```
In [1]:  import pandas as pd
         from sqlalchemy import create_engine
         df=pd.read_csv('customer.csv')
         df.head()
```

Out[1]:

	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 [2]:  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 [3]:  df.describe()
```

	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

describe gives the summary statistics of numerical columns only

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

Out[4]:

	Customer ID	Age	Gender	Item Purchased	Category	Purchase Amount (USD)	L
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	

include='all' gives summary statistics of non integer values also.

In [5]: `df.isnull()`

Out[5]:

	Customer ID	Age	Gender	Item Purchased	Category	Purchase Amount (USD)	Location	Size
0	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False
...	...	...	...	...	...	...	...	...
3895	False	False	False	False	False	False	False	False
3896	False	False	False	False	False	False	False	False
3897	False	False	False	False	False	False	False	False
3898	False	False	False	False	False	False	False	False
3899	False	False	False	False	False	False	False	False

3900 rows × 18 columns

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

```
In [8]: df.isnull().sum()
```

```
Out[8]: 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 [10]: df.columns=df.columns.str.lower()
         df.columns=df.columns.str.replace(' ','_')
         df=df.rename(columns={'purchase_amount_(usd)':'purchase_amount'})
```

```
In [11]: df.columns
```

```
Out[11]: 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 [12]: #create a column age_group
         labels=['Young adult','Adult','Middle-aged','Senior']
         df['age_group']=pd.qcut(df['age'],q=4,labels=labels)
```

```
In [13]: df[['age','age_group']].head(5)
```

Out[13]:

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

```
In [14]:  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_mappi
```

create column purchase_frequency_days.

first we created a dictionary called 'frequency_mapping' that maps each textual frequency to the equivalent number of days.

```
In [15]:  df[['purchase_frequency_days','frequency_of_purchases']].head(10)
```

Out[15]:

	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 [16]:  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 [17]:
```python
(df['discount_applied']==df['promo_code_used']).all()
```

Out[17]: np.True_

In [18]:
```python
df=df.drop('promo_code_used',axis=1)
```

In [19]:
```python
df.columns
```

Out[19]: 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 [21]:
```python
pip install pymysql sqlalchemy
```

Requirement already satisfied: pymysql in c:\users\manvika\appdata\local\progra
ms\python\python313\lib\site-packages (1.1.2)
Requirement already satisfied: sqlalchemy in c:\users\manvika\appdata\local\pro
grams\python\python313\lib\site-packages (2.0.46)
Requirement already satisfied: greenlet>=1 in c:\users\manvika\appdata\local\pr
ograms\python\python313\lib\site-packages (from sqlalchemy) (3.3.1)
Requirement already satisfied: typing-extensions>=4.6.0 in c:\users\manvika\app
data\local\programs\python\python313\lib\site-packages (from sqlalchemy) (4.1
4.1)
Note: you may need to restart the kernel to use updated packages.
[notice] A new release of pip is available: 25.2 -> 26.0.1
[notice] To update, run: python.exe -m pip install --upgrade pip

In [22]:
```python
pip install psycopg2-binary sqlalchemy
```

```
Requirement already satisfied: psycopg2-binary in c:\users\manvika\appdata\loca
l\programs\python\python313\lib\site-packages (2.9.11)
Requirement already satisfied: sqlalchemy in c:\users\manvika\appdata\local\pro
grams\python\python313\lib\site-packages (2.0.46)
Requirement already satisfied: greenlet>=1 in c:\users\manvika\appdata\local\pr
ograms\python\python313\lib\site-packages (from sqlalchemy) (3.3.1)
Requirement already satisfied: typing-extensions>=4.6.0 in c:\users\manvika\app
data\local\programs\python\python313\lib\site-packages (from sqlalchemy) (4.1
4.1)
Note: you may need to restart the kernel to use updated packages.
[notice] A new release of pip is available: 25.2 -> 26.0.1
[notice] To update, run: python.exe -m pip install --upgrade pip
```

In [23]: `pip install mysql-connector-python sqlalchemy pymysql pandas`

```
Requirement already satisfied: mysql-connector-python in c:\users\manvika\appda
ta\local\programs\python\python313\lib\site-packages (9.5.0)
Requirement already satisfied: sqlalchemy in c:\users\manvika\appdata\local\pro
grams\python\python313\lib\site-packages (2.0.46)
Requirement already satisfied: pymysql in c:\users\manvika\appdata\local\progra
ms\python\python313\lib\site-packages (1.1.2)
Requirement already satisfied: pandas in c:\users\manvika\appdata\local\program
s\python\python313\lib\site-packages (2.3.1)
Requirement already satisfied: greenlet>=1 in c:\users\manvika\appdata\local\pr
ograms\python\python313\lib\site-packages (from sqlalchemy) (3.3.1)
Requirement already satisfied: typing-extensions>=4.6.0 in c:\users\manvika\app
data\local\programs\python\python313\lib\site-packages (from sqlalchemy) (4.1
4.1)
Requirement already satisfied: numpy>=1.26.0 in c:\users\manvika\appdata\local\
programs\python\python313\lib\site-packages (from pandas) (2.3.2)
Requirement already satisfied: python-dateutil>=2.8.2 in c:\users\manvika\appda
ta\local\programs\python\python313\lib\site-packages (from pandas) (2.9.0.post
0)
Requirement already satisfied: pytz>=2020.1 in c:\users\manvika\appdata\local\p
rograms\python\python313\lib\site-packages (from pandas) (2025.2)
Requirement already satisfied: tzdata>=2022.7 in c:\users\manvika\appdata\loca
l\programs\python\python313\lib\site-packages (from pandas) (2025.2)
Requirement already satisfied: six>=1.5 in c:\users\manvika\appdata\local\progr
ams\python\python313\lib\site-packages (from python-dateutil>=2.8.2->pandas)
(1.17.0)
Note: you may need to restart the kernel to use updated packages.
[notice] A new release of pip is available: 25.2 -> 26.0.1
[notice] To update, run: python.exe -m pip install --upgrade pip
```

In [24]: `!pip install jupytext`

```
Requirement already satisfied: jupytext in c:\users\manvika\appdata\local\progr
ams\python\python313\lib\site-packages (1.19.1)
Requirement already satisfied: markdown-it-py>=1.0 in c:\users\manvika\appdata\
local\programs\python\python313\lib\site-packages (from jupytext) (4.0.0)
Requirement already satisfied: mdit-py-plugins in c:\users\manvika\appdata\loca
l\programs\python\python313\lib\site-packages (from jupytext) (0.5.0)
Requirement already satisfied: nbformat in c:\users\manvika\appdata\local\progr
ams\python\python313\lib\site-packages (from jupytext) (5.10.4)
Requirement already satisfied: packaging in c:\users\manvika\appdata\local\prog
rams\python\python313\lib\site-packages (from jupytext) (25.0)
Requirement already satisfied: pyyaml in c:\users\manvika\appdata\local\program
s\python\python313\lib\site-packages (from jupytext) (6.0.2)
Requirement already satisfied: mdurl~=0.1 in c:\users\manvika\appdata\local\pro
grams\python\python313\lib\site-packages (from markdown-it-py>=1.0->jupytext)
(0.1.2)
Requirement already satisfied: fastjsonschema>=2.15 in c:\users\manvika\appdat
a\local\programs\python\python313\lib\site-packages (from nbformat->jupytext)
(2.21.1)
Requirement already satisfied: jsonschema>=2.6 in c:\users\manvika\appdata\loca
l\programs\python\python313\lib\site-packages (from nbformat->jupytext) (4.2
5.0)
Requirement already satisfied: jupyter-core!=5.0.*,>=4.12 in c:\users\manvika\a
ppdata\local\programs\python\python313\lib\site-packages (from nbformat->jupyte
xt) (5.8.1)
Requirement already satisfied: traitlets>=5.1 in c:\users\manvika\appdata\loca
l\programs\python\python313\lib\site-packages (from nbformat->jupytext) (5.1
4.3)
Requirement already satisfied: attrs>=22.2.0 in c:\users\manvika\appdata\local\
programs\python\python313\lib\site-packages (from jsonschema>=2.6->nbformat->ju
pytext) (25.3.0)
Requirement already satisfied: jsonschema-specifications>=2023.03.6 in c:\user
s\manvika\appdata\local\programs\python\python313\lib\site-packages (from jsons
chema>=2.6->nbformat->jupytext) (2025.4.1)
Requirement already satisfied: referencing>=0.28.4 in c:\users\manvika\appdata\
local\programs\python\python313\lib\site-packages (from jsonschema>=2.6->nbform
at->jupytext) (0.36.2)
Requirement already satisfied: rpds-py>=0.7.1 in c:\users\manvika\appdata\loca
l\programs\python\python313\lib\site-packages (from jsonschema>=2.6->nbformat-
>jupytext) (0.26.0)
Requirement already satisfied: platformdirs>=2.5 in c:\users\manvika\appdata\lo
cal\programs\python\python313\lib\site-packages (from jupyter-core!=5.0.*,>=4.1
2->nbformat->jupytext) (4.3.8)
Requirement already satisfied: pywin32>=300 in c:\users\manvika\appdata\local\p
rograms\python\python313\lib\site-packages (from jupyter-core!=5.0.*,>=4.12->nb
format->jupytext) (311)
[notice] A new release of pip is available: 25.2 -> 26.0.1
[notice] To update, run: python.exe -m pip install --upgrade pip
```