Darshan
UNIVERSITY
योग: कर्मसु कौशलम्

Data Mining

Lab - 4

Name: Smit Maru

Enrollment No: 23010101161

Step 1. Import the necessary libraries

In [2]:

```
import pandas as pd
```

Step 2. Import the dataset from this address.

In [3]:

```
url = "https://raw.githubusercontent.com/justmarkham/DAT8/master/data/chipotle.tsv"
data = pd.read_csv(url , sep="\t")
print(data)
```

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js

```
      order_id  quantity                              item_name  \
0            1         1           Chips and Fresh Tomato Salsa   
1            1         1                                   Izze   
2            1         1                       Nantucket Nectar   
3            1         1  Chips and Tomatillo-Green Chili Salsa   
4            2         2                           Chicken Bowl   
...        ...       ...                                    ...   
4617      1833         1                          Steak Burrito   
4618      1833         1                          Steak Burrito   
4619      1834         1                     Chicken Salad Bowl   
4620      1834         1                     Chicken Salad Bowl   
4621      1834         1                     Chicken Salad Bowl   

                                     choice_description item_price  
0                                                   NaN     $2.39   
1                                          [Clementine]     $3.39   
2                                               [Apple]     $3.39   
3                                                   NaN     $2.39   
4     [Tomatillo-Red Chili Salsa (Hot), [Black Beans...    $16.98   
...                                                 ...        ...  
4617  [Fresh Tomato Salsa, [Rice, Black Beans, Sour ...    $11.75   
4618  [Fresh Tomato Salsa, [Rice, Sour Cream, Cheese...    $11.75   
4619  [Fresh Tomato Salsa, [Fajita Vegetables, Pinto...    $11.25   
4620  [Fresh Tomato Salsa, [Fajita Vegetables, Lettu...     $8.75   
4621  [Fresh Tomato Salsa, [Fajita Vegetables, Pinto...     $8.75   

[4622 rows x 5 columns]
```

Step 3. Assign it to a variable called chipo.

In [4]:
```python
chipo = data
```

Step 4. See the first 10 entries

In [5]:
```python
print(chipo.head(10))
```

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js

```
   order_id  quantity                              item_name  \
0         1         1           Chips and Fresh Tomato Salsa   
1         1         1                                   Izze   
2         1         1                       Nantucket Nectar   
3         1         1  Chips and Tomatillo-Green Chili Salsa   
4         2         2                           Chicken Bowl   
5         3         1                           Chicken Bowl   
6         3         1                          Side of Chips   
7         4         1                          Steak Burrito   
8         4         1                       Steak Soft Tacos   
9         5         1                          Steak Burrito   

                                  choice_description item_price  
0                                                NaN     $2.39   
1                                       [Clementine]     $3.39   
2                                            [Apple]     $3.39   
3                                                NaN     $2.39   
4  [Tomatillo-Red Chili Salsa (Hot), [Black Beans...    $16.98   
5  [Fresh Tomato Salsa (Mild), [Rice, Cheese, Sou...    $10.98   
6                                                NaN     $1.69   
7  [Tomatillo Red Chili Salsa, [Fajita Vegetables...    $11.75   
8  [Tomatillo Green Chili Salsa, [Pinto Beans, Ch...     $9.25   
9  [Fresh Tomato Salsa, [Rice, Black Beans, Pinto...     $9.25   
```

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

In [6]:

```
# Solution 1
temp = chipo.shape
print(temp[0])
```

```
4622
```

In [7]:

```
# Solution 2
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
```

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

In [8]:

```
temp = chipo.shape
print(temp[1])
```

```
5
```

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js

Step 7. Print the name of all the columns.

In [9]:

```
chipo.columns
```

Out[9]:

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

Step 8. How is the dataset indexed?

In [10]:

```
chipo.index
```

Out[10]:

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

Step 9. Number of Unique Items ?

In [11]:

```
data["item_name"].nunique()
```

Out[11]:

```
50
```

Step 10. Which was the most-ordered item?

In [12]:

```
data.groupby('item_name').sum(numeric_only=True).sort_values('quantity', ascending
```

Out[12]:

	order_id	quantity
item_name		
Chicken Bowl	713926	761

Step 11. How many items were orderd in total?

In [13]:

```
data['quantity'].sum()
```

Out[13]:

```
4972
```

Step 12. Turn the item price into a float

12.a Check the item price type

In [14]:

```
data['item_price'].dtype
```

Out[14]:

```
dtype('O')
```

Step 12.b. Create a lambda function and change the type of item price

In [15]:

```
data['item_price'] = data['item_price'].apply(lambda x : float(x.replace("$","")))
```

Step 12.c. Check the item price type

In [16]:

```
data['item_price'].dtype
```

Out[16]:

```
dtype('float64')
```

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

In [17]:

```
revenue = (data['item_price']*data['quantity']).sum()
print(revenue)
print(revenue.dtype)
```

```
39237.02
float64
```

Step 15. How many orders were made ?

In [18]:

```
data['order_id'].nunique()
```

Out[18]:

```
1834
```

Step 17. How many different choice descriptions are there?

In [19]:

```
data['choice_description'].nunique()
```

Out[19]:

```
1043
```

Step 18. What items have been ordered more than 100 times?

In [20]:

```
most_orderd = data.groupby('item_name')['quantity'].sum()
print(most_orderd.count())
most_orderd = most_orderd[most_orderd > 100]
print(most_orderd)
```

```
50
item_name
Bottled Water                   211
Canned Soda                     126
Canned Soft Drink               351
Chicken Bowl                    761
Chicken Burrito                 591
Chicken Salad Bowl              123
Chicken Soft Tacos              120
Chips                           230
Chips and Fresh Tomato Salsa    130
Chips and Guacamole             506
Side of Chips                   110
Steak Bowl                      221
Steak Burrito                   386
Name: quantity, dtype: int64
```

Step 19. What is the average revenue amount per order?

In [21]:

```python
# Solution 1
temp = data['order_id'].nunique()
print(revenue/temp)
```

```
21.39423118865867
```