**Key Analysis**

**No major problems found, but watch out for:**

* **Super Sales:** About 1% of your sales bring in a lot more money than the others. These "high-flyer" transactions are worth a closer look to understand what makes them so successful.
* **No Losses:** Good news! All your sales are making money, and you don't have any negative or zero revenue.

**Best Sells**

* **Top Products:** Certain chemical products (like ChemProduct\_70, ChemProduct\_96, ChemProduct\_9, and ChemProduct\_29) are your star performers. They have excellent profit margins (up to 59%) and bring in good revenue.
* **Regional Performance:** All regions are doing well with similar profit margins (around 34-35%). South America has slightly lower margins, but still contributes significantly to profit.

**Most Profitable**

* **Best Customers:** The "Tannery" customer type, specifically in the "Beamhouse" product category, is your most profitable segment, with the highest average profit margin.
* **Lowest Profit:** The "Automotive" customer type, in the "Finishing" product category, has the lowest profit margin.
* **Monthly Trends:** October 2024 was your best month for profit margins. However, November 2024 saw a noticeable dip, which could be due to things like special promotions, discounts, or holidays.

1. Model Evaluation Metrics (RMSE, MAE, R2)

After running "Cell 9: Model Evaluation", you will see the RMSE, MAE, and R2 scores for both XGBoost and Linear Regression models for both Quantity and Revenue.

RMSE (Root Mean Squared Error):

Interpretation: This measures the average magnitude of the errors. It's the square root of the average of the squared differences between prediction and actual observation.

What to look for: Lower RMSE values indicate a better fit. Since it's in the same units as the target variable (e.g., units of Quantity, units of Revenue), it's easy to understand the typical error size.

Example: If your sales quantity RMSE is 10, it means, on average, your predictions are off by about 10 units.

MAE (Mean Absolute Error):

Interpretation: This measures the average of the absolute differences between predictions and actual observations.

What to look for: Like RMSE, lower MAE values indicate a better fit. MAE is less sensitive to large errors (outliers) compared to RMSE because it doesn't square the differences. It provides a more straightforward average error magnitude.

R2 (R-squared) Score:

Interpretation: This represents the proportion of the variance in the dependent variable (Quantity or Revenue) that is predictable from the independent variables (features).

What to look for: R2 values range from 0 to 1 (or can be negative for very poor models).

An R2 of 1 indicates that the model explains all the variability of the response data around its mean.

An R2 of 0 indicates that the model explains no variability of the response data around its mean.

Higher R2 values indicate a better fit.

Example: An R2 of 0.85 means 85% of the variability in sales quantity (or revenue) can be explained by your model's features.

Comparing XGBoost and Linear Regression:

Generally, XGBoost is expected to outperform Linear Regression in complex, non-linear datasets, as it's a powerful ensemble method (gradient boosting trees). You'll likely see lower RMSE/MAE and higher R2 for XGBoost, indicating it captures more nuances in the data.

Linear Regression provides a simple, interpretable baseline. While it might not achieve the highest accuracy, its coefficients directly show the linear relationship between features and the target.