

# Data Engineering Capstone Project – Week 2 Activities

---

## Objectives

The target project activities should be:

### **1. Continue With Selecting Dataset(s)**

As a team you should have at this point identified a broad topic for your project.

You may be continuing to research available datasets to suit the topic. An option to determine whether a dataset is suitable is to load it, or a subset of the data, into Power BI and perform initial explorations of the data through Power BI only.

### **2. Store Selected Dataset(s)**

Any dataset(s) that you have selected you should store in a database system. You can choose MySQL, Oracle, MongoDB or anything else.

You may use this as an opportunity to revisit some of the ideas around relational data modelling, or MongoDB query language.

If using MongoDB, you may choose to interface to it through VS Code rather than, or in addition to, using MongoDB Compass.

When the dataset is stored in a database, you should connect Power BI to that database and load the data, or a subset of it into Power BI directly from the database.

### **3. Continue Hypothesis/question selection**

To analyze the data, some common approaches to initial exploratory data analysis are:

1. Define some hypotheses about the data that you will attempt to prove true or false
2. Define some questions about the data that you will attempt to answer
3. Define some metrics about the data that you will attempt to measure

You may begin the tasks above through Power BI, SQL, MongoDB queries or otherwise. We will push this further with some of the analysis tools in Modules 3-5.