Project Proposal

September 30, 2019

Group member: Jia Tang (jt854), Yijia Cui (yc2576)

Problem Introduction

In this project, we will determine the best-fit hospitals for patients. In order to do so, we will examine the patients' medical conditions and their social economic status, as well as the hospitals' conditions. To be more specific, we will first look into the patients' age, diagnosis, severity of symptoms, and the level of emergency when they are committed to the hospitals. We will also look at the patients' race, ethnicity, and whether or not they have medical care or health insurance. Then we will examine the length of stay and the average costs of patients with different symptoms across different hospitals. We will investigate the possible relationships between these aspects.

One of the significance of our proposed project is that it helps patients with their choices of hospitals. It can save the patients' time and energy by providing them with data support for their choices. Moreover, our proposed project will also improve the efficiency of hospitals. By classifying the hospitals, this project is able to bring to the hospitals more patients whose symptoms match the hospitals' specialty, and less patients whose symptoms do not, which enable us to

Description of Dataset

In this project, the dataset we will use is the SPARCS Hospital Inpatient Discharges dataset for hospitals in New York State. Providing the basic information of all patients, this dataset includes the age, race, gender and ethnicity of the patients discharged from 2012. The data set also provides information about length of stay, type of admission, patient deposition, CCS Diagnosis, severity of illness, risk of mortality, payment method, which allows us to the possible relationship between them clearly. Since the dataset also includes the data of total charges and total cost corresponding to each patient, we can also investigate the costs covered by insurance, which can help us to investigate which hospital saved money best. Furthermore, the dataset also allows us to compute other features, including ratio of severity of illness, with 1 representing minor and 4 presenting extreme, to length of stay for each patient in corresponding hospital.

Link to data:

https://health.data.ny.gov/Health/Hospital-Inpatient-Discharges-SPARCS-De-Identified/u4ud-w55t/data