<u>Group Name:</u> ILoveFantasyHockey

<u>Group Members:</u> Hania Dziurdzik - haniakd2 (Captain), Erin Dowdy - eedowdy2

CS 410 Final Project Proposal

We chose the **Free Topics** option. We want to analyze text data found on the UIUC Reddit page to find relevant locations mentioned in the forum given a specific query (e.g. "restaurant" or "lecture hall") . It would be a convenient resource for people who are unfamiliar with the area to quickly search for and visualize new places similar to ones they've seen before. We will also make an algorithm to analyze which locations are the most relevant to the user's provided query. Once sorted by relevance, we will display the locations on a map, with the most often mentioned locations displayed most prominently so that users can see the places in the area related to their query that others bring up most.

We plan to use Python for our project. Python is ideally suited for this project because it has the versatility to be used for all of the different forms of analysis that we need to do. We plan to use the Python libraries Selenium and BeautifulSoup for website scraping, MetaPy for NLP, and Pandas, GeoPandas, and MatPlotLib for mapping and visualization. Additionally, both of us are very familiar with working in Python.

The main tasks for this project are acquiring and formatting data into a usable format (i.e. scraping the website), creating the algorithm for ranking the locations by relevance, generating the map, and designing the visuals. We will spend 7 hours researching and brainstorming our approach, 10 hours working on website scraping and data formatting, 15 hours designing and implementing the algorithm, and 8 hours working on the visualization.