

## week3-project2-1

January 2, 2024

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

```
[2]: df = pd.read_csv("C:\\Users\\Dipen\\Downloads\\chipotle.csv")
```

```
[3]: df
```

```
[3]:
```

	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]

```
[4]: df.head(20)#gives first 20 values
```

```
[4]:
```

	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
10	5	1	Chips and Guacamole
11	6	1	Chicken Crispy Tacos
12	6	1	Chicken Soft Tacos
13	7	1	Chicken Bowl
14	7	1	Chips and Guacamole
15	8	1	Chips and Tomatillo-Green Chili Salsa
16	8	1	Chicken Burrito
17	9	1	Chicken Burrito
18	9	2	Canned Soda
19	10	1	Chicken 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
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
10	NaN	\$4.45
11	[Roasted Chili Corn Salsa, [Fajita Vegetables,...	\$8.75
12	[Roasted Chili Corn Salsa, [Rice, Black Beans,...	\$8.75
13	[Fresh Tomato Salsa, [Fajita Vegetables, Rice,...	\$11.25
14	NaN	\$4.45
15	NaN	\$2.39
16	[Tomatillo-Green Chili Salsa (Medium), [Pinto ...	\$8.49
17	[Fresh Tomato Salsa (Mild), [Black Beans, Rice...	\$8.49
18	[Sprite]	\$2.18
19	[Tomatillo Red Chili Salsa, [Fajita Vegetables...	\$8.75

```
[5]: df.tail(20)#gives last 20 values
```

```
[5]:
```

	order_id	quantity	item_name \
4602	1827	1	Barbacoa Burrito
4603	1827	1	Barbacoa Burrito
4604	1828	1	Chicken Bowl
4605	1828	1	Chips and Guacamole
4606	1828	1	Canned Soft Drink
4607	1829	1	Steak Burrito
4608	1829	1	Veggie Burrito
4609	1829	1	Canned Soft Drink
4610	1830	1	Steak Burrito
4611	1830	1	Veggie Burrito
4612	1831	1	Carnitas Bowl
4613	1831	1	Chips
4614	1831	1	Bottled Water
4615	1832	1	Chicken Soft Tacos
4616	1832	1	Chips and Guacamole
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
4602	[Tomatillo Green Chili Salsa]	\$9.25
4603	[Tomatillo Green Chili Salsa]	\$9.25
4604	[Fresh Tomato Salsa, [Rice, Black Beans, Chees...	\$8.75
4605	NaN	\$4.45
4606	[Coke]	\$1.25
4607	[Tomatillo Green Chili Salsa, [Rice, Cheese, S...	\$11.75
4608	[Tomatillo Red Chili Salsa, [Fajita Vegetables...	\$11.25
4609	[Sprite]	\$1.25
4610	[Fresh Tomato Salsa, [Rice, Sour Cream, Cheese...	\$11.75
4611	[Tomatillo Green Chili Salsa, [Rice, Fajita Ve...	\$11.25
4612	[Fresh Tomato Salsa, [Fajita Vegetables, Rice,...	\$9.25
4613	NaN	\$2.15
4614	NaN	\$1.50
4615	[Fresh Tomato Salsa, [Rice, Cheese, Sour Cream]]	\$8.75
4616	NaN	\$4.45
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

```
[6]: df.shape #is a tuple of array dimensions that tells the number of rows and
         ↪ columns of a given DataFrame.
```

```
[6]: (4622, 5)
```

```
[7]: df.columns #gives the columns name from the given dataset
```

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

```
[8]: df.isnull().sum() # returns the number of missing values from the given dataset.
```

```
[8]: order_id          0  
     quantity         0  
     item_name        0  
     choice_description  1246  
     item_price        0  
     dtype: int64
```

```
[9]: #Finding missing value  
     missing_values = df.isnull()
```

```
[10]: print(missing_values)
```

	order_id	quantity	item_name	choice_description	item_price
0	False	False	False	True	False
1	False	False	False	False	False
2	False	False	False	False	False
3	False	False	False	True	False
4	False	False	False	False	False
...	...	...	...	...	...
4617	False	False	False	False	False
4618	False	False	False	False	False
4619	False	False	False	False	False
4620	False	False	False	False	False
4621	False	False	False	False	False

```
[4622 rows x 5 columns]
```

```
[11]: #Tackeling missing values  
     df_cleaned = df.dropna(inplace=True)  
     print(df_cleaned)
```

```
None
```

```
[12]: print(df)
```

	order_id	quantity	item_name \
1	1	1	Izze
2	1	1	Nantucket Nectar

4	2	2	Chicken Bowl
5	3	1	Chicken Bowl
7	4	1	Steak Burrito
...	...	...	...
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
1	[Clementine]	\$3.39
2	[Apple]	\$3.39
4	[Tomatillo-Red Chili Salsa (Hot), [Black Beans...	\$16.98
5	[Fresh Tomato Salsa (Mild), [Rice, Cheese, Sou...	\$10.98
7	[Tomatillo Red Chili Salsa, [Fajita Vegetables...	\$11.75
...	...	...
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

[3376 rows x 5 columns]

```
[13]: # Remove non-numeric characters and convert to numeric
df['item_price'] = pd.to_numeric(df['item_price'].str.replace('[^\d.]', ''),
                                errors='coerce')
```

C:\Users\Dipen\AppData\Local\Temp\ipykernel\_22628\2375342477.py:2:

FutureWarning: The default value of regex will change from True to False in a future version.

```
df['item_price'] = pd.to_numeric(df['item_price'].str.replace('[^\d.]', ''),
                                errors='coerce')
```

```
[14]: df
```

```
[14]:
```

	order_id	quantity	item_name \
1	1	1	Izze
2	1	1	Nantucket Nectar
4	2	2	Chicken Bowl
5	3	1	Chicken Bowl
7	4	1	Steak Burrito
...	...	...	...
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
1	[Clementine]	3.39
2	[Apple]	3.39
4	[Tomatillo-Red Chili Salsa (Hot), [Black Beans...	16.98
5	[Fresh Tomato Salsa (Mild), [Rice, Cheese, Sou...	10.98
7	[Tomatillo Red Chili Salsa, [Fajita Vegetables...	11.75
...	...	...
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

[3376 rows x 5 columns]

```
[15]: # Display the data types of each column
print(df.dtypes)
```

```
order_id          int64
quantity          int64
item_name         object
choice_description object
item_price        float64
dtype: object
```

```
[16]: #identify the duplicates entries in each column
```

```
[17]: duplicates=df[df.duplicated()]
```

```
[18]: print(duplicates)
```

	order_id	quantity	item_name \
238	103	1	Steak Burrito
248	108	1	Canned Soda
297	129	1	Steak Burrito
381	165	1	Canned Soft Drink
484	205	1	Chicken Bowl
567	233	1	Canned Soft Drink
684	282	1	Canned Soft Drink
773	319	1	Chicken Bowl
908	376	1	Steak Burrito
924	381	1	Chicken Soft Tacos
1117	462	1	Chicken Burrito
1203	491	1	Chicken Burrito

1245	507	1	Canned Soft Drink
1381	561	1	Canned Soft Drink
1420	576	1	Barbacoa Bowl
1421	576	1	Barbacoa Salad Bowl
1502	610	1	Chicken Burrito
1612	651	1	Izze
1791	724	1	Canned Soft Drink
1792	724	1	Canned Soft Drink
2078	838	1	Chicken Soft Tacos
2127	856	1	Chicken Burrito
2307	926	1	Chicken Bowl
2318	926	1	Chicken Bowl
2362	940	1	Canned Soft Drink
2448	972	1	Chicken Bowl
2553	1011	1	Canned Soft Drink
2554	1011	1	Canned Soft Drink
2662	1057	1	Chicken Soft Tacos
2717	1078	1	Chicken Bowl
2833	1125	1	Chicken Burrito
2973	1182	1	Chicken Bowl
2977	1184	1	Steak Bowl
3637	1454	1	Chicken Soft Tacos
3716	1484	1	Canned Soda
3764	1507	1	Chicken Burrito
3829	1534	1	Chicken Burrito
4112	1647	1	Chicken Bowl
4236	1691	1	Chicken Bowl
4463	1778	1	Canned Soft Drink
4603	1827	1	Barbacoa Burrito

	choice_description	item_price
238	[Tomatillo Red Chili Salsa, [Rice, Black Beans...	11.75
248	[Mountain Dew]	1.09
297	[Tomatillo Green Chili Salsa, [Rice, Cheese, G...	11.75
381	[Coke]	1.25
484	[Fresh Tomato Salsa, [Fajita Vegetables, Rice,...	8.75
567	[Diet Coke]	1.25
684	[Coke]	1.25
773	[Tomatillo-Green Chili Salsa (Medium), Black B...	8.49
908	[Roasted Chili Corn Salsa (Medium), [Rice, Faj...	8.99
924	[Tomatillo-Red Chili Salsa (Hot), Cheese]	8.49
1117	[Fresh Tomato Salsa, [Rice, Black Beans, Chees...	11.25
1203	[Fresh Tomato Salsa, [Cheese, Sour Cream, Lett...	8.75
1245	[Lemonade]	1.25
1381	[Coke]	1.25
1420	[Roasted Chili Corn Salsa]	9.25
1421	[Roasted Chili Corn Salsa]	9.39
1502	[Fresh Tomato Salsa, [Fajita Vegetables, Rice,...	8.75

1612	[Blackberry]	3.39
1791	[Coke]	1.25
1792	[Coke]	1.25
2078	[Fresh Tomato Salsa, [Cheese, Sour Cream, Lett...	8.75
2127	[Fresh Tomato Salsa, [Rice, Black Beans, Chees...	11.25
2307	[Roasted Chili Corn Salsa, [Fajita Vegetables,...	8.75
2318	[Roasted Chili Corn Salsa, [Fajita Vegetables,...	8.75
2362	[Coke]	1.25
2448	[Tomatillo Red Chili Salsa, [Rice, Black Beans...	8.75
2553	[Coke]	1.25
2554	[Coke]	1.25
2662	[Tomatillo Red Chili Salsa, [Fajita Vegetables...	11.25
2717	[Fresh Tomato Salsa, [Rice, Black Beans, Chees...	11.25
2833	[Fresh Tomato Salsa, [Rice, Cheese, Lettuce]]	8.75
2973	[Tomatillo-Red Chili Salsa (Hot), [Rice, Black...	8.49
2977	[Fresh Tomato Salsa, [Rice, Pinto Beans, Chees...	9.25
3637	[Fresh Tomato Salsa, Rice]	8.75
3716	[Dr. Pepper]	1.09
3764	[Fresh Tomato Salsa, [Rice, Black Beans, Chees...	8.75
3829	[Fresh Tomato Salsa, [Rice, Cheese, Sour Cream...	8.75
4112	[Fresh Tomato Salsa, [Fajita Vegetables, Rice,...	11.25
4236	[White Rice]	8.50
4463	[Lemonade]	1.25
4603	[Tomatillo Green Chili Salsa]	9.25

```
[19]: # Display the count of entries for each category, including duplicates
category_counts1 = df['order_id'].value_counts()
print(category_counts1)
```

```
926      23
1483     14
691      11
205      11
1786     10
..
1205      1
1207      1
433       1
1210      1
917       1
Name: order_id, Length: 1833, dtype: int64
```

```
[20]: # Calculate the total number of duplicate values of order_id column
total_duplicate_values1 = len(category_counts1)
print(total_duplicate_values1)
```

```
1833
```



```
[21]: # Calculating the mean without handling duplicates of order_id column
mean_without_duplicates = df['order_id'].mean()
print({mean_without_duplicates})
```

```
{920.3320497630332}
```

```
[22]: # Calculating the mean after dropping duplicates
mean_with_duplicates_handled = df['order_id'].drop_duplicates().mean()
print({mean_with_duplicates_handled})
```

```
{917.4249863611566}
```

```
[23]: # Dropping duplicates from 'item_name' column permanently
df['order_id'] = df['order_id'].drop_duplicates()
```

```
[24]: df
```

```
[24]:
```

	order_id	quantity	item_name \
1	1.0	1	Izze
2	NaN	1	Nantucket Nectar
4	2.0	2	Chicken Bowl
5	3.0	1	Chicken Bowl
7	4.0	1	Steak Burrito
...	...	...	...
4617	1833.0	1	Steak Burrito
4618	NaN	1	Steak Burrito
4619	1834.0	1	Chicken Salad Bowl
4620	NaN	1	Chicken Salad Bowl
4621	NaN	1	Chicken Salad Bowl

	choice_description	item_price
1	[Clementine]	3.39
2	[Apple]	3.39
4	[Tomatillo-Red Chili Salsa (Hot), [Black Beans...	16.98
5	[Fresh Tomato Salsa (Mild), [Rice, Cheese, Sou...	10.98
7	[Tomatillo Red Chili Salsa, [Fajita Vegetables...	11.75
...	...	...
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

```
[3376 rows x 5 columns]
```

```
[ ]:
```

```
[25]: # Display the count of entries for each category, including duplicates
category_counts2 = df['item_name'].value_counts()
print(category_counts2)
```

```
Chicken Bowl          726
Chicken Burrito       553
Steak Burrito         368
Canned Soft Drink     301
Steak Bowl            211
Chicken Soft Tacos    115
Chicken Salad Bowl    110
Canned Soda           104
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
Chicken Crispy Tacos   47
Carnitas Soft Tacos    40
Steak Crispy Tacos     35
Steak Salad Bowl      29
Nantucket Nectar      27
Barbacoa Soft Tacos    25
Izze                  20
Veggie Salad Bowl     18
Barbacoa Crispy Tacos  11
Barbacoa Salad Bowl   10
Chicken Salad          9
Carnitas Crispy Tacos  7
Veggie Soft Tacos      7
Burrito                6
Carnitas Salad Bowl    6
Veggie Salad           6
Steak Salad            4
Bowl                   2
Crispy Tacos           2
Salad                  2
Veggie Crispy Tacos    1
Carnitas Salad         1
Name: item_name, dtype: int64
```

```
[26]: # Calculate the total number of duplicate values of order_id column
total_duplicate_values2 = len(category_counts2)
print(total_duplicate_values2)
```

38

```
[27]: # Counting unique items without handling duplicates
unique_items = df['item_name'].nunique()
print({unique_items})
```

{38}

```
[28]: # Counting unique items after dropping duplicates
unique_items_after_dropping = df['item_name'].drop_duplicates().nunique()
print({unique_items_after_dropping})
```

{38}

```
[29]: # Display the count of entries for each category, including duplicates
category_counts3= df['item_price'].value_counts()
print(category_counts3)
```

8.75	730
11.25	521
9.25	398
8.49	311
11.75	284
1.25	264
8.99	246
10.98	99
1.09	84
6.49	53
11.48	51
3.39	45
17.50	40
2.50	27
11.89	26
2.18	19
8.69	19
16.98	19
9.39	17
8.19	14
22.50	13
18.50	12
7.40	11
21.96	10
11.08	10
17.98	7
3.75	7
23.50	6
10.58	5
5.00	3

```

26.25    3
6.78     2
8.50     2
22.16    2
35.00    2
27.75    2
23.78    2
35.25    1
16.38    1
33.75    1
4.36     1
12.98    1
26.07    1
22.96    1
32.94    1
22.20    1
11.49    1
Name: item_price, dtype: int64

```

```

[30]: # Calculate the total number of duplicate values of order_id column
total_duplicate_values3 = len(category_counts3)
print(total_duplicate_values3)

```

```
47
```

```

[31]: # Calculating the mean without handling duplicates of order_id column
mean_without_duplicates = df['item_price'].mean()
print({mean_without_duplicates})

```

```
{9.011321090047451}
```

```

[32]: # Calculating the mean after dropping duplicates
mean_with_duplicates_handled = df['item_price'].drop_duplicates().mean()
print({mean_with_duplicates_handled})

```

```
{14.496382978723403}
```

```

[33]: # Dropping duplicates from 'item_name' column permanently
df['item_price'] = df['item_price'].drop_duplicates()

```

```
[34]: df
```

```

[34]:   order_id  quantity  item_name \
1         1.0         1         Izze
2         NaN         1  Nantucket Nectar
4         2.0         2    Chicken Bowl
5         3.0         1    Chicken Bowl
7         4.0         1    Steak Burrito

```

...	...	...	...
4617	1833.0	1	Steak Burrito
4618	NaN	1	Steak Burrito
4619	1834.0	1	Chicken Salad Bowl
4620	NaN	1	Chicken Salad Bowl
4621	NaN	1	Chicken Salad Bowl

	choice_description	item_price
1	[Clementine]	3.39
2	[Apple]	NaN
4	[Tomatillo-Red Chili Salsa (Hot), [Black Beans...	16.98
5	[Fresh Tomato Salsa (Mild), [Rice, Cheese, Sou...	10.98
7	[Tomatillo Red Chili Salsa, [Fajita Vegetables...	11.75
...	...	...
4617	[Fresh Tomato Salsa, [Rice, Black Beans, Sour ...	NaN
4618	[Fresh Tomato Salsa, [Rice, Sour Cream, Cheese...	NaN
4619	[Fresh Tomato Salsa, [Fajita Vegetables, Pinto...	NaN
4620	[Fresh Tomato Salsa, [Fajita Vegetables, Lettu...	NaN
4621	[Fresh Tomato Salsa, [Fajita Vegetables, Pinto...	NaN

[3376 rows x 5 columns]

```
[35]: #Finding missing value
missing_values = df.isnull()
print(missing_values)
```

	order_id	quantity	item_name	choice_description	item_price
1	False	False	False	False	False
2	True	False	False	False	True
4	False	False	False	False	False
5	False	False	False	False	False
7	False	False	False	False	False
...	...	...	...	...	...
4617	False	False	False	False	True
4618	True	False	False	False	True
4619	False	False	False	False	True
4620	True	False	False	False	True
4621	True	False	False	False	True

[3376 rows x 5 columns]

```
[36]: # Counting missing values in each column
missing_values_count = df.isnull().sum()
print(missing_values_count)
```

order_id	1543
quantity	0
item_name	0

```
choice_description      0
item_price              3329
dtype: int64
```

```
[37]: #Tackeling missing values
df_cleaned = df.dropna(inplace=True)
print(df_cleaned)
```

None

```
[38]: df
```

```
[38]:
```

	order_id	quantity	item_name \
1	1.0	1	Izze
4	2.0	2	Chicken Bowl
5	3.0	1	Chicken Bowl
7	4.0	1	Steak Burrito
11	6.0	1	Chicken Crispy Tacos
13	7.0	1	Chicken Bowl
16	8.0	1	Chicken Burrito
21	11.0	1	Barbacoa Burrito
135	60.0	2	Chicken Salad Bowl
147	66.0	1	Steak Burrito
148	67.0	2	Steak Burrito
150	68.0	2	Chicken Burrito
281	123.0	2	Steak Salad Bowl
322	140.0	1	Steak Burrito
353	152.0	2	Steak Burrito
409	178.0	3	Chicken Bowl
445	193.0	3	Bowl
476	204.0	1	Chicken Burrito
510	214.0	1	Burrito
606	250.0	1	Steak Salad Bowl
728	301.0	1	Steak Burrito
1254	511.0	4	Chicken Burrito
1514	616.0	3	Chicken Bowl
1718	692.0	2	Chicken Burrito
2235	901.0	4	Canned Soda
2954	1175.0	2	Steak Burrito
3334	1336.0	3	Steak Bowl
3480	1398.0	3	Carnitas Bowl
4235	1691.0	1	Chicken Bowl

	choice_description	item_price
1	[Clementine]	3.39
4	[Tomatillo-Red Chili Salsa (Hot), [Black Beans...	16.98
5	[Fresh Tomato Salsa (Mild), [Rice, Cheese, Sou...	10.98

7	[Tomatillo Red Chili Salsa, [Fajita Vegetables...	11.75
11	[Roasted Chili Corn Salsa, [Fajita Vegetables,...	8.75
13	[Fresh Tomato Salsa, [Fajita Vegetables, Rice,...	11.25
16	[Tomatillo-Green Chili Salsa (Medium), [Pinto ...	8.49
21	[[Fresh Tomato Salsa (Mild), Tomatillo-Green C...	8.99
135	[Tomatillo Green Chili Salsa, [Sour Cream, Che...	22.50
147	[Tomatillo-Green Chili Salsa (Medium), [Rice, ...	11.48
148	[Tomatillo-Red Chili Salsa (Hot), [Rice, Chees...	17.98
150	[Tomatillo Red Chili Salsa, [Rice, Black Beans...	17.50
281	[Tomatillo Red Chili Salsa, [Black Beans, Chee...	23.78
322	[Fresh Tomato (Mild), [Lettuce, Guacamole, Ric...	11.08
353	[Fresh Tomato (Mild), [Lettuce, Guacamole, Ric...	22.16
409	[[Fresh Tomato Salsa (Mild), Tomatillo-Green C...	32.94
445	[Braised Carnitas, Pinto Beans, [Sour Cream, C...	22.20
476	[Fresh Tomato (Mild), [Guacamole, Lettuce, Ric...	10.58
510	[Adobo-Marinated and Grilled Chicken, Pinto Be...	7.40
606	[Fresh Tomato Salsa, [Pinto Beans, Cheese, Gua...	11.89
728	[Roasted Chili Corn (Medium), [Sour Cream, Che...	8.69
1254	[Fresh Tomato Salsa, [Fajita Vegetables, Rice,...	35.00
1514	[Fresh Tomato Salsa, [Rice, Black Beans, Chees...	26.25
1718	[Tomatillo-Red Chili Salsa (Hot), [Lettuce, Gu...	21.96
2235	[Sprite]	4.36
2954	[[Tomatillo-Green Chili Salsa (Medium), Roaste...	22.96
3334	[Tomatillo Green Chili (Medium), [Rice, Black ...	26.07
3480	[Roasted Chili Corn Salsa, [Fajita Vegetables,...	35.25
4235	[White Rice]	8.50

```
[39]: #Examine the Quantity and Item Price columns
print(df['quantity'].unique())
```

```
[1 2 3 4]
```

```
[40]: print(df['item_price'].unique())
```

```
[ 3.39 16.98 10.98 11.75  8.75 11.25  8.49  8.99 22.5  11.48 17.98 17.5
23.78 11.08 22.16 32.94 22.2 10.58  7.4  11.89  8.69 35.  26.25 21.96
 4.36 22.96 26.07 35.25  8.5 ]
```

```
[41]: # Display unique values in the 'Choice Description' column
unique_choices = df['choice_description'].unique()
print(unique_choices)
```

```
['[Clementine]'
 '[Tomatillo-Red Chili Salsa (Hot), [Black Beans, Rice, Cheese, Sour Cream]]'
 '[Fresh Tomato Salsa (Mild), [Rice, Cheese, Sour Cream, Guacamole, Lettuce]]'
 '[Tomatillo Red Chili Salsa, [Fajita Vegetables, Black Beans, Pinto Beans,
Cheese, Sour Cream, Guacamole, Lettuce]]'
 '[Roasted Chili Corn Salsa, [Fajita Vegetables, Rice, Black Beans, Cheese, Sour
```

```

Cream]]'
' [Fresh Tomato Salsa, [Fajita Vegetables, Rice, Cheese, Sour Cream,
Guacamole]]'
' [Tomatillo-Green Chili Salsa (Medium), [Pinto Beans, Cheese, Sour Cream]]'
' [[Fresh Tomato Salsa (Mild), Tomatillo-Green Chili Salsa (Medium), Tomatillo-
Red Chili Salsa (Hot)], [Rice, Cheese, Sour Cream, Lettuce]]'
' [Tomatillo Green Chili Salsa, [Sour Cream, Cheese, Guacamole]]'
' [Tomatillo-Green Chili Salsa (Medium), [Rice, Black Beans, Sour Cream,
Guacamole]]'
' [Tomatillo-Red Chili Salsa (Hot), [Rice, Cheese, Sour Cream, Lettuce]]'
' [Tomatillo Red Chili Salsa, [Rice, Black Beans, Sour Cream, Cheese, Lettuce]]'
' [Tomatillo Red Chili Salsa, [Black Beans, Cheese, Guacamole]]'
' [Fresh Tomato (Mild), [Lettuce, Guacamole, Rice, Cheese]]'
' [[Fresh Tomato Salsa (Mild), Tomatillo-Green Chili Salsa (Medium), Roasted
Chili Corn Salsa (Medium)], [Black Beans, Rice, Fajita Veggies, Cheese,
Guacamole, Lettuce]]'
' [Braised Carnitas, Pinto Beans, [Sour Cream, Cheese, Cilantro-Lime Rice]]'
' [Fresh Tomato (Mild), [Guacamole, Lettuce, Rice, Fajita Veggies, Sour Cream,
Cheese]]'
' [Adobo-Marinated and Grilled Chicken, Pinto Beans, [Sour Cream, Salsa, Cheese,
Cilantro-Lime Rice, Guacamole]]'
' [Fresh Tomato Salsa, [Pinto Beans, Cheese, Guacamole, Lettuce]]'
' [Roasted Chili Corn (Medium), [Sour Cream, Cheese]]'
' [Fresh Tomato Salsa, [Fajita Vegetables, Rice, Black Beans, Cheese, Lettuce]]'
' [Fresh Tomato Salsa, [Rice, Black Beans, Cheese, Sour Cream, Lettuce]]'
' [Tomatillo-Red Chili Salsa (Hot), [Lettuce, Guacamole, Sour Cream, Cheese,
Rice, Black Beans]]'
' [Sprite]'
' [[Tomatillo-Green Chili Salsa (Medium), Roasted Chili Corn Salsa (Medium),
Tomatillo-Red Chili Salsa (Hot)], [Black Beans, Rice, Cheese, Sour Cream,
Guacamole, Lettuce]]'
' [Tomatillo Green Chili (Medium), [Rice, Black Beans, Sour Cream, Cheese]]'
' [Roasted Chili Corn Salsa, [Fajita Vegetables, Rice, Black Beans, Cheese, Sour
Cream, Guacamole, Lettuce]]'
' [White Rice]'

```

```

[42]: #Concatenate multiple descriptions into one string
df['Aggregated_Choice'] = df.groupby('item_name')['choice_description'].
    ↪transform(lambda x: '; '.join(x))

```

```

[43]: #this method keeps only the first description for each item
df['First_Choice'] = df.groupby('item_name')['choice_description'].
    ↪transform('first')

```

```

[44]: df.head(10)

```



```

[44]:      order_id  quantity      item_name \
1         1.0        1           Izze
4         2.0        2       Chicken Bowl
5         3.0        1       Chicken Bowl
7         4.0        1       Steak Burrito
11        6.0        1  Chicken Crispy Tacos
13        7.0        1       Chicken Bowl
16        8.0        1       Chicken Burrito
21       11.0        1       Barbacoa Burrito
135       60.0        2  Chicken Salad Bowl
147       66.0        1       Steak Burrito

                                choice_description  item_price \
1                                [Clementine]           3.39
4  [Tomatillo-Red Chili Salsa (Hot), [Black Beans...      16.98
5  [Fresh Tomato Salsa (Mild), [Rice, Cheese, Sou...      10.98
7  [Tomatillo Red Chili Salsa, [Fajita Vegetables...      11.75
11 [Roasted Chili Corn Salsa, [Fajita Vegetables,...       8.75
13 [Fresh Tomato Salsa, [Fajita Vegetables, Rice,...      11.25
16 [Tomatillo-Green Chili Salsa (Medium), [Pinto ...       8.49
21 [[Fresh Tomato Salsa (Mild), Tomatillo-Green C...       8.99
135 [Tomatillo Green Chili Salsa, [Sour Cream, Che...      22.50
147 [Tomatillo-Green Chili Salsa (Medium), [Rice, ...      11.48

                                Aggregated_Choice \
1                                [Clementine]
4  [Tomatillo-Red Chili Salsa (Hot), [Black Beans...
5  [Tomatillo-Red Chili Salsa (Hot), [Black Beans...
7  [Tomatillo Red Chili Salsa, [Fajita Vegetables...
11 [Roasted Chili Corn Salsa, [Fajita Vegetables,...
13 [Tomatillo-Red Chili Salsa (Hot), [Black Beans...
16 [Tomatillo-Green Chili Salsa (Medium), [Pinto ...
21 [[Fresh Tomato Salsa (Mild), Tomatillo-Green C...
135 [Tomatillo Green Chili Salsa, [Sour Cream, Che...
147 [Tomatillo Red Chili Salsa, [Fajita Vegetables...

                                First_Choice
1                                [Clementine]
4  [Tomatillo-Red Chili Salsa (Hot), [Black Beans...
5  [Tomatillo-Red Chili Salsa (Hot), [Black Beans...
7  [Tomatillo Red Chili Salsa, [Fajita Vegetables...
11 [Roasted Chili Corn Salsa, [Fajita Vegetables,...
13 [Tomatillo-Red Chili Salsa (Hot), [Black Beans...
16 [Tomatillo-Green Chili Salsa (Medium), [Pinto ...
21 [[Fresh Tomato Salsa (Mild), Tomatillo-Green C...
135 [Tomatillo Green Chili Salsa, [Sour Cream, Che...
147 [Tomatillo Red Chili Salsa, [Fajita Vegetables...

```

```
[45]: #cross checking the order_id
      #check for null values
      null_order_ids = df[df['order_id'].isnull()]
      print(null_order_ids)
```

Empty DataFrame

Columns: [order\_id, quantity, item\_name, choice\_description, item\_price, Aggregated\_Choice, First\_Choice]  
Index: []

```
[46]: # Check for duplicate Order IDs
      duplicate_order_ids = df[df.duplicated('order_id')]
      print(duplicate_order_ids)
```

Empty DataFrame

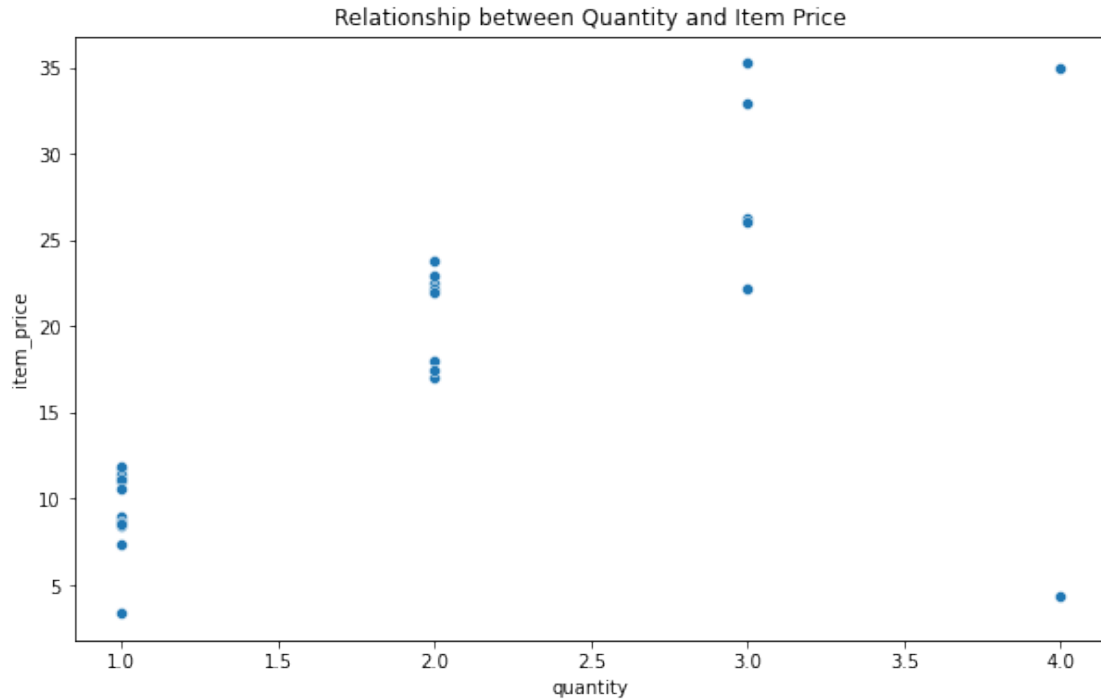
Columns: [order\_id, quantity, item\_name, choice\_description, item\_price, Aggregated\_Choice, First\_Choice]  
Index: []

[47]:

```
Input In [47]
def find_special_characters(item_name):
    ~
```

**IndentationError:** unexpected indent

```
[52]: plt.figure(figsize=(10, 6))
      sns.scatterplot(x='quantity', y='item_price', data=df)
      plt.title('Relationship between Quantity and Item Price')
      plt.xlabel('quantity')
      plt.ylabel('item_price')
      plt.show()
```



```
[61]: # Calculating the correlation between 'quantity' and 'item_price'
correlation = df['quantity'].corr(df['item_price'])
print(correlation)
```

0.7014176320697094

```
[62]: #For categorical columns (e.g., Item Name), consider encoding or transforming
      ↪ them into a format suitable for analysis.
df_encoded = pd.get_dummies(df, columns=['item_name'], prefix='item')
```

```
[63]: print(df_encoded)
```

	order_id	quantity	choice_description \
1	1.0	1	[Clementine]
4	2.0	2	[Tomatillo-Red Chili Salsa (Hot), [Black Beans...
5	3.0	1	[Fresh Tomato Salsa (Mild), [Rice, Cheese, Sou...
7	4.0	1	[Tomatillo Red Chili Salsa, [Fajita Vegetables...
11	6.0	1	[Roasted Chili Corn Salsa, [Fajita Vegetables,...
13	7.0	1	[Fresh Tomato Salsa, [Fajita Vegetables, Rice,...
16	8.0	1	[Tomatillo-Green Chili Salsa (Medium), [Pinto ...
21	11.0	1	[[Fresh Tomato Salsa (Mild), Tomatillo-Green C...
135	60.0	2	[Tomatillo Green Chili Salsa, [Sour Cream, Che...
147	66.0	1	[Tomatillo-Green Chili Salsa (Medium), [Rice, ...
148	67.0	2	[Tomatillo-Red Chili Salsa (Hot), [Rice, Chees...
150	68.0	2	[Tomatillo Red Chili Salsa, [Rice, Black Beans...

281	123.0	2	[Tomatillo Red Chili Salsa, [Black Beans, Chee...
322	140.0	1	[Fresh Tomato (Mild), [Lettuce, Guacamole, Ric...
353	152.0	2	[Fresh Tomato (Mild), [Lettuce, Guacamole, Ric...
409	178.0	3	[[Fresh Tomato Salsa (Mild), Tomatillo-Green C...
445	193.0	3	[Braised Carnitas, Pinto Beans, [Sour Cream, C...
476	204.0	1	[Fresh Tomato (Mild), [Guacamole, Lettuce, Ric...
510	214.0	1	[Adobo-Marinated and Grilled Chicken, Pinto Be...
606	250.0	1	[Fresh Tomato Salsa, [Pinto Beans, Cheese, Gua...
728	301.0	1	[Roasted Chili Corn (Medium), [Sour Cream, Che...
1254	511.0	4	[Fresh Tomato Salsa, [Fajita Vegetables, Rice,...
1514	616.0	3	[Fresh Tomato Salsa, [Rice, Black Beans, Chees...
1718	692.0	2	[Tomatillo-Red Chili Salsa (Hot), [Lettuce, Gu...
2235	901.0	4	[Sprite]
2954	1175.0	2	[[Tomatillo-Green Chili Salsa (Medium), Roaste...
3334	1336.0	3	[Tomatillo Green Chili (Medium), [Rice, Black ...
3480	1398.0	3	[Roasted Chili Corn Salsa, [Fajita Vegetables,...
4235	1691.0	1	[White Rice]

	item_price	Aggregated_Choice \
1	3.39	[Clementine]
4	16.98	[Tomatillo-Red Chili Salsa (Hot), [Black Beans...
5	10.98	[Tomatillo-Red Chili Salsa (Hot), [Black Beans...
7	11.75	[Tomatillo Red Chili Salsa, [Fajita Vegetables...
11	8.75	[Roasted Chili Corn Salsa, [Fajita Vegetables,...
13	11.25	[Tomatillo-Red Chili Salsa (Hot), [Black Beans...
16	8.49	[Tomatillo-Green Chili Salsa (Medium), [Pinto ...
21	8.99	[[Fresh Tomato Salsa (Mild), Tomatillo-Green C...
135	22.50	[Tomatillo Green Chili Salsa, [Sour Cream, Che...
147	11.48	[Tomatillo Red Chili Salsa, [Fajita Vegetables...
148	17.98	[Tomatillo Red Chili Salsa, [Fajita Vegetables...
150	17.50	[Tomatillo-Green Chili Salsa (Medium), [Pinto ...
281	23.78	[Tomatillo Red Chili Salsa, [Black Beans, Chee...
322	11.08	[Tomatillo Red Chili Salsa, [Fajita Vegetables...
353	22.16	[Tomatillo Red Chili Salsa, [Fajita Vegetables...
409	32.94	[Tomatillo-Red Chili Salsa (Hot), [Black Beans...
445	22.20	[Braised Carnitas, Pinto Beans, [Sour Cream, C...
476	10.58	[Tomatillo-Green Chili Salsa (Medium), [Pinto ...
510	7.40	[Adobo-Marinated and Grilled Chicken, Pinto Be...
606	11.89	[Tomatillo Red Chili Salsa, [Black Beans, Chee...
728	8.69	[Tomatillo Red Chili Salsa, [Fajita Vegetables...
1254	35.00	[Tomatillo-Green Chili Salsa (Medium), [Pinto ...
1514	26.25	[Tomatillo-Red Chili Salsa (Hot), [Black Beans...
1718	21.96	[Tomatillo-Green Chili Salsa (Medium), [Pinto ...
2235	4.36	[Sprite]
2954	22.96	[Tomatillo Red Chili Salsa, [Fajita Vegetables...
3334	26.07	[Tomatillo Green Chili (Medium), [Rice, Black ...
3480	35.25	[Roasted Chili Corn Salsa, [Fajita Vegetables,...
4235	8.50	[Tomatillo-Red Chili Salsa (Hot), [Black Beans...

	First_Choice \
1	[Clementine]
4	[Tomatillo-Red Chili Salsa (Hot), [Black Beans...
5	[Tomatillo-Red Chili Salsa (Hot), [Black Beans...
7	[Tomatillo Red Chili Salsa, [Fajita Vegetables...
11	[Roasted Chili Corn Salsa, [Fajita Vegetables,...
13	[Tomatillo-Red Chili Salsa (Hot), [Black Beans...
16	[Tomatillo-Green Chili Salsa (Medium), [Pinto ...
21	[[Fresh Tomato Salsa (Mild), Tomatillo-Green C...
135	[Tomatillo Green Chili Salsa, [Sour Cream, Chee...
147	[Tomatillo Red Chili Salsa, [Fajita Vegetables...
148	[Tomatillo Red Chili Salsa, [Fajita Vegetables...
150	[Tomatillo-Green Chili Salsa (Medium), [Pinto ...
281	[Tomatillo Red Chili Salsa, [Black Beans, Chee...
322	[Tomatillo Red Chili Salsa, [Fajita Vegetables...
353	[Tomatillo Red Chili Salsa, [Fajita Vegetables...
409	[Tomatillo-Red Chili Salsa (Hot), [Black Beans...
445	[Braised Carnitas, Pinto Beans, [Sour Cream, C...
476	[Tomatillo-Green Chili Salsa (Medium), [Pinto ...
510	[Adobo-Marinaded and Grilled Chicken, Pinto Be...
606	[Tomatillo Red Chili Salsa, [Black Beans, Chee...
728	[Tomatillo Red Chili Salsa, [Fajita Vegetables...
1254	[Tomatillo-Green Chili Salsa (Medium), [Pinto ...
1514	[Tomatillo-Red Chili Salsa (Hot), [Black Beans...
1718	[Tomatillo-Green Chili Salsa (Medium), [Pinto ...
2235	[Sprite]
2954	[Tomatillo Red Chili Salsa, [Fajita Vegetables...
3334	[Tomatillo Green Chili (Medium), [Rice, Black ...
3480	[Roasted Chili Corn Salsa, [Fajita Vegetables,...
4235	[Tomatillo-Red Chili Salsa (Hot), [Black Beans...

	item_Barbacoa Burrito	item_Bowl	item_Burrito	item_Canned Soda \
1	0	0	0	0
4	0	0	0	0
5	0	0	0	0
7	0	0	0	0
11	0	0	0	0
13	0	0	0	0
16	0	0	0	0
21	1	0	0	0
135	0	0	0	0
147	0	0	0	0
148	0	0	0	0
150	0	0	0	0
281	0	0	0	0
322	0	0	0	0
353	0	0	0	0

409	0	0	0	0
445	0	1	0	0
476	0	0	0	0
510	0	0	1	0
606	0	0	0	0
728	0	0	0	0
1254	0	0	0	0
1514	0	0	0	0
1718	0	0	0	0
2235	0	0	0	1
2954	0	0	0	0
3334	0	0	0	0
3480	0	0	0	0
4235	0	0	0	0

	item_Carnitas Bowl	item_Chicken Bowl	item_Chicken Burrito \
1	0	0	0
4	0	1	0
5	0	1	0
7	0	0	0
11	0	0	0
13	0	1	0
16	0	0	1
21	0	0	0
135	0	0	0
147	0	0	0
148	0	0	0
150	0	0	1
281	0	0	0
322	0	0	0
353	0	0	0
409	0	1	0
445	0	0	0
476	0	0	1
510	0	0	0
606	0	0	0
728	0	0	0
1254	0	0	1
1514	0	1	0
1718	0	0	1
2235	0	0	0
2954	0	0	0
3334	0	0	0
3480	1	0	0
4235	0	1	0

	item_Chicken Crispy Tacos	item_Chicken Salad Bowl	item_Izze \
1	0	0	1

4	0	0	0
5	0	0	0
7	0	0	0
11	1	0	0
13	0	0	0
16	0	0	0
21	0	0	0
135	0	1	0
147	0	0	0
148	0	0	0
150	0	0	0
281	0	0	0
322	0	0	0
353	0	0	0
409	0	0	0
445	0	0	0
476	0	0	0
510	0	0	0
606	0	0	0
728	0	0	0
1254	0	0	0
1514	0	0	0
1718	0	0	0
2235	0	0	0
2954	0	0	0
3334	0	0	0
3480	0	0	0
4235	0	0	0

	item_Steak Bowl	item_Steak Burrito	item_Steak Salad Bowl
1	0	0	0
4	0	0	0
5	0	0	0
7	0	1	0
11	0	0	0
13	0	0	0
16	0	0	0
21	0	0	0
135	0	0	0
147	0	1	0
148	0	1	0
150	0	0	0
281	0	0	1
322	0	1	0
353	0	1	0
409	0	0	0
445	0	0	0
476	0	0	0

510	0	0	0
606	0	0	1
728	0	1	0
1254	0	0	0
1514	0	0	0
1718	0	0	0
2235	0	0	0
2954	0	1	0
3334	1	0	0
3480	0	0	0
4235	0	0	0

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]: