In [2]:
```
import pandas as pd
df= pd.read_csv("customer_shopping_behavior.csv")
```

df.head()

In [4]:
```
df.tail()
```

Out[4]:

	Customer ID	Age	Gender	Item Purchased	Category	Purchase Amount (USD)	Location	Size	Colc
3895	3896	40	Female	Hoodie	Clothing	28	Virginia	L	Turquois
3896	3897	52	Female	Backpack	Accessories	49	Iowa	L	Whit
3897	3898	46	Female	Belt	Accessories	33	New Jersey	L	Gree
3898	3899	44	Female	Shoes	Footwear	77	Minnesota	S	Brow
3899	3900	52	Female	Handbag	Accessories	81	California	M	Beig

In [5]:
```
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 [6]:
```
df.describe()
```

Out[6]:

	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 [9]:
```
df.describe(include="all")
```

Out[9]:

	Customer ID	Age	Gender	Item Purchased	Category	Purchase Amount (USD)	Location	S
count	3900.000000	3900.000000	3900	3900	3900	3900.000000	3900	3
unique	NaN	NaN	2	25	4	NaN	50	
top	NaN	NaN	Male	Blouse	Clothing	NaN	Montana	
freq	NaN	NaN	2652	171	1737	NaN	96	1
mean	1950.500000	44.068462	NaN	NaN	NaN	59.764359	NaN	N
std	1125.977353	15.207589	NaN	NaN	NaN	23.685392	NaN	N
min	1.000000	18.000000	NaN	NaN	NaN	20.000000	NaN	N
25%	975.750000	31.000000	NaN	NaN	NaN	39.000000	NaN	N
50%	1950.500000	44.000000	NaN	NaN	NaN	60.000000	NaN	N
75%	2925.250000	57.000000	NaN	NaN	NaN	81.000000	NaN	N
max	3900.000000	70.000000	NaN	NaN	NaN	100.000000	NaN	N

In [10]:
```
df.isnull().sum()
```

Out[10]:
```
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 [15]:
```
df["Review Rating"]= df.groupby("Category")["Review Rating"].transform(lambda x:x.f
```

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             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 [17]:
```
df.columns= df.columns.str.lower()
```

In [18]:
```
df.columns
```

Out[18]:
```
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 [20]:
```
df.columns= df.columns.str.replace(" ","_")
```

In [21]:
```
df.columns
```

Out[21]:
```
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 [22]:
```
df= df.rename(columns={"purchase_amount_(usd)":"purchase_amount"})
```

In [23]:
```
df.columns
```

Out[23]:
```
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 [24]:
```
df.head()
```

Out[24]:

	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 [27]:
```
#new column age group is created
labels=["Young","Adult","Middle-aged","old"]
df["age_group"]=pd.qcut(df["age"] ,q=4 , labels=labels)
```

In [28]:
```
df.columns
```

Out[28]:
```
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', 'age_group'],
      dtype='object')
```

In [32]:
```
df[["age","age_group"]].head()
```

Out[32]:

	age	age_group
0	55	Middle-aged
1	19	Young
2	50	Middle-aged
3	21	Young
4	45	Middle-aged

In [34]:
```
df.head(10)
```

Out[34]:

	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
5	6	46	Male	Sneakers	Footwear	20	Wyoming
6	7	63	Male	Shirt	Clothing	85	Montana
7	8	27	Male	Shorts	Clothing	34	Louisiana
8	9	26	Male	Coat	Outerwear	97	West Virginia
9	10	57	Male	Handbag	Accessories	31	Missouri

In [38]:
```
#purchase frequency days
frequency_mapping ={
    "Fortnightly" : 14,
    "Weekly" : 7,
    "Monthly": 30,
    "Quaterly": 90,
    "Bi-Weekly": 14,
    "Annually": 360,
    "Every 3 Months" : 90
}
df["purchase_frequency_days"]= df["frequency_of_purchases"].map(frequency_mapping)
```

In [39]:
```
df[["frequency_of_purchases","purchase_frequency_days"]].head()
```

Out[39]:

	frequency_of_purchases	purchase_frequency_days
0	Fortnightly	14.0
1	Fortnightly	14.0
2	Weekly	7.0
3	Weekly	7.0
4	Annually	360.0

In [40]:
```
df.columns
```

Out[40]:
```
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', 'age_group',
       'purchase_frequency_days'],
      dtype='object')
```

In [43]:
```
df[['discount_applied', 'promo_code_used']].head()
```

Out[43]:

	discount_applied	promo_code_used
0	Yes	Yes
1	Yes	Yes
2	Yes	Yes
3	Yes	Yes
4	Yes	Yes

In [46]:
```
(df['discount_applied']==df[ 'promo_code_used']).head(10)
```

Out[46]:
```
0    True
1    True
2    True
3    True
4    True
5    True
6    True
7    True
8    True
9    True
dtype: bool
```

In [48]:
```
(df['discount_applied']==df[ 'promo_code_used']).all()
```

Out[48]:
```
np.True_
```

In [51]:
```
df=df.drop('promo_code_used', axis=1)
```

In [52]:
```
df.columns
```

Out[52]:
```
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 [53]:
```
!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:\anaconda3\lib\site-packages (2.0.43)
Requirement already satisfied: greenlet>=1 in c:\anaconda3\lib\site-packages (from s
qlalchemy) (3.2.4)
Requirement already satisfied: typing-extensions>=4.6.0 in c:\anaconda3\lib\site-pac
kages (from sqlalchemy) (4.15.0)
Downloading pymysql-1.1.2-py3-none-any.whl (45 kB)
Installing collected packages: pymysql
Successfully installed pymysql-1.1.2
```

In [54]:
```
from sqlalchemy import create_engine

# MySQL connection
username = "root"
password = "playwithdata"
host = "localhost"
port = "3306"
database = "customer_behavior"

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

# Write DataFrame to MySQL
table_name = "customer"   # choose any table name
df.to_sql(table_name, engine, if_exists="replace", index=False)

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

Out[54]:

	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