```
In [3]: import pandas as pd
```

```
In [4]: df= pd.read_csv("supermarket_sales - Sheet1.csv")
```

```
In [5]: df.head()
```

Out[5]:

	Invoice ID	Branch	City	Customer type	Gender	Product line	Unit price	Quantity	Tax 5%	
0	750-67-8428	A	Yangon	Member	Female	Health and beauty	74.69	7	26.1415	548.
1	226-31-3081	C	Naypyitaw	Normal	Female	Electronic accessories	15.28	5	3.8200	80.
2	631-41-3108	A	Yangon	Normal	Male	Home and lifestyle	46.33	7	16.2155	340.
3	123-19-1176	A	Yangon	Member	Male	Health and beauty	58.22	8	23.2880	489.
4	373-73-7910	A	Yangon	Normal	Male	Sports and travel	86.31	7	30.2085	634.

```
In [6]: df.shape
```

Out[6]: (1000, 17)

```
In [7]: df.isnull().sum()
```

```
Out[7]: 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
```

```
In [8]:  df.dtypes
```

```
Out[8]:  Invoice ID                 object
         Branch                     object
         City                       object
         Customer type              object
         Gender                     object
         Product line               object
         Unit price                float64
         Quantity                    int64
         Tax 5%                    float64
         Total                     float64
         Date                       object
         Time                       object
         Payment                    object
         cogs                      float64
         gross margin percentage   float64
         gross income              float64
         Rating                    float64
         dtype: object
```

```
In [9]:  df['Quantity'] = df['Quantity'].astype(object)
```

```
In [10]:  df.dtypes
```

```
Out[10]:  Invoice ID                 object
          Branch                     object
          City                       object
          Customer type              object
          Gender                     object
          Product line               object
          Unit price                float64
          Quantity                   object
          Tax 5%                    float64
          Total                     float64
          Date                       object
          Time                       object
          Payment                    object
          cogs                      float64
          gross margin percentage   float64
          gross income              float64
          Rating                    float64
          dtype: object
```

```
In [11]:  df['Gender'].value_counts()
```

```
Out[11]:  Gender
          Female    501
          Male      499
          Name: count, dtype: int64
```

```
In [12]:  df['Gender'] = df['Gender'].map({'Male':0,'Female':1})
```

```
In [13]: df.head()
```

Out[13]:

	Invoice ID	Branch	City	Customer type	Gender	Product line	Unit price	Quantity	Tax 5%	
0	750-67-8428	A	Yangon	Member	1	Health and beauty	74.69	7	26.1415	548.
1	226-31-3081	C	Naypyitaw	Normal	1	Electronic accessories	15.28	5	3.8200	80.
2	631-41-3108	A	Yangon	Normal	0	Home and lifestyle	46.33	7	16.2155	340.
3	123-19-1176	A	Yangon	Member	0	Health and beauty	58.22	8	23.2880	489.
4	373-73-7910	A	Yangon	Normal	0	Sports and travel	86.31	7	30.2085	634.

```
In [ ]:
```