Project Topic Idea: Measuring the success of MA school districts

1. Problem Statement

Massachusetts is widely regarded as having one of the best public school systems in the United States; however, the state also has recurring issues regarding equitable access to good schools and higher education. In our project, we want to use data concerning school districts across Massachusetts in order to better understand which school district features can lead to higher rates of students attending college. We will use machine learning models to predict what proportion of students in a school district will attend college or university based on certain features of the school district.

2. Significance of the Problem

Access to higher education is extremely important for individuals looking to attain a high quality of life. If there is not equitable access to higher education, then there cannot be broader social equity in our communities. The problem tackled by our project is the inequitable access to higher education in school districts across Massachusetts, and the insights made could help to determine which aspects of our school districts Massachusetts officials should focus on in order to increase rates of students attending higher education in districts across the state. For example, users can take advantage of the ML model to compare differing school district features such as teacher salaries or expenditure per student to see which changes have the most impact in increasing predicted rates of students attending college. These insights could prove useful for future educational policies or funding allocation in state government.

3. Dataset(s)

We are using data from the Massachusetts Department of Elementary and Secondary Education. One of the department's responsibilities is to collect data to inform local and state decisions, which is the information we are tapping into for this project. Specifically, we used data from the state and district profiles. We downloaded different datasets as .xlsx files and merged them on the district code column.

- Graduates Attending Institutions of Higher Education (District) - All Students - All Colleges and Universities (Attend Range: 16 Months)
- Enrollment By Race/Gender Report (District)
- Selected Populations Report (District)
- Teacher Salaries Report
- Per Pupil Expenditures, All Funds
- Student Discipline Data Report - All Offenses - All Students
    - Cells that contain 0.0% may be percentages that are more than 0.0% but less than 0.05%

Dataset File

Variable name in file	Description	Feature/ Outcome
Attending Coll./Univ. (%)	The percentage of high school graduates attending an institution of higher education. The types include private two-year, private four-year, public two-year, and public four year institutions.	Outcome
African American	The percent of students in the district this year that are African American	Feature
Asian	The percent of students in the district this year that are Asian	Feature
Hispanic	The percent of students in the district this year that are Hispanic	Feature
White	The percent of students in the district this year that are White	Feature
Native American	The percent of students in the district this year that are Native American	Feature
Native Hawaiian, Pacific Islander	The percent of students in the district this year that are Native Hawaiian/Pacific Islander	Feature
Multi-Race, Non-Hispanic	The percent of students in the district this year that are Multi-Race and Non-Hispanic	Feature
Males	The percent of students in the district this year that are Male	Feature
Females	The percent of students in the district this year that are Female	Feature
First Language Not English %	The percent of students in the district this year whose first language is a language other than English	Feature
English Language Learner %	The percent of students in the district this year who are English learners, defined as "a student whose first language is a language other than English who is unable to perform ordinary classroom work in English."	Feature

Students With Disabilities %	The percent of students in the district this year who have an Individualized Education Program (IEP)	Feature
High Needs %	The percent of students in the district this year who are designated as either low income (prior to School Year 2015), economically disadvantaged (starting in School Year 2015), or ELL, or former ELL, or a student with disabilities	Feature
Salary Totals	The total of teacher salaries for the district this year	Feature
Average Salary	The average of teacher salaries for the district this year determined as the Salary Totals divided by the number of full-time equivalent teachers	Feature
FTE Count	The count of full-time equivalent teachers for the district this year	Feature
In-District Expenditures	All of the district's operating expenditures for in-district programs this year	Feature
Total In-district FTEs	The average count of enrollment across this school year, for pupils enrolled at local schools in the district	Feature
In-District Expenditures per Pupil	The total in-district expenditures divided by in-district FTE average membership for this year	Feature
Total Expenditures	All of the district's operating expenditures for in-district programs and out-of-district placements.	Feature
Total Pupil FTEs	The average count of enrollment across this school year for students enrolled at local schools in the district and those publicly-funded students enrolled at other districts, including charter schools, special education collaboratives, and private special education schools	Feature
Total Expenditures per Pupil	The total expenditures divided by total FTE average membership for the district in this year	Feature
Students	The count of students in the district this year	Feature
Students Disciplined	The count of students disciplined in the district this year	Feature

% In-School Suspension	The percent of students disciplined with an in-school suspension	Feature
% Out-of-School Suspension	The percent of students disciplined with an out-of-school suspension	Feature
% Expulsion	The percent of students disciplined with an expulsion	Feature
% Removed to Alternate Setting	The percent of students disciplined with a removal to an alternate setting	Feature

*** Some data descriptions are partial quotes from the Massachusetts Department of Education's description of the categories
GitHub Repo

Based on what we discussed regarding machine learning (Week 07 Day 02), does your dataset include a set of feature variables and one outcome variable that you can use for a supervised machine learning task?

Our dataset contains several different sets of feature variables. Some examples of accessible feature variables include the racial percentages of students in each district, discipline data for each district, and information about in-district expenditures per pupil. We also have one output variable that we can use for a supervised machine learning task — in this case, the variable is the percentage of Massachusetts high school graduates attending an institution of higher education.

Due to the wide scope of the sets of feature variables that we have available to us, we have the opportunity to input several different chosen feature variables about a given school district — therefore, we have multiple possible regression calculations available. In each case, the ML task will perform this regression calculation to output the percentage of students in the given school district that will attend college or university.