Joseph Day, Shelby Mahaffie, Sherry Shenker

January 26, 2016

CMSC 122 Project Proposal: Chicago High School Finder

Selecting a high school to attend as a student living in Chicago is not a simple endeavor. Options include traditional neighborhood schools, magnet and selective enrollment schools, charter schools, and private schools. Choosing a high school requires understanding the options that are available based on a student's location, understanding eligibility for schools that select students based on complex preference algorithms, and understanding the quality of each of these options. Unfortunately, the data necessary to acquire this understanding is dispersed across many websites, and is often times hidden and in a very non-user-friendly format. Thus, many of the benefits that could arise from the city offering students a wide range of options are lost because these options are too difficult to understand for the families that stand to benefit the most. Our goal is to create a tool that synthesizes multiple sources on high schools in the city to allow families to easily view critical, unbiased information on schools that are available to their children, filter schools based on criteria that is important to them, and access data on enrollment processes, selectivity, and financial aid if applicable.

In order to achieve this goal, we will build a web interface that allows families to input information regarding their school preferences and geographic location and generates a listing of schools that best match their needs. The focus is on presenting information that is most valuable to the specific user in a way that can be used directly for decision making. The tool will have the capability to filter for schools that fall in a variety of categories, such as ones that offer ESL programs, vocational-education or private schools with low tuitions. Moreover, families will be able to compare schools based on metrics of achievement and school community through intuitive representations of the data. Finally, we will utilize information on the time required to travel to each school from a student's home and direct users to information on the enrollment process.

We plan to get most of our data from the Chicago Public Schools (CPS) website, which publishes spreadsheets that include high school graduation rates, college and career readiness indicators, state exam scores, parent satisfaction survey results, and rates of graduation, college enrollment, and persistence, for all public and charter schools in the city. The CPS website also

has an interface that can list the neighborhood schools whose boundaries include a specific address. We also plan to scrape data from *Chicago Magazine*, which publishes an annual guide to Chicago private schools with information on financial aid, student-to-teacher ratio, religious affiliation, acceptance rate, and college acceptance rate for each school.

We will use Google Maps to determine how long it would take a student to get to a particular school using public transportation in time for the start of the school day. Other potential sources of data that we are considering include statistics on crime rates in the neighborhoods of the schools and on the bus/train routes that the student would take to get to school, data published by the CPS Office of Access and Enrollment on the admission scores of selective high schools, and Illinois High School Sports Association data on school sport programs.

Our first steps, which we hope to complete by the end of 5th week, will be to convert all of the CPS data we plan to use to CSV files and scrape the private school information from the *Chicago Magazine* website. Sixth week, we plan to clean the data and make a SQL table, with unique school ID numbers as a key, that contains all of the relevant information on each school. Next, we will create an algorithm that filters schools and decides the order that schools should be listed in the search results, incorporating Google Maps data. Finally, around 9th week, we will begin creating the web interface described above.

In addition to expanding our knowledge of Python, SQL, and web-scraping, we will have to acquire a few new skillsets not covered in the course to build our project. We will be learning how to use the Google Maps API, creating a web interface using a Python library, and generating our own algorithm to rank schools that are displayed as results in the user's search.