

DATA MODEL

E-COMMERCE

PLATFORM

by Waiwit Charoenwiatpong 6410414003

AGENDA

01

E-COMMERCE PLATFORM,
KPI, AND REQUIREMENTS

02

OPERATIONAL
DATA

03

ANALYTICS DOMAIN

04

PIPELINE DESIGN,
CLASS DIAGRAM, AND
IMPLEMENTATION

05

DEMO

06

CONCLUSION

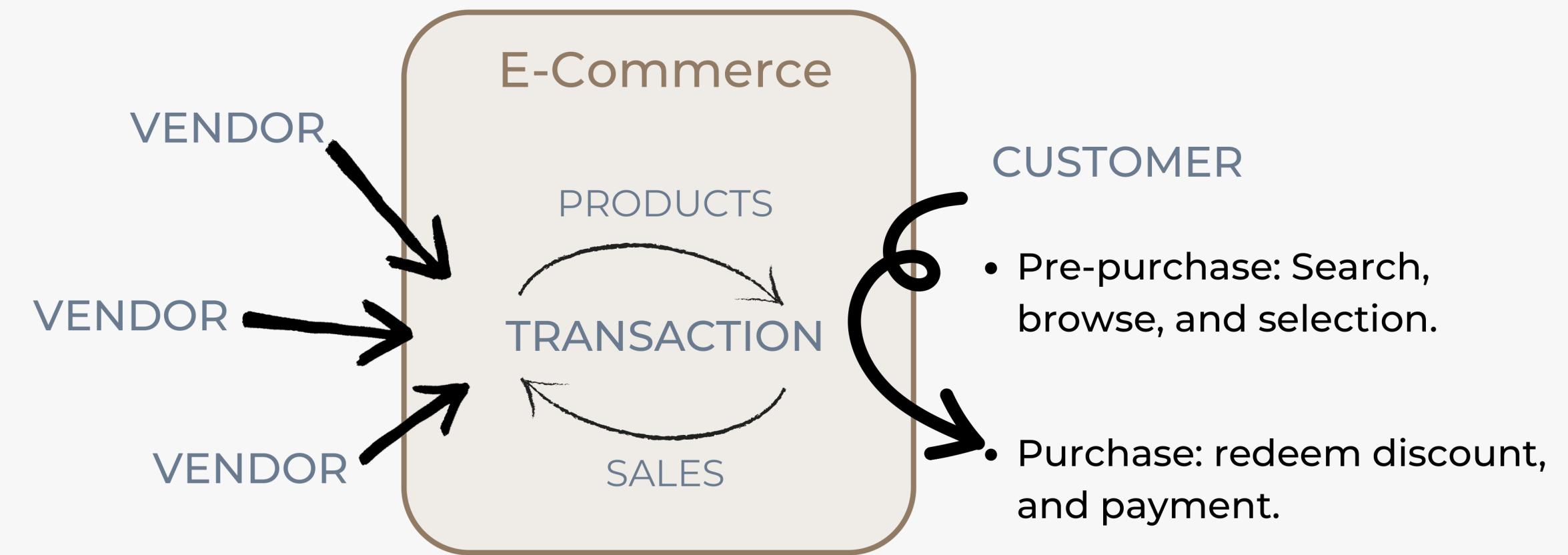
E-COMMERCE PLATFORM



E-COMMERCE PLATFORM

Key characteristics

- Mimic characteristics of marketplace
- Capture customers purchasing behaviors
- Capture transactions between vendors and customers



PLATFORM FEATURES & WORKFLOW

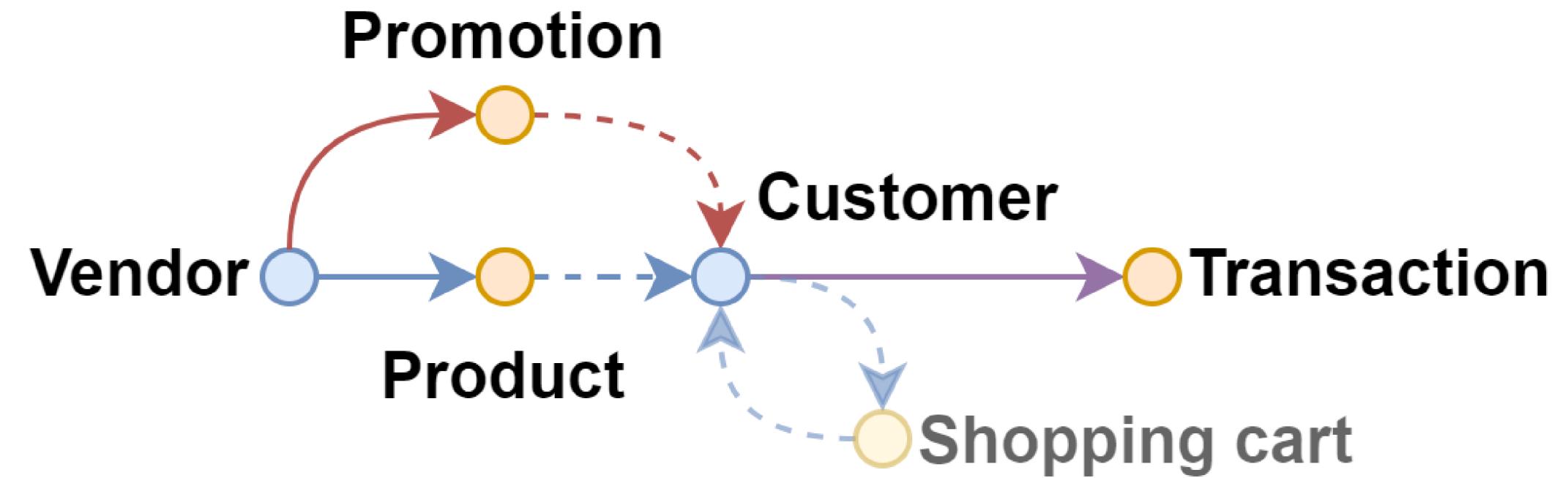
Key Features

Vendor:

- Vendor info
- Inventory management
- Promotion offering

Customer

- Customer info
- Shopping transactions
- Redeemed transactions



○ Actor

○ Entity

→ Dataflow of product

→ Dataflow of promotion

→ Dataflow of product & promotion

BUSINESS METRICS AND KPIS

Economy

Measure economic growth.

- Market size: number of vendors and customers
- Value of trading in platform
- Retention cost: e.g., cost of promotion

Vendor

Monitor growth potential.

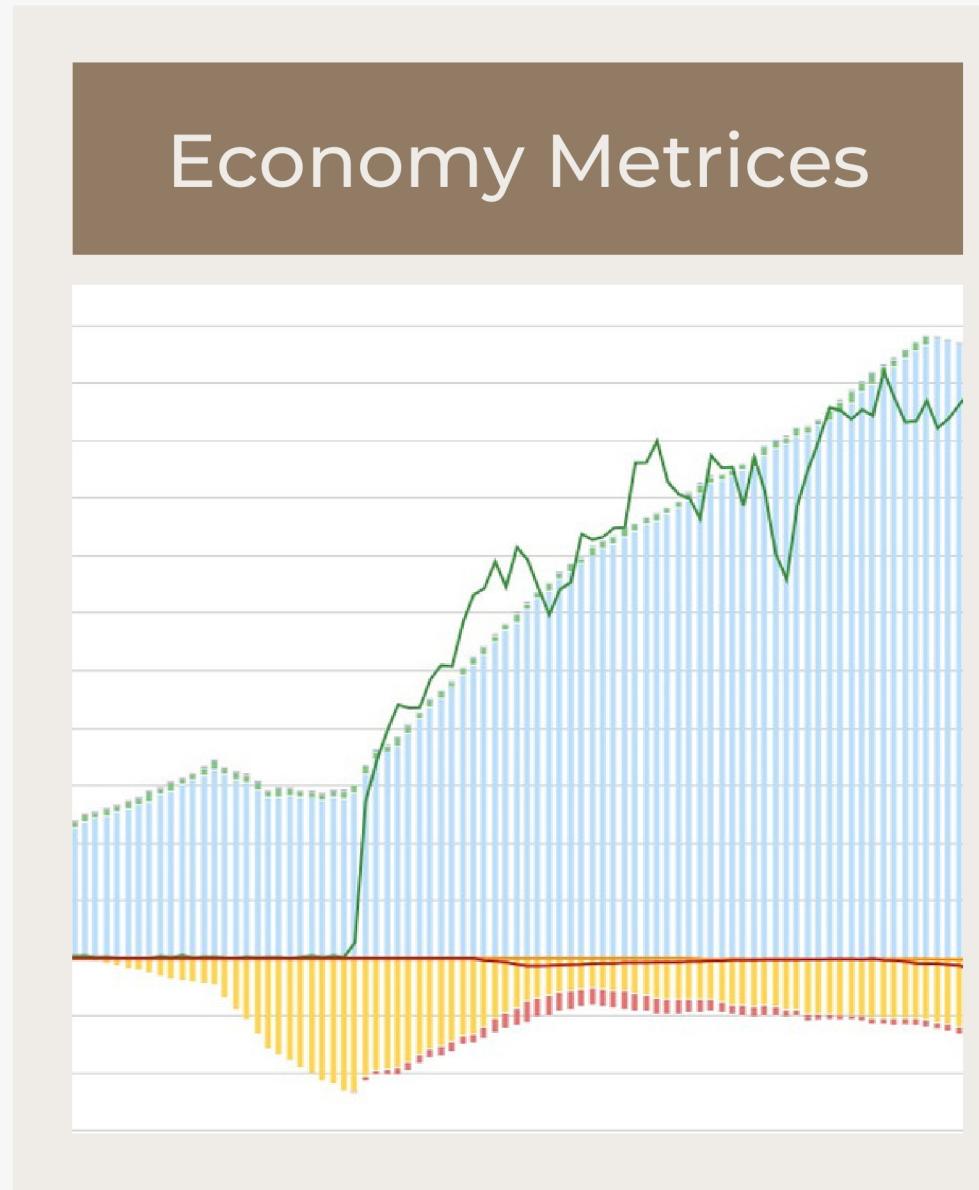
- Daily income and moving averages.
- Trend of sold products and new products.
- Reliability: rating and comment

Customer

Find potential customers.

- Classify customers from behaviors - e.g., purchasing amount, gap-between-purchase, and frequently purchased products

BUSINESS REPORT - SAMPLE





BUSINESS REQUIREMENTS

Design data models for operational data and for data analysis; and implement data pipeline (POC).



Operational Data Domain

Design data model that capture data related to business activities: vendor info, customer info, sales products, transaction.



Data Analytics Domain

Design data model that supports business metrics and KPIs: platform economy, vendor profile, customer profile



Implement Data Pipeline

Design & implement data pipeline from-end-to-end (ingestion to reporting)

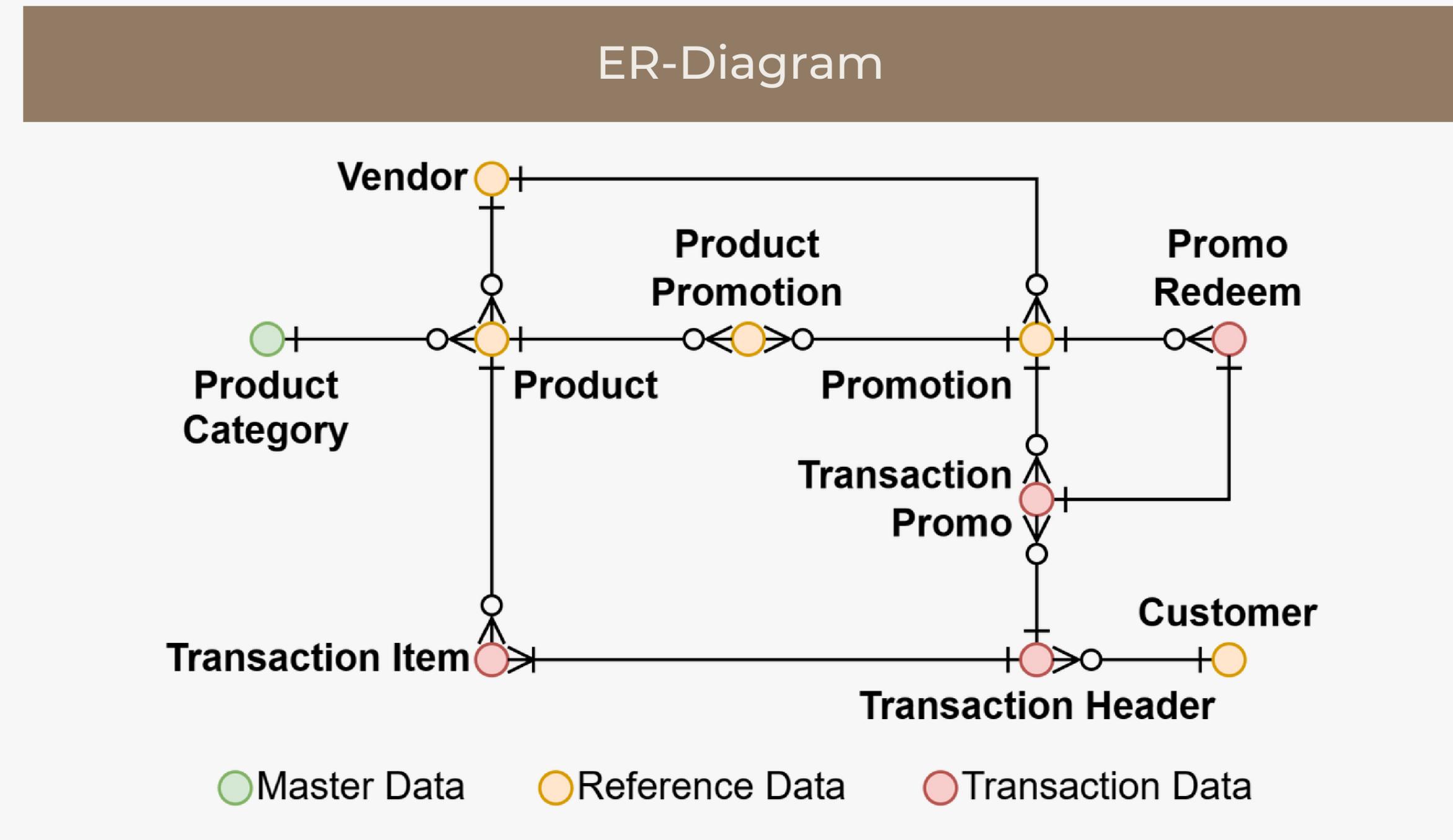
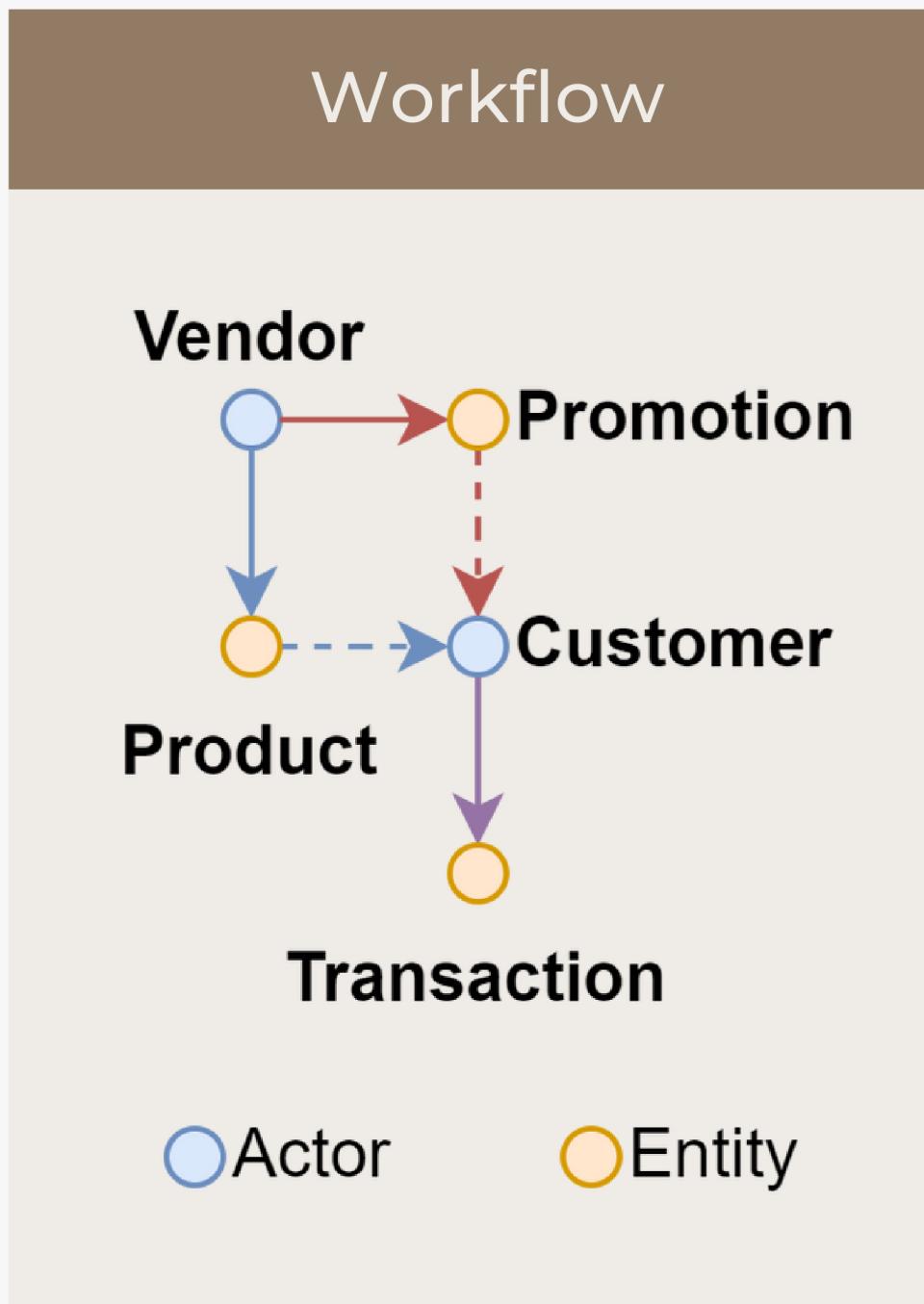


OPERATIONAL DATA

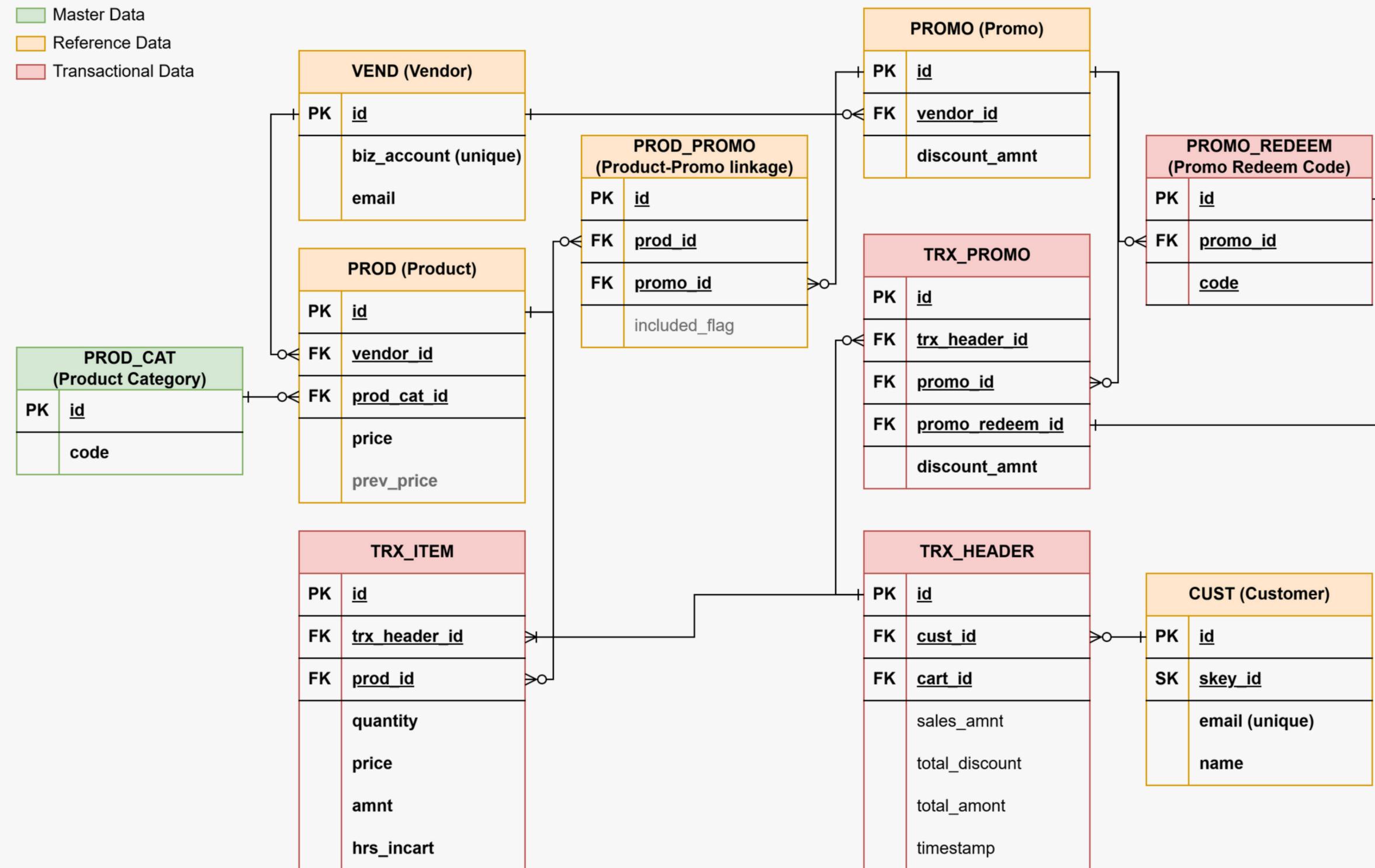
Data modeling design and ER-Diagram that represents data associated with trading on platform



OPS. WORKFLOW AND ER-DIAGRAM



Master Data
 Reference Data
 Transactional Data



Operational Data ER-Diagram: Reveal relationship and some fields of data entities.

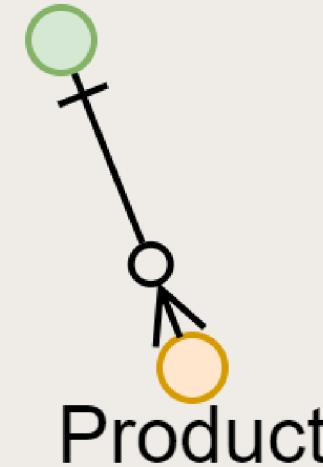
PRODUCT CATEGORY

- DB Entity Name:

PROD_CAT

- Relation with others:

Product Category



Attr.	Key	Type	Description
id	PK	string	Product category id - 'CAT0000'
code		string	Category code - 'electronics.smartphone'
label		string	Category label - 'Smartphone'
desc		string	Category description
created_date		date	Created date (YYYY-MM-DD) - '2022-11-01'
updated_date		date	Updated date (YYYY-MM-DD) - '2022-11-01'
deactivated_flag		bool	Flag deactivated category - 'FALSE'

PRODUCT CATEGORY (CONT'D)

- Data Sample

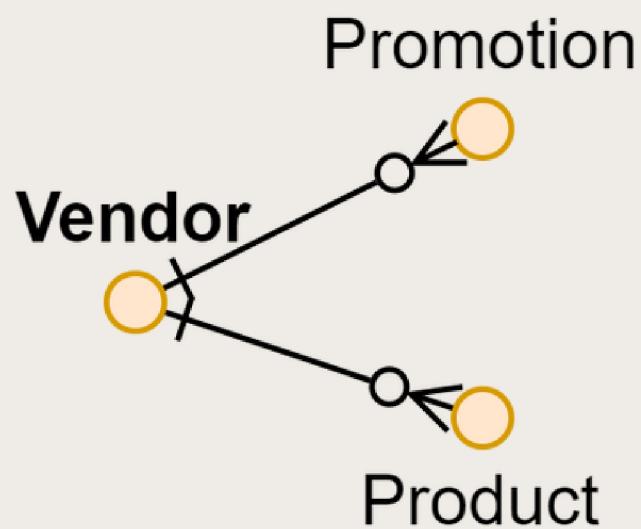
id	code	label	desc	created_dt	updated_dt	deactivated_flag
CAT0001	accessories.bag	Bag	Accessories bag	2022-11-01	2022-11-01	FALSE
CAT0002	accessories.umbrella	Umbrella	Accessories umbrella	2022-11-01	2022-11-01	FALSE
CAT0003	accessories.wallet	Wallet	Accessories wallet	2022-11-01	2022-11-01	FALSE
CAT0004	apparel.belt	Belt	Apparel belt	2022-11-01	2022-11-01	FALSE
CAT0005	apparel.costume	Costume	Apparel costume	2022-11-01	2022-11-01	FALSE
CAT0006	apparel.dress	Dress	Apparel dress	2022-11-01	2022-11-01	FALSE
CAT0007	apparel.jacket	Jacket	Apparel jacket	2022-11-01	2022-11-01	FALSE
CAT0008	apparel.scarf	Scarf	Apparel scarf	2022-11-01	2022-11-01	FALSE
CAT0009	apparel.shirt	Shirt	Apparel shirt	2022-11-01	2022-11-01	FALSE
CAT0010	apparel.shoes	Shoes	Apparel shoes	2022-11-01	2022-11-01	FALSE
CAT0011	apparel.shoes.ballet	Ballet Shoes	Apparel shoes ballet	2022-11-01	2022-11-01	FALSE

VENDOR

- DB Entity Name:

VEND

- Relation with others:



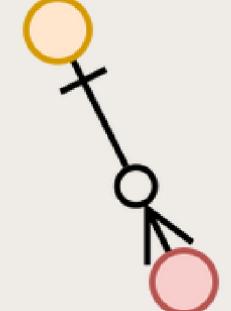
Attr.	Key	Type	Description
id	PK	string	Vendor id - 'VEND0000'
biz_account		string	Vendor's unique business account
email		string	Vendor's business email
biz_alias		string	Vendor's unique business alias
official_flag		bool	Flag official vendor - 'FALSE'
registered_dt		datetime	Registered DT (YYYY-MM-DD h24:mm:ss)
updated_dt		datetime	Updated DT (YYYY-MM-DD h24:mm:ss)

VENDOR (CONT'D)

- Data Sample

id	biz_account	email	biz_alias	prev_biz_alias	phone	street_addr	state	zip_code	country	prev_addr	official_flag	registered_dt	updated_dt
VEND0001	Kirlin_1	Kirlin_1@vend.com	Kirlin and Sons		843138706	351 Darlene Green	Queensland	9209	Australia		TRUE	2022-11-04 4:10:44	2022-11-04 4:10:44
VEND0002	sporer_2	sporer_2@vend.com	Sporer-Keebler		657402165	3638 Marsh Union	Queensland	4002	Australia		TRUE	2022-11-07 21:38:27	2022-11-07 21:38:27
VEND0003	swaniawski_3	swaniawski_3@vend.com	Swaniawski, Nitzsche and Welch		860238257	9333 Valentine Poin	New South Wales	1710	Australia		FALSE	2022-11-08 2:58:49	2022-11-08 2:58:49
VEND0004	haley_4	haley_4@vend.com	Haley Group		431665303	32941 Krystal Mill Ap	New South Wales	2780	Australia		FALSE	2022-11-08 16:05:48	2022-11-08 16:05:48
VEND0005	johnston_5	johnston_5@vend.com	Johnston-Casper		962138374	5783 Evan Roads Ap	Queensland	9632	Australia		FALSE	2022-11-08 19:29:08	2022-11-08 19:29:08
VEND0006	daugherty_6	daugherty_6@vend.com	Daugherty LLC		17235499	76752 David Lodge	Queensland	4816	Australia		FALSE	2022-11-13 19:11:38	2022-11-13 19:11:38
VEND0007	romaguera_7	romaguera_7@vend.com	Romaguera Ltd		618691259	010 Weaver Land	South Australia	5528	Australia		FALSE	2022-11-15 7:42:25	2022-11-15 7:42:25
VEND0008	reichel_8	reichel_8@vend.com	Reichel LLC		358772912	350 Stacy Glens	Tasmania	7374	Australia		FALSE	2022-11-16 18:42:12	2022-11-16 18:42:12
VEND0009	goyette_9	goyette_9@vend.com	Goyette, Howell and Collier		433095629	4138 David Fall	Western Australia	6858	Australia		TRUE	2022-11-16 22:12:50	2022-11-16 22:12:50
VEND0010	kilback_10	kilback_10@vend.com	Kilback Group		931703207	7921 Robert Port Su	Western Australia	6678	Australia		FALSE	2022-11-17 13:55:18	2022-11-17 13:55:18
VEND0011	feil_11	feil_11@vend.com	Feil, Hilpert and Koss		996888716	43235 Mckenzie Vie	New South Wales	2077	Australia		FALSE	2022-11-18 0:40:26	2022-11-18 0:40:26
VEND0012	gottlieb_12	gottlieb_12@vend.com	Gottlieb Group		234918950	45654 Hess Rest	New South Wales	2514	Australia		FALSE	2022-11-18 13:56:50	2022-11-18 13:56:50
VEND0013	connelly_13	connelly_13@vend.com	Connelly-Carter		674063249	244 Abbott Parkway	Victoria	3470	Australia		TRUE	2022-11-19 11:46:52	2022-11-19 11:46:52
VEND0014	bechtelar_14	bechtelar_14@vend.com	Bechtelar-Rippin		177016567	977 Rita Grove Apt.	New South Wales	2575	Australia		FALSE	2022-11-20 10:17:54	2022-11-20 10:17:54
VEND0015	lubowitz_15	lubowitz_15@vend.com	Lubowitz-Walter		628639996	4038 Smith Avenue	Western Australia	6749	Australia		FALSE	2022-11-21 20:40:00	2022-11-21 20:40:00
VEND0016	welch_16	welch_16@vend.com	Welch, Rath and Koepp		765449869	5097 Jodi Vista Suite	New South Wales	2725	Australia		TRUE	2022-11-22 4:40:37	2022-11-22 4:40:37
VEND0017	hickle_17	hickle_17@vend.com	Hickle Group		560454237	838 Franklin Prairie	Victoria	3040	Australia		FALSE	2022-11-24 17:18:50	2022-11-24 17:18:50
VEND0018	lang_18	lang_18@vend.com	Lang, Towne and Schuppe		786198321	03030 White Lakes	Western Australia	6050	Australia		FALSE	2022-11-25 4:37:53	2022-11-25 4:37:53
VEND0019	morissette_19	morissette_19@vend.com	Morissette LLC		618443244	65417 Walsh Radial	New South Wales	2774	Australia		TRUE	2022-11-25 12:48:22	2022-11-25 12:48:22
VEND0020	prosacco_20	prosacco_20@vend.com	Prosacco LLC		848798080	444 Robert Mews	Western Australia	6114	Australia		TRUE	2022-11-25 17:24:53	2022-11-25 17:24:53
VEND0021	corwin_21	corwin_21@vend.com	Corwin-Romaguera		452667626	9345 Spencer Junct	New South Wales	2009	Australia		TRUE	2022-11-26 2:59:28	2022-11-26 2:59:28
VEND0022	tillman_22	tillman_22@vend.com	Tillman LLC		638447911	28711 Kristine Junct	South Australia	5331	Australia		FALSE	2022-11-26 12:40:26	2022-11-26 12:40:26
VEND0023	veum_23	veum_23@vend.com	Veum-Koelpin		101193493	181 Moreno Light Ap	Queensland	4487	Australia		FALSE	2022-11-28 12:32:05	2022-11-28 12:32:05

CUSTOMER

• DB Entity Name:	CUST
• Relation with others:	<p>Customer</p>  <pre> classDiagram class Customer class TransactionHeader Customer "1" -- "0..1" TransactionHeader </pre>
	Transaction Header

Attr.	Key	Type	Description
id	PK	string	Customer id - 'CID0000'
skey_id	SK	index	Customer surrogate key
email		string	Customer's email
name		string	Customer name
registered_dt		datetime	Registered DT (YYYY-MM-DD h24:mm:ss)
updated_dt		datetime	Updated DT (YYYY-MM-DD h24:mm:ss)
validated_flag		bool	Flag validated records - 'TRUE'

CUSTOMER (CONT'D)

- Data Sample

Id	skey_id	email	phone	name	gender	dob	street_addr	zip_code	state	country	alt_email	registered_dt	updated_dt	validated_flag
CID00001	1	Bockman_1@mail.com	229116393	Eli Bockman	Male	1980-12-16	6 Meadow Vale Court	2153	New South Wales	Australia		2022-12-27 0:09:28	2022-12-27 0:09:28	TRUE
CID00002	2	Calton_2@mail.com	357303004	Sheila-kathryn Calton	Female	1977-05-13	17979 Del Mar Point	2448	New South Wales	Australia		2022-12-27 0:37:38	2022-12-27 0:37:38	TRUE
CID00003	3	Merali_3@mail.com	359821528	Fina Merali	Female	1976-02-23	4 Delaware Trail	2210	New South Wales	Australia		2022-12-27 0:40:27	2022-12-27 0:40:27	TRUE
CID00004	4	Lind_4@mail.com	359191980	Mala Lind	Female	1973-03-10	97736 7th Trail	2023	New South Wales	Australia		2022-12-27 2:07:59	2022-12-27 2:07:59	TRUE
CID00005	5	Flattman_5@mail.com	352682613	Sawyer Flattman	Male	1994-07-21	44339 Golden Leaf Avenue	4557	Queensland	Australia		2022-12-27 3:57:24	2022-12-27 3:57:24	TRUE
CID00006	6	Kitteman_6@mail.com	304076754	Rayshell Kitteman	Female	1983-03-25	9 McBride Trail	2760	New South Wales	Australia		2022-12-27 4:01:26	2022-12-27 4:01:26	TRUE
CID00007	7	Radage_7@mail.com	213180742	Erroll Radage	Male	2000-07-13	9861 New Castle Avenue	2428	New South Wales	Australia		2022-12-27 4:30:48	2022-12-27 4:30:48	TRUE
CID00008	8	Parr_8@mail.com	358928994	Harlin Parr	Male	1977-02-27	52 Moland Street	3331	Victoria	Australia		2022-12-27 5:22:51	2022-12-27 5:22:51	TRUE
CID00009	9	Neasham_9@mail.com	359635727	Marjie Neasham	Female	1967-07-06	092 2nd Alley	2135	New South Wales	Australia		2022-12-27 6:03:16	2022-12-27 6:03:16	TRUE
CID00010	10	Keyson_10@mail.com	354689763	Sorcha Keyson	Female	2001-04-15	59 Spaight Circle	2233	New South Wales	Australia		2022-12-27 6:05:30	2022-12-27 6:05:30	TRUE
CID00011	11	Firth_11@mail.com	224254270	Basile Firth	Male	1980-08-13	032 Bartelt Crossing	2444	New South Wales	Australia		2022-12-27 9:43:35	2022-12-27 9:43:35	TRUE
CID00012	12	Cammocke_12@mail.com	571465035	Mile Cammocke	Male	1980-09-20	18 Jenna Center	2650	New South Wales	Australia		2022-12-27 10:42:13	2022-12-27 10:42:13	TRUE
CID00013	13	Skpsey_13@mail.com	659325070	Kim Skpsey	Female	1977-12-03	3 Cordella Plaza	2153	New South Wales	Australia		2022-12-27 10:53:15	2022-12-27 10:53:15	TRUE
CID00014	14	Assaf_14@mail.com	498830437	Geoff Assaf	Male	1976-12-02	28 5th Center	4413	Queensland	Australia		2022-12-27 11:19:33	2022-12-27 11:19:33	TRUE
CID00015	15	Ginnelly_15@mail.com	601150499	Trixi Ginnelly	Female	1978-06-10	52 Carey Alley	4740	Queensland	Australia		2022-12-27 11:25:25	2022-12-27 11:25:25	TRUE
CID00016	16	Klees_16@mail.com	457063652	Garvin Klees	Male	1978-09-25	96 Texas Plaza	3218	Victoria	Australia		2022-12-27 11:47:45	2022-12-27 11:47:45	TRUE
CID00017	17	Zellmer_17@mail.com	490662865	Fee Zellmer	Male	1973-09-30	48 Eagan Avenue	4868	Queensland	Australia		2022-12-27 11:59:59	2022-12-27 11:59:59	TRUE
CID00018	18	Sancraft_18@mail.com	490884647	Mona Sancraft	Female	1968-06-22	9 Buell Park	4116	Queensland	Australia		2022-12-27 13:07:16	2022-12-27 13:07:16	TRUE
CID00019	19	Vanichkin_19@mail.com	486131013	Marion Vanichkin	Female	1995-04-20	2951 Petterle Place	2756	New South Wales	Australia		2022-12-27 13:25:57	2022-12-27 13:25:57	TRUE
CID00020	20	Stonnell_20@mail.com	653844173	Lurette Stonnell	Female	1977-11-09	78 Del Sol Hill	2535	New South Wales	Australia		2022-12-27 14:51:12	2022-12-27 14:51:12	TRUE
CID00021	21	Dwerryhouse_21@mail.com	228374387	Laurie Dwerryhouse	Female	1985-12-22	54 Sage Plaza	4726	Queensland	Australia		2022-12-27 15:58:27	2022-12-27 15:58:27	TRUE
CID00022	22	Smalley_22@mail.com	456000414	Hunfredo Smalley	Male	1979-04-16	38726 Ilene Crossing	2212	New South Wales	Australia		2022-12-27 16:28:28	2022-12-27 16:28:28	TRUE
CID00023	23	Jerche_23@mail.com	456282700	Tomasine Jerche	Female	1981-10-27	3496 Brown Drive	2207	New South Wales	Australia		2022-12-27 17:05:01	2022-12-27 17:05:01	TRUE
CID00024	24	Coupe_24@mail.com	213153151	Basilius Coupe	Male	1976-04-14	7188 Cody Way	3350	Victoria	Australia		2022-12-27 18:48:33	2022-12-27 18:48:33	TRUE

PRODUCT

<ul style="list-style-type: none"> • DB Entity Name:
PROD
<ul style="list-style-type: none"> • Relation with others:
<pre> graph LR Product --> Vendor Product --> ProdCat[Product Category] Product --> TransItem[Transaction Item] Product --> Promotion Product --> Product2[Product] </pre> <p>The diagram illustrates the relationships of the Product entity. It is connected to four other entities: Vendor, Product Category, Transaction Item, and Promotion. Additionally, there is a self-referencing relationship indicated by a line connecting the Product entity to itself.</p>

Attr.	Key	Type	Description
id	PK	index	Product id
vendor_id	FK	string	Vendor id - 'VEND0000'
prod_cat_id	FK	string	Product category id - 'CAT0000'
name		string	Product name
brand		string	Product brand
desc		string	Product description
price		float64	price of product (#.##) - '100.00'

PRODUCT (CONT'D)

Attr.	Type	Description
features	json	Features of product
search_kw	json	Search key words
prev_price	float64	previous price (#.##) - '100.00'
created_dt	datetime	Created DT (YYYY-MM-DD h24:mm:ss)
updated_dt	datetime	Updated DT (YYYY-MM-DD h24:mm:ss)
deactivated_flag	bool	Deactivated product flag

PRODUCT (CONT'D)

- Data Sample

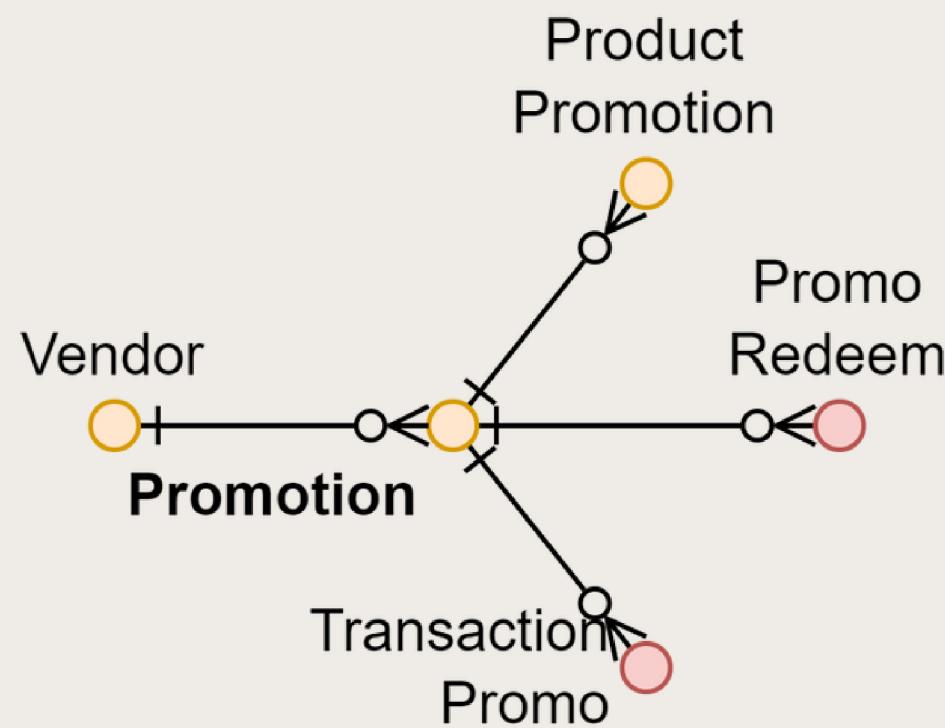
id	vendor_id	prod_cat_id	name	brand	desc	price	features	search_kw	prev_price	created_dt	updated_dt	deactivated_flag
1001588	VEND0002	CAT0094	Meizu Smartphone	meizu	Meizu Smartphone: e	128.31	{"brand": "meizu"}	[{"meizu", "electronics", "smartphone"}]		2022-11-15 0:01:01	2022-11-15 0:01:01	FALSE
1001589	VEND0003	CAT0094	Samsung Smartphone	samsung	Samsung Smartphone: e	77.14	{"brand": "samsung"}	[{"samsung", "electronics", "smartphone"}]		2022-11-15 0:06:30	2022-11-15 0:06:30	FALSE
1001590	VEND0001	CAT0094	Samsung Smartphone	samsung	Samsung Smartphone: e	409.02	{"brand": "samsung"}	[{"samsung", "electronics", "smartphone"}]		2022-11-15 0:14:20	2022-11-15 0:14:20	FALSE
1001591	VEND0003	CAT0094	Samsung Smartphone	samsung	Samsung Smartphone: e	370.41	{"brand": "samsung"}	[{"samsung", "electronics", "smartphone"}]		2022-11-15 0:25:41	2022-11-15 0:25:41	FALSE
1001592	VEND0002	CAT0094	Samsung Smartphone	samsung	Samsung Smartphone: e	409.02	{"brand": "samsung"}	[{"samsung", "electronics", "smartphone"}]		2022-11-15 0:43:25	2022-11-15 0:43:25	FALSE
1001593	VEND0003	CAT0094	Samsung Smartphone	samsung	Samsung Smartphone: e	409.02	{"brand": "samsung"}	[{"samsung", "electronics", "smartphone"}]		2022-11-15 0:58:57	2022-11-15 0:58:57	FALSE
1001594	VEND0004	CAT0094	Samsung Smartphone	samsung	Samsung Smartphone: e	414.53	{"brand": "samsung"}	[{"samsung", "electronics", "smartphone"}]		2022-11-15 1:13:03	2022-11-15 1:13:03	FALSE
1001595	VEND0005	CAT0094	Philips Smartphone	philips	Philips Smartphone: e	82.37	{"brand": "philips"}	[{"philips", "electronics", "smartphone"}]		2022-11-15 1:28:35	2022-11-15 1:28:35	FALSE
1001596	VEND0001	CAT0094	Meizu Smartphone	meizu	Meizu Smartphone: e	205.53	{"brand": "meizu"}	[{"meizu", "electronics", "smartphone"}]		2022-11-15 1:43:51	2022-11-15 1:43:51	FALSE
1001597	VEND0005	CAT0094	Sony Smartphone	sony	Sony Smartphone: e	275.55	{"brand": "sony"}	[{"sony", "electronics", "smartphone"}]		2022-11-15 2:00:40	2022-11-15 2:00:40	FALSE
1001598	VEND0002	CAT0094	Sony Smartphone	sony	Sony Smartphone: e	252.39	{"brand": "sony"}	[{"sony", "electronics", "smartphone"}]		2022-11-15 2:17:14	2022-11-15 2:17:14	FALSE
1001599	VEND0005	CAT0094	Meizu Smartphone	meizu	Meizu Smartphone: e	128.31	{"brand": "meizu"}	[{"meizu", "electronics", "smartphone"}]		2022-11-15 2:31:43	2022-11-15 2:31:43	FALSE
1001600	VEND0002	CAT0094	Meizu Smartphone	meizu	Meizu Smartphone: e	179.79	{"brand": "meizu"}	[{"meizu", "electronics", "smartphone"}]		2022-11-15 2:48:17	2022-11-15 2:48:17	FALSE
1001601	VEND0001	CAT0094	Sony Smartphone	sony	Sony Smartphone: e	306.42	{"brand": "sony"}	[{"sony", "electronics", "smartphone"}]		2022-11-15 3:02:28	2022-11-15 3:02:28	FALSE
1001602	VEND0001	CAT0094	Apple Smartphone	apple	Apple Smartphone: e	514.76	{"brand": "apple"}	[{"apple", "electronics"}]	515.67	2022-11-15 3:16:49	2023-02-20 9:04:11	FALSE
1001603	VEND0001	CAT0094	Apple Smartphone	apple	Apple Smartphone: e	721	{"brand": "apple"}	[{"apple", "electronics", "smartphone"}]		2022-11-15 3:33:12	2022-11-15 3:33:12	FALSE
1001604	VEND0005	CAT0094	Apple Smartphone	apple	Apple Smartphone: e	771.96	{"brand": "apple"}	[{"apple", "electronics", "smartphone"}]		2022-11-15 3:46:04	2022-11-15 3:46:04	FALSE
1001605	VEND0005	CAT0094	Apple Smartphone	apple	Apple Smartphone: e	643.23	{"brand": "apple"}	[{"apple", "electronics", "smartphone"}]		2022-11-15 3:58:03	2022-11-15 3:58:03	FALSE
1001606	VEND0004	CAT0094	Apple Smartphone	apple	Apple Smartphone: e	642.69	{"brand": "apple"}	[{"apple", "electronics", "smartphone"}]		2022-11-15 4:12:48	2022-11-15 4:12:48	FALSE
1001607	VEND0004	CAT0094	Apple Smartphone	apple	Apple Smartphone: e	864.63	{"brand": "apple"}	[{"apple", "electronics", "smartphone"}]		2022-11-15 4:26:41	2022-11-15 4:26:41	FALSE
1001608	VEND0004	CAT0094	Apple Smartphone	apple	Apple Smartphone: e	581.22	{"brand": "apple"}	[{"apple", "electronics"}]	581.48	2022-11-15 4:43:09	2023-02-21 4:44:11	FALSE

PROMOTION

- DB Entity Name:

PROMO

- Relation with others:



Attr.	Key	Type	Description
id	PK	index	Promotion id
vendor_id	FK	string	Vendor id - 'VEND0000'
label		string	Promotion Label
discount_amnt		float64	Discount amount
quota		int	Quantity redeem-code
lunch_schedule_dt		datetime	Start promotion datetime
expiration_dt		datetime	Expiration datetime of promotion

PROMOTION (CONT'D)

Attr.	Type	Description
created_dt	datetime	Created DT (YYYY-MM-DD h24:mm:ss)
updated_dt	datetime	Updated DT (YYYY-MM-DD h24:mm:ss)

PROMOTION (CONT'D)

- Data Sample

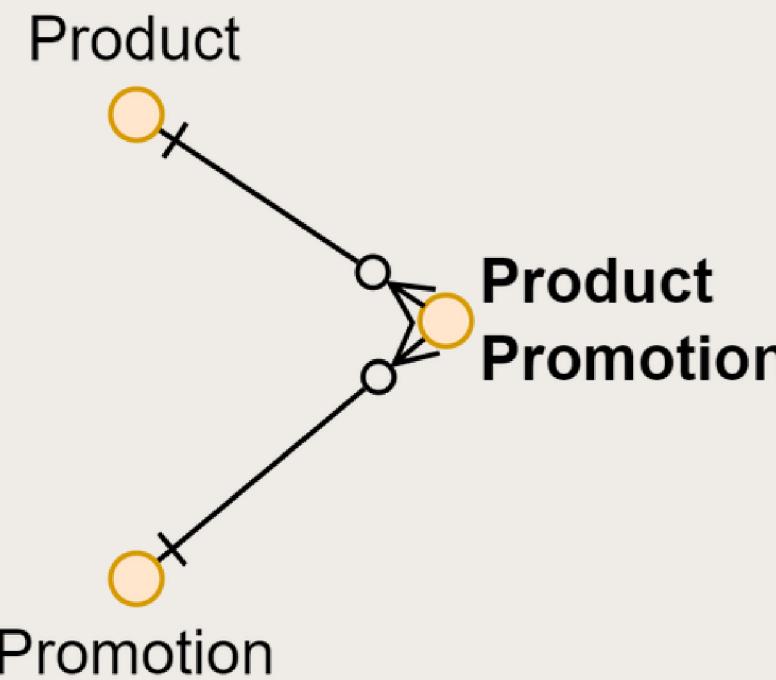
id	vendor_id	label	discount_amnt	quota	lunch_schedule_dt	expiration_dt	created_dt	updated_dt
60932	VEND0002	SK offer	45	500	2023-03-15 0:00:01	2023-04-30 23:59:59	2023-03-09 6:15:10	2023-03-09 6:15:10
60933	VEND0016	Welch voucher	50	300	2023-03-15 0:00:01	2023-04-15 23:59:59	2023-03-12 6:25:03	2023-03-12 6:25:03
60934	VEND0013	Carter Discount	40	900	2023-03-15 0:00:01	2023-04-30 23:59:59	2023-03-13 21:00:30	2023-03-13 21:00:30

PRODUCT PROMOTION

- DB Entity Name:

PROD_PROMO

- Relation with others:



Attr.	Key	Type	Description
id	PK	index	linkage-id
prod_id	FK	FK. index	Product id
promo_id	FK	FK. index	Promotion id
included_flag		bool	Including / excluding product
deactivated_flag		bool	Deactivate linkage
created_dt		datetime	Create dt (YYYY-MM-DD h24:mm:ss)
updated_dt		datetime	Updated dt (YYYY-MM-DD h24:mm:ss)

PRODUCT PROMOTION (CONT'D)

- Data Sample

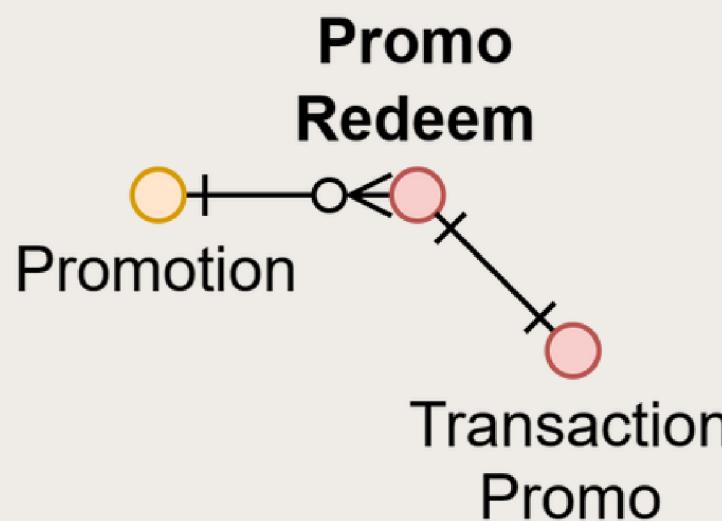
id	prod_id	promo_id	included_flag	deactivated_flag	created_dt	updated_dt
6976	1002106	60932	TRUE	FALSE	2023-03-09 7:25:11	2023-03-09 7:25:11
6977	1003683	60933	TRUE	FALSE	2023-03-12 7:13:31	2023-03-12 7:13:31
6978	1002864	60933	TRUE	FALSE	2023-03-12 7:52:53	2023-03-12 7:52:53
6979	1002178	60934	TRUE	FALSE	2023-03-13 21:41:33	2023-03-13 21:41:33
6980	1002105	60934	TRUE	FALSE	2023-03-13 22:48:52	2023-03-13 22:48:52

PROMO REDEEM

- DB Entity Name:

PROMO_REDEEM

- Relation with others:



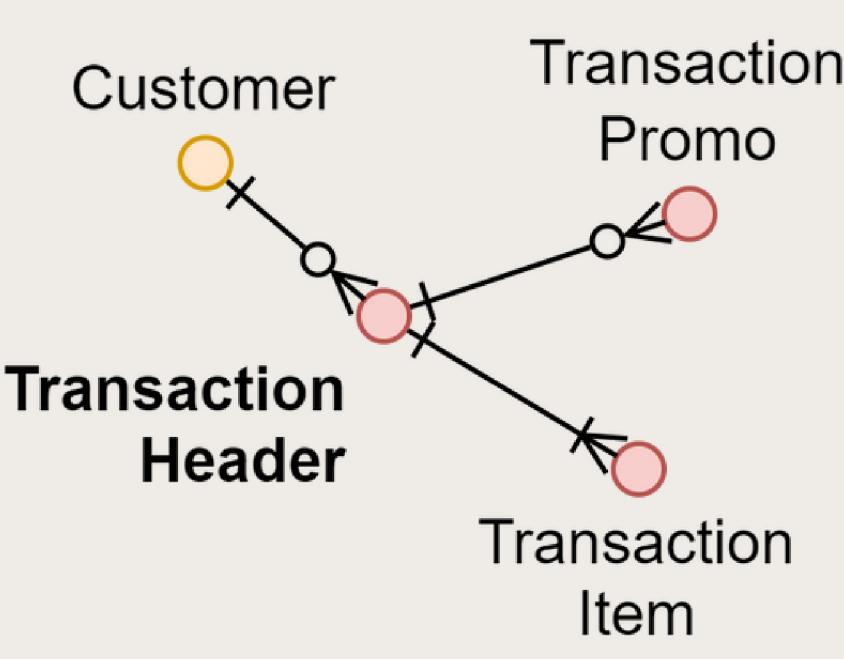
Attr.	Key	Type	Description
id	PK	index	Promo-Redeem id
promo_id	FK	FK. index	Promotion id
code		string	Promotion redeem-code
created_dt		datetime	Redeem create (YYYY-MM-DD h24:mm:ss)
expiration_dt		datetime	Redeem expire (YYYY-MM-DD h24:mm:ss)
activated_dt		datetime	Redeem activate (YYYY-MM-DD h24:mm:ss)

PROMO REDEEM (CONT'D)

- Data Sample

id	promo_id	code	created_dt	expiration_dt	activated_dt
6721	60932	18d17bc7	2023-03-09 6:15:10	2023-04-30 23:59:59	
6722	60932	ee89433d	2023-03-09 6:15:10	2023-04-30 23:59:59	
6723	60932	16b00bf7	2023-03-09 6:15:10	2023-04-30 23:59:59	
6724	60932	69162fa6	2023-03-09 6:15:10	2023-04-30 23:59:59	
6725	60932	6cb5f128	2023-03-09 6:15:10	2023-04-30 23:59:59	
6726	60932	854cfe12	2023-03-09 6:15:10	2023-04-30 23:59:59	
6727	60932	e42e9d68	2023-03-09 6:15:10	2023-04-30 23:59:59	2023-03-29 11:39:41
6728	60932	5c260f0e	2023-03-09 6:15:10	2023-04-30 23:59:59	
6729	60932	28e7e289	2023-03-09 6:15:10	2023-04-30 23:59:59	

TRANSACTION HEADER

• DB Entity Name:	Attr.			Description
TRX_HEADER	Attr.	Key	Type	
• Relation with others:	id	PK	index	Transaction header id
	customer_id	FK	string	Customer id
	cart_id		N/A	Shopping cart id (<i>Not Available</i>)
	sales_amnt		float64	Sales value without discount (#.##)
	total_discount		float64	Total discount value (#.##)
	total_amnt		float64	Discounted sales value (#.##)
	timestamp		datetime	Transaction time(YYYY-MM-DD h24:mm:ss)

TRANSACTION HEADER (CONT'D)

- Data Sample

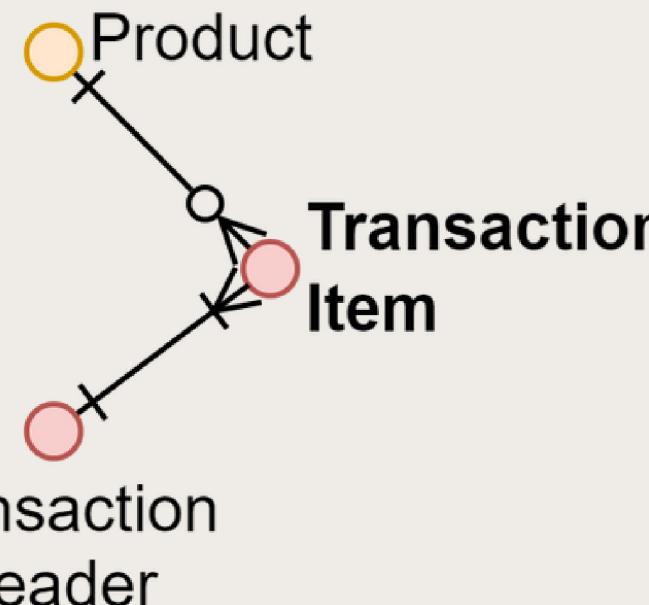
id	cust_id	cart_id	sales_amnt	total_discount	total_amnt	timestamp
236097	CID00127		33.2	0	33.2	2023-01-08 0:00:01
236098	CID00267		251.74	0	251.74	2023-01-08 0:06:06
236099	CID00175		1081.98	0	1081.98	2023-01-08 0:24:21
236100	CID00165		908.62	0	908.62	2023-01-08 0:30:26
236101	CID00240		1518.11	0	1518.11	2023-01-08 1:06:56
236104	CID00226		102.38	0	102.38	2023-01-08 1:19:06
236105	CID00227		193.03	0	193.03	2023-01-08 1:49:31
236106	CID00176		2328.68	0	2328.68	2023-01-08 1:55:36
236108	CID00034		621.98	0	621.98	2023-01-08 2:01:41
236110	CID00255		921.49	0	921.49	2023-01-08 2:13:51
236111	CID00237		197.55	0	197.55	2023-01-08 2:19:56

TRANSACTION ITEM

- DB Entity Name:

TRX_ITEM

- Relation with others:



Transaction Header

Attr.	Key	Type	Description
id	PK	index	Transaction item id
trx_header_id	FK	FK. index	Transaction header id
prod_id	FK	FK. index	Product id
quantity		int	Quantity of purchasing, product-unit
price		float64	Purchasing product-price
amnt		float64	Sales value: quantity x price
hrs_incart		N/A	Hours in shopping cart (Not Available)

TRANSACTION ITEM (CONT'D)

- Data Sample

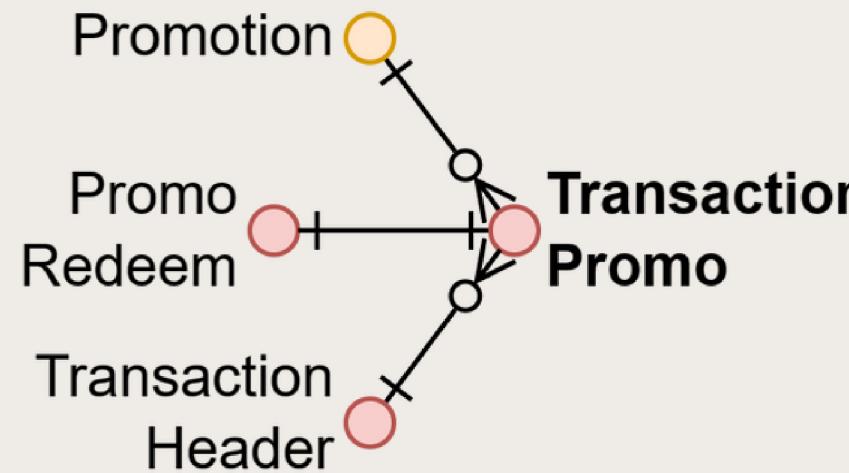
id	trx_header_id	prod_id	quantity	price	amnt	hrs_incart
186957	236097	1005711	1	33.2	33.2	
186958	236098	1002864	1	251.74	251.74	
186959	236099	1001888	1	1081.98	1081.98	
186960	236100	1003229	1	908.62	908.62	
186961	236101	1002008	1	566.01	566.01	
186962	236101	1004551	1	51.46	51.46	
186963	236101	1002266	1	900.64	900.64	
186964	236104	1005693	1	102.38	102.38	
186965	236105	1003754	1	193.03	193.03	
186966	236106	1002360	1	1747.79	1747.79	
186967	236106	1002738	1	580.89	580.89	

TRANSACTION PROMO

- DB Entity Name:

TRX_PROMO

- Relation with others:



Attr.	Key	Type	Description
id	PK	index	Transaction promo id
trx_header_id	FK	FK. index	Transaction header id
promo_id	FK	FK. index	Promotion id
promo_redeem_id	FK	FK. Index	Promo-Redeem id
discount_amnt		float64	Discount amount (#.##)

TRANSACTION PROMO (CONT'D)

- Data Sample

id	trx_header_id	promo_id	promo_redeem_id	discount_amnt
87385	290212	60934	8333	40
87386	290234	60934	7592	40
87387	290261	60934	8293	40
87388	290306	60933	7262	50
87389	290306	60934	8069	40
87390	290337	60934	7644	40
87391	290388	60934	8314	40
87392	290381	60934	8204	40

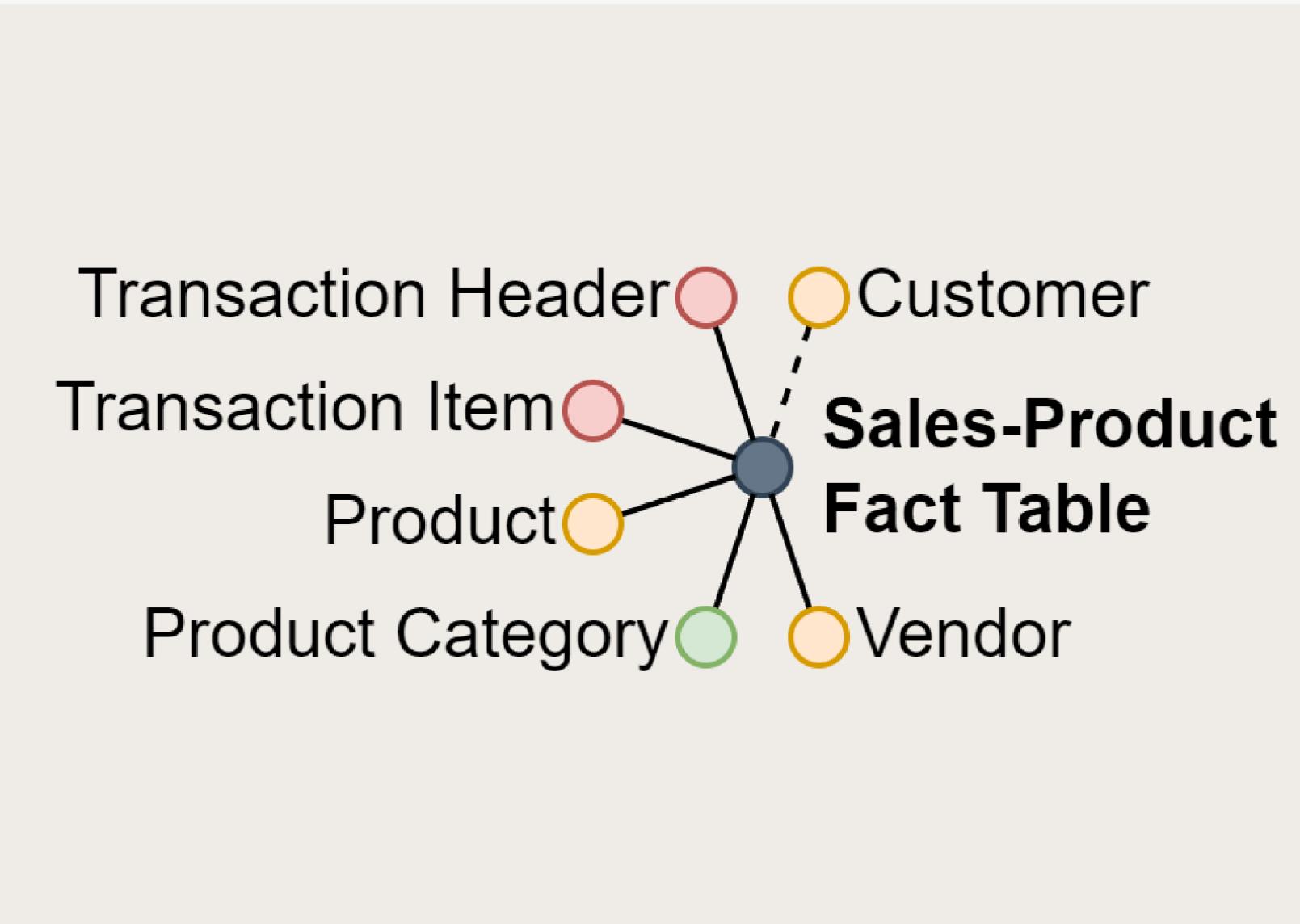
ANALYTICS DOMAIN

Integrated data modeling for
analytics purposes

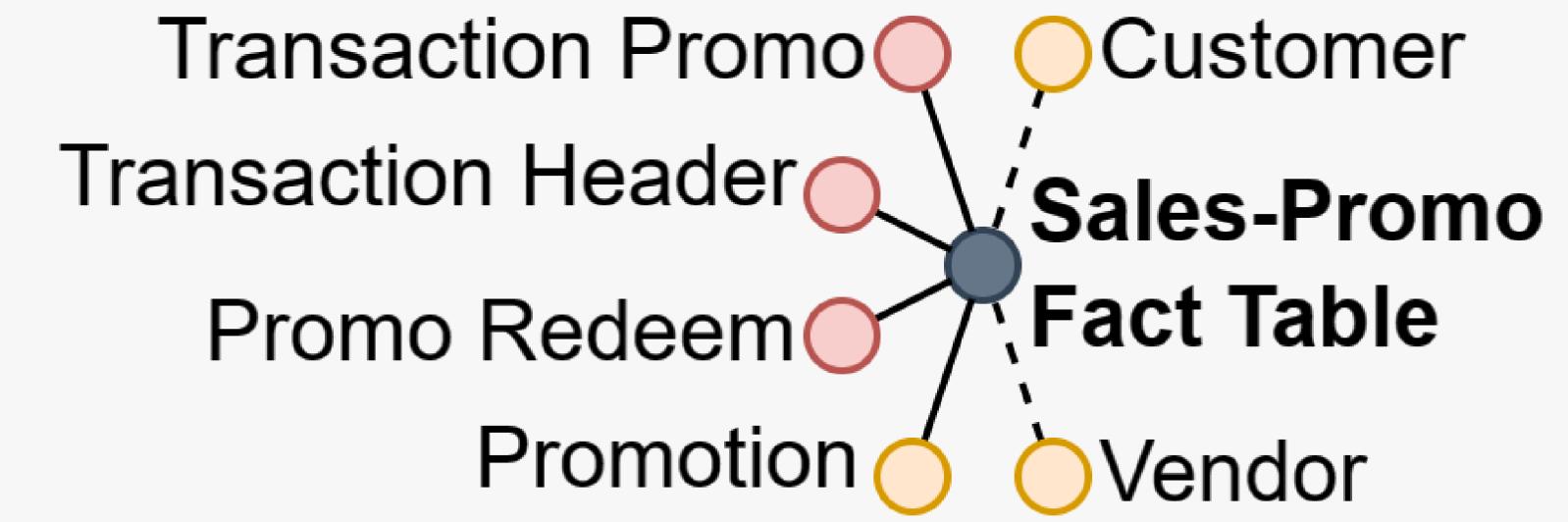


DATA WAREHOUSE

Sales-Product Fact Table



Sales-Promo Fact Table

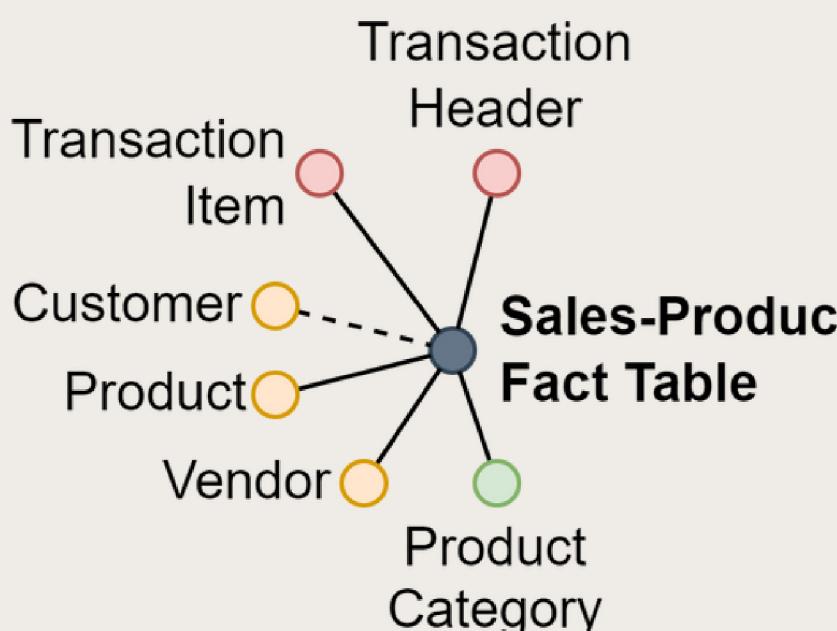


SALES-PRODUCT FACT TABLE

- DW Entity Name:

SALES_PROD_FACT

- Source and relevant entities:



Attr.	Key	Type	Description
trx_header_id	FK	FK, index	Transaction Header id
cust_id	FK	string	Customer id
trx_item_id	FK	FK, index	Transaction item id
prod_id	FK	FK, index	Product id
prod_cat_id	FK	string	Product category id
vendor_id	FK	string	Vendor id
trx_header_timestamp		datetime	Transaction timestamp

SALES-PRODUCT FACT TABLE (CONT'D)

Attr.	Type	Description
prod_name	string	Product Name
prod_brand	string	Product Brand
prod_cat_label	string	Product Category
trx_item_quantity	int	Purchased item qty
trx_item_price	float64	Purchased item price
trx_item_sales_amnt	float64	Sales amount

Attr.	Type	Description
trx_item_hrs_in_cart	N/A	<i>Not Available</i>
vend_official_flag	bool	Flag if official vendor

SALES-PRODUCT FACT TABLE (CONT'D)

- Data Sample

trx_header_id	cust_id	trx_item_id	prod_id	prod_cat_id	vendor_id	trx_header_timestamp	prod_name	prod_brand	prod_cat_code	prod_cat_label	trx_item_quantity	trx_item_price	trx_item_sales_amnt	trx_item_hrs_incart	official_flag
236097	CID00127	186957	1005711	CAT0028	VEND0006	1/8/2023 0:00	Aqua Water Heater	aqua	appliances.environmental	Water Heater	1	33.2	33.2		FALSE
236098	CID00267	186958	1002864	CAT0068	VEND0016	1/8/2023 0:06	Lenovo Notebook	lenovo	computers.notebook	Notebook	1	251.74	251.74		TRUE
236099	CID00175	186959	1001888	CAT0094	VEND0009	1/8/2023 0:24	Apple Smartphone	apple	electronics.smartphones	Smartphone	1	1081.98	1081.98		TRUE
236100	CID00165	186960	1003229	CAT0066	VEND0011	1/8/2023 0:30	Pulser Desktop	pulser	computers.desktop	Desktop	1	908.62	908.62		FALSE
236101	CID00240	186961	1002008	CAT0094	VEND0002	1/8/2023 1:06	Huawei Smartphone	huawei	electronics.smartphones	Smartphone	1	566.01	566.01		TRUE
236101	CID00240	186962	1004551	CAT0041	VEND0032	1/8/2023 1:06	Elenberg Microwave	elenberg	appliances.kitchen.appliances	Microwave	1	51.46	51.46		FALSE
236101	CID00240	186963	1002266	CAT0094	VEND0005	1/8/2023 1:06	Samsung Smartphone	samsung	electronics.smartphones	Smartphone	1	900.64	900.64		FALSE
236104	CID00226	186964	1005693	CAT0028	VEND0056	1/8/2023 1:19	Haier Water Heater	haier	appliances.environmental	Water Heater	1	102.38	102.38		FALSE
236105	CID00227	186965	1003754	CAT0098	VEND0028	1/8/2023 1:49	Haier Tv	haier	electronics.video_tv	Tv	1	193.03	193.03		FALSE
236106	CID00176	186966	1002360	CAT0094	VEND0006	1/8/2023 1:55	Apple Smartphone	apple	electronics.smartphones	Smartphone	1	1747.79	1747.79		FALSE
236106	CID00176	186967	1002738	CAT0068	VEND0009	1/8/2023 1:55	Hp Notebook	hp	computers.notebook	Notebook	1	580.89	580.89		TRUE
236108	CID00034	186968	1006594	CAT0087	VEND0075	1/8/2023 2:01	Jbl Headphone	jbl	electronics.audio.headphones	Headphone	1	33.21	33.21		FALSE
236108	CID00034	186969	1001665	CAT0094	VEND0005	1/8/2023 2:01	Apple Smartphone	apple	electronics.smartphones	Smartphone	1	588.77	588.77		FALSE
236110	CID00255	186970	1003272	CAT0066	VEND0024	1/8/2023 2:13	Pulser Desktop	pulser	computers.desktop	Desktop	1	921.49	921.49		TRUE
236111	CID00237	186971	1002088	CAT0094	VEND0004	1/8/2023 2:19	Xiaomi Smartphone	xiaomi	electronics.smartphones	Smartphone	1	197.55	197.55		FALSE
236112	CID00161	186972	1001665	CAT0094	VEND0005	1/8/2023 2:26	Apple Smartphone	apple	electronics.smartphones	Smartphone	1	588.77	588.77		FALSE
236113	CID00252	186973	1005756	CAT0028	VEND0062	1/8/2023 2:32	Ariston Water Heater	ariston	appliances.environmental	Water Heater	1	122.18	122.18		FALSE
236114	CID00221	186974	1005403	CAT0027	VEND0075	1/8/2023 3:02	Dauscher Vacuum	dauscher	appliances.environmental	Vacuum	1	33.21	33.21		FALSE
236115	CID00117	186975	1005693	CAT0028	VEND0056	1/8/2023 3:08	Haier Water Heater	haier	appliances.environmental	Water Heater	1	102.38	102.38		FALSE
236116	CID00112	186976	1004155	CAT0044	VEND0008	1/8/2023 3:20	Atlant Refrigerators	atlant	appliances.kitchen.refrigerators	Refrigerators	1	283.12	283.12		FALSE
236117	CID00257	186977	1005105	CAT0047	VEND0043	1/8/2023 3:32	Samsung Washer	samsung	appliances.kitchen.washers	Washer	1	463.15	463.15		TRUE
236117	CID00257	186978	1005353	CAT0027	VEND0004	1/8/2023 3:32	Elenberg Vacuum	elenberg	appliances.environmental	Vacuum	1	33.44	33.44		FALSE
236119	CID00046	186979	1004545	CAT0041	VEND0031	1/8/2023 3:39	Gorenje Microwave	gorenje	appliances.kitchen.appliances	Microwave	1	295.99	295.99		FALSE
236119	CID00046	186980	1002105	CAT0094	VEND0013	1/8/2023 3:39	Samsung Smartphone	samsung	electronics.smartphones	Smartphone	1	254.82	254.82		TRUE
236119	CID00046	186981	1004921	CAT0047	VEND0017	1/8/2023 3:39	Hotpoint-Ariston Washer	hotpoint-ariston	appliances.kitchen.washers	Washer	1	275.37	275.37		FALSE
236122	CID00204	186982	1003452	CAT0071	VEND0010	1/8/2023 3:45	Acer Monitor	acer	computers.peripherals	Monitor	1	514.79	514.79		FALSE

SALES-PROMO FACT TABLE

• DW Entity Name:
SALES_PROMO_FACT
• Source and relevant entities:

```

graph TD
    SPFT((Sales-Promo Fact Table)) --- TH((Transaction Header))
    SPFT --- P((Promo))
    SPFT --- C((Customer))
    SPFT --- V((Vendor))
    SPFT --- PR((Promo Redeem))
    
```

Attr.	Key	Type	Description
trx_header_id	FK	FK. index	Transaction Header id
cust_id	FK	FK. index	Customer id
promo_id	FK	FK. index	Promotion id
trx_promo_id	FK	FK. index	Transaction promo id
promo_redeem_id	FK	FK. index	Promo-Redeem id
vendor_id	FK	FK. index	Vendor id
trx_header_timestamp		datetime	Transaction timestamp

SALES-PROMO FACT TABLE (CONT'D)

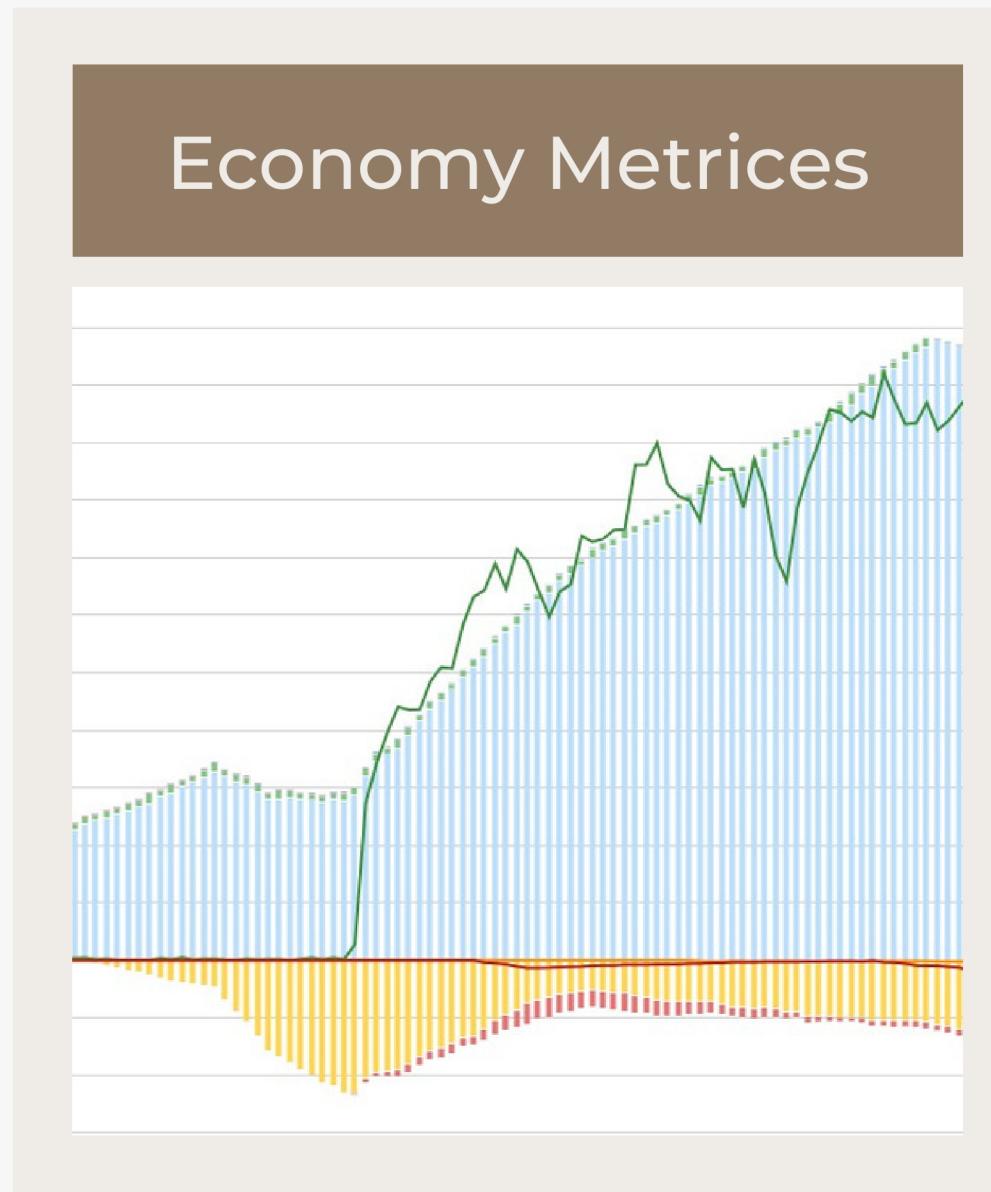
Attr.	Type	Description
promo_label	string	Promotion Label
trx_promo_discount_amnt	float64	Discount amount (#.##)
promo_redeem_activated_dt	datetime	Promo-redeem activation date time (YYYY-MM-DD h24:mm:ss)

SALES-PROMO FACT TABLE (CONT'D)

- Data Sample

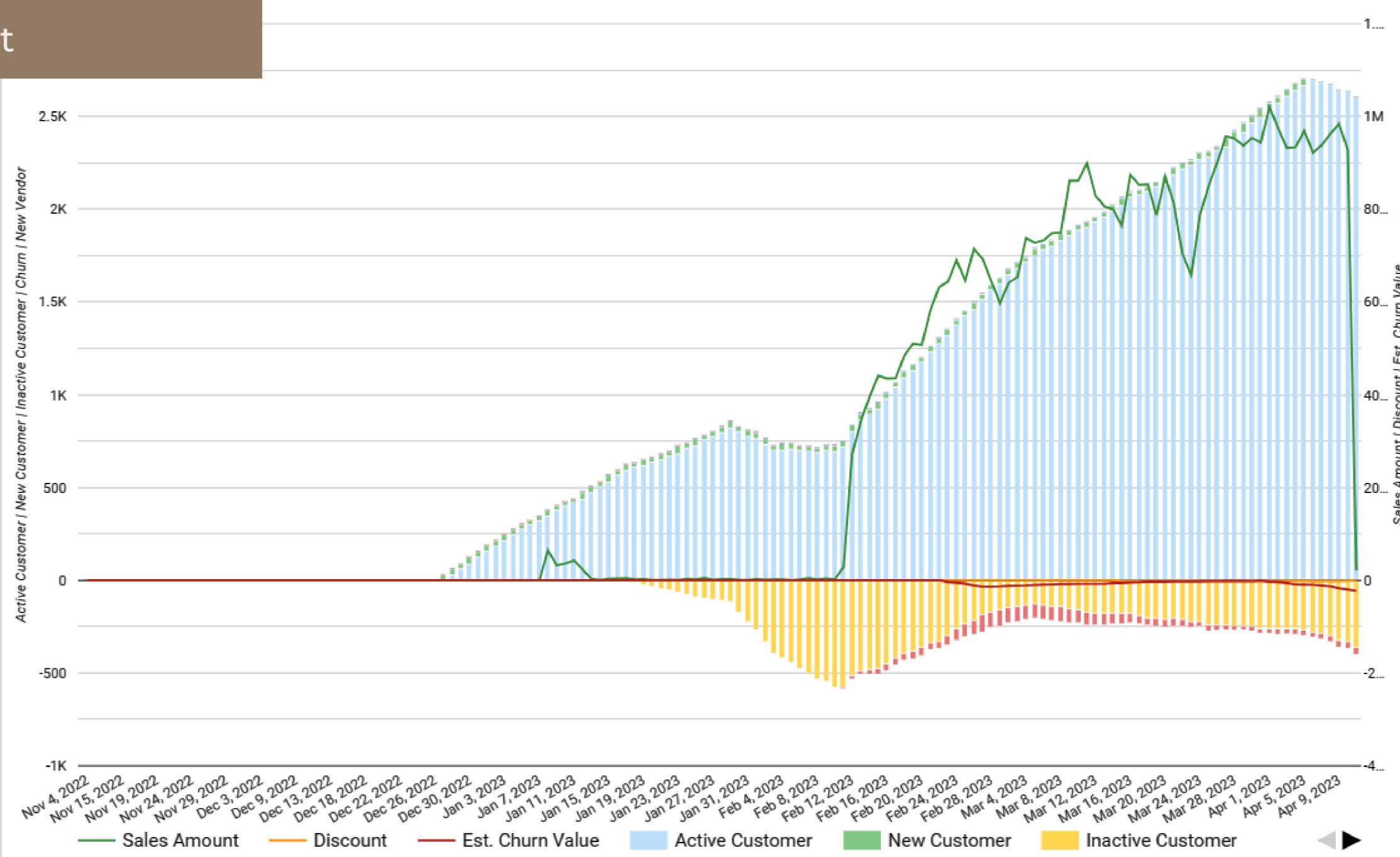
trx_header_id	cust_id	promo_id	trx_promo_id	promo_redeem_id	vendor_id	trx_header_timestamp	promo_label	trx_promo_discount	promo_redeem_acti
292971	CID00375	60932	87430	6839	VEND0002	3/16/2023 5:48	SK offer	45	3/16/2023 5:48
293081	CID01669	60932	87432	6758	VEND0002	3/16/2023 6:55	SK offer	45	3/16/2023 6:55
293263	CID00480	60932	87436	7178	VEND0002	3/16/2023 8:50	SK offer	45	3/16/2023 8:50
294110	CID00976	60932	87449	6799	VEND0002	3/16/2023 17:28	SK offer	45	3/16/2023 17:28
294124	CID01150	60932	87451	6957	VEND0002	3/16/2023 17:40	SK offer	45	3/16/2023 17:40
294146	CID01019	60932	87453	7214	VEND0002	3/16/2023 17:52	SK offer	45	3/16/2023 17:52
294643	CID01418	60932	87460	7155	VEND0002	3/16/2023 22:56	SK offer	45	3/16/2023 22:56
295212	CID00496	60932	87472	7215	VEND0002	3/17/2023 4:37	SK offer	45	3/17/2023 4:37
295463	CID00375	60932	87477	7049	VEND0002	3/17/2023 7:15	SK offer	45	3/17/2023 7:15
295652	CID01920	60932	87478	7186	VEND0002	3/17/2023 8:58	SK offer	45	3/17/2023 8:58
295718	CID00287	60932	87480	6787	VEND0002	3/17/2023 9:35	SK offer	45	3/17/2023 9:35
296190	CID00246	60932	87493	6945	VEND0002	3/17/2023 14:33	SK offer	45	3/17/2023 14:33
296796	CID00999	60932	87502	7187	VEND0002	3/17/2023 20:20	SK offer	45	3/17/2023 20:20
298069	CID01013	60932	87520	6793	VEND0002	3/18/2023 9:12	SK offer	45	3/18/2023 9:12
298303	CID00663	60932	87528	6914	VEND0002	3/18/2023 11:20	SK offer	45	3/18/2023 11:20
299441	CID00718	60932	87539	7157	VEND0002	3/18/2023 22:23	SK offer	45	3/18/2023 22:23

DATA MART



ECONOMY-METRICS

- Economic report



ECONOMY-METRICES (CONT'D)

• DW Entity Name:	Attr.	Type	Description
ECON_METRIX	date	date	Date on e-commerce platform (last 360 days)
• Source and relevant entities:	n_vendor_new	int	Number of new vendors on date
Transaction Header	n_cust_new	int	Number of new customers on date
Vendor	n_cust_active	int	Number of active customers on date
Customer	n_cust_inactive	int	Number of inactive customers on date
	n_cust_churn	int	Number of churn customers on date
	est_churn_amnt	float64	Estimated marketing value of churn (#.##)

ECONOMY-METRICS (CONT'D)

Attr.	Type	Description
daily_sales_amnt	float64	Daily total sales (#.##)
daily_sales_amnt_avg_15d	float64	Moving average daily total sales 15 days (#.##)
daily_sales_amnt_avg_30d	float64	Moving average daily total sales 30 days (#.##)
daily_sales_amnt_avg_45d	float64	Moving average daily total sales 45 days (#.##)
daily_discount	float64	Daily total discount (#.##)
daily_discount_avg_15d	float64	Moving average daily total discount 15 days (#.##)
daily_discount_avg_30d	float64	Moving average daily total discount 30 days (#.##)
daily_discount_avg_45d	float64	Moving average daily total discount 45 days (#.##)

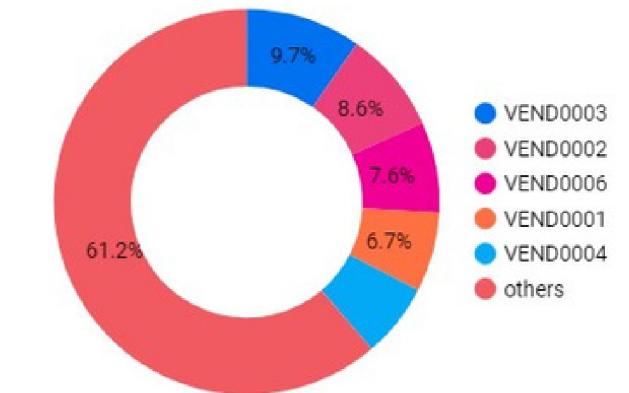
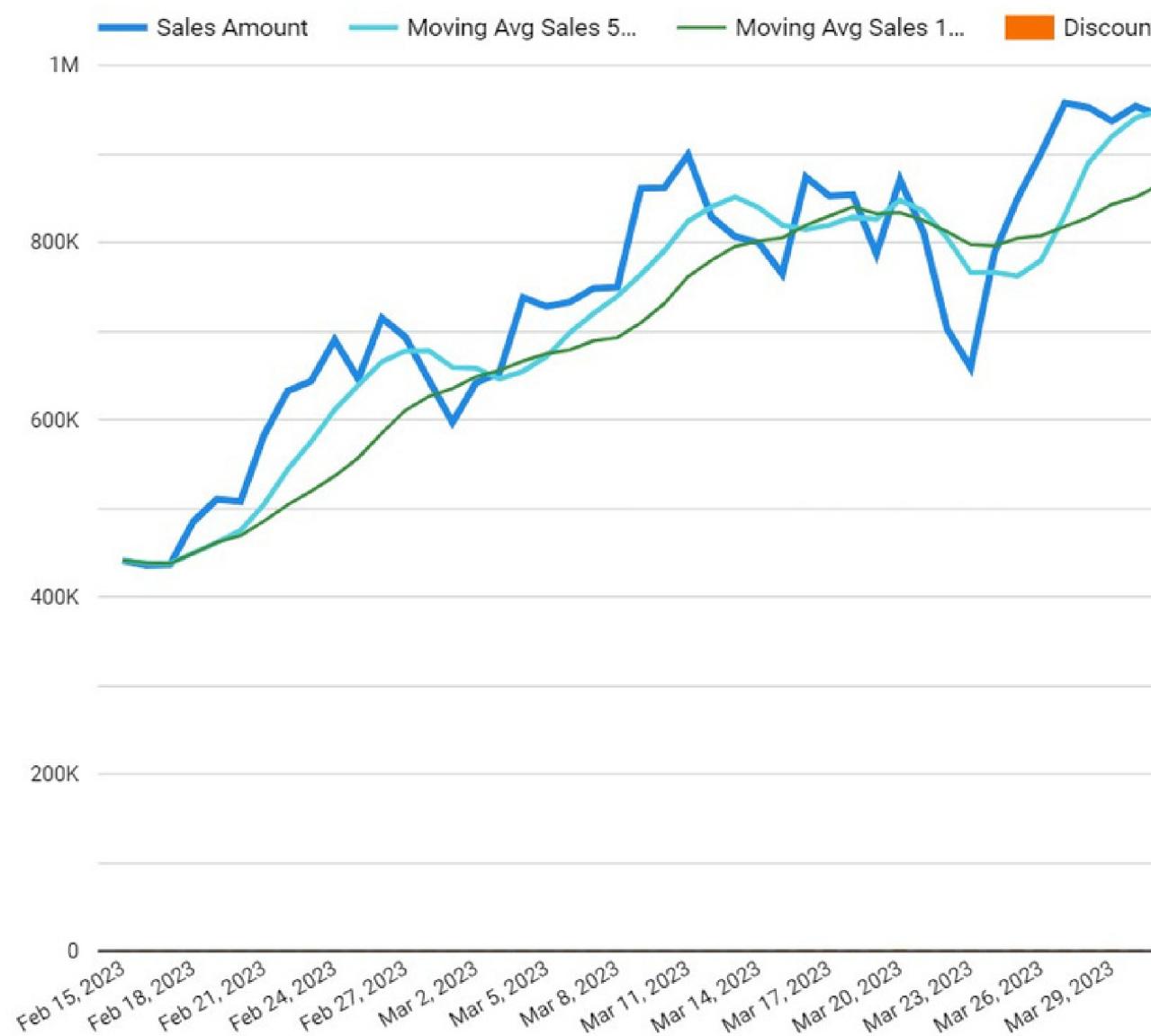
ECONOMY-METRICS (CONT'D)

- Data Sample

	n_vendor_new	n_cust_new	n_cust_active	n_cust_inactive	n_cust_churn	est_churn_amnt	daily_sales_amn	daily_sales_amn	daily_sales_amn	daily_sales_amn	daily_discount	daily_discount_a	daily_discount_a	daily_discount_a	daily_discount_a
-03-10	7	22	1888	159	71	7350.45	861777.82	713174.77	547204.33	365492.16	0	0	0	0	0
-03-11	5	32	1900	171	69	7214.79	898881.63	727099.62	577028.34	385427.37	0	0	0	0	0
-03-12	3	26	1929	178	65	7238.6	828613.17	739284.38	604579.56	403724.65	0	0	0	0	0
-03-13	5	25	1956	180	62	7238.6	806388.41	745404.57	630517.82	421621.45	0	0	0	0	0
-03-14	7	34	1987	178	58	5614.43	799767.08	752522.83	648079.49	439342.24	0	0	0	0	0
-03-15	10	39	2021	181	55	5506.74	764350.09	760513.32	662124.66	456267.73	1490	99.33	49.67		
-03-16	13	22	2065	178	53	4540.85	873924.44	778989.53	678059.02	475659.82	1805	219.67	109.83		
-03-17	3	24	2080	190	48	4116.38	852394.86	793051.72	691780.69	494590.41	1900	346.33	173.17		
-03-18	5	26	2098	201	43	2882.06	853817.96	806442.19	705729.48	513513.19	1345	436	218		
-03-19	2	24	2122	202	44	2882.06	786774.06	809718.46	717423.12	530968.2	1390	528.67	264.33		
-03-20	5	38	2136	212	44	2882.06	870896.96	819276.95	730294.13	550280.81	2320	683.33	341.67		
-03-21	6	37	2185	204	41	2753.45	811337.62	824522.65	740339.63	568270.33	1430	778.67	389.33		
-03-22	9	27	2218	211	38	2771.45	702591.04	821479.46	746842.56	583883.46	1430	874	437		
-03-23	11	24	2237	225	32	2771.45	658450.96	815411.49	749379.38	598465.83	1505	974.33	487.17		
-03-24	5	35	2267	223	28	2310.99	787135.14	810473.42	754547.57	615856.65	1190	1053.67	526.83		
-03-25	4	30	2282	243	28	2117.96	850029.68	809690.21	761432.49	634699.62	1985	1186	593		
-03-26	7	18	2317	239	27	2092.23	901351.02	809854.83	768477.23	654637.17	1720	1300.67	650.33		
-03-27	12	43	2337	242	22	1553.09	957300.9	818434.01	778859.2	675864.38	1645	1410.33	705.17		
-03-28	9	41	2379	245	20	1553.09	952255.71	828158.5	786781.54	696398.05	1775	1528.67	764.33		
-03-29	7	49	2415	251	19	1553.09	936990.97	837306.76	794914.8	711155.25	2120	1670	835		
-03-30	7	41	2462	248	24	2038.02	953609.65	849924.06	805218.69	724724.46	1665	1681.67	890.5		
-03-31	12	44	2492	261	22	879.81	943564.65	854566.75	816778.14	736894.93	1375	1653	936.33		

VENDOR-PROFILE

- Vendor Profile Dashboard



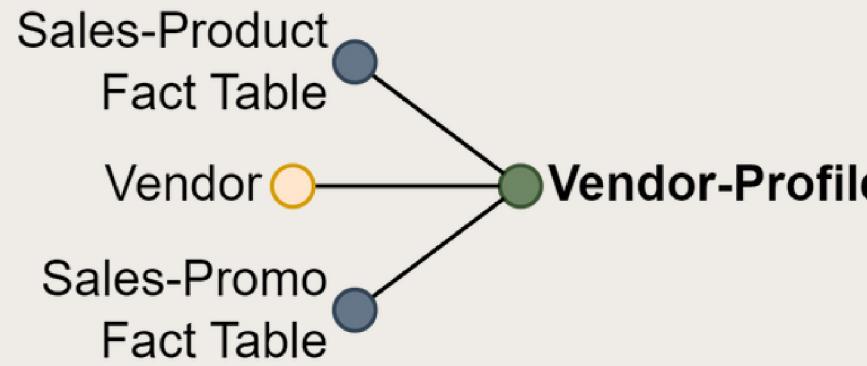
<input checked="" type="checkbox"/> id	Sales Amount
<input type="text"/> Type to search	
<input checked="" type="checkbox"/> VEND0003	3.2M
<input checked="" type="checkbox"/> VEND0002	2.8M
<input checked="" type="checkbox"/> VEND0006	2.5M
<input checked="" type="checkbox"/> VEND0001	2.2M
<input checked="" type="checkbox"/> VEND0004	2.1M
<input checked="" type="checkbox"/> VEND0008	1.7M

VENDOR-PROFILE (CONT'D)

- DW Entity Name:

VEND_PROFILE

- Source and relevant entities:



Attr.	Type	Description
id	str	Vendor id
date	date	Vendor transaction date (last 45 days)
sales_amnt	float64	Vendor's total sales amount (#.##)
discount_amnt	float64	Vendor's total discount amount (#.##)
sales_mvavg05	float64	Moving average sales amount 5 days (#.##)
sales_mvavg10	float64	Moving average sales amount 10 days (#.##)
dc_mvavg05	float64	Moving average discount 5 days (#.##)
dc_mvavg10	float64	Moving average discount 10 days (#.##)

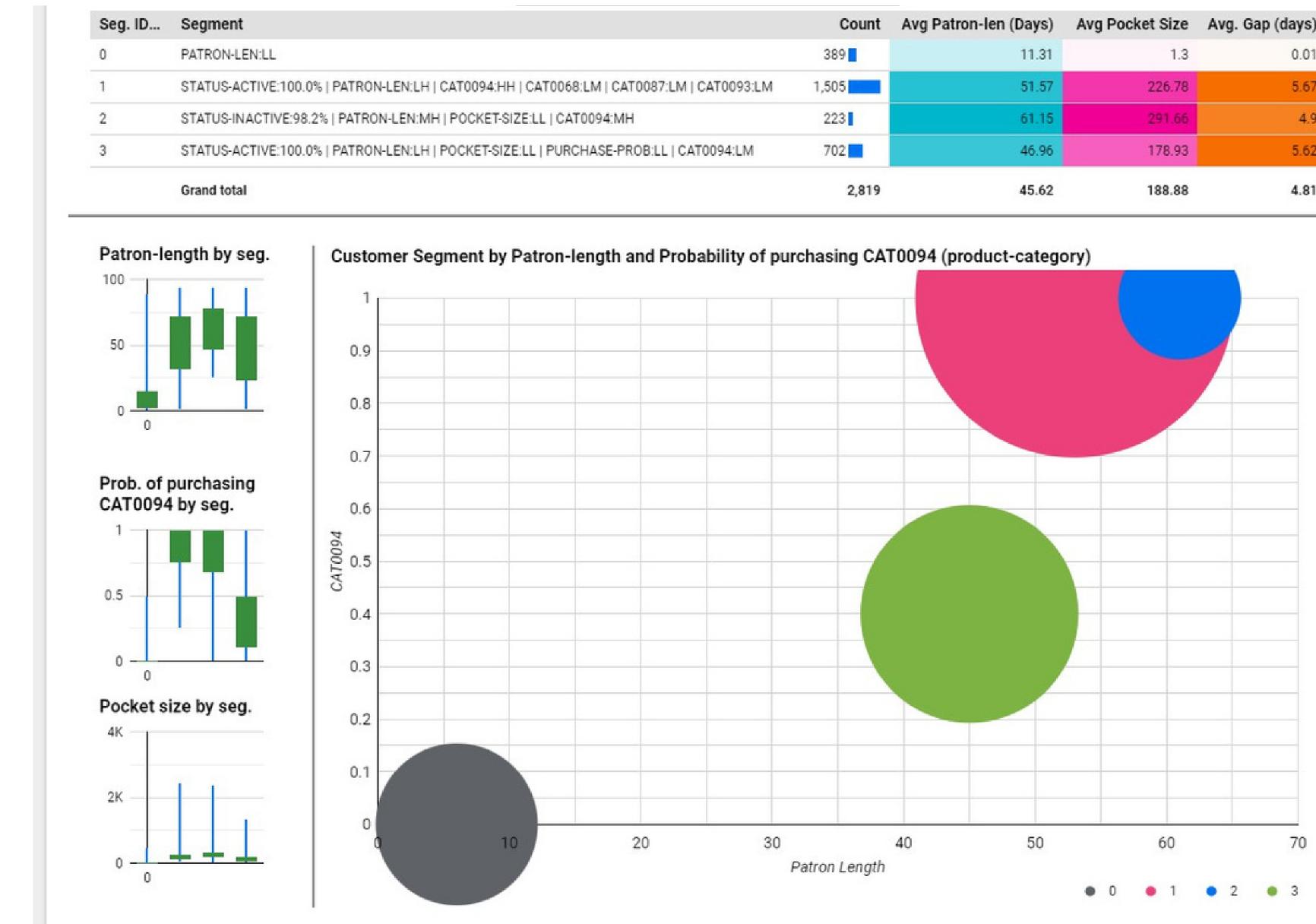
VENDOR-PROFILE (CONT'D)

- Data Sample

id	date	sales_amnt	discount_amnt	sales_mvavg05	sales_mvavg10	dc_mvavg05	dc_mvavg10
VEND0001	2023-02-15	35738.14	0	35738.14	35738.14	0	0
VEND0001	2023-02-16	32134.63	0	33936.38	33936.38	0	0
VEND0001	2023-02-17	26274.71	0	31382.49	31382.49	0	0
VEND0001	2023-02-18	27699.16	0	30461.66	30461.66	0	0
VEND0001	2023-02-19	28107.66	0	29990.86	29990.86	0	0
VEND0001	2023-02-20	41158.29	0	31074.89	31852.1	0	0
VEND0001	2023-02-21	40623.47	0	32772.66	33105.15	0	0
VEND0001	2023-02-22	39968.55	0	35511.43	33963.08	0	0
VEND0001	2023-02-23	38336.91	0	37638.98	34449.06	0	0
VEND0001	2023-02-24	35561.35	0	39129.71	34560.29	0	0
VEND0001	2023-02-25	48804.14	0	40658.88	35866.89	0	0
VEND0001	2023-02-26	40980.53	0	40730.3	36751.48	0	0
VEND0001	2023-02-27	48333.15	0	42403.22	38957.32	0	0

CUSTOMER PROFILE & SEGMENT

- Customer Segment Dashboard

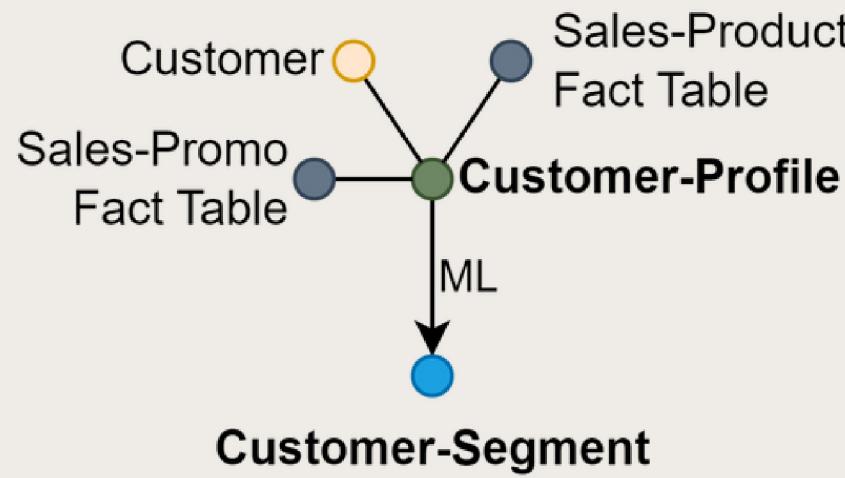


CUSTOMER-PROFILE

- DW Entity Name:

CUST_PROFILE

- Source and relevant entities:



Attr.	Type	Description
id	str	Customer id
reg_date	date	registered date
patron_length	int	patron-length from register-date (days)
latest_purchase	date	Last transaction date - within last 90 days
recent_flag	string	Recency status: active, inactive, churn, unknown
pkt_sz	float64	Median of daily purchasing value (#.##)
dc_amnt	float64	Median of daily discounted value (#.##)
gap	float64	Median of gap between purchase (days)

CUSTOMER-PROFILE (CONT'D)

Attr.	Type	Description
p_purchase	float64	Probability of purchasing within 90 days
p_purchase_dc	float64	Probability of discounted purchase in purchasing within 90 days
category_p	json	Probability of purchased product-categories (75%-100%)

Additional Field for Customer Segment, ML result		
CAT[0000]	float	Probability of each purchased categories, named columns as category id e.g. CAT0001
segment_id	int	Customer segment id (result of ML)
segment	str	Label of segment

CUSTOMER-PROFILE (CONT'D)

- Data Sample

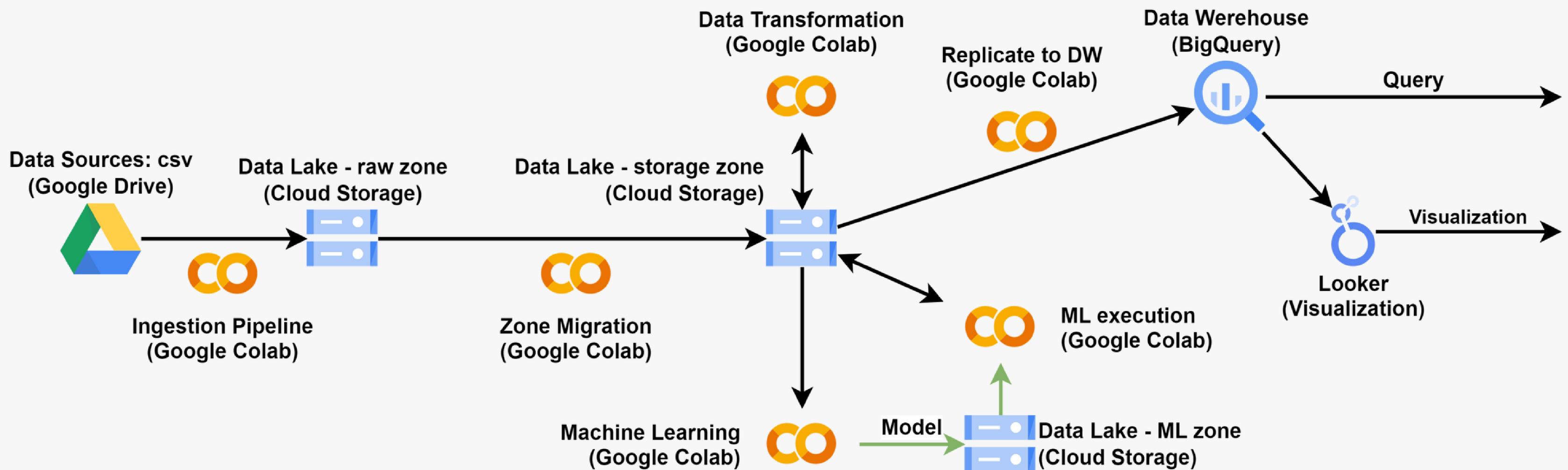
id	reg_date	patron_length	latest_purchase	recent_flag	pkt_sz	dc_amnt	gap	p_purchase	p_purchase_dc	category_p
CID00001	2022-12-27	94	2023-03-03	inactive	105.68	0	2	0.06	0	{"category_prob": "C"}
CID00002	2022-12-27	94	2023-02-28	inactive	174.89	0	7	0.04	0	{"category_prob": "C"}
CID00003	2022-12-27	94	2023-03-31	active	128.45	0	2.5	0.1	0	{"category_prob": "C"}
CID00004	2022-12-27	94	2023-03-29	active	180.13	0	12	0.08	0	{"category_prob": "C"}
CID00005	2022-12-27	94	2023-03-31	active	151.61	0	2	0.13	0	{"category_prob": "C"}
CID00006	2022-12-27	94	2023-03-29	active	239.27	0	2.5	0.1	0	{"category_prob": "C"}
CID00007	2022-12-27	94	2023-03-31	active	313.78	0	1	0.18	0	{"category_prob": "C"}
CID00008	2022-12-27	94	2023-03-29	active	339.61	0	34	0.04	0	{"category_prob": "C"}
CID00009	2022-12-27	94	2023-03-27	active	310.14	0	1	0.09	0	{"category_prob": "C"}
CID00010	2022-12-27	94	2023-03-31	active	87.25	0	4	0.1	0	{"category_prob": "C"}
CID00011	2022-12-27	94	2023-03-30	active	146.46	0	11	0.07	0	{"category_prob": "C"}
CID00012	2022-12-27	94	2023-03-20	active	267.32	45	3.5	0.17	0.13	{"category_prob": "C"}
CID00013	2022-12-27	94	2023-03-25	active	161.93	0	11.5	0.08	0	{"category_prob": "C"}
CID00014	2022-12-27	94	2023-03-12	active	218.8	0	1.5	0.1	0	{"category_prob": "C"}
CID00015	2022-12-27	94	2023-03-27	active	255.86	40	2	0.14	0.08	{"category_prob": "C"}
CID00016	2022-12-27	94	2023-03-27	active	178.87	0	16	0.07	0	{"category_prob": "C"}
CID00017	2022-12-27	94	2023-03-31	active	71.82	0	3	0.1	0	{"category_prob": "C"}
CID00018	2022-12-27	94	2023-03-15	active	173.36	40	11	0.06	0.2	{"category_prob": "C"}

DATA PIPELINE

Data Pipeline Design, Object-Class Model for Pipeline control, and metrics

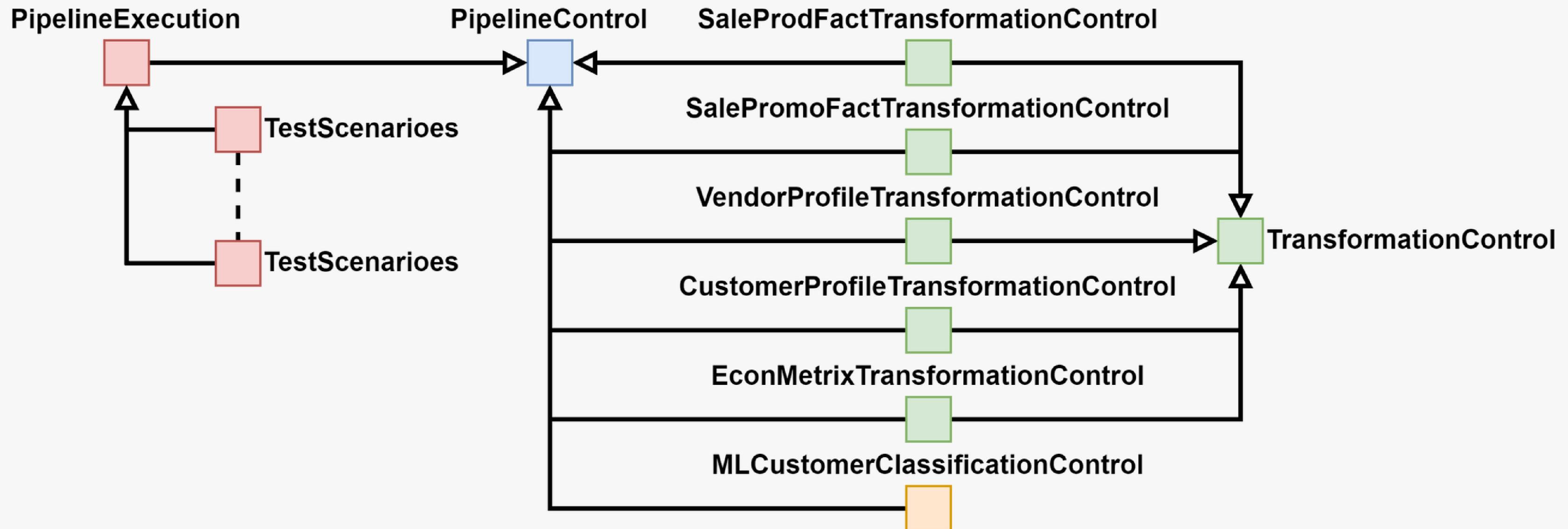


DATA PIPELINE DESIGN

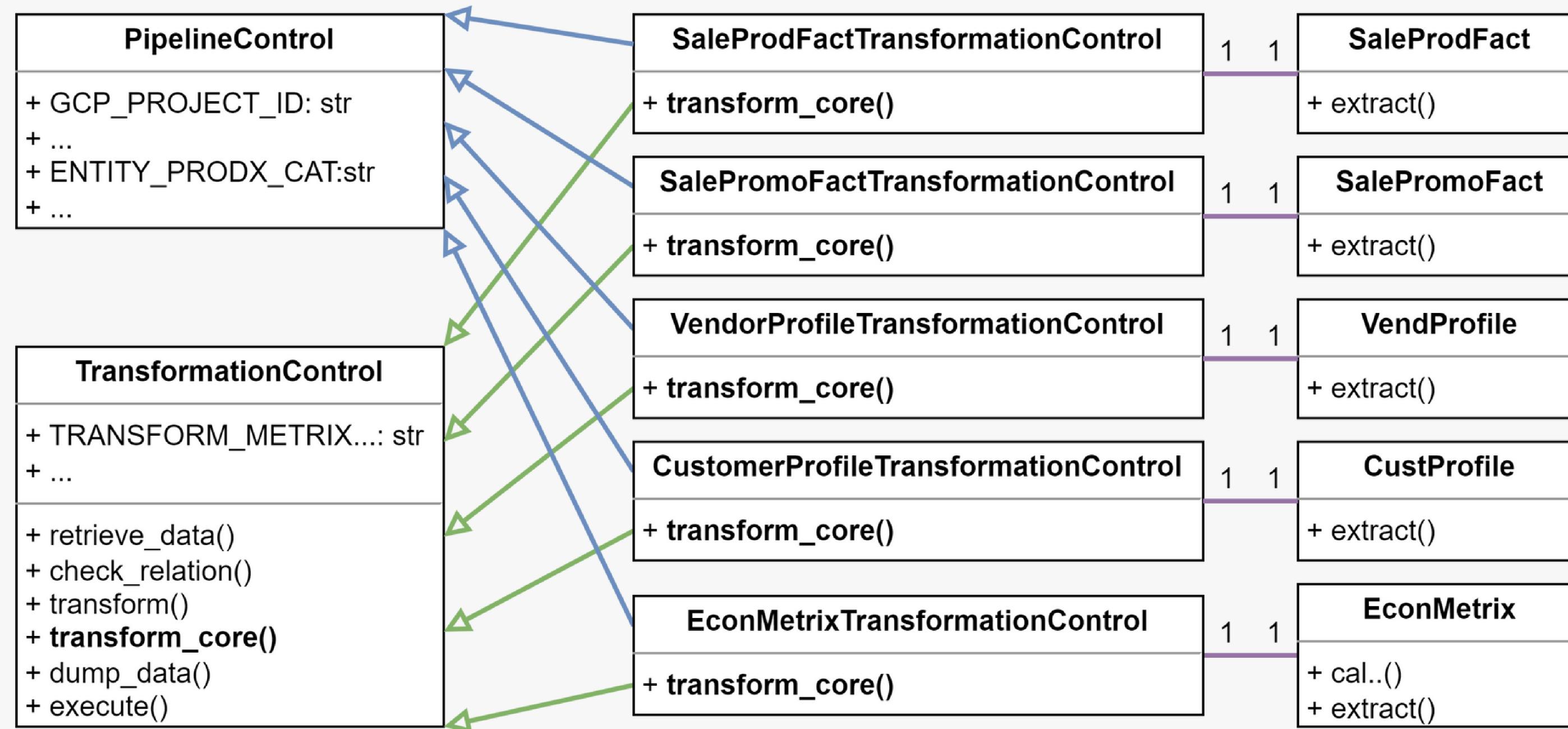


PIPELINE CONTROL CLASSES

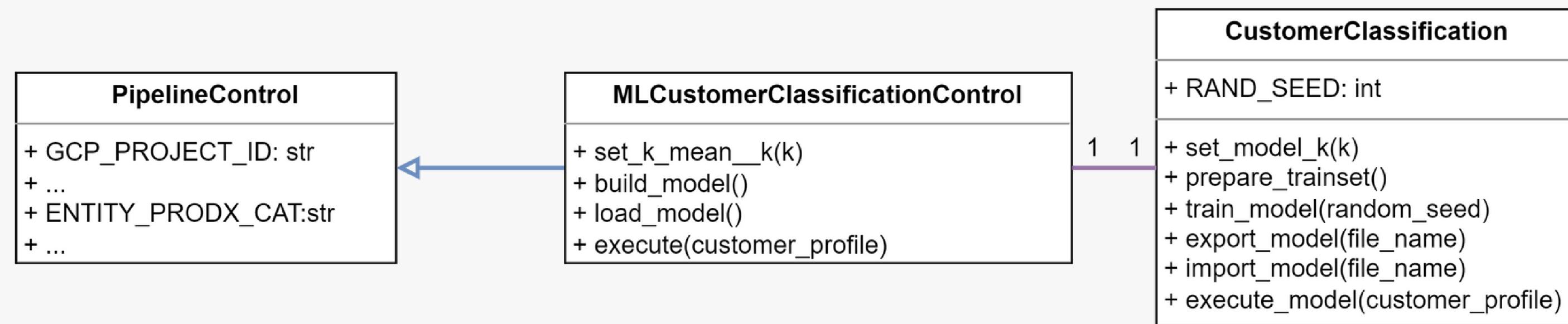
- Class Inheritance



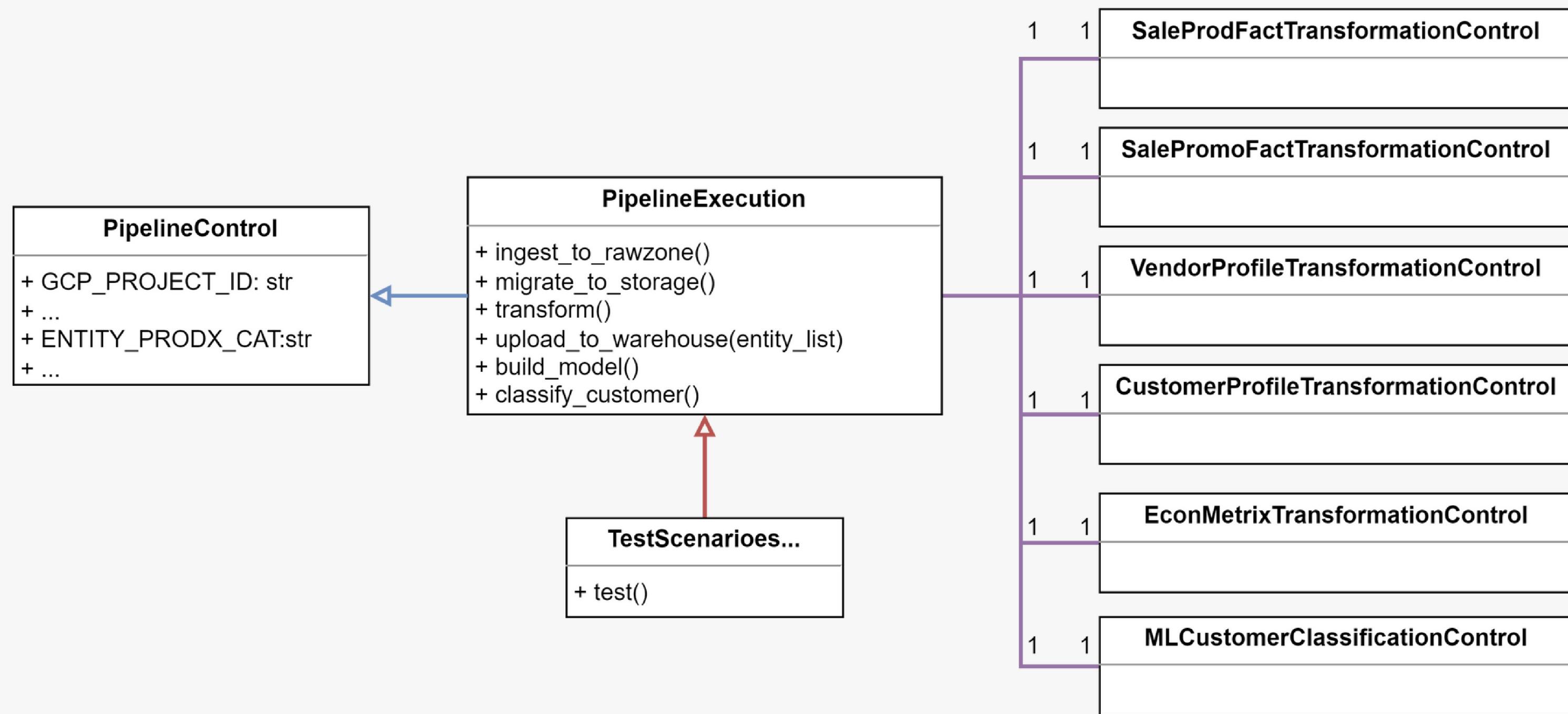
TRANSFORM CTRL CLASS DIAGRAM



ML CONTROL CLASS DIAGRAM



PIPELINE CTRL CLASS DIAGRAM



PIPELINE PROCESS AND METRICS

Sequence	Process	Metrics Detail
Ingestion	Ingestion	Validate source
Zone Migration	Zone Migration	Validate source, check structure/data change, and verify migration
Transformation	Transformation	Verify source extraction, data relation, and transformation result
Replicate to DW	Replicate to DW	Validate source existence, and loading status
	ML Training	Verify source extraction, model building, and storing result
	ML Execution	Verify source extraction, model loading and execution, and storing result

```
graph TD; Ingestion((Ingestion)) --> ZoneMigration((Zone Migration)); ZoneMigration --> Transformation((Transformation)); Transformation --> ReplicateDW((Replicate to DW)); ReplicateDW --> MLExecution((ML Execution)); MachineLearning((Machine Learning)) -.-> MLTraining((ML Training)); MachineLearning -.-> MLExecution;
```

INGESTION METRICS

	label	local_csv	err_flag	err_desc	timestamp
0	PROD_CAT	mockup_ecomm_prod_cat.csv	0	None	2023-06-11 13:41:45.777331
1	VEND	mockup_ecomm_vend.csv	0	None	2023-06-11 13:41:47.483593
2	PROD	mockup_ecomm_prod.csv	0	None	2023-06-11 13:41:50.443787
3	CUST	mockup_ecomm_cust.csv	0	None	2023-06-11 13:41:52.855409
4	PROMO	mockup_ecomm_promo.csv	0	None	2023-06-11 13:41:54.160972
5	PROD_PROMO	mockup_ecomm_prod_promo.csv	0	None	2023-06-11 13:41:55.478926
6	PROMO_REDEEM	mockup_ecomm_promo_redeem.csv	0	None	2023-06-11 13:41:57.163656
7	TRX_HEADER	mockup_ecomm_trx_header.csv	0	None	2023-06-11 13:41:59.914374
8	TRX_ITEM	mockup_ecomm_trx_item.csv	0	None	2023-06-11 13:42:03.192795
9	TRX_PROMO	mockup_ecomm_trx_promo.csv	0	None	2023-06-11 13:42:03.931327

ZONE-MIGRATION METRICS

	label	migrate_mode	reason	unchanged_row	insert_row	update_row	delete_row
0	PROD_CAT	full-dump	None	None	122	None	None
1	VEND	full-dump	None	None	693	None	None
2	PROD	full-dump	None	None	13335	None	None
3	CUST	full-dump	None	None	3000	None	None
4	PROMO	full-dump	None	None	3	None	None
5	PROD_PROMO	full-dump	None	None	5	None	None
6	PROMO_REDEEM	full-dump	None	None	1700	None	None
7	TRX_HEADER	full-dump	None	None	20126	None	None
8	TRX_ITEM	full-dump	None	None	123507	None	None
9	TRX_PROMO	full-dump	None	None	1028	None	None

TRANSFORMATION METRICS

		mandatory	verified_mandatory	optional	verified_optional	df_mandatory	df_optional
	proc-label	metrice					
SALES_PROD_FACT	source-pulse-check	5	5	1	1	entity ty pulse extract err_...	entity ty pulse extract err_flag e...
	extraction-check	5	5	1	1	entity ty pulse extract err_...	entity ty pulse extract err_flag e...
	relationship-check	4	4	1	1	tgt_entity ref_entity tgt_fk ref_pk...	tgt_entity ref_entity tgt_fk ref_pk ...
	transformation-status	1	1	None	None	trx_header_id cust_id trx_item_id ...	None
	load-status	1	1	None	None	entity verified err_flag reason ...	None
SALES_PROMO_FACT	source-pulse-check	4	4	2	2	entity ty pulse extract er...	entity ty pulse extract err_flag e...
	extraction-check	4	4	2	2	entity ty pulse extract er...	entity ty pulse extract err_flag e...
	relationship-check	4	4	2	2	tgt_entity ref_entity tgt_fk...	tgt_entity ref_entity tgt_fk ref_pk ...
	transformation-status	1	1	None	None	trx_header_id cust_id promo_id trx_p...	None
	load-status	1	1	None	None	entity verified err_flag reason...	None
VEND_PROFILE	source-pulse-check	3	3	0	0	entity ty pulse extract...	Empty DataFrame Columns: [entity, ty, pulse, e...
	extraction-check	3	3	0	0	entity ty pulse extract...	Empty DataFrame Columns: [entity, ty, pulse, e...
	relationship-check	2	2	0	0	tgt_entity ref_entity tgt_fk ref_...	Empty DataFrame Columns: [tgt_entity, ref_enti...

REPLICATION TO DW METRICS

	entity	src_pulse	tgt_dataset	uploaded	err_flag	err_desc
0	PROD_CAT	1	madt7102-termproject-ecommerce.reference_master	1	0	None
1	VEND	1	madt7102-termproject-ecommerce.reference_master	1	0	None
2	PROD	1	madt7102-termproject-ecommerce.reference_master	1	0	None
3	CUST	1	madt7102-termproject-ecommerce.reference_master	1	0	None
4	PROMO	1	madt7102-termproject-ecommerce.reference_master	1	0	None
5	SALES_PROD_FACT	1	madt7102-termproject-ecommerce.fact_dimension	1	0	None
6	SALES_PROMO_FACT	1	madt7102-termproject-ecommerce.fact_dimension	1	0	None
7	VEND_PROFILE	1	madt7102-termproject-ecommerce.data_mart	1	0	None
8	ECON_METRIX	1	madt7102-termproject-ecommerce.data_mart	1	0	None

MACHINE LEARNING METRICS

Training

	progress	err_flag	err_desc
src-extract	1	0	None
build-model	1	0	None
store-model	1	0	None

Execution

	progress	err_flag	err_desc
src-extract	1.0	0.0	NaN
load-model	1.0	0.0	NaN
run	1.0	0.0	NaN
store-result	1.0	0.0	NaN

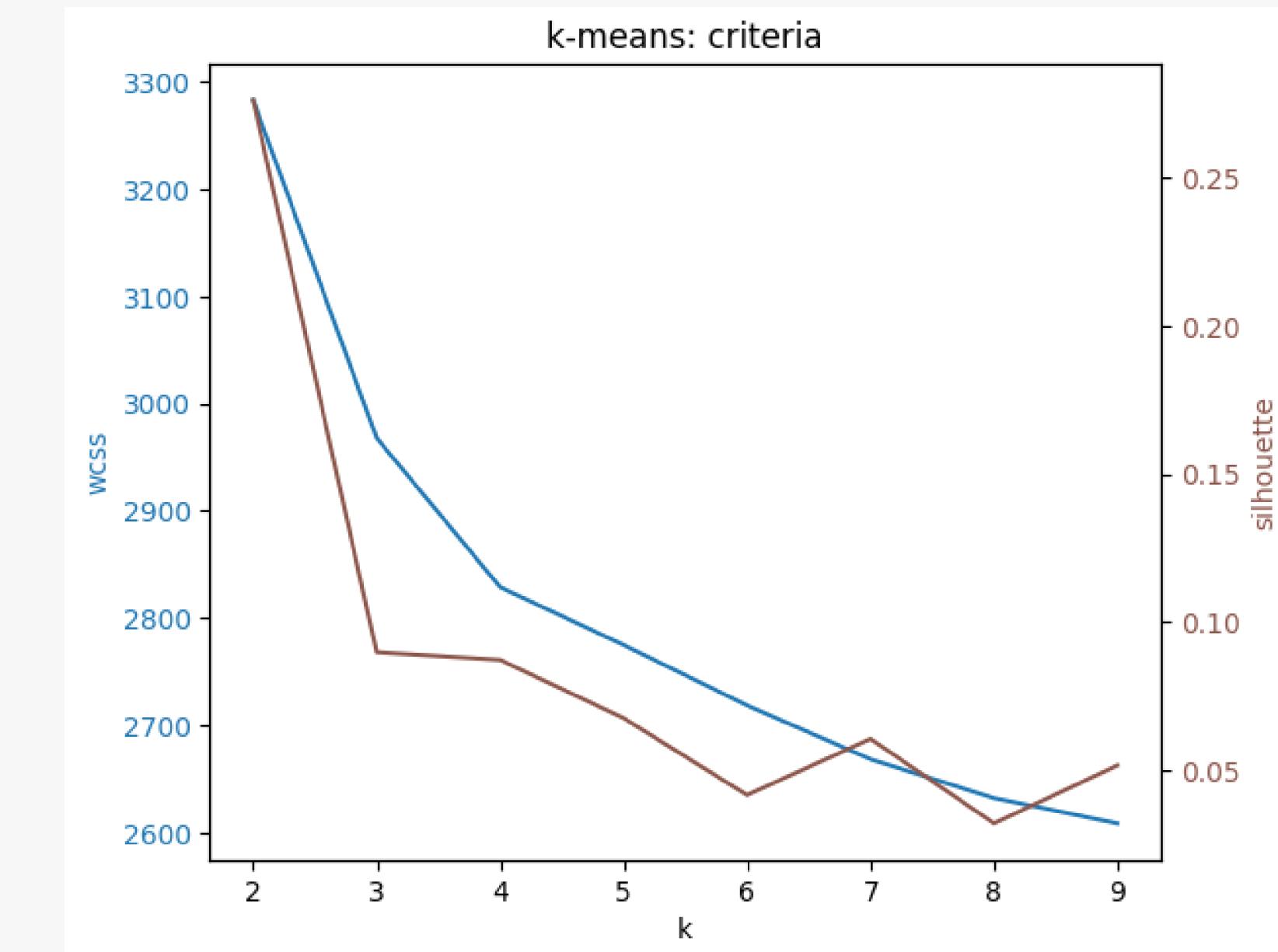
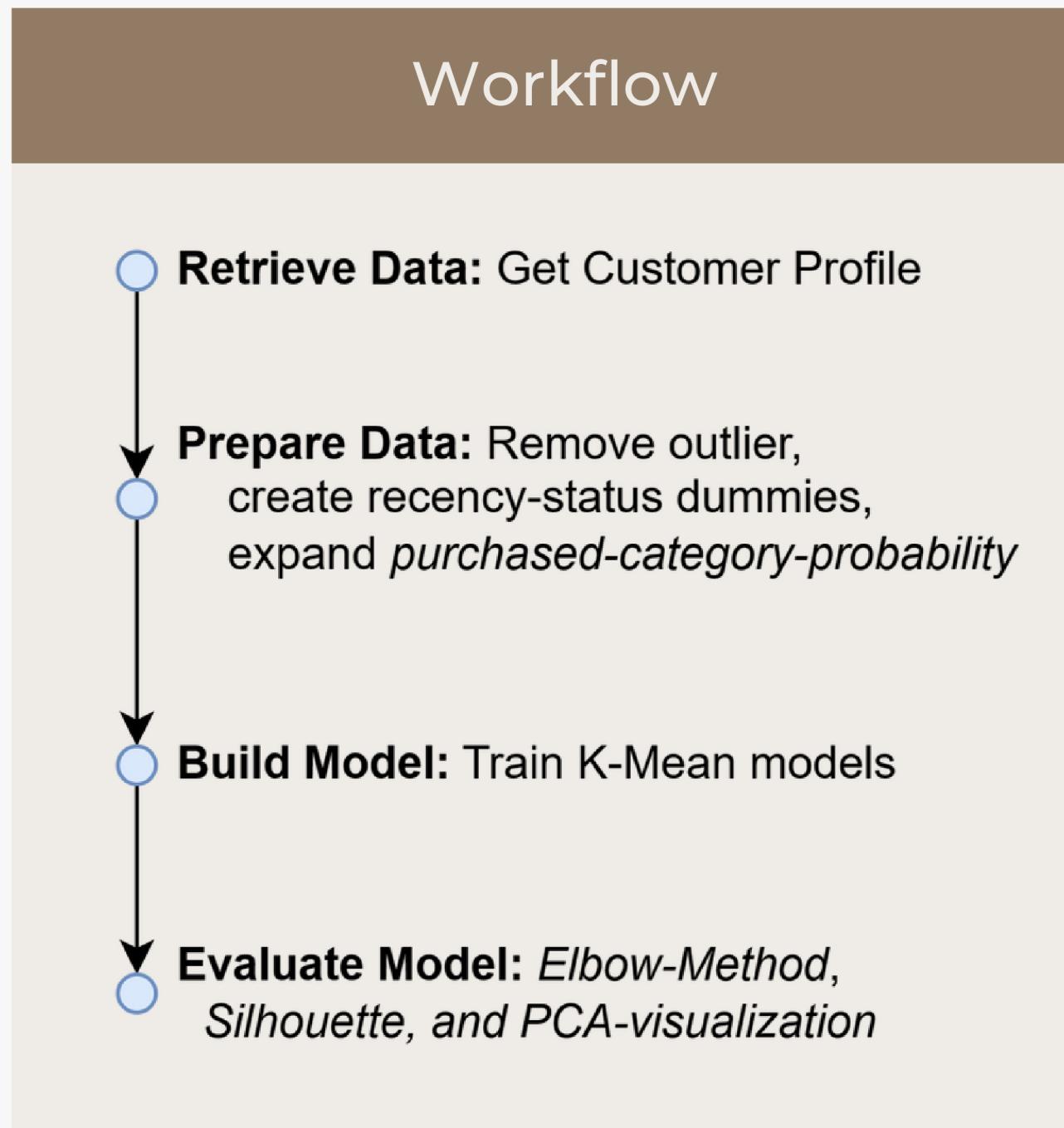
Upload Result

	entity	src_pulse	tgt_dataset	uploaded	err_flag	err_desc
0	CUSTOMER_SEGMENT	1	madt7102-termproject-ecommerce.data_mart	1	0	None

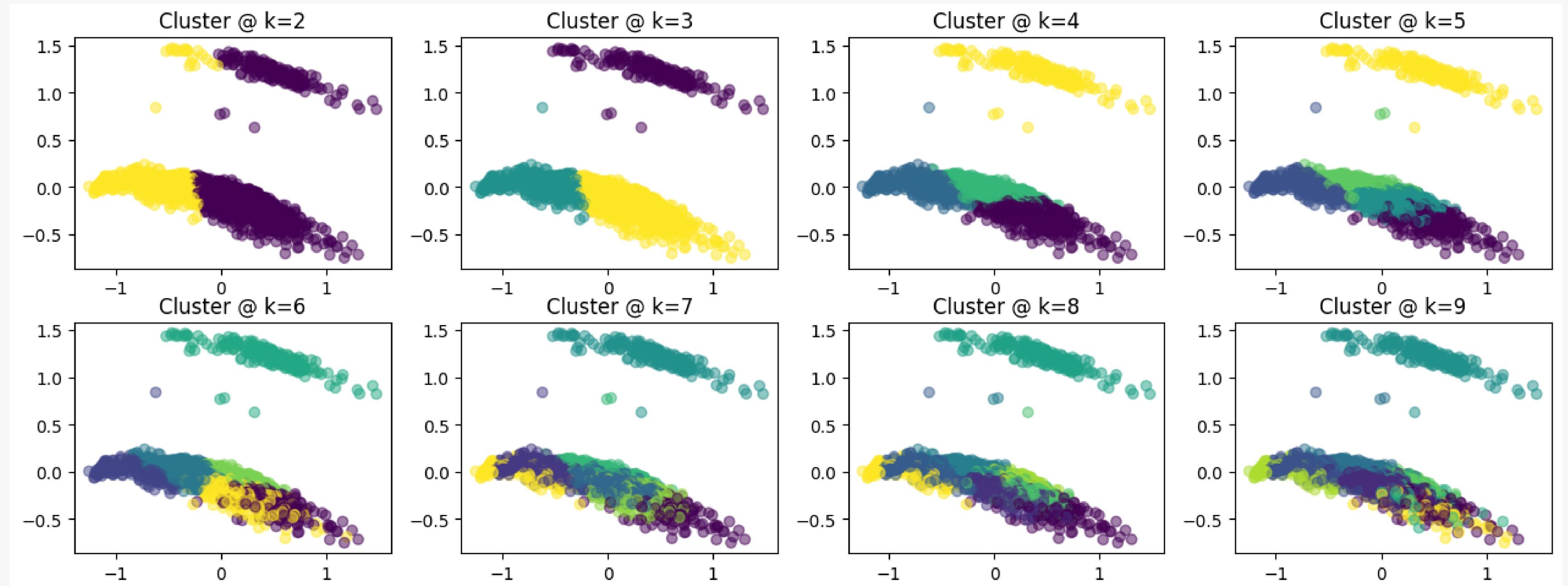
ML: CUSTOMER SEGMENT



ML-MODEL DEVELOPMENT



Model Evaluation: Elbow method and Silhouette



Model Evaluation: PCA Visualization

A stylized yellow brushstroke logo resembling the word "Result".

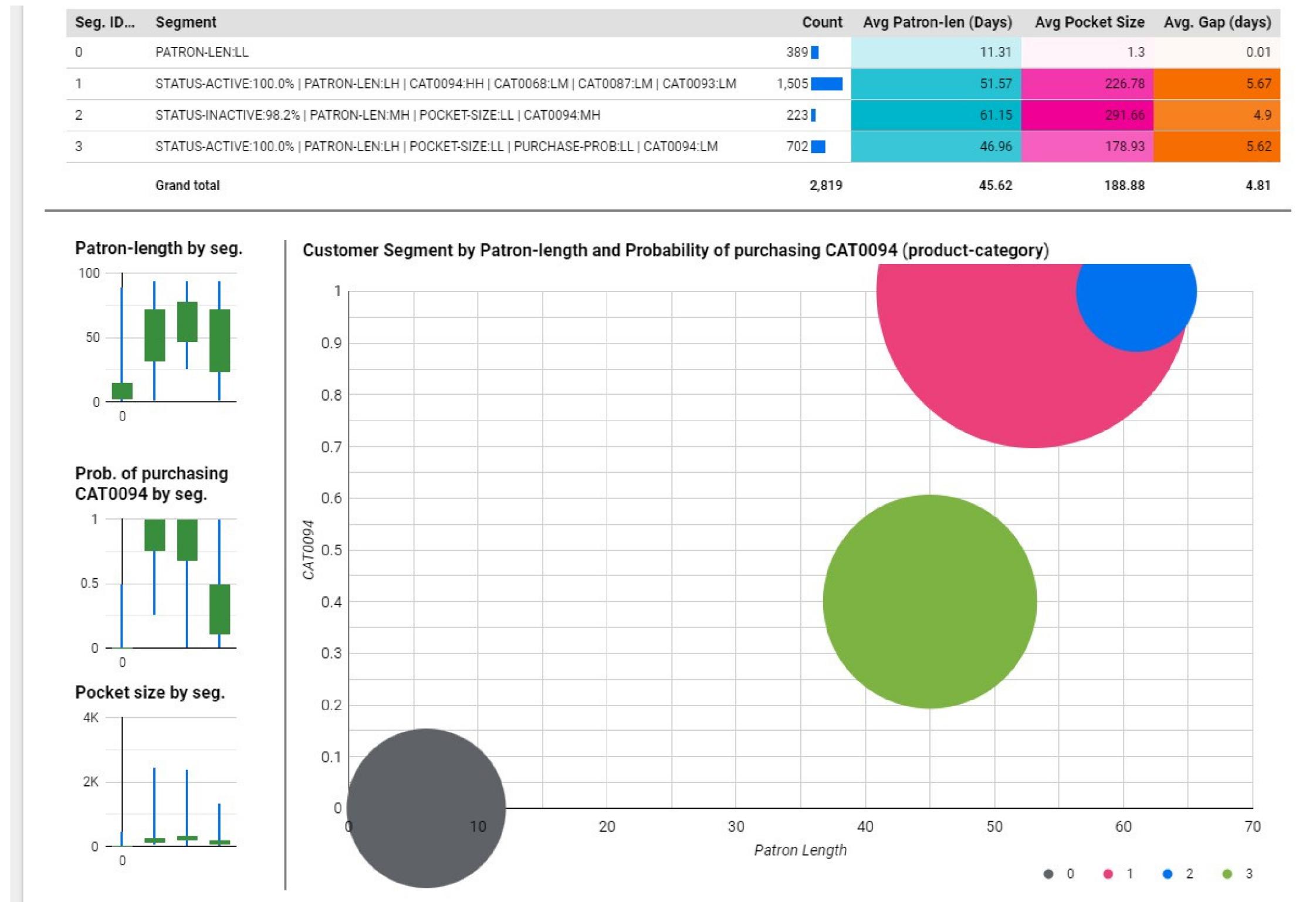
CUSTOMER SEGMENT USING K-MEAN

With K-Mean classification model

It is possible to classify customer to these 4 segments.

Note: Segment key attributes reveals in labels.

Seg. ID	Segment
0	PATRON-LEN:LL
1	STATUS-ACTIVE:100.0% PATRON-LEN:LH CAT0094:HH CAT0068:LM CAT0087:LM CAT0093:LM
2	STATUS-INACTIVE:98.2% PATRON-LEN:MH POCKET-SIZE:LL CAT0094:MH
3	STATUS-ACTIVE:100.0% PATRON-LEN:LH POCKET-SIZE:LL PURCHASE-PROB:LL CAT0094:LM



Visualize customer segment information on Looker

DEMO & CONCLUSION



Conclusion

KEYS FOR DESIGN & IMPLEMENTATION

Remarks, key points, and suggestions for data model design and implementation.

Business Understanding

Business Knowledge is mandatory - leads to purpose of data collection, analytics requirements and implementation

Data Management Strategies

Define effectively ways to manage data - e.g., data availabilities, data change management, etc., and strategic metrics

Tech & Implementation Knowledge

Cloud technology and technical knowledge is required for determining whether data model and strategies are practical.