Ex1 - Filtering and Sorting Data

This time we are going to pull data directly from the internet. Special thanks to: https://github.com/justmarkham for sharing the dataset and materials.

Step 1. Import the necessary libraries

```
import pandas as pd
```

Step 2. Import the dataset from this address.

Step 3. Assign it to a variable called chipo.

```
chipo = pd.read_csv('chipotle.tsv', sep='\t')
```

```
chipo.head()
```

	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

Các bước tiếp theo: Tạo mã bằng chipo | Xem các đồ thị được đề xuất | New interactive sheet

```
chipo.info()
```

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

Step 4. How many products cost more than $10.00?

```
chipo['item_price'] = chipo['item_price'].apply(lambda x: float(x[1:]))
```

```
products_over_10 = chipo[chipo['item_price'] > 10.00]
```

```
num_products_over_10 = products_over_10.shape[0]
```

```
print(f"Number of products costing more than $10.00: {num_products_over_10}")
```

```
Number of products costing more than $10.00: 1130
```

Step 5. What is the price of each item?

print a data frame with only two columns item_name and item_price

```
item_prices = chipo[['item_name', 'item_price']]
print(item_prices)
```

```
                                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
...                                    ...         ...
4617                        Steak Burrito       11.75
4618                        Steak Burrito       11.75
4619                   Chicken Salad Bowl       11.25
4620                   Chicken Salad Bowl        8.75
4621                   Chicken Salad Bowl        8.75

[4622 rows x 2 columns]
```

Step 6. Sort by the name of the item

```
sorted_items = chipo.sort_values(by='item_name')
print(sorted_items[['item_name', 'item_price']])
```

```
               item_name  item_price
3389  6 Pack Soft Drink       12.98
341   6 Pack Soft Drink        6.49
1849  6 Pack Soft Drink        6.49
1860  6 Pack Soft Drink        6.49
2713  6 Pack Soft Drink        6.49
...                 ...         ...
2384  Veggie Soft Tacos        8.75
781   Veggie Soft Tacos        8.75
2851  Veggie Soft Tacos        8.49
1699  Veggie Soft Tacos       11.25
1395  Veggie Soft Tacos        8.49

[4622 rows x 2 columns]
```

Step 7. What was the quantity of the most expensive item ordered?

```
most_expensive_price = chipo['item_price'].max()
```

```
most_expensive_item = chipo[chipo['item_price'] == most_expensive_price]
```

```
quantity_most_expensive = most_expensive_item['quantity'].values[0]
```

```
print(f"Quantity of the most expensive item ordered: {quantity_most_expensive}")
```

```
Quantity of the most expensive item ordered: 15
```

Step 8. How many times was a Veggie Salad Bowl ordered?

```
veggie_salad_bowl_orders = chipo[chipo['item_name'] == 'Veggie Salad Bowl']
num_veggie_salad_bowl_orders = veggie_salad_bowl_orders.shape[0]
print(f"Number of times Veggie Salad Bowl was ordered: {num_veggie_salad_bowl_orders}")
```

```
Number of times Veggie Salad Bowl was ordered: 18
```

Step 9. How many times did someone order more than one Canned Soda?

```
canned_soda_orders = chipo[(chipo['item_name'] == 'Canned Soda') & (chipo['quantity'] > 1)]
num_canned_soda_orders = canned_soda_orders.shape[0]
print(f"Number of times someone ordered more than one Canned Soda: {num_canned_soda_orders}")
```

Number of times someone ordered more than one Canned Soda: 20