CUNY MSDS DATA 608 – Final Proposal

Nicholas Schettini & Niteen Kumar

**Data Source**

The UCI Machine Learning Repository: Diabetes 130-US Hospitals for years 1999-2008 Data Set contains 100000 instances of data, with 55 attributes.  The dataset represents 10 years of clinical data at 130 US hospitals. The data contains attributes such as patient’s race, gender, age, time in hospital, number of lab tests performed, number of in-patient, outpatient, and emergency visits for prior year, etc.

<http://archive.ics.uci.edu/ml/datasets/diabetes+130-us+hospitals+for+years+1999-2008>

**Background**

Hospitals are penalized when patients are discharged and readmitted within 30 days.  Readmissions are associated with negative patient and financial outcomes. A readmission can be defined in multiple ways, including:

* Patients who are readmitted to the same hospital, or another applicable acute care hospital for any reason.
* Readmissions to any applicable acute care hospital are counted, no matter what the principal diagnosis was.

Using the dataset, we will attempt to create an interactive EDA tool that could help a hospital (business) to better understand their patient population, and what factors might determine if a patient will be readmitted.

**Tools**

We will create an app using Plotly Dash to create interactive visualizations of our dataset. The visualizations will include various forms of EDA – including:

* Patient demographics (age, sex, etc.)
* Medication history and how medications impact a patient being readmitted to the hospital
* Patient medical history