

# Predicting Trending Items with H&M Sales Data

Demand forecasting for fashion apparel retail sales.

Author: Wells Wang

Contact: wellswang0216@gmail.com



### The Problem

**Fast-fashion companies struggle to control inventory. By predicting trending items within a few days of launch, companies can respond to high demands in inventory.**

The Solution

**Binary classification model with  
input of first n days sales.**

**Demand forecasting + rules.**



# Potential Impact of the Solution



## Increased Sales and Revenue

By predicting trending items after a certain period since launch, H&M can ensure that they have enough stock to meet demand. This will lead to increased sales and revenue for the company.



## Improved Customer Satisfaction

By having the trending items in stock, H&M can improve customer satisfaction by ensuring that customers can find what they are looking for. This will lead to increased customer loyalty and repeat business.

# Dataset Overview, Data Preprocessing, and EDA

## Customer

Age 

Membership 

Newsletter

Postal Code 

## Article

Color 

Categories 

Detail Description

## Transaction

Date 

Price 

Units 

Sales Channel 

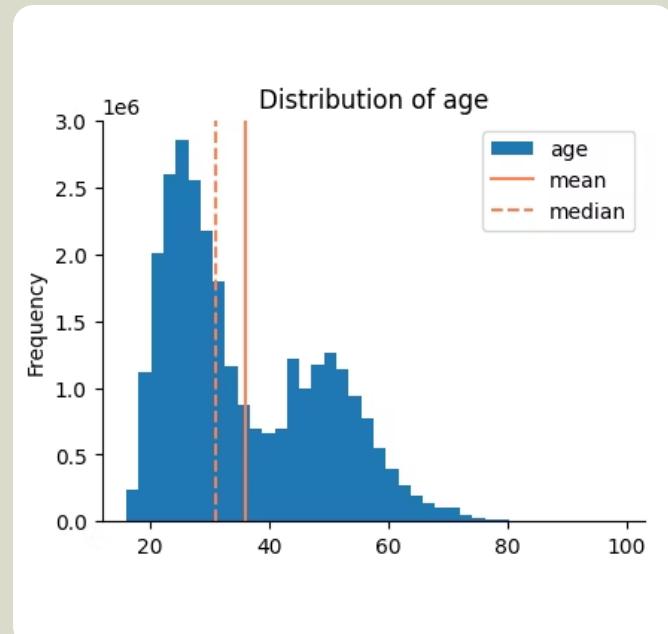
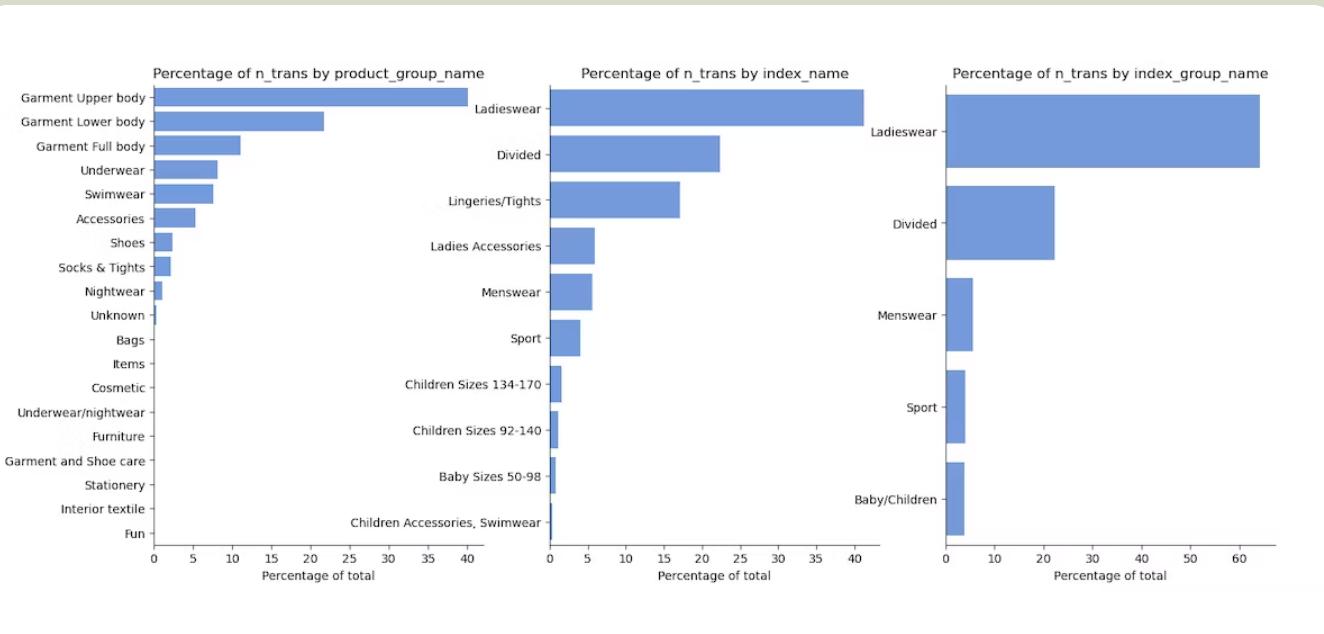
## Data Preprocessing

- Missing values and data type check
- Duplicate information among the columns
- Merging datasets → Shape: (28813419, 25)

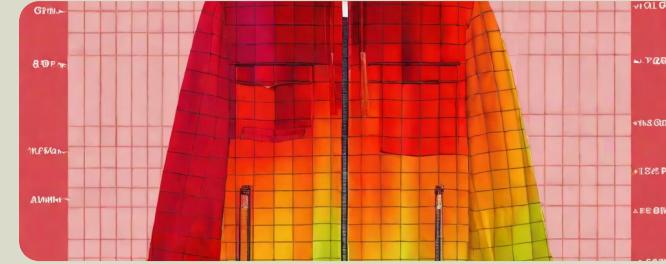
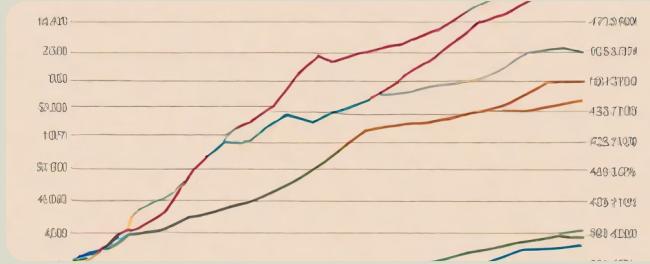
## Preliminary EDA

- Two age ranges can be identified with the most number of transactions: 23-33 and 45-55.
- We can spot a seasonal trend for the number of transactions by month, peaking in June each year.

Monthly transaction number from 2018 - 2019



# Next Steps...



## Data Processing

- We haven't thoroughly looked into all the categorical variables, maybe we can determine some key variables and drop those that are redundant after we figure out the hierarchies.
- We may want to look into individual product groups and see if there are differences in customer demographics. (Kmeans)

## Feature Engineering

- Transactions are only aggregated with the number of units of each article. It cannot tell if a customer bought multiple different items in one order. We can retrieve this information by grouping t\_dat and customer\_id. The assumption here is a customer would only make one order maximum each day.
- Hierarchies can be seen from categorical variables that are related to products and might need feature engineering.
- As we are doing demand forecasting for the project, we need to create a lag feature for the time series

## Baseline Modeling

The next step in baseline modeling will be to create a simple predictive model using the selected features and a basic algorithm.

- We might select logistics regression as the binary classification baseline
- We might select SARIMA as the demand forecasting baseline.

# Thank You



Placeholder sample image