super-store-sales

May 2, 2023

[12]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sb
df=pd.read_csv('supermarket_sales - Sheet1.csv')
```

[111]:
```
df.head(10)
```

[111]:

	Invoice ID	Branch	City	Customer type	Gender	Product line	Unit price	Quantity	Tax 5%	Total	Date	Time	Payment	cogs	gross margin percentage	gross income	Rating
0	750-67-8428	A	Yangon	Member	Female	Health and beauty	74.69	7	26.1415	548.9715	1/5/2019	13:08	Ewallet	522.83	4.761905	26.1415	9.1
1	226-31-3081	C	Naypyitaw	Normal	Female	Electronic accessories	15.28	5	3.8200	80.2200	3/8/2019	10:29	Cash	76.40	4.761905	3.8200	9.6
2	631-41-3108	A	Yangon	Normal	Male	Home and lifestyle	46.33	7	16.2155	340.5255	3/3/2019	13:23	Credit card	324.31	4.761905	16.2155	7.4
3	123-19-1176	A	Yangon	Member	Male	Health and beauty	58.22	8	23.2880	489.0480	1/27/2019	20:33	Ewallet	465.76	4.761905	23.2880	8.4
4	373-73-7910	A	Yangon	Normal	Male	Sports and travel	86.31	7	30.2085	634.3785	2/8/2019	10:37	Ewallet	604.17	4.761905	30.2085	5.3
5	699-14-3026	C	Naypyitaw	Normal	Male	Electronic accessories	85.39	7	29.8865	627.6165	3/25/2019	18:30	Ewallet	597.73	4.761905	29.8865	4.1
6	355-53-5943	A	Yangon	Member	Female	Electronic accessories	68.84	6	20.6520	433.6920	2/25/2019	14:36	Ewallet	413.04	4.761905	20.6520	5.8
7	315-22-5665	C	Naypyitaw	Normal	Female	Home and lifestyle	73.56	10	36.7800	772.3800	2/24/2019	11:38	Ewallet	735.60	4.761905	36.7800	8.0
8	665-32-9167	A	Yangon	Member	Female	Health and beauty	36.26	2	3.6260	76.1460	1/10/2019	17:15	Credit card	72.52	4.761905	3.6260	7.2
9	692-92-5582	B	Mandalay	Member	Female	Food and beverages	54.84	3	8.2260	172.7460	2/20/2019	13:27	Credit card	164.52	4.761905	8.2260	5.9

```
5  18:30      Ewallet  597.73                4.761905       29.8865     4.1
6  14:36      Ewallet  413.04                4.761905       20.6520     5.8
7  11:38      Ewallet  735.60                4.761905       36.7800     8.0
8  17:15  Credit card   72.52                4.761905        3.6260     7.2
9  13:27  Credit card  164.52                4.761905        8.2260     5.9
```

[15]:
```
df.columns
```

[15]:
```
Index(['Invoice ID', 'Branch', 'City', 'Customer type', 'Gender',
       'Product line', 'Unit price', 'Quantity', 'Tax 5%', 'Total', 'Date',
       'Time', 'Payment', 'cogs', 'gross margin percentage', 'gross income',
       'Rating'],
      dtype='object')
```

[17]:
```
df.isnull().sum()
```

[17]:
```
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
```

[18]:
```
df.info()
```

```
<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
```

[19]:
```
df.max()
```

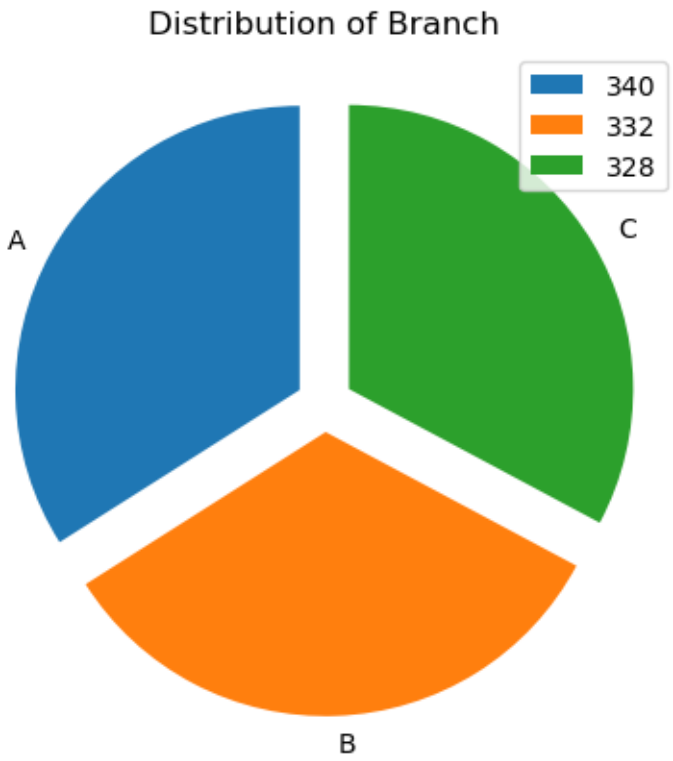
[19]:
```
Invoice ID                    898-04-2717
Branch                                  C
City                               Yangon
Customer type                      Normal
Gender                               Male
Product line             Sports and travel
Unit price                           99.96
Quantity                                10
Tax 5%                               49.65
Total                              1042.65
Date                              3/9/2019
Time                                 20:59
Payment                            Ewallet
cogs                                 993.0
gross margin percentage           4.761905
gross income                         49.65
Rating                                10.0
dtype: object
```

[25]:
```
df['Branch'].value_counts()
```

[25]:
```
A    340
B    332
C    328
Name: Branch, dtype: int64
```

[117]:
```
x=df['Branch'].value_counts()
plt.pie(x,labels=['A','B','C'],startangle=90,explode=[0.1,0.1,0.1])
plt.title('Distribution of Branch')
plt.legend(x)
```

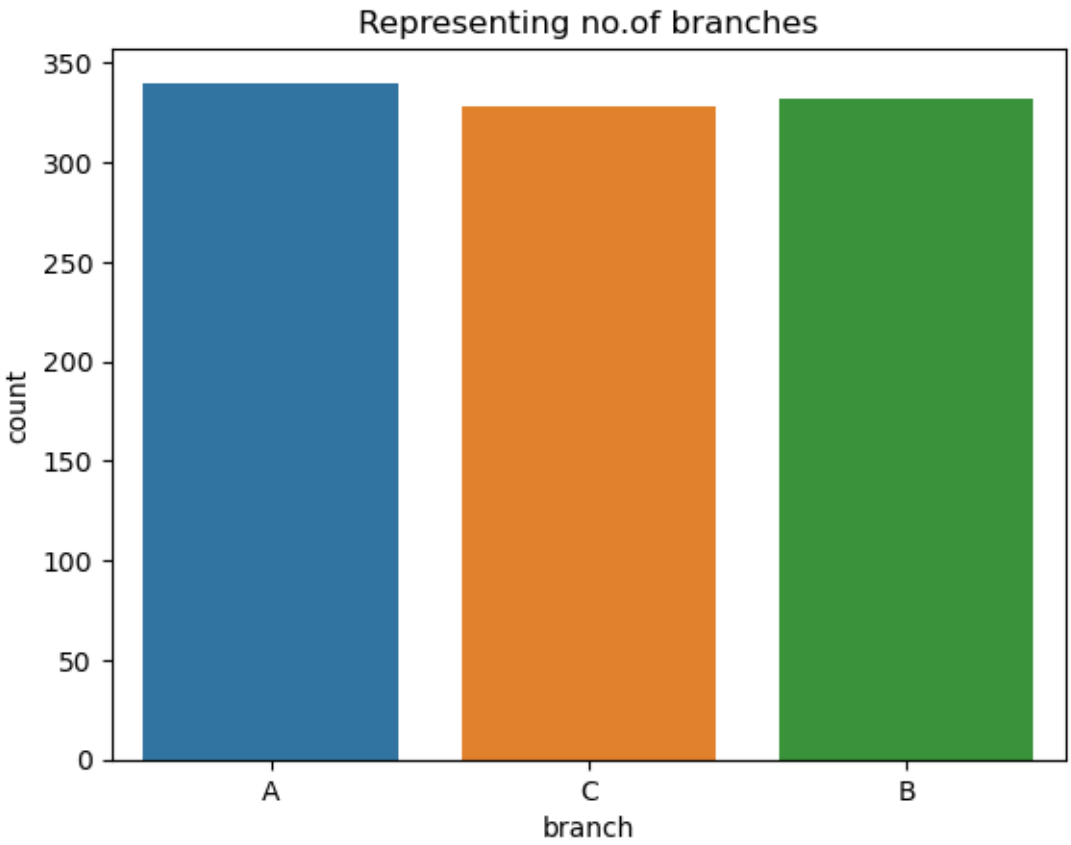
[117]: <matplotlib.legend.Legend at 0x269d2f9d6a0>



[40]:
```
x=df['Branch']
sb.countplot(x)
plt.xlabel('branch')
plt.title('Representing no.of branches')
```

```
C:\Users\Mokshogna Teja\anaconda3\lib\site-packages\seaborn\_decorators.py:36:
FutureWarning: Pass the following variable as a keyword arg: x. From version
0.12, the only valid positional argument will be `data`, and passing other
arguments without an explicit keyword will result in an error or
misinterpretation.
  warnings.warn(
```

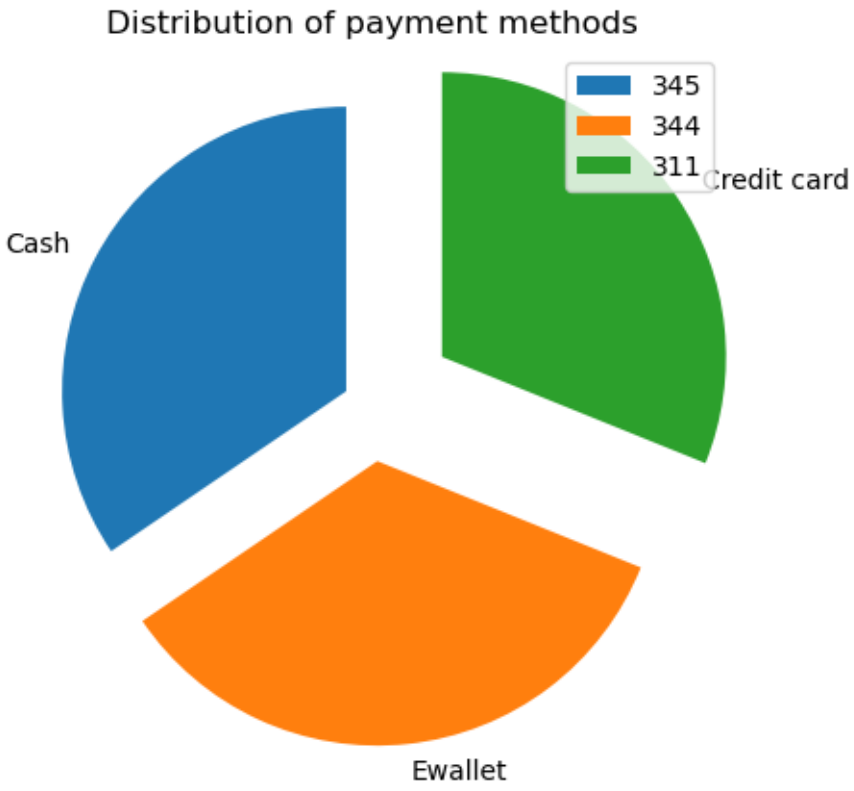
[40]: Text(0.5, 1.0, 'Representing no.of branches')



[70]:
```
x=df['Payment'].value_counts()
```

[105]:
```
x=df['Payment'].value_counts()
plt.pie(x,labels=['Cash','Ewallet','Credit card'],startangle=90,explode=[0.1,0.
↪2,0.3])
plt.title('Distribution of payment methods')
plt.legend(x)
```

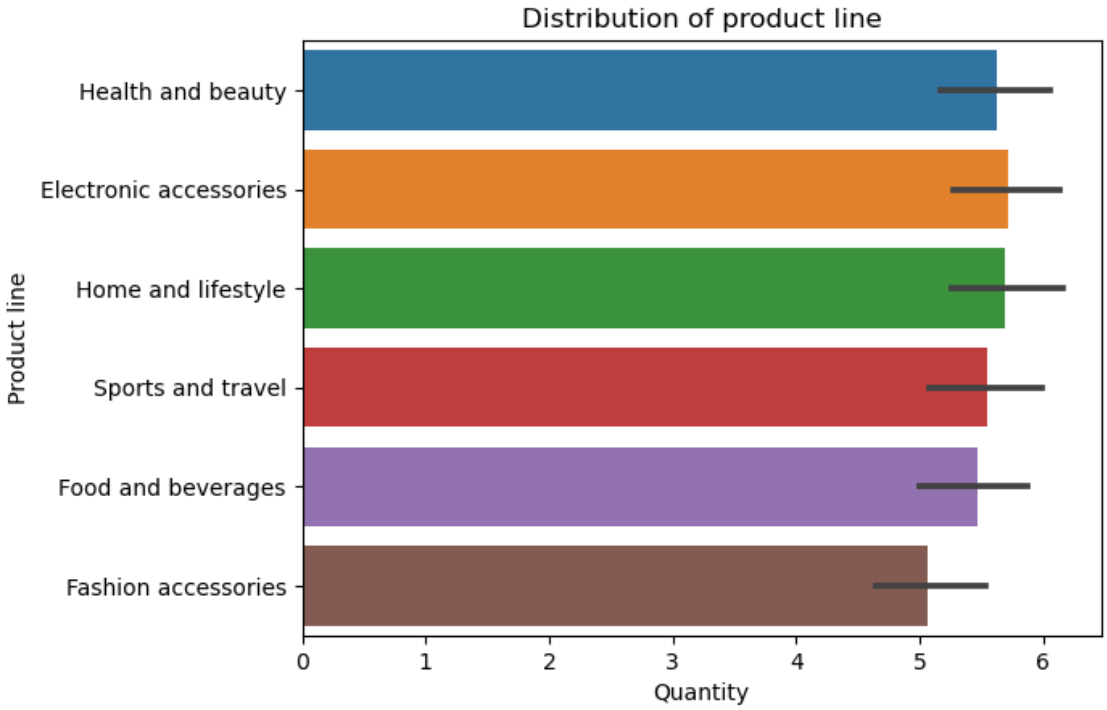
[105]:
```
<matplotlib.legend.Legend at 0x269d2c7d880>
```



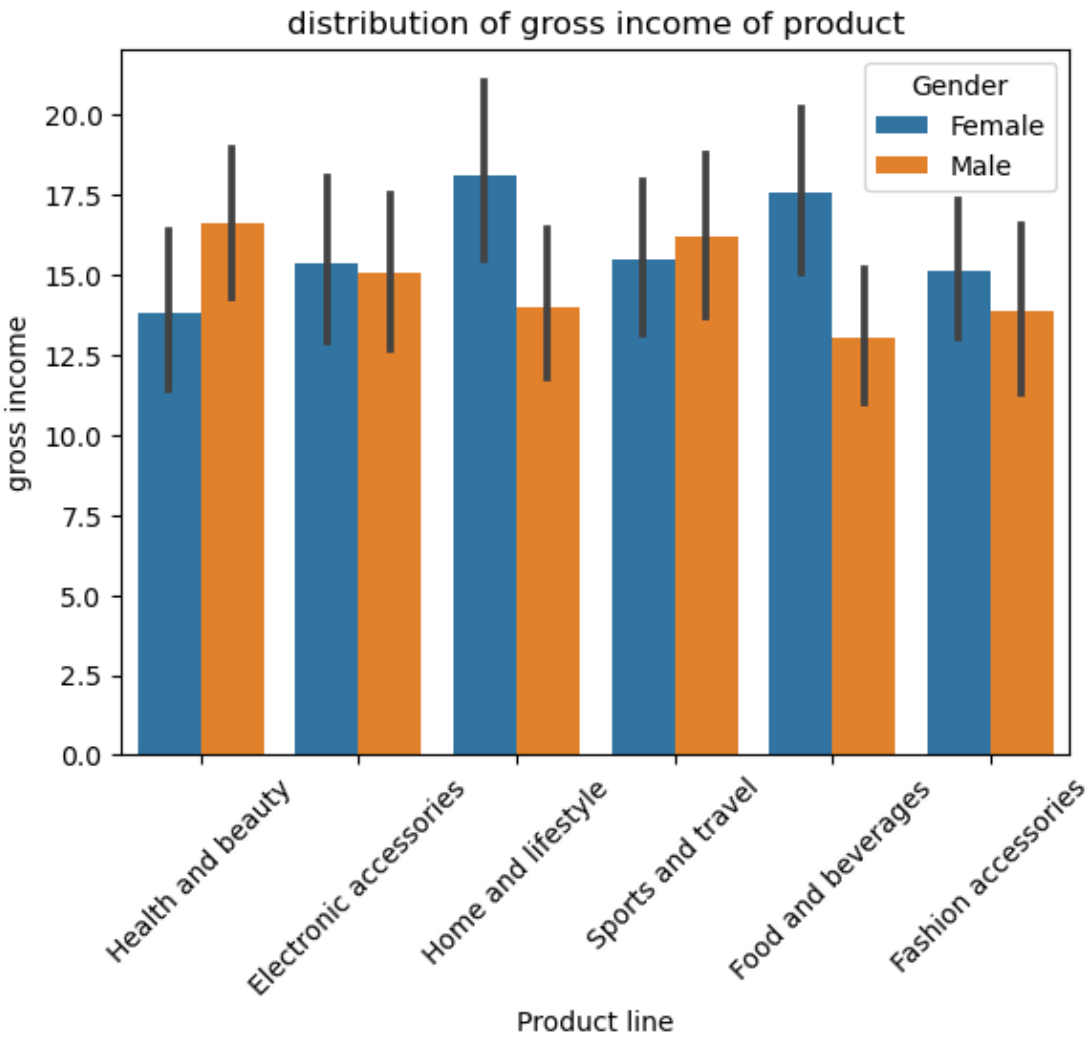
[89]:
```
y=df['Product line']
x=df['Quantity']
sb.barplot(x,y)
plt.xlabel('Quantity')
plt.ylabel('Product line')
plt.title('Distribution of product line')
```

```
C:\Users\Mokshogna Teja\anaconda3\lib\site-packages\seaborn\_decorators.py:36:
FutureWarning: Pass the following variables as keyword args: x, y. From version
0.12, the only valid positional argument will be `data`, and passing other
arguments without an explicit keyword will result in an error or
misinterpretation.
  warnings.warn(
```

[89]:
```
Text(0.5, 1.0, 'Distribution of product line')
```

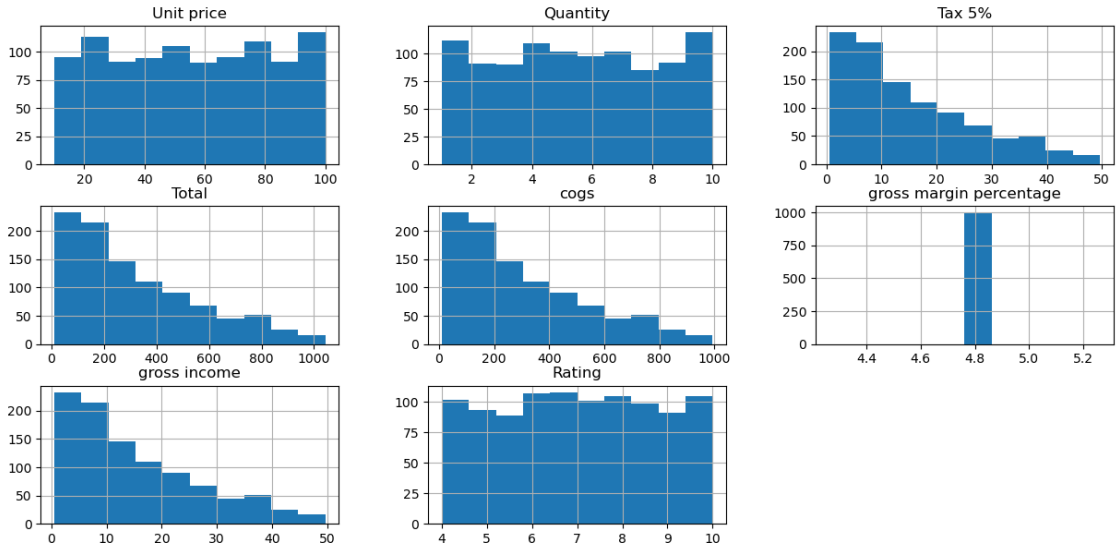


[93]:
```
sb.barplot(data = df, x = "Product line", y = "gross income", hue = "Gender")
plt.xticks(rotation = 45)
plt.title('distribution of gross income of product')
plt.show()
```



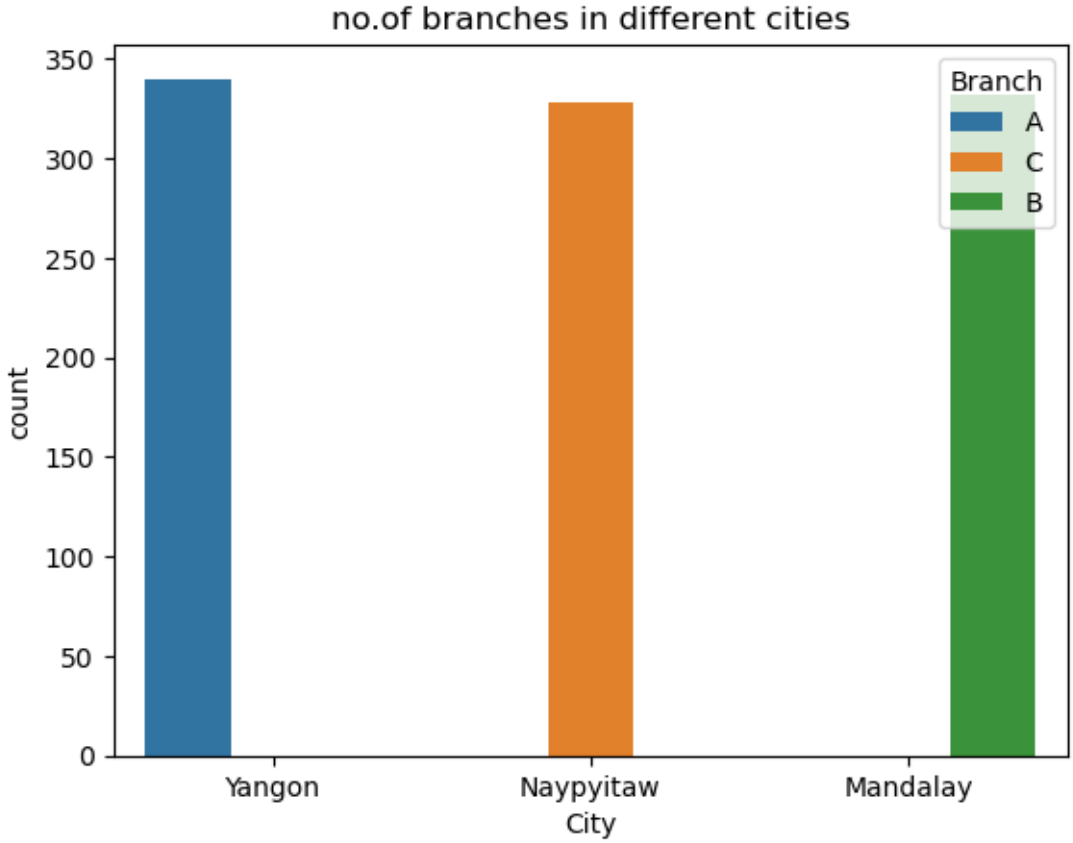
[94]:
```
df.hist(figsize = (15, 7))
```

[94]:
```
array([[<AxesSubplot:title={'center':'Unit price'}>,
        <AxesSubplot:title={'center':'Quantity'}>,
        <AxesSubplot:title={'center':'Tax 5%'}>],
       [<AxesSubplot:title={'center':'Total'}>,
        <AxesSubplot:title={'center':'cogs'}>,
        <AxesSubplot:title={'center':'gross margin percentage'}>],
       [<AxesSubplot:title={'center':'gross income'}>,
        <AxesSubplot:title={'center':'Rating'}>, <AxesSubplot:>]],
      dtype=object)
```



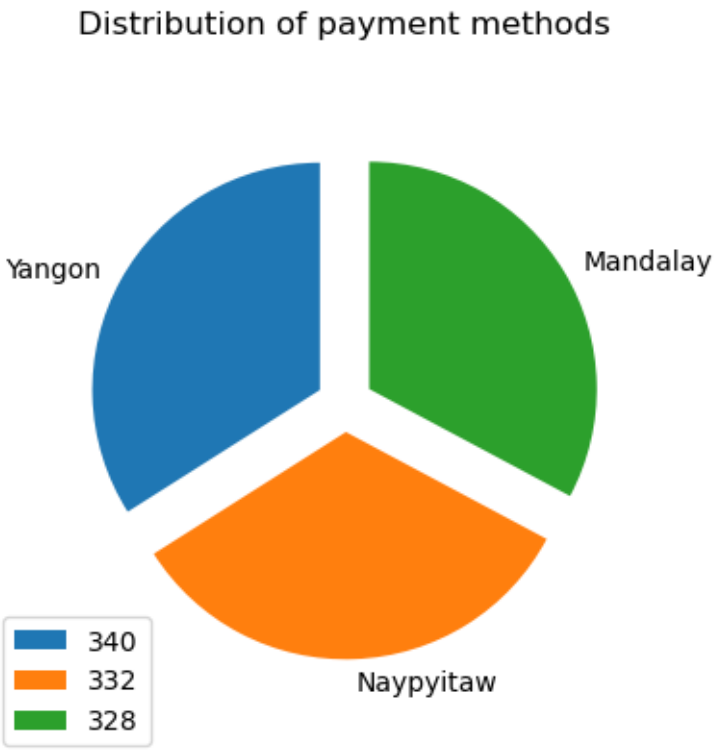
[102]:
```
sb.countplot(data= df, x='City', hue= 'Branch')
plt.title('no.of branches in different cities')
```

[102]: Text(0.5, 1.0, 'no.of branches in different cities')



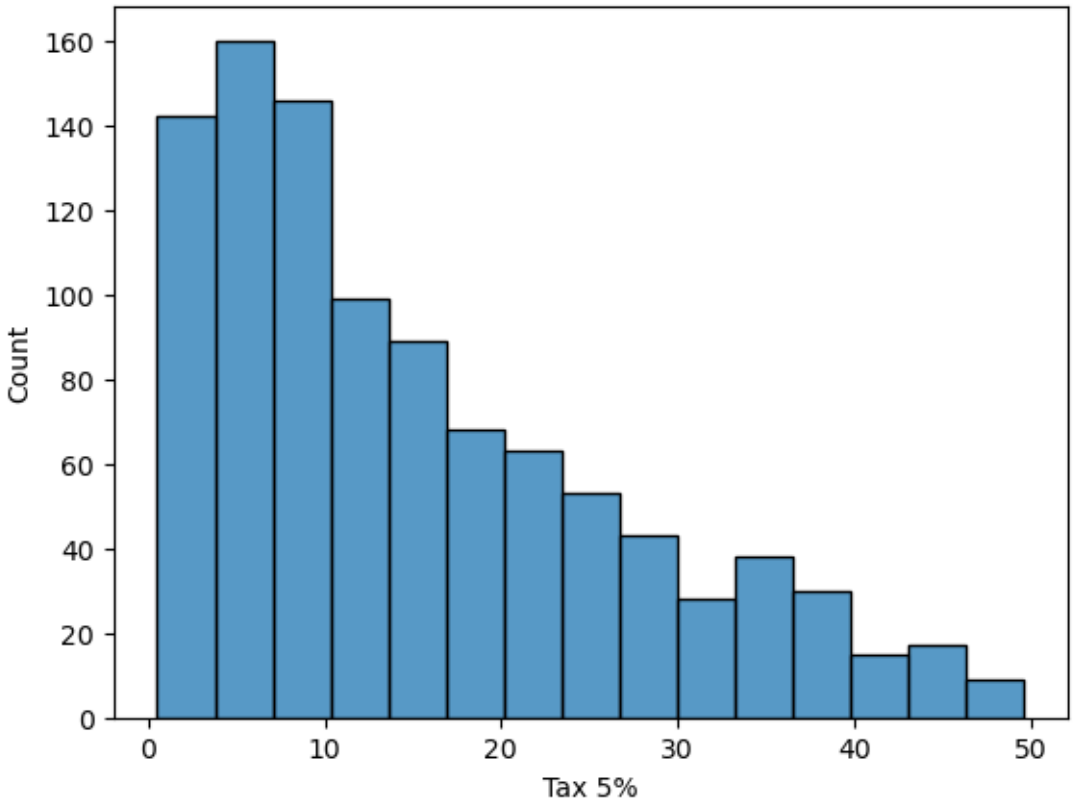
[114]:
```
x=df['City'].value_counts()
plt.pie(x,labels=['Yangon','Naypyitaw','Mandalay'],startangle=90,explode=[0.1,0.
↪1,0.1],radius=0.8)
plt.title('Distribution of payment methods')
plt.legend(x)
```

[114]: <matplotlib.legend.Legend at 0x269d2e0f220>

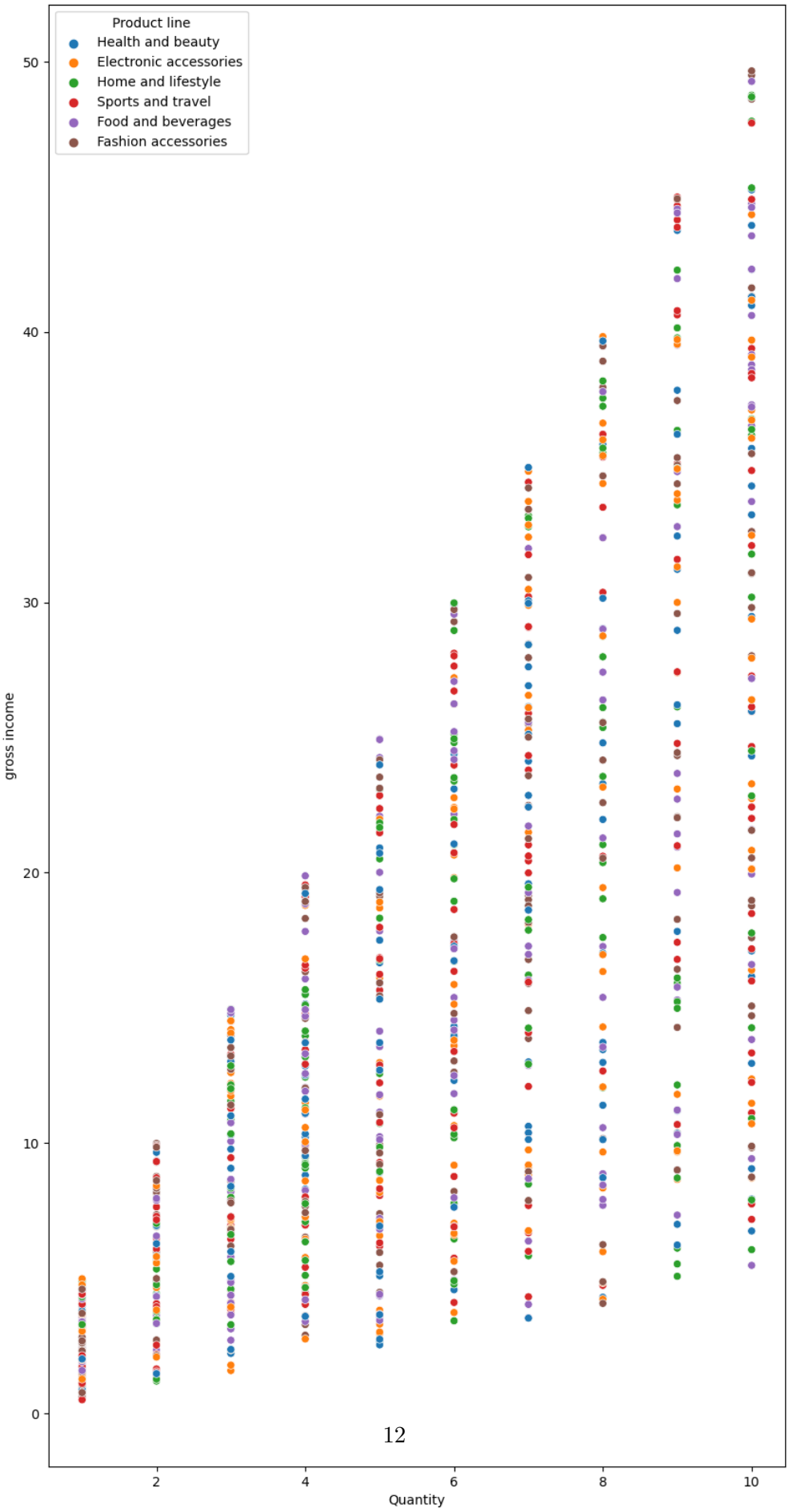


[115]:
```
y=df['Tax 5%']
sb.histplot(y)
```

[115]: <AxesSubplot:xlabel='Tax 5%', ylabel='Count'>



[127]:
```
sb.scatterplot(data=df,x='Quantity',y='gross income',hue='Product line')
plt.rcParams["figure.figsize"]=(5,10)
plt.show()
```

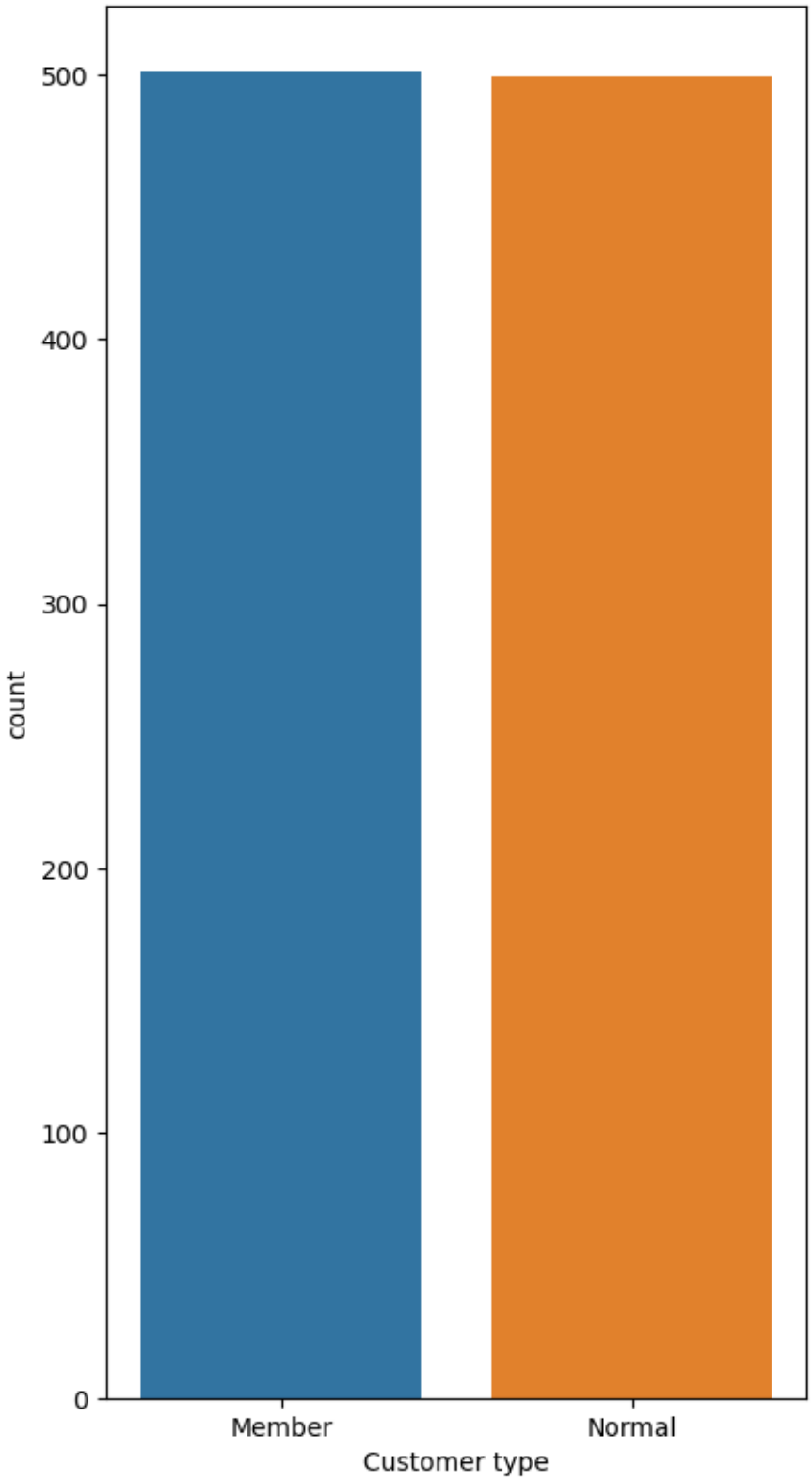

Product line
Health and beauty
Electronic accessories
Home and lifestyle
Sports and travel
Food and beverages
Fashion accessories
gross income
Quantity
0
10
20
30
40
50
2
4
6
8
10

[137]:

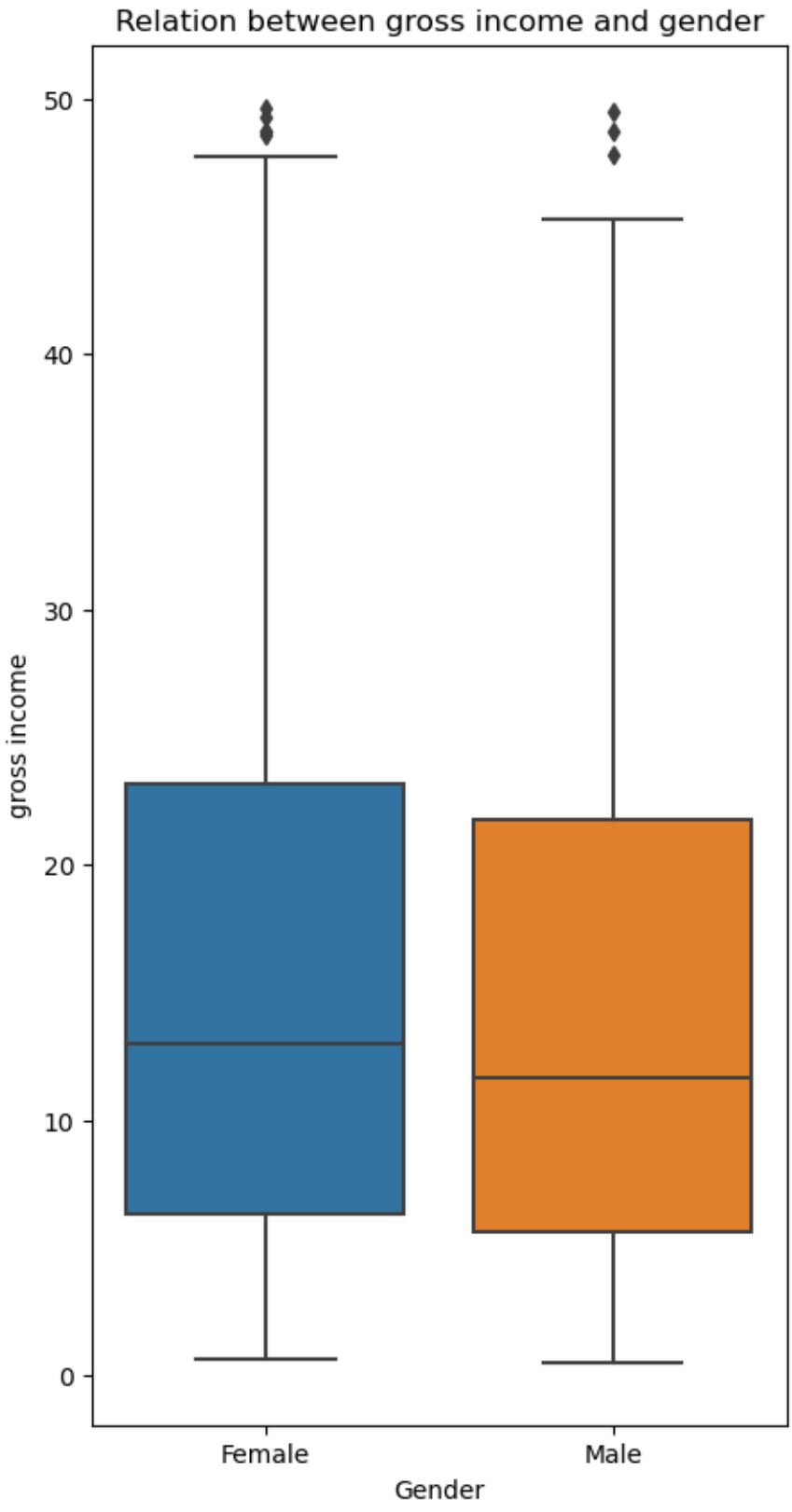
```
  File "C:\Users\Mokshogna Teja\AppData\Local\Temp\ipykernel_28872\1288981918.
↪py", line 6
    plt.show(
             ^
SyntaxError: unexpected EOF while parsing
```

[138]:
```
sb.countplot(data= df, x='Customer type')
```

[138]:
```
<AxesSubplot:xlabel='Customer type', ylabel='count'>
```


500
400
300
200
100
0
count
Member
Normal
Customer type

[141]:
```
sb.boxplot(x=df['Gender'],y=df['gross income'])
plt.title('Relation between gross income and gender')
plt.show()
```


Relation between gross income and gender
gross income
50
40
30
20
10
0
Female
Male
Gender

[ ]: