IDS DATASCIENCE FOR GOOD : PASSNYC

Team:

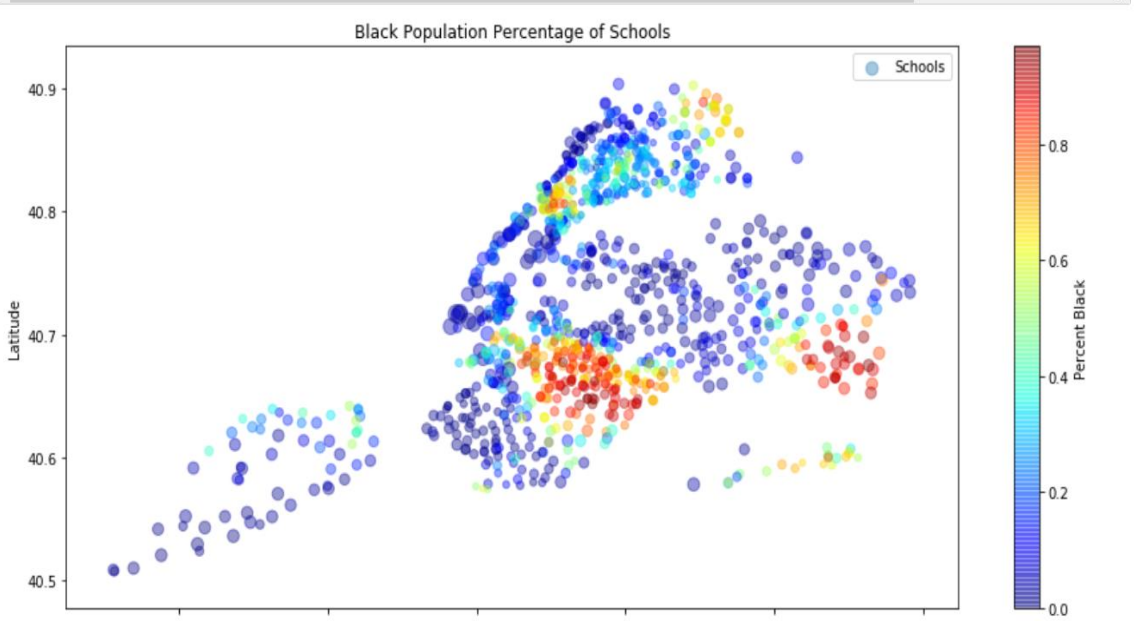
Srinath M P: PES1201801946

Sai Eashwar K S: PES1201801910

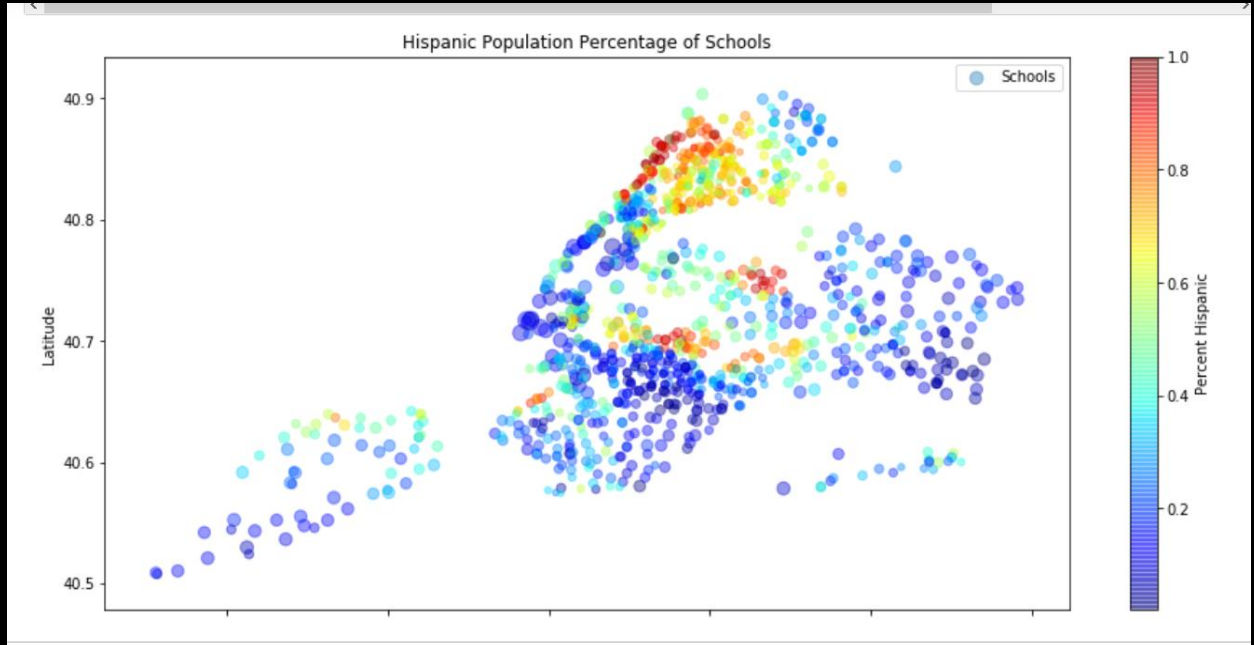
CLEANING

- Removed outliers
- Removed '%' symbol in many columns and converted them to float
- Removed '$', ',' and spaces in 'School Income Estimate'
- Filled missing numerical values with the mean of the column

VISUALIZATIONS



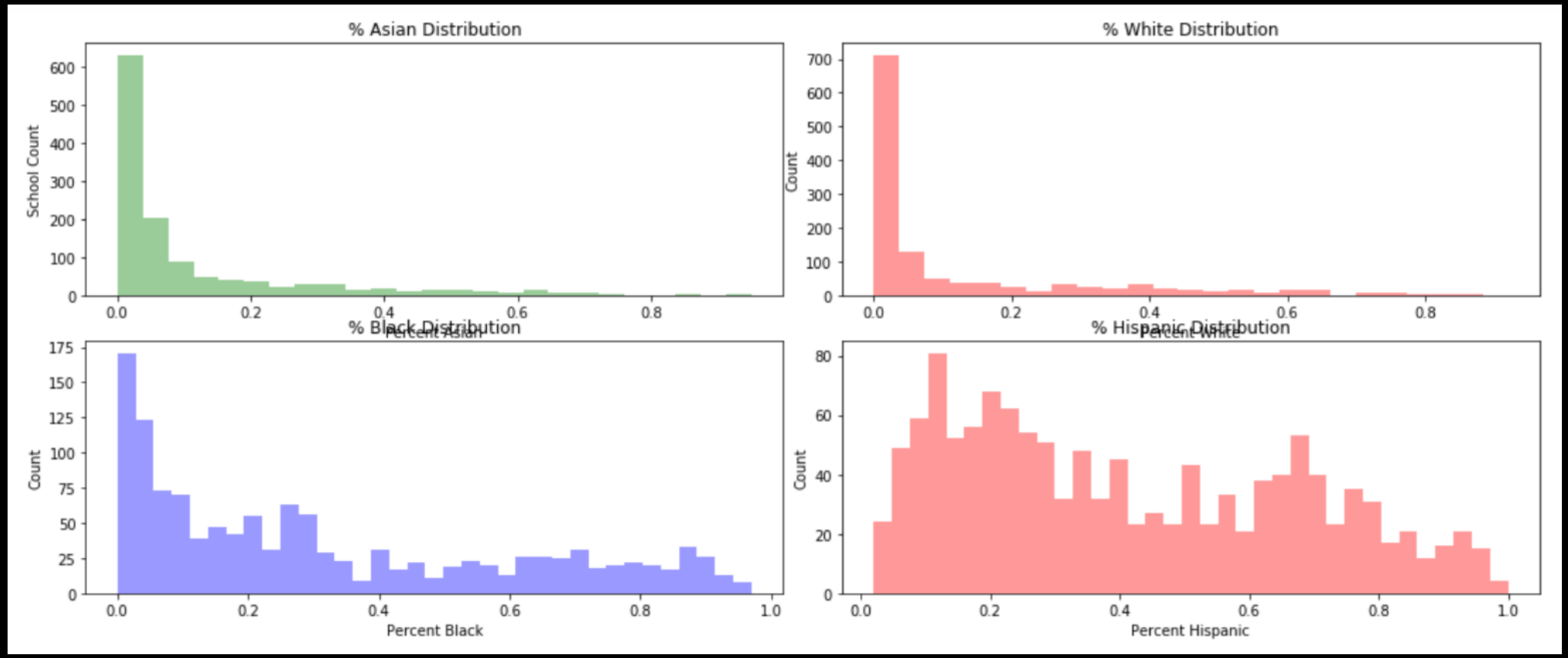
Central New York contains most of the Schools with Blacks: 1) As we can see here that the Majority of the black population are living in Central New York . 2)If you compare this to the Hispanic Population, they are totally isolated from each other.(Red)



Majority of the hispanic population are living in Upper New York. Here we notice that there are a few Hispanics in Central New York while there are a majority of Blacks in Central New York.

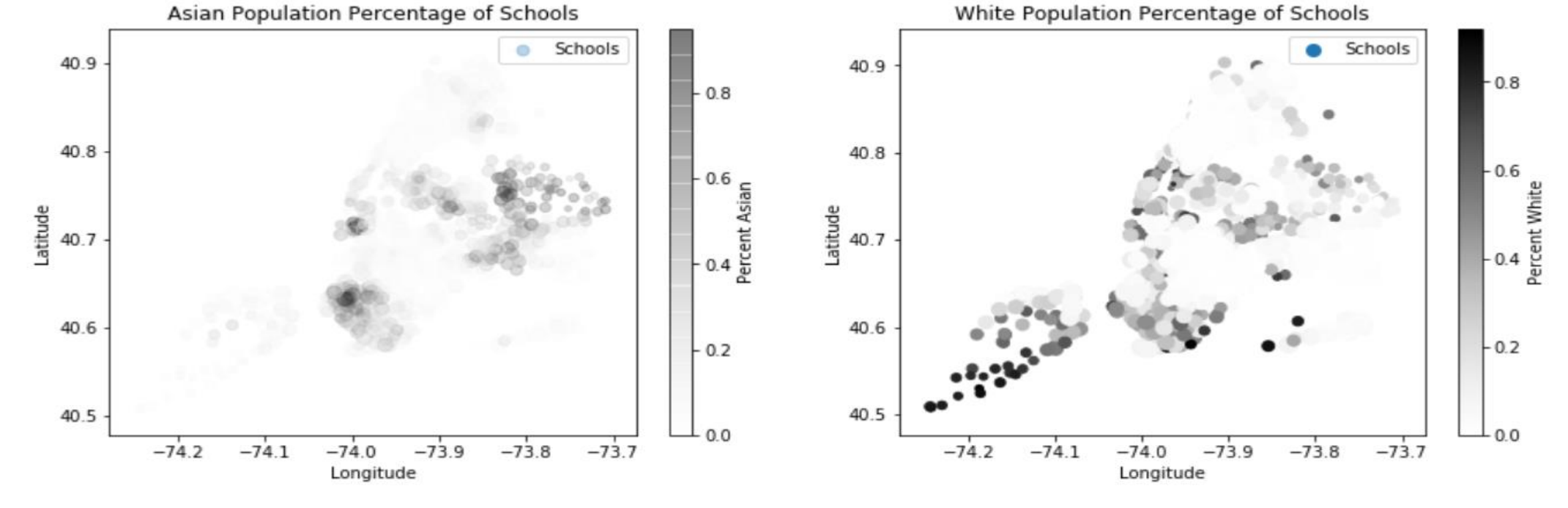
We can see that in Central and Upper New York, there is very less diversity

VISUALIZATIONS



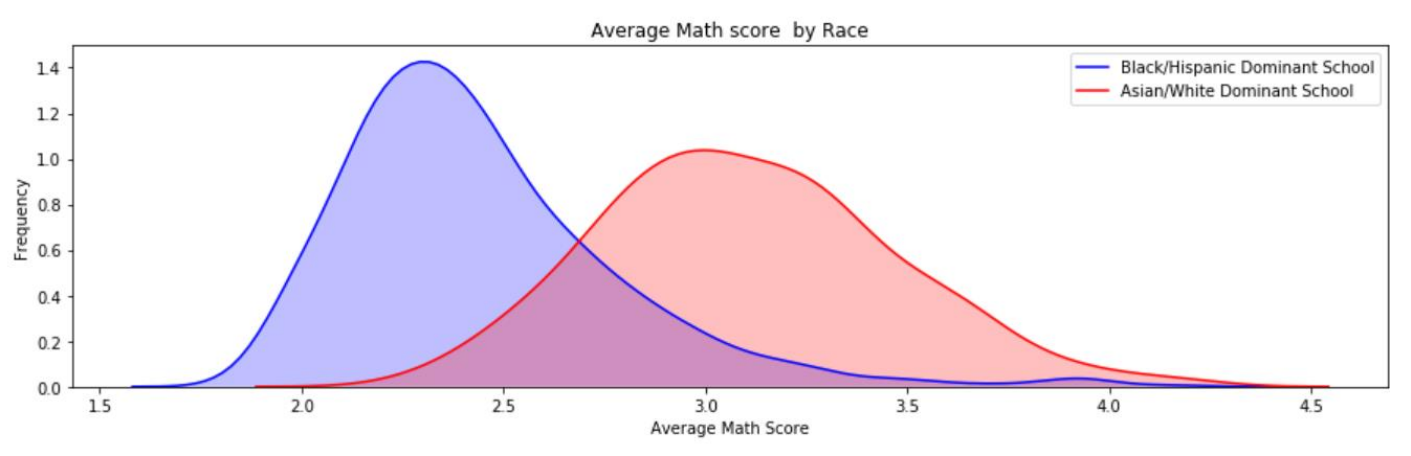
We can see that the population is as follows : 1.Hispanic 2.Black 3.Asian 4.White
Also we can see that majority of Whites and Asians are representing approximately 10% of the school's population

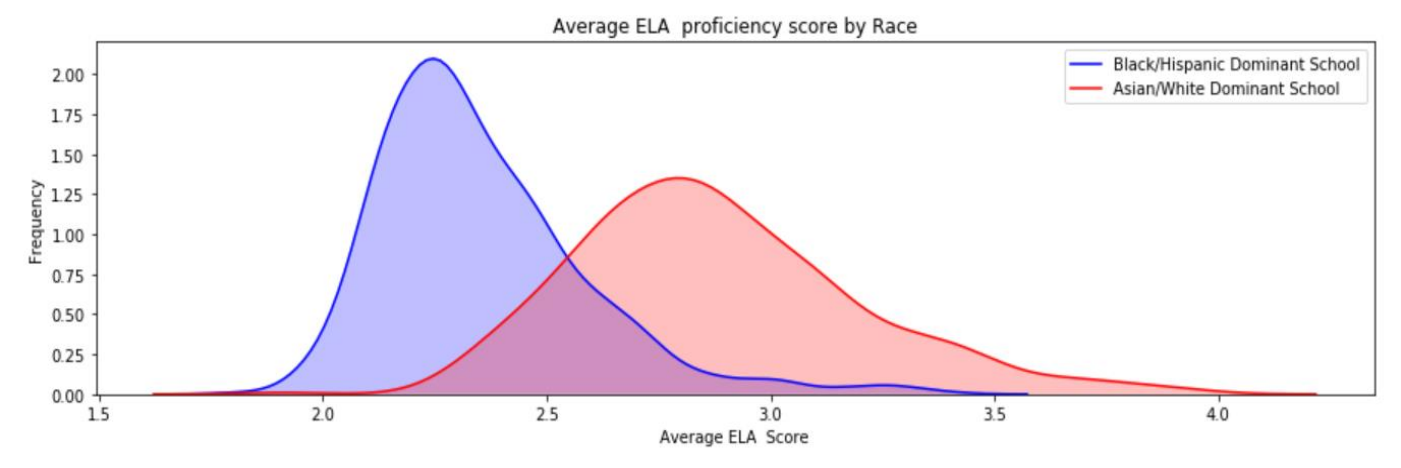
VISUALIZATIONS



- No school has a very high percentage of Asians, The percentage of Asians is relatively high in Central NY
- The White Population has a very noticeable high density in south-eastern and a few southern schools

VISUALIZATIONS





We can infer the following:

- Black and Hispanic dominant schools have an Average ELA Score of 2.35
- Black and Hispanic dominant schools have an Average Math Score of 2.44.
- White and Asian dominant schools have an Average ELA Score of 3.04
- White and Asian dominant schools have an Average Math Score of 3.32

We can also see that the scores in White and Asian dominant schools is more spread out

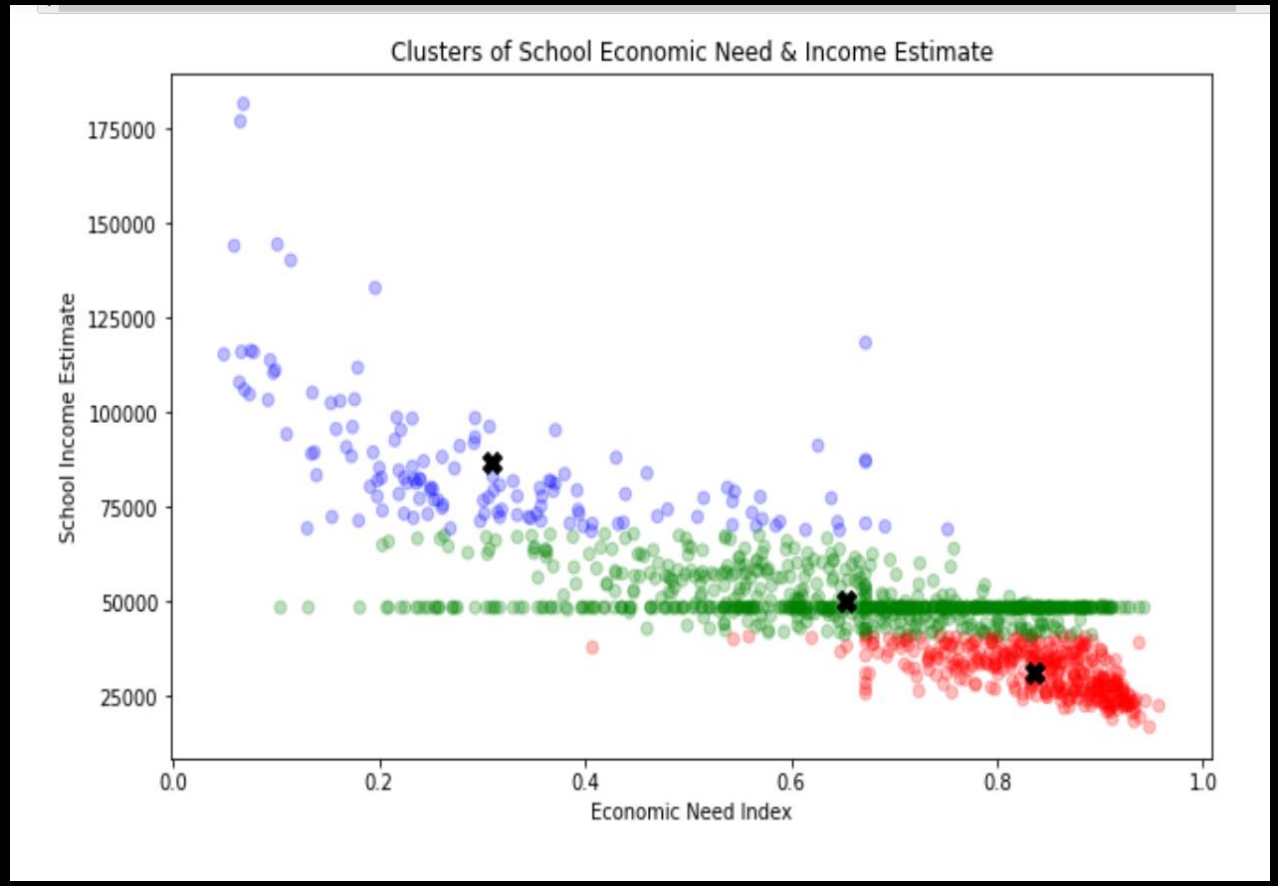
VISUALIZATIONS

We can see three clusters here:
1: The red represents the schools with greater economic need but lower income(hence are very much in need of Income)

2: The green represents the schools with moderate economic need and moderate Income estimate(hence are in moderate need of Income)

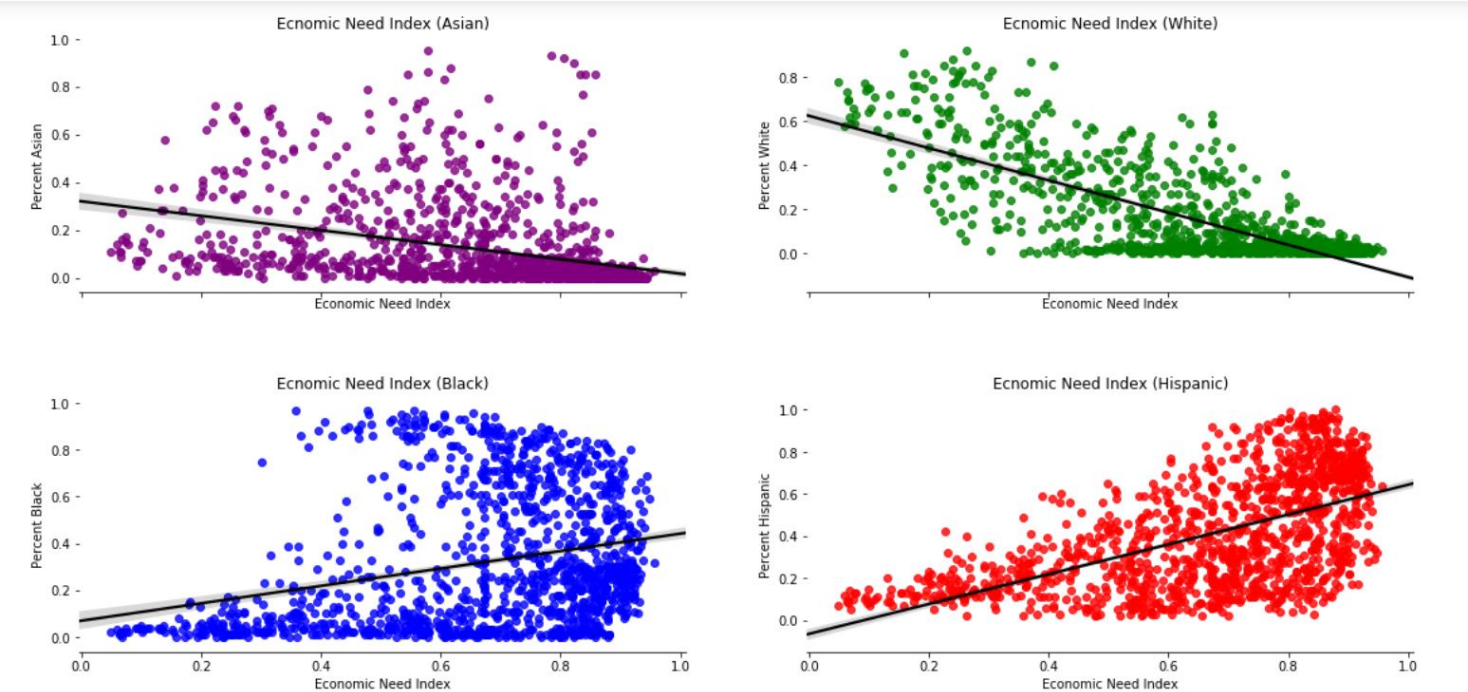
3: The blue represents the schools with low economic need ,but have surplus income. Hence are in no need of any additional income

Hence we can visually see that schools with lower income have greater Economic Need

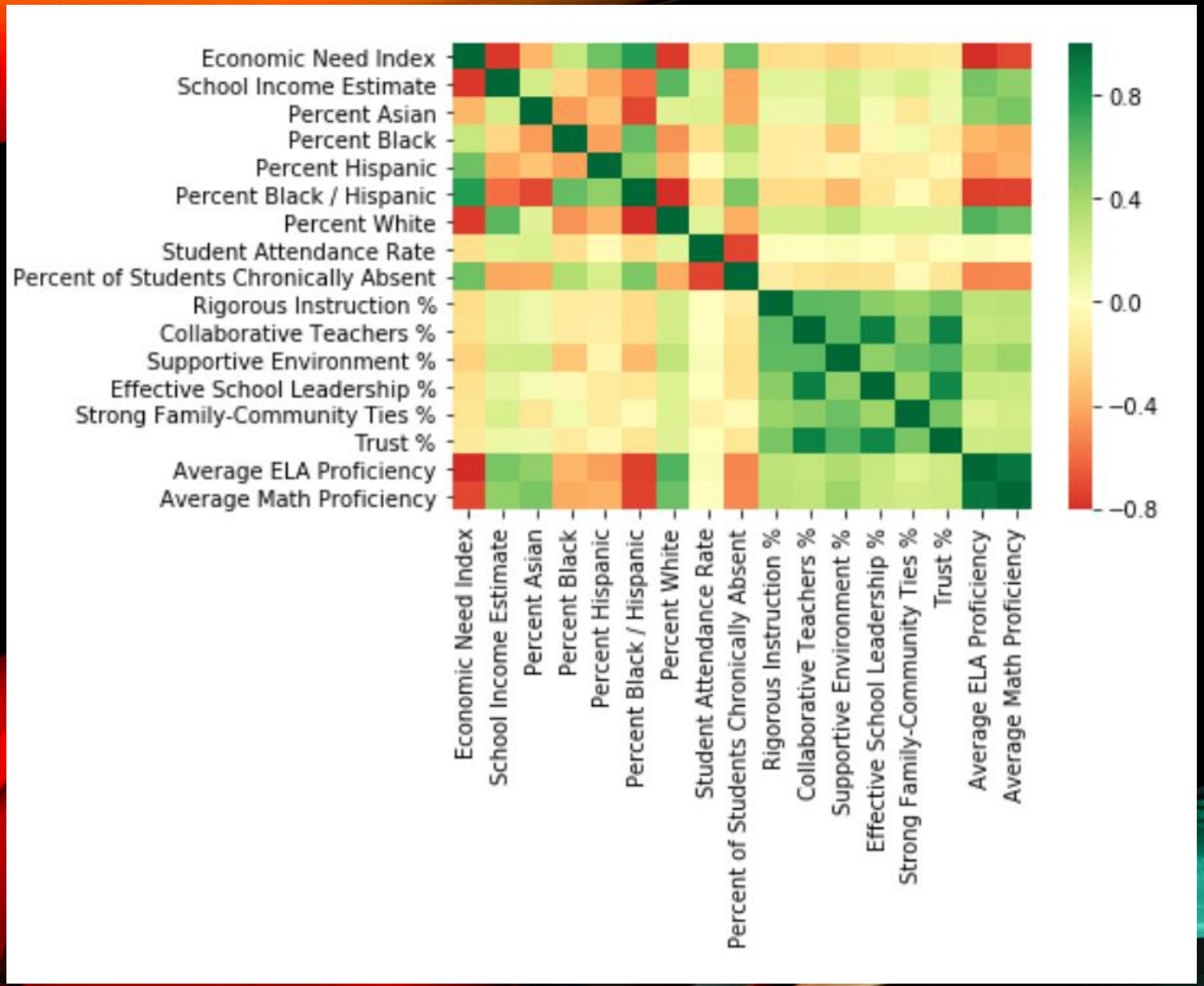


HYPOTHESIS TESTING

- M1 : mean Economic Need Index of schools with higher White and Asian populations
- M2 : mean Economic Need Index of schools with higher Black and Hispanic populations
- H0 : M2 – M1 <= 0; H1 : M2 – M1 > 0
- P-value ~ 0
- Conclusion : Schools with higher Black and Hispanic populations have higher Economic Need Index

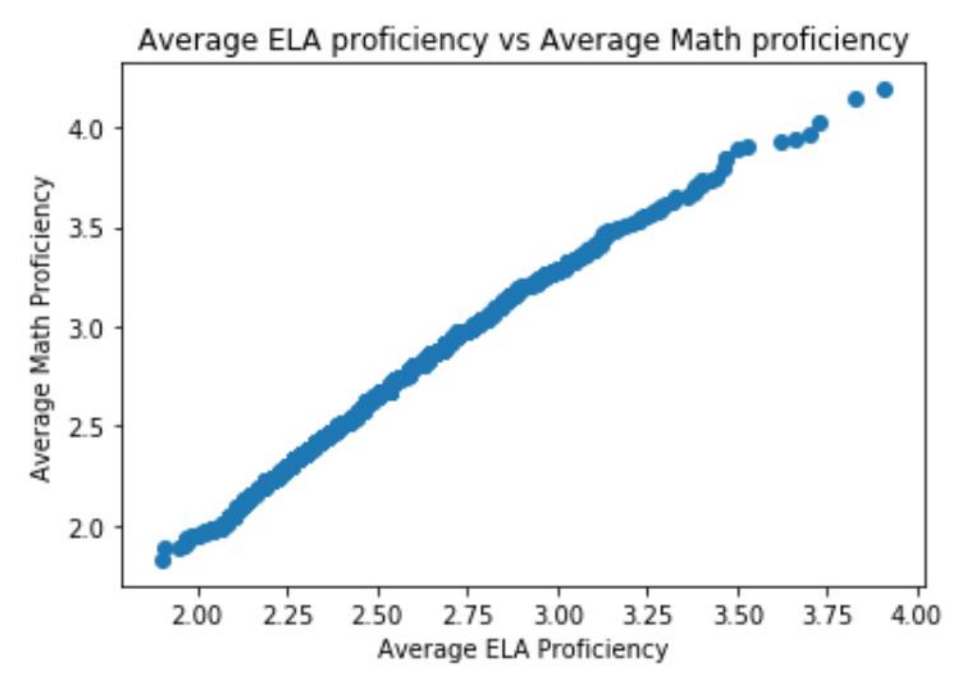


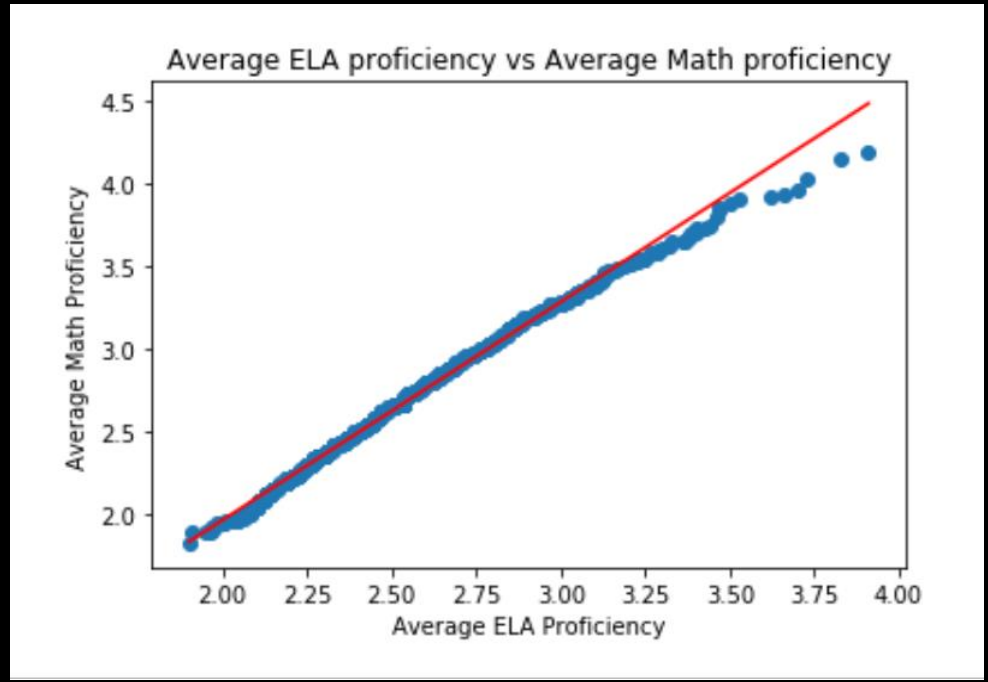
Schools with a Higher White and Asian** Population tend to have a Lower Economic Need Index Score. Schools with a Higher Hispanic and Black Population tend to have a Higher Economic Need Index Score.



CORRELATION MATRIX

SIMPLE LINEAR REGRESSION





Predicted Average Math Proficiency = -0.668983982402051 + 1.3191596541420 52 * Average ELA Proficiency
Correlation coefficient = 0.9975