Homework 1: Getting and Knowing your Data

- Step 1. Import the necessary libraries (tips: import pandas)

In [7]:
```
import pandas as pd
```

- Step 2. Import the dataset from chipotle.csv

In [8]:
```
df = pd.read_csv('chipotle.csv')

df
```

Out[8]:

	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;Ri...	$16.98
...	...	...	...	...	...
4617	1833	1	Steak Burrito	Fresh Tomato Salsa;Rice;Black Beans;Sour Cream...	$11.75
4618	1833	1	Steak Burrito	Fresh Tomato Salsa;Rice;Sour Cream;Cheese;Lett...	$11.75
4619	1834	1	Chicken Salad Bowl	Fresh Tomato Salsa;Fajita Vegetables;Pinto Bea...	$11.25
4620	1834	1	Chicken Salad Bowl	Fresh Tomato Salsa;Fajita Vegetables;Lettuce	$8.75
4621	1834	1	Chicken Salad Bowl	Fresh Tomato Salsa;Fajita Vegetables;Pinto Bea...	$8.75

4622 rows × 5 columns

- Step 3. See the first 10 entries

In [9]:
```
df.head(10)
```

Out[9]:

	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;Ri...	$16.98
5	3	1	Chicken Bowl	Fresh Tomato Salsa (Mild);Rice;Cheese;Sour Cre...	$10.98
6	3	1	Side of Chips	NaN	$1.69
7	4	1	Steak Burrito	Tomatillo Red Chili Salsa;Fajita Vegetables;Bl...	$11.75
8	4	1	Steak Soft Tacos	Tomatillo Green Chili Salsa;Pinto Beans;Cheese...	$9.25
9	5	1	Steak Burrito	Fresh Tomato Salsa;Rice;Black Beans;Pinto Bean...	$9.25

- Step 4. What is the number of observations in the dataset?

In [10]:
```
len(df)
```

Out[10]:
```
4622
```

- Step 5. Print the name of all the columns.

In [11]:
```
df.columns
```

Out[11]:
```
Index(['order_id', 'quantity', 'item_name', 'choice_description',
       'item_price'],
      dtype='object')
```

- Step 6. How many items were orderd in total? (tips: sum of quantity)

In [12]:
```
df['quantity'].sum()
```

Out[12]:
```
4972
```

- Step 7. Turn the item_price into a **float** (tips: remove '$'; convert string to floag (function: astype))

In [13]:
```
df['item_price'] = df['item_price'].str[1:].astype('float')

df['item_price']
```

Out[13]:
```
0        2.39
1        3.39
2        3.39
3        2.39
4       16.98
        ...
4617    11.75
4618    11.75
4619    11.25
4620     8.75
4621     8.75
Name: item_price, Length: 4622, dtype: float64
```

- Step 8. How much was the **revenue** for the period in the dataset? (tips: revenue = quantity * item_price)

In [14]:
```
print('Total Revenue was: $' + str((df['quantity']* df['item_price']).sum())
```

```
Total Revenue was: $39237.02
```

- Step 9. How many **different** items are sold? (tips: drop_duplicates)

In [15]:
```
len(df['item_name'].drop_duplicates())
```

Out[15]:
```
50
```

- Step 10. How many products cost more than $10.00?

In [16]:
```
len(df[df['item_price'] > 10])
```

Out[16]:
```
1130
```

- Step 11. What is the price of each "Chicken Bowl"?

In [17]:
```
df_chick_bowl = df[df['item_name'] == 'Chicken Bowl']

df_chick_bowl = df_chick_bowl[['item_name','choice_description','item_price'

df_chick_bowl.drop_duplicates()
```

Out[17]:

	item_name	choice_description	item_price
4	Chicken Bowl	Tomatillo-Red Chili Salsa (Hot);Black Beans;Ri...	16.98
5	Chicken Bowl	Fresh Tomato Salsa (Mild);Rice;Cheese;Sour Cre...	10.98
13	Chicken Bowl	Fresh Tomato Salsa;Fajita Vegetables;Rice;Chee...	11.25
19	Chicken Bowl	Tomatillo Red Chili Salsa;Fajita Vegetables;Bl...	8.75
26	Chicken Bowl	Roasted Chili Corn Salsa (Medium);Pinto Beans;...	8.49
...	...	...	...
4540	Chicken Bowl	Tomatillo Green Chili Salsa;Fajita Vegetables;...	11.25
4553	Chicken Bowl	Roasted Chili Corn Salsa;Black Beans;Sour Crea...	11.25
4589	Chicken Bowl	Fresh Tomato Salsa;Rice;Black Beans;Sour Cream...	11.25
4595	Chicken Bowl	Tomatillo Green Chili Salsa;Rice;Black Beans	8.75
4599	Chicken Bowl	Roasted Chili Corn Salsa;Cheese;Lettuce	8.75

358 rows × 3 columns

- Step 12. Sort by the item_price of "Chicken Bowl"

In [18]:
```
df_chick_bowl.sort_values(by='item_price')
```

Out[18]:

	item_name	choice_description	item_price
2918	Chicken Bowl	Fresh Tomato (Mild);Rice;Sour Cream;Cheese	8.19
3664	Chicken Bowl	Fresh Tomato (Mild);Lettuce;Fajita Veggies;Pin...	8.19
3101	Chicken Bowl	Fresh Tomato (Mild);Lettuce;Fajita Veggies;Bla...	8.19
2548	Chicken Bowl	Tomatillo-Red Chili Salsa (Hot);Rice;Black Bea...	8.49
3424	Chicken Bowl	Roasted Chili Corn Salsa (Medium);Tomatillo-Gr...	8.49
...	...	...	...
4423	Chicken Bowl	Fresh Tomato Salsa;Rice;Fajita Vegetables;Blac...	22.50
1429	Chicken Bowl	Fresh Tomato Salsa;Rice;Sour Cream;Guacamole	22.50
2510	Chicken Bowl	Roasted Chili Corn Salsa;Rice;Black Beans;Chee...	22.50
1514	Chicken Bowl	Fresh Tomato Salsa;Rice;Black Beans;Cheese;Sou...	26.25
409	Chicken Bowl	Fresh Tomato Salsa (Mild);Tomatillo-Green Chil...	32.94

726 rows × 3 columns

- Step 13. What is the most expensive item ordered?

In [19]:
```
df.sort_values(by = "item_price", ascending = False).head(1)
```

Out[19]:

	order_id	quantity	item_name	choice_description	item_price
3598	1443	15	Chips and Fresh Tomato Salsa	NaN	44.25

- Step 14. How many times was a "Steak Burrito" ordered?

In [20]:
```
len(df[df['item_name'] == "Steak Burrito"])
```

Out[20]:
```
368
```

- Step 15. How many times did someone order more than one Canned Soda?

In [155…
```
len(df[(df['item_name'] == "Canned Soda") & (df['quantity'] > 1)])
```

Out[155]:
```
20
```