

Regional Sales Analysis

Exploratory Data Analysis (2014–2018)

Prepared by Wolfgang Alkier



Problem Statement

- Sales teams lack a clear, data-driven view of performance by region, product, channel, and customer.
- This limits their ability to spot growth opportunities and optimize resources.

Approach

- Loaded all Excel tabs into a structured dataset.
- Cleaned and standardized dates, numbers, and categories.
- Created revenue and profit measures.
- Conducted categorical and numerical analysis.
- Compared 2017 actuals vs. budgets.

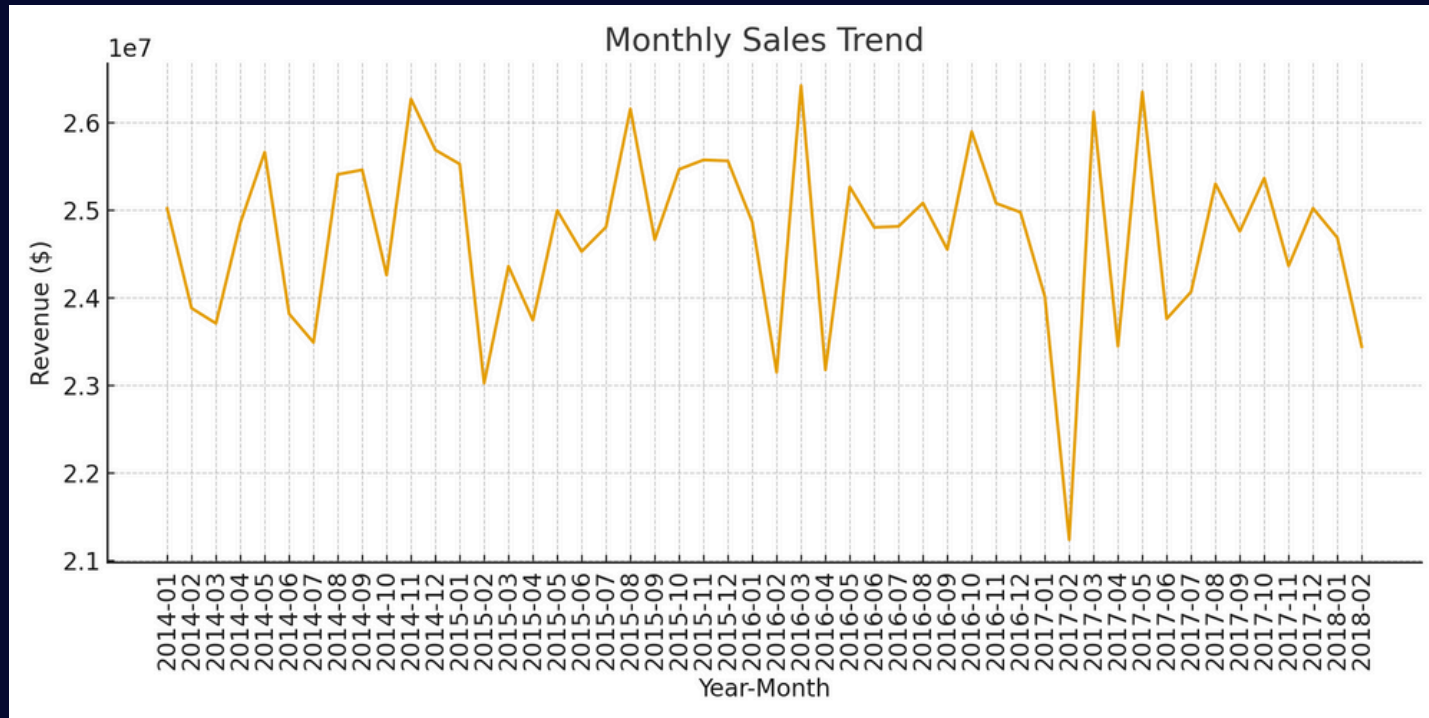
Data Overview

- Dataset: Regional Sales Dataset.xlsx
- Tabs used: Sales Orders (~64k records, 2014–2018), Customers (customer IDs and names), Regions (region codes and names), Products (product IDs, SKUs, names), 2017 Budgets (budgeted revenue by product)
- Key fields in Sales Orders: order_date, channel, warehouse_code, delivery_region_index, product_description_index, order_quantity, unit_price, line_total, total_unit_cost
- Derived metrics:
 - $\text{profit} = \text{line_total} - \text{total_unit_cost}$
 - $\text{margin\%} = \text{profit} / \text{line_total} * 100$ (flat across categories → mentioned only in sidenotes)

Project Workflow

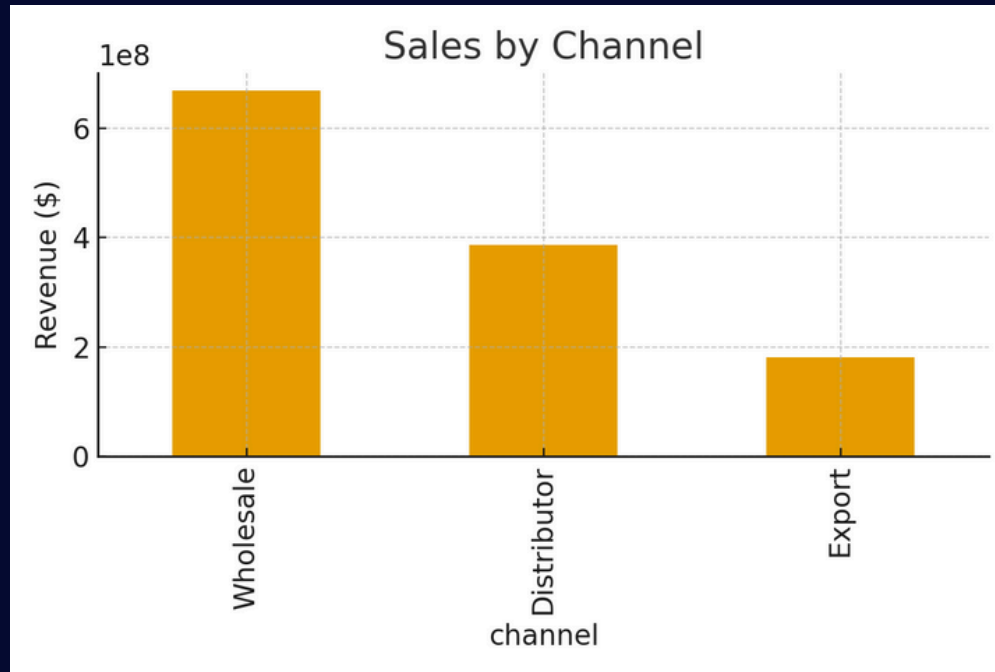
- Data Collection & Cleaning
 - Loaded Excel sheets
 - Standardized dates, numbers, categories
 - Created profit & margin fields
- Exploratory Data Analysis (EDA)
 - Trends over time
 - Sales by channel, region, product, warehouse
 - Customer Pareto analysis
 - Outlier detection
 - Channel × Region × Revenue
 - Budget vs Actual comparison
- Insights Generation & Recommendations
 - Identified concentration risks
 - Highlighted top contributors & gaps
 - Actionable strategies for products, regions, channels, customers, warehouses

Monthly Sales Trend



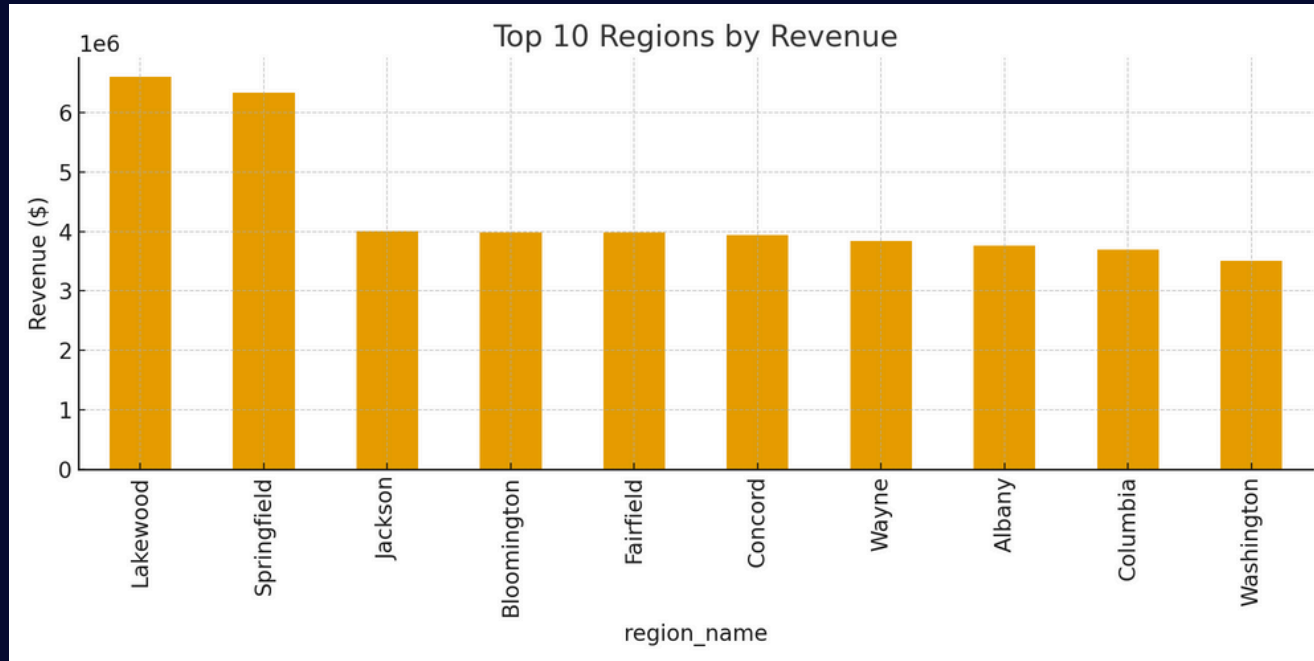
- Monthly revenue is stable, ~\$23–26M per month across 2014–2018.
- Clear seasonal swings: some months spike higher (likely end-of-quarter or holiday demand).
- A few isolated dips → worth checking supply or recording issues.

Sales by Channel



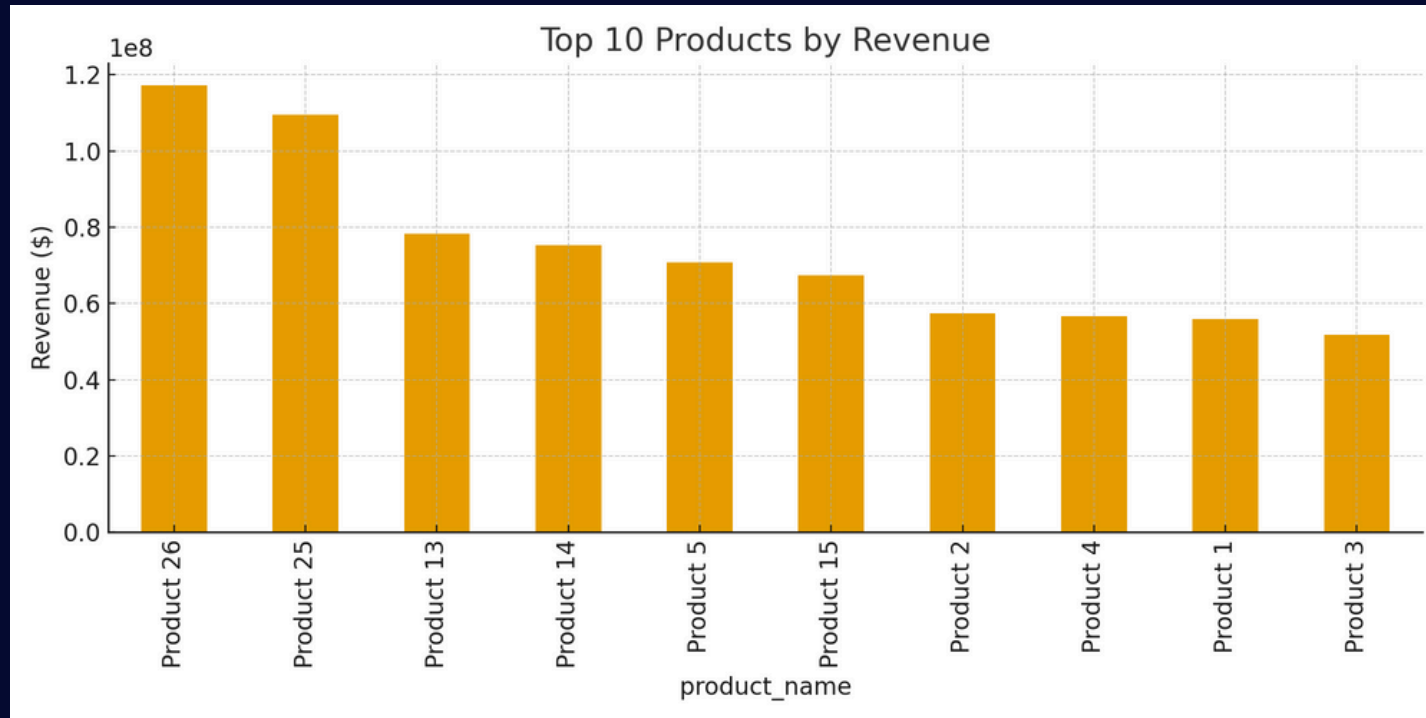
- Wholesale dominates revenue and order volume.
- Distributor and Export are smaller contributors.
- Sidenote: margins are essentially the same across channels, so revenue share tells the full story.

Top 10 Regions by Revenue



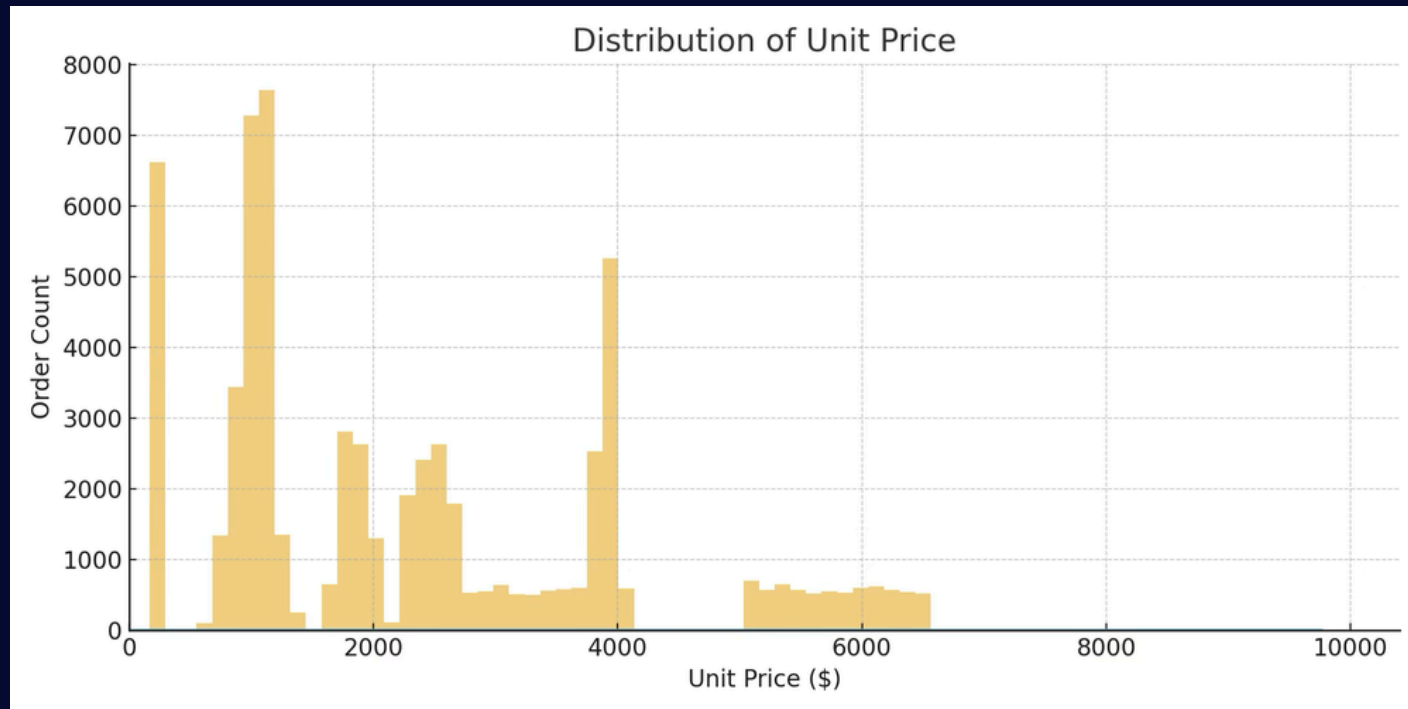
- Two regions clearly stand out with much higher revenue than the rest.
- The majority of smaller regions contribute marginally.
- Again, margin% is identical

Top 10 Products by Revenue



- Strong concentration: a handful of SKUs generate the majority of revenue.
- Long tail of products contributes very little.
- Focus on availability and promotion of top products.

Distribution of unit price



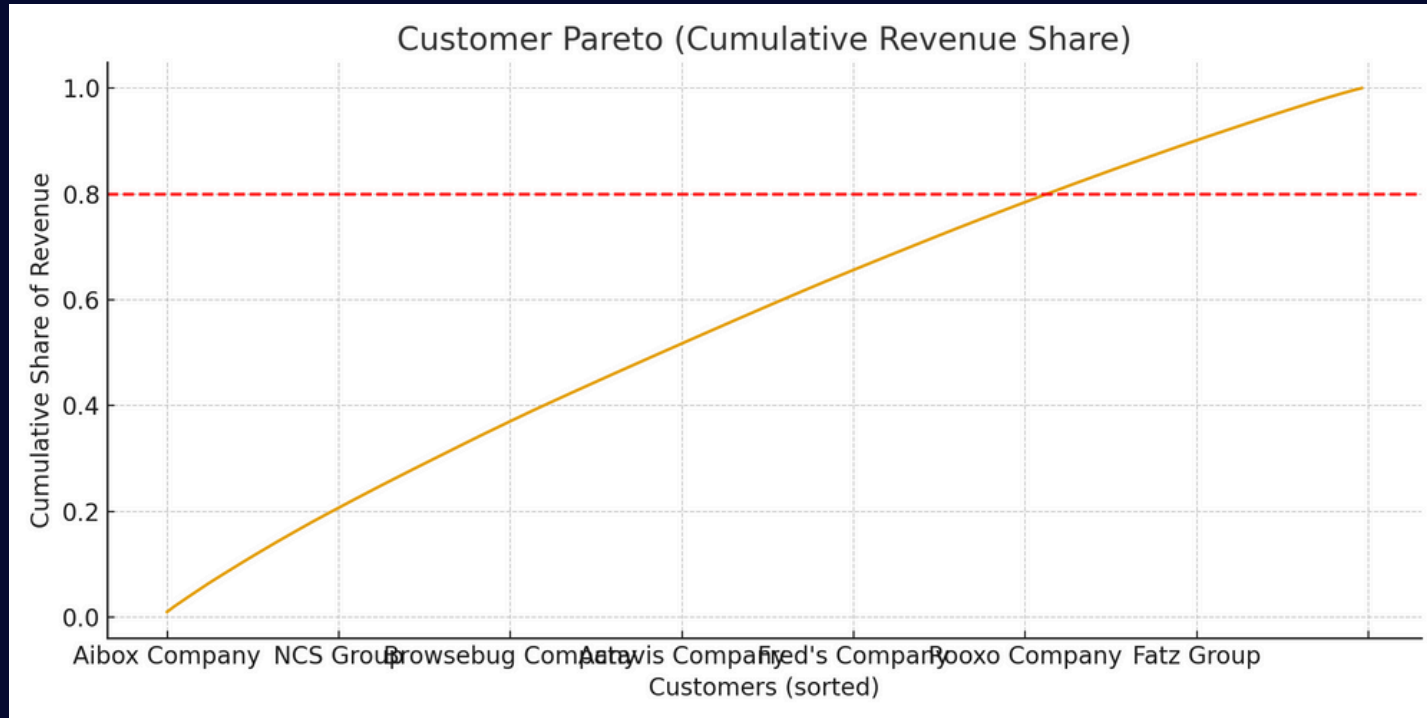
- Lower-priced products are ordered far more frequently, as seen by the tall peaks in the lower unit price ranges.
- Higher-priced products are less common, with only a few orders at very high price levels, indicating niche or special sales.

Sales by Warehouse



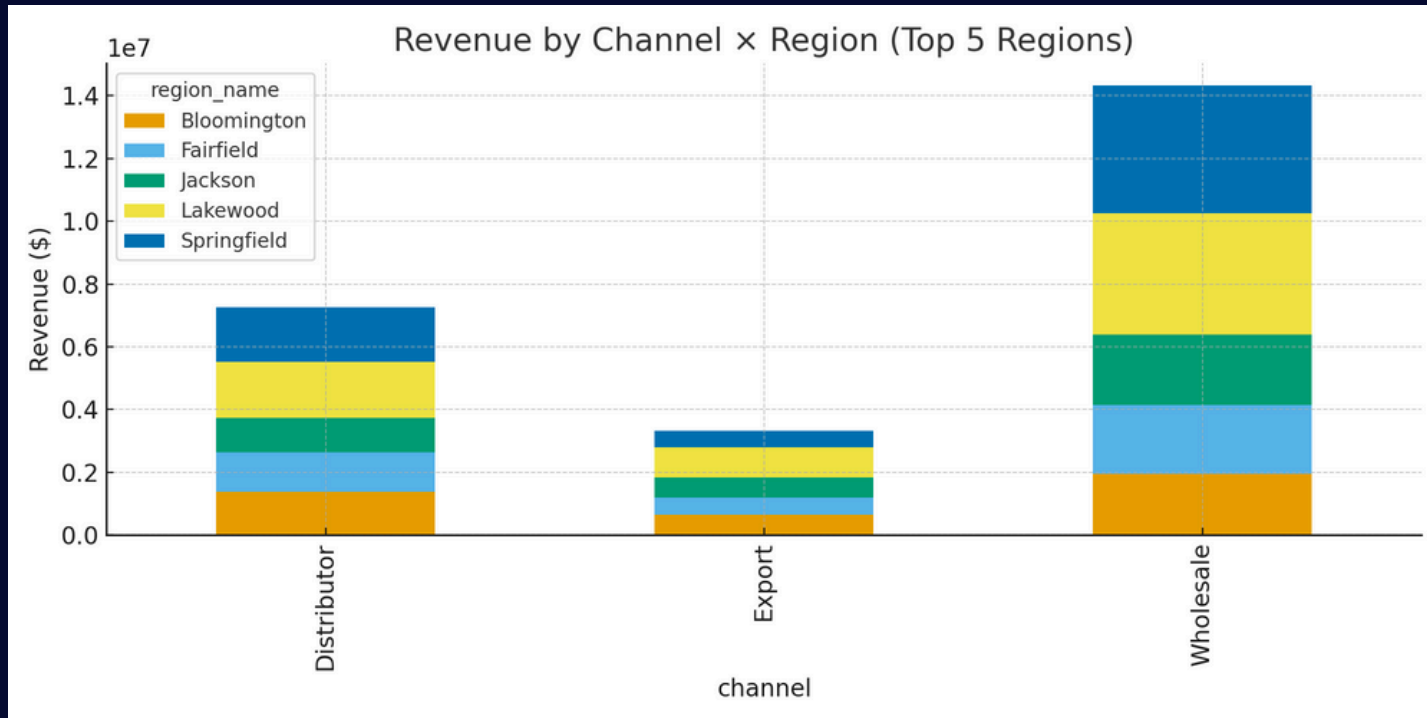
- The business is clearly dependent on 1 WH
- Important for logistics: if a major warehouse fails, business impact is very high.

Customer Pareto



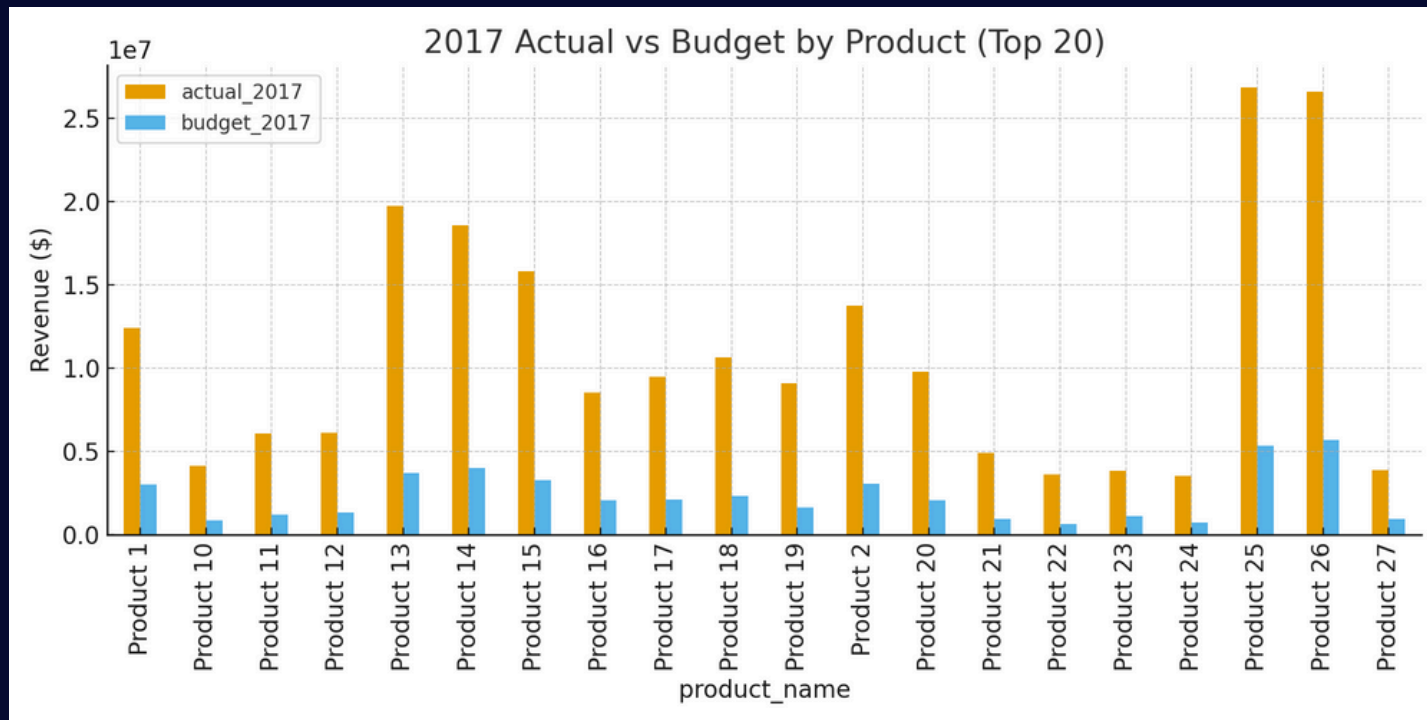
- To reach 80% of sales revenue, you already need almost three-quarters the customer base.
- This is not a “classic” Pareto 80/20 distribution (where 20% of customers generate 80% of sales).
- Instead, your sales are much more evenly spread across customers.

Channel × Region Structure



- Certain channels dominate specific regions.
- Example: Wholesale is dominant everywhere, but Export has relative importance in a few regions.
- Helps align go-to-market strategies with geography.

Actual vs Budget by Product 2017



Most products exceeded budget expectations in 2017 — especially Products 25, 26, 13, and 14, which performed 3–5x above target. However, a few products remained close to or below their planned budgets, signaling areas where demand did not materialize as expected.

Key Insights

1. Concentration risk:

- Channels: Wholesale-heavy.
- Products: few SKUs dominate.
- Regions: only 2 strong regions.
- Warehouses: 1 carries most of the load.

2. Seasonality: predictable monthly swings, but topline is steady year over year.

3. Budgets: actual vs. planned shows which SKUs under/over delivered → immediate feedback loop for planning.

4. Margins: broadly constant, so not a driver of insights (just a sidenote).

Recommendations

- Channel mix: Explore growth options for Distributor/Export to reduce Wholesale dependency.
- Regional focus: Double down on the 2 top regions, but develop strategies to expand mid-tier regions.
- Product strategy: Protect availability of top SKUs; improve marketing of long-tail products.
- Customer strategy: for growth, you can't just rely on a handful of top accounts; you need to activate a broader base.
- Warehouse ops: Ensure resilience of top warehouses; replicate best practices in lower performers.
- Planning: Use 2017 budget vs. actual gaps to improve forecasting accuracy.

Thanks for your attention

Codecademy Data & AI Course, Batch 1
September 2025