In [265]:
```
# Prince Ogwu
#Introduction To Pandas Week1

```

In [ ]:
```
Questions1. Which was the most-ordered item?
Questions2. For the most-ordered item, how many items were ordered?
Questions3. What was the most ordered item in the choice_description colum
Questions4. How many items were ordered in total?
Questions5. Turn the item price into a float
Questions6. How much was the revenue for the period in the dataset?
Questions7. How many orders were made in the period?
Questions8. What is the average revenue amount per order?
```

In [ ]:
```

```

In [183]:
```
%matplotlib inline
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
from scipy import stats

print('Sucesfully imported neccessary library')
```

```
Sucesfully imported neccessary library
```

In [ ]:
```

```

url= 'https://raw.githubusercontent.com/justmarkham/DAT8/master/data/chipotle.tsv' (https://raw.githubusercontent.com/justmarkham/DAT8/master/data/chipotle.tsv') chipo = pd.read_csv(url, sep = '\t')

In [185]:
```
# Reading data from remote link
url = "https://bit.ly/2UUomN0"
chipo = pd.read_csv(url, sep = '\t')
```

In [196]:
```
#Get information about our dataFrame
```

```
<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   float64
dtypes: float64(1), int64(2), object(2)
memory usage: 180.7+ KB
```

In [39]:
```
#View top10 rows in our dataframe
```

Out[39]:

	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

In [197]:
```
#Checking for duplicates
```

Out[197]: 59

In [198]:
```
#There are 59 Duplicate rows which would be dropped.
chipo.drop_duplicates(keep= 'first', inplace=True)
```

In [199]:
```

```

Out[199]: (4563, 5)

In [201]:
```
#Chcking our data for missing value
```

Out[201]:
```
order_id                0
quantity                0
item_name               0
choice_description    1228
item_price              0
dtype: int64
```

```
## We observe that choice_description has a total of 1228 missing values
```

In [202]:
```
#To continue with this analysis, we have to replace the NaN values with No
chipo.fillna('Not Available', inplace=True)


```

Questions1. Which was the most-ordered item?

In [203]:

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

In [204]:

```
0    Chicken Bowl
Name: item_name, dtype: object
```

Chicken Bowl was the most ordered item

In [ ]:

In [73]:

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

In [115]:

In [113]:

```
#Most ordered items placed.
chipo.groupby(['item_name'])['quantity'].sum().nlargest()
```

Out[113]:

```
item_name
Chicken Bowl           752
Chicken Burrito        584
Chips and Guacamole    501
Steak Burrito          383
Canned Soft Drink      340
Name: quantity, dtype: int64
```

Chicken Bowl was the most ordered item and a total of 752 items were placed.

In [ ]:

In [263]:
```
#Most ordered items placed.

chipo.groupby(['choice_description'])['quantity'].count().nlargest()
```

Out[263]:
```
choice_description
Not Available                                                  122
8
[Diet Coke]                                                     13
3
[Coke]                                                          11
5
[Sprite]                                                         7
7
[Fresh Tomato Salsa, [Rice, Black Beans, Cheese, Sour Cream, Lettuce]]      4
1
Name: quantity, dtype: int64
```

In [ ]:

In [ ]:

In [ ]:

In [233]:

Out[233]:
```
4913
```

In [ ]:

In [ ]:
```
  """ #convert points column from object to float
chipo['item_price'] = chipo['item_price'].astype(float)

#view updated DataFrame
```

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

In [235]:

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 4563 entries, 0 to 4621
Data columns (total 5 columns):
 #   Column              Non-Null Count  Dtype
---  ------              --------------  -----
 0   order_id            4563 non-null   int64
 1   quantity            4563 non-null   int64
 2   item_name           4563 non-null   object
 3   choice_description  4563 non-null   object
 4   item_price          4563 non-null   float64
dtypes: float64(1), int64(2), object(2)
memory usage: 213.9+ KB
```

In [ ]:
```

```

In [ ]:

```
```

In [221]:

```
revenue= chipo.item_price.sum()

```

```
Total revenue generated is 34177.25
```

In [ ]:

```
```

In [ ]:

```
```

In [220]:

```
orders_placed= len(chipo.item_name)

```

```
Total orders made in the period is 4563
```

In [ ]:

```
```

In [ ]:

```
```

In [218]:

```
average_amount_per_order = revenue/orders_placed
```

Out[218]:

```
7.490083278544817
```

In [ ]:

```
```

In [143]:

```
 chipo.item_name.nunique()
```

Out[143]:

```
50
```

In [ ]:

```
```