

Capstone 1 Project Proposal: Texas Education

What is the problem you want to solve?

I would like to predict the percentage of students who will earn a college degree within four years after graduating high school from a certain school district. This problem strictly applies to students graduating high school in the major regions of Texas (Houston, Austin, San Antonio, Richardson (Dallas), Austin) that will attend a Texas college.

Who is your client and why do they care about this problem?

The clients are parents considering a move to or within the major regions of Texas. They'd like for their child to be prepared to earn a college degree and live a more comfortable life. In specifying they want their child to graduate within four years, the clients are looking to avoid a situation in which their child takes more time to earn their degree (more tuition money spent) or even fails out (worst case scenario with minimal return on investment). In strictly focusing on Texas colleges, the clients are also looking to avoid expensive out-of-state tuition.

What data are you using? How will you acquire the data?

I will download district-level SAT, ACT, AP, and Wealth/ADA datasets from the Public Education Information Management System ("PEIMS") on the Texas Education Agency's website. Datasets on college enrollment and college graduation will be downloaded from the Texas Public Education Information Resource ("TPEIR") website.

How will you solve this problem?

I will clean/wrangle all the datasets and then merge them together into one dataset that contains each school district's historical features (SAT, ACT, AP, Wealth/ADA, College Enrollment) and the resulting percentage of students who earned their degree within four years. With historical feature and target data in place, I will attempt to build a model that can predict the percentage of students who will earn their degree within four years, based on the features of the school district they attended high school in.