Capstone Project 1: Project Proposal

Based on the feedback on my initial ideas and discussing them further with my mentor, we picked **Pump it Up challenge** from **Drivendata** as my Capstone Project 1.

- **What is the problem?**
  In this challenge, we are using datasets provided to us by Taarifa and the Tanzanian Ministry of Water to predict which pumps are functional, which need some repairs, and which don't work at all.

- **The client and why do they care about this problem:**
  The client is Tanzanian Ministry of Water and a smart understanding of which waterpoints will fail can improve maintenance operations and ensure that clean, potable water is available to communities across Tanzania.

- **What data are you using? How will you acquire the data?**
  The data is provided to us by Taarifa and the Tanzanian Ministry of Water Dataset (Train values, Train labels and Test values) are loaded on URLs below:
  - Train values: http://s3.amazonaws.com/drivendata/data/7/public/4910797b-ee55-40a7-8668-10efd5c1b960.csv
  - Train labels:http://s3.amazonaws.com/drivendata/data/7/public/0bf8bc6e-30d0-4c50-956a-603fc693d966.csv
  - Test values:http://s3.amazonaws.com/drivendata/data/7/public/702ddfc5-68cd-4d1d-a0de-f5f566f76d91.csv

- **My approach to solve this problem:**
  My approach is to find those features which would be important in predicting the result. For example, the location and the population around any particular water point would play a vital role in the status of that water point. Also, there are some features with missing values and I need to be deciding whether to drop that feature, fill the missing values with the mean and median, or to assign them to unknown category.

- **Deliverables?**
  I'll be writing my code in Python (Jupyter Notebook) and I will also have a slide deck to present my findings.