**Data Scientist Task 2 – Business Driver Impact Modelling**

# Introduction

The purpose of this task is to assess your ability to analyse a problem, build statistical models to capture the relationship between different actions and business metrics, create effective data visualisations, and present insights and recommendations that can shape future data products and drive future strategic decisions. For that, we provide you with a task that is very representative of the type of work we do here at ASOS.

# The problem

At ASOS we want to be able to understand the direction and scale of impact from different business drivers, both external and internal to the company, on the number of product units sold in different regions. As a dynamic business operating in a dynamic market, it is non-trivial to isolate the effect of a single driver, yet the ability to do so would shape decisions made in many areas of the business. For example, we could update our trading events to better stimulate demand and/or modify our warehouse processes to ensure there is adequate supply.

# The task

You are tasked with analysing and modelling the relationship between the drivers and the number of product units sold in one of our key markets. To do so, you are provided with the daily number of units sold in the market between Jan 2019 and Nov 2020, as well as information about various potential drivers of the market during the said period. While the focus of the task is to understand and model the relationship based on past events, any models proposed should also be able to project the expected number of units to be sold in the near future, guiding short-term business decisions.

You will present the work to fellow scientists and business stakeholders, who are interested in the recommended actions and their potential impact, as well as how you get to those numbers.

**As part of your submission, we would like to understand…**

* How you interrogated the data and what assumptions you made
* What kind of approach have you used and a discussion of its pros and cons, and any reason why a particular approach was suitable or not suitable in this case
* If the potential business drivers provided have any impact on the number of product units sold in the face of trend and seasonality fluctuations, and if so, the scale and direction of the impact
* How you evaluate the model's fit to both the past and future number of product units sold
* Based on the output of the model, what business recommendation(s) would you provide; and
* From what you know about the business, what other data sources would you use to enhance your model, and how would you set out to collect the data

# What are we looking for?

* Good management of time and resources – we prefer a basic, functioning solution rather than an incomplete but ambitious one
* Understanding of why such analysis/modelling is important and the business opportunities it could bring
* A solid analysis/modelling approach with a good explanation and evaluation strategy
* Visualisations that are capable of summarising complex relationships within the data
* Effective communication of both the technical analysis/modelling and final business recommendation, with an understanding of communicating to an audience of various technical abilities
* A discussion on what tools you have used for analysis, modelling and visualisation – the team primarily interacts with a Python-based code stack, but we are also happy to discuss solutions using R if you are more comfortable working with it for this exercise; and
* A discussion on the kind of experiments you would run to validate the trading decisions that are made based on outputs from the model.

**Preparation for a potential final interview:**

If you are invited for a final interview, we will spend some time during the interview going through your solution. We are interested in how you approached the problem instead of holding you against a set of marking criteria. To prepare for this please make sure you have sent us beforehand the following for review:

* The code and/or notebook of your solution and any analysis you have done leading up to it; and
* A presentation that you will be presenting back at the potential final interview.

**Some important remarks…**

* We are not looking for a model with the best fit to the past data or best predictive ability (although having such a model is a plus), but one that is also interpretable – one that can enable a decision maker to easily tell that making a trading decision X / event Y happened has caused, or is caused by, a Z change in number of units sold in the market.
* It is fine to make assumptions that may simplify your work. In fact, it might be necessary when it comes to communicating the results effectively. If that is the case, make sure you explicitly address them in the presentation you are submitting for review and have given a thought about what to do in the case that they were not correct.

# Data provided

**DS\_task2\_data.csv** (64 KB) – The CSV file contains a dataset with 28 columns, giving the daily number of product units sold and records of various potential business drivers between 2019-01-01 and 2020-11-17. These include measurements of the weather and governmental responses to the COVID-19 pandemic, as well as a log of some trading events and business operation metrics:

|  |  |  |
| --- | --- | --- |
| **Column name** | **Data type** | **Notes** |
| date | date |  |
| units\_sold | integer | Number of product units sold |
| GR\_stringency\_index | float | A score reflecting the stringency of the local government’s response to the COVID-19 outbreak. Adapted from [Oxford COVID-19 Government Response Tracker](https://www.bsg.ox.ac.uk/research/research-projects/coronavirus-government-response-tracker). |
| weather\_temperature | float | Average temperature of the region. |
| weather\_status | string | Indicator on whether the region is mainly cloudy, rainy, or clear. |
| marketing\_spend | float | Spend on marketing activities in the region. |
| delivery\_(standard|express)\_days | float | Number of days for standard and express deliveries to be processed. |
| trading\_global\_\* | binary  (0 or 1) | Trading events include (but are not limited to) sales, promotions, and discount code events.  Columns represents whether a corresponding trading event that are available to customers in all markets was held during the day. |
| trading\_local\_\* | binary  (0 or 1) | Whether a corresponding trading event that are available only to customers in the local region was held during the day. |