

34.51°

Introduction to Analytics Core Data Store (ACDS)



Agenda

1. Analytics Core Data Store
2. Data Quality and Quirks
3. Golden Rules for using the ACDS
4. 84.51° Data Landscape

Customer Sales Data Product

Part of the Data Products Value Stream,
which is part of the Data Platform Program,
which is part of Enterprise Capabilities.



To connect: CSDPLeads_DL@8451.com

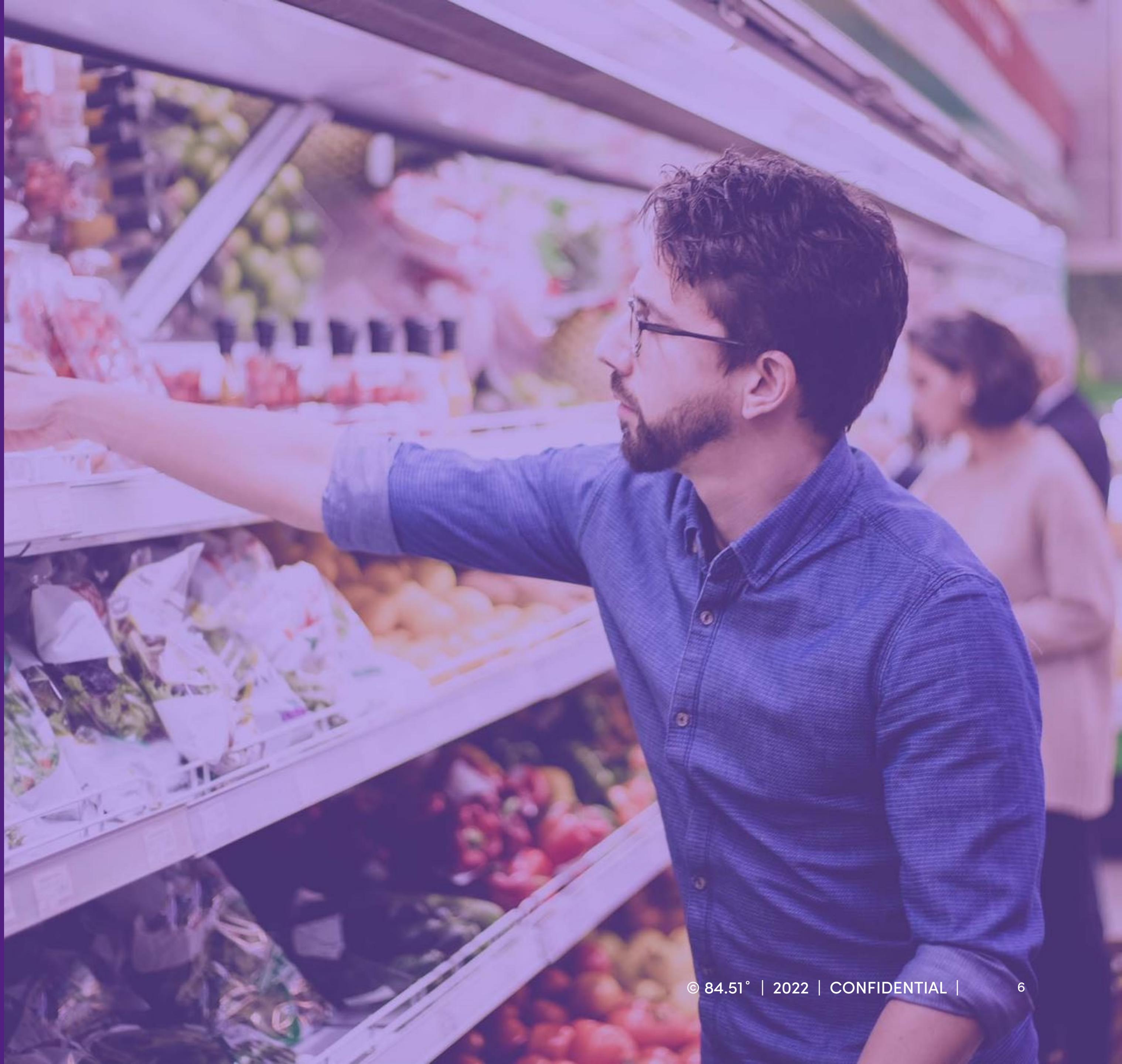
Analytics Core Data Store (ACDS)

Marts, Dimensions, Facts, and Derived

Types of Data

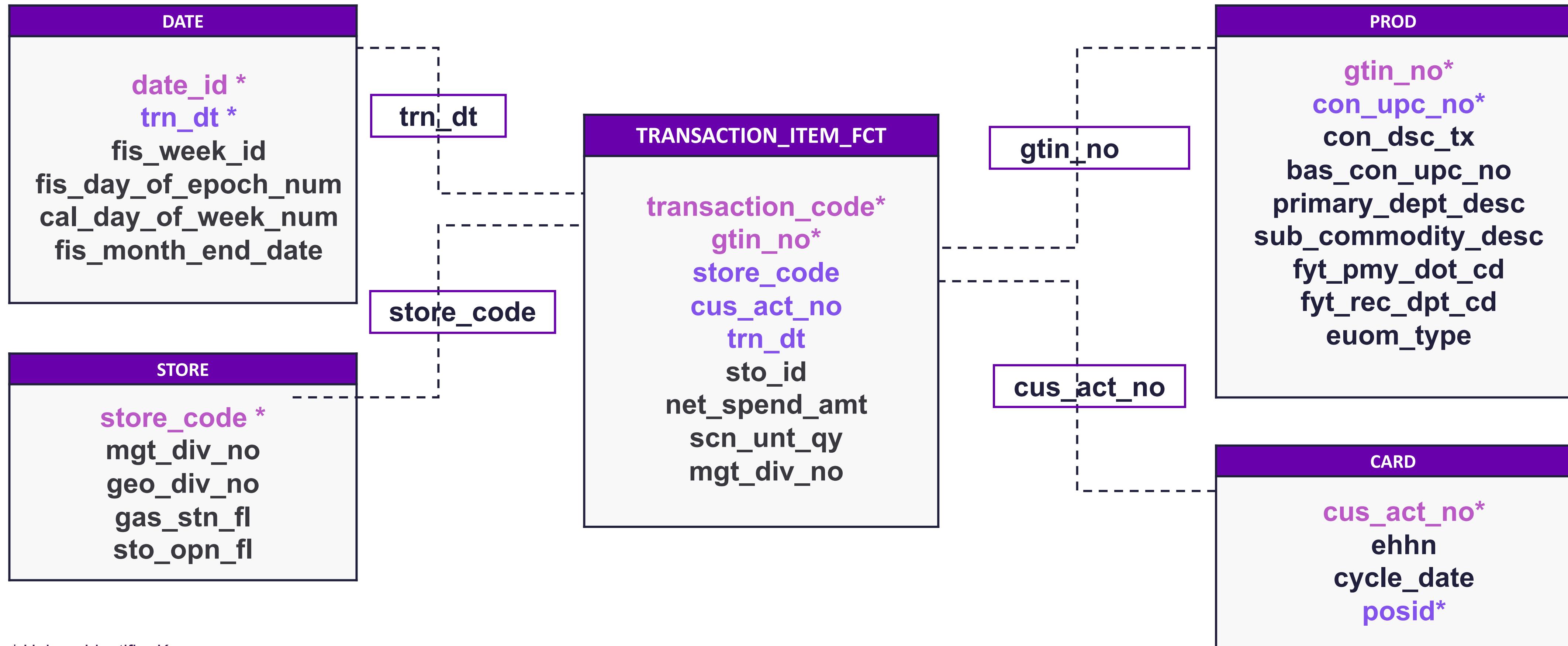
- **Data Mart** – a collection of data assets that can be used together to answer business questions, comprised individual data assets that are either facts or dimensions, typically built with a specific purpose in mind
- **Facts**– a data asset containing recorded values of some transactional or business process
- **Dimension** – a data asset containing information about the attributes of a given domain (product, store, date, household)
- **Derived** – for example, segmentations are considered somewhere in between fact and dimensions, these are considered as historic records, tied to a specific point in time

A story of a boy named Bernie



ACDS is a Data Mart

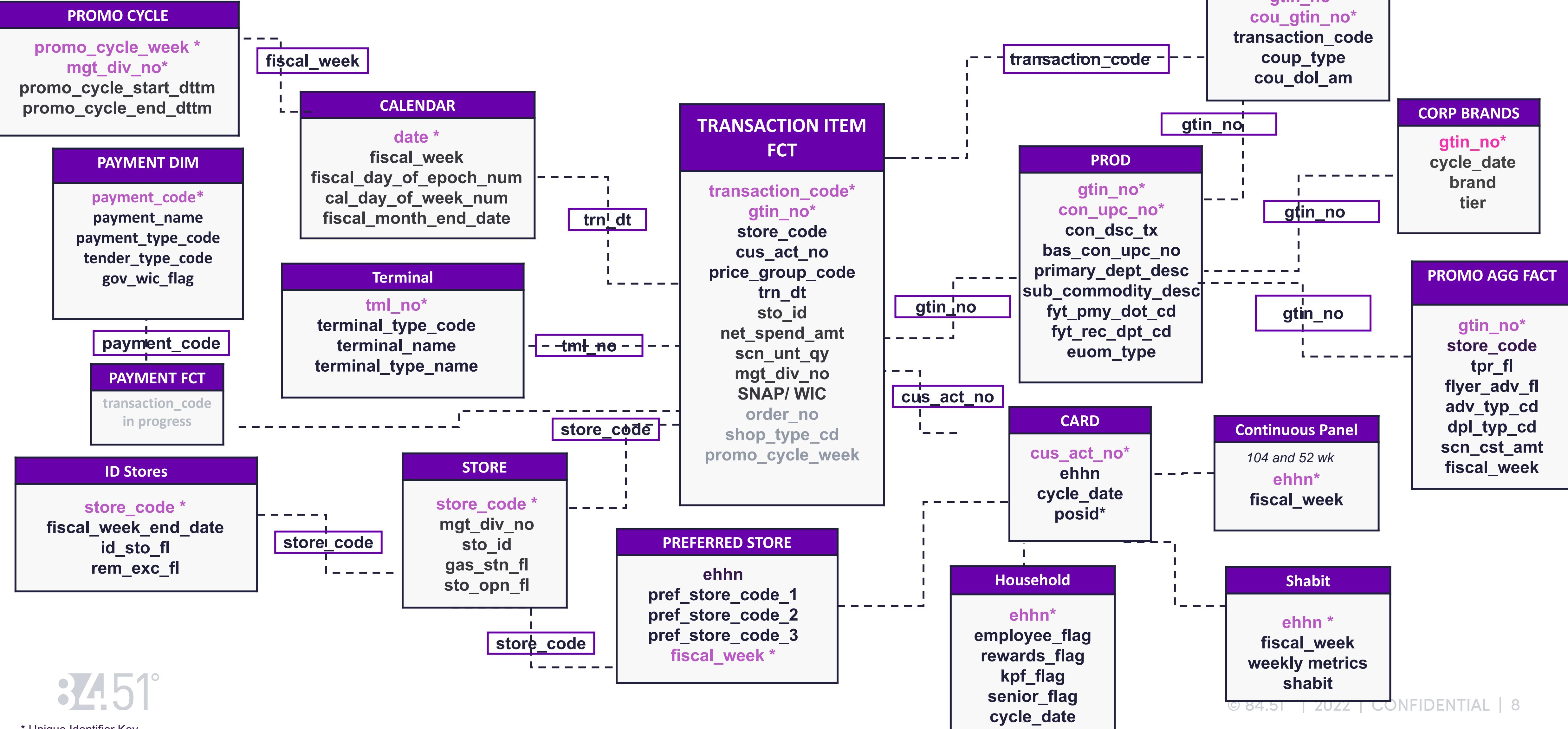
This is a sample of the data available in each table



* Unique Identifier Key

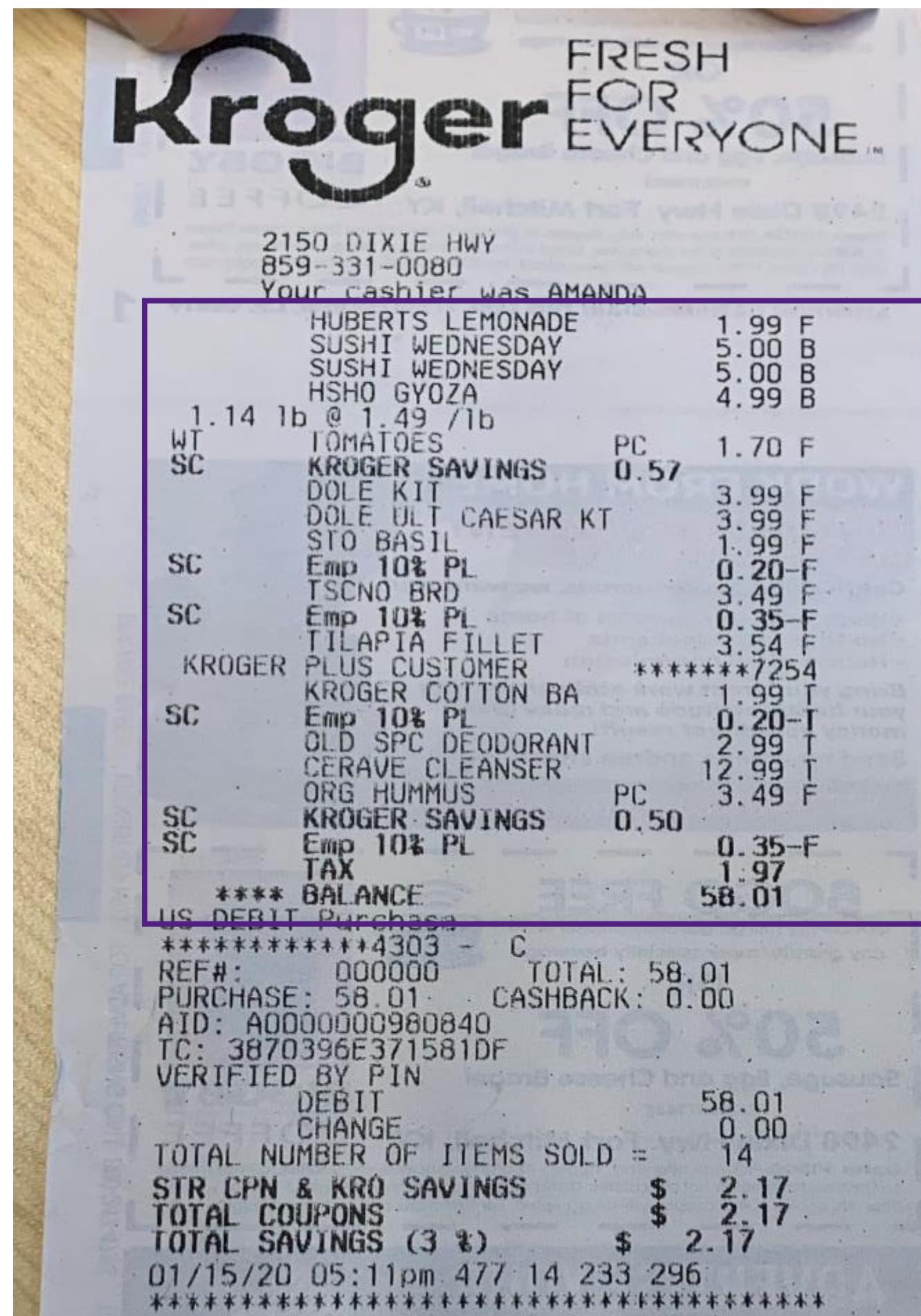
There is a lot in the ACDS!

Compressing images



Transaction Item Fact

Transactional item level details describing what a customer purchased.



mgt_div_no	net_spend_amt	scn_unt_qty	gtin_no	trn_tm	transaction_code	store_code	trn_dt
014	22.16	1	000111036838	17:19:00	014_00477_122_221_20200115_17:19:00	014_00477	20200115
014	1.99	1	0004900007035	17:11:00	014_00477_14_233_20200115_17:11:00	014_00477	20200115
014	3.54	1	0023889540000	17:11:00	014_00477_14_233_20200115_17:11:00	014_00477	20200115
014	3.14	1	0004157304669	17:11:00	014_00477_14_233_20200115_17:11:00	014_00477	20200115
014	4.99	1	0066955603002	17:11:00	014_00477_14_233_20200115_17:11:00	014_00477	20200115
014	3.99	1	0007143001702	17:11:00	014_00477_14_233_20200115_17:11:00	014_00477	20200115
014	2.99	1	0001204400025	17:11:00	014_00477_14_233_20200115_17:11:00	014_00477	20200115
014	1.79	1	000111035091	17:11:00	014_00477_14_233_20200115_17:11:00	014_00477	20200115
014	3.14	1	000111019439	17:11:00	014_00477_14_233_20200115_17:11:00	014_00477	20200115
014	3.99	1	0007143000088	17:11:00	014_00477_14_233_20200115_17:11:00	014_00477	20200115
014	12.99	1	0360600053775	17:11:00	014_00477_14_233_20200115_17:11:00	014_00477	20200115
014	1.70	1	0082692000009	17:11:00	014_00477_14_233_20200115_17:11:00	014_00477	20200115
014	10.00	2	0066955604007	17:11:00	014_00477_14_233_20200115_17:11:00	014_00477	20200115
014	1.79	1	000111001812	17:11:00	014_00477_14_233_20200115_17:11:00	014_00477	20200115

Transaction Item Fact

Krogerepedia Metadata and Data Dictionary

Key: transaction_code, gtin_no

Provides details about every item scanned at Kroger.

To speed up development and minimize cost, we should use the sample mart (subset of our transaction data).

- This a simple sample (1% of households) and should not be used as a statistical sample.
- Default parameter in Effodata

Product Dimension

Additional details about the product.



gtin_no	con_dsc_tx	bas_con_upc_no	con_siz_tx	primary_dept_desc	recap_dept_desc	sub_dept_desc	commodity_desc	sub_commodity_desc
0023889540000	TILAPIA FLT FR FH RW SC	0023889540000	1 RW	MEAT	SEAFOOD	FRESH SEAFOOD	SEAFOOD - TILAPIA	TILAPIA - FILLET
0066955603002	HSHO PORK GYOZA	0066955603002	4 OZ	DELI/BAKE	FRESH FOOD DEST	FRESH FOOD DEST	FRESH SUSHI	SUSHI - SNACKS & SALADS
0004157304669	LABREA PANE TSCNO BRD SLC	0004157304669	15.5 OZ	DELI/BAKE	BAKERY	BAKERY	BREAD	BREAD:ARTISAN
0066955604007	HSHO 5 SUSHI WEDNESDAY	0066955604007	7 OZ	DELI/BAKE	FRESH FOOD DEST	FRESH FOOD DEST	FRESH SUSHI	SUSHI - IN STORE PREPARED
0001204400025	OLD SPC HE PURSPRT AP/DEO	0001204400025	3 OZ	DRUG/GM	HBC	PERSONAL CARE	DEODORANTS	SOLID/STICK AP
0007143001702	DOLE ULT CAESAR SALAD KIT	0007143001702	11.3 OZ	PRODUCE	FRESH PRODUCE	VEGETABLES	SALAD MIX	KITS
0007143000088	DOLE GREEK CHP SALAD KIT	0007143000088	10.3 OZ	PRODUCE	FRESH PRODUCE	VEGETABLES	SALAD MIX	KITS
0360600053775	CRVE FACIAL CLEANSER	0360600053775	12 FO	DRUG/GM	HBC	BEAUTY	HAND BODY & FACIAL PRODUCTS	FACE
0082692000009	TOMATOES BUNCHED	0000000004664	11 LB	PRODUCE	FRESH PRODUCE	VEGETABLES	TOMATOES	TOMATOES TOV BULK RED
000111001812	STO BASIL	000111001812	0.5 OZ	PRODUCE	FRESH PRODUCE	VEGETABLES	HERBS/GARLIC	HERBS ORGANIC PACKAGED
0004900007035	HUBTS STRAWBERRY LEMONADE	0004900007035	16 FO	GROCERY	NATURAL FOODS	NATURAL FOODS	NF JUICE	NON-CARB JCE(OVER 50% JCE)
000111019439	ST ORIGIONAL HUMMUS ORG	000111019439	8 OZ	DELI/BAKE	DELI PACKAGED	PACKAGED DELI	HUMMUS/SALSA/DIPS	SAL: HUMMUS
000111035091	KRO TRPL CTTN BLL JMB SRC	000111035091	200 CT	DRUG/GM	HBC	HEALTH	FIRST AID PRODUCTS	COTTON

Product Dimension

Krogerpedia Data Dictionary

- Key: gtin_no
- Information about products from PID (Kroger), but not comprehensive of all product attributes.
- Contains enterprise product hierarchy. We use the PID hierarchy. Learn more [here](#).
- Learn more about UPCs [here](#).
 - Scan UPC
 - Base UPC
 - PLU (global and retailer)
 - Markdown
 - RPB

Store Dimension

Additional details about where the customer shopped.



store_code	sto_nam_tx	van_nam_tx	sto_str_adr_tx	sto_cty_nam_tx	sto_sta_cd
014_00477	KROGER	FT. MITCHELL	2150 DIXIE HWY	FORT MITCHELL	KY

Store Dimension

Krogerpedia Data Dictionary

- Key: store_code
 - Watch out! Store numbers are not unique!
- Use geo_div_no, instead of mgt_div_no, to identify division.
- Do not use sto_opn_flg to determine if a store is open. Instead join the store dimension to transactions and pull stores with valid transactions within a recent timeframe.

Card Dimension

Krogerpedia Data Dictionary

- Key: cus_act_no
- You should pull the card to household relationship from the transaction table, rather than card dimension.
- Kroger householding
 - Based off name and address
 - Card to person has a many to one relationship
 - Person to household has a many to one relationship
 - Card to household has a many to one relationship
 - The relationship can change over time
 - [Learn more here](#)



Date Dimension

Krogerpedia Data Dictionary

- Key: date
- Contains information to relate dates to Kroger fiscal calendar
- Check out [kayday](#), an internal package for handling Kroger dates. Note, much of [kayday](#)'s functionality is inherited by Effodata and KPI Aggregator.

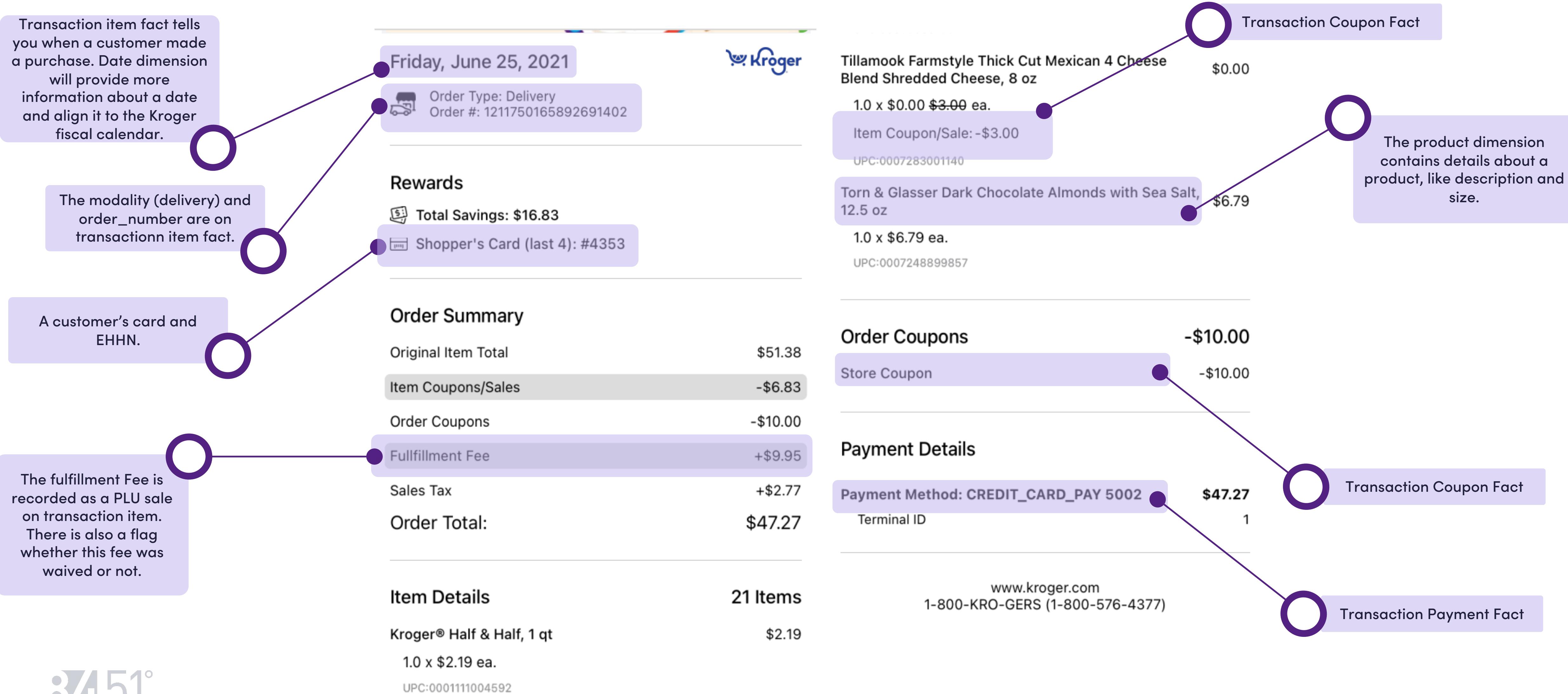


The image shows a 2021 Kroger calendar. The left side is a dark blue panel with the Kroger logo and the year 2021. The right side is a light blue panel featuring a cartoon illustration of a family (a man, a woman, and two children) sitting on a grassy hill under a blue sky with bees. Below the calendar is a horizontal bar with various holidays listed.

MONTH	S	M	T	W	T	F	S	PERIOD
FEB.	2	3	4	5	6	7	8	Q1
	9	10	11	12	13	14	15	
	16	17	18	19	20	21	22	P1
	23	24	25	26	27	28	29	
MAR.	1	2	3	4	5	6	7	
	8	9	10	11	12	13	14	P2
	15	16	17	18	19	20	21	
	22	23	24	25	26	27	28	
	29	30	31					
APR.	5	6	7	8	9	10	11	P3
	12	13	14	15	16	17	18	
	19	20	21	22	23	24	25	
	26	27	28	29	30			
MAY	3	4	5	6	7	8	9	P4
	10	11	12	13	14	15	16	
	17	18	19	20	21	22	23	
	24	25	26	27	28	29	30	Q2
JUNE	1	2	3	4	5	6	7	P5
	7	8	9	10	11	12	13	
	14	15	16	17	18	19	20	
	21	22	23	24	25	26	27	
	28	29	30					
JULY	5	6	7	8	9	10	11	P6
	12	13	14	15	16	17	18	
	19	20	21	22	23	24	25	
	26	27	28	29	30	31		
AUG.	2	3	4	5	6	7	8	P7
	9	10	11	12	13	14	15	
	16	17	18	19	20	21	22	
	23	24	25	26	27	28	29	
	30	31						Q3
SEPT.	1	2	3	4	5	6	7	P8
	13	14	15	16	17	18	19	
	20	21	22	23	24	25	26	
	27	28	29	30				
OCT.	4	5	6	7	8	9	10	P9
	11	12	13	14	15	16	17	
	18	19	20	21	22	23	24	
	30	31						Q4
NOV.	1	2	3	4	5	6	7	P10
	15	16	17	18	19	20	21	
	22	23	24	25	26	27	28	
	29	30						
DEC.	4	5	6	7	8	9	10	P11
	11	12	13	14	15	16	17	
	18	19	20	21	22	23	24	
	25	26	27	28	29	30	31	
	1	2	3	4	5	6	7	Q4
JAN.	8	9	10	11	12	13	14	P12
	15	16	17	18	19	20	21	
	22	23	24	25	26	27	28	
	29	30						
2021	1	2	3	4	5	6	7	P13
	10	11	12	13	14	15	16	
	17	18	19	20	21	22	23	
	24	25	26	27	28	29	30	
	24	25	26	27	28	29	30	

MONTH	WEEK OF YEAR	S	M	T	W	T	F	S	PERIOD
JAN.	49	X	3	4	5	6	7	8	9
	50	Y	10	11	12	13	14	15	16
	51	Z	17	18	19	20	21	22	23
	52	A	24	25	26	27	28	29	30
FEB.	1	B	31						
	2	C	1	2	3	4	5	6	7
	3	D	14	15	16	17	18	19	20
	4	E	21	22	23	24	25	26	27
	5	F	28						
MAR.	5	F	1	2	3	4	5	6	7
	6	G	7	8	9	10	11	12	13
	7	H	14	15	16	17	18	19	20
	8	I	21	22	23	24	25	26	27
	9	J	28	29	30	31			
APR.	9	J							
	10	K	4	5	6	7	8	9	10
	11	L	11	12	13	14	15	16	17
	12	M	18	19	20	21	22	23	24
	13	N	25	26	27	28	29	30	
MAY	13	N							
	14	O	2	3	4	5	6	7	8
	15	P	9	10	11	12	13	14	15
	16	Q	16	17	18	19	20	21	22
	17	R	23	24	25	26	27	28	29
	18	S	30	31					
JUN.	18	S	1	2	3	4	5	6	7
	19	T	6	7	8	9	10	11	12
	20	U	13	14	15	16	17	18	19
	21	V	20	21	22	23	24	25	26
	22	W	27	28	29	30			
JUL.	22	W							
	23	X	4	5	6	7	8	9	10
	24	Y	11	12	13	14	15	16	17
JULY	24	Y							
JULY	25	Z	16	17	18	19	20	21	22
	26	A	23	24	25	26	27	28	29
JULY	26	A							
JULY	27	B	1	2	3	4	5	6	7
	28	C	8	9	10	11	12	13	14
JULY	28	C							
JULY	29	D	15	16	17	18	19	20	21
	30	E	22	23	24	25	26	27	28
JULY	30	E							
JULY	31	F	29	30	31				
JULY	31	F							
JULY	32	G	5	6	7	8	9	10	11
	33	H	12	13	14	15	16	17	18
	34	I	19	20	21	22	23	24	25
	35	J	26	27	28	29	30		
	36	K	3	4	5	6	7	8	9
	37	L	10	11	12	13	14	15	16
	38	M	17	18	19	20	21	22	23
	39	N	24	25	26	27	28	29	30
	40	O	31						
	41</td								

Which ACDS asset contains each data point?



Data Quality and Quirks

Data Quality

The perception or assessment of data's fitness to serve its purpose in the context we are using it

Concepts of Data Quality

- **Correctness** – was the data sent captured correctly and accurately?
- **Consistency** – Is the data consistent across systems?
- **Completeness** – did we get the data we expected? Both volume and content
- **Timeliness** – can we get the data on-time? Can we produce it on-time?
- **Metadata** – data about the data

The main ways we achieve 'quality' are:

- File Monitoring
- Threshold based quality checks – Statistical Process Control (SPC)

We regularly apply quality checks to the ACDS data at each phase of its life cycle (raw, source, and target)

Data Quality is a team sport, if you see something, say something by raising a question in Krogerpedia!

Straggler Data

The 'straggle' is real...

Data that arrives later than we would have expected from Kroger

Caused by:

- Inclement weather taking down systems... clouds!
- POS issues... technology!
- In-store conditions... someone forgets, internet is broken, etc.



We **can not** prevent them from happening, but we **can** understand them and make users of our data aware

How do we define and measure stragglers? How late is late?

Date Diff (cycle – tran dt)	% of records
0	0.07%
1	97.94%
2	1.68%
3+	0.31%

- When we load the ACDS on Sundays, we will be missing ~2% of transactions from that entire week
- You can completely avoid this issue by using data from the week before the most recent publish of ACDS
- Recent POS changes (transition from TLOG to KLOG) will likely improve the tardiness of data, as data is streamed from the stores to KTD. 84.51° still receives the data once daily.

Conclusion:

A difference of one day is on time, everything we receive later than one day we consider late

Stragglers – Impact and conclusion

- When do we alert users about stragglers?
confluence...<https://confluence.kroger.com/confluence/display/8451DG/Weekly+Reporting>
- Most of the data we receive arrives when we expect it to
- Stragglers are present, and we can not reasonably prevent them from happening
- We will make you aware in cases when stragglers will significantly impact your analysis
- Straggler data is that which we receive more than one day after it had been transacted at the POS
- Overall, we receive most (97%+) of the data when we expect to receive it
- Some divisions are tardier than others...
- Every day of the week appears to be equally impacted by stragglers

Skeletons

"If you can't get rid of the family skeleton, you may as well make it dance" -- George Bernard Shaw

Occasionally, we receive some fact data (transactions, coupons, payments) that we are unable to find a corresponding record for in our dimensions (products or households)

These occurs due the timing we receive the household and product file.

Examples:

- Card-HH - cards that do not yet have a household assigned to them in Kroger's system



What do skeletons look like?

Households... in trans item fact

cus_act_no	ehhn	trn_dt	net_spend_amt	transaction_code	store_code
42344350172	None	20200113	3.00	016_00525_510_17_20200113_09:59:00	016_00525
42344350172	None	20200113	12.99	016_00525_510_17_20200113_09:59:00	016_00525
42344350172	None	20200113	3.00	016_00525_510_17_20200113_09:59:00	016_00525
42344350172	None	20200113	3.99	016_00525_510_17_20200113_09:59:00	016_00525
42344350172	None	20200113	7.49	016_00525_510_17_20200113_09:59:00	016_00525
42344350172	None	20200113	2.49	016_00525_510_17_20200113_09:59:00	016_00525
42344350172	None	20200113	1.99	016_00525_510_17_20200113_09:59:00	016_00525
42344350172	None	20200113	3.99	016_00525_510_17_20200113_09:59:00	016_00525
46801660514	None	20200113	1.00	097_00249_501_185_20200113_22:18:00	097_00249
46801660514	None	20200113	1.50	097_00249_501_185_20200113_22:18:00	097_00249

Products...

Skeleton products show up just like every other product does!

We do not create skeleton records in the product dimensions, if you need to include them in your analysis (i.e. for some total enterprise metric), you would need to do a left outer join to product dimension in order to keep all the transactions (both skeleton and non)

* For example code you can use to see how many skeletons exist in your data please reference [here...](#)

Skeleton Conclusions

Their presence impacts how we use our data in order to answer questions from the business

Since data coming in daily is more likely to have records not found in prior week's weekly dimension, we expect to find more skeleton records in the daily data

- Daily data lives in a separate bucket, this data gets re-stated (Sunday) using the updated dimensions from the weekly load to account for the prior week's skeletons

Daily v Weekly

- You should default to using the weekly publish of ACDS data. You must have valid need and understand the increased risk and inconsistencies before using daily data.
- The risk of stragglers and skeletons is increased with daily transactions.
- If you must use daily data, please be sure to contact the Customer Sales Data Product
- More information:
<https://confluence.kroger.com/confluence/display/8451DG/Daily+v+Weekly+Data>

Data Quirks : Overall conclusion and impact

What do I do with all this knowledge?

- Both quirks discussed are due to the frequency at which we receive data from Kroger (i.e. timeliness issues)

- Stragglers are data we receive late
- Skeletons caused by the timing at which we receive dimension data

Stragglers we are forced to live with... we can only make you aware of them

Skeletons we can correct (eventually) when we re-state the daily data to update skeleton records

Due to this, we recommend using the weekly data as the default

The Golden Rules of ACDS

The Golden Rules

Golden rules are a set of data standards developed to ensure that our analytic practices are accurate and consistent across the business, including Kroger.

Customer Exclusions

For analyses that include customer analysis, exclude non-card and store card transactions. While these are valid transactions, including these transactions would result in overstated “per household” metrics.

Product Exclusions

To align with how Kroger views their business, exclude products that do not align with Kroger reporting standards, such as gift cards, fuel, and pharmacy. Including would inflate results. Exclude non-customer items and empty items, these items will inflate results. Exclude undefined and miscellaneous UPCS, in order to have a true hierarchy.

Store Exclusions

Only include transactions from divisions that align with Kroger reporting standards. This will exclude divisions with low card-capture rates (as we cannot perform customer analysis) and divisions whose operations are not fully incorporated into Kroger.

Fuel Exclusions

Exclude fuel, by recap department rather than store. Food purchased at gas stations is part of customer’s behavior. Also, exclude third party fuel stores, as Kroger no longer owns these.

Use Effodata to consistently apply the Golden Rules in your analyses.

Standard Customer Exclusions

Customer Exclusion Golden Rules

Rule: If your analysis does not include a customer component (i.e. not calculating household metrics), include non-card and store card transactions. If your analysis does include a customer component, exclude non-card and store card transactions. These are automatically applied if you apply the “customer exclusions” in ACDS

Reason: Non-card and store card transactions do represent valid transactions; however, these transactions cannot be tied back to a specific household. Including these transactions in an analysis with a customer component would result in overstated “per household” metrics.



Standard Store Exclusions

Store Exclusion Documentation



Rule: For standard analyses, only include transactions from divisions that align with Kroger reporting standards. This will result in the exclusion of divisions with low card-capture rates (e.g. Food 4 Less, Ruler)

Reason: Our analyses should align as closely as possible to how Kroger views their business. Additionally, for divisions with low card-capture rates (or no loyalty card at all), it is not feasible to perform any customer analysis.

Geographic Division and Kroger FCs (Ocado)



In the summer of 2021, in partnership with UK based company Ocado, Kroger launched automated warehouses to fulfill pickup and delivery orders. Ocado transactions are all part of management division 540, 541, and 542.

Rule: For standard analyses, use geographic division (geo_div_no) to identify and include or exclude division.

Reason: Ocado (mgt_div_no 540) represents many geographic areas, however we do reporting as a customer would experience shopping.

For more information and updates, refer to [this Confluence page](#).

Standard Product Exclusions

Product Exclusion Documentation

Rule: For standard analyses, exclude products that do not align with Kroger reporting standards. These include products such as gift cards, fuel, and pharmacy.

Reason: Our analyses should align as closely as possible to how Kroger views their business, and including these items will inflate results. For example, with vendor gift card sales (e.g. Macy's, iTunes, etc.), Kroger does not receive the full proceeds of the card.



PRODUCT LIST GENERATION

UPC Input Package

Rule: For analyses based on an input UPC list, use the UPC Input Package to generate the appropriate mappings.

Reason: This is a standard way of mapping a UPC list to our data and will allow you to see what gets mapped and what does not. This can then be communicated back to the CSA and client.

Note: Make sure the start and end dates entered into the tool cover the entirety of your analysis period. If you cut it short (e.g. not providing enough time to account for a pre period), you will only include a subset of product matching. This will result in spikes/drop-offs in metrics over time.

Standard Fuel Exclusions

Fuel Exclusion Documentation

Rule: For standard analyses, no want to include fuel in analysis

Reason: Kroger does not include fuel in its reporting and fuel prices fluctuate quite a bit.

Note, Fuel Exclusions are also incorporated into Product and Store Exclusions where applicable.

:



Joining to Transaction

Effodata Documentation

Rule: When performing joins onto transactions tables, an ACDS query etc (or any tables for that matter), always be sure to join on the distinct keys of your base tables.

Reason: Joining by non-distinct keys can cause duplication in your resulting datasets which can ultimately result in overstated calculations.

Pro-tip: Use Effodata to join dimension records onto fact tables to ensure you are following best practice.

Calculating Basic Metrics and Returns

KPI Aggregator Documentation

Rule: When calculating sales and units, always calculate the net amount, taking negative values (i.e. returns) into consideration. When calculating visits and households, do not take negative or zero values into consideration.

Reason: Though rare, returns do happen and should be accounted for when calculating sales and units. The return transaction, however, should not be counted as it is not a true visit (and would inflate that value).

Pro-tip: Use KPI Aggregator to calculate basic metrics and ensure you align to standard metric definitions. There are parameters to appropriately handle return transactions.

As Is/As Was/As Of

Documentation

We have three dimensions in our transaction data that have relationships that can change over time: products (Kroger hierarchy), stores (division hierarchy), and households (card to household relationship).

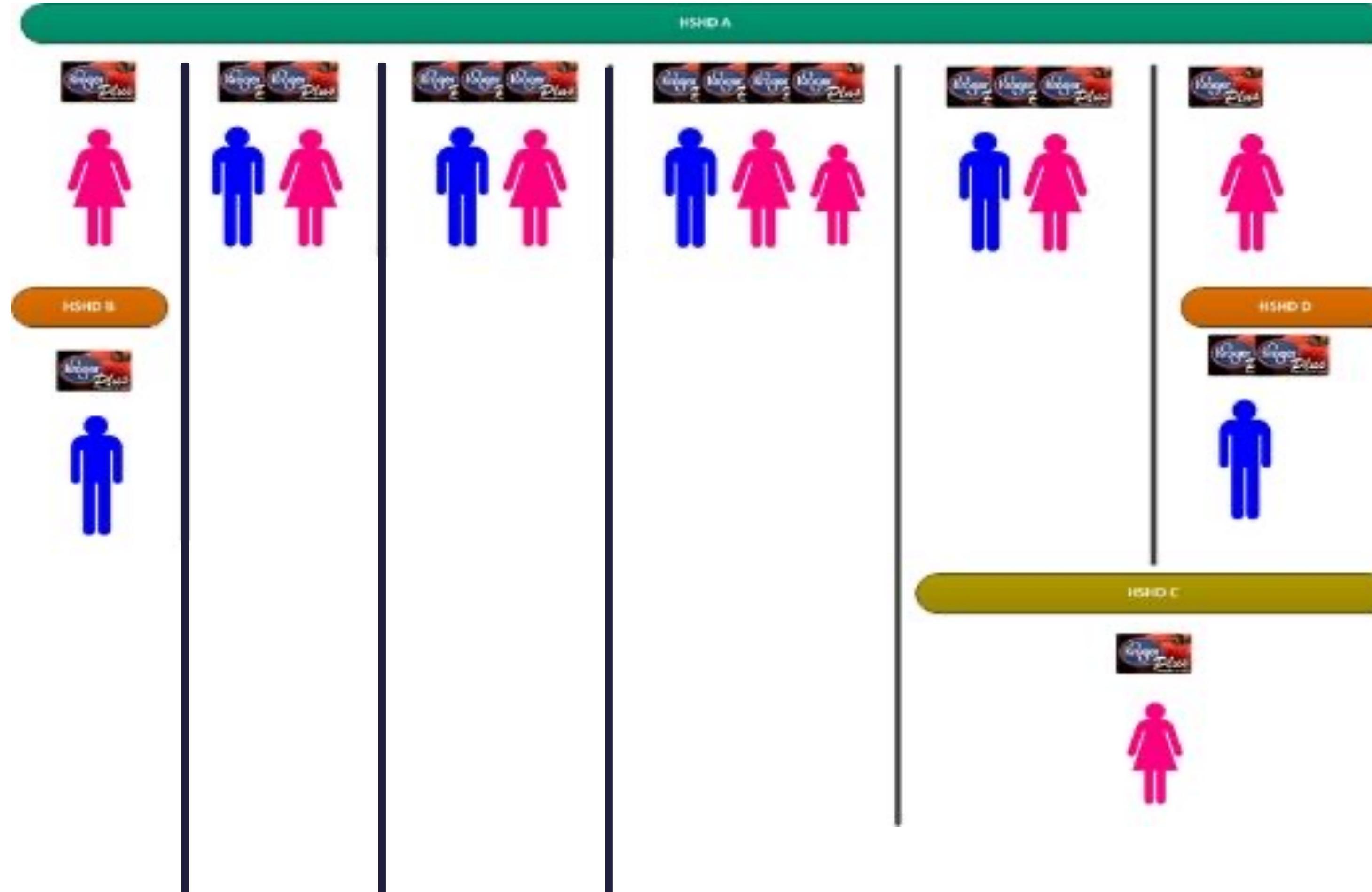
We have three ways in which we can handle these changes:

- As Is: uses the most current relationship ('as is' today) for all reporting
- As Of: choose any point in time and use the relationship 'as of' that point (what was used in the Exadata CDM)
- As Was: uses what the relationship was at the time of the transaction

Dimension	Certified Relationship	Not Certified/Not Supported
Household	AS WAS	AS IS/AS OF
Product	AS IS	AS WAS/AS OF
Store	AS IS	AS WAS/AS OF

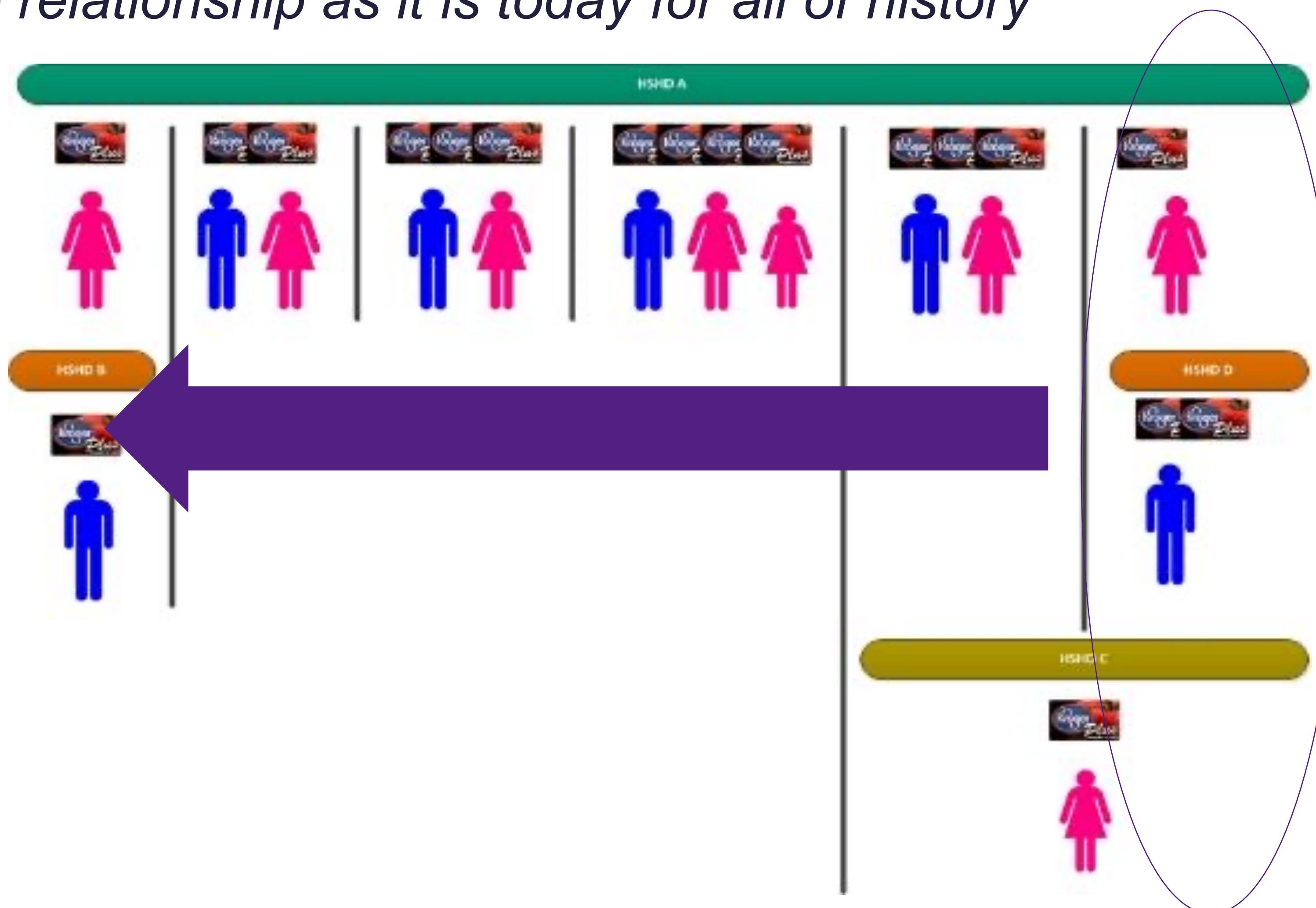
EXAMPLE HOUSEHOLD OVER TIME

Days of Our Lives Edition



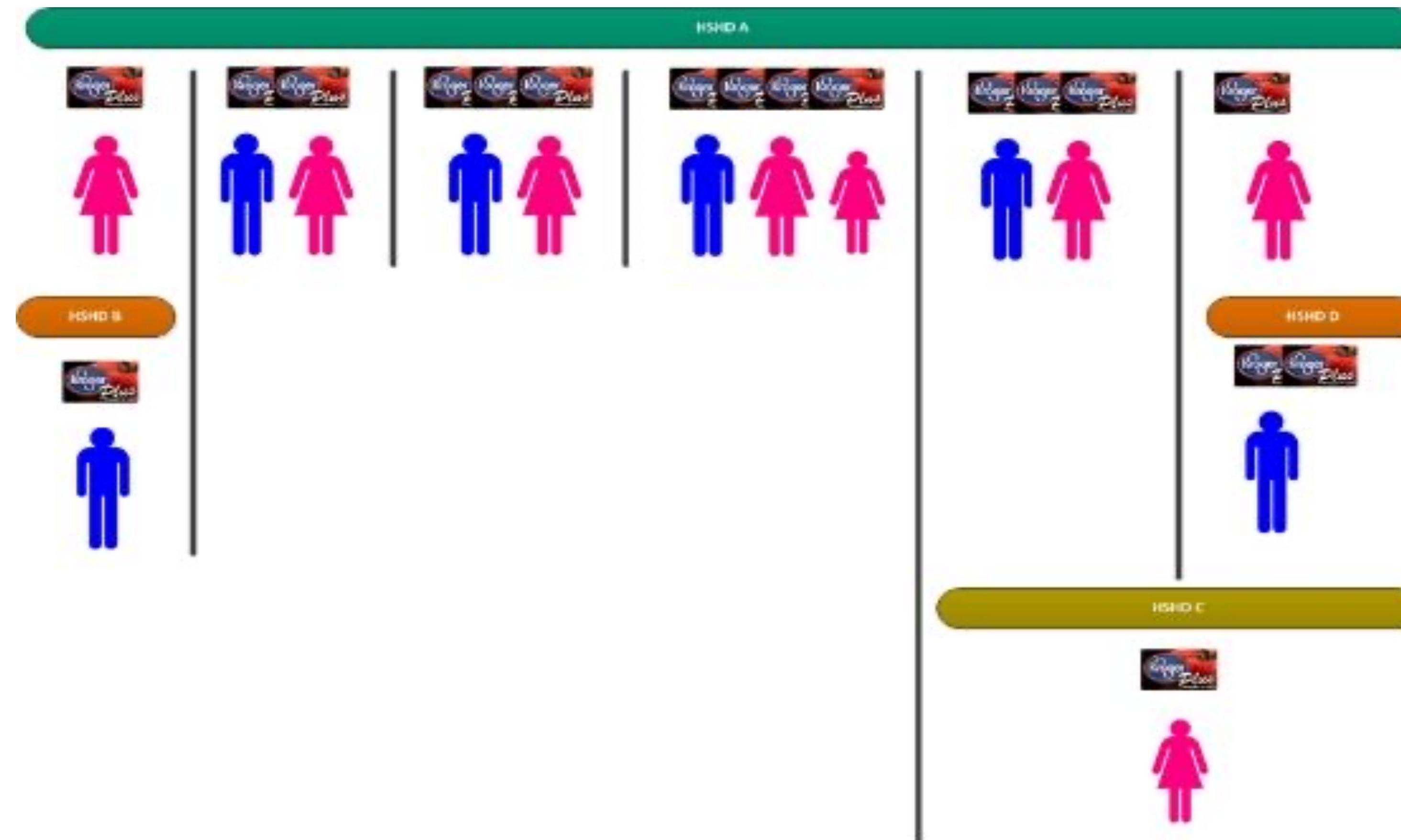
AS IS SCENARIO

Uses the relationship as it is today for all of history



AS WAS SCENARIO – THIS IS WHAT ACDS DOES

Uses the relationship as it was at the time of the transaction



Segmentations

Seg Package Documentation

Rule: For standard segmentations, use the Seg package.

Reason: the Seg package only has mainly certified segmentations, thus ensuring you are using the trusted segmentations that have passed Steward muster.

Rule: For standard analyses, left join segmentation information onto households so that households with missing information are not excluded from results.

Reason: Customer behavior is valid regardless of whether or not the customer has an established segmentation.

Preferred Store

Rule: Do not use preferred store as a way to include or exclude transactions and do not aggregate data by preferred store.

Reason: Using preferred store to extract transactions will exclude sales that households have outside of their preferred store. Additionally, aggregating metrics based on preferred store will result in sales being falsely assigned to stores. Both will lead to a misrepresentation of household behavior at Kroger.

Note: Use of preferred store is acceptable when selecting households for projects specific to stores or geographic locations. Preferred Store is available using the Seg package.

ID Stores

<https://krogerpedia.kroger.com/asset/73b52877-98f0-43eb-bfc2-c6541da88627?type=dataset2>

Rule: When conducting YOY analyses for Kroger, limit the store universe to ID stores. When conducting YOY analyses for brands and CPGs, do not limit to ID stores (proprietary to Kroger).

Reason: In order to provide an accurate assessment of changes in key metrics over time, it is necessary to follow the same group of stores for the entire time period of interest (similar to continuous panel but for stores).

Note: For time periods longer than one quarter, Kroger calculates results at the quarter/period level using the 5-quarter flag and aggregates them together. For example, to calculate IDs for 2016 FY, transactions must be limited to IDs for each of the four individual quarters, and then aggregated together.

STORE LEVEL ANALYSES EXCLUSION

Rule: We now have a Micro Fulfillment Center that will come through as a “normal” store in our data (014_00789), having opened on January 21st 2019. For store level analyses and reporting, exclude this store. For other analyses (ex. any customer level analysis), include the store’s data.

Reason: Store 014_00789 is special in that it’s not a traditional store. It is a special store meant only to service Clicklist / Pickup orders – a customer cannot actually shop it. However, these are still valid transactions in our data and should be included in a customer’s behavior. However, for store level analyses and reporting, the store isn’t directly comparable to other stores in the division since it can only handle Pickup orders.

Modality and Other E-Commerce Fields

[Confluence Documentation](#)

[Effodata Example Docs](#)

We should use modality to instore, pickup, and delivery transactions. Historically we used order_type to do so, but this field will be deprecated in May 2023.

Customer_fulfillment and application_name are also available to provide more granularity, but should not be used for commercial deliverables.

Golden Rules and Modality

The golden rules can be applied to digital transactions in the same way they're applied to normal transactions. One thing to note is that delivery and pickup fees are filtered out by golden rules. They fall under recap department 34, which gets filtered out by pid_fyt_rec_dpt_cd_excl: pid_fyt_rec_dpt_cd in ('30', '34', '58', '78')

You can learn more about the business logic of these golden rules at <https://github.8451.com/FoundationalComponents/GoldenRules>.

```
5]: df_ecomm_transactions = acds.get_transactions(start_date="20210501",
                                                end_date="20210507",
                                                apply_golden_rules=golden_rules(),
                                                query_filters=[ "modality in ('PICKUP', 'DELIVERY')",
                                                                "customer_fulfillment in ('INSTORE', 'WALGREENS', 'INSTACARTSEAMLESS')"])
```

```
5]: # Any fees after applying golden rules?
df_ecomm_transactions.filter(f.col('gtin_no').isin(fee_gtin_all)).count()
```

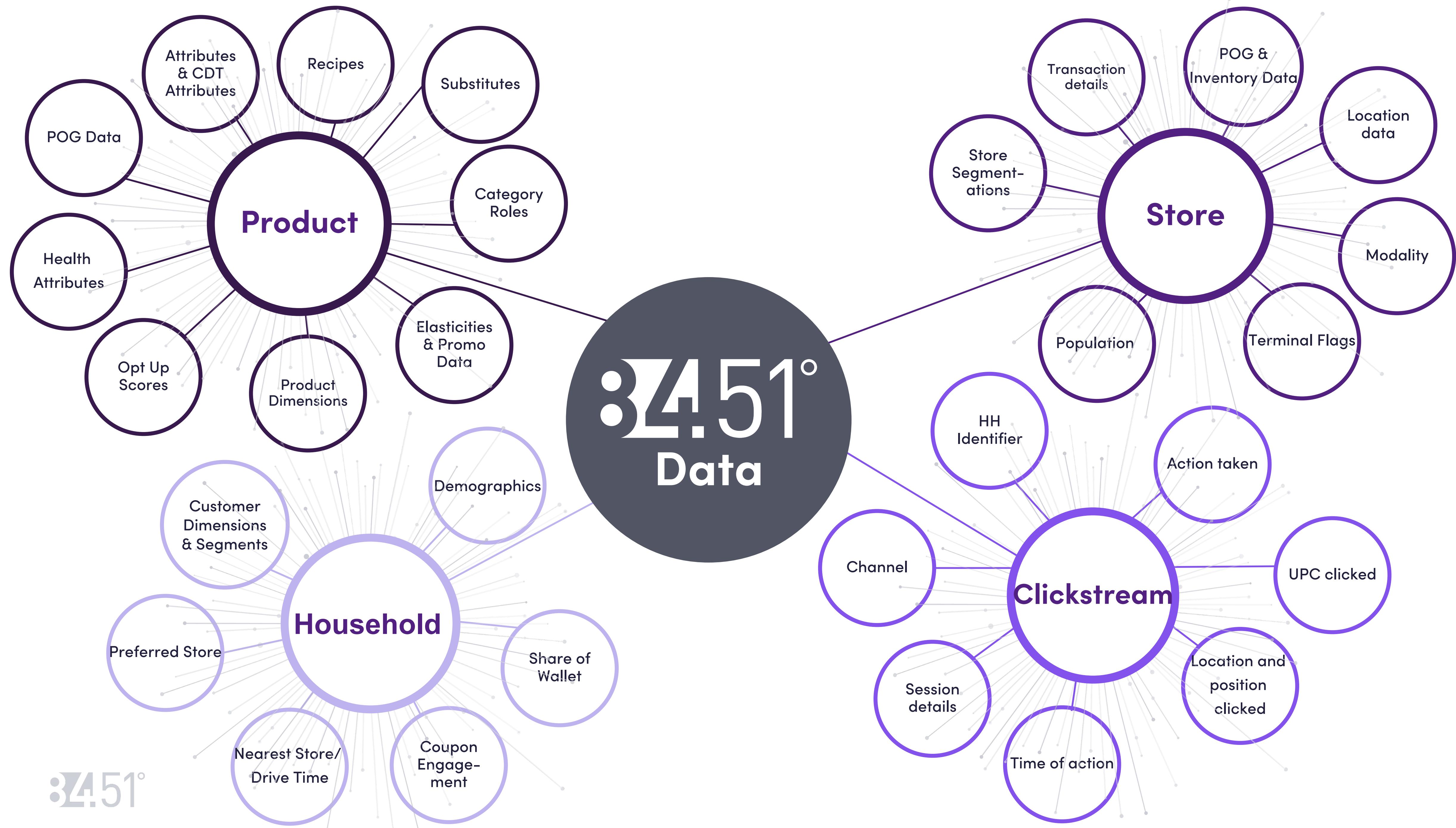
STORE LEVEL ANALYSES EXCLUSION FOR PICKUP AND DELIVERY

Rule: There are 3 stores (014_00359, 014_00901, 014_00789) that should be EXCLUDED from store level Pickup and Delivery analyses but should be INCLUDED in other (ex. Household) Pickup and Delivery analyses

Reason: Store 014_00789 is our Micro Fulfillment Center and 014_00359 and 014_00901 are stores where Pickup transactions can occur, but customers cannot actually pick up an order at these stores. However, these are still valid transactions in our data and should be included in a customer's behavior. However, for store level analyses and reporting, these stores should be excluded since they are considered as different by Kroger

84.51 Data Landscape

84.51° Data



84.51°

Discovering Data in Krogerpedia

Azure Learning Center Documentation

Krogerpedia is a data catalog application developed by KTD and is a place for you to find the data you need. Krogerpedia is intended to be used by both technical and business users to help answer the following questions:

- What data does Kroger and/or 84.51° have?
- Where can I go to find the data I need?
- Who are the appropriate contacts for this data?
- When there are multiple sources of data, which one should I use?
- Is this data fit for my purpose?

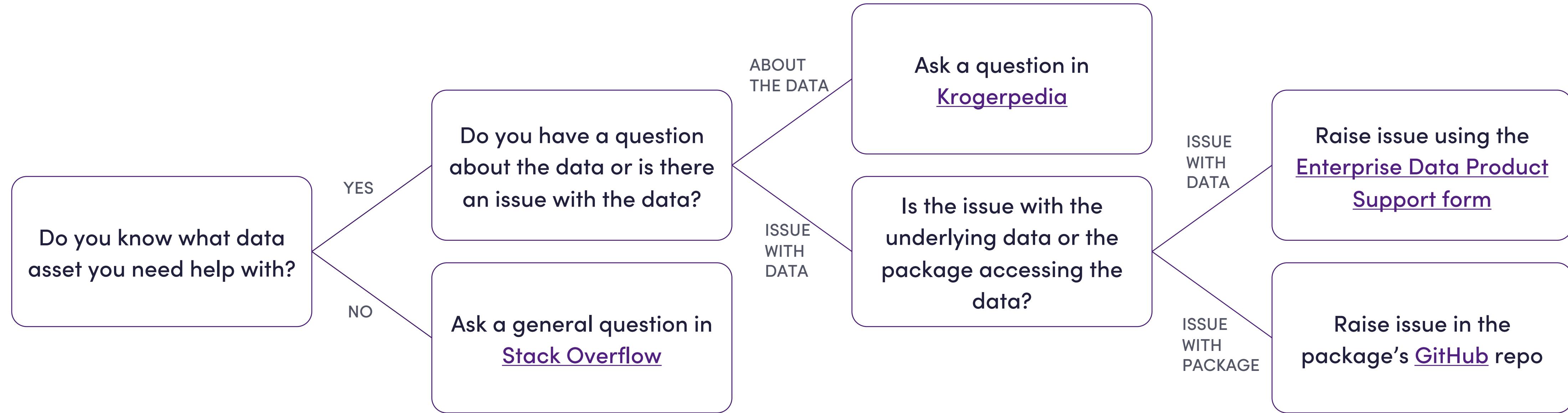
Visit Krogerpeida to discover additional data assets that may be helpful to your projects!

If you are a producer of data that others will consume, you should register the data in Krogerpedia. See the Learning Center docs for more information!

The home-made application will be transitioning to Alation in 2023. More details to come!

Other Data Foundations

Getting Support



*This support pattern applies to all EC Data Products and Services

Appendix

Data Product Pattern

Security

Saving and sharing files