**Final Project, Part 1: *Proposal***

David Susman

10/9/18

General Assembly Data Science

Hospital organizations are conscious of their patient admission numbers, but more so if a patient is readmitted within 30 days of begin discharged. Reimbursement from both government and private payers as well as ratings for their organization can both be affected by their readmission rate. Due to that hospitals are being more proactive in attempting to prevent patient readmissions, especially with populations that can be maintained easily. One of those populations of interest are those patients that have being diagnosed with diabetes.

An important and standard metric in monitoring the diabetic population is ensuring that they have had at least one hemoglobin A1c (HbA1c) measurement within the past year. The purpose for this in to ensure that the level for the diabetic are within a certain range, where based on the result may lead to changes in medication, dosing and further care. If a patient is maintaining their A1c levels, once diagnosed with diabetes, the cost for a hospital is minimal versus the much higher expenses around hospital readmission. I hypothesize that using various clinical data point such as the testing and results of the HbA1c test will reduce hospital readmission rates for the diabetic population.

At IBM Watson Health we used predefined algorithms to define rates of readmission, and apply them along a specific organizations population. My goal is firstly to prove the positive impact that a simple test can have on readmission rates, as well as I hope to create a model to calculate risk around possible readmission cost and rates particularly around diabetics. I am working with my companies offering team to see if I can define and create my own deidentified patient data set, based on a parameters specific to patients with diabetes, displaying their clinical, demographic and financial data. I am not entirely sure if this will be possible at the moment, but I hope to have this wrapped up by the end of the week. As a backup though I have found a clinical dataset defined as a diabetic population (diabetic\_data.csv) that may need to be curated a bit with what appears to be a decent amount of missing data points. It also does not have any data around the cost associated to the hospital as well as does not go in patient history all that in depth. It also comes with a data dictionary to define a few of the non-standard code sets in the file IDs\_mapping.csv.