# Graduate Data Science Training Proposal

# Overview

Worley has the aspiration to be known as the leading digital company in energy, chemicals and resources by 2025. Training Worley graduates in data science will significantly contribute to this goal.

Data Science can be broken down into three main areas:

* Automation of processes
* Display and investigation of data (dashboards etc.)
* Using data to create models for prediction

The proposed training will cover each of these elements and the tools required to complete each element.

# Coding Language

Python will be the main coding language as it is Open Sourced, easily learned, has a large userbase and industry standard for data science.

# Budget

Assuming 40 graduates are interested in taking part in the course:

# Format

The format will be mixed with in-person training at 240 St Georges Tce and over Microsoft. Teams. Course Jupyter Notebooks will be hosted on Binder. This means time is not wasted installing Python and setting up the correct environment.

# Content

* Week 1 - Introduction to python
  + Python types – list, string, int, float, dictionary, set, tuple
  + Operations
  + Flow control – for, if, while
  + Error catching
  + Creating text files
  + Creating a function
  + Importing libraries
  + Classes
* Week 2 – Intro to Numpy + Matplotlib
  + Numpy array
  + Vectorised operations
  + Np.where
  + Axis object
  + Scatter graph
  + Bar chart
  + Histogram
* Week 3 – Intro to pandas
  + Creating DataFrame
  + Indexing DataFrame
  + Conditional Slicing
  + Groupby
  + Appling functions
  + Speed testing %%timeit
  + Project example – degradation assessment
* Week 4 – Intro to SciPy and Scikit learn, creating models
  + Cirstd, cirmean
  + Random forest model
  + OCR with tesseract
  + Image recognition model
* Week 5 - VBA
  + Language basics
  + Creating a macro
  + Debugging
  + Creating a toolbar