Analysis of National Graduation Rate

Subhabrata Mukherjee

Oct 2017

Introduction

The objective of this analysis:

- Analyze graduation data across all states in United States
- Compare graduation rate to discover pattern by socio-economical factors
- Create predictive models to forecast graduation rate for a school district

Data sets

- EDFacts data for school year 2011 and 2014
  - Graduation data for 10,000+ school district
  - Math & reading/language art assessments by race/ethnicity
- Census data for school year 2011 and 2014
  - Small Area Income and Poverty Estimates for all 13,000+ school district
- Elementary/Secondary data for school year 2011 and 2014
 - Basic statistical data on U.S. schools

DATA EXPLORATION

Graduation rate by Race and Year

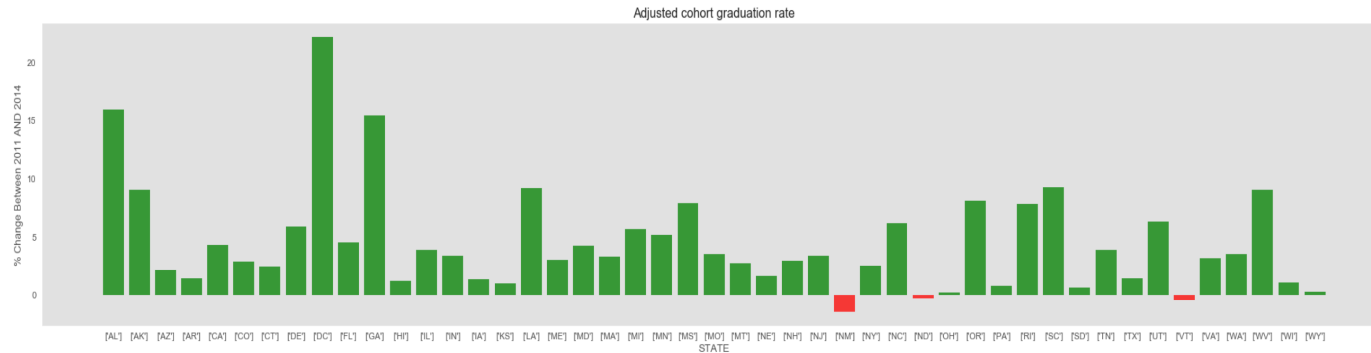Adjusted cohort graduation rate by race and year
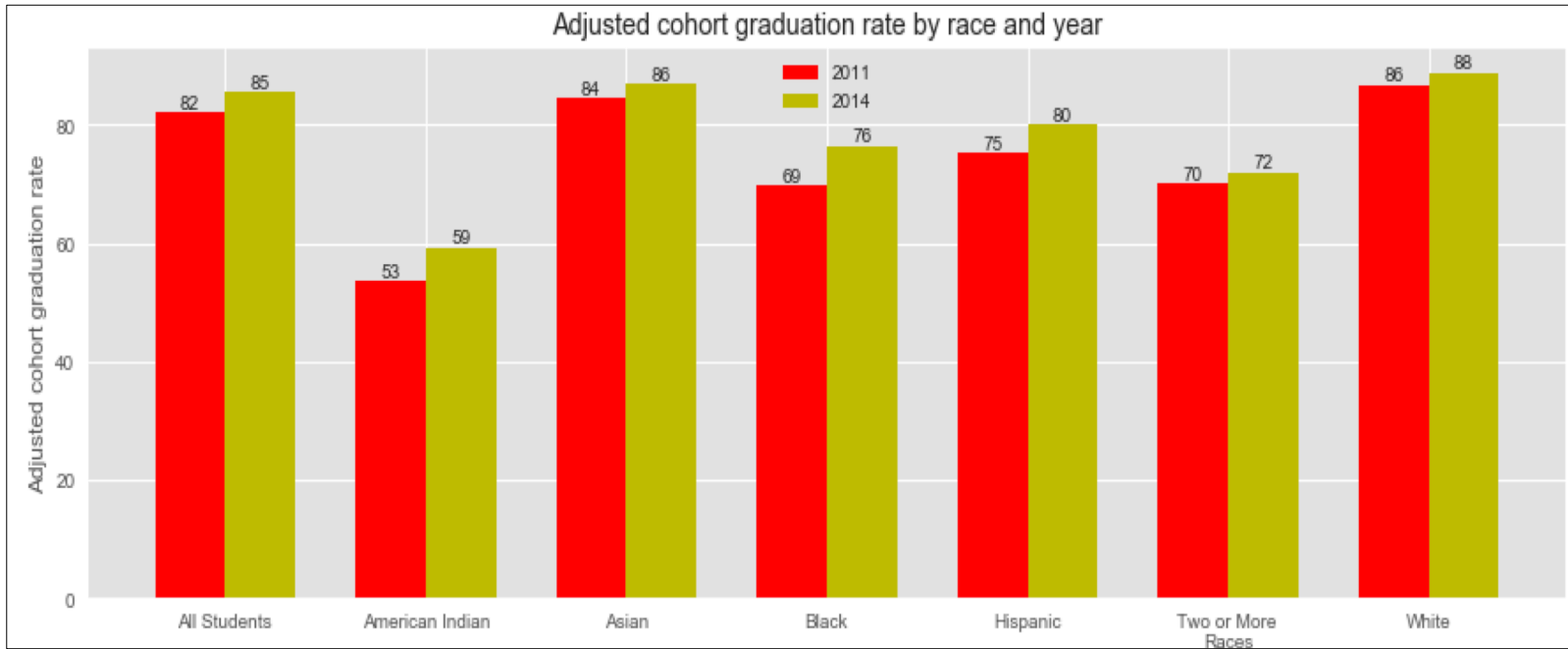
- ❑ In 2014-15 Graduation rate increased to 85%, an increase from 82% in 2011-12.
- ❑ Graduation rate of white students: 88.0%
- ❑ Graduation rate of Asian students: 86.0%
- ❑ Graduation rate of Hispanic students: 80.0%
- ❑ Graduation rate of Black students: 76.0%
- ❑ Graduation rate of American Indian students: 59.0%

Graduation rate by Race and Year
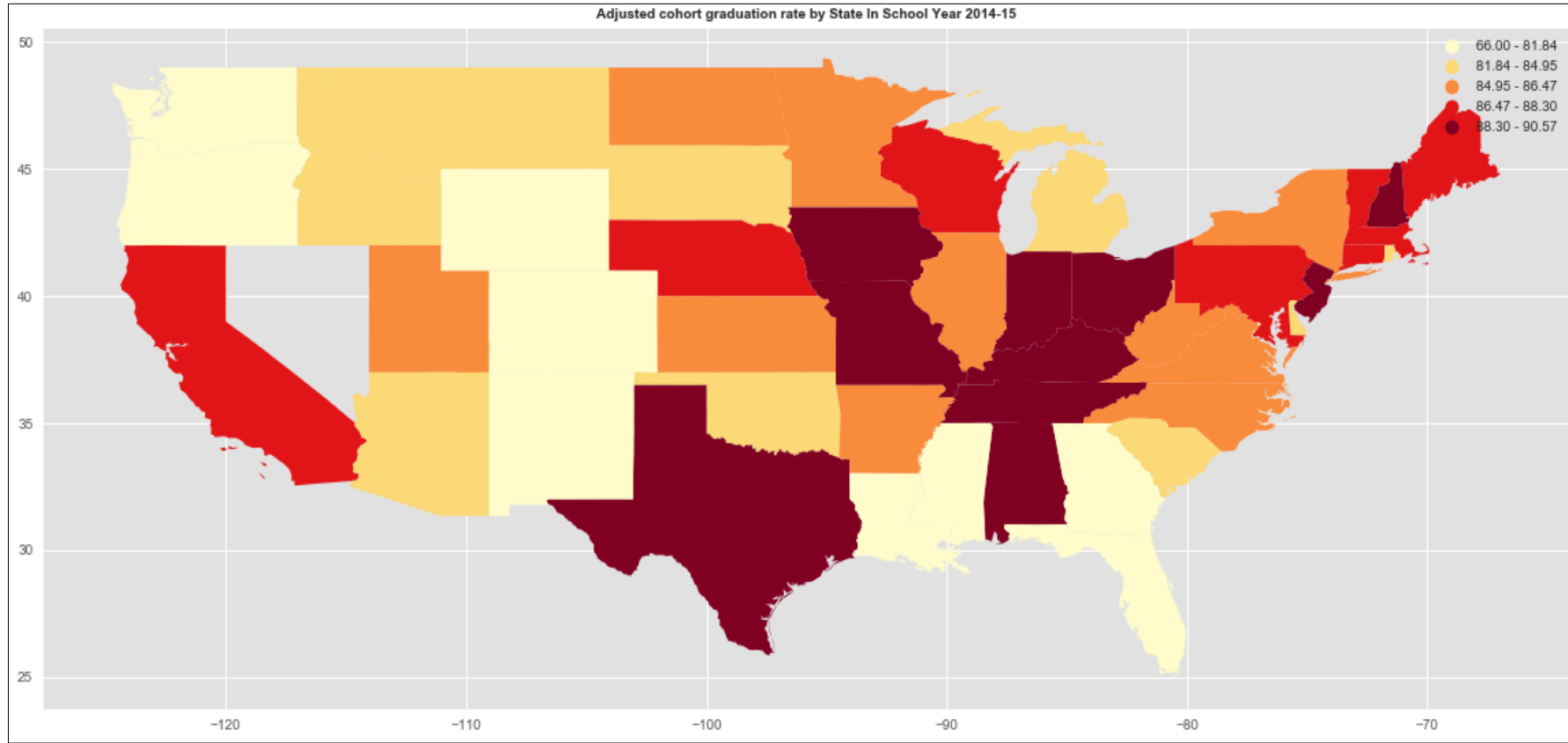
Adjusted cohort graduation rate by State in School Year 2014-15

School district in Northeast and South region is doing better than the school district in Midwest and west region

Graduation rate by State and Year

Adjusted cohort graduation rate

- ❑ States with negative growth are New Mexico, North Dakota, Vermont
- ❑ Three of the top five states having lowest graduation rate matched in 2011 and 2014
- ❑ Two of the top five states having highest graduation match in 2011 and 2014

Top five states having lowest graduation rate in 2011 and 2014:

2011	
STATE	**% Graduated**
DISTRICT OF COLUMBIA	54.01
NEVADA	64.38
OREGON	69.01
ALASKA	69.08
GEORGIA	70.60

| 2014	
STATE	**% Graduated**
DISTRICT OF COLUMBIA	66.00
NEW MEXICO	69.81
OREGON	74.62
ALASKA	75.33
FLORIDA	77.88

Top five states having highest graduation rate in 2011 and 2014:

2011	
STATE	**% Graduated**
WISCONSIN	87.29
VERMONT	87.78
OHIO	88.13
IOWA	88.67
TEXAS	88.98

2014	
STATE	**% Graduated**
MISSOURI	89.76
INDIANA	89.87
IOWA	89.91
TEXAS	90.25
TENNESSEE	90.57

State wise co-relation

- ❑ A strong correlation of 0.9 in adjusted cohort graduation rates for both years
- ❑ Three states are outliers
  - ❑ New Mexico
  - ❑ Georgia
  - ❑ Alabama
- ❑ New Mexico is the only state which is showing negative growth in adjusted cohort graduation rate.

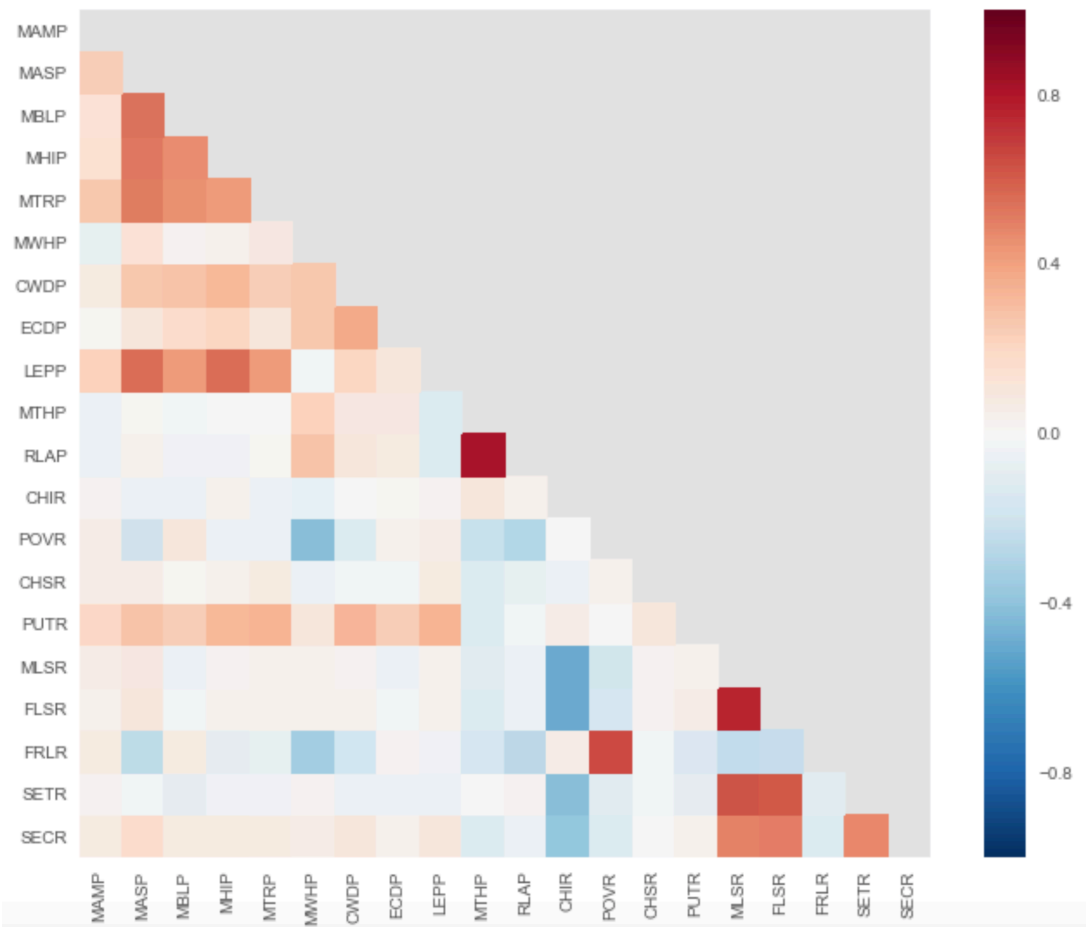# Graduation Rate vs All Race
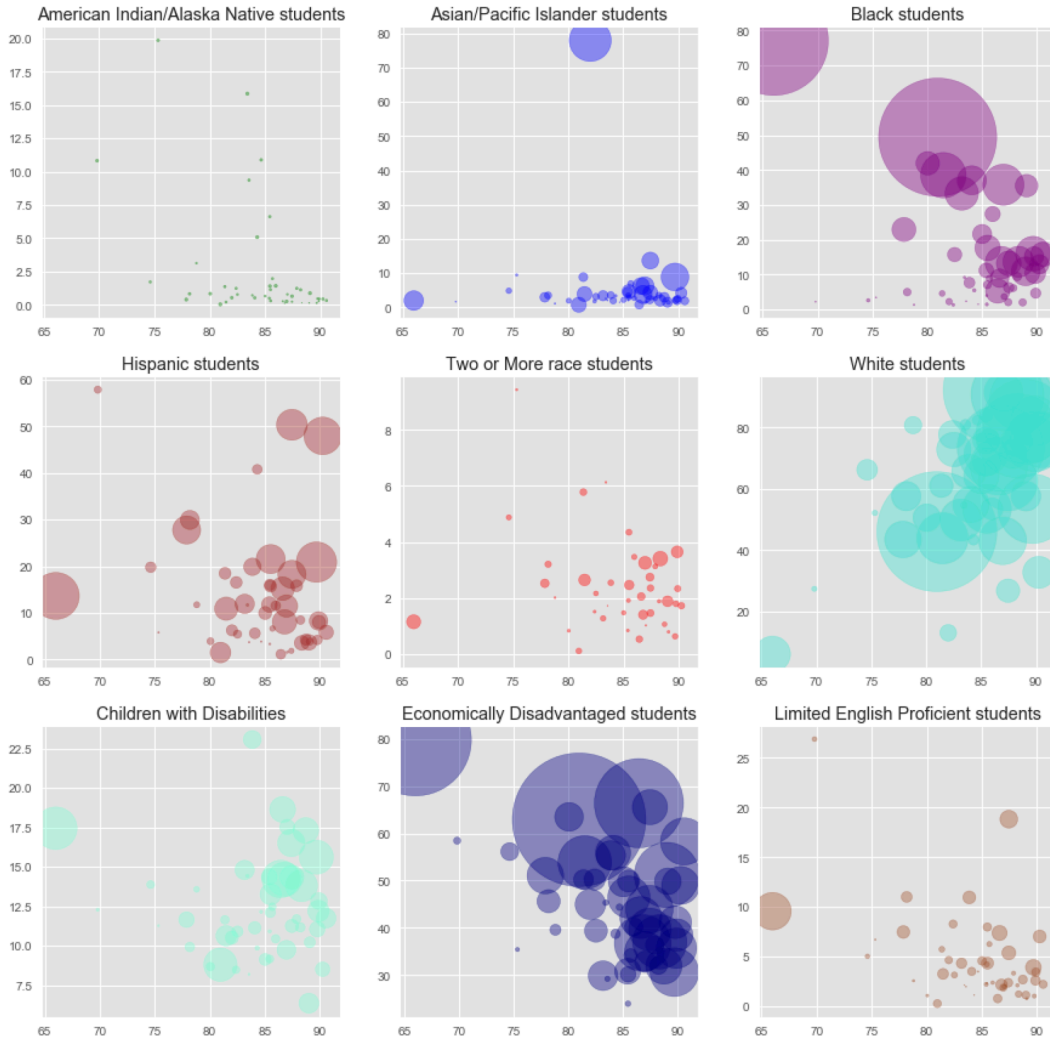
Relationships found:
- ❑ As % of Black students increases the graduation rate increases
- ❑ As % of White students increases the graduation rate increases
- ❑ As % of Asian students increases the graduation rate remains same

PREDICTIVE MODELS

Linear Regression

| Selected 12 features	
% Passed - American Indian	% Passed - Limited English proficient
% Passed - Asian	% Proficient - Math
% Passed - Black	% Proficient - ELA
% Passed - Hispanic	% of Children under poverty
% Passed - White	% of Students on free/reduced lunch
% Passed - Economically disadvantaged	% of secondary teachers

Correlation matrix shows some relationships but none strong enough to skew the model

Linear Regression


Actual vs. predicted graduation rate
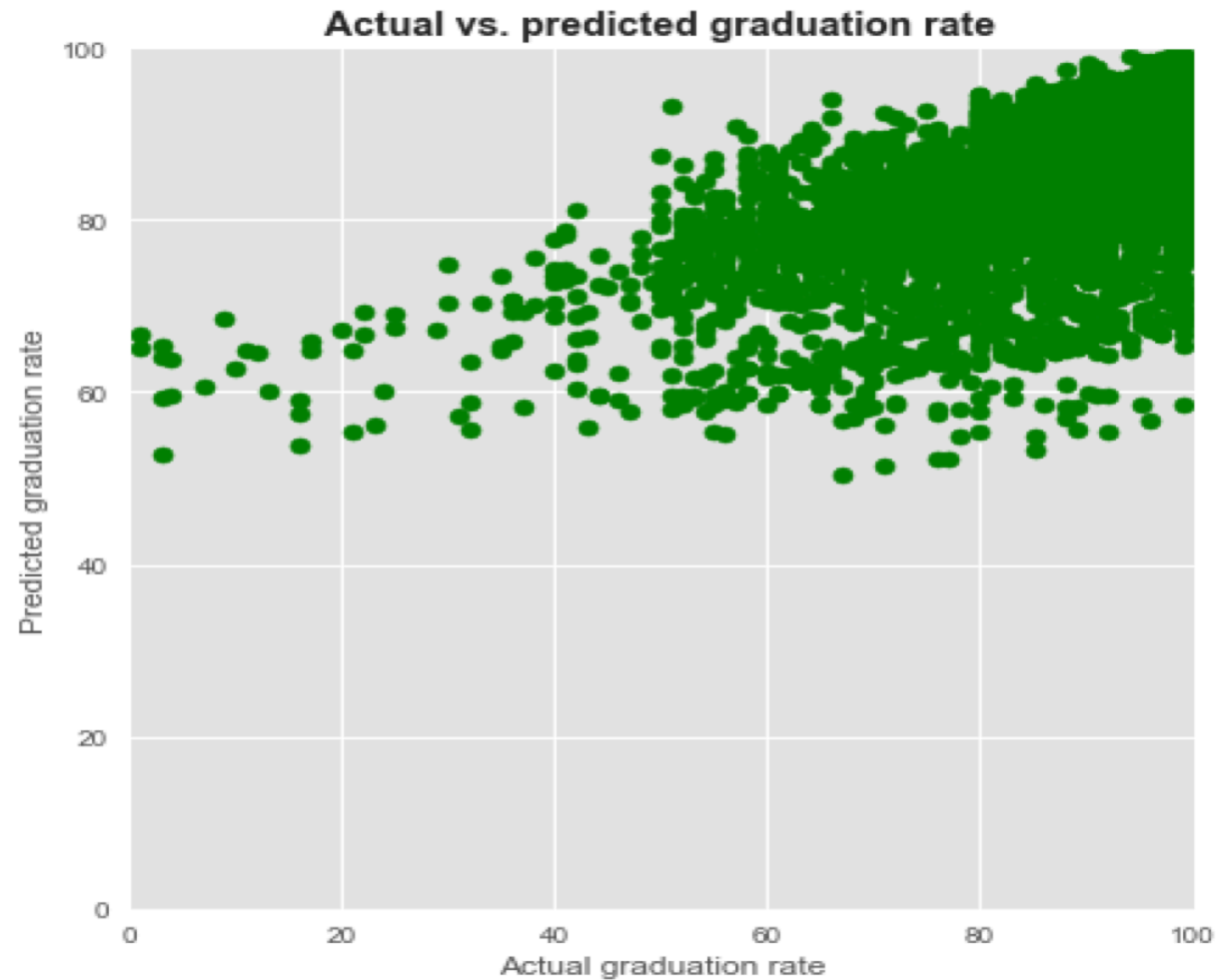
- ❑ Model explains 45% of the variance in the data – a moderate value for fit
- ❑ Training data has a MSE = 63.76 and test data has a MSE = 64.42
- ❑ Linear regression is not a good choice for this data set

Random Tree Regression

Feature importance of Random Forest regression on Graduation Rate
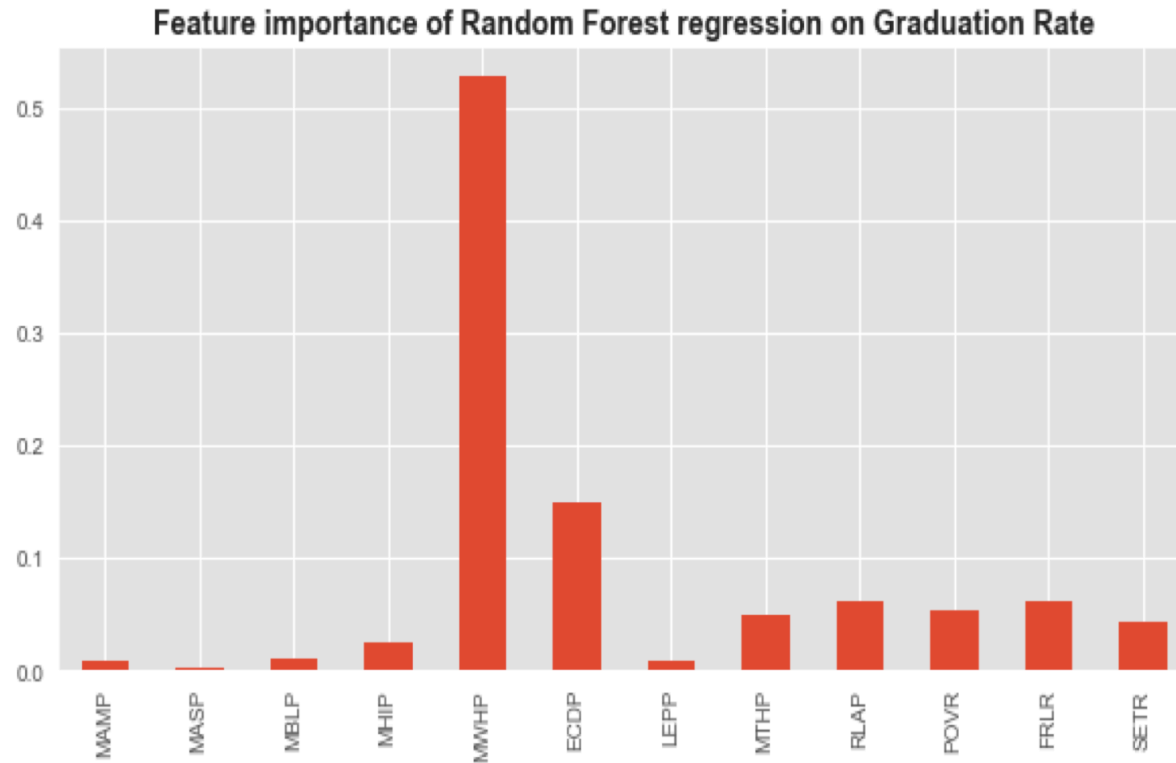
❑ Model explains 67% of data's variance but does not generalize well to the test set
❑ % of White students and % of students Economically disadvantages are most important features during training

Gradient Boosting Regression

Variable Importance
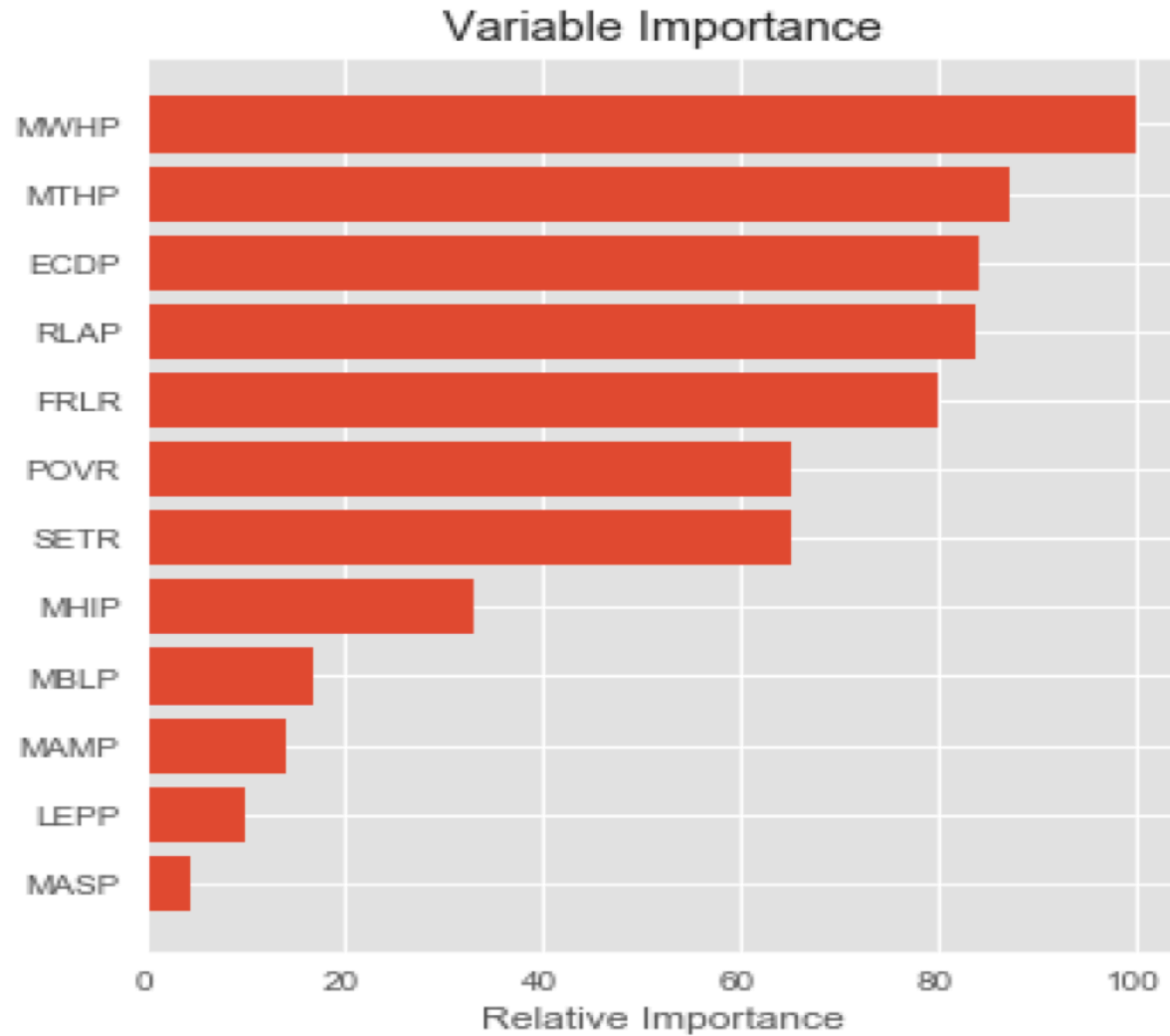
❑ Model explains 70% of data's variance but does not generalize well to the test set
❑ % of White students and % of students proficient in Math are most important features during training

Recommendations and future work

- **For the District:**
  - Conduct a study to determine the root cause why the school district in Northeast and South region is doing better than the school district in Midwest and west region. Is there any evidence that re-zoning these districts will allow the overall graduation rate to increase further?

- **For State:**
  - Investigate why New Mexico, North Dakota, Vermont are showing negative growth in graduation rate. Using smaller geographic areas as suggested above, plot graduation rate to determine if these states are showing any trend by location. Use this data to create a corrective action to improve the graduation rate.

Future opportunities to continue this work include further training of the regression model on smaller district areas with more accurate dataset to determine if the regional division holds on a larger data set.