**Memo**

CHECKSUM: 1537

To: Manager

Cc: Professor Mintz

From: Ilya Gustafson

Date: August 4, 2025

Re: ETL Data Transformation

**Introduction/Background**: This memo provides an overview of what was done to the yearly data to input it into the staging table for later usage. We had three years of product order data (2022, 2023, and 2024) that needed to be combined into one staging table for business analysis. Each year's data came in different formats with different fields, so I had to clean and transform the data to make it consistent across all years.

**Document How Each Year was Transformed**:

* **2022:** The 2022 data was the most straightforward to work with, including all basic information needed. I renamed "Product" to "Product\_Name", "Per-Unit Price" to "unitPrice", and "Order Total" to "orderTotal" for consistency. Since customer tracking wasn't implemented in 2022, I added customer\_id as NULL and tagged all records with year 2022.
* 2**023:** The 2023 data was more challenging due to shipping problems that year, with orders split into two shipments (Quantity\_1 and Quantity\_2 fields). I combined these quantity fields to get total items ordered and calculated the order total by multiplying per-unit price by total quantity. Geographic information wasn't tracked in 2023, so I set Country and State as NULL while using the available Customer\_ID information. All records were tagged with year 2023.
* **2024:** The 2024 data included a new discount policy for orders of 90+ units (10% discount). I calculated the final order total by subtracting the Quantity Discount from the Order Subtotal, used all available geographic information, set customer\_id as NULL since tracking was discontinued, and tagged all records with year 2024.

**Conclusion**: The data from all three years has been successfully transformed and loaded into the staging table with proper calculations for order totals and quantities. The next steps that need to be done include running comprehensive quality assurance tests to verify data accuracy and completeness across all years, creating user documentation for the business team that explains the data structure and any limitations such as missing customer IDs in certain years, and beginning development of the analytical reports and dashboards that will enable management to track product sales trends across different regions and time periods.