**Data Science Evaluation**

There are three exercises in this evaluation, each of which is intended to help us understand your analytical strengths. We’re more interested in how you approach the problem and reach an answer than getting the absolute best score or solution, so feel free to apply your skills and knowledge as you feel is appropriate and be sure to document all of the steps that you took to reach a solution.

Although we designed these questions without a time limit, please record how long you spent on each part and include it in your submission.

Please submit your solution in a .pdf file- including any code, graphs and analysis.

**Detecting a Marketing Promotion**

In the data file named ***marketing\_promo.csv*** you will find data related to traffic and order activity over a period of several months. During this period a promotion offering free shipping was launched on the website. All visitors were exposed to the promotion and once started it remained active for all of the remaining days in the data you have available.

1. Based on the available data, deduce which day the promotion was launched.

**Clustering**

In the data file named ***cluster\_data.cv*** you will find a dataset, where each row represents a data point. The dataset is artificially generated.

1. Explain how you think the data was generated, including how many clusters are in the data and their centroid positions.

**Predicting Customer Spending**

In the zip file named ***user\_transactions.csv*** you will find a dataset containing the purchase history of five thousandcustomers who made their first purchase on Farfetch in 2012, along with a file named ***user\_dat.csv*** containing data about them*.* It can be assumed that the first order of the customer in the dataset was their first order on Farfetch.

1. Fit an appropriate distribution to the order values, detailing any steps you took to arrive at the result. Give the name and parameter values of the fitted distribution.
2. Produce a histogram of the order values and add your fitted distribution in Q1 to demonstrate the fit.
3. Predict the total spend of all of the customers in the dataset in 2014.
4. Explain the strengths and weaknesses of the approach you have taken in Q3 and the assumptions you have made. How could the prediction be improved?

*Note: it is not required that you use all of the supplementary data, this is provided to support ideas you may have for modelling.*