

# Homework 8

Fill out the following information (each category below should be on a separate line): ¶

Name: Dulce Carrillo

Date Submitted: 03/14/2017

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

```
In [4]: 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 [5]: sales.columns
```

```
Out[5]: Index(['account number', 'name', 'sku', 'category', 'quantity', 'unit price',
              'ext price', 'date'],
              dtype='object')
```

Changing column names:

```
In [9]: sales = sales.rename (columns={'account number': 'acct_num', 'unit price': 'uni
t_price', 'ext price': 'ext_price'})
```

In [10]: `sales.head()`

Out[10]:

	acct_num	name	sku	category	quantity	unit_price	ext_price	date
0	296809	Carroll PLC	QN-82852	Belt	13	44.48	578.24	2014-09-27 07:13:03
1	98022	Heidenreich-Bosco	MJ-21460	Shoes	19	53.62	1018.78	2014-07-29 02:10:44
2	563905	Kerluke, Reilly and Bechtelar	AS-93055	Shirt	12	24.16	289.92	2014-03-01 10:51:24
3	93356	Waters-Walker	AS-93055	Shirt	5	82.68	413.40	2013-11-17 20:41:11
4	659366	Waelchi-Fahey	AS-93055	Shirt	18	99.64	1793.52	2014-01-03 08:14:27

In [ ]:

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

```
In [11]: sales.iloc[:,[1,3,4,5]]
```

Out[11]:

	<b>name</b>	<b>category</b>	<b>quantity</b>	<b>unit_price</b>
<b>0</b>	Carroll PLC	Belt	13	44.48
<b>1</b>	Heidenreich-Bosco	Shoes	19	53.62
<b>2</b>	Kerluke, Reilly and Bechtelar	Shirt	12	24.16
<b>3</b>	Waters-Walker	Shirt	5	82.68
<b>4</b>	Waelchi-Fahey	Shirt	18	99.64
<b>5</b>	Kerluke, Reilly and Bechtelar	Shirt	17	52.82
<b>6</b>	Cole-Eichmann	Shoes	18	15.28
<b>7</b>	Hegmann and Sons	Shoes	7	78.78
<b>8</b>	Senger, Upton and Breitenberg	Shoes	17	38.19
<b>9</b>	Kerluke, Reilly and Bechtelar	Shirt	12	26.98
<b>10</b>	Cole-Eichmann	Shirt	19	60.22
<b>11</b>	Hegmann and Sons	Belt	6	13.12
<b>12</b>	Kihn, McClure and Denesik	Shoes	4	59.69
<b>13</b>	Ernser, Cruickshank and Lind	Shirt	12	97.25
<b>14</b>	Koelpin PLC	Shoes	9	81.44
<b>15</b>	Waters-Walker	Shoes	18	53.33
<b>16</b>	Kerluke, Reilly and Bechtelar	Shirt	4	35.62
<b>17</b>	Koelpin PLC	Shirt	17	98.23
<b>18</b>	Kihn, McClure and Denesik	Belt	15	69.52
<b>19</b>	Carroll PLC	Shirt	12	80.12
<b>20</b>	Volkman, Goyette and Lemke	Belt	13	81.19
<b>21</b>	Senger, Upton and Breitenberg	Shoes	2	48.15
<b>22</b>	Kerluke, Reilly and Bechtelar	Shoes	1	54.94
<b>23</b>	Berge LLC	Belt	4	57.75
<b>24</b>	Carroll PLC	Shirt	14	47.68
<b>25</b>	Huel-Haag	Belt	2	41.40
<b>26</b>	Kerluke, Reilly and Bechtelar	Shirt	6	58.52
<b>27</b>	Volkman, Goyette and Lemke	Shoes	12	96.62
<b>28</b>	Kunze Inc	Shirt	12	93.67
<b>29</b>	Gorczyan-Hahn	Shirt	12	72.63
<b>...</b>	...	...	...	...
<b>970</b>	Kihn, McClure and Denesik	Belt	15	51.35

	name	category	quantity	unit_price
971	Kerluke, Reilly and Bechtelar	Shoes	16	90.96
972	Kilback-Gerlach	Shoes	12	22.81
973	Waters-Walker	Shirt	19	28.76
974	Heidenreich-Bosco	Shirt	2	35.71
975	Carroll PLC	Shoes	9	32.75
976	Kihn, McClure and Denesik	Shoes	2	92.23
977	Hegmann and Sons	Shoes	12	18.24
978	Davis, Kshlerin and Reilly	Shirt	6	49.81
979	Huel-Haag	Belt	2	22.38
980	Senger, Upton and Breitenberg	Belt	10	25.65
981	Kuphal, Zieme and Kub	Belt	12	37.50
982	Gorczy-Hahn	Shirt	20	72.35
983	Kihn, McClure and Denesik	Shoes	2	48.21
984	Kilback-Gerlach	Shirt	1	91.96
985	Senger, Upton and Breitenberg	Shoes	7	62.27
986	Waters-Walker	Shirt	20	99.74
987	Heidenreich-Bosco	Belt	15	82.82
988	Berge LLC	Shoes	8	85.04
989	Cole-Eichmann	Shoes	10	73.24
990	Waters-Walker	Shirt	9	68.62
991	Huel-Haag	Belt	1	57.33
992	Kunze Inc	Shirt	9	27.16
993	Waelchi-Fahey	Shoes	10	56.56
994	Carroll PLC	Belt	20	14.37
995	Kihn, McClure and Denesik	Belt	11	60.50
996	Kunze Inc	Shoes	16	19.66
997	Waters-Walker	Shoes	13	90.95
998	Huel-Haag	Shoes	9	98.22
999	Heidenreich-Bosco	Shoes	14	74.83

1000 rows × 4 columns

### Subset the dataframe to contain only shirt sales

```
In [12]: Shirt_df = sales[['name', 'category', 'quantity', 'unit_price']]
        Shirt_df = Shirt_df[Shirt_df['category']== 'Shirt']
```

```
In [14]: Shirt_df.head()
```

```
Out[14]:
```

	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

### Calculate the total cost per shirt sale

```
In [15]: Shirt_df['Shirt_sales'] = Shirt_df.quantity * Shirt_df.unit_price
```

```
In [16]: Shirt_df.head()
```

```
Out[16]:
```

	name	category	quantity	unit_price	Shirt_sales
2	Kerluke, Reilly and Bechtelar	Shirt	12	24.16	289.92
3	Waters-Walker	Shirt	5	82.68	413.40
4	Waelchi-Fahey	Shirt	18	99.64	1793.52
5	Kerluke, Reilly and Bechtelar	Shirt	17	52.82	897.94
9	Kerluke, Reilly and Bechtelar	Shirt	12	26.98	323.76

```
In [20]:
```

### Group the shirt sales by company name

```
In [24]: Shirt_by_company = Shirt_df.groupby('name').sum()
```

In [25]: `Shirt_by_company.head()`

Out[25]:

	quantity	unit_price	Shirt_sales
name			
<b>Berge LLC</b>	166	1226.54	9670.24
<b>Carroll PLC</b>	257	1098.93	13717.61
<b>Cole-Eichmann</b>	236	1226.75	14528.01
<b>Davis, Kshlerin and Reilly</b>	161	828.51	7533.03
<b>Ernser, Cruickshank and Lind</b>	262	1500.25	16944.19

In [26]: `Shirt_by_company = Shirt_df.groupby('name', as_index=False).sum()`

### Graph the top 10 shirt sales

In [28]: `top_sellers = Shirt_by_company.sort_values(by='Shirt_sales', ascending=False).head(10)`

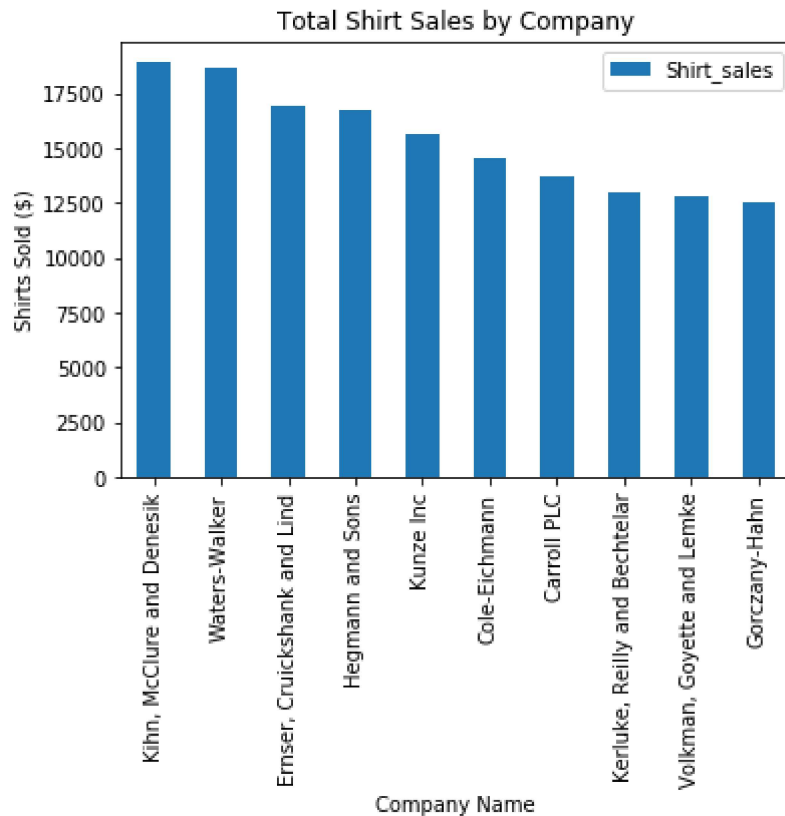
In [29]: `top_sellers`

Out[29]:

	name	quantity	unit_price	Shirt_sales
<b>11</b>	Kihn, McClure and Denesik	288	1653.58	18956.35
<b>19</b>	Waters-Walker	288	1603.36	18633.71
<b>4</b>	Ernser, Cruickshank and Lind	262	1500.25	16944.19
<b>7</b>	Hegmann and Sons	278	1528.84	16774.47
<b>14</b>	Kunze Inc	260	1439.92	15638.87
<b>2</b>	Cole-Eichmann	236	1226.75	14528.01
<b>1</b>	Carroll PLC	257	1098.93	13717.61
<b>10</b>	Kerluke, Reilly and Bechtelar	269	1038.53	12958.23
<b>17</b>	Volkman, Goyette and Lemke	220	1136.25	12791.27
<b>5</b>	Gorczy-Hahn	237	1132.22	12576.83

```
In [31]: 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[31]: <matplotlib.text.Text at 0x16091f0>



## 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 [ ]:

In [ ]: