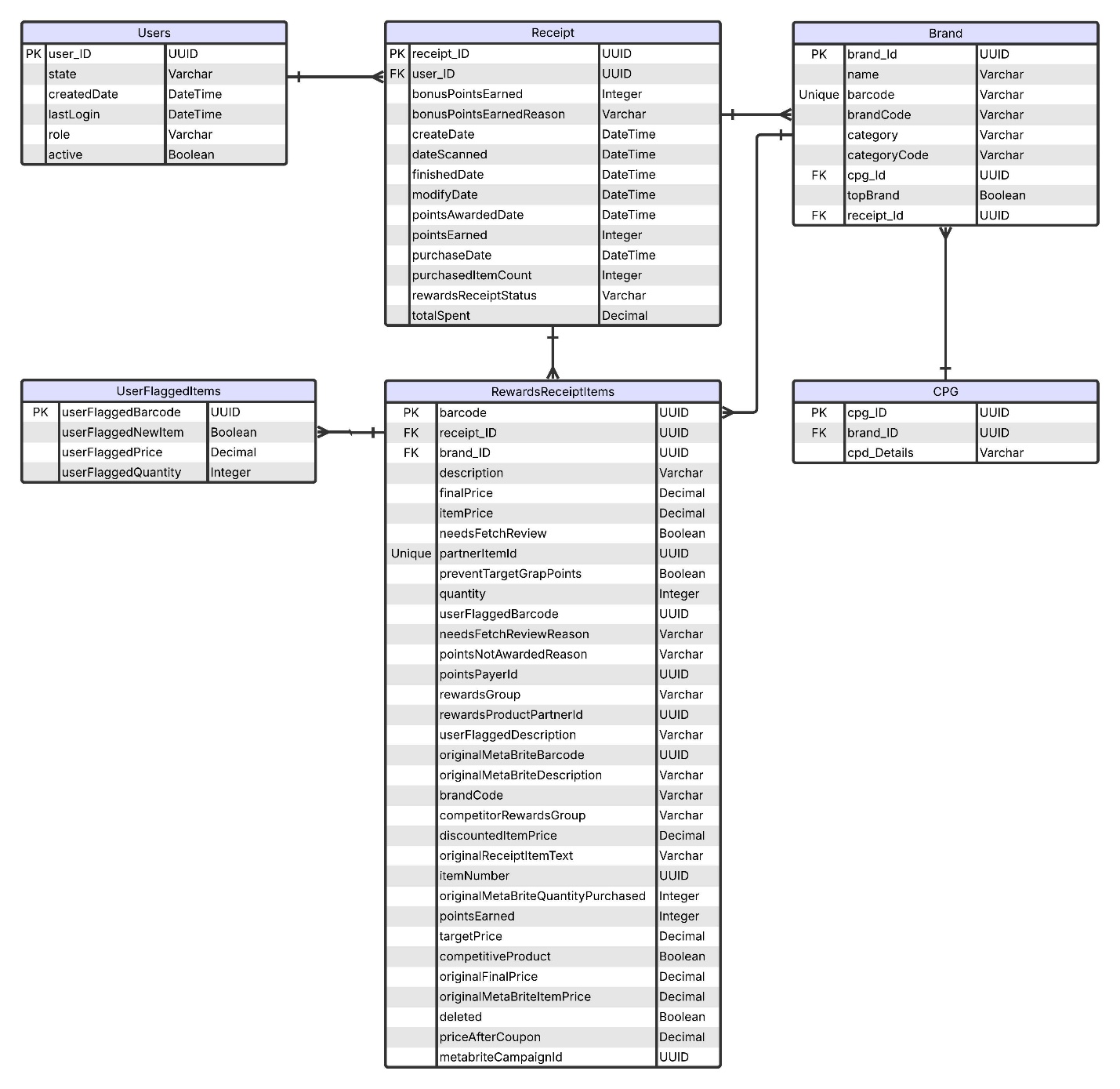
**Fetch Rewards-Analytical Engineer Internship**

# **First: Review Existing Unstructured Data and Diagram a New Structured Relational Data Model**



# **Second: Write queries that directly answer predetermined questions from a business stakeholder**

1. What are the top 5 brands by receipts scanned for most recent month?

SELECT b.name, COUNT(DISTINCT r.receipt\_id) AS receipt\_count

FROM Brand b

JOIN RewardsReceiptItems rri ON b.brand\_id = rri.brand\_id

JOIN Receipts r ON rri.receipt\_id = r.receipt\_id

WHERE strftime('%Y-%m', r.dateScanned) = (

SELECT strftime('%Y-%m', MAX(dateScanned))

FROM Receipts)

GROUP BY b.name

ORDER BY receipt\_count DESC

LIMIT 5;

1. How does the ranking of the top 5 brands by receipts scanned for the recent month compare to the ranking for the previous month?

SELECT b.name, COUNT(DISTINCT r.receipt\_id) AS receipt\_count

FROM Brand b

JOIN RewardsReceiptItems rri ON b.brand\_id = rri.brand\_id

JOIN Receipts r ON rri.receipt\_id = r.receipt\_id

WHERE strftime('%Y-%m', r.dateScanned) = (

SELECT strftime('%Y-%m', date(MAX(dateScanned), '-1 month'))

FROM Receipts)

GROUP BY b.name

ORDER BY receipt\_count DESC

LIMIT 5;

1. How does the ranking of the top 5 brands by receipts scanned for the recent month compare to the ranking for the previous month?

SELECT rewardsReceiptStatus, AVG(totalSpent) AS average\_spend

FROM Receipts

WHERE RewardsReceiptStatus IN ('FINISHED', 'REJECTED')

GROUP BY rewardsReceiptStatus;

1. How does the ranking of the top 5 brands by receipts scanned for the recent month compare to the ranking for the previous month?

SELECT rewardsReceiptStatus, SUM(purchasedItemCount) AS total\_items\_purchased

FROM Receipts

WHERE rewardsReceiptStatus IN ('FINISHED', 'REJECTED')

GROUP BY rewardsReceiptStatus;

1. How does the ranking of the top 5 brands by receipts scanned for the recent month compare to the ranking for the previous month?

SELECT b.name, SUM(r.totalSpent) AS total\_spend

FROM Brand b

JOIN RewardsReceiptItems rri ON b.brand\_id = rri.brand\_id

JOIN Receipts r ON rri.receipt\_id = r.receipt\_id

WHERE r.createDate >= date('now', '-6 months')

GROUP BY b.name

ORDER BY total\_spend DESC

LIMIT 1;

1. How does the ranking of the top 5 brands by receipts scanned for the recent month compare to the ranking for the previous month?

SELECT b.name, COUNT(DISTINCT r.receipt\_id) AS transaction\_count

FROM Brand b

JOIN RewardsReceiptItems rri ON b.brand\_id = rri.brand\_id

JOIN Receipts r ON rri.receipt\_id = r.receipt\_id

WHERE r.createDate >= date('now', '-6 months')

GROUP BY b.name

ORDER BY transaction\_count DESC

LIMIT 1;

# **Third: Evaluate Data Quality Issues in the Data Provided**