**ETL Project Report: Sepsis and Hospital Data**

**Team members: Karenetta Sullivan**

**Data Sources:**

*Hospital Data*

[*https://data.medicare.gov/data/hospital-compare*](https://data.medicare.gov/data/hospital-compare)

Scraping Data

<https://www.cdc.gov/sepsis/what-is-sepsis.html>

<https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/HAC/Hospital-Acquired-Conditions.html>

**Summary:**

Pulled CSV files for hospital data from the data Medicare website for complications and deaths in hospitals. I decided it would be interesting to link the data and review the hospitals with the highest number of Sepsis deaths in the Chicago area.

The goal of this project is to create a SQL table and join the data using SQL Alchemy.

**Technical Steps (please see the Python notebook attached with all our codes in it):**

**Extract:**

I downloaded three CSV files from the Medicare.gov website.

**Transform:**

I dropped columns from the hospital data files, the columns were also contained in the state data file. I then merged the hospital and state files using an inner join on the state column.

**Load:**

A hospital database was created in MySQL and I used the import wizard to store all data csv files.