By Yiwei Yang

Fetch Rewards Data Analysis

Agenda

1. **Key Data Quality Issues & Outstanding Questions**
2. **Data trends & Interesting Business Insights(Power users analysis)**
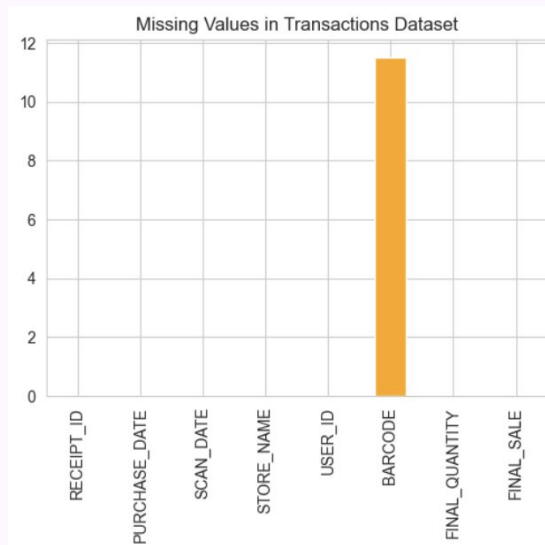3. **Recommendations & Next Steps**

Data Issues: Missing Data
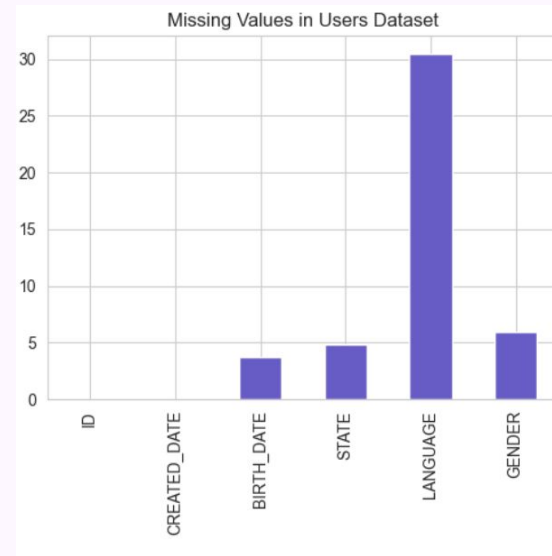
Percentage of data missing in three datasets

Products dataset

Missing Values in Products Dataset

Transactions dataset

Missing Values in Transactions Dataset

Users dataset

Missing Values in Users Dataset

Other Data Issues

Duplicate records

- **171** duplicate transactions that may distort sales and user behavior insights.
- **215** duplicate products that can create inconsistencies in product-level reporting.

Table referencing

- Nearly **50K** transactions reference non-existent users, meaning that we have missing or incorrect user data.
- **19K** transactions reference non-existent products, there are possible barcode mismatches.

Data Types

- Many columns are stored as object, even though they should be numeric or datetime. This can cause calculation errors, slow processing, and incorrect filtering.

Outstanding Questions

Data cleaning & Validation

- Are we going to delete all missing values or find other ways such as filling with means/backward filling?
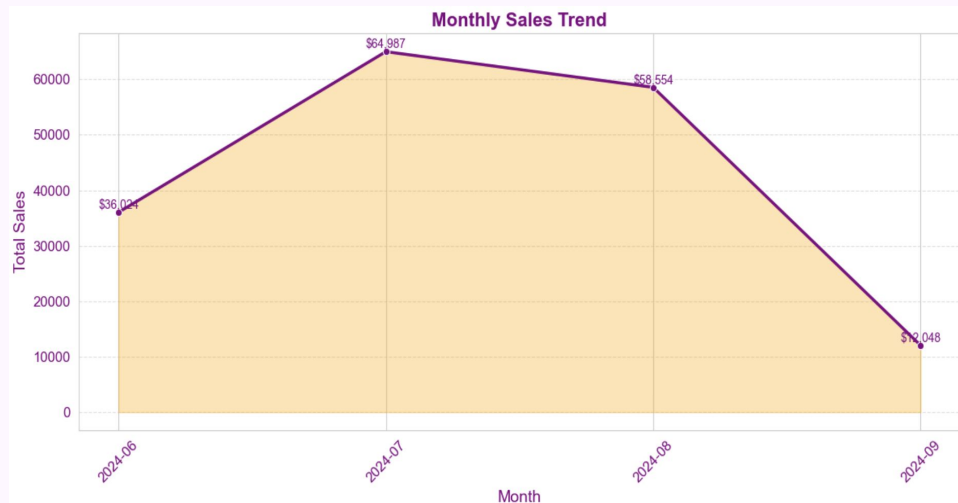
Feature usage

- Is **categorie_4** in products table actively used? We missed over **80%** of the data

Referencing problem

- Should transactions with **missing references** be excluded from key analyses, or is there a plan to correct them?

Data Trends

Monthly Sales Trend

- Sales **peaked in July** at **$64,987**, followed by a slight decline in **August ($58,554)**.
- **Sharp drop** in **September**, down to **$12,048**, suggesting potential **seasonal or operational impacts**.
- **Insight:** There's an opportunity to **investigate factors behind the September dip** and develop strategies to sustain mid-year performance.

Total Sales by Product Category

Category Growth Over Time
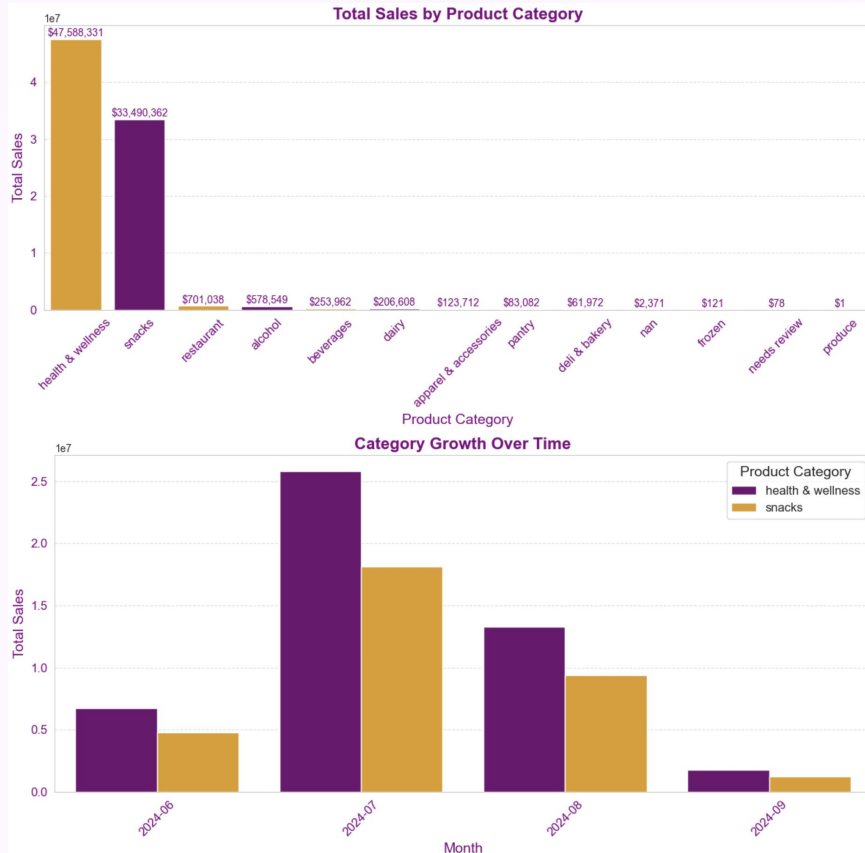
1. **Top Categories by Sales:**
   - **Health & Wellness** leads with **$47,588,331** in total sales.
   - **Snacks** follow closely with **$33,490,362** in total sales.
2. **Emerging Categories:**
   - **Restaurants** and **Alcohol** categories show potential, with **$701,038** and **$578,549** in sales respectively.
3. **Underperforming Categories:**
   - **Frozen**, **Produce**, and categories marked **'Needs Review'** show minimal sales, indicating potential data issues or limited market reach.
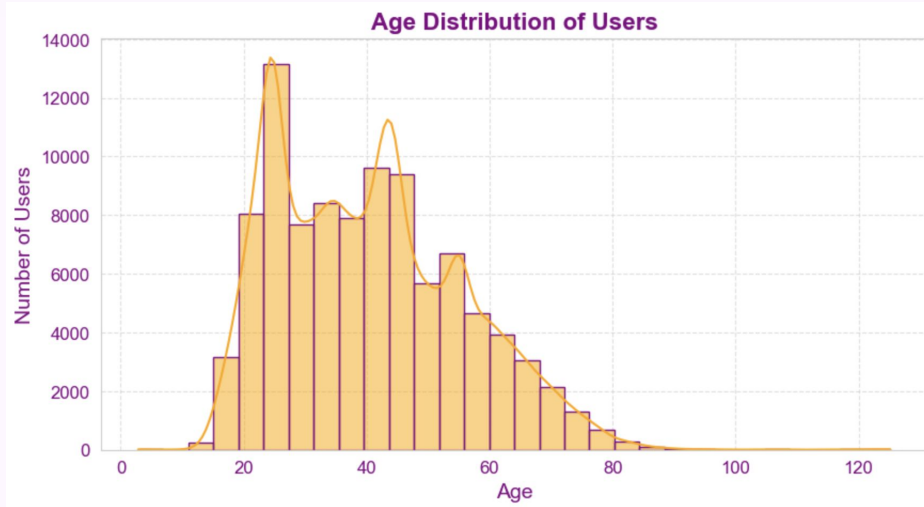
Age Distribution of Users:

- Majority of users are between **25 and 45 years old**, peaking at **25 and 40**.
- **Older demographic (60+)** shows lower engagement, indicating **opportunities for targeted marketing**.

1. **Top States by Transactions:**
   - **Florida (FL)** leads in transactions, followed by **New York (NY)** and **Pennsylvania (PA)**.
2. **Regional Patterns:**
   - States like **California (CA)** and **Texas (TX)**, while expected to perform high, rank lower than eastern states like **FL** and **PA**.
3. **Opportunity Areas:**
   - **Midwest and Southern states** show lower transaction counts, suggesting **potential for targeted marketing efforts**.

Number of Transactions by State
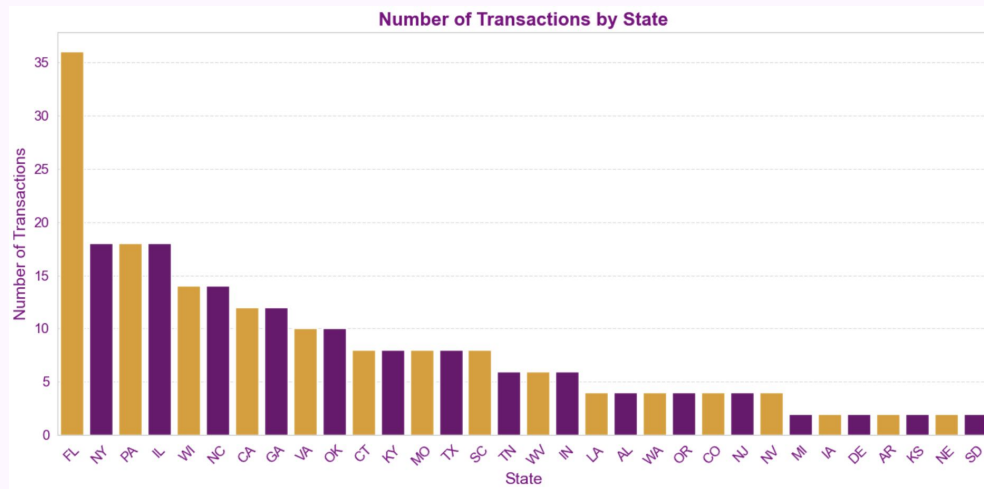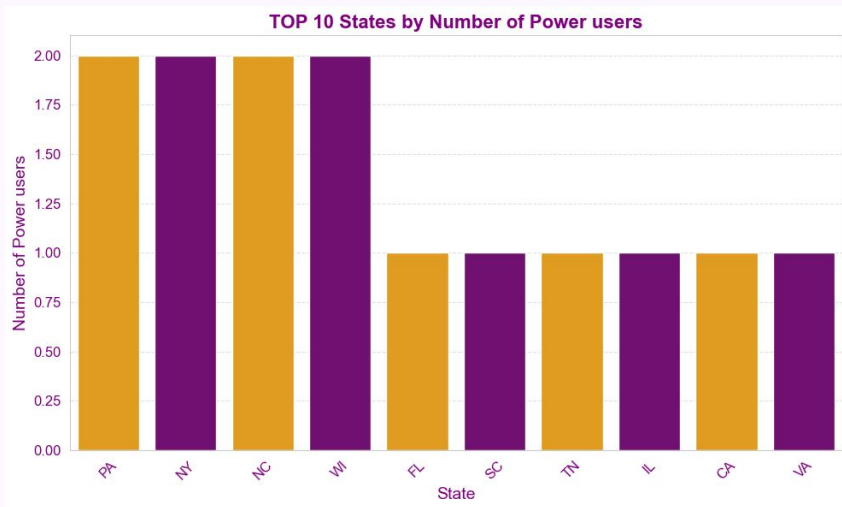
Power Users Analysis

Who are power users?

Users in the top **20% of spenders** who have made transactions on **more than 1 day** in 4 months.

High revenue generator +Engagement

- **2.08%** of total users

- Generate **32.35%** of total sales

Power Users Analysis

TOP 10 States by Number of Power users

Geographic Distribution:

- **Pennsylvania (PA)**, **New York (NY)**, **North Carolina (NC), Wisconsin(WI)** lead in the number of power users.
- Compare to the total users geography distribution, **Florida (FL) and Illinois(IL)** have fewer power users.

Power Users Analysis

Distribution of Days having transactions of Power users
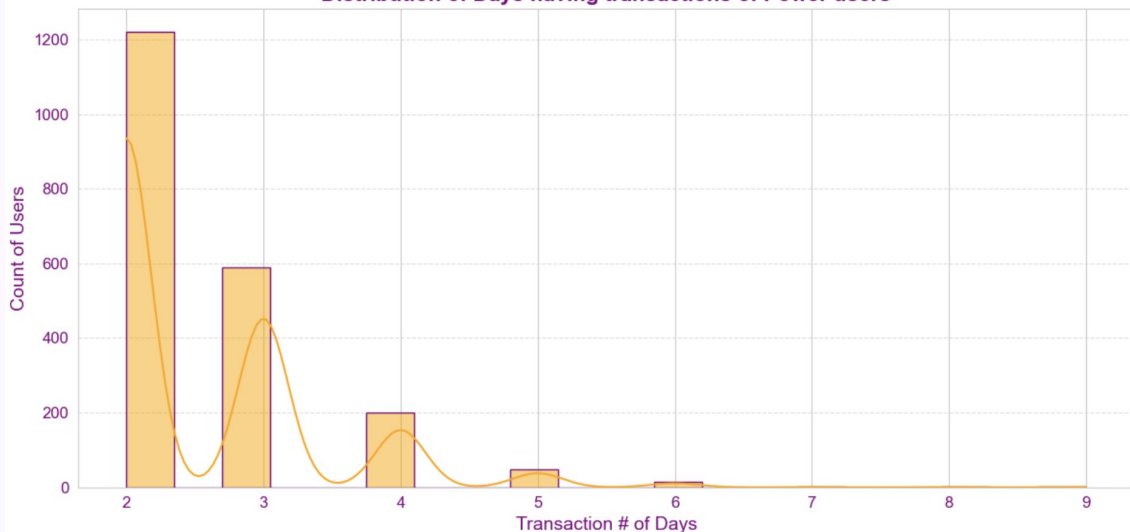
Transaction Frequency:

- While power users contribute **32.35%** of total revenue, **the majority have transactions on only 2 days**.
- This suggests that even our most valuable users are not consistently engaged over time.

Power Users Analysis

Opportunities for Improvement:

- **Loyalty Programs:** Introduce personalized rewards or loyalty tiers to encourage frequent engagement.
- **Behavioral Triggers:** Use targeted notifications or reminders based on purchase patterns to drive more consistent usage.
- **In-App Features:** Gamify the app experience or offer exclusive perks to motivate daily or weekly engagement.

Recommendations & Next steps

Address Data Quality Issues

Duplicate transactions and correct **invalid user/product references**.

Leverage Power Users

Focus on **retention strategies** for power users in **high-performing states**.

Launch **targeted promotions** to boost engagement

Data Enrichment

Collect more complete **user demographic data** for better segmentation

Expand High-Performing Categories:

Invest in the **Health & Wellness** and **Snacks** categories to sustain growth.

Explore marketing opportunities for **emerging categories** like **Restaurants** and **Alcohol**.

Data Validation

Validate and clean **transaction data**.

Fill in **missing product information**.

Targeted marketing for older demographics & Midwest and Southern states