

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

```
In [3]: df=pd.read_csv(r'E:\SQL Files\customer_shopping_behavior.csv')
```

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

```
Out[4]:
```

	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
0	1	55	Male	Blouse	Clothing	53	Kentucky	L	Gray	Winter	3.1	Yes	Express	Yes	Yes	14	Venmo	Fortnightly
1	2	19	Male	Sweater	Clothing	64	Maine	L	Maroon	Winter	3.1	Yes	Express	Yes	Yes	2	Cash	Fortnightly
2	3	50	Male	Jeans	Clothing	73	Massachusetts	S	Maroon	Spring	3.1	Yes	Free Shipping	Yes	Yes	23	Credit Card	Weekly
3	4	21	Male	Sandals	Footwear	90	Rhode Island	M	Maroon	Spring	3.5	Yes	Next Day Air	Yes	Yes	49	PayPal	Weekly
4	5	45	Male	Blouse	Clothing	49	Oregon	M	Turquoise	Spring	2.7	Yes	Free Shipping	Yes	Yes	31	PayPal	Annually

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

```
Out[5]:
```

	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
count	3900.000000	3900.000000	3900	3900	3900	3900.000000	3900	3900	3900	3900	3863.000000	3900	3900	3900	3900.000000	3900	3900	
unique	Nan	Nan	2	25	4	Nan	50	4	25	4	Nan	2	6	2	2	Nan	6	7
top	Nan	Nan	Male	Blouse	Clothing	Nan	Montana	M	Olive	Spring	Nan	No	Free Shipping	No	No	Nan	PayPal	Every 3 Months
freq	Nan	Nan	2652	171	1737	Nan	96	1755	177	999	Nan	2847	675	2223	2223	Nan	677	584
mean	1950.500000	44.068462	Nan	Nan	Nan	59.764359	Nan	Nan	Nan	Nan	3.750065	Nan	Nan	Nan	25.351538	Nan	Nan	
std	1125.977353	15.207589	Nan	Nan	Nan	23.685392	Nan	Nan	Nan	Nan	0.716983	Nan	Nan	Nan	14.447125	Nan	Nan	
min	1.000000	18.000000	Nan	Nan	Nan	20.000000	Nan	Nan	Nan	Nan	2.500000	Nan	Nan	Nan	1.000000	Nan	Nan	
25%	975.750000	31.000000	Nan	Nan	Nan	39.000000	Nan	Nan	Nan	Nan	3.100000	Nan	Nan	Nan	13.000000	Nan	Nan	
50%	1950.500000	44.000000	Nan	Nan	Nan	60.000000	Nan	Nan	Nan	Nan	3.800000	Nan	Nan	Nan	25.000000	Nan	Nan	
75%	2925.250000	57.000000	Nan	Nan	Nan	81.000000	Nan	Nan	Nan	Nan	4.400000	Nan	Nan	Nan	38.000000	Nan	Nan	
max	3900.000000	70.000000	Nan	Nan	Nan	100.000000	Nan	Nan	Nan	Nan	5.000000	Nan	Nan	Nan	50.000000	Nan	Nan	

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

```
Out[6]:
```

	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
count	3900.000000	3900.000000	3900	3900	3900	3900.000000	3900	3900	3900	3900	3863.000000	3900	3900	3900	3900.000000	3900	3900	
unique	Nan	Nan	2	25	4	Nan	50	4	25	4	Nan	2	6	2	2	Nan	6	7
top	Nan	Nan	Male	Blouse	Clothing	Nan	Montana	M	Olive	Spring	Nan	No	Free Shipping	No	No	Nan	PayPal	Every 3 Months
freq	Nan	Nan	2652	171	1737	Nan	96	1755	177	999	Nan	2847	675	2223	2223	Nan	677	584
mean	1950.500000	44.068462	Nan	Nan	Nan	59.764359	Nan	Nan	Nan	Nan	3.750065	Nan	Nan	Nan	25.351538	Nan	Nan	
std	1125.977353	15.207589	Nan	Nan	Nan	23.685392	Nan	Nan	Nan	Nan	0.716983	Nan	Nan	Nan	14.447125	Nan	Nan	
min	1.000000	18.000000	Nan	Nan	Nan	20.000000	Nan	Nan	Nan	Nan	2.500000	Nan	Nan	Nan	1.000000	Nan	Nan	
25%	975.750000	31.000000	Nan	Nan	Nan	39.000000	Nan	Nan	Nan	Nan	3.100000	Nan	Nan	Nan	13.000000	Nan	Nan	
50%	1950.500000	44.000000	Nan	Nan	Nan	60.000000	Nan	Nan	Nan	Nan	3.800000	Nan	Nan	Nan	25.000000	Nan	Nan	
75%	2925.250000	57.000000	Nan	Nan	Nan	81.000000	Nan	Nan	Nan	Nan	4.400000	Nan	Nan	Nan	38.000000	Nan	Nan	
max	3900.000000	70.000000	Nan	Nan	Nan	100.000000	Nan	Nan	Nan	Nan	5.000000	Nan	Nan	Nan	50.000000	Nan	Nan	

```
In [8]: df['Review Rating']=df.groupby('Category')['Review Rating'].transform(lambda x : x.fillna(x.median()))
```

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

```
Out[10]:
```

	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
count	3900.000000	3900.000000	3900	3900	3900	3900.000000	3900	3900	3900	3900	3863.000000	3900	3900	3900	3900.000000	3900	3900	
unique	Nan	Nan	2	25	4	Nan	50	4	25	4	Nan	2	6	2	2	Nan	6	7
top	Nan	Nan	Male	Blouse	Clothing	Nan	Montana	M	Olive	Spring	Nan	No	Free Shipping	No	No	Nan	PayPal	Every 3 Months
freq	Nan	Nan	2652	171	1737	Nan	96	1755	177	999	Nan	2847	675	2223	2223	Nan	677	584
mean	1950.500000	44.068462	Nan	Nan	Nan	59.764359	Nan	Nan	Nan	Nan	3.750065	Nan	Nan	Nan	25.351538	Nan	Nan	
std	1125.977353	15.207589	Nan	Nan	Nan	23.685392	Nan	Nan	Nan	Nan	0.716983	Nan	Nan	Nan	14.447125	Nan	Nan	
min	1.000000	18.000000	Nan	Nan	Nan	20.000000	Nan	Nan	Nan	Nan	2.500000	Nan	Nan	Nan	1.000000	Nan	Nan	
25%	975.750000	31.000000	Nan	Nan	Nan	39.000000	Nan	Nan	Nan	Nan	3.100000	Nan	Nan	Nan	13.000000	Nan	Nan	
50%	1950.500000	44.000000	Nan	Nan	Nan	60.000000	Nan	Nan	Nan	Nan	3.800000	Nan	Nan	Nan	25.000000	Nan	Nan	
75%	2925.250000	57.000000	Nan	Nan	Nan	81.000000	Nan	Nan	Nan	Nan	4.400000	Nan	Nan	Nan	38.000000	Nan	Nan	
max	3900.000000	70.000000	Nan	Nan	Nan	100.000000	Nan	Nan	Nan	Nan	5.000000	Nan	Nan	Nan	50.000000	Nan	Nan	

```
In [11]: (df['Discount Applied']==df['Promo Code Used']).all()
```

```
Out[11]: np.True_
```

```
In [14]: !pip install pymysql sqlalchemy
import pandas as pd
from sqlalchemy import create_engine

username="root"
password="mahi7781"
host="localhost"
port="3306"
database="customer_behavior"

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

table_name ="customer"
df.to_sql(table_name, engine, if_exists="replace", index=False)

pd.read_sql("select* from customer limit 5",engine)
```

[notice] A new release of pip is available: 25.3 -> 26.0.1
[notice] To update, run: python.exe -m pip install --upgrade pip
Requirement already satisfied: pymysql in c:\users\welcome\appdata\local\programs\python\python314\lib\site-packages (1.1.2)
Requirement already satisfied: sqlalchemy in c:\users\welcome\appdata\local\programs\python\python314\lib\site-packages (2.0.46)
Requirement already satisfied: greenlet>=1 in c:\users\welcome\appdata\local\programs\python\python314\lib\site-packages (from sqlalchemy) (3.3.1)
Requirement already satisfied: typing-extensions>=4.6.0 in c:\users\welcome\appdata\local\programs\python\python314\lib\site-packages (from sqlalchemy) (4.15.0)

```
Out[14]:
```

	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
0	1	55	Male	Blouse	Clothing	53	Kentucky	L	Gray	Winter	3.1	Yes	Express	Yes	Yes	14	Venmo	Fortnightly
1	2	19	Male	Sweater	Clothing	64	Maine	L	Maroon	Winter	3.1	Yes	Express	Yes	Yes	2	Cash	Fortnightly
2	3	50	Male	Jeans	Clothing	73	Massachusetts	S	Maroon	Spring	3.1	Yes	Free Shipping	Yes	Yes	23	Credit Card	Weekly
3	4	21	Male	Sandals	Footwear	90	Rhode Island	M	Maroon	Spring	3.5	Yes	Next Day Air	Yes	Yes	49	PayPal	Weekly
4	5	45	Male	Blouse	Clothing	49	Oregon	M	Turquoise	Spring	2.7	Yes	Free Shipping	Yes	Yes	31	PayPal	Annually

```
In [15]: df.to_csv("clean_customer_data.csv", index=False)
```

