

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

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

Out[2]:

	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 [3]: df.shape
```

Out[3]: (3900, 18)

```
In [4]: 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 [5]: df.describe(include='all')

Out[5]:

	Customer ID	Age	Gender	Item Purchased	Category	Purchase Amount (USD)	Location	S
count	3900.000000	3900.000000	3900	3900	3900	3900.000000	3900	3900
unique	Nan	Nan	2	25	4	Nan	50	17
top	Nan	Nan	Male	Blouse	Clothing	Nan	Montana	Nan
freq	Nan	Nan	2652	171	1737	Nan	96	17
mean	1950.500000	44.068462	Nan	Nan	Nan	59.764359	Nan	Nan
std	1125.977353	15.207589	Nan	Nan	Nan	23.685392	Nan	Nan
min	1.000000	18.000000	Nan	Nan	Nan	20.000000	Nan	Nan
25%	975.750000	31.000000	Nan	Nan	Nan	39.000000	Nan	Nan
50%	1950.500000	44.000000	Nan	Nan	Nan	60.000000	Nan	Nan
75%	2925.250000	57.000000	Nan	Nan	Nan	81.000000	Nan	Nan
max	3900.000000	70.000000	Nan	Nan	Nan	100.000000	Nan	Nan

In []:

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

```
Out[6]: 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 [8]: df['Review Rating']=df.groupby('Category')['Review Rating'].transform(lambda x: x.fillna(x.mean()))
```

```
In [43]: df.isnull().sum()
```

```
Out[43]: 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 [33]: df.columns = [ ' '.join(col.split('_')) for col in df.columns]
```

```
In [34]: df.columns
```

```
Out[34]: Index(['customerid', 'age', 'gender', 'itempurchased', 'category',
       'purchaseamountusd', 'location', 'size', 'color', 'season',
       'reviewrating', 'subscriptionstatus', 'shippingtype', 'discountapplied',
       'promocodeused', 'previouspurchases', 'paymentmethod',
       'frequencyofpurchases'],
      dtype='object')
```

```
In [41]: df = df.rename(columns={
    'customerid': 'customer_id',
    'itempurchased': 'item_purchased',
    'purchaseamountusd': 'purchase_amount_usd',
    'reviewrating': 'review_rating',
    'subscriptionstatus': 'subscription_status',
    'shippingtype': 'shipping_type',
    'discountapplied': 'discount_applied',
    'promocodeused': 'promo_code_used',
    'previouspurchases': 'previous_purchases',
    'paymentmethod': 'payment_method',
    'frequencyofpurchases': 'frequency_of_purchases'
})
```

```
In [42]: df.columns
```

```
Out[42]: Index(['customer_id', 'age', 'gender', 'item_purchased', 'category',
       'purchase_amount_usd', '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 [44]: labels=['young_Adult','Adult','Middle_aged','Senior']
df['age_group']=pd.qcut(df['age'],q=4,labels=labels)
```

```
In [45]: df[['age', 'age_group']].head(10)
```

```
Out[45]:   age  age_group
```

0	55	Middle_aged
1	19	young_Adult
2	50	Middle_aged
3	21	young_Adult
4	45	Middle_aged
5	46	Middle_aged
6	63	Senior
7	27	young_Adult
8	26	young_Adult
9	57	Middle_aged

```
In [46]: 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 [47]: df[['purchase_frequency_days', 'frequency_of_purchases']].head(10)
```

```
Out[47]: 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 [48]: df[['discount_applied', 'promo_code_used']].head(10)
```

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

```
Out[49]: True
```

```
In [50]: df = df.drop('promo_code_used', axis=1)
```

```
In [51]: df.columns
```

```
Out[51]: Index(['customer_id', 'age', 'gender', 'item_purchased', 'category',
       'purchase_amount_usd', '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 [52]: !pip install pymysql sqlalchemy
```

```
Collecting pymysql
  Downloading pymysql-1.1.2-py3-none-any.whl.metadata (4.3 kB)
Requirement already satisfied: sqlalchemy in c:\users\hp\anaconda3\lib\site-packages (2.0.34)
Requirement already satisfied: typing-extensions>=4.6.0 in c:\users\hp\anaconda3\lib\site-packages (from sqlalchemy) (4.11.0)
Requirement already satisfied: greenlet!=0.4.17 in c:\users\hp\anaconda3\lib\site-packages (from sqlalchemy) (3.0.1)
Downloading pymysql-1.1.2-py3-none-any.whl (45 kB)
Installing collected packages: pymysql
Successfully installed pymysql-1.1.2
```

```
In [55]: from sqlalchemy import create_engine
import pandas as pd
```

```
# MySQL connection details
username = "root"
password = "Sanjiv%402123" # @ replaced with %40
```

```
host = "localhost"
port = "3306"
database = "customer_behavior"

engine = create_engine(
    f"mysql+pymysql://{{username}}:{{password}}@{{host}}:{{port}}/{{database}}"
)

# Write DataFrame to MySQL
df.to_sql("customer", engine, if_exists="replace", index=False)

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

Out[55]:

	customer_id	age	gender	item_purchased	category	purchase_amount_usd	location
0	1	55	Male	Blouse	Clothing	53	Kentuc
1	2	19	Male	Sweater	Clothing	64	Mai
2	3	50	Male	Jeans	Clothing	73	Massachuse
3	4	21	Male	Sandals	Footwear	90	Rhode Islan
4	5	45	Male	Blouse	Clothing	49	Oreg

In []: