assi3a

January 27, 2026

```
[1]: import pandas as pd
     import numpy as np
```

```
[5]: df = pd.read_csv("retail_store_sales.csv")
```

```
[7]: df.head()
```

```
[7]:   Transaction ID Customer ID       Category          Item  Price Per Unit  \
    0    TXN_6867343     CUST_09     Patisserie   Item_10_PAT            18.5
    1    TXN_3731986     CUST_22  Milk Products  Item_17_MILK            29.0
    2    TXN_9303719     CUST_02       Butchers   Item_12_BUT            21.5
    3    TXN_9458126     CUST_06      Beverages   Item_16_BEV            27.5
    4    TXN_4575373     CUST_05           Food   Item_6_FOOD            12.5

       Quantity  Total Spent  Payment Method Location Transaction Date  \
    0      10.0        185.0  Digital Wallet   Online       2024-04-08
    1       9.0        261.0  Digital Wallet   Online       2023-07-23
    2       2.0         43.0     Credit Card   Online       2022-10-05
    3       9.0        247.5     Credit Card   Online       2022-05-07
    4       7.0         87.5  Digital Wallet   Online       2022-10-02

      Discount Applied
    0             True
    1             True
    2            False
    3              NaN
    4            False
```

```
[9]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 12575 entries, 0 to 12574
Data columns (total 11 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   Transaction ID    12575 non-null  object
 1   Customer ID       12575 non-null  object
 2   Category          12575 non-null  object
 3   Item              11362 non-null  object
 4   Price Per Unit    11966 non-null  float64
 5   Quantity          11971 non-null  float64
 6   Total Spent       11971 non-null  float64
 7   Payment Method    12575 non-null  object
 8   Location          12575 non-null  object
 9   Transaction Date  12575 non-null  object
 10  Discount Applied  8376 non-null   object
dtypes: float64(3), object(8)
memory usage: 1.1+ MB
```

[13]:

```
df[['Category', 'Total Spent']].head()
```

[13]:

```
         Category  Total Spent
0      Patisserie        185.0
1   Milk Products        261.0
2        Butchers         43.0
3       Beverages        247.5
4            Food         87.5
```

[15]:

```
grouped = df.groupby("Category")["Total Spent"]
```

[17]:

```
summary_stats = grouped.agg(
    Mean="mean",
    Median="median",
    Minimum="min",
    Maximum="max",
    Standard_Deviation="std"
)
```

[19]:

```
summary_stats
```

[19]:

```
                                          Mean  Median  Minimum  Maximum  \
Category
Beverages                           131.716243  109.50      5.0    410.0
Butchers                            139.116310  119.25      5.0    410.0
Computers and electric accessories  129.107989  110.00      5.0    410.0
Electric household essentials       134.441623  110.50      5.0    410.0
Food                                129.271400  110.00      5.0    410.0
Furniture                           128.072131  100.50      5.0    410.0
Milk Products                       119.042961   92.00      5.0    410.0
Patisserie                          126.416031  104.00      5.0    410.0

                                    Standard_Deviation
Category
Beverages                                    98.473704
Butchers                                    100.344679
Computers and electric accessories           88.682576
Electric household essentials                95.619428
```

```
Food                                         90.599147
Furniture                                    98.268320
Milk Products                                92.871603
Patisserie                                   91.166165
```

```
[21]: total_spent_list = df.groupby("Category")["Total Spent"].apply(list)
```

```
[24]: total_spent_list
```

```
[24]: Category
      Beverages                             [247.5, nan, 126.0, 287.0, 100.0, 255.5,
      46.5,...
      Butchers                              [43.0, 109.5, 72.0, 134.0, 190.0, 276.5,
      165.0...
      Computers and electric accessories    [165.0, 66.0, 11.0, 234.0, 316.0, 255.5,
      335.0...
      Electric household essentials         [23.0, 66.5, 315.0, 92.5, 70.0, 76.0,
      12.5, 62...
      Food                                  [87.5, 40.0, 45.5, 18.5, 245.0, 200.0,
      196.0, ...
      Furniture                             [nan, 27.5, nan, 49.0, nan, 410.0, 92.5,
      190.0...
      Milk Products                         [261.0, 52.0, 55.0, 275.0, 105.0, 192.5,
      96.0,...
      Patisserie                            [185.0, 200.0, 237.0, 232.0, 182.5, 100.0,
      nan...
      Name: Total Spent, dtype: object
```

```
[50]: total_spent_dict = total_spent_list.to_dict()
```

```
[29]: summary_stats.to_csv("category_wise_total_spent_statistics.csv")
```