# ANI Data Engineering: Capstone Project Brief

## Project Outline

This training course will help you to understand modern tools & techniques for storing, processing, analyzing and visualizing data. The broad aim of the capstone project is to exercise these tools & techniques to communicate actionable insights or "tell a story" through data. The theory and practical steps required to achieve this will be covered during the training.

The project will be undertaken as a team, and so a significant aspect will be to communicate and work effectively as a unit. Teams will be assigned at the beginning of the course. Requests to be on a team with specific colleagues will be considered and facilitated where possible.

The data used can be any work appropriate dataset, either public domain, or taken from a work scenario. Bear in mind if considering using Allstate data, that there will be appropriate approvals required to use Allstate data in this context. These approvals can take a significant amount of time, and so can delay getting started on the project. There may also be some obfuscation or anonymization of the data required, which can result in difficulties using it effectively.

It is expected that all project groups do some initial exploration to identify potential datasets. There are many sources of public domain datasets that you can make use of, for example:

- Datasets from Kaggle https://www.kaggle.com/datasets

- Datasets from Quandl https://www.quandl.com/

- The world economic outlook database
https://www.imf.org/external/pubs/ft/weo/2020/01/weodata/index.aspx

Often, interesting insights can be developed by combining information from multiple datasets and sources. This allows unique insights to be gained, while also throwing up interesting real-world challenges in managing and processing data from multiple sources.

In each "off" week a session will be scheduled to cover capstone project activity. In these sessions, guidance will be given on the high-level goals for that stage of the project. Time will be allocated during these sessions to discuss project activity with your instructors and work through any issues.

The final activity of the training course will be to present your work to instructors and colleagues. Each group will be given 20mins of presentation time, with further time allotted for questions from the audience. These presentations will be in a separate session at least a week after the final training module has finished.