CSN 515: Data Mining & Warehousing
Recommendation System for Apparel Store

Group Details

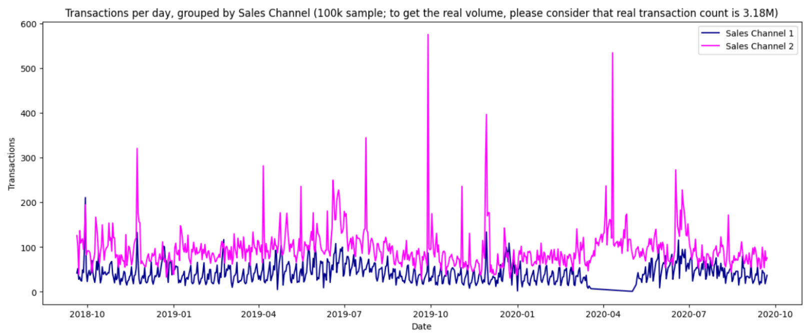
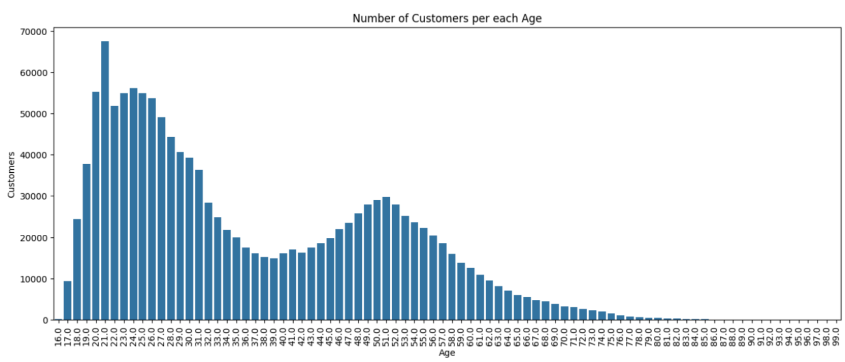
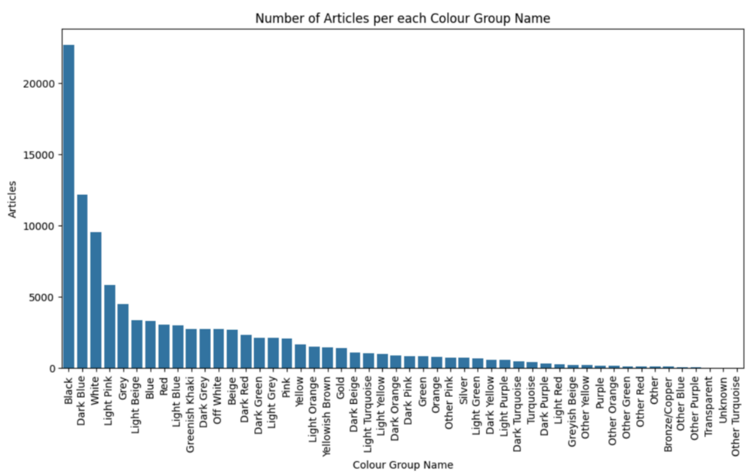
Group Number : 2

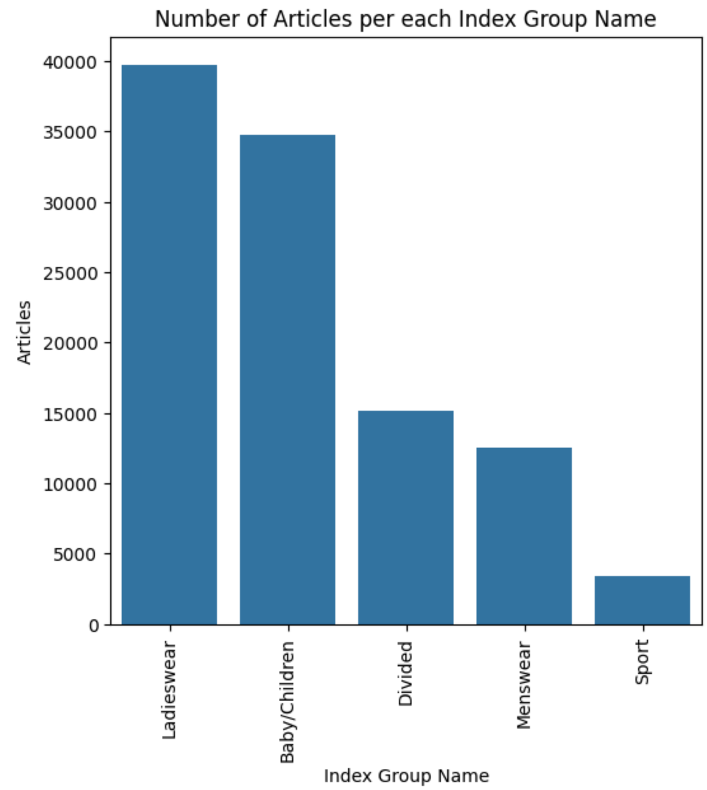
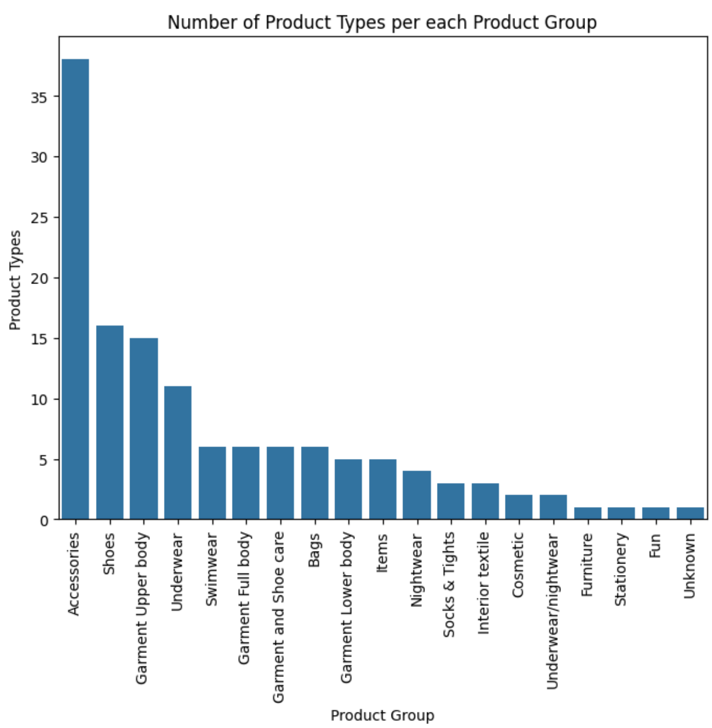
Group Members:			
Enrollment No.	**Name**	**Contribution**	**Page No.**
20114013	Anshul Gharde	Association rule using Apriori	6
20114035	Dighiya Sanidhya	Last week's most popular articles	5
20114068	Patare Abhishek	Content based Recommendation	7
20114079	Rahul Kurkure	Customer wise most popular articles	5
20114082	Prabhas Sagiraju	Cluster based Recommendation	10

Problem Statement:

P2 : Use data mining to make recommendations for items to buy from an apparel store.

Exploratory Data Analysis:

Figure	Description
	It shows the volume of sales over the 2 year period. We can see the sales have peaked during festive seasons of April(wedding season), October(Dussehra and diwali) and December(Christmas and new year).
	It describes the age distribution of the customers. As, H&M is a fashion brand, we can see that the majority of the customers are from the 20-30 year demographic.
	It describes the number of articles sold per each color group name. We can see that black colored clothes are the most popular by a significant margin.

<table>
<tr>
<td>
</td>
<td>It describes the number of articles per index group name. We can see that ladieswear leads the list followed closely by baby/children wear and then menswear.</td>
</tr>
<tr>
<td>
</td>
<td>It describes the number of product types per each product group. We can see that there are a wide variety of products in the accessories section ranging from spectacles to socks.</td>
</tr>
</table>

General Data-Preprocessing:

***Data cleaning*:**

Handling null values: There were only 1% data points with null values so we dropped them.

For content based reduction some articles in the product inventory did not have textual description so we didn't consider those items for content base reduction.

Data integration:

We had data about customer's articles and transactions so we have to integrate by performing joins.

Data/attribute reduction:

Many attributes in the original dataset were irrelevant so we have to reduce the subset of those attributes.

Data transformation:

First we had to convert the transaction date from string to relevant date time format.

For apriori association rule analysis article ids in the transaction didn't give nice results so we had to construct a new attribute/item by concatenating the product type and color.

Proposed Recommendation Methodology:

General Ideas

1. Last week's most popular articles

We are recommending the most popular item based on last week's transaction history of all the customers, by simply counting the most frequently occurring item (Article_id) from transactions.csv file.

This method can be used to recommend the most trending items in the apparel store, which is likely to be purchased by all the customers.

	article_id	prod_name	product_type_name	product_group_name	department_name	index_name	detail_desc
0	108775015	Strap top	Vest top	Garment Upper body	Jersey Basic	Ladieswear	Jersey top with narrow shoulder straps.
1	108775044	Strap top	Vest top	Garment Upper body	Jersey Basic	Ladieswear	Jersey top with narrow shoulder straps.
2	108775051	Strap top (1)	Vest top	Garment Upper body	Jersey Basic	Ladieswear	Jersey top with narrow shoulder straps.
6	111565001	20 den 1p Stockings	Underwear Tights	Socks & Tights	Tights basic	Lingeries/Tights	Semi shiny nylon stockings with a wide, reinfo...
7	111565003	20 den 1p Stockings	Socks	Socks & Tights	Tights basic	Lingeries/Tights	Semi shiny nylon stockings with a wide, reinfo...
8	111586001	Shape Up 30 den 1p Tights	Leggings/Tights	Garment Lower body	Tights basic	Lingeries/Tights	Tights with built-in support to lift the botto...
9	111593001	Support 40 den 1p Tights	Underwear Tights	Socks & Tights	Tights basic	Lingeries/Tights	Semi shiny tights that shape the tummy, thighs...
10	111609001	200 den 1p Tights	Underwear Tights	Socks & Tights	Tights basic	Lingeries/Tights	Opaque matt tights. 200 denier.
27	123173001	Control Top 50 den 1p Tights	Leggings/Tights	Garment Lower body	Tights basic	Lingeries/Tights	50 denier tights with reinforcement at the top...
33	129085001	Pirate Leggings (1)	Leggings/Tights	Garment Lower body	Jersey Basic	Ladieswear	3/4-length leggings in stretch jersey with an ...
34	129085026	Pirate Leggings (1)	Leggings/Tights	Garment Lower body	Jersey Basic	Ladieswear	3/4-length leggings in stretch jersey with an ...
35	129085027	Pirate Leggings (1)	Leggings/Tights	Garment Lower body	Jersey Basic	Ladieswear	3/4-length leggings in stretch jersey with an ...
36	130035001	Black Umbrella	Umbrella	Items	Other items	Ladies Accessories	Umbrella with a telescopic handle and matching...
39	144993001	Mama 100 den 1p Tights	Underwear Tights	Socks & Tights	Tights basic	Lingeries/Tights	Matt tights with an elasticated waist and extr...
40	145872001	Dorian l/s basic	Sweater	Garment Upper body	Men Sport Woven	Sport	Long-sleeved sports top in fast-drying, breath...

2. Customer wise most popular articles

We are recommending items to a customer based on his/her previous purchases by grouping the transactions based on customer id and finding the most frequently occurring item (Article_id) in those groups.

This method can provide recommendations based on the customer's previous trend (last 2 years) assuming that the customer is likely to buy an item which he/she is buying repeatedly.

Top 10 recommendations for customer with customer_id = '00000dbacae5abe5e23885899a1fa44253a17956c6d1c3d25f88aa139fdfc657'

	customer_id	product_code_x	detail_desc
31140481	00000dbacae5abe5e23885899a1fa44253a17956c6d1c3...	568601	Fitted jacket in woven fabric with notch lapel...
17074660	00000dbacae5abe5e23885899a1fa44253a17956c6d1c3...	797065	Fitted jacket in woven fabric with notch lapel...
4212358	00000dbacae5abe5e23885899a1fa44253a17956c6d1c3...	625548	Padded jacket with a detachable hood, stand-up...
4212359	00000dbacae5abe5e23885899a1fa44253a17956c6d1c3...	176209	Short, padded jacket with a jersey-lined hood ...
4212360	00000dbacae5abe5e23885899a1fa44253a17956c6d1c3...	627759	Padded parka in woven fabric with a soft, brus...
16728347	00000dbacae5abe5e23885899a1fa44253a17956c6d1c3...	656719	Tailored trousers in a stretch weave with two ...
14479579	00000dbacae5abe5e23885899a1fa44253a17956c6d1c3...	745232	Short 5-pocket skirt in washed, stretch denim ...
14479578	00000dbacae5abe5e23885899a1fa44253a17956c6d1c3...	607642	Top in a crêpe weave with a V-shaped opening a...
19160879	00000dbacae5abe5e23885899a1fa44253a17956c6d1c3...	785186	Long-sleeved top in soft jersey with a draped ...
19357366	00000dbacae5abe5e23885899a1fa44253a17956c6d1c3...	812683	Long dress in an airy weave with a round, gath...

3. Making Association rules using Apriori algorithm

We are recommending items to the customers that are related to the items purchased by the customer previously. The items that are closely associated to each other are found by apriori algorithm.We are only considering the last one month's transaction because apriori algorithm is a memory intensive process.

Analogous to apriori terminology, Our transactions are the list of product names ordered by a customer in a day and items are the product names itself. So, we transformed the transaction table by grouping it on (customer_id, date) and listed the product names corresponding to them in a different column. Now, we have transformed this transaction table into a 1 hot encoded format where each cell value is a boolean value specifying whether that item is present in that transaction or not.

Once the 1 hot encoded format is ready, we will apply apriori algorithm to get all frequent itemsets that are having min_support. Then, we find association rules that are having min_confidence.

When we performed apriori algorithm on product names, we haven't found enough association rules as it was a much more specific attribute. So we have shifted to a more generic attribute pair (product type name + color).

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric	antecedents_length	consequents_length
0	(Black-Trousers)	(Black-Sweater)	0.119929	0.067071	0.012521	0.104405	1.556618	0.004477	1.041685	0.406310	1	1
1	(Black-Sweater)	(Black-Trousers)	0.067071	0.119929	0.012521	0.186684	1.556618	0.004477	1.082077	0.383290	1	1
2	(Black-Trousers)	(Black-Top)	0.119929	0.066872	0.010177	0.084862	1.269032	0.002158	1.019659	0.240887	1	1
3	(Black-Top)	(Black-Trousers)	0.066872	0.119929	0.010177	0.152194	1.269032	0.002158	1.038057	0.227190	1	1
4	(Black-Trousers)	(Blue-Trousers)	0.119929	0.106008	0.027003	0.225156	2.123943	0.014289	1.153770	0.601290	1	1
5	(Blue-Trousers)	(Black-Trousers)	0.106008	0.119929	0.027003	0.254723	2.123943	0.014289	1.180864	0.591927	1	1
6	(Black-Trousers)	(Grey-Trousers)	0.119929	0.043380	0.012285	0.102437	2.361364	0.007083	1.065796	0.655079	1	1
7	(Grey-Trousers)	(Black-Trousers)	0.043380	0.119929	0.012285	0.283197	2.361364	0.007083	1.227772	0.602660	1	1
8	(Blue-Trousers)	(Grey-Trousers)	0.106008	0.043380	0.011864	0.111912	2.579790	0.007265	1.077168	0.684986	1	1
9	(Grey-Trousers)	(Blue-Trousers)	0.043380	0.106008	0.011864	0.273480	2.579790	0.007265	1.230511	0.640141	1	1

Novelty and Innovation :

4. Content based recommendation :

Content-based recommendation suggests items to users based on the intrinsic features and characteristics of the items themselves. It relies on item representation, user profiles, and feature extraction, often using techniques like TF-IDF for textual content.

Recommendations are generated by calculating similarity between the user profile and item vectors in a feature space, making it suitable for new users and providing transparent, feature-driven suggestions.

The Preprocessing steps involved in this are described below step by step.

1. Remove the articles that do not have detailed description
2. Concatenate all text columns to create a document collection
3. Preprocessing the concatenated text
   a. Lower casing
   b. Removing punctuation
   c. Tokenization : splits the processed text into a list of individual words (tokens) based on whitespace. This step is essential for further analysis and manipulation of the text at the word level
   d. Stopword removal : removes common English stop words from the list of tokens. Stopwords are frequently used words (e.g., "the," "and," "is") that are often excluded in text analysis as they do not contribute much to the meaning of the text.

e. Joining Tokens : Concatenating tokens generated above.

After the preprocessing is done, we move on to the process of TF-IDF vectorization and finding content-based recommendation using cosine similarity

TF-IDF Vectorization:

1. Term Frequency-Inverse Document Frequency (TF-IDF):
    a. TF-IDF is a numerical statistic that reflects the importance of a word in a document relative to a collection of documents (corpus).
    b. It consists of two components: Term Frequency (TF) measures how often a term appears in a document, and Inverse Document Frequency (IDF) emphasizes the rarity of the term across the entire corpus.
2. Term Frequency (TF):
    a. TF is the ratio of the number of times a term appears in a document to the total number of terms in that document.
    b. Formula:

$$TF(t, d) = \frac{Number\ of\ times\ term\ t\ appears\ in\ document\ d}{Total\ number\ of\ terms\ in\ document\ d}$$

3. Inverse Document Frequency (IDF):
    a. IDF is the logarithmically scaled inverse fraction of the documents that contain the term.
    b. Formula:

$$IDF(t, D) = log(\frac{Total\ number\ of\ documents\ in\ corpus\ \ D}{Number\ of\ documents\ containing\ term\ \ t + 1})$$

    c. Adding 1 in the denominator prevents division by zero.
4. TF-IDF Calculation:
    a. The TF-IDF score for a term in a document is calculated by multiplying its TF and IDF scores:

$$TF\text{-}IDF(t, d, D) = TF(t, d) \times IDF(t, D)$$

5. Vectorization:

a. Each document in the corpus is represented as a vector in a high-dimensional space, where each dimension corresponds to a unique term.
b. The TF-IDF scores for terms in a document form the components of the vector.

Content-Based Recommendation Using Cosine Similarity:

1. *Cosine Similarity*:
   a. Cosine similarity measures the cosine of the angle between two non-zero vectors in an inner product space.
   b. For document vectors, cosine similarity is used to quantify the similarity between two documents in the TF-IDF vector space.
2. *Calculating Cosine Similarity*:
   a. Given two document vectors A and B, the cosine similarity

   $$cosine\_similarity(A, B) = \frac{A.B}{||A||.||B||}$$

   b. A.B is the dot product of the vectors, and ||A|| and ||B|| are their respective Euclidean norms.
3. *Interpretation of Cosine Similarity*:
   a. Cosine similarity ranges from -1 (completely dissimilar) to 1 (identical).
   b. Higher cosine similarity indicates greater similarity between documents.
4. *Recommendation Process*:
   a. To recommend items, the system calculates cosine similarity between the vector representation of the user's profile and the vectors of available items.
   b. Items with higher cosine similarity scores are recommended as they are deemed more similar to the user's preferences.

By combining TF-IDF vectorization with cosine similarity, content-based recommendation systems can effectively match user preferences with relevant items in a corpus.

Top 6 articles similar to article_id '584200005'

```
article_id                           584200005
prod_name            Bianca off shoulder blouse
product_type_name                       Blouse
Name: 19282, dtype: object
```

	article_id	prod_name	product_type_name	product_group_name	department_name	index_name	detail_desc
19282	584200005	Bianca off shoulder blouse	Blouse	Garment Upper body	Tops Woven	Divided	Off-the-shoulder blouse in a cotton weave with...
19283	584200006	Bianca off shoulder blouse	Blouse	Garment Upper body	Tops Woven	Divided	Off-the-shoulder blouse in a cotton weave with...
93610	860646001	VIOLET LS BLOUSE	Blouse	Garment Upper body	Tops Woven	Divided	Cold shoulder blouse in a crêpe weave with nar...
55871	713699004	MOLLY OFF SHOULDER	Blouse	Garment Upper body	Tops Woven	Divided	Short, off-the-shoulder blouse in a crêpe weav...
55870	713699002	MOLLY OFF SHOULDER	Blouse	Garment Upper body	Tops Woven	Divided	Short, off-the-shoulder blouse in a crêpe weav...
55872	713699005	MOLLY OFF SHOULDER	Blouse	Garment Upper body	Tops Woven	Divided	Short, off-the-shoulder blouse in a crêpe weav...

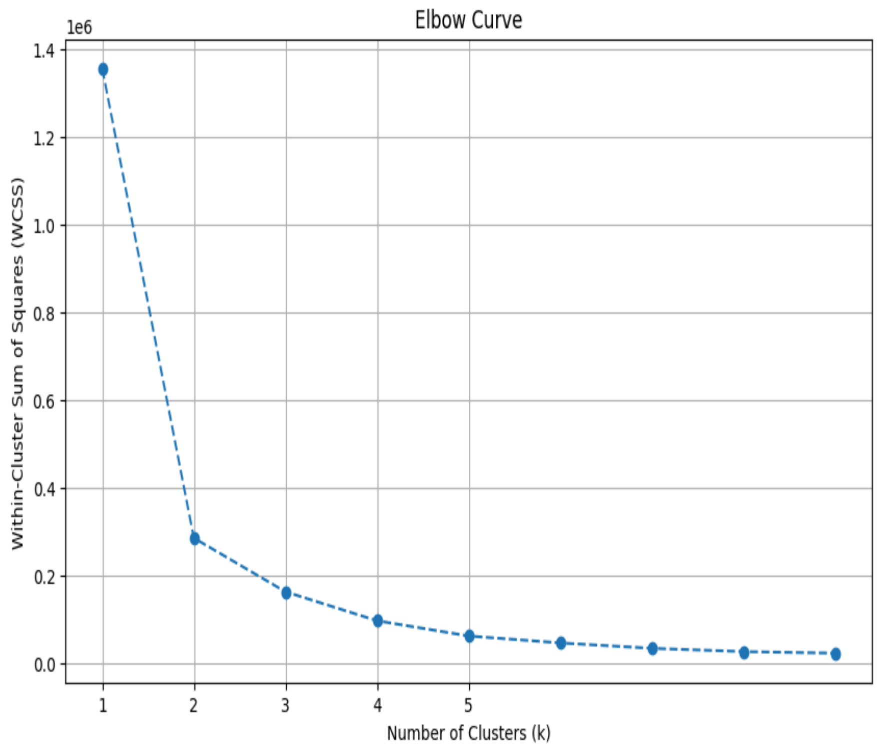
5. Clustering based on age group

Age-based clustering in an apparel store recommendation system is effective because individuals within similar age groups often share common fashion preferences, lifestyle choices, and clothing needs.

By grouping customers based on age, the system can tailor recommendations to specific age-related trends and styles, enhancing the likelihood of providing relevant and appealing apparel suggestions for each cluster. This approach leverages the correlation between age and fashion preferences to deliver more personalized and targeted recommendations to diverse customer segments.

Steps :

1. Extracting the relevant features (custome_id, age) from table customer.csv
2. Some rows did not have the age field but their proportion was very less (about 1%), so we dropped those rows
3. Then we applied k-means with values of k ranging from 1 to 9
4. Then we plotted Within-Cluster Sum of Squares (WCSS) for each k



From here we can see after k = 4 there is not much change Within-Cluster Sum of Squares (WCSS) therefore we are choosing k = 4 for clustering.

Correspond cluster centers are [59.56582223], [31.17848297], [46.42478339], [22.23711723]]

These cluster centers represent the clusters with centroid age groups as 22, 31, 46, 59.

Now we applied the initial three recommendation methods on each cluster separately.

Association rules for the cluster whose center is 60

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric	antecedents_length	consequents_length
0	(Black-Trousers)	(Black-Sweater)	0.120085	0.062607	0.010324	0.085975	1.373236	0.002806	1.025565	0.308886	1	1
1	(Black-Sweater)	(Black-Trousers)	0.062607	0.120085	0.010324	0.164904	1.373236	0.002806	1.053670	0.289946	1	1
2	(Black-Trousers)	(Blue-Trousers)	0.120085	0.087979	0.020088	0.167285	1.901424	0.009523	1.095238	0.538777	1	1
3	(Blue-Trousers)	(Black-Trousers)	0.087979	0.120085	0.020088	0.228332	1.901424	0.009523	1.140277	0.519811	1	1
4	(Black-Trousers)	(Grey-Trousers)	0.120085	0.040877	0.010515	0.087565	2.142164	0.005607	1.051169	0.605947	1	1
5	(Grey-Trousers)	(Black-Trousers)	0.040877	0.120085	0.010515	0.257241	2.142164	0.005607	1.184658	0.555906	1	1

Association rules for the cluster whose center is 46

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric	antecedents_length	consequents_length
0	(Black-Trousers)	(Black-Sweater)	0.136269	0.072961	0.014753	0.108264	1.483861	0.004811	1.039589	0.377528	1	1
1	(Black-Sweater)	(Black-Trousers)	0.072961	0.136269	0.014753	0.202205	1.483861	0.004811	1.082647	0.351746	1	1
2	(Blue-Trousers)	(Black-Sweater)	0.121337	0.072961	0.010476	0.086336	1.183318	0.001623	1.014639	0.176312	1	1
3	(Black-Sweater)	(Blue-Trousers)	0.072961	0.121337	0.010476	0.143580	1.183318	0.001623	1.025972	0.167111	1	1
4	(Black-Trousers)	(Black-Top)	0.136269	0.068324	0.011649	0.085486	1.251188	0.002339	1.018766	0.232433	1	1
5	(Black-Top)	(Black-Trousers)	0.068324	0.136269	0.011649	0.170499	1.251188	0.002339	1.041265	0.215482	1	1
6	(Black-Trousers)	(Blue-Trousers)	0.136269	0.121337	0.033272	0.244167	2.012309	0.016738	1.162509	0.582425	1	1
7	(Blue-Trousers)	(Black-Trousers)	0.121337	0.136269	0.033272	0.274216	2.012309	0.016738	1.190066	0.572527	1	1
8	(Black-Trousers)	(Grey-Trousers)	0.136269	0.053675	0.016258	0.119306	2.222739	0.008943	1.074521	0.636894	1	1
9	(Grey-Trousers)	(Black-Trousers)	0.053675	0.136269	0.016258	0.302891	2.222739	0.008943	1.239019	0.581306	1	1
10	(Blue-Trousers)	(Grey-Trousers)	0.121337	0.053675	0.016362	0.134846	2.512274	0.009849	1.093823	0.685079	1	1
11	(Grey-Trousers)	(Blue-Trousers)	0.053675	0.121337	0.016362	0.304831	2.512274	0.009849	1.263956	0.636097	1	1

Association rules for the cluster whose center is 31

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric	antecedents_length	consequents_length
0	(Black-Bra)	(Black-Underwear bottom)	0.045228	0.050377	0.010755	0.237802	4.720398	0.008477	1.245900	0.825489	1	1
1	(Black-Underwear bottom)	(Black-Bra)	0.050377	0.045228	0.010755	0.213495	4.720398	0.008477	1.213943	0.829965	1	1
2	(Black-Trousers)	(Black-Dress)	0.117953	0.090944	0.012490	0.105889	1.164323	0.001763	1.016714	0.160005	1	1
3	(Black-Dress)	(Black-Trousers)	0.090944	0.117953	0.012490	0.137335	1.164323	0.001763	1.022468	0.155251	1	1
4	(Black-Leggings/Tights)	(Black-Top)	0.069937	0.070572	0.012103	0.173051	2.452131	0.007167	1.123925	0.636722	1	1
5	(Black-Top)	(Black-Leggings/Tights)	0.070572	0.069937	0.012103	0.171494	2.452131	0.007167	1.122579	0.637157	1	1
6	(Black-Trousers)	(Black-Leggings/Tights)	0.117953	0.069937	0.010275	0.087114	1.245602	0.002026	1.018816	0.223543	1	1
7	(Black-Leggings/Tights)	(Black-Trousers)	0.069937	0.117953	0.010275	0.146922	1.245602	0.002026	1.033959	0.212002	1	1
8	(Black-Trousers)	(Black-Sweater)	0.117953	0.063696	0.012800	0.108514	1.703634	0.005286	1.050274	0.468251	1	1
9	(Black-Sweater)	(Black-Trousers)	0.063696	0.117953	0.012800	0.200948	1.703634	0.005286	1.103868	0.441117	1	1
10	(White-Sweater)	(Black-Sweater)	0.047892	0.063696	0.010570	0.220695	3.464829	0.007519	1.201461	0.747169	1	1
11	(Black-Sweater)	(White-Sweater)	0.063696	0.047892	0.010570	0.165937	3.464829	0.007519	1.141531	0.759780	1	1
12	(Black-T-shirt)	(White-T-shirt)	0.057106	0.044493	0.011522	0.201763	4.534745	0.008981	1.197022	0.826690	1	1
13	(White-T-shirt)	(Black-T-shirt)	0.044493	0.057106	0.011522	0.258963	4.534745	0.008981	1.272397	0.815776	1	1
14	(Black-Trousers)	(Black-Top)	0.117953	0.070572	0.011290	0.095713	1.356253	0.002965	1.027802	0.297801	1	1
15	(Black-Top)	(Black-Trousers)	0.070572	0.117953	0.011290	0.159974	1.356253	0.002965	1.050023	0.282619	1	1
16	(Black-Trousers)	(Blue-Trousers)	0.117953	0.103465	0.029223	0.247752	2.394543	0.017019	1.191807	0.660264	1	1
17	(Blue-Trousers)	(Black-Trousers)	0.103465	0.117953	0.029223	0.282443	2.394543	0.017019	1.229236	0.649594	1	1
18	(Black-Trousers)	(Grey-Trousers)	0.117953	0.044044	0.013040	0.110549	2.510005	0.007845	1.074772	0.682043	1	1
19	(Grey-Trousers)	(Black-Trousers)	0.044044	0.117953	0.013040	0.296062	2.510005	0.007845	1.253018	0.629312	1	1
20	(Blue-Trousers)	(Grey-Trousers)	0.103465	0.044044	0.012443	0.120266	2.730627	0.007886	1.086643	0.706926	1	1
21	(Grey-Trousers)	(Blue-Trousers)	0.044044	0.103465	0.012443	0.282525	2.730627	0.007886	1.249569	0.662984	1	1

Association rules for the cluster whose center is 22

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric	antecedents_length	consequents_length
0	(Black-Leggings/Tights)	(Black-Top)	0.049739	0.067823	0.010517	0.211436	3.117439	0.007143	1.182118	0.714776	1	1
1	(Black-Top)	(Black-Leggings/Tights)	0.067823	0.049739	0.010517	0.155060	3.117439	0.007143	1.124648	0.728643	1	1
2	(Black-Trousers)	(Black-Sweater)	0.111847	0.068195	0.012014	0.107410	1.575046	0.004386	1.043934	0.411076	1	1
3	(Black-Sweater)	(Black-Trousers)	0.068195	0.111847	0.012014	0.176165	1.575046	0.004386	1.078071	0.391818	1	1
4	(White-Sweater)	(Black-Sweater)	0.057673	0.068195	0.011265	0.195328	2.864264	0.007332	1.157994	0.690705	1	1
5	(Black-Sweater)	(White-Sweater)	0.068195	0.057673	0.011265	0.165190	2.864264	0.007332	1.128792	0.698505	1	1
6	(Black-Trousers)	(Blue-Trousers)	0.111847	0.106691	0.024787	0.221611	2.077120	0.012853	1.147638	0.583868	1	1
7	(Blue-Trousers)	(Black-Trousers)	0.106691	0.111847	0.024787	0.232320	2.077120	0.012853	1.156931	0.580498	1	1
8	(Black-Trousers)	(Grey-Trousers)	0.111847	0.038064	0.010217	0.091351	2.399918	0.005960	1.058644	0.656778	1	1
9	(Grey-Trousers)	(Black-Trousers)	0.038064	0.111847	0.010217	0.268424	2.399918	0.005960	1.214027	0.606401	1	1
10	(Blue-Trousers)	(Grey-Trousers)	0.106691	0.038064	0.010312	0.096648	2.539099	0.006250	1.064852	0.678556	1	1
11	(Grey-Trousers)	(Blue-Trousers)	0.038064	0.106691	0.010312	0.270900	2.539099	0.006250	1.225221	0.630145	1	1

Conclusion:

We explored five different recommendation systems tailored to the dataset from H&M, an apparel store. Through a combination of traditional and innovative approaches, we aimed to enhance the customer experience by providing personalized recommendations and insights into consumer preferences.

We began with two straightforward yet effective methods: Last week's most popular articles and Customer-wise most popular articles. By analyzing recent trends and individual customer preferences, we were able to offer recommendations that reflect both current market demand and personalized interests.

Moving beyond simple popularity-based recommendations, we delved into the method of Association Rules using the Apriori algorithm. By identifying frequent itemsets and deriving meaningful associations between articles, we uncovered hidden patterns in customer purchasing behavior, enabling targeted cross-selling and upselling strategies.

Further, as a novelty and innovation, we introduced two methods of recommendations. First, a Content-based recommendation system, leveraging article attributes such as category, color, size and description to suggest items similar to those a customer has shown interest in. This method enhances discoverability and encourages exploration of H&M's diverse product range.

Secondly, we explored Clustering based on age group, recognizing that demographic segmentation can provide valuable insights into consumer behavior. By grouping customers based on age and analyzing their preferences within each cluster, we gained a deeper understanding of age-specific trends and preferences, informing targeted marketing campaigns and product recommendations.