1. Obtaining the history of occupational fields that a specific university or college students go into after college. (Selected approach)

Feasibility:  
a. This data might be proprietary and not publicly available and hence difficult to create the model from it  
b. However, if obtained can be easily implemented. Higher probability weights can be added to SOC codes which has been often explored by alumni of college

Breaking this case into two:

1. Data that contains both fields of study/major and employment outcome
2. Data that contains only employment outcome
3. Data that contains both fields of study/major and employment outcome

For this case, the original crosswalk from the DOE-BLS dataset can be used in conjunction with the school’s crosswalk.

The school’s crosswalk can only be used in this case, but then the major/field of study has to be converted into CIP codes by hand and that might not be a great setup.

Thus the plan is to use the CIP code, get the title of the CIP code and then match the CIP code title to majors/field of study from the schools to which it matches. For example, the 2010 CIP code for ‘Biology’ is 26.0101. Say in the school’s specific data there are fields of study from the school that are dubbed ‘Biology’ and ‘Computational Biology’, then we can say that the CIP code for ‘Biology’ in the DOE data is the same for ‘Biology’ and ‘Computational Biology’ in the school.

Then there are two ways of ranking SOC data:

1. By number of employment in the field from the BLS data (more students are likely to get into the field with the most employment)
2. SOC data generated can be a combination of the SOC fields of employment and the school’s field of employment. Then, the school’s field of employment is given a higher preference. Fields of employment from the school can be converted into SOC using the same method from above: matching the school field’s name with any SOC data field that has similar naming to the school’s field.
3. Data that contains only employment outcome  
   For this case, the original crosswalk from the DOE-BLS dataset can be used in conjunction with the school’s data.

The CIP code is used to get the SOC codes from the original crosswalk and the for the SOC data generated, the fields of study can be weighted higher by checking if the SOC code title has any semblance to any field of employment data from the school. Weighting by number of employment in the field from the BLS data can be used as an additional weighting factor.

1. Looking at job hirings at a specific distance from the school.  
     
   This would be a new model which would directly connect college name/college code to SOC code based on job postings around the school at some distance (maybe < 100 miles). This is based on the assumption that students are most likely to be hired around their school college  
     
   Feasibility:  
   a. Dataset can be scrapped from job posting websites (might require more technical skills)  
   b. This can be joined with the CIP-SOC crosswalk such that higher probability weights can be added to SOC codes of job postings which are near the college.
2. Added weights of final salaries (salary multiplier) of some specific colleges for SOC data.  
     
   Some graduates from some colleges tend to generate higher salaries than some other graduates because of the school they attended.   
     
   Feasibility:  
   a. Rank of school can determine salary multiplier to raise/lower

b. Location of school leads to higher or lower living costs causing salary multiplier to fluctuate