In [1]:
```
import pandas as pd
```

In [2]:
```
csvdata = pd.read_csv('D:/Study/3-TYBT/DSL/Assignment-1/Data.csv')
```

In [11]:
```
csvdata.head(10)
```

Out[11]:

| | Country | Gender | Mean | N= |
|---|---|---|---|---|
| **0** | AT | Male | 7.3 | 471 |
| **1** | NaN | Female | 7.3 | 570 |
| **2** | NaN | Both | 7.3 | 1041 |
| **3** | BE | Male | 7.8 | 468 |
| **4** | NaN | Female | 7.8 | 542 |
| **5** | NaN | Both | 7.8 | 1010 |
| **6** | BG | Male | 5.8 | 416 |
| **7** | NaN | Female | 5.8 | 555 |
| **8** | NaN | Both | 5.8 | 971 |
| **9** | CY | Male | 7.8 | 433 |

In [4]:
```
from pandas import ExcelWriter
from pandas import ExcelFile
```

In [8]:
```

```

In [9]:
```
exceldata = pd.read_excel('D:/Study/3-TYBT/DSL/Assignment-1/Data1.xls')
```

In [12]:
```
exceldata.head(10)
```

Out[12]:

| | OrderDate | Region | Rep | Item | Units | Unit Cost | Total |
|---|---|---|---|---|---|---|---|
| **0** | 2014-09-01 | Central | Smith | Desk | 2 | 125.00 | 250.00 |
| **1** | 2015-06-17 | Central | Kivell | Desk | 5 | 125.00 | 625.00 |
| **2** | 2015-09-10 | Central | Gill | Pencil | 7 | 1.29 | 9.03 |
| **3** | 2015-11-17 | Central | Jardine | Binder | 11 | 4.99 | 54.89 |
| **4** | 2015-10-31 | Central | Andrews | Pencil | 14 | 1.29 | 18.06 |
| **5** | 2014-02-26 | Central | Gill | Pen | 27 | 19.99 | 539.73 |
| **6** | 2014-10-05 | Central | Morgan | Binder | 28 | 8.99 | 251.72 |
| **7** | 2015-12-21 | Central | Andrews | Binder | 28 | 4.99 | 139.72 |
| **8** | 2014-02-09 | Central | Jardine | Pencil | 36 | 4.99 | 179.64 |
| **9** | 2015-08-07 | Central | Kivell | Pen Set | 42 | 23.95 | 1005.90 |

In [15]:
```
exceldata = pd.read_excel("D:/Study/3-TYBT/DSL/Assignment-1/Data1.xls", index_col ="Rep")
```

In [19]:
```
first = exceldata[["Units", "Total"]]
```

In [20]:
```
print(first)
```

```
          Units    Total
Rep
```

```
Smith         2   250.00
Kivell        5   625.00
Gill          7     9.03
Jardine      11    54.89
Andrews      14    18.06
Gill         27   539.73
Morgan       28   251.72
Andrews      28   139.72
Jardine      36   179.64
Kivell       42  1005.90
Gill         46   413.54
Kivell       50   999.50
Jardine      50   249.50
Gill         53    68.37
Morgan       55   686.95
Andrews      66   131.34
Smith        67    86.43
Andrews      75   149.25
Gill         80   719.20
Smith        87  1305.00
Jardine      90   449.10
Morgan       90   449.10
Jardine      94  1879.06
Kivell       96   479.04
Jones         4    19.96
Parent       15   299.85
Jones        16   255.84
Howard       29    57.71
Jones        35   174.65
Jones        60   299.40
Jones        60   539.40
Jones        62   309.38
Jones        64   575.36
Parent       74  1183.26
Parent       81  1619.19
Jones        95   189.05
Howard       96   479.04
Sorvino       3   825.00
Sorvino       7   139.93
Thompson     32    63.68
Sorvino      56   167.44
Thompson     57  1139.43
Sorvino      76   151.24
```

In [21]:
```
exceldata.sort_values("Rep", axis = 0, ascending = True,
                      inplace = True, na_position ='last')
```

In [23]:
```
exceldata
```

Out[23]:

| Rep | OrderDate | Region | Item | Units | Unit Cost | Total |
|---|---|---|---|---|---|---|
| **Andrews** | 2014-04-18 | Central | Pencil | 75 | 1.99 | 149.25 |
| **Andrews** | 2015-10-31 | Central | Pencil | 14 | 1.29 | 18.06 |
| **Andrews** | 2015-04-10 | Central | Pencil | 66 | 1.99 | 131.34 |
| **Andrews** | 2015-12-21 | Central | Binder | 28 | 4.99 | 139.72 |
| **Gill** | 2015-05-14 | Central | Pencil | 53 | 1.29 | 68.37 |
| **Gill** | 2015-05-31 | Central | Binder | 80 | 8.99 | 719.20 |
| **Gill** | 2015-09-10 | Central | Pencil | 7 | 1.29 | 9.03 |
| **Gill** | 2014-02-26 | Central | Pen | 27 | 19.99 | 539.73 |
| **Gill** | 2015-01-15 | Central | Binder | 46 | 8.99 | 413.54 |
| **Howard** | 2015-04-27 | East | Pen | 96 | 4.99 | 479.04 |

| Rep | OrderDate | Region | Item | Units | Unit Cost | Total |
|---|---|---|---|---|---|---|
| **Howard** | 2014-07-12 | East | Binder | 29 | 1.99 | 57.71 |
| **Jardine** | 2015-12-04 | Central | Binder | 94 | 19.99 | 1879.06 |
| **Jardine** | 2014-05-05 | Central | Pencil | 90 | 4.99 | 449.10 |
| **Jardine** | 2015-11-17 | Central | Binder | 11 | 4.99 | 54.89 |
| **Jardine** | 2014-02-09 | Central | Pencil | 36 | 4.99 | 179.64 |
| **Jardine** | 2015-03-24 | Central | Pen Set | 50 | 4.99 | 249.50 |
| **Jones** | 2015-07-04 | East | Pen Set | 62 | 4.99 | 309.38 |
| **Jones** | 2014-10-22 | East | Pen | 64 | 8.99 | 575.36 |
| **Jones** | 2014-09-18 | East | Pen Set | 16 | 15.99 | 255.84 |
| **Jones** | 2014-01-06 | East | Pencil | 95 | 1.99 | 189.05 |
| **Jones** | 2014-04-01 | East | Binder | 60 | 4.99 | 299.40 |
| **Jones** | 2015-02-18 | East | Binder | 4 | 4.99 | 19.96 |
| **Jones** | 2014-08-15 | East | Pencil | 35 | 4.99 | 174.65 |
| **Jones** | 2014-06-08 | East | Binder | 60 | 8.99 | 539.40 |
| **Kivell** | 2015-08-07 | Central | Pen Set | 42 | 23.95 | 1005.90 |
| **Kivell** | 2014-01-23 | Central | Binder | 50 | 19.99 | 999.50 |
| **Kivell** | 2014-11-25 | Central | Pen Set | 96 | 4.99 | 479.04 |
| **Kivell** | 2015-06-17 | Central | Desk | 5 | 125.00 | 625.00 |
| **Morgan** | 2014-06-25 | Central | Pencil | 90 | 4.99 | 449.10 |
| **Morgan** | 2015-07-21 | Central | Pen Set | 55 | 12.49 | 686.95 |
| **Morgan** | 2014-10-05 | Central | Binder | 28 | 8.99 | 251.72 |
| **Parent** | 2014-07-29 | East | Binder | 81 | 19.99 | 1619.19 |
| **Parent** | 2014-11-08 | East | Pen | 15 | 19.99 | 299.85 |
| **Parent** | 2014-12-29 | East | Pen Set | 74 | 15.99 | 1183.26 |
| **Smith** | 2014-09-01 | Central | Desk | 2 | 125.00 | 250.00 |
| **Smith** | 2015-02-01 | Central | Binder | 87 | 15.00 | 1305.00 |
| **Smith** | 2014-12-12 | Central | Pencil | 67 | 1.29 | 86.43 |
| **Sorvino** | 2015-08-24 | West | Desk | 3 | 275.00 | 825.00 |
| **Sorvino** | 2015-03-07 | West | Binder | 7 | 19.99 | 139.93 |
| **Sorvino** | 2014-03-15 | West | Pencil | 56 | 2.99 | 167.44 |
| **Sorvino** | 2015-09-27 | West | Pen | 76 | 1.99 | 151.24 |
| **Thompson** | 2014-05-22 | West | Pencil | 32 | 1.99 | 63.68 |
| **Thompson** | 2015-10-14 | West | Binder | 57 | 19.99 | 1139.43 |

In [27]:
```
exceldata.dropna(inplace = True)
```

In [30]:
```
desc = exceldata['Region'].describe()
desc
```

Out[30]:
```
count          43
unique          3
top       Central
freq           24
Name: Region, dtype: object
```

In [33]:
```
desc = exceldata['Item'].describe()
desc
```

Out[33]:
```
count         43
unique         5
top       Binder
freq          15
Name: Item, dtype: object
```

In [34]:
```
exceldata.dtypes
```

Out[34]:
```
OrderDate    datetime64[ns]
Region               object
Item                 object
Units                 int64
Unit Cost           float64
Total               float64
dtype: object
```

In [35]:
```
exceldata.Total.dtypes
```

Out[35]:
```
dtype('float64')
```

In [36]:
```
unique_value = exceldata["Item"].nunique()
unique_value
```

Out[36]:
```
5
```

In [49]:
```
pd.options.display.float_format = '{:.1f}'.format
```

In [50]:
```
print(exceldata)
```

```
          OrderDate   Region     Item  Units  Unit Cost  Total
Rep
Andrews  2014-04-18  Central   Pencil     75        2.0  149.2
Andrews  2015-10-31  Central   Pencil     14        1.3   18.1
Andrews  2015-04-10  Central   Pencil     66        2.0  131.3
Andrews  2015-12-21  Central   Binder     28        5.0  139.7
Gill     2015-05-14  Central   Pencil     53        1.3   68.4
Gill     2015-05-31  Central   Binder     80        9.0  719.2
Gill     2015-09-10  Central   Pencil      7        1.3    9.0
Gill     2014-02-26  Central      Pen     27       20.0  539.7
Gill     2015-01-15  Central   Binder     46        9.0  413.5
Howard   2015-04-27     East      Pen     96        5.0  479.0
Howard   2014-07-12     East   Binder     29        2.0   57.7
Jardine  2015-12-04  Central   Binder     94       20.0 1879.1
Jardine  2014-05-05  Central   Pencil     90        5.0  449.1
Jardine  2015-11-17  Central   Binder     11        5.0   54.9
Jardine  2014-02-09  Central   Pencil     36        5.0  179.6
Jardine  2015-03-24  Central  Pen Set     50        5.0  249.5
Jones    2015-07-04     East  Pen Set     62        5.0  309.4
Jones    2014-10-22     East      Pen     64        9.0  575.4
Jones    2014-09-18     East  Pen Set     16       16.0  255.8
Jones    2014-01-06     East   Pencil     95        2.0  189.1
Jones    2014-04-01     East   Binder     60        5.0  299.4
Jones    2015-02-18     East   Binder      4        5.0   20.0
Jones    2014-08-15     East   Pencil     35        5.0  174.7
Jones    2014-06-08     East   Binder     60        9.0  539.4
Kivell   2015-08-07  Central  Pen Set     42       23.9 1005.9
Kivell   2014-01-23  Central   Binder     50       20.0  999.5
Kivell   2014-11-25  Central  Pen Set     96        5.0  479.0
Kivell   2015-06-17  Central     Desk      5      125.0  625.0
Morgan   2014-06-25  Central   Pencil     90        5.0  449.1
Morgan   2015-07-21  Central  Pen Set     55       12.5  687.0
```

```
Morgan   2014-10-05  Central   Binder     28        9.0  251.7
Parent   2014-07-29     East   Binder     81       20.0 1619.2
Parent   2014-11-08     East      Pen     15       20.0  299.8
Parent   2014-12-29     East  Pen Set     74       16.0 1183.3
Smith    2014-09-01  Central     Desk      2      125.0  250.0
Smith    2015-02-01  Central   Binder     87       15.0 1305.0
Smith    2014-12-12  Central   Pencil     67        1.3   86.4
Sorvino  2015-08-24     West     Desk      3      275.0  825.0
Sorvino  2015-03-07     West   Binder      7       20.0  139.9
Sorvino  2014-03-15     West   Pencil     56        3.0  167.4
Sorvino  2015-09-27     West      Pen     76        2.0  151.2
Thompson 2014-05-22     West   Pencil     32        2.0   63.7
Thompson 2015-10-14     West   Binder     57       20.0 1139.4
```

In [52]:
```
exceldata.dropna(inplace = True)
before = exceldata.dtypes
exceldata["Units"]= exceldata["Units"].astype(float)
exceldata["Total"]= exceldata["Total"].astype(int)
after = exceldata.dtypes
```

In [ ]:
```

```

In [54]:
```
before
```

Out[54]:
```
OrderDate    datetime64[ns]
Region               object
Item                 object
Units               float64
Unit Cost           float64
Total                 int32
dtype: object
```

In [56]:
```
exceldata
```

Out[56]:

| Rep | OrderDate | Region | Item | Units | Unit Cost | Total |
|---|---|---|---|---|---|---|
| **Andrews** | 2014-04-18 | Central | Pencil | 75.0 | 2.0 | 149 |
| **Andrews** | 2015-10-31 | Central | Pencil | 14.0 | 1.3 | 18 |
| **Andrews** | 2015-04-10 | Central | Pencil | 66.0 | 2.0 | 131 |
| **Andrews** | 2015-12-21 | Central | Binder | 28.0 | 5.0 | 139 |
| **Gill** | 2015-05-14 | Central | Pencil | 53.0 | 1.3 | 68 |
| **Gill** | 2015-05-31 | Central | Binder | 80.0 | 9.0 | 719 |
| **Gill** | 2015-09-10 | Central | Pencil | 7.0 | 1.3 | 9 |
| **Gill** | 2014-02-26 | Central | Pen | 27.0 | 20.0 | 539 |
| **Gill** | 2015-01-15 | Central | Binder | 46.0 | 9.0 | 413 |
| **Howard** | 2015-04-27 | East | Pen | 96.0 | 5.0 | 479 |
| **Howard** | 2014-07-12 | East | Binder | 29.0 | 2.0 | 57 |
| **Jardine** | 2015-12-04 | Central | Binder | 94.0 | 20.0 | 1879 |
| **Jardine** | 2014-05-05 | Central | Pencil | 90.0 | 5.0 | 449 |
| **Jardine** | 2015-11-17 | Central | Binder | 11.0 | 5.0 | 54 |
| **Jardine** | 2014-02-09 | Central | Pencil | 36.0 | 5.0 | 179 |
| **Jardine** | 2015-03-24 | Central | Pen Set | 50.0 | 5.0 | 249 |
| **Jones** | 2015-07-04 | East | Pen Set | 62.0 | 5.0 | 309 |
| **Jones** | 2014-10-22 | East | Pen | 64.0 | 9.0 | 575 |

| Rep | OrderDate | Region | Item | Units | Unit Cost | Total |
|---|---|---|---|---|---|---|
| **Jones** | 2014-09-18 | East | Pen Set | 16.0 | 16.0 | 255 |
| **Jones** | 2014-01-06 | East | Pencil | 95.0 | 2.0 | 189 |
| **Jones** | 2014-04-01 | East | Binder | 60.0 | 5.0 | 299 |
| **Jones** | 2015-02-18 | East | Binder | 4.0 | 5.0 | 19 |
| **Jones** | 2014-08-15 | East | Pencil | 35.0 | 5.0 | 174 |
| **Jones** | 2014-06-08 | East | Binder | 60.0 | 9.0 | 539 |
| **Kivell** | 2015-08-07 | Central | Pen Set | 42.0 | 23.9 | 1005 |
| **Kivell** | 2014-01-23 | Central | Binder | 50.0 | 20.0 | 999 |
| **Kivell** | 2014-11-25 | Central | Pen Set | 96.0 | 5.0 | 479 |
| **Kivell** | 2015-06-17 | Central | Desk | 5.0 | 125.0 | 625 |
| **Morgan** | 2014-06-25 | Central | Pencil | 90.0 | 5.0 | 449 |
| **Morgan** | 2015-07-21 | Central | Pen Set | 55.0 | 12.5 | 686 |
| **Morgan** | 2014-10-05 | Central | Binder | 28.0 | 9.0 | 251 |
| **Parent** | 2014-07-29 | East | Binder | 81.0 | 20.0 | 1619 |
| **Parent** | 2014-11-08 | East | Pen | 15.0 | 20.0 | 299 |
| **Parent** | 2014-12-29 | East | Pen Set | 74.0 | 16.0 | 1183 |
| **Smith** | 2014-09-01 | Central | Desk | 2.0 | 125.0 | 250 |
| **Smith** | 2015-02-01 | Central | Binder | 87.0 | 15.0 | 1305 |
| **Smith** | 2014-12-12 | Central | Pencil | 67.0 | 1.3 | 86 |
| **Sorvino** | 2015-08-24 | West | Desk | 3.0 | 275.0 | 825 |
| **Sorvino** | 2015-03-07 | West | Binder | 7.0 | 20.0 | 139 |
| **Sorvino** | 2014-03-15 | West | Pencil | 56.0 | 3.0 | 167 |
| **Sorvino** | 2015-09-27 | West | Pen | 76.0 | 2.0 | 151 |
| **Thompson** | 2014-05-22 | West | Pencil | 32.0 | 2.0 | 63 |
| **Thompson** | 2015-10-14 | West | Binder | 57.0 | 20.0 | 1139 |

In [57]:
```
print(exceldata.info())
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 43 entries, Andrews to Thompson
Data columns (total 6 columns):
 #   Column     Non-Null Count  Dtype         
---  ------     --------------  -----         
 0   OrderDate  43 non-null     datetime64[ns]
 1   Region     43 non-null     object        
 2   Item       43 non-null     object        
 3   Units      43 non-null     float64       
 4   Unit Cost  43 non-null     float64       
 5   Total      43 non-null     int32         
dtypes: datetime64[ns](1), float64(2), int32(1), object(2)
memory usage: 1.7+ KB
None
```

In [58]:
```
print(exceldata.describe())
```

```
       Units  Unit Cost  Total
count   43.0       43.0   43.0
```

```
mean    49.3       20.3  456.0
std     30.1       47.3  447.1
min      2.0        1.3    9.0
25%     27.5        4.0  144.0
50%     53.0        5.0  299.0
75%     74.5       18.0  600.0
max     96.0      275.0 1879.0
```

In [62]:
```
x=pd.read_excel('D:/Study/3-TYBT/DSL/Assignment-1/Data1.xls',usecols=['Units','Total'])
x["Units"].fillna(int(x["Units"].mean()),inplace=True)
```

In [63]:
```
x
```

Out[63]:

| | Units | Total |
|---|---|---|
| **0** | 2 | 250.0 |
| **1** | 5 | 625.0 |
| **2** | 7 | 9.0 |
| **3** | 11 | 54.9 |
| **4** | 14 | 18.1 |
| **5** | 27 | 539.7 |
| **6** | 28 | 251.7 |
| **7** | 28 | 139.7 |
| **8** | 36 | 179.6 |
| **9** | 42 | 1005.9 |
| **10** | 46 | 413.5 |
| **11** | 50 | 999.5 |
| **12** | 50 | 249.5 |
| **13** | 53 | 68.4 |
| **14** | 55 | 687.0 |
| **15** | 66 | 131.3 |
| **16** | 67 | 86.4 |
| **17** | 75 | 149.2 |
| **18** | 80 | 719.2 |
| **19** | 87 | 1305.0 |
| **20** | 90 | 449.1 |
| **21** | 90 | 449.1 |
| **22** | 94 | 1879.1 |
| **23** | 96 | 479.0 |
| **24** | 4 | 20.0 |
| **25** | 15 | 299.8 |
| **26** | 16 | 255.8 |
| **27** | 29 | 57.7 |
| **28** | 35 | 174.7 |
| **29** | 60 | 299.4 |
| **30** | 60 | 539.4 |

| | Units | Total |
|---|---|---|
| **31** | 62 | 309.4 |
| **32** | 64 | 575.4 |
| **33** | 74 | 1183.3 |
| **34** | 81 | 1619.2 |
| **35** | 95 | 189.1 |
| **36** | 96 | 479.0 |
| **37** | 3 | 825.0 |
| **38** | 7 | 139.9 |
| **39** | 32 | 63.7 |
| **40** | 56 | 167.4 |
| **41** | 57 | 1139.4 |
| **42** | 76 | 151.2 |

In [ ]: