Education and Immigration in California

California has been beacon for immigrants from all over the world since its induction into the union in 1850. Over a quarter of the states population are immigrants. Education is essential for social mobility. For Americans born into born into families in the bottom quintile of income, the greatest predictor of whether they will stay in that quintile is whether they graduate high school. We see differences in academic achievement of native born versus immigrant students. There are a number of possible explanations, the simplest being, that immigrants have less wealth that is correlated to academic success. There is also however great variance in the success of immigrants. By understanding how social and demographic factors affect the educational outcomes of immigrants, we can create a model to predict the success of an immigrant student for all the schools in california,The ultimate goal of this project is to improve educational outcomes for migrant students. This is accomplished in two ways, the first is to use the model to identify schools performing significantly above projection. These schools should be more closely examined to determine how other schools might try to emulate that success. The second is to advise migrants considering where to settle based on the educational future of their children.

There are several factor i would like to account for when creating my model.

- Standardized Testing Scores, to judge the academic success of any school, we will be looking at the result of the yearly star test that has averages for each school as well as for exclusively migrant students.
- Income, wealth is a huge factor in determining academic success, when comparing schools we want to keep in mind whether they fall in a wealthy area that will likely affect our scores.
- Per Pupil Spending, Because schools public schools are funded primarily through property taxes, area income and per pupil spending are correlated. We are still concerned with Per Pupil Spending because it will help indicate schools with greater academic resources.
- Staff Educational Demographics, information about how long the average teacher at each district has been teaching may give us some indication of the quality of teachers.
- Staff Racial Demographics, there have been studies that indicates people of color have better medical outcomes when their medical staff are racial similar, this will allow us to see how immigrants' education is affected by the demographics of the their instructors
- Free Meal Program, food insecurity is far more common in immigrants than in the native born population. This will allow us to see how public benefits like food assistance can affect education.
- Sanctuary Status, This will examine how a counties sanctuary status can impact the education of the migrants that live there.

Overview

Name	URL	Description
California Assessment of Student Performance and Progress (CAASPP) Research File	https://caaspp.cde.ca.gov/caaspp2016/ResearchFileList.aspx?rf=True&ps=True	This file contains the average test scores for every school in california. This file contains the averages scores of the migrant students at those schools. To protect student privacy the file does not have the averages for classes of students with less than 10 people. This file also contains an Entities file that list all schools with their zip codes.
Staff Data Files	https://www.cde.ca.gov/ds/sd/df/	This will contain an anonymized list of all the faculty, each entry has that employees race and highest degree.
Current Expense of Education	https://www.cde.ca.gov/ds/fd/ec/currentexpense.asp	This file contains the annual budget of each school district divided by the the amount of pupils.
Student Poverty – Free or Reduced Price Meals Data	https://www.cde.ca.gov/ds/sd/sd/fsspfrpm.asp	This file contains the totals for students eligible for reduced price and free meals per schools. The key features from this file are the percentage eligible for reduced priced meals, and the percentage that actually utilizes the program.
Sanctuary Counties List	https://cis.org/Map-Sanctuary-Cities-Counties-and-States	A list of which counties have a policy to not hand over criminals to immigration enforcement.

California Assessment of Student Performance and Progress (CAASPP) Research File

Column Name	Type	Description
District Code	int	Unique id for the district that school belongs to.
School Code	int	Unique id for the school
Grade	int	The grade of the test average
Subgroup ID	int	This allows us to select only tests from immigrant students.
Mean Scale Score	int	The average score for the group, this will be our target variable when predicting student success.

Current Expense of Education

Column Name	Type	Description
CDS	int	District Code, used to join on
Current Expense ADA	int	An estimation of the total number of students taught in a year
EDP 365	int	Total expenses for that district for the year
Current Expense Per ADA	int	Yearly expenditure for a single student

Student Poverty – Free or Reduced Price Meals Data

Column Name	Type	Description
School Code	int	Unique id for the school
Percent (%) Eligible FRPM (K-12)	int	Percentage of students who could receive meal assistance
Enrollment(K-12)	int	The total number of students at the school
FRPM Count (K-12)	int	Total number of students who do utilize meal assistance, this will be used to calculate the percentage of students who utilize meal assistance

Staff Data Files

Column Name	Type	Description
District Code	int	Unique id for district
Ethnic Group	int	Ethnicity of employee, will be used for computing the ethnic makeup of a district
YearsTeaching	int	Number of years the teacher has been teaching, will be used to aggregate for a district value

Using these features, the project will create a regression to predict the mean scale score for the migrant subgroup for each school. The training and testing sets will have a layout as follows.

Training Data Set

The training file should contain the following information for each school with at least one class of ten migrants.

Column Name	Type	Derived	Description	Origin	renamed
Migrant_Score	int	yes	This column represents the average score for migrants at each school where the data is available	California Assessment of Student Performance and Progress (CAASPP) Research File	Mean Scale Score
All_Score	int	yes	This column represents the average score for every student at every schools where the migrant data is also available	California Assessment of Student Performance and Progress (CAASPP) Research File	Mean Scale Score
Per_Student_Spending	int	no	This column is the per student spending of the school's district	Current Expense of Education	Current Expenses Per ADA
Food_Aid_Capable_Percentage	int	yes	The percent of children at the school eligible for food assistance	Student Poverty – Free or Reduced Price Meals Data	

Food_Aid_Actual_Percentage	int	yes	The percent of children at a school utilizing food assistance	Student Poverty – Free or Reduced Price Meals Data	
School_Size	int	no	Total size of the school in students	Student Poverty – Free or Reduced Price Meals Data	Enrollment(K-12)
Average_Years_Teaching	int	yes	The average amount of experience teachers have in the schools district	Staff Data Files	
Sanctuary	int	yes	Whether the school is in a sanctuary county		Sanctuary Counties List
Percent_Ethnicity_X	int	yes	For the 10 integer values that Ethnic Group could take, there is a column that gives what percentage of that school's district's staff are that ethnicity	Staff Data Files	

It is important to note that this training set does not include records for every california school. There are a great number of schools for which there are not enough migrant students to release the data. All other data besides migrant scores is available for every school.

There are two desirable outcomes from this model. The first requires us to feed in a sample of our migrant schools list into the regression. Next, having the model predict the Migrant_Score for the schools that weren't used for training, those schools that exceed their predicted results by a statistically significant margin are candidates for further study. These schools are likely to have other features that would account for their success.

The intended user for this sort of model is a school or district official looking to improve the educational outcomes for their migrant students by simulating those districts and schools whose performance exceed their demographics.

The second use for this regression is to give a more complete picture of how different schools would educate migrants. There are many schools in california where it is not publicly available how well the migrant students are doing. This would complicate the role of a concerned migrant parent who wants to ensure their child is in a school district where people from their demographic background can succeed. By this time training our model on the entire migrant schools training set and having the model predict the Migrant_Score of schools using only the school information and the All_Score, an estimation can be made for the success of migrant children at each school. This gives the migrant parent more security when making decisions about their child's educational future.