

Data Processing & AI Report

Cleaning Summary

Rows before: 10

Rows after cleaning: 10

Rows after transformations: 10

Removed duplicates: 0

Filled missing numeric: 0

Filled missing text: 0

Removed invalid rows: 0

Transformations Summary

Created 'total' column, grouped and aggregated data, filtered invalid/zero totals, and sorted by total.

Top Grouped Data (Preview)

product	total_sales	avg_total	count_rows
chair	240.0	240.00	1
desk	300.0	300.00	1
headphones	80.0	80.00	1
laptop	750.0	750.00	1
monitor	200.0	200.00	1
mouse	50.0	50.00	1
notebook	82.5	41.25	2
pen	84.0	42.00	2

AI Analysis

```
sdk_http_response=HttpResponse(  
    headers=<dict len=11>  
) candidates=[Candidate(  
    avg_logprobs=-0.6865237871285547,  
    content=Content(  
        parts=[  
            Part(  
                text="""## Sales Data Analysis Report
```

Date: October 26, 2023

1. Dataset Summary:

The dataset contains information on 10 customer orders, including the order ID, date, customer name, product, product category, price, quantity, city, and total amount. The data spans a relatively short timeframe. The categories available are: 'electronics', 'furniture', and 'office supplies'.

2. Key Patterns and Insights:

* **Electronics Dominance:** Based on the limited data, electronics appear to be the top-selling category by value due to the sale of a laptop. However, more data is needed to confirm if this is consistent.

* **Varying Order Sizes:** The quantity of products ordered varies significantly, ranging from 1 to 50 items. This suggests different customer needs (individual vs. bulk).

* **Product Diversity:** The dataset includes a diverse set of products, reflecting a range of customer needs within the available categories.

3. Anomalies and Data Quality Issues:

- * **Small Sample Size:** A dataset of 10 rows is insufficient for drawing reliable conclusions. Any insights derived from this data should be considered preliminary.
- * **Date Format:** The 'date' column is stored in a numeric timestamp format. This is not human-readable and needs to be converted to a standard date format for easier analysis and interpretation.
- * **Inconsistent Total Calculation:** Examining the head, it appears that total is not calculated from `price` * `quantity`. `order_id` 1001 has price 1.2 and quantity 20, but a total of 24, not $1.2 \times 20 = 24$. This should be fixed.
- * **Missing Totals Calculation:** The group `pen` has two sales which likely differ in `price` and/or `quantity` but have the same total of 84 across both.

4. Business Recommendations:

- * **Data Collection Improvement:** Prioritize expanding the dataset to include a larger sample size covering a longer period. This will enable more robust analysis and accurate identification of trends.
- * **Date Format Conversion:** Convert the 'date' column to a standard date format (e.g., YYYY-MM-DD) for improved readability and analysis.
- * **Investigate Total Calculation:** Investigate the data quality issue where the total cost may be incorrect or inconsistent with the price and quantity. Fix this to give correct figures.
- * **Targeted Marketing:**
 - * Conduct further analysis to understand the demographics and purchasing behavior of customers buying electronics, particularly high-value items like laptops.
 - * Explore opportunities to offer bundled deals or promotions on frequently purchased items like pens and notebooks.

Disclaimer: These recommendations are based on a limited dataset. A more comprehensive analysis is needed to validate these findings and develop more effective business strategies.

```
"""
),
],
role='model'
),
finish_reason=<FinishReason.STOP: 'STOP'>
)]         create_time=None           model_version='gemini-2.0-flash'           prompt_feedback=None
response_id='TvMvadLZBfq9juMPk9XjqAo' usage_metadata=GenerateContentResponseUsageMetadata(
candidates_token_count=611,
candidates_tokens_details=[

    ModalityTokenCount(
        modality=<MediaModality.TEXT: 'TEXT'>,
        token_count=611
    ),
],
prompt_token_count=1227,
prompt_tokens_details=[

    ModalityTokenCount(
        modality=<MediaModality.TEXT: 'TEXT'>,
        token_count=1227
    ),
],
total_token_count=1838
) automatic_function_calling_history=[] parsed=None
```

Visualization

