Introduction of company:

The company H&M which is an initialism for Hennes & Mauritz from the fashion domain is chosen. H&M is the second largest multinational wholesaler in the fashion area. Founder of H&M is Erling Persson and it started in 1947. H&M initially sold women's apparel in Sweden and expanded to sell different types of clothes, shoes, accessories and cosmetics. The mission of H&M is to make fashion accessible and enjoyable for all while the vision is to lead the change towards circular and climate positive fashion while being a fair and equal company. Their products are always up-to-date, unique and suit all age groups. They always improve their product quality and design to meet customers' expectations and requirements. People can find their ideal clothes with the most reasonable price at H&M shops. H&M drives for continuous improvement, creativity and innovation in their products.

Dataset used:

Association Rules

The dataset used in this project is taken from data flair which was published by Arvel Miller in 2021. Two datasets which are H&M Sales 2018 and 2019 are taken and then merged into one excel file in order to perform data analysis. This dataset contains 240 rows with 15 attributes including category, sub-category, sales and quantity which are quite important for the analysis.

Sentiment Analysis

The data used for Sentiment Analysis is the data collected from Twitter by using Rapid Miner. This dataset has 12 attributes and 1000 rows.

Justification of Choice of Dataset

Association Rules

Since H&M is a famous and familiar brand, many people get to use their products. Hence, we are wondering how they can ensure their products are always ready for stock and the relationship between their products. Besides, there are no missing values found in this dataset and the attributes included are easy to understand and suitable for this project. Therefore, H&M Sales 2018 and 2019 are chosen for analysis.

Table 1.1: Definitions of the Attributes in the Dataset H&M Sales 2018 and 2019

Attributes	Definition
Order ID	Unique ID for each order of the products
Order Date	The date that customers complete the transaction
Ship Mode	Way of shipping products
Customer ID	Unique ID for each customer
Country	Country of H&M store
City	City of H&M store
State	State of H&M store
Region	Region of H&M store
Zone	Zone of H&M store
Product ID	Unique ID for each product
Category	Categories of all the products
Sub-Category	Subcategories of all the products
Sales	Total sales of each product
Quantity	Amount of each product sold
Discount	Discount of each product
Profit	Profit of each product

Sentiment Analysis

Customers' expectations can be determined through customer feedback. Due to the widely used of social media, customers tend to give comments through their personal accounts. Therefore, data collected from Twitter is suitable for analysing customer satisfaction levels.

Table 1.2: Definitions of the Attributes in the Dataset Collected from Twitter

No	Variable	Description	Data Type
1	Created-At	The time and date that the tweet was created	Date
2	From-User	The user who creates the tweet	Nominal
3	From-User-Id	The user id that creates the tweet	Real
4	To-User	The user who receives the tweet	Nominal
5	To-User-Id	The user id that receives the tweet	Real
6	Language	The language used in the tweet	Nominal
7	Source	The source of the tweet	Nominal
8	Text	The content of the tweet	Nominal
9	Geo-Location-Latitude	The latitude location from which the tweet was sent	Nominal
10	Geo-Location-Longitude	The longitude location from which the tweet was sent	Nominal
11	Retweet-Count	Number of times the tweet was retweeted	Integer
12	Id	User ID	Real

Business problem:

Many fashion companies including H&M face problems in inventory management. Sometimes, their products are unavailable or face overproduction for certain products. It may cause the company in running loss or the customers have a bad impression on H&M due to out of stock issues. Hence, they would like to conduct inventory control on their product so that there are sufficient products to meet demand without out of stock or excess inventory.

Since H&M provide a lot of choices in clothing such as dresses, jackets, jeans and trousers, it is hard for them to identify the products that contribute the highest profits. By knowing the best selling and customer's buying patterns, H&M is able to put them together at the shelf in order to persuade the people to buy and save their time in finding different products at different places. Besides, H&M companies wish to place the best seller at the most conspicuous place so that the customers can find the clothes easily and thus increase sales.

E-commerce plays an important role in the retail industry as it can provide a flexible and convenient shopping experience to the consumers. The presence of e-commerce allows the company to provide their product and services to a wide range of customers. It also allows the customers to buy products and services anytime and anywhere without going to the physical store. H&M companies wish to utilise e-commerce platforms such as Shein, Lazada, Zalora and so on to promote their product in order to boost their sales. Therefore, it is vital to understand the customers' preference on the product categories. The goal of this project is to help H&M companies to have a better understanding on the relationship of their products as well as increase sales.

Customer satisfaction has also been one of the concerns of H&M as the company needs to have insight into customer satisfaction to perform continuous improvement on the company's overall business decisions, including the business strategies. Thus, H&M has signed up for several official accounts on social media platforms such as Facebook, Twitters, Instagram, YouTube, and etc, in order to receive feedback and suggestions from customers to understand the customers' opinions toward H&M and boost the sales.

Objective of Analysis

1. To identify the product that has the highest turnover rate in order to keep track of inventory
2. To determine the buying pattern of customers.
3. To generate online shopping recommendation items to customers
4. To analyse the customer satisfaction level

Related work

Research is conducted by M. Hamdani [5] with the aim to increase product sales by finding association rules from product purchase transactions of Indomaret Tanjung Anom using the Apriori Algorithm. Association rule mining which is a data mining technique is applied to determine rules or combinations of items or frequent itemsets. From the result of the strength level of association rules, the association rules generated by the Apriori algorithm have a higher level of strength than those generated by the FP-growth algorithm. However, the discussion without FP-growth in detail becomes the gap and inconsistency of this research because it leads to the lack of information about the use of FP-growth. Based on the research conducted, association rule mining using the Apriori Algorithm successfully determines the buying patterns through a transaction in order to increase sales by combining goods used as sales packages or bundling which is the key finding of this study.

Similar research is done by Y. Kurnia [4] through analysis of customers' purchasing patterns that can be used as recommendations in determining the sales strategy and developing the right target promotion for O! Fish restaurants in order to boost sales. The gap or inconsistency in the review is the limitation of the transaction data usage by the author. The data used is the transaction data that occurs within 5 days only which may lead to some inaccuracy in finding the association rules for a restaurant that operates for the long term. From this study, the importance of determining the minimum support value and minimum confidence value is investigated. If a lower minimum support value and minimum confidence value are set, then the accuracy value produced will be low. This is because items that have a weak association in customer purchases will be displayed in the results also since more association rules will be generated due to the low minimum support and confidence value. In contrast, the higher the minimum support value and minimum confidence value, the higher the accuracy values will be generated because the rules of a strong purchasing association will be raised in the results.

Another study is conducted by Y. A. Ünvan [6] which focused on conducting a market basket analysis using association rules. In the study, two association rules algorithms, Apriori and FP Growth were tried to analyse the supermarket sales data. The gap or inconsistency found is that the researcher did not study well on the nature of the dataset, causing the Apriori algorithm to not produce any output. This is because the dataset is categorical. Thus, the researcher opted to utilise the FP Growth algorithm and the conviction value was considered to determine the top 10 rules. As a result, the best rule is the rule with 21.06 conviction and 100% confidence value. The rule states that a customer who buys milk, sweet relish and pepperoni pizza will also buy eggs. The product placement in the supermarket is made according to the rules to boost sales and increase revenue. The key finding of the review is that if market basket analysis is interpreted correctly, it is definitely profitable.

In addition to association rules, V. Umarani and M. Subathra [7] have investigated classification techniques to predict customer buying patterns based on gender, age and salary. Classification is a supervised machine learning algorithm that classifies the new observations based on the training data. In the research, K-nearest Neighbour (KNN) and Decision Tree algorithms are applied on the social network dataset to predict the customer buying patterns. The distance between the test data and all of the training points is calculated by the KNN algorithm. Subsequently, the correct class for the test data is predicted. The KNN algorithm is applied to the dataset with 10-fold cross validation. The classification report generated for the model indicates that KNN has an accuracy score that is not very high.

The Decision Tree algorithm is also applied by V. Umarani and M. Subathra [7] in the research. It is an algorithm that recursively divides the data into subsets according to the most important attribute at each node of the tree. The Decision Tree model is applied to the social network dataset with 10-fold cross validation as well. The dataset is divided into 10 equal-sized subsets to be used in cross validation. Based on the classification report generated, the Decision Tree model performs better than the KNN model with respect to classification average accuracy and performance issues. The gap or inconsistency in the review is that the results may be different if the parameters are fine tuned. To increase the accuracy in predicting the buying patterns, the parameters of the classifiers should be optimised. In addition, the results will also vary based on the nature of the dataset. The key finding of the review is that the choice of algorithm depends on the application as well as the nature of the data.

A research is conducted by Olivera Grljevic [2] to highlight the importance of customer satisfaction, the impact of social media on consumer behaviour, and propose sentiment analysis as a valuable tool for businesses to extract insights from social media content. The study conducted sentiment analysis on different sources of data (Amazon, IMDB, Yelp) using different classifiers. The performance varied depending on the specific data and classifier chosen, emphasising the importance of selecting appropriate algorithms for different contexts. The gap or inconsistency found is the difficulties arising from negation, irony, and implicitly expressed sentiment. Overall, the key finding is that sentiment analysis is important in understanding customer opinions and utilising social media data to improve business strategies.

Shubham Agarwal and Madhavi Damle [1] have done research on sentiment analysis to evaluate influencer marketing. The study determined the appropriate influencer marketing parameters using sentiment analysis. The researcher utilised secondary data from reliable secondary sources and employed conceptual demonstration to analyse the data and draw conclusions. The gap or inconsistency is that there may be fake followers or bots. Bots are computer programmes that are designed to induce behaviour similar to humans, which can be misleading. The key finding is that sentiment analysis can aid in making decisions regarding influencer selection, enhancing the impact and assessment of influencer marketing efforts.

Based on the study from related works by many researchers, the most common business intelligence technique used is Association Rules Mining. M. Hamdani [5], Y. Kurnia [4] and Y. A. Ünvan [6] have analysed the items purchased simultaneously in order to identify customers' buying patterns by using the Apriori Algorithm which is a type of Association Rules with the purpose of boosting sales, determining promotional development strategy or facilitating sales packages. M. Hamdani [4] suggests that association rules discovered can be a reference to the stock of goods and automate the calculation of the analysis of goods sales. Other than that, the study by V. Umarani and M. Subathra [7] suggests that the Decision Tree algorithm performs better in predicting customer buying patterns. However, the choice of algorithm depends on the application and nature of the data. Additionally, the research done by Olivera Grljevic [2] proposes sentiment analysis as a tool for businesses to extract insights from social media content, and emphasises the importance of choosing suitable algorithms for different contexts. Lastly, Shubham Agarwal and Madhavi Damle [1] emphasises the value of sentiment analysis in decision making and enhancing the effectiveness of influencer marketing strategies.

Data exploration

Association Rules

R programming is used to explore the dataset 'H&M_Sales'. The summary of the dataset before data preprocessing is shown in Figure 4.1.

Figure 4.1: Summary of H&M_Sales before data preprocessing

```
  OrderID            OrderDate                        Ship_Mode          CustomerID          Country
 Length:249         Min.   :2018-01-16 00:00:00.00   Length:249         Length:249         Length:249
 Class :character   1st Qu.:2018-10-11 00:00:00.00   Class :character   Class :character   Class :character
 Mode  :character   Median :2019-05-28 00:00:00.00   Mode  :character   Mode  :character   Mode  :character
                    Mean   :2019-03-27 02:30:21.69
                    3rd Qu.:2019-10-12 00:00:00.00
                    Max.   :2019-12-28 00:00:00.00
     City              State               Zone            ProductID           Category
 Length:249         Length:249         Length:249         Length:249         Length:249
 Class :character   Class :character   Class :character   Class :character   Class :character
 Mode  :character   Mode  :character   Mode  :character   Mode  :character   Mode  :character



 Sub_Category           Sales              Quantity        Discount           Profit
 Length:249         Min.   :   1.248   Min.   : 1.00   Min.   :0.0000   Min.   :-1665.052
 Class :character   1st Qu.:  20.016   1st Qu.: 2.00   1st Qu.:0.0000   1st Qu.:    1.204
 Mode  :character   Median :  66.284   Median : 3.00   Median :0.2000   Median :    7.250
                    Mean   : 238.835   Mean   : 3.94   Mean   :0.1778   Mean   :   -5.233
                    3rd Qu.: 208.560   3rd Qu.: 5.00   3rd Qu.:0.2000   3rd Qu.:   21.259
                    Max.   :8159.952   Max.   :14.00   Max.   :0.8000   Max.   :  585.552
```

From Figure 4.1, the dataset consists of 249 observations and 15 features. There are 10 string-type features, which are OrderID, Ship_Mode, CustomerID, Country, City, State, Zone, ProductID, Category, and Sub_Category. The OrderDate is a date-type feature, while the Sales, Quantity, Discount, and Profit are numerical data. Since the information of the string-type data is not displayed, the string-type features except the OrderID and ProductID are then converted into factors for better understanding.

Figure 4.2: Summary of H&M_Sales after data preprocessing

```
   OrderID             OrderDate                          Ship_Mode      CustomerID                Country
 Length:249         Min.   :2018-01-16 00:00:00.00   First Class   : 46   Length:249         United States:249
 Class :character   1st Qu.:2018-10-11 00:00:00.00   Second Class  : 54   Class :character
 Mode  :character   Median :2019-05-28 00:00:00.00   Standard Class:149   Mode  :character
                    Mean   :2019-03-27 02:30:21.69
                    3rd Qu.:2019-10-12 00:00:00.00
                    Max.   :2019-12-28 00:00:00.00

            City           State         Zone       ProductID              Category     Sub_Category
 Los Angeles   : 27   California:52   Central:83   Length:249         Accessories: 46   Dresses :32
 Chicago       : 14   Texas     :27   East   :47   Class :character   Clothing   :145   Formals :29
 Houston       : 12   Ohio      :24   North  :11   Mode  :character   Footwear   : 58   Bags    :26
 Philadelphia  : 12   Illinois  :19   South  :29                                        Tops    :25
 New York City: 11    New York  :18   West   :79                                        Jackets :23
 San Francisco: 11    Minnesota :13                                                     Sneakers:23
 (Other)      :162    (Other)   :96                                                     (Other) :91
     Sales              Quantity          Discount            Profit
 Min.   :   1.248   Min.   : 1.00   Min.   :0.0000   Min.   :-1665.052
 1st Qu.:  20.016   1st Qu.: 2.00   1st Qu.:0.0000   1st Qu.:    1.204
 Median :  66.284   Median : 3.00   Median :0.2000   Median :    7.250
 Mean   : 238.835   Mean   : 3.94   Mean   :0.1778   Mean   :   -5.233
 3rd Qu.: 208.560   3rd Qu.: 5.00   3rd Qu.:0.2000   3rd Qu.:   21.259
 Max.   :8159.952   Max.   :14.00   Max.   :0.8000   Max.   :  585.552
```

Based on Figure 4.2, it is observed that no missing value exists in the dataset, but there are unusual statistics represented for the Profit. The minimum value and the mean of Profit are negative values, which indicates H&M might suffer from loss in the years 2018 and 2019. Also, the dataset only consists of the data collected from the United States, and the items sold are categorised into Accessories, Clothing, and Footwear.

Sentiment Analysis

During the process of data cleaning and preprocessing, the unnecessary columns are removed as they do not contribute to the objectives of the study. Besides, the unnecessary words, characters and stop words in the Twitter text are removed. Then, the duplicate data and missing values are also eliminated from the dataset.

Figure 4.3: Result of H&M Twitter data after preprocessing

Row No.	Lower Case...	Created-At	From-User	Retweet-Co...
1	xdinary heroes	May 15, 2023...	Xdinary Heroes	2922
2	weather anta...	May 14, 2023...	Massimo	1199
3	lady peter o...	May 15, 2023...	SportsDokita ...	1240
4	inspiration	May 15, 2023...	Rocelyn Fort...	0
5	shut real safe...	May 15, 2023...	ابوحبيبه	0
6	disposition s...	May 15, 2023...	Antaro abuba...	0
7	crazy hm rn p...	May 15, 2023...	estele⁷	0
8	excellence vi...	May 15, 2023...	januar dwi ra...	0
9	blessings wri...	May 15, 2023...	Unique Armst...	0
10	young would...	May 15, 2023...	fann pangera...	0
11	learned don...	May 15, 2023...	innocent larki	0
12	claimed pepe	May 15, 2023...	Yuri Hanniel	0
13	agreed respe...	May 15, 2023...	Zyeria Jones	0

ExampleSet (983 examples, 0 special attributes, 4 regular attributes)

Based on Figure 4.3, it can be seen that only 983 examples are left with 4 attributes. The data will be used for sentiment analysis later.

Figure 4.4: Summary of H&M Twitter data

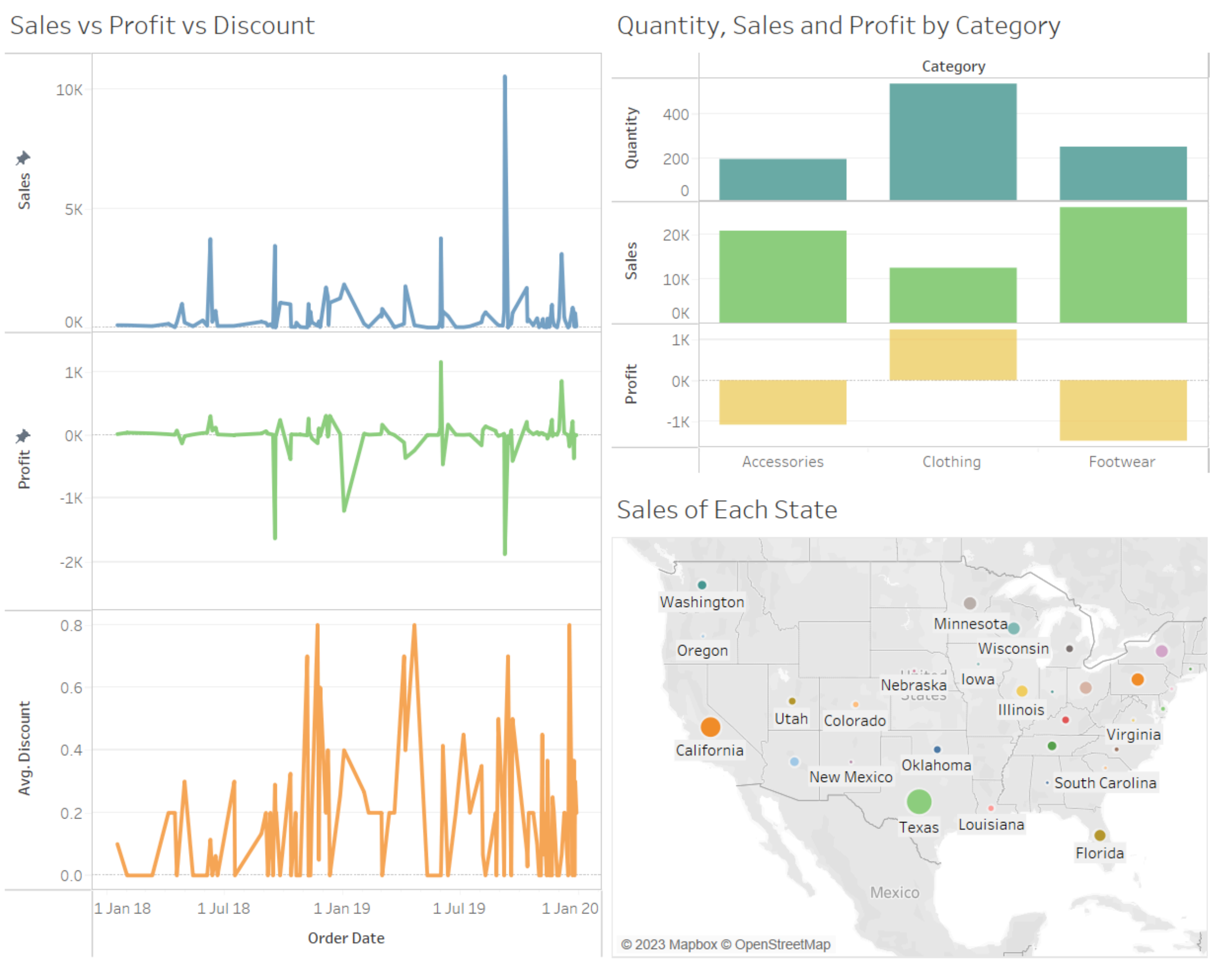
Name	Type	Missing	Statistics		
Lower Case Text	Nominal	0	Least 🤷♂ (1)	Most 003rohit [...] govt... (1)	Values 003rohit [...] bhi govt... (1), 1656 ces [...] cb 2184
Created-At	Date-time	0	Earliest date May 14, 2023 10:54 PM	Latest date May 15, 2023 9:43 PM	Duration 0d 22h 49m 2s
From-User	Nominal	0	Least ◆ eli ◆ (1)	Most Teguh fathi. (16)	Values Teguh fathi. (16), Zyeria Jones (16), ...[197 more]
Retweet-Count	Integer	0	Min 0	Max 23452	Average 110.728

Based on Figure 4.4, there are no missing values found in the data.

Descriptive analytics

Association Rules

Figure 5.1: Dashboard regarding the profitability of H&M

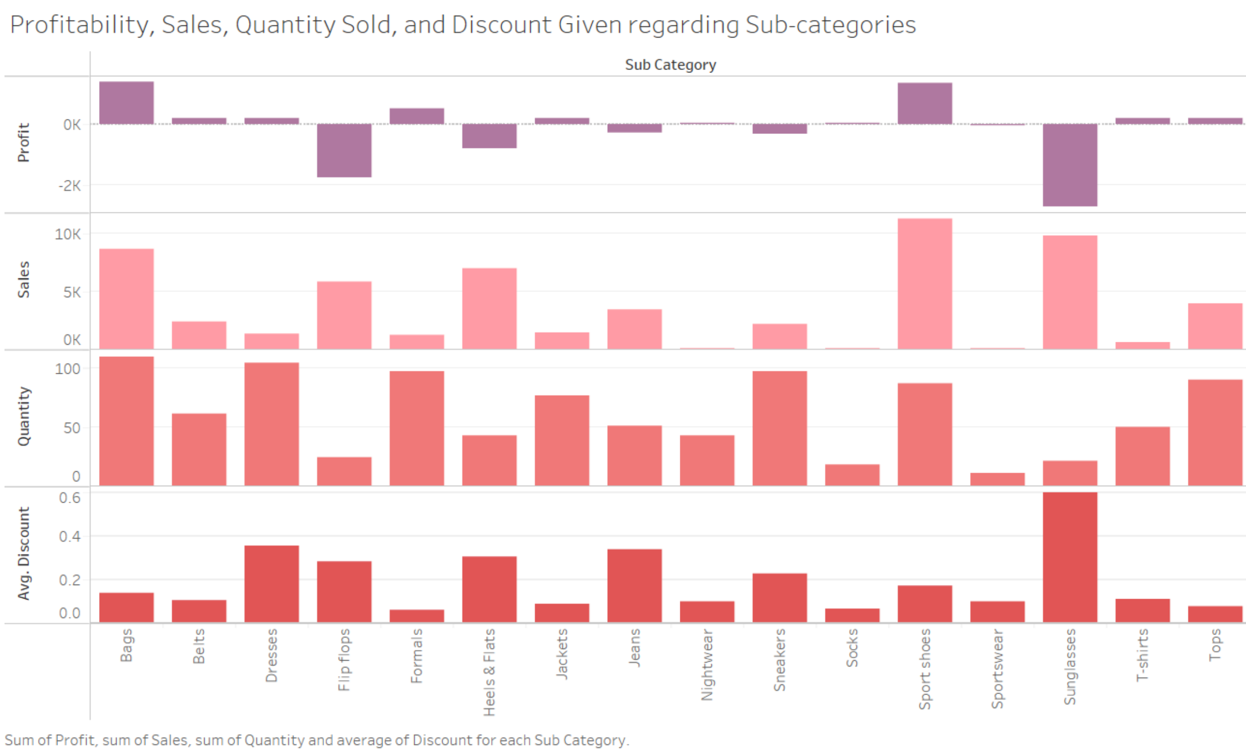


A comparison between the Sales, Profit, and Discount from January 2018 to December 2019 is visualised to see if the three features are related to each other in Figure 5.1. Typically, when the discount given in each transaction is high, but the sales are low, the company will bear a slight loss as shown in November 2018 (the region before January 2019). However, when the discount given is high and the sales are also high, the company will bear a huge amount of loss, as shown in September 2019 (the region after July 2019).

From the bar plots of Quantity, Sales and Profit by Category, it is observed that the category that appears the most in the dataset is Clothing, as its quantity was the highest and it was the only category that helps the company generate profit, although with the lowest sales. Conversely, the other two categories achieved higher sales with fewer units sold but resulted in the company's loss.

Furthermore, investigation on the Sales of Each State graph displays that the 3 states that contributed the top 3 sales are Texas, California, and Minnesota as they show the largest circles which indicate the numbers of sales were the highest. Besides, there are also several states that did not present any sales records.

Figure 5.2: Sub-category vs Profit, Sales, Quantity, and Discount



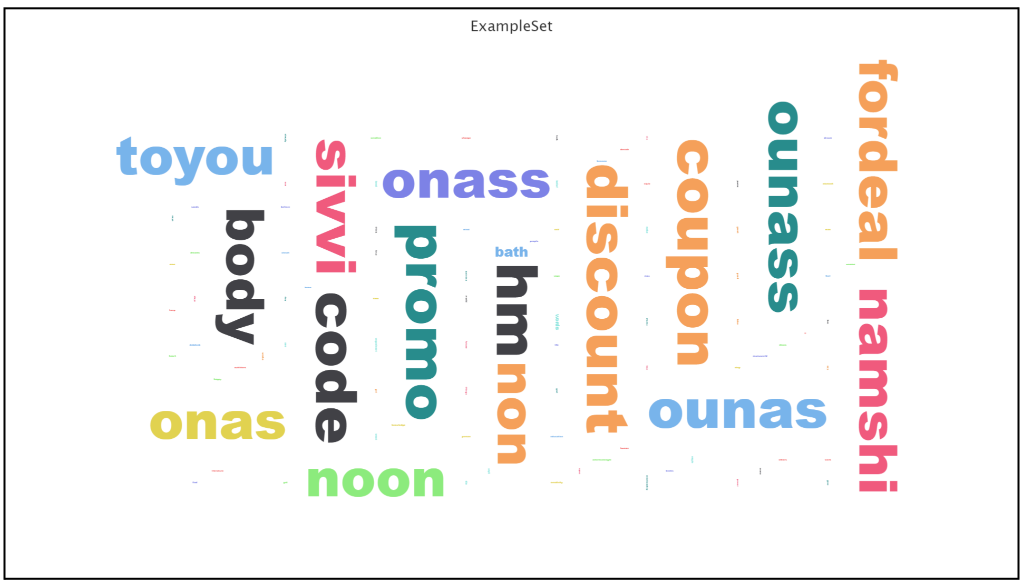
Observing Figure 5.2, the highest profit earned by the company was achieved via the selling of Bags. To highlight, the Bags was the best selling product of the company as its total quantity sold over the 2 years was over 100 units, and the sales hit by selling Bags was also quite high, with few portions of discount given.

Moreover, Sport shoes recorded the highest sales that resulted in a relatively high profit, but the number of quantity sold was just moderate, which was less than 100 units. On the other hand, the sales of Sunglasses was the second highest, however, it led to the largest amount of loss. The large number of sales and losses might be explained by the highest amount of discount given, which was an average of 60% discount.

The other products that were sold but led to the loss of the company were Flip flops, Heels & flats, Jeans, and Sneakers. The loss caused by selling these products was probably due to the amount of discount given was quite impressive.

Sentiment Analysis

Figure 5.3: Wordcloud



As expected, hm which indicates H&M is one of the words that most frequently occur in the tweet. When a promotion is conducted, customer satisfaction achieves and hence customers may tend to share information or feelings through a comment. Hence, the words promo, code, discount coupon, toyou and fordeal may be widely used. Sivvi, Namshi and Ounass are the Middle East's premier luxury websites, featuring the world's most exclusive brands in a globally relevant yet locally driven editorial environment.

Data analysis

Association Rules

The Apriori algorithm is applied to discover usable patterns in the co-occurrence of the items in transactions, and determine reliable association rules.

Figure 6.1: Summary of the transactions

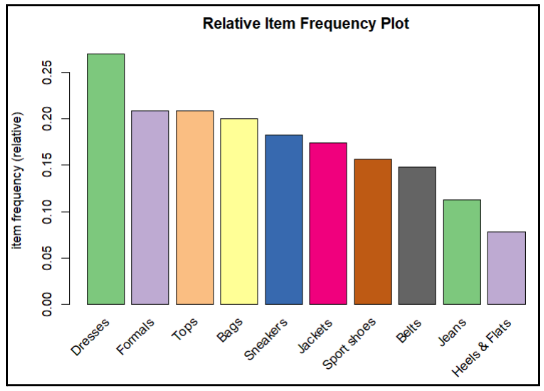
```
transactions as itemMatrix in sparse format with
 115 rows (elements/itemsets/transactions) and
 16 columns (items) and a density of 0.125

most frequent items:
 Dresses   Formals      Tops      Bags Sneakers   (Other)
      31        24        24        23       21       107

element (itemset/transaction) length distribution:
sizes
 1  2  3  4  5  6
58 29 12  7  4  5

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
      1       1       1       2       2       6
```

Figure 6.2: Top 10 frequent items



According to Figure 6.1, regrouping the transaction based on OrderID, OrderDate, and CustormerID, the 249 observations are grouped into 115 transactions. The most frequent items that appear in the transactions are Dresses (31), Formal (24), Tops (24), Bags (23), and Sneakers (21). Moreover, the majority of the transactions consist of only one item.

Figure 6.3: Rules involving Sunglasses.

	lhs		rhs	support	confidence	coverage	lift	count
[1]	{Flip flops, Socks}	=>	{Sunglasses}	0.008695652	1.0	0.008695652	38.33333	1
[2]	{Belts, Flip flops}	=>	{Sunglasses}	0.008695652	1.0	0.008695652	38.33333	1
[3]	{Belts, Flip flops, Socks}	=>	{Sunglasses}	0.008695652	1.0	0.008695652	38.33333	1
[4]	{Bags, Flip flops, Socks}	=>	{Sunglasses}	0.008695652	1.0	0.008695652	38.33333	1
[5]	{Bags, Belts, Socks}	=>	{Sunglasses}	0.008695652	1.0	0.008695652	38.33333	1
[6]	{Bags, Belts, Flip flops}	=>	{Sunglasses}	0.008695652	1.0	0.008695652	38.33333	1
[7]	{Jackets, Jeans, Sport shoes}	=>	{Sunglasses}	0.008695652	1.0	0.008695652	38.33333	1
[8]	{Bags, Jeans, Sport shoes}	=>	{Sunglasses}	0.008695652	1.0	0.008695652	38.33333	1
[9]	{Jackets, Jeans, Tops}	=>	{Sunglasses}	0.008695652	1.0	0.008695652	38.33333	1
[10]	{Bags, Jeans, Tops}	=>	{Sunglasses}	0.008695652	1.0	0.008695652	38.33333	1
[11]	{Bags, Sport shoes, Tops}	=>	{Sunglasses}	0.008695652	1.0	0.008695652	38.33333	1
[12]	{Bags, Jackets, Sport shoes}	=>	{Sunglasses}	0.008695652	1.0	0.008695652	38.33333	1
[13]	{Bags, Jackets, Tops}	=>	{Sunglasses}	0.008695652	1.0	0.008695652	38.33333	1
[14]	{Bags, Belts, Flip flops, Socks}	=>	{Sunglasses}	0.008695652	1.0	0.008695652	38.33333	1
[15]	{Jackets, Jeans, Sport shoes, Tops}	=>	{Sunglasses}	0.008695652	1.0	0.008695652	38.33333	1
[16]	{Bags, Jeans, Sport shoes, Tops}	=>	{Sunglasses}	0.008695652	1.0	0.008695652	38.33333	1
[17]	{Bags, Jackets, Jeans, Sport shoes}	=>	{Sunglasses}	0.008695652	1.0	0.008695652	38.33333	1
[18]	{Bags, Jackets, Jeans, Tops}	=>	{Sunglasses}	0.008695652	1.0	0.008695652	38.33333	1
[19]	{Bags, Jackets, Sport shoes, Tops}	=>	{Sunglasses}	0.008695652	1.0	0.008695652	38.33333	1
[20]	{Bags, Jackets, Jeans, Sport shoes, Tops}	=>	{Sunglasses}	0.008695652	1.0	0.008695652	38.33333	1
[21]	{Belts, Socks}	=>	{Sunglasses}	0.008695652	0.5	0.017391304	19.16667	1
[22]	{Bags, Socks}	=>	{Sunglasses}	0.008695652	0.5	0.017391304	19.16667	1
[23]	{Nightwear, Sneakers}	=>	{Sunglasses}	0.008695652	0.5	0.017391304	19.16667	1
[24]	{Jackets, Nightwear}	=>	{Sunglasses}	0.008695652	0.5	0.017391304	19.16667	1
[25]	{Jeans, Sport shoes}	=>	{Sunglasses}	0.008695652	0.5	0.017391304	19.16667	1
[26]	{Jeans, Tops}	=>	{Sunglasses}	0.008695652	0.5	0.017391304	19.16667	1
[27]	{Jackets, Sport shoes}	=>	{Sunglasses}	0.008695652	0.5	0.017391304	19.16667	1
[28]	{Formals, Nightwear, Sneakers}	=>	{Sunglasses}	0.008695652	0.5	0.017391304	19.16667	1
[29]	{Formals, Jackets, Nightwear}	=>	{Sunglasses}	0.008695652	0.5	0.017391304	19.16667	1
[30]	{Dresses, Formals, Nightwear}	=>	{Sunglasses}	0.008695652	0.5	0.017391304	19.16667	1
[31]	{Jackets, Nightwear, Sneakers}	=>	{Sunglasses}	0.008695652	0.5	0.017391304	19.16667	1
[32]	{Dresses, Nightwear, Sneakers}	=>	{Sunglasses}	0.008695652	0.5	0.017391304	19.16667	1
[33]	{Dresses, Jackets, Nightwear}	=>	{Sunglasses}	0.008695652	0.5	0.017391304	19.16667	1
[34]	{Jeans, Sport shoes, Tops}	=>	{Sunglasses}	0.008695652	0.5	0.017391304	19.16667	1
[35]	{Bags, Jackets, Jeans}	=>	{Sunglasses}	0.008695652	0.5	0.017391304	19.16667	1
[36]	{Jackets, Sport shoes, Tops}	=>	{Sunglasses}	0.008695652	0.5	0.017391304	19.16667	1
[37]	{Formals, Jackets, Nightwear, Sneakers}	=>	{Sunglasses}	0.008695652	0.5	0.017391304	19.16667	1
[38]	{Dresses, Formals, Nightwear, Sneakers}	=>	{Sunglasses}	0.008695652	0.5	0.017391304	19.16667	1
[39]	{Dresses, Formals, Jackets, Nightwear}	=>	{Sunglasses}	0.008695652	0.5	0.017391304	19.16667	1
[40]	{Dresses, Jackets, Nightwear, Sneakers}	=>	{Sunglasses}	0.008695652	0.5	0.017391304	19.16667	1
[41]	{Dresses, Formals, Jackets, Sneakers}	=>	{Sunglasses}	0.008695652	0.5	0.017391304	19.16667	1
[42]	{Dresses, Formals, Jackets, Nightwear, Sneakers}	=>	{Sunglasses}	0.008695652	0.5	0.017391304	19.16667	1

Based on Figure 5.2, since the selling of Sunglasses was the main contributor to the loss of the company, the itemsets involving sunglasses are investigated and shown in Figure 6.3. From Figure 6.3, there are 42 rules involving Sunglasses. It is clear that the values of lift of these rules are high, up to 38.333, implying these rules are very interesting, and the rules are quite reliable as their confidence values are above 0.5. However, the supports of these rules are very low, not more than 1%, which means these itemsets just occur only once in the transactions. Hence, these rules are not worth further analysis.

In order to discover meaningful and interesting underlying rules in the transactions, several criteria with different parameters are set:

1. Minimum support is 2% with 30% confidence.
2. Minimum support is 2% with 50% confidence.
3. Minimum support is 2% with 60% confidence.
4. Minimum support is 3% with 60% confidence.

With minimum support of 2%, the itemset is said to exist at least three times in the transaction. Also, as the confidence value increases, the likeliness of the co-occurrence of the items also increases.

In Figure 6.4, it can be observed that Dresses and Sneakers are mutually dependent as when either item is observed, the other item will also be observed. The itemsets related to the Dresses and Sneakers also appear most frequently among the transactions, hence it is no wonder the support of the itemsets are the highest.

In addition, the itemset with the strongest association, 1.0000 also found to be related to the Dresses and Sneakers, which is when Belts and Sneakers are observed, Dresses will also be observed. The most interesting rule is also involving the Dresses and Sneakers, which is when dresses and formals are observed, sneakers will also be observed, with the highest lift, 4.3125, and the rule is 75% reliable.

Comparing Figure 6.5 and Figure 6.6 to Figure 6.4, one of the differences discovered is that as the threshold of confidence increases, the number of rules that satisfy the threshold decreases. Also, the itemsets related to Dresses and Sneakers are no longer the itemsets that occur most frequently when the minimum confidence changes.

From Figure 6.7, it is obvious that the increase in the minimum support has led to a change in the most interesting rule and, again decrease in the number of rules. The most interesting rule from Criteria 4 has now become when Dresses and Formals are observed, Sneakers will be observed, with a lift of 3.6508, and is 66.67% likely to occur.

Figure 6.4: Rules from Criterion 1

	lhs		rhs	support	confidence	coverage	lift	count
[1]	{Bags, Formals}	=>	{Jackets}	0.02608696	0.7500000	0.03478261	4.312500	3
[2]	{Belts, Sneakers}	=>	{Dresses}	0.02608696	1.0000000	0.02608696	3.709677	3
[3]	{Dresses, Formals}	=>	{Sneakers}	0.03478261	0.6666667	0.05217391	3.650794	4
[4]	{Formals, Sneakers}	=>	{Jackets}	0.02608696	0.6000000	0.04347826	3.450000	3
[5]	{Nightwear}	=>	{Formals}	0.03478261	0.6666667	0.05217391	3.194444	4
[6]	{Dresses, Jackets}	=>	{Sneakers}	0.03478261	0.5714286	0.06086957	3.129252	4
[7]	{Flip flops}	=>	{Bags}	0.02608696	0.6000000	0.04347826	3.000000	3
[8]	{Formals, Sneakers}	=>	{Dresses}	0.03478261	0.8000000	0.04347826	2.967742	4
[9]	{Jackets, Sneakers}	=>	{Dresses}	0.03478261	0.8000000	0.04347826	2.967742	4
[10]	{Jackets, Sneakers}	=>	{Formals}	0.02608696	0.6000000	0.04347826	2.875000	3
[11]	{Dresses, Formals}	=>	{Jackets}	0.02608696	0.5000000	0.05217391	2.875000	3
[12]	{Nightwear}	=>	{Bags}	0.02608696	0.5000000	0.05217391	2.500000	3
[13]	{Belts, Dresses}	=>	{Sneakers}	0.02608696	0.4285714	0.06086957	2.346939	3
[14]	{Formals, Jackets}	=>	{Sneakers}	0.02608696	0.4285714	0.06086957	2.346939	3
[15]	{Formals, Jackets}	=>	{Bags}	0.02608696	0.4285714	0.06086957	2.142857	3
[16]	{Dresses, Sneakers}	=>	{Jackets}	0.03478261	0.3636364	0.09565217	2.090909	4
[17]	{Dresses, Jackets}	=>	{Formals}	0.02608696	0.4285714	0.06086957	2.053571	3
[18]	{Jackets}	=>	{Bags}	0.06956522	0.4000000	0.17391304	2.000000	8
[19]	{Bags}	=>	{Jackets}	0.06956522	0.3478261	0.20000000	2.000000	8
[20]	{Dresses}	=>	{Sneakers}	0.09565217	0.3548387	0.26956522	1.943164	11
[21]	{Sneakers}	=>	{Dresses}	0.09565217	0.5238095	0.18260870	1.943164	11
[22]	{Heels & Flats}	=>	{Jackets}	0.02608696	0.3333333	0.07826087	1.916667	3
[23]	{Nightwear}	=>	{Dresses}	0.02608696	0.5000000	0.05217391	1.854839	3
[24]	{Heels & Flats}	=>	{Sneakers}	0.02608696	0.3333333	0.07826087	1.825397	3
[25]	{Bags, Jackets}	=>	{Formals}	0.02608696	0.3750000	0.06956522	1.796875	3
[26]	{Dresses, Sneakers}	=>	{Formals}	0.03478261	0.3636364	0.09565217	1.742424	4
[27]	{Jackets}	=>	{Formals}	0.06086957	0.3500000	0.17391304	1.677083	7
[28]	{Formals, Jackets}	=>	{Dresses}	0.02608696	0.4285714	0.06086957	1.589862	3
[29]	{Belts}	=>	{Dresses}	0.06086957	0.4117647	0.14782609	1.527514	7
[30]	{Jackets}	=>	{Dresses}	0.06086957	0.3500000	0.17391304	1.298387	7

Figure 6.5: Rules from Criterion 2

	lhs		rhs	support	confidence	coverage	lift	count
[1]	{Bags, Formals}	=>	{Jackets}	0.02608696	0.7500000	0.03478261	4.312500	3
[2]	{Belts, Sneakers}	=>	{Dresses}	0.02608696	1.0000000	0.02608696	3.709677	3
[3]	{Dresses, Formals}	=>	{Sneakers}	0.03478261	0.6666667	0.05217391	3.650794	4
[4]	{Formals, Sneakers}	=>	{Jackets}	0.02608696	0.6000000	0.04347826	3.450000	3
[5]	{Nightwear}	=>	{Formals}	0.03478261	0.6666667	0.05217391	3.194444	4
[6]	{Dresses, Jackets}	=>	{Sneakers}	0.03478261	0.5714286	0.06086957	3.129252	4
[7]	{Flip flops}	=>	{Bags}	0.02608696	0.6000000	0.04347826	3.000000	3
[8]	{Formals, Sneakers}	=>	{Dresses}	0.03478261	0.8000000	0.04347826	2.967742	4
[9]	{Jackets, Sneakers}	=>	{Dresses}	0.03478261	0.8000000	0.04347826	2.967742	4
[10]	{Jackets, Sneakers}	=>	{Formals}	0.02608696	0.6000000	0.04347826	2.875000	3
[11]	{Dresses, Formals}	=>	{Jackets}	0.02608696	0.5000000	0.05217391	2.875000	3
[12]	{Nightwear}	=>	{Bags}	0.02608696	0.5000000	0.05217391	2.500000	3
[13]	{Sneakers}	=>	{Dresses}	0.09565217	0.5238095	0.18260870	1.943164	11
[14]	{Nightwear}	=>	{Dresses}	0.02608696	0.5000000	0.05217391	1.854839	3

Figure 6.6: Rules from Criterion 3

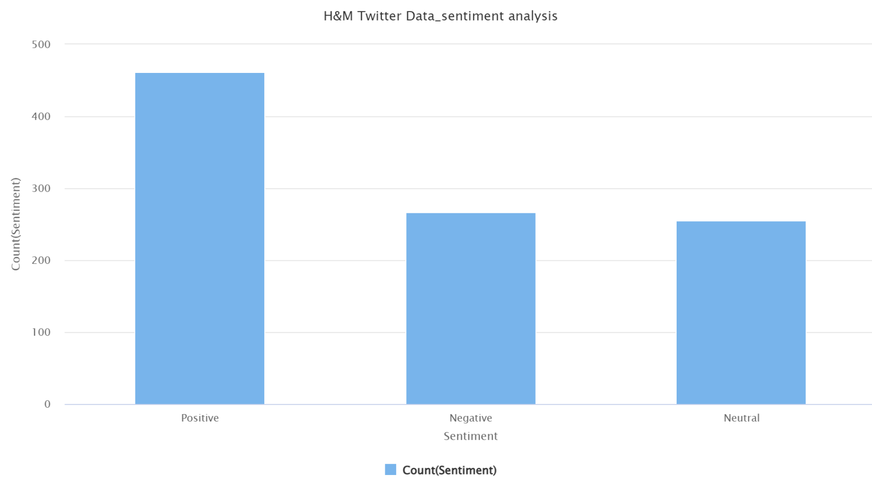
	lhs		rhs	support	confidence	coverage	lift	count
[1]	{Bags, Formals}	=>	{Jackets}	0.02608696	0.7500000	0.03478261	4.312500	3
[2]	{Belts, Sneakers}	=>	{Dresses}	0.02608696	1.0000000	0.02608696	3.709677	3
[3]	{Dresses, Formals}	=>	{Sneakers}	0.03478261	0.6666667	0.05217391	3.650794	4
[4]	{Formals, Sneakers}	=>	{Jackets}	0.02608696	0.6000000	0.04347826	3.450000	3
[5]	{Nightwear}	=>	{Formals}	0.03478261	0.6666667	0.05217391	3.194444	4
[6]	{Flip flops}	=>	{Bags}	0.02608696	0.6000000	0.04347826	3.000000	3
[7]	{Formals, Sneakers}	=>	{Dresses}	0.03478261	0.8000000	0.04347826	2.967742	4
[8]	{Jackets, Sneakers}	=>	{Dresses}	0.03478261	0.8000000	0.04347826	2.967742	4
[9]	{Jackets, Sneakers}	=>	{Formals}	0.02608696	0.6000000	0.04347826	2.875000	3

Figure 6.7: Rules from Criterion 4

	lhs		rhs	support	confidence	coverage	lift	count
[1]	{Dresses, Formals}	=>	{Sneakers}	0.03478261	0.6666667	0.05217391	3.650794	4
[2]	{Nightwear}	=>	{Formals}	0.03478261	0.6666667	0.05217391	3.194444	4
[3]	{Formals, Sneakers}	=>	{Dresses}	0.03478261	0.8000000	0.04347826	2.967742	4
[4]	{Jackets, Sneakers}	=>	{Dresses}	0.03478261	0.8000000	0.04347826	2.967742	4

Sentiment Analysis

Figure 6.8: Bar Chart for Sentiment Analysis



Based on Figure 6.8, it is observed that the overall comments that gathered from Twitter about H&M move towards a more positive sentiment which is 461 positive comments in total. Meanwhile, the negative sentiment is higher than the neutral sentiment which are 267 and 255 respectively. This suggests that while there is a positive shift in sentiment, there still exists a significant portion of negative sentiment expressed by Twitter users regarding H&M.

Discussion

Association Rules

Association rule is useful in identifying the relationships between the products as mentioned in the literature review. The Apriori algorithm can help H&M to effectively determine the associations between the products, enabling them to generate personalised recommendations for customers and efficiently manage inventory.

There are several rules with different criteria generated in Figure 6.4 to Figure 6.7. By comparing all the rules, the confidence of the rules from criteria 3 and 4 are at least 0.6667 which means the rules are quite reliable and able to represent the strong association among the subcategories. Then, the highest lift is found in the criteria 3. Therefore, the most interesting and meaningful rule is based on criteria 3 which sets the support threshold to 2% with 60% confidence. This decision is supported by the research conducted by Y. Kurnia [2]. If a minimum support value and a high minimum confidence value are given, the accuracy values generated will also be good because only items with strong associations will be raised in the results.

{Bags, Formals} => {Jackets} appears in about 2.61% of the total transaction. There is a 75% chance that jackets will be bought if bags and formals had been bought by the customer. Confidence of 0.75 suggests a strong association between bags, formals and jackets. Furthermore, there is 100% of confidence that the customers will buy belts and sneakers if they decide to buy dresses. Therefore, bags, formals and jackets are the best combination as there are strong relationships between each other. Same goes to the combination of belts, sneakers and dresses, it is suggested that these three items tend to occur together in the transaction. The companies can personalise the marketing and offer tailored promotions to the customers. H&M companies can provide product recommendations according to the rules through e-commerce platforms in order to improve customer loyalty and increase sales. This decision is supported by the research done by V. Umarani and M. Subathra [6], where the sales and revenue increase directly when making product placement according to the rules interpreted. For example, sending targeted messages to encourage or persuade them to come back and buy again. H&M can also pop-up advertisements on these groups of products to increase the exposure level of jackets and dresses.

In addition, sport shoes have the highest quantity sold from Figure 5.2 but dresses occur most frequently in the transaction shown in Figure 6.2. Therefore, dresses and sport shoes can be considered best-selling. H&M companies can put these two items at the most conspicuous place like near to the door. The companies can also optimise their inventory management by stocking up on dresses, sport shoes and avoid keeping stock on heels and flats.

Sentiment Analysis

Sentiment Analysis is used to detect customer satisfaction efficiently. Based on Figure 6.8, customers can be said to be satisfied with H&M products since overall feedback gathered from Twitter about H&M moves towards a more positive sentiment. On the other hand, from Figure 6.9, 카나 is considered as the influencer since its retweet count is the highest among all the comments. 카나 is suggested to be the most beneficial for a business to partner with and engage from their audience because 카나's tweet has a higher chance of being explored by the public.

Figure 7.1: The highest retweet count shown in RapidMiner

Row No.	Lower Case...	Created-At	From-User	Retweet... ↓
617	ateezofficial ...	May 14, 2023...	□□	23452

Figure 7.2: The name of the influencer shown in Excel file

challenges	2023-05-14 23:00:27	Jada Laroc	.0	16577628(	.1	Positive
ateezoffici	2023-05-14 23:00:26	카나	23452.0	16577628(	.0	Neutral
sterilising	2023-05-14 23:00:25	ivan idder	0	16577628(	-.4	Negative

Conclusion

In recent times, the fashion industry has experienced significant growth and witnessed the emergence of numerous new companies. This situation brings challenges to H&M because H&M needs to compete with many other companies to maintain its customer base and attract new customers. Hence, the aim of this study is to keep track of inventory by identifying the product that has the highest turnover rate, determining the association between the product and generating online shopping recommendations to customers. By achieving these aims, H&M can enhance its competitiveness and meet the evolving demands of the fashion market.

Based on the research and discussion conducted, the Association Rules technique was successfully implemented to identify the product that has the highest sales and determine the association between the product. According to the paper reviewed, Apriori Algorithm is the most common type of Association Rules in data mining that is used to find the frequent item sets efficiently and hence it is being applied in this study. Association rules that discovered well in providing information on the item combination rules to generate online shopping recommendation items to customers. Furthermore, Sentiment Analysis was applied with the purpose of analysing the reviews to obtain information about customer satisfaction.

From the association rules generated, the most appropriate criteria is the criteria with a minimum support value is 2% and a minimum confidence value is 60%. The item combination involved bags, formals and jackets as well as the combination included belts, sneakers and dresses are the best combination as there are strong relationships between items in particular combinations. Meanwhile, sports shoes that have the highest sales and dresses that are most often purchased simultaneously with other products are both considered as top-selling products. Based on the Sentiment Analysis result, the overall customer feedback tends to be positive indicating that the customer satisfaction level is high.

In conclusion, the rapid growth in the amount of data is due to the development of information technology. Data mining can be defined as a technique for extracting interesting information or hidden patterns in large amounts of data that support decision-making. As such, the application of detecting the sales pattern from a large database will be easier, which will make improvements to the marketing plan and strategy such as promotion, product bundling, recommendation and so on. Identifying the product that has the highest turnover rate can help in keeping track of inventory and beneficially improve the inventory system. All of these will result in better customer satisfaction and lead to economic growth for a company.

Reference:

1. Agarwal, S., & Damle, M. (2020). Sentiment Analysis to Evaluate Influencer Marketing: Exploring To Identify the Parameters of Influence. *Palarch's Journal Of Archaeology of Egypt/Egyptology, 17*(6), 4784-4800. ISSN 1567-214x

2. Grljević, O., & Bošnjak, Z. (2018). Sentiment analysis of customer data. *Strategic Management, 23*(3), 38-49. https://doi.org/10.5937/straman1803038g

3. *H&M mission and vision statement analysis.* (n.d.). Edrawsoft. https://www.edrawmind.com/article/hm-mission-and-vision-statement-analysis.html#:~:text=The%20H%26M%20mission%20statement%20is,you%20encounter%20on%20the%20streets

4. Kurnia, Y., Isharianto, Y., Giap, Y. C., Hermawan, A., & Riki, R. (2019). Study of application of data mining market basket analysis for knowing sales pattern (association of items) at the O! Fish restaurant using apriori algorithm. *Journal of Physics*, *1175*, 012047. https://doi.org/10.1088/1742-6596/1175/1/012047

5. Santoso, M. H. (2021). Application of Association Rule Method Using Apriori Algorithm to Find Sales Patterns Case Study of Indomaret Tanjung Anom. *Brilliance*, *1*(2), 54–66. https://doi.org/10.47709/brilliance.v1i2.1228

6. Ünvan, Y. A. (2020). Market basket analysis with association rules. *Communications in Statistics - Theory and Methods*, *50*(7), 1615-1628. https://doi.org/10.1080/03610926.2020.1716255

7. Umarani, V., Subathra, M. (2021). Investigation of KNN and Decision Tree Induction Modelin Predicting Customer Buying Pattern. *Proceedings of the First International Conference on Combinatorial and Optimization, ICCAP 2021, December 7-8 2021, Chennai, India.* https://doi.org/10.4108/eai.7-12-2021.2314593