# Data Science Skills Assessment

Greetings budding developer. This document describes the assessment to see where your strengths in lie, so that we can assess your fit.

In this scenario, we have been hired by an electricity network to help process Drone based LiDAR data of their electricity lines. Your task is to create a Python package that can identify how many wires are present in a LiDAR point cloud and generate 3D catenary models of these wires.

## Project tasks and deliverables

For this project, you are given LiDAR point cloud datasets (in .parquet files found [here](https://enea.egnyte.com/fl/FPjBkTV4MYh3)) and are tasked to generate best fit Catenary models on each wire within the dataset. As a reminder, a Catenary is a curve that describes the shape of a flexible cable subject to its own weight. It is described by the following equation:

Where c is a curvature parameter, x is the distance along the wire is the x value of the trough, y is the elevation of the wire, and is the lowest elevation of the wire.

Note that each of the lidar files present different issues to work through to improve your algorithm. Also note that the above equation is 2-D whereas the point clouds are 3-D, additional parameters will be required to adapt the above.

We encourage you to use your creativity for this task but here are a few suggested steps to get started:

* Cluster the point clouds per wire, you can use libraries such as scikit-learn to do so.
* Find a plane of best fit for each wire cluster.
* Find the best fitting catenary within the above plane.

This code must be shared via a GitHub, GitLab, or Bitbucket repository. Along with the code, we would like you to write a short summary in the README of your repo documenting the steps of your algorithm.

As a summary, we are looking for you to:

* Implement a solution as a Python package.
* Ensure it is performant, maintainable and reusable
* Share your work via a GitHub / GitLab / Bitbucket repository
* Include instructions on how to set up and run your code

We will then ask you in the interview to walk us through your solution and results and will ask various follow-up questions to understand your approach

## How you will be assessed based on your code

We will be focusing on your coding abilities as well as your knowledge of the steps in a software pipeline. Notably we are looking for:

* **Code quality and organisation:** Can we easily tell what is going on in your code via appropriate module organisation and documentation?
* **Reproducibility and ease of running your solution:** Can we run your code without any issues?
* **Software engineering practices:** Are you using best practices to make your code easier to share and maintain e.g. structure, maintainability, version control.
* **Performance and scalability:** How easy it would be to run the code at scale.
* **Clarity of communication *–*** in your documentation and presentation.
* **Approach:** How you approached this technical problem, and the logical structure of the analysis.