HW8

March 2, 2017

1 Homework 8

Fill out the following information (each category below should be on a separate line): Name:
Sarah Rambeau
Date Submitted: 3/2/2017

Use the following data set to answer the questions for your homework:

```
In [1]: import pandas as pd
        import matplotlib.pyplot as plt

        # run plots in the notebook
        %matplotlib inline

        url = "http://pbpython.com/extras/sample-salesv2.csv"

        sales = pd.read_csv(url)

In [2]: sales.columns = ['acct_num', 'name', 'sku', 'category', 'quantity', 'unit_p

In [3]: sales.head()
```

Out[3]:

	acct_num	name	sku	category	quantity	\
0	296809	Carroll PLC	QN-82852	Belt	13	
1	98022	Heidenreich-Bosco	MJ-21460	Shoes	19	
2	563905	Kerluke, Reilly and Bechtelar	AS-93055	Shirt	12	
3	93356	Waters-Walker	AS-93055	Shirt	5	
4	659366	Waelchi-Fahey	AS-93055	Shirt	18	

	unit_price	ext_price	date
0	44.48	578.24	2014-09-27 07:13:03
1	53.62	1018.78	2014-07-29 02:10:44
2	24.16	289.92	2014-03-01 10:51:24
3	82.68	413.40	2013-11-17 20:41:11
4	99.64	1793.52	2014-01-03 08:14:27

Subset the dataframe to contain only the name, category, quantity and unit price columns

```
In [4]: new_df = sales[['name', 'category', 'quantity', 'unit_price']]
        new_df.head(20)
```

Out[4]:

	name	category	quantity	unit_price
0	Carroll PLC	Belt	13	44.48
1	Heidenreich-Bosco	Shoes	19	53.62
2	Kerluke, Reilly and Bechtelar	Shirt	12	24.16
3	Waters-Walker	Shirt	5	82.68
4	Waelchi-Fahey	Shirt	18	99.64
5	Kerluke, Reilly and Bechtelar	Shirt	17	52.82
6	Cole-Eichmann	Shoes	18	15.28
7	Hegmann and Sons	Shoes	7	78.78
8	Senger, Upton and Breitenberg	Shoes	17	38.19
9	Kerluke, Reilly and Bechtelar	Shirt	12	26.98
10	Cole-Eichmann	Shirt	19	60.22
11	Hegmann and Sons	Belt	6	13.12
12	Kihn, McClure and Denesik	Shoes	4	59.69
13	Ernser, Cruickshank and Lind	Shirt	12	97.25
14	Koelpin PLC	Shoes	9	81.44
15	Waters-Walker	Shoes	18	53.33
16	Kerluke, Reilly and Bechtelar	Shirt	4	35.62
17	Koelpin PLC	Shirt	17	98.23
18	Kihn, McClure and Denesik	Belt	15	69.52
19	Carroll PLC	Shirt	12	80.12

Subset the dataframe to contain only shirt sales

```
In [5]: new_df = new_df[new_df['category']== 'Shirt']
        new_df.head(20)
```

Out[5]:

	name	category	quantity	unit_price
2	Kerluke, Reilly and Bechtelar	Shirt	12	24.16
3	Waters-Walker	Shirt	5	82.68
4	Waelchi-Fahey	Shirt	18	99.64
5	Kerluke, Reilly and Bechtelar	Shirt	17	52.82
9	Kerluke, Reilly and Bechtelar	Shirt	12	26.98
10	Cole-Eichmann	Shirt	19	60.22
13	Ernser, Cruickshank and Lind	Shirt	12	97.25
16	Kerluke, Reilly and Bechtelar	Shirt	4	35.62
17	Koelpin PLC	Shirt	17	98.23
19	Carroll PLC	Shirt	12	80.12
24	Carroll PLC	Shirt	14	47.68
26	Kerluke, Reilly and Bechtelar	Shirt	6	58.52
28	Kunze Inc	Shirt	12	93.67
29	Gorczany-Hahn	Shirt	12	72.63
30	Kilback-Gerlach	Shirt	2	90.80
31	Hamill-Hackett	Shirt	6	42.96

```
32                Kilback-Gerlach    Shirt         1       71.50
33    Davis, Kshlerin and Reilly     Shirt         1       74.43
37                  Waters-Walker    Shirt        12       94.34
40                  Gorczany-Hahn    Shirt        11       11.21
```

Calculate the total cost per shirt sale

```
In [6]: new_df['shirt_sales'] = new_df.quantity * new_df.unit_price
        new_df.head(20)
```

```
Out[6]:                            name category  quantity  unit_price  shirt_sal
2   Kerluke, Reilly and Bechtelar    Shirt        12       24.16       289.
3                   Waters-Walker    Shirt         5       82.68       413.
4                   Waelchi-Fahey    Shirt        18       99.64      1793.
5   Kerluke, Reilly and Bechtelar    Shirt        17       52.82       897.
9   Kerluke, Reilly and Bechtelar    Shirt        12       26.98       323.
10                  Cole-Eichmann    Shirt        19       60.22      1144.
13   Ernser, Cruickshank and Lind    Shirt        12       97.25      1167.
16  Kerluke, Reilly and Bechtelar    Shirt         4       35.62       142.
17                    Koelpin PLC    Shirt        17       98.23      1669.
19                    Carroll PLC    Shirt        12       80.12       961.
24                    Carroll PLC    Shirt        14       47.68       667.
26  Kerluke, Reilly and Bechtelar    Shirt         6       58.52       351.
28                      Kunze Inc    Shirt        12       93.67      1124.
29                  Gorczany-Hahn    Shirt        12       72.63       871.
30                Kilback-Gerlach    Shirt         2       90.80       181.
31                 Hamill-Hackett    Shirt         6       42.96       257.
32                Kilback-Gerlach    Shirt         1       71.50        71.
33    Davis, Kshlerin and Reilly     Shirt         1       74.43        74.
37                  Waters-Walker    Shirt        12       94.34      1132.
40                  Gorczany-Hahn    Shirt        11       11.21       123.
```

Group the shirt sales by company name

```
In [7]: shirts_company = new_df.groupby('name', as_index=False).sum()
        shirts_company.head(20)
```

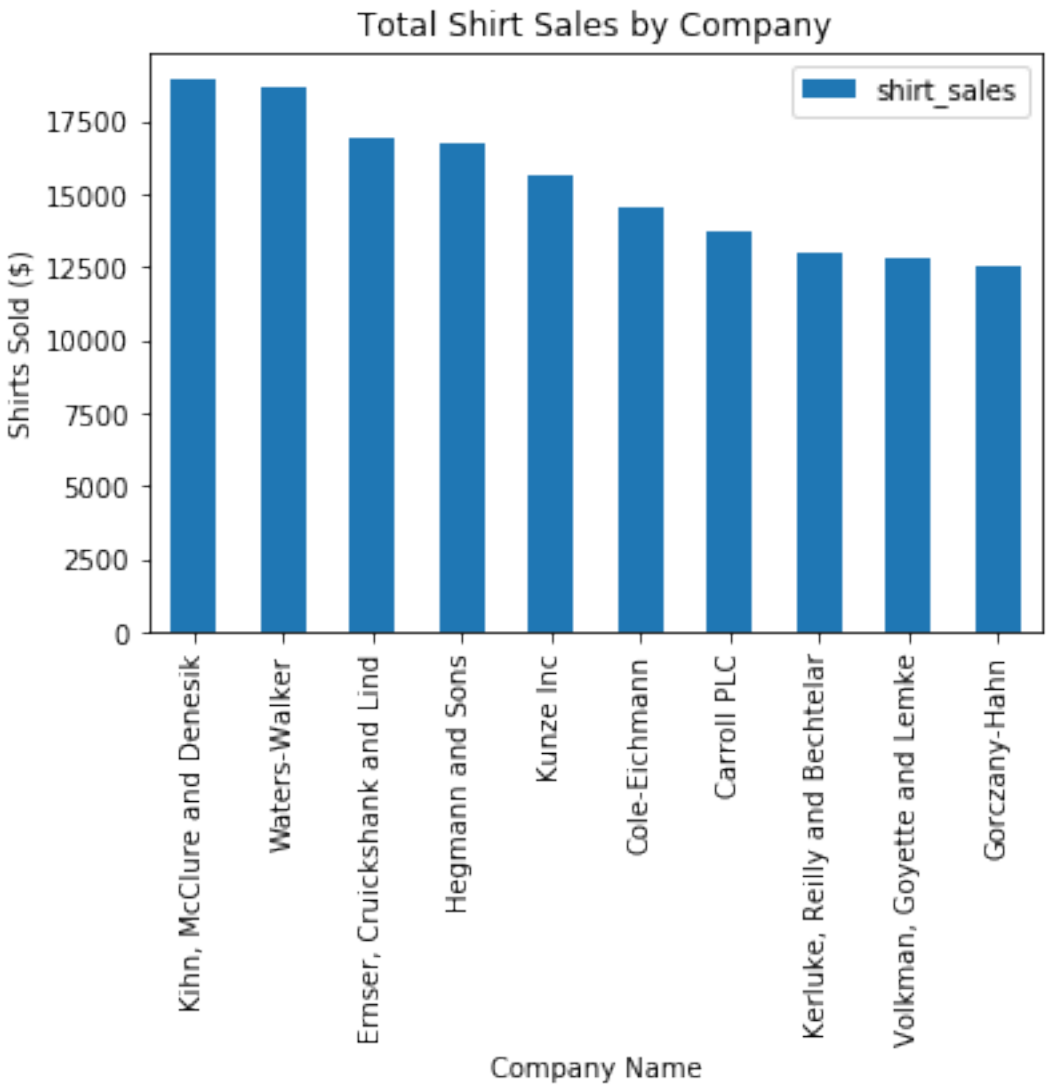
```
Out[7]:                            name  quantity  unit_price  shirt_sales
0                     Berge LLC       166     1226.54      9670.24
1                   Carroll PLC       257     1098.93     13717.61
2                  Cole-Eichmann      236     1226.75     14528.01
3     Davis, Kshlerin and Reilly      161      828.51      7533.03
4   Ernser, Cruickshank and Lind      262     1500.25     16944.19
5                  Gorczany-Hahn      237     1132.22     12576.83
6                 Hamill-Hackett      148     1091.55      8880.04
7               Hegmann and Sons      278     1528.84     16774.47
8              Heidenreich-Bosco       92      582.24      5965.25
9                     Huel-Haag      200     1146.17     11944.01
```

```
10  Kerluke, Reilly and Bechtelar       269     1038.53     12958.23
11      Kihn, McClure and Denesik       288     1653.58     18956.35
12                Kilback-Gerlach       163     1052.53      9904.85
13                    Koelpin PLC       132      786.07      7908.28
14                      Kunze Inc       260     1439.92     15638.87
15          Kuphal, Zieme and Kub       252     1167.28     12101.14
16  Senger, Upton and Breitenberg       144      939.38      7659.70
17     Volkman, Goyette and Lemke       220     1136.25     12791.27
18                  Waelchi-Fahey       201     1057.67     11689.05
19                  Waters-Walker       288     1603.36     18633.71
```

Graph the top 10 shirt sales

```
In [8]: top_sellers = shirts_company.sort_values(by='shirt_sales', ascending=False)
        shirt_plot = top_sellers.plot(kind="bar",
                                      title="Total Shirt Sales by Company",
                                      x="name",
                                      y="shirt_sales")
        shirt_plot.set_xlabel("Company Name")
        shirt_plot.set_ylabel("Shirts Sold ($)")

Out[8]: <matplotlib.text.Text at 0x1c692131518>
```



1.0.1 To turn in your homework:

- Save this notebook as a PDF (or html if you can't get PDF output working)
- Upload the file to GitHub
- Provide the URL to this file on your GitHub repo in canvas

In [ ]: