Project Proposal

**1. What are the names and NetIDs of all your team members? Who is the captain? The captain will have more administrative duties than team members.**

Tony Law | [tonylaw2@illinois.edu](mailto:tonylaw2@illinois.edu)

Morgan Whitman | [whitman5@illinois.edu](mailto:whitman5@illinois.edu)

Aareana Reza | [areza5@illinois.edu](mailto:areza5@illinois.edu)

Team Name: TAM

**2. What is your free topic? Please give a detailed description. What is the task? Why is it important or interesting? What is your planned approach? What tools, systems or datasets are involved? What is the expected outcome? How are you going to evaluate your work?**

Hospital Transparency and Treatment Search System.

Healthcare costs make up one of individuals and employers largest expenses and prices continue to rise. New legislation in recent years has required hospitals to provide medical services price transparency by 2023. This data is accessible on hospital websites but not easily accessible to individuals searching by service.

Planned Approach:

* Find datasets/APIs that provide a list of hospital price transparency
* Scrapper functionality to extract useful data from hospital websites
* Search functionality to allow users to sort through and filter data based on a query
* Provide a recommendation functionality to match results based on similar content

Datasets:

* “Cost of care” and “price transparency” found on some hospital websites. Hospitals post prices on websites. Many times, in an Excel file format. (e.g., https://www.rush.edu/patients-visitors/billing/cost-care, https://www.uchicagomedicine.org/patients-visitors/patient-information/billing/price-transparency)
* CPT codes / Procedure code mapping/interchanging
* Kaggle Dataset hospital-price-transparency link

Expected Outcome:

A system that responds to a user query/search of medical service with a list of hospitals that are ranked according to reported prices and their ratings

Work Evaluation

To evaluate the project outcome, we plan to cross check our results with a specific known subset of hospitals/services to verify results

**3. Which programming language do you plan to use?**

Python

**4. Please justify that the workload of your topic is at least 20\*N hours, N being the total number of students in your team. You may list the main tasks to be completed, and the estimated time cost for each task.**

The project effort will be split into these main components:

* Find appropriate datasets and websites to extract/scrape data from 10 hrs.
* Data cleaning, mapping and transformation to make it ready for the search logic 10 hrs.
* Implementation of data parsers 10 hrs.
* Implementation of search functionality 20 hrs.
* Testing 10 hrs.
* Bonus feature: Additional functionality to implement recommendation system 10 hrs.