**Data Science Task**

You are trying to determine the 7-year survival of prostate cancer patients. A patient survived if they are still alive 7 years after diagnosis. This means that a patient is counted as dead whether or not the death was due to their cancer. You have been given details about the patients and their cancers to help you with your prediction.

**DIRECTIONS**

You have been given two datasets. The set labeled ‘train’ has details of patients, the state of their cancer at time of diagnosis, and some information about the progression of their disease. You will use this data to train any models or create any rules you consider relevant. The second data set, labeled ‘test’, is the set you will predict and submit to finish. There is also a file called ‘Data Dictionary.xlsx’ which contains information about the data in the dataset.

Here are some specific instructions:

* The test data:
  + Do NOT change the starting column names. If you do we cannot score your data
  + Final data set must be in the form ‘.csv’.
  + In the columns 'survival\_1\_year' and ‘survival\_7\_years’, 0 means dead and 1 means alive/survived
  + Populate the column ‘survival\_7\_years’ with your prediction. This must be a 0 or a 1, not a probability.
* Send us any code you used to train and test your models
* A description and explanation of your prediction process and your final model.
  + This can be in any form, but a person looking at your explanation and your code should be able to understand both the why and the how of your approach.

We will measure the performance of your model/rules by accuracy. In other words, we will take predictions from your scored data and compute the following: