Data Science Apprenticeship   
Block Release Day November 2021  
Programming Against an Interface

# Introduction

The aim of this session is to give you experience in programming using an “software interface”, that is, where different teams write different pieces of code, with an agreed way of communicating between them. We will also use GitHub to manage the code.

# Getting into Teams

Get into a team of around 4 people. Once you have settled on your Team, let us know who is in your team and we will assign you a letter. Then, split into two sub-teams. One sub-team is called preprocessing, the other data analysis. You will work on separate parts of the code, but not share ideas until the end.

# Working on the Task

Clone the project at <https://github.com/colingjNottingham/TeamA> (or TeamB, or TeamC,… according to the letter that you have been assigned). Then, if you are in the preprocessing subteam, load *preprocessing.ipynb* into Jupyter Notebook and work through it. If you are in the data analysis team, load *dataAnalysis.ipynb* into Jupyter Notebook and work through it.

At the end, if you are in the preprocessing subteam, upload your *cleanedCustomerData.csv* file to the GitHub project for your Team. If you are in the data analysis subteam, upload the completed *dataAnalysis.ipynb*

# Final Comparison

The teams should now get together and see if the *dataAnalysis.ipynb* code created by the data analysis subteam works on the *cleanedCustomerData.csv* file created by the preprocessing subteam.