Slack Message / Email to Stakeholders

Subject: Data Quality Insights & Key Business Findings from Analysis

Greetings,

I've completed an in-depth analysis of our users, transactions, and products datasets, identifying key data quality issues, outstanding questions, interesting trends, business insights, and next steps that can inform future decisions. Below is a summary of my findings:

1. **Data Quality Issues & Outstanding Questions**
   a. **Missing & Inconsistent Data:**
      i. **Users Dataset:** BIRTH_DATE, STATE, LANGUAGE, and GENDER have significant missing values, affecting demographic-based insights.
      ii. **Transactions Dataset:** BARCODE is missing in **5,762 (~11.5%)** records, and FINAL_SALE is missing in **12,500 (25%)** of transactions.
      iii. **Products Dataset: 226K+** missing BRAND/MANUFACTURER values and some barcodes linked to multiple manufacturers/brands, leading to potential misclassification.
   b. **Duplicate & Anomalous Transactions:**
      i. **171** duplicate transactions removed.
      ii. Cases where FINAL_QUANTITY = 0 but FINAL_SALE has a value (possible refunds, weight-based pricing, or data entry errors).
   c. **String & Categorical Data Issues:**
      i. GENDER values are inconsistent (non_binary vs. non-binary), and STATE names require validation.
2. **Outstanding Questions:**
   a. Should missing BARCODE values be manually mapped, or should these transactions be excluded?
   b. Do we treat FINAL_QUANTITY = 0 but FINAL_SALE > 0 as valid, refunds, or do we flag them as potential data issues?
   c. Can we validate whether some barcodes are correctly assigned to multiple brands/manufacturers, or do we need to standardize them?
   d. Is FINAL_QUANTITY expected to be a whole number, or does it represent weight/volume for some products?
3. **Interesting Trends: Fetch's YoY Growth / Decline**
   a. Fetch saw rapid growth from **2017 to 2020**, peaking at **16,886** new users in **2020 (+137.97% YoY)**.
   b. Growth slowed in **2021 (+13.52%)** and rebounded in **2022 (+39.87%)** but declined sharply in **2023 (-42.37%)** and **2024 (-24.79%)**.
   c. **Possible Causes:**
      i. Market saturation & increased competition.
      ii. Decreased user acquisition efforts.
      iii. Changes in consumer shopping behavior post-pandemic.

     d. **Potential Actions:**
          i. Analyze acquisition channels & optimize marketing spend.
          ii. Improve user retention efforts (e.g., personalized incentives, loyalty rewards).
          iii. Compare Fetch's decline to industry trends to determine if this is platform-specific or market-wide.

4. **Business Insight: Power Users & Brand Engagement**
   a. Tostitos is the leading brand in the Dips & Salsa category, ranking #1 in both sales **($260.99)** and purchase frequency **(72 transactions)**.
   b. CVS leads in total sales among long-term users **($72)**, showing strong engagement with pharmacy/household essentials.
   c. Dove consistently ranks among top-scanned and top-purchased brands, reinforcing high engagement in personal care.
   d. **Potential Actions:**
      i. Target Tostitos, CVS, and Dove for exclusive Fetch promotions.
      ii. Use brand engagement insights to develop category-based rewards.
      iii. Create targeted loyalty campaigns for high-frequency & high-value brands.

5. **Next Steps & Request for Input** - to improve data reliability and optimize future insights, I'd appreciate input on the following:
   a. **Data Quality Adjustments:** Should we manually map missing barcodes or exclude these transactions?
   b. **Anomaly Validation:** How do we handle cases where FINAL_QUANTITY = 0 but FINAL_SALE > 0?
   c. **User Growth Strategy:** Do we have insights on how competitor growth compares to Fetch's recent decline?

Would love your feedback on these points. Let's discuss how we can refine our strategy based on these findings.

Best,
Prachi Yadav
Senior Data Analyst