

# N Brown Data Science Recruitment Exercise - Size Prediction

November 8, 2018

## 1 Project context

We sell about GBP 3m worth of clothes each day across all our brands. About 15-20% of those come back to us as returns. Returns are very costly to the clothing industry – they have to be manually inspected. The main reason for customers returning items is that the garment was either too big or too small. Reducing the amount of returns would help our margins and improve customer experience.

**Could we help customers by predicting what size they would need?**

## 2 Project description

### 2.1 What information could we use to predict someone's dress size?

In your opinion, what factors/variables can help us predict a customer's size? Please provide justification for your suggestions.

Can you identify how easy it is to obtain the suggested data?

### 2.2 What can we learn from the data that we have?

We have provided you with a dataset `uc_data_train.csv`. It contains 80,000 rows and each row corresponds to a different customer. There are 14 features associated with each customer.

Using your preferred programming language or software tool, perform an exploratory analysis on this dataset. For example, investigate variable distributions, correlations, etc. Summarise everything we can learn from this dataset.

Can you come up with some rules that could help predict dress size? For example, "People over 50 tend to have larger dress sizes than those younger than 50." Summarise all your findings and observations in a report, which is aimed at a technical audience. Also, provide any code you have generated.

### 2.3 Can you predict customer's womenswear size?

Using the dataset `uc_data_train.csv`, come up with a method to predict womenswear sizes.

Test your method by applying it to dataset `uc_data_test.csv` and generating the missing womenswear sizes. Provide the updated csv file and any code you have generated.

What methods do you know for checking whether a prediction is good. How good is your prediction? Include your working and justifications in the report. State clearly any assumptions you may have made.

## 2.4 How can we use the new prediction?

Give your recommendation on how your new prediction can be implemented both from technical, and commercial perspective.

Do you think there may be any barriers to the successful implementation?

Are there any risks associated with the implementation?

## 3 Outputs

Please provide the following evidence for your work

- a report of your chosen format detailing all your solutions. The report is aimed at a technical audience. It should not exceed 4 pages;
- the updated dataset `uc_data_test.csv` containing your size predictions;
- any code you have generated.