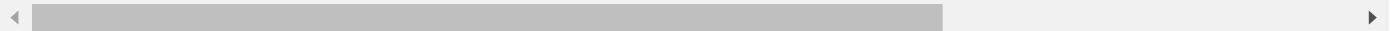


Author: Hasnat Kabir Shakil

Data

	Invoice ID	Branch	City	Customer type	Gender	Product line	Unit price	Quantity	Tax 5%	Total	Date	Time
0	750-67-8428	A	Yangon	Member	Female	Health and beauty	74.69	7	26.1415	548.9715	1/5/2019	13:08
1	226-31-3081	C	Naypyitaw	Normal	Female	Electronic accessories	15.28	5	3.8200	80.2200	3/8/2019	10:29
2	631-41-3108	A	Yangon	Normal	Male	Home and lifestyle	46.33	7	16.2155	340.5255	3/3/2019	13:23
3	123-19-1176	A	Yangon	Member	Male	Health and beauty	58.22	8	23.2880	489.0480	1/27/2019	20:33
4	373-73-7910	A	Yangon	Normal	Male	Sports and travel	86.31	7	30.2085	634.3785	2/8/2019	10:37



```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 17 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   Invoice ID      1000 non-null    object  
 1   Branch          1000 non-null    object  
 2   City            1000 non-null    object  
 3   Customer type   1000 non-null    object  
 4   Gender          1000 non-null    object  
 5   Product line    1000 non-null    object  
 6   Unit price     1000 non-null    float64 
 7   Quantity        1000 non-null    int64  
 8   Tax 5%          1000 non-null    float64 
 9   Total           1000 non-null    float64 
 10  Date            1000 non-null    object  
 11  Time            1000 non-null    object  
 12  Payment         1000 non-null    object  
 13  cogs            1000 non-null    float64 
 14  gross margin percentage 1000 non-null    float64 
 15  gross income   1000 non-null    float64 
 16  Rating          1000 non-null    float64 

dtypes: float64(7), int64(1), object(9)
memory usage: 132.9+ KB
```

Descriptive Statistics

	Unit price	Quantity	Tax 5%	Total	cogs	gross margin percentage	gross income	Rating
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	1.000000e+03	1000.000000	1000.000000
mean	55.672130	5.510000	15.379369	322.966749	307.58738	4.761905e+00	15.379369	6.97270
std	26.494628	2.923431	11.708825	245.885335	234.17651	6.131498e-14	11.708825	1.71858
min	10.080000	1.000000	0.508500	10.678500	10.17000	4.761905e+00	0.508500	4.00000
25%	32.875000	3.000000	5.924875	124.422375	118.49750	4.761905e+00	5.924875	5.50000
50%	55.230000	5.000000	12.088000	253.848000	241.76000	4.761905e+00	12.088000	7.00000
75%	77.935000	8.000000	22.445250	471.350250	448.90500	4.761905e+00	22.445250	8.50000
max	99.960000	10.000000	49.650000	1042.650000	993.00000	4.761905e+00	49.650000	10.00000

Missing values

```

Invoice ID          0
Branch             0
City               0
Customer type      0
Gender             0
Product line       0
Unit price         0
Quantity           0
Tax 5%             0
Total              0
Date               0
Time               0
Payment            0
cogs               0
gross margin percentage 0
gross income        0
Rating              0
dtype: int64

```

There is no missing values

Outliers

- We calculate skewness of each column using skew() function.
- The skewness of each of our columns is zero / 0.
- That is, we can say that 0 indicates that the distribution is perfectly symmetric.
- And, there are no outliers in our data.

```

Unit price          0.007077
Quantity            0.012941
Tax 5%              0.892570
Total               0.892570
cogs                0.892570
gross margin percentage 0.000000
gross income         0.892570
Rating              0.009010
dtype: float64

```

Let's analyze and try to find some insights

% of each Product:

Fashion accessories	0.18
Food and beverages	0.17
Electronic accessories	0.17
Sports and travel	0.17
Home and lifestyle	0.16
Health and beauty	0.15

Name: Product line, dtype: float64

```
% of each payment method:  
Ewallet      0.34  
Cash         0.34  
Credit card   0.31  
Name: Payment, dtype: float64
```

sort the total product sales

```
Product line  
Food and beverages      56145.0  
Sports and travel        55123.0  
Electronic accessories    54338.0  
Fashion accessories       54306.0  
Home and lifestyle        53862.0  
Health and beauty         49194.0  
Name: Total, dtype: float64
```

Behavior of each gender

	Quantity	Total	Rating
Gender			
Female	5.726547	335.095659	6.964471
Male	5.292585	310.789226	6.980962

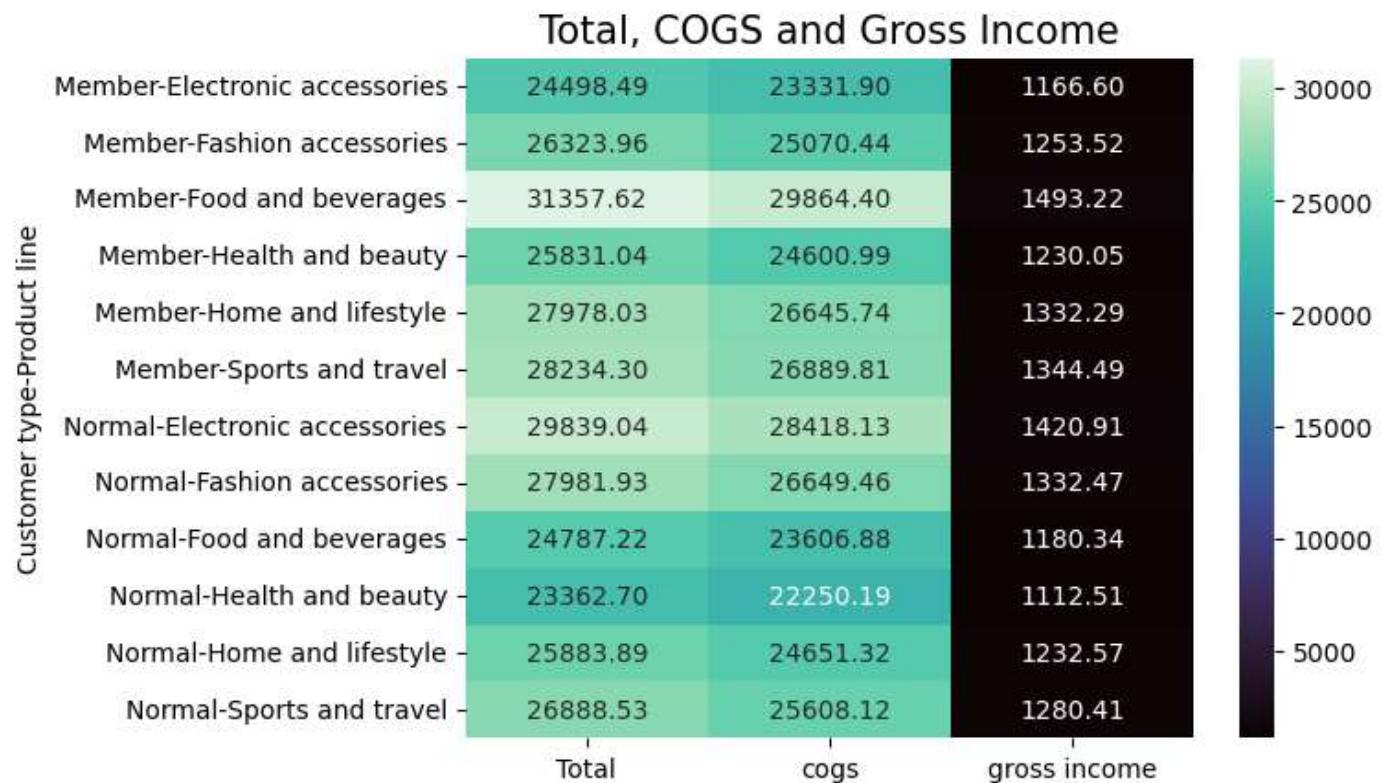
Total count of products by each gender

	Gender	Female	Male	Total
Product line				
Electronic accessories		84	86	170
Fashion accessories		96	82	178
Food and beverages		90	84	174
Health and beauty		64	88	152
Home and lifestyle		79	81	160
Sports and travel		88	78	166
Total		501	499	1000

Total, COGS and Gross-Income with Pivot Table

Customer type	Product line	Total	cogs	gross income
Member	Electronic accessories	24498.4950	23331.90	1166.5950
	Fashion accessories	26323.9620	25070.44	1253.5220
	Food and beverages	31357.6200	29864.40	1493.2200
	Health and beauty	25831.0395	24600.99	1230.0495
	Home and lifestyle	27978.0270	26645.74	1332.2870
	Sports and travel	28234.3005	26889.81	1344.4905
Normal	Electronic accessories	29839.0365	28418.13	1420.9065
	Fashion accessories	27981.9330	26649.46	1332.4730
	Food and beverages	24787.2240	23606.88	1180.3440
	Health and beauty	23362.6995	22250.19	1112.5095
	Home and lifestyle	25883.8860	24651.32	1232.5660
	Sports and travel	26888.5260	25608.12	1280.4060

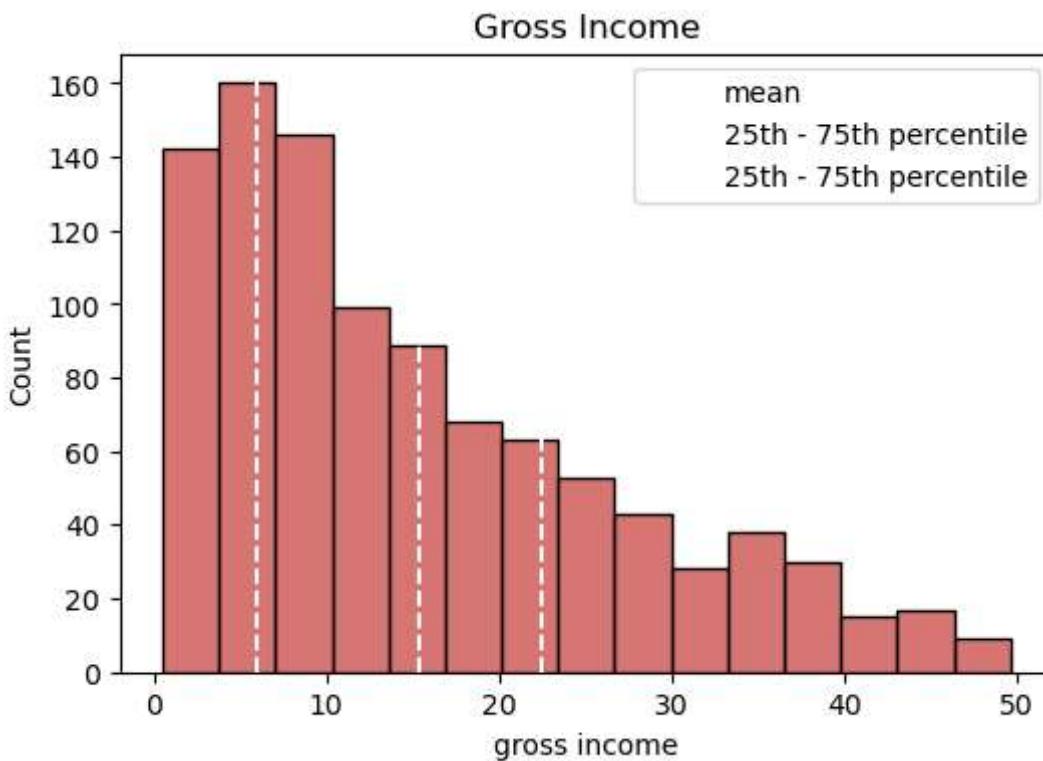
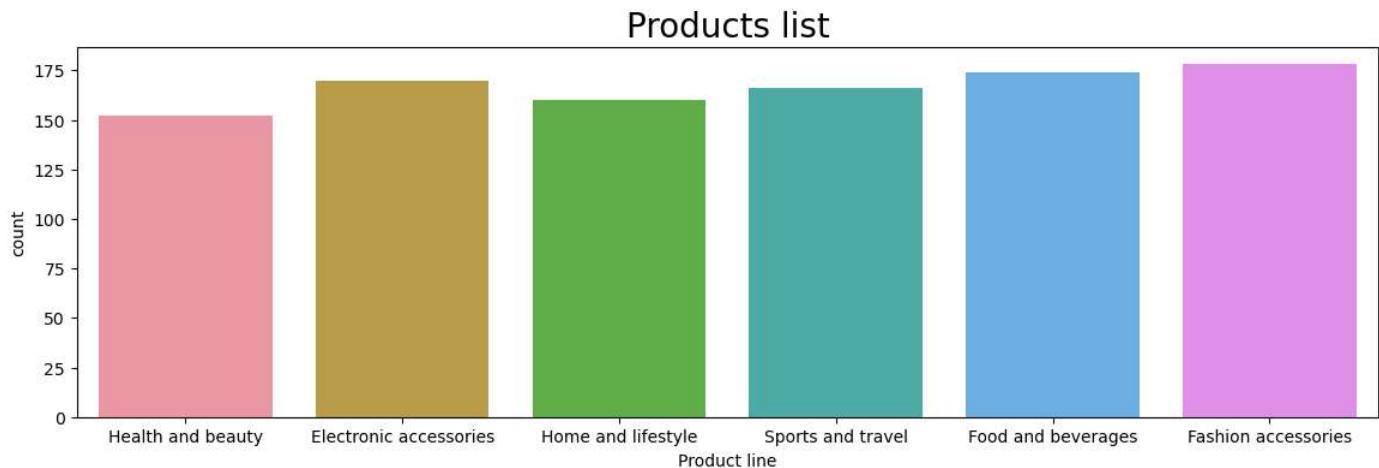
Pivot-Table Heatmap

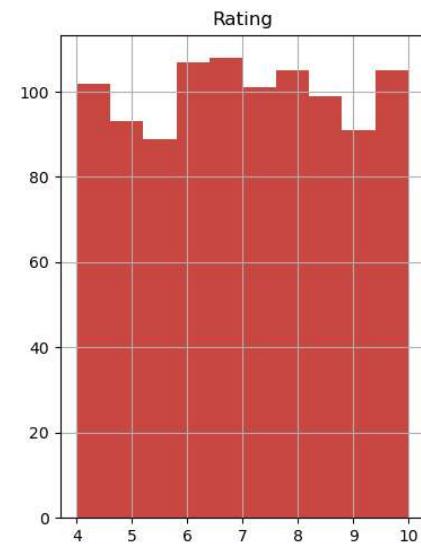
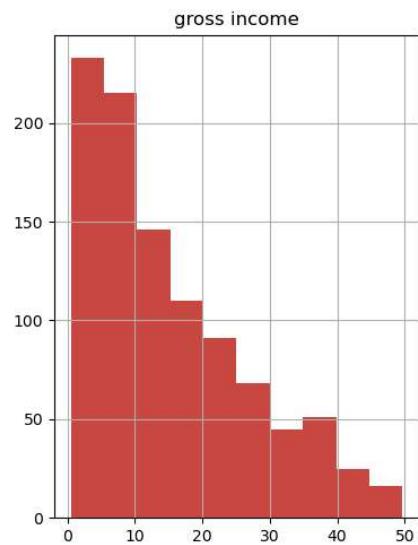
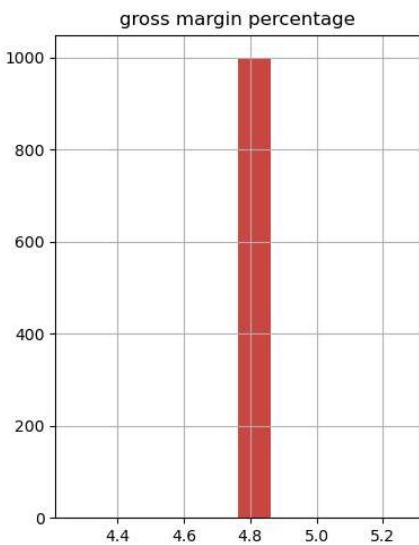
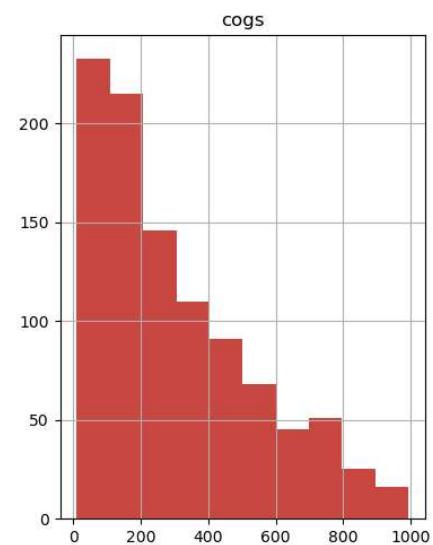
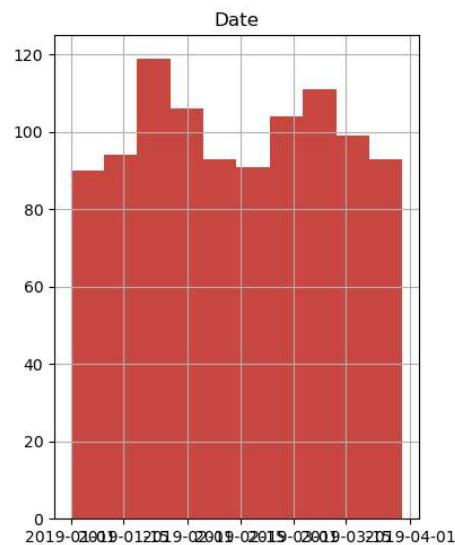
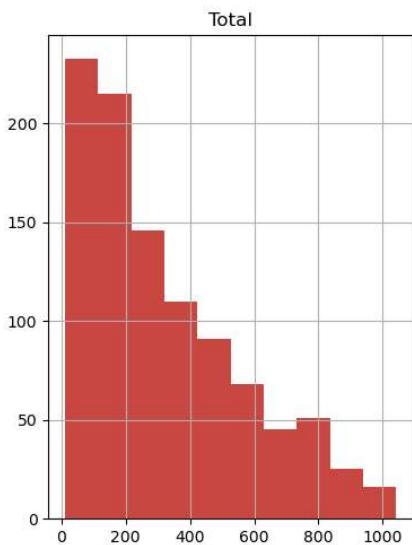
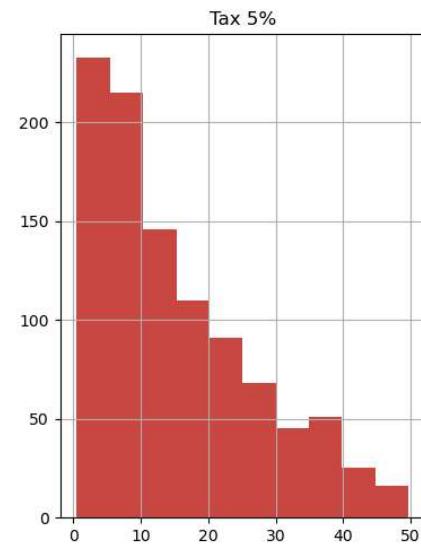
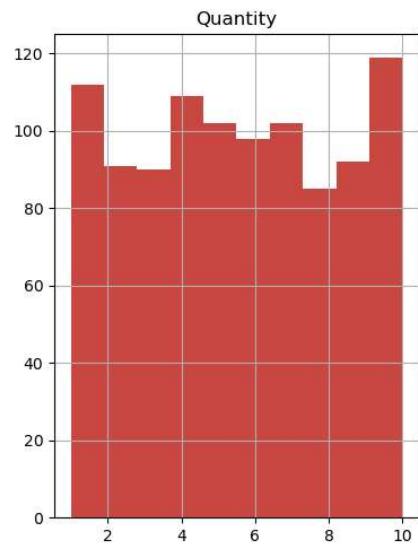
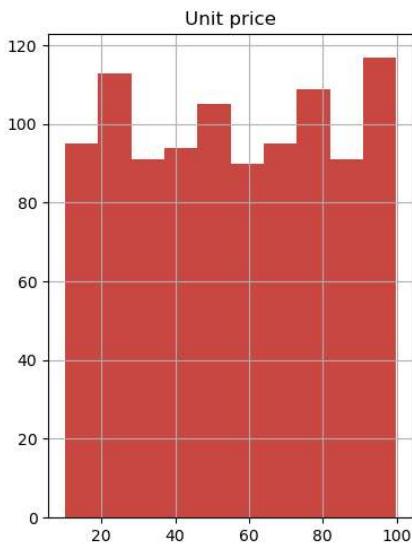


Explore the Dataset

Understanding and analyze each feature, data distribution

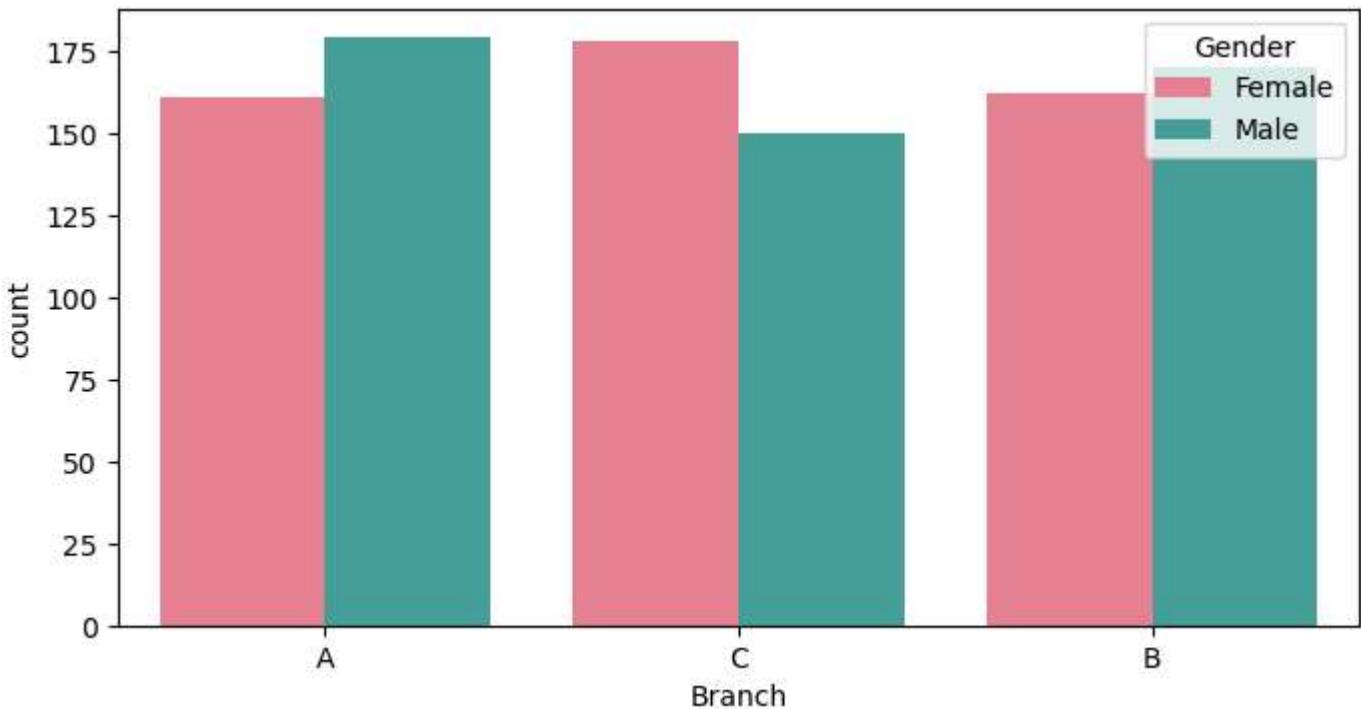
- Univariate
- Bivariate





Number of male and female in each branch

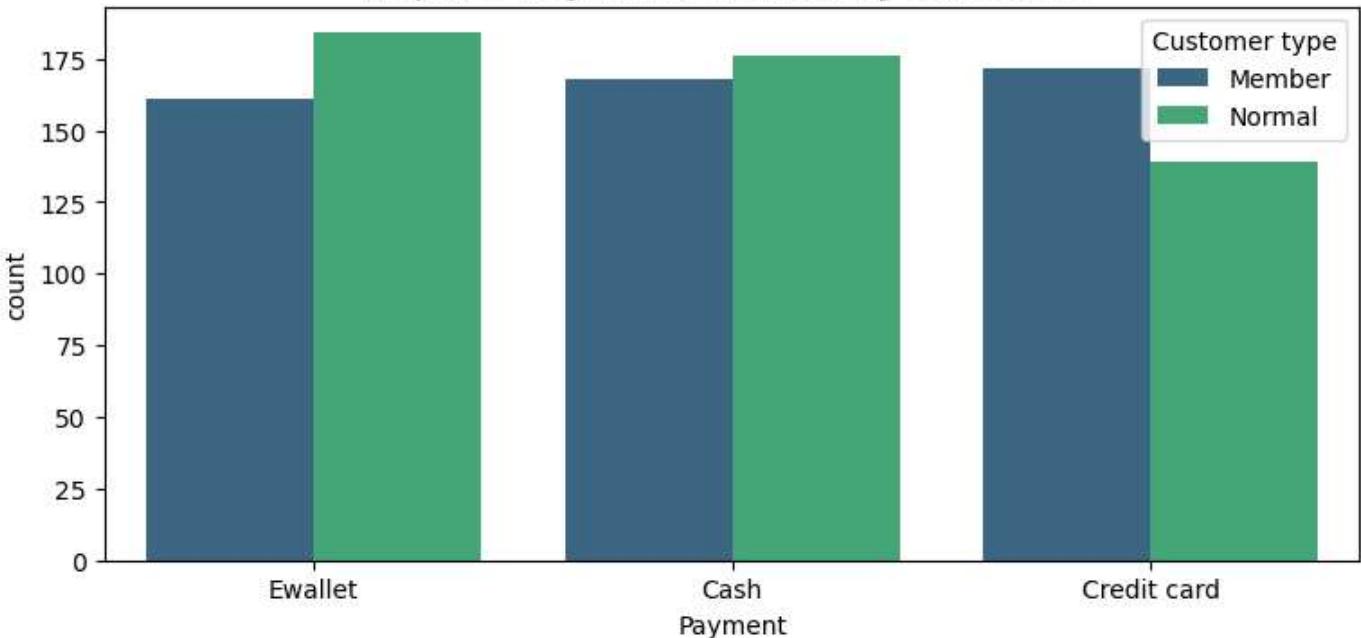
```
<AxesSubplot:xlabel='Branch', ylabel='count'>
```



Customer type and their payment method

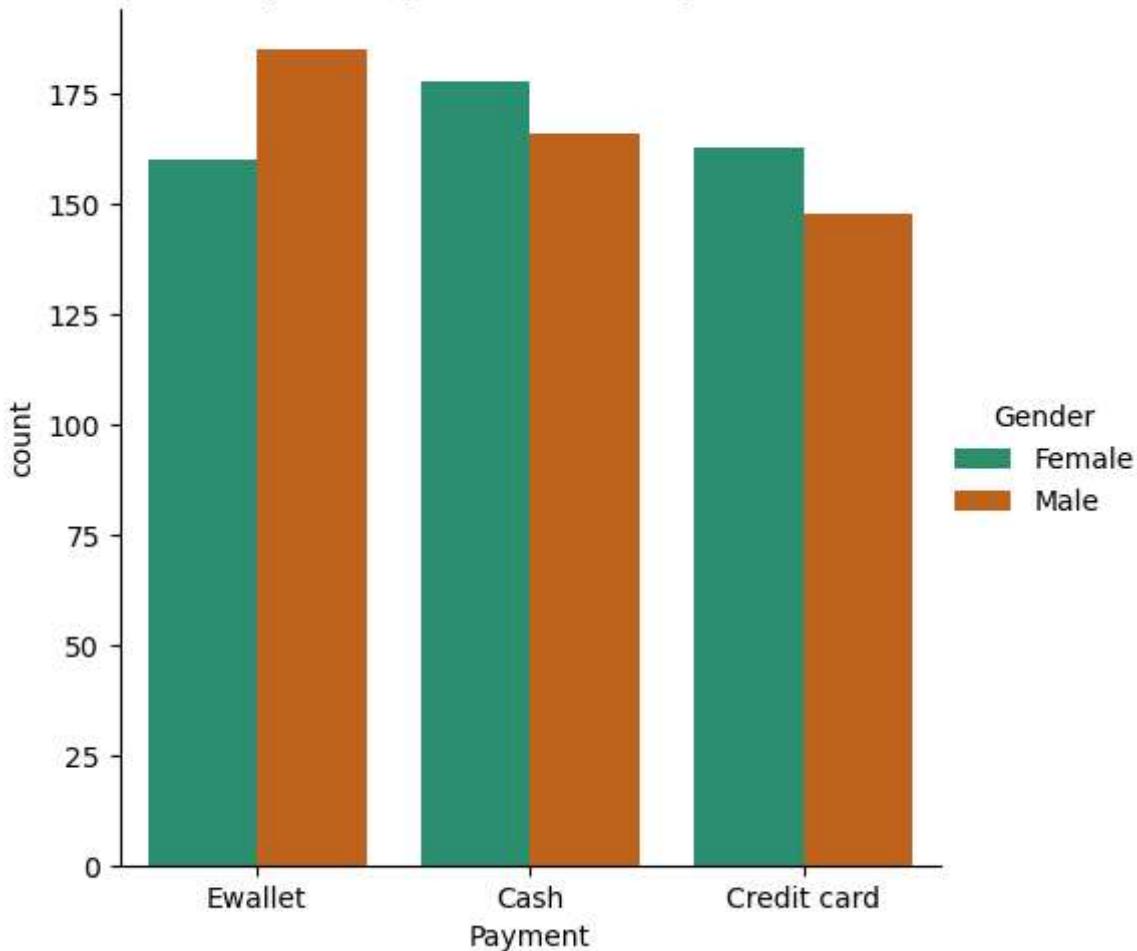
```
Text(0.5, 1.0, 'Popular Payment method by Customer')
```

Popular Payment method by Customer

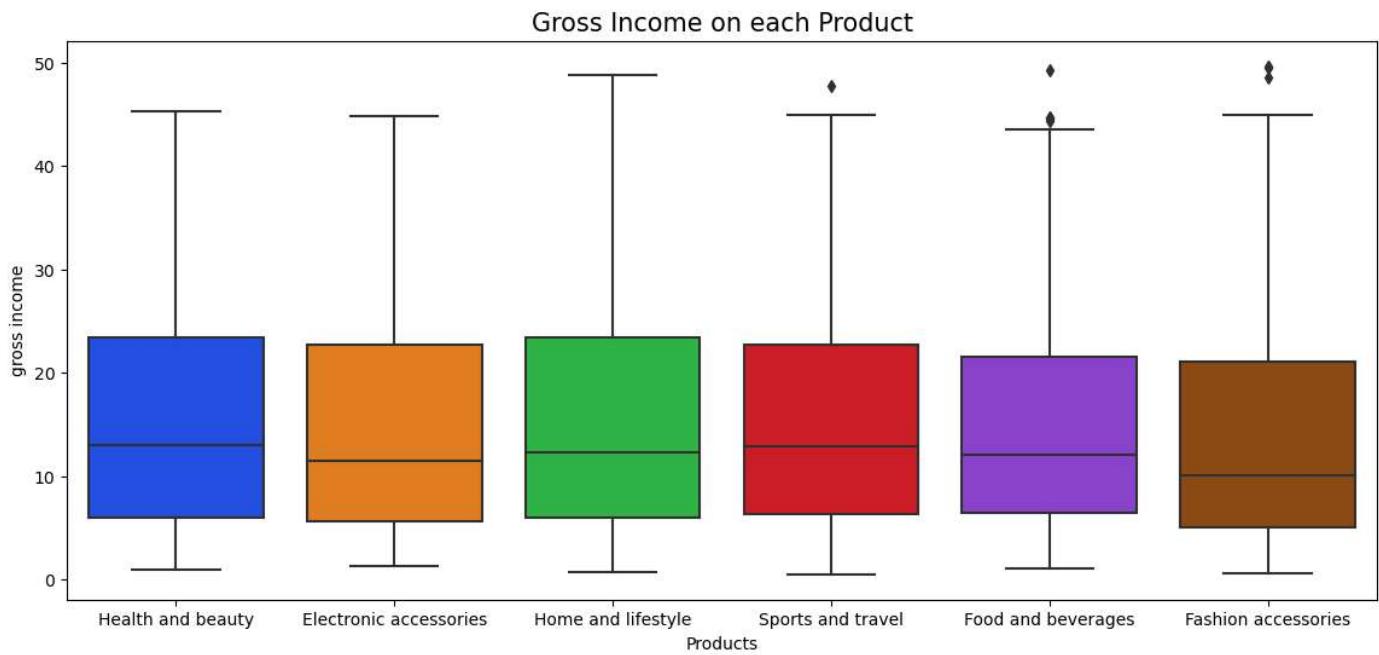


```
Text(0.5, 1.0, 'Popular Payment method by Gender')
```

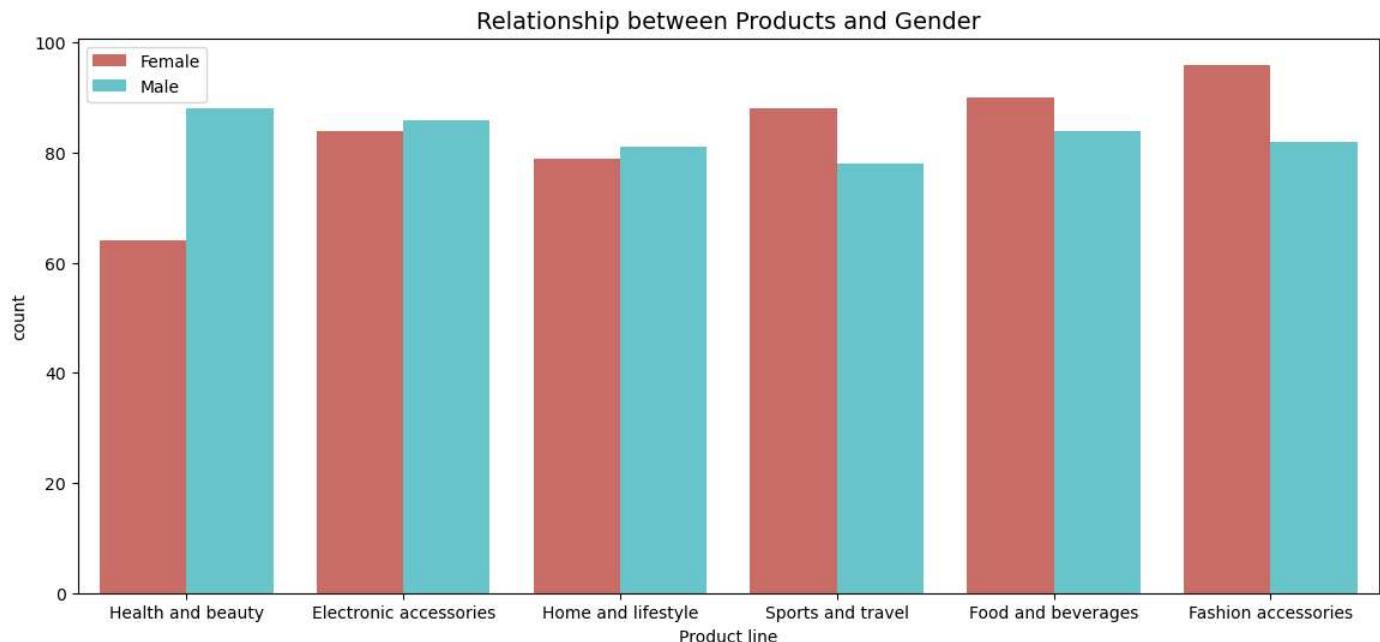
Popular Payment method by Gender



'Gross Income' on every product



Relationship between categorical variables

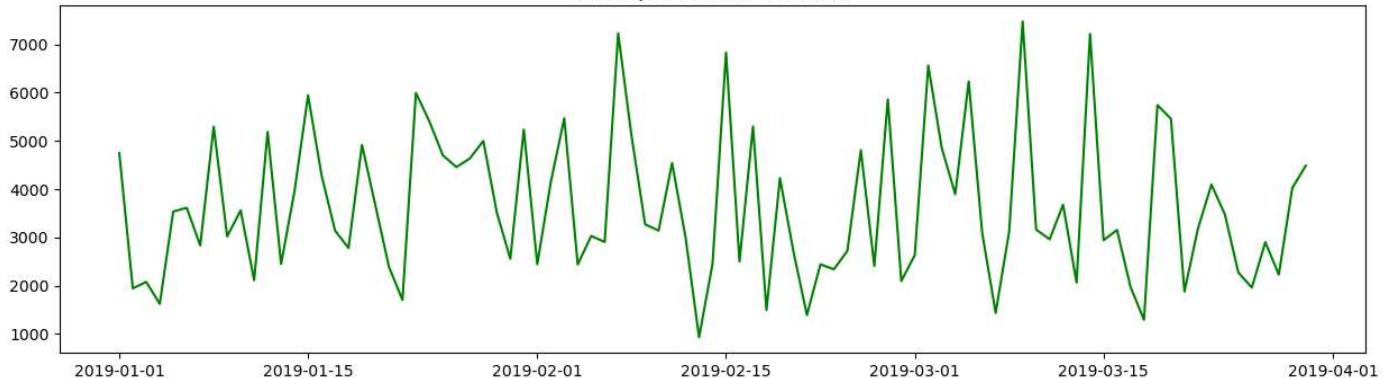


Gross-Income in each branch

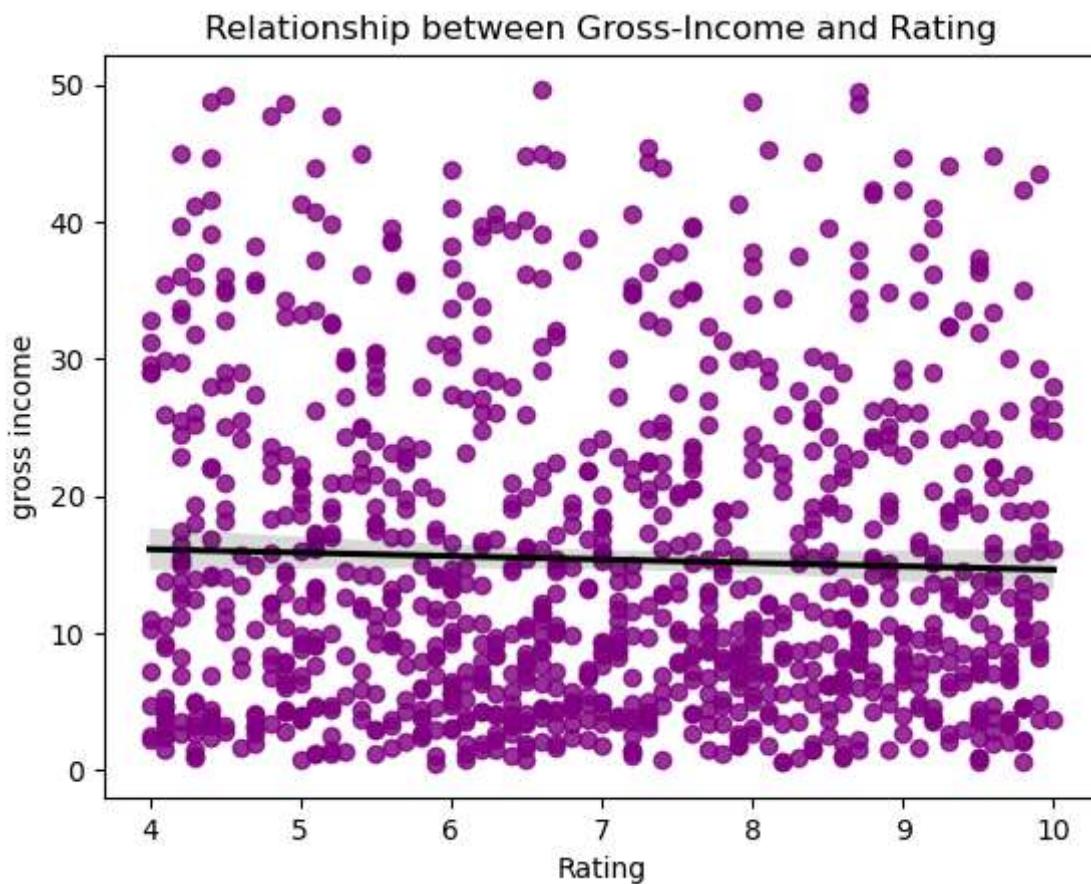


Time-period Distribution

Time period distribution

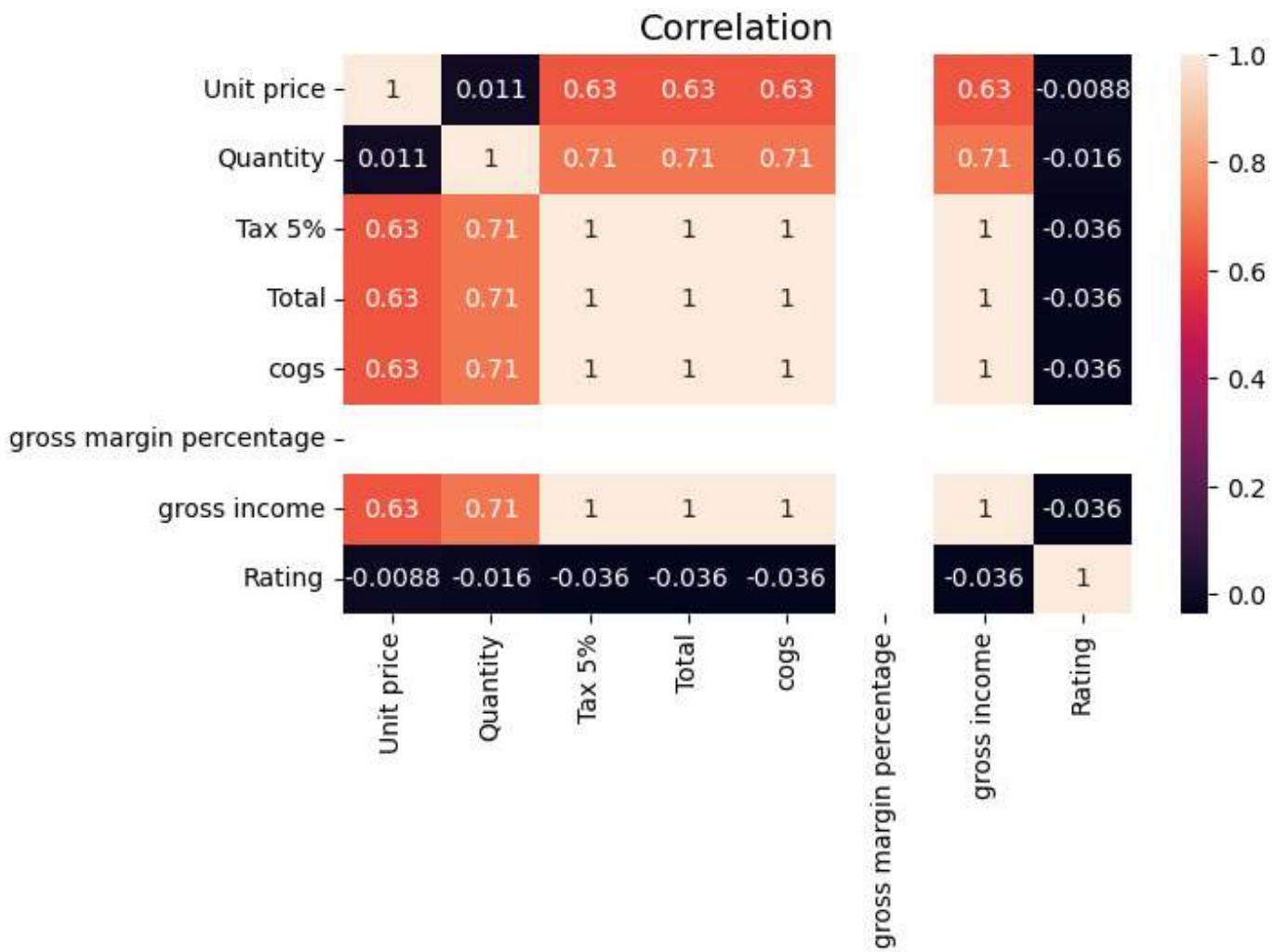


Relationship between rating and gross income

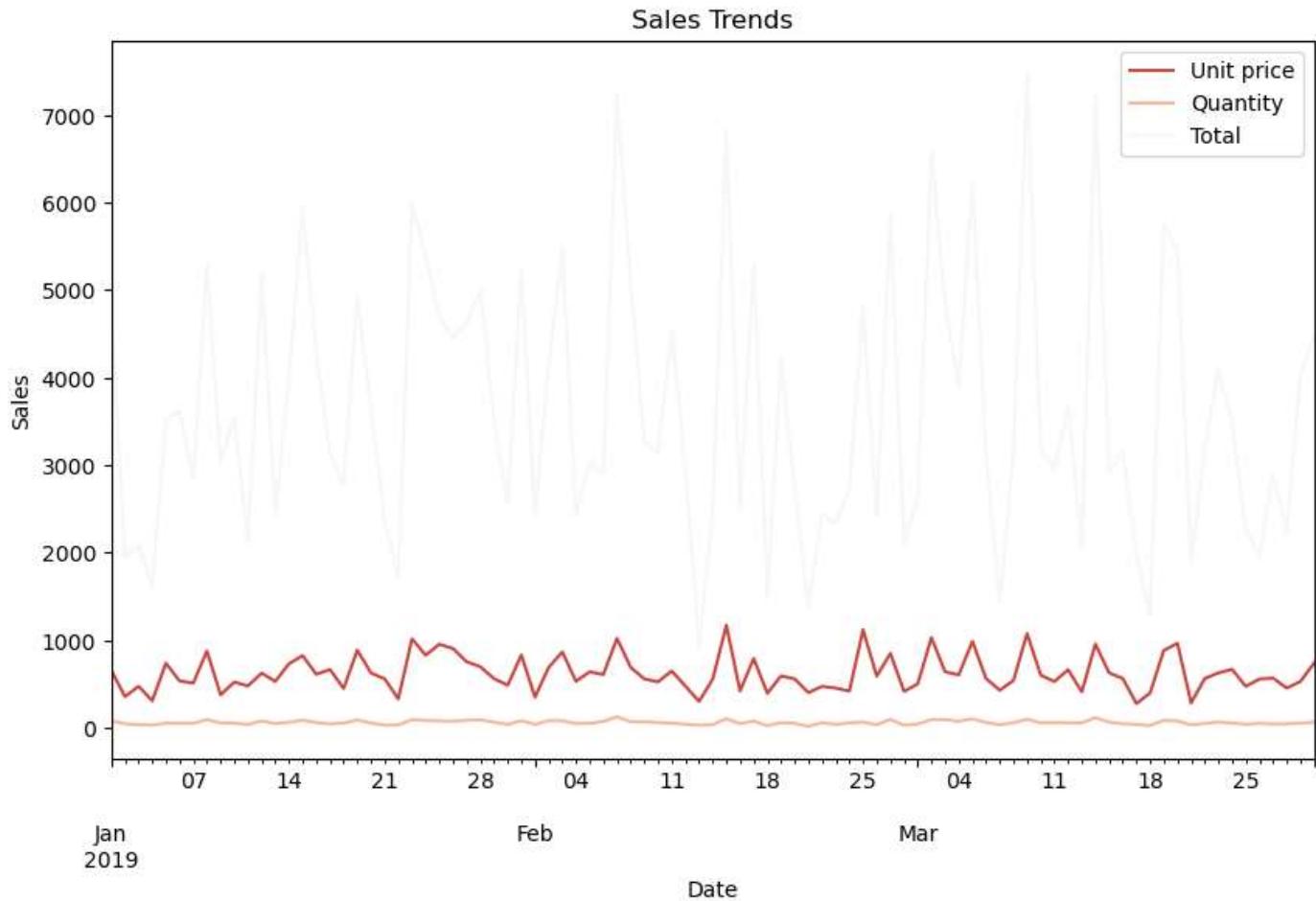


correlation

	Unit price	Quantity	Tax 5%	Total	cogs	gross margin percentage	gross income	Rating
Unit price	1.000000	0.010778	0.633962	0.633962	0.633962	NaN	0.633962	-0.008778
Quantity	0.010778	1.000000	0.705510	0.705510	0.705510	NaN	0.705510	-0.015815
Tax 5%	0.633962	0.705510	1.000000	1.000000	1.000000	NaN	1.000000	-0.036442
Total	0.633962	0.705510	1.000000	1.000000	1.000000	NaN	1.000000	-0.036442
cogs	0.633962	0.705510	1.000000	1.000000	1.000000	NaN	1.000000	-0.036442
gross margin percentage	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
gross income	0.633962	0.705510	1.000000	1.000000	1.000000	NaN	1.000000	-0.036442
Rating	-0.008778	-0.015815	-0.036442	-0.036442	-0.036442	NaN	-0.036442	1.000000



Sales Trends



percentage of total ratings for each rating

