Ex2 - Getting and Knowing your 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

In [1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

Step 2. Import the dataset from this address.

Step 3. Assign it to a variable called chipo.

In [2]:
```
chipo = pd.read_csv("chipotle.tsv", sep='\t')
```

Step 4. See the first 10 entries

In [3]:
```
chipo.head(10)
```

Out[3]:

	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

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

In [4]:
```
# Solution 1
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
```

In [5]:
```
# Solution 2
chipo.shape
```

Out[5]:
```
(4622, 5)
```

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

In [6]:
```
chipo.columns.nunique()
```

Out[6]:
```
5
```

In [7]:
```
chipo.shape
```

Out[7]:
```
(4622, 5)
```

Step 7. Print the name of all the columns.

In [8]:
```
chipo.columns
```

Out[8]:
```
Index(['order_id', 'quantity', 'item_name', 'choice_description',
       'item_price'],
      dtype='object')
```

Step 8. How is the dataset indexed?

In [9]:
```
chipo.index
```

Out[9]:
```
RangeIndex(start=0, stop=4622, step=1)
```

Step 9. Which was the most-ordered item?

In [10]:
```
chipo.groupby("item_name").sum().sort_values(by="quantity", ascending=False).head(1)
```

Out[10]:

	order_id	quantity
item_name		
Chicken Bowl	713926	761

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

In [11]:
```
chipo.groupby("item_name").sum().sort_values(by="quantity", ascending=False).head(1)
```

Out[11]:

	order_id	quantity
item_name		
Chicken Bowl	713926	761

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

In [12]:
```
chipo.head(2)
```

Out[12]:

	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

In [13]:
```
chipo.groupby("choice_description").sum().sort_values(by="quantity", ascending=False).head(1)
```

Out[13]:

	order_id	quantity
choice_description		
[Diet Coke]	123455	159

Step 12. How many items were orderd in total?

In [14]:
```
chipo["quantity"].sum()
```

Out[14]:
```
4972
```

Step 13. Turn the item price into a float

Step 13.a. Check the item price type

In [15]:
```
chipo["item_price"].dtype
```

Out[15]:
```
dtype('O')
```

Step 13.b. Create a lambda function and change the type of item price

In [16]:
```
chipo.dropna()
chipo["item_price"] = chipo["item_price"].apply(lambda x: float(x[1:-1]))
```

Step 13.c. Check the item price type

In [17]:
```
chipo["item_price"].dtype
```

Out[17]:
```
dtype('float64')
```

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

In [18]:
```
chipo.head(2)
```

Out[18]:

	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

In [19]:
```
(chipo["quantity"] * chipo["item_price"]).sum()
```

Out[19]:
```
39237.02
```

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

In [20]:
```
chipo["order_id"].value_counts().count()
```

Out[20]:
```
1834
```

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

In [21]:
```
chipo['revenue'] = chipo['quantity'] * chipo['item_price']
chipo.groupby(by=['order_id']).sum().mean()['revenue']
```

Out[21]:
```
21.394231188658654
```

Step 17. How many different items are sold?

In [22]:
```
chipo.item_name.value_counts()
```

Out[22]:
```
Chicken Bowl                                726
Chicken Burrito                             553
Chips and Guacamole                         479
Steak Burrito                               368
Canned Soft Drink                           301
Chips                                       211
Steak Bowl                                  211
Bottled Water                               162
Chicken Soft Tacos                          115
Chips and Fresh Tomato Salsa                110
Chicken Salad Bowl                          110
Canned Soda                                 104
Side of Chips                               101
Veggie Burrito                               95
Barbacoa Burrito                             91
Veggie Bowl                                  85
Carnitas Bowl                                68
Barbacoa Bowl                                66
Carnitas Burrito                             59
Steak Soft Tacos                             55
6 Pack Soft Drink                            54
Chips and Tomatillo Red Chili Salsa          48
Chicken Crispy Tacos                         47
Chips and Tomatillo Green Chili Salsa        43
Carnitas Soft Tacos                          40
Steak Crispy Tacos                           35
Chips and Tomatillo-Green Chili Salsa        31
Steak Salad Bowl                             29
Nantucket Nectar                             27
Barbacoa Soft Tacos                          25
Chips and Roasted Chili Corn Salsa           22
Chips and Tomatillo-Red Chili Salsa          20
Izze                                         20
Veggie Salad Bowl                            18
Chips and Roasted Chili-Corn Salsa           18
Barbacoa Crispy Tacos                        11
Barbacoa Salad Bowl                          10
Chicken Salad                                 9
Veggie Soft Tacos                             7
Carnitas Crispy Tacos                         7
Burrito                                       6
Veggie Salad                                  6
Carnitas Salad Bowl                           6
Steak Salad                                   4
Salad                                         2
Bowl                                          2
Crispy Tacos                                  2
Veggie Crispy Tacos                           1
Chips and Mild Fresh Tomato Salsa             1
Carnitas Salad                                1
Name: item_name, dtype: int64
```

In [23]:
```
chipo.item_name.value_counts().count()
```

Out[23]:
```
50
```