Assignment 8 – College Scorecard Analytics Project

Nov. 22nd, 2016

Summary:

A recently completed study of colleges in the US collected information regarding school and student demographics, degrees awarded by subject area as well as financial information including average degree cost by income bracket, student tuition and scholarships, repayment rates, and post-college average earnings.

As the Data Scientist on this project, you have broad discretion in the selection of source data, correlation of meaningful attributes, predictive model selection, and presentation of findings from this analysis.

Your task is to complete the following analysis activities:

1. Retrieve and review student study data (https://collegescorecard.ed.gov/data/)

2. Perform exploratory data analysis
   - 2.1. Count of colleges by degree length and region
   - 2.2. Categorize student graduation counts by degree name

3. Study evaluation
   - 3.1. Describe study type and approach
   - 3.2. Bias analysis: populations measured, access to scholarships, etc.
   - 3.3. Hypothesis analysis

4. Summary analysis – Value-add analysis for education
   - 4.1. Identify highest “value” in terms of (cost – salary) per degree
   - 4.2. Identify highest “value” in terms of (cost – salary) per college

5. Predictive model – Student graduation rates
   - 5.1. Evaluate the primary features that affect college program completion
   - 5.2. Develop a simple model to predict student graduation (free to use any model)

6. Results presentation
   - 6.1. Summarize your approach, code segments, and results analysis in an (.Rmd) file.
   - 6.2. Create custom infographic to visualize your analysis results