In [14]:
```
import pandas as pd
df = pd.read_csv("D:/intern elevate/Task_Project/Retail/orders.csv",)
df.head(20)
```

Out[14]:

	Order Id	Order Date	Ship Mode	Segment	Country	City	State	Postal Code	Regio
0	1	01-03-2023	Second Class	Consumer	United States	Henderson	Kentucky	42420	Soutl
1	2	15-08-2023	Second Class	Consumer	United States	Henderson	Kentucky	42420	Soutl
2	3	10-01-2023	Second Class	Corporate	United States	Los Angeles	California	90036	Wes
3	4	18-06-2022	Standard Class	Consumer	United States	Fort Lauderdale	Florida	33311	Soutl
4	5	13-07-2022	Standard Class	Consumer	United States	Fort Lauderdale	Florida	33311	Soutl
5	6	13-03-2022	Not Available	Consumer	United States	Los Angeles	California	90032	Wes
6	7	28-12-2022	Standard Class	Consumer	United States	Los Angeles	California	90032	Wes
7	8	25-01-2022	Standard Class	Consumer	United States	Los Angeles	California	90032	Wes
8	9	23-03-2023	Not Available	Consumer	United States	Los Angeles	California	90032	Wes
9	10	16-05-2023	Standard Class	Consumer	United States	Los Angeles	California	90032	Wes
10	11	31-03-2023	Not Available	Consumer	United States	Los Angeles	California	90032	Wes
11	12	25-12-2023	Not Available	Consumer	United States	Los Angeles	California	90032	Wes
12	13	11-02-2022	Standard Class	Consumer	United States	Concord	North Carolina	28027	Soutl
13	14	18-07-2023	Standard Class	Consumer	United States	Seattle	Washington	98103	Wes
14	15	09-11-2023	unknown	Home Office	United States	Fort Worth	Texas	76106	Centra

	Order Id	Order Date	Ship Mode	Segment	Country	City	State	Postal Code	Regio
15	16	18-06-2022	Standard Class	Home Office	United States	Fort Worth	Texas	76106	Centra
16	17	04-02-2022	Standard Class	Consumer	United States	Madison	Wisconsin	53711	Centra
17	18	04-08-2023	Second Class	Consumer	United States	West Jordan	Utah	84084	Wes
18	19	23-01-2022	Second Class	Consumer	United States	San Francisco	California	94109	Wes
19	20	11-01-2022	Second Class	Consumer	United States	San Francisco	California	94109	Wes

In [15]:
```
df['Ship Mode'].unique()
```

Out[15]:
```
array(['Second Class', 'Standard Class', 'Not Available', 'unknown',
       'First Class', nan, 'Same Day'], dtype=object)
```

In [16]:
```
df = pd.read_csv("D:/intern elevate/Task_Project/Retail/orders.csv",na_values=['
df['Ship Mode'].unique()
```

Out[16]:
```
array(['Second Class', 'Standard Class', nan, 'First Class', 'Same Day'],
      dtype=object)
```

In [17]:
```
df.columns
```

Out[17]:
```
Index(['Order Id', 'Order Date', 'Ship Mode', 'Segment', 'Country', 'City',
       'State', 'Postal Code', 'Region', 'Category', 'Sub Category',
       'Product Id', 'cost price', 'List Price', 'Quantity',
       'Discount Percent'],
      dtype='object')
```

In [20]:
```
df.columns=df.columns.str.lower()
df.columns=df.columns.str.replace(' ','_')
df. columns
```

Out[20]:
```
Index(['order_id', 'order_date', 'ship_mode', 'segment', 'country', 'city',
       'state', 'postal_code', 'region', 'category', 'sub_category',
       'product_id', 'cost_price', 'list_price', 'quantity',
       'discount_percent'],
      dtype='object')
```

In [21]:
```
df.head(5)
```

Out[21]:

	order_id	order_date	ship_mode	segment	country	city	state	postal_co
0	1	01-03-2023	Second Class	Consumer	United States	Henderson	Kentucky	42
1	2	15-08-2023	Second Class	Consumer	United States	Henderson	Kentucky	42
2	3	10-01-2023	Second Class	Corporate	United States	Los Angeles	California	90
3	4	18-06-2022	Standard Class	Consumer	United States	Fort Lauderdale	Florida	33
4	5	13-07-2022	Standard Class	Consumer	United States	Fort Lauderdale	Florida	33

In [24]:
```
df['discount']=df['list_price']*df['discount_percent']*.01
df['sale_price']= df['list_price']-df['discount']
df['profit']=df['sale_price']-df['cost_price']
df
```

Out[24]:

	order_id	order_date	ship_mode	segment	country	city	state	pos
0	1	01-03-2023	Second Class	Consumer	United States	Henderson	Kentucky	
1	2	15-08-2023	Second Class	Consumer	United States	Henderson	Kentucky	
2	3	10-01-2023	Second Class	Corporate	United States	Los Angeles	California	
3	4	18-06-2022	Standard Class	Consumer	United States	Fort Lauderdale	Florida	
4	5	13-07-2022	Standard Class	Consumer	United States	Fort Lauderdale	Florida	
...	...	...	...	...	...	...	...	
9989	9990	18-02-2023	Second Class	Consumer	United States	Miami	Florida	
9990	9991	17-03-2023	Standard Class	Consumer	United States	Costa Mesa	California	
9991	9992	07-08-2022	Standard Class	Consumer	United States	Costa Mesa	California	
9992	9993	19-11-2022	Standard Class	Consumer	United States	Costa Mesa	California	
9993	9994	17-07-2022	Second Class	Consumer	United States	Westminster	California	

9994 rows × 19 columns

In [25]:
```
df.dtypes
```

Out[25]:
```
order_id              int64
order_date           object
ship_mode            object
segment              object
country              object
city                 object
state                object
postal_code           int64
region               object
category             object
sub_category         object
product_id           object
cost_price            int64
list_price            int64
quantity              int64
discount_percent      int64
discount            float64
sale_price          float64
profit              float64
dtype: object
```

In [29]:
```
pd.to_datetime(df['order_date'],format="%d-%m-%Y")
```

Out[29]:
```
0      2023-03-01
1      2023-08-15
2      2023-01-10
3      2022-06-18
4      2022-07-13
          ...
9989   2023-02-18
9990   2023-03-17
9991   2022-08-07
9992   2022-11-19
9993   2022-07-17
Name: order_date, Length: 9994, dtype: datetime64[ns]
```

In [30]:
```
df['order_date']=pd.to_datetime(df['order_date'],format="%d-%m-%Y")
df.dtypes
```

Out[30]:
```
order_id                     int64
order_date          datetime64[ns]
ship_mode                   object
segment                     object
country                     object
city                        object
state                       object
postal_code                  int64
region                      object
category                    object
sub_category                object
product_id                  object
cost_price                   int64
list_price                   int64
quantity                     int64
discount_percent             int64
discount                   float64
sale_price                 float64
profit                     float64
dtype: object
```

In [34]:
```
df.drop(columns=['list_price','cost_price','discount_percent'],inplace=True)
```

```
---------------------------------------------------------------------------
KeyError                                  Traceback (most recent call last)
Cell In[34], line 1
----> 1 df.drop(columns=['list_price','cost_price','discount_percent'],inplace=Tr
ue)

File C:\ProgramData\anaconda3\Lib\site-packages\pandas\core\frame.py:5581, in Dat
aFrame.drop(self, labels, axis, index, columns, level, inplace, errors)
   5433 def drop(
   5434     self,
   5435     labels: IndexLabel | None = None,
   (...)
   5442     errors: IgnoreRaise = "raise",
   5443 ) -> DataFrame | None:
   5444     """
   5445     Drop specified labels from rows or columns.
   5446
   (...)
   5579             weight  1.0     0.8
   5580     """
-> 5581     return super().drop(
   5582         labels=labels,
   5583         axis=axis,
   5584         index=index,
   5585         columns=columns,
   5586         level=level,
   5587         inplace=inplace,
   5588         errors=errors,
   5589     )

File C:\ProgramData\anaconda3\Lib\site-packages\pandas\core\generic.py:4788, in N
DFrame.drop(self, labels, axis, index, columns, level, inplace, errors)
   4786 for axis, labels in axes.items():
   4787     if labels is not None:
-> 4788         obj = obj._drop_axis(labels, axis, level=level, errors=errors)
   4790 if inplace:
   4791     self._update_inplace(obj)

File C:\ProgramData\anaconda3\Lib\site-packages\pandas\core\generic.py:4830, in N
DFrame._drop_axis(self, labels, axis, level, errors, only_slice)
   4828         new_axis = axis.drop(labels, level=level, errors=errors)
   4829     else:
-> 4830         new_axis = axis.drop(labels, errors=errors)
   4831     indexer = axis.get_indexer(new_axis)
   4833 # Case for non-unique axis
   4834 else:

File C:\ProgramData\anaconda3\Lib\site-packages\pandas\core\indexes\base.py:7070,
in Index.drop(self, labels, errors)
   7068 if mask.any():
   7069     if errors != "ignore":
-> 7070         raise KeyError(f"{labels[mask].tolist()} not found in axis")
   7071     indexer = indexer[~mask]
   7072 return self.delete(indexer)

KeyError: "['list_price', 'cost_price', 'discount_percent'] not found in axis"
```

In [35]:
```
df
```

Out[35]:

	order_id	order_date	ship_mode	segment	country	city	state	pos
0	1	2023-03-01	Second Class	Consumer	United States	Henderson	Kentucky	
1	2	2023-08-15	Second Class	Consumer	United States	Henderson	Kentucky	
2	3	2023-01-10	Second Class	Corporate	United States	Los Angeles	California	
3	4	2022-06-18	Standard Class	Consumer	United States	Fort Lauderdale	Florida	
4	5	2022-07-13	Standard Class	Consumer	United States	Fort Lauderdale	Florida	
...	...	...	...	...	...	...	...	
9989	9990	2023-02-18	Second Class	Consumer	United States	Miami	Florida	
9990	9991	2023-03-17	Standard Class	Consumer	United States	Costa Mesa	California	
9991	9992	2022-08-07	Standard Class	Consumer	United States	Costa Mesa	California	
9992	9993	2022-11-19	Standard Class	Consumer	United States	Costa Mesa	California	
9993	9994	2022-07-17	Second Class	Consumer	United States	Westminster	California	

9994 rows × 16 columns

In [36]:
```
df.to_csv(r'D:\intern elevate\Task_Project\Retail\clean\order.csv')
```

In [ ]:
```

```