HHA 550 Applied Healthcare Analytics

**Python – Group Project Requirements**

Scenario:

You have been assigned a new project from the Director of Clinical Informatics. She would like you to and your team to use the dataset provided to determine the leading variables of readmission so that the hospital can start taking corrective action and reduce the 30-day readmission rate.

Dataset:

<https://archive.ics.uci.edu/ml/datasets/Diabetes+130-US+hospitals+for+years+1999-2008>

Objectives

1. Create four groups of students
   1. Each group must email the professor with the students names in each group
2. Download the data
3. Clean the data
   1. Perform the ETL process to normalize the data
4. Put the dataset into the Python program
5. Using the dataset provided find the fewest variables out of the dataset that will be required to predict readmission
   1. Show what variables they are and why they are important
      1. Correlation
      2. Linear regression
      3. Logistic regression
      4. ANOVA
      5. Multilevel linear modes
      6. Clusters
      7. Etc…
   2. Show your work in the presentation
      1. Code
      2. Algorithms
      3. Visuals
      4. Anything else to support your findings
6. Present your findings and how you came to your conclusion
   1. Everyone must participate in the project and the presentation

Presentation Audience:

Executive IT and nursing staff that have a basic understanding of informatics and healthcare data but need to know in the basic of terms how you came about your findings so that they trust what you are presenting

Special Notes:

You and your group will be presenting all 3 projects at the same time (infrastructure, Python, Tableau)