MEASURES OF CENTRAL TENDENCY

When starting to analyze data, the first question asked is always "where is the center?".

There are three different ways to measure this:
- **Mean**
  - Also known as the '**average**'
  - It is the sum of all data, divided by the number of data points
- **Mode**
  - The '**most frequently occurring number**' in the data set
- **Median**
  - If the data points are ordered from smallest to largest, the **median is the number that is in the middle**
  - It is the point that separates the higher half of the data from the lower half

When would each measure be used?
- **Mean**
  - If the rules for mode and median do not apply, 'mean' would be used
- **Mode**
  - When there are categories that are being compared
- **Median**
  - When a sample of the population is being measured (when measuring only occurs for a few weeks)

Why is this important?
The measure of central tendency will give a baseline to reference against as the project progresses. This baseline will be compared against the improvement measure so it is easy to see if the change has been an improvement.

Example:
Suppose the following sample is in the data set: 2,1,3,2,9,7,4

Mean: 2+1+3+2+9+7+4 = 28; 28/7 data points = 4   The mean is 4
Mode: The most frequently occurring number is '2'. It occurs twice, while all the others occur only once
Median: First, order from smallest to largest; 1,2,2,3,4,7,9; the number '3' is the middle point, so it is the median

The measures of tendency are 4, 2 and 3. This was a small sample size, only 7 data points, and that with a larger sample size, the mean, mode and median will tend to move closer together. For very large sample sizes, they become almost indistinguishable.