Dataset 1: Cross-Sectional Child Income Statistics by College Tier and Parent Income Percentile

Member: Alia

Dataset description: The data consists of 1’500 rows divided into 100 blocks of 15 rows. Each block represents a parent’s mean income percentile and is broken down into 15 different selectivity and type combinations. In total, there are 4 variables that represents parents’ information and 11 variables that give information about their kids. All the data consists of numerical values and the selectivity and type combinations are coded into 15 classes.

Why is this dataset interesting?

Given the variables, we can explore the financial mobility across generations. Furthermore, we can obtain predictions in the form of frequentist probabilities about a child future earnings given their parents' earning and college degree. Also, we believe that this dataset can lead us to further investigations about the U.S. educational system and if college opportunities are fair or biased.

How to access or collect the dataset

Dataset can be found here below.

https://opportunityinsights.org/data/?\_\_cf\_chl\_f\_tk=vqPGKqmdKzzg3F7z4tVKRgeoLPq7.MAlMgu12saA4Vs-1642604618-0-gaNycGzNB\_0

What potential questions do you want to explore?

- What is the relationship between variables?

- How does income impact the marital status and what are their interrelations?

- How does it vary by gender? (Explore other datasets)

Dataset 2: Cross-Sectional Statistics on Children’s Income Distributions by College Tier

Member: Marinie

Dataset description: The data consists of 1365 rows divided into 91 blocks of 15 rows. Each block represents a kid’s earning percentile and is broken down into one of 15 selectivity and type combinations. Percentile is starting from the 9th percentile. In total, there are 7 numerical variables that all give kids information. Similarly, to the first dataset, all the data consists of numerical values and the selectivity and type combinations are coded into the 7 variables.

Why is this dataset interesting?

This dataset is interesting because it provides further details about the future of the kids in the previous dataset. It is a subset of the first dataset from the kids’ perspective. Hence, it allows us to complete our analysis of the first dataset. In particular, it helps us explore the relationship between the college type and income and to compare with the previous generation.

How to access or collect the dataset

Dataset can be found here below.

https://opportunityinsights.org/data/?\_\_cf\_chl\_f\_tk=vqPGKqmdKzzg3F7z4tVKRgeoLPq7.MAlMgu12saA4Vs-1642604618-0-gaNycGzNB\_0

What potential questions do you want to explore?

* How long has this been since graduation?
* What was the major of the kids and how does that impact their income?
* How was the data collected and what was the geographical scope?