Practice Project 6 - Dataframe - Advanced groupby & Filter operations

Table of Contents

1.Groupby operation with aggregation.

In [ ]:
```
import pandas as pd

# Read the CSV file into a DataFrame
df = pd.read_csv('data/data.csv')

# Group by 'category' column and calculate the sum of 'quantity' and 'price'
result = df.groupby('category').agg({'quantity': 'sum', 'price': 'sum'})

# Print the result
print(result)
```

```
          quantity       price
category
A              248  308.641752
B              236  318.870926
C              286  321.370045
```

2.Groupby operation with multiple aggregation functions.

In [ ]:
```
import pandas as pd

# Read the CSV file into a DataFrame
df = pd.read_csv('data/data.csv')

# Group by 'category' column and calculate the sum, mean, and count of 'quantity'
result = df.groupby('category')['quantity'].agg(['sum', 'mean', 'count'])
```

```
# Print the result
print(result)
```

```
          sum       mean  count
category
A         248  35.428571      7
B         236  33.714286      7
C         286  47.666667      6
```

3.Filter operation using groupby.

In [ ]:

```
import pandas as pd

# Read the CSV file into a DataFrame
df = pd.read_csv('data/data.csv')

# Group by 'category' column and filter the groups where the sum of 'quantity' is greater than 100
result = df.groupby('category').filter(lambda x: x['quantity'].sum() > 100)

# Print the filtered DataFrame
print(result)
```

```
   category  quantity      price
0         B        34  33.971175
1         C        47  60.972681
2         A         2  14.357632
3         B         6  87.066915
4         A        68  18.152921
5         B        33  13.826712
6         A        37  18.313960
7         B         3  75.365077
8         B        58  45.595639
9         B         4  42.892304
10        A        48  34.771635
11        B        98  20.153104
12        A         7  59.936505
13        A        17  90.011499
14        C        12  62.963619
15        C        15  50.981245
16        C        61  27.310937
17        C        84  69.427472
18        A        69  73.097601
19        C        67  49.714092
```

4.Filter operation using multiple conditions

In [ ]:

```
import pandas as pd

# Read the CSV file into a DataFrame
df = pd.read_csv('data/data.csv')

# Filter the DataFrame based on multiple conditions
result = df[(df['category'] == 'A') & (df['price'] > 50)]

# Print the filtered DataFrame
print(result)
```

```
   category  quantity      price
12        A         7  59.936505
13        A        17  90.011499
18        A        69  73.097601
```