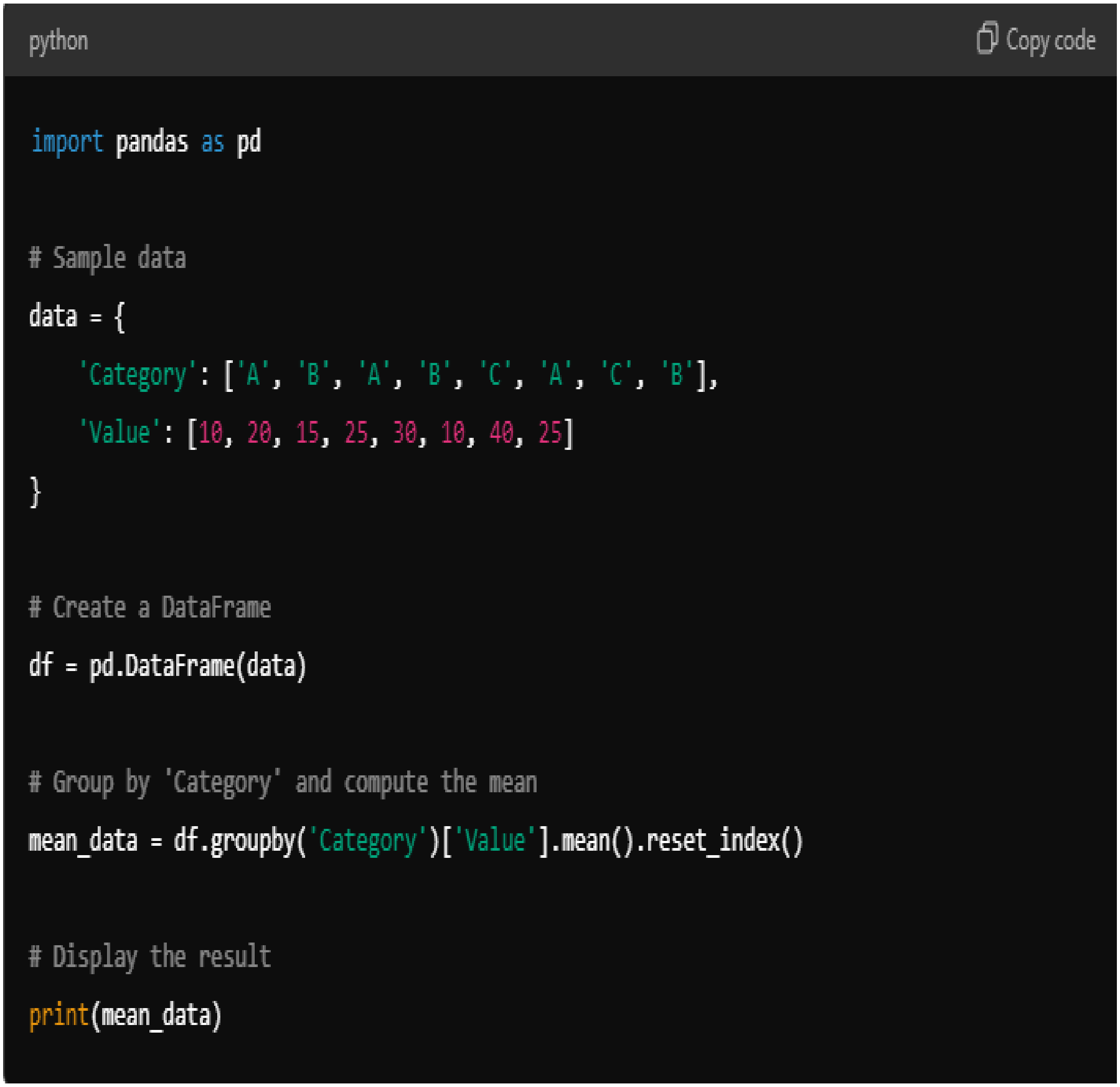
EXPERIMENT NO. 3

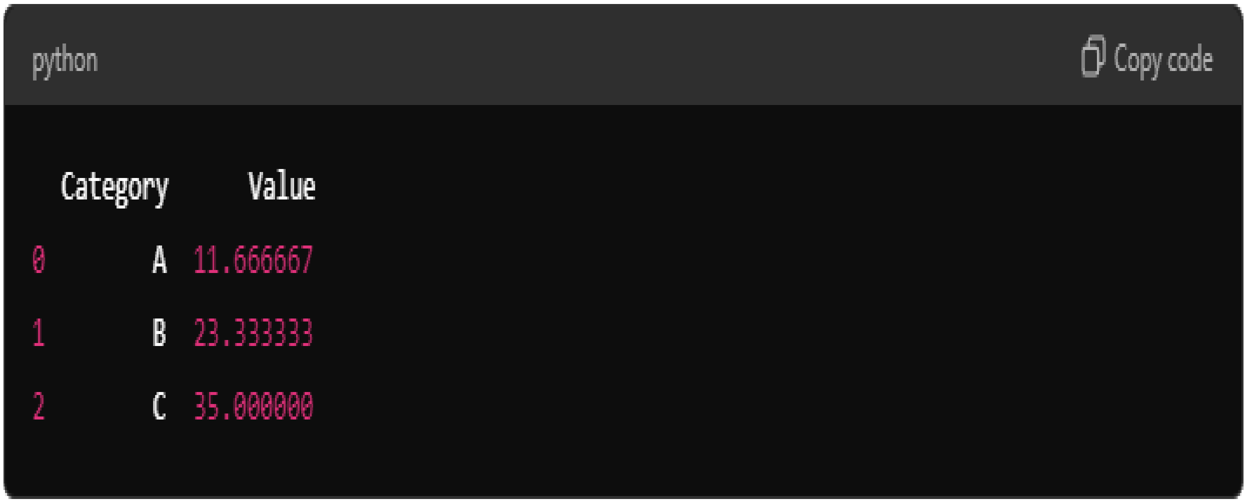
Data Aggregation: Aggregate data by grouping rows based on a specific attribute and computing summary statistics, such as mean, median, or count.

1. **Compute Mean**



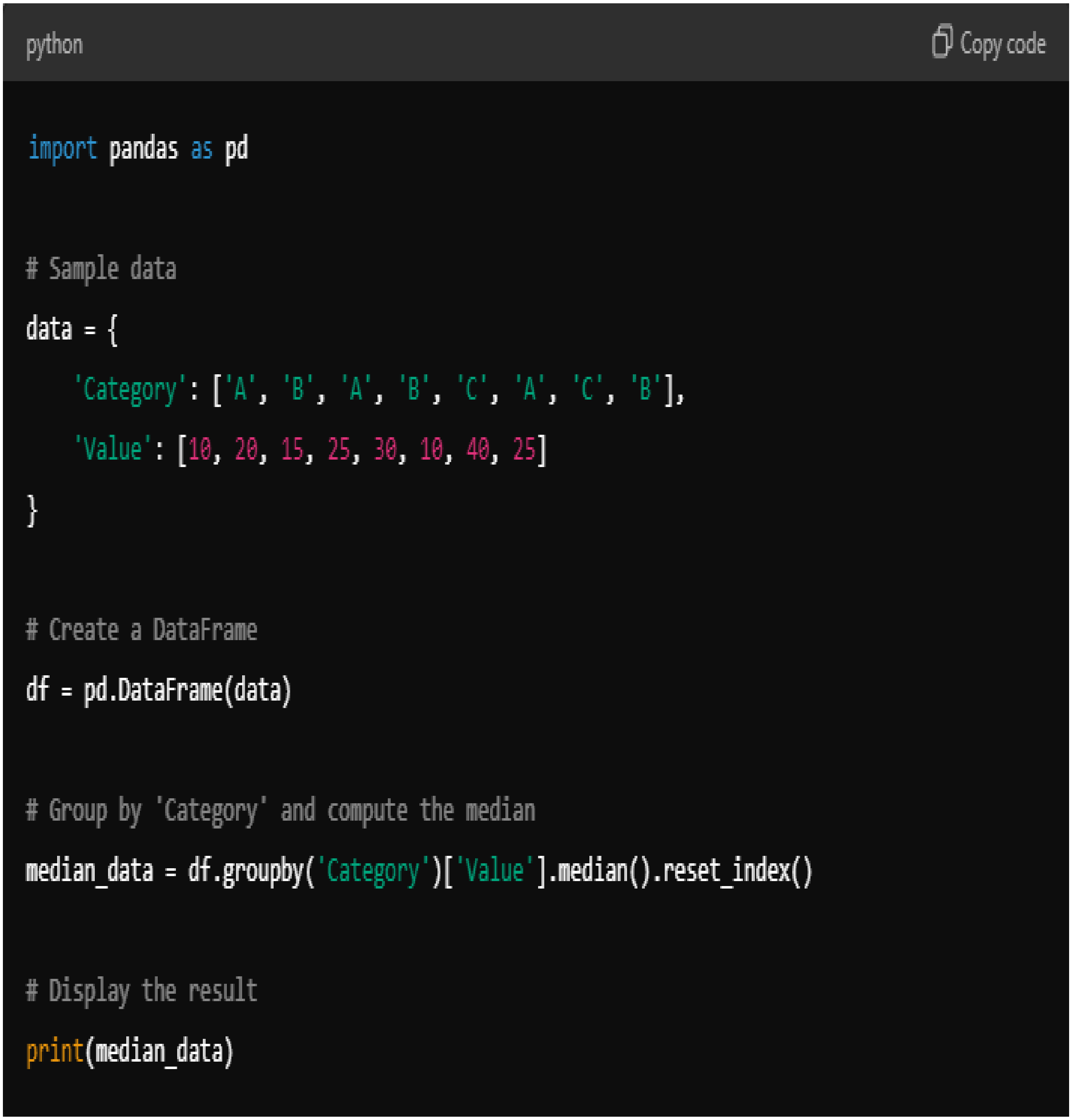
```python
import pandas as pd

# Sample data
data = {
    'Category': ['A', 'B', 'A', 'B', 'C', 'A', 'C', 'B'],
    'Value': [10, 20, 15, 25, 30, 10, 40, 25]
}

# Create a DataFrame
df = pd.DataFrame(data)

# Group by 'Category' and compute the mean
mean_data = df.groupby('Category')['Value'].mean().reset_index()

# Display the result
print(mean_data)
```

Output:



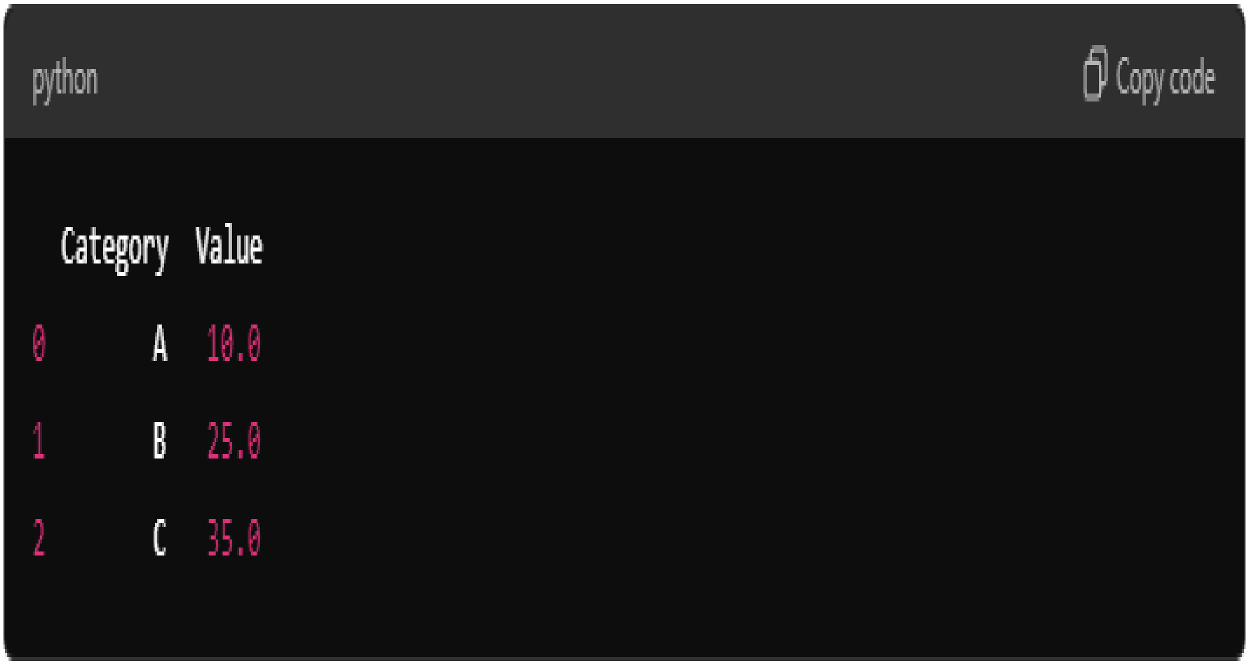
```
  Category      Value
0        A  11.666667
1        B  23.333333
2        C  35.000000
```

2. **Compute Median**



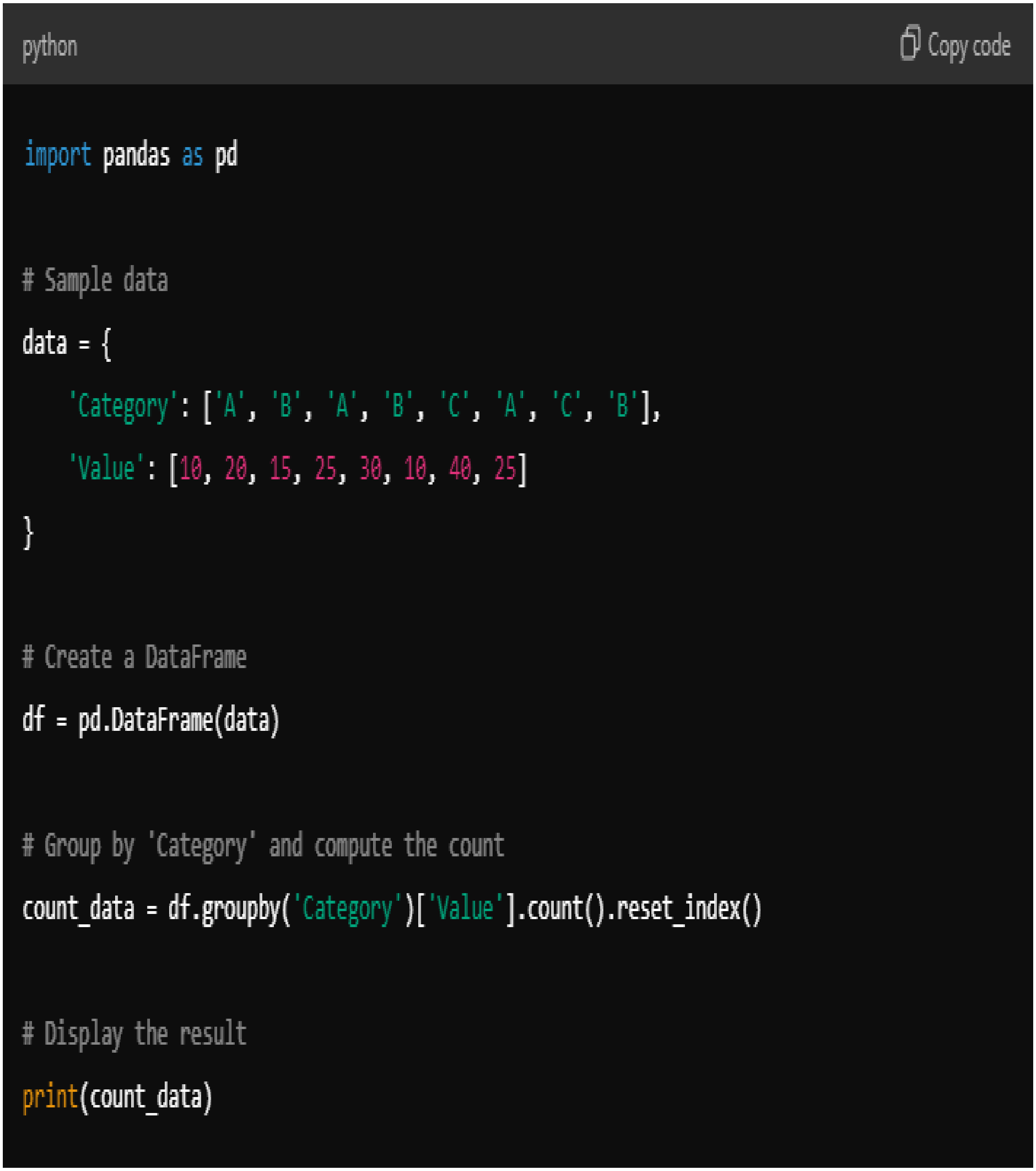
```python
import pandas as pd

# Sample data
data = {
    'Category': ['A', 'B', 'A', 'B', 'C', 'A', 'C', 'B'],
    'Value': [10, 20, 15, 25, 30, 10, 40, 25]
}

# Create a DataFrame
df = pd.DataFrame(data)

# Group by 'Category' and compute the median
median_data = df.groupby('Category')['Value'].median().reset_index()

# Display the result
print(median_data)
```

Output:



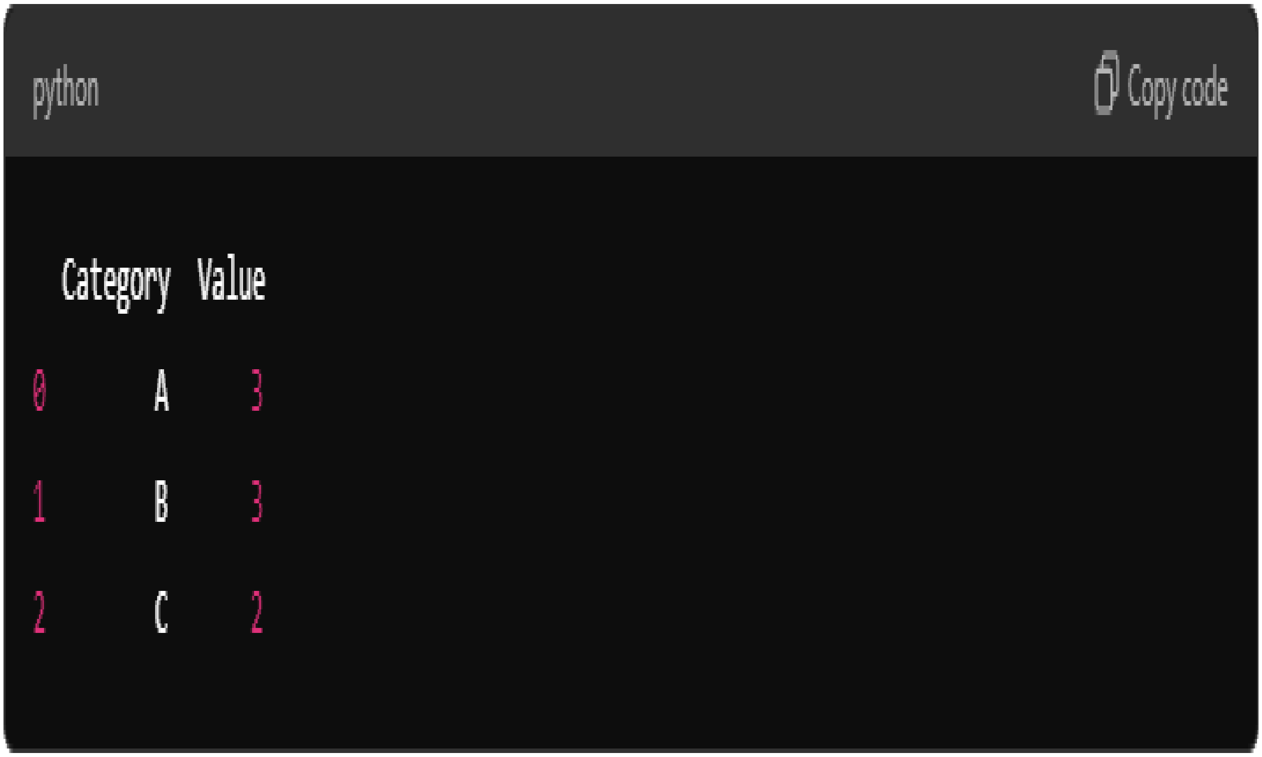
```
  Category  Value
0        A   10.0
1        B   25.0
2        C   35.0
```

3. **Compute Count**



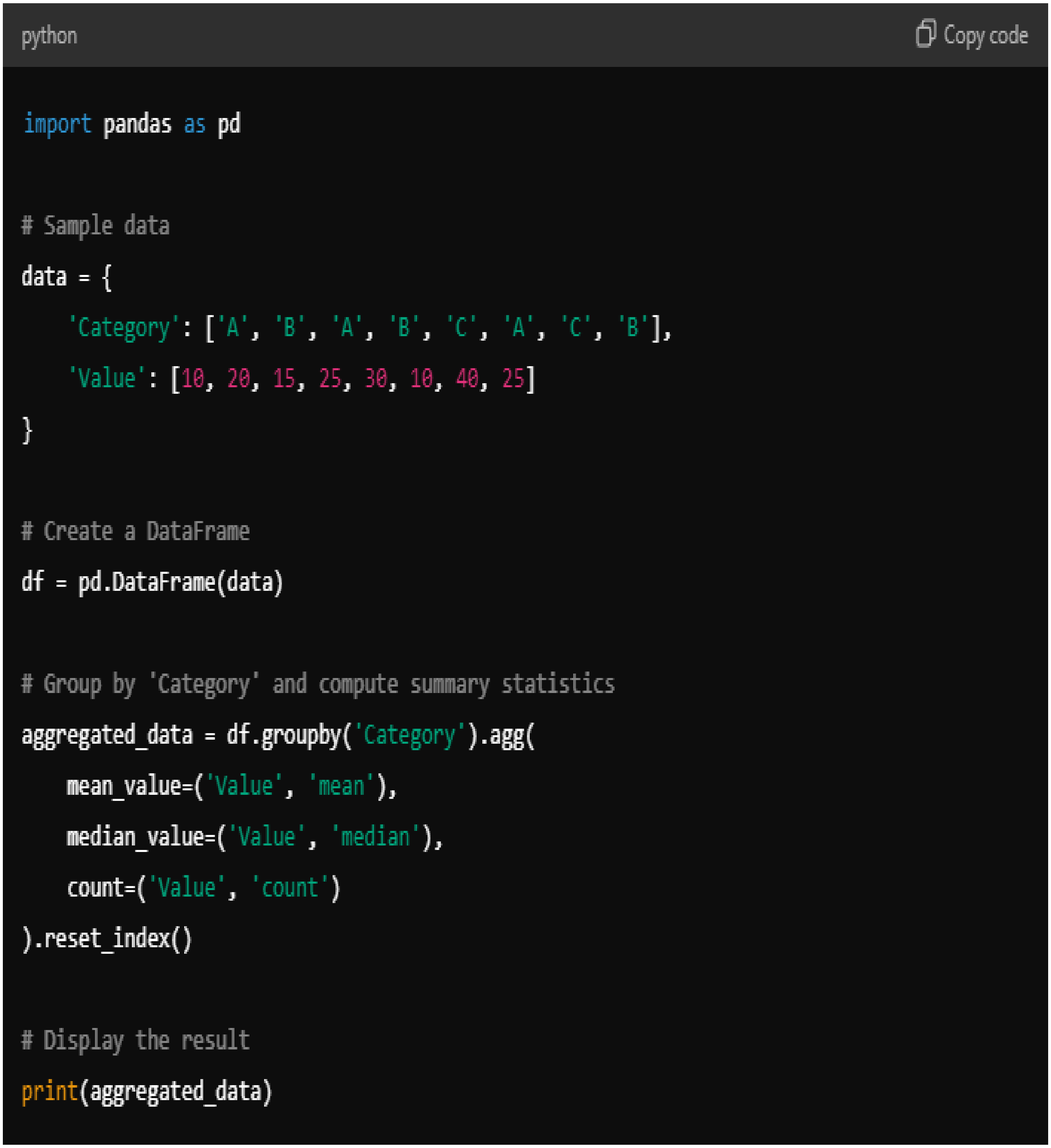
```python
import pandas as pd

# Sample data
data = {
    'Category': ['A', 'B', 'A', 'B', 'C', 'A', 'C', 'B'],
    'Value': [10, 20, 15, 25, 30, 10, 40, 25]
}

# Create a DataFrame
df = pd.DataFrame(data)

# Group by 'Category' and compute the count
count_data = df.groupby('Category')['Value'].count().reset_index()

# Display the result
print(count_data)
```

Output:



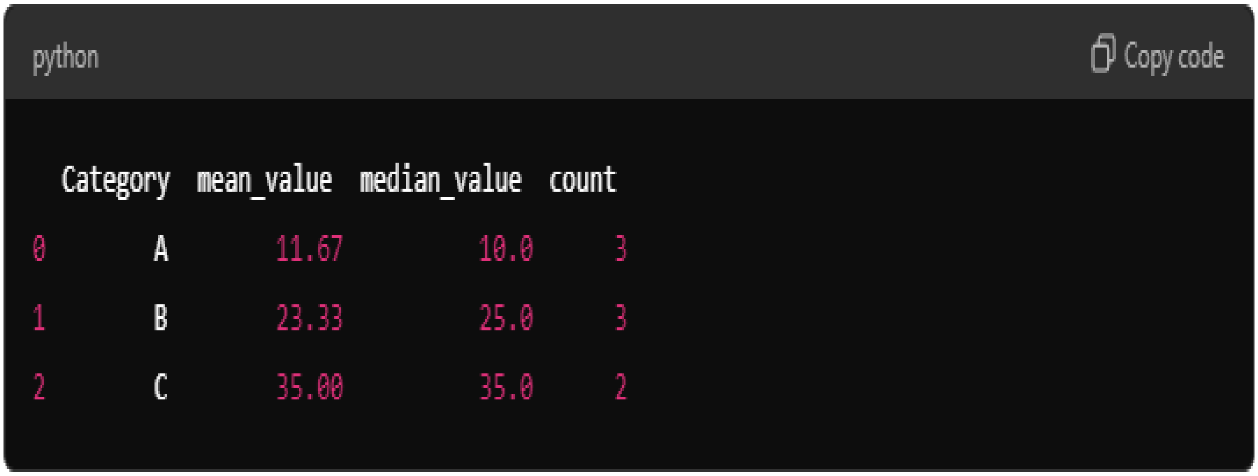
```
  Category  Value
0        A      3
1        B      3
2        C      2
```

Aggregate data by grouping rows based on a specific attribute and computing summary statistics like mean, median, and count:



```python
import pandas as pd

# Sample data
data = {
    'Category': ['A', 'B', 'A', 'B', 'C', 'A', 'C', 'B'],
    'Value': [10, 20, 15, 25, 30, 10, 40, 25]
}

# Create a DataFrame
df = pd.DataFrame(data)

# Group by 'Category' and compute summary statistics
aggregated_data = df.groupby('Category').agg(
    mean_value=('Value', 'mean'),
    median_value=('Value', 'median'),
    count=('Value', 'count')
).reset_index()

# Display the result
print(aggregated_data)
```

Output:



```
  Category  mean_value  median_value  count
0        A       11.67          10.0      3
1        B       23.33          25.0      3
2        C       35.00          35.0      2
```