FETCH TAKE HOME ASSESSMENT

This report provides an analysis of the provided data to answer key business questions and uncover insights regarding Fetch's user behavior, brand performance, and year-over-year growth. The assessment covers:

Objectives:
Identify data quality issues and propose resolutions.
Understand field relationships and potential data challenges.
Present insights effectively using Power BI charts and reports.
Summarize key findings and provide actionable recommendations.

1. Data Quality Issues Identified:

Users Table:
Data Inconsistencies:
Date formats in CREATED_DATE and BIRTH_DATE were inconsistent (some contained time zones, others didn't).
Duplicate user records.

Transactions Table:
Data Inconsistencies:
FINAL_SALE values included non-numeric characters.
Some transactions had FINAL_QUANTITY = zero, which could indicate anomalies.
Purchase date discrepancies (scan date occurring before purchase date).

Products Table:
Data Inconsistencies:
Duplicated barcodes with varying product descriptions.

Data Cleaning Steps Taken
Missing Data Handling:
Filled missing TOTAL_SPEND with 0.
Set missing categorical fields (STATE, LANGUAGE, GENDER) to "UNKNOWN".
Dropped rows with critical missing values (e.g., missing USER_ID).
Data Type Standardization:
Converted FINAL_SALE and FINAL_QUANTITY to numeric.
Standardized date formats to YYYY-MM-DD.
Duplicate Removal:
Removed duplicate USER_ID and BARCODE entries based on business rules.
Consistency Checks:
Ensured purchase dates were always earlier than scan dates.

2. SQL Queries to Answer Business Questions

*** Queries are present in Queries Folder**

Open-Ended Questions
As a Sr. Data Analyst at Fetch, I have made the following assumptions to derive better insights and provide valuable recommendations that contribute to improving the company's growth and strategic decision-making.

I also took the initiative to match only users present in both the Users and Transactions tables, as this alignment is crucial to accurately assess the company's growth.

1. Who Are Fetch's Power Users?
Identify Fetch's power users based on their purchasing behavior. A power user is defined as someone who belongs to the top 10% of total spenders.

Assumptions:
The users_spend table contains user data along with their total spend (TOTAL_SPEND) and total quantity (TOTAL_QUANTITY).
Power users are those whose total spending falls within the top 10% of all users.
Spending distribution follows a continuous trend, and the threshold for power users is determined using the 90th percentile.

2. Which Is the Leading Brand in the Dips & Salsa Category?
Determine the leading brand in the "Dips & Salsa" category based on total sales and the number of items sold.

Assumptions:
The transactions table contains purchase records linked to product details through the barcode field.
The products table categorizes products using the category_2 column, where "Dips & Salsa" is a valid category.
The leading brand is defined as the one with the highest total sales (FINAL_SALE). In case of a tie, the brand with the highest quantity (FINAL_QUANTITY) sold will be chosen.

3. At What Percent Has Fetch Grown Year Over Year?
Calculate Fetch's year-over-year (YoY) growth percentage based on total sales in each year.
Assumptions:
The transactions table contains purchase records with a purchase_date column in YYYY-MM-DD format.
Growth is calculated using the formula:
(current year's sales − previous year's sales/ previous year's sales) × 100
If sales data for the previous year is unavailable, growth should return NULL.

3. Communicate with stakeholders

Construct an email or slack message that is understandable to a product or business leader who is not familiar with your day-to-day work. Summarize the results of your investigation. Include:

- Key data quality issues and outstanding questions about the data
- One interesting trend in the data
    - Use a finding from part 2 or come up with a new insight
- Request for action: explain what additional help, info, etc. you need to make sense of the data and resolve any outstanding issues

Subject: Insights from Fetch Data Analysis & Next Steps
Hi,

I hope you're doing well.

As part of our data analysis efforts, I've reviewed the available datasets and identified key findings related to Fetch's user behavior, brand performance, and sales growth. Below is a summary of our analysis, along with actionable recommendations and outstanding questions that require further clarification.

1. Key Data Quality Issues Identified
During the data exploration process, several data quality issues were observed that may impact analysis accuracy and business decision-making:
Missing Data:
BIRTH_DATE (10% missing), STATE (13% missing), LANGUAGE (40% missing) in the users table.
BARCODE missing in 5,762 transactions, leading to unmatched product details.
Inconsistent Data Formats:
Varying formats in CREATED_DATE and BIRTH_DATE fields (presence of time zone differences).
Duplicate Records:
Some duplicated user IDs and barcodes affect transactional accuracy.
Sales Anomalies:
Transactions with FINAL_QUANTITY recorded as "zero," which might indicate erroneous data entry or return transactions.

Outstanding Questions:
Can we obtain additional data sources to fill missing values (e.g., state, language)?
Are barcode mismatches due to system errors or retailer data inconsistencies?
What could be causing the high number of unmatched users in transactions?

2. Interesting Trends Discovered:
One key trend observed in the analysis is:

Power Users Contribution:

The top 10% of users (based on total spend) contribute to over 50% of Fetch's total revenue.
These power users are primarily from Florida (FL), Pennsylvania (PA), and New York (NY), with a strong preference for health and wellness products.
Spanish-speaking users account for a significant portion of spending in these regions.

Recommendation:
We suggest introducing personalized loyalty programs and targeted promotions to further engage power users in these key regions.

3. New Insight from the Data:
Top-Performing Brand in "Dips & Salsa" Category:

Our analysis revealed that Tostitos is the market leader in the "Dips & Salsa" category.
This category sees the highest sales in the Midwest region, indicating a growing demand.

Recommendation:
We recommend increasing product availability and running targeted marketing campaigns for Tostitos in high-performing states to maximize sales.

4. Request for Action:

To further refine our analysis and unlock deeper insights, the following support is needed:

Data Enrichment:
Request updated user demographic details to reduce missing values in the STATE and LANGUAGE columns.
Collaboration with the product team to validate missing barcode entries and categorize unmatched transactions.

Data Validation Guidelines:
Clarification on how to handle zero quantity transactions and sales anomalies for more accurate reporting.

Business Objectives Alignment:
A brief discussion to align our analysis with upcoming business goals (e.g., focusing on new user acquisition vs. retention).
Please let me know a convenient time to discuss these findings and next steps.
Looking forward to your thoughts.

Best regards,
Aditya K
Sr. Data Analyst

Related Docs and Files:
1. Insights Report (PDF):
A detailed report summarizing key insights, business trends, and recommendations.
Includes data quality assessment, SQL queries, and assumptions made during the analysis.

2. Power BI Dashboard:
An interactive dashboard showcasing:
User spending behavior and power users.
Brand performance and sales trends.
Year-over-year growth analysis.
Filters are provided for deeper exploration.

3. Data Files:
Cleaned Data: Processed datasets with missing values handled, date formats standardized, and duplicates removed.
Uncleaned Data: Raw datasets provided for reference to ensure transparency.

4. Python Code for Data Cleaning:
The data cleaning process was performed using Python scripts to ensure consistency and accuracy across all datasets.