Methodology at a Glance: Data Sources and Data Elements for Corresponding Cohorts

**IPEDS (2008/2016 Cohorts)**
- Institution type: Highest Degree, Control, No Distance Education
- SAT/ACT
- Average net price for Title IV-aided high-income students
- Percent Undergraduates Age 20-21
- Percent of full-time first-time undergraduates receiving Pell grants
- Percent women (First-time Full-time students Undergraduate Degree/certificate-seeking)
- Full-time first-time degree/certificate-seeking undergraduate enrollment and enrollment (admission) yield
- Percent Asian/Native Hawaiian/Other Pacific Islander total - derived (First-time Full-time students UG Degree/certificate-seeking)
- Percent admitted – total
- Freshmen Enrollment (100s)
- Percent Hispanic Full-time students Undergraduate Degree/certificate-seeking First-time
- Cost of attendance change between 2000 and 2013

**College Scorecard**
- Percent of dependent students by income background
- Percent of aided students by income background
- Percent of non-traditional aged students at entry
- Percent of ethnic distribution in zip code area (via US Census)
- Percent of college graduates in zip code area (via US Census)
- Median household income of earnings cohort
- Poverty rate of earnings cohort (via US Census)
- Cumulative loan debt at selected percentiles
- Student undergraduate enrollment by ethnicity/race
- Instructional expenditures per full-time equivalent student
- Percent of all undergraduate students receiving a federal student loan
- Percentage of independent students by selected income level
- Median debt for independent students and for graduated students
- Share of married students
- Percent of first-generation students in earnings cohort
- Unemployment rate (via US Census)

**Opportunity Insights**
- Mobility Rate: Percent of students who have parents in the Bottom 20% of the income distribution and reach the Top 20% of the income distribution
"Value-Added" Calculation for Each Component Score

Retention rate trend: The difference-in-difference value between the actual and predicted trend in first-year student retention for the 2013 and the 2015 cohort period. The predicted rates are based on 12 data elements that explain 75 percent of the variation in institutional student retention.

Graduation rate trend: The difference-in-difference value between the actual and predicted trend in the six-year graduation rate for the 2008 and 2010 starting cohort period. The predicted rates are based on 11 data elements that explain 88 percent of the variation in institutional student graduation rate.

Loan Repayment Rate: The difference value between the actual and predicted three-year loan repayment rate for the 2010-11 academic year graduating cohort (as defined by the National Student Loan Data System). The predicted rate is based on 16 data elements that explain 81 percent of the variation in institutional student loan repayment rate.

Income Mobility Rate: The difference value between the actual and predicted student income mobility rate, which is the percent of students born 1980 through 1982 who have parents in the bottom 20% of the income distribution and reach the top 20% of the income distribution in 2014. The selected institution is the one attended most frequently by the student at the age of 19 through 22. The predicted rate is based on 13 data elements that explain 70 percent of the variation in institutional student income mobility rate.

Calculation of the Predicted Outcomes
A range of statistical algorithms described here control for the above-listed data elements, and they estimate each of the four outcomes separately. The final estimation model for each outcome is determined by an iterative process that aims to maximize the explained outcome variance with the most parsimonious model after exclusion of statistical outliers where deemed necessary.

For further inquiries, contact Sergei Herzog