

## Data Science Take-home Challenge

Thank you for applying for a Data Science position in Sainsbury's Tech. With this technical challenge, we want to understand how you approach one of the kinds of problem that you might tackle in Sainsbury's.

### The business context

This challenge focuses on one of our general merchandise brands, Argos.

The Argos range is advertised in a product catalogue. Historically this was a paper catalogue, however since 2020 the publication has been fully digital (the transition from print to digital has no bearing on the task and can be ignored).

New editions of the catalogue are published twice annually. Prior to publication of each new catalogue edition, the Argos merchandising teams conduct a product 'range refresh' to determine which products to retain, which products to discontinue, and which new products to include. At each range refresh we see about 30% of product lines replaced by new products.

As items are sold, Argos replenishes the stock of products in stores. Failing to replenish stock in stores leads to lost sales. However, there are also costs associated with clearing unsold, discontinued items.

Many products have low sales volumes in some locations (less than one item per month). Therefore, to make intelligent decisions about whether to restock items, it is important to know the likelihood of these products being discontinued at the next range refresh.

### The task

We would like you to develop a solution that could be used by the replenishment team to predict, at any given point in time, which products will be discontinued after the next product range refresh.

To solve this problem, you can use any library or tool as long as it belongs in the Python 3 ecosystem and it is open source/free for public use. Please return all the files that are necessary to invoke your model, and also all the notebooks or scripts that were used for the EDA and creating your model.

### How we assess your solution

As data scientists in Sainsbury's we are responsible for the end-to-end solution of a problem, from the scoping and exploratory data analysis (EDA), model development, all the way to building a final, production-ready solution that our engineers can attach to their pipelines. Therefore, we ask you to build a solution that reflects this end-to-end process, and to provide us with the required files to understand your development process and your solution. All steps are important.

We recognise that this is a challenging problem, and we don't expect a perfect solution. However, we are keen to understand your way of thinking and your problem-solving skills. We therefore ask that, as you develop your solution, wherever you make a simplifying assumption, choose a simpler method, or neglect some part of your codebase, you call out what you would have done differently when developing a 'real-life' product.

## The available data

Argos has a set of historical data at product-level for several previous range refreshes. It also has historical data for the product-level target variable. The target variable is binary (TRUE or FALSE) and indicates if a product was discontinued after the incoming range refresh.

This data has been shared with you in two files, plus a data dictionary. If you have any difficulty accessing the following files, please do contact us to allow the data to be shared with you:

- ProductDetails.csv
- CatalogueDiscontinuation.csv
- DataDictionary.xlsx

## Thank you

We want to thank you in advance for the time you invest in solving this challenge. We hope you enjoy working on it and look forward to reviewing your solution.

**Please let us know as soon as possible if you have any difficulties with completing the test within the given time window.**