```
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	Regioi
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)
```

5/12/25,

3:36 PM					Unt	itled				
Out[21]:	or	der_id d	order_date	ship_mode	segment	country	city		state po	stal_co
	0	1	01-03- 2023	Second Class	Consumer	United States	Henderson	Kent	tucky	424
	1	2	15-08- 2023	Second Class	Consumer	United States	Henderson	Kent	tucky	424
	2	3	10-01- 2023	Second Class	Corporate	United States	Los Angeles	Calif	ornia	901
	3	4	18-06- 2022	Standard Class	Consumer	United States	Fort Lauderdale	Flo	orida	33
	4	5	13-07- 2022	Standard Class	Consumer	United States	Fort Lauderdale	Flo	orida	33:
	4									•
Tn [2/1].	df['d	iscount'	]=df['list	:_price']*d	_		t']*.01			
In [24]:	df['s	ale_prio	e']= df['] df['sale_p	list_price' price']-df[		_				
Out[24]:	df['s	ale_prio	df['sale_p		'cost_prio	_	try	city	state	pos
	df['s	ale_pric rofit']= <b>order_ic</b>	df['sale_p	orice']-df[  te ship_mo  3- Seco	'cost_prio	ent coun			<b>state</b> Kentucky	
	df['s df['p df	ale_pric rofit']= order_ic	df['sale_rdf order_da	te ship_mo 3- Seco 23 Cla	de segm and Consur	ent coun mer Uni Sta	ted Hende	erson		,
	df['s df['p df	ale_pric rofit']= order_ic	d order_da  1 01-0 20	te ship_mo 3- Seco 23 Cla 8- Seco 23 Cla 1- Seco	de segm and Consur	ent coun mer Uni Sta mer Sta	ted Hende ted Hende ted Hende	erson	Kentucky	,
	df['sdf['pdf]	ale_pric rofit']=  order_ic	d order_da  1 01-0 20 2 15-0 2 10-0	te ship_mo 3- Seco 23 Cla 8- Seco 23 Cla 1- Seco 23 Cla 6- Standa	de segm and Consur ass and Corpor	ent coun mer Uni Sta mer Uni Sta trate Uni	ted Hendertes  ted Hendertes  ted ted Los Angles	erson erson geles Fort	Kentucky	,
	df['s df['p df  0	ale_pric rofit']=  order_ic	d order_da  1 01-0 20 2 15-0 20 3 10-0 20 4 18-0	te ship_mo 3- Seco 23 Cla 8- Seco 23 Cla 1- Seco 23 Cla 6- Standa 22 Cla 7- Standa	de segm  de segm  do Consur  and Consur  and Corpor  and Corpor  ard ass  Consur	ent coun mer Uni Sta mer Uni Sta trate Uni Sta mer Uni	ted Hendeltes ted Hendeltes ted Los Angites ted Laudeltes	erson erson geles Fort rdale Fort	Kentucky  Kentucky  California	,
	df['s df['p df  0  1 2	ale_pric rofit']=  order_ic	d order_da  1 01-0 20 2 15-0 2 20 3 10-0 2 20 4 18-0 2 20 5 13-0	te ship_mo 3- Seco 23 Cla 8- Seco 23 Cla 1- Seco 23 Cla 6- Standa 22 Cla 7- Standa	de segm  and Consur  and Corpor  ard Consur  ard Consur  ard Consur  ard Consur	ent coun mer Uni Sta mer Uni Sta trate Uni Sta mer Uni	ted Hendeltes Hendeltes Los Angletes Laudeltes Laudelted	erson erson geles Fort rdale Fort	Kentucky  Kentucky  California  Florida	,
	df['s df['p df	ale_pric rofit']=  order_ic	d order_da  1 01-0 20 2 15-0 2 20 3 10-0 2 20 4 13-0 2 20	te ship_mo 3- Seco 23 Cla 8- Seco 23 Cla 1- Seco 23 Cla 6- Standa 22 Cla 7- Standa 22 Cla	de segm and Consur ass and Corpor ass ard Consur ard ass ard ard ass ard ard ass ard ard ass ard ard ass	ent coun mer Uni Sta mer Uni Sta mer Uni Sta	ted Hendel ted Hendel ted Los Ang ted Laudel ted Laudel ted Laudel	erson  geles  Fort rdale  Fort rdale	Kentucky  Kentucky  California  Florida	
	df['s df['p df']	ale_pric rofit']=  order_ic	df['sale_r df['sale_r d order_da 1 01-0 20 2 15-0 20 3 10-0 3 20 4 18-0 20 6 13-0 20 17-0	te ship_mo 3-	de segm and Consur and Consur and Corpor ard Consur	ent coun mer Uni Sta mer Uni	ted Hender ted Hender ted Los Anguers ted Lauder ted La	erson  geles  Fort rdale  Fort rdale   Miami	Kentucky  Kentucky  California  Florida  Florida	

9994 rows × 19 columns

9992

9993

9994

07-08-

2022

19-11-

17-07-

2022

2022

In [25]: df.dtypes

Standard

Standard

Class

Class

Class

Second

Consumer

Consumer

Consumer

United

States

United

States

United

States

Costa Mesa California

Westminster California

Costa Mesa

California

9991

9992

9993

```
Out[25]: order_id
                                int64
          order_date
                               object
          ship mode
                               object
          segment
                               object
          country
                               object
                               object
          city
          state
                               object
                                int64
          postal_code
                               object
          region
                               object
          category
          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
         pd.to datetime(df['order date'],format="%d-%m-%Y")
In [29]:
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]
         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
                                       object
          state
          postal_code
                                       int64
          region
                                       object
          category
                                       object
          sub_category
                                       object
          product_id
                                       object
          cost_price
                                       int64
          list_price
                                        int64
                                        int64
          quantity
          discount_percent
                                        int64
                                      float64
          discount
          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,
   (\ldots)
   5442
            errors: IgnoreRaise = "raise",
   5443 ) -> DataFrame | None:
  5444
   5445
            Drop specified labels from rows or columns.
   5446
   (\ldots)
   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,
                inplace=inplace,
   5587
   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":
                raise KeyError(f"{labels[mask].tolist()} not found in axis")
-> 7070
   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 city segment country state pos 2023-03-Second United 0 1 Consumer Henderson Kentucky Class States 01 2023-08-Second United 2 1 Consumer Henderson Kentucky Class States 15 2023-01-Second United 2 3 Corporate Los Angeles California Class States 10 2022-06-Standard United Fort 3 4 Consumer Florida Class States Lauderdale 18 2022-07-Standard United Fort 4 5 Consumer Florida 13 Class Lauderdale States 2023-02-Second United 9989 9990 Miami Florida Consumer Class States 18 2023-03-Standard United 9990 9991 Consumer Costa Mesa California Class 17 States Standard 2022-08-United Consumer 9991 9992 Costa Mesa California Class 07 States 2022-11-Standard United 9993 9992 Consumer California Costa Mesa 19 Class States 2022-07-Second United 9993 9994 Consumer Westminster California Class 17 States 9994 rows × 16 columns

In [36]: df.to\_csv(r'D:\intern elevate\Task\_Project\Retail\clean\order.csv')

In [ ]: