This dataset is originally from the National Institute of Diabetes and Digestive and Kidney Diseases. The objective is to predict based on diagnostic measurements whether a patient has diabetes.

Several constraints were placed on the selection of these instances from a larger database. In particular, all patients here are females at least 21 years old of Pima Indian heritage.

**DATA OVERVIEW:**

* Pregnancies: Number of times pregnant
* Glucose: Plasma glucose concentration 2 hours in an oral glucose tolerance test
* BloodPressure: Diastolic blood pressure (mm Hg)
* SkinThickness: Triceps skin fold thickness (mm)
* Insulin: 2-Hour serum insulin (mu U/ml)
* BMI: Body mass index (weight in kg/ (height in m) ^2)
* DiabetesPedigreeFunction: Diabetes pedigree function
* Age: Age (years)
* Outcome: Class variable (0 or 1)

### **Objective:**

* Understand the Dataset & cleanup (if required).
* Build classification models to predict based on diagnostic measurements whether a patient has diabetes
* Also evaluate the models & compare their respective scores like R2.
* Remove the target feature from the dataset and build a clustering model to cluster the data.