

Hey,

I took a deep dive into our data, and here are some key findings:

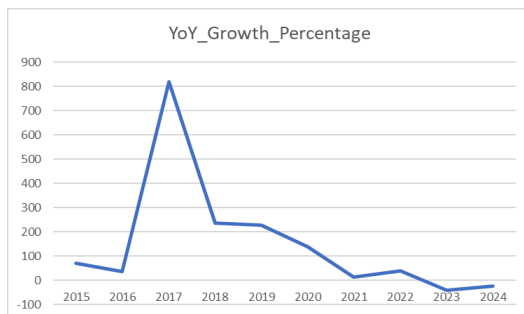
! Major Data Quality Issues:

1. **Missing Data:** **50%** of transactions are missing a product match, and **99%** can't be linked to a user. Implications include:
 - Inaccurate reporting & business insights on product performance and user behavior
 - Operational challenges (e.g., customer service, compliance, and audits)
 - Inaccurate predictive models, invalid dashboards and broken ETLs
 - Potential revenue loss & missed opportunities for marketing or targeted offers
2. **Products Data:** **26%** of the products have missing **manufacturer and brand**.
3. **Inconsistent Barcodes:** Some barcodes have unusual lengths or appear duplicated across different products, potentially leading to incorrect mapping, reporting and analysis



Interesting Trend:

1. **YoY user growth** - Peaked at **820% in 2017**, slowed to **236%-138% (2018-2021)**, continued to drop since with the past two years demonstrating **negative YoY user growth**. Suggesting a significant early-stage adoption, followed by stabilization and slower growth.
2. YoY averages at **151.54%**, but it has been highly fluctuating. See below:



An alternative approach would be to calculate **CAGR** that is designed to smooth out fluctuations.

? Questions & Next Steps:

1. **Data issues are a high priority** since they impact data quality and prevent us from relying on our data. Can you advise who would be the right person to take it with?
2. **YoY user growth:**
 - a. Does the YoY align with business plans and priorities?

- b. Do we have any insights into the factors behind the 2017 user surge (marketing, product, external trends)? Could we replicate it?
- c. If valuable, we can analyze breakdowns by users, marketing sources ect. And further explore CAGR as a key metric to monitor.

Would love your input on this! I'll set up a meeting to dive deeper.