# SEAMS Data Analysis Exercise

You have been provided with a dataset – Events.csv. This represents a count of pipe asset failures over the past three years grouped by region, diameter, and age.

**Note that this dataset does not represent individual pipe assets.**

The data is a grouped summary of individual assets and has the following fields:

* Region – the region in which the assets are installed.
* Diameter – the diameter of the pipe assets (mm).
* Age – the asset age (years).
* Length – the **sum** pipe length (km).
* Events – a **total** count of failure events over the **past three years.**

Given this data, please perform the items at 1 – 3 below. **Throughout the analysis, please consider the failure rate (failures/km/year).**

1. Summarise the data. Help us to understand what affects the number of asset failures i.e. how the rate of failure changes by the different attributes.
2. Determine an approach for predicting asset failures. If possible, you should consider the impact of age, diameter, and region.
3. Prepare a short (around 10 minutes) presentation to communicate your findings.

You can use any software or technique to perform the analysis. Your presentation must be in a common format (PowerPoint, PDF, HTML etc.) such that we can easily display it during the interview.