Notes:

- On a daily basis, I am using Snowflake so the SQL that I have used to write the query below is the standard version of SQL: ANSI.
- I used the Postgres database to upload the data and analyze it. Some of the errors while uploading the data are below,
    - In receipts.json file,
        - Error: invalid input syntax for type json
        - DETAIL: Token "PLATE" is invalid.
        - Example: "description":"522 9\" PLATE","discountedItemPrice":"0.77","finalPrice":"0.77","itemPrice":"0.77"

        - ERROR: invalid input syntax for type json
        - DETAIL: Token "tall" is invalid.
        - Example: {"barcode":"043000005217","description":"Black and White Easter Bunny 16 1/2\" tall by no3

    - In brands.json file,
        - ERROR: invalid input syntax for type json
        - DETAIL: Token "Big" is invalid.
        - CONTEXT: JSON data, line 1: ...7e2aa60b873d6b0666d1"},"name":"General Mills "Big...
        - COPY brands_import, line 750, column receipts_json: "{"_id":{"$oid":"5dc07e2aa60b873d6b0666d1"},"name":"General Mills \"Big G\" Cereal","cpg":{"$ref":"Cogs..."

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

```
WITH BRANDS_PRICE AS (SELECT BD.NAME
                           , SUM(ID.FINALPRICE) AS TOT_AMT
                      FROM USERS_DIM UD
                           JOIN RECEIPTS_SUMMARY_FACT RS ON UD._ID = RS.USER_ID
                           JOIN RECEIPTS_FACT RF ON RS.RECEIPTS_ID = RF.RECEIPTS_ID
                           JOIN ITEMS_DIM ID ON RF.ITEMS_ID = ID._ID
                           JOIN BRANDS_DIM BD ON RF.BRAND_ID = BD._ID
                      WHERE CREATEDDATE >= DATEADD(month, -6, SYSDATE())
                      GROUP BY 1)
SELECT A.NAME
FROM (SELECT NAME
           , RANK() OVER(ORDER BY TOT_AMT DESC) AS RANK
      FROM BRANDS_PRICE) A
WHERE A.RANK = 1;
```

- **Evaluate the data quality issues in the data provided**

```
--Duplicate rows
SELECT ID, COUNT(ID)
FROM USERS
GROUP BY ID
HAVING COUNT(ID)>1;

--Barcode is assigned to multiple brands
SELECT BARCODE, count(BARCODE)
FROM BRANDS
GROUP BY BARCODE
HAVING count(BARCODE)>1;

SELECT *
FROM BRANDS
WHERE BARCODE = '"511111004790"';

--Multiple barcode, category, categorycode & cpg id for same brand
SELECT *
FROM BRANDS
WHERE NAME IN ('"Health Magazine"', '"Caleb"s Kola"');
```

- **Communicate with Stakeholders**

Hello,

This is Nischal from the Analytics team. I was going through our Receipts & Brand data and I noticed that a barcode is assigned to multiple brands and also the brand has multiple barcodes assigned to it. Should the barcode be unique to each brand and should each brand only have one barcode assigned to it?

I found this issue by counting the number of times a barcode has been used in our brand data and also, if you look into “Health Magazine” and “Caleb’s Kola” brand data in the screenshot below, you can see multiple barcodes assigned to these brands.

We would like to understand how we should handle barcodes in more detail because this has a negative downstream impact on our reports, which may lead to inaccurate assumptions while making business decisions as well as issues with data storage. The information behind how barcode codes should be assigned to a brand and the number of barcodes, category codes and CPG reference id will help streamline the process and optimize our data systems, while saving resources for our team.

Please let me know if you have any questions. I would be happy to further discuss and resolve this issue as soon as we can. Thanks!

Nischal

id	barcode	category	categorycode	cpg	name	topbrand
"5a4d23dae4b0bcb2c74ea77e"	"511111000518"	"Beverages"	(null)	{"$ref":"Cogs","$id":{"$oid":"5332f5fbe4b03c9a25efd0ba"}}	"Caleb's Kola"	false
"5d6415d5a3a018514994f429"	"511111605058"	"Magazines"	(null)	{"$ref":"Cogs","$id":{"$oid":"5d5d4fd16d5f3b23d1bc7905"}}	"Health Magazine"	(null)
"5f298852be37ce7958c5952d"	"511111915287"	"Magazines"	"MAGAZINES"	{"$ref":"Cogs","$id":{"$oid":"5d66b9dcee7f2d201c7281cd"}}	"Health Magazine"	(null)
"5d601d74a3a018514994f422"	"511111004912"	"Snacks"	(null)	{"$ref":"Cogs","$id":{"$oid":"53e10d6368abd3c7065097cc"}}	"Caleb's Kola"	(null)