Logo, company name

Description automatically generated

**ISSS621 Data Science For Business**

Prepared by: ISSS621 (Group 2)

|  |  |
| --- | --- |
| Adrian Tan Tze En  Jeremy Don Any Yi Liang  Lim Weisheng Vincent  Sim Chin Tiong Billy  Sruthi Basani | [adrian.tan.2022@mitb.smu.edu.sg](mailto:adrian.tan.2022@mitb.smu.edu.sg)  [jeremyang.2022@mitb.smu.edu.sg](mailto:jeremyang.2022@mitb.smu.edu.sg)  [vincent.lim.2022@mitb.smu.edu.sg](mailto:vincent.lim.2022@mitb.smu.edu.sg) [billy.sim.2022@mitb.smu.edu.sg](mailto:billy.sim.2022@mitb.smu.edu.sg)  [sruthib.2022@mitb.smu.edu.sg](mailto:sruthib.2022@mitb.smu.edu.sg) |

Contents

[1 Business Problem 1](#_Toc136039500)

[1.1 Background 1](#_Toc136039501)

[1.2 Translating Business Question into Data Science Task 1](#_Toc136039502)

[2 Dataset 1](#_Toc136039503)

[3 Design Solution 2](#_Toc136039504)

[3.1) Customer Segmentation: 2](#_Toc136039505)

[3.2) Inventory Management: 2](#_Toc136039506)

[3.3) Market Basket Analysis: 3](#_Toc136039507)

[3.4) Dynamic Product Clustering: 3](#_Toc136039508)

[3.5) Personalized Marketing Strategy: 3](#_Toc136039509)

[3.6) Cross-Selling and Upselling: 3](#_Toc136039510)

[3.7) Personalized Pricing: 3](#_Toc136039511)

[4 Interpret Result into Actionable Intelligence 3](#_Toc136039512)

[4.1 Product 4](#_Toc136039513)

[4.2 Inventory management 4](#_Toc136039514)

[4.3 Customer Types and Advertising 4](#_Toc136039515)

[4.4 Pricing 4](#_Toc136039516)

[5 Envisaged Closed-Loop System 4](#_Toc136039517)

[6 Potential Governance Issue 5](#_Toc136039518)

[References 6](#_Toc136039519)

# Business Problem

## Background

H&M (Hennes & Mauritz) is a multinational clothing company based in Sweden. As of 23 June 2022, H&M Group operates in 75 geographical markets with 4,801 stores under the various company brands. It is also the second largest international clothing retailer after Inditex. In 2021, H&M had a global market share of 7.4%. The company's main competitors are Zara, Uniqlo, and Gap. H&M is a successful company that has grown rapidly over the years and the company's success is due to its focus on fast fashion. Fast fashion refers to a business model and a trend in the fashion industry characterized by the rapid production and consumption of inexpensive clothing. It involves the quick turnover of fashion trends, allowing consumers to purchase trendy and affordable garments that mimic the latest styles seen on fashion runways or worn by celebrities.

## Translating Business Question into Data Science Task

H&M primarily targets young adults with their stylish and affordable clothing. They cater to a diverse range of sizes and styles, making them a popular choice for people of all body types. Given the highly competitive nature of the fast-fashion industry and the already low-price points, it is crucial for fast fashion retailers like H&M to avoid unfavorable product cycles and enhance its profitability. H&M needs to understand and predict market demands accurately. Data can enable H&M to develop a more agile and efficient supply chain, improve customer targeting, optimize inventory management, set appropriate pricing, increase average order value, and align its merchandise with local customer preferences. In this project, we will focus on showcasing how H&M can leverage its data through strategies such as Customer Segmentation, Inventory Management, and Market Basket Analysis to boost sales and ultimately improve profit margins.

# Dataset

Our primary dataset was sourced from Kaggle, provided by H&M. We will be working with three sub-datasets in csv format, namely:

* **articles.csv**: This 25-columns dataset with 105,542 rows, contains information about the items available for sale, including product type (e.g. Vest top, Socks), color, and etc.

A picture containing text, screenshot, line, diagram

Description automatically generatedA picture containing text, screenshot, line, diagram

Description automatically generated

Exhibit 1: Distribution of select data columns of the articles.csv dataset.

* **customers.csv**: This 7-columns, 1.37 million rows dataset is a database of customers, identified by their customer\_id and contains information such as their age, residential postal code, their member status, access to fashion news (none, regularly, monthly), and etc.

A picture containing text, screenshot, diagram, rectangle

Description automatically generatedA graph of a member status

Description automatically generated with low confidenceA picture containing text, screenshot, diagram, line

Description automatically generated

Exhibit 2: Distribution of select data columns of the customers.csv dataset.

* **transactions\_train.csv**: This is a database consisting of 5 columns and 31.8 million rows, representing transactional data of the customers. Each row represents a single item transaction that contains information on the date of transaction, customer\_id, article\_id, price (normalized) and the sales channel of which the transaction took place. The columns customer\_id and article\_id are unique identifiers of the customers.csv and articles.csv datasets respectively.

A picture containing text, screenshot, diagram, rectangle

Description automatically generatedA picture containing text, screenshot, diagram, line

Description automatically generated

Exhibit 3: Distribution of select data columns of the transactions\_train.csv dataset.

**Features engineering**

To improve the usability and relevance of the datasets for subsequent analyses, the team processed and aggregated relevant rows from the base datasets. For example, to facilitate customer segmentation, we aggregated transactions by customer\_id, calculated the sum of transacted prices, and examined which product and color types are most frequently bought by each customer.

A picture containing text, screenshot, font, diagram

Description automatically generated

Exhibit 4: Blue font data columns the customer dataset were featured engineered from the other two datasets.

# Design Solution

## 3.1) Customer Segmentation:

We will employ machine learning techniques to segment H&M customers into distinct categories based on features such as club member status, fashion news frequency, and purchase history. This will allow us to identify high-value and growth customer segments, which can then be targeted with personalized promotional campaigns.

## 3.2) Inventory Management:

We will build predictive models to forecast demand for various products. These forecasts will help H&M optimize its inventory management strategy, ensuring that the right mix of products is available at the right time. Additionally, we can use these forecasts to optimize product placement within stores, improving the overall shopping experience.

## 3.3) Market Basket Analysis:

We will perform market basket analysis to identify pairs or groups of products that are frequently purchased together. These insights can be used to provide personalized product recommendations, thereby increasing basket size and enhancing customer experience. Additionally, these insights can inform marketing strategies, such as the creation of product bundles that encourage additional purchases.

## 3.4) Dynamic Product Clustering:

We will further improve the relevance of our promotional campaigns and product recommendations by dynamically clustering H&M's products. Using unsupervised machine learning techniques, we can group similar products based on characteristics such as type, color, and graphical appearance. This will allow us to tailor our marketing and inventory strategies to the specific preferences of different customer segments.

## 3.5) Personalized Marketing Strategy:

Building on the customer segmentation and product clustering, we can develop a personalized marketing strategy. This will involve targeting specific customer segments with promotions and recommendations for the product clusters that most closely match their preferences.

## 3.6) Cross-Selling and Upselling:

Based on customer segmentation and market basket analysis, we can design cross-selling and upselling strategies. Cross-selling refers to selling different products to the customer (e.g., accessories for purchased clothes), while upselling encourages the purchase of a higher-value alternative to the item being viewed or purchased.

## 3.7) Personalized Pricing:

By understanding customer preferences and willingness to pay through segmentation and past transactions, we can develop personalized pricing strategies. This could involve offering dynamic pricing to different customer segments or even individual customers.

**Performance Metrics**: We will use different metrics to evaluate the performance of each component of our solution. For classification tasks, such as customer segmentation, we can use accuracy or F1 score. For regression tasks, such as demand forecasting, we can use mean absolute error or root mean squared error. For unsupervised tasks, such as product clustering, we can use silhouette score.

Additionally, we will consider metrics such as Average Order Value and Net Promoter Score, or customized metrics that we can use to evaluate how effective and successful the solutions will be.

# Interpret Result into Actionable Intelligence

The project aims to aid H&M in making business decisions that would improve its profitability, hence the models selected will cover these 4 aspects.

## 4.1 Product

All the different products of H&M will first be categorized into different product types based on machine learning methods such as Dynamic Product Clustering. This is done to better understand how the products in H&M are related to one another based on the different customer segments and will provide ease of use when applying the many different models.

## 4.2 Inventory management

Applying machine learning techniques to the transaction data, we will be able to identify the demand of products, through this, management could make better decisions with regards to inventory management and reduce costs by holding only enough products to meet the expected demand.

## 4.3 Customer Types and Advertising

By segregating the customers into segments through Customer Segmentation will help identify groups of customers with certain traits, traits that could be used when it comes to targeted advertising. Based on this result, H&M can use Personalized Marketing Strategies to tailor their advertisements and promotions to specific customer segments

Furthermore, from the results of Market Basket Analysis and Cross/Up Selling, H&M could potentially advertise products that complement the products customers have purchased or recommend higher value alternatives that would better suit their needs to the customers. These methods combined will aid H&M in improving customer retention and acquisition.

## 4.4 Pricing

Through the model, Personalized Pricing, we will be able to predict the prices each customer segment is willing to pay for products. Through the understanding of how each segment will react to prices of individual products, H&M could maximize its returns by pricing its products based on segments of customers that are more resistant to price changes.

# Envisaged Closed-Loop System

According to research by Accenture (Accenture Interactive, 2018), 91% of all consumers prefer to shop with brands who can recognize and provide them with relevant offers and recommendations. The same study revealed that 83% of consumers are prepared to share their data to enjoy a more personalized experience. These statistics underscore the significance of Customer 360 view which consolidates all customer data into a singular, unified perspective. This approach promotes enhanced engagement and personalized service for customers.

For our project, we aim to achieve a closed-loop system to monitor and enhance interaction with customers and provision of personalized service. To achieve this, we will look at the ingestion of various data sources to create a comprehensive Customer 360 view. This includes data generated through H&M’s direct interactions with customers, such as click-through rates, browsing patterns and sales data. Furthermore, we will be looking at utilizing external data sources. For instance, social media content can provide valuable insights into the effectiveness of specific marketing campaigns, contributing to our understanding of customers’ behaviors and preferences. We can also enrich our data sources using demographic information, economic factors or other third-party data that can enhance our understanding of customers’ needs.

With the plethora of data sources, it will allow for a comprehensive Customer 360 view to provide personalized services as well as a tight feedback loop that allows H&M to understand the effect of their marketing campaigns.

# Potential Governance Issue

While the benefits of a comprehensive Customer 360 view are significant, it is also important to address potential governance concerns associated with bias and fairness. Bias in data collection, processing, or analysis can lead to skewed results, which can affect the quality of personalized experiences. Bias can occur due to an over-reliance on certain types of data, neglecting certain customer groups, or inadvertently incorporating societal biases present in the data sources.

To mitigate these risks, we can take proactive steps to identify and eliminate bias from the data collection, processing, and decision-making processes. For instance, diverse data sets covering different customer groups and demographics should be collected to reduce the likelihood of a bias perspective.

While personalized pricing can be a powerful tool for maximizing profitability, it could have ethical and fairness issues. Charging different prices to different customers for a similar product, based on their willingness to pay, might be seen as discriminatory. For this reason, dynamic pricing needs to be implemented in a transparent and fair manner.

As companies accumulate larger volumes of customer data and leverage an increasing array of data analytics tools to enhance business processes, operational efficiency, and financial performance, it becomes imperative for them to address the escalating cost required to uphold data integrity. This is particularly significant due to the intensified focus on privacy and data security by global regulatory bodies such as the European Data Protection Board. Companies must carefully evaluate the advantages derived from expanded data collection against the concurrent rise in compliance costs associated with handling such data.

# References

Accenture Interactive. (2018). 2018 Personalisation Pulse Check. Retrieved from Accenture: <https://www.accenture.com/t20171031T052156Z__w__/us-en/_acnmedia/PDF-34/Accenture-Pulse-Check-Dive-Key-Findings-Personalized-Experiences.pdf>

**Data Source –**

https://www.kaggle.com/competitions/h-and-m-personalized-fashion-recommendations/data