

What is the problem you want to solve?

The public, private, and home schools provide education in the United States. State governments set overall educational standards, often mandate standardized tests for K–12 public school systems, and supervise, usually through a board of regents, state colleges and universities. Funding comes from the state, local, and the federal government. Private schools are generally free to determine their own curriculum and staffing policies, with voluntary accreditation available through independent regional accreditation authorities. About 87% of school-age children attend public schools, about 10% attend private schools, and roughly 3% are home-schooled.

The United States spends more per student on education than any other country. In 2014, the Pearson/Economist Intelligence Unit rated US education as 14th best in the world, just behind Russia. In 2015, the Programme for International Student Assessment rated U.S. high school students #40 Globally in Math and #24 in Science and Reading.

It has been alleged, since the 1950s and especially in recent years that American schooling is undergoing a crisis in which academic performance is behind other countries, such as Russia, Japan, or China, in core subjects. Congress passed the National Defense Education Act in 1958 in an attempt to rectify these problems and a series of other legislative acts in later decades such as No Child Left Behind. According to the Organization for Economic Cooperation and Development, however, American students of 2012 ranked 25th in math, 17th in science, and 14th in reading compared with students in 27 other countries. In 2013, Amanda Ripley published *The Smartest Kids in the World (And How They Got That Way)*, a comparative study of how the American education system differs from top-performing countries such as Finland and South Korea.

Despite the demonstrated link between economic growth and education standards, high schools and colleges sharply disagree about the college readiness of high school graduates, in that 44% of college faculty believe that incoming students are not ready for writing at the college level, while 90% of high school teachers believe exiting students are well prepared.

The first school compulsory attendance law in the United States was enacted in Massachusetts in 1852. Other states soon followed, but it was not until the early 20th century that all states in the union had such a law. In the 19th and early 20th century the enforcement of compulsory school attendance laws by the school or government officials was usually quite sketchy, but in those years, the student who dropped out of school had a reasonably good chance of finding a job, since there was high demand for semi-skilled and even unskilled laborers. The nature of the American workplace as it now exists puts the young person without a high school diploma who is seeking a job with an adequate wage in a precarious situation. Even a high school diploma may not be sufficient to put the young person on a good path with the opportunity to move up the employment ladder. Additional education may be required, whether it is a community college, a technical certification program, or a four-year college.

Every year, over 1.2 million students drop out of high school in the United States alone. That's a student every 26 seconds or 7,000 a day. About 25% of high school freshmen fail to graduate from high school on time. The U.S., which had some of the highest graduation rates of any developed country, now ranks 22nd out of 27 developed countries. In 2010, 38 states had higher graduation rates. Vermont had the highest rate, with 91.4% graduating. And Nevada had the lowest with 57.8% of students graduating.

For this project, my focuses is on major educational issues in the United States and analyze the root cause of high school dropout issue and provided a detailed report on the following:

1. A summarized visualization of high school dropout factories by location
2. A comparison of 3-5 most high-school graduation rate district against 3-5 least high-school graduation rate district
3. Identifying the most prominent race in each district having least high-school graduation rate
4. Identifying the district with the most change in high-school graduation rate and potential reasons for the change
5. Building a predictive model of most high-school graduation rate in each district using machine learning
6. Finally, after addressing #5, I wanted to identify the most salient features/variables used by the model for predicting most high-school graduation rate, within the limitations of my dataset

Who is your client and why do they care about this problem? In other words, what will your client DO or DECIDE based on your analysis that they wouldn't have otherwise?

There are two different types of clients that could be interested in the findings from this project. The first type of clients would be the US online and print media that cover socioeconomic and urban issues. These clients are magazines that take an active interest in stories driven by socially relevant issues and are backed by data analytics, for creating awareness within the public while simultaneously enhancing the quality of their readership. For example, US online media such as US News and Gates Foundation would fall under this category. I also anticipate interest from Government funded bodies and non-profits offering job placement services, and subsidized education services for youth and adults.

What data are you going to use for this? How will you acquire this data?

National Center For Educational Statistics has a portal from where you can download the raw data. Apart from that, I will be using the Civil Rights Data Collection from ED.gov. I used 2 data sets consisting of Regulatory Adjusted Cohort Graduation Rate from US Department Of Education portal for School Year 2011 to 2013. Each dataset was provided as a raw dataset in CSV format for 2011 and 2013, which are imported as Pandas data frame.

In brief, outline your approach to solving this problem (knowing that this might change later).

1. I will study the four-year adjusted cohort graduation rate in 2011 and 2013 and identify top 3-5 school district having highest graduation rate and bottom 3-5 school district having lowest graduation rate.
2. Analyze the student population and investigate which among the following variables which are the having most significant impact:
 - a. **Location:**
 - i. By Regions: West, Southwest, Midwest, Southeast and Northeast
 - ii. By State
 - b. **Demographics:** Analysis of enrollment data in Public School by Major racial and ethnic groups to find out if the drop out rate is high due to any particular race
 - i. American Indian/Alaska Native
 - ii. Asian/Pacific Islander
 - iii. Hispanic
 - iv. Black
 - v. White
 - vi. Two or More Races
 - c. **Advanced Placement:** Analysis of Students enrolling for Advanced Placement course to find out if drop out rate is having any relation with the AP courses selected by students
 - i. Advanced Placement Course and Test Taking
 - ii. Advanced Placement Course Taking by Subject
 - d. **Athletics:** Analyze the Interscholastic Athletics data to find if there is any relation with the drop out rate
 - e. **Bullying and Harassment:** Analysis of bullying and harassment data to find out if there are students who are dropping due to excessive bullying and harassment in school
 - i. Bullying and Harassment on basis of Disability
 - ii. Bullying and Harassment on basis of Race, Color or National Origin
 - iii. Bullying and Harassment on basis of Sex
 - f. **Retention:** Analysis of the number of students retained has any relationship with the low graduation rate
 - g. **Course Enrollment:** Analysis of the courses enrolled by students in high school to find out if there is any relation in low graduation rate due to course selection
 - i. Algebra I & Geometry
 - ii. High School Math and Science
 - h. **Children with Disabilities:** Analysis of the number of disabled students to find out if there are students who are dropping due to disabilities
 - i. **Discipline:** Analysis of number of disciplinary action taken has any impact on the low graduation rate
 - j. **Finance:** Analysis of school expenses towards teacher's salary is related to the quality of education and hence may relate to the low graduation rate
 - i. Total Personnel salaries at school level
 - ii. Personnel salaries at school level for teachers only

- k. **Single Sex classes:** Analysis of number of single sex classes can lead to a relationship of low graduation rate
- l. **Staff:** Analysis of teaching staff to find any relationship with the low graduation rate
 - i. Total No of Teachers
 - ii. Teachers with State certification
 - iii. No of classroom teachers in their first year of teaching
 - iv. No of classroom teachers in their second year of teaching
 - v. No of High school counselors
 - vi. Teachers absent more than 10 days of the school year
- m. **Tests:** Analysis of number of students assessments in ELA and Math to find out if the low graduation rate is attributed towards any particular subject
 - i. Assessments in Reading/Language Arts
 - ii. Assessments in Mathematics
- n. School crime
- o. Alcohol, Illicit Drugs and Cigarettes consumption
- p. Federal Funding
- 3. Initial Data Exploration and checking the missing values
- 4. Normalizing the four-year adjusted cohort graduation rate if required
- 5. Exploring further into the factors contributing low graduation rate
- 6. Comparison of districts having most and least the cohort graduation rate
- 7. Find districts with maximum change in the cohort graduation rate
- 8. Identify prominent age groups and economic factors
- 9. Prediction using machine learning
- 10. Model Selection and evaluation

What are your deliverables? Typically, this would include code, along with a paper and/or a slide deck.

- 1. The Python source code
- 2. Final paper with detailed analysis with limitations and future extensions
- 3. A presentation of key findings and recommendations