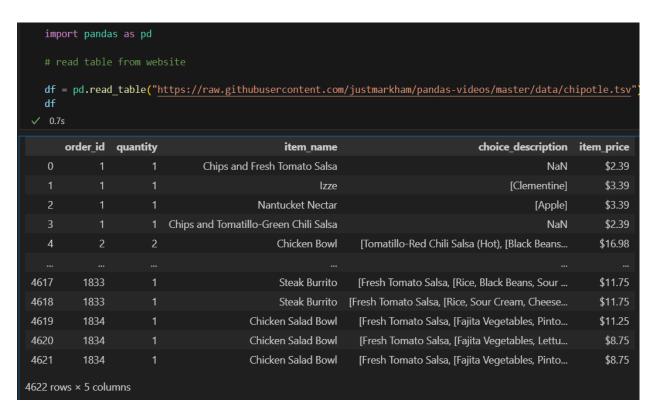
Practical two pandas: -

Link table:

https://raw.githubusercontent.com/justmarkham/pandasvideos/master/data/chipotle.tsv



Data cleaning

```
df.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4622 entries, 0 to 4621
Data columns (total 5 columns):
    Column
                        Non-Null Count Dtype
     order_id
                        4622 non-null
                                         int64
     quantity
                        4622 non-null
                                        int64
     item_name
                        4622 non-null
                                        object
     choice description 3376 non-null
                                        object 💪
     item_price
                         4622 non-null
                                         object
dtypes: int64(2), object(3)
memory usage: 180.7+ KB
```

كلهم مش متسواين يعني فيه

Missing values

فيه 1246 فيه

Missing values

```
# drop all rows that have na values
df = df.dropna()
```

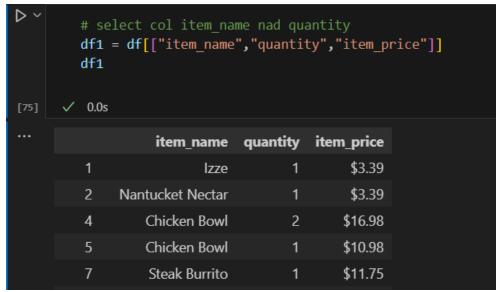
\Rightarrow After drop all rows

Check duplicated

There 41 rows are duplicated

عاوزين نعرف كل منتج اشتري كام مره بس تلاحظ ان فيه تلاته كولوم واحد -1 للاسم و التاني للكمية والتالت للسعر فلازم نعمل

Select => item_name, quantity and item_price



فيه علامه الدولار لازم تشيلها عشان تعرف تجمع الأرقام طب ازاى

ار بي	<u> </u>	ليها عسال تعرف تجمع الأرقاد	ر درم ــ	رحا ،سودر				
D ~	<pre># select col item_name nad quantity df1 = df[["item_name","quantity","item_price"]] # remove dolar sign to make arithmetic operation => for in # cast for str first to can remove \$ then cast to intger df1["item_price"] = df1["item_price"].str.replace("\$"," ") df1["item_price"] = df1["item_price"].astype(float) df1</pre>							
[88]	✓ 0.0)s						
		item_name	quantity	item_price				
	0	Chips and Fresh Tomato Salsa	1	2.39				
	1	lzze	1	3.39				
	2	Nantucket Nectar	1	3.39				
	3	Chips and Tomatillo-Green Chili Salsa	1	2.39				
	4	Chicken Bowl	2	16.98				
	4617	Steak Burrito	1	11.75				
	4618	Steak Burrito	1	11.75				
	4619	Chicken Salad Bowl	1	11.25				
	4620	Chicken Salad Bowl	1	8.75				
	4621	Chicken Salad Bowl	1	8.75				

pandas.DataFrame.replace

DataFrame.replace(to_replace=None, value=_NoDefault.no_default, *,
inplace=False, limit=None, regex=False, method=_NoDefault.no_default)

بتتعامل مع النصوص بس ف لازم تحول الحاجه دي ل نص <= Replace بتعامل مع النصوص بس ف لازم تحول الحاجه دي ل نص حالت العالم المعالم المعالم

طیب احنا عاوزین مجموع کل الکمیه لکل عنصر معین و هنا هنستخدم Groupby وبعدها sum

▶ ~ [89]	<pre># sum of all quantity to know how many each item was brought df2 = df1.groupby("item_name").sum() df2.head(10) </pre> <pre> 0.0s</pre>									
		quantity	item_price							
	item_name									
	6 Pack Soft Drink	55	356.95							
	Barbacoa Bowl	66	672.36							
	Barbacoa Burrito	91	894.75							
	Barbacoa Crispy Tacos	12	120.21							
	Barbacoa Salad Bowl	10	106.40							
	Barbacoa Soft Tacos	25	250.46							
	Bottled Water	211	302.56							
	Bowl	4	29.60							
	Burrito	6	44.40							
	Canned Soda	126	137.34							

بعدها عاوزين نرتبهم علي حسب الكولوم بتاع الكميه و يكون تنازلي sort_value(By = ["col_name"], ascending =False)

```
DataFrame.sort_values(by, *, axis=0, ascending=True, inplace=False, kind='quicksort', na_position='last', ignore_index=False, key=None) #
```

Sort by quantity

[90]	<pre># sort the data frame b df3 = df2.sort_values(b df3.head(10) ✓ 0.0s</pre>			ending=Fal
		quantity	item_price	
	item_name	,		
	Chicken Bowl	761	7342.73	
	Chicken Burrito	591	5575.82	
	Chips and Guacamole	506	2201.04	
	Steak Burrito	386	3851.43	
	Canned Soft Drink	351	438.75	
	Chips	230	494.34	
	Steak Bowl	221	2260.19	
	Bottled Water	211	302.56	
	Chips and Fresh Tomato Salsa	130	361.36	
	Canned Soda	126	137.34	

Sorting by item_price

