Revenue Structure & Concentration Analysis

EXECUTIVE SUMMARY

Javier Gallegos
SQL · Excel · Tableau · GitHub

Evaluates revenue growth dynamics, customer concentration risk, and internal SKU dependency using HHI, effective counts, and Pareto concentration modeling.

Primary structural exposure is **product-level fragility within select categories**, not customer dependency. Customer & category level concentration both score **low risk**. SKU compression in high revenue categories warrants targeted monitoring.

0.034
CUSTOMER HHI
Low concentration — no single customer exceeds ≈9% of total revenue

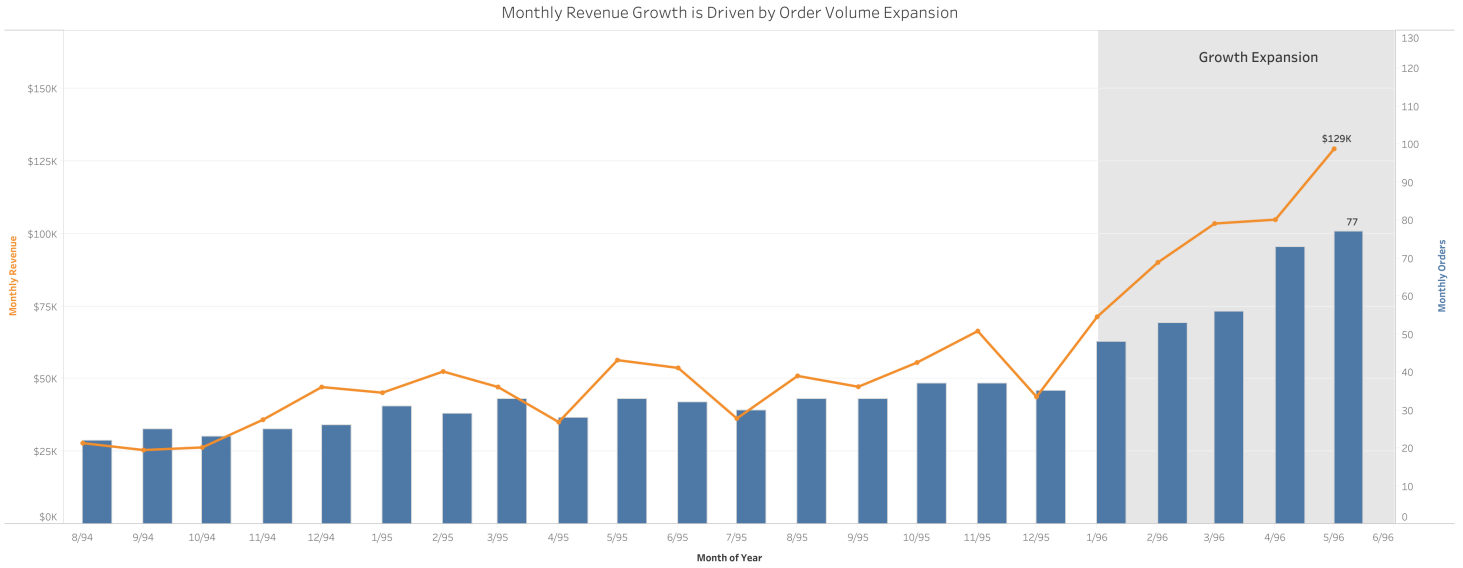
0.142
CATEGORY HHI
Balanced category exposure; top 2 categories contribute ≈ 40% of revenue

3–4
EFFECTIVE SKUs
Driving category 1 revenue despite 12 total SKUs in the category.

INSIGHT 1 · REVENUE GROWTH

Monthly Revenue Growth Driven by Order Volume Expansion

GROWTH ACCELERATING

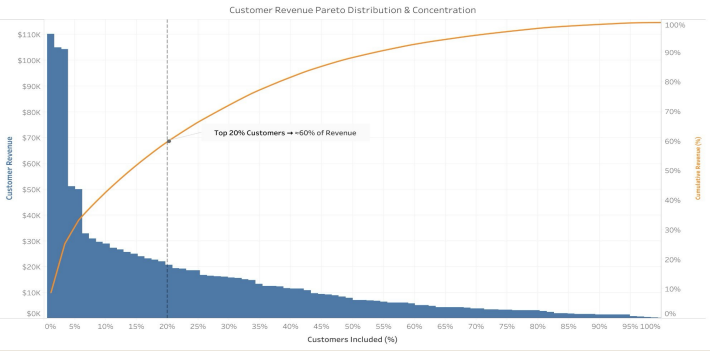


Revenue accelerated sharply in the last five months, with **monthly orders reaching 77 and revenue hitting $129k**. AOV remained stable and new customer acquisition stayed minimal, confirming **growth is primarily volume-driven** rather than price inflation.

INSIGHT 2 · CUSTOMER CONCENTRATION

Pareto Distribution & Concentration

LOW RISK

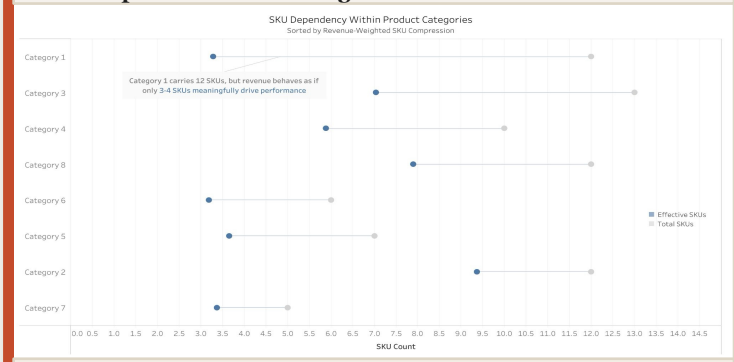


Top 20% of customers generate about **60% of revenue**, indicating a mild Pareto curve. Customer **HHI of 0.034** confirms **low dependency risk** at the customer level.

INSIGHT 3 · SKU DEPENDENCY

SKU Compression Within Categories

MODERATE RISK



Category 1 (**≈21% of revenue**) shows the highest compression gap (**HHI: 0.304**). SKU-level dependency is localized to select categories – **not customer or category-level**.

STRUCTURAL RISK SUMMARY

LOW	Customer Concentration — HHI 0.034
LOW	Category-Level Concentration — HHI 0.142
MODERATE	SKU Dependency within select categories (1, 6, 5)

Primary exposure is **product-level fragility**, not customer or category dependency. Risk is localized at the SKU level within high-revenue category.

STRATEGIC RECOMMENDATIONS

1. Monitor high-compression categories for overreliance on leading SKUs *(Categories 1, 6, 5)*
2. Evaluate SKU rationalization where redundancy exists
3. Protect top-performing SKUs in compressed categories from supply or pricing disruptions
4. Continue customer diversification to sustain the current low-concentration structure

Business Context, Revenue Growth & Concentration Analysis

Business Context & Objective

As revenue has accelerated in the last 5 months, management required a structural assessment of growth sustainability and concentration risk. The objective was to determine how revenue expansion occurred and whether expansion was supported by diversified revenue foundations or dependent on concentrated customer, category, or SKU-level exposure. My approach decomposes revenue drivers using SQL-based growth analysis, Pareto concentration modeling, and Herfindahl-Hirschman Index (HHI) measurement to quantify customer, category, and product-level dependency. The analysis isolates structural fragility points and identifies targeted risk mitigation opportunities.

Revenue Growth Mechanics

Monthly revenue accelerated sharply in early 1996; however, decomposition of the drivers revealed that growth was not acquisition-driven. New customer counts remained minimal during the acceleration period, and average order value (AOV) remained within its historical range.

Instead, revenue expansion was driven primarily by increased purchasing frequency among existing active customers. Orders per active customers rose from an average of 1.28 during the final five months of 1995 to 1.45 during the first five months of 1996, reaching as high as 1.6 in March.

Additionally, 88 of 89 customers in the base are repeat buyers, with the sole one-time customer contributing only $100 in total revenue. Taken together, revenue growth reflects deeper engagement within the existing customer base rather than expansion of the customer footprint.

This engagement-driven dynamic culminated in the strongest revenue expansion in the past 22 months, with monthly revenue reaching $129k and order volume peaking at 77 in the most recent month.

Customer Concentration & Dependency

To evaluate exposure at the account level, revenue concentration was assessed using Pareto segmentation and the Herfindahl–Hirschman Index (HHI).

A Pareto analysis indicates that the top 20% of customers generate ≈60% of total revenue. While this demonstrates natural revenue hierarchy, it does not indicate structural dependency. No individual customer exceeds ≈9% of total revenue share.

Customer-level concentration measured via HHI equals 0.034, which falls well within the low-concentration range. Although the company serves 88 customers, revenue behaves as if distributed across roughly 30 equally sized "effective" customers, indicating moderate hierarchy but no dominant "whale" account.

The long tail of smaller accounts contributes incremental diversification rather than risk. Overall, customer dependency exposure is limited and does not represent the primary structural vulnerability within the revenue base.

Geographic Structure

Revenue is distributed across 21 countries, with the United States (19.4%) and Germany (18.1%) contributing the largest shares. The top three countries collectively account for approximately 48% of revenue, and no single country exceeds 20%.

Intra-country analysis reveals that certain markets depend heavily on anchor accounts. Austria's revenue, for example, is largely driven by a single leading customer, whereas Germany and the United States exhibit more distributed revenue across multiple accounts. This illustrates that macro geographic diversification does not necessarily imply local diversification within each region.

Overall, exposure at the geographic level is moderate but not systemic, with concentration risk occurring at the account-within-country layer rather than at the country level itself.

Product Portfolio Structure & SKU Dependency

To evaluate structural exposure within the product portfolio, revenue concentration was measured at both the category and SKU-within-category levels using the Herfindahl–Hirschman Index (HHI) and effective count translation.

At the company level, revenue is well diversified across product categories (Category HHI = 0.143). No single category exceeds 21% of total revenue share, and the top half (4 of 8 categories) collectively account for 66%. This indicates balanced macro diversification with no excessive category-level dependency.

However, concentration analysis at the SKU-within-category level reveals localized structural compression.

Category 1 (≈21% revenue share), for example, contains 12 SKUs but exhibits an effective SKU count of only 3.29 (Category SKU HHI = 0.304). Although 12 products exist physically, revenue behaves as if driven by only 3–4 economically meaningful SKUs. Similar patterns appear in Categories 5 and 6.

This distinction is critical: while customer-level and category-level concentration risk is low, dependency risk emerges within specific SKU clusters. The portfolio appears diversified at an aggregate level, yet revenue within select categories is disproportionately driven by a small subset of products

In summary, structural exposure is not systemic across categories but concentrated within specific SKU clusters. The primary risk identified in this analysis is product-level dependency within otherwise diversified categories.

Cross-sell & Market Basket Analysis

To evaluate potential growth levers beyond purchasing frequency and structural concentration, product-level co-purchasing behavior was assessed using market basket methodology (support, confidence, and lift).

The strongest observed pairing (Products 21 and 61) exhibited high lift with confidence levels of 0.20 (A→B) and 0.33 (B→A). However, support remained limited at approximately 0.96% of total orders, and the pair contributed only ≈0.43% of total company revenue. Expected co-occurrence counts were near minimum stability thresholds, indicating low economic scale. Across all evaluated combinations, no product pair exceeded 1% of total revenue contribution.

Accordingly, revenue dynamics appear structurally independent across products. Growth is driven by intensified purchasing of individual SKUs rather than portfolio-level bundling effects.

Conclusion of Analytical Section

> Revenue expansion is engagement-driven and supported by strong customer retention. Customer and category level concentration risk is low, with no dominant accounts or systemic regional exposure. However, SKU-level analysis reveals localized dependency within select categories, where revenue is disproportionately driven by a small subset of products. Structural resilience is broad, but product-level fragility represents the primary exposure.

Technical Architecture & Analytical Framework

Data Structure & SQL Design

The analysis was built using a layered SQL structure designed to standardize revenue calculations and reduce duplication across queries. Foundational base views were created by joining transactional tables — *Orders*, *Order Details*, *Customers*, *Products*, and *Categories* — into unified revenue layers. These layers served as the validated source of truth for all downstream analysis. Three primary revenue base layers were constructed:

- **Line-item revenue layer (order-line grain):**
  Revenue was calculated at the OrderID & ProductID level by joining Orders to Order Details, then joining Customer, Product, and Vendors tables. This established the most granular economic unit of analysis.
- **Order and customer revenue layers:**
  Line-item revenue was aggregated to the OrderID level and subsequently to the CustomerID level, creating standardized order-level and customer-level revenue bases used across concentration, growth, and dependency analysis.

Rather than writing isolated queries, the analysis was structured into layered analytical components. Each analytical layer was built on top of the revenue foundations, with intermediate metrics — such as revenue share, ranking, and cumulative distributions — defined once and applied consistently across concentration and dependency analyses.

Window functions were used to compute revenue shares, running totals for Pareto segmentation, and partitioned rankings across customers, countries, and product categories, ensuring consistent measurement across all hierarchy levels.

Concentration Measurement Methodology

Revenue concentration was evaluated using the **Herfindahl–Hirschman Index (HHI)**, a widely used concentration metric that measures how revenue is distributed across entities such as customers, categories, or SKUs. HHI is calculated as the sum of squared revenue shares, producing higher values when revenue is concentrated among a small number of entities and lower values when revenue is broadly distributed. Because raw HHI values can be difficult to interpret intuitively, results were translated into **effective counts** using: **Effective Count = 1 / HHI.**

This conversion expresses concentration as the number of equally sized entities that would generate the same level of concentration. Applying this framework consistently across customers, product categories, and SKUs allows concentration risk to be evaluated at multiple hierarchical levels. This makes it possible to distinguish between macro diversification (e.g., revenue distributed across categories) and localized dependency within specific customer or product clusters.

While revenue appears well diversified at the macro level — with category concentration remaining low — deeper analysis reveals localized compression within certain categories. For example, **Category 1 contains 12 SKUs but exhibits an HHI of 0.304**, which translates to an effective SKU count of approximately **3.29 SKUs**. Although twelve products exist within the category, revenue behaves economically as if only **three to four SKUs** drive majority of the category performance. Similar compression patterns appear in **Categories 5 and 6**, though to a lesser degree.

This pattern indicates that the primary exposure within the portfolio is **SKU-level dependency within otherwise diversified categories**, rather than concentration at the customer or category level.

Market Basket & Cross-Sell Framework

Market basket analysis was implemented using **self-joins on the line-item revenue layer** to generate product pairs purchased within the same transaction. Line items were joined on OrderID, producing all product combinations appearing within an order. For each pair, three association metrics were calculated:

- **Support**: share of orders containing the pair
- **Confidence**: conditional probability of purchasing one product given another
- **Lift**: ratio of observed co-occurrence to expected frequency under independence

Lift values greater than one indicate that products appear together more frequently than expected under independence. To ensure practical relevance, pair-level revenue contribution and expected counts were also evaluated. The most notable pair (**Product 21 & 61**) exhibited **0.96% support**, confidence of **0.20 / 0.33**, and lift above one, but contributed only **0.43% of total revenue**, indicating limited structural impact. As a result, the analysis did not identify statistically and economically meaningful product associations or emergent co-purchasing behavior.

Assumptions & Limitations

The analysis assumes that historical purchasing patterns are representative of underlying customer behavior and that the available transactional dataset captures the full set of completed orders during the observation period. Revenue calculations are based on recorded order line values and do not account for returns, cancellations, or external market factors that may influence demand.

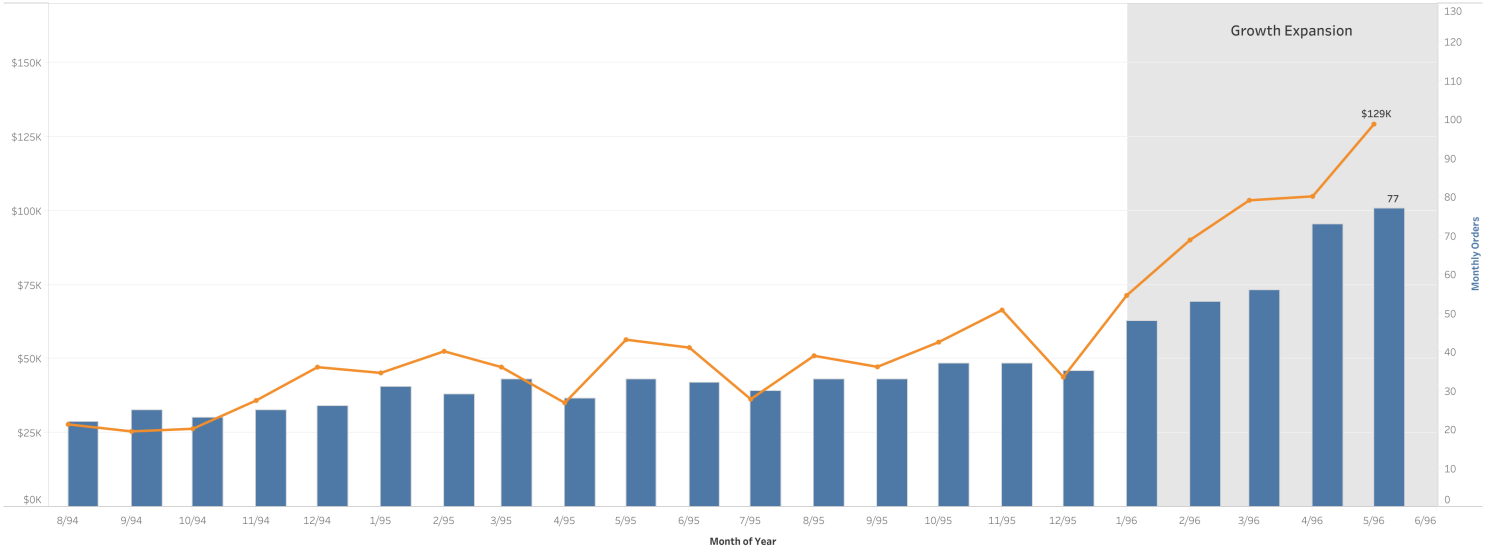
Market basket analysis evaluates co-purchasing behavior within the same transaction and therefore does not capture sequential purchasing relationships occurring across separate orders. Additionally, concentration metrics such as HHI reflect revenue distribution within the observed dataset and may change if the product catalog or customer base expands over time.

Resources

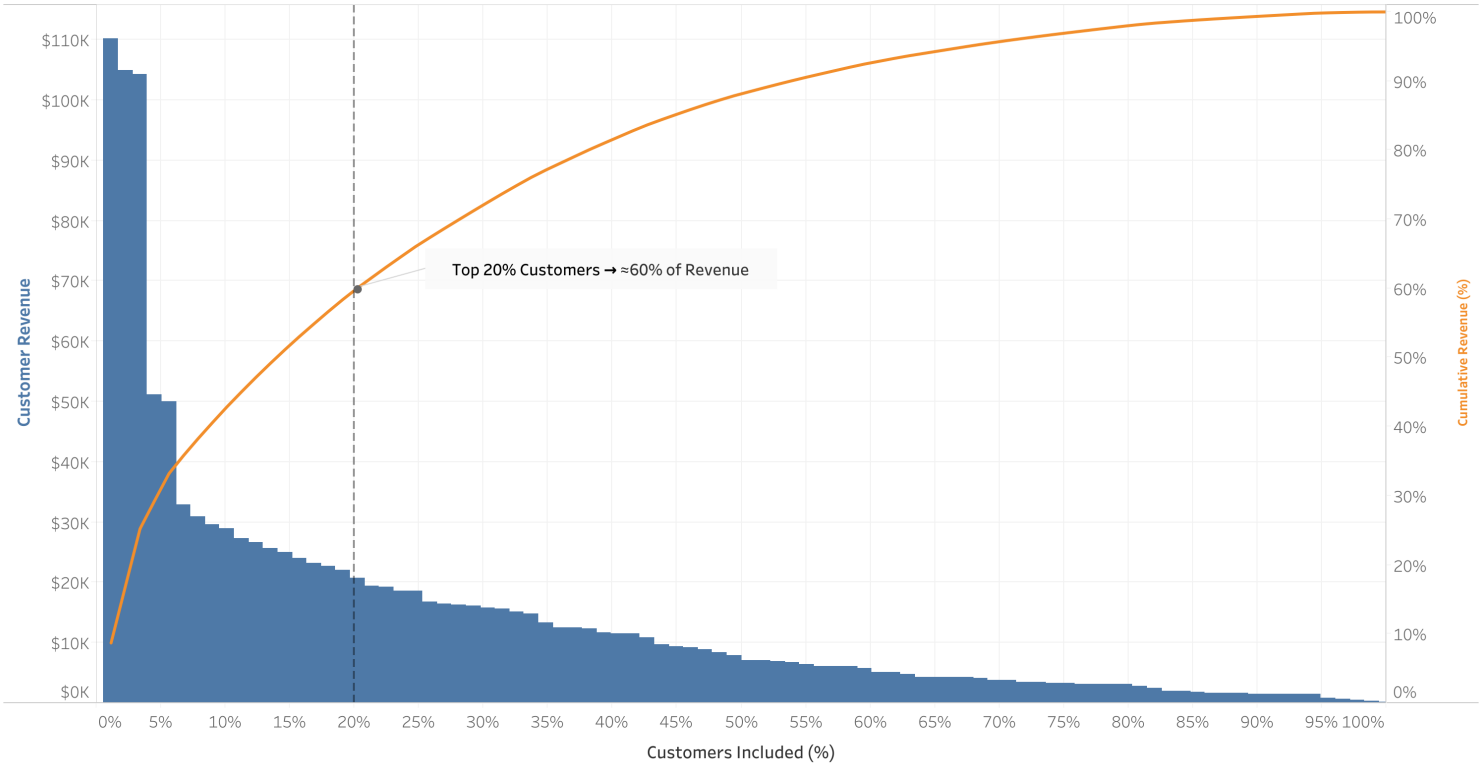
This analysis was conducted using the Northwind transactional dataset, with all computations implemented in SQL. Concentration and dependency metrics were evaluated using the Herfindahl–Hirschman Index (HHI) framework, while product association analysis followed standard market basket methodology using support, confidence, and lift.

- **Project landing page:** https://jdesk99.github.io/SQL-Revenue-Concentration-Analysis-git/
- **All SQL queries, SQL query outputs, the SQLite database file, and supporting Excel file used in this analysis are available at:** https://github.com/jdesk99/SQL-Revenue-Concentration-Analysis-git

Monthly Revenue Growth is Driven by Order Volume Expansion



Customer Revenue Pareto Distribution & Concentration



SKU Dependency Within Product Categories

Sorted by Revenue-Weighted SKU Compression