

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

```
In [2]: df = pd.read_csv('customer_shopping_behavior.csv')
```

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

```
Out[3]:
```

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



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

```
Out[4]: 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 [5]: df.describe()
```

Out[5]:

	Customer ID	Age	Purchase Amount (USD)	Review Rating	Previous Purchases
<b>count</b>	3900.000000	3900.000000	3900.000000	3863.000000	3900.000000
<b>mean</b>	1950.500000	44.068462	59.764359	3.750065	25.351538
<b>std</b>	1125.977353	15.207589	23.685392	0.716983	14.447125
<b>min</b>	1.000000	18.000000	20.000000	2.500000	1.000000
<b>25%</b>	975.750000	31.000000	39.000000	3.100000	13.000000
<b>50%</b>	1950.500000	44.000000	60.000000	3.800000	25.000000
<b>75%</b>	2925.250000	57.000000	81.000000	4.400000	38.000000
<b>max</b>	3900.000000	70.000000	100.000000	5.000000	50.000000

In [6]: *# Imputing missing values in Review Rating column with the median rating of the pro*

```
df['Review Rating'] = df.groupby('Category')['Review Rating'].transform(lambda x: x
```

In [7]: `df.isnull().sum()`

```
Out[7]: 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  0
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 [8]: *# Renaming columns according to snake casing for better readability and documentati*

```
df.columns = df.columns.str.lower()
df.columns = df.columns.str.replace(' ', '_')
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', 'promo_code_used', 'previous_purchases',
              'payment_method', 'frequency_of_purchases'],
              dtype='object')
```

```
In [10]: # create a new column age_group
labels = ['Young Adult', 'Adult', 'Middle-aged', 'Senior']
df['age_group'] = pd.qcut(df['age'], q=4, labels = labels)
```

```
In [ ]:
```

```
In [11]: # create new column purchase_frequency_days

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_mapping)
```

```
In [12]: df[['purchase_frequency_days', 'frequency_of_purchases']].head(10)
```

```
Out[12]:
```

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

```
In [13]: df[['discount_applied', 'promo_code_used']].head(10)
```

Out[13]:

	discount_applied	promo_code_used
0	Yes	Yes
1	Yes	Yes
2	Yes	Yes
3	Yes	Yes
4	Yes	Yes
5	Yes	Yes
6	Yes	Yes
7	Yes	Yes
8	Yes	Yes
9	Yes	Yes

In [14]: `(df['discount_applied'] == df['promo_code_used']).all()`

Out[14]: True

In [15]: `# Dropping promo code used column`  
`df = df.drop('promo_code_used', axis=1)`

In [16]: `df.columns`

Out[16]: 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', 'age\_group', 'purchase\_frequency\_days'],  
dtype='object')

In [26]: `# !pip install pymysql sqlalchemy`

In [36]: `from sqlalchemy import create_engine  
from urllib.parse import quote_plus  
import pymysql  
import pandas as pd`  
`username = "root"  
password = quote_plus("DhoniCSK@858585") # encodes special characters  
host = "localhost"  
port = "3306"  
database = "customer_behavior"`  
`engine = create_engine(f"mysql+pymysql://{username}:{password}@{host}:{port}/{database}")  
# Write DataFrame to MySQL  
table_name = "customer" # choose any table name  
df.to_sql(table_name, engine, if_exists="replace", index=False)`

```
# Test connection
try:
    with engine.connect() as connection:
        print("✅ Successfully connected to MySQL!")
except Exception as e:
    print("❌ Connection failed:", e)
```

✅ Successfully connected to MySQL!

```
In [37]: # Read back sample
pd.read_sql("SELECT * FROM customer LIMIT 5;", engine)
```

Out[37]:

	customer_id	age	gender	item_purchased	category	purchase_amount	location
0	1	55	Male	Blouse	Clothing	53	Kentucky
1	2	19	Male	Sweater	Clothing	64	Maine
2	3	50	Male	Jeans	Clothing	73	Massachusetts
3	4	21	Male	Sandals	Footwear	90	Rhode Island
4	5	45	Male	Blouse	Clothing	49	Oregon



In [ ]:

In [ ]: