



H&M Shopping Recommendation System

Jeffrey, Lu, Ingrid



Objectives ✕

How will our recommendation systems benefit to H&M?



Reduce churn



Increase user satisfaction



**Increase sales,
AOV & AOQ**

(Average Order Value & Average Order Quantity)



**Reduce return for
sustainability**

H&M Dataset ×

Over 31.8M
products sold

9M
transactions

from
1.4M
customers

from 2018-2020

×

105K+
Products over
131
Product types

21
Garment Types with
50
Colour Types

Customers

Age, Membership,
Active Status, Postal Code

Products

Product Type, Colour,
Garment

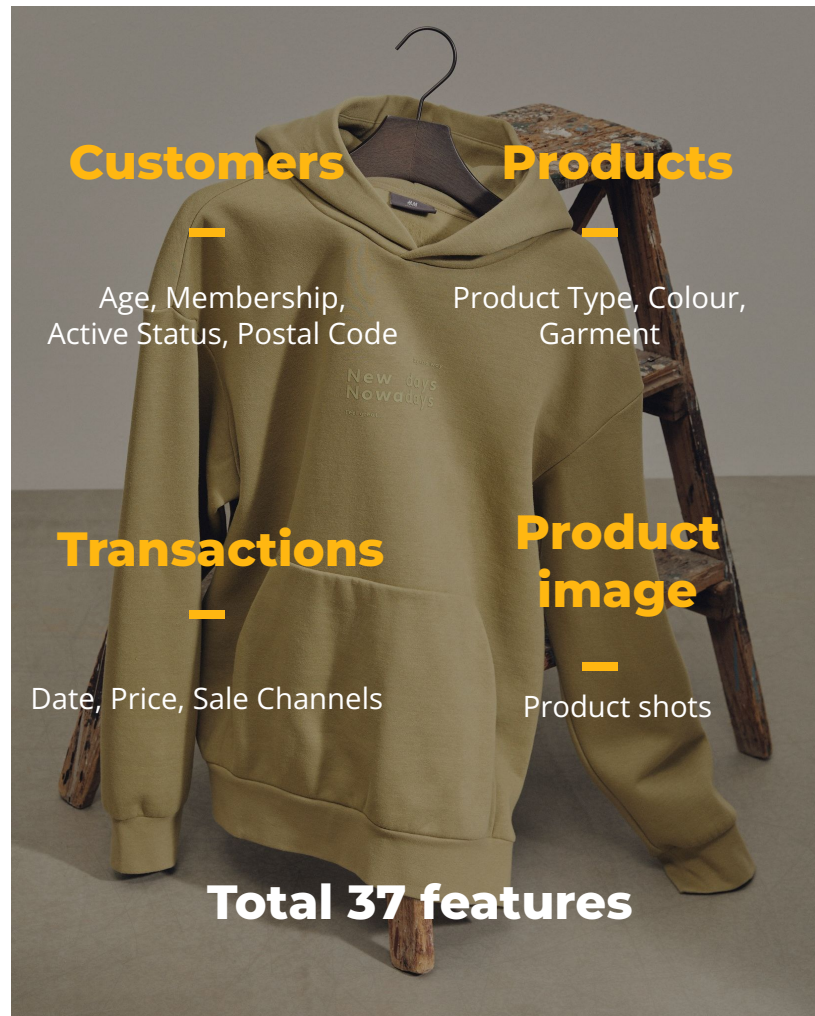
Transactions

Date, Price, Sale Channels

Product
image

Product shots

Total 37 features



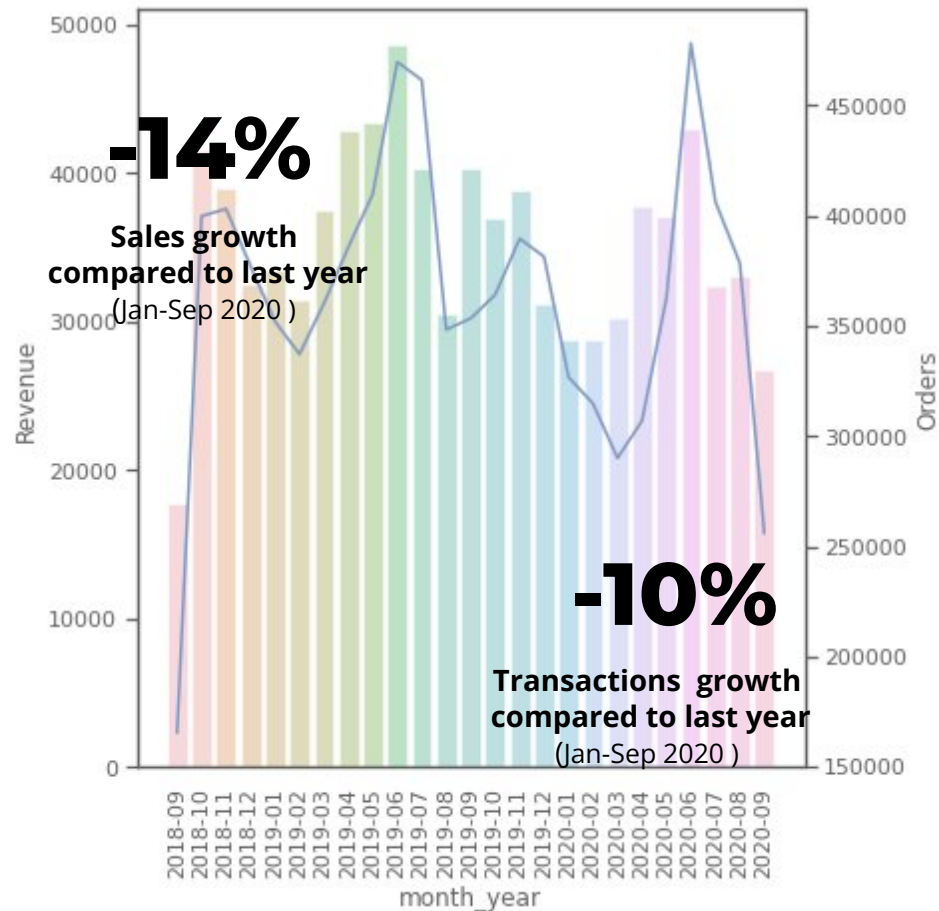
01 ×

Exploratory Data Analysis

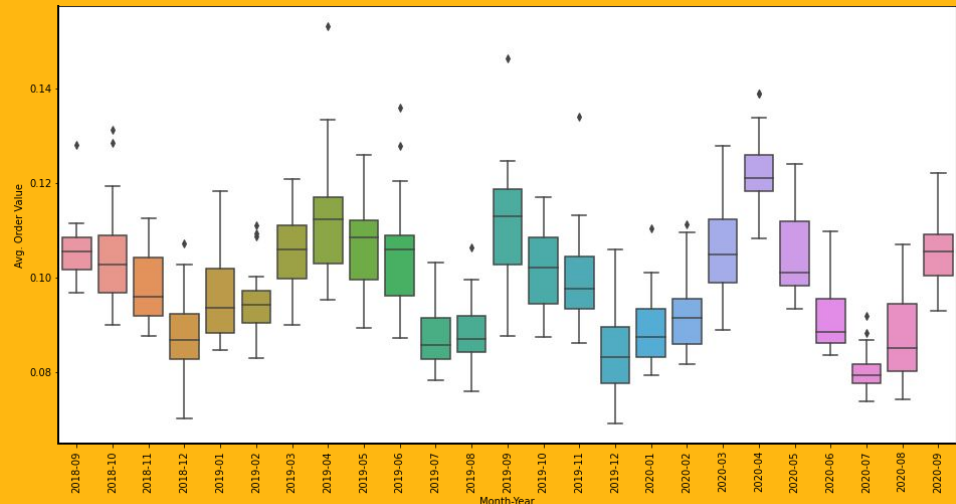




Our Business



Average Order Value



26%

Customers purchased less
than AOQ
(2018- 2020)

Avg. 3-4

items per order
(2018- 2020)

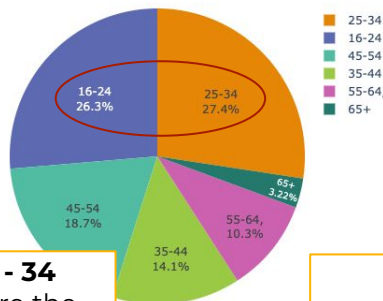


Our Customers

Customer Age Distribution

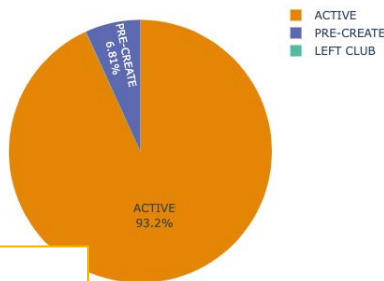


% by age range



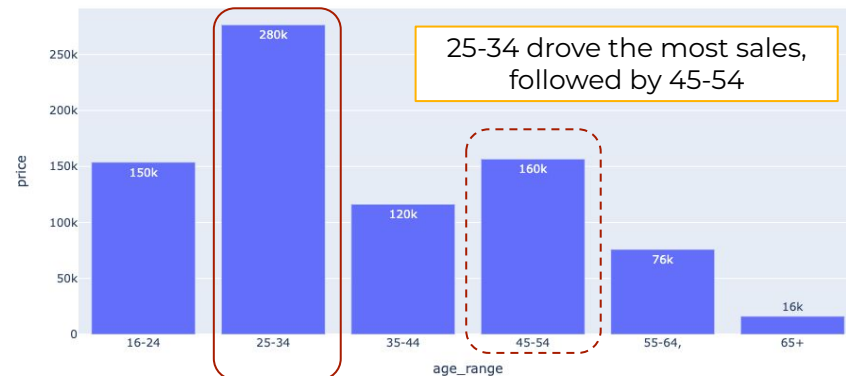
Age 16 - 34
(~54%) are the
key target
customer

Membership Club Status



93%
Active Users

Total Sales per Age Range



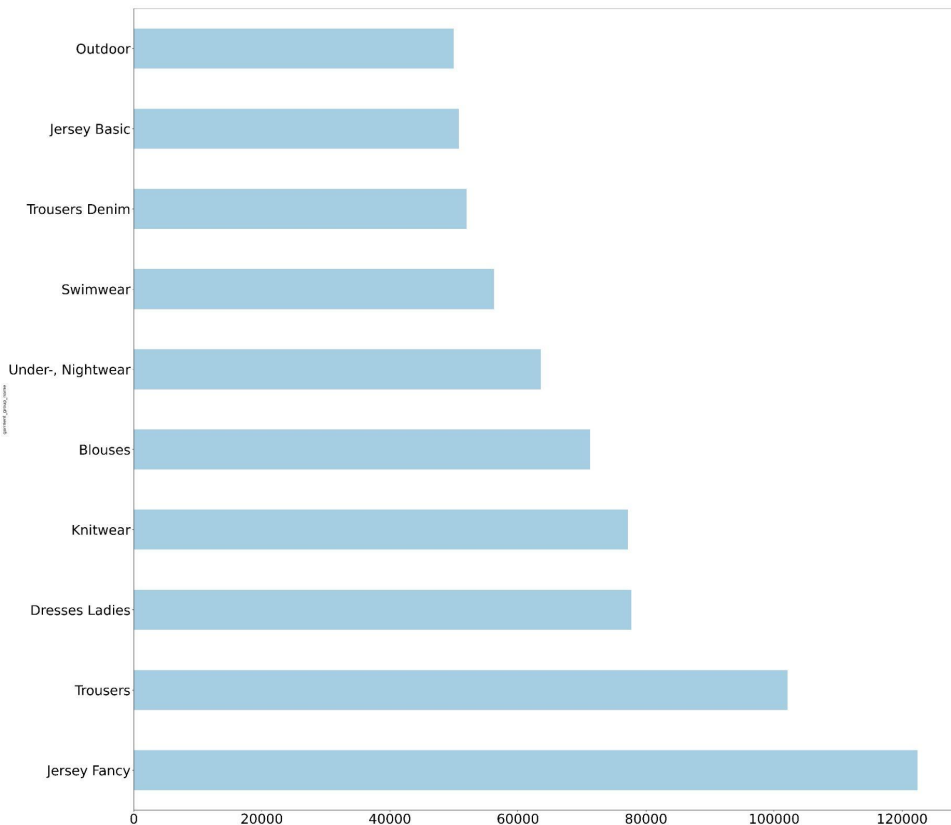
Avg. per Customer Sales



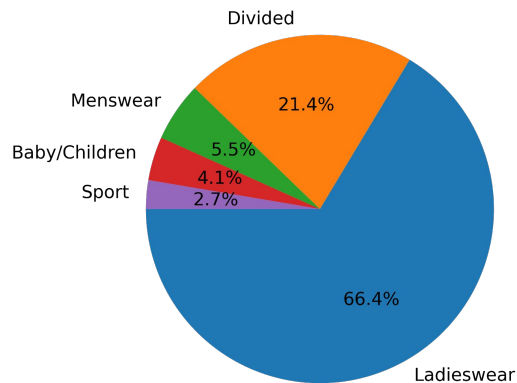


Our Products ✕

Top Sales by Garment Group (Sep 2018-Sep 2020)



Top Sales by Index Group (Sep 2018-Sep 2020)



Top Sales by Item (Sep 2018-Sep 2020)



Jade HW Skinny
Denim TRS



Jade HW Skinny
Denim TRS



Mariette Blazer

02x

Recommendation System

YOU MIGHT ALSO LIKE



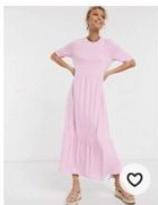
ASOS DESIGN shirred tiered maxi dress in mono sp...
£39.00



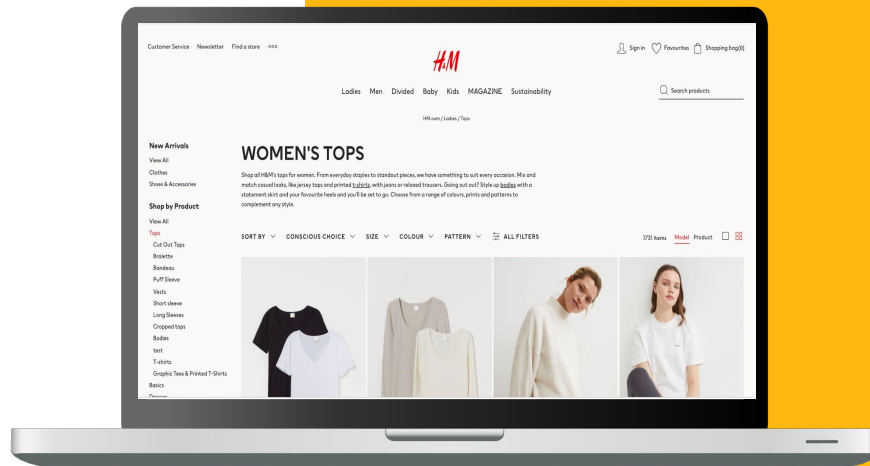
Vintage Supply midaxi smock dress with open back...
£50.00



ASOS DESIGN smock maxi dress in blue daisy print
£38.00



ASOS DESIGN tiered smock t-shirt midi dress in p...
£22.00





Recommendation System ✕

1

Content-based Filtering

Based on customers' previously purchased items

2

Collaborative Filtering

Based on user similarity

3

LSTM Predictive Model

Predict the coming hot items



On Products

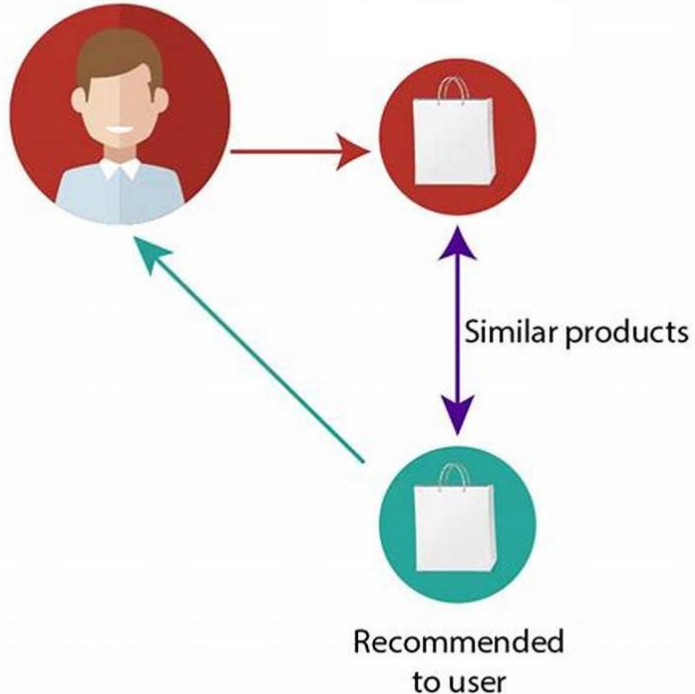


On Users



On Trend

Content-Based Filtering



Latest 3 months data with **customer id & customers' last purchase item** (size:525K)

Extract and dummy 6 item features from 28K unique products

Similarity Distance Measure By **Euclidean Distance**

Product group name

Colour value

Colour

Graphical appearance

Section

Garment

Collaborative Filtering

Alternating least square(ALS)

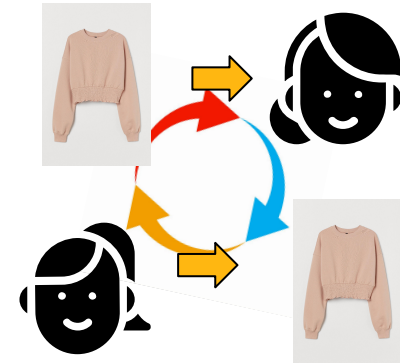


Dataset Characteristic:

- ❖ Implicit dataset (without explicit ranking)

Application:

1. Data Preprocessing
 - ❖ Take the latest 3 month data for model application (3,647,973x5)
 - ❖ Create additional column 'Frequency' - No. of times the same user purchased a specific item
 - ❖ Reassign customer_id from 0 to 530342 since matrix only takes numeric data
2. Model Building
 - ❖ Columns: Customer_id, Article_id, Frequency
 - ❖ Use of Implicit Library
 - csr_matrix (Compressed Sparse Row matrix)
 - Use of ALS to build model



LSTM Model



Find out the top 10 products for each day in 2 year and make a product_list

Filter out the the transaction data according to the product_list

Make dummies according to product_list
If product was top 10 of that day:1
otherwise: 0

Compile every 60 days data which are 3 dimensional arrays into the list

Build and Train the LSTM Model with
Epochs =10 Dropout=0.2

Input the last 60 days data in the model
and predict the top 10 products in the
future



DEMO On Streamlit

