

Brands

1. Missing or Incorrect data – brandCode, category, categoryCode have NULL data, or have inconsistent naming

```
SELECT
    'brandCode' AS FieldName,
    COUNT(*) AS TotalNulls
FROM [fetch].[dbo].[brands]
WHERE brandCode IS NULL
UNION ALL
SELECT
    'category' AS FieldName,
    COUNT(*) AS TotalNulls
FROM [fetch].[dbo].[brands]
WHERE category IS NULL
UNION ALL
SELECT
    'categoryCode' AS FieldName,
    COUNT(*) AS TotalNulls
FROM [fetch].[dbo].[brands]
WHERE categoryCode IS NULL
UNION ALL
SELECT
    'cpg' AS FieldName,
    COUNT(*) AS TotalNulls
FROM [fetch].[dbo].[brands]
WHERE categoryCode IS NULL
UNION ALL
SELECT
    'topBrand' AS FieldName,
    COUNT(*) AS TotalNulls
FROM [fetch].[dbo].[brands]
WHERE categoryCode IS NULL
UNION ALL
SELECT
    'name' AS FieldName,
    COUNT(*) AS TotalNulls
FROM [fetch].[dbo].[brands]
WHERE categoryCode IS NULL;
```

2. Duplicate data in barcode field

```
SELECT barcode, COUNT(*) AS DuplicateCount
FROM [fetch].dbo.brands
GROUP BY barcode
HAVING COUNT(*) > 1;
```

Users

1. There are a lot of duplicates _ids, more than half of the data set is duplicated

```
SELECT _id, COUNT(*) AS DuplicateCount
FROM [fetch].dbo.users
GROUP BY _id
HAVING COUNT(*) > 1;
```

2. There is missing data in 'state', a lot of active users do not have a state associated with them

```
SELECT * FROM [fetch].dbo.users WHERE state is NULL;
```

3. There are some active users that do not have a 'lastLogin' timestamp.

```
SELECT *  
FROM [fetch].dbo.users  
WHERE active = 1 AND lastLogin IS NULL;
```

4. There are active users where the signUpSource is NULL

```
SELECT  
    'signUpSource ' AS FieldName ,  
    COUNT(*) AS TotalNulls  
FROM [fetch].[dbo].[users]  
WHERE signUpSource IS NULL;
```

Receipts

1. Many NULLs in various fields meaning not all data were captured, or was captured incorrectly, such as in pointsAwardedDate and/or finishedDate

```
SELECT  
    'finishedDate ' AS FieldName,  
    COUNT(*) AS TotalNulls  
FROM [fetch].[dbo].[receipts]  
WHERE finishedDate IS NULL  
UNION ALL  
SELECT  
    'pointsAwardedDate ' AS FieldName ,  
    COUNT(*) AS TotalNulls  
FROM [fetch].[dbo].[receipts]  
WHERE pointsAwardedDate IS NULL;
```

2. There are multiple barcodes associated with "ITEMS NOT FOUND", meaning there is missing information

```
SELECT barcode, description FROM [fetch].dbo.receipts WHERE description like '%ITEM NOT FOUND%'
```

3. Barcode format is not consistent.

```
SELECT barcode  
FROM receipt_items  
WHERE barcode LIKE '%[a-zA-Z]%';
```

Possible data enhancements, such as include add pointsAwardedDate for better tracking/monitoring of users points for auditing.