

CPG BRANDS AD-HOC INSIGHTS

DATA ANALYST SQL PROJECT

CREATED BY SANG LU (SANDRA)



AGENDA



Introduction to Company



Problem Statement



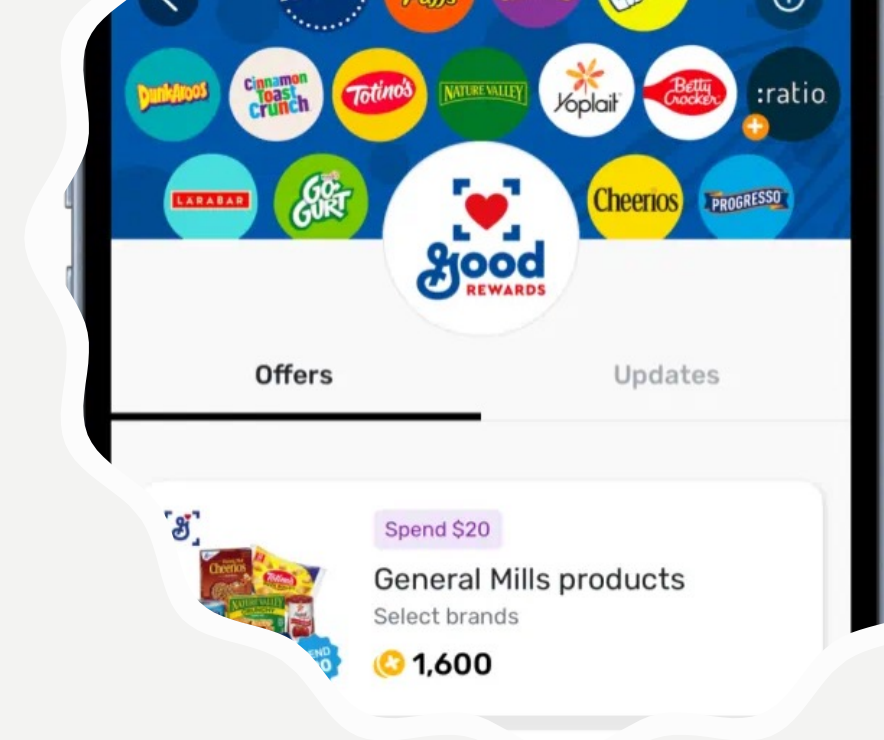
Ad-Hoc Request and Insights

INTRODUCTION TO COMPANY

Fetch Rewards is America's leading consumer-engagement platform that rewards shoppers for buying the brands they love. The Fetch app gives users the easiest way to save on everyday purchases by simply scanning their receipt.

Fetch helps brands activate and engage with America's consumers through billions of item-level receipts, driving business-changing results for leading CPG brands, retailers and restaurants.

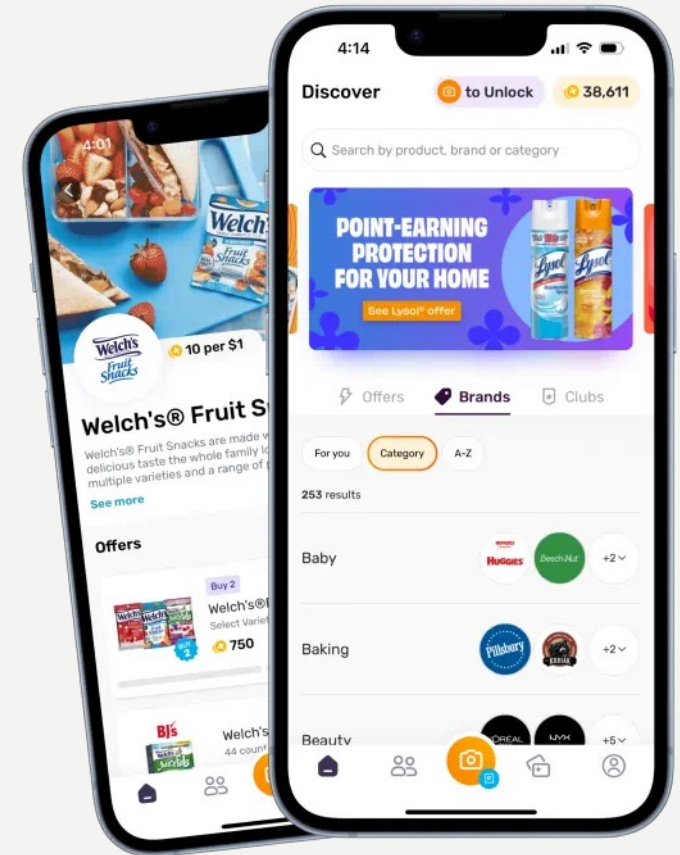
Domain: Consumer Goods



PROBLEM STATEMENT

The problem at hand is to analyze and compare various aspects of brand performance based on receipts scanned and user data. The focus is on understanding the recent month's data and comparing it to the previous month's data. Additionally, we aim to determine the average spend and total number of items purchased for receipts with different reward status. Furthermore, we need to identify the brand with the highest spend and the brand with the most transactions among users created within the past 6 months.

Overall, the problem revolves around analyzing brand performance metrics, such as receipts scanned, average spend, total items purchased, and user behavior within specific timeframes. The objective is to provide insights into brand rankings, user spending patterns, and identify key brand performance indicators based on the available data.



AD HOC REQUEST AND INSIGHTS



Review unstructured JSON data and diagram a new structured relational data model



Generated a SQL query to answer several business questions



Evaluate data quality issues in the provided data



Present this insight to business stakeholders

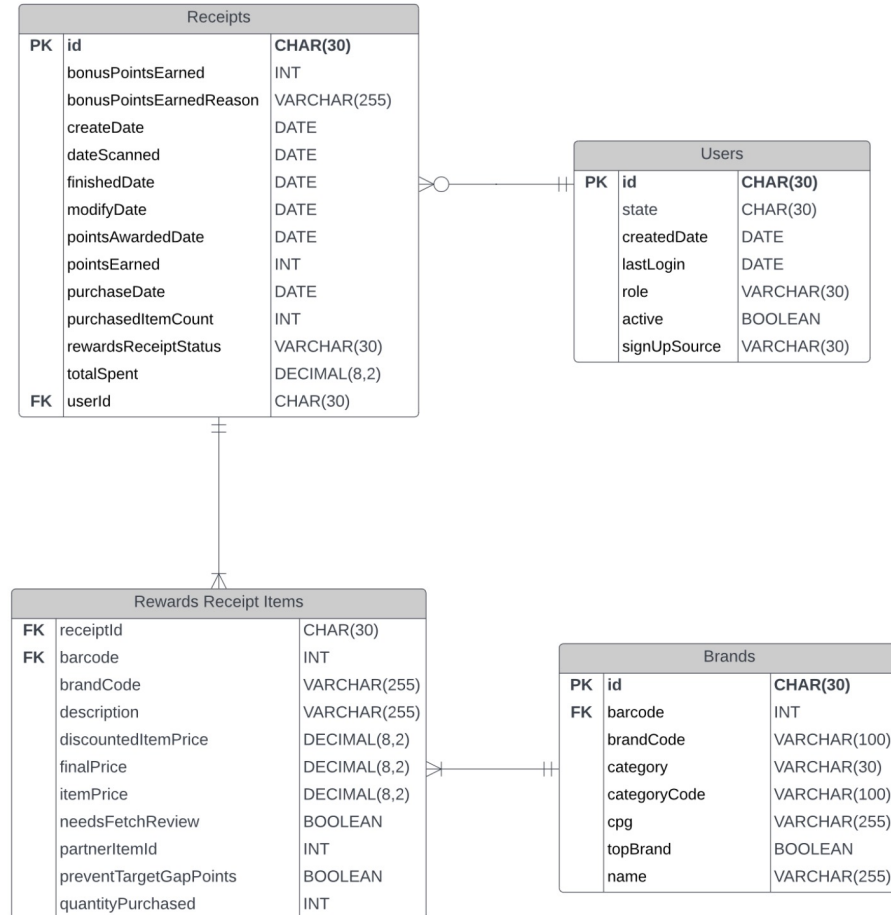
STANDARDIZE UNSTRUCTURED DATA

```
{ "_id":  
  {"$oid": "601ac115be37ce2ead437551"}, "barcode": "511111019862", "category": "Baking", "categoryCode": "BAKING", "cpg": {"$id":  
  {"$oid": "601ac114be37ce2ead437550"}, "$ref": "Cogs"}, "name": "test brand @1612366101024", "topBrand": false}  
  {"_id":  
    {"$oid": "601c5460be37ce2ead43755f"}, "barcode": "511111519928", "brandCode": "STARBUCKS", "category": "Beverages", "categoryCode": "BEVERAGES", "cpg": {"$id":  
      {"$oid": "5332f5fbe4b03c9a25efd0ba"}, "$ref": "Cogs"}, "name": "Starbucks", "topBrand": false}  
    {"_id": {"$oid": "601ac142be37ce2ead43755d"}, "barcode": "511111819905", "brandCode": "TEST BRANDCODE @1612366146176", "category": "Baking", "categoryCode": "BAKING", "cpg": {"$id":  
      {"$oid": "601ac142be37ce2ead437559"}, "$ref": "Cogs"}, "name": "test brand @1612366146176", "topBrand": false}
```




	id	barcode	brandcode	category	categoryCode	name	.	cpg
1	601ac115be37ce2ead437551	511111019862	<null>	Baking	BAKING	test brand @1612366101024	0	601ac114be37ce2ead437550
2	601c5460be37ce2ead43755f	511111519928	STARBUCKS	Beverages	BEVERAGES	Starbucks	0	5332f5fbe4b03c9a25efd0ba
3	601ac142be37ce2ead43755d	511111819905	TEST BRANDCODE @1612366146176	Baking	BAKING	test brand @1612366146176	0	601ac142be37ce2ead437559
4	601ac142be37ce2ead43755a	511111519874	TEST BRANDCODE @1612366146051	Baking	BAKING	test brand @1612366146051	0	601ac142be37ce2ead437559
5	601ac142be37ce2ead43755e	511111319917	TEST BRANDCODE @1612366146827	Candy & Sweets	CANDY_AND_SWEETS	test brand @1612366146827	0	5332fa12e4b03c9a25efd1e7
6	601ac142be37ce2ead43755b	511111719885	TEST BRANDCODE @1612366146091	Baking	BAKING	test brand @1612366146091	0	601ac142be37ce2ead437559
7	601ac142be37ce2ead43755c	511111219897	TEST BRANDCODE @1612366146133	Baking	BAKING	test brand @1612366146133	0	601ac142be37ce2ead437559

DATA MODEL



QUESTIONS

- what are the **top 5 brands by receipts** scanned for most recent month?
 - How does the **ranking of the top 5 brands by receipts** scanned for the recent month compared to the ranking for the previous month?
 - When considering **average spend** from receipts with '**rewardsReceiptStatus**' of 'Accepted' or 'Rejected', which is greater?
 - When considering **total number of items** purchased from receipts with '**rewardsReceiptStatus**' of 'Accepted' or 'Rejected', which is greater?
 - Which brand has the **most spend** among **users** who were created within the past 6 months?
 - Which brand has the **most transactions** among **users** who were created within the past 6 months?
- 

QUESTION 1



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

```
-- Find scanned receipts' ids for the most recent month from 2021-02-01 (included) to 2021-03-01 (included)
with MostRecentScannedReceipts AS (select id
                                   from receipts
                                   where dateScanned between "2021-02-01" and "2021-03-01"),

-- brandCode of items listed on the receipts
itemsList AS (select brandCode
                 from items
                 where receiptId in (select id from MostRecentScannedReceipts)
                 and brandCode is not null)

select brandCode, count(*) as Numbers
from itemsList
group by 1
order by 2 desc
limit 5;
```

Output

	 brandCode	 Numbers
1	BRAND	3
2	MISSION	2
3	VIVA	1

QUESTION 2



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

```
-- Find scanned receipts' ids for the previous month from 2021-01-01 (included) to 2021-01-30 (included)
with PreviousMonthScannedReceipts AS (select id
                                       from receipts
                                       where dateScanned between "2021-01-01" and "2021-01-30"),

-- brandCode of items listed on the receipts
itemsList AS (select brandCode
                  from items
                  where receiptId in (select id from PreviousMonthScannedReceipts)
                  and brandCode is not null)

select brandCode, count(*) as Numbers
from itemsList
group by 1
order by 2 desc
limit 5;
```

Output



	 brandCode	 Numbers
1	HY-VEE	291
2	BEN AND JERRYS	180
3	PEPSI	93
4	KROGER	89
5	KLEENEX	88

QUESTION 3

When considering average spend from receipts with 'rewardsReceiptStatus' of 'Accepted' or 'Rejected', which is greater?

```
select round(avg(case when rewardsReceiptStatus = "FINISHED" then totalSpent end), 2) as average_spend_accepted,  
       round(avg(case when rewardsReceiptStatus = "REJECTED" then totalSpent end), 2) as average_spend_rejected  
from receipts;
```

Output



	 average_spend_accepted ▾	 average_spend_rejected ▾
1	80.85	23.33

QUESTION 4

When considering total number of items purchased from receipts with 'rewardsReceiptStatus' of 'Accepted' or 'Rejected', which is greater?

```
select sum(case when rewardsReceiptStatus = "FINISHED" then purchasedItemCount end) as total_items_accepted,  
       sum(case when rewardsReceiptStatus = "REJECTED" then purchasedItemCount end) as total_items_rejected  
from receipts;
```

Output

	 total_items_accepted ▾	 total_items_rejected ▾
1	24552	519

QUESTION 5



Which brand has the most spend among users who were created within the past 6 months?

```
-- Find the users who created within the past 6 months from 2020-09-01 to 2021-03-01
with newRegisteredUsers as (select id, min(createdDate) as earlist_createdDate
                             from users
                             group by 1
                             having earlist_createdDate between "2020-09-01" and "2021-03-01"),

-- Find the receipts scanned by those users' Id
receiptsOfUsers as (select id
                     from receipts
                     where userId in (select id from newRegisteredUsers))

-- brandCode of items listed on the receipts
select brandCode, sum(finalPrice * quantityPurchased) as totalSpent
from items
where receiptId in (select id from receiptsOfUsers)
  and brandCode is not null
group by 1
order by 2 desc
limit 1;
```

Output

	 brandCode ▾	 totalSpent ▾
1	HEMPLER'S	5611.11

QUESTION 6

Which brand has the most transactions among users who were created within the past 6 months?

```
with newRegisteredUsers as (select id, min(createdDate) as earlist_createdDate
                             from users
                             group by 1
                             having earlist_createdDate between "2020-09-01" and "2021-03-01"),

-- Find the receipts scanned by those users' Id
receiptsOfUsers as (select id
                    from receipts
                    where userId in (select id from newRegisteredUsers))

select brandCode, count(*) as transactions
from items
where receiptId in (select id from receiptsOfUsers)
   and brandCode is not null
group by 1
order by 2 desc
limit 1;
```

Output

	 brandCode ▾	 transactions ▾
1	HY-VEE	291

THANK YOU!