In [1]:
```
import pandas as pd
import numpy as np
```

In [2]:
```
# Handling null values
df = pd.read_csv(r"C:\Users\RYAN\Downloads\orders.csv",na_values=['Not Available', 'unknown'])
```

In [3]:
```
df.head()
```

Out[3]:

	Order Id	Order Date	Ship Mode	Segment	Country	City	State	Postal Code	Region	Category	Sub Category	Product Id	cost price	List Price
0	1	2023-03-01	Second Class	Consumer	United States	Henderson	Kentucky	42420	South	Furniture	Bookcases	FUR-BO-10001798	240	260
1	2	2023-08-15	Second Class	Consumer	United States	Henderson	Kentucky	42420	South	Furniture	Chairs	FUR-CH-10000454	600	730
2	3	2023-01-10	Second Class	Corporate	United States	Los Angeles	California	90036	West	Office Supplies	Labels	OFF-LA-10000240	10	10
3	4	2022-06-18	Standard Class	Consumer	United States	Fort Lauderdale	Florida	33311	South	Furniture	Tables	FUR-TA-10000577	780	960
4	5	2022-07-13	Standard Class	Consumer	United States	Fort Lauderdale	Florida	33311	South	Office Supplies	Storage	OFF-ST-10000760	20	20

In [4]:
```
df.columns
```

Out[4]:
```
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 [5]:
```
df['Ship Mode'].unique()
```

Out[5]:
```
array(['Second Class', 'Standard Class', nan, 'First Class', 'Same Day'],
      dtype=object)
```

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

Out[6]:
```
(9994, 16)
```

In [7]:
```
df.describe()
```

Out[7]:

	Order Id	Postal Code	cost price	List Price	Quantity	Discount Percent
count	9994.000000	9994.000000	9994.000000	9994.000000	9994.000000	9994.000000
mean	4997.500000	55190.379428	201.189714	229.756854	3.789574	3.484090
std	2885.163629	32063.693350	537.743203	623.245839	2.225110	1.114211
min	1.000000	1040.000000	0.000000	0.000000	1.000000	2.000000
25%	2499.250000	23223.000000	20.000000	20.000000	2.000000	2.000000
50%	4997.500000	56430.500000	50.000000	50.000000	3.000000	3.000000
75%	7495.750000	90008.000000	180.000000	210.000000	5.000000	4.000000
max	9994.000000	99301.000000	18110.000000	22640.000000	14.000000	5.000000

In [8]:
```
df.isnull().sum()
```

Out[8]:
```
Order Id           0
Order Date         0
Ship Mode          6
Segment            0
Country            0
City               0
State              0
Postal Code        0
Region             0
Category           0
Sub Category       0
Product Id         0
cost price         0
List Price         0
Quantity           0
Discount Percent   0
dtype: int64
```

In [9]:
```
df.head()
```

Out[9]:

	Order Id	Order Date	Ship Mode	Segment	Country	City	State	Postal Code	Region	Category	Sub Category	Product Id	cost price	List Price
0	1	2023-03-01	Second Class	Consumer	United States	Henderson	Kentucky	42420	South	Furniture	Bookcases	FUR-BO-10001798	240	260
1	2	2023-08-15	Second Class	Consumer	United States	Henderson	Kentucky	42420	South	Furniture	Chairs	FUR-CH-10000454	600	730
2	3	2023-01-10	Second Class	Corporate	United States	Los Angeles	California	90036	West	Office Supplies	Labels	OFF-LA-10000240	10	10
3	4	2022-06-18	Standard Class	Consumer	United States	Fort Lauderdale	Florida	33311	South	Furniture	Tables	FUR-TA-10000577	780	960
4	5	2022-07-13	Standard Class	Consumer	United States	Fort Lauderdale	Florida	33311	South	Office Supplies	Storage	OFF-ST-10000760	20	20

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

Out[10]:
```
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 [11]:
```
# changing datatype
df['Order Date'] = pd.to_datetime(df['Order Date'],format="%Y-%m-%d")
```

In [12]:
```
df.dtypes
```

Out[12]:
```
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
dtype: object
```

In [13]:
```
# Renaming columns using str function
df.columns=df.columns.str.lower()
df.columns=df.columns.str.replace(' ','_')
```

In [14]:
```
df.head()
```

Out[14]:

	order_id	order_date	ship_mode	segment	country	city	state	postal_code	region	category	sub_category	produ
0	1	2023-03-01	Second Class	Consumer	United States	Henderson	Kentucky	42420	South	Furniture	Bookcases	FUR 1000
1	2	2023-08-15	Second Class	Consumer	United States	Henderson	Kentucky	42420	South	Furniture	Chairs	FUR 1000
2	3	2023-01-10	Second Class	Corporate	United States	Los Angeles	California	90036	West	Office Supplies	Labels	OFI 1000
3	4	2022-06-18	Standard Class	Consumer	United States	Fort Lauderdale	Florida	33311	South	Furniture	Tables	FUR 1000
4	5	2022-07-13	Standard Class	Consumer	United States	Fort Lauderdale	Florida	33311	South	Office Supplies	Storage	OFI 1000

In [15]:
```
# To create a new column discount
df['discount']=df['list_price'] * (df['discount_percent'] / 100)
df['discount']
```

Out[15]:
```
0        5.2
1       21.9
2        0.5
3       19.2
4        1.0
        ...
9989     1.2
9990     3.6
9991     5.2
9992     0.9
9993     7.2
Name: discount, Length: 9994, dtype: float64
```

In [16]:
```
df.head()
```

Out[16]:

	order_id	order_date	ship_mode	segment	country	city	state	postal_code	region	category	sub_category	produ
0	1	2023-03-01	Second Class	Consumer	United States	Henderson	Kentucky	42420	South	Furniture	Bookcases	FUR 1000
1	2	2023-08-15	Second Class	Consumer	United States	Henderson	Kentucky	42420	South	Furniture	Chairs	FUR 1000
2	3	2023-01-10	Second Class	Corporate	United States	Los Angeles	California	90036	West	Office Supplies	Labels	OFI 1000
3	4	2022-06-18	Standard Class	Consumer	United States	Fort Lauderdale	Florida	33311	South	Furniture	Tables	FUR 1000
4	5	2022-07-13	Standard Class	Consumer	United States	Fort Lauderdale	Florida	33311	South	Office Supplies	Storage	OFI 1000

In [17]:
```
# To find selling price
df['selling_price']=df['list_price']-df['discount']
```

In [18]:
```
# To find the profit made
df['profit']=df['selling_price']-df['cost_price']
```

In [19]:
```
df.head()
```

Out[19]:

	order_id	order_date	ship_mode	segment	country	city	state	postal_code	region	category	sub_category	produ
0	1	2023-03-01	Second Class	Consumer	United States	Henderson	Kentucky	42420	South	Furniture	Bookcases	FUR 1000
1	2	2023-08-15	Second Class	Consumer	United States	Henderson	Kentucky	42420	South	Furniture	Chairs	FUR 1000
2	3	2023-01-10	Second Class	Corporate	United States	Los Angeles	California	90036	West	Office Supplies	Labels	OFI 1000
3	4	2022-06-18	Standard Class	Consumer	United States	Fort Lauderdale	Florida	33311	South	Furniture	Tables	FUR 1000
4	5	2022-07-13	Standard Class	Consumer	United States	Fort Lauderdale	Florida	33311	South	Office Supplies	Storage	OFI 1000

In [20]:
```
df.dtypes
```

Out[20]:
```
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
selling_price              float64
profit                     float64
dtype: object
```

In [21]:
```
df['ship_mode'].isnull().sum()
```

Out[21]:
```
np.int64(6)
```

In [22]:
```
df[df['ship_mode'].isnull()]
```

Out[22]:

	order_id	order_date	ship_mode	segment	country	city	state	postal_code	region	category	sub_category	pr
5	6	2022-03-13	NaN	Consumer	United States	Los Angeles	California	90032	West	Furniture	Furnishings	1
8	9	2023-03-23	NaN	Consumer	United States	Los Angeles	California	90032	West	Office Supplies	Binders	1
10	11	2023-03-31	NaN	Consumer	United States	Los Angeles	California	90032	West	Furniture	Tables	1
11	12	2023-12-25	NaN	Consumer	United States	Los Angeles	California	90032	West	Technology	Phones	1
14	15	2023-11-09	NaN	Home Office	United States	Fort Worth	Texas	76106	Central	Office Supplies	Appliances	1
118	119	2023-07-19	NaN	Corporate	United States	Bristol	Tennessee	37620	South	Office Supplies	Binders	1

In [23]:
```
# Dropping unwanted columns
df.drop(columns=['list_price','discount_percent'],inplace=True)
```

In [24]:
```
df.head()
```

Out[24]:

	order_id	order_date	ship_mode	segment	country	city	state	postal_code	region	category	sub_category	produ
0	1	2023-03-01	Second Class	Consumer	United States	Henderson	Kentucky	42420	South	Furniture	Bookcases	FUR 1000
1	2	2023-08-15	Second Class	Consumer	United States	Henderson	Kentucky	42420	South	Furniture	Chairs	FUR 1000
2	3	2023-01-10	Second Class	Corporate	United States	Los Angeles	California	90036	West	Office Supplies	Labels	OFI 1000
3	4	2022-06-18	Standard Class	Consumer	United States	Fort Lauderdale	Florida	33311	South	Furniture	Tables	FUR 1000
4	5	2022-07-13	Standard Class	Consumer	United States	Fort Lauderdale	Florida	33311	South	Office Supplies	Storage	OFI 1000

In [35]:
```
import sqlalchemy as sal
engine = sal.create_engine('mysql+mysqldb://root:Ryanantony%4007@localhost:3306/shop')
conn=engine.connect()
```

In [36]:
```
df.to_sql(
    name='sales',
    con=conn,
    index=False,
    if_exists='append'  # or 'append', 'fail'
)
```

Out[36]: 9994

In [ ]: