**Data for ASOS UCL Graph NN Project**

The training data for this project consists of a sample of orders data from Sept 2021 and accepted returns data extending from Sept 2021 – Nov 2021. The test set contains orders for the following month, Oct 2021 and accepted returns from Oct 2021 – Dec 2021. The data is structured in three files: customer\_nodes consisting of basic (anonymised) customer account information including history of purchases and returns, product\_nodes with variant information and history of purchases/returns for the product, and an event\_table – listing the interactions (orders) between customers and products and whether or not the event lead to a return.

All students will have signed an NDA with ASOS.

**Customer\_nodes**

Schema:

* hash(customerId): Anonymised customerId
* yearOfBirth
* isMale: binary indicating whether customer isMale (0 = female)
* premier:
* salesPerCustomer: number of sales made by customer in the training period
* returnsPerCustomer: number of returns made by customer in the training period
* customerReturnRate: ratio of returnsPerCustomer/salesPerCustomer
* customer\_level\_(return\_code)\_ratio: a series of columns giving the proportion of returns made by the customer in each of the return codes (e.g. return\_code=‘Too Big’)
* country\_x: one-hot encoded column for each of the top 10 countries from which orders are made (e.g. country\_Germany, any country outside the top 10 are combined in ‘other’)

Date range:

* 1st September 2021 – 1st October 2021

Table

Description automatically generated

**Product\_nodes**

Schema:

* productID
* variantID : specific ID for product variant (for a particular product, there exists a different variant for each colour, size etc.)
* avgGbpPrice: average price for which the product was purchased (in GBP) over the training period
* avgDiscountValue: average amount (in GBP) that the product was discounted by when purchased
* salesPerProduct: number of times the product was purchased in the training period
* returnsPerProduct: number of times the product was returned in the training period
* productReturnRate: ratio of returnsPerProduct/salesPerProduct
* product\_level\_(return\_code)\_ratio: a series of columns giving the proportion of returns for the product in each of the return codes (e.g. return\_code=‘Too Big’)
* brand\_x: one-hot encoded column for each of the top 10 brands ordered in the training period (e.g. brand\_Nike, any brand outside the top 10 are combined in ‘other’)

Date range:

* 1st September 2021 – 1st October 2021

Table

Description automatically generated

**Event\_table**

Schema:

* variantID
* hash(customerId)
* isReturned: 0/1 label indicating whether the product was returned by the customer

(in each case, one item is bought/returned)

Table

Description automatically generated