In [1]:
```
import pandas as pd
df= pd.read_csv('D:/data/customer_shopping_behavior.csv')
df.head()
```

Out[1]:

	Customer_ID	Age	Gender	Item_Purchased	Category	Purchase_Amount	
0	1	55	Male	Blouse	Clothing	53	
1	2	19	Male	Sweater	Clothing	64	
2	3	50	Male	Jeans	Clothing	73	Ma
3	4	21	Male	Sandals	Footwear	90	R
4	5	45	Male	Blouse	Clothing	49	

In [2]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3900 entries, 0 to 3899
Data columns (total 17 columns):
 #   Column                  Non-Null Count  Dtype  
---  ------                  --------------  -----  
 0   Customer_ID             3900 non-null   int64  
 1   Age                     3900 non-null   int64  
 2   Gender                  3900 non-null   object 
 3   Item_Purchased          3900 non-null   object 
 4   Category                3900 non-null   object 
 5   Purchase_Amount         3900 non-null   int64  
 6   Location                3900 non-null   object 
 7   Size                    3900 non-null   object 
 8   Color                   3900 non-null   object 
 9   Season                  3900 non-null   object 
 10  Review_Rating           3863 non-null   float64
 11  Subscription_Status     3900 non-null   object 
 12  Shipping_Type           3900 non-null   object 
 13  Discount_Applied        3900 non-null   object 
 14  Previous_Purchases      3900 non-null   int64  
 15  Payment_Method          3900 non-null   object 
 16  Frequency_of_Purchases  3900 non-null   object 
dtypes: float64(1), int64(4), object(12)
memory usage: 518.1+ KB
```

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

Out[3]:

	Customer_ID	Age	Purchase_Amount	Review_Rating	Previous_Pu
count	3900.000000	3900.000000	3900.000000	3863.000000	390
mean	1950.500000	44.068462	59.764359	3.750065	2
std	1125.977353	15.207589	23.685392	0.716983	1
min	1.000000	18.000000	20.000000	2.500000	
25%	975.750000	31.000000	39.000000	3.100000	1
50%	1950.500000	44.000000	60.000000	3.800000	2
75%	2925.250000	57.000000	81.000000	4.400000	3
max	3900.000000	70.000000	100.000000	5.000000	5

In [4]:
```
df.isnull().sum()
```

Out[4]:
```
Customer_ID                0
Age                        0
Gender                     0
Item_Purchased             0
Category                   0
Purchase_Amount            0
Location                   0
Size                       0
Color                      0
Season                     0
Review_Rating             37
Subscription_Status        0
Shipping_Type              0
Discount_Applied           0
Previous_Purchases         0
Payment_Method             0
Frequency_of_Purchases     0
dtype: int64
```

In [5]:
```
df['Review_Rating'] = df.groupby('Category')['Review_Rating'].transform(lambda
df.head()
```

Out[5]:

	Customer_ID	Age	Gender	Item_Purchased	Category	Purchase_Amount	
0	1	55	Male	Blouse	Clothing	53	
1	2	19	Male	Sweater	Clothing	64	
2	3	50	Male	Jeans	Clothing	73	Ma
3	4	21	Male	Sandals	Footwear	90	R
4	5	45	Male	Blouse	Clothing	49	

In [6]:
```
df.isnull().sum()
```

Out[6]:
```
Customer_ID               0
Age                       0
Gender                    0
Item_Purchased            0
Category                  0
Purchase_Amount           0
Location                  0
Size                      0
Color                     0
Season                    0
Review_Rating             0
Subscription_Status       0
Shipping_Type             0
Discount_Applied          0
Previous_Purchases        0
Payment_Method            0
Frequency_of_Purchases    0
dtype: int64
```

In [7]:
```
df.columns = df.columns.str.lower()
df.columns = df.columns.str.replace(' ','_')
```

In [ ]:
```
```

In [8]:
```
df = df.rename(columns = {'purchase_amount_(usd)':'purchase_amount'})
```

In [9]:
```
df.columns
```

Out[9]:
```
Index(['customer_id', 'age', 'gender', 'item_purchased', 'category',
       'purchase_amount', 'location', 'size', 'color', 'season',
       'review_rating', 'subscription_status', 'shipping_type',
       'discount_applied', 'previous_purchases', 'payment_method',
       'frequency_of_purchases'],
      dtype='object')
```

In [10]:
```
labels = ['young adult','adult','middle-aged','senior']
df['age_group'] = pd.qcut(df['age'],q=4,labels = labels)
```

In [11]:
```
df[['age','age_group']].head()
```

Out[11]:

	age	age_group
0	55	middle-aged
1	19	young adult
2	50	middle-aged
3	21	young adult
4	45	middle-aged

In [12]:
```
df.head(1)
```

Out[12]:

	customer_id	age	gender	item_purchased	category	purchase_amount	loca
0	1	55	Male	Blouse	Clothing	53	Kent

In [13]:
```
df['frequency_of_purchases'].unique()
```

Out[13]:
```
array(['Fortnightly', 'Weekly', 'Annually', 'Quarterly', 'Bi-Weekly',
       'Monthly', 'Every 3 Months'], dtype=object)
```

In [14]:
```
# create frequency_of_purchases_day
frequency_mapping = {
    'Fortnightly':14,
    'Weekly':7,
    'Annually':365,
    'Quarterly':90,
    'Bi-Weekly':14,
    'Monthly':30,
    'Every 3 Months':90
}

df['frequency_of_purchases_days'] = df['frequency_of_purchases'].map(frequency
```

In [15]:
```
df[['frequency_of_purchases_days','frequency_of_purchases']].head()
```

Out[15]:

	frequency_of_purchases_days	frequency_of_purchases
0	14	Fortnightly
1	14	Fortnightly
2	7	Weekly
3	7	Weekly
4	365	Annually

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

Out[16]:

	customer_id	age	gender	item_purchased	category	purchase_amount	
0	1	55	Male	Blouse	Clothing	53	
1	2	19	Male	Sweater	Clothing	64	
2	3	50	Male	Jeans	Clothing	73	Mass
3	4	21	Male	Sandals	Footwear	90	Rh
4	5	45	Male	Blouse	Clothing	49	

In [17]:
```
pip install pymysql sqlalchemy
```

```
Requirement already satisfied: pymysql in c:\users\hp\anaconda3\lib\site-packag
es (1.1.2)
Requirement already satisfied: sqlalchemy in c:\users\hp\anaconda3\lib\site-pac
kages (2.0.30)
Requirement already satisfied: typing-extensions>=4.6.0 in c:\users\hp\anaconda
3\lib\site-packages (from sqlalchemy) (4.11.0)
Requirement already satisfied: greenlet!=0.4.17 in c:\users\hp\anaconda3\lib\si
te-packages (from sqlalchemy) (3.0.1)
Note: you may need to restart the kernel to use updated packages.
```

In [19]:
```
from sqlalchemy import create_engine
username = 'root'
password ='Sunil1234'
host = 'localhost'
port = '3306'
database = 'customer_shopping'


# Example: username=root, password=Sunil123, database=customer_shopping
engine = create_engine("mysql+pymysql://root:Sunil1234@localhost/customer_shop


table_name ='mytable'

df.to_sql(
    name='customers',   # MySQL table name
    con=engine,
    if_exists='append', # 'append' adds data, 'replace' overwrites table
    index=False         # Do not write DataFrame index as a column
)

pd.read_sql("select * from mytable limit 5;",engine)
```

Out[19]:

	customer_id	age	gender	item_purchased	category	purchase_amount	
0	1	55	Male	Blouse	Clothing	53	
1	2	19	Male	Sweater	Clothing	64	
2	3	50	Male	Jeans	Clothing	73	Mass
3	4	21	Male	Sandals	Footwear	90	Rh
4	5	45	Male	Blouse	Clothing	49	

In [20]:
```
df.to_sql(
    name='mytable',
    con=engine,
```

```
    if_exists='replace',   # creates table automatically
    index=False
)
```

Out[20]: 3900

In [ ]: