The dataset contains 30 variables (1 response and 29 other variables) on 1508 colleges and universities in the United States. This dataset was derived from a larger collection of measures on schools in the United States (https://collegescorecard.ed.gov/data/). The variables collected here can be grouped into three categories: school identifiers, school characteristics, and student population/applicant characteristics. Below is a summary of the variables in the data:

|  |  |  |
| --- | --- | --- |
| Variable Name | Characteristic | Detailed Description of Variable |
| UNITID | Unit ID for institution | Unique numerical identifier of institution |
| INSTNM | Name of institution | Actual name of institution (text) |
| STABBR | State Postcode | Two-character label for state in which institution is located (i.e. factor variable) |
| NUMBRANCH | Number of branch campuses | Numeric value for number of satellite campuses/affiliations   * 1 indicates there is only one main campus |
| CONTROL | Control of institution | Factor indicating whether the institution is public (1), private non-profit (2), or private for-profit (3) |
| REGION | Regional location of institution | Factor variable with the following levels:   * 0 = US Service Schools * 1 = New England (CT, ME, MA, NH, RI, VT) * 2 = Mid East (DE, DC, MD, NJ, NY, PA) * 3 = Great Lakes (IL, IN, MI, OH, WI) * 4 = Plains (IA, KS, MN, MO, NE, ND, SD) * 5 = Southeast (AL, AR, FL, GA, KY, LA, MS, NC, SC, TN, VA, WV) * 6 = Southwest (AZ, NM, OK, TX) * 7 = Rocky Mountains (CO, ID, MT, UT, WY) * 8 = Far West (AK, CA, HI, NV, OR, WA) * 9 = Outlying Areas (AS, FM, GU, MH, MP, PR, PW, VI) |
| HBCU | Historically Black College and University | Factor variable where levels are no (0) or yes (1) |
| PBI | Predominantly Black University | Factor variable indicating whether currently predominantly serves black community (1) or not (0) |
| TRIBAL | Tribal college and university | Factor variable indicating whether university for native American tribes (1) or not (0) |
| HSI | Hispanic-serving institution | Factor variable indicating whether university serves Hispanic communities (1) or not (0) |
| WOMENONLY | Women-only College | Factor variable indicating women-only institution (1) or not (0) |
| ADM\_RATE | Admission rate (Response) | Defined as number of admitted students out of total number of undergraduate applications |
| COSTT4\_A | Average cost of attendance per academic year | average annual total cost of attendance, including tuition and fees, books and supplies, and living expenses for all full-time, first-time, degree/certificate-seeking undergraduates |
| AVGFACSAL | Average faculty salary | Calculated as the total salary outlays divided by the number of months worked for all full-time nonmedical instructional staff. |
| PFTFAC | Proportion of full-time faculty members | Calculated as the number of full-time nonmedical faculty divided by the total number of nonmedical faculty. Proportions are expressed as decimals |
| PCTPELL | Percentage of undergraduates receiving Pell grant (financial aid) | calculated as the quotient of the number of Pell grant recipients divided by the count of all undergraduates for either a fall enrollment cohort. Proportions are expressed as decimals |
| UG25ABV | Percentage of undergraduates aged 25 and above | A numeric variable |
| INC\_PCT\_LO | Percentage of aided students whose family income is between $0-$30,000 | A numeric variable indicating whether financial aid is awarded primarily to low income students |
| PAR\_ED\_PCT\_1STGEN | Percentage of first-generation students | A numeric variable |
| FEMALE | Proportion of student body that is female | A numeric variable |
| MD\_FAMINC | Median family income of students | A numeric variable |
| PCT\_WHITE | Percent of the population from students' zip codes that is White, via Census data | Numeric variable describing demographics of student’s home neighbourhood. |
| PCT\_BLACK | Percent of the population from students' zip codes that is Black, via Census data | Numeric variable describing demographics of student’s home neighbourhood. |
| PCT\_ASIAN | Percent of the population from students' zip codes that is Asian, via Census data | Numeric variable describing demographics of student’s home neighbourhood. |
| PCT\_HISPANIC | Percent of the population from students' zip codes that is Hispanic, via Census data | Numeric variable describing demographics of student’s home neighbourhood. |
| PCT\_BA | Percent of the population from students' zip codes with a bachelor's degree over the age 25, via Census data | Numeric variable describing education levels of student’s home neighbourhood. |
| PCT\_GRAD\_PROF | Percent of the population from students' zip codes over 25 with a professional degree, via Census data | Numeric variable describing education levels of student’s home neighbourhood. |
| PCT\_BORN\_US | Percent of the population from students' zip codes that was born in the US, via Census data | Numeric variable describing demographics of student’s home neighbourhood. |
| POVERTY\_RATE | Poverty rate, via Census data | Numeric variable describing income levels of student’s home neighbourhood. |
| UNEMP\_RATE | Unemployment rate, via Census data | Numeric variable describing income levels of student’s home neighbourhood. |