

```
In [7]: import pandas as pd
df=pd.read_csv('customer_shopping_behavior.csv')
```

```
In [8]: df.head()
```

Out[8]:

	Customer ID	Age	Gender	Item Purchased	Category	Purchase Amount (USD)	Location	Siz
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 [11]: 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 [12]: df.describe()
```

Out[12]:

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

In [14]: `df.describe(include='all')`

Out[14]:

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

In [16]: `df.isnull().sum()`

```
Out[16]: 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 [20]: df['Review Rating']=df.groupby('Category')['Review Rating'].transform(Lambda
```

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

```
Out[21]: 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 [26]: df.columns=df.columns.str.lower()
         df.columns=df.columns.str.replace(' ','_')
         df=df.rename(columns={'purchase_amount_(usd)': 'purchase_amount'})
```

```
In [27]: df.columns
```

```
Out[27]: 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 [30]: #create a column age_group
labels=['young adult','adult','middle-aged','senior']
df['age_group']=pd.qcut(df['age'],q=4,labels=labels)
```

```
In [31]: df[['age','age_group']].head(10)
```

```
Out[31]:
```

	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 [32]: #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_map
```

```
In [33]: df[['purchase_frequency_days','frequency_of_purchases']].head(10)
```

Out[33]:

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

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

In [34]: `df[['discount_applied','promo_code_used']].head(10)`

Out[34]:

	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

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

Out[36]: `np.True_`

In [37]: `df=df.drop('promo_code_used',axis=1)`

In [38]: `df.columns`

Out[38]: `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 [46]: pip install psycopg2-binary sqlalchemy
```

```
Requirement already satisfied: psycopg2-binary in c:\users\welcome\anaconda3\lib\site-packages (2.9.11)
Requirement already satisfied: sqlalchemy in c:\users\welcome\anaconda3\lib\site-packages (2.0.43)
Requirement already satisfied: greenlet>=1 in c:\users\welcome\anaconda3\lib\site-packages (from sqlalchemy) (3.2.4)
Requirement already satisfied: typing-extensions>=4.6.0 in c:\users\welcome\anaconda3\lib\site-packages (from sqlalchemy) (4.15.0)
Note: you may need to restart the kernel to use updated packages.
```

```
In [13]: from sqlalchemy import create_engine
```

```
username='postgres'
password='123456789'
host='localhost'
port='5432'
database='customer_behavior'

engine=create_engine(f"postgresql+psycopg2://{username}:{password}@{host}:{port}/{database}")
table_name="customer"
df.to_sql(table_name,engine,if_exists="replace",index=False)

print(f"data successfully loaded into table '{table_name}' in database '{database}'")
```

```
data successfully loaded into table 'customer' in database 'customer_behavior'.
```

```
In [1]: print("downloading this file")
```

```
downloading this file
```

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```