



# CENTRAL RETAIL

x

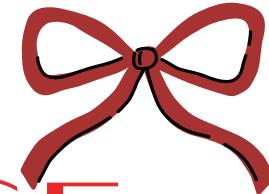
# DotProduct.

*Presenting*  
SIP Case Competition



## Introduction and Motivation

# CENTRAL SURPRISE



### **Overview of the main objectives**

**Main Goal:** Growing Customer Base for Central Retail Store

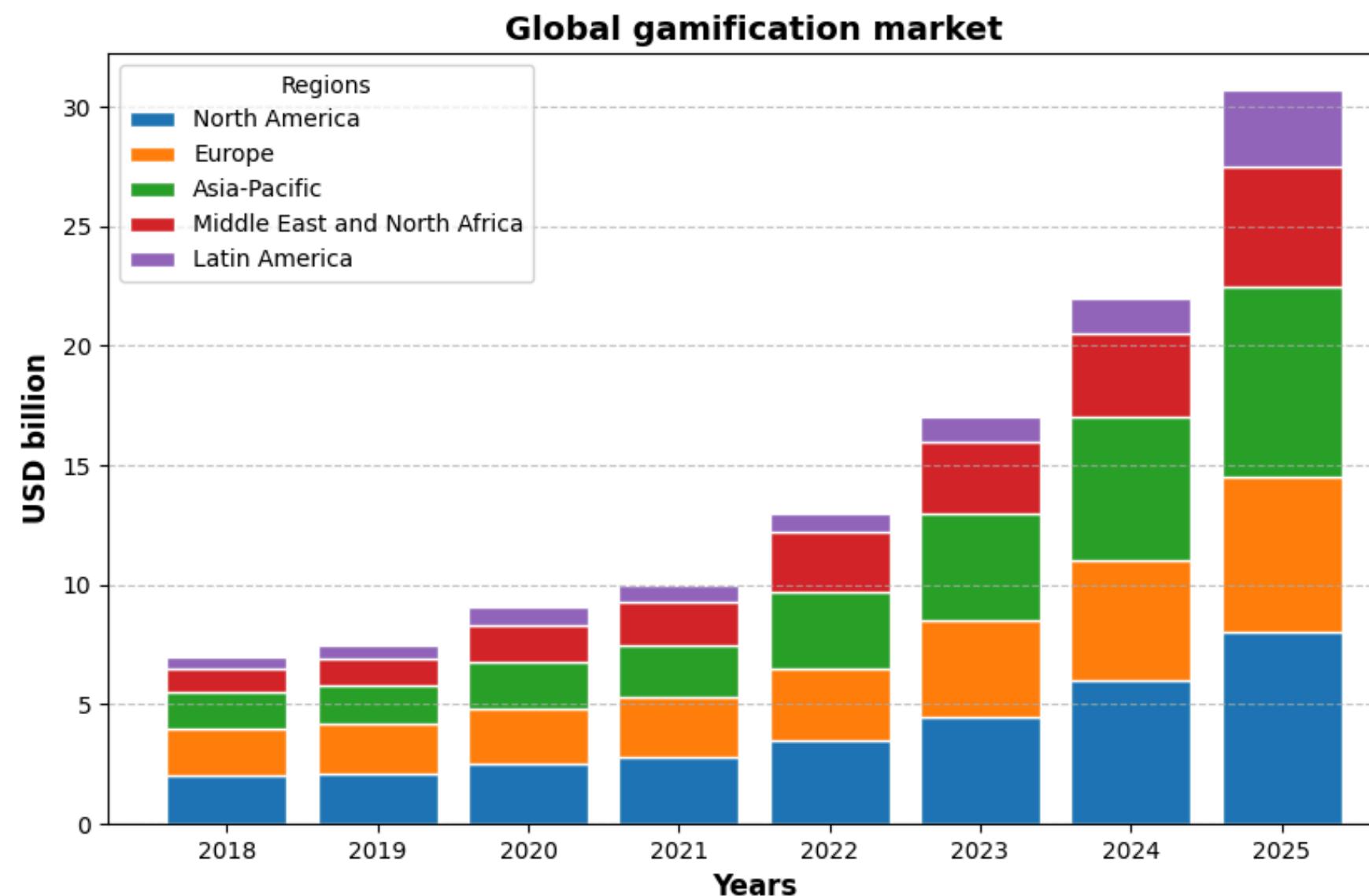
### **Problem:**

Why would customers buy from Central Retail Store?

### **Solution:**

Gamification of retail service exclusively of Central member

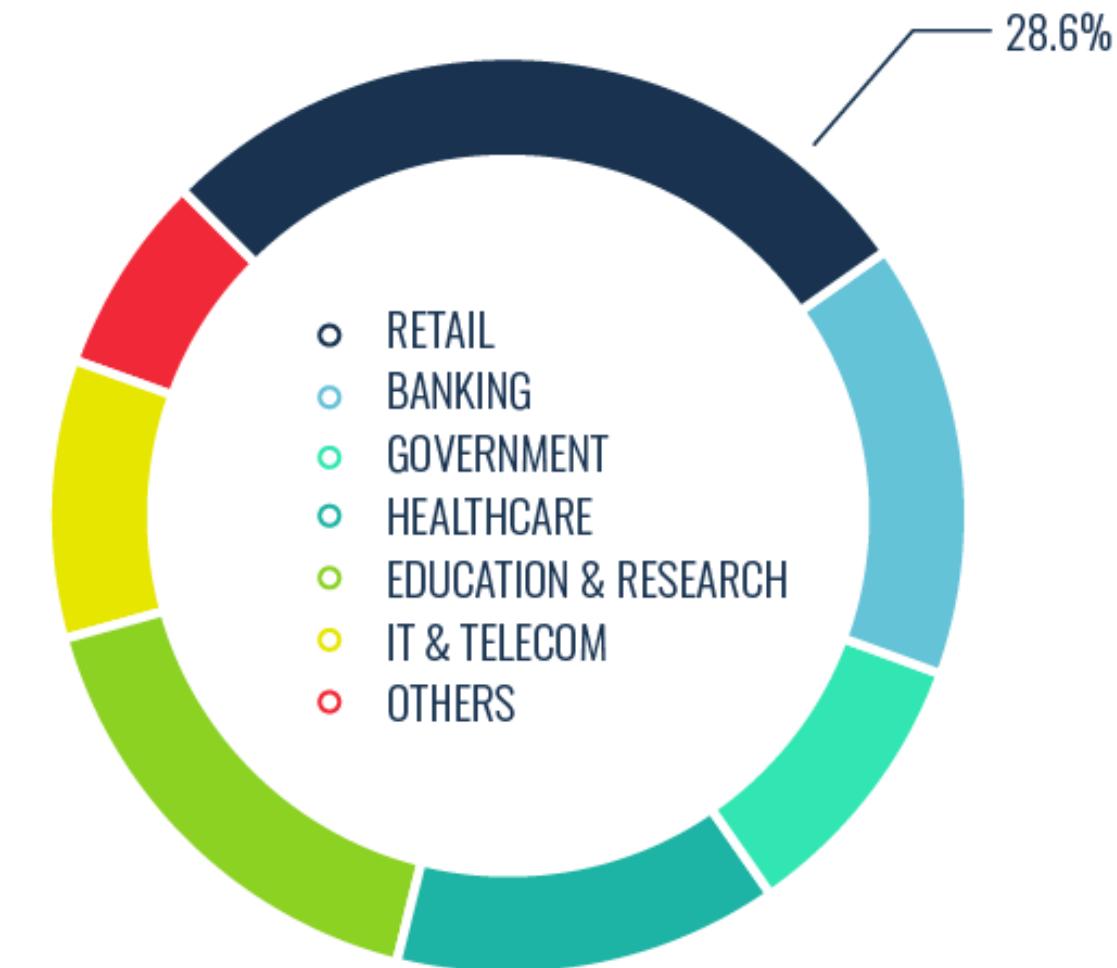
# Why Gamification?



- Global gamification market grow from 9.1 billion in 2020 to an expected USA **30.7 billion USD by 2025**
- 70% of global companies use gamification in one way or another

Source: (Clock, May 27, 2023).

**Global Market Share by Industry**



- Gamification has been successfully adopted by 28.6% of the retail industry in recent times.
- Combining this trend with advances in machine learning is likely to result in continued success.

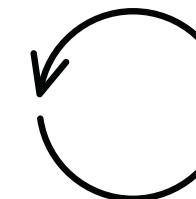
Source: (Fortune, Business Insights, Jun, 2020).

# CENTRAL SURPRISE FEATURE 1#

## Reward Rush

Randomly Select Top Product Combinations

## SEQUENTIAL ANALYSIS MODEL



## Changes weekly

### POTS 'n' PANS

Unlock delicious rewards with Pots 'n' Pans Bonus Points! Earn points on every purchase and redeem them for discounts, special offers, and more.



### BONUS 100 Points

### HATS 'n' HANDBAG

Hats 'n' Handbags Bonus Points! Exclusive stylish purchase earns you points towards fabulous rewards. Start collecting today!



### BONUS 200 Points

## CENTRAL SURPRISE FEATURE 2#

### CENTRAL SURPRISE (Mystery Boxes)

#### Recommendation System Model

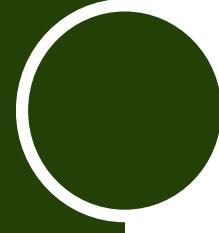
Provide reasonable products  
option to keep customer engage



#### Exclusive Reward



#### Rare Reward



#### Uncommon Reward (Selectable)

Changes daily



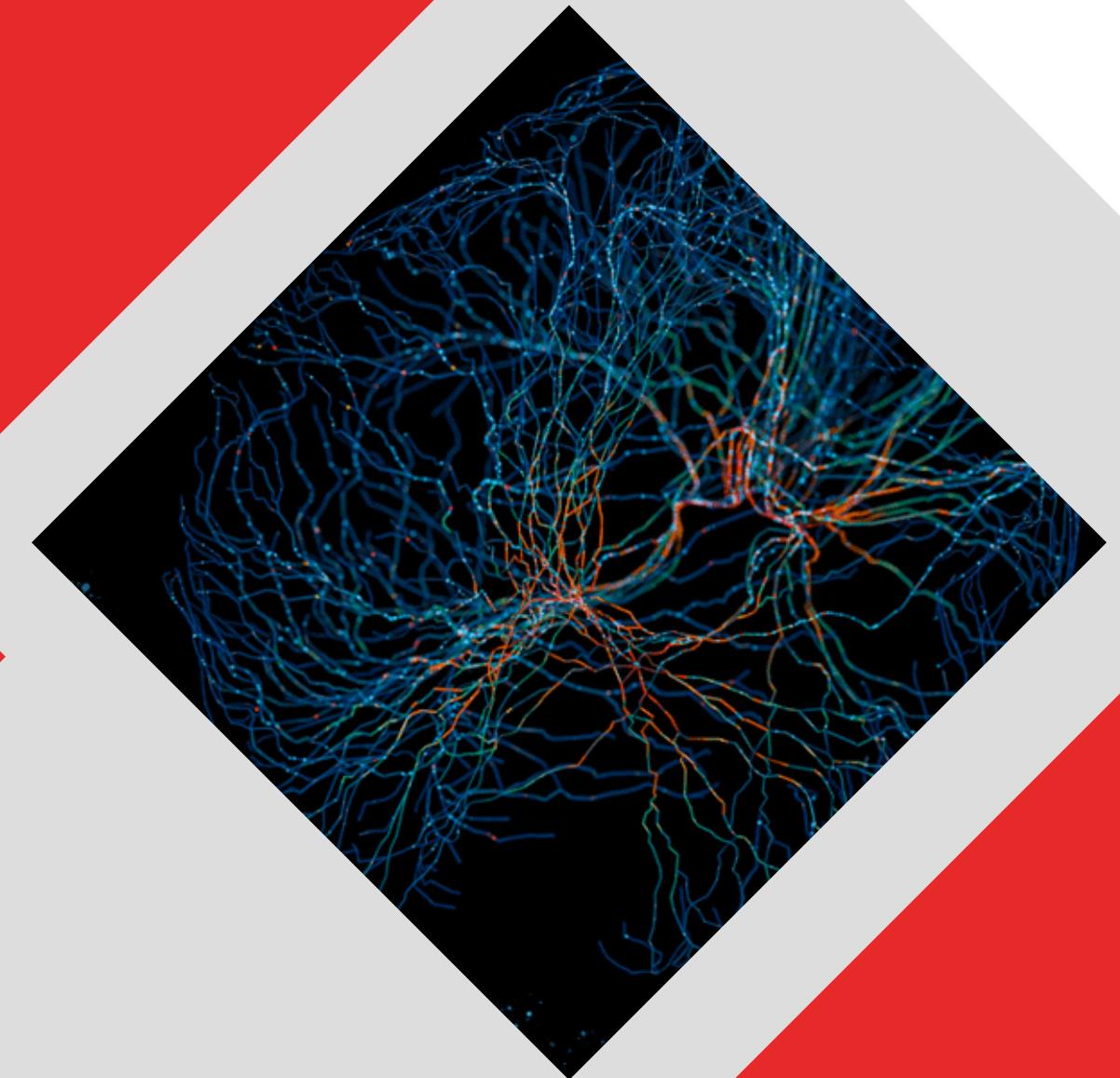
#### Common Reward (Many Items)

Changes daily

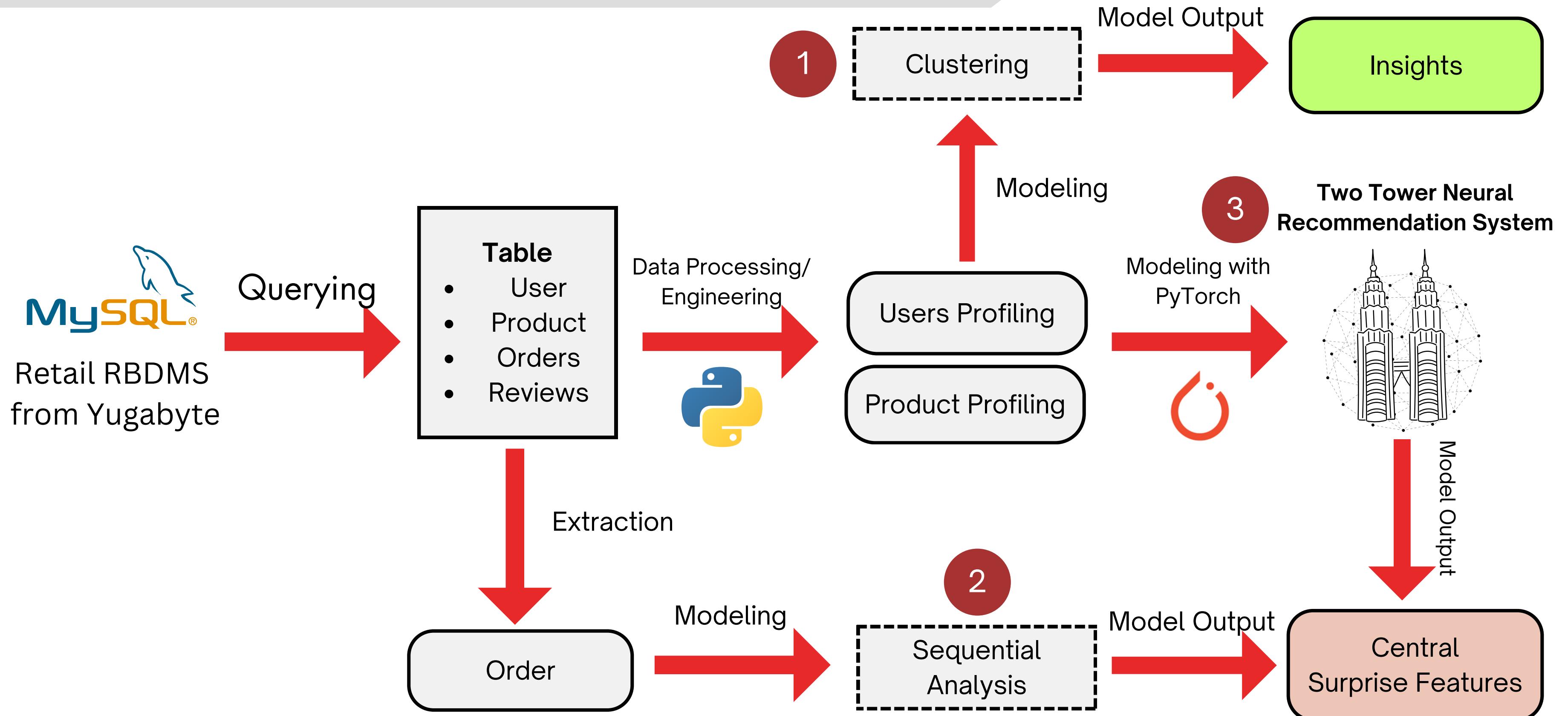




# Implementations



# Data Pipeline For CENTRAL SURPRISE

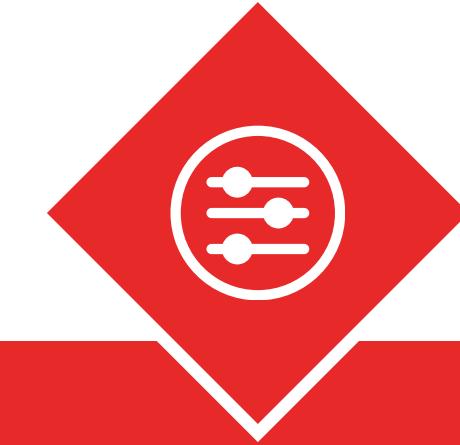


# Data Processing/Engineering



## Users Profiling

- id
- name
- state
- latitude
- longitude
- source
- customer duration
- age
- total spent per product category
- discount usage proportion
- email provider

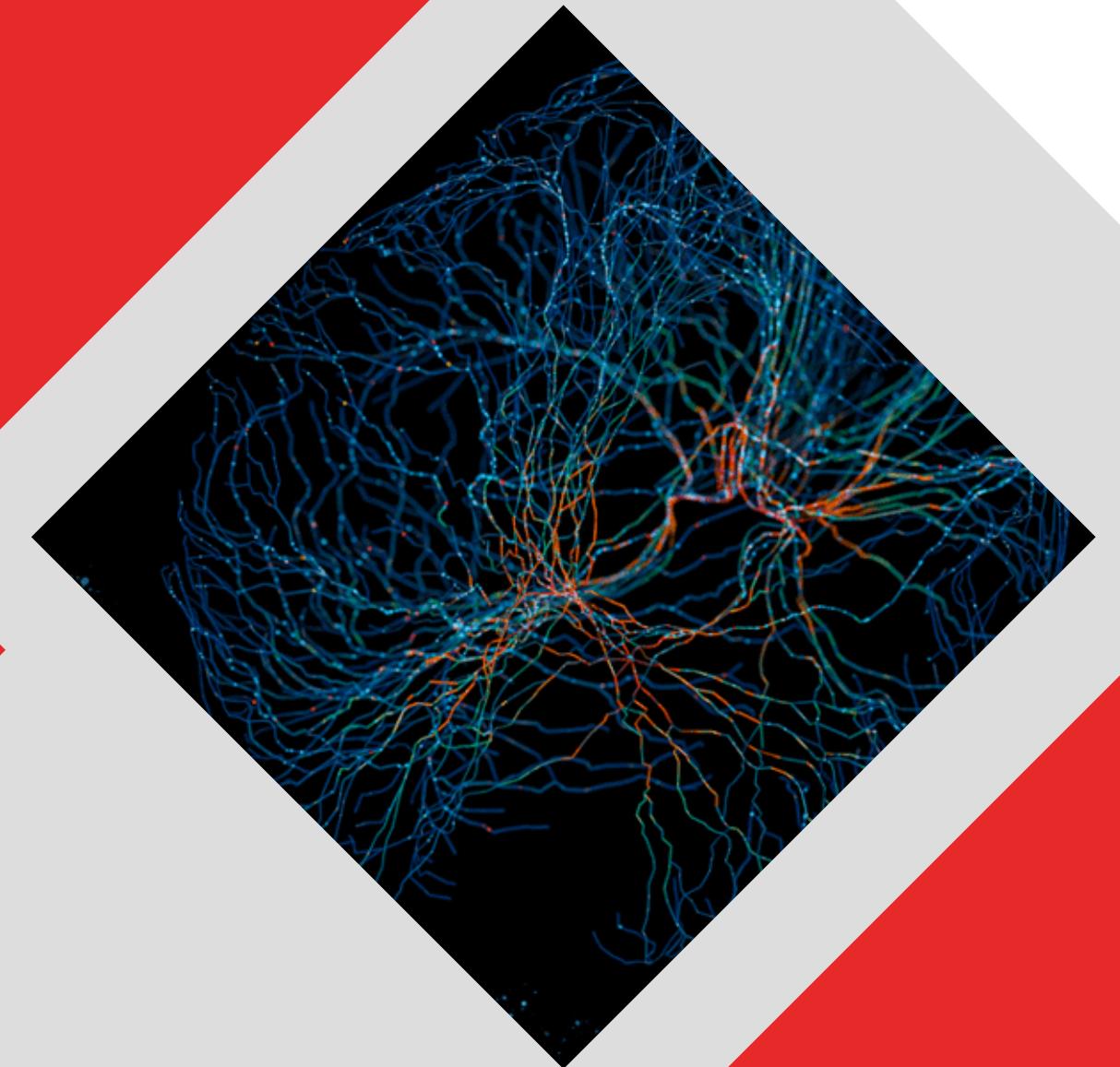


## Product Profiling

- id
- category
- price
- title
- vendor
- average\_rating
- review\_length
- review\_count



# 1. Clustering



# Customer Clustering

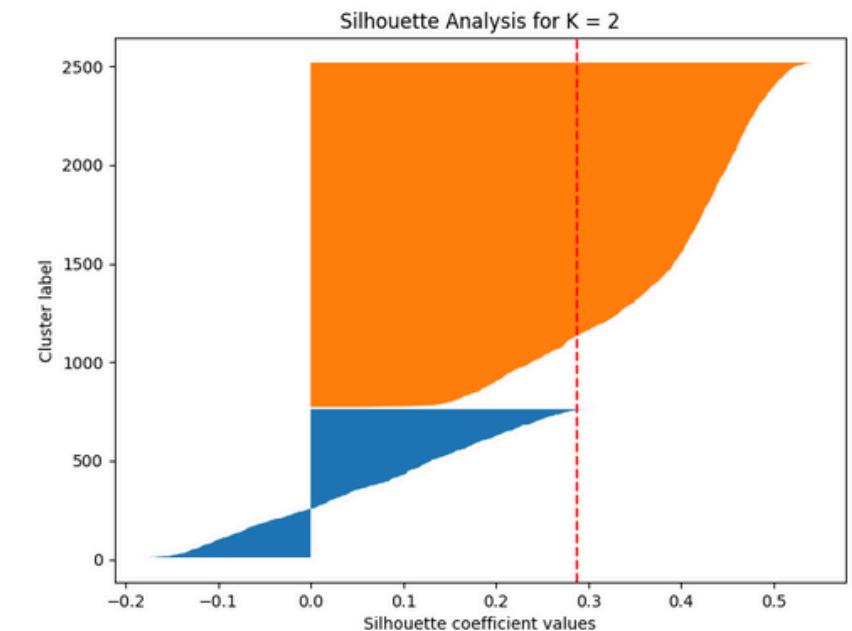
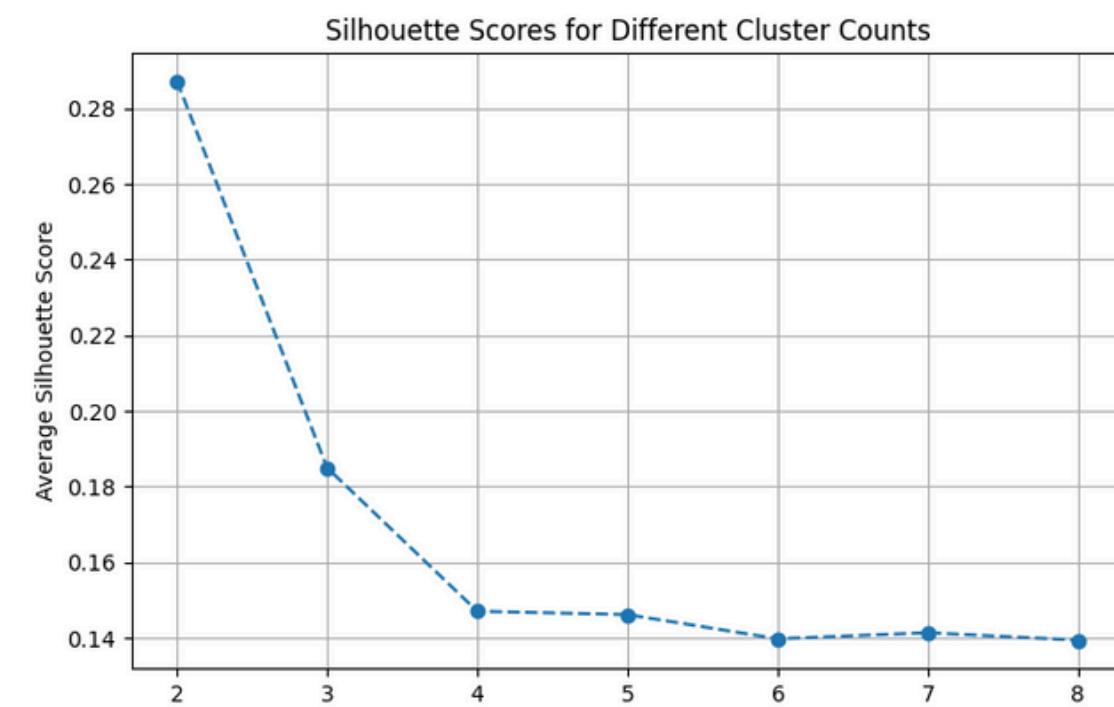
- **Finding insights from Data**
  - **Main Question:** Does discount usage, mail type, website source differ between low-spending and high-spending customers?
  - Drop discount usage, mail type, website source variables from User Profiles data before clustering
  - Standard Scale data

# Customer Clustering

## K-Means Clustering

Number of Clusters = 2

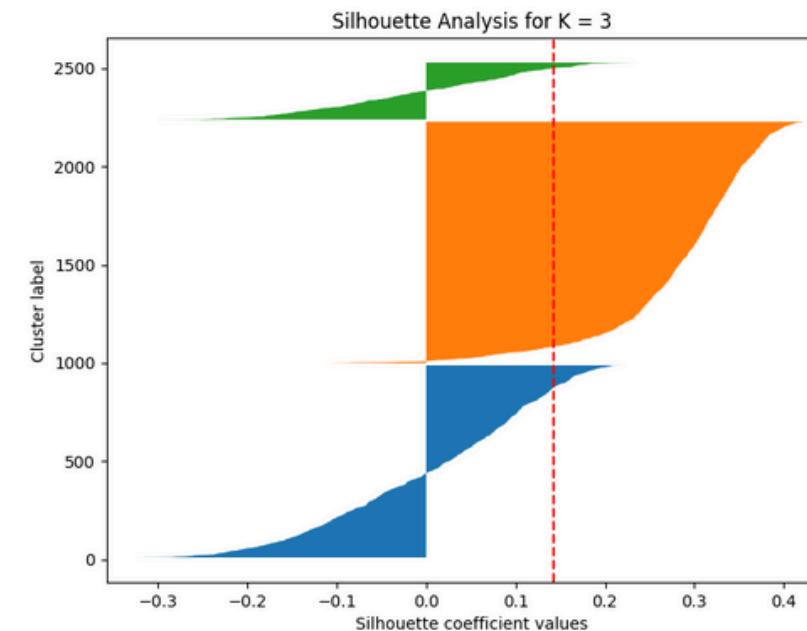
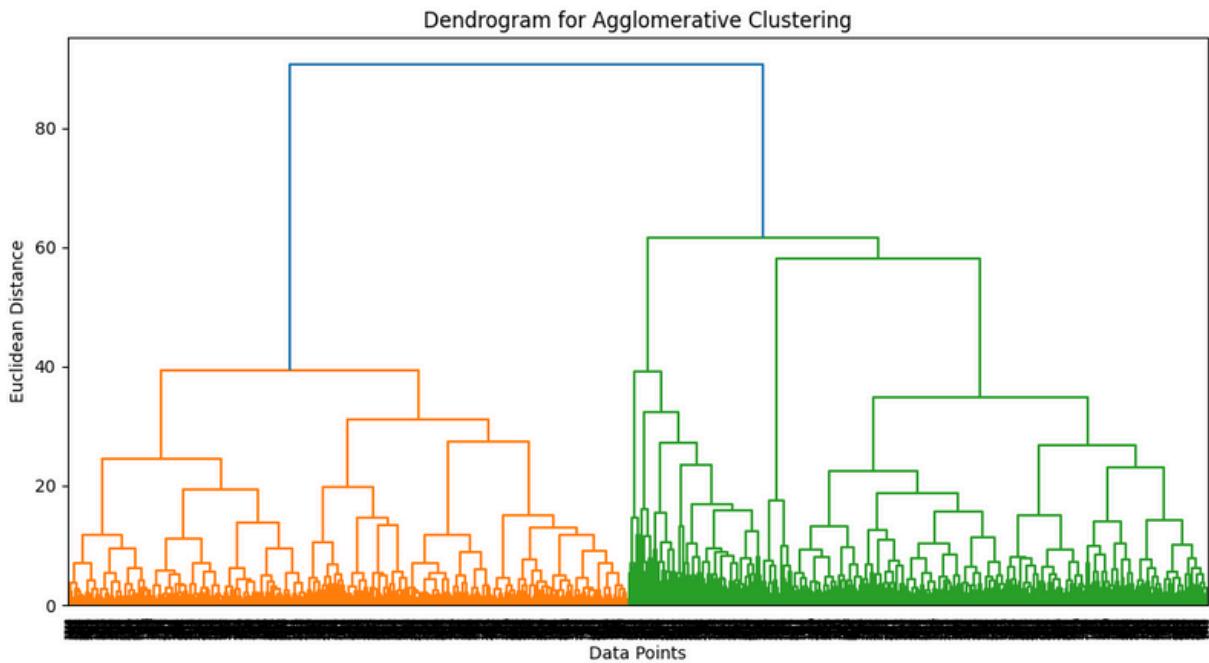
Average silhouette score = 0.287



## Agglomerative Clustering

Number of Clusters = 3

Average silhouette score = 0.143



# Customer Clustering

## K-Means Clustering

Low-spending  
High-spending

Latitude	Longitude	Day Duration	Age	Total Spent (Doohickey)	Total Spent (Gadget)	Total Spent (Gizmo)	Total Spent (Widget)	Total Orders	Discount Usage Proportion	Source Affiliate	Source Facebook	Source Google	Source Organic	Source Twitter	Email Provider (Gmail)	Email Provider (Hotmail)	Email Provider (Yahoo)	Size
39.649	-95.024	2677.68	44.99	110.894	162.817	146.2	170.388	3.347	0.062	0.209	0.202	0.199	0.19	0.2	0.333	0.338	0.329	1747
40.414	-95.567	2652.274	45.972	731.964	1066.69	940.184	1059.147	17.147	0.102	0.187	0.211	0.207	0.202	0.193	0.355	0.296	0.349	753

## Agglomerative Clustering

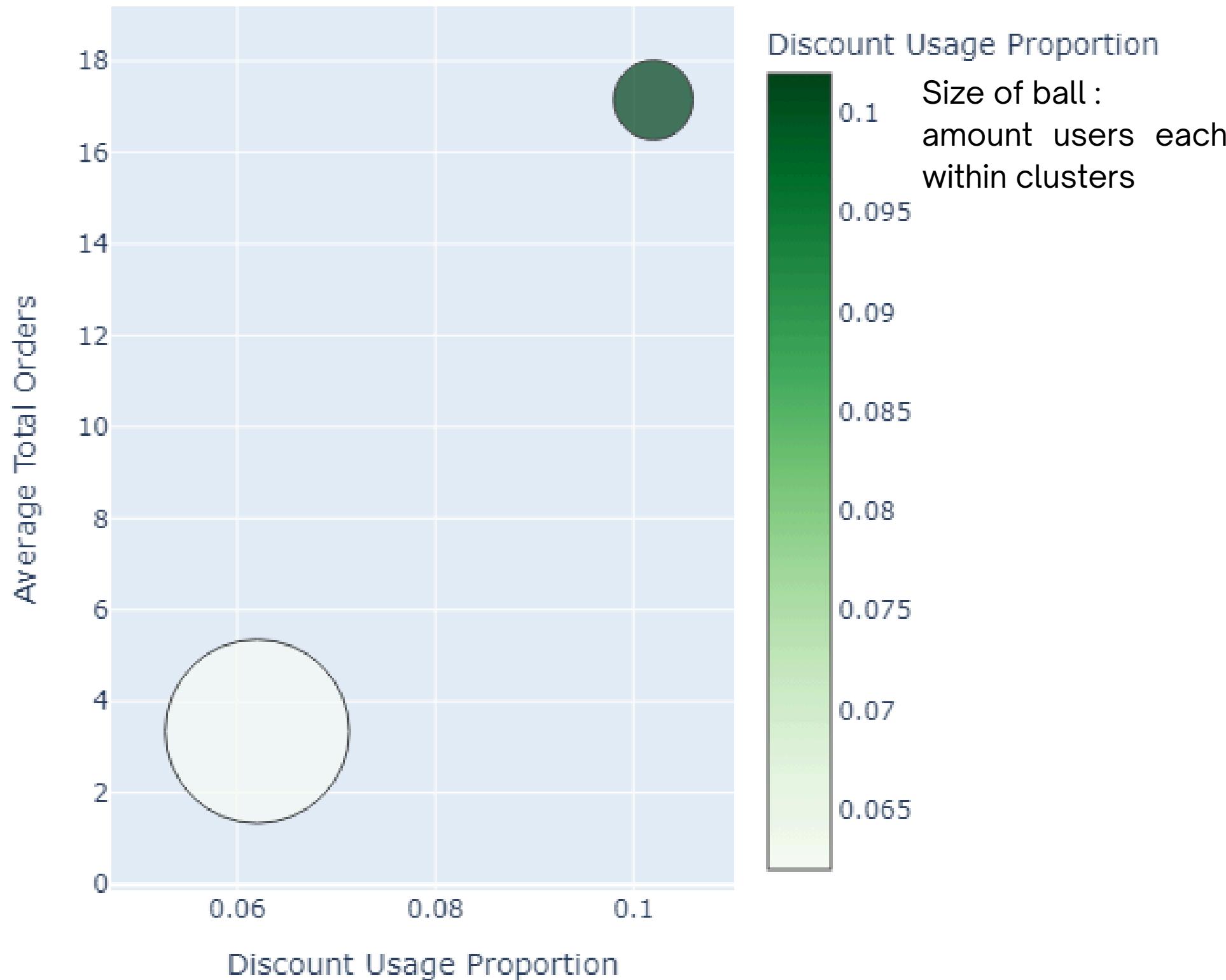
Low-spending  
Medium-spending  
High-spending

Latitude	Longitude	Day Duration	Age	Total Spent Doohickey	Total Spent Gadget	Total Spent Gizmo	Total Spent Widget	Total Orders	Discount Usage Proportion	Source Affiliate	Source Facebook	Source Google	Source Organic	Source Twitter	Email Provider Gmail	Email Provider Hotmail	Email Provider Yahoo	Size
39.2	-93.766	2687.744	44.572	54.295	77.415	68.692	73.316	1.64	0.045	0.198	0.211	0.207	0.182	0.201	0.352	0.334	0.314	1230
41.072	-97.458	2672.327	45.943	407.96	562.861	494.737	621.097	10.886	0.102	0.212	0.196	0.192	0.203	0.196	0.325	0.318	0.358	979
38.739	-93.556	2587.409	46.1	957.82	1516.832	1355.78	1364.168	20.914	0.101	0.186	0.206	0.206	0.21	0.192	0.337	0.313	0.351	291

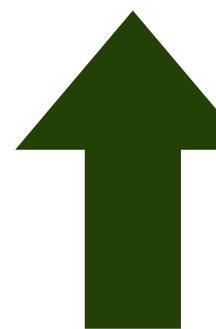
- Provide Discounts may turn low-spending customers to high-spending customers

# Customer Clustering

Spending Level relationship with Discount Usage Rate



## Insight



Discount Rate



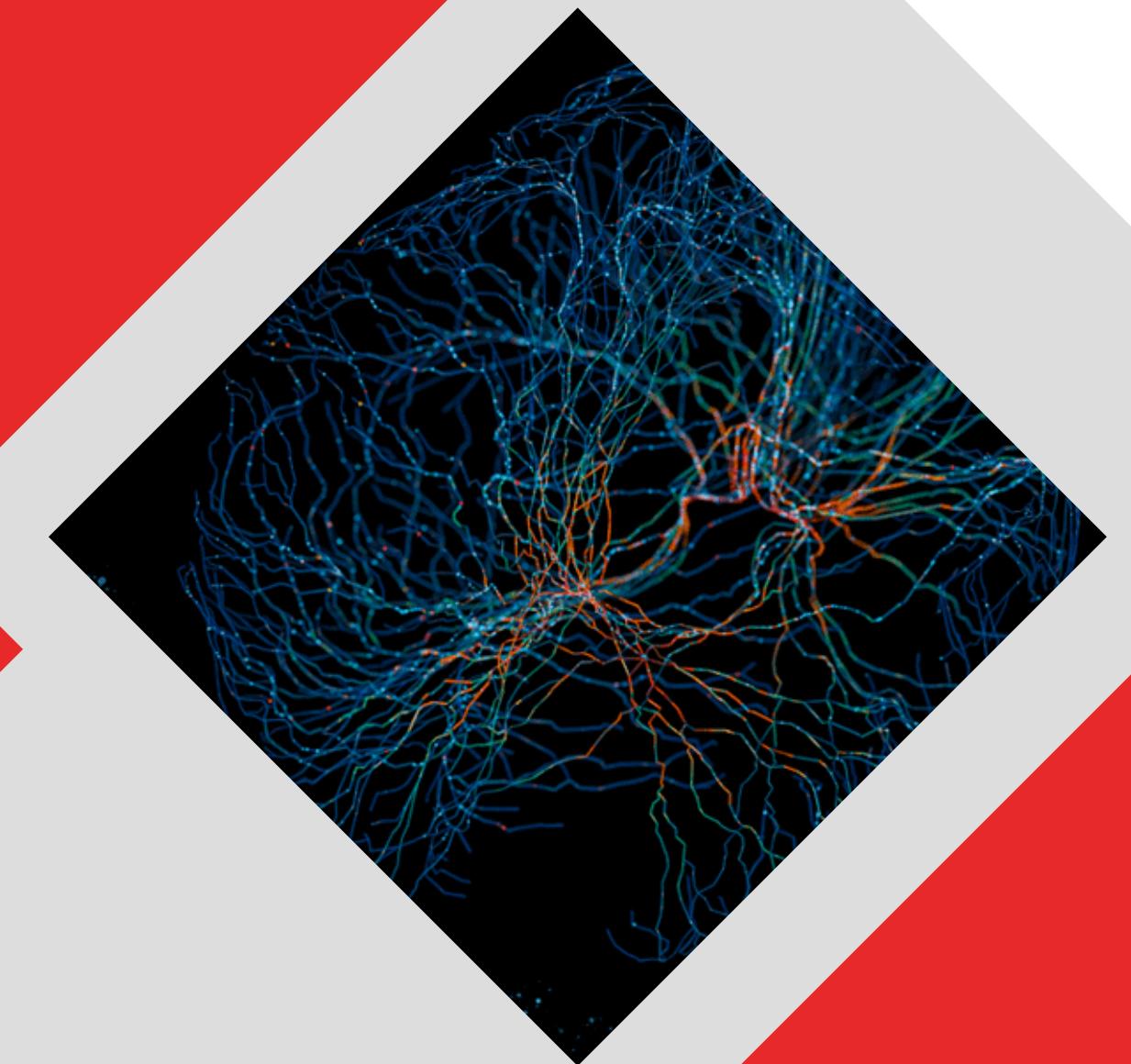
Customer Orders

## Strategy

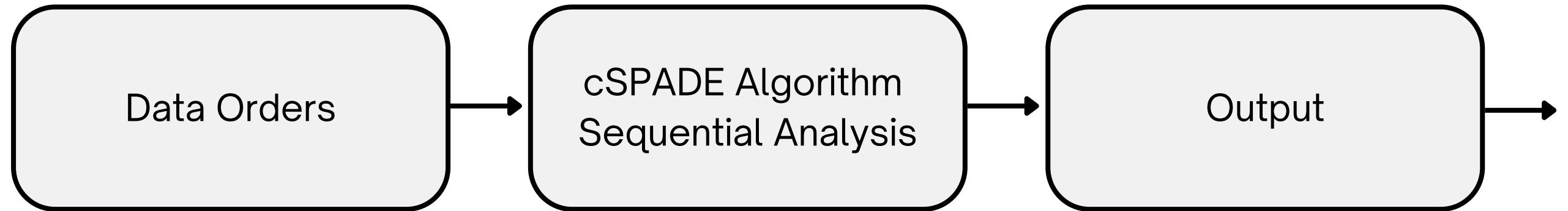
We can increase in the amount of orders significantly by using promotion



# 2. Sequential Analysis



# Sequential Analysis



Most common products that are bought in sequence

```
"sequence","support","pattern"
"<{77},{169}>",0.00859106529209622,2
"<{80},{115}>",0.00744558991981672,2
"<{49},{67}>",0.00744558991981672,2
"<{35},{97}>",0.00687285223367698,2
"<{53},{88}>",0.00687285223367698,2
"<{17},{86}>".0.00687285223367698,2
```

- Data: Orders
- Use Customer\_id, Product\_id ,Created at
- Checking and filtering for same product repeated purchase
- Re-label variable name so cSPADE can understand data

## Reward Rush

### POTS 'n' PANS

Unlock delicious rewards with Pots 'n' Pans Bonus Points! Earn points on every purchase and redeem them for discounts, special offers, and more.



### BONUS 100 Points

### HATS 'n' HANDBAG

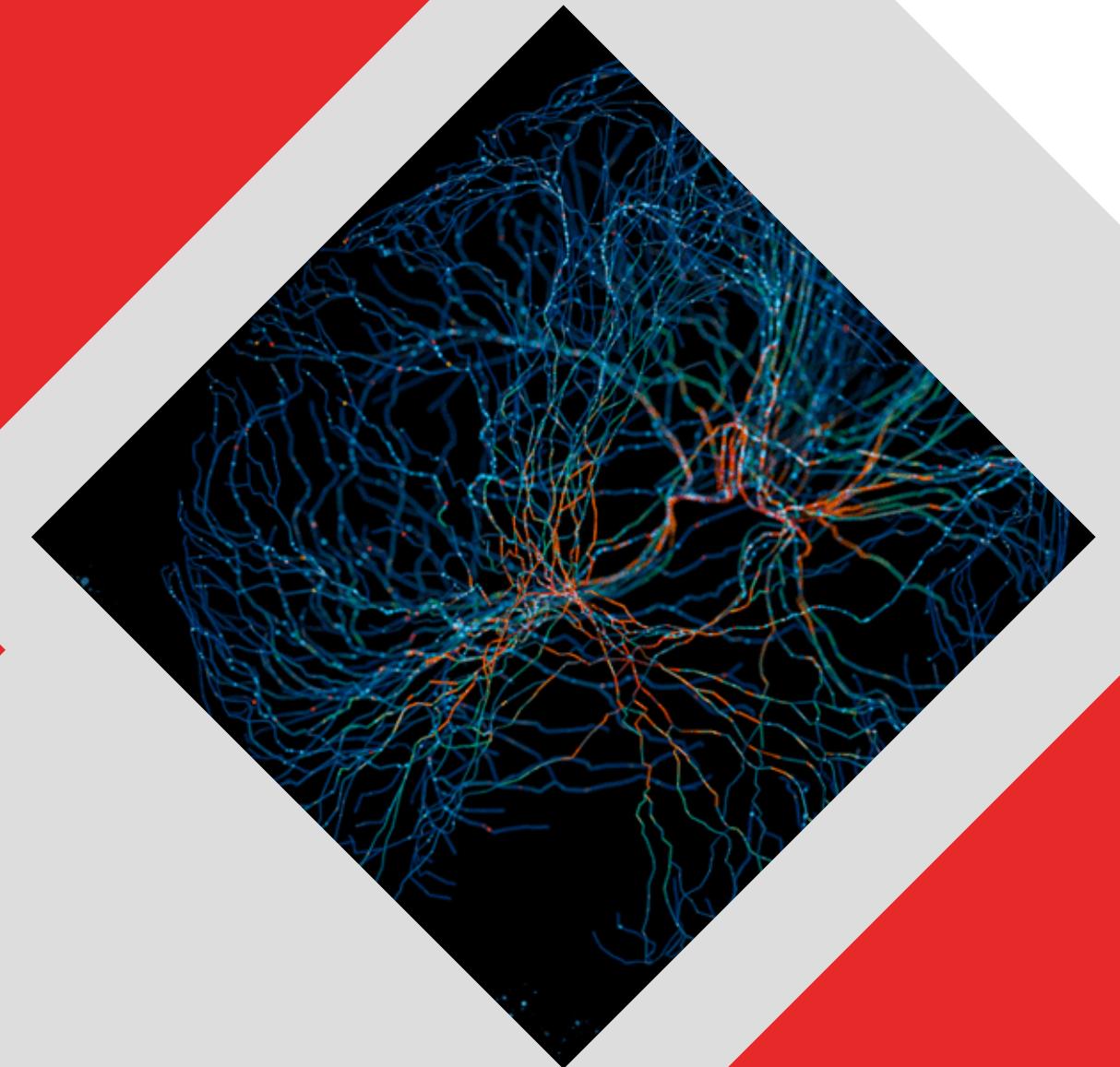
Hats 'n' Handbags Bonus Points! Exclusive stylish purchase earns you points towards fabulous rewards. Start collecting today!



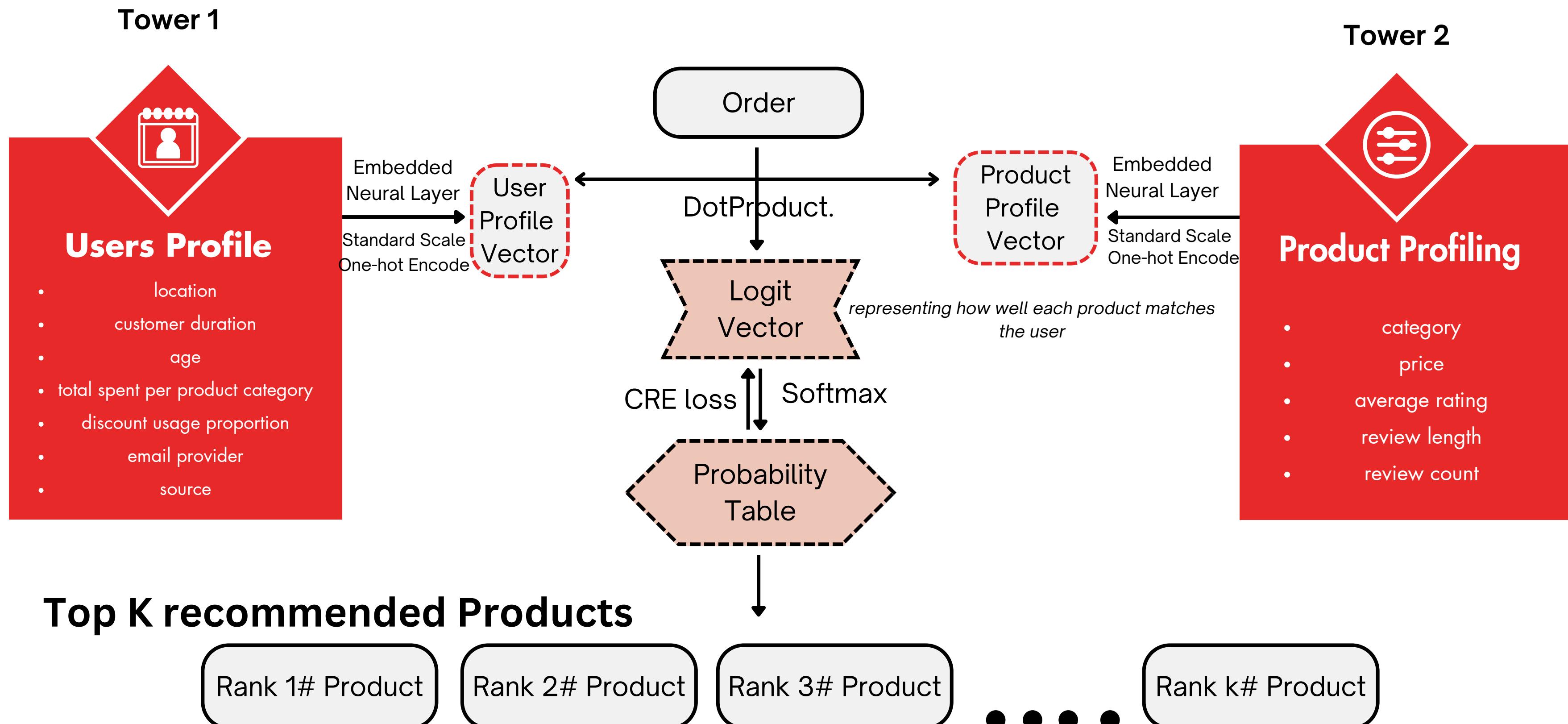
### BONUS 200 Points



# 3. Two Tower Recommendation System



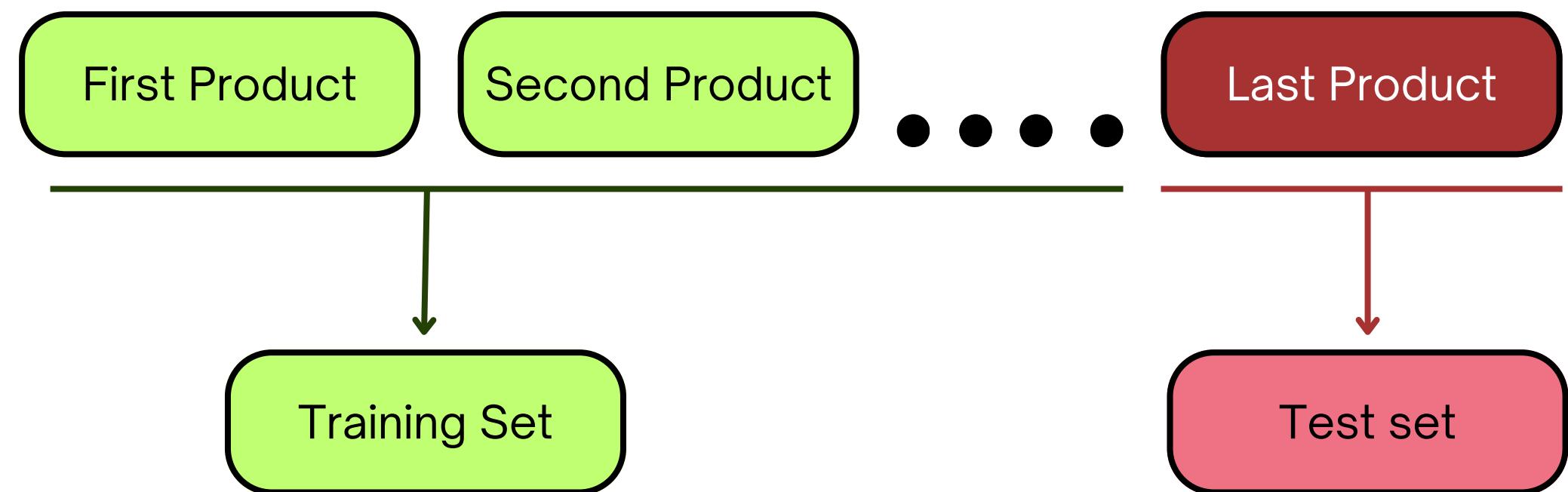
# Two Tower Neural Recommendation System



# Model Evaluation

## Train-Test Split

Example: User 0 interaction



- To mimics real world situation

## Result

Evaluate on Test Set

$k = 10$ , Hit Rate = 6%

$k = 20$ , Hit Rate = 11%

$k = 25$ , Hit Rate = 15%



Unsatisfactory Result

## Hit-Rate

*evaluates how well a recommendation model performs by checking if the correct (ground truth) product appears in the Top-K recommended items.*

$$\frac{\text{Number of users with the correct item in Top-K}}{\text{Total number of users}}$$

## Baseline

Random Guessing	$\sim 0.5\% (K/\text{Total Items})$
Popularity-Based	5% – 20%
Recent Interaction	10% – 30%
Collaborative Filtering	20% – 40%
Matrix Factorization (SVD, ALS)	30% – 50%

# Two Tower Neural Recommendation System

## Problem

- Low Hit rate Problem
  - Caused by using synthetic data

## Solution

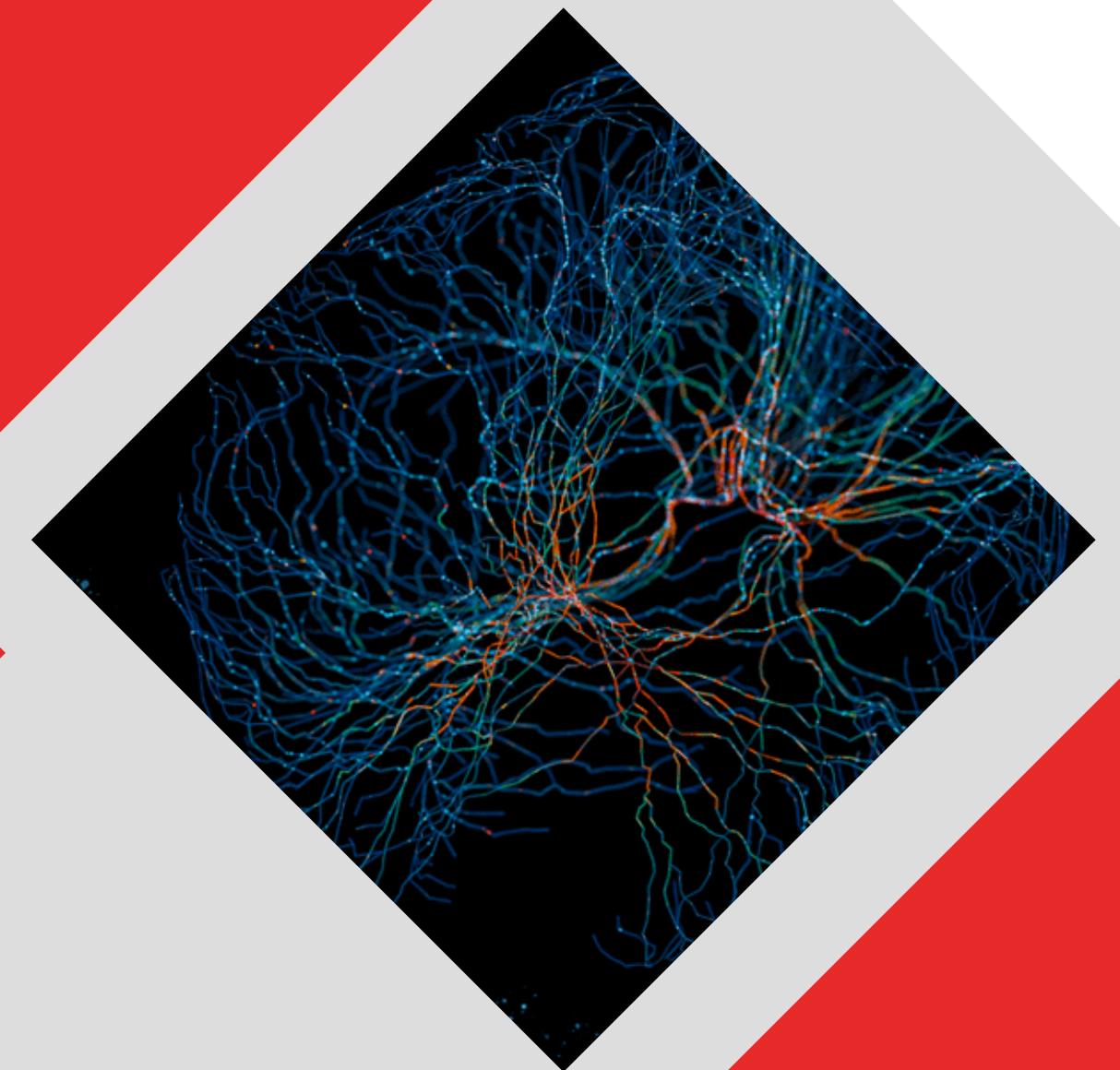
- Include generic items discount in Central Surprise rewards
  - Ex. All items discount,Gift Card
- Change item pool in Central Surprise rewards daily
  - Hit rate per week increase to 55%
- Add Selectable Items to Choose

## Alternative Solution

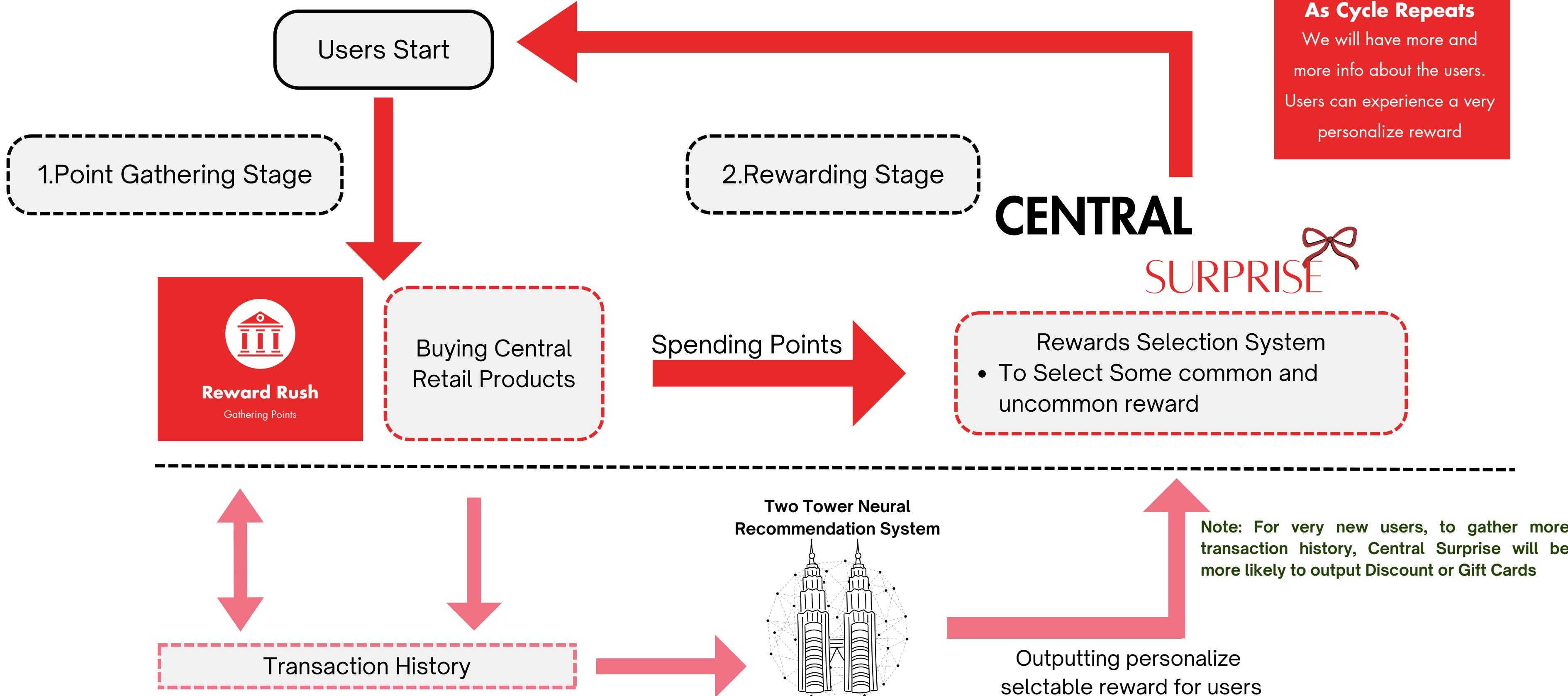
- By using Central Retail Database with millions of record, The model is likely to improve significantly



# Conclusion



# Users Experience Cycle





# Thank You

