SUMMARY

Superstore USA Data Analysis

1. Dataset Overview

The Superstore USA dataset contains retail transaction records, including details like order priority, sales, profit, product categories, and customer information. The analysis focuses on understanding data quality, order prioritization, and drawing insights from sales operations.

---

2. Data Cleaning & Preparation

a. Handling Missing Values

The column *Product Base Margin* had missing values. Instead of removing rows, the missing entries were filled using the average value of that column. This approach preserved data volume while maintaining consistency.

b. Fixing Inconsistencies in Order Priority

The *Order Priority* column had inconsistent formatting, such as extra spaces and varied casing (e.g., "critical" vs. "Critical"). These inconsistencies led to duplicated categories in visualizations. The values were standardized to ensure clean and accurate grouping, which improved the reliability of charts and summaries.

---

3. Exploratory Analysis

a. Order Priority Distribution

A bar chart was used to understand the distribution of order priorities. It showed that the levels – Critical, High, Medium, Low, and Not Specified – are fairly balanced. However, the presence of "Not Specified" suggests that some records may have been missing or incomplete in terms of operational classification.

b. Pie Chart Visualization

The pie chart helped visualize the proportion of each priority level. The distribution was again fairly even, which suggests that the store manages orders with varying urgency but without a clear skew toward any single level. This may imply consistent handling across different urgency levels.

---

4. Key Insights

- **Balanced Order Priorities**: No single priority category dominates, indicating uniform order handling policies.

- **Data Standardization Matters**: Small inconsistencies in text formatting caused duplication in plots, highlighting the importance of cleaning text data.

- **Incomplete Data Flags**: The "Not Specified" priority group suggests that some orders may lack essential classification, which can affect operations or performance analysis.

- **Visual Representation Added Clarity**: Charts helped validate the data cleaning process and exposed underlying issues like label duplication.

---

5. Recommendations & Next Steps

- **Audit Orders with 'Not Specified' Priority**: Investigate if these were errors or exceptions and ensure data entry protocols prevent such gaps.

- **Drill Down by Product and Region**: Understanding how priorities vary by product category or customer region may reveal logistical or strategic patterns.

- **Track Changes Over Time**: Analyzing priority distribution across months or years could help identify operational shifts or seasonal demands.