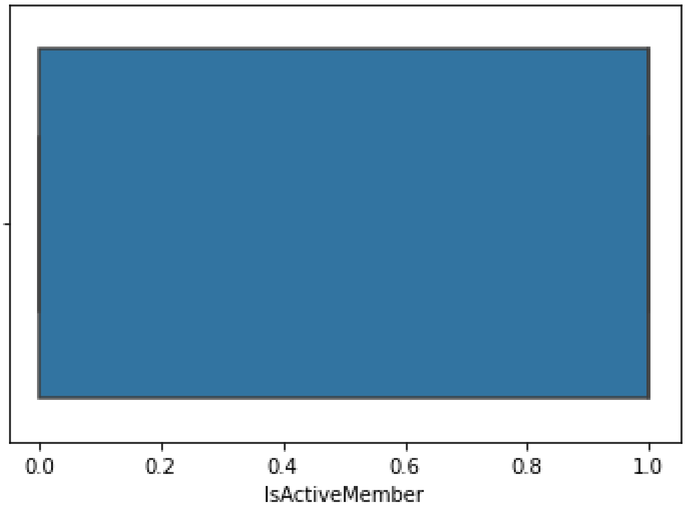
```
sns.boxplot(df.HasCrCard)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fedee060290>
```



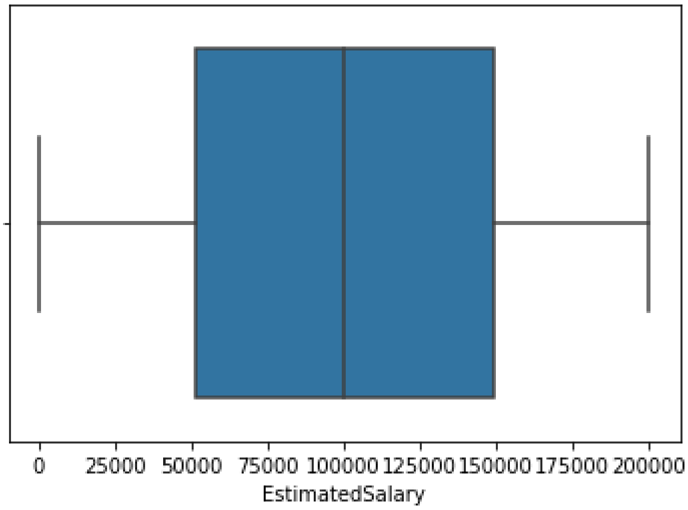
```
sns.boxplot(df.IsActiveMember)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fedee02d490>
```



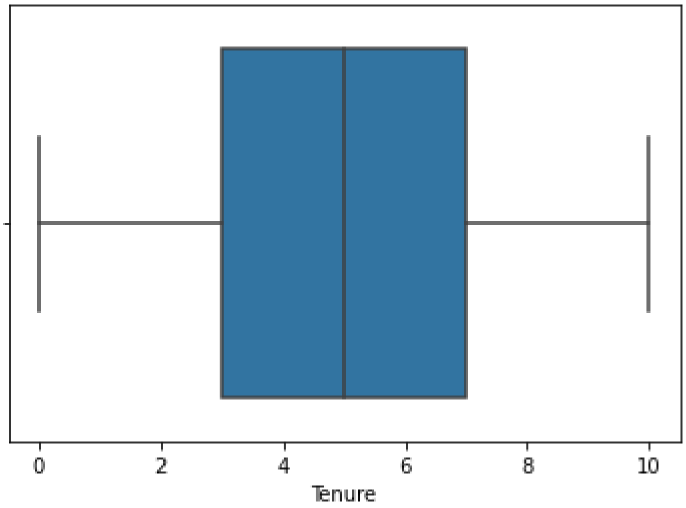
```
sns.boxplot(df.EstimatedSalary)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fededfa6190>
```



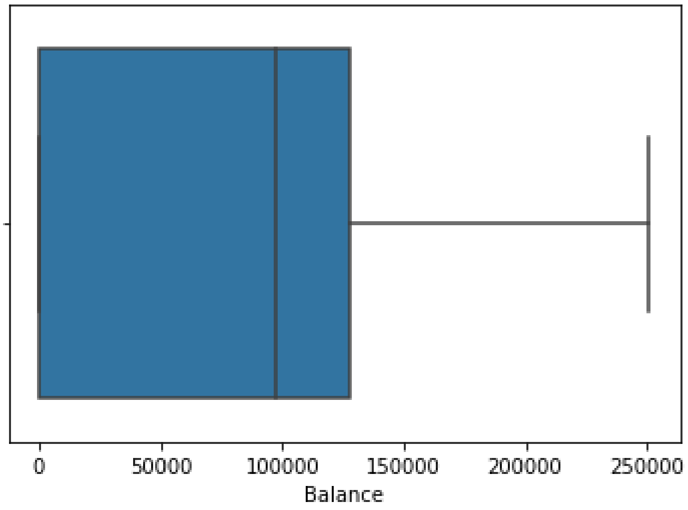
```
sns.boxplot(df.Tenure)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fededf7aa10>
```



```
sns.boxplot(df.Balance)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fedee03f350>
```



```
#Outlier Removal
```

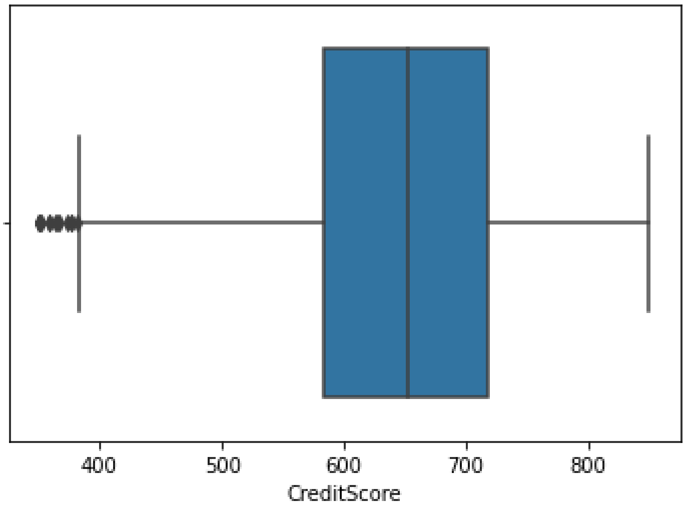
```
def outlier_credit_score(df):
```

```
  File "<ipython-input-45-109cb1ec7f34>", line 1
    def outlier_credit_score(df):
                                 ^
SyntaxError: unexpected EOF while parsing
```

SEARCH STACK OVERFLOW

```
sns.boxplot(df.CreditScore)
```

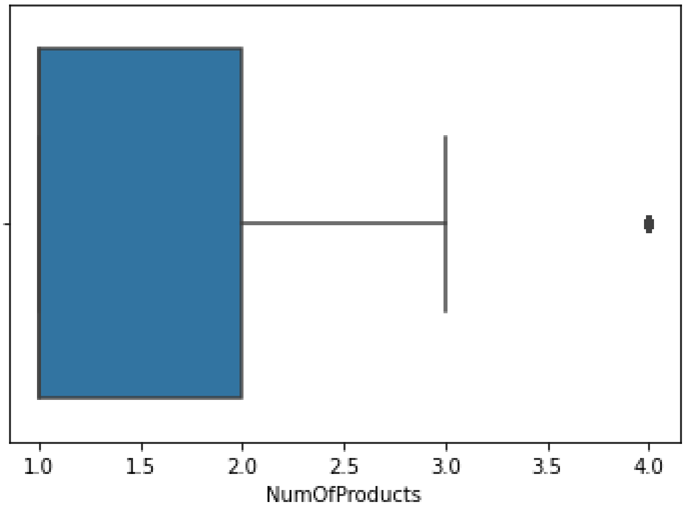
```
<matplotlib.axes._subplots.AxesSubplot at 0x7fedede784d0>
```



```
def outlier_NOP(df):
      File "<ipython-input-47-a02a39560060>", line 1
        def outlier_NOP(df):
```

```
sns.boxplot(df.NumOfProducts)
```

```
    <matplotlib.axes._subplots.AxesSubplot at 0x7fededdd5490>
```



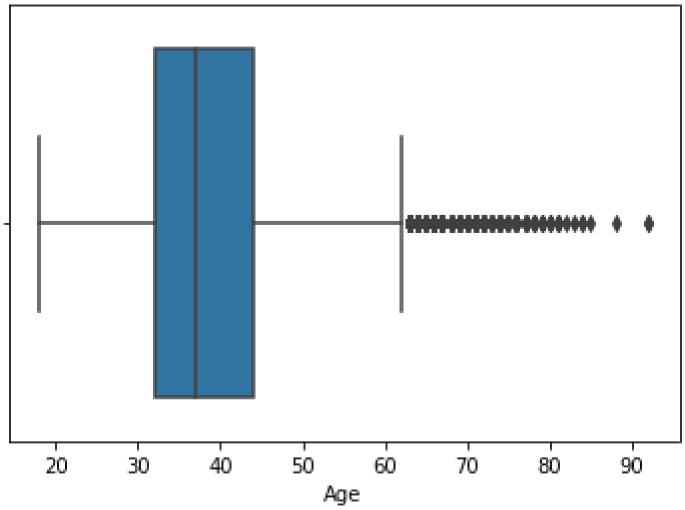
```
def outlier_age(df):
```

```
      File "<ipython-input-49-9f66786fd25b>", line 1
        def outlier_age(df):
                            ^
    SyntaxError: unexpected EOF while parsing
```

SEARCH STACK OVERFLOW

```
sns.boxplot(df.Age)
```

```
    <matplotlib.axes._subplots.AxesSubplot at 0x7fededdd5c50>
```



```
df.info()
```

```
    <class 'pandas.core.frame.DataFrame'>
    RangeIndex: 10000 entries, 0 to 9999
    Data columns (total 14 columns):
     #   Column           Non-Null Count  Dtype
```

```
 ---  ------          --------------  -----
  0   RowNumber        10000 non-null  int64
  1   CustomerId       10000 non-null  int64
  2   Surname          10000 non-null  object
  3   CreditScore      10000 non-null  int64
  4   Geography        10000 non-null  object
  5   Gender           10000 non-null  object
  6   Age              10000 non-null  int64
  7   Tenure           10000 non-null  int64
  8   Balance          10000 non-null  float64
  9   NumOfProducts    10000 non-null  int64
  10  HasCrCard        10000 non-null  int64
  11  IsActiveMember   10000 non-null  int64
  12  EstimatedSalary  10000 non-null  float64
  13  Exited           10000 non-null  int64
 dtypes: float64(2), int64(9), object(3)
 memory usage: 1.1+ MB
```

```
df.head(2)
```

	RowNumber	CustomerId	Surname	CreditScore	Geography	Gender	Age	Tenure	Bal
0	1	15634602	Hargrave	619	France	Female	42	2	
1	2	15647311	Hill	608	Spain	Female	41	1	838

```
df.drop(['CustomerId','RowNumber','Surname'],axis=1,inplace=True)
```

```
df.head(2)
```

	CreditScore	Geography	Gender	Age	Tenure	Balance	NumOfProducts	HasCrCard
0	619	France	Female	42	2	0.00	1	1
1	608	Spain	Female	41	1	83807.86	1	0

```
from sklearn.preprocessing import LabelEncoder
```

```
df.head(2)
```

	CreditScore	Geography	Gender	Age	Tenure	Balance	NumOfProducts	HasCrCard
0	619	France	Female	42	2	0.00	1	1
1	608	Spain	Female	41	1	83807.86	1	0

```
X=df.drop('Exited',axis=1)
y=df.Exited
```

```
X
```

	CreditScore	Geography	Gender	Age	Tenure	Balance	NumOfProducts	HasCrCa
0	619	France	Female	42	2	0.00	1	
1	608	Spain	Female	41	1	83807.86	1	
2	502	France	Female	42	8	159660.80	3	
3	699	France	Female	39	1	0.00	2	
4	850	Spain	Female	43	2	125510.82	1	
...	...	...	...	...	...	...	...	
9995	771	France	Male	39	5	0.00	2	
9996	516	France	Male	35	10	57369.61	1	
9997	709	France	Female	36	7	0.00	1	
9998	772	Germany	Male	42	3	75075.31	2	
9999	792	France	Female	28	4	130142.79	1	

10000 rows × 10 columns

```
from sklearn.preprocessing import StandardScaler
sc=StandardScaler()
X = sc.fit_transform(X)
```

```
---------------------------------------------------------------------------
ValueError                                Traceback (most recent call last)
<ipython-input-59-b4989add7a59> in <module>
      1 from sklearn.preprocessing import StandardScaler
      2 sc=StandardScaler()
----> 3 X = sc.fit_transform(X)

                              5 frames
/usr/local/lib/python3.7/dist-packages/pandas/core/generic.py in __array__(self,
dtype)
   1991
   1992     def __array__(self, dtype: NpDtype | None = None) -> np.ndarray:
-> 1993         return np.asarray(self._values, dtype=dtype)
   1994
   1995     def __array_wrap__(

ValueError: could not convert string to float: 'France'
```

SEARCH STACK OVERFLOW

```
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,
                                               random_state=42)
```

```
x_train.shape, x_test.shape, y_train.shape, y_test.shape
```

```
((8000, 10), (2000, 10), (8000,), (2000,))
```

✓ 0s completed at 11:12 AM ● ×