Homework #0203

In [1]:
```
import pandas as pd
import numpy as np
import os
```

Read in the data.

In [2]:
```
# Go to the directory where the data file is located.
# os.chdir(r'~~')                # Please, replace the path with your own.
```

In [3]:
```
df = pd.read_csv('data_sales.csv', header='infer')
```

In [4]:
```
df.shape
```

Out[4]:
```
(43, 5)
```

In [5]:
```
df.head(5)
```

Out[5]:

	Date	Region	Item	UnitPrice	Units
0	4-Jul-2014	East	Pen Set	4.99	62
1	12-Jul-2014	East	Binder	1.99	29
2	21-Jul-2014	Central	Pen Set	12.49	55
3	29-Jul-2014	East	Binder	19.99	81
4	7-Aug-2014	Central	Pen Set	23.95	42

Answer the following questions.

1). Append a new variable $Amount = UnitPric \times Units$.

In [8]:
```
df['Amount'] = df.UnitPrice * df.Units
```

2). Average unit price for each region. Use the groupby() method.

In [40]:
```
AverageUprice = df.groupby("Region")['UnitPrice'].mean()
print(AverageUprice)
```

```
Region
Central    18.018750
East        9.143846
West       53.658333
Name: UnitPrice, dtype: float64
```

3). Average unit price for each region. Use the pivot_table() method.

In [25]:
```
AverageUPrice = df.pivot_table(df, index='Region')
print(AverageUPrice)
```

```
              Amount  UnitPrice      Units
Region
Central   464.127917  18.018750  49.958333
East      461.699231   9.143846  53.153846
West      414.453333  53.658333  38.500000
```

4). Average unit price and units for each region in one code sentence. Use the groupby() method.

In [26]:
```
df.groupby('Region')[["UnitPrice", "Units"]].mean()
```

Out[26]:

Region	UnitPrice	Units
Central	18.018750	49.958333
East	9.143846	53.153846
West	53.658333	38.500000

5). Average unit price and units for each region in one code sentence. Use the pivot_table() method.

In [29]:
```
df.pivot_table(df,index='Region')[["UnitPrice", "Units"]]
```

Out[29]:

Region	UnitPrice	Units
Central	18.018750	49.958333
East	9.143846	53.153846
West	53.658333	38.500000

6). Total units for each region and item type in one code sentence. Use the pivot_table() method. Fill the missing values with 0.

In [39]:
```
df.pivot_table(df,index=['Item','Region'], aggfunc='sum')['Units']
```

Out[39]:
```
Item     Region
Binder   Central    424
         East       234
         West        64
Desk     Central      7
         West         3
Pen      Central     27
         East       175
         West        76
Pen Set  Central    243
         East       152
Pencil   Central    498
         East       130
         West        88
Name: Units, dtype: int64
```

7). Total sales amount for each region and item type in one code sentence. Use the pivot_table() method. Fill the missing values with 0.

In [ ]:
```
df.pivot_table(df,index='Region')[[ "Units"]]
```