College majors and employment relevance

February 6, 2022

What is your data set?

The dataset represents a list of 174 college majors available in the USA, with relevant information about the labor force and basic information regarding employment, originally being part of an analysis done by FiveThirtyEight [1] about the economic implications of deciding what major to study. As higher education becomes more expensive every year, and fewer students choose their major just by considering their strengths and passions, purely out of socio-economical reasoning, it is detrimental to have a statistical approach to major selection. With this idea in mind, a set of 20 variables available about each major will lead the analysis that has as main goal to indicate the differences and similarities between majors, and also the financial benefits. Out of the 20 variables in the dataset, the major name and the major category are categorical variables, while total number of people in the major, male graduates, female graduates, proportion of graduates that are female, number of full time employed, number of part time employed, number of unemployed, unemployment rate, median earnings, 25th percentile earnings, 75th percentile earnings, number of graduates with jobs requiring college degree, number of graduates with jobs that don't require college degree and number of low-wage jobs are numerical variables.

What makes this data set specifically an interesting multivariate data set?

The dataset has enough information to allow us to detail what majors are alike and what majors are improving the chances for a graduate to have a wage that affords more than the basic needs. Additionally, it allows grouping majors by demographics and answer questions such as *What majors put you on a trajectory for low-income jobs ?* or *Are some majors more susceptible for unemployment than others ?*. Since the major category is also part of the dataset, it would also be interesting to know if the variables existing in the dataset can predict which major category a particular major is part of. Given that there is a constant debate between STEM students and humanities students, can we draw conclusions about the life prospects, just by purely analysing characteristics about these degrees.

Describe one or two possible directions of analysis

A few directions that could possibly be explored are clustering the observations based on variables present in the dataset. One idea that could be particularly interesting is to know if the majors can be clustered by a combination of full time employment and median income or by

gender and number of low wage jobs. Given that in reality a dataset including all the available majors at a global level could include many more datapoints, dimensionality reduction can be explored to summarize the entire dataset with fewer number of variables. Furthermore, regression analysis could be used to infer if any of the parameters about a major are a leading cause to low-wages or to unemployment. If so, are there any changes that colleges offering those degrees could do in order to make the majors more attractive to the general population ?

Bibliography

[1] fivethirtyeight.com *The Economic Guide To Picking A College Major*, Ben Casselman