Assignment 2
ZARA Sales
Nourhan Mersal

I selected the Zara sales dataset because it directly relates to e-commerce and retail sales transactions, making it highly relevant to business analysis.
Dataset Scope

This dataset contains 252 unique order IDs detailing product sales from Zara stores on February 19, 2024, recorded between 8:50:06 and 9:10:52.

Descriptive Analysis

	Sales Volume	price
mean	1823.7	86.3
median	1839.5	79.9
mode	1466.0	89.9
standard deviation	697.7	52.1
minimum	529.0	8.0
maximum	2989.0	439.0
Range	2460.0	431.0
Q1	1243.0	49.9
Q3	2398.8	109.0
IQR	1155.8	59.1
Outliers (**Lower)**	-490.6	-38.8
Outliers (**upper)**	4132.4	197.7

Sales Volume:

The **Sales Volume** data are **clean and well-behaved**, as **no outliers** were detected.

- **No Extreme Values:** All data points fall within the expected, natural range, indicating a homogeneous distribution.
- **Data Reliability:** The absence of extreme values suggests the data is likely reliable and free from significant errors (e.g., typos or measurement mistakes) that would create anomalous spikes or dips.
- **Normal Variation:** The variations observed are standard and do not indicate any special anomalies or "signals" that require immediate investigation.

Item Price

The Item Price data shows positive skewness and contains four upper outliers, but no lower outliers were found.

- **Potential Data Errors:** The outliers *may* indicate data entry errors that resulted in unusually high price points.
- **Unique Product Representation:** Alternatively, these outliers could be legitimate values representing the sales of genuinely high-end or luxury products within the dataset.

Insights:

1. **Pivot 2:** Product Position Products located in the Aisle generate the highest sales, accounting for 38% of total sales volume. This suggests that these items hold a prime placement or are in high demand.
2. **Pivot 3:** Sales by Section The Men's section sales significantly dominate the transactions, representing 86.21% of sales volume compared to 13.79% for Women's. The PLAID TIE DYE OVERSHIRT is the top-selling men's item, and the GATHERED WAIST KNIT SWEATER is the top-selling women's item.
3. **Pivot 4:** Pricing Strategy Sales volume correlates strongly with the mode price (the most common price point). The highest sales volumes are achieved by products priced at this accessible point. Conversely, the highest-priced items have the lowest sales volume. This pattern suggests an effective pricing strategy where current pricing aligns well with customer willingness to buy.
4. **Pivot 5:** Promotion Impact Promotions do not appear to be driving sales, as the sales volume for items without a promotion is higher than for items with one. This finding warrants further investigation into the effectiveness and targeting of current promotional campaigns.

More insights:
- **Total Revenue:** $38,988,476.48
- **Average Item Price:** $86.30
- **Average Sales Volume per Second:** 1.762 units
- **The item generating the Highest-Revenue is the** VINTAGE EFFECT LEATHER BOMBER JACKET (priced at $299).
- **The Top-Selling Product** (by volume) is the PLAID TIE DYE OVERSHIRT.
- **The Top-Selling Category** (by volume) is Men.

Metric	Sales Volume (Units)	Price ($)	Insight/Implication
Median vs. Mean	Median (1839.5) ≈ Mean (1823.7)	Mean (86.3) > Median (79.9)	Sales Volume is Symmetrical: Consistent demand. Price is Right-Skewed: The few higher-priced items (outliers) pull the average price up.
Mode Price	Mode Price (89.9) aligns with Highest Volume	89.9	Validated "Sweet Spot": The most common price point is also the most successful in driving volume.
Standard Deviation	SD (697.7) is high relative to the mean	SD (52.1) is high relative to the mean	High Variability: Both units sold and pricing show large spreads, indicating a diverse product mix and fluctuating demand within the short window.
Outliers	No Outliers	4 Upper Outliers	No Volume Anomalies: Sales performance is internally consistent. Pricing Extremes Exist: Four products skew the overall price analysis (likely luxury/unique items).

Key Business Decisions and Actions:

1. **Men's Dominance (86.21%):** Review inventory levels for top-selling men's items (like the PLAID TIE DYE OVERSHIRT) to prevent stock-outs.
2. **Women's Underperformance (13.79%):** Investigate and Reallocate,Is the target customer profile for women's wear not active during this 8:50-9:10 AM window? Analyze historical data to find their peak purchasing time. If performance is consistently low, reallocate marketing budget *away* from this time slot or aggressively promote top-selling women's items during this time to test for volume lift.
3. **Highest Selling Women's Item (GATHERED WAIST KNIT SWEATER):**Give this specific item a prime position (e.g., Aisle slot or prominent banner) and ensure its price is competitive.
4. **Aisle" Position:** Maintain and analyze the current product mix within the "Aisle" position. Test moving lower-performing men's items into the "Aisle" to see if sales lift, leveraging the established customer behavior
5. **High-Revenue Item:**Analyze margins on this high-revenue item. Ensure it is given premium online positioning (e.g., above the fold) and is included in any high-value customer retargeting campaigns.
6. **Highest Volume at Mode Price ($89.9):** Review pricing for products near but not at the mode price (e.g., $79.9 or $99.9). Consider moving high-volume items closer to $89.9 to maximize demand. This is your price elasticity ceiling for volume.
7. **Sales Volume without Promotion is Higher:** non promotional item sales more, Immediately investigate the type of promotions used. Are they *discounting* low-demand items?
8. **Upper Outliers (4 items above $197.7):** Audit the 4 items to confirm they are legitimate prices (not errors). If legitimate, create a "Luxury/Premium" segment for these products and exclude them from general promotional analysis to avoid skewing overall results.

Data Analysis Steps:

1. Removed Duplicates by checking the Product ID.
2. Identified all Data Types for all columns in a separate worksheet.
3. Separated Date and Time from the 'scraped_at' column to analyze time.
4. Calculated Key Metrics (mean, median, mode, etc.) for Sales Volume and Item Price.
5. Identified and Flagged Outliers using the IQR method to check for extreme values in both columns.
6. Calculated Revenue (Sales Volume × Item Price).
7. Used Pivot Tables to quickly summarize the data and extract the key business insights.