Data Frames

August 17, 2021

1 Data Frames setting(working on data)

[4]:
```
import pandas as pd
```

2 Method 1: Specify full path

[9]:
```
stats=pd.read_csv('C:\\Users\\ddaya\\OneDrive\\Documents\\Python programming\\
                   P1-OfficeSupplies.csv')
```

[10]:
```
stats
```

[10]:

	OrderDate	Region	Rep	Item	Units	Unit Price
0	4-Jul-14	East	Richard	Pen Set	62	4.99
1	12-Jul-14	East	Nick	Binder	29	1.99
2	21-Jul-14	Central	Morgan	Pen Set	55	12.49
3	29-Jul-14	East	Susan	Binder	81	19.99
4	7-Aug-14	Central	Matthew	Pen Set	42	23.95
5	15-Aug-14	East	Richard	Pencil	35	4.99
6	24-Aug-14	West	James	Desk	3	275.00
7	1-Sep-14	Central	Smith	Desk	2	125.00
8	10-Sep-14	Central	Bill	Pencil	7	1.29
9	18-Sep-14	East	Richard	Pen Set	16	15.99
10	27-Sep-14	West	James	Pen	76	1.99
11	5-Oct-14	Central	Morgan	Binder	28	8.99
12	14-Oct-14	West	Thomas	Binder	57	19.99
13	22-Oct-14	East	Richard	Pen	64	8.99
14	31-Oct-14	Central	Rachel	Pencil	14	1.29
15	8-Nov-14	East	Susan	Pen	15	19.99
16	17-Nov-14	Central	Alex	Binder	11	4.99
17	25-Nov-14	Central	Matthew	Pen Set	96	4.99
18	4-Dec-14	Central	Alex	Binder	94	19.99
19	12-Dec-14	Central	Smith	Pencil	67	1.29
20	21-Dec-14	Central	Rachel	Binder	28	4.99
21	29-Dec-14	East	Susan	Pen Set	74	15.99
22	6-Jan-15	East	Richard	Pencil	95	1.99
23	15-Jan-15	Central	Bill	Binder	46	8.99

24	23-Jan-15	Central	Matthew	Binder	50	19.99
25	1-Feb-15	Central	Smith	Binder	87	15.00
26	9-Feb-15	Central	Alex	Pencil	36	4.99
27	18-Feb-15	East	Richard	Binder	4	4.99
28	26-Feb-15	Central	Bill	Pen	27	19.99
29	7-Mar-15	West	James	Binder	7	19.99
30	15-Mar-15	West	James	Pencil	56	2.99
31	24-Mar-15	Central	Alex	Pen Set	50	4.99
32	1-Apr-15	East	Richard	Binder	60	4.99
33	10-Apr-15	Central	Rachel	Pencil	66	1.99
34	18-Apr-15	Central	Rachel	Pencil	75	1.99
35	27-Apr-15	East	Nick	Pen	96	4.99
36	5-May-15	Central	Alex	Pencil	90	4.99
37	14-May-15	Central	Bill	Pencil	53	1.29
38	22-May-15	West	Thomas	Pencil	32	1.99
39	31-May-15	Central	Bill	Binder	80	8.99
40	8-Jun-15	East	Richard	Binder	60	8.99
41	17-Jun-15	Central	Matthew	Desk	5	125.00
42	25-Jun-15	Central	Morgan	Pencil	90	4.99

3 Method 2:Change the working Directory

[11]:
```
import os
```

[12]:
```
print(os.getcwd())
```

```
C:\Users\ddaya\Documents\Python Programs
```

[13]:
```
os.chdir('C:\\Users\\ddaya\\OneDrive\\Documents\\Python programming')
```

[14]:
```
print(os.getcwd())
```

```
C:\Users\ddaya\OneDrive\Documents\Python programming
```

[15]:
```
stats=pd.read_csv('P1-OfficeSupplies.csv')
```

[16]:
```
stats
```

[16]:

	OrderDate	Region	Rep	Item	Units	Unit Price
0	4-Jul-14	East	Richard	Pen Set	62	4.99
1	12-Jul-14	East	Nick	Binder	29	1.99
2	21-Jul-14	Central	Morgan	Pen Set	55	12.49
3	29-Jul-14	East	Susan	Binder	81	19.99
4	7-Aug-14	Central	Matthew	Pen Set	42	23.95
5	15-Aug-14	East	Richard	Pencil	35	4.99
6	24-Aug-14	West	James	Desk	3	275.00
7	1-Sep-14	Central	Smith	Desk	2	125.00
8	10-Sep-14	Central	Bill	Pencil	7	1.29

```
9   18-Sep-14     East  Richard  Pen Set     16       15.99
10  27-Sep-14     West    James      Pen     76        1.99
11   5-Oct-14  Central   Morgan   Binder     28        8.99
12  14-Oct-14     West   Thomas   Binder     57       19.99
13  22-Oct-14     East  Richard      Pen     64        8.99
14  31-Oct-14  Central   Rachel   Pencil     14        1.29
15   8-Nov-14     East    Susan      Pen     15       19.99
16  17-Nov-14  Central     Alex   Binder     11        4.99
17  25-Nov-14  Central  Matthew  Pen Set     96        4.99
18   4-Dec-14  Central     Alex   Binder     94       19.99
19  12-Dec-14  Central    Smith   Pencil     67        1.29
20  21-Dec-14  Central   Rachel   Binder     28        4.99
21  29-Dec-14     East    Susan  Pen Set     74       15.99
22   6-Jan-15     East  Richard   Pencil     95        1.99
23  15-Jan-15  Central     Bill   Binder     46        8.99
24  23-Jan-15  Central  Matthew   Binder     50       19.99
25   1-Feb-15  Central    Smith   Binder     87       15.00
26   9-Feb-15  Central     Alex   Pencil     36        4.99
27  18-Feb-15     East  Richard   Binder      4        4.99
28  26-Feb-15  Central     Bill      Pen     27       19.99
29   7-Mar-15     West    James   Binder      7       19.99
30  15-Mar-15     West    James   Pencil     56        2.99
31  24-Mar-15  Central     Alex  Pen Set     50        4.99
32   1-Apr-15     East  Richard   Binder     60        4.99
33  10-Apr-15  Central   Rachel   Pencil     66        1.99
34  18-Apr-15  Central   Rachel   Pencil     75        1.99
35  27-Apr-15     East     Nick      Pen     96        4.99
36   5-May-15  Central     Alex   Pencil     90        4.99
37  14-May-15  Central     Bill   Pencil     53        1.29
38  22-May-15     West   Thomas   Pencil     32        1.99
39  31-May-15  Central     Bill   Binder     80        8.99
40   8-Jun-15     East  Richard   Binder     60        8.99
41  17-Jun-15  Central  Matthew     Desk      5      125.00
42  25-Jun-15  Central   Morgan   Pencil     90        4.99
```

[17]:
```
stats=pd.read_csv('P1-UK-Bank-Customers.csv')
```

[18]:
```
stats
```

[18]:
```
      Customer ID       Name   Surname  Gender  Age            Region  \
0       100000001      Simon     Walsh    Male   21           England
1       400000002    Jasmine    Miller  Female   34  Northern Ireland
2       100000003       Liam     Brown    Male   46           England
3       300000004     Trevor      Parr    Male   32             Wales
4       100000005    Deirdre   Pullman  Female   38           England
…             …          …         …       …  …                 …
4009    200004010        Sam     Lewis    Male   64          Scotland
```

```
4010    200004011      Keith    Hughes    Male   52          Scotland
4011    200004012     Hannah  Springer  Female   50          Scotland
4012    200004013  Christian      Reid    Male   51          Scotland
4013    300004014    Stephen       May    Male   33             Wales

     Job Classification Date Joined    Balance
0          White Collar   05.Jan.15  113810.15
1           Blue Collar   06.Jan.15   36919.73
2          White Collar   07.Jan.15  101536.83
3          White Collar   08.Jan.15    1421.52
4           Blue Collar   09.Jan.15   35639.79
…                  …         …         …
4009              Other   30.Dec.15   19711.66
4010        Blue Collar   30.Dec.15   56069.72
4011              Other   30.Dec.15   59477.82
4012        Blue Collar   30.Dec.15     239.45
4013        Blue Collar   30.Dec.15   30293.19

[4014 rows x 9 columns]
```

4 1. Full Data

[19]:
```
stats
```

[19]:
```
      Customer ID       Name   Surname  Gender  Age            Region  \
0       100000001      Simon     Walsh    Male   21           England
1       400000002    Jasmine    Miller  Female   34  Northern Ireland
2       100000003       Liam     Brown    Male   46           England
3       300000004     Trevor      Parr    Male   32             Wales
4       100000005    Deirdre   Pullman  Female   38           England
…             …          …       …       …  …               …
4009    200004010        Sam     Lewis    Male   64          Scotland
4010    200004011      Keith    Hughes    Male   52          Scotland
4011    200004012     Hannah  Springer  Female   50          Scotland
4012    200004013  Christian      Reid    Male   51          Scotland
4013    300004014    Stephen       May    Male   33             Wales

     Job Classification Date Joined    Balance
0          White Collar   05.Jan.15  113810.15
1           Blue Collar   06.Jan.15   36919.73
2          White Collar   07.Jan.15  101536.83
3          White Collar   08.Jan.15    1421.52
4           Blue Collar   09.Jan.15   35639.79
…                  …         …         …
4009              Other   30.Dec.15   19711.66
4010        Blue Collar   30.Dec.15   56069.72
```

```
4011              Other   30.Dec.15   59477.82
4012        Blue Collar   30.Dec.15     239.45
4013        Blue Collar   30.Dec.15   30293.19

[4014 rows x 9 columns]
```

5 2. Number of rows

[20]:
```
len(stats)
```

[20]:
```
4014
```

6 3. See Columns

[21]:
```
stats.columns
```

[21]:
```
Index(['Customer ID', 'Name', 'Surname', 'Gender', 'Age', 'Region',
       'Job Classification', 'Date Joined', 'Balance'],
      dtype='object')
```

7 4. Number of Columns

[22]:
```
len(stats.columns)
```

[22]:
```
9
```

8 5. Top Rows

[23]:
```
stats.head(6) # Remember the brackets
```

[23]:
```
   Customer ID     Name  Surname  Gender  Age            Region  \
0    100000001    Simon    Walsh    Male   21           England
1    400000002  Jasmine   Miller  Female   34  Northern Ireland
2    100000003     Liam    Brown    Male   46           England
3    300000004   Trevor     Parr    Male   32             Wales
4    100000005  Deirdre  Pullman  Female   38           England
5    300000006      Ava  Coleman  Female   30             Wales

  Job Classification Date Joined    Balance
0       White Collar   05.Jan.15  113810.15
1        Blue Collar   06.Jan.15   36919.73
2       White Collar   07.Jan.15  101536.83
3       White Collar   08.Jan.15    1421.52
4        Blue Collar   09.Jan.15   35639.79
5        Blue Collar   09.Jan.15  122443.77
```

9 6. Bottom Rows

[24]:
```
stats.tail() # Or stats.tail(10)
```

[24]:
```
      Customer ID       Name   Surname  Gender  Age    Region  \
4009    200004010        Sam     Lewis    Male   64  Scotland
4010    200004011      Keith    Hughes    Male   52  Scotland
4011    200004012     Hannah  Springer  Female   50  Scotland
4012    200004013  Christian      Reid    Male   51  Scotland
4013    300004014    Stephen       May    Male   33     Wales

     Job Classification Date Joined   Balance
4009              Other   30.Dec.15  19711.66
4010        Blue Collar   30.Dec.15  56069.72
4011              Other   30.Dec.15  59477.82
4012        Blue Collar   30.Dec.15    239.45
4013        Blue Collar   30.Dec.15  30293.19
```

10 7. Information on the columns

[25]:
```
stats.info() #  Like the str function in R
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4014 entries, 0 to 4013
Data columns (total 9 columns):
 #   Column              Non-Null Count  Dtype
---  ------              --------------  -----
 0   Customer ID         4014 non-null   int64
 1   Name                4014 non-null   object
 2   Surname             4014 non-null   object
 3   Gender              4014 non-null   object
 4   Age                 4014 non-null   int64
 5   Region              4014 non-null   object
 6   Job Classification  4014 non-null   object
 7   Date Joined         4014 non-null   object
 8   Balance             4014 non-null   float64
dtypes: float64(1), int64(2), object(6)
memory usage: 282.4+ KB
```

[26]:
```
# 8. get stats on the columns
```

[27]:
```
stats.describe() # Like summary() in R
```

[27]:
```
        Customer ID          Age        Balance
count  4.014000e+03  4014.000000    4014.000000
mean   1.696831e+08    38.611111   39766.448274
std    8.865374e+07     9.819121   29859.489192
```

```
min    1.000000e+08    15.000000      11.520000
25%    1.000020e+08    31.000000   16115.367500
50%    1.000038e+08    37.000000   33567.330000
75%    2.000031e+08    45.000000   57533.930000
max    4.000038e+08    64.000000  183467.700000
```

[28]:
```
stats.describe().transpose()
```

[28]:
```
              count          mean           std           min           25%  \
Customer ID  4014.0  1.696831e+08  8.865374e+07  1.000000e+08  1.000020e+08
Age          4014.0  3.861111e+01  9.819121e+00  1.500000e+01  3.100000e+01
Balance      4014.0  3.976645e+04  2.985949e+04  1.152000e+01  1.611537e+04

                      50%           75%          max
Customer ID  1.000038e+08  2.000031e+08  400003848.0
Age          3.700000e+01  4.500000e+01         64.0
Balance      3.356733e+04  5.753393e+04     183467.7
```

11 Renaming columns of a DataFrame

[29]:
```
stats.columns
```

[29]:
```
Index(['Customer ID', 'Name', 'Surname', 'Gender', 'Age', 'Region',
       'Job Classification', 'Date Joined', 'Balance'],
      dtype='object')
```

[30]:
```
stats.columns=['CustomerID', 'Name', 'Surname', 'Gender', 'Age', 'Region',
               'JobClassification', 'DateJoined', 'Balance']
```

[31]:
```
stats.head()
```

[31]:
```
   CustomerID     Name  Surname  Gender  Age            Region  \
0   100000001    Simon    Walsh    Male   21           England
1   400000002  Jasmine   Miller  Female   34  Northern Ireland
2   100000003     Liam    Brown    Male   46           England
3   300000004   Trevor     Parr    Male   32             Wales
4   100000005  Deirdre  Pullman  Female   38           England

  JobClassification DateJoined    Balance
0      White Collar  05.Jan.15  113810.15
1       Blue Collar  06.Jan.15   36919.73
2      White Collar  07.Jan.15  101536.83
3      White Collar  08.Jan.15    1421.52
4       Blue Collar  09.Jan.15   35639.79
```

12 Subsetting Data Frames in Pandas

13 Three Parts Buckle up:

14 - Rows

15 - Columns

16 - Combine the Both

[32]:
```
stats.head()
```

[32]:
```
   CustomerID     Name  Surname  Gender  Age            Region  \
0   100000001    Simon    Walsh    Male   21           England
1   400000002  Jasmine   Miller  Female   34  Northern Ireland
2   100000003     Liam    Brown    Male   46           England
3   300000004   Trevor     Parr    Male   32             Wales
4   100000005  Deirdre  Pullman  Female   38           England

  JobClassification DateJoined    Balance
0      White Collar  05.Jan.15  113810.15
1       Blue Collar  06.Jan.15   36919.73
2      White Collar  07.Jan.15  101536.83
3      White Collar  08.Jan.15    1421.52
4       Blue Collar  09.Jan.15   35639.79
```

17 Part 1. Rows

[33]:
```
stats[21:26]
```

[33]:
```
    CustomerID      Name   Surname  Gender  Age    Region JobClassification  \
21   200000022     Jason    Butler    Male   58  Scotland       Blue Collar
22   300000023   Deirdre  McDonald  Female   41     Wales      White Collar
23   200000024      Carl     Quinn    Male   52  Scotland       Blue Collar
24   100000025  Jennifer    Hughes  Female   38   England      White Collar
25   200000026   Richard    Fraser    Male   55  Scotland       Blue Collar

   DateJoined   Balance
21  18.Jan.15  21252.97
22  18.Jan.15  66785.78
23  19.Jan.15   6580.81
24  20.Jan.15  20505.32
25  21.Jan.15  43249.26
```

[34]:
```
stats[110:120]
```

[34]:
```
     CustomerID     Name    Surname  Gender  Age            Region  \
110   300000111    Sonia  Robertson  Female   25             Wales
111   300000112   Nathan   Paterson    Male   26             Wales
112   400000113      Tim   Hardacre    Male   29  Northern Ireland
113   400000114    Fiona      Mills  Female   18  Northern Ireland
114   400000115     Ruth     Oliver  Female   43  Northern Ireland
115   100000116   Alison   Johnston  Female   36           England
116   100000117      Amy    McGrath  Female   40           England
117   200000118     Adam    McGrath    Male   52          Scotland
118   100000119  Vanessa      Lyman  Female   18           England
119   100000120   Andrea    Dickens  Female   31           England

    JobClassification DateJoined    Balance
110             Other  16.Mar.15   70799.64
111      White Collar  16.Mar.15   20627.65
112      White Collar  16.Mar.15   82229.52
113             Other  16.Mar.15   51171.29
114       Blue Collar  16.Mar.15   37889.38
115             Other  19.Mar.15   21806.30
116             Other  25.Mar.15   12415.50
117             Other  28.Mar.15   67682.92
118      White Collar  31.Mar.15   33524.41
119      White Collar  31.Mar.15  136370.38
```

[35]:
```
stats[4010:]
```

[35]:
```
      CustomerID       Name   Surname  Gender  Age    Region  \
4010   200004011      Keith    Hughes    Male   52  Scotland
4011   200004012     Hannah  Springer  Female   50  Scotland
4012   200004013  Christian      Reid    Male   51  Scotland
4013   300004014    Stephen       May    Male   33     Wales

     JobClassification DateJoined   Balance
4010       Blue Collar  30.Dec.15  56069.72
4011             Other  30.Dec.15  59477.82
4012       Blue Collar  30.Dec.15    239.45
4013       Blue Collar  30.Dec.15  30293.19
```

[36]:
```
stats[:6] # Same as head()
```

[36]:
```
   CustomerID     Name  Surname  Gender  Age            Region  \
0   100000001    Simon    Walsh    Male   21           England
1   400000002  Jasmine   Miller  Female   34  Northern Ireland
2   100000003     Liam    Brown    Male   46           England
3   300000004   Trevor     Parr    Male   32             Wales
4   100000005  Deirdre  Pullman  Female   38           England
5   300000006      Ava  Coleman  Female   30             Wales
```

```
  JobClassification DateJoined    Balance
0      White Collar  05.Jan.15  113810.15
1       Blue Collar  06.Jan.15   36919.73
2      White Collar  07.Jan.15  101536.83
3      White Collar  08.Jan.15    1421.52
4       Blue Collar  09.Jan.15   35639.79
5       Blue Collar  09.Jan.15  122443.77
```

18 Quick Exercise(resfersher)

19 1. Reverse The DataFrame

[37]:
```
stats[::-1]# Or stats[199:100:-1]
```

[37]:
```
      CustomerID       Name   Surname  Gender  Age            Region  \
4013   300004014    Stephen       May    Male   33             Wales
4012   200004013  Christian      Reid    Male   51          Scotland
4011   200004012     Hannah  Springer  Female   50          Scotland
4010   200004011      Keith    Hughes    Male   52          Scotland
4009   200004010        Sam     Lewis    Male   64          Scotland
…            …          …       …      …   …                …
4      100000005    Deirdre   Pullman  Female   38           England
3      300000004     Trevor      Parr    Male   32             Wales
2      100000003       Liam     Brown    Male   46           England
1      400000002    Jasmine    Miller  Female   34  Northern Ireland
0      100000001      Simon     Walsh    Male   21           England

     JobClassification DateJoined    Balance
4013       Blue Collar  30.Dec.15   30293.19
4012       Blue Collar  30.Dec.15     239.45
4011             Other  30.Dec.15   59477.82
4010       Blue Collar  30.Dec.15   56069.72
4009             Other  30.Dec.15   19711.66
…                  …         …          …
4          Blue Collar  09.Jan.15   35639.79
3         White Collar  08.Jan.15    1421.52
2         White Collar  07.Jan.15  101536.83
1          Blue Collar  06.Jan.15   36919.73
0         White Collar  05.Jan.15  113810.15

[4014 rows x 9 columns]
```

20 2. Get only every 20th Rows

[38]:
```
stats[::20]
```

[38]:
```
      CustomerID       Name     Surname  Gender  Age    Region  \
0      100000001      Simon       Walsh    Male   21   England
20     300000021      Boris    Johnston    Male   37     Wales
40     100000041     Edward       Terry    Male   27   England
60     100000061      Kylie      Howard  Female   35   England
80     100000081       Joan    Buckland  Female   36   England
…            …          …           …       …  …         …
3920   100003921      Isaac    Buckland    Male   37   England
3940   200003941     Joshua  Sutherland    Male   59  Scotland
3960   100003961    Leonard       Grant    Male   48   England
3980   200003981  Elizabeth       James  Female   43  Scotland
4000   300004001  Gabrielle      Duncan  Female   34     Wales

     JobClassification DateJoined    Balance
0         White Collar  05.Jan.15  113810.15
20               Other  16.Jan.15   31778.90
40         Blue Collar  01.Feb.15   51412.60
60        White Collar  12.Feb.15    4586.23
80        White Collar  16.Mar.15   59935.75
…                  …          …          …
3920      White Collar  24.Dec.15   35743.13
3940             Other  25.Dec.15    2114.65
3960             Other  27.Dec.15   72061.71
3980             Other  28.Dec.15   19695.66
4000      White Collar  29.Dec.15   92083.79

[201 rows x 9 columns]
```

21 Part 2. Columns

[39]:
```
stats.columns
```

[39]:
```
Index(['CustomerID', 'Name', 'Surname', 'Gender', 'Age', 'Region',
       'JobClassification', 'DateJoined', 'Balance'],
      dtype='object')
```

[40]:
```
stats.head()
```

[40]:
```
   CustomerID     Name  Surname  Gender  Age            Region  \
0   100000001    Simon    Walsh    Male   21           England
1   400000002  Jasmine   Miller  Female   34  Northern Ireland
2   100000003     Liam    Brown    Male   46           England
3   300000004   Trevor     Parr    Male   32             Wales
```

```
4   100000005  Deirdre  Pullman  Female   38          England

  JobClassification DateJoined    Balance
0      White Collar  05.Jan.15  113810.15
1       Blue Collar  06.Jan.15   36919.73
2      White Collar  07.Jan.15  101536.83
3      White Collar  08.Jan.15    1421.52
4       Blue Collar  09.Jan.15   35639.79
```

[41]:
```
stats['Name']
```

[41]:
```
0           Simon
1         Jasmine
2            Liam
3          Trevor
4         Deirdre
          …
4009          Sam
4010        Keith
4011       Hannah
4012    Christian
4013      Stephen
Name: Name, Length: 4014, dtype: object
```

[42]:
```
stats['Name'].head()
```

[42]:
```
0      Simon
1    Jasmine
2       Liam
3     Trevor
4    Deirdre
Name: Name, dtype: object
```

[43]:
```
stats[['Name','Surname']].head() # In R you would be passing a vector:
↪c('Name','Surname')
```

[43]:
```
       Name  Surname
0    Simon    Walsh
1  Jasmine   Miller
2     Liam    Brown
3   Trevor     Parr
4  Deirdre  Pullman
```

22 Quick Access requires the name to be one word(or Column)

[44]:
```
stats.Name.head() # or stats.Name
```

[44]:
```
0      Simon
1    Jasmine
2       Liam
3     Trevor
4    Deirdre
Name: Name, dtype: object
```

23 Part 3. Combining Both

[45]:
```
stats[4:8][['Name','Surname']]
```

[45]:

	Name	Surname
4	Deirdre	Pullman
5	Ava	Coleman
6	Dorothy	Thomson
7	Lisa	Knox

[46]:
```
stats[['Name','Surname']][4:8]  # df2=stats[['Name','Surname']]
                                # df2[4:8]
```

[46]:

	Name	Surname
4	Deirdre	Pullman
5	Ava	Coleman
6	Dorothy	Thomson
7	Lisa	Knox

[47]:
```
df2=stats[['Name','Surname']]
df2[4:8]
```

[47]:

	Name	Surname
4	Deirdre	Pullman
5	Ava	Coleman
6	Dorothy	Thomson
7	Lisa	Knox

24 Basic Operations with DataFrames

25 Mathematical Operation:

[48]:
```
stats.head()
```

[48]:
```
   CustomerID     Name  Surname  Gender  Age            Region  \
0   100000001    Simon    Walsh    Male   21           England
1   400000002  Jasmine   Miller  Female   34  Northern Ireland
2   100000003     Liam    Brown    Male   46           England
3   300000004   Trevor     Parr    Male   32             Wales
4   100000005  Deirdre  Pullman  Female   38           England

  JobClassification DateJoined    Balance
0      White Collar  05.Jan.15  113810.15
1       Blue Collar  06.Jan.15   36919.73
2      White Collar  07.Jan.15  101536.83
3      White Collar  08.Jan.15    1421.52
4       Blue Collar  09.Jan.15   35639.79
```

[49]:
```
Result=stats['Balance*2']=stats.Balance*2
Result.head()
```

[49]:
```
0    227620.30
1     73839.46
2    203073.66
3      2843.04
4     71279.58
Name: Balance, dtype: float64
```

26 Add Column:

[50]:
```
stats['Balance*2']=stats.Balance*2
```

[51]:
```
stats.head()
```

[51]:
```
   CustomerID     Name  Surname  Gender  Age            Region  \
0   100000001    Simon    Walsh    Male   21           England
1   400000002  Jasmine   Miller  Female   34  Northern Ireland
2   100000003     Liam    Brown    Male   46           England
3   300000004   Trevor     Parr    Male   32             Wales
4   100000005  Deirdre  Pullman  Female   38           England

  JobClassification DateJoined    Balance  Balance*2
0      White Collar  05.Jan.15  113810.15  227620.30
1       Blue Collar  06.Jan.15   36919.73   73839.46
2      White Collar  07.Jan.15  101536.83  203073.66
3      White Collar  08.Jan.15    1421.52    2843.04
4       Blue Collar  09.Jan.15   35639.79   71279.58
```

[52]:
```
# comparison to R
stats.['xyz']=[1,2,3,4,5]# Error No Recycling option
```

```
  File "<ipython-input-52-70bd73c3ba16>", line 2
    stats.['xyz']=[1,2,3,4,5]# Error No Recycling option
          ^
SyntaxError: invalid syntax
```

27 Removing a column

[53]:
```
stats.head()
```

[53]:
```
   CustomerID     Name  Surname  Gender  Age            Region  \
0   100000001    Simon    Walsh    Male   21           England
1   400000002  Jasmine   Miller  Female   34  Northern Ireland
2   100000003     Liam    Brown    Male   46           England
3   300000004   Trevor     Parr    Male   32             Wales
4   100000005  Deirdre  Pullman  Female   38           England

  JobClassification DateJoined    Balance  Balance*2
0      White Collar  05.Jan.15  113810.15  227620.30
1       Blue Collar  06.Jan.15   36919.73   73839.46
2      White Collar  07.Jan.15  101536.83  203073.66
3      White Collar  08.Jan.15    1421.52    2843.04
4       Blue Collar  09.Jan.15   35639.79   71279.58
```

[54]:
```
stats.drop('Balance*2',1).head() # 1 is vertical and 0 is Horiz.
```

[54]:
```
   CustomerID     Name  Surname  Gender  Age            Region  \
0   100000001    Simon    Walsh    Male   21           England
1   400000002  Jasmine   Miller  Female   34  Northern Ireland
2   100000003     Liam    Brown    Male   46           England
3   300000004   Trevor     Parr    Male   32             Wales
4   100000005  Deirdre  Pullman  Female   38           England

  JobClassification DateJoined    Balance
0      White Collar  05.Jan.15  113810.15
1       Blue Collar  06.Jan.15   36919.73
2      White Collar  07.Jan.15  101536.83
3      White Collar  08.Jan.15    1421.52
4       Blue Collar  09.Jan.15   35639.79
```

[55]:
```
stats.head()
```

[55]:
```
   CustomerID     Name  Surname  Gender  Age            Region  \
0   100000001    Simon    Walsh    Male   21           England
1   400000002  Jasmine   Miller  Female   34  Northern Ireland
```

```
2   100000003     Liam    Brown    Male   46           England
3   300000004   Trevor     Parr    Male   32             Wales
4   100000005  Deirdre  Pullman  Female   38           England

  JobClassification DateJoined    Balance  Balance*2
0      White Collar  05.Jan.15  113810.15  227620.30
1       Blue Collar  06.Jan.15   36919.73   73839.46
2      White Collar  07.Jan.15  101536.83  203073.66
3      White Collar  08.Jan.15    1421.52    2843.04
4       Blue Collar  09.Jan.15   35639.79   71279.58
```

[56]:
```
stats=stats.drop('Balance*2',1) # Drop Permanently
```

[57]:
```
stats.head()
```

[57]:
```
    CustomerID     Name  Surname  Gender  Age            Region  \
0   100000001    Simon    Walsh    Male   21           England
1   400000002  Jasmine   Miller  Female   34  Northern Ireland
2   100000003     Liam    Brown    Male   46           England
3   300000004   Trevor     Parr    Male   32             Wales
4   100000005  Deirdre  Pullman  Female   38           England

  JobClassification DateJoined    Balance
0      White Collar  05.Jan.15  113810.15
1       Blue Collar  06.Jan.15   36919.73
2      White Collar  07.Jan.15  101536.83
3      White Collar  08.Jan.15    1421.52
4       Blue Collar  09.Jan.15   35639.79
```

28 Filtering DataFrames

29 Filtering is about Rows

[58]:
```
stats.Age>70
```

[58]:
```
0       False
1       False
2       False
3       False
4       False
        …
4009    False
4010    False
4011    False
4012    False
4013    False
```

```
Name: Age, Length: 4014, dtype: bool
```

[59]:
```
Filter=stats.Age>63
```

[60]:
```
Filter
```

[60]:
```
0       False
1       False
2       False
3       False
4       False
        …
4009     True
4010    False
4011    False
4012    False
4013    False
Name: Age, Length: 4014, dtype: bool
```

[61]:
```
stats[Filter] # Conceptually this is just like R
```

[61]:

	CustomerID	Name	Surname	Gender	Age	Region	JobClassification	DateJoined	Balance
631	200000632	Nicholas	Allan	Male	64	Scotland	Blue Collar	22.May.15	15909.96
841	200000842	Matt	Manning	Male	64	Scotland	Blue Collar	10.Jun.15	3775.44
1498	200001499	Cameron	Ellison	Male	64	Scotland	Blue Collar	03.Aug.15	71106.33
1586	200001587	Jake	Ellison	Male	64	Scotland	Other	11.Aug.15	14242.57
1608	200001609	Yvonne	Dickens	Female	64	Scotland	Other	13.Aug.15	23522.36
1639	200001640	Christopher	Underwood	Male	64	Scotland	Blue Collar	16.Aug.15	47115.07
2040	200002041	Abigail	Fraser	Female	64	Scotland	Blue Collar	13.Sep.15	24729.53
2055	200002056	Joshua	Carr	Male	64	Scotland	Other	14.Sep.15	14558.13
2310	200002311	Kevin	Howard	Male	64	Scotland			
2352	200002353	Leonard	Lyman	Male	64	Scotland			
2752	200002753	Ian	Hunter	Male	64	Scotland			
2772	200002773	Alan	Watson	Male	64	Scotland			
3676	200003677	Anthony	Lewis	Male	64	Scotland			
4008	200004009	Alison	Quinn	Female	64	Scotland			
4009	200004010	Sam	Lewis	Male	64	Scotland			

```
2310       Blue Collar  26.Sep.15   10325.52
2352       Blue Collar  28.Sep.15  139415.88
2752       Blue Collar  23.Oct.15   92921.99
2772       Blue Collar  24.Oct.15   24994.57
3676       Blue Collar  12.Dec.15   48456.48
4008             Other  30.Dec.15   73503.90
4009             Other  30.Dec.15   19711.66
```

30 Let‘s use in Practice now

[62]:
```
Filter2=stats.Balance<5000
```

[63]:
```
Filter2
```

[63]:
```
0       False
1       False
2       False
3        True
4       False
        …
4009    False
4010    False
4011    False
4012     True
4013    False
Name: Balance, Length: 4014, dtype: bool
```

[64]:
```
stats[Filter2]
```

[64]:
```
      CustomerID       Name   Surname  Gender  Age            Region  \
3      300000004     Trevor      Parr    Male   32             Wales
14     300000015  Madeleine  Marshall  Female   36             Wales
15     100000016   Nicholas    Newman    Male   42           England
17     200000018   Samantha   Coleman  Female   42          Scotland
26     400000027     Rachel   McGrath  Female   37  Northern Ireland
…          …        …       …     …   …                 …
3947   100003948       Owen     Baker    Male   18           England
3955   300003956       Jane    Duncan  Female   34             Wales
3987   100003988    Theresa   Forsyth  Female   30           England
4006   100004007     Rachel    Davies  Female   34           England
4012   200004013  Christian      Reid    Male   51          Scotland

     JobClassification DateJoined  Balance
3         White Collar  08.Jan.15  1421.52
14               Other  12.Jan.15  2846.03
15        White Collar  14.Jan.15  2116.85
17               Other  14.Jan.15  3801.69
```

```
26        White Collar  23.Jan.15  3967.20
…                …        …       …
3947       Blue Collar  26.Dec.15  3858.90
3955       Blue Collar  26.Dec.15  4478.46
3987      White Collar  29.Dec.15  4570.98
4006       Blue Collar  30.Dec.15  4561.22
4012       Blue Collar  30.Dec.15   239.45

[324 rows x 9 columns]
```

[65]:
```
stats[stats.Balance<100]
```

[65]:
```
      CustomerID     Name   Surname  Gender  Age            Region  \
74     400000075   Olivia      Dowd  Female   30  Northern Ireland
774    200000775   Warren   Roberts    Male   37          Scotland
1319   100001320     Jane      King  Female   38           England
2045   400002046    Megan      Hart  Female   19  Northern Ireland
3467   300003468  Stewart  Johnston    Male   41             Wales
3496   300003497  Rebecca    Howard  Female   30             Wales
3884   100003885  Abigail     Mills  Female   29           England

     JobClassification DateJoined  Balance
74         White Collar  12.Feb.15    21.03
774         Blue Collar  01.Jun.15    69.01
1319       White Collar  22.Jul.15    11.52
2045        Blue Collar  13.Sep.15    69.78
3467        Blue Collar  30.Nov.15    77.46
3496        Blue Collar  02.Dec.15    96.26
3884       White Collar  23.Dec.15    98.68
```

31 More then one filters

[66]:
```
stats[Filter & Filter2]
```

[66]:
```
     CustomerID  Name  Surname Gender  Age    Region JobClassification  \
841   200000842  Matt  Manning   Male   64  Scotland       Blue Collar

    DateJoined  Balance
841  10.Jun.15  3775.44
```

[67]:
```
stats[(stats.Age>63) & (stats.Balance<5000)] # Same as stats[Filter & Filter2]
```

[67]:
```
     CustomerID  Name  Surname Gender  Age    Region JobClassification  \
841   200000842  Matt  Manning   Male   64  Scotland       Blue Collar

    DateJoined  Balance
841  10.Jun.15  3775.44
```

32 AnotherOne:

[68]:
```
stats[stats.Region=='England']
```

[68]:
```
      CustomerID      Name    Surname  Gender  Age   Region JobClassification  \
0      100000001     Simon      Walsh    Male   21  England      White Collar
2      100000003      Liam      Brown    Male   46  England      White Collar
4      100000005   Deirdre    Pullman  Female   38  England       Blue Collar
6      100000007   Dorothy    Thomson  Female   34  England       Blue Collar
9      100000010   Dominic       Parr    Male   42  England      White Collar
…            …         …          …       …  …        …                 …
4003   100004004      Jane   Hemmings  Female   28  England       Blue Collar
4004   100004005      John   Hamilton    Male   45  England      White Collar
4005   100004006  Kimberly       Gray  Female   44  England             Other
4006   100004007    Rachel     Davies  Female   34  England       Blue Collar
4007   100004008       Sam  Sanderson    Male   28  England       Blue Collar

     DateJoined    Balance
0      05.Jan.15  113810.15
2      07.Jan.15  101536.83
4      09.Jan.15   35639.79
6      11.Jan.15   42879.84
9      12.Jan.15   10912.45
…          …          …
4003   30.Dec.15   68518.55
4004   30.Dec.15    8435.91
4005   30.Dec.15   64470.77
4006   30.Dec.15    4561.22
4007   30.Dec.15   42128.29

[2159 rows x 9 columns]
```

[69]:
```
# How to get unique()
stats.Region.unique()
```

[69]:
```
array(['England', 'Northern Ireland', 'Wales', 'Scotland'], dtype=object)
```

33 Quick Exercise:

34 Find out everything about Matt Manning

[70]:
```
stats[(stats.Name=='Matt')&(stats.Surname=='Manning')]
```

[70]:
```
     CustomerID  Name  Surname Gender  Age    Region JobClassification  \
841   200000842  Matt  Manning   Male   64  Scotland       Blue Collar

    DateJoined  Balance
```

```
841   10.Jun.15   3775.44
```

35 Accessing Individual Elements

36 1) .at for lables Important: even integers are treated as labels

37 2) .iat for interger location

[71]:
```
stats.head()
```

[71]:
```
   CustomerID     Name  Surname  Gender  Age            Region  \
0   100000001    Simon    Walsh    Male   21           England
1   400000002  Jasmine   Miller  Female   34  Northern Ireland
2   100000003     Liam    Brown    Male   46           England
3   300000004   Trevor     Parr    Male   32             Wales
4   100000005  Deirdre  Pullman  Female   38           England

  JobClassification DateJoined    Balance
0      White Collar  05.Jan.15  113810.15
1       Blue Collar  06.Jan.15   36919.73
2      White Collar  07.Jan.15  101536.83
3      White Collar  08.Jan.15    1421.52
4       Blue Collar  09.Jan.15   35639.79
```

[72]:
```
stats.iat[2,1]
```

[72]:
```
'Liam'
```

[73]:
```
stats.at[2,'Name']
```

[73]:
```
'Liam'
```

38 Why we need?

[74]:
```
sub10=stats[::1000]
```

[75]:
```
sub10
```

[75]:
```
      CustomerID        Name  Surname  Gender  Age            Region  \
0      100000001       Simon    Walsh    Male   21           England
1000   400001001       Grace   Duncan  Female   31  Northern Ireland
2000   100002001  Bernadette      Ince  Female   43           England
3000   100003001        Matt  Abraham    Male   47           England
4000   300004001   Gabrielle   Duncan  Female   34             Wales

     JobClassification DateJoined    Balance
```

```
0          White Collar  05.Jan.15  113810.15
1000              Other  26.Jun.15   32162.34
2000       White Collar  11.Sep.15   57739.46
3000        Blue Collar  03.Nov.15    8576.46
4000       White Collar  29.Dec.15   92083.79
```

[76]:
```
sub10.iat[1,0] # It's counting 0 axis
```

[76]:
```
400001001
```

[77]:
```
sub10.at[1000,'CustomerID'] # It's sees index
```

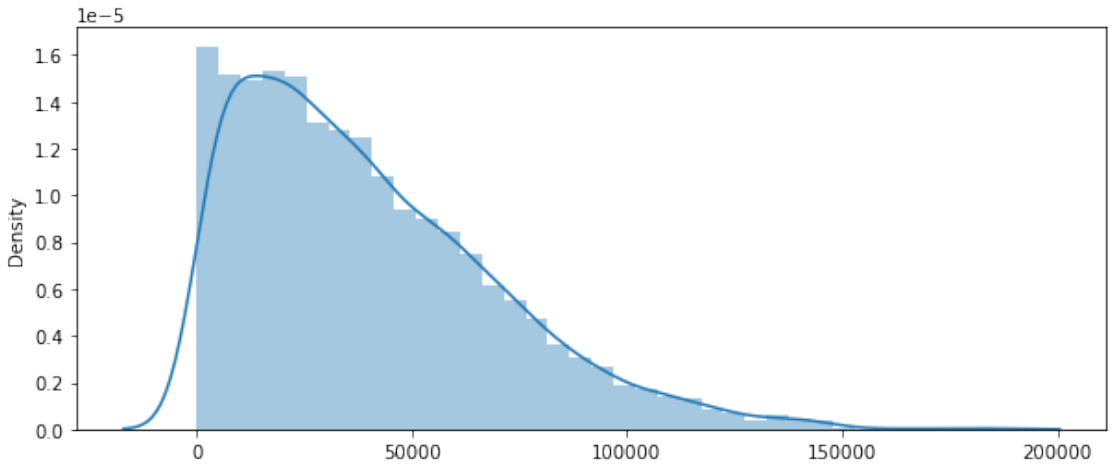
[77]:
```
400001001
```

39 Introduction to Seaborn

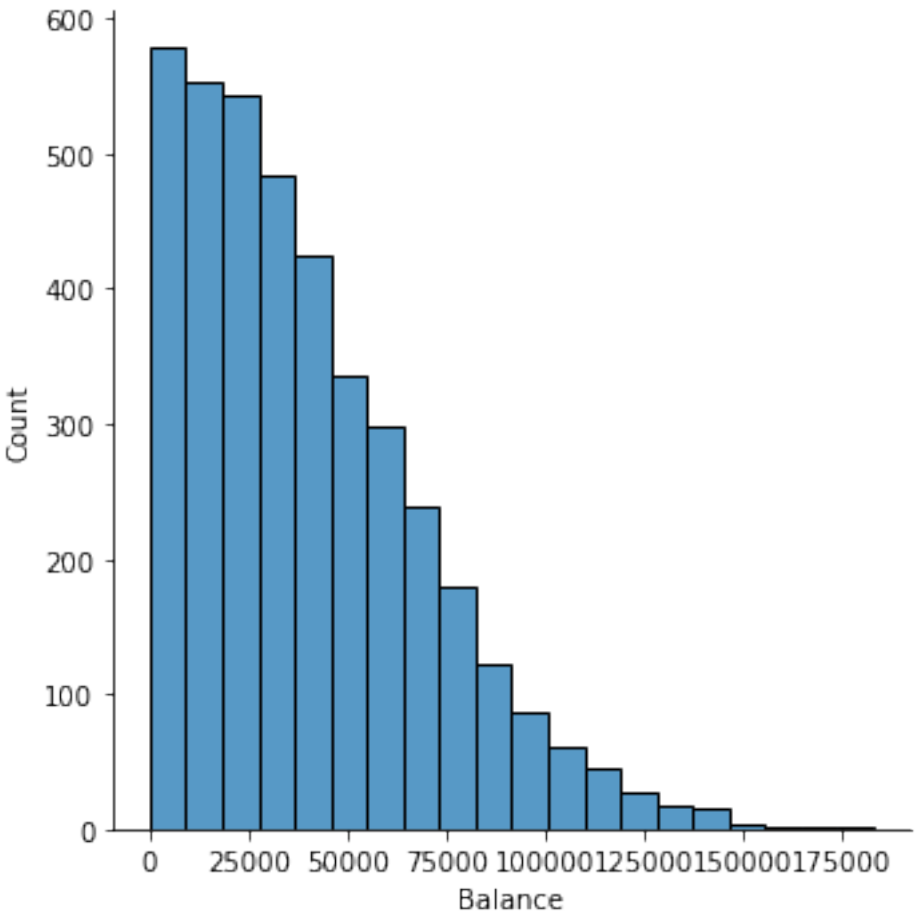
[78]:
```
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')

%matplotlib inline
plt.rcParams['figure.figsize']=10,4
```

40 Distribution:

[79]:
```
Vis1=sns.distplot([stats.Balance])
```



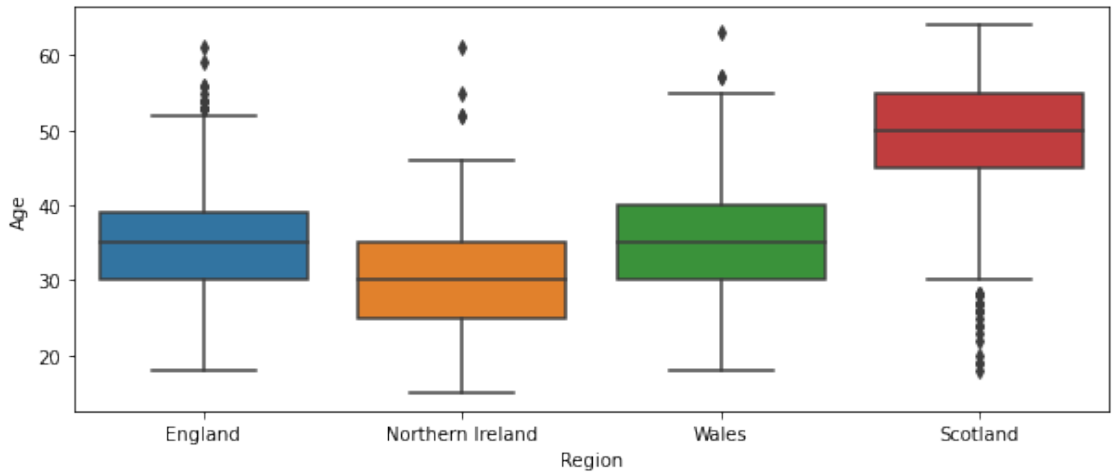
[80]:
```
Vis1=sns.displot(stats["Balance"],bins=20)
```



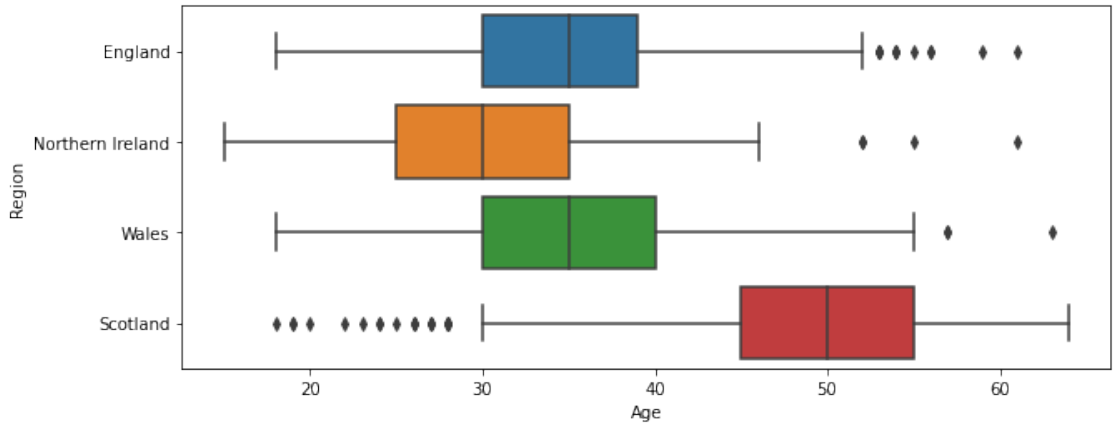
41 BoxPlots:

Google: Seaborn gallery

[81]:
```
vis2=sns.boxplot(data=stats,x="Region",y="Age")
```



[82]:
```
vis3=sns.boxplot(data=stats,x="Age",y="Region")
```



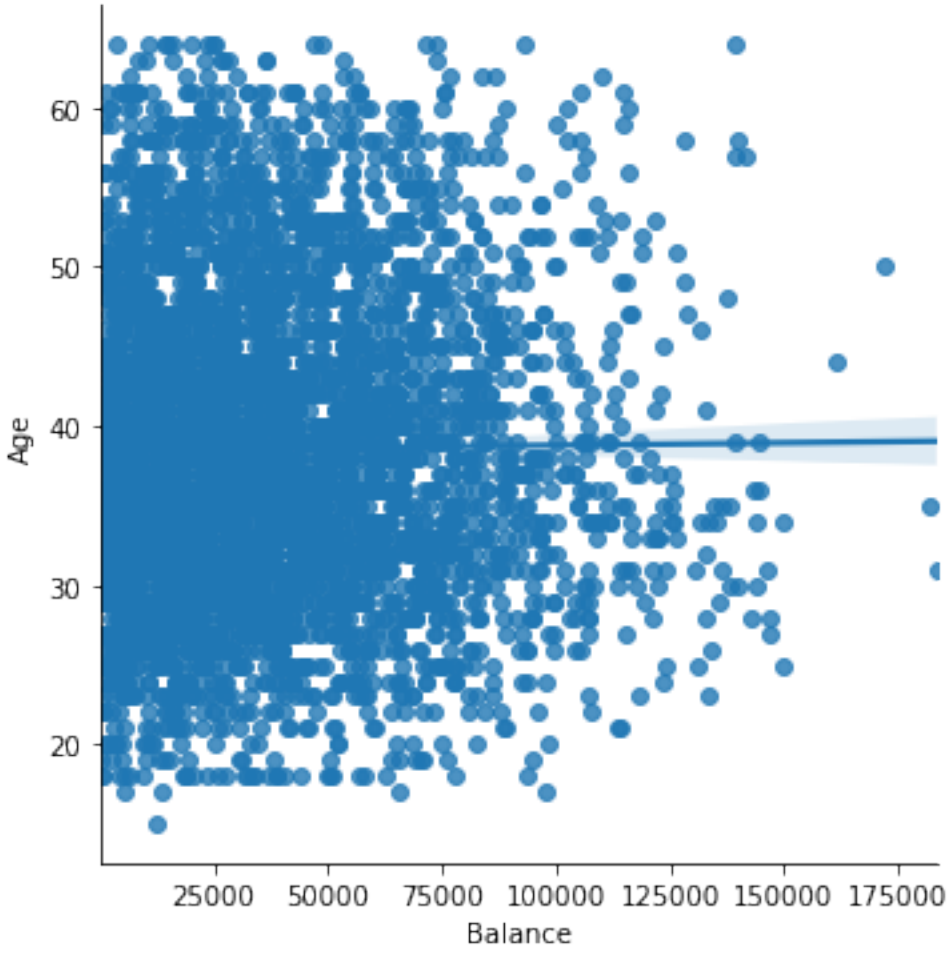
[83]:
```
stats.head()
```

[83]:

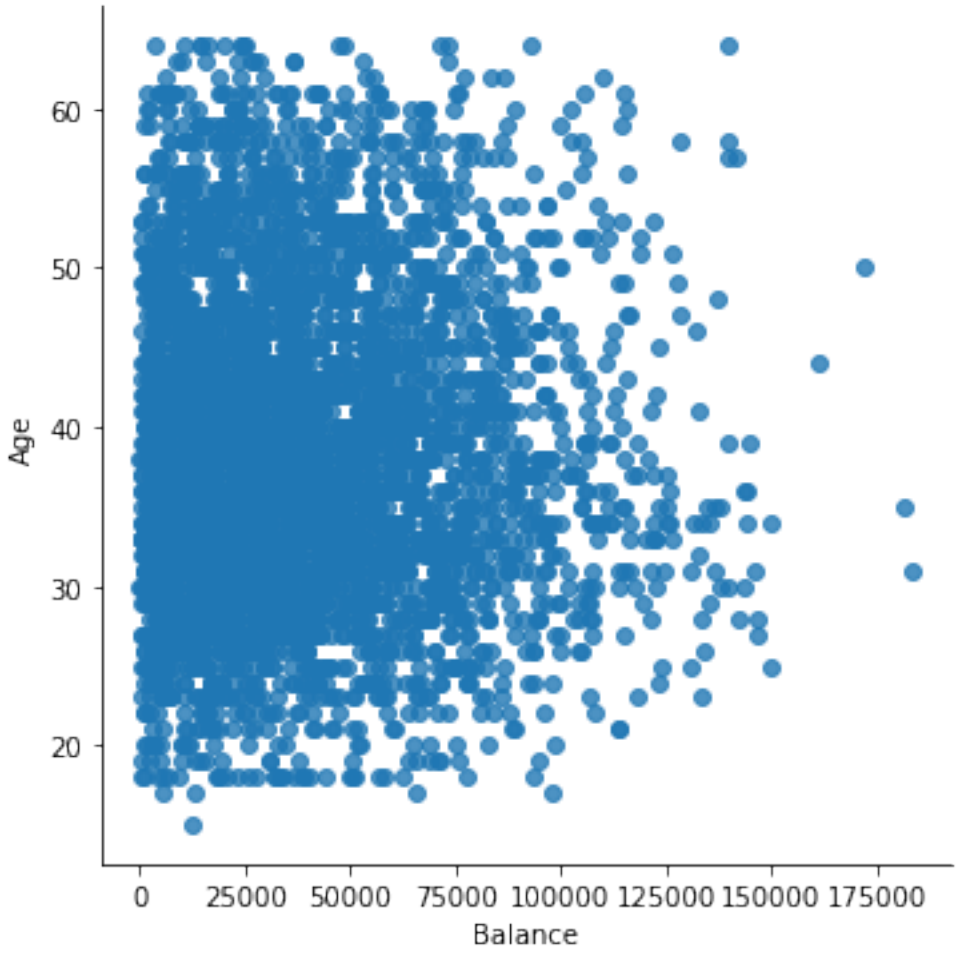
	CustomerID	Name	Surname	Gender	Age	Region	JobClassification	DateJoined	Balance
0	100000001	Simon	Walsh	Male	21	England	White Collar	05.Jan.15	113810.15
1	400000002	Jasmine	Miller	Female	34	Northern Ireland	Blue Collar	06.Jan.15	36919.73
2	100000003	Liam	Brown	Male	46	England	White Collar	07.Jan.15	101536.83
3	300000004	Trevor	Parr	Male	32	Wales			
4	100000005	Deirdre	Pullman	Female	38	England			

```
3      White Collar  08.Jan.15    1421.52
4       Blue Collar  09.Jan.15   35639.79
```

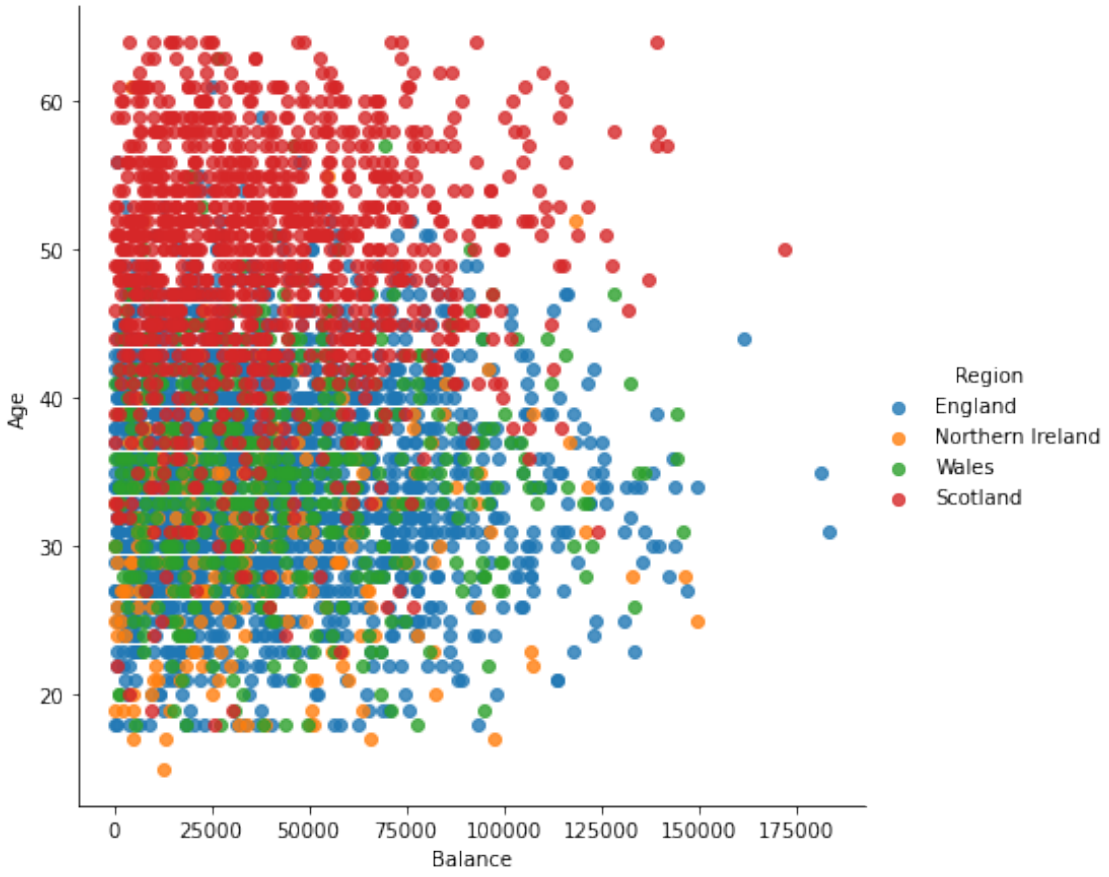
[84]:
```
vis4=sns.lmplot(data=stats,x='Balance',y='Age')
# Or vis4=sns.lmplot,x='Balance',y='Age',data=stats)
```



[85]:
```
vis4=sns.lmplot(data=stats,x='Balance',y='Age',fit_reg=False)
```



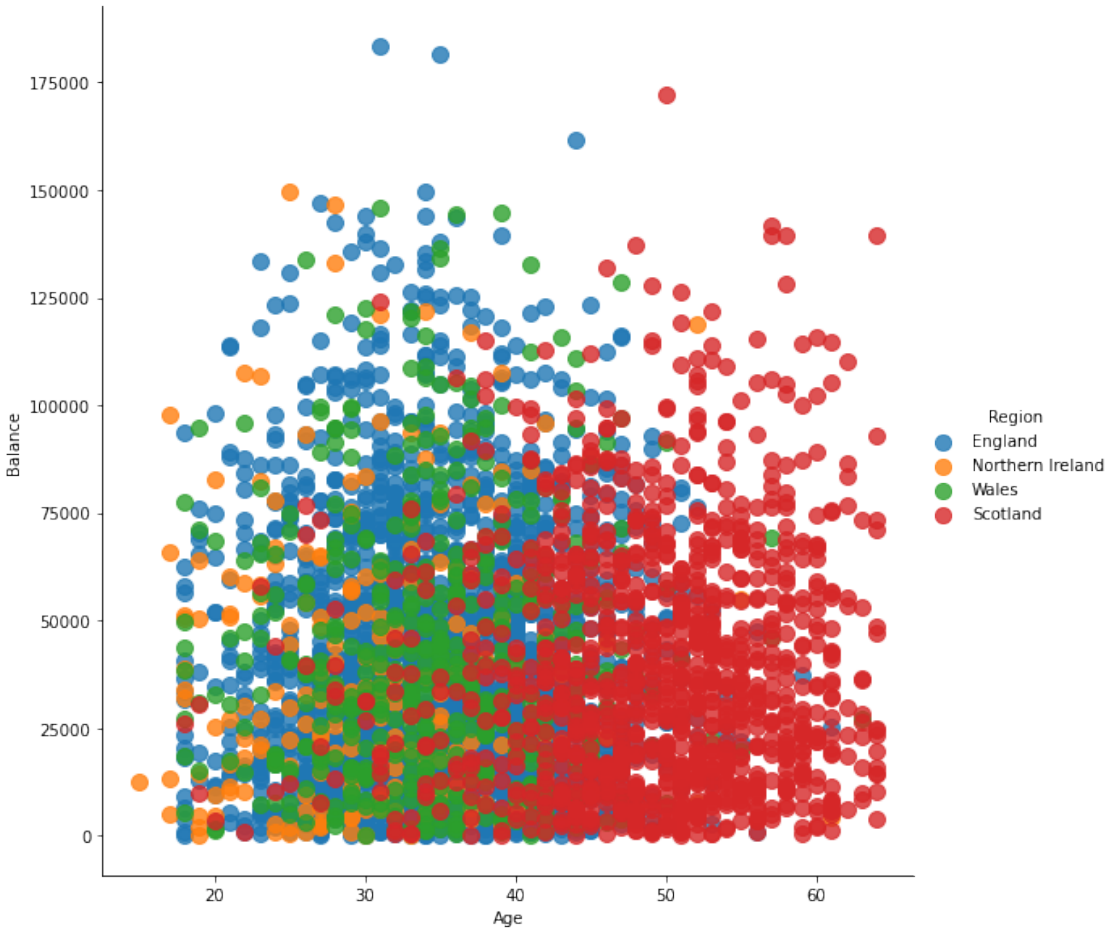
[86]:
```
vis4=sns.
 ↪lmplot(data=stats,x='Balance',y='Age',fit_reg=False,hue='Region',size=6)
#hue=color
```



42 Marker Size

[87]:
```
vis4=sns.lmplot(data=stats,x='Balance',y='Age',fit_reg=False,hue='Region',\
                size=6,scatter_kws={'s':100})
```

[88]:
```
vis4=sns.lmplot(data=stats,x='Age',y='Balance',fit_reg=False,hue='Region',\
                size=8,scatter_kws={'s':100})
```


175000
150000
125000
100000
75000
50000
25000
0
Balance
20
30
40
50
60
Age
Region
England
Northern Ireland
Wales
Scotland