**Capstone Project Proposal**

**Yongjiang Liu**

It is not uncommon for people to work on industries not directly related to their courses of study in college. Take finance industry for instance, we saw people of all different backgrounds, from computer science to astronomy. However, is finding jobs outside your field a good choice? Which industries are more immigrant-friendly (by saying immigrant, I am talking about those who enter an industry without a relevant education background, e.g. a chemical engineer in an administration position)? Which industries pay locals more (same logic as “immigrant”, here “locals” means people with relevant education background)? With the help of data, we are able to answer those questions.

Our data are collected from the [IPUMS Higher Ed portal](https://highered.ipums.org/highered/), which is an institution that provides data from three “leading surveys for studying the science and engineering (STEM) workforce in the United States.” These surveys cover questions about educational choices, demographics, employment outcomes and more over a period from 1993 to 2013. [1] Most of the variables needed in this study can be found on their website, such as the “major of participant’s highest degree”, “graduation year” and “annual salary”. The dataset we are going to work with contains more than 80,000 observations, covering 4 types of degrees and more than 10 fields of majors.

First, we will clean the dataset and group observations by different industries. Analysis will be carried out under each specific industry. Visualization of those data could give us intuitive explanations to our questions. Following that, we will dive deeper with various regressions and correlation analysis, aiming to find out the main factor that causes the pay gap between “locals” and “immigrants”, assuming these gaps do exist. With machine learning algorithm, our model could tell if it is a good idea for an individual to take educational-irrelevant jobs.

Educational institutions and consulting companies would be our target clients. Supported by this study, educational institutions such as universities can guide their students with better career paths. On the other hand, companies focus on providing college application consulting services would be more confident to advice their customers.

Furthermore, we can answer more complicated questions using this dataset: which variables influence employee’s income the most, given one specific industry? Could the answer lie in the degree he/she is holding, or the city he/she was born in? Further investigations will be covered in this study.

If interested in the idea of this study, you are more than welcome to take a look at what we have in this repository. Feel free to raise any questions, as it will inspire as well as help us answer more questions.

[1] Jeremy Singer-Vine, December 07, 2016, Data Is Plural - 2016.12.07 edition, https://tinyletter.com/data-is-plural/letters/data-is-plural-2016-12-07-edition