**Importing Libraries**

**import** numpy **as** np

**import** pandas **as** pd

**import** seaborn **as** sns

**import** matplotlib.pyplot **as** plt

**Loading Dataset**

df**=**pd**.**read\_csv(r'C:\Users\LENOVO\Downloads\Churn\_Modelling.csv')

df**.**shape

(10000, 14)

df**.**head()

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Row** | **Cust** | **Sur** | **Credi** | **Geog** | **Ge** | **A** | **Te** | **Bala** | **NumOf** | **HasC** | **IsActiv** | **Estimat** | **Ex** |  |
|  | **Num** | **omer** | **na** | **tScor** | **raph** | **nd** | **g** | **nu** | **Produc** | **rCar** | **eMemb** | **edSalar** | **ite** |  |
|  | **nce** |  |
|  | **ber** | **Id** | **me** | **e** | **y** | **er** | **e** | **re** | **ts** | **d** | **er** | **y** | **d** |  |
|  |  |  |
|  |  | 1563 | Har |  | Fran | Fe | 4 |  |  |  |  |  | 101348. |  |  |
| **0** | 1 | grav | 619 | mal | 2 | 0.00 | 1 | 1 | 1 | 1 |  |
| 4602 | ce | 2 | 88 |  |
|  |  | e |  | e |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  | 1564 |  |  | Spai | Fe | 4 |  | 8380 |  |  |  | 112542. |  |  |
| **1** | 2 | Hill | 608 | mal | 1 | 1 | 0 | 1 | 0 |  |
| 7311 | n | 1 | 7.86 | 58 |  |
|  |  |  |  | e |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  | 1561 | Oni |  | Fran | Fe | 4 |  | 1596 |  |  |  | 113931. |  |  |
| **2** | 3 | 502 | mal | 8 | 60.8 | 3 | 1 | 0 | 1 |  |
| 9304 | o | ce | 2 | 57 |  |
|  |  |  | e |  | 0 |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  | 1570 | Bon |  | Fran | Fe | 3 |  |  |  |  |  | 93826.6 |  |  |
| **3** | 4 | 699 | mal | 1 | 0.00 | 2 | 0 | 0 | 0 |  |
| 1354 | i | ce | 9 | 3 |  |
|  |  |  | e |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  | 1573 | Mit |  | Spai | Fe | 4 |  | 1255 |  |  |  | 79084.1 |  |  |
| **4** | 5 | chel | 850 | mal | 2 | 10.8 | 1 | 1 | 1 | 0 |  |
| 7888 | n | 3 | 0 |  |
|  |  | l |  | e |  | 2 |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
| df**.**tail() | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | **Row** | **Cust** | **Sur** | **Cred** | **Geo** | **Ge** | **A** | **Te** | **Bala** | **NumOf** | **Has** | **IsActiv** | **Estima** | **Ex** |  |
|  | **Num** | **omer** | **na** | **itSco** | **grap** | **nd** | **g** | **nu** | **Produc** | **CrC** | **eMemb** | **tedSala** | **ite** |  |
|  | **nce** |  |
|  | **ber** | **Id** | **me** | **re** | **hy** | **er** | **e** | **re** | **ts** | **ard** | **er** | **ry** | **d** |  |
|  |  |  |
| **9** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **9** | 9996 | 1560 | Obij | 771 | Fran | Ma | 3 | 5 | 0.00 | 2 | 1 | 0 | 96270.6 | 0 |  |
| **9** | 6229 | iaku | ce | le | 9 | 4 |  |
|  |  |  |  |  |  |  |  |  |
| **5** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Row** | **Cust** | **Sur** | **Cred** | **Geo** | **Ge** | **A** | **Te** | **Bala** | **NumOf** | **Has** | **IsActiv** | **Estima** | **Ex** |  |
|  | **Num** | **omer** | **na** | **itSco** | **grap** | **nd** | **g** | **nu** | **Produc** | **CrC** | **eMemb** | **tedSala** | **ite** |  |
|  | **nce** |  |
|  | **ber** | **Id** | **me** | **re** | **hy** | **er** | **e** | **re** | **ts** | **ard** | **er** | **ry** | **d** |  |
|  |  |  |
| **9** |  |  | Joh |  |  |  |  |  | 573 |  |  |  |  |  |  |
| **9** |  | 1556 |  | Fran | Ma | 3 |  |  |  |  | 101699. |  |  |
| 9997 | nsto | 516 | 10 | 69.6 | 1 | 1 | 1 | 0 |  |
| **9** | 9892 | ce | le | 5 | 77 |  |
|  | ne |  |  | 1 |  |  |  |  |  |
| **6** |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **9** |  |  |  |  |  | Fe |  |  |  |  |  |  |  |  |  |
| **9** |  | 1558 |  |  | Fran | 3 |  |  |  |  |  | 42085.5 |  |  |
| 9998 | Liu | 709 | ma | 7 | 0.00 | 1 | 0 | 1 | 1 |  |
| **9** | 4532 | ce | 6 | 8 |  |
|  |  |  | le |  |  |  |  |  |  |  |
| **7** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **9** |  |  | Sab |  | Ger |  |  |  | 750 |  |  |  |  |  |  |
| **9** |  | 1568 |  | Ma | 4 |  |  |  |  | 92888.5 |  |  |
| 9999 | bati | 772 | man | 3 | 75.3 | 2 | 1 | 0 | 1 |  |
| **9** | 2355 | le | 2 | 2 |  |
|  | ni |  | y |  | 1 |  |  |  |  |  |
| **8** |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **9** |  |  |  |  |  | Fe |  |  | 130 |  |  |  |  |  |  |
| **9** |  | 1562 | Wal |  | Fran | 2 |  |  |  |  | 38190.7 |  |  |
| 10000 | 792 | ma | 4 | 142. | 1 | 1 | 0 | 0 |  |
| **9** | 8319 | ker | ce | 8 | 8 |  |
|  |  | le |  | 79 |  |  |  |  |  |
| **9** |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

**Visualizations**

**1. Univariate**

sns**.**distplot(df['EstimatedSalary'],hist**=True**)

C:\Users\LENOVO\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 flexibility) or `histplot` (an axes-level function for histograms).

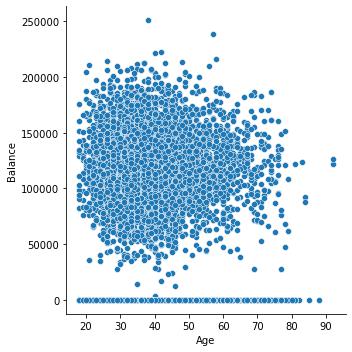
warnings.warn(msg, FutureWarning) <AxesSubplot:xlabel='EstimatedSalary', ylabel='Density'>



**2. Bivariate**

sns**.**relplot(x**=**'Age',y**=**'Balance',data**=**df)

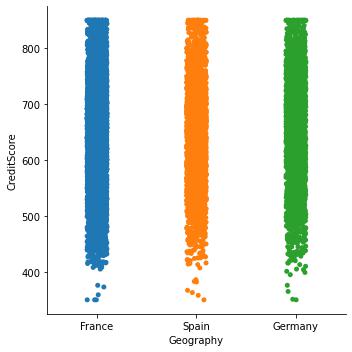
<seaborn.axisgrid.FacetGrid at 0x1fd190d2070>



sns**.**catplot(x**=**'Geography',y**=**'CreditScore',data**=**df)

*# for categorical data*

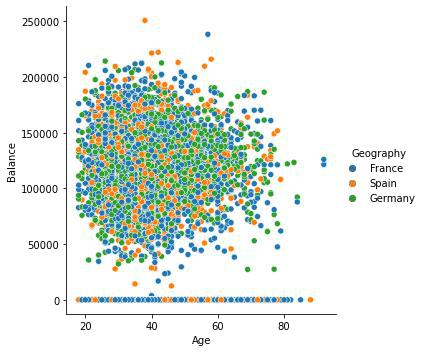
<seaborn.axisgrid.FacetGrid at 0x1fd194e8d90>



**3. Multivariate**

sns**.**relplot(x**=**'Age',y**=**'Balance',hue**=**'Geography',data**=**df)

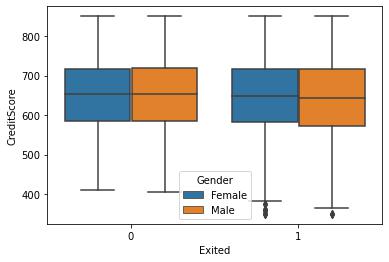
<seaborn.axisgrid.FacetGrid at 0x1fd1956e580>



sns**.**boxplot(x**=**'Exited',y**=**'CreditScore',hue**=**'Gender',data**=**df)

*# for categorical data*

<AxesSubplot:xlabel='Exited', ylabel='CreditScore'>



**Descriptive Statistics**

df**.**describe()

**RowN**

**umbe**

**r**

**co** 10000.

**un** 00000 **t**

1. 5000.5

**ea**0000

**n**

**st** 2886.8

1. 9568

**mi** 1.0000

1. 0
2. 2500.7

**%**5000

**Custo**

**merId**

1.0000

00e+0

4

1.5690

94e+0

7

7.1936

19e+0

4

1.5565

70e+0

7

1.5628

53e+0

7

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Credit** |  | **Tenur** | **Balanc** | **NumOf** |  |
| **Age** | **Product** |  |
| **Score** | **e** | **e** |  |
|  | **s** |  |
|  |  |  |  |  |
| 10000. | 10000. | 10000. | 10000. | 10000.0 |  |
| 00000 | 00000 | 00000 |  |
| 000000 | 00000 |  |
| 0 | 0 | 0 |  |
|  |  |  |
| 650.52 | 38.921 | 5.0128 | 76485. | 1.53020 |  |
| 8800 | 800 | 00 | 889288 | 0 |  |
| 96.653 | 10.487 | 2.8921 | 62397. | 0.58165 |  |
| 299 | 806 | 74 | 405202 | 4 |  |
| 350.00 | 18.000 | 0.0000 | 0.0000 | 1.00000 |  |
| 0000 | 000 | 00 | 00 | 0 |  |
| 584.00 | 32.000 | 3.0000 | 0.0000 | 1.00000 |  |
| 0000 | 000 | 00 | 00 | 0 |  |

**HasC**

**rCard**

10000

.0000

0

0.705

50

0.455

84

0.000

00

0.000

00

|  |  |
| --- | --- |
| **IsActive** | **Estimat** |
| **Membe** | **edSalar** |
| **r** | **y** |
| 10000.0 | 10000.0 |
| 00000 | 00000 |
| 0.51510 | 100090. |
| 0 | 239881 |
| 0.49979 | 57510.4 |
| 7 | 92818 |
| 0.00000 | 11.5800 |
| 0 | 00 |
| 0.00000 | 51002.1 |
| 0 | 10000 |

**Exited**

10000.

00000

0

0.2037

00

0.4027

69

0.0000

00

0.0000

00

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **RowN** | **Custo** | **Credit** |  | **Tenur** | **Balanc** | **NumOf** | **HasC** | **IsActive** | **Estimat** |  |  |
|  | **umbe** | **Age** | **Product** | **Membe** | **edSalar** | **Exited** |  |
|  | **merId** | **Score** | **e** | **e** | **rCard** |  |
|  | **r** |  | **s** | **r** | **y** |  |  |
|  |  |  |  |  |  |  |  |  |
| **50** | 5000.5 | 1.5690 | 652.00 | 37.000 | 5.0000 | 97198. | 1.00000 | 1.000 | 1.00000 | 100193. | 0.0000 |  |
| 74e+0 |  |
| **%** | 0000 | 0000 | 000 | 00 | 540000 | 0 | 00 | 0 | 915000 | 00 |  |
| 7 |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
| **75** | 7500.2 | 1.5753 | 718.00 | 44.000 | 7.0000 | 127644 | 2.00000 | 1.000 | 1.00000 | 149388. | 0.0000 |  |
| 23e+0 | .24000 |  |
| **%** | 5000 | 0000 | 000 | 00 | 0 | 00 | 0 | 247500 | 00 |  |
| 7 | 0 |  |
|  |  |  |  |  |  |  |  |  |  |  |
| **m** | 10000. | 1.5815 | 850.00 | 92.000 | 10.000 | 250898 | 4.00000 | 1.000 | 1.00000 | 199992. | 1.0000 |  |
| 69e+0 | .09000 |  |
| **ax** | 00000 | 0000 | 000 | 000 | 0 | 00 | 0 | 480000 | 00 |  |
| 7 | 0 |  |
|  |  |  |  |  |  |  |  |  |  |  |

**Handling the missing(null) 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

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

*# Since no null values are found no need to handle*

**Split the data into dependent and independent variables**

x**=**df**.**iloc[:,3:13]**.**values

print(x**.**shape)

y**=**df**.**iloc[:,13:14]**.**values

print(y**.**shape)

(10000, 10)

(10000, 1)

**Finding and Replacing Outliers**

df**.**skew()

C:\Users\LENOVO\AppData\Local\Temp\ipykernel\_3336\1665899112.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.

df.skew()

RowNumber 0.000000

CustomerId 0.001149

CreditScore -0.071607

Age 1.011320

Tenure 0.010991

Balance -0.141109

NumOfProducts 0.745568

HasCrCard -0.901812

IsActiveMember -0.060437

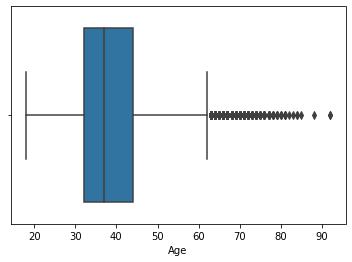
EstimatedSalary 0.002085

Exited 1.471611

dtype: float64

sns**.**boxplot(df["Age"])

<AxesSubplot:xlabel='Age'>



q1 **=** df["Age"]**.**describe()["25%"]

q3 **=** df["Age"]**.**describe()["75%"]

iqr **=** q3**-**q1

l\_b **=** q1 **-**(1.5**\***iqr)

u\_b **=** q3 **+** (1.5**\***iqr)

df[df["Age"]**<**l\_b]

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Row** | **Cust** | **Sur** | **Credi** | **Geog** | **Ge** | **A** | **Te** | **Bal** | **NumOf** | **HasC** | **IsActiv** |
| **Num** | **omer** | **na** | **tScor** | **raph** | **nd** | **g** | **nu** | **anc** | **Produc** | **rCar** | **eMemb** |
| **ber** | **Id** | **me** | **e** | **y** | **er** | **e** | **re** | **e** | **ts** | **d** | **er** |

df[df["Age"]**>**u\_b]

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Row** | **Cust** | **Surna** | **Cred** | **Geo** | **Ge** | **A** | **Te** | **Bal** | **NumO** | **Has** | **IsActiv** |  |
|  | **Num** | **omer** | **itSco** | **grap** | **nd** | **g** | **nu** | **anc** | **fProdu** | **CrC** | **eMemb** |  |
|  | **me** |  |
|  | **ber** | **Id** | **re** | **hy** | **er** | **e** | **re** | **e** | **cts** | **ard** | **er** |  |
|  |  |  |
| **5** |  | 1562 |  |  | Spai | Fe | 6 |  |  |  |  |  |  |
| 59 | T'ien | 511 | ma | 4 | 0.00 | 1 | 1 | 0 |  |
| **8** | 3944 | n | 6 |  |
|  |  |  | le |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
| **8** |  | 1580 | Nduk |  | Spai | Fe | 7 |  |  |  |  |  |  |
| 86 | 652 | ma | 10 | 0.00 | 2 | 1 | 1 |  |
| **5** | 5254 | aku | n | 5 |  |
|  |  | le |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
| **1** |  | 1580 | Dunb |  | Spai | Fe | 6 |  |  |  |  |  |  |
| **0** | 105 | 670 | ma | 1 | 0.00 | 1 | 1 | 1 |  |
| 4919 | abin | n | 5 |  |
| **4** |  |  | le |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |

**Estimat edSalar y**

**Estima tedSala ry**

1643.1

1

114675

.75

177655

.68

**Ex**

**ite**

**d**

**Ex**

**ite**

**d**

1

0

1

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Row** | **Cust** | **Surna** | **Cred** | **Geo** | **Ge** | **A** | **Te** | **Bal** | **NumO** | **Has** | **IsActiv** | **Estima** | **Ex** |  |
| **Num** | **omer** | **itSco** | **grap** | **nd** | **g** | **nu** | **anc** | **fProdu** | **CrC** | **eMemb** | **tedSala** | **ite** |  |
| **me** |  |
| **ber** | **Id** | **re** | **hy** | **er** | **e** | **re** | **e** | **cts** | **ard** | **er** | **ry** | **d** |  |
|  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **1** |  | 1558 | Macle |  |
| **5** | 159 |  |
| 9975 | an |  |
| **8** |  |  |
|  |  |  |  |
| **1** |  | 1578 |  |  |
| **8** | 182 | Hsia |  |
| 9669 |  |
| **1** |  |  |  |
|  |  |  |  |
| **...** | ... | ... | ... |  |
| **9** |  |  |  |  |
| **7** | 9754 | 1570 | Chied |  |
| **5** | 5174 | ozie |  |
|  |  |
| **3** |  |  |  |  |
| **9** |  |  |  |  |
| **7** | 9766 | 1577 | Thom |  |
| **6** | 7067 | as |  |
|  |  |
| **5** |  |  |  |  |
| **9** |  |  | Chuk |  |
| **8** |  | 1581 |  |
| 9833 | wujek |  |
| **3** | 4690 |  |
|  | wu |  |
| **2** |  |  |  |
|  |  |  |  |
| **9** |  |  |  |  |
| **8** | 9895 | 1570 | Vagin |  |
| **9** | 4795 |  |
|  |  |  |
| **4** |  |  |  |  |
| **9** |  |  |  |  |
| **9** | 9937 | 1565 | Parks |  |
| **3** | 3037 |  |
|  |  |  |
| **6** |  |  |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Fran | Fe | 7 |  |
| 646 | ma |  |
| ce | 3 |  |
|  | le |  |
|  |  |  |  |
| 510 | Fran | Ma | 6 |  |
| ce | le | 5 |  |
|  |  |
| ... | ... | ... ... | |  |
|  | Ger | Ma | 6 |  |
| 656 | man |  |
| le | 8 |  |
|  | y |  |
|  |  |  |  |
| 445 | Fran | Ma | 6 |  |
| ce | le | 4 |  |
|  |  |
|  | Ger | Fe | 6 |  |
| 595 | man | ma |  |
| 4 |  |
|  | y | le |  |
|  |  |  |
|  | Fran | Fe | 7 |  |
| 521 | ma |  |
| ce | 7 |  |
|  | le |  |
|  |  |  |  |
| 609 | Fran | Ma | 7 |  |
| ce | le | 7 |  |
|  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 972 |  |  |  |
| 6 | 59.2 | 1 | 0 | 1 |
|  | 5 |  |  |  |
| 2 | 0.00 | 2 | 1 | 1 |
| ... | ... | ... | ... | ... |
|  | 153 |  |  |  |
| 7 | 545. | 1 | 1 | 1 |
|  | 11 |  |  |  |
|  | 136 |  |  |  |
| 2 | 770. | 1 | 0 | 1 |
|  | 67 |  |  |  |
|  | 105 |  |  |  |
| 2 | 736. | 1 | 1 | 1 |
|  | 32 |  |  |  |
| 6 | 0.00 | 2 | 1 | 1 |
| 1 | 0.00 | 1 | 0 | 1 |

104719

.66

48071.

61

...

186574

.68

43678.

06

89935.

73

49054.

10

18708.

76

0

0

...

0

0

1

0

0

359 rows × 14 columns

*#Replace the outlier*

outlier\_list **=** list(df[df["Age"] **>** u\_b]["Age"])

outlier\_list

[66,

75,

65,

73,

65,

72,

67,

67,

79,

80,

68,

75,

66,

66,

70,

63,

72,

64,

64,

70,

67,

82,

63,

69,

65,

69,

64,

65,

74,

67,

66,

67,

63,

70,

71,

72,

67,

74,

76,

66,

63,

66,

68,

67,

63,

71,

66,

69,

73,

65,

66,

64,

69,

64,

77,

74,

65,

70,

67,

69,

67,

74,

69,

74,

74,

64,

63,

63,

70,

74,

65,

72,

77,

66,

65,

74,

88,

63,

71,

63,

64,

67,

70,

68,

72,

71,

66,

75,

67,

73,

69,

76,

63,

85,

67,

74,

76,

66,

69,

66,

72,

63,

71,

63,

74,

67,

72,

72,

66,

84,

71,

66,

63,

74,

69,

84,

67,

64,

68,

66,

77,

70,

67,

79,

67,

76,

73,

66,

67,

64,

73,

76,

72,

64,

71,

63,

70,

65,

66,

65,

80,

66,

63,

63,

63,

63,

66,

74,

69,

63,

64,

76,

75,

68,

69,

77,

64,

66,

74,

71,

67,

68,

64,

68,

70,

64,

75,

66,

64,

78,

65,

74,

64,

64,

71,

77,

79,

70,

81,

64,

68,

68,

63,

79,

66,

64,

70,

69,

71,

72,

66,

68,

63,

71,

72,

72,

64,

78,

75,

65,

65,

67,

63,

68,

71,

73,

64,

66,

71,

69,

71,

66,

76,

69,

73,

64,

64,

75,

73,

71,

72,

63,

67,

68,

73,

67,

64,

63,

92,

65,

75,

67,

71,

64,

66,

64,

66,

67,

77,

92,

67,

63,

66,

66,

68,

65,

72,

71,

76,

63,

67,

67,

66,

67,

63,

65,

70,

72,

77,

74,

72,

73,

77,

67,

71,

64,

72,

81,

76,

69,

68,

74,

64,

64,

71,

68,

63,

67,

63,

64,

76,

63,

63,

68,

67,

72,

70,

81,

67,

73,

66,

68,

71,

66,

63,

75,

69,

64,

69,

70,

71,

71,

66,

70,

63,

64,

65,

63,

67,

71,

67,

65,

66,

63,

73,

66,

64,

72,

71,

69,

67,

64,

81,

73,

63,

67,

74,

83,

69,

71,

78,

63,

70,

69,

72,

70,

63,

74,

80,

69,

72,

67,

76,

71,

67,

71,

78,

63,

63,

68,

64,

70,

78,

69,

68,

64,

64,

77,

77]

outlier\_dict **=** {}**.**fromkeys(outlier\_list,u\_b)

outlier\_dict

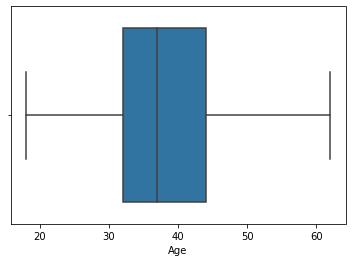
{66: 62.0,

1. 62.0,
2. 62.0,
3. 62.0,
4. 62.0,
5. 62.0,
6. 62.0,
7. 62.0,
8. 62.0,
9. 62.0,
10. 62.0,
11. 62.0,
12. 62.0,
13. 62.0,
14. 62.0,
15. 62.0,
16. 62.0,
17. 62.0,
18. 62.0,
19. 62.0,
20. 62.0,
21. 62.0,
22. 62.0,
23. 62.0,
24. 62.0}

df["Age"] **=** df["Age"]**.**replace(outlier\_dict)

sns**.**boxplot(df["Age"])

<AxesSubplot:xlabel='Age'>



df[df["Age"]**>**u\_b]

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Row** | **Cust** | **Sur** | **Credi** | **Geog** | **Ge** | **A** | **Te** | **Bal** | **NumOf** | **HasC** | **IsActiv** | **Estimat** | **Ex** |
| **Num** | **omer** | **na** | **tScor** | **raph** | **nd** | **g** | **nu** | **anc** | **Produc** | **rCar** | **eMemb** | **edSalar** | **ite** |
| **ber** | **Id** | **me** | **e** | **y** | **er** | **e** | **re** | **e** | **ts** | **d** | **er** | **y** | **d** |

**Check for Categorical columns and perform encoding.**

**from** sklearn.compose **import** ColumnTransformer

**from** sklearn.preprocessing **import** OneHotEncoder

ct**=**ColumnTransformer([('oh',OneHotEncoder(),[1,2])],remainder**=**'passthrough')

x**=**ct**.**fit\_transform(x)

print(x**.**shape)

(10000, 13)

* *saving the data* **import** joblib joblib**.**dump(ct,"churnct.pkl")

['churnct.pkl']

**Split the data into training and testing**

**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**=**0)

print(x\_train**.**shape)

print(x\_test**.**shape)

(8000, 13)

(2000, 13)

**Scale the independent variables**

**from** sklearn.preprocessing **import** StandardScaler

sc**=**StandardScaler()

x\_train**=**sc**.**fit\_transform(x\_train)

x\_test**=**sc**.**transform(x\_test)

joblib**.**dump(sc,"churnsc.pkl")

['churnsc.pkl']