

HW8

March 8, 2017

1 Homework 8

Fill out the following information (each category below should be on a separate line): Name: Yongqi Yan Date Submitted: 3/8/2017

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

```
In [6]: 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 [7]: sales.head()
```

```
Out[7]:
```

	account number	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 [15]: pd.DataFrame(sales, columns=['name', 'category', 'quantity', 'unit price'])
```

Out [15]:

	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
20	Volkman, Goyette and Lemke	Belt	13		81.19
21	Senger, Upton and Breitenberg	Shoes	2		48.15
22	Kerluke, Reilly and Bechtelar	Shoes	1		54.94
23	Berge LLC	Belt	4		57.75
24	Carroll PLC	Shirt	14		47.68
25	Huel-Haag	Belt	2		41.40
26	Kerluke, Reilly and Bechtelar	Shirt	6		58.52
27	Volkman, Goyette and Lemke	Shoes	12		96.62
28	Kunze Inc	Shirt	12		93.67
29	Gorczy-Hahn	Shirt	12		72.63
..
970	Kihn, McClure and Denesik	Belt	15		51.35
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 x 4 columns]

Subset the dataframe to contain only shirt sales

```
In [14]: sales[sales.category=='Shirt']
```

```
Out[14]:
```

	account number	name	sku	category \
2	563905	Kerluke, Reilly and Bechtelar	AS-93055	Shirt
3	93356	Waters-Walker	AS-93055	Shirt
4	659366	Waelchi-Fahey	AS-93055	Shirt
5	563905	Kerluke, Reilly and Bechtelar	AS-93055	Shirt
9	563905	Kerluke, Reilly and Bechtelar	KV-99194	Shirt
10	995267	Cole-Eichmann	KV-99194	Shirt
13	555594	Ernser, Cruickshank and Lind	FK-71853	Shirt
16	563905	Kerluke, Reilly and Bechtelar	KV-99194	Shirt
17	201259	Koelpin PLC	KV-99194	Shirt
19	296809	Carroll PLC	VG-32047	Shirt
24	296809	Carroll PLC	AS-93055	Shirt
26	563905	Kerluke, Reilly and Bechtelar	VG-32047	Shirt
28	850140	Kunze Inc	KV-99194	Shirt
29	115138	Gorczy-Hahn	VG-32047	Shirt
30	711951	Kilback-Gerlach	AS-93055	Shirt
31	676847	Hamill-Hackett	AS-93055	Shirt
32	711951	Kilback-Gerlach	AS-93055	Shirt
33	305803	Davis, Kshlerin and Reilly	AS-93055	Shirt
37	93356	Waters-Walker	AS-93055	Shirt
40	115138	Gorczy-Hahn	VG-32047	Shirt
41	201259	Koelpin PLC	KV-99194	Shirt
49	555594	Ernser, Cruickshank and Lind	FK-71853	Shirt
50	995267	Cole-Eichmann	KV-99194	Shirt
52	758133	Kihn, McClure and Denesik	FK-71853	Shirt
54	201259	Koelpin PLC	FK-71853	Shirt
55	555594	Ernser, Cruickshank and Lind	KV-99194	Shirt

61	995267	Cole-Eichmann	AS-93055	Shirt
62	115138	Gorczyany-Hahn	AS-93055	Shirt
65	115138	Gorczyany-Hahn	KV-99194	Shirt
66	555594	Ernser, Cruickshank and Lind	VG-32047	Shirt
..
918	995267	Cole-Eichmann	AS-93055	Shirt
920	563905	Kerluke, Reilly and Bechtelar	VG-32047	Shirt
923	93356	Waters-Walker	KV-99194	Shirt
925	850140	Kunze Inc	KV-99194	Shirt
926	115138	Gorczyany-Hahn	AS-93055	Shirt
935	305803	Davis, Kshlerin and Reilly	FK-71853	Shirt
936	201259	Koelpin PLC	KV-99194	Shirt
941	995267	Cole-Eichmann	KV-99194	Shirt
942	555594	Ernser, Cruickshank and Lind	AS-93055	Shirt
943	524021	Hegmann and Sons	KV-99194	Shirt
944	524021	Hegmann and Sons	AS-93055	Shirt
945	711951	Kilback-Gerlach	KV-99194	Shirt
947	296809	Carroll PLC	VG-32047	Shirt
948	93356	Waters-Walker	AS-93055	Shirt
949	995267	Cole-Eichmann	KV-99194	Shirt
950	676847	Hamill-Hackett	FK-71853	Shirt
954	995267	Cole-Eichmann	AS-93055	Shirt
955	93356	Waters-Walker	AS-93055	Shirt
957	929400	Senger, Upton and Breitenberg	VG-32047	Shirt
958	734922	Berge LLC	VG-32047	Shirt
962	555594	Ernser, Cruickshank and Lind	KV-99194	Shirt
965	98022	Heidenreich-Bosco	AS-93055	Shirt
973	93356	Waters-Walker	AS-93055	Shirt
974	98022	Heidenreich-Bosco	VG-32047	Shirt
978	305803	Davis, Kshlerin and Reilly	AS-93055	Shirt
982	115138	Gorczyany-Hahn	VG-32047	Shirt
984	711951	Kilback-Gerlach	VG-32047	Shirt
986	93356	Waters-Walker	KV-99194	Shirt
990	93356	Waters-Walker	AS-93055	Shirt
992	850140	Kunze Inc	KV-99194	Shirt

	quantity	unit	price	ext price	date
2	12		24.16	289.92	2014-03-01 10:51:24
3	5		82.68	413.40	2013-11-17 20:41:11
4	18		99.64	1793.52	2014-01-03 08:14:27
5	17		52.82	897.94	2013-12-04 02:07:05
9	12		26.98	323.76	2014-05-20 00:21:28
10	19		60.22	1144.18	2014-03-10 06:23:31
13	12		97.25	1167.00	2014-09-19 13:20:00
16	4		35.62	142.48	2014-10-05 23:38:16
17	17		98.23	1669.91	2014-01-26 01:52:36
19	12		80.12	961.44	2014-05-24 16:03:28
24	14		47.68	667.52	2014-01-12 00:28:09

26	6	58.52	351.12	2014-03-14	02:28:57
28	12	93.67	1124.04	2014-02-19	06:03:09
29	12	72.63	871.56	2014-02-07	03:30:14
30	2	90.80	181.60	2014-03-26	20:56:14
31	6	42.96	257.76	2013-11-14	22:16:18
32	1	71.50	71.50	2014-04-14	16:58:07
33	1	74.43	74.43	2014-09-20	09:49:19
37	12	94.34	1132.08	2014-09-04	13:39:04
40	11	11.21	123.31	2013-12-10	20:13:01
41	4	29.79	119.16	2013-11-07	17:43:31
49	6	65.40	392.40	2014-09-17	19:15:27
50	10	23.79	237.90	2013-11-17	23:56:36
52	11	13.06	143.66	2014-03-01	08:58:41
54	14	57.69	807.66	2014-05-26	16:29:02
55	10	83.05	830.50	2014-01-05	10:16:57
61	7	41.75	292.25	2014-07-08	15:01:53
62	13	22.23	288.99	2014-01-13	18:31:57
65	10	85.12	851.20	2014-05-17	03:23:30
66	17	75.04	1275.68	2014-02-09	16:56:59
..
918	20	96.14	1922.80	2014-08-02	19:08:15
920	16	64.75	1036.00	2014-02-11	09:04:11
923	2	96.80	193.60	2014-07-13	09:44:29
925	8	35.98	287.84	2014-05-20	14:38:17
926	9	14.71	132.39	2014-03-01	13:31:22
935	20	11.96	239.20	2014-09-04	18:27:56
936	12	91.33	1095.96	2014-03-01	09:46:48
941	13	97.04	1261.52	2014-06-17	08:28:31
942	6	29.88	179.28	2014-09-04	14:54:23
943	6	62.36	374.16	2014-01-07	03:14:00
944	15	88.44	1326.60	2014-10-15	01:54:05
945	16	79.71	1275.36	2014-08-30	06:23:27
947	12	79.85	958.20	2014-01-16	11:38:27
948	2	47.74	95.48	2014-03-22	02:00:16
949	18	97.33	1751.94	2014-01-16	12:27:14
950	18	98.39	1771.02	2014-01-06	08:14:31
954	7	80.69	564.83	2014-01-10	21:42:10
955	15	84.23	1263.45	2014-07-18	04:09:08
957	10	67.36	673.60	2013-11-02	03:09:24
958	14	93.45	1308.30	2014-02-10	07:13:32
962	14	93.56	1309.84	2014-10-21	14:14:21
965	4	22.33	89.32	2014-05-22	10:55:51
973	19	28.76	546.44	2013-12-19	19:43:07
974	2	35.71	71.42	2014-07-03	03:16:33
978	6	49.81	298.86	2013-11-01	16:31:25
982	20	72.35	1447.00	2014-08-11	19:02:14
984	1	91.96	91.96	2013-10-23	15:08:31
986	20	99.74	1994.80	2014-04-18	15:39:12

```

990          9      68.62      617.58  2014-08-12 08:06:21
992          9      27.16      244.44  2014-02-08 06:22:34

```

[404 rows x 8 columns]

Calculate the total cost per shirt sale

```
In [10]: sum(sales['ext price']) /sum(sales['quantity'])
```

```
Out[10]: 56.142902788535977
```

Group the shirt sales by company name

```
In [18]: shirtSale=sales[sales.category=='Shirt']
        shirtSale.sort_values(by='name',ascending=False)
```

```
Out[18]:
```

	account	number	name	sku	category	quantity	unit	price	\
37		93356	Waters-Walker	AS-93055	Shirt	12		94.34	
683		93356	Waters-Walker	AS-93055	Shirt	9		95.16	
432		93356	Waters-Walker	VG-32047	Shirt	13		45.97	
123		93356	Waters-Walker	VG-32047	Shirt	17		86.18	
130		93356	Waters-Walker	FK-71853	Shirt	16		21.43	
359		93356	Waters-Walker	FK-71853	Shirt	17		58.55	
167		93356	Waters-Walker	AS-93055	Shirt	16		75.48	
906		93356	Waters-Walker	KV-99194	Shirt	9		70.81	
617		93356	Waters-Walker	VG-32047	Shirt	17		71.32	
915		93356	Waters-Walker	FK-71853	Shirt	6		48.34	
633		93356	Waters-Walker	VG-32047	Shirt	13		45.42	
923		93356	Waters-Walker	KV-99194	Shirt	2		96.80	
531		93356	Waters-Walker	KV-99194	Shirt	19		47.53	
549		93356	Waters-Walker	KV-99194	Shirt	7		59.50	
254		93356	Waters-Walker	VG-32047	Shirt	2		75.16	
955		93356	Waters-Walker	AS-93055	Shirt	15		84.23	
990		93356	Waters-Walker	AS-93055	Shirt	9		68.62	
986		93356	Waters-Walker	KV-99194	Shirt	20		99.74	
973		93356	Waters-Walker	AS-93055	Shirt	19		28.76	
671		93356	Waters-Walker	KV-99194	Shirt	12		66.41	
506		93356	Waters-Walker	KV-99194	Shirt	11		83.22	
649		93356	Waters-Walker	KV-99194	Shirt	20		49.97	
948		93356	Waters-Walker	AS-93055	Shirt	2		47.74	
3		93356	Waters-Walker	AS-93055	Shirt	5		82.68	
330		659366	Waelchi-Fahey	FK-71853	Shirt	17		75.46	
765		659366	Waelchi-Fahey	VG-32047	Shirt	11		25.74	
216		659366	Waelchi-Fahey	FK-71853	Shirt	15		89.45	
179		659366	Waelchi-Fahey	FK-71853	Shirt	2		12.68	
862		659366	Waelchi-Fahey	VG-32047	Shirt	14		85.44	
652		659366	Waelchi-Fahey	KV-99194	Shirt	13		32.24	
..		

655	296809	Carroll PLC	VG-32047	Shirt	11	33.31
282	296809	Carroll PLC	VG-32047	Shirt	10	91.46
798	296809	Carroll PLC	VG-32047	Shirt	16	82.36
886	296809	Carroll PLC	KV-99194	Shirt	13	43.03
677	296809	Carroll PLC	AS-93055	Shirt	19	67.10
596	296809	Carroll PLC	AS-93055	Shirt	18	55.23
851	296809	Carroll PLC	VG-32047	Shirt	11	14.92
264	296809	Carroll PLC	VG-32047	Shirt	15	63.45
753	296809	Carroll PLC	FK-71853	Shirt	3	55.60
227	296809	Carroll PLC	AS-93055	Shirt	1	58.80
712	734922	Berge LLC	KV-99194	Shirt	6	88.31
713	734922	Berge LLC	VG-32047	Shirt	11	58.29
162	734922	Berge LLC	AS-93055	Shirt	10	93.77
411	734922	Berge LLC	VG-32047	Shirt	1	93.10
869	734922	Berge LLC	FK-71853	Shirt	4	59.00
153	734922	Berge LLC	KV-99194	Shirt	7	64.02
319	734922	Berge LLC	VG-32047	Shirt	2	33.92
329	734922	Berge LLC	KV-99194	Shirt	7	46.71
666	734922	Berge LLC	FK-71853	Shirt	6	17.54
958	734922	Berge LLC	VG-32047	Shirt	14	93.45
861	734922	Berge LLC	KV-99194	Shirt	13	21.76
477	734922	Berge LLC	FK-71853	Shirt	4	63.74
611	734922	Berge LLC	KV-99194	Shirt	16	24.13
172	734922	Berge LLC	AS-93055	Shirt	11	81.53
470	734922	Berge LLC	FK-71853	Shirt	7	58.95
605	734922	Berge LLC	AS-93055	Shirt	11	36.86
534	734922	Berge LLC	FK-71853	Shirt	17	49.12
571	734922	Berge LLC	AS-93055	Shirt	8	94.37
369	734922	Berge LLC	VG-32047	Shirt	4	96.25
722	734922	Berge LLC	FK-71853	Shirt	7	51.72

	ext price	date
37	1132.08	2014-09-04 13:39:04
683	856.44	2013-12-03 07:38:11
432	597.61	2014-07-05 08:56:46
123	1465.06	2014-08-13 15:42:49
130	342.88	2014-05-25 00:38:58
359	995.35	2014-05-04 23:02:26
167	1207.68	2014-08-23 04:04:11
906	637.29	2014-01-30 11:06:55
617	1212.44	2014-09-16 10:45:19
915	290.04	2014-10-07 11:12:45
633	590.46	2013-12-29 21:47:14
923	193.60	2014-07-13 09:44:29
531	903.07	2014-03-10 16:34:16
549	416.50	2014-04-29 12:04:58
254	150.32	2013-11-26 12:30:16
955	1263.45	2014-07-18 04:09:08

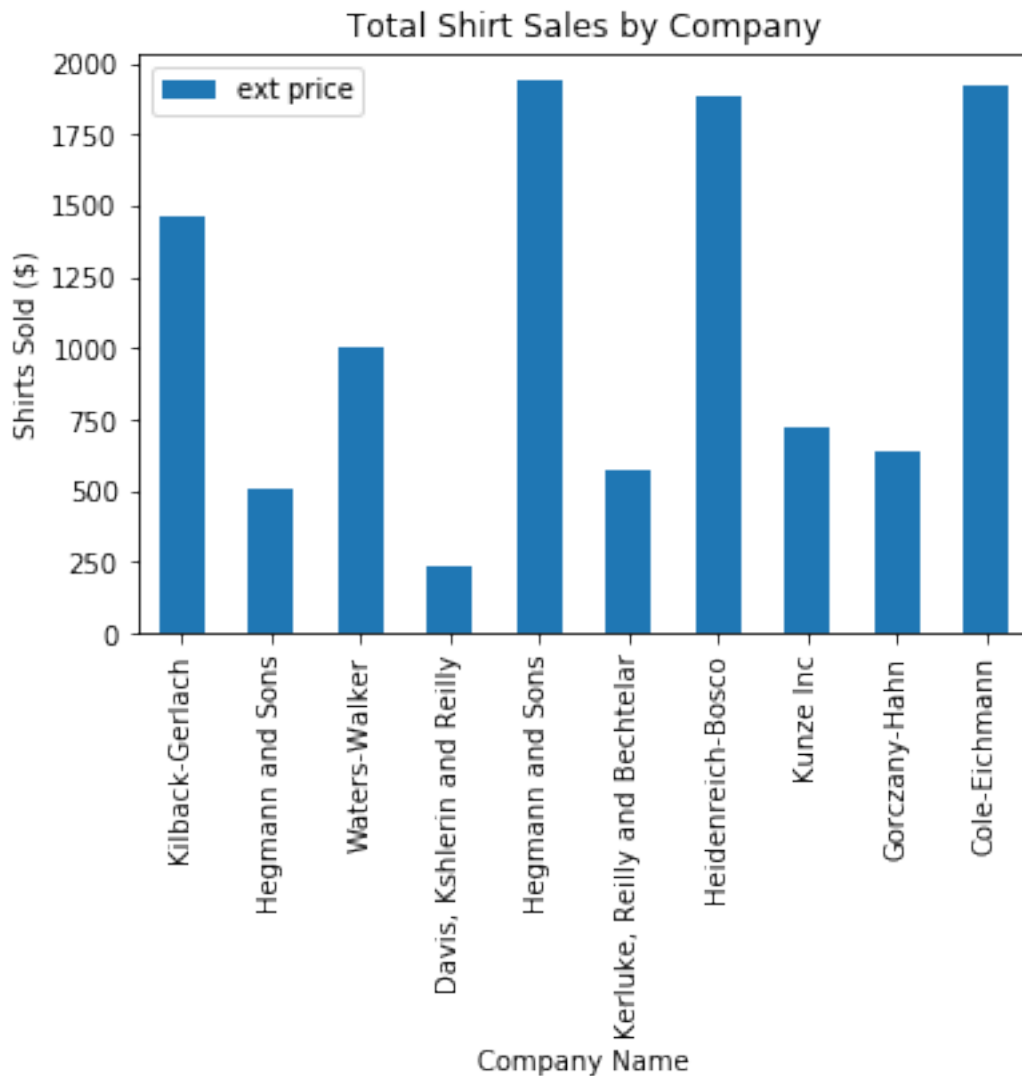
990	617.58	2014-08-12	08:06:21
986	1994.80	2014-04-18	15:39:12
973	546.44	2013-12-19	19:43:07
671	796.92	2014-04-02	00:13:13
506	915.42	2014-01-03	19:04:02
649	999.40	2014-04-28	20:35:33
948	95.48	2014-03-22	02:00:16
3	413.40	2013-11-17	20:41:11
330	1282.82	2013-11-17	15:36:57
765	283.14	2014-02-06	22:09:58
216	1341.75	2014-09-09	09:21:01
179	25.36	2014-04-27	03:38:16
862	1196.16	2014-06-04	14:34:07
652	419.12	2013-12-22	05:25:56
..
655	366.41	2014-07-29	15:09:19
282	914.60	2013-11-24	08:11:02
798	1317.76	2014-03-24	12:29:30
886	559.39	2014-01-20	19:43:09
677	1274.90	2014-06-20	16:14:34
596	994.14	2013-12-02	17:42:51
851	164.12	2014-01-14	05:53:11
264	951.75	2014-10-12	09:08:07
753	166.80	2013-11-27	18:35:28
227	58.80	2014-09-06	13:00:55
712	529.86	2013-12-23	16:08:03
713	641.19	2014-04-23	09:49:39
162	937.70	2014-08-17	09:21:15
411	93.10	2013-12-25	10:00:14
869	236.00	2014-08-15	07:13:09
153	448.14	2013-11-05	03:55:12
319	67.84	2014-06-27	15:53:05
329	326.97	2014-06-15	15:52:58
666	105.24	2014-08-19	04:51:46
958	1308.30	2014-02-10	07:13:32
861	282.88	2014-09-16	19:43:28
477	254.96	2014-08-31	19:57:31
611	386.08	2013-12-20	01:56:23
172	896.83	2014-09-01	17:46:10
470	412.65	2014-03-14	04:16:55
605	405.46	2014-01-01	08:13:10
534	835.04	2014-02-24	20:49:28
571	754.96	2013-12-31	01:14:11
369	385.00	2014-07-09	19:18:49
722	362.04	2013-12-07	09:43:31

[404 rows x 8 columns]

Graph the top 10 shirt sales

```
In [22]: shirtSaleTop10=shirtSale.sort_values(by='quantity',ascending=False).head(10)
shirt_plot= shirtSaleTop10.plot(kind="bar",
                                title="Total Shirt Sales by Company",
                                x="name",
                                y="ext price")
shirt_plot.set_xlabel("Company Name")
shirt_plot.set_ylabel("Shirts Sold ($)")
```

Out [22]: <matplotlib.text.Text at 0x7f0684900128>



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 []: