

01

RECOMMENDATION SYSTEM FOR



June 2022

H&M Team



JORDANO GARZA

Ing. Negocios y
Tecnología



ALBERTO MONTÁN

Ing. Mecatrónica



AXEL GARCÍA

Ing. Físico

Contents

- BUSINESS UNDERSTANDING
- DATA EXPLORATION
- DATA PREPARATION
- MODEL BUILDING
- HYPER-PARAMETER TUNING
AND MODEL EVALUATION
- RESULT / OUTCOMES

Bussines Understanding

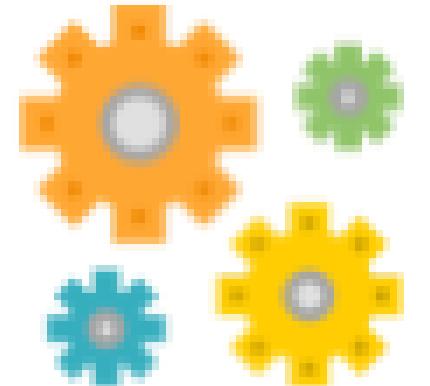


RECOMMENDER SYSTEM FOR H&M | JUNE
2022

Bussines Understanding

kaggle

Previous
Transactions



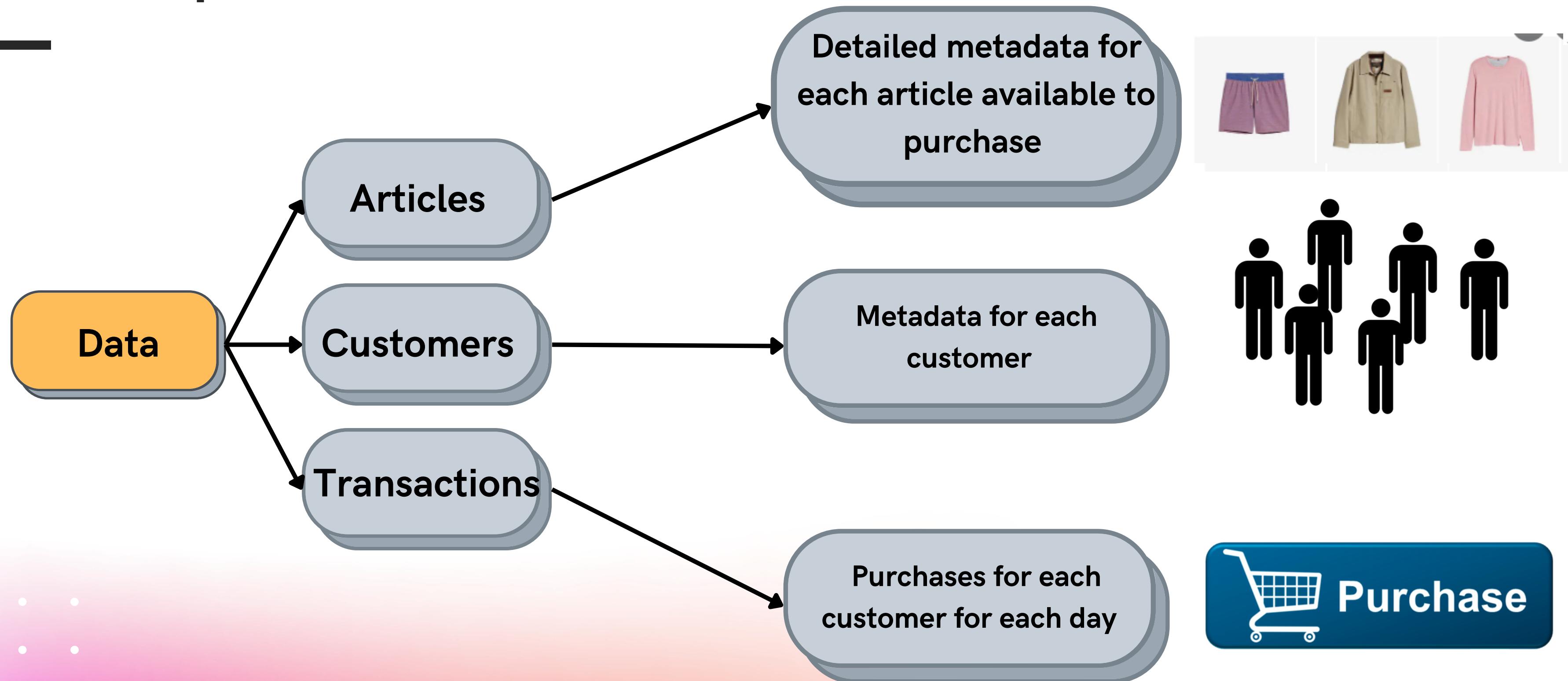
RECOMMENDER
SYSTEM



Top articles for
an individual

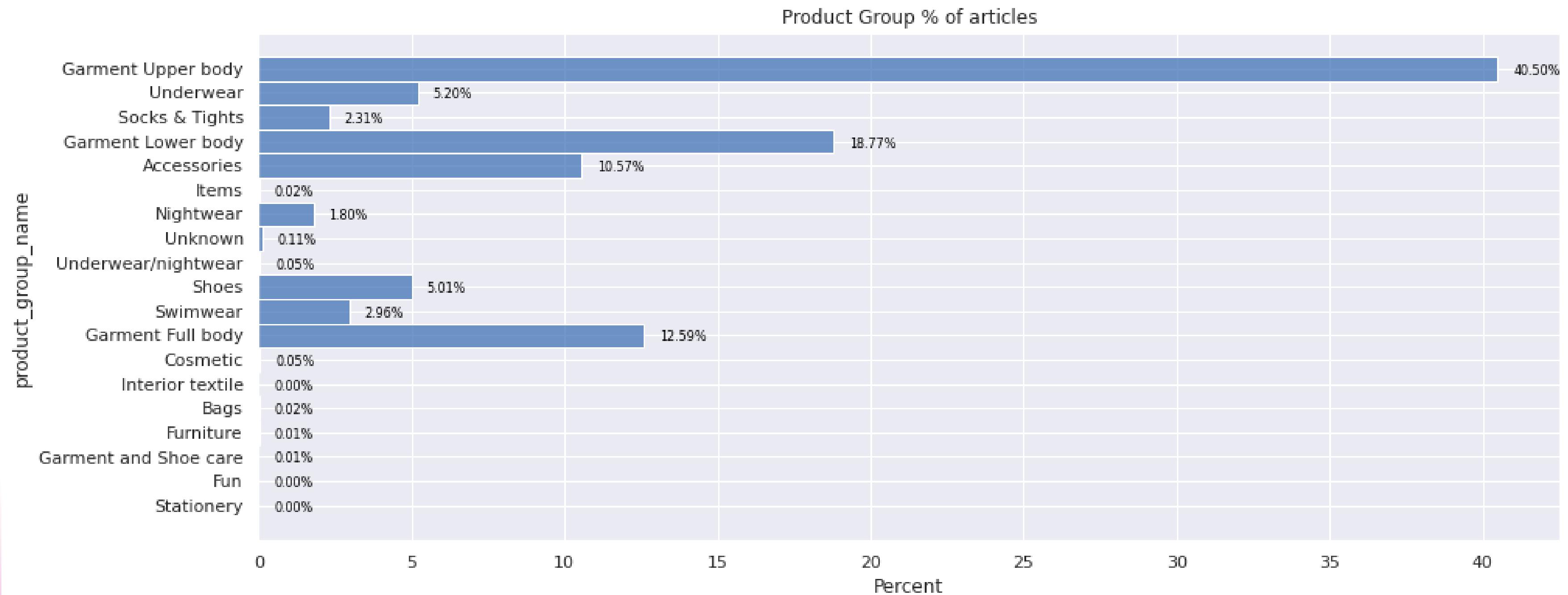


Data Exploration



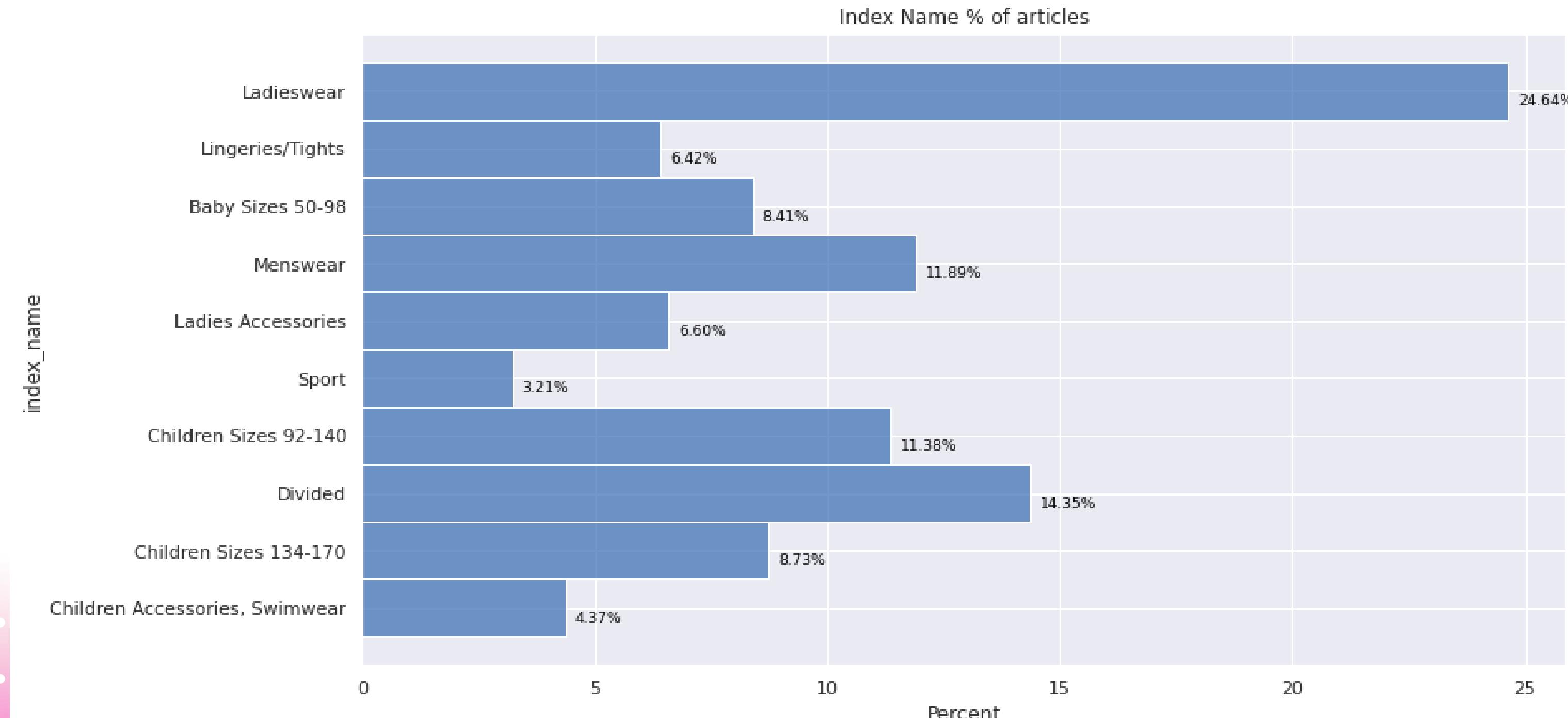
Data Exploration:

Articles



Data Exploration:

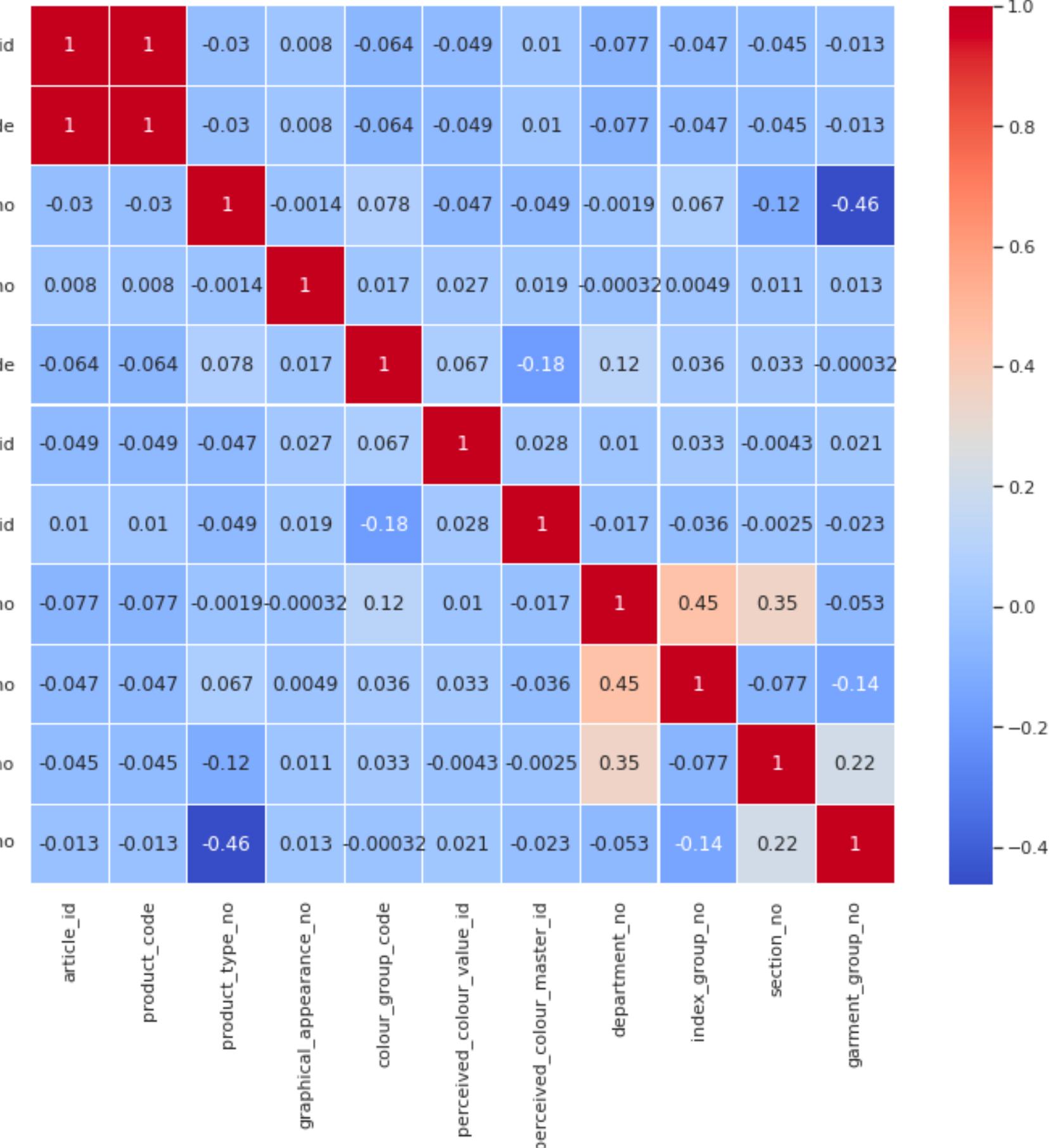
Articles



Data Exploration: Articles

Product code strongly correlated with the target "article_ID"

Correlation Matrix for Articles Dataset



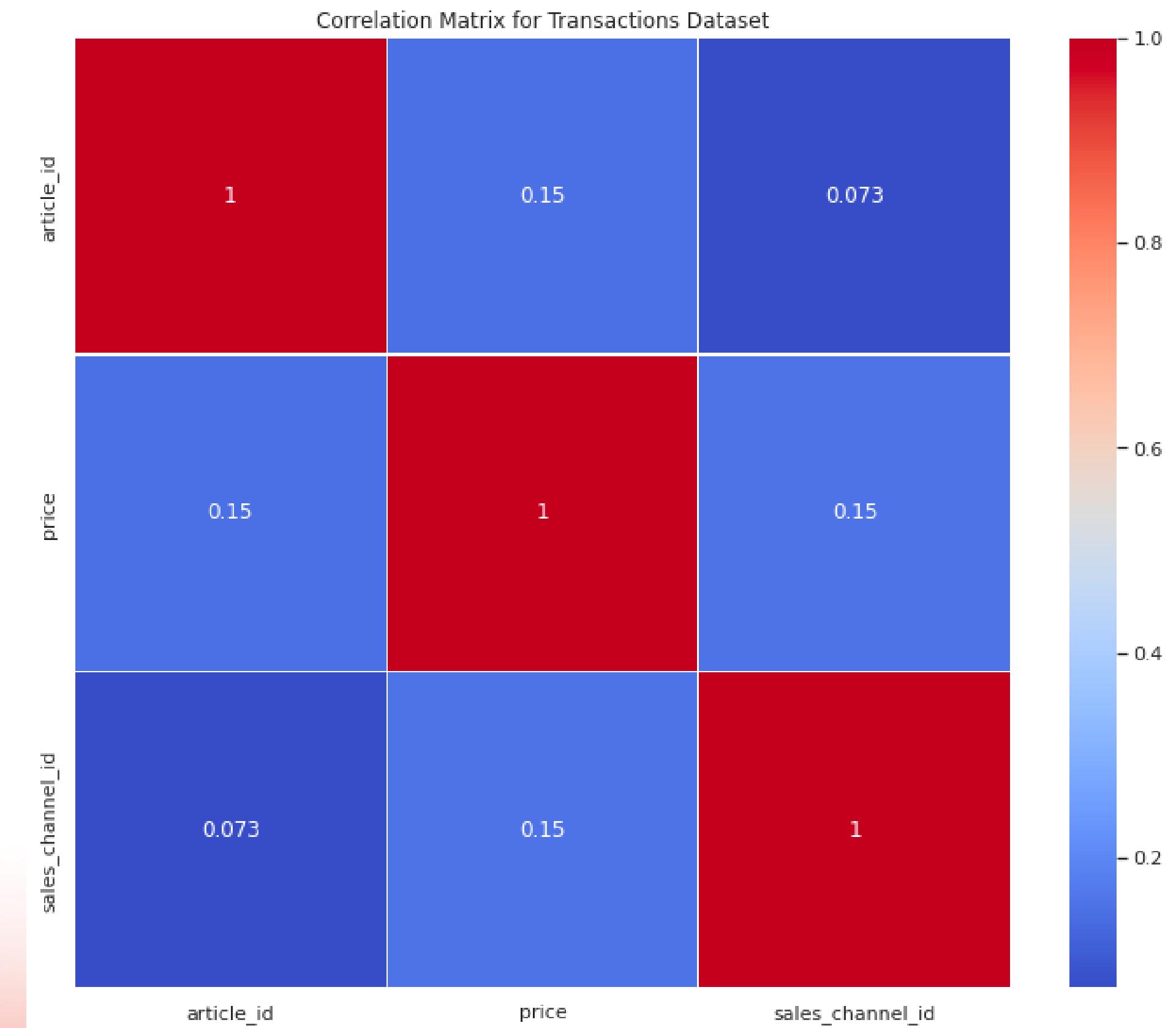
Data Exploration: Customers



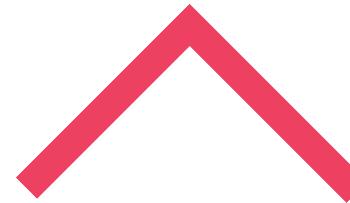
Target customer seems to
be in between 18-30
years old

Data Exploration: Transactions

- Around 18 million entries and 5 features.
- Transactions will be the unique dataset for training and doing predictions.
- There were no correlations between features.



Data preparation



*Transactions
Dataset

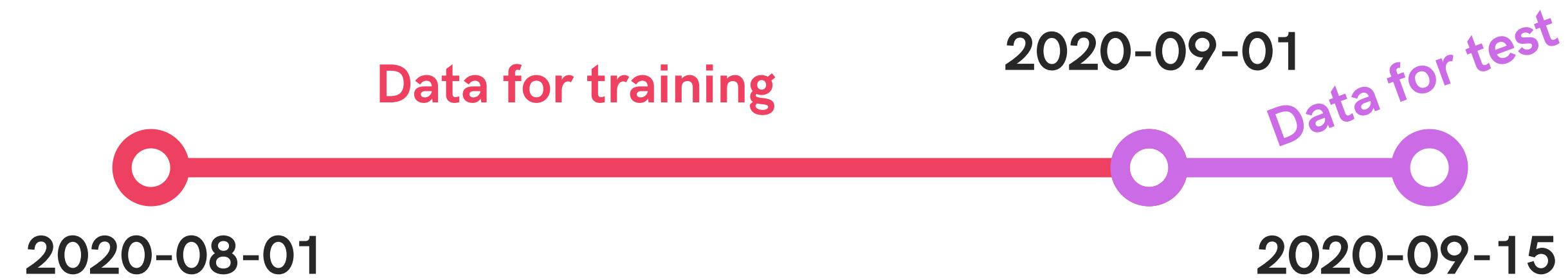
Train and test models



Data preparation



- 254163 Customers available

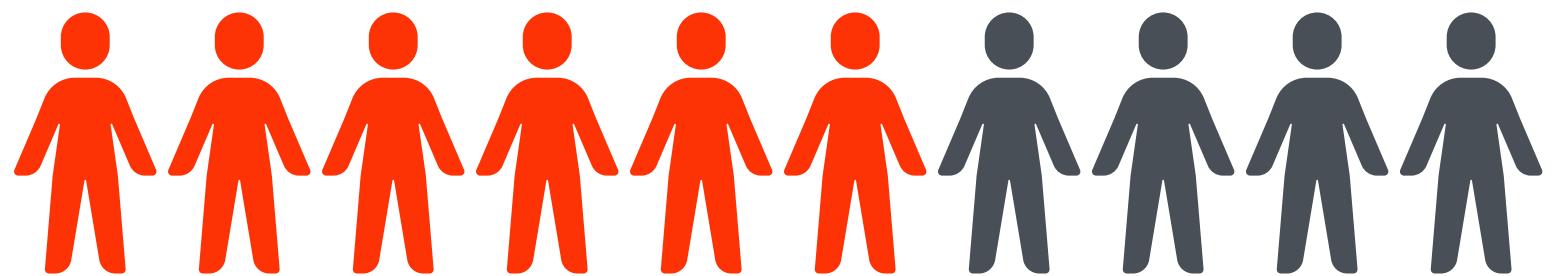


Data preparation

Customers after filtering

- 254163 Original customers
- 29978 Customers after first filter
- 13462 Customers after second filter

Filter the dataset



- Minimum number of items
- Same items on training and test

Data preparation

Dataset Columns

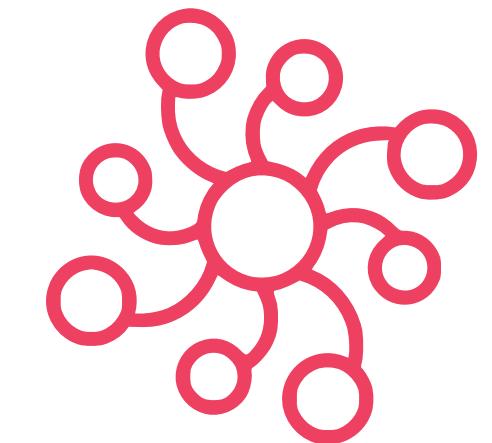
- userID: Unique identifier for each customer
- itemID: Unique identifier for each item
- rating: Whether the customer has purchased the item or not

Final dataset format

	userID	itemID	rating
0	0	572998001	1
1	0	572998007	1
2	0	684824006	1
3	0	713253003	1
4	0	808685002	1

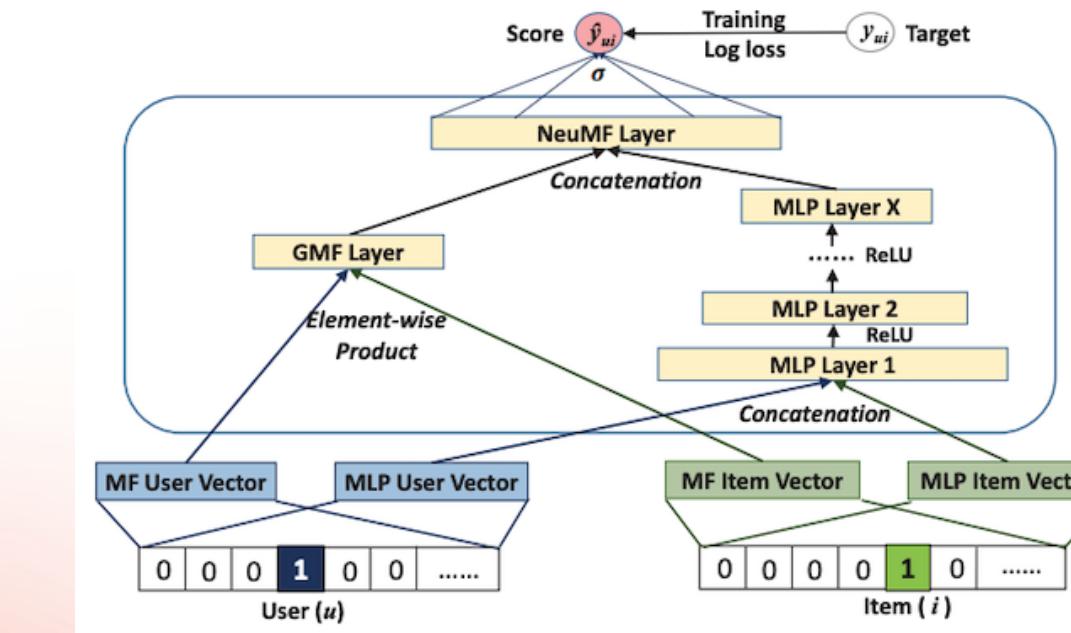
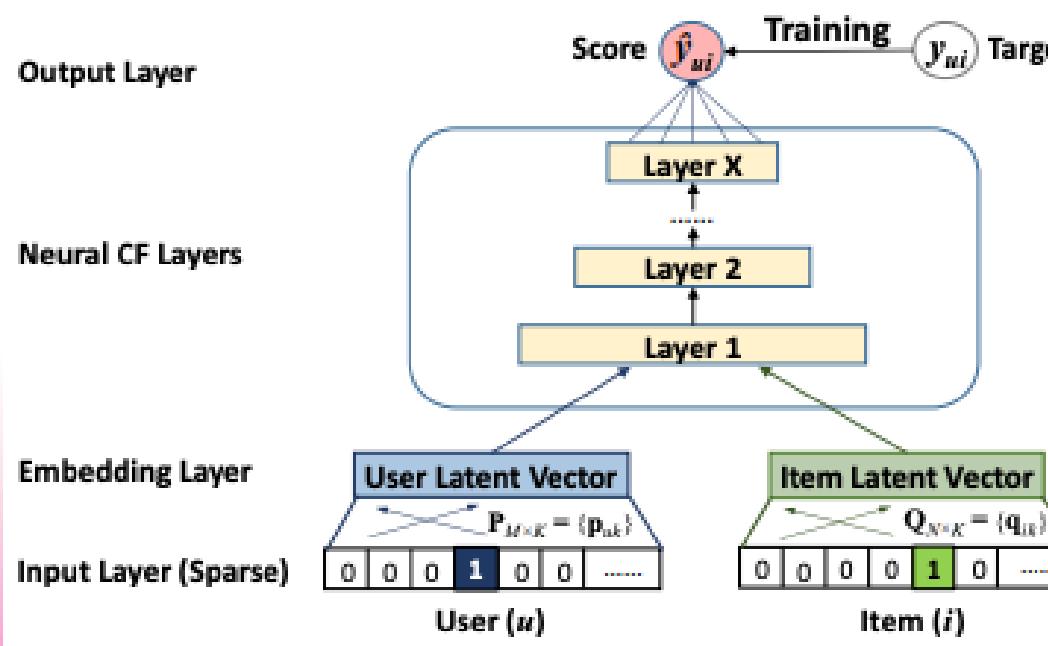
Model building

Models based on collaborative filtering technique

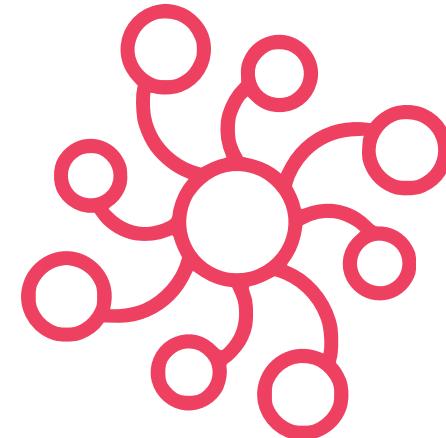


Neural Collaborative Filtering
(NCF)

Bilateral Variational Autoencoder (BiVAE)



Model building



Initial configuration

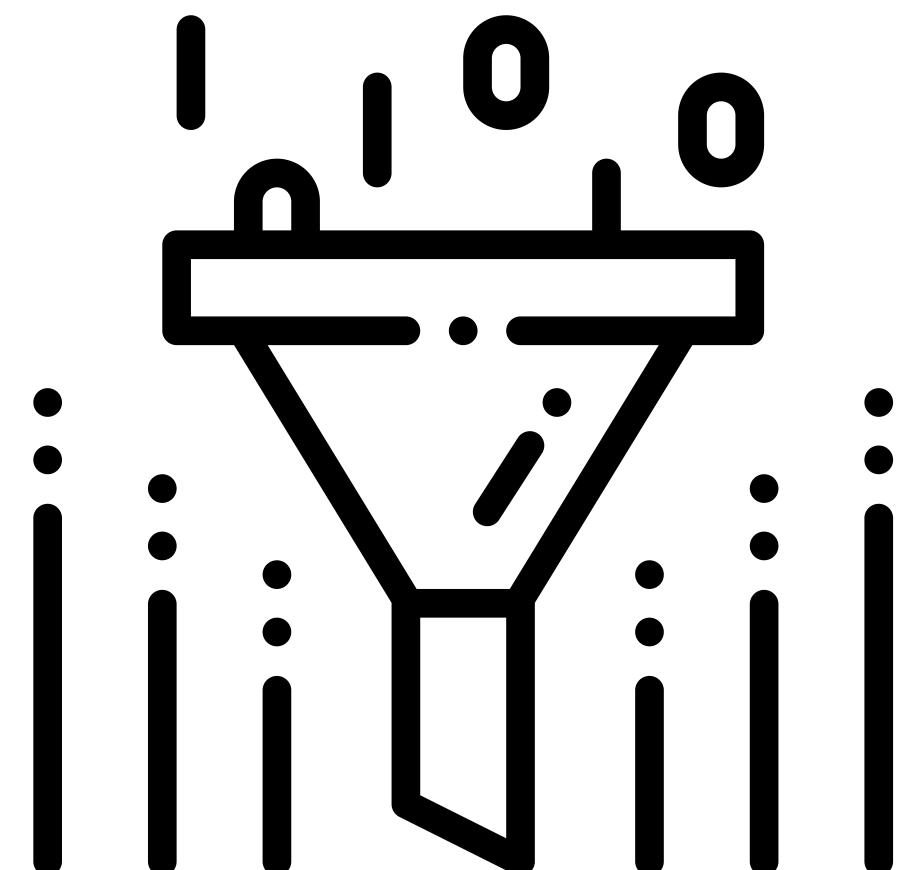
NCF

- Epochs: 5
- Batch size: 64
- Top K: 12
- Learning rate: 0.001

ViBAE

- Epochs: 5
- Batch size: 128
- Top K: 12
- Learning rate: 0.001

Hyper-parameter tuning and model evaluation



Five tests were performed for NCF and BiVAE models.

For NCF, these parameters are , TOP_K, # epochs, batch size, seed and learning rate.

And for BiVAE are latent_dim, enconders_dims, TOP_K, # epochs, batch size and learing rate

The evaluation will be displayed on a comparative table next for each method.

NFC hyper- parameters

	TOP_K	EPOCHS	BATCH SIZE	SEED	LEARINING RATE
Original run	12	5	64	42	1e-3
2nd	12	5	128	42	0.005
3th	12	10	128	42	0.009
4th	12	15	256	42	1e-4
5th	12	20	64	42	1e-2

NFC metrics

	Original run	2nd	3rd	4th	5th
mAP	0.008121	0.005587	0.002979	0.002388	0.000636
Precision@K	0.011667	0.008333	0.010000	0.008333	0.003333
Recall@K	0.035690	0.027803	0.014457	0.013194	0.003485
Training time	1465.63	1221.647	2303.655	3405.22	5270.619

BiVAE hyper- parameters

	LATENT_DI M	ENCODER_DI MS	TOP_K	NUM_EPOCHS	BATCH_SIZE	LEARNING_ RATE
1st	50	100	12	5	128	0.001
2nd	40	90	12	10	64	0.005
3th	60	110	12	15	256	0.009
4th	100	200	12	20	512	0.01
5th	5	10	12	5	64	0.0001

BiVAE metrics

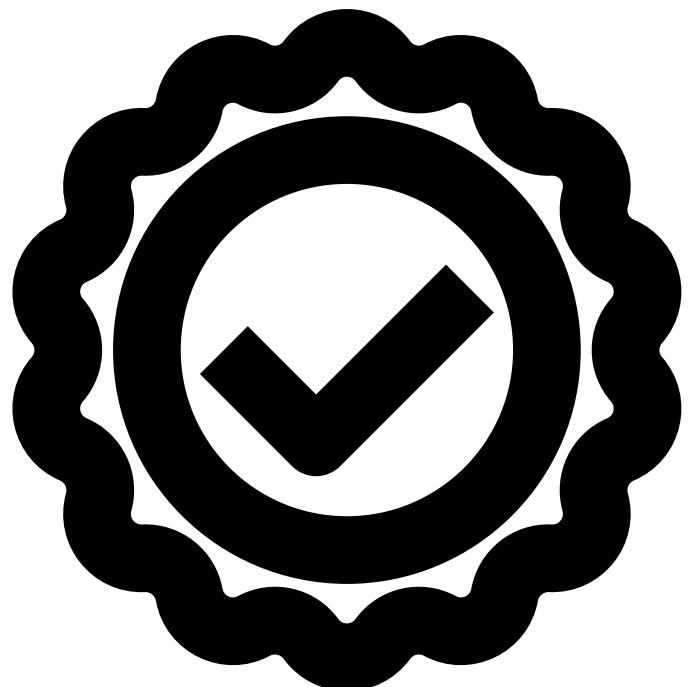
	mAP	Recall	Precision
1st	0.004321	0.030485	0.006667
2nd	0.005767	0.024318	0.005000
3th	0.013306	0.041250	0.006667
4th	0.000833	0.002500	0.001667
5th	0.000000	0.000000	0.000000

Best model

BiVAE with mAP value of **0.0133**



Results / Outcomes



The outcomes from this analysis were a dataframe in which every customer have potential articles that he/she could buy in the future.

	userID	itemID	prediction	rank
131	0	753778008	0.994701	8.0
173	0	832330003	0.996724	4.0
242	0	901955002	0.992990	12.0
279	0	776237031	0.996834	2.0
542	0	866383006	0.997796	1.0
...
979700	49	863980001	0.961343	10.0
980279	49	840607010	0.980818	2.0
980450	49	808029001	0.973376	6.0
982610	49	887937001	0.973914	5.0
983260	49	840607001	0.960832	12.0

CONCLUSION

References

- He, X., Liao, L., Zhang, H., Nie, L., Hu, X., \& Chua, T. S. (2017, April). Neural collaborative filtering. In Proceedings of the 26th international conference on world wide web (pp. 173-182)
- Truong, Q. T., Salah, A., \& Lauw, H. W. (2021, March). Bilateral variational autoencoder for collaborative filtering. In Proceedings of the 14th ACM International Conference on Web Search and Data Mining (pp. 292-300)