# Channel 4 Chief Data Scientist Technical Task

## Instructions



Whiskey contextual moment appears in Homeland @ 8mins 46 seconds

The next ad break at 11 mins features a Haig Club whiskey advert

## Introduction

The goal of this project is to identify moments in Channel 4 programmes that could be of commercial interest.

In this dataset, we applied a 3rd-party object recognition model to images extracted from video from Channel 4 programmes at 1-second intervals. In effect, we took a screenshot of the video every second, and fed it to AWS Rekognition for images. That model outputs continuous 'confidence' scores (between 0-1) for each of its 1000+ pre-defined object & scene types, e.g. Person, Palm Tree, Beach. These confidence scores are the *Input Features* to the wider model we're building.

We are interested in identifying moments in the video content that correspond to a few categories of interest to us, e.g. Alcohol, Food, Hot Drinks, Phones. These are the binary *Output Labels* for the model we're building. We have provided you with them here as ground truth labels for every observation in the dataset for training - in production, your job would be to infer them.

To recap: the goal of this project is to build a product that identifies when the Output Labels are present, e.g. a moment in time where alcohol was present on screen, based on the Rekognition confidence scores.

## Dataset

Download from:

- http://bit.ly/2ACYjjU

We have provided you with a CSV (sorted by `show_name` and `time_offset`). The most important columns are:

- `show_name` is the name of the Channel 4 programme, e.g. Hollyoaks, Fresh Meat - there is one episode for each show.
- `time_offset` is the timestamp (in seconds since the beginning of the television episode of that show).
- `r_abies` … `r_zoo` are the *Input Features*, produced by AWS Rekognition (see Introduction) for the screenshot of the video at `time_offset`.
- `h_alcohol_str` … `h_phone_str` are the *Output Labels* that you're trying to predict.

## Questions for you

1. Pick a single Output Label (e.g. Alcohol or Phones) and build a model that tackles it.
   - **For this, we are expecting you to write a small amount of code and run it**. We would prefer but don't require that you use Python. Please provide your code, though we probably won't run it.
   - **Describe in a couple of paragraphs (perhaps with a figure) what you were trying to achieve, how you went about it, your results, and your conclusions.**
   - **If you had to cut corners in your implementation, what were they? Why were they justified? What would you like to have done?**
   - For sharing with us, a Jupyter Notebook, website or PDF would be fine.
   - N.B. This is a big problem, and we know it's not feasible to make serious headway on it in a few hours, or to write production code - demonstrate your technical competence with a narrow, deep dive in one area to give us a hint of what you can do. Show that you can set it up as a machine learning problem, sensibly & correctly apply a relevant model, in good coding style, and communicate about it. Highlight how things could be improved.

**[When you have completed Question #1, please send it to Channel 4 via your recruiter, and then we'll be in touch to arrange a time for the discussion around Question #2.]**

2.  Imagine you're the Technical Lead on this project. **Prepare to discuss the larger project,** including its goals, pitfalls, the approach you might take etc. **You do not need to write anything down - we'll talk about it out loud.** For instance, we might talk about:
    - What questions would you ask the Product Owner?
    - How would you improve or extend this dataset?
    - What kinds of modelling approaches would you expect to work best and why?
    - How would you evaluate your success?
    - How would you put this into production?
    - How would you manage the project? What kind of team would you want for it? What milestones would you set?
    - What issues do you see that could make this project difficult? How well do you think the modelling will work?
    - How would you go about estimating revenue and costs for this over a 3-year period? What help would you need?
    - How would you decide whether you think this project is a good idea?
    - Any problems or confusions with this Technical Task?

N.B. This should take at most a few hours in total for Questions 1 and 2. If it's taking you longer than that, cut corners on the coding, and spend more time thinking & writing.

## Contact details

If anything is unclear or you have a question, you're welcome to contact Greg Detre <gdetre@channel4.co.uk>.