Capstone Two: Project Proposal

What are the primary factors that have contributed to the closing of 4-year undergraduate colleges and universities in the US in the last decade?

According to the "Number of higher education institutions in the U.S. 1980-2017" data, between 2012 and 2017, the number of higher ed institutions decreased by roughly 400. At least 2 colleges in my region that have closed in the last several years. It could be beneficial for colleges and universities to know what factors have led to the closing of other institutions, particularly as they face financial difficulties from the coronavirus, which compound the previously existing demographic trends that challenge college enrollment. Insight into which factors that have contributed most significantly in the closing of other colleges could allow current college administrators to make informed decisions about how to protect the future of their institution.

The prospective client is the Council of Christian Colleges and Universities (CCCU). They provide data, support and insight for 180 member colleges and universities. Based on the analysis, they hope to get a sense of the strengths and weaknesses of their member institutions, identify institutions that may be in danger of closing, and help to provide those institutions with planning, data and support to enable them to make wise decisions regarding the school's future.

The primary data source will be the US Department of Education's College Scorecard, available from Kaggle and from https://collegescorecard.ed.gov/data/. There is also an API available at https://api.data.gov/ed/collegescorecard/. Secondary data may be taken from the Statista dataset "Number of higher education institutions in the U.S. 1980-2017" at https://www.statista.com/statistics/240833/higher-education-institutions-in-the-us-by-type/.

Proposed process:

1. Find the ~400 schools that have closed in the last decade, explore their distributions and identify common features
2. Identify features of similar schools that have not closed in the last decade to create a dataset containing both closed and opened schools
3. Subset the data further to look at primarily 4-year, undergraduate institutions. These will likely be private schools, possibly with a religious affiliation.
4. Choose potentially relevant features to examine. (The dataset contains many features, which will need to be narrowed.)
5. Choose an appropriate modeling technique to model the data.

Resulting products:

- GitHub repo with code for each step in the project
- A written report
- A slide deck for an oral presentation