Weekly Pandas Challenge - Data In Motion

Name : MOHAMMED WASIM R D | Linkedin : https://www.linkedin.com/in/mdwasimrd

Dataset Link :
https://raw.githubusercontent.com/justmarkham/DAT8/master/data/chipotle.tsv

In [155…

```
#Importing the dataset

import pandas as pd
chipotle = pd.read_csv( "https://raw.githubusercontent.com/justmarkham/DAT8/m
chipotle.head(5)
```

Out[155…

	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

In [156…

```
df = pd.DataFrame(chipotle)
print(df)
```

```
      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]
```

In [157…

```
df.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 [158…

```
#Dimensions of the dataset

len(df)
```

Out[158…

```
4622
```

1) Which was the most-ordered item?

In [159…

```
most_ordered = df.groupby(['item_name']).sum()
item = most_ordered.sort_values(['quantity'], ascending = False)
item.head(10)
```

Out[159…

item_name	order_id	quantity
Chicken Bowl	713926	761
Chicken Burrito	497303	591
Chips and Guacamole	449959	506
Steak Burrito	328437	386
Canned Soft Drink	304753	351
Chips	208004	230
Steak Bowl	193752	221
Bottled Water	175944	211
Chips and Fresh Tomato Salsa	100419	130
Canned Soda	76396	126

The most ordered item was Chicken Bowl

2) For the most-ordered item, how many items were ordered?

As we can see from the above output that Chicken Bowl was ordered 761 times.

3) What was the most ordered item in the choice_description column?

In [160…

```
cd_item = df.groupby(['choice_description']).sum()
item1 = cd_item.sort_values(['quantity'], ascending = False)
print(item1)
```

choice_description	order_id	quantity
[Diet Coke]	123455	159
[Coke]	122752	143
[Sprite]	80426	89
[Fresh Tomato Salsa, [Rice, Black Beans, Cheese...	43088	49
[Fresh Tomato Salsa, [Rice, Black Beans, Cheese...	36041	42
...	...	...
[Roasted Chili Corn Salsa, [Fajita Vegetables, ...	577	1
[Roasted Chili Corn Salsa, [Fajita Vegetables, ...	585	1
[Roasted Chili Corn Salsa, [Fajita Vegetables, ...	235	1
[Roasted Chili Corn Salsa, [Guacamole, Sour Cre...	987	1
[[Tomatillo-Red Chili Salsa (Hot), Tomatillo-Gr...	1299	1

[1043 rows x 2 columns]

Diet Coke was the most ordered item in choice_description column.

4) How many items were orderd in total?

In [161…

```
total_items=df.quantity.sum()
total_items
```

Out[161…

```
4972
```

There were total 4972 items ordered in total

5) Turn the item price into a float

In [162…

```
new_chipo=df
new_chipo.iloc[:, 4].replace("\$", "", inplace=True, regex=True)
new_chipo.head(10)
```

Out[162…

	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

Comparing both Chipotle and New_chipo we can see that item_price has been changed to float

6) How much was the revenue for the period in the dataset?

In [163…

```
df['order_value'] = quantity* item_price
df[:10]
```

Out[163…

	order_id	quantity	item_name	choice_description	item_price	order_value
0	1	1	Chips and Fresh Tomato Salsa	NaN	2.39	2.39
1	1	1	Izze	[Clementine]	3.39	3.39
2	1	1	Nantucket Nectar	[Apple]	3.39	3.39
3	1	1	Chips and Tomatillo-Green Chili Salsa	NaN	2.39	2.39
4	2	2	Chicken Bowl	[Tomatillo-Red Chili Salsa (Hot), [Black Beans...	16.98	33.96
5	3	1	Chicken Bowl	[Fresh Tomato Salsa (Mild), [Rice, Cheese, Sou...	10.98	10.98
6	3	1	Side of Chips	NaN	1.69	1.69
7	4	1	Steak Burrito	[Tomatillo Red Chili Salsa, [Fajita Vegetables...	11.75	11.75
8	4	1	Steak Soft Tacos	[Tomatillo Green Chili Salsa, [Pinto Beans, Ch...	9.25	9.25
9	5	1	Steak Burrito	[Fresh Tomato Salsa, [Rice, Black Beans, Pinto...	9.25	9.25

In [164…

```
df2 = df['order_value'].sum()
df2
```

Out[164…

```
39237.02
```

7) How many orders were made in the period?

In [165…

```
orders = chipotle.order_id.value_counts().count()
orders
```

Out[165…

```
1834
```

8) What is the average revenue amount per order?

In [166…
```
avg_amt = df.groupby(['order_id']).sum()
print(avg_amt)
```

```
          quantity  order_value
order_id
1                4        11.56
2                2        33.96
3                2        12.67
4                2        21.00
5                2        13.70
...            ...          ...
1830             2        23.00
1831             3        12.90
1832             2        13.20
1833             2        23.50
1834             3        28.75

[1834 rows x 2 columns]
```

In [167…
```
avg_amt.mean()
```

Out[167…
```
quantity        2.711014
order_value    21.394231
dtype: float64
```

From the above results we can see that the average order value is 21.394231

9) How many different items are sold?

In [168…
```
diff_items = df.item_name.value_counts().count()
diff_items
```

Out[168…
```
50
```

In [ ]:

In [ ]: