Q1. Import the necessary libraries

In [2]:
```
import pandas as pd
```

Q 2. Read Data from tsv file

In [73]:
```
file= pd.read_csv("c.csv")
file
file.describe()
```

Out[73]:

	fd	quantity
count	4622.000000	4622.000000
mean	927.254868	1.075725
std	528.890796	0.410186
min	1.000000	1.000000
25%	477.250000	1.000000
50%	926.000000	1.000000
75%	1393.000000	1.000000
max	1834.000000	15.000000

In [98]:
```
file.fillna("nothing",inplace=True)
file
```

Out[98]:

	fd	quantity	item_name	choice_description	item_price
1	1	1	Izze	[Clementine]	$3.39
2	1	1	Nantucket Nectar	[Apple]	$3.39
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
7	4	1	Steak Burrito	[Tomatillo Red Chili Salsa, [Fajita Vegetables...	$11.75
...	...	...	...	...	...
4617	1833	1	Steak Burrito	[Fresh Tomato Salsa, [Rice, Black Beans, Sour ...	$11.75
4618	1833	1	Steak Burrito	[Fresh Tomato Salsa, [Rice, Sour Cream, Cheese...	$11.75
4619	1834	1	Chicken Salad Bowl	[Fresh Tomato Salsa, [Fajita Vegetables, Pinto...	$11.25
4620	1834	1	Chicken Salad Bowl	[Fresh Tomato Salsa, [Fajita Vegetables, Lettu...	$8.75
4621	1834	1	Chicken Salad Bowl	[Fresh Tomato Salsa, [Fajita Vegetables, Pinto...	$8.75

3376 rows × 5 columns

Q3. See the first 10 entries

In [99]:
```
file[0:10]
```

Out[99]:

	fd	quantity	item_name	choice_description	item_price
1	1	1	Izze	[Clementine]	$3.39
2	1	1	Nantucket Nectar	[Apple]	$3.39
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
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
11	6	1	Chicken Crispy Tacos	[Roasted Chili Corn Salsa, [Fajita Vegetables,...	$8.75
12	6	1	Chicken Soft Tacos	[Roasted Chili Corn Salsa, [Rice, Black Beans,...	$8.75
13	7	1	Chicken Bowl	[Fresh Tomato Salsa, [Fajita Vegetables, Rice,...	$11.25

Q4. What is the information in the dataset?

In [100]:

```
file.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 3376 entries, 1 to 4621
Data columns (total 5 columns):
 #   Column              Non-Null Count  Dtype 
---  ------              --------------  ----- 
 0   fd                  3376 non-null   int64 
 1   quantity            3376 non-null   int64 
 2   item_name           3376 non-null   object
 3   choice_description  3376 non-null   object
 4   item_price          3376 non-null   object
dtypes: int64(2), object(3)
memory usage: 287.3+ KB
```

Q5. How many columns in the dataset?

In [101]:

```
file.shape
print("total columns are ",file.shape[1])
```

```
total columns are  5
```

Q6. Print the name of all the columns.

In [102]:

```
 file.columns
```

Out[102]:

```
Index(['fd', 'quantity', 'item_name', 'choice_description', 'item_price'], dt
ype='object')
```

Q7. How is the dataset indexed?

In [103]:

```
file.index
```

Out[103]:

```
Int64Index([   1,    2,    4,    5,    7,    8,    9,   11,   12,   13,
            ...
            4609, 4610, 4611, 4612, 4615, 4617, 4618, 4619, 4620, 4621],
           dtype='int64', length=3376)
```

Q8. Which was the most-ordered item?

```
In [104]: grouped = file.groupby('item_name')['quantity'].sum()

          most_ordered_item = grouped.idxmax()

          most_ordered_item
```

```
Out[104]: 'Chicken Bowl'
```

Q9. What was the most ordered item in the choice_description column?

```
In [105]: grouped = file.groupby('choice_description')['quantity'].sum()

          most_ordered_item = grouped.idxmax()

          most_ordered_item
```

```
Out[105]: '[Diet Coke]'
```

Q10. How many items were orderd in total?

```
In [106]: file['quantity'].sum()
```

```
Out[106]: 3590
```

Q11. Check the item price type

```
In [107]: file.item_price.dtype
```

```
Out[107]: dtype('O')
```

Q12. Change the data type of item price column

In [121]:
```
file['item_price'].astype('str')
```

Out[121]:
```
1        $3.39
2        $3.39
4       $16.98
5       $10.98
7       $11.75
         ...
4617    $11.75
4618    $11.75
4619    $11.25
4620     $8.75
4621     $8.75
Name: item_price, Length: 3376, dtype: object
```

Q13. How much revenue was generated for the period in the dataset?

In [108]:
```
count = (file['item_price'].str.startswith('$')).sum()
count
```

Out[108]:
```
3376
```

Q14. What is the average revenue amount per order?

In [122]:
```


total_revenue = file['item_price'].sum()
total_orders = file['quantity'].count()

# average_revenue_per_order = total_revenue / total_orders

print('The average revenue amount per order is:', average_revenue_per_order)


```

```
---------------------------------------------------------------------------
TypeError                                 Traceback (most recent call last)
Input In [122], in <cell line: 4>()
      1 total_revenue = file['item_price'].sum()
      2 total_orders = file['quantity'].count()
----> 4 average_revenue_per_order = total_revenue / total_orders
      6 print('The average revenue amount per order is:', average_revenue_per
_order)

TypeError: ufunc 'true_divide' not supported for the input types, and the inp
uts could not be safely coerced to any supported types according to the casti
ng rule ''safe''
```

Q15. How many different items are sold?

In [110]:

```
t=file['item_name'].nunique()
print('total items sold are',t)
```

```
total items sold are 38
```