Assumptions and Key Findings

Assumptions Made During Data Cleaning

Transactions Table:

- Duplicate transactions were removed, assuming they were mistakenly repeated rather than purchased twice by the user.

- Transactions with FINAL_SALE or FINAL_QUANTITY values of 0 were deleted. This was based on the observation that other records with the same receipt_id and barcode had correct FINAL_SALE and FINAL_QUANTITY values, suggesting these 0 values resulted from data entry errors.

Products Table:

- Two brands shared the same barcode. Assuming the first entry was correct, the second entry was removed to maintain data consistency.

Key Insights from the Data

- Sales Distribution: FINAL_SALE values range from 0 to 462, with the majority of transactions falling between 0 and 100.

- Top Retailers by Sales Amount: Walmart leads in sales volume, followed by Costco and Sam's Club.

- Purchase Frequency: Users frequent Walmart most often, followed by Dollar General and ALDI.

- Demographics:

    o Gender: There are more registered female users than male users on Fetch.

    o Location: Fetch users are predominantly located in Texas, with high concentrations also in Florida, California, and New York.

    o Age: The majority of Fetch users are aged between 20 and 40.

    o Language: English is the primary language for most users.

- Data Completeness:

    o Several major manufacturers, including COTY, Procter & Gamble, General Mills, PepsiCo, and Unilever, have items missing barcodes.

    o Notable brands such as Coca-Cola, Annie's Homegrown Grocery, Dove, Barefoot, and others also have missing barcodes.

- Product Category: The Health & Wellness category has the largest number of products, surpassing other categories in product variety.

Additional Insights

- **Top 5 Brands by Receipts Scanned Among Users 21 and Over**:

- **NERDS CANDY** - 3 receipts
- **DOVE** - 3 receipts
- **TRIDENT** - 2 receipts
- **SOUR PATCH KIDS** - 2 receipts
- **MEIJER** - 2 receipts

- **Top 5 Brands by Sales Among Users with Accounts Older Than Six Months**:
  - **CVS** - $72.00
  - **DOVE** - $30.91
  - **TRIDENT** - $23.36
  - **COORS LIGHT** - $17.48
  - **TRESEMMÉ** - $14.58

- **Fetch's Power Users (Users with Highest Distinct Transaction Counts)**:
  - **User ID 64e62de5ca929250373e6cf5** - 10 transactions
  - **User ID 62925c1be942f00613f7365e** - 10 transactions
  - **User ID 64063c8880552327897186a5**- 9 transactions
  - **User ID 6327a07aca87b39d76e03864**- 7 transactions
  - **User ID 624dca0770c07012cd5e6c03**- 7 transactions

- **Leading Brand in the Dips & Salsa Category**:
  - **TOSTITOS** with total sales of $181.30

- **Year-over-Year Growth of Fetch (Total Users)**:
  - Fetch's user base has shown significant growth, particularly between 2017 and 2020, with growth percentages as high as 820% in 2017 and 236.6% in 2018.
  - In 2023, Fetch experienced a decline in growth by -42.3%, followed by another drop in 2024 with -24.8%.