Data Science Take-home Test

Thank you for applying for the Data Scientist position in Sainsbury's Tech. We are excited about innovation, solving challenging problems and adding business value across all our brands; Sainsbury's groceries, Argos, Tu and Habitat. With this take-home technical test, we want to understand how you approach one of the kinds of problem that you might tackle in Sainsbury's.

The business problem

The semi-annual refresh of the Argos product range sees about 30% of product lines replaced by new products. Throughout the year, Argos replenishes stock of products in stores as long as the expected sales revenue is sufficient to justify the investment. Some products sell less often than once per month in some locations. Therefore, in order to confidently evaluate the expected sales revenue of a product in a location, it is important to know the likelihood of it being discontinued at the next range refresh.

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.

What we ask from you

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 then 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.

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. While we are interested to see the final result, we are also keen on understanding your way of thinking and approaching the problem. All steps are important.

We would also like you to summarise your findings in a presentation (in a .pdf format) and attach it to the rest of the files. During the technical interview, we would like you to present your findings in a 10-minute presentation, as if to a non-technical stakeholder, e.g. the Head of Replenishment.

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:
ProductDetails.csv
CatalogueDiscontinuation.csv
DataDictionary.xlsx

If you have any difficulty accessing these files, please do contact us to allow the data to be shared with you.
This data set has been personalised for you, and can be identified. Please do not share this data, the problem, or your solution with others.