

# Improving Sales for Fast Fashion

## **ISSS621 – Data Science for Business**



# Presentation Agenda



Business  
Context



Raw Data &  
Insights



Design Solution



Interpret Results  
into Actionable  
Items



Closed Loop  
Recommendation



Potential  
Governance

# Business Context



# Fast Fashion

- The value of the fast fashion market worldwide was estimated to be worth over 106 billion U.S. dollars in 2022.
- Fast fashion refers to a business model and a trend in the fashion industry characterized by the rapid production and consumption of inexpensive clothing.



# Business Problem



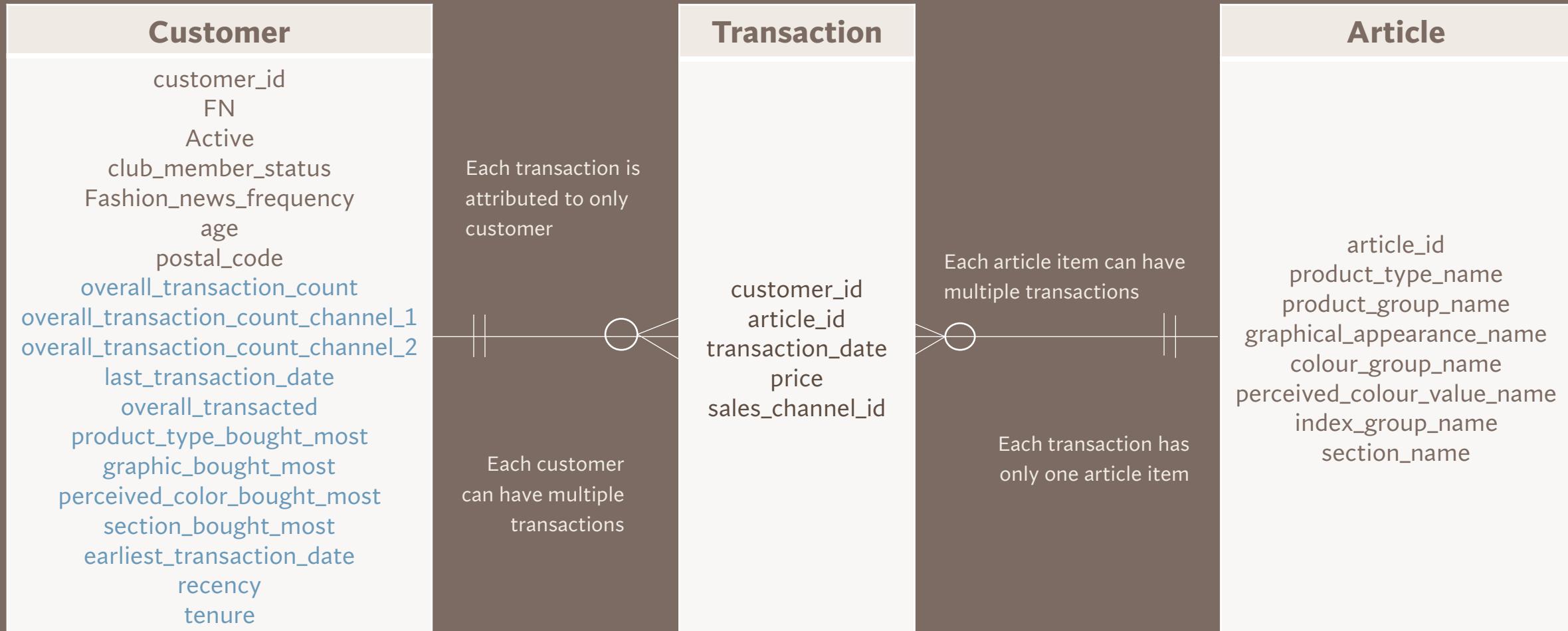
We are tasked by H&M to provide data science solutions to the below business problems:

1. As sales performance for the 35-44 age group has been poor, H&M wants to focus on improving sales performance for this age group
  - Resources are limited and hence H&M wants to perform segmentation to better understand this group to perform targeted marketing.
2. With supply chain disruptions leading to higher costs for the business, H&M is looking at streamlining their inventory management.
  - H&M is looking at minimizing excess inventory and associated cost such as storage cost and markdowns.
3. With large volume of transaction data generated, H&M wanted to explore opportunities for up-selling and cross selling using the past 3 months.
  - H&M is looking out for opportunities to bundle product together for promotional campaigns.



# Raw Data & Insights

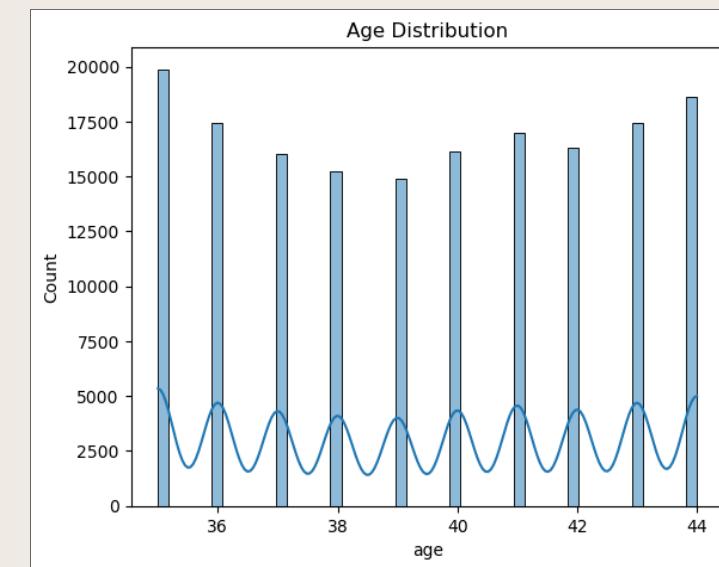
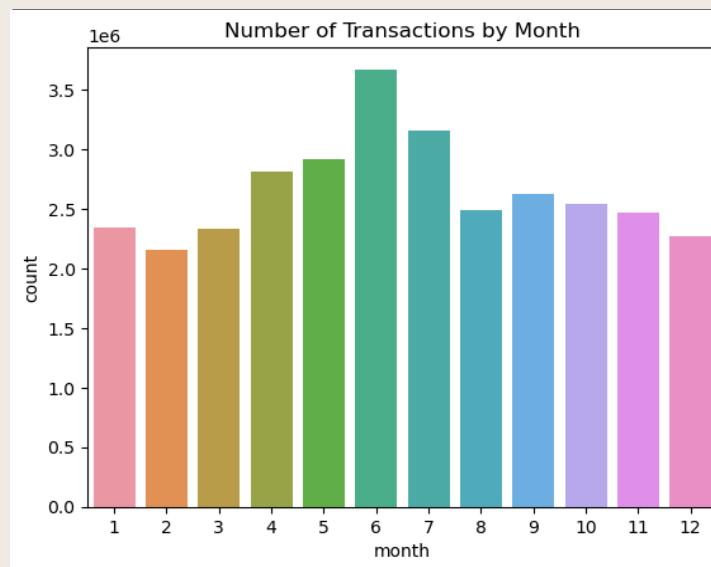
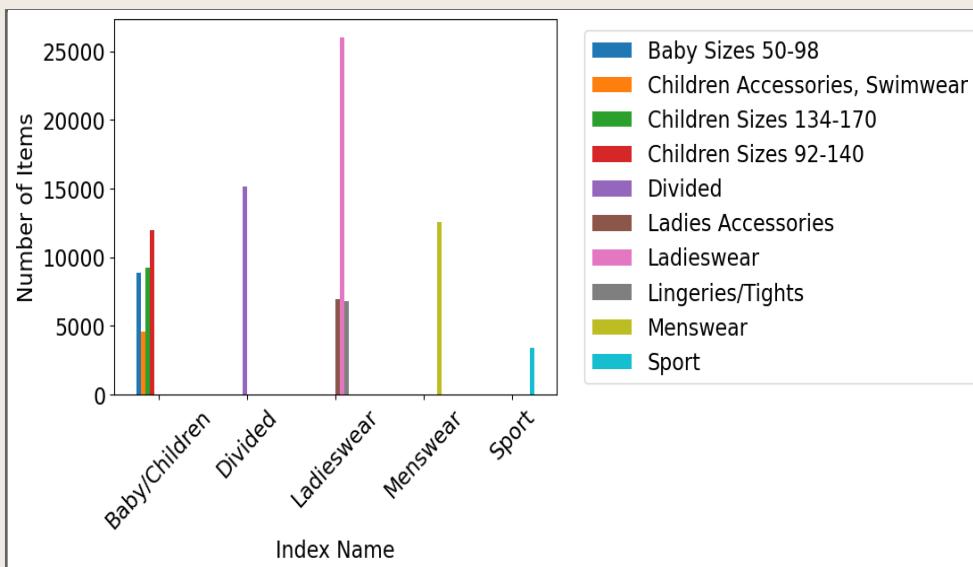
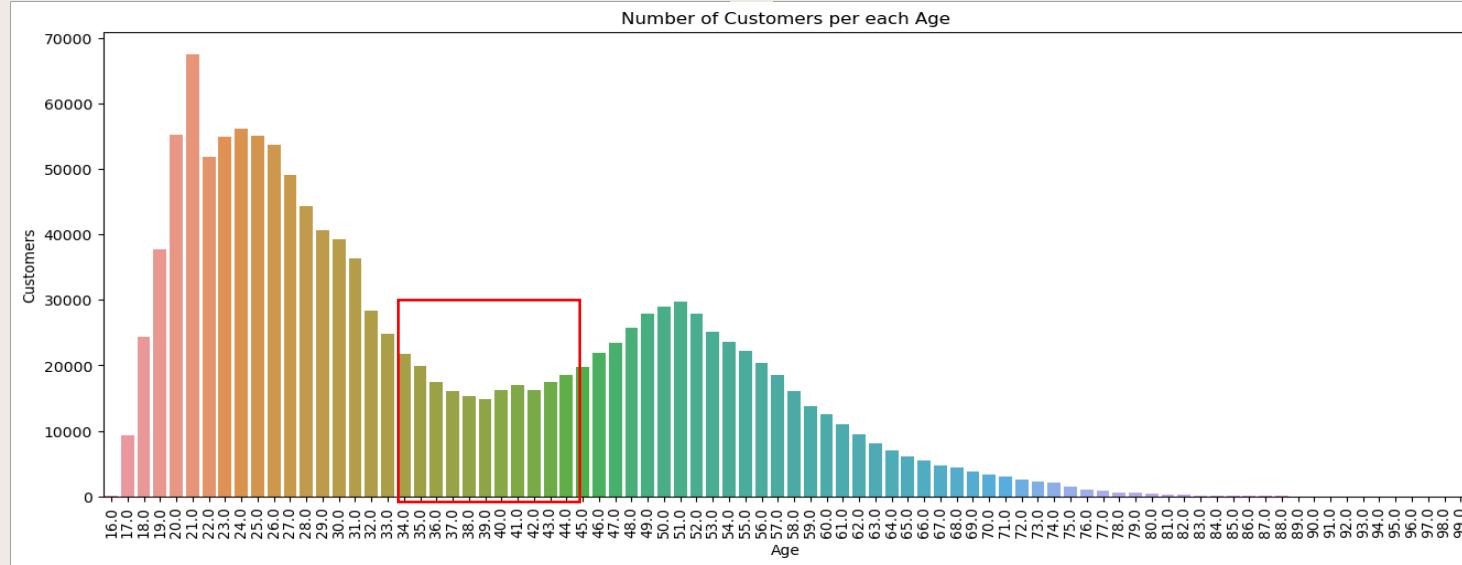
# Overview Of The Three Main Datasets



Created features are denoted in Blue

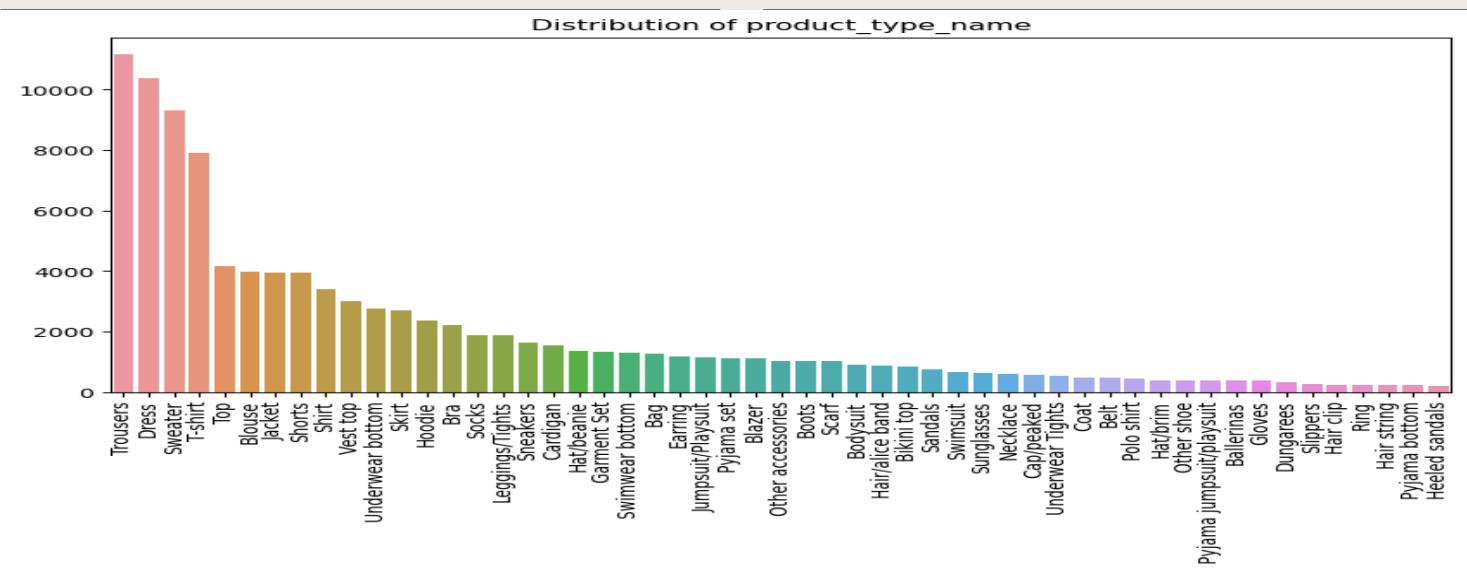
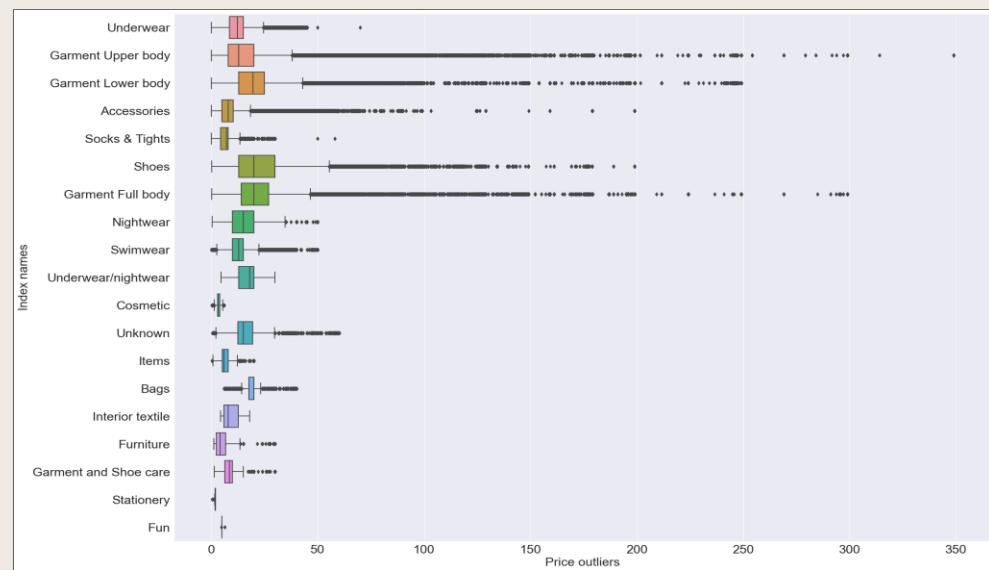
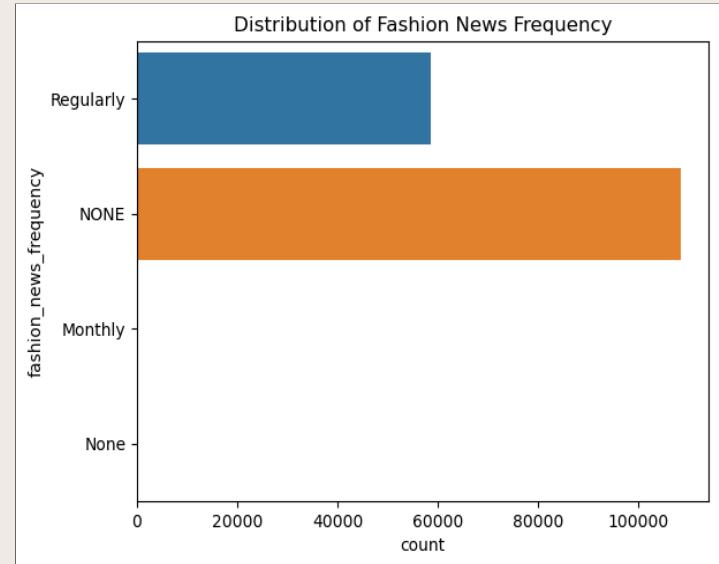
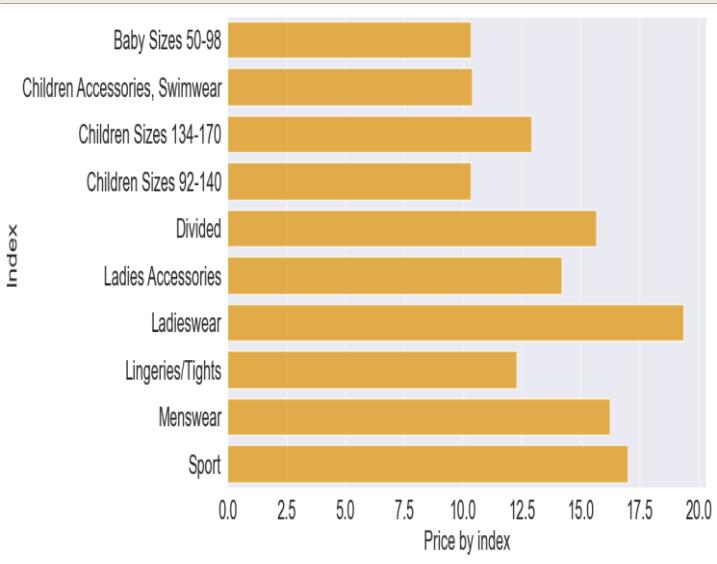
# Insights - P1

- Ladieswear accounts for significant sales in H&M, and Sports accounts for the least
- Ladieswear also account for the highest price by index
- Transactions per month are highest in June, which seems seasonal
- Customers in the age range of 35-44 seems low, which seems to increase after 44. We will be targeting this specific age group for further analysis and strategies



# Insights – P2

- Trousers and dress accounts for most of the transactions, which should be part of ladieswear
- Large number of customers are not receiving fashion news





# Design Solutions

# Solutions



## Customer segmentation

Segmentation will allow H&M to identify high-value or at-risk customers who can then be targeted with specific campaigns



## Market basket analysis

Frequent pattern mining can help to identify products purchased together; insights can be used to provide purchase recommendations and formulating marketing strategies



## Personalized pricing

Understanding customers preferences and willing to pay can help H&M to develop personalized or dynamic pricing



## Cross-selling and upselling

Design cross and upsell strategies; selling related products and recommending higher value products to customers



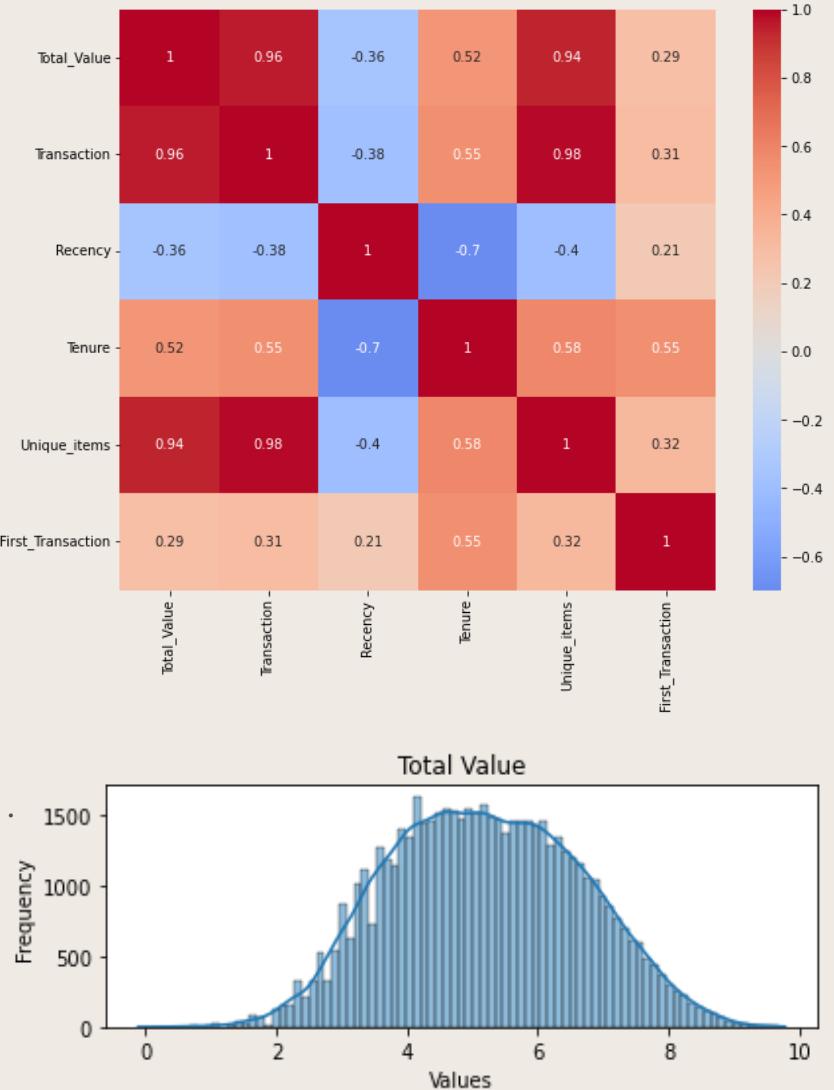
## Inventory management

Utilize predictive models to forecast demand for key products to help H&M optimize inventory management strategy

# Customer Segmentation



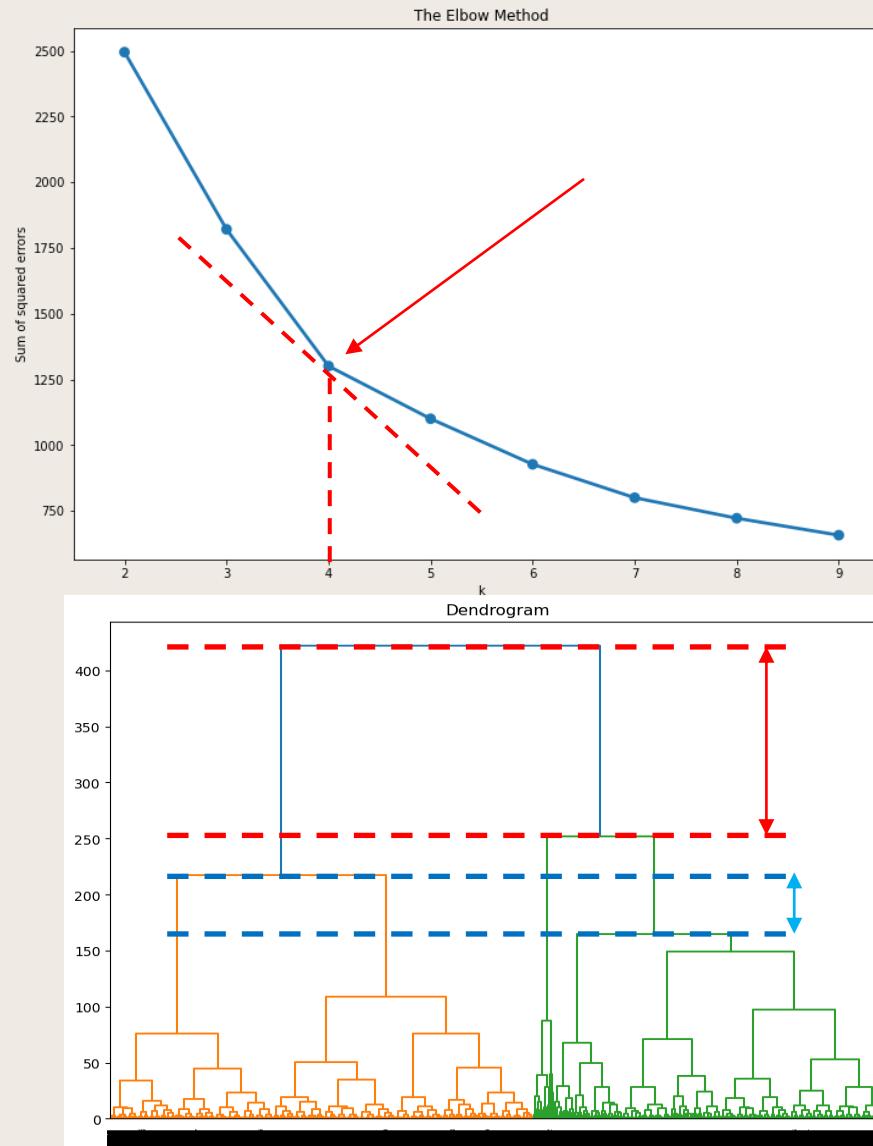
# Customer Segmentation



## Feature Engineering

- To better understand the characteristics of customer segments (360 view)
- New Features included:
  - Total Value
  - Total Transactions
  - Number of Unique Items
  - Days Since First Transaction
  - Days Since Last Transaction
- Transformation of Variables
  - Log transformation of Total Value to reduce the effect of outliers for k-means clustering

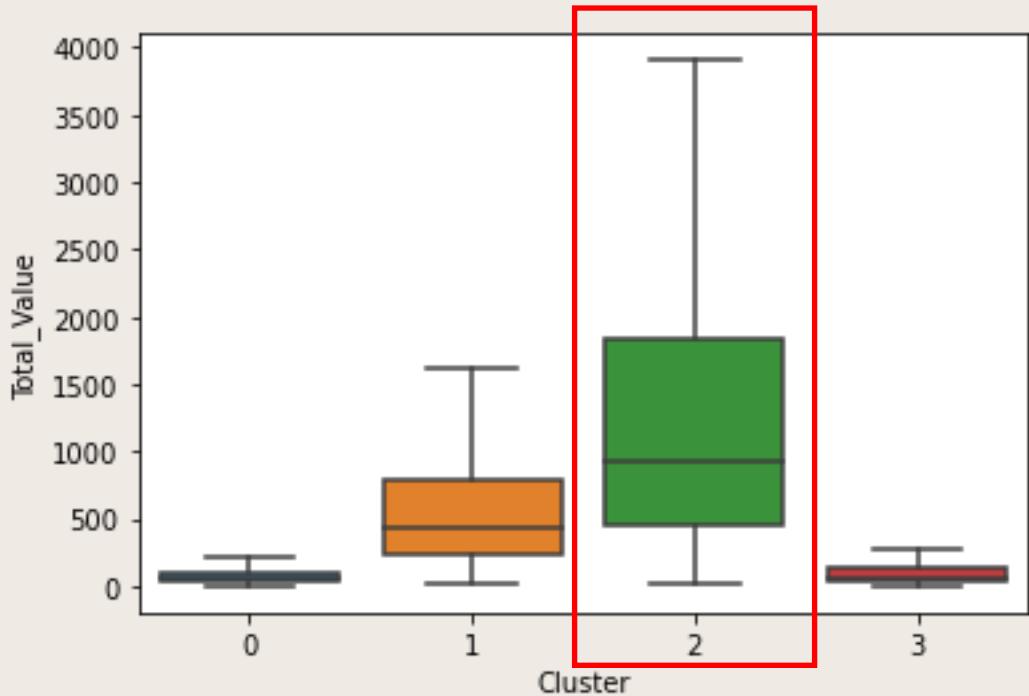
# Customer Segmentation



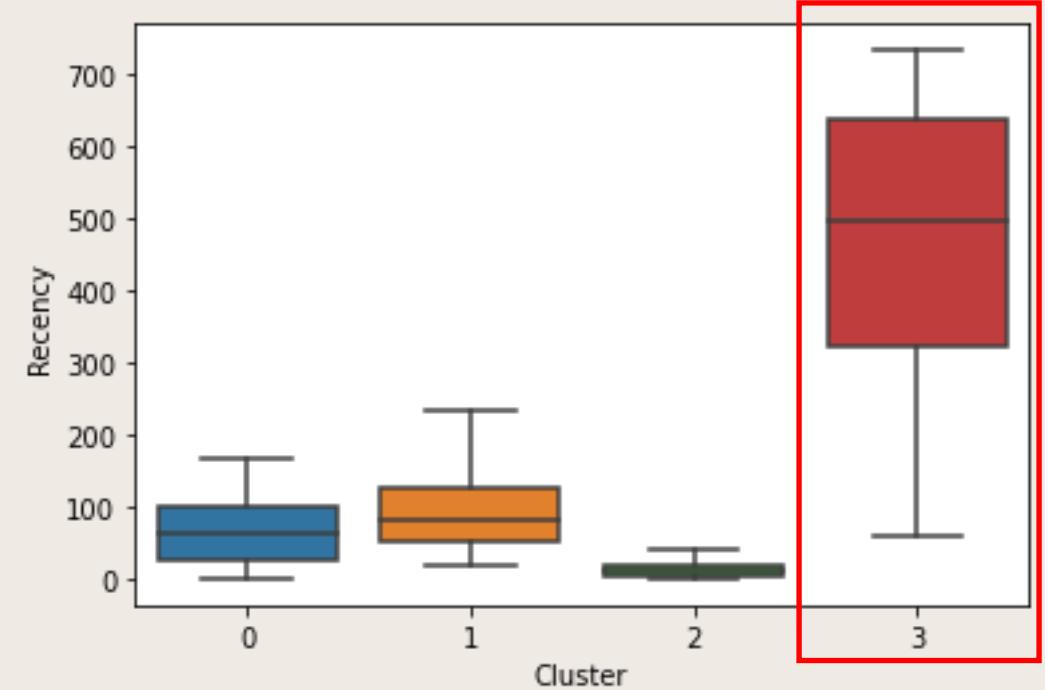
## Finding the Suitable k- value

- **From the Plot of SSE v.s. Number of Clusters:**
  - Using Elbow Method, it shows the 4 Clusters would be suitable as we can see that this is the point where the rate of decrease in SSE slows down (marginal improvement for higher k-values)
- **Dendrogram**
  - From the longest vertical dist, 2 Clusters would be suitable. But this would limit the differentiation of the customer segments.
  - Hence the second longest vertical distance is looked at: 4 Clusters
- k-value of **4** is selected

# Customer Segmentation

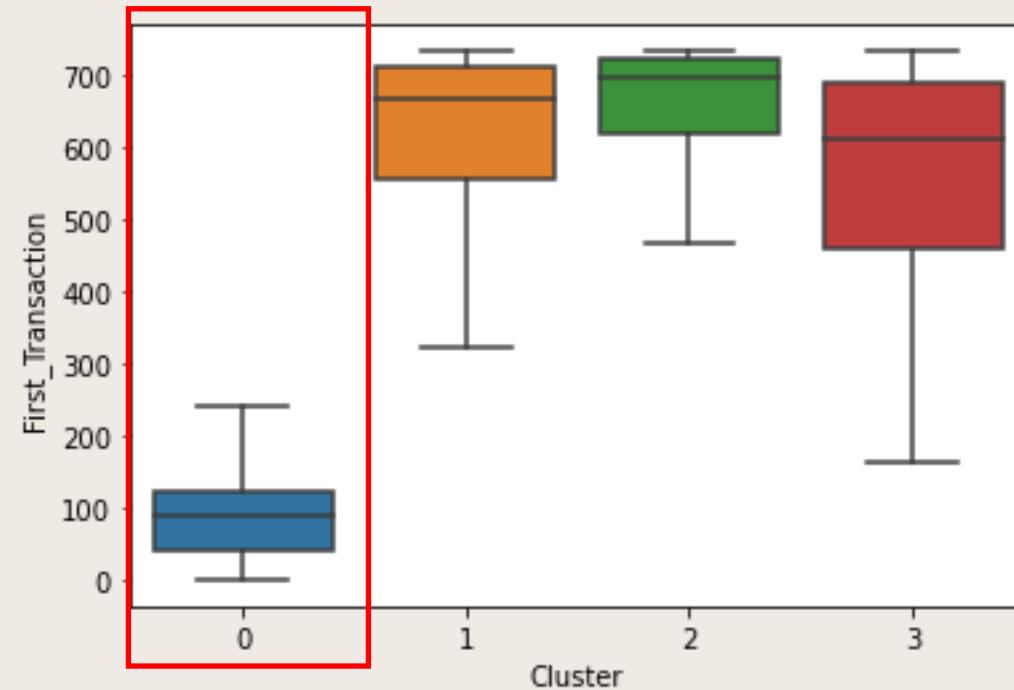
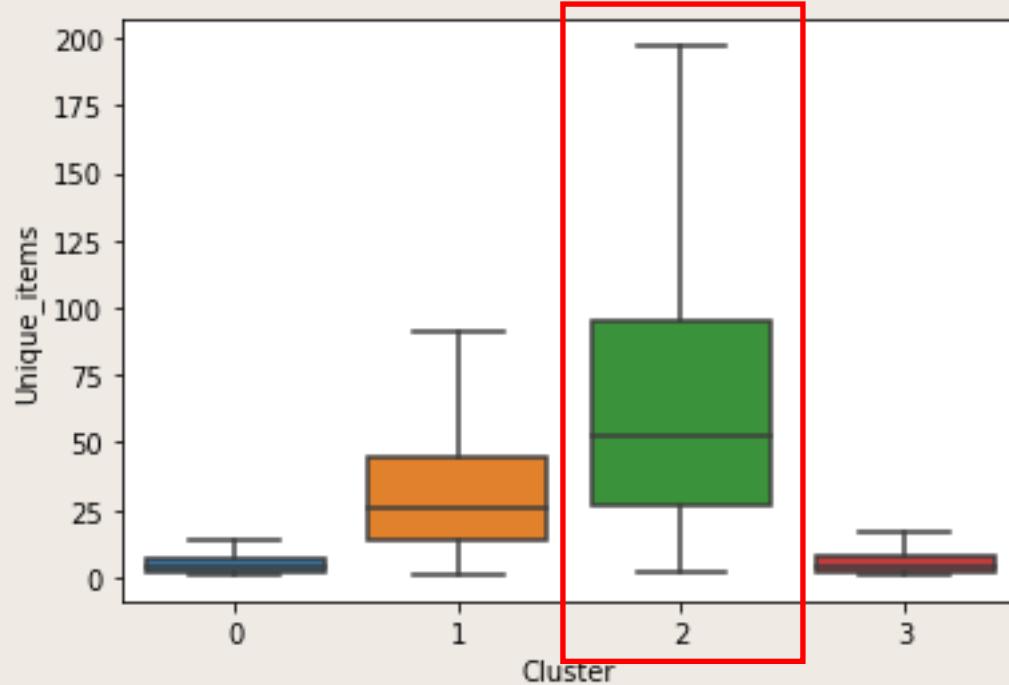


- **Cluster 2** is the high-value customer group with much higher revenue contribution to H&M
- Cluster 2 should be prioritized for marketing campaigns to grow this group



- In terms of Recency, **Cluster 3** has the least recent purchases
- Median customer's last purchase is 495 days ago
- Cluster 3 represents churned customers and should be of lowest priority for promotion campaigns

# Customer Segmentation



- Similar to total value, **Cluster 2** purchases the greatest number of unique products for all clusters.
- From the days since first transaction, we can tell that **Cluster 0** is the newly acquired customers for H&M
- This represents opportunity for H&M to grow this segment by promotion campaigns and recommending a wider product range to Cluster 0

# Market Basket Analysis

**Business question: What products are frequently purchased together by our target customers (age 35-44 years)**

## Frequent pattern mining pre-processing

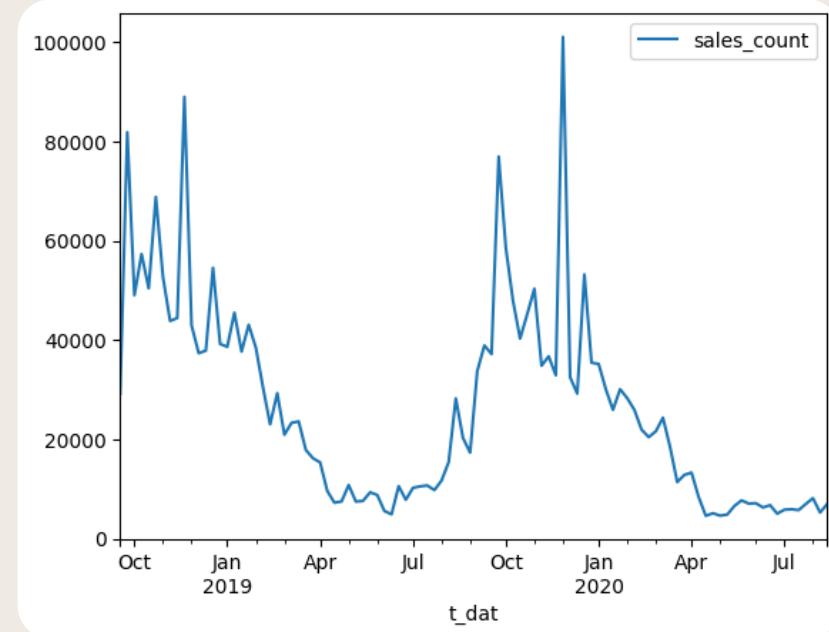
- Filter to extract transactions made by our target age group
- Zoomed in on transactions in the last 3 months of the data set (due to computation resource limitation)
- Conduct frequent pattern mining using the Apriori Algorithm on the product\_type\_name (e.g. dress, trousers, T-shirts etc.)
- We used a minimum support of 0.05 (table on left showing only top 20 item-sets for illustration)

Item sets	Support	Confidence
Dress, Trousers	0.146	0.429 / 0.385
T-shirt, Trousers	0.134	0.353 / 0.490
Trousers, Top	0.117	0.501 / 0.309
Sweater, Trousers	0.113	0.553 / 0.312
Dress, T-shirts	0.106	0.312 / 0.389
Dress, Top	0.106	0.311 / 0.452
T-shirt, Top	0.096	0.351 / 0.410
Dress, Blouse	0.092	0.269 / 0.545
Vest top, Trousers	0.090	0.453
Blouse, Trousers	0.089	0.531
Bikini top, Swimwear bottom	0.084	0.789
Vest top, Dress	0.084	0.421
Sweater, Dress	0.083	0.406
T-shirt, Sweater	0.083	0.303 / 0.406
T-shirt, Vest top	0.083	0.303 / 0.414
Sweater, Top	0.078	0.385 / 0.335
Vest top, Top	0.077	0.384 / 0.327
Shirt, Dress	0.074	0.575
Shorts, Trousers	0.071	0.470
Blouse, Top	0.070	0.416 / 0.299

# Inventory Management

**Business question: How much inventory to keep stock for each product?**

Sales trend of Sweaters from Sep 2018 to Aug 2020



There are **seasonal trends** in sales volume!

How should we measure forecasting performance?

Note: There were insufficient weekly data points to utilize neural networks or a 52-week lag SARIMA.

**ARIMA Model**

$$\left(1 - \sum_{i=1}^p \varphi_i L^i\right) (1 - L)^d X_t = \delta + \left(1 + \sum_{i=1}^q \theta_i L^i\right) \varepsilon_t$$

Where:

p is the order (# of time lags of the autoregressive model)

d is the degree of differencing (# times the data have had past values subtracted)

q is the order of the moving-average model

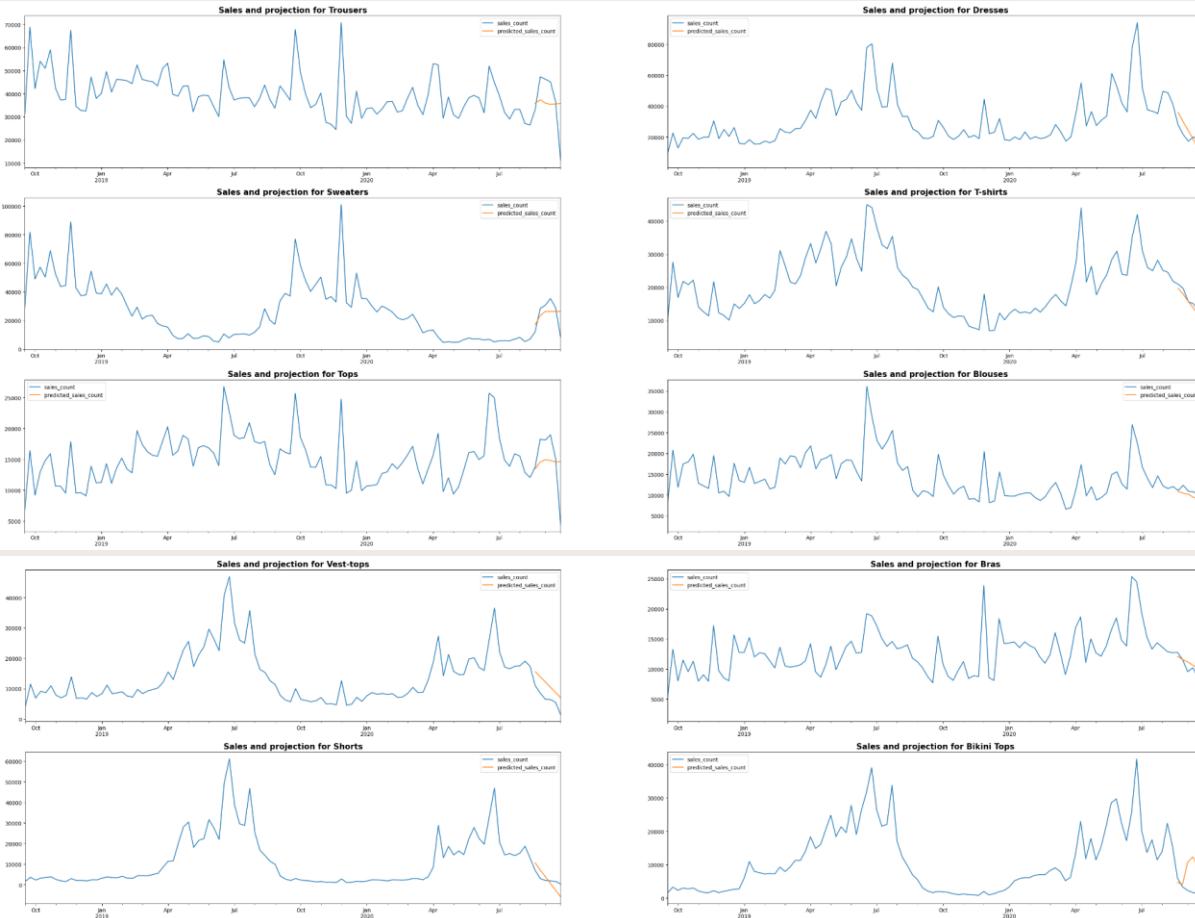
Utilize the Grid Search method to find optimal values for hyper parameters p, d and q, given that this is a regression problem, we optimize by minimizing Mean Square Error.

$$\text{Minimize } MSE(p, q, r) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Where:

n is the total number of weekly sales\_count observations

# Inventory Management



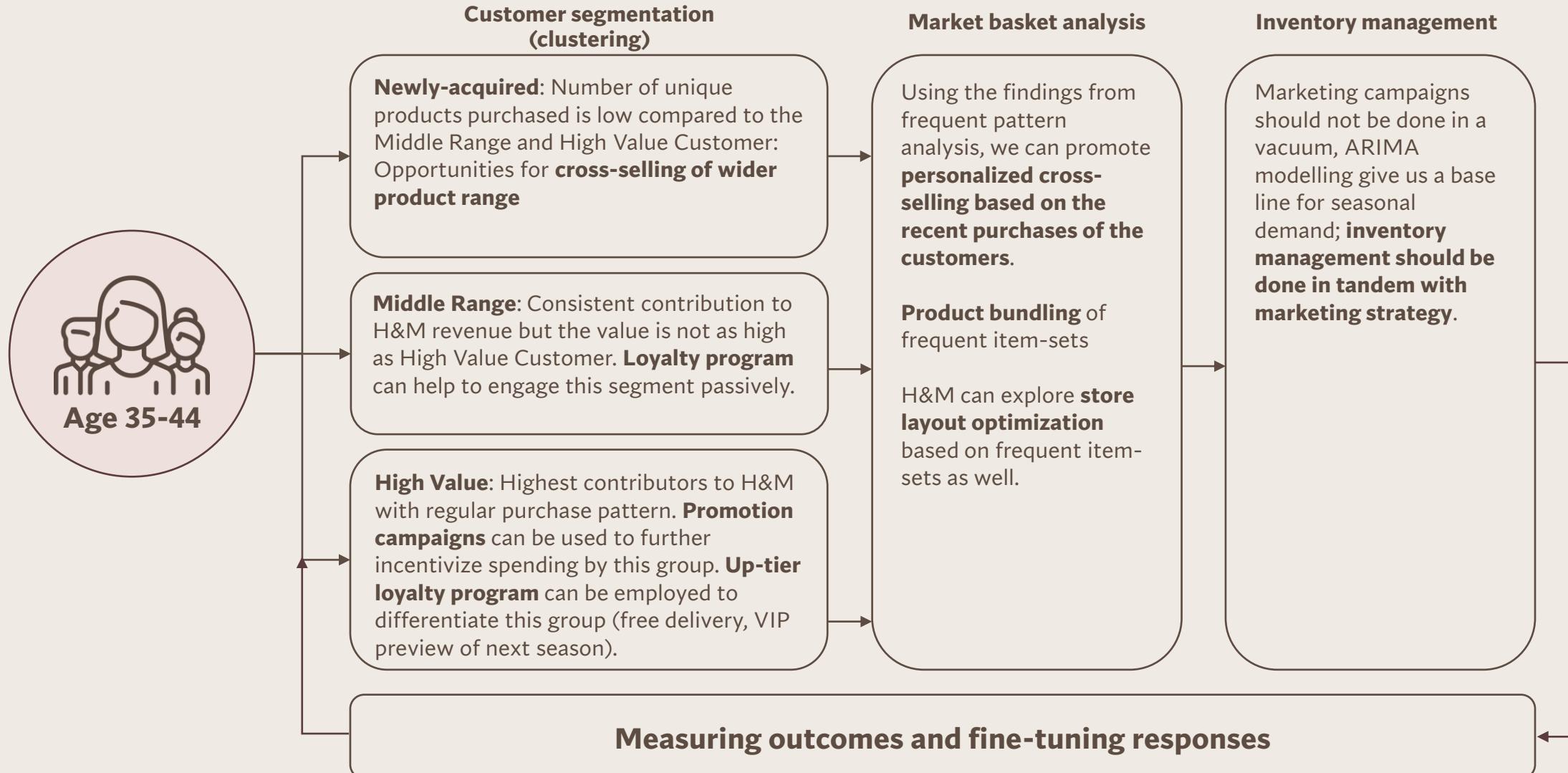
	Items	RMSE	Avg. sales count	RSME/Sales count
0	Trousers	12363.244	36604.000	0.337757
1	Dresses	6052.645	18336.833	0.330081
2	Sweaters	8974.770	23984.000	0.374198
3	T-shirts	2789.380	14646.833	0.190443
4	Tops	4912.223	14786.333	0.332214
5	Blouses	2559.023	9733.000	0.262922
6	Vest Tops	4843.598	6534.333	0.741254
7	Bras	3113.557	9202.000	0.338357
8	Shorts	3933.692	2646.833	1.486188
9	Bikini Tops	6759.750	2491.333	2.713306

- While the ARIMA model did a decent job forecasting demand for most of the products, it fared less ideally for certain items such as Vest Tops, Shorts and Bikini Tops.
- One main reason is because of the **lack of weekly datapoints** to capture the seasonality of item sales. This underscore the importance of good quality data for data scientists to base their models on.
- Nonetheless, this exercise gave us a baseline expected sales trend on which we can adjust according to our recommended actions to target customers between age 35-44.



# Interpret Results into Actionable Items

# Translating Data Science into Business Actions



# Cross-selling and Product Bundles



# Cross-selling and Product Bundles

- For Product Bundles, H&M will be offering a 5% off the total price for up to 10,000 transactions for each bundle
- Campaign Budget is \$X

## Objective 1: Maximise Revenue

- Compute the product (Max\_Revenue) of **Confidence** (Item A, B) and **Combined Price** of Item A and B

$$\text{Max_Revenue} = \text{Confidence}(A, B) * 0.95 \text{ Price}(A + B)$$

$$\text{Cost} = 0.05 (\text{Price}(A+B)) * \text{Max Volume}$$

- Sort the item sets in descending order based on their Max\_Revenue value
- To determine the k-number of itemsets for bundling, we will check if the cost exceed the budget X by checking in order of the Max\_Revenue value

## Objective 2: Maximise Sales Volume

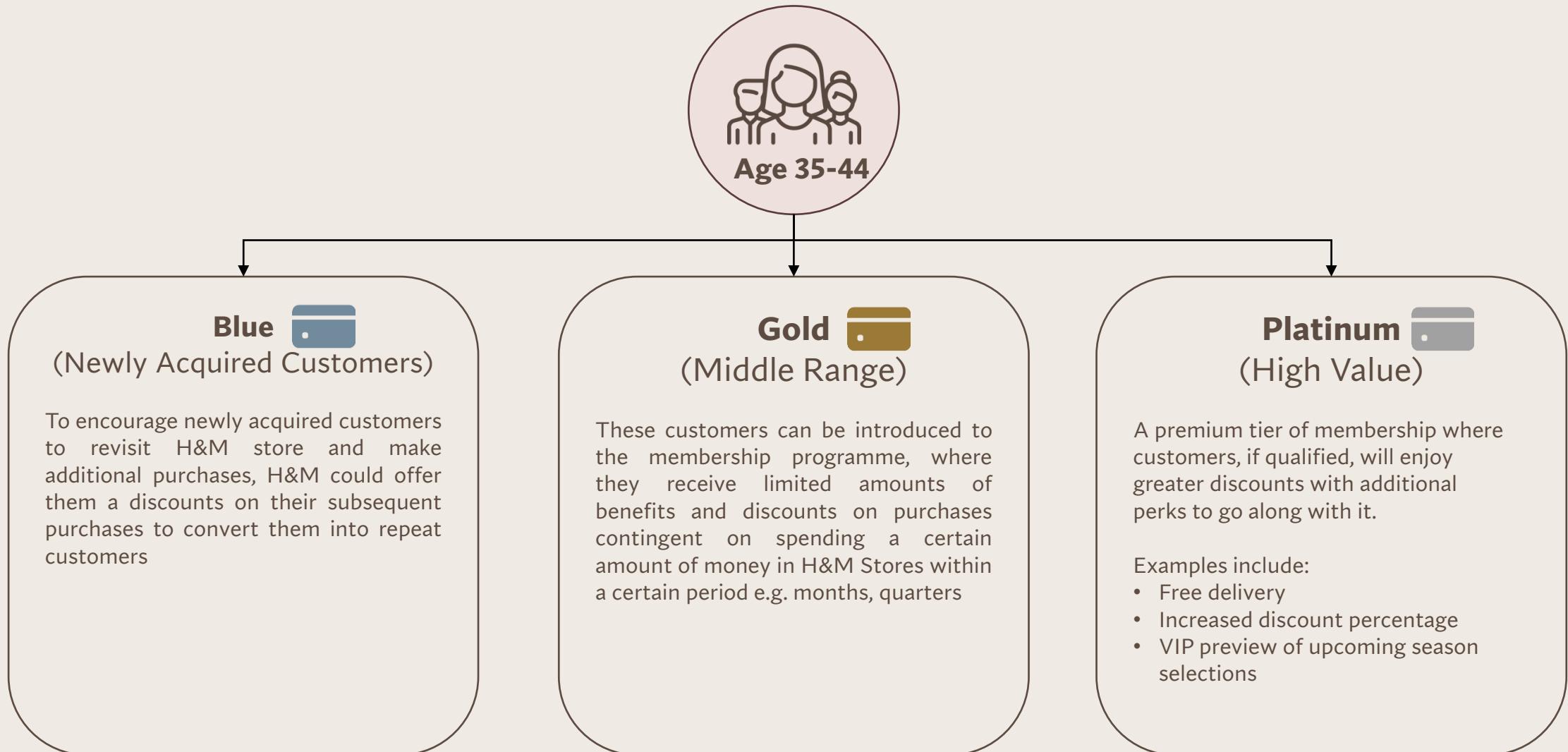
- Compute the product (Max\_Volume) of **Confidence** (Item A, B) and **Support** (Item A, B)

$$\text{Max_Volume} = \text{Confidence}(A, B) * \text{Support}(A, B)$$

$$\text{Cost} = 0.05 (\text{Price}(A+B)) * \text{Max Volume}$$

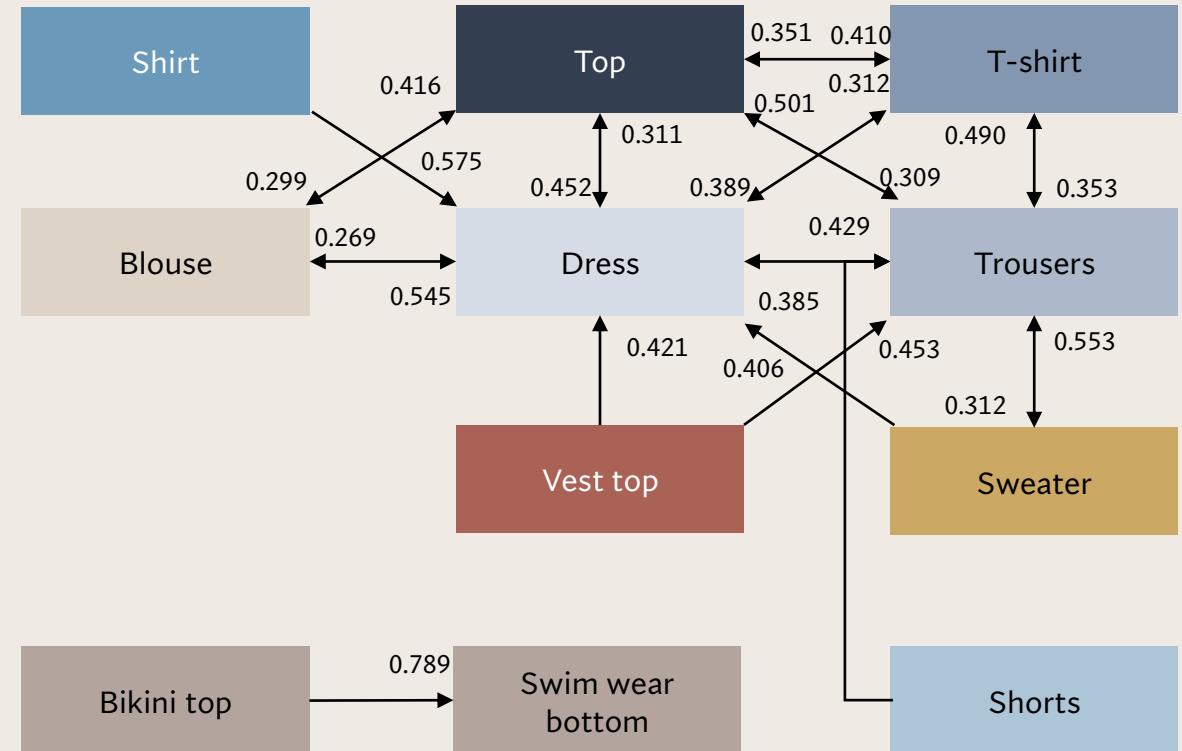
- Sort the item sets in descending order based on their Max\_Volume value
- To determine the k-number of itemsets for bundling, we will check if the cost exceed the budget X by checking in order of the Max\_Volume value

# Personalized Pricing – Tier membership



# Store Layout Optimization

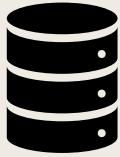
Item sets	Support	Confidence
Dress, Trousers	0.146	0.429 / 0.385
T-shirt, Trousers	0.134	0.353 / 0.490
Trousers, Top	0.117	0.501 / 0.309
Sweater, Trousers	0.113	0.553 / 0.312
Dress, T-shirts	0.106	0.312 / 0.389
Dress, Top	0.106	0.311 / 0.452
T-shirt, Top	0.096	0.351 / 0.410
Dress, Blouse	0.092	0.269 / 0.545
Vest top, Trousers	0.090	0.453
Blouse, Trousers	0.089	0.531
Bikini top, Swimwear bottom	0.084	0.789
Vest top, Dress	0.084	0.421
Sweater, Dress	0.083	0.406
T-shirt, Sweater	0.083	0.303 / 0.406
T-shirt, Vest top	0.083	0.303 / 0.414
Sweater, Top	0.078	0.385 / 0.335
Vest top, Top	0.077	0.384 / 0.327
Shirt, Dress	0.074	0.575
Shorts, Trousers	0.071	0.470
Blouse, Top	0.070	0.416 / 0.299



A photograph of a person sitting cross-legged on a bed, surrounded by folded laundry. The person is wearing a striped shirt, blue jeans, and a black watch. A white callout bubble is overlaid on the image, containing the text.

# Closed Loop Recommendation

# Closed-Loop Data Ecosystem



**Data Collection** – Online and In-Store Transactions, we can track customers purchases through our loyalty programs and click and interaction of the customer in online purchases.

We can also get customer satisfaction data through surveys and reviews to get a 360 view. Data sources can also include social media platforms; Instagram, TikTok

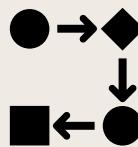


**Feedback** – Based on the data collection done on the solutions we provided, we can track the impact and whether it worked as predicted.

This can include sales volume, revenue and customer satisfaction



**Actionable Insights** – Leveraging on our analysis, tailor strategies to maximise profitability and customer satisfaction



**Iteration** – We use the data collected and feedback to adjust our further analysis and actions

# Closed-Loop Data Ecosystem

After segmenting our customers, conducting MBA based on our analysis and insights, it could reveal a segment of customers who make small but frequent purchases and often shop for certain categories of items together.

Based on the analysis of the performance of the proposed strategy, we will identify if any modifications or adjustments is necessary, in the event where we conclude the strategies are unviable, we could at this stage come up with new strategies for consideration

By analyzing transaction data, we identify opportunities to cross-sell/bundle products to our customers. A tiered membership programme could also be implemented on top of that to further increase customer engagement and loyalty

We would then proceed to track the outcomes of this specific segment, which includes sales revenue, sales volume, customer feedback and other measures

# Governance

SALE

# Need For Data Governance at H&M



The need for good data governance isn't new – any retailer knows how mismanaged data can impact a business. Establishing a data governance framework and roadmap – Future steps on how the data organization functions



As of today, H&M Group has nearly 5,000 stores in 71 markets worldwide, welcoming nearly 5 billion visitors each year. With such impressive statistics, H&M Group is a major player in this industry



Unlike the past where H&M existing models only considers selling garments, new features and products has been introduced since then such as pastries, manicures and garment repairs. Therefore, H&M requires an overhaul of its models to provide flexible and dynamic data solutions



Despite implementation of disruptive technologies in key development areas being burdened with significant time lag due to the bureaucracies of a large corporation, developing a data strategy with innovative initiatives to aid H&M achieve its business vision brings us much excitement

# Data Strategy

## Data Principles & Execution



**“Grow”** refers to the need to grow their assets, better understand their customer to be able to provide better offerings, expand their product descriptions and make automated decisions based on them.



Gobal operations poses challenges for H&M in terms of handling data solutions and providing data services mainly due to constant changes in the organization especially the incorporation of different regulations of different markets



**“Care”** refers to having proper data governance in place, understanding their assets and their quality..



As H&M Group is not a closed ecosystem, they have different customers and import product data from them, so they need to **“Share”** data internally and externally while not compromising on the privacy .



On strategic level, working with awareness. Rising awareness for the importance of data management is also of crucial importance. C-level leaders and people working with the systems understand why data is important



H&M Group must make sure they have scalable and established data quality platform. They must work on establishing a future-proof stable MDM solution, including tools, methodology and processes needed to support the business.

# Recommendations for Privacy laws



## Situation

- Global operations poses challenges for H&M in terms of handling data solutions and providing data services mainly due to constant changes in the organization especially the incorporation of different regulations of different markets when transitioning from a Nordic brand to a Global brand.
- There's been increasing regulatory guidance, forcing retailers to reassess and reconsider their current data practices With significant amounts of data coming in from multiple systems and channels, how do you ensure you protect consumer privacy, abide by industry and government regulations, and effectively manage and utilize your data?



## Solution

- A comprehensive data inventory allows retail businesses to understand where is their entire data throughout the enterprise.
- The data inventory assesses the sources of the information gathered, how the data is used across the enterprise, and the types of data assets shared with 3rd parties.
- By achieving compliance across regulations, retailers obtain complete visibility and categorize, correlate, and map all types of business-critical data. This permits the retail companies to sort their data through policies and also be able to manage risks associated.

# Thank you

