

Project 3 Assignment

Project 3

Instructions: For the final project you'll do data exploration with the colleges data set. For each of the questions include code and any relevant output.

Feature engineering

1. Use feature engineering to construct 3 additional variables to include in your association mining.
2. Plot your new features to view their distributions.

rules

3. Mine for rules with data that includes your new features from 1. Filter the rules for your features and describe the rules you uncover. **Note:** you might need to try different values of support and confidence to generate rules with your newly created features.

Clustering

4. How many universities are in the cluster of grant receivers? What is the median and range of `family_median_income` of these universities?
5. Upon review you're informed that there are too many universities receiving grants. They really like the cluster approach and suggest you make 5 clusters instead. After redo-ing the analysis, how many universities are in the new cluster should receive the grant? What is the median and range of `family_median_income` of these universities and how does it compare to your answer in 4?
6. What is one finding you've uncovered that surprised you, and what method did you use to find it? Feel free to use any technique we learned in class.