

Predicting Trending Items with H&M Sales Data

Demand forecasting for fashion apparel retail sales.

Author: Wells Wang

Contact: wellswang0216@gmail.com



The Problem

Can “Trending” of products be quantified? If we can label, analyze, and predict “trending”, we can learn from it and add values to our product line.

The Solution

Binary classification model with trending items.

Demand forecasting + rules.



Potential Impact of the Solution



Increased Sales and Revenue



Improved Customer Satisfaction

Dataset Overview & Data Preprocessing

Customer

Age 

Membership 

Newsletter

Postal Code 

Article

Color 

Categories     

Detail Description

Transaction

Date 

Price 

Units 

Sales Channel 

Blazer 024
0241602018.jpg



Beanie 068
0682261020.jpg



Blouse 040
0407031005.jpg



Jacket 028
0282832012.jpg



Monthly transaction number from 2018 - 2019



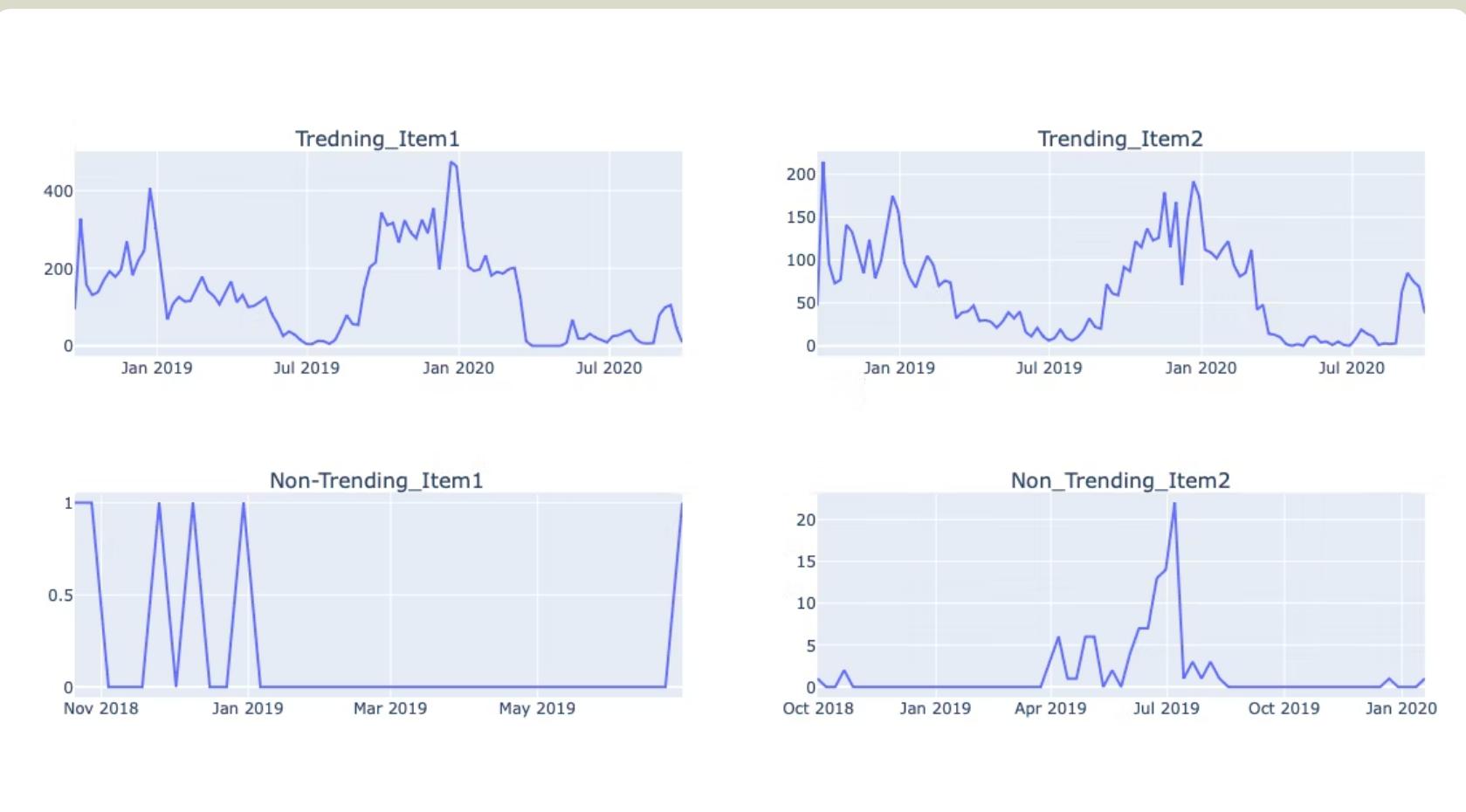


Trending is defined as observing a high volume of sales combined with high week-over-week growth.

We use specific thresholds to set precisely what "high" means.

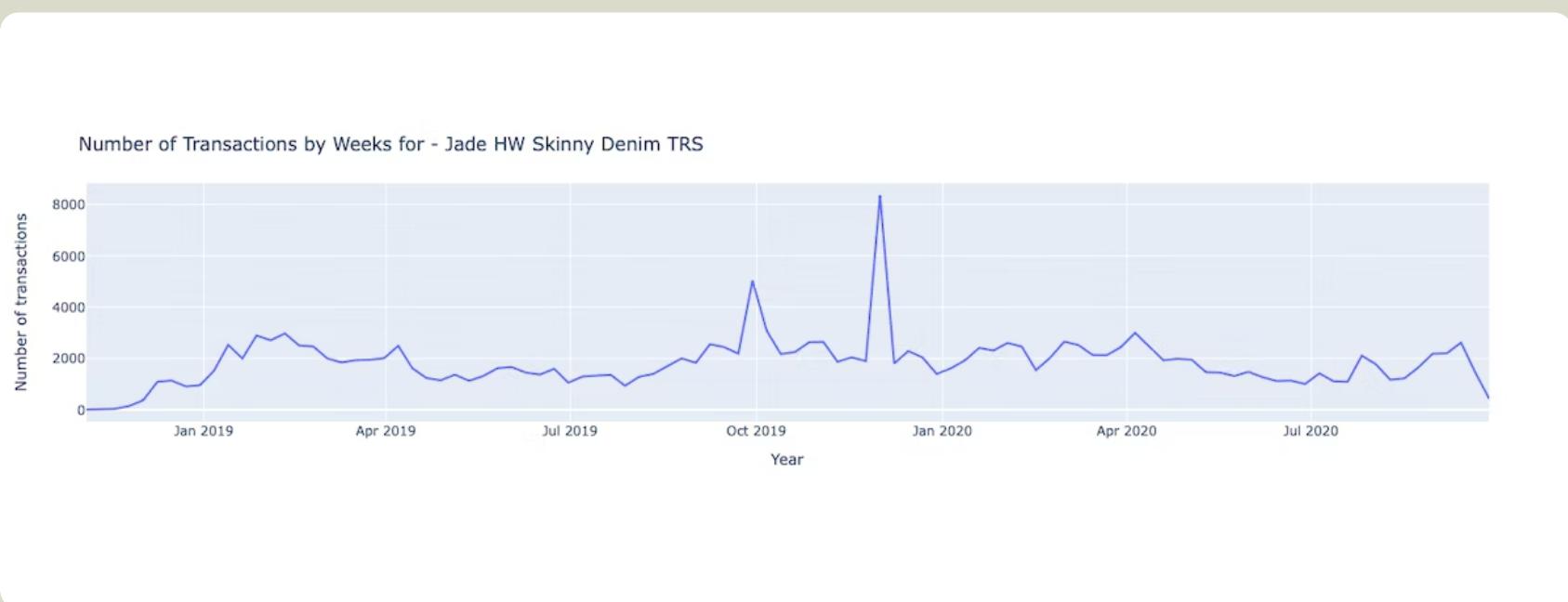
EDA

Trending Items
v.s.
Non-Trending
Items



Top Selling
Item

Number of
transactions by
weeks



Glimpse

Jade HW
Skinny Denim
TRS



Model Performance

We've selected logistic regression as our baseline model for this trending (binary) classification task. As we have an imbalanced target variable SMOTE method is also used for upsampling a more balanced training sets. The best model that optimize in favour of recall has a recall score of 0.87.

36K

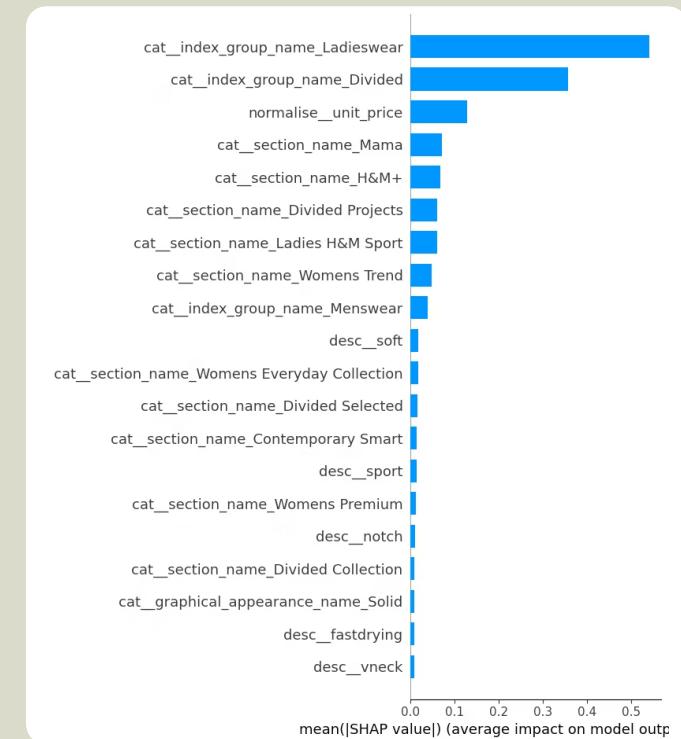
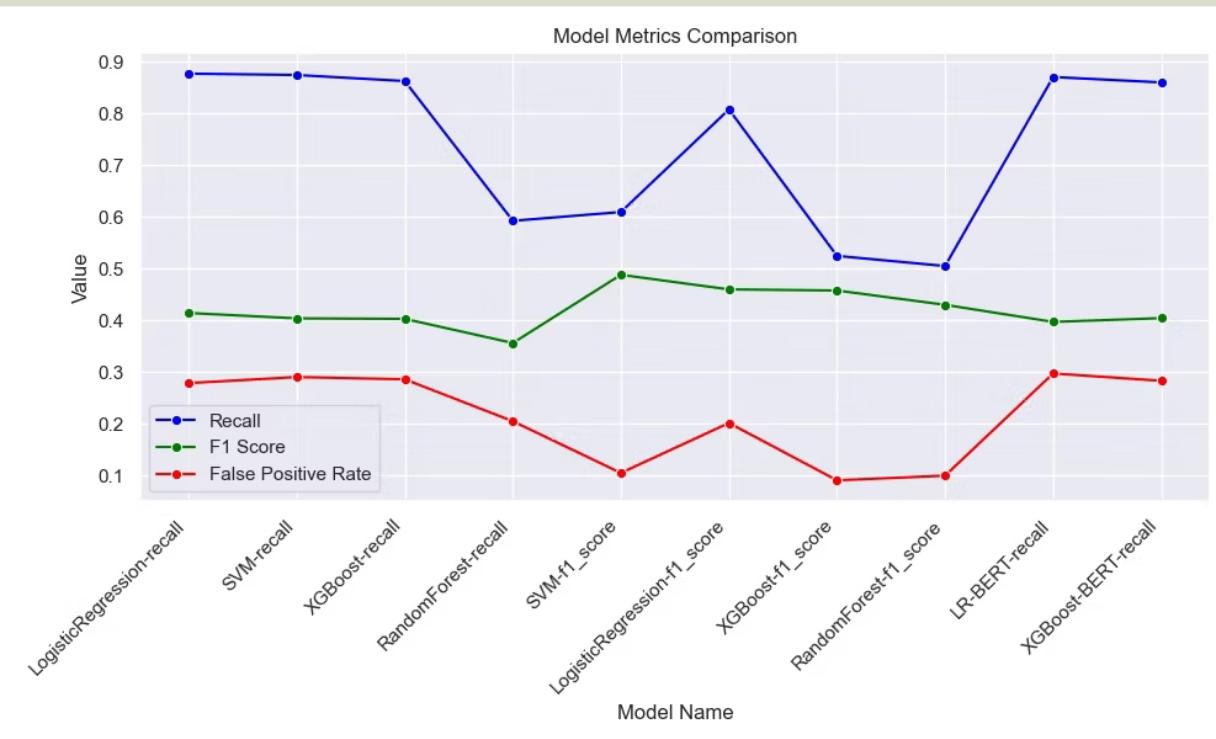
Products

0.4

F1 Score

0.87

Recall



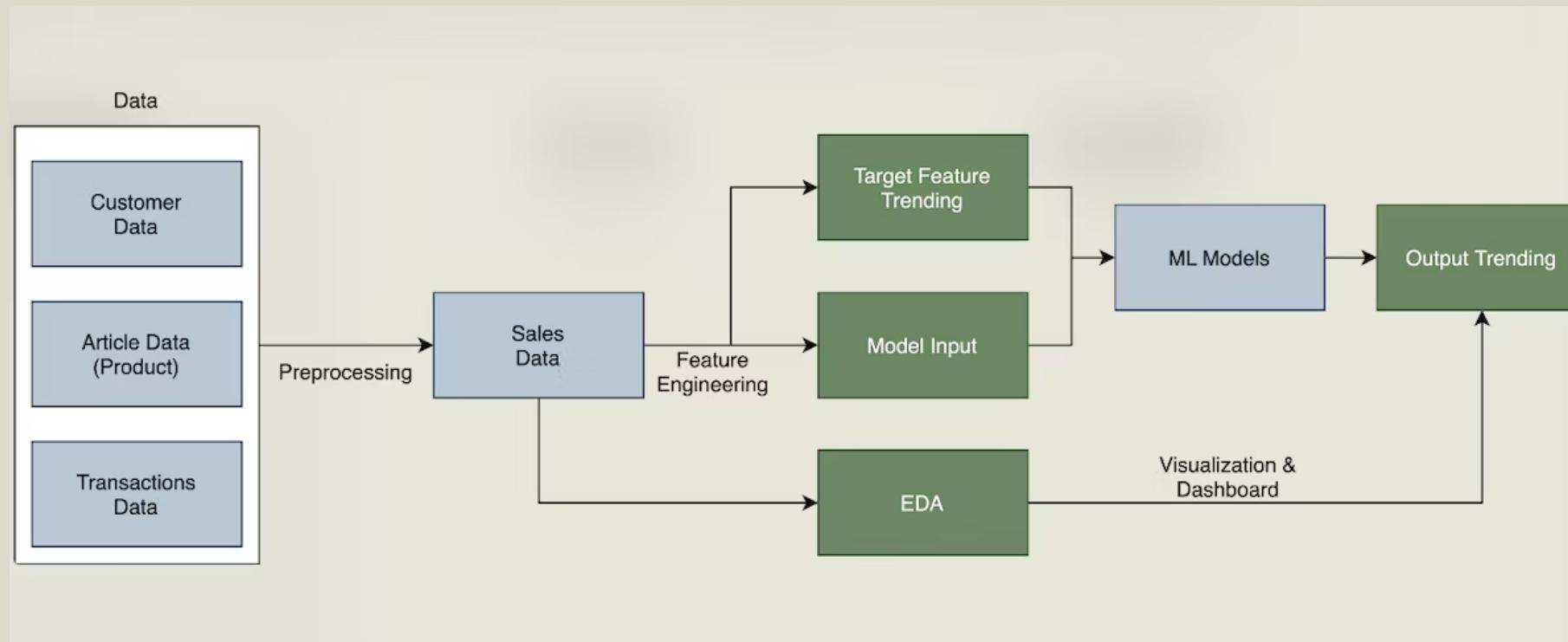
Model Interpretation

With shapely values, we created a BeeSwarm plot to highlight important relationships between features and impact on the model (predicting the trending target)



Pipeline

We built a pipeline that wrap up vectorization of product description and product name, one-hot encoding, smote, and modelling. With the pipeline, we can easily replicate the work on our desired case study. Furthermore, this could also be used when new data comes in for model updates.



Case Study

Vest top

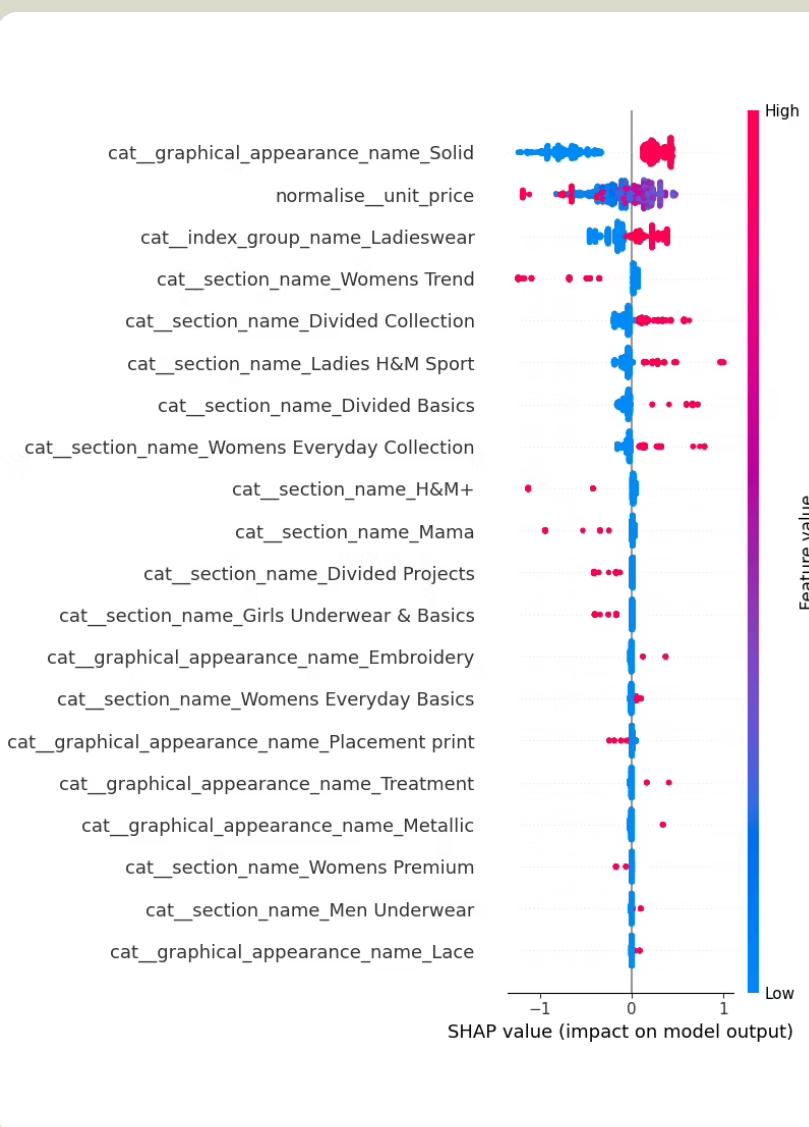
Product Type

0.72

Recall

0.50

F1 score



Vest top 010
0108775015.jpg



Vest top 010
0108775051.jpg



Vest top 019
0194270044.jpg



Vest top 019
0194270046.jpg



Vest top 010
0108775044.jpg



Vest top 019
0194270002.jpg



Vest top 019
0194270045.jpg



Vest top 020
0200761022.jpg



Trousers

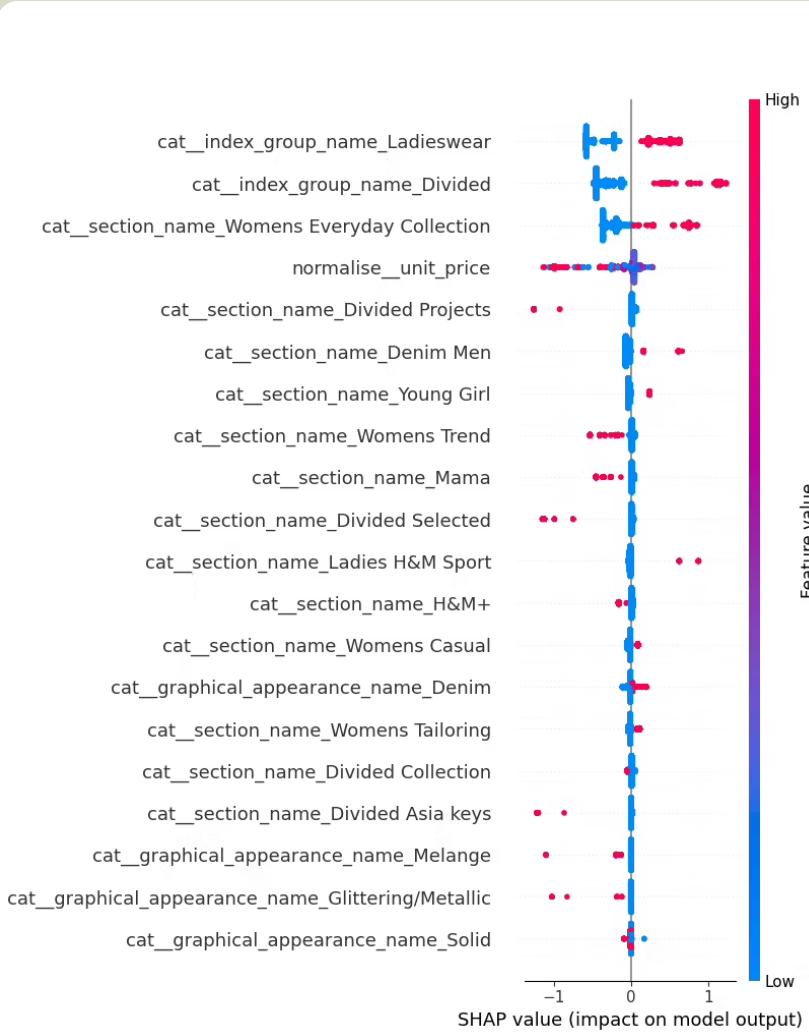
Product Type

0.92

Recall

0.45

F1 score



Trousers 011
0118458003.jpg



Trousers 011
0118458028.jpg



Trousers 011
0118458034.jpg



Trousers 011
0118458039.jpg

Trousers 011
0118458004.jpg



Trousers 011
0118458029.jpg



Trousers 011
0118458038.jpg

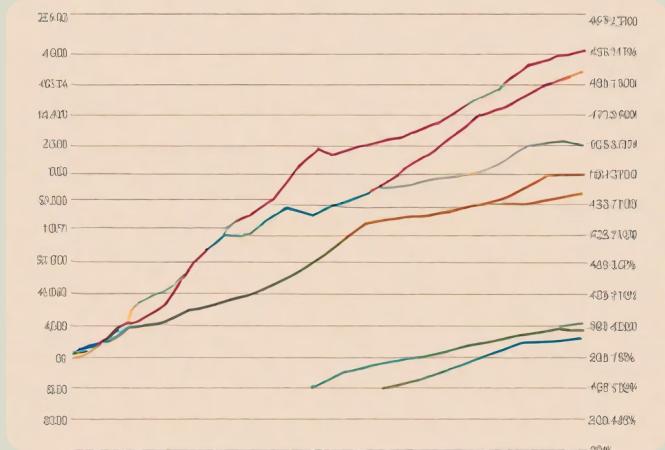


Trousers 011
0118458034.jpg



Trousers 015
0150959011.jpg

Next Steps...



EDA

Interactive Dashboard



Feature Engineering

More Complex trending (target)

Integrate trending identification to pipeline



Advanced Modelling

Time series analysis for peak detection

Representation learning

ML / NN for trending classification



Placeholder sample image

Demo



Wells Wang

Data Analyst | SQL, Python, Tableau

