In [24]:
```
import pandas as pd
from sqlalchemy import create_engine
```

In [25]:
```
df = pd.read_csv("consumer_data.csv")
print(df.shape)
df.head()
```

```
(3900, 18)
```

Out[25]:

	Customer ID	Age	Gender	Item Purchased	Category	Purchase Amount (USD)	Location	Size
0	1	55	Male	Blouse	Clothing	53	Kentucky	L
1	2	19	Male	Sweater	Clothing	64	Maine	L
2	3	50	Male	Jeans	Clothing	73	Massachusetts	S
3	4	21	Male	Sandals	Footwear	90	Rhode Island	M
4	5	45	Male	Blouse	Clothing	49	Oregon	M

In [26]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3900 entries, 0 to 3899
Data columns (total 18 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 (USD)   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  Promo Code Used         3900 non-null   object 
 15  Previous Purchases      3900 non-null   int64  
 16  Payment Method          3900 non-null   object 
 17  Frequency of Purchases  3900 non-null   object 
dtypes: float64(1), int64(4), object(13)
memory usage: 548.6+ KB
```

In [27]:
```
df.isnull().sum()
```

Out[27]:
```
Customer ID                0
Age                        0
Gender                     0
Item Purchased             0
Category                   0
Purchase Amount (USD)      0
Location                   0
Size                       0
Color                      0
Season                     0
Review Rating             37
Subscription Status        0
Shipping Type              0
Discount Applied           0
Promo Code Used            0
Previous Purchases         0
Payment Method             0
Frequency of Purchases     0
dtype: int64
```

In [28]:
```
df['Review Rating'] = df['Review Rating'].fillna(df['Review Rating'].mode()[0]
```

In [29]:
```
df.columns = df.columns.str.lower()
df.columns = df.columns.str.replace(' ', '_')
df = df.rename(columns={'purchase_amount_(usd)': 'purchase_amount'})
```

In [30]:
```
df.columns
```

Out[30]:
```
Index(['customer_id', 'age', 'gender', 'item_purchased', 'category',
       'purchase_amount', 'location', 'size', 'color', 'season',
       'review_rating', 'subscription_status', 'shipping_type',
       'discount_applied', 'promo_code_used', 'previous_purchases',
       'payment_method', 'frequency_of_purchases'],
      dtype='object')
```

In [31]:
```
frequency_mapping = {
    'Fortnightly': 14,
    'Weekly': 7,
    'Monthly': 30,
    'Quarterly': 90,
    'Bi-Weekly': 14,
    'Annually': 365,
    'Every 3 Months': 90
}

df['purchase_frequency_days'] = df['frequency_of_purchases'].map(frequency_map
```

In [32]:
```
df[['purchase_frequency_days','frequency_of_purchases']].head(10)
```

Out[32]:

	purchase_frequency_days	frequency_of_purchases
0	14	Fortnightly
1	14	Fortnightly
2	7	Weekly
3	7	Weekly
4	365	Annually
5	7	Weekly
6	90	Quarterly
7	7	Weekly
8	365	Annually
9	90	Quarterly

In [33]:

```
from sqlalchemy.engine import URL
from sqlalchemy import create_engine, exc

# MySQL connection
username = "root"
password = "P@55w0rd"
host = "127.0.0.1"
port = 3306
# target database
database = "customer_behavior"

# Build URL safely using SQLAlchemy
url = URL.create(
    "mysql+pymysql",
    username=username,
    password=password,
    host=host,
    port=port,
    database=database,
)
engine = create_engine(url)

# Write DataFrame to MySQL (handle missing DB gracefully)
try:
    table_name = "customer"   # choose any table name
    df.to_sql(table_name, engine, if_exists="replace", index=False)
    print(pd.read_sql("SELECT * FROM customer LIMIT 5;", engine))
except exc.OperationalError as e:
    print("OperationalError while connecting/writing to the DB:", e)
    print("Hint: ensure the database exists and the server is running, or crea
```

	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
3	4	21	Male	Sandals	Footwear	90
4	5	45	Male	Blouse	Clothing	49

	location	size	color	season	review_rating	subscription_status
0	Kentucky	L	Gray	Winter	3.1	Yes
1	Maine	L	Maroon	Winter	3.1	Yes
2	Massachusetts	S	Maroon	Spring	3.1	Yes
3	Rhode Island	M	Maroon	Spring	3.5	Yes
4	Oregon	M	Turquoise	Spring	2.7	Yes

	shipping_type	discount_applied	promo_code_used	previous_purchases
0	Express	Yes	Yes	14
1	Express	Yes	Yes	2
2	Free Shipping	Yes	Yes	23
3	Next Day Air	Yes	Yes	49
4	Free Shipping	Yes	Yes	31

	payment_method	frequency_of_purchases	purchase_frequency_days
0	Venmo	Fortnightly	14
1	Cash	Fortnightly	14
2	Credit Card	Weekly	7
3	PayPal	Weekly	7
4	PayPal	Annually	365