TAKE HOME ASSESSMENT

PRACHI YADAV

About Myself

A Data Analyst who loves to **SELECT** insights, **JOIN** ideas, and **GROUP** BY logic!

- I specialize in SQL, Python, and data visualization, turning raw data into actionable insights.
- I believe in clean queries and clean data—no NULL insights here!
- My approach? Ask the right questions, optimize for efficiency, and always keep learning.
- Today, I'll walk you through my take-home challenge—the logic, the insights, and my decision-making process.

Let's 'EXECUTE' this presentation!

Challenge Overview

Goal: Analyze transaction, user, and product data to derive insights.

Dataset: Provided datasets included:

- Users (demographics)
- Transactions (purchase data)
- Products (categories & brands)

Key Objectives:

- Identify data trends and anomalies.
- Perform exploratory analysis and generate actionable insights.
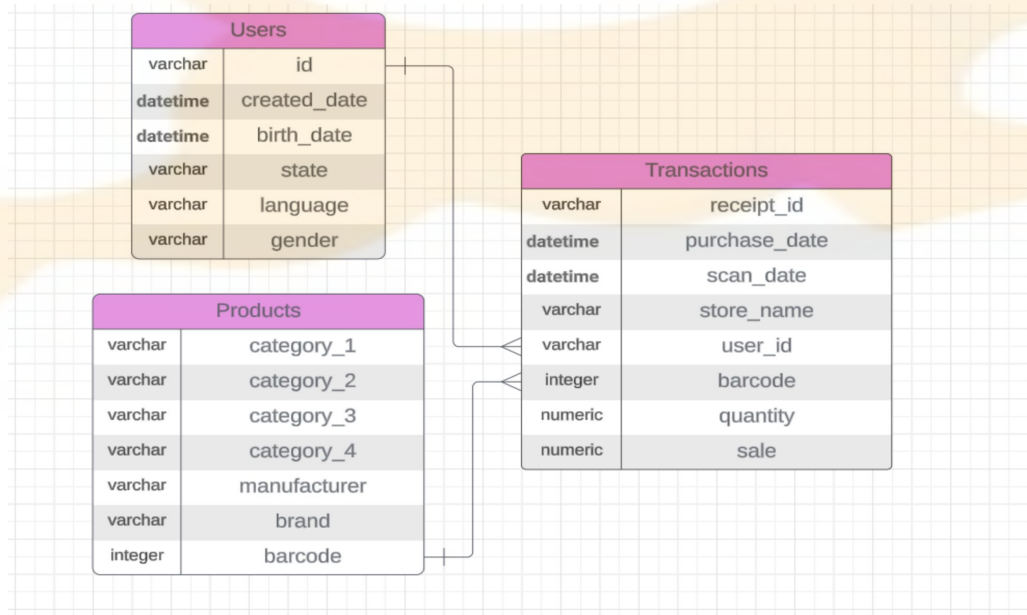- Address data quality issues and make business recommendations.

Data Overview

Users Data → Contains customer demographics (age, state, gender, language, etc.)
Transactions Data → Records purchase details (receipt ID, store, barcode, quantity, sale amount, etc.)
Products Data → Links barcodes to product categories, brands, and manufacturers.

ERD:

Understanding Fetch's Business Context

Fetch's Data Model: Uses receipt transactions to analyze brand performance & user behavior.

Key Business Considerations:

- How do users interact with Fetch? (Purchases, receipt scans)
- Which brands and categories perform best?
- What anomalies or gaps exist in the data?

My Approach:

- Understand business questions before diving into the data.
- Identify missing and inconsistent data.
- Extract meaningful insights for decision-making.

Data Exploration & Cleaning

Data Issues Identified:

- **Users Data:** Missing values in birthdate, state, language, gender.
- **Transactions Data:** 11.5% missing barcodes, 25% missing final sales.
- **Products Data:** 226K records missing brand or manufacturer info.

Cleaning Approach:

- Standardized data types (dates, categories).
- Removed duplicates.
- Retained missing values rather than making assumptions.

Key Assumptions in Data Cleaning & Analysis

- Retained missing data unless necessary to remove
- Did not impute categorical data (state, gender) due to lack of strong correlations
- Handled missing sales values by keeping them rather than assuming a value
- Defined power users based on transaction activity, not Fetch points
- Used `created_date` as a proxy for Fetch's user growth
- Assumed that decimal values in `quantity` were valid

Key Insights from Analysis

1) **Top 5 Brands by Receipts Scanned (Users Aged 21+)**

   <u>**Brands Identified**</u>: Nerds Candy, Dove, Hershey's, Coca-Cola, Great Value

   <u>**Insights:**</u>

   - High frequency of scans indicates **everyday purchases** like snacks, personal care, and beverages.
   - Strong **brand familiarity** and **repeat purchase behavior** among users.

Business Impact:

   Opportunity for Loyalty Campaigns:

   - Launch **bonus point offers** for frequent scans.
   - Create **brand-specific challenges** to boost engagement.
   - Strengthen **partnerships with high-engagement brands**.

2)Top 5 Brands by Sales (Users with Older Accounts)

Brands Identified: CVS, Trident, Dove, Coors Light, Quaker

Insights:

- High sales driven by **long-term users**, indicating strong **customer loyalty**.
- Mix of essentials and consumables shows **consistent, high-value purchases**.

Business Impact:
Retail-Specific Incentives:

- Offer **exclusive rewards** for repeat purchases with top brands.
- Create **multi-brand campaigns** targeting loyal users.
- Leverage data to strengthen **retail partnerships**.

3) Sales by Generation (Health & Wellness Category)

Key Findings:

- **Generation X dominates sales** in the Health & Wellness category.
- **Minimal engagement from Gen Z and the Silent Generation.**

Insights:

- Gen X likely driven by **fitness, supplements, and self-care priorities**.
- Low Gen Z engagement could indicate a **data gap** or **market opportunity**.

Business Impact:
   Personalized Wellness Offers:

- Target Gen X with **custom rewards** for wellness products.
- Explore **targeted campaigns for Gen Z**, using trends like eco-friendly or mental wellness products.
- Investigate **potential data gaps** to ensure accurate insights

Business Trends & Open Questions

User Growth Trend:

- **Growth peaked in 2020 (+138%) but declined in 2023 (-42%).**
- Potential causes: **Market saturation, reduced marketing efforts, increased competition.**

Unresolved Questions:

- Should we **manually fill missing barcodes** or exclude transactions?
- If `quantity = 0` but `sale > 0`, is it a refund or a system issue?
- Should each product have **one barcode per brand**, or is duplication valid?

SQL Approach & Techniques Used

Key SQL Techniques Used

Common Table Expressions (CTEs) – Used for modular queries and better readability.
INNER JOIN – Combined `users`, `transactions`, and `products` to analyze brand purchases.
Aggregation Functions – Used `GROUP BY`, `COUNT(*)`, and `SUM(FINAL_SALE)` for insights.

Why?

- Improved query structure and readability.
- Ensured data consistency and accurate insights

Advanced SQL & Handling Edge Cases

Window Functions – Used `RANK()`, `DENSE_RANK()` for brand ranking, `LAG()` for YoY growth.
CASE WHEN Handling – Managed missing barcodes and inconsistent sales data.
Query Optimization – Can use indexed columns to improve performance and CTEs instead of subqueries.

Why?

- Optimized for large datasets.
- Improved performance and trend analysis.

Recommendations & Next Steps

Business Recommendations:

- **Leverage top brands** (Nerds, Dove) for targeted campaigns.
- **Create loyalty incentives** for **power users & frequent buyers.**
- **Investigate missing sales data** before making pricing decisions.
- **Address growth slowdown** by exploring new user acquisition strategies.

Next Steps:

- Validate barcode data inconsistencies.
- Test different marketing strategies for engagement.
- Reassess Fetch's **user acquisition model post-2022.**

Questions?

Should I 'COMMIT' and call it a day,

or

'ROLLBACK' and try again?