Getting and Knowing your Data

Step 1. Import the necessary libraries

```
import pandas as pd
import numpy as np
```

Step 2. Import the dataset from this address.

```
data = pd.read_csv('/content/chipotle.csv')
```

Step 3. Assign it to a variable called chipo.

```
chipo = pd.read_csv('chipotle.csv')
```

Step 4. See the first 10 entries

```
chipo.head(10)
```

	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
5	3	1	Chicken Bowl	[Fresh Tomato Salsa (Mild), [Rice, Cheese, Sou...	$10.98
6	3	1	Side of Chips	NaN	$1.69
7	4	1	Steak Burrito	[Tomatillo Red Chili Salsa, [Fajita Vegetables...	$11.75
8	4	1	Steak Soft Tacos	[Tomatillo Green Chili Salsa, [Pinto Beans, Ch...	$9.25
9	5	1	Steak Burrito	[Fresh Tomato Salsa, [Rice, Black Beans, Pinto...	$9.25

Next steps: View recommended plots New interactive sheet

Step 5. What is the number of observations in the dataset?

```
chipo.shape    # rows and columns
```

```
(4622, 5)
```

Step 6. What is the number of columns in the dataset?

```
chipo.shape[1]
```

```
5
```

Step 7. Print the name of all the columns.

```
chipo.columns
```

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

Step 8. How is the dataset indexed?

```
chipo.index
```

```
RangeIndex(start=0, stop=4622, step=1)
```

Step 9. Which was the most-ordered item?

```
chipo['item_name'].value_counts().head(1)
```

item_name	count
Chicken Bowl	726

dtype: int64

Step 10. For the most-ordered item, how many items it was ordered?

```
chipo['quantity'][chipo['item_name'] == 'Chicken Bowl'].sum()
```

```
761
```

Step 11. What was the most ordered item in the choice_description column?

```
chipo['choice_description'].value_counts().head(1)
```

choice_description	count
[Diet Coke]	134

dtype: int64

Step 12. How many items were orderd in total?

```
chipo['order_id'].count()
```

```
4622
```

Step 13. Turn the item price into a float

```
chipo.dtypes
```

	0
order_id	int64
quantity	int64
item_name	object
choice_description	object
item_price	object

dtype: object

```
def price_converter(x):
#     print(x)
    x = x.strip('$')
#     print(x)
    x = float(x)
    return x
```

```
chipo['item_price'] = chipo['item_price'].apply(price_converter)
```

```
chipo.dtypes
```

	0
order_id	int64
quantity	int64
item_name	object

```
data['item_price'].head()
```

	item_price
0	$2.39
1	$3.39
2	$3.39
3	$2.39
4	$16.98

dtype: object

Step 14. How much was the revenue for the period in the dataset?

```
revenue = chipo['item_price'].sum()
revenue
```

```
34500.16
```

Step 15. How many orders were made in the period?

```
total_orders = chipo['order_id'].count()
total_orders
```

```
4622
```

Step 16. What is the average revenue amount per order?

```
revenue/total_orders
```

```
7.464335785374297
```

Step 17. How many different items are sold?

```
chipo['item_name'].nunique()
```

```
50
```