



LIVE
LEARN
LOVE

Customer segmentation and Product Recommendation

P R E S E N T A T I O N

 <https://sg.hdi.com>



Content List



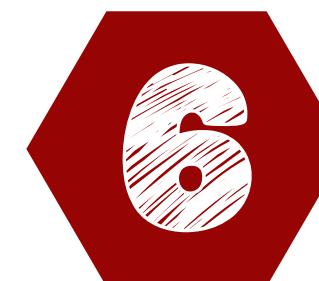
Business Overview



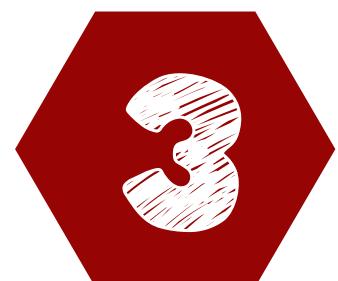
Data Preparation



Analysis Objective



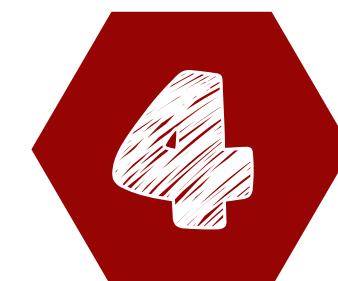
Customer Segmentation



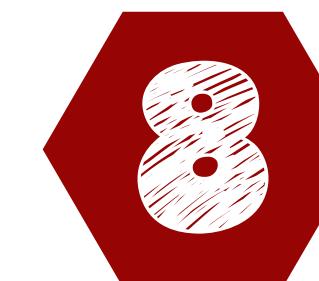
**Data Pre-processing
(Cleansing data)**



Product Recommendation



**Exploratory Data Analysis
(EDA)**



Actions and Summary



1

A large, stylized number '1' is enclosed within a dark red hexagonal shape. The '1' has a white, textured interior with a diagonal hatching pattern.

BUSINESS OVERVIEW

HD International

A world-class distribution company with a foundation of love and care.

HDI is a **Social Network Marketing organisation** that offers a wide range of premium bee-based nutritional and skin care products as well as a unique opportunity to learn about entrepreneurship. The Company has 200,000 more Enterprisers with 6 representative offices in America, Singapore, Malaysia, Indonesia , Hong Kong and the Philippines.

Apart from Business Networks, the Company also has other businesses, such as:

- 1 Property and investment development
- 2 Advertising / Media
- 3 Stock / Financial Broker
- 4 Research and development



2

ANALYSIS OBJECTIVES

Analysis Objectives

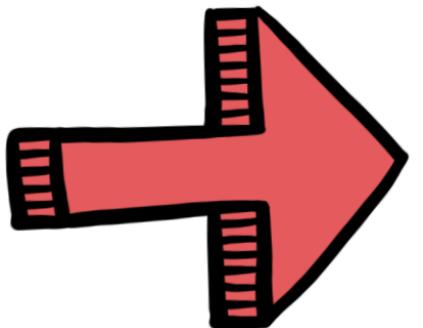
For this project, we will focus on **network marketing business** by analysing membership data and transaction data during 2021 - 2023 to increase sales and membership through personalized marketing by recommending the right products to the right customers.

To achieve the objective, it is important to comprehend customer behavior and divide them into suitable groups based on common characteristics, and purchasing patterns to take appropriate actions of each customer groups such as suggest different marketing strategies and recommend suitable products for each customer within the groups. Therefore, we will create a personalized marketing through **customer segmentation** and **product recommendation** system.

Customer Segmentation



Grouping similar customers together, based on purchasing behavior, demographic, preference and other information.



Product Recommendation

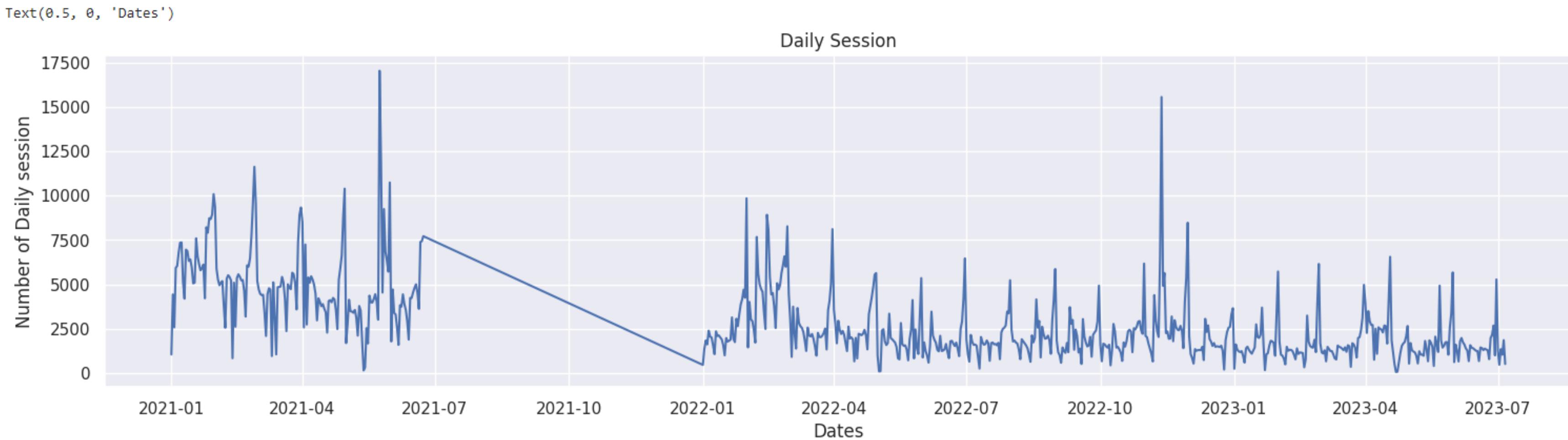


Product listings that are customized for individual customer based on their behavior and preferences, or the behavior and preferences of similar customers to suggest products that customers might like to buy or engage with.

3

DATA PRE-PROCESSING

Overview Analysis – Transaction



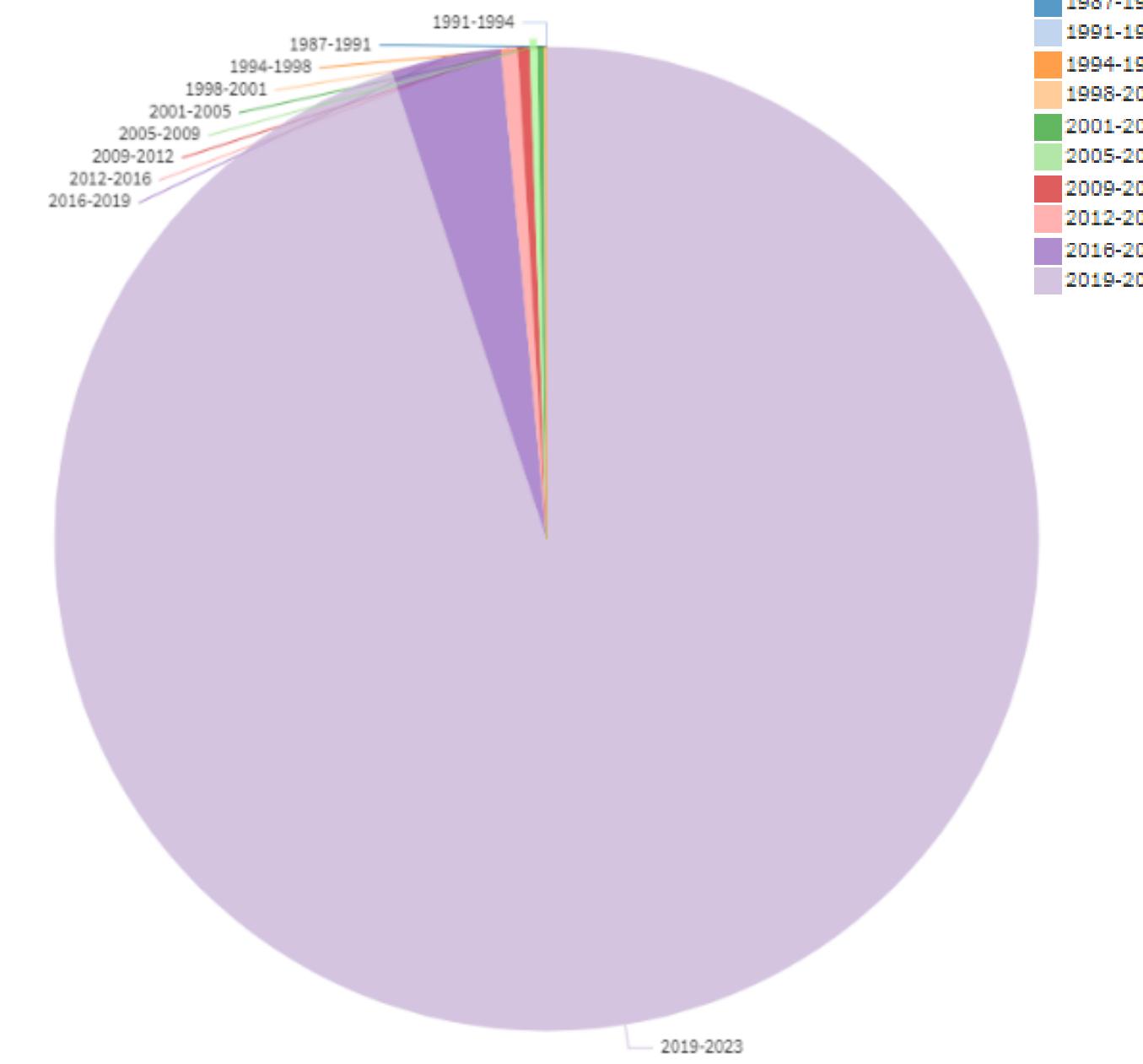
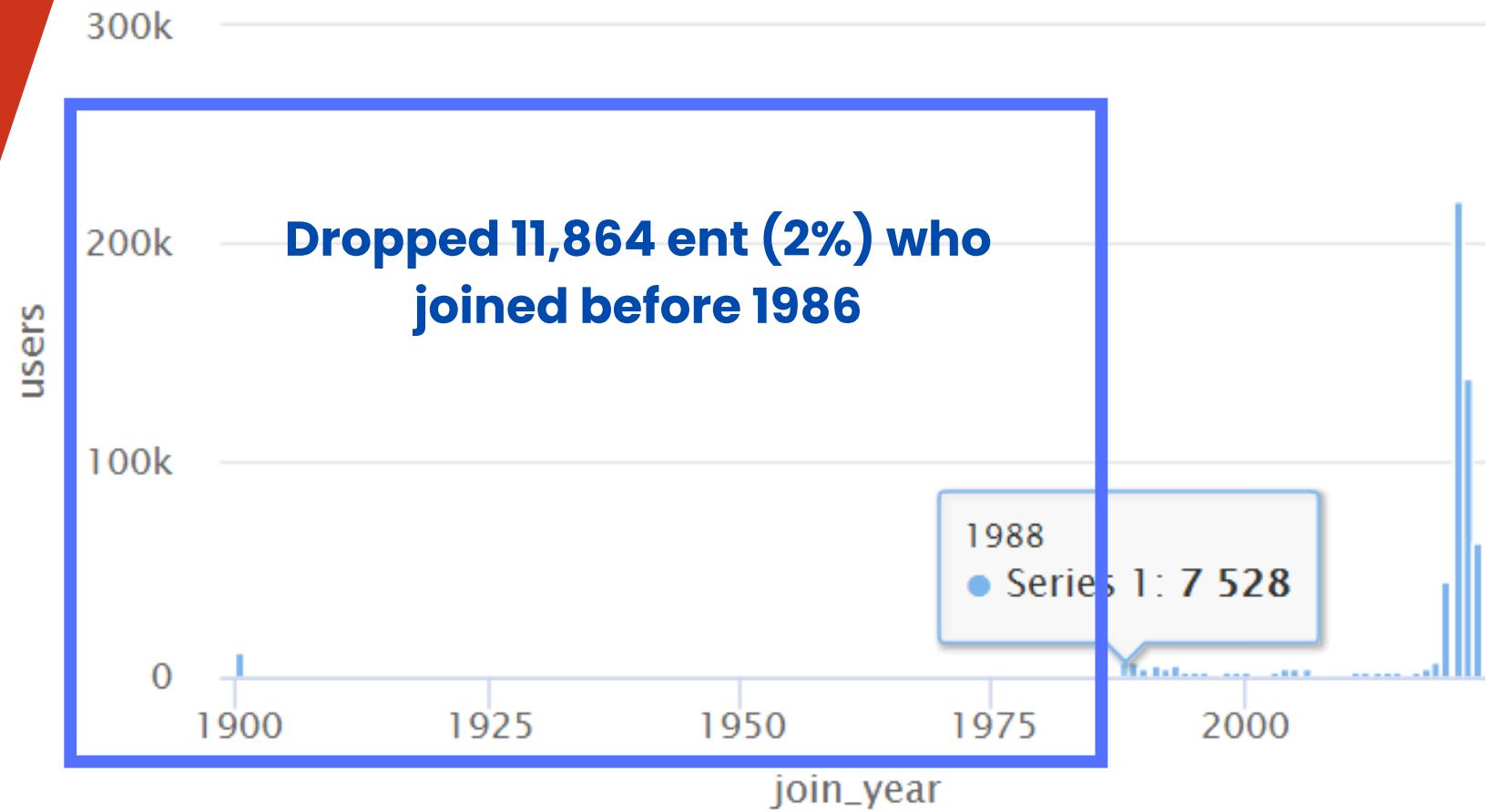
Transaction Table

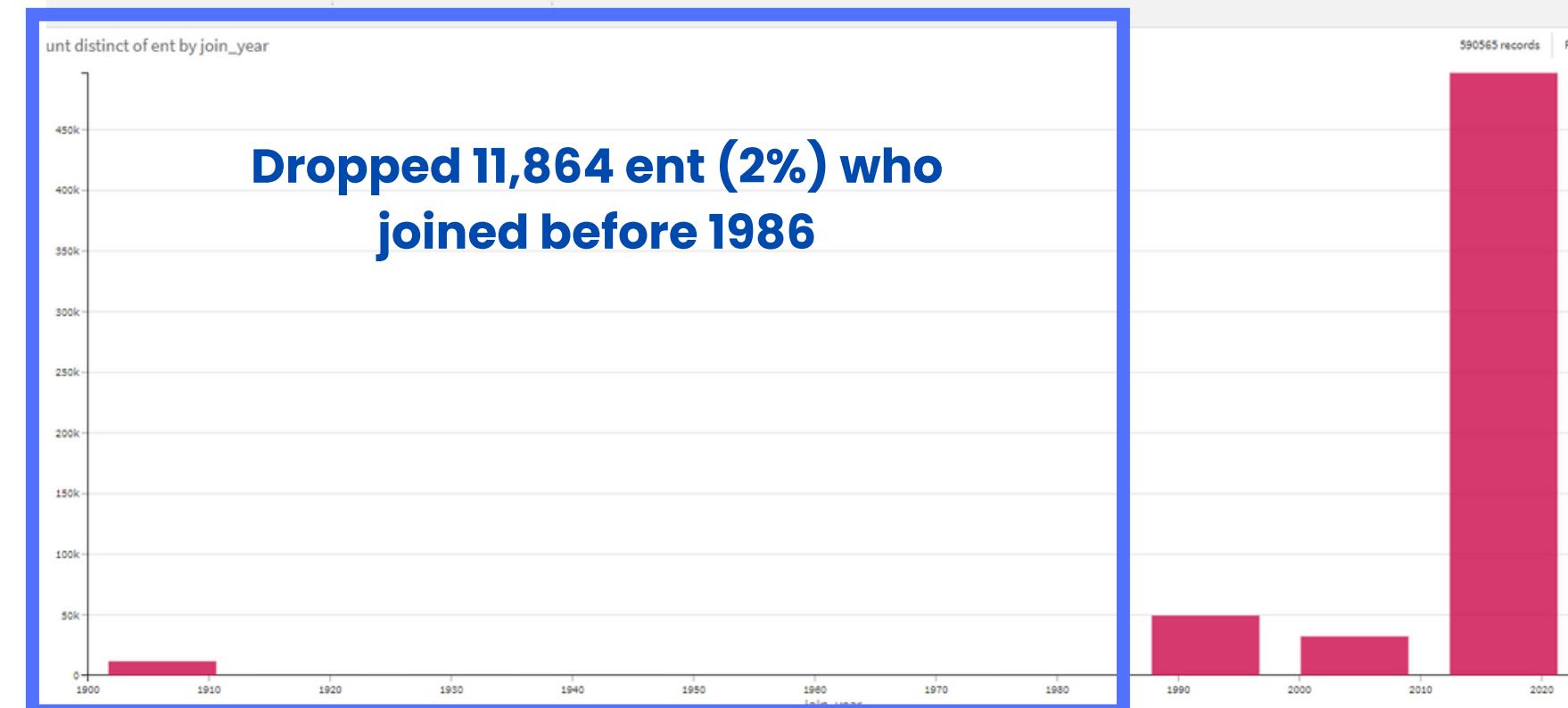
Year	2021	2022	2023
Transaction	1,048,575	995,632	362,109
Ent	325,627	272,550	101,271

The reasons why we don't use 2021 transaction:

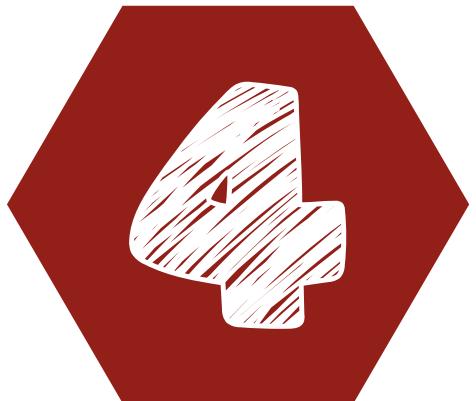
- 1 Missing transaction of FY2021 during second half year.
- 2 Rapidly changing of customer behavior.

Overview Analysis - Member





ตัด User ออก เนื่องจากบริษัทเริ่มก่อตั้งที่ปี 1986 (11,864 ent - 2%) และตัด user ที่ไม่ transaction 2022-2023 โดย user ที่เหลือ = 263,482 user (55.38%)



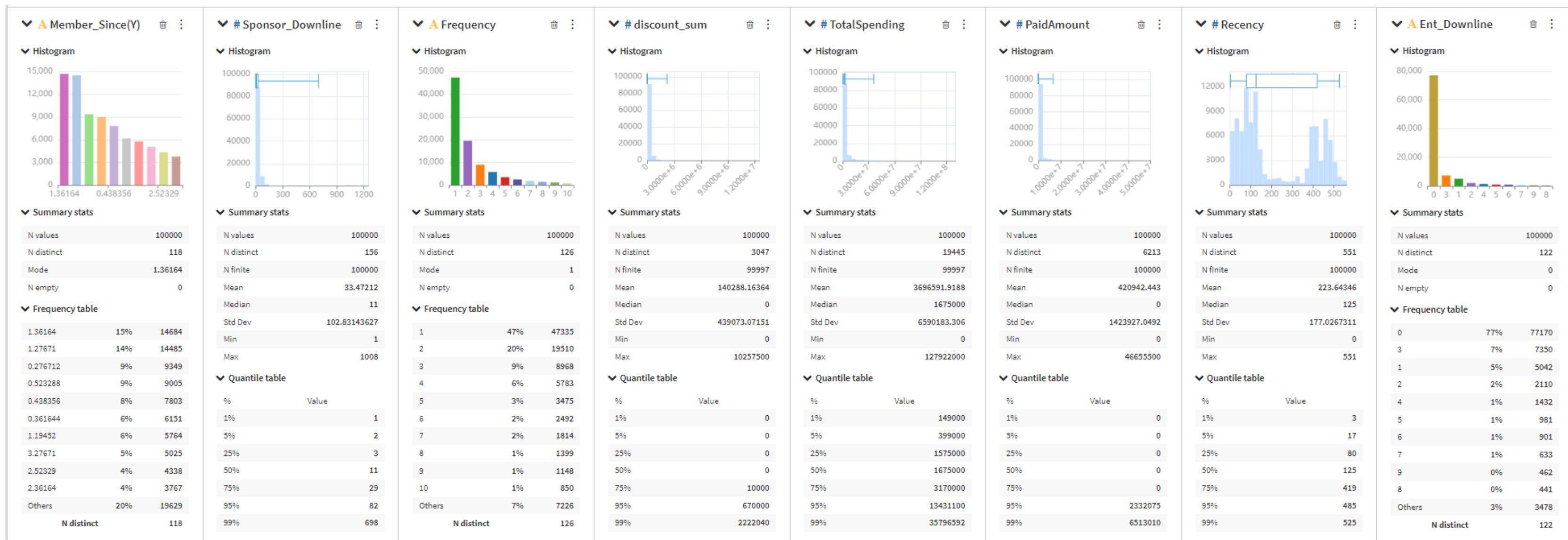
EXPLORATORY DATA ANALYSIS (EDA)

Exploratory Data Analysis



LIVE
LEARN
LOVE

Sanity Check



Understanding dataset



Data Type



Date Time Format

Verify Data



Outlier check



Missing Value

5

DATA PREPARATION

Features Preparation (segmentation)

ORIGINAL_STATUS

Distinct Customers code

MEMBER_SINCE (Y)

Today Date(July 2023) - Registered date

SPONSOR_DOWNLINE

Count ent by Sponsor

ENT_DOWNLINE

Count others ent who undering focus ent

TOTALSPENDING

Sum of total_spending by ent

FREQUENCY

Count transaction

RECENCY

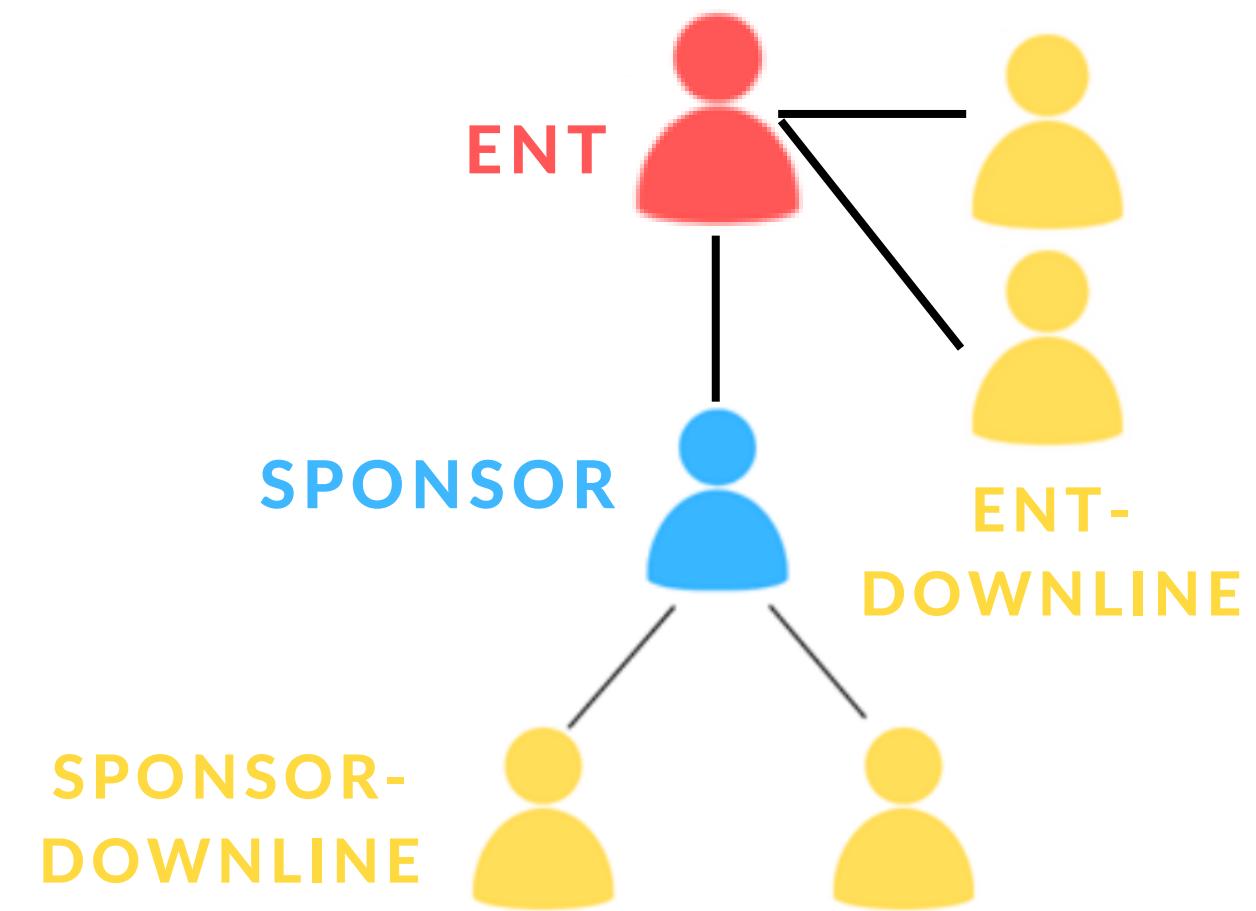
Last date - Payment date

DISCOUNT_SUM

Sum of discount by ent

PAIDAMOUNT

Sum of paid_amount by ent

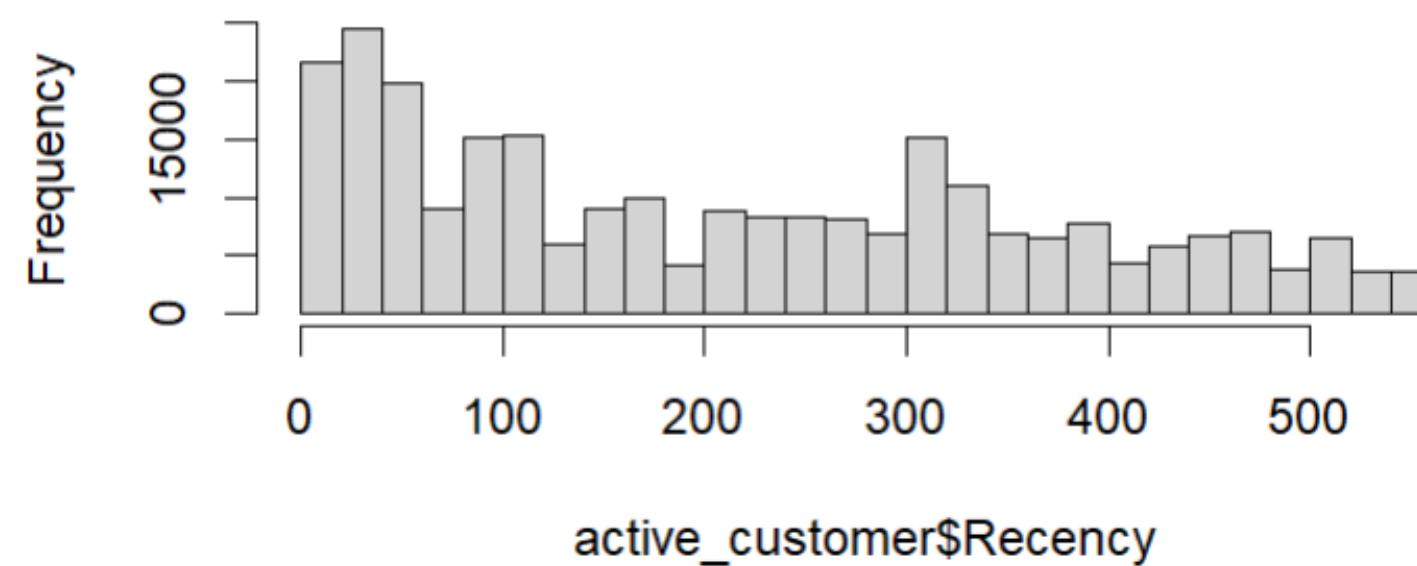


Exploratory Data Analysis

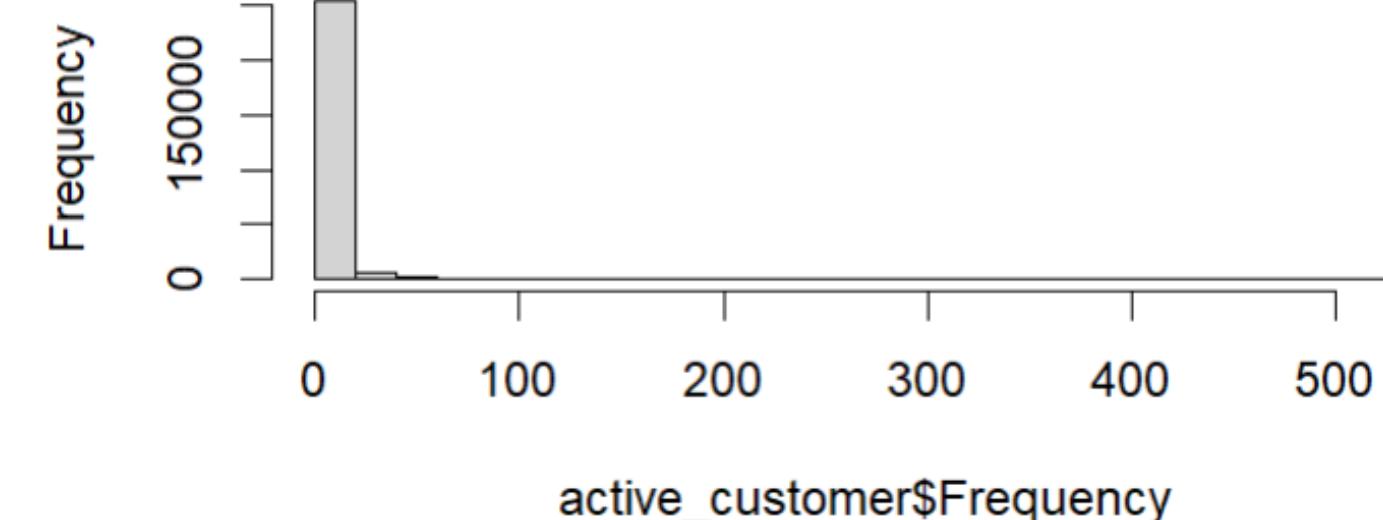
Correlation matrix on 8 variables (Pearson) No split ▾

	Member_Since(Y)	Sponsor_Downline	Frequency	discount_sum	TotalSpending	PaidAmount	Recency	Ent_Downline
Member_Since(Y)	1.000	-0.074	0.403	0.307	0.441	0.148	-0.457	0.247
Sponsor_Downline	-0.074	1.000	-0.000	-0.023	-0.038	-0.000	0.189	0.036
Frequency	0.403	-0.000	1.000	0.607	0.815	0.436	-0.244	0.472
discount_sum	0.307	-0.023	0.607	1.000	0.781	0.446	-0.135	0.435
TotalSpending	0.441	-0.038	0.815	0.781	1.000	0.520	-0.278	0.489
PaidAmount	0.148	-0.000	0.436	0.446	0.520	1.000	0.024	0.269
Recency	-0.457	0.189	-0.244	-0.135	-0.278	0.024	1.000	-0.137
Ent_Downline	0.247	0.036	0.472	0.435	0.489	0.269	-0.137	1.000

Histogram of active_customer\$Recency



Histogram of active_customer\$Frequency

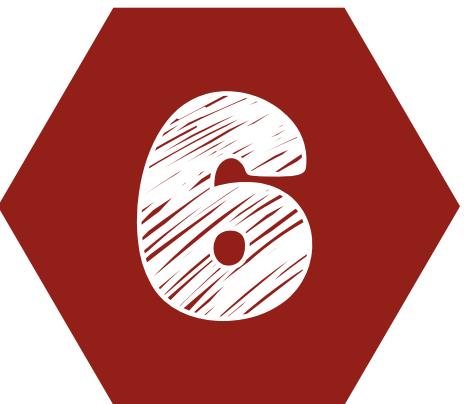


Customer Single View



LIVE
LEARN
LOVE

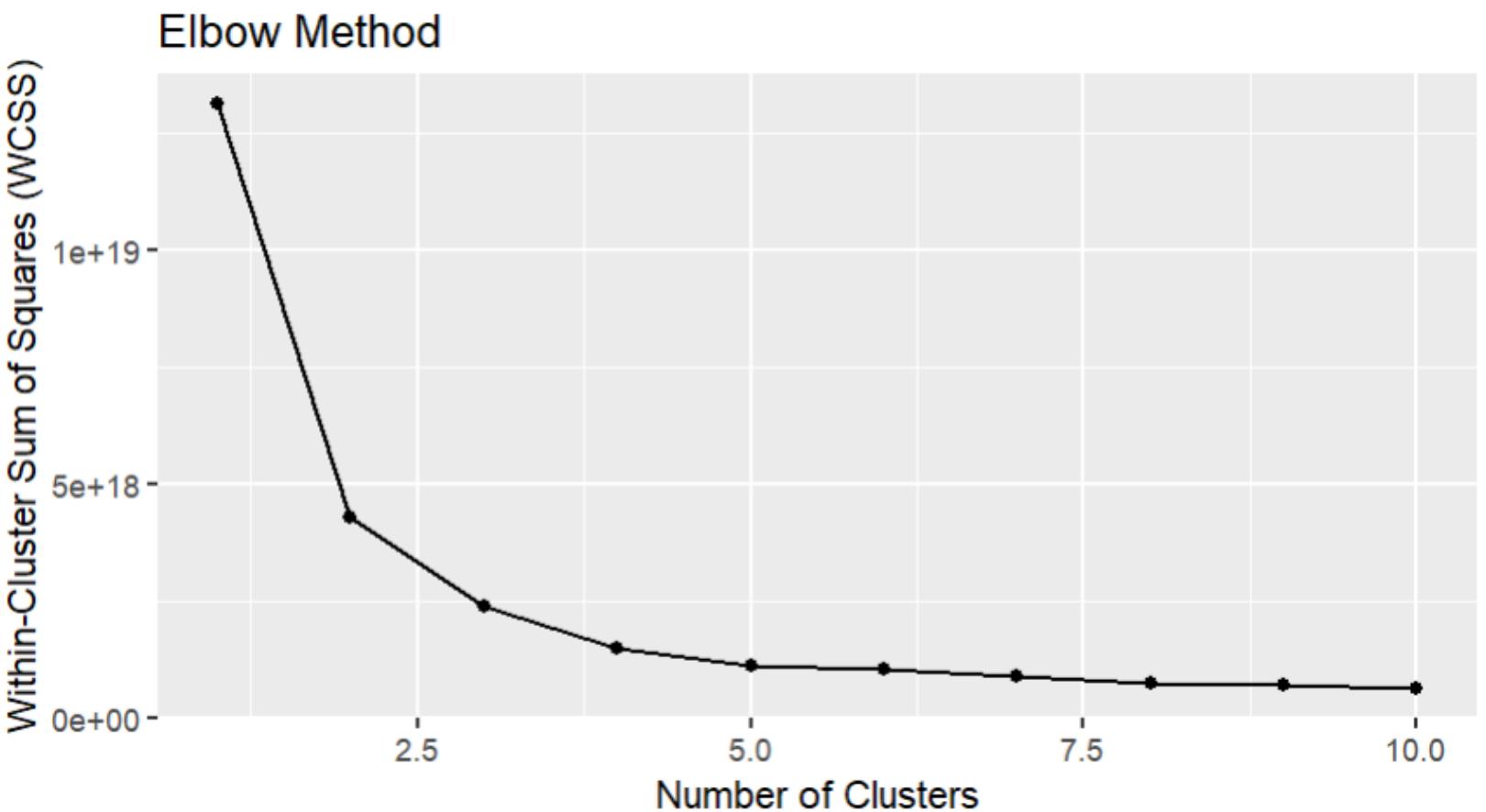
col_0	ent	original_status	join_month	join_year	sponsor	join_date	join_date_parsed	Member_Since(Y)	Sponsor_Downline	Frequency	discount_sum	TotalSpending	PaidAmount	Recency	Ent_Downline	
	string	string	bigint	bigint	string	string	date	double	bigint	bigint	double	double	double	bigint	bigint	
Text	Text	Text	Integer	Integer	Text	Date (unparsed)	Date	Decimal	Integer	Integer	Decimal	Decimal	Decimal	Integer	Integer	
39	TCC41J3QZE0	-		1	2020	TC4Z3J17	2020-01-01	2020-01-01T00:00:00.000Z	3.526027397260274	36	4	0.0	2440000.0	610000.0	96	1
40	TCC4RC4RRE7	-		1	2020	T3341WZK	2020-01-01	2020-01-01T00:00:00.000Z	3.526027397260274	107	9	0.0	2994000.0	0.0	47	2
41	TCC4RCWEZJI	-		1	2020	TCC444EQ4Z5	2020-01-01	2020-01-01T00:00:00.000Z	3.526027397260274	90	6	2500.0	2065500.0	0.0	25	0
42	TCC4R4CJRCI	-		1	2020	T3QWWQ12	2020-01-01	2020-01-01T00:00:00.000Z	3.526027397260274	8	9	336500.0	3065500.0	0.0	21	0
43	TCC4JE34J30	-		1	2023	TRE4RRE2	2023-01-01	2023-01-01T00:00:00.000Z	0.5232876712328767	31	1	0.0	50000.0	0.0	387	0
44	TCC4Q1C44W0	-		1	2022	TCC4RQZ4JWY	2022-01-01	2022-01-01T00:00:00.000Z	1.5232876712328767	13	8	2500.0	2886500.0	249000.0	9	0
45	TCC4WR1Q3Z5	-		1	2021	TZWECERB	2021-01-01	2021-01-01T00:00:00.000Z	2.5232876712328767	3	3	0.0	1663000.0	0.0	41	0
46	TCC4WRWZWZ5	-		1	2021	TC13WWW5	2021-01-01	2021-01-01T00:00:00.000Z	2.5232876712328767	12	7	0.0	2112000.0	0.0	8	0
47	TCC4WRZ41ZB	-		1	2021	TZ3Z4R3P	2021-01-01	2021-01-01T00:00:00.000Z	2.5232876712328767	57	4	0.0	1592000.0	0.0	36	0
48	TCC4WU43WQ0	-		1	2021	T3R14437	2021-01-01	2021-01-01T00:00:00.000Z	2.5232876712328767	90	6	2500.0	1625500.0	0.0	209	0
49	TCC4WUQ3JZ0	-		1	2021	TZZEW3QK	2021-01-01	2021-01-01T00:00:00.000Z	2.5232876712328767	72	10	0.0	7669000.0	0.0	23	0
50	TCC4WUWQ4ZU	-		1	2022	TCC44WEJZ0	2022-01-01	2022-01-01T00:00:00.000Z	1.5232876712328767	8	7	0.0	1882000.0	159000.0	24	0
51	TCC4WZECJJK	-		1	2021	TCC44R1CCE2	2021-01-01	2021-01-01T00:00:00.000Z	2.5232876712328767	2	2	0.0	1503000.0	0.0	287	0
52	TCC4WZEQCJK	-		1	2021	TCC4RWECCZF	2021-01-01	2021-01-01T00:00:00.000Z	2.5232876712328767	26	11	0.0	4063000.0	0.0	66	1
53	TCC4WZRQ31F	-		1	2021	TCC44JZCWWB	2021-01-01	2021-01-01T00:00:00.000Z	2.5232876712328767	2	10	0.0	4389500.0	572500.0	4	0
54	TCC4WZZ1JE0	-		1	2021	TCC4JWCZEEN	2021-01-01	2021-01-01T00:00:00.000Z	2.5232876712328767	1	2	0.0	1803000.0	0.0	62	0
55	TCC43113E15	-		1	2021	TCC4QRQZWJF	2021-01-01	2021-01-01T00:00:00.000Z	2.5232876712328767	4	4	0.0	2845000.0	0.0	103	0
56	TCC431WJ43K	-		1	2021	TCC4J3WQ332	2021-01-01	2021-01-01T00:00:00.000Z	2.5232876712328767	1	5	2500.0	1631500.0	0.0	49	0
57	TCC43RQ3QCP	-		1	2021	TCC4QJZEQJ2	2021-01-01	2021-01-01T00:00:00.000Z	2.5232876712328767	27	9	22500.0	3385500.0	0.0	3	0
58	TCC43RW4ERP	-		1	2021	TCC4WZJC3JU	2021-01-01	2021-01-01T00:00:00.000Z	2.5232876712328767	1	13	0.0	4355000.0	0.0	38	0
59	TCC43RW43P	-		1	2021	TCC4J34ERQB	2021-01-01	2021-01-01T00:00:00.000Z	2.5232876712328767	10	15	17500.0	4525500.0	0.0	55	0
60	TCC43J4QERO	-		1	2021	TCC4QQQEJC5	2021-01-01	2021-01-01T00:00:00.000Z	2.5232876712328767	39	5	0.0	2745000.0	0.0	49	0
61	TCC43JE4JWI	-		1	2021	TCCCCC4K	2021-01-01	2021-01-01T00:00:00.000Z	2.5232876712328767	235	3	0.0	1755000.0	0.0	55	0
62	TCC43JEZJWB	-		1	2021	TCC4W1QJC3P	2021-01-01	2021-01-01T00:00:00.000Z	2.5232876712328767	6	13	0.0	2784000.0	0.0	32	1
63	TCC43J11Q3F	-		1	2021	TCC4WRC3W3B	2021-01-01	2021-01-01T00:00:00.000Z	2.5232876712328767	1	5	10000.0	1691000.0	282000.0	158	0
64	TCC43J1RWJY	-		1	2021	TCC4J44C1CP	2021-01-01	2021-01-01T00:00:00.000Z	2.5232876712328767	56	5	2500.0	1608500.0	0.0	56	0
65	TCC43J1QER2	-		1	2021	TCC4WC34E3P	2021-01-01	2021-01-01T00:00:00.000Z	2.5232876712328767	62	9	0.0	2703000.0	0.0	32	0
66	TCC43J13ECK	-		1	2021	TCC4QJRW1WB	2021-01-01	2021-01-01T00:00:00.000Z	2.5232876712328767	1	7	17500.0	4428500.0	2349000.0	60	



CUSTOMER SEGMENTATION

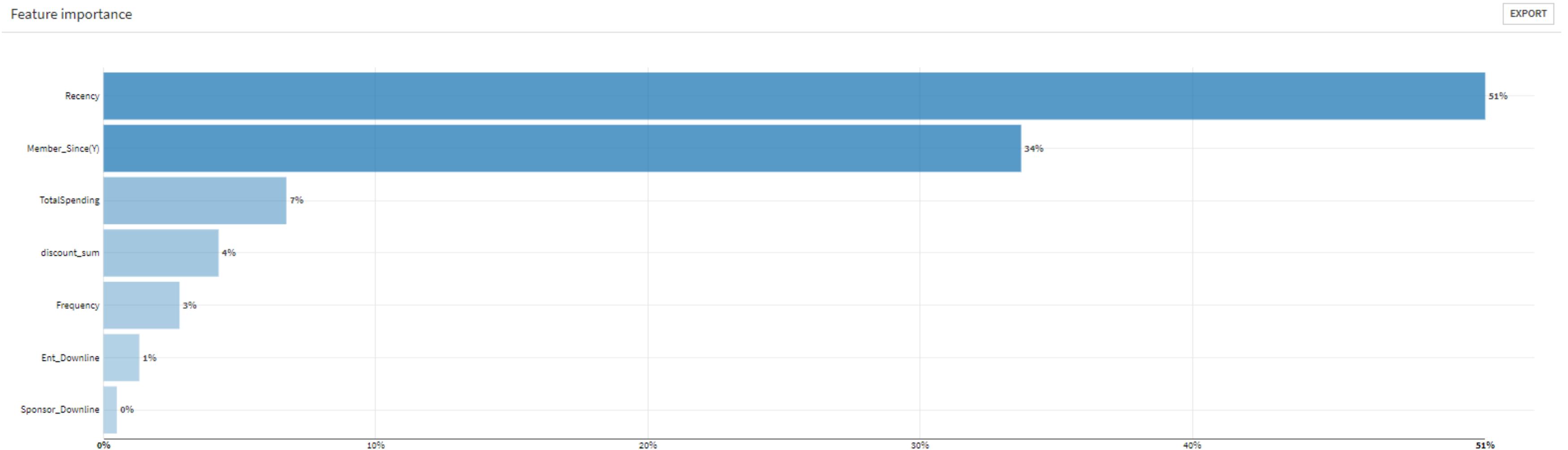
K-mean clustering & N calculation

<input type="checkbox"/> SESSION 2		
<input type="checkbox"/> ● KMeans (k=7) (s2)	0.333	☆
<input type="checkbox"/> ● KMeans (k=5) (s2)	0.428	☆
<input checked="" type="checkbox"/> ● KMeans (k=4) (s2)	0.431	☆
<input type="checkbox"/> ● KMeans (k=3) (s2)	0.433	★
<input checked="" type="checkbox"/> ● KMeans (k=6) (s2)	0.359	☆



we decide to split segment in **3 tiers** due to the Elbow method and the silhouette score

Features Importance (segmentation)



**STANDARD
RESCALING**

**MIN - MAX
RESCALING**

RECENCY
DISCOUNT_SUM

MEMBER_SINCE (Y)
ENT_DOWNLINE

TOTALSPENDING

SPONSOR_DOWNLINE
FREQUENCY

Interpreting Results

KMeans (k=3) (s2) 

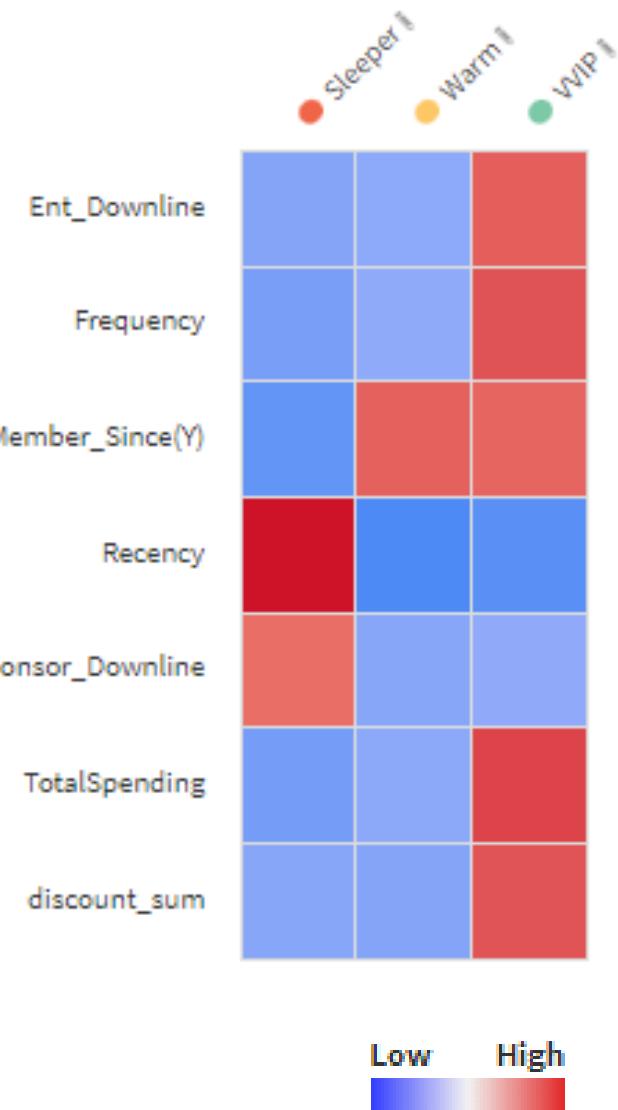
KMeans (k=3)

Drop outliers

Sleeper  116626 (44.26%)

Warm  133596 (50.70%)

VVIP  10707 (4.06%)



Sleeper



- Highest recency (**Avg. 368 days**)
- Highest sponsor_downline (**Avg. 48 ents**)
- Low frequency (**Avg. 2 time**)
- Low Total_spending (**Avg. 1.89 MB**)

Warm



- High Member_since (**Avg. 2 Years**)
- Low recency (**Avg. 94 days**)
- Low frequency (**Avg. 4 time**)
- Low Total_spending (**Avg. 3.29 MB**)

VVIP

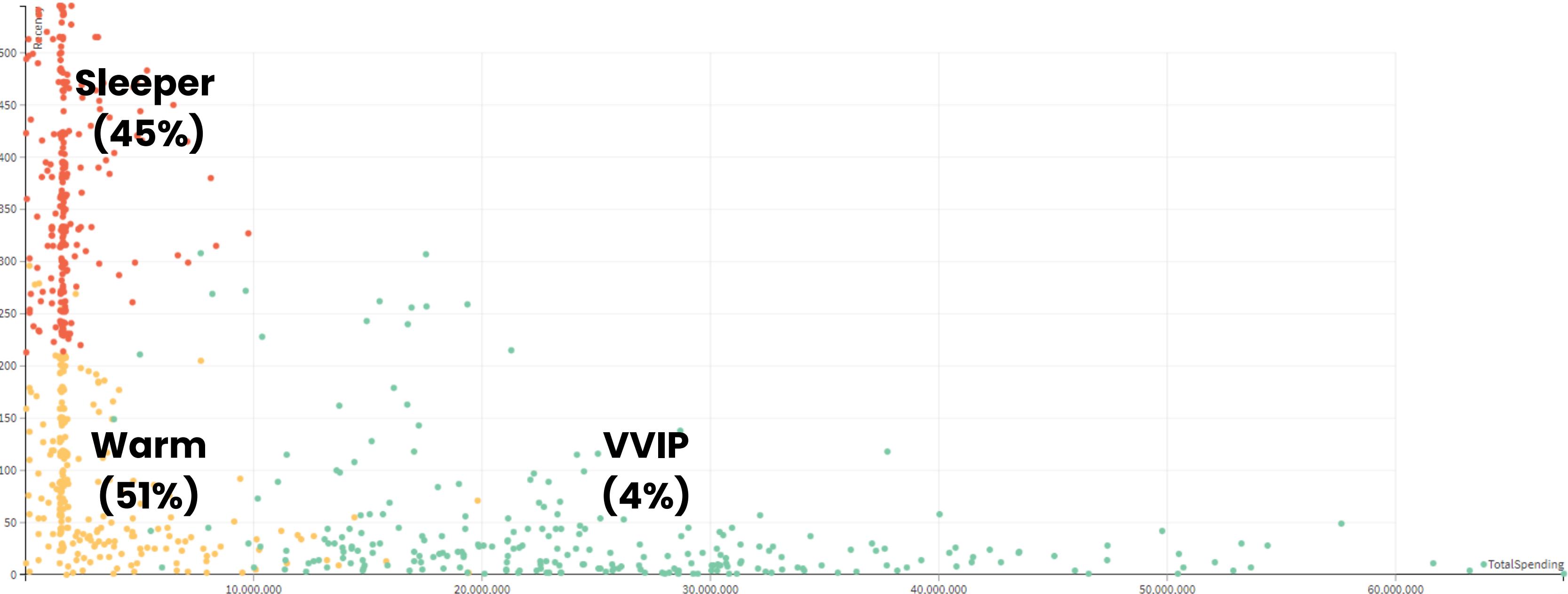


- High Frequency (**Avg. 25 time**)
- High Member_since (**Avg. 4 years**)
- High Total_spending (**Avg. 11.41 MB**)
- High Ent_Downline (**Avg. 12 pers**)

Scatter plot

Scatter plot

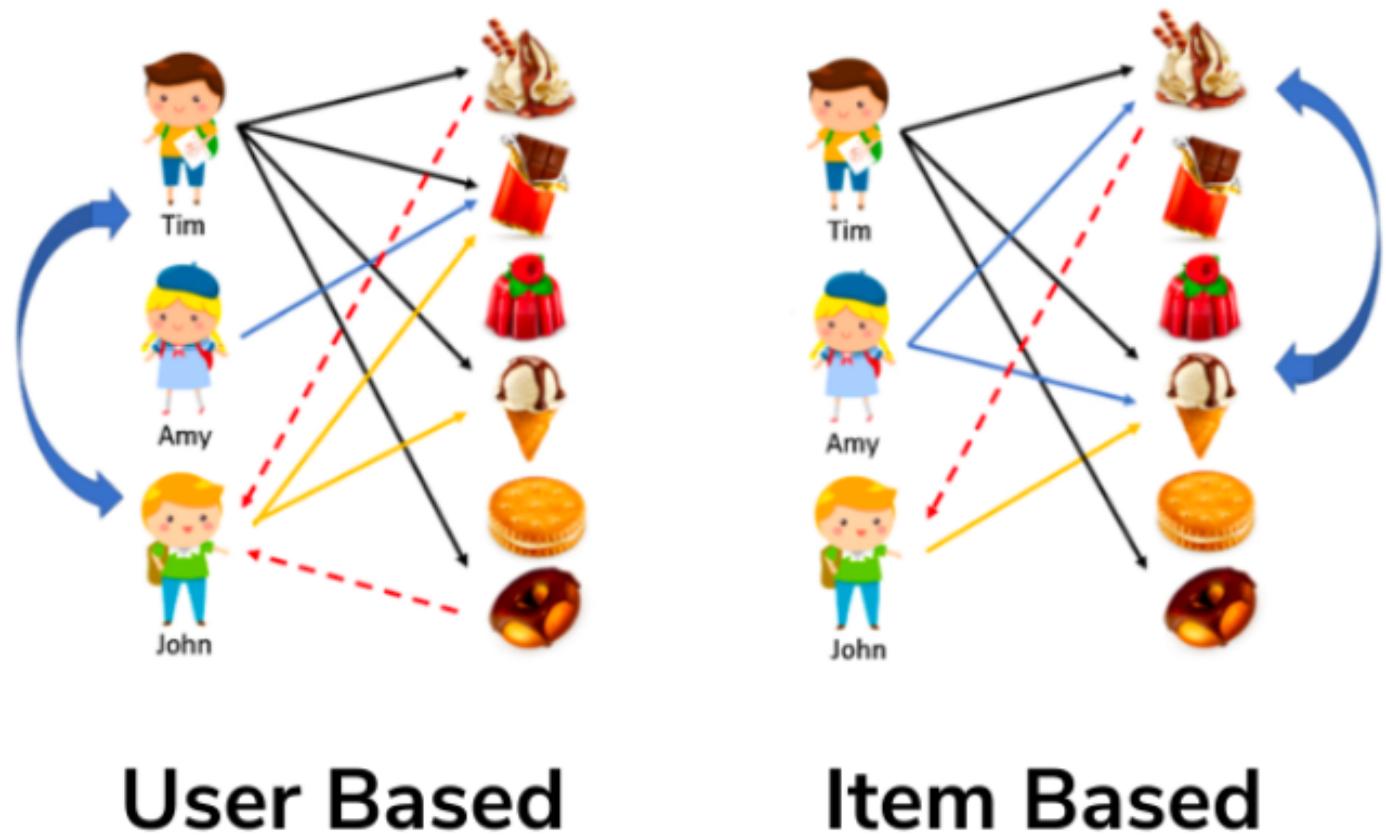
↑ Vertical axis: Recency → Horizontal axis: TotalSpending



7

PRODUCT RECOMMENDATION

Product Recommendation : User-Based Collaborative Filtering



User who have similar behavior purchase (Look - A - Like)

The recommendation is done by finding similar items based on how users interact with them. For example, items that have the same purchase by a user in are considered similar. To find the item-item similarity matrix.

Product Recommendation : Data preparation



Items



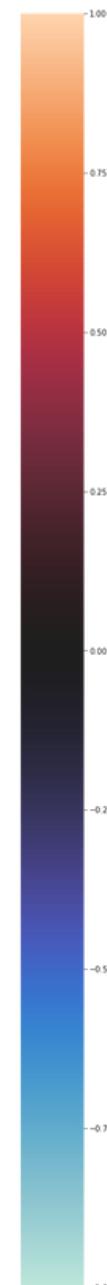
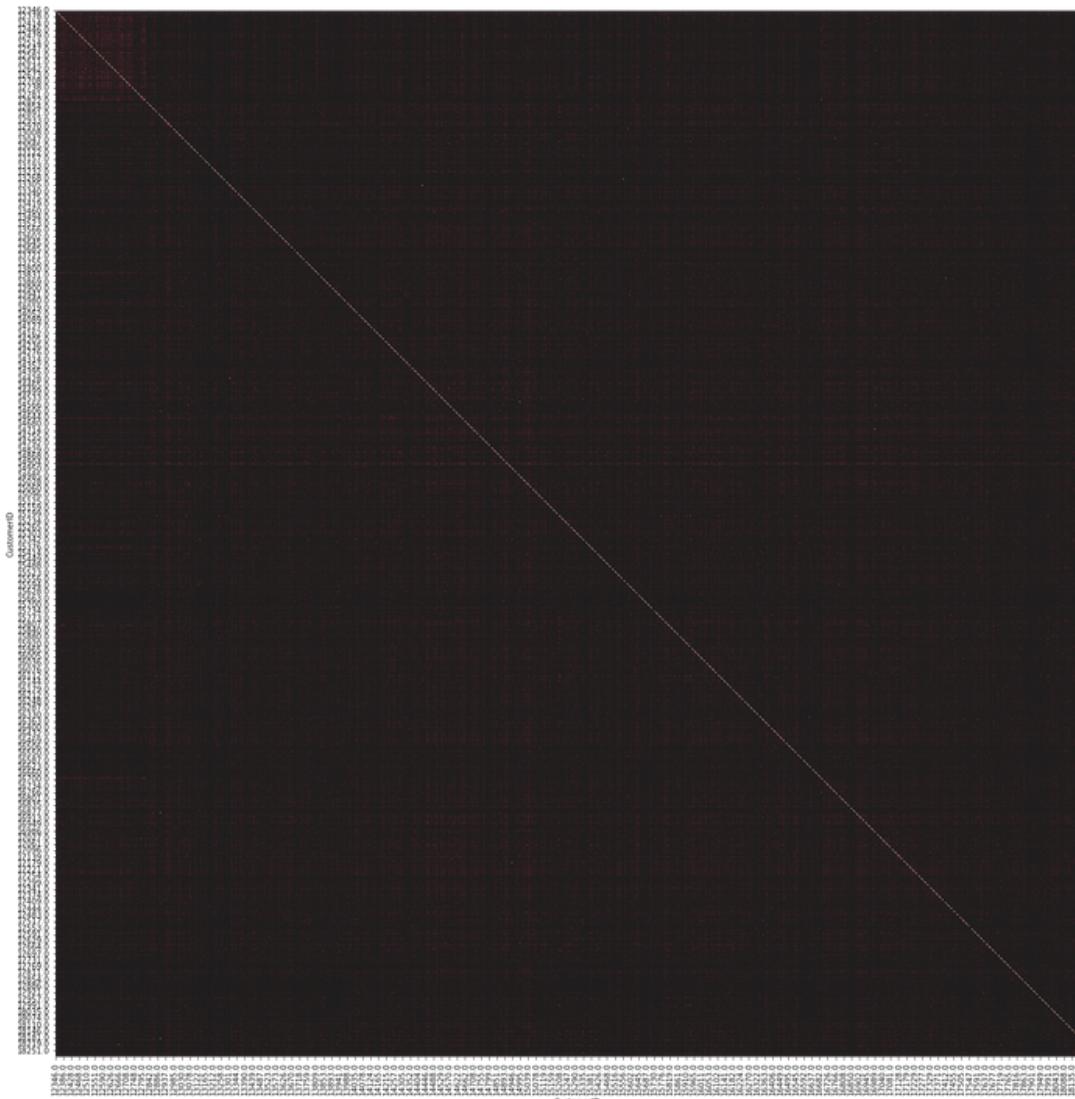
Users

ent	0CECCR	Unknown_2	Unknown_3	Unknown_4	5C4CCE	6CQC41	8C4CCR	9C4CCE	AC4CCR
1 884R4EJZ0	1	0	0	0	0	0	0	0	0
2 884RC4REI	0	0	0	0	0	1	1	1	1
3 8C4ECCCC44F	0	0	0	0	0	0	0	0	0
4 8C4ECCCC3Y	0	0	0	0	0	0	0	0	0
5 8CC1CCCC3E0	0	0	0	0	0	0	0	0	0
6 8CCCCCC4R1F	0	0	0	0	0	2	4	4	4
7 BC113ZJP	0	0	0	0	0	0	0	0	0
8 BC11JW3U	0	0	0	0	0	1	0	0	0
9 BC14413K	0	0	0	0	0	0	0	0	0
10 BC1CW3Z5	0	0	0	0	0	0	0	0	0
11 BC1CWJE7	1	0	0	0	0	0	0	0	0

We prepare the data by cross the user and product to gather and we use the **cosine** to find user basket similarity.

Product Recommendation: Interpret result

Product similarity



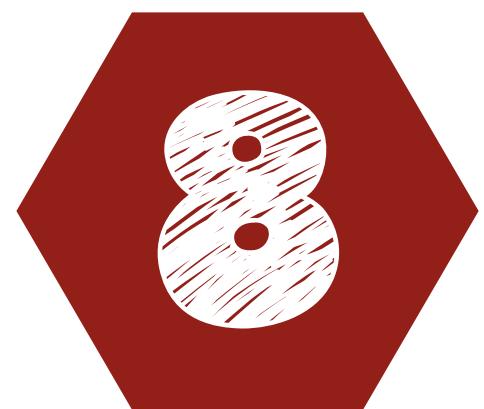
```
links_filtered = links.loc[(links['value'] > 0.5) & (links['edge1'] != links['edge2'])]
links_filtered = links_filtered.sort_values(by='value', ascending=False)
print(links_filtered)
```

	edge1	edge2	value
103680	KC4CCQ	KC4CCW	0.897085
104218	KC4CCW	KC4CCQ	0.897085
239569	PC4CCZ	XCJCCZ	0.849837
137349	XCJCCZ	PC4CCZ	0.849837
103537	KC4CCQ	KC4CCJ	0.848668
27141	KC4CCJ	KC4CCQ	0.848668
159360	5C4C43	5C4C4Z	0.832050
192178	5C4C4Z	5C4C43	0.832050
27142	KC4CCJ	KC4CCW	0.819892
104076	KC4CCW	KC4CCJ	0.819892
148497	BC4C4J	8WCCE	0.754512
147421	8WCCE	BC4C4J	0.754512
25380	KC4C41	KC4C44	0.707107



We select top 5 products with high similarity score to recommend to user from the top purchase product

We can also apply to our segment



ACTIONS AND SUMMARY

Actions - Customer Segmentation

Sleeper (45%)

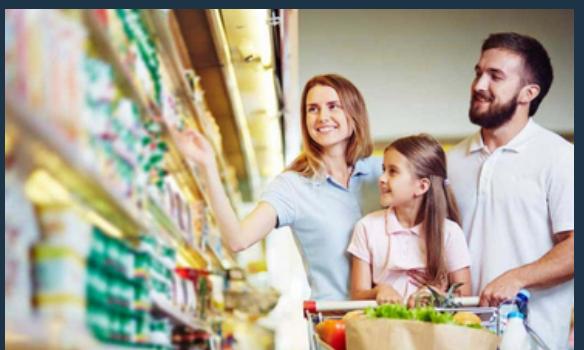


SNOOZE & ALARM

- Identify Churn Indicators (snooze Quarterly)
- Retargeting campaign and/or promotion (special gift if spending amount reach target)
- Offer personalized and proactive customer communication/product recommendation

- Encourage sponsor to take care customers (downline)
- Providing excellent customer service and a positive experience
- Focus on Customer Success and Continuous Improvement

Warm (51%)



Build a Path – EDUCATE – CONNECT

- Cohort Analysis to reduce customer churn
- Provide excellent customer service

- Offer personalized and proactive customer communication/product recommendation
- Continuously contact customer to retain relationship eg. offer class for influencers 101 or online channels for new members can setup their own HDI net work path

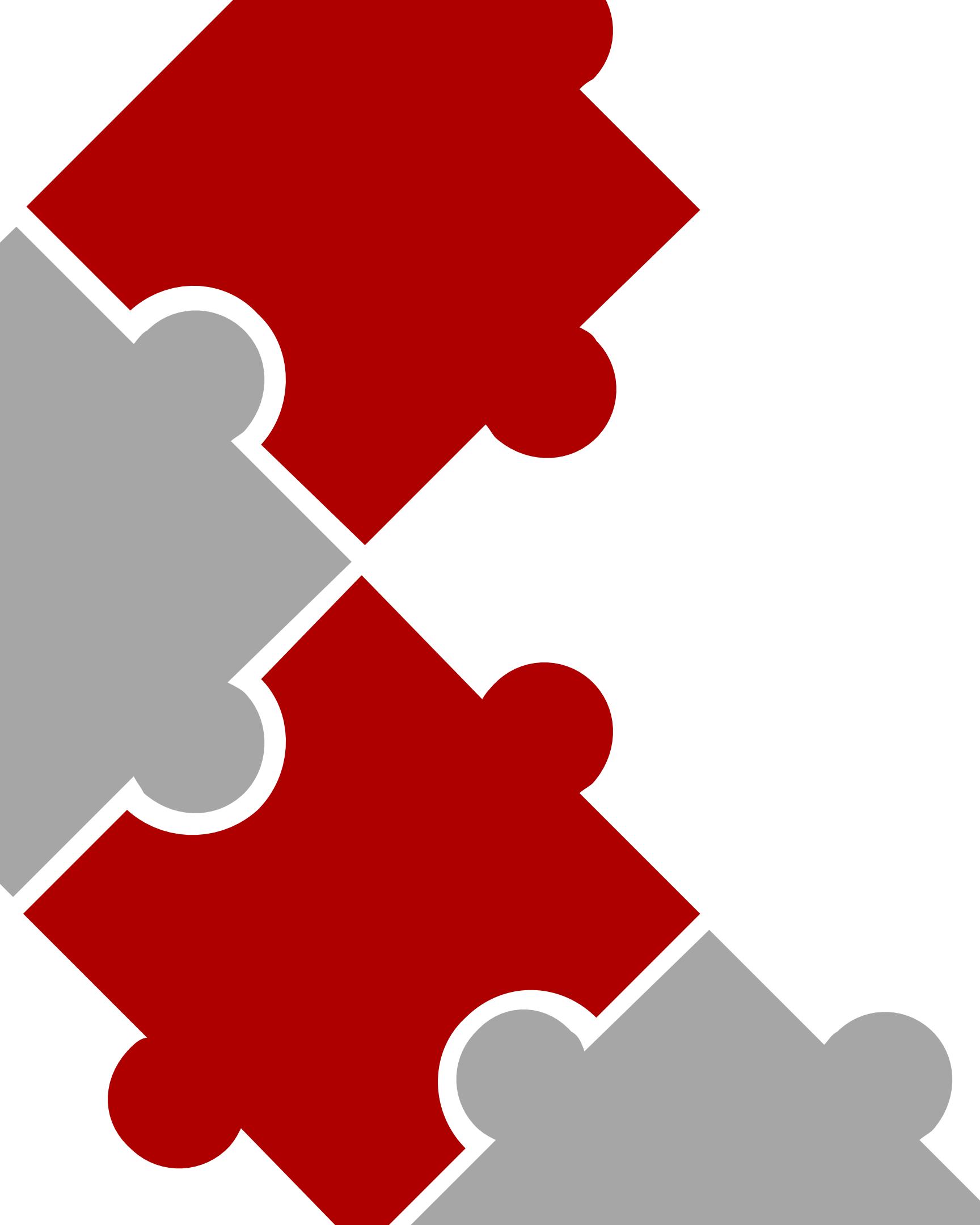
VVIP (4%)



Spotlight & Reward

- Offer personalized and proactive customer communication/product recommendation
- Provide excellent customer service – Customer satisfaction
- VIP Events and Experiences & Reward for loyalty customer

- Long-term relationship management
- Increase Customer Lifetime Value Cross-selling/Up-selling
- Cohort Analysis to reduce customer churn
- Encourage sponsor to take care customers (downline)



OUR TEAM

KHEM WIPASARAPHONG
6510424003

PIRIYA TAWAIJEAN
6510424004

THADSANAN THARARAK
6510424019

CHANAPAT EAKAMONWIWAT
6510424020

WARAPORN JINO
6510424025



Thank You



<https://sg.hdi.com>