**YOOX Net-A-Porter - Senior Data Scientist – Technical Exercise**

**Candidate Brief:**

Being able to identify customers at risk of lapse would enable us to take action to keep those customers shopping with us. You are asked to predict which of our customers will not shop with us again.

**Task:**

To create this prediction, you are given 18 months of shopping history (1989-07-01 – 1990-12-31) for each customer. Using this information we would like you to

- explore which variables contain information for predicting customer lapse.

- make a lapse prediction for each account

You will need to submit back:

1. A document outlining key findings, approach, and any limitations.

2. The code you write in generating your prediction. Please make this as readable as possible and suitable to run on a private test data set.

**Data Provided:**

You have been given two datasets:

**transactions.csv** – contains item level shopping history for all customers for a period of 18 months prior to the prediction date.

**account.csv** – contains account level data, and the target variable ‘lapsed\_next\_period’ (you are predicting the ‘lapsed\_next\_period’ column for each account in this file)

Below outlines the attributes contained in each datasets:

***Transactions***

**customer\_id** – A unique id representing each customer

**order\_id** – A unique id representing each order (a single order can contain multiple products)

**order\_date –** The date the order was placed

**var\_1, var2 –** Categorical variables related to the type of order

**product\_id** – A unique id representing each product

**product\_type\_id –** A categorical variable related to the type of the product.

**designer\_id –** A unique id representing the designer of each product.

**gross\_spend –** The value of the item purchased

**net\_spend -** The value of the item purchased**. Zero if the customer has return the item.**

***Account***

**customer\_id -**A unique id representing each customer

**var\_3, var\_4, var\_5 –** Categorical variables relating to customer shopping habits.

**var\_6 *–*** A continuous variable related to customer shopping habits.

**lapsed\_next\_period –** A categorical variable indicating whether the customer will lapse in the next period. **This is the variable you need to predict.**