Chloe Bergsma-Safar
January 20, 2022

Web Scraping & Linear Regression Project Proposal

Question/need:

The framing question of my analysis is, what features can best predict the price of undergraduate tuition and fees at U.S. private non-profit colleges and universities offering Bachelor's degrees?

High school students, their guardians, and any adult considering obtaining a Bachelor's degree from a private school may benefit from this model because it can help them determine the value of a given school.

Data Description:

I plan to use data from the National Center of Education Statistics. I will use web scraping techniques to obtain the data.

An individual unit of analysis in this project is a private non-profit institution offering Bachelor's degrees. The features I expect to work with are those that can be scraped with Python's BeautifulSoup package, including:

1. total faculty
2. total undergraduate enrollment
3. student-to-faculty ratio
4. price of books and supplies
5. % of undergraduate students receiving aid
6. number of undergraduate students receiving grant or scholarship aid
7. number of undergraduate students receiving Pell grants
8. number of students submitting SAT scores
9. % of students submitting SAT scores
10. number of students submitting ACT scores
11. % of students submitting ACT scores
12. SAT Evidence-Based Reading and Writing 25th percentile score
13. SAT Evidence-Based Reading and Writing 75th percentile score
14. SAT Math 25th percentile score
15. SAT Math 75th percentile score
16. ACT Composite 25th percentile score
17. ACT Composite 75th percentile score
18. ACT English 25th percentile score
19. ACT English 75th percentile score
20. ACT Math 25th percentile score
21. ACT Math 75th percentile score

The target I will predict using linear regression is the price of an institution's tuition and fees.

Tools:

To meet the tools requirement for this project, I intend to use BeautifulSoup. I also plan on using the following additional Python packages to complete my analysis:

- Pandas
- Numpy
- Seaborn
- Scikit-learn