Market Basket Analysis & Recommendation System

FiILZAHANTI NUHA RAMADHANI

Overview

The H&M Group is a brand and business family with 53 online markets and approximately 4,850 stores. H&M's online store provides customers with a wide range of products to choose from.

Product recommendations are essential for improving the shopping experience. More importantly, assisting customers in making the best decisions has a positive impact on sustainability because it reduces returns and thus reduces transportation emissions.

Outline

01
OVERVIEW

03
EXPLORATORY DATA
ANALYSIS

05
RECOMMENDATION
SYSTEM

02
DATA

04
MARKET BASKET
ANALYSIS

06
RECOMMENDATIONS

DATA

ARTICLES

	0	1
article_id	108775015	108775044
product_code	108775	108775
prod_name	Strap top	Strap top
product_type_no	253	253
product_type_name	Vest top	Vest top
product_group_name	Garment Upper body	Garment Upper body
graphical_appearance_no	1010016	1010016
graphical_appearance_name	Solid	Solid
colour_group_code	9	10
colour_group_name	Black	White
perceived_colour_value_id	4	3
perceived_colour_value_name	Dark	Light
perceived_colour_master_id	5	9
perceived_colour_master_name	Black	White
department_no	1676	1676
department_name	Jersey Basic	Jersey Basic
index_code	A	A
index_name	Ladieswear	Ladieswear
index_group_no	1	1
index_group_name	Ladieswear	Ladieswear
section_no	16	16
section_name	Womens Everyday Basics	Womens Everyday Basics
garment_group_no	1002	1002
garment_group_name	Jersey Basic	Jersey Basic
detail_desc	Jersey top with narrow shoulder straps.	Jersey top with narrow shoulder straps.

CUSTOMERS

	customer_id	FN	Active	club_member_status	fashion_news_frequency	age	postal_code
0	00000dbacae5abe5e23885899a1fa44253a17956c6d1c3...	NaN	NaN	ACTIVE	NONE	49.0	52043ee2162cf5aa7ee79974281641c6f11a68d276429a...
1	0000423b00ade91418cceaf3b26c6af3dd342b51fd051e...	NaN	NaN	ACTIVE	NONE	25.0	2973abc54daa8a5f8ccfe9362140c63247c5eee03f1d93...
2	000058a12d5b43e67d225668fa1f8d618c13dc232df0ca...	NaN	NaN	ACTIVE	NONE	24.0	64f17e6a330a85798e4998f62d0930d14db8db1c054af6...
3	00005ca1c9ed5f5146b52ac8639a40ca9d57aeff4d1bd2...	NaN	NaN	ACTIVE	NONE	54.0	5d36574f52495e81f019b680c843c443bd343d5ca5b1c2...

TRANSACTION_TRAIN

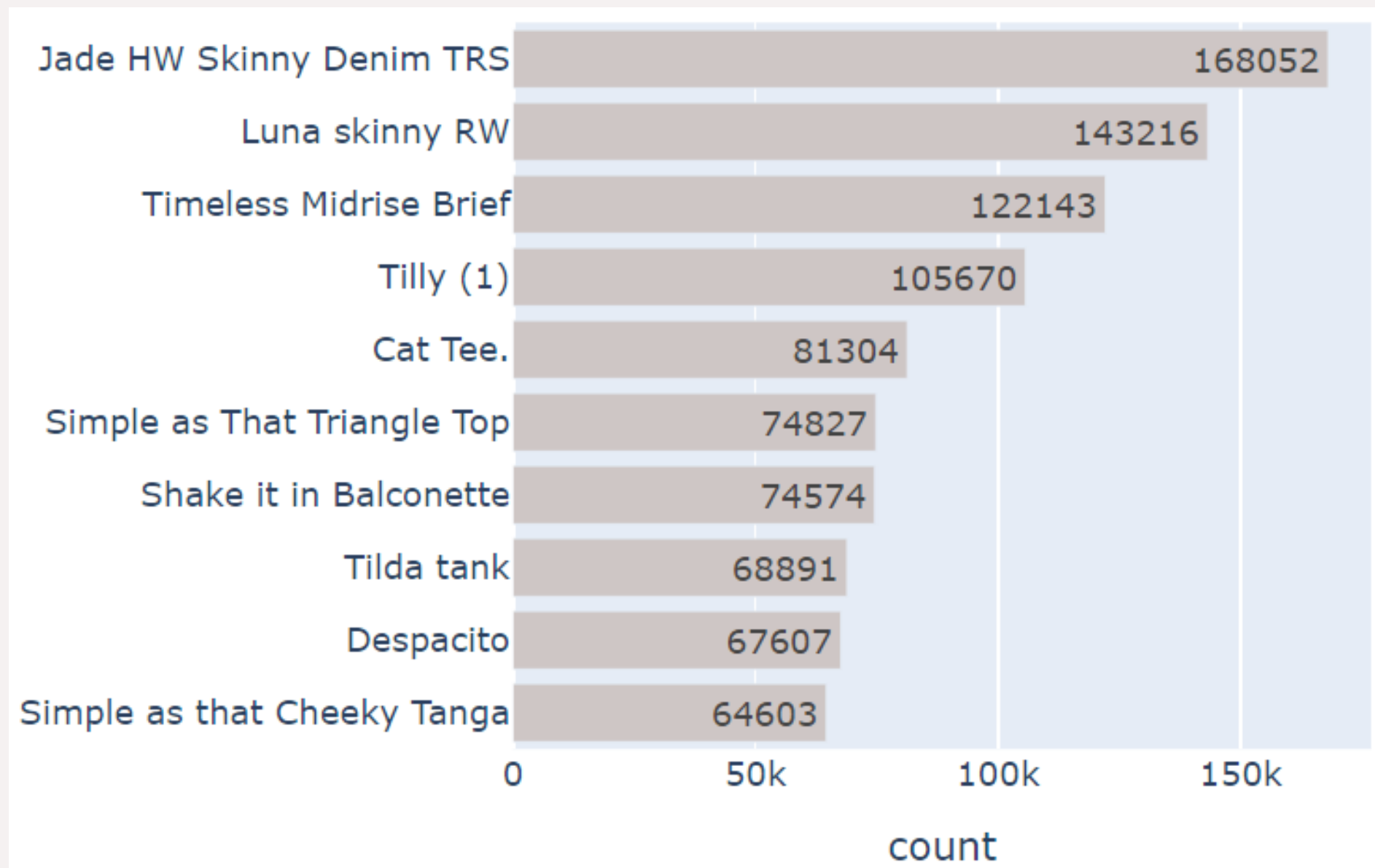|	t_dat	customer_id	article_id	price	sales_channel_id
0	2018-09-20	000058a12d5b43e67d225668fa1f8d618c13dc232df0ca...	663713001	0.050831	2
1	2018-09-20	000058a12d5b43e67d225668fa1f8d618c13dc232df0ca...	541518023	0.030492	2
2	2018-09-20	00007d2de826758b65a93dd24ce629ed66842531df6699...	505221004	0.015237	2
3	2018-09-20	00007d2de826758b65a93dd24ce629ed66842531df6699...	685687003	0.016932	2

ARTICLES = 1054 ROWS
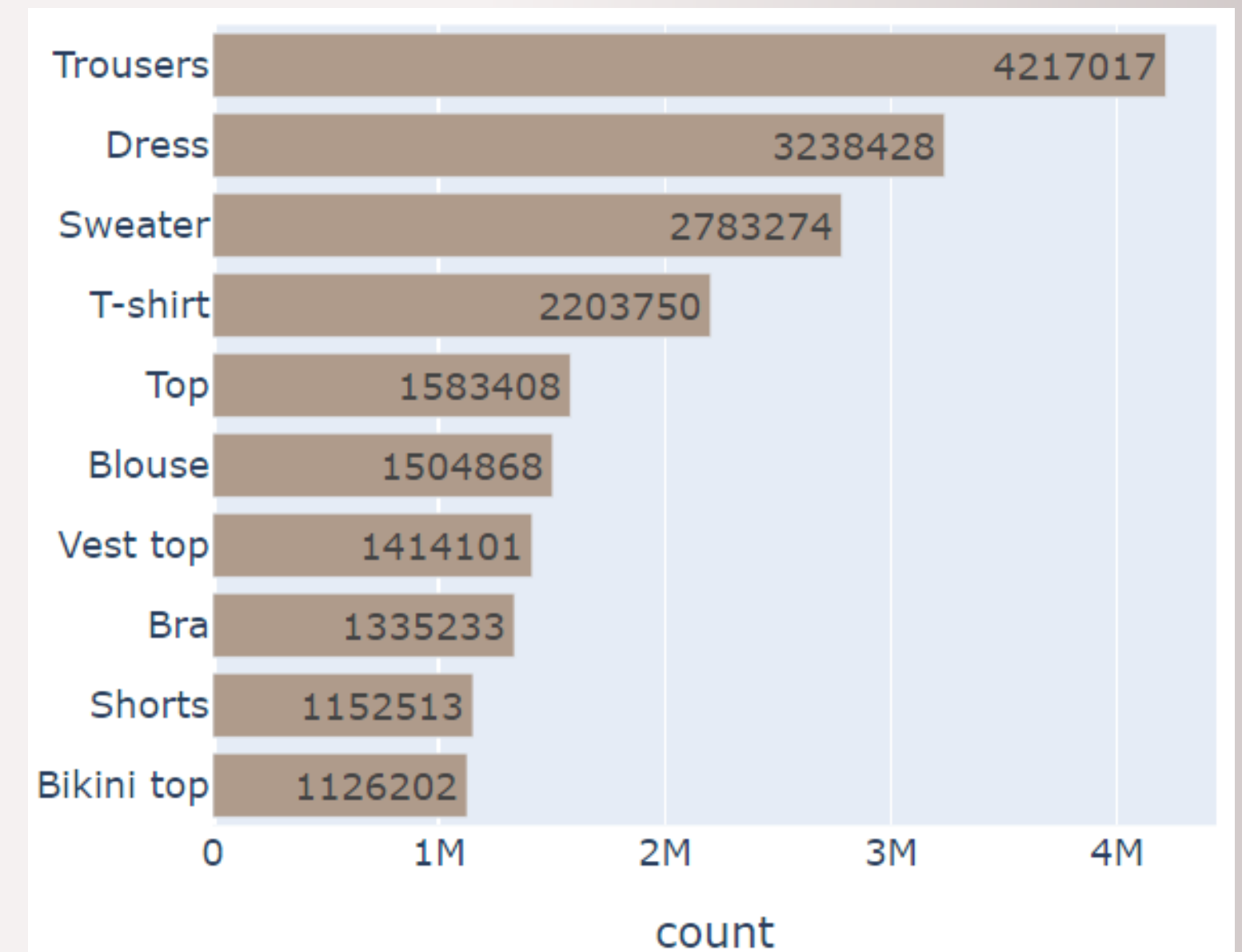**CUSTOMERS** = 1371980 ROWS
TRANSACTION_TRAIN = 31788324 ROWS

EXPLORATORY DATA ANALYSIS (EDA)
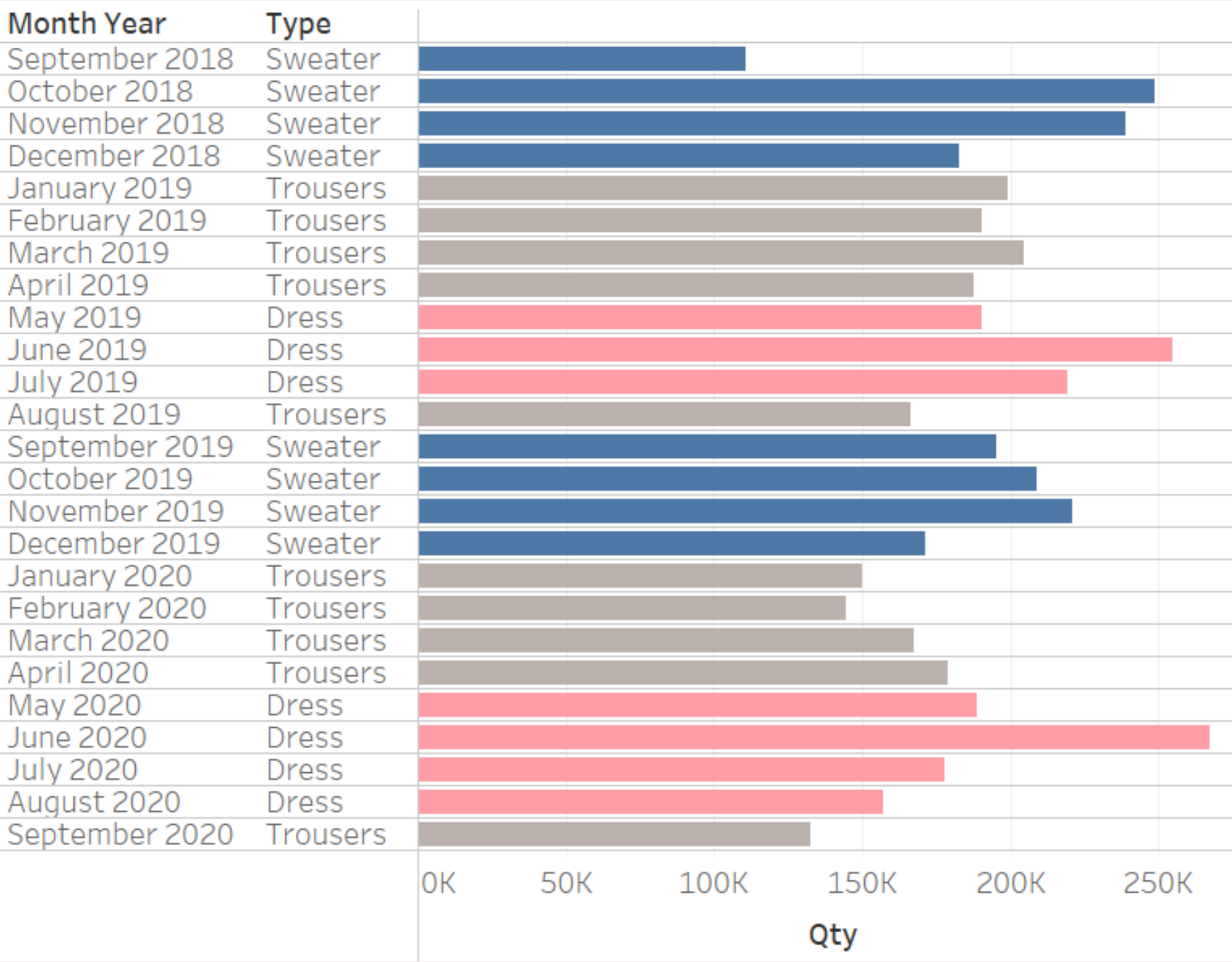
Top 10 Bestselling Products

Product	count
Jade HW Skinny Denim TRS	168052
Luna skinny RW	143216
Timeless Midrise Brief	122143
Tilly (1)	105670
Cat Tee.	81304
Simple as That Triangle Top	74827
Shake it in Balconette	74574
Tilda tank	68891
Despacito	67607
Simple as that Cheeky Tanga	64603

Top 10 Bestselling Product Types

Product Type	count
Trousers	4217017
Dress	3238428
Sweater	2783274
T-shirt	2203750
Top	1583408
Blouse	1504868
Vest top	1414101
Bra	1335233
Shorts	1152513
Bikini top	1126202

EXPLORATORY DATA ANALYSIS (EDA)

Bestselling Product Types Each Month

Month Year	Type
September 2018	Sweater
October 2018	Sweater
November 2018	Sweater
December 2018	Sweater
January 2019	Trousers
February 2019	Trousers
March 2019	Trousers
April 2019	Trousers
May 2019	Dress
June 2019	Dress
July 2019	Dress
August 2019	Trousers
September 2019	Sweater
October 2019	Sweater
November 2019	Sweater
December 2019	Sweater
January 2020	Trousers
February 2020	Trousers
March 2020	Trousers
April 2020	Trousers
May 2020	Dress
June 2020	Dress
July 2020	Dress
August 2020	Dress
September 2020	Trousers

Qty: 0K 50K 100K 150K 200K 250K

Observations

Sweater, trousers, or dress are the most popular products each month from 2018 to 2020.

EXPLORATORY DATA ANALYSIS (EDA)

Top 5 Best Selling Product Types from August 2018 - August 2020
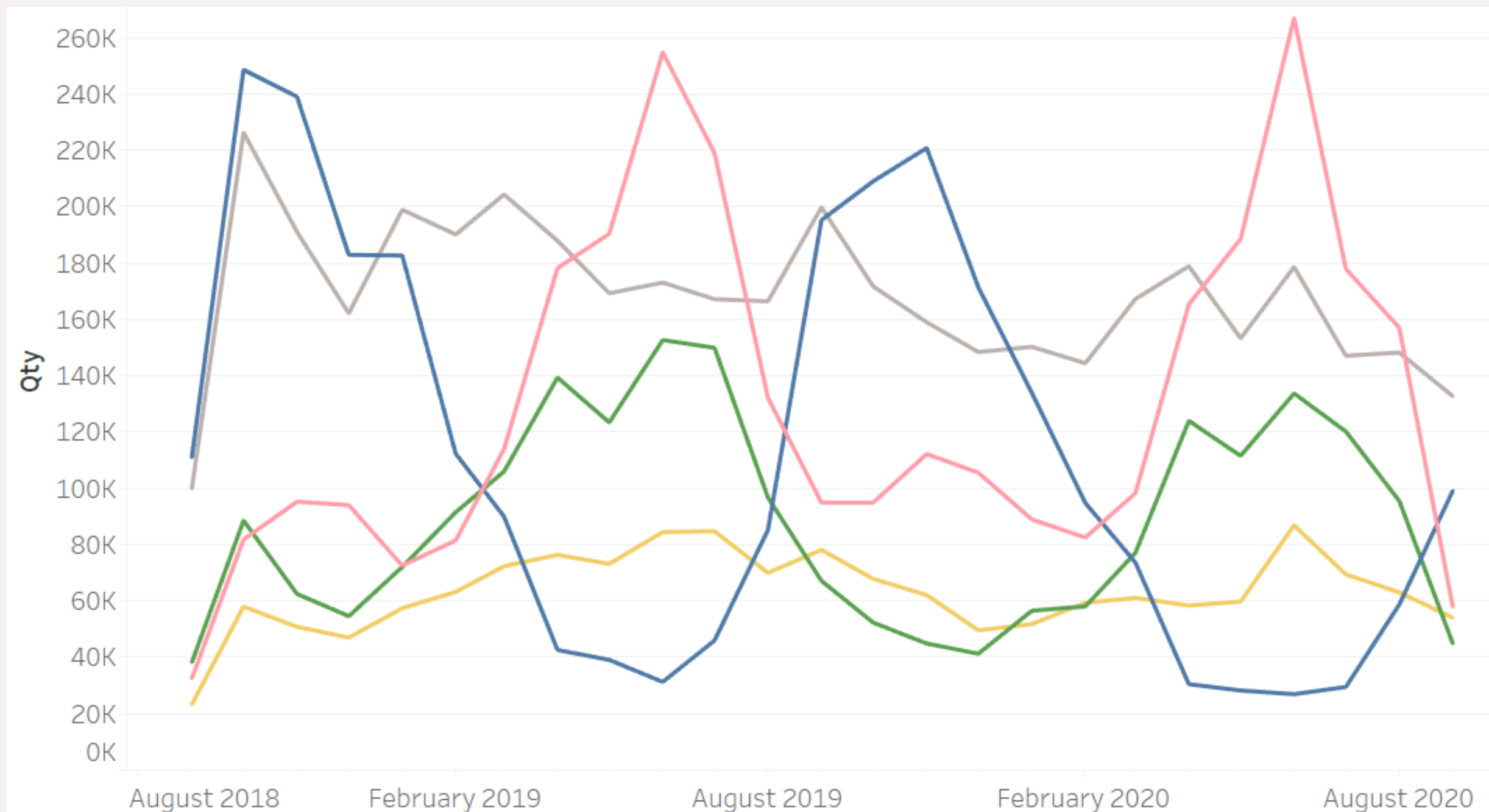
Observations
The top five products of all time have different sales patterns.

SALES PATTERN OF PRODUCT TYPES FROM JANUARY TO DECEMBER

Additional information:

H&M Group is a global fashion brand with 53 online markets. The postal codes are anonymized. Since the postal codes might be from different countries or climate zones, the sales pattern of the same type of product may differ.

GROUPING TYPE OF PRODUCTS

Observations

1. Products in group A mostly sold well in September - February and products in group B mostly have high sales in around May - August
2. Customers bought more products in autumn-winter than in spring-summer.
3. There are type of products that sold well nearly all the time (e.g. hair ties, hair string, and keychain).

Recommendation

To recommend products to customers, **consider new products** and when certain type of products will be in **high demand**. It is critical to generate recommendations **using the most appropriate data.**

MARKET BASKET ANALYSIS

Market basket analysis (MBA) is a data mining techinique. The primary goal of MBA is to provide information to retailers in order to help them better understand their customers' purchasing habits, allowing them to make better-informed decisions.

SUPPORT

Probability a product appears in transaction data

CONFIDENCE

Probability customers purchase product B, given they have purchased product A. Support(A&B) / Support(B)

LIFT

Support(A&B)/ Support(B)*Support(A)

MARKET BASKET ANALYSIS

	FP Growth	Apriori
Speed	Faster	Slower
Candidates	X	V
Memory	Small	Large

1. PURCHASED TOGETHER PRODUCTS
2. PRODUCT PLACEMENT
3. PRODUCT BUNDLES
4. OFFER SPECIAL DEALS
5. RECOMMENDATION PRODUCT BOUGHT TOGETHER

www.techtarget.com/searchcustomerexperience/definition/market-basket-analysis

towardsdatascience.com/fp-growth-frequent-pattern-generation-in-data-mining-with-python-implementation-244e561ab1c3

MARKET BASKET ANALYSIS

MARCH-MAY 2020 DATA

	Invoice_no	prod_name
0	106601135	[Dixie]
1	106601171	[Shirtdress Lou]
2	106601220	[Kendrick Como, Kendrick tie thong]
3	106601222	[Mona, Bob v-neck 2-pack, Pingu spring, Twenty...
4	106601267	[TEA HOOD S.1]
...	...	...
11923	107560416	[Eleonor button dress, Olivia long linen shirt]
11924	107560424	[Sadie Shirt]
11925	107560631	[Lazer Razer High Rise Hipster, Lazer Razer Pa...
11926	107560915	[Olivia woven Shorts, Dessie HW Belted Woven S...
11927	107560962	[Sporty Spice PRICE Swimsuit]

11928 rows × 2 columns

MAY 2020 DATA

	Invoice_no	prod_name
0	107198512	[V-neck Strap Top.]
1	107198783	[Charlotte bralette laguna opt2, Coachella Pus...
2	107198842	[TILDA HIP BELT, Alicia wrap skirt]
3	107198952	[PEACH STRAP]
4	107199508	[Flipflop solid]
...	...	...
5417	107560416	[Eleonor button dress, Olivia long linen shirt]
5418	107560424	[Sadie Shirt]
5419	107560631	[Lazer Razer High Rise Hipster, Lazer Razer Pa...
5420	107560915	[Olivia woven Shorts, Dessie HW Belted Woven S...
5421	107560962	[Sporty Spice PRICE Swimsuit]

5422 rows × 2 columns

MARKET BASKET ANALYSIS

MARCH-MAY 2020 DATA
PREDICT JUNE 2020

RULES= 215

ESTIMATED
ADDITIONAL REVENUE

16.56

REVENUE

>654.2

ESTIMATED INCREASE

>2.53%

	antecedent	consequent	confidence	lift	support
0	{Baby Love Cheeky Brief}	{Baby Love Triangle Top}	0.727273	1084.363636	0.000671
1	{Baby Love Triangle Top}	{Baby Love Cheeky Brief}	1.000000	1084.363636	0.000671
2	{Ginger Top}	{Ginger Highwaist}	0.545455	813.272727	0.000503
3	{Ginger Highwaist}	{Ginger Top}	0.750000	813.272727	0.000503
4	{Ursula Cheeky V Brief}	{Ursula Top}	0.571429	757.333333	0.000671
...	...	...	...	...	...

PREDICTION

	invoice_no	prod_name_x	prediction	lift	prod_name_y	price	total_price	price_est
8376	107993163	[Susan tank dress, Eleonor button dress]	0	0.000000	0	0.000000	0.053356	0.00000
7196	107931720	[Shake it in Balconette]	Swish HR wide side Brief	32.237838	Swish HR wide side Brief	0.006763	0.025407	0.00218
3316	107715147	[PIXI PINNAFORE, Price TEE TVP]	Money tee TVP	28.673077	Money tee TVP	0.004220	0.014373	0.00121
8278	107987321	[Gemini skirt.]	0	0.000000	0	0.000000	0.020322	0.00000
3789	107736191	[Mademoiselle dress, Chiba Dress, Dragonfly dr...	0	0.000000	0	0.000000	0.067339	0.00000

MARKET BASKET ANALYSIS

MAY 2020 DATA
PREDICT JUNE 2020

RULES= 487

	antecedent	consequent	confidence	lift	support
0	{Tropicana Brazilian CON}	{Tropicana Top}	1.000000	1355.500000	0.000553
1	{Tropicana Top}	{Tropicana Brazilian CON}	0.750000	1355.500000	0.000553
2	{Tropic Like Its Hot Brief}	{Tropic Like Its Hot Top}	1.000000	1084.400000	0.000553
3	{Tropic Like Its Hot Top}	{Tropic Like Its Hot Brief}	0.600000	1084.400000	0.000553
4	{Bondi Rescue Tie Brief}	{Bondi Rescue Bandeau}	0.600000	1084.400000	0.000553
...	...	...	...	...	...

ESTIMATED ADDITIONAL REVENUE

24.777

REVENUE

>654.2

ESTIMATED INCREASE

>3.787%

PREDICTION

	invoice_no	prod_name_x	prediction	lift	prod_name_y	price	total_price	price_est
5198	107815294	[Desert Top, Class Omar ring pk, Simple as tha...	Timeless Cheeky Brief	37.136986	Timeless Cheeky Brief	0.004051	0.033508	0.001504
4177	107757168	[Liza coral padded wire]	0	0.000000	0	0.000000	0.016932	0.000000
6177	107869665	[SF Michelle dress, Mia]	0	0.000000	0	0.000000	0.067763	0.000000
3100	107706784	[New Girl Cheeky Tanga, Timeless Midrise Brief...	Superwoman Superpush	28.574440	Superwoman Superpush	0.010661	0.111780	0.003046
1223	107624181	[ED Jafar, ED House]	0	0.000000	0	0.000000	0.037424	0.000000

MARKET BASKET ANALYSIS

DATA USED TO GENERATE RECOMMENDATIONS IN JUNE 2020

	MARCH-MAY 2020	**MAY 2020**
DATA	LARGE	SMALL
SPEED	SLOW	FAST
ADDITIONAL REVENUE	LESS	MORE

RECOMMENDATION:

Use **previous month's data** to generate rules for cross-selling or recommending products to customers for the next month.

Product Recommendation

NEW ITEMS

POPULAR ITEMS

CONTENT-BASED RECOMMENDATION

COLLABORATIVE FILTERING
Item-based and user-based recommendation

HYBRID FILTERING

REAL-WORLD RECOMMENDATION SYSTEM

1. THE RETRIEVAL STAGE	2. THE RANKING STAGE
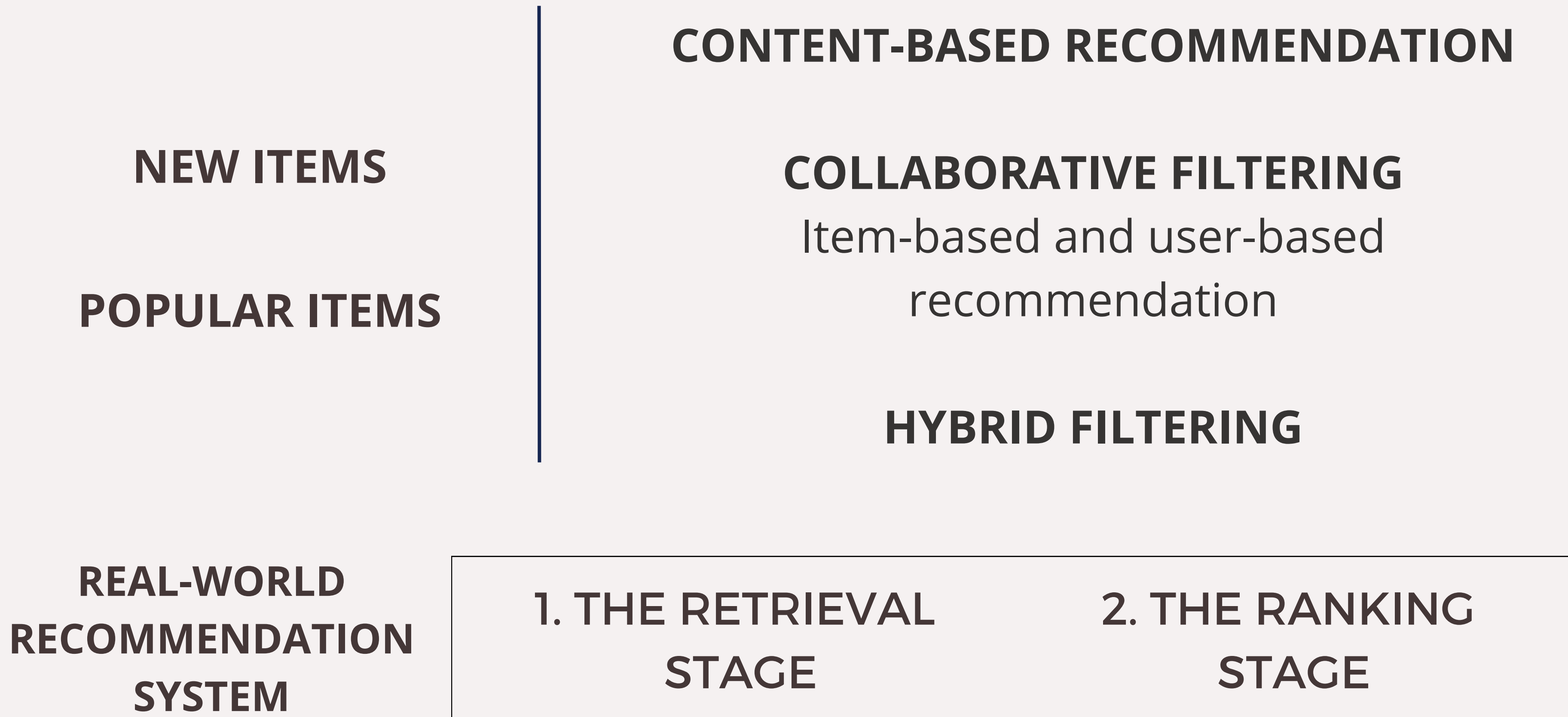
A Simple Recommendation System

1
- REPURCHASED ITEMS
- PREVIOUSLY PURCHASED ITEMS

2
USER-ITEM RECOMMENDATION (TENSORFLOW RECOMMENDER)

3
LAST WEEK'S POPULAR ITEMS

Evaluation Metric

MEAN AVERAGE PRECISION @ 12 (MAP@12)

$$MAP@12 = \frac{1}{U}\sum_{u=1}^{U}\sum_{k=1}^{\min\,(n,12)} P(k) \times rel(k)$$

P(K) = Precision at K

rel(K) = Rel at K (Indicator if the item at rank K is relevant)

Result

	customer_id	prediction
0	00000dbacae5abe5e23885899a1fa44253a17956c6d1c3...	0568601043 0858856005 0568601023 0927225001 09...
1	0000423b00ade91418cceaf3b26c6af3dd342b51fd051e...	0797988005 0832331003 0878190005 0809238001 07...
2	000058a12d5b43e67d225668fa1f8d618c13dc232df0ca...	0794321007 0805000008 0805000007 0794321011 07...
3	00005ca1c9ed5f5146b52ac8639a40ca9d57aeff4d1bd2...	0919365008 0677930066 0864216003 0736923009 09...
4	00006413d8573cd20ed7128e53b7b13819fe5cfc2d801f...	0791587015 0896152002 0730683050 0896152001 09...
...	...	...
1371975	ffffbbf78b6eaac697a8a5dfbfd2bfa8113ee5b403e474...	0557599022 0713997002 0720125039 0740922009 07...
1371976	ffffcd5046a6143d29a04fb8c424ce494a76e5cdf4fab5...	0818754002 0762846031 0762846026 0889460003 09...
1371977	ffffcf35913a0bee60e8741cb2b4e78b8a98ee5ff2e6a1...	0762846027 0689365050 0794819001 0884081001 07...
1371978	ffffd7744cebcf3aca44ae7049d2a94b87074c3d4ffe38...	0714790020 0906612001 0928088001 0714790028 09...
1371979	ffffd9ac14e89946416d80e791d064701994755c3ab686...	0685813018 0685813043 0685813042 0784926002 08...

1371980 rows × 2 columns

MAP@12: 0.0225

Recommendations

- In some months from March - August, the sales of products in group B didn't reach 50% of the highest sales of the products.
- To recommend products to customers, consider new products and when certain type of products will be in high demand. It is critical to generate recommendations using the most appropriate data.
- Offer special deals, product bundles based on the rules generated from the market basket analysis using previous month's data.
- Personalized product recommendations using hybrid recommendation system

References

1. www.kaggle.com/code/pierrelouisdanieau/recommender-system-associations-rules/notebook#4.-Association-Rules-modelling-:-Fp-growth-algorithm (Recommender System Associations Rules | Kaggle)
2. www.techtarget.com/searchcustomerexperience/definition/market-basket-analysis (What is market basket analysis? Definition from WhatIs.com)
3. www.sciencedirect.com/science/article/pii/S1877050916305208 (Market Basket Analysis: Identify the Changing Trends of Market Data Using Association Rule Mining - ScienceDirect)
4. towardsdatascience.com/fp-growth-frequent-pattern-generation-in-data-mining-with-python-implementation-244e561ab1c3 (FP Growth — Frequent Pattern Generation in Data Mining with Python Implementation | by Chonyy | Towards Data Science)
5. www.tensorflow.org/recommenders/examples/basic_retrieval (Recommending movies: retrieval | TensorFlow Recommenders)