##Project Question##

**What is the question you hope to answer?**

Can we forecast the onset of Diabetes Mellitus in native female Pima Indians?

**Corollary questions**

Given eight variables that have been recognized as risk factors for Diabetes in native female Pima Indians, can we make a determination of the onset of the disease?

*What are the eight variables?* The eight variables of concern are:

1. Number of times pregnant
2. Plasma Glucose Concentration at 2 Hours in an Oral Glucose Tolerance Test (GTIT).
3. Diastolic Blood Pressure (mm Hg).
4. Triceps Skin Fold Thickness (mm).
5. 2-Hour Serum Insulin Uh/ml).
6. Body Mass Index (Weight in kg / (Height in in)).
7. Diabetes Pedigree Function.
8. Age (in years)

**What data are you planning to use to answer that question?**

The population for this study was the Pima Indian population near Phoenix, Arizona. That population has been under continuous study since 1965 by the National Institute of Diabetes and Digestive and Kidney Diseases because of its high incidence rate of diabetes. Each community resident over 5 years of age was asked to undergo a standardized examination every two years, which included an oral glucose tolerance test. Diabetes was diagnosed according to World Health Organization Criteria; that is, if the 2 hour post-load plasma glucose was at least 200 mg/dl (11.1 mmol/l) at any survey examination or if the Indian Health Service Hospital serving the community found a glucose concentration of at least 200 mg/dl during the course of routine medical care. The data set (found [here in Kaggle](https://www.kaggle.com/uciml/pima-indians-diabetes-database)).

**What do you know about the data so far?**

Small dataset of 8 variables and about ~750 rows. I chose a relatively small dataset since this is my first real coding/data science project.

**Why did you choose this topic?**

A close family member suffered from Diabetes and was very stressful to manage (daily insulin shot, careful diet etc). Would love to develop an algorithm to better diagnose disease.