

Data Engineering Capstone Project – Week 1 Activities

Objectives

The target project activities should be:

1. Identify A Broad Theme

As a team you should identify a broad topic that you think would make sense for your project. Sometimes the best approach to deciding on a topic can be to identify some questions or hypotheses that you will attempt to explore or answer.

You may choose to use data related to your day-to-day activities. This can produce a very valuable project with potential for future use. However, take consideration for any potential delays related to data access approvals.

A simple example for demonstration:

Property Prices Vs Happiness

We will explore the question of whether there is a correlation between self-reported happiness in a country/region and house prices in that country/region. Some questions we will attempt to answer:

- 1. Are property prices correlated with reported happiness*
- 2. Are there other factors that may also correlate with either or both*
- 3. Is there a difference in this correlation across geographic regions*
- 4. Does the correlation change below or above a certain level of GDP*

2. Identify One or More Datasets

This will feed into the decision of a broad topic. You will need access to datasets that will facilitate the topic that you choose. Similarly, evaluating the type of data that is available will likely help to give ideas on what topics might suit.

3. Initial Exploration of Data

Once you have an idea of what datasets might be good candidates for the project, you should try to perform some initial exploration of the data with Power BI.

This can be a good opportunity to reinforce the Power BI material that was covered, while also gaining some insight as to the quality and depth of the data you are dealing with.