



PRODUCT RECOMMENDATION AND MARKET BASKET ANALYSIS

Filzahanti Nuha Ramadhani



OUTLINE

01

**DATASET
INTRODUCTION**

02

**DATA
EXPLORATION**

04

**MARKET BASKET
ANALYSIS**

03

**PRODUCT
RECOMMENDATION**

05

RECOMMENDATIONS

DATASET

H&M Personalized Fashion
Recommendations | Kaggle

53 ONLINE MARKETS
APPROX. 4850 STORES



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

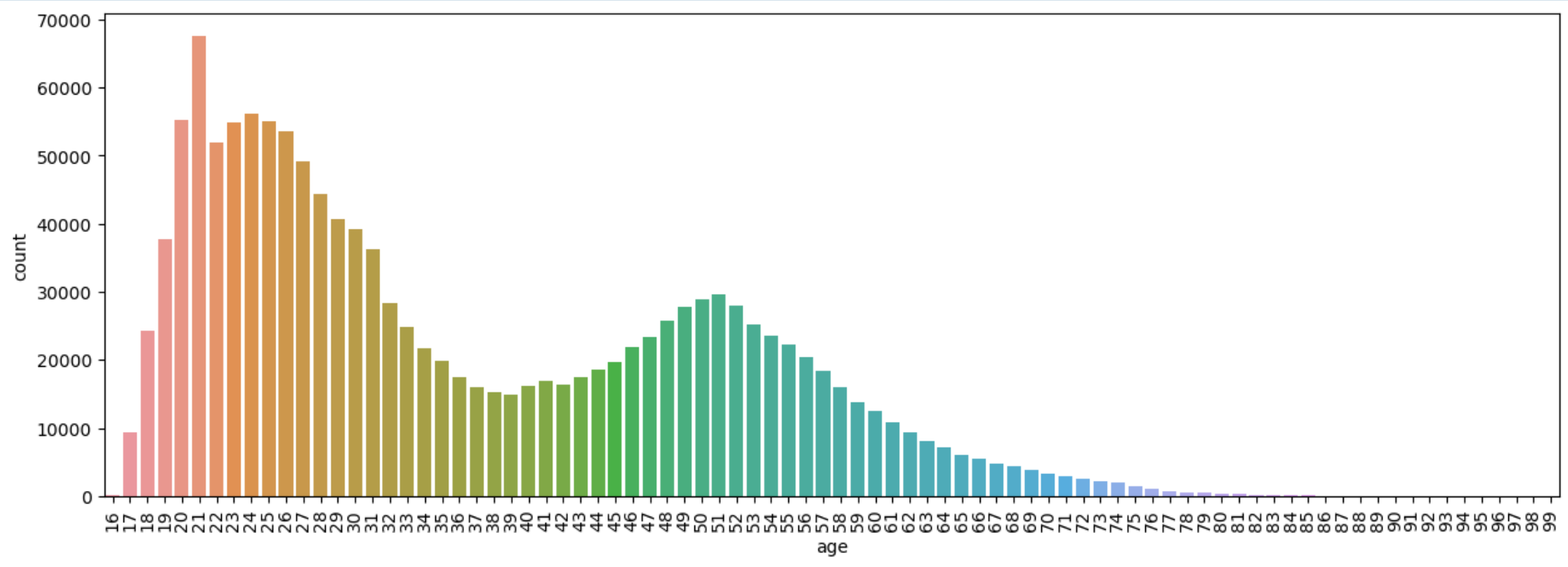
	0	1
article_id	108775015	108775044
product_code	108775	108775
prod_name	Strap top	Strap top
product_type_no	253	253
product_type_name	Vest top	Vest top
product_group_name	Garment Upper body	Garment Upper body
graphical_appearance_no	1010016	1010016
graphical_appearance_name	Solid	Solid
colour_group_code	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...

DATA EXPLORATION

AGE DISTRIBUTION



UNIQUE POSTAL CODE

352899

ARTICLE ID

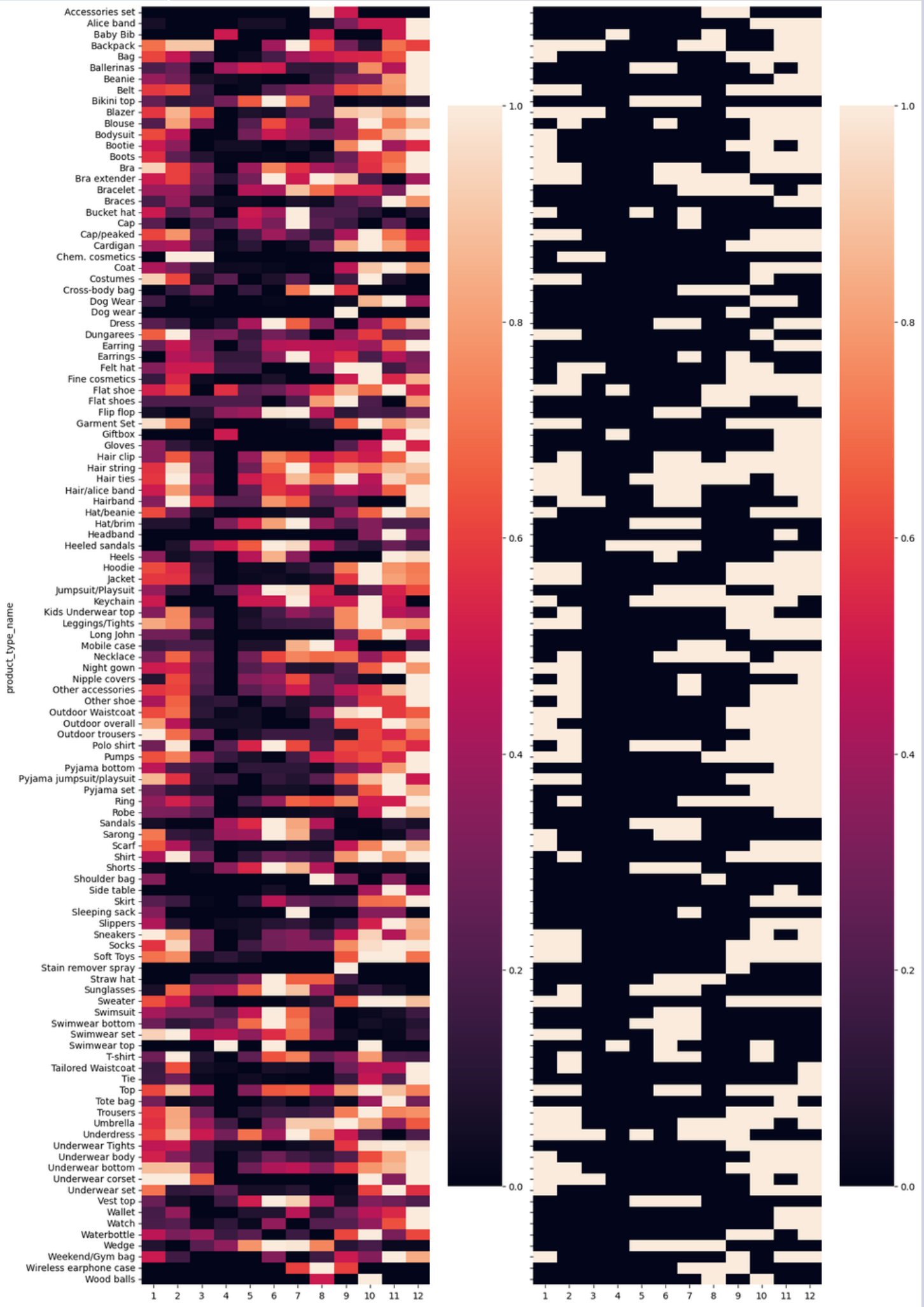
105542

PRODUCT CODE

47224

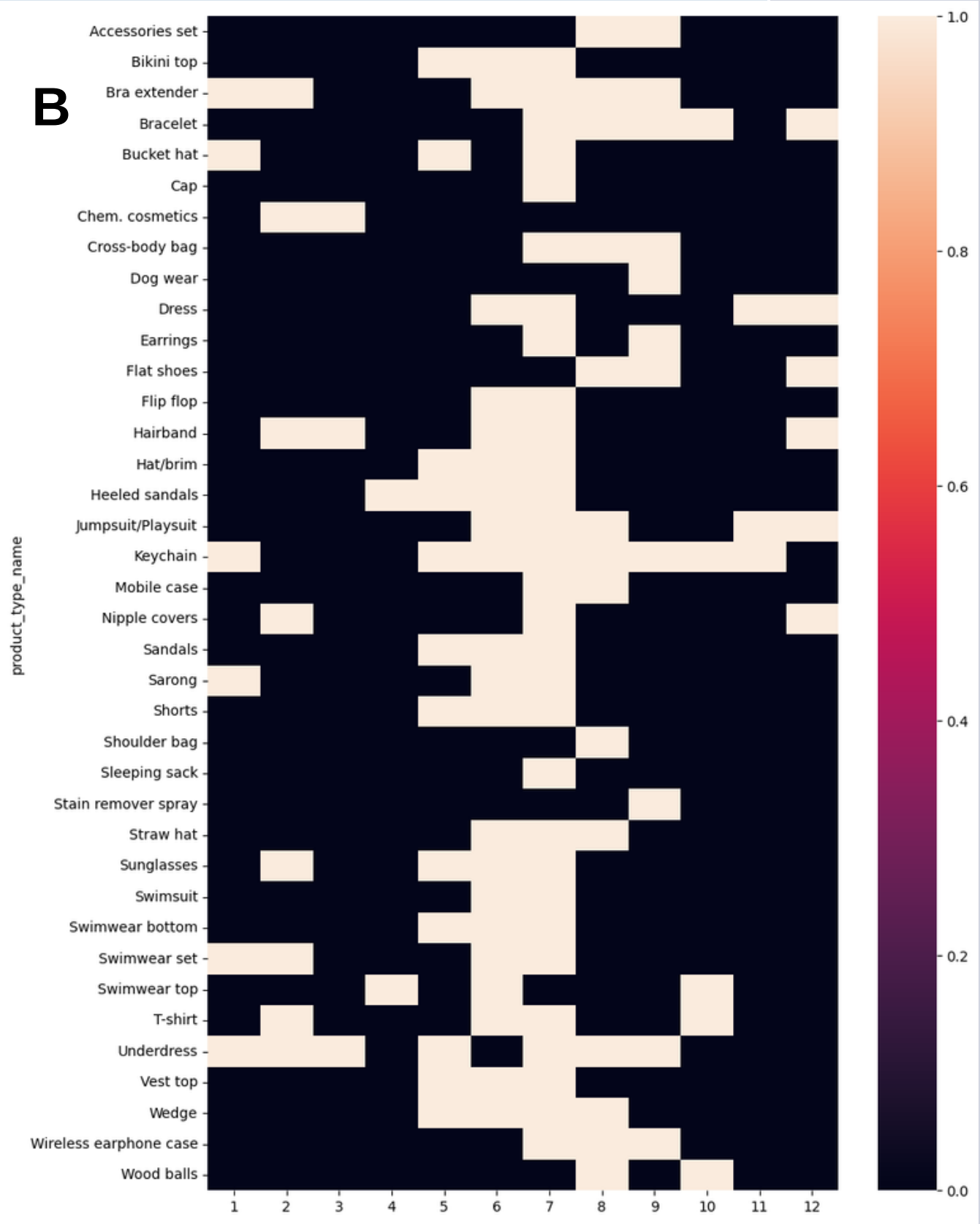
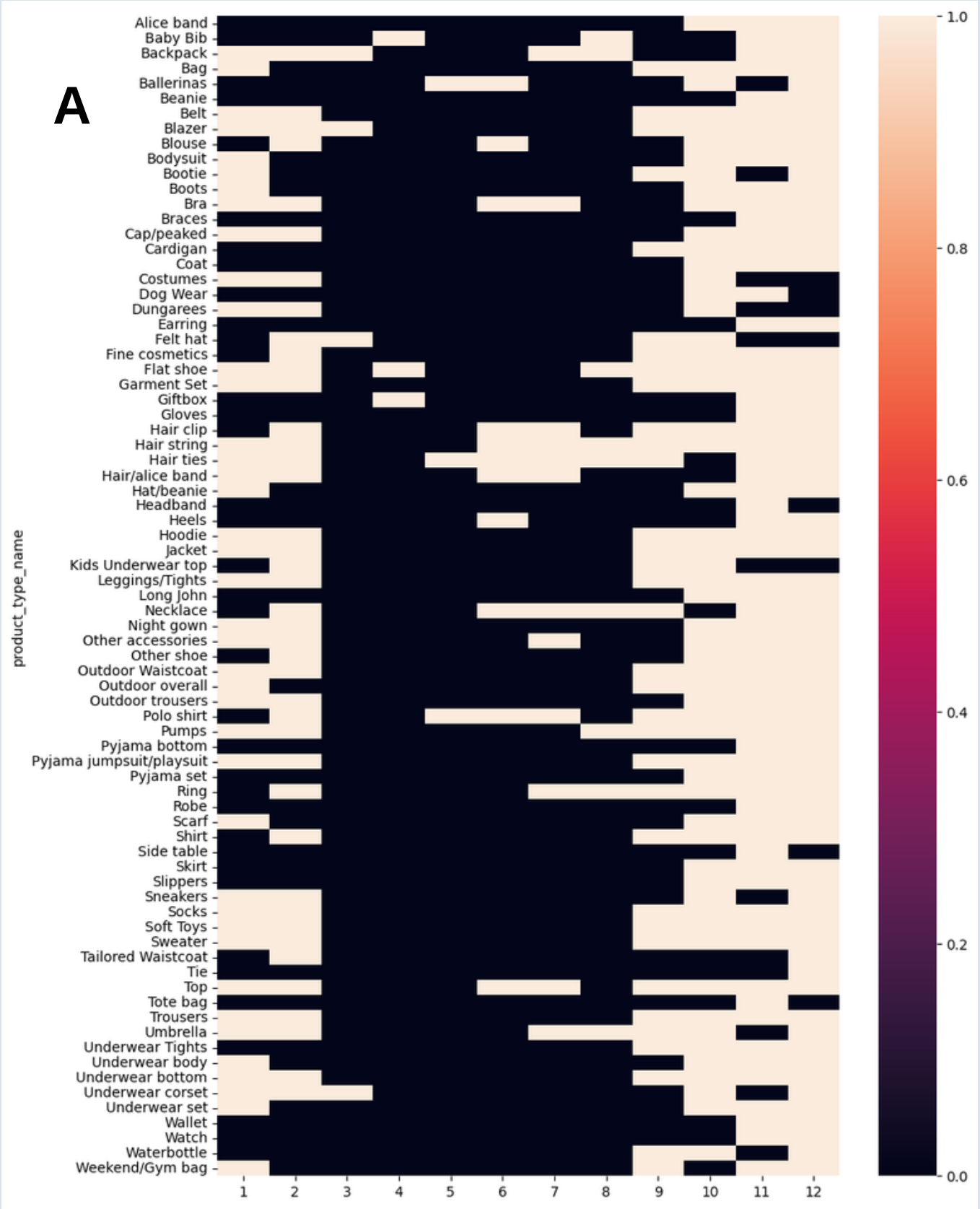
INVOICE NO

9080179



DATA EXPLORATION

OBSERVATIONS



1. Products in group A mostly sold well in September - February
2. Products in group B mostly have high sales in around May - August
3. There are winter clothing in group A, but not all of them are associated with winter
4. Customers bought products mostly in autumn and winter

PRODUCT RECOMMENDATION

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 PRODUCT RECOMMENDATION

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

MAP@12 **0.02077**

DATA EXPLORATION II

MARCH 2020 - JUNE 2020

	postal_code	transaction
0	2c29ae653a9282cce4151bd87643c907644e09541abc28...	199523
1	1f5bd429acc88fbbf24de844a59e438704aa8761bc7b99...	1439

Product Name	March	April	May	June	July	August	Grand T..
Tilly (1)	76	45	124	213	252	256	966
Tilda tank	59	65	121	207	148	123	723
Cat Tee.	41	26	66	175	167	145	620
Nora T-shirt	33	13	101	155	120	134	556
Jade HW Skinny ..	77	50	77	67	133	151	555
Timeless Midrise..	20	40	97	162	145	83	547
Simple as That T..	31	23	126	204	118	30	532
Luna skinny RW	76	22	87	100	122	122	529
Pluto RW slacks ..	81	12	83	118	91	117	502
Charlie Top	82	3	88	115	117	96	501
Jen tee	65	17	86	167	75	77	487
Melrose	47	16	49	115	124	111	462
Skinny Ankle R...	117	2	73	87	126	54	459
Primo slacks	61	39	71	93	100	95	459
Simple as that C..	34	11	113	171	85	44	458
Brit Baby Tee	27	14	67	70	125	133	436
7p Basic Shaftless	46	12	56	81	127	108	430
SUPREME RW ti..	83	16	44	58	123	104	428
Sorrel	10	5	71	120	128	94	428
Tequila denim			40	121	163	73	397
Mariette Blazer	83		39	62	72	111	367
Eleven top	5	1	28	111	142	77	364
Therese tee	37	31	32	82	102	66	350
Lemon HW			11	120	144	73	348
Thyme ton	4	22	64	138	79	40	347

MARKET BASKET ANALYSIS

Market basket analysis (MBA) is a data mining technique.

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.
 $\text{Support}(A\&B) / \text{Support}(B)$

LIFT

$\text{Support}(A\&B) /$
 $\text{Support}(B) * \text{Support}(A)$

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

MARKET BASKET ANALYSIS

DATA MARCH-MAY 2020

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 x 2 columns		

DATA MAY 2020

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 x 2 columns		

MARKET BASKET ANALYSIS

DATA MARCH-MAY 2020

PREDICT JUNE 2020

RULES= 215

	antecedent	consequent	confldence	llft	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	llft	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

ESTIMATED REVENUE

16.56

REVENUE

>654.2

ESTIMATED INCREASE

>2.53%

MARKET BASKET ANALYSIS

DATA MAY 2020
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
...

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

ESTIMATED REVENUE

24.777

REVENUE

>654.2

ESTIMATED INCREASE

>3.787%

REFERENCES

1. [WWW.KAGGLE.COM/CODE/PIERRELOUISDANIEAU/RECOMMENDER-SYSTEM-ASSOCIATIONS-RULES/NOTEBOOK#4.-ASSOCIATION-RULES-MODELLING-:-FP-GROWTH-ALGORITHM](https://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](https://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](https://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](https://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 CHONY Y | TOWARDS DATA SCIENCE)
5. [WWW.TENSORFLOW.ORG/RECOMMENDERS/EXAMPLES/BASIC_RETRIEVAL](https://www.tensorflow.org/recommenders/examples/basic_retrieval)
(RECOMMENDING MOVIES: RETRIEVAL | TENSORFLOW RECOMMENDERS)

RECOMMENDATIONS

- IN SOME MONTHS FROM MARCH - AUGUST, THE SALES OF PRODUCTS IN GROUP B DIDN'T REACH 50% OF HIGHEST SALES OF TYPES OF PRODUCT.
- OFFER SPECIAL DEALS, PRODUCT BUNDLES BASED ON RULES GENERATED FROM MARKET BASKET ANALYSIS USING PREVIOUS MONTH'S DATA.