In [2]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

url='https://raw.githubusercontent.com/justmarkham/DAT8/master/data/ch
df=pd.read_csv(url,sep='\t')
df.head()
```

Out[2]:

	order_id	quantity	item_name	choice_description	item_price
0	1	1	Chips and Fresh Tomato Salsa	NaN	$2.39
1	1	1	Izze	[Clementine]	$3.39
2	1	1	Nantucket Nectar	[Apple]	$3.39
3	1	1	Chips and Tomatillo-Green Chili Salsa	NaN	$2.39
4	2	2	Chicken Bowl	[Tomatillo-Red Chili Salsa (Hot), [Black Beans...	$16.98

In [17]:
```
print(len(df))
print(df.shape)
df.columns
```

```
4622
(4622, 5)
```

Out[17]:
```
Index(['order_id', 'quantity', 'item_name', 'choice_description',
       'item_price'],
      dtype='object')
```

In [19]:
```
#the most frequent
df.groupby('item_name')['quantity'].sum().sort_values()[::-1]
```

Out[19]:
```
item_name
Chicken Bowl                           761
Chicken Burrito                        591
Chips and Guacamole                    506
Steak Burrito                          386
Canned Soft Drink                      351
Chips                                  230
Steak Bowl                             221
Bottled Water                          211
Chips and Fresh Tomato Salsa           130
Canned Soda                            126
Chicken Salad Bowl                     123
Chicken Soft Tacos                     120
Side of Chips                          110
Veggie Burrito                          97
Barbacoa Burrito                        91
```

```
Veggie Bowl                                 87
Carnitas Bowl                               71
Barbacoa Bowl                               66
Carnitas Burrito                            60
Steak Soft Tacos                            56
6 Pack Soft Drink                           55
Chips and Tomatillo Red Chili Salsa         50
Chicken Crispy Tacos                        50
Chips and Tomatillo Green Chili Salsa       45
Carnitas Soft Tacos                         40
Steak Crispy Tacos                          36
Chips and Tomatillo-Green Chili Salsa       33
Steak Salad Bowl                            31
Nantucket Nectar                            29
Chips and Tomatillo-Red Chili Salsa         25
Barbacoa Soft Tacos                         25
Chips and Roasted Chili Corn Salsa          23
Izze                                        20
Chips and Roasted Chili-Corn Salsa          18
Veggie Salad Bowl                           18
Barbacoa Crispy Tacos                       12
Barbacoa Salad Bowl                         10
Chicken Salad                                9
Veggie Soft Tacos                            8
Carnitas Crispy Tacos                        8
Burrito                                      6
Carnitas Salad Bowl                          6
Veggie Salad                                 6
Steak Salad                                  4
Bowl                                         4
Salad                                        2
Crispy Tacos                                 2
Chips and Mild Fresh Tomato Salsa            1
Veggie Crispy Tacos                          1
Carnitas Salad                               1
Name: quantity, dtype: int64
```

In [8]:
```
#amount of different items
len(df['item_name'].unique())
```

Out[8]: 50

In [9]:
```
#total amount of items
df['quantity'].sum()
```

Out[9]: 4972

In [10]:
```
#transform price into floats
#1 sposob
def clean(x):
    return float(x[1:])
df['item_price'].apply(clean)
```

Out[10]:

0	2.39
1	3.39
2	3.39
3	2.39
4	16.98
5	10.98
6	1.69
7	11.75
8	9.25
9	9.25
10	4.45
11	8.75
12	8.75
13	11.25
14	4.45
15	2.39
16	8.49
17	8.49
18	2.18
19	8.75
20	4.45
21	8.99
22	3.39
23	10.98
24	3.39
25	2.39
26	8.49
27	8.99
28	1.09
29	8.49
...	
4592	11.75
4593	11.75
4594	11.75
4595	8.75
4596	4.45
4597	1.25
4598	1.50
4599	8.75
4600	4.45
4601	1.25
4602	9.25
4603	9.25
4604	8.75
4605	4.45
4606	1.25
4607	11.75
4608	11.25
4609	1.25
4610	11.75
4611	11.25
4612	9.25
4613	2.15

```
4614     1.50
4615     8.75
4616     4.45
4617    11.75
4618    11.75
4619    11.25
4620     8.75
4621     8.75
Name: item_price, Length: 4622, dtype: float64
```

In [21]:
```
#transform price into floats
#2 sposob
df['item_price'].apply(lambda x: float(x[1:])).head()
#df.head() # samo df has not changed
```

Out[21]:
```
0     2.39
1     3.39
2     3.39
3     2.39
4    16.98
Name: item_price, dtype: float64
```

In [23]:
```
#change colomn of price
df['item_price_float']=df['item_price'].apply(lambda x: float(x[1:]))
df.head()
```

Out[23]:

	order_id	quantity	item_name	choice_description	item_price	item_price_float
0	1	1	Chips and Fresh Tomato Salsa	NaN	$2.39	2.39
1	1	1	Izze	[Clementine]	$3.39	3.39
2	1	1	Nantucket Nectar	[Apple]	$3.39	3.39
3	1	1	Chips and Tomatillo-Green Chili Salsa	NaN	$2.39	2.39
4	2	2	Chicken Bowl	[Tomatillo-Red Chili Salsa (Hot), [Black Beans...	$16.98	16.98

In [24]:
```
#the price of each item
df[['item_name','item_price']].drop_duplicates()
```

Out[24]:

	item_name	item_price
0	Chips and Fresh Tomato Salsa	$2.39
1	Izze	$3.39
2	Nantucket Nectar	$3.39
3	Chips and Tomatillo-Green Chili Salsa	$2.39
4	Chicken Bowl	$16.98

5	Chicken Bowl	$10.98
6	Side of Chips	$1.69
7	Steak Burrito	$11.75
8	Steak Soft Tacos	$9.25
9	Steak Burrito	$9.25
10	Chips and Guacamole	$4.45
11	Chicken Crispy Tacos	$8.75
12	Chicken Soft Tacos	$8.75
13	Chicken Bowl	$11.25
16	Chicken Burrito	$8.49
18	Canned Soda	$2.18
19	Chicken Bowl	$8.75
21	Barbacoa Burrito	$8.99
23	Chicken Burrito	$10.98
26	Chicken Bowl	$8.49
27	Carnitas Burrito	$8.99
28	Canned Soda	$1.09
31	Steak Burrito	$8.99
33	Carnitas Bowl	$8.99
34	Bottled Water	$1.09
38	Chips and Tomatillo Green Chili Salsa	$2.95
39	Barbacoa Bowl	$11.75
40	Chips	$2.15
44	Chicken Salad Bowl	$8.75
49	Chips and Guacamole	$3.99
...	...	...
3195	Barbacoa Bowl	$8.69
3252	Carnitas Burrito	$8.69
3276	Barbacoa Burrito	$11.48
3294	Bottled Water	$4.50
3334	Steak Bowl	$26.07
3354	Steak Bowl	$18.50
3375	Chicken Burrito	$16.98
3389	6 Pack Soft Drink	$12.98

3480	Carnitas Bowl	$35.25
3502	Steak Salad	$8.69
3598	Chips and Fresh Tomato Salsa	$44.25
3599	Bottled Water	$10.50
3601	Veggie Burrito	$33.75
3634	Chicken Burrito	$26.25
3658	Chicken Soft Tacos	$21.96
3750	Carnitas Salad	$8.99
3823	Chicken Burrito	$16.38
3887	Side of Chips	$13.52
3889	Veggie Soft Tacos	$16.98
3890	Carnitas Crispy Tacos	$17.98
3891	Chicken Crispy Tacos	$16.98
3973	Canned Soft Drink	$5.00
3989	Chicken Salad	$8.19
4152	Bottled Water	$15.00
4235	Chicken Bowl	$8.50
4237	Chips and Guacamole	$8.50
4354	Steak Soft Tacos	$18.50
4489	Chips and Guacamole	$17.80
4509	Chips	$1.99
4510	Barbacoa Bowl	$11.49

209 rows × 2 columns

In [14]:
```
#year revenue
(df['quantity']*df['item_price_float']).sum() #scalar product
```

Out[14]: 39237.02

In [15]:
```
#amount of orders
len(df['order_id'].unique())
```

Out[15]: 1834

In [26]:
```
#average cheque
df['total']=df['quantity']*df['item_price_float']
df.groupby('order_id').head()
df.groupby('order_id')['total'].sum().mean()
```

Out[26]:
```
21.394231188658654
```

In [28]:
```
dfa=pd.DataFrame({"date": np.random.choice(pd.date_range(start="201701
                  "order_id": np.arange(1800)})
dfa.head()
```

Out[28]:

	date	order_id
0	2017-01-01	0
1	2017-01-10	1
2	2017-01-04	2
3	2017-01-14	3
4	2017-01-15	4

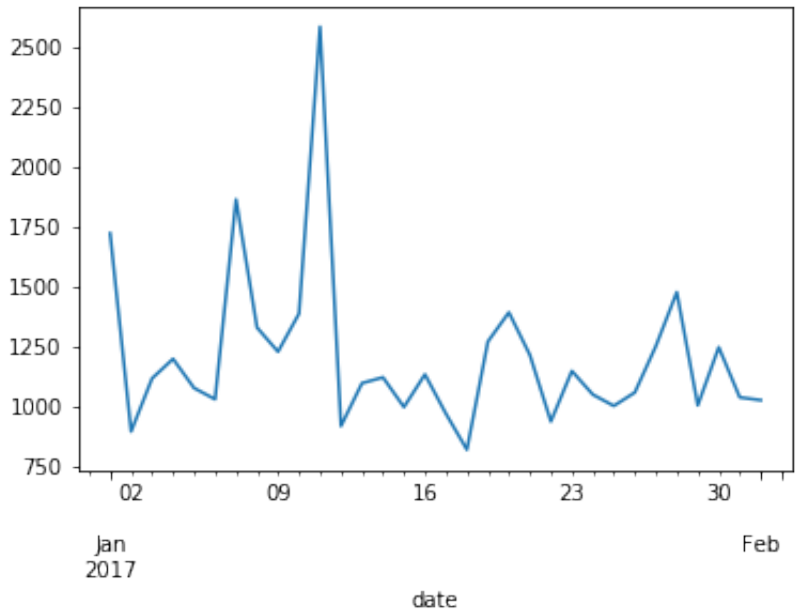
In [29]:
```
#let's glue them: by default how=inner
df=df.merge(dfa,how='inner')
df.head()
```

Out[29]:

	order_id	quantity	item_name	choice_description	item_price	item_price_float	total	dat
0	1	1	Chips and Fresh Tomato Salsa	NaN	$2.39	2.39	2.39	2017 01-1
1	1	1	Izze	[Clementine]	$3.39	3.39	3.39	2017 01-1
2	1	1	Nantucket Nectar	[Apple]	$3.39	3.39	3.39	2017 01-1
3	1	1	Chips and Tomatillo-Green Chili Salsa	NaN	$2.39	2.39	2.39	2017 01-1
4	2	2	Chicken Bowl	[Tomatillo-Red Chili Salsa (Hot), [Black Beans...	$16.98	16.98	33.96	2017 01-0

In [37]:
```
# group income by dates and plot
df.groupby('date')['total'].sum().plot()
```

Out[37]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x7f2dc67845d0>
```



In [41]:
```
#how many Veggie Salad Bowl
df[df['item_name']=='Veggie Salad Bowl']
```

Out[41]:

	order_id	quantity	item_name	choice_description	item_price	item_price_float	total	
186	83	1	Veggie Salad Bowl	[Fresh Tomato Salsa, [Fajita Vegetables, Rice,...	$11.25	11.25	11.25	2 C
295	128	1	Veggie Salad Bowl	[Fresh Tomato Salsa, [Fajita Vegetables, Lettu...	$11.25	11.25	11.25	2 C
455	195	1	Veggie Salad Bowl	[Fresh Tomato Salsa, [Fajita Vegetables, Rice,...	$11.25	11.25	11.25	2 C
496	207	1	Veggie Salad Bowl	[Fresh Tomato Salsa, [Rice, Lettuce, Guacamole...	$11.25	11.25	11.25	2 C
960	394	1	Veggie Salad Bowl	[Fresh Tomato Salsa, [Fajita Vegetables, Lettu...	$8.75	8.75	8.75	2 C
1316	536	1	Veggie Salad Bowl	[Fresh Tomato Salsa, [Fajita Vegetables, Rice,...	$8.75	8.75	8.75	2 C
1884	760	1	Veggie Salad Bowl	[Fresh Tomato Salsa, [Fajita Vegetables, Rice,...	$11.25	11.25	11.25	2 C
2156	869	1	Veggie Salad Bowl	[Tomatillo Red Chili Salsa, [Fajita	$11.25	11.25	11.25	2 C

				Vegetables...				
2223	896	1	Veggie Salad Bowl	[Roasted Chili Corn Salsa, Fajita Vegetables]	$8.75	8.75	8.75	2 0
2269	913	1	Veggie Salad Bowl	[Fresh Tomato Salsa, [Fajita Vegetables, Rice,...	$8.75	8.75	8.75	2 0
2683	1066	1	Veggie Salad Bowl	[Roasted Chili Corn Salsa, [Fajita Vegetables,...	$8.75	8.75	8.75	2 0
3223	1289	1	Veggie Salad Bowl	[Tomatillo Red Chili Salsa, [Fajita Vegetables...	$11.25	11.25	11.25	2 0
3293	1321	1	Veggie Salad Bowl	[Fresh Tomato Salsa, [Rice, Black Beans, Chees...	$8.75	8.75	8.75	2 0
4109	1646	1	Veggie Salad Bowl	[Tomatillo Red Chili Salsa, [Fajita Vegetables...	$11.25	11.25	11.25	2 0
4201	1677	1	Veggie Salad Bowl	[Fresh Tomato Salsa, [Fajita Vegetables, Black...	$11.25	11.25	11.25	2 0
4261	1700	1	Veggie Salad Bowl	[Fresh Tomato Salsa, [Fajita Vegetables, Rice,...	$11.25	11.25	11.25	2 0

In [42]:
```
df[df['item_name']=='Veggie Salad Bowl']['quantity'].sum()
```

Out[42]: 16

In [ ]:
```

```