December 2020

Final Project: Is College Worth It?

Dataset

Minnesota Population Center. IPUMS Higher Ed:Version 1.0 [dataset]. Minneapolis, MN: University of Minnesota, 2016. https://doi.org/10.18128/D100.V1.0

By: Viraj Joshi, vj3675

01

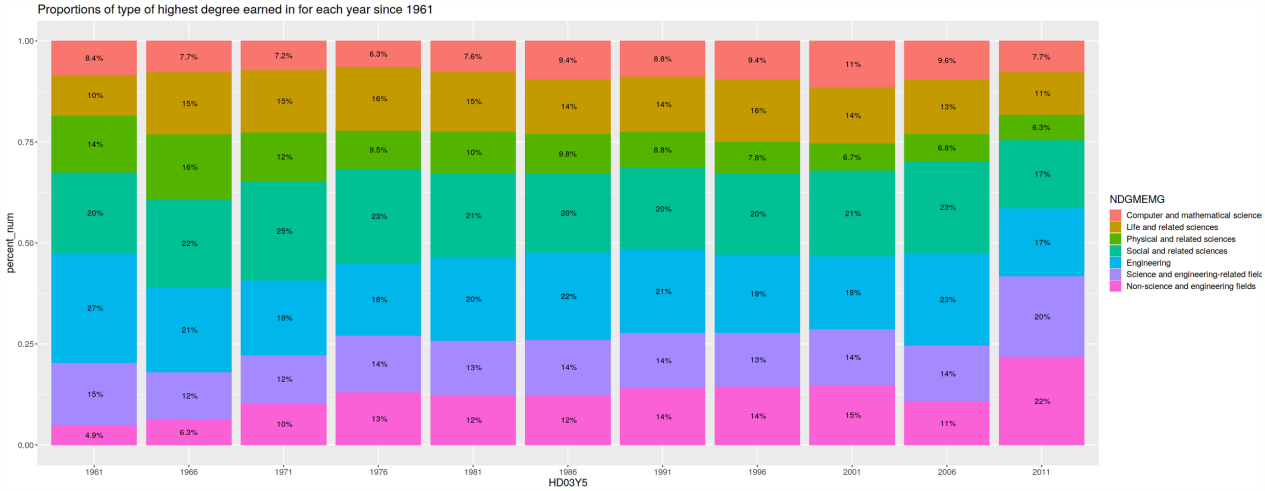
Basic Analysis

Dataset

The dataset is comprised of surveys, the 2013 SESTAT SDR and the 2013 SESTAT NSCG.The SESTAT is a subset of all three surveys(NSCG, NSRCG, SDR) conducted by the National Science Foundation(NSF). Individuals who have science/ engineering degrees or occupations are in this database. This implies that the NSCG survey can have individuals without a bachelor and/or occupation in S&E.

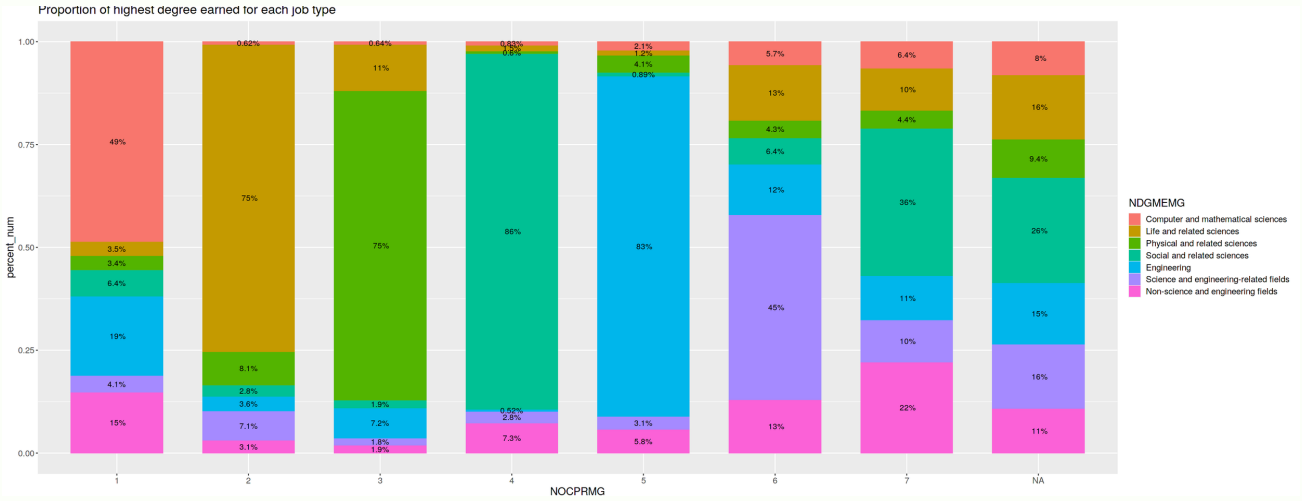
Education

The ebb and flow of the proportions of higher degrees over time could be connected to claims of job satisfaction and salary.



Degree Relevance

What is the distribution of majors for each occupation?



If people work in the field that they were trained for, their degree is relevant. One expects the ith NOCPRMG (job code) category to have the ith NDGEMEMG(degree trained for) as the dominant major. In fact, fields 2-4 are mainly dominated by greater than 50% of those who obtained degrees in preparation for the occupation.

Salary Model

Variables that affect SALARY:

- Employer sector
- highest degree trained for
- School awarding highest degree located in the US
- Job Satisfaction
- Satisfaction with Salary
- years since obtained bachelor's degree
- 10+ more

Developing a model

Dropping Observations: Those that are unemployed,not in the labor force and looking for work, and employed people that had no salary(615 people) were dropped. The last type of observation is of particular interest to justify removal because they mainly work for business or industry but are unpaid. It seems to only weaken the ability of predictor used for employees who do have a salary. To combine the concerns of a minority who take on unpaid roles (e.g those in retirement) with the vast majority taking on paid roles for vastly different reasons does not seem appropriate.

Variables were selected through an iterative method that focused on raising metrics of accuracy and explained variability.

However good (or bad) this model may be, it could be the case that salary cannot be explained by linear regression, and more continuous,numerical data was needed. This model provides an acceptable balance between complexity and accuracy.

Which career path to maximize SALARY?

If we look at the most common combinations(which implies the most realistic situations) of variables relevant to starting a career for the top 20% of earners predicted by the model, it would be an indication of what could earn the most in a career.

For the following median salaries and relevant combinations

- $96223.04 - Physcial and Related Sciences, Government, School awarding highest degree located in the US,Management and Administration,benefits available, health insurance available,Job required technical expertise: natural sciences, Doctorate Earned
- $92304.60 - Engineering, Government, School awarding highest degree located in the US,Research and Development,benefits available,health insurance available,Job required technical expertise: natural sciences,Doctorate Earned
- $103199.68 - Engineering, Government, School awarding highest degree located in the US,Management and Administration,benefits available,health insurance available,Job required technical expertise: natural sciences,Doctorate Earned

The groups here are by no means the groups with the highest median, but investigating salary groups with 1 or 2 people who have a certain combination of factors that give them a higher salary is not useful in determining a feasible plan to maximize Salary.

Job Satisfaction Model

Variables that affect Job Satisfaction:

- Satisfaction with Salary
- Satisfaction with Independence
- Satisfaction with Security...

Developing a model

Dropping Observations: Those that are unemployed and those not in the labor force and looking for work are dropped the dataset. It is not reasonable to predict job satisfaction for those with no jobs. Unlike Regression 1, I keep those that are employed and have a salary of 0 because they do have a opinion of satisfaction.

Variable selection was determined by selecting all varaibles related to satisfaction.

The model, while achieving an upper bound accuracy of .92 , the model performs worse than a far simpler model, simply outputting 1 to all inputs, will achieve ~99% accuracy because ~99% are satisfied with their job in this dataset. This suggests the model, while initially impressive, is not useful.

Which career path would maximize JOBSATIS?

The model would tell one which combination of satisfaction variables would result in a satisfied job. This would allow one to pick the remaining variables that determine career path. Some factors to look for when starting a career are if a job has insurance/vacation, highest degree trained for, etc.

The top 3 most common/attainable career plans to maximize satisfaction,listed in the combination formed above and based on the chosen factor combinations, are as follows

1. Engineering, engineering, highest degree is bachelor degree, business or industry employer sector, highest degree awarded in US ,job has health insurance
2. Non-science and engineering, Social and related sciences, highest degree is bachelor degree, business or industry employer sector, highest degree awarded in US ,job has health insurance
3. Science and engineering-related, Science and engineering-related, highest degree is professional degree, business or industry employer sector, highest degree awarded in US ,job has health insurance

However, these recommendations are not to say that these groups have the highest rates of job satisfaction within their own factor combination because we excluded those that were unsatisfied in their combination pool. Instead, we can say that they occur the most frequently within the pool of those who are satisfied with their job, and thus, the best factors one could optimize to maximize job satisfaction.

Fact Check Article

Gallup: Does Higher Learning = Higher Job Satisfaction?

https://news.gallup.com/poll/6871/does-higher-learning-higher-job-satisfaction.aspx

1 Claim #1: Education level has very little to do with job satisfaction, or satisfaction with income and time flexibility.

The analysis of the claim will ignore the analysis of high school graduates with job satisfaction in the article because the surveys used in our dataset imply the participant has at least a bachelors(NSCG+SDR). There are three parts to this claim.

i) Education level has very little to do with job satisfaction

Since my recommendation to maximize job satisfaction involves 4/7 job types as well as 3/4 levels of education in the survey, it does not seem like any education level alone implies satisfaction or dissatisfaction. Just as the article states "educational achievement...seems to have very little to do with overall job satisfaction", the data provided in the NSCG and SDR surveys shows that each distribution for the highest level of education attained(DGRDG) with respect to job satisfaction(JOBSATIS) reveals similar proportions of satisfaction in each each DGRDG category (Graph 1) and as a result, implies the same conclusion as the article.For example, one could include DGRDG in the logistic regression model, and while some levels are significant, most are not and do not improve AUROC or accuracy. However, while the claim itself is valid, it is important to note, like we did in the regression analysis, that the addition of other factors associated with education level could change the rates of those being satisfied with their job,and as a result, the probability of a given person being satisfied. This is evident in that some combinations of factors and being satisfied are rarer than others and an indication that with a larger context, education level does somewhat deal with job satisfaction.

ii) Education level has very little to do with satisfaction and time flexibility

To determine satisfaction with time flexibility for a given highest education level attained,the variables HRSWKGR (number of hours worked per week) and WKSWKGR (number of weeks worked per year) were used. Much of HRSWKGR and WKSWKGR factors were not relevant to predicting satisfaction and did not improve model accuracy, so they were left out of the model. Furthermore, the data from the surveys shows that each DGRDG distribution with respect to HRSWKGR/WKSWKGR and JOBSATIS reveals similar proportions in each category on the x-axis(Graph 2/3) and as a result, implies the same conclusion as the article. However, we also claim the last point of the previous paragraph replaced with the appropriate variable.

iii) Education level has very little to do with satisfaction and income

To determine satisfaction with income for a given highest education level attained, the variable SATSAL were used. The data from the surveys shows that each DGRDG distribution with respect to SATSAL and JOBSATIS reveals similar proportions in each category (Graph 4) and as a result, implies the same conclusion as the article. However, the minimum salary threshold that the article states ,where the association between highest education level attained and SATSAL disappears, could be present for samples where that salary threshold for any DGRDG level is not met.

Broken down by DGRDG, Graphs 2-4 corroborate the article's claim that "income and time flexibility doesn't seem to have much to do with your educational attainment".