**Exploratory Data Analysis (EDA) Report**

**Project Context:**

This analysis is the first phase of a broader project investigating business performance for an online clothing store. The dataset covers transactions from January to June 2024 across five product categories, multiple subcategories, four regions, and two customer types.

The goal was to develop a foundational understanding of the dataset’s structure, value distributions, and initial patterns in sales performance. Key variables explored include unit price, discount, sales value, and units sold.

**Summary Statistics**

I began by reviewing the overall distribution of key numerical variables:

* Units Sold per transaction ranged from 1 to 4, with an average of 2.5 units.
* Unit Price ranged from $5.00 to $149.94, with an average price of approximately $37.49.
* Discounts were modest across transactions, averaging 8.2%, with a maximum of 30%.
* Sales per transaction ranged from $3.87 to $580.19, with an average of $85.93.

This distribution suggests a healthy mix of low- and high-value transactions, with pricing and discounting strategies appearing to be moderate and consistent.

**Frequency Counts**

I examined the counts of different categories in the dataset:

|  |  |
| --- | --- |
| Categories | Counts |
| Headwear | 1,364 |
| Bodywear | 1,350 |
| Legwear | 1,331 |
| Handwear | 1,318 |
| Footwear | 1,285 |

The frequency of is relatively even, indicating balanced representation across product categories.

**Aggregate Metrics**

To better understand performance, I calculated the total sales, units sold, and discounts key.

|  |  |
| --- | --- |
| Categories | Counts |
| Total Sales | $571,237.59 |
| Total Units sold | 16,625 |
| Total Discount | 547.75 |

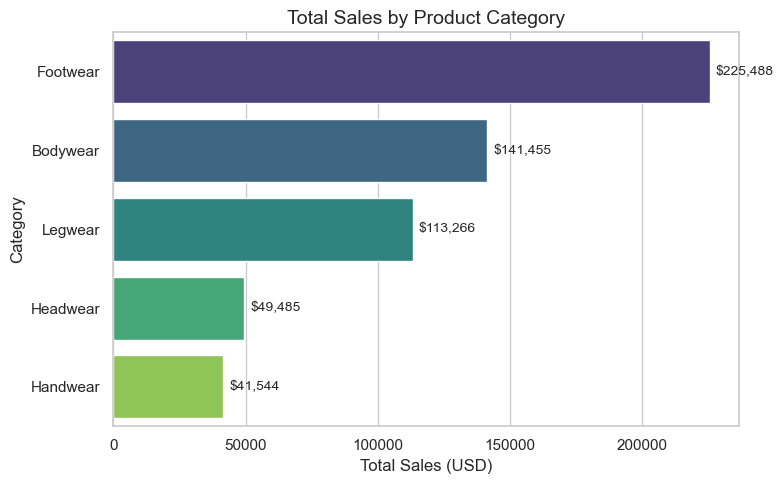
Note: Discounts are recorded as proportions (e.g., 0.10 = 10%). The total reflects the cumulative discount proportion applied across all transactions, it’s not a dollar value.

These figures provide a high-level benchmark for overall business activity in the first half of 2024.

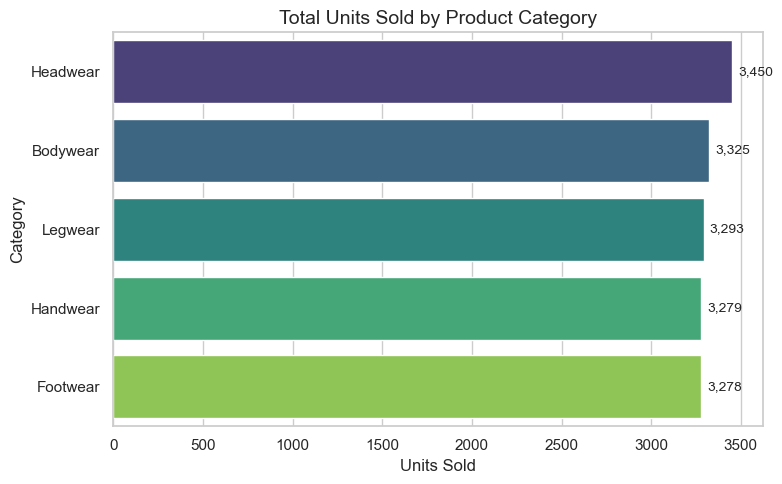
**Category-Based Sales Analysis**

I also examined how product categories performed in terms of revenue, pricing, and units sold.

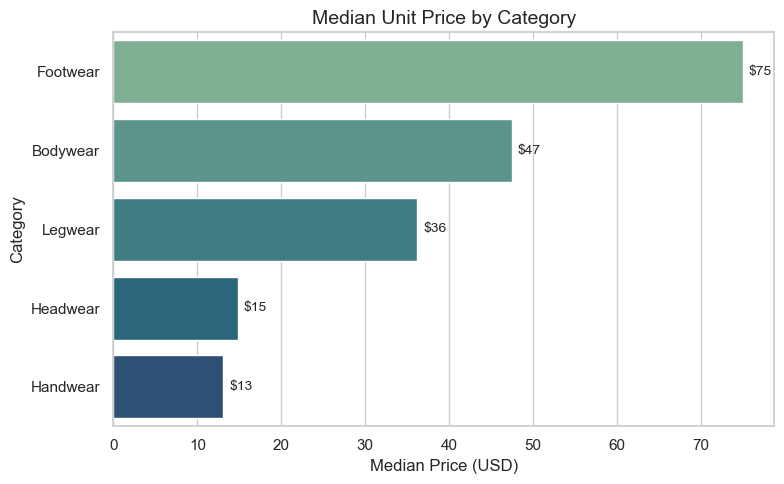
**Total Sales by Category**



**Units Sold by Category**



**Median Unit Price by Category**



**Total Discount by Category (Proportion)**

|  |  |
| --- | --- |
| Categories | Counts |
| Headwear | 0.08 |
| Bodywear | 0.08 |
| Legwear | 0.08 |
| Handwear | 0.07 |
| Footwear | 0.07 |

**Insight:**

* Footwear outperformed all categories in total sales, despite having fewer units sold than Headwear or Bodywear. This is due to significantly higher prices, not higher volume or deeper discounts.
* Headwear and Handwear, though frequently sold, contribute far less to revenue due to their low-price points.
* Discounts were fairly uniform across categories, suggesting discounting wasn't a key factor in category-level performance.
* These findings reinforce that product pricing and perceived value, rather than quantity or discounts giving, are likely driving revenue differences.

**Conclusion**

This EDA laid a strong foundation by helping me understand the structure and performance dynamics of the clothing store's data. I’ve identified pricing and category value as major drivers of performance.