Opportunity and Significance

- With so many colleges and universities, it can be difficult to find the right one.
- Factors such as cost, location, and course transferability can further compound confusion.
- What is needed is a tool that can allow prospective students make more informed decisions when deciding which school to attend.

Technical Approach, Accomplishments and Results

- Used Python to scrape course information from approximately 500 institutions of higher learning.
- Used Gensim, a machine learning library, to analyze and categorize course information based on content.
- Created a web application using the Django framework which displays courses based on location, cost, and type.

Next Steps for Development and Test

- Continue to optimize web scraper in order to gather data from more schools.
- Continue to improve the accuracy of the machine learning algorithm.

Commercialization Plan & Partners

- This project will be used by the Amesite company where it will continue to be refined before it is ultimately deployed on the web.

Technical Objectives

- Survey the type, availability, and cost of all 10X math courses available online, using publicly available web data.
- Compare costs of courses (by any residency status) and accreditation.
- Map the location of each course to the location of a user.
- Use machine learning to analyze course content and categorize each based on similarity.

Related Work and State of Practice

The realm of the project explores uncharted territory in the world of college content aggregation while building upon the groundwork already laid in the fields of data mining and machine learning.

Special thanks to our professors Dr. Khayyam Hashmi and Dr. Sam Bryfczynski, and our client Dr. Anne Marie Sastry.