H&M

Project 1 Technical Report

Data Driven Marketing Decisions

Section 001

ABSTRACT

Recommendation for H&M to improve their sales and profitability based on different customer segments

Team 2

Abhyas Ramadugu

Derik Boghosian

Janvi Ramavat

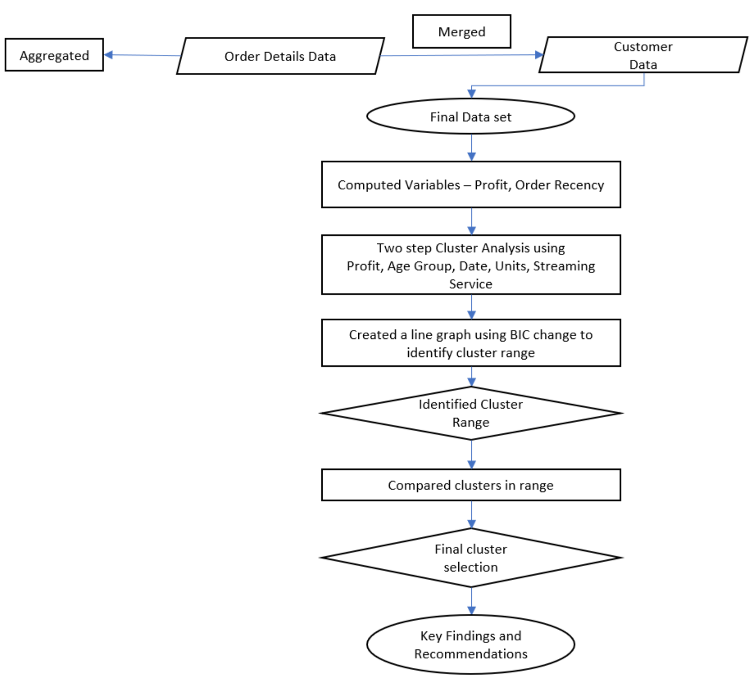
Mausam Shah

Contents

Introduction

H&M is a multi-channel clothing, accessories, and home company that combines fashion, quality, price, and sustainability. It is present internationally in 74 markets and offers clothing collections for women, men, teenagers, children, and babies, as well as home accessories and furnishings. The H&M Group has steadily increased its revenue and profits from 2017 to 2019. Losses were at a maximum in 2020, with a 20% loss compared to 2019. Although not at its peak capacity, the company has seen decent growth coming into 2022. Shareholders currently feel optimistic about the future. H&M clothing sales and order quantity among women and men are not evenly split down the middle. Compared to Zara, H&M focuses more on female consumers. Our data includes transactional and customer data which we then used to compute different recency, frequency, and monetary variables. We processed the data through aggregation and merging of two data sets. We performed cluster analysis on this data to solve the managerial problem of increasing sales and profitability. A detailed flow of the work can be found below.

Flowchart



Flowchart 1 – Steps for our overall analysis

Data

Overview and Source

The data provided to us by the company has two data sets. The data is from January 2019 to August 2021.

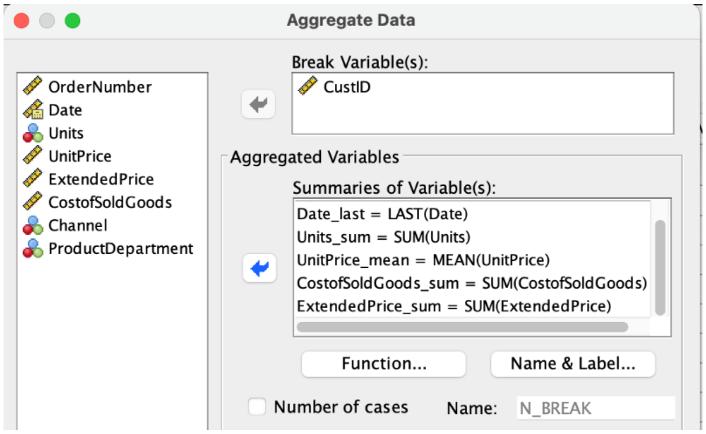
1. The first dataset contains 80,000+ rows of transactional data from 42,000 randomly selected US customers. This sample is a representative of all customers of the company. The variables in this dataset are: OrderNumber, CustID, Order_Date, Units, UnitPrice, Revenue, CostofSoldGoods, Channel. ProductDepartment
2. Additional customer data has the following variables: CustID, Zipcode, Gender, AgeGroup, Income, HHSize, YearsofEducation, HomeOwnership, PetOwner, HasCableSubscription, NumberStreamingSubs

Data Processing

Oftentimes, we frequent the stores that we have a liking to, whether it is online shopping or retail. Every time we make a purchase at the same store a new row is added in a sequential basis with the same customer ID. Our original transactional data was not useful for the type of analysis conducted in the study. And this is why we had to transform the data by aggregating order data by customer ID and merging additional customer information into the same dataset.

This study focuses on customer data. Hence to run a successful cluster analysis we aggregate the data in such a way that each customer has only one row.

- Import data from excel into SPSS
- Data>Aggregate



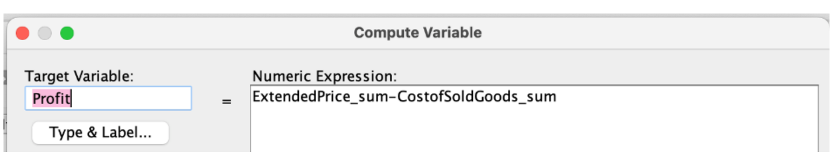
- The Break Variable is the variable that we need to identify of the new data file. In this case it will be Cust ID
- The rest will go to aggregated variables. Relevant aggregating functions were applied. For example, SUM for revenue and LAST for Date.

Screenshot 1 – SPSS Data Aggregation

Variables Computed

- To calculate Profit:
    - o Transform>Compute Variable
    - o Enter "Profit" under Target Variable
    - o To calculate profit, the numeric expression will be:

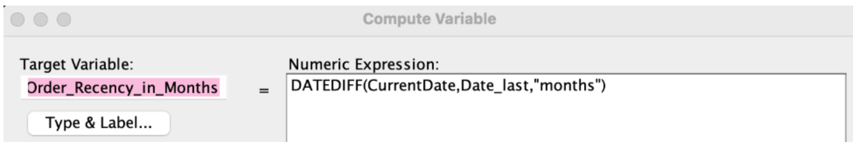
      ExtendedPrice_sum-CostofSoldGoods_sum



Screenshot 2 – SPSS Profit Variable

- To calculate recency (this computation will give us the number of months between the last two orders from a single customer)
    - o Change Variable Type of *Order data* into *"Date" mm/dd/yyyy*
    - o Create CurrentDate variable and set the variable type as "Date"

- Transform -> Compute variable -> Date Creation -> Date.mdy -> Click OK (In this study, CurrentDate has been set to 08/31/2021, the last order date in the data)
- Transform -> Compute Variable -> Date Arithmetic -> DateDiff(CurrentDate, Date_Last,"months")->Click OK

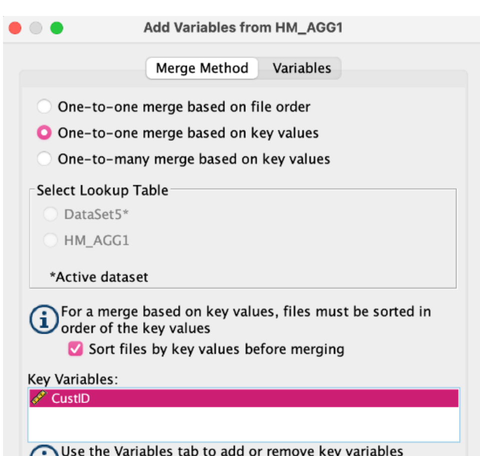


Screenshot 3 – SPSS Date Recency Variable

Merging the Data

The goal is to merge the two data files such that for each customer, we have one row, including both purchase data and demographics. We merged two data sets using the following steps:

- Open both data files in SPSS
- Sort both data files based on the variable that connects rows in both files. For example, CustID
- Delete any empty rows
- Go to the file that you perceive as primary, and you want the variables on that file to be listed first.
- Go to Data >> Merge Files >> Add Variables:
- Select the file that you want to merge with, then click continue



- Choose the items listed below and then click OK:
    - One-to-one merge based on key values
    - Sort files before merging

- The key variable that is common in both files, and you want to use to connect cases in two files (Customer Id)

Screenshot 4 – SPSS Merging Files

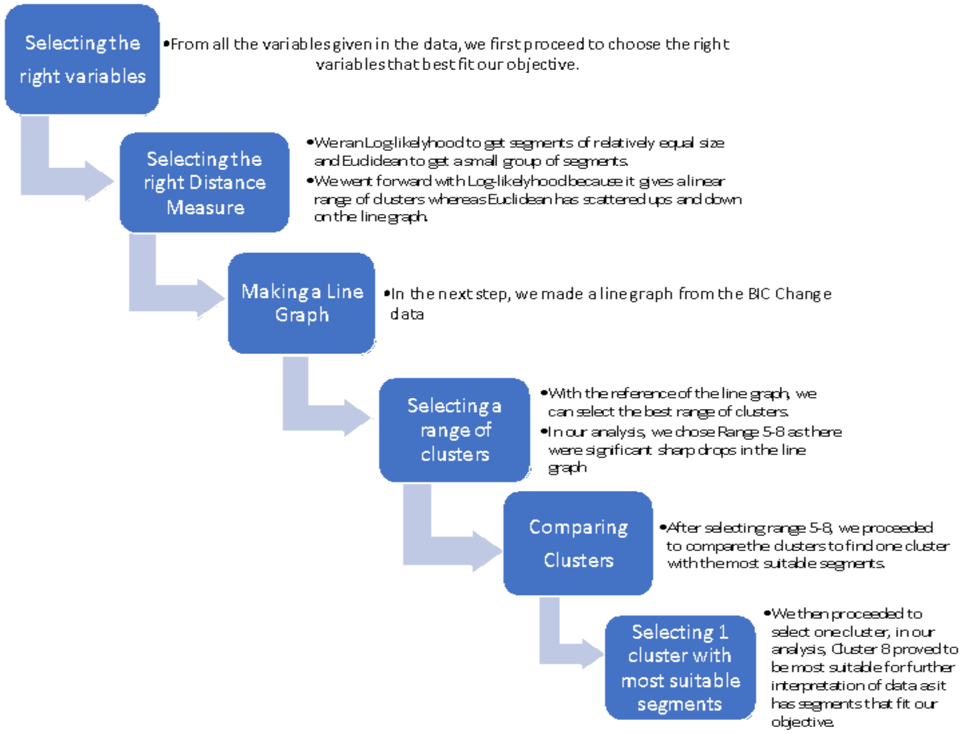
Variables used for our study

Variable Name	Data Type	Calculation (If Any)	Relevance to the study
Profit*	Numeric	ExtendedPrice - CostofSoldGoods	To study the current purchasing capacity
Order_Recency_in_Months*	Numeric	DateDiff(CurrentDate, OrderDate,"months")	Recency factor can help offer tiered customer service
AgeGroup	String	N/A	As a clothing company, following gender trends is vital.
UnitPrice	Numeric	N/A	Increasing items per order is key to increase earning potential.
ZipCode	Numeric	N/A	Based on geographic locations, retail experiences can be tailored.
NumberStreamingSubs	Numeric	N/A	Extensive exposure to pop culture can affect shopping behavior. Also helps to keep trendy designs to satisfy such customers.

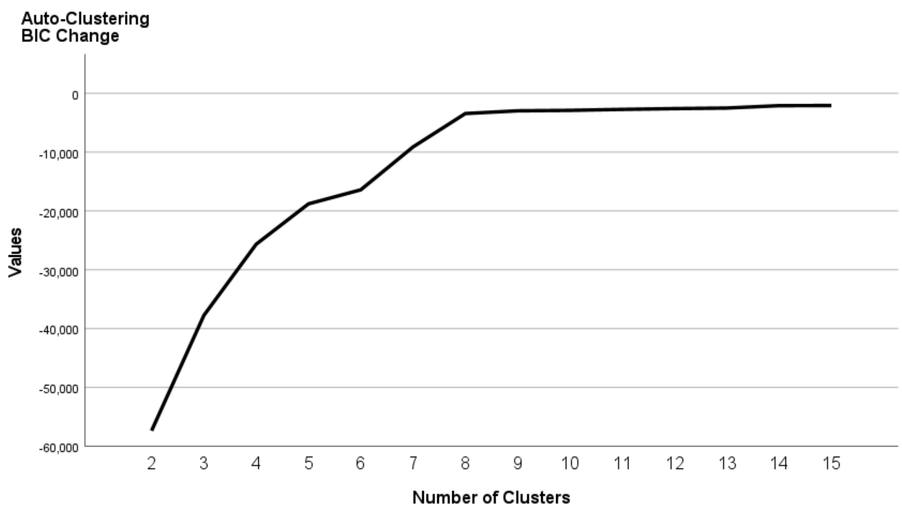
Table 1 – Variables used in our study

Data Analysis

Cluster Analysis



Graphic 1 – Cluster analysis flowchart



Using the variables above, we ran our cluster analysis and created the BIC Change graph.

Cluster Range Chosen- 6-8

Graph 1 – BIC Change graph

As shown in Graph 1, we ran Cluster Analysis using the variables mentioned in Table 1. We carried out Log-likelihood to run the analysis. After receiving the data, we made a line graph using BIC Change. As presented in Figure 1.1, we can see that there are sharp drops from ranges 6 to 8. Hence, we came to the conclusion that Cluster range 6-8 would best suitable for our analysis and further interpretation of data.

Cluster Group 6

Cluster Distribution

		N	% of Combined	% of Total
Cluster	1	6829	16.1%	16.1%
	2	4962	11.7%	11.7%
	3	6951	16.4%	16.4%
	4	6833	16.1%	16.1%
	5	7249	17.1%	17.1%
	6	9628	22.7%	22.7%
	Combined	42452	100.0%	100.0%
Excluded Cases		1		0.0%
Total		42453		100.0%

Table 2.0 – 6 clusters distribution

Centroids

		Unit_Price_Mean		Order_Recency_in_months		Profit		NumberStreamingSubs	
		Mean	Std. Deviation	Mean	Std. Deviation	Mean	Std. Deviation	Mean	Std. Deviation
Cluster	1	75.0299	41.39698	14.5811	11.25951	123.2819	155.16169	1.84	1.319
	2	72.7560	46.72721	16.5635	11.49092	53.7004	62.59342	1.68	1.274
	3	76.4095	42.69406	14.4878	20.69426	148.2612	257.80594	1.85	1.302
	4	76.7617	43.93302	14.5459	11.39410	124.3527	156.01350	1.82	1.310
	5	68.4484	42.69510	16.4761	11.57345	43.1670	56.88208	1.61	1.219
	6	72.1383	41.51721	15.7386	11.46170	76.5274	106.24869	1.69	1.239
	Combined	73.4891	43.01515	15.3780	13.42510	95.1273	153.97634	1.75	1.278

Table 2.1 – 6 cluster centroids based on variables

Cluster Group 7

Cluster Distribution

		N	% of Combined	% of Total
Cluster	1	6796	16.0%	16.0%
	2	4962	11.7%	11.7%
	3	6789	16.0%	16.0%
	4	247	0.6%	0.6%
	5	6781	16.0%	16.0%
	6	7249	17.1%	17.1%
	7	9628	22.7%	22.7%
	Combined	42452	100.0%	100.0%
Excluded Cases		1		0.0%
Total		42453		100.0%

Table 3.0 – 7 clusters distribution

Centroids

		Unit_Price_Mean		Order_Recency_in_months		Profit		NumberStreamingSubs	
		Mean	Std. Deviation	Mean	Std. Deviation	Mean	Std. Deviation	Mean	Std. Deviation
Cluster	1	75.0303	41.48029	14.6236	11.25993	119.4150	145.08096	1.83	1.319
	2	72.7560	46.72721	16.5635	11.49092	53.7004	62.59342	1.68	1.274
	3	76.2783	43.12130	14.5457	11.34890	117.8536	140.79073	1.83	1.302
	4	80.9453	17.88071	9.8097	92.55550	1235.2159	553.96160	2.52	1.133
	5	76.7234	44.05509	14.6166	11.39409	118.7741	142.69310	1.81	1.310
	6	68.4484	42.69510	16.4761	11.57345	43.1670	56.88208	1.61	1.219
	7	72.1383	41.51721	15.7386	11.46170	76.5274	106.24869	1.69	1.239
	Combined	73.4891	43.01515	15.3780	13.42510	95.1273	153.97634	1.75	1.278

Table 3.1 – 7 cluster centroids based on variables

Cluster Group 8

Cluster Distribution

		N	% of Combined	% of Total
Cluster	1	6796	16.0%	16.0%
	2	4962	11.7%	11.7%
	3	6789	16.0%	16.0%
	4	247	0.6%	0.6%
	5	6781	16.0%	16.0%
	6	7249	17.1%	17.1%
	7	3885	9.2%	9.2%
	8	5743	13.5%	13.5%
	Combined	42452	100.0%	100.0%
Excluded Cases		1		0.0%
Total		42453		100.0%

Table 4.0 – 8 clusters distribution

Centroids

		Unit_Price_Mean		Order_Recency_in_months		Profit		NumberStreamingSubs	
		Mean	Std. Deviation	Mean	Std. Deviation	Mean	Std. Deviation	Mean	Std. Deviation
Cluster	1	75.0303	41.48029	14.6236	11.25993	119.4150	145.08096	1.83	1.319
	2	72.7560	46.72721	16.5635	11.49092	53.7004	62.59342	1.68	1.274
	3	76.2783	43.12130	14.5457	11.34890	117.8536	140.79073	1.83	1.302
	4	80.9453	17.88071	9.8097	92.55550	1235.2159	553.96160	2.52	1.133
	5	76.7234	44.05509	14.6166	11.39409	118.7741	142.69310	1.81	1.310
	6	68.4484	42.69510	16.4761	11.57345	43.1670	56.88208	1.61	1.219
	7	60.6341	32.59451	17.3761	11.63172	45.6556	66.90174	.51	.559
	8	79.9206	44.95732	14.6309	11.21143	97.4115	121.72960	2.49	.886
	Combined	73.4891	43.01515	15.3780	13.42510	95.1273	153.97634	1.75	1.278

Table 4.1 – 8 cluster centroids based on variables

Cluster Group Comparison

We compared the means of each variable for all clusters in our range. The next step is to compare the Clusters from the Range and select the most suitable cluster. Upon comparing Cluster 6 and 7, we were able to rule out clusters that had no significant differences. We then proceeded to compare Cluster 7 with 8. At the end, we chose cluster 8 because it had a diverse group of segments which had the greatest potential to increase sales and profits for H&M.

Choosing the most suitable Cluster

			Unit_Price_Mean		Order_Recency_in_months		Profit		NumberStreamingSubs	
		N	Mean	Std. Deviation	Mean	Std. Deviation	Mean	Std. Deviation	Mean	Std. Deviation
Cluster	1	6796	75.03	41.4803	14.624	11.25993	119.415	145.081	1.83	1.319
	2	4962	72.756	46.7272	16.564	11.49092	53.7004	62.59342	1.68	1.274
	3	6789	76.278	43.1213	14.546	11.3489	117.8536	140.7907	1.83	1.302
	4	247	80.945	17.8807	9.8097	92.5555	1235.2159	553.9616	2.52	1.133
	5	6781	76.723	44.0551	14.617	11.39409	118.7741	142.6931	1.81	1.31
	6	7249	68.448	42.6951	16.476	11.57345	43.167	56.88208	1.61	1.219
	7	3885	60.634	32.5945	17.376	11.63172	45.6556	66.90174	0.51	0.559
	8	5743	79.921	44.9573	14.631	11.21143	97.4115	121.7296	2.49	0.886
	Combined	42452	73.489	43.0152	15.378	13.4251	95.1273	153.9763	1.75	1.278

Cluster group 8 proves to be the most suitable cluster among others because it aligns with our objective to find segments that have the highest potential to get profitable. Compared to other clusters in the range, Cluster 8 has segments that have high profits as well as segments that can be potentially profitable if a marketing strategy is applied. Hence, we went ahead with Cluster 8 for our analysis.

Key Findings

From our total eight cluster analysis, we have selected three clusters to further consider and develop implications for. The top cluster of importance is cluster number eight. This specific segment holds a large number of individuals that we can focus on for potential growth of sales and an increase of profits. Furthermore, the UnitPrice for this cluster is relatively high at $79.92.

Where we see an opportunity for growth is in their ordering recency. Currently this variable is at about 14.63 months, we hope to decrease the number of months between orders by spending more money on collaborations with film and television production companies and partnerships with celebrities. This becomes more important when we discuss the next variable of importance, the number of streaming subscriptions. The eighth cluster has the highest number of total subscriptions at 2.49, which poses an opportunity to expose more watchers to H&M's goods through targeted advertisements. This means that these consumers are constantly being exposed to advertisements and characters in content who could be wearing H&M merchandise. Lastly, cluster number eight has a profit variable coming in at $97.41, which is relatively average among all the clusters, we hope to see this rise through our managerial suggestions being put into effect.

Our second most integral cluster is number five. This cluster has over 6700 individuals (even more than cluster eight). Their unit price is only a few dollars less at $76.72. Their ordering recency is also similarly low like cluster eight's, at about 14.62 months. The profit with this segment is a bit higher at $118.77. Finally, this cluster also has a high average streaming subscription at 1.81. Since these customers are spending more than some others, we feel that providing them with special events, luxury exclusive product lines, and personal shoppers could help to increase potential spending, profits, and experiences.

Finally, we find that cluster four is also of great importance and interest. We do not see a huge opportunity or potential for growth or profit increases here. However, we see that they are integral to retain. Although the smallest cluster in number of individuals, they bring it the highest profits by far at over $1200. We must ensure that they continue to shop and remain our loyal customers. We would suggest doing so by making them feel valued. This could mean personalized birthday merchandise from an H&M line or free deliveries for life.