PES UNIVERSITY, Bangalore

(Established under Karnataka Act No. 16 of 2013)

UE18CS203

B.Tech, Sem III
Session : Aug-Dec, 2019

UE18CS203 – INTRODUCTION TO DATA SCIENCE

REPORT
ON
DATASCIENCE For Good:PASSNYC

SECTION :

#	SRN	Name	Contact No.	Email ID	Sign
1.	PES1201801946	Srinath M P	7349503609	srinathmp77@gmail.com	
2.	PES1201801910	Sai Eashwar KS	8660388306	saieashwar.ks	

				@gmail.com	
-	-	-	-	-	-

ABOUT THE DATA SET

PASSNYC uses public data to identify students within New York City's under-performing school districts and, through consulting and collaboration with partners, aims to increase the diversity of students taking the Specialized High School Admissions Test (SHSAT).

The dataset has a size of 1272 rows and 161 columns.

In this section, a short description of the meaning of each column can be found. Along with this, the type of data in the column is indicated.

1. Adjusted Grade
2. New?
3. Other Location Code in LCGMS
4. School Name
5. SED Code -State Education Department
6. Location Code
7. District
8. Latitude
9. Longitude
10. Address (Full)
11. City
12. Zip
13. Grades - The range of grade levels in this school
14. Grade Low - Lowest grade level in this school

15. Grade High - Highest grade level in this school
16. Community School?
17. Economic Need Index - (%temp housing) + (% HRA eligible *0.5) + (% free lunch eligible *0.5). The higher the index, the higher the need
18. School Income Estimate
19. Percent ELLELL = English Language Learners
20. Percent Asian
21. Percent Black
22. Percent Hispanic
23. Percent Black / Hispanic
24. Percent White
25. Student Attendance Rate - total number of days attended by all students / total number of days on register for all students
26. Percent of Students Chronically Absent - Missing 10% of school days - or 18 days+ per year in a 180-day school year
27. Rigorous Instruction % - How well the curriculum and instruction engage students, build critical-thinking skills, and are aligned to the Common Core
28. Rigorous Instruction Rating - - How well the curriculum and instruction engage students, build critical-thinking skills, and are aligned to the Common Core
29. Collaborative Teachers % - How well teachers participate in opportunities to develop, grow, and contribute to the continuous improvement of the school community
30. Collaborative Teachers Rating - - How well teachers participate in opportunities to develop, grow, and contribute to the continuous improvement of the school community
31. Supportive Environment % - How well the school establishes a culture where students feel safe, challenged to grow, and supported to meet high expectations
32. Supportive Environment Rating - - How well the school establishes a culture where students feel safe, challenged to grow, and supported to meet high expectations
33. Effective School Leadership % - How well school leadership inspires the school community with a clear instructional vision and effectively distributes leadership to realize this vision
34. Effective School Leadership Rating - - How well school leadership inspires the school community with a clear instructional vision and effectively distributes leadership to realize this vision
35. Strong Family-Community Ties % - How well the school forms effective partnerships with families to improve the school
36. Strong Family-Community Ties Rating - - How well the school forms effective partnerships with families to improve the school

37. Trust % - Whether the relationships between administrators, educators, students, and families are based on trust and respect
38. Trust Rating - Whether the relationships between administrators, educators, students, and families are based on trust and respect
39. Student Achievement Rating - Weighted Average Score + the Closing the Achievement Gap Additional Points - 4: Exceeding Target, 3: Meeting Target, 2: Approaching Target, 1: Not Meeting Target
40. Average ELA Proficiency - ELA = English Language Arts Performance Levels = 1 (insufficient), 2 (partial but insufficient), 3 (sufficient), and 4 (more than sufficient) - reflect the extent to which students demonstrate the level of understanding expected at their grade level, based on the New York State P-12 Common Core Learning Standards
41. Average Math Proficiency
42. Grade 3 ELA - All Students Tested - -- No. of students tested for ELA in 3rd grade
43. Grade 3 ELA 4s - All Students - -- No. of students in 3rd grade who scored a 4 in ELA
44. Grade 3 ELA 4s - American Indian or Alaska Native - -- No. of students in 3rd grade with this particular background who scored a 4 in ELA
45. Grade 3 ELA 4s - Black or African American - -- No. of students in 3rd grade with this particular background who scored a 4 in ELA
46. Grade 3 ELA 4s - Hispanic or Latino - -- No. of students in 3rd grade with this particular background who scored a 4 in ELA
47. Grade 3 ELA 4s - Asian or Pacific Islander - -- No. of students in 3rd grade with this particular background who scored a 4 in ELA
48. Grade 3 ELA 4s - White - -- No. of students in 3rd grade with this particular background who scored a 4 in ELA
49. Grade 3 ELA 4s - Multiracial - -- No. of students in 3rd grade with this particular background who scored a 4 in ELA
50. Grade 3 ELA 4s - Limited English Proficient - -- No. of students in 3rd grade with this particular background who scored a 4 in ELA
51. Grade 3 ELA 4s - Economically Disadvantaged - -- No. of students in 3rd grade with this particular background who scored a 4 in ELA
52. Grade 3 Math - All Students tested - -- No. of students in 3rd grade tested for Math
53. Grade 3 Math 4s - All Students - -- No. of students in 3rd grade who scored a 4 in math
54. Grade 3 Math 4s - American Indian or Alaska Native - -- No. of students in 3rd grade with this particular background who scored a 4 in math
55. Grade 3 Math 4s - Black or African American - -- No. of students in 3rd grade with this particular background who scored a 4 in math

56. Grade 3 Math 4s - Hispanic or Latino - -- No. of students in 3rd grade with this particular background who scored a 4 in math
57. Grade 3 Math 4s - Asian or Pacific Islander - -- No. of students in 3rd grade with this particular background who scored a 4 in math
58. Grade 3 Math 4s - White - -- No. of students in 3rd grade with this particular background who scored a 4 in math
59. Grade 3 Math 4s - Multiracial - -- No. of students in 3rd grade with this particular background who scored a 4 in math
60. Grade 3 Math 4s - Limited English Proficient - -- No. of students in 3rd grade with this particular background who scored a 4 in math
61. Grade 3 Math 4s - Economically Disadvantaged - -- No. of students in 3rd grade with this particular background who scored a 4 in math
62. Grade 4 ELA - All Students Tested - -- No. of students tested for ELA in 4th grade
63. Grade 4 ELA 4s - All Students - -- No. of students in 4th grade who scored a 4 in ELA
64. Grade 4 ELA 4s - American Indian or Alaska Native - -- No. of students in 4th grade with this particular background who scored a 4 in ELA
65. Grade 4 ELA 4s - Black or African American - -- No. of students in 4th grade with this particular background who scored a 4 in ELA
66. Grade 4 ELA 4s - Hispanic or Latino - -- No. of students in 4th grade with this particular background who scored a 4 in ELA
67. Grade 4 ELA 4s - Asian or Pacific Islander - -- No. of students in 4th grade with this particular background who scored a 4 in ELA
68. Grade 4 ELA 4s - White - -- No. of students in 4th grade with this particular background who scored a 4 in ELA
69. Grade 4 ELA 4s - Multiracial - -- No. of students in 4th grade with this particular background who scored a 4 in ELA
70. Grade 4 ELA 4s - Limited English Proficient - -- No. of students in 4th grade with this particular background who scored a 4 in ELA
71. Grade 4 ELA 4s - Economically Disadvantaged - -- No. of students in 4th grade with this particular background who scored a 4 in ELA
72. Grade 4 Math - All Students Tested - -- No. of students in 4th grade tested for Math
73. Grade 4 Math 4s - All Students - -- No. of students in 4th grade who scored a 4 in math
74. Grade 4 Math 4s - American Indian or Alaska Native - -- No. of students in 4th grade with this particular background who scored a 4 in math
75. Grade 4 Math 4s - Black or African American - -- No. of students in 4th grade with this particular background who scored a 4 in math

76. Grade 4 Math 4s - Hispanic or Latino - -- No. of students in 4th grade with this particular background who scored a 4 in math
77. Grade 4 Math 4s - Asian or Pacific Islander - -- No. of students in 4th grade with this particular background who scored a 4 in math
78. Grade 4 Math 4s - White - -- No. of students in 4th grade with this particular background who scored a 4 in math
79. Grade 4 Math 4s - Multiracial - -- No. of students in 4th grade with this particular background who scored a 4 in math
80. Grade 4 Math 4s - Limited English Proficient - -- No. of students in 4th grade with this particular background who scored a 4 in math
81. Grade 4 Math 4s - Economically Disadvantaged - -- No. of students in 4th grade with this particular background who scored a 4 in math
82. Grade 5 ELA - All Students Tested - -- No. of students tested for ELA in 5th grade
83. Grade 5 ELA 4s - All Students - -- No. of students in 5th grade who scored a 4 in ELA
84. Grade 5 ELA 4s - American Indian or Alaska Native - -- No. of students in 5th grade with this particular background who scored a 4 in ELA
85. Grade 5 ELA 4s - Black or African American - -- No. of students in 5th grade with this particular background who scored a 4 in ELA
86. Grade 5 ELA 4s - Hispanic or Latino - -- No. of students in 5th grade with this particular background who scored a 4 in ELA
87. Grade 5 ELA 4s - Asian or Pacific Islander - -- No. of students in 5th grade with this particular background who scored a 4 in ELA
88. Grade 5 ELA 4s - White - -- No. of students in 5th grade with this particular background who scored a 4 in ELA
89. Grade 5 ELA 4s - Multiracial - -- No. of students in 5th grade with this particular background who scored a 4 in ELA
90. Grade 5 ELA 4s - Limited English Proficient - -- No. of students in 5th grade with this particular background who scored a 4 in ELA
91. Grade 5 ELA 4s - Economically Disadvantaged - -- No. of students in 5th grade with this particular background who scored a 4 in ELA
92. Grade 5 Math - All Students Tested - -- No. of students in 5th grade tested for Math
93. Grade 5 Math 4s - All Students - -- No. of students in 5th grade who scored a 4 in math
94. Grade 5 Math 4s - American Indian or Alaska Native - -- No. of students in 5th grade with this particular background who scored a 4 in math
95. Grade 5 Math 4s - Black or African American - -- No. of students in 5th grade with this particular background who scored a 4 in math

96. Grade 5 Math 4s - Hispanic or Latino - -- No. of students in 5th grade with this particular background who scored a 4 in math
97. Grade 5 Math 4s - Asian or Pacific Islander - -- No. of students in 5th grade with this particular background who scored a 4 in math
98. Grade 5 Math 4s - White - -- No. of students in 5th grade with this particular background who scored a 4 in math
99. Grade 5 Math 4s - Multiracial - -- No. of students in 5th grade with this particular background who scored a 4 in math
100. Grade 5 Math 4s - Limited English Proficient - -- No. of students in 5th grade with this particular background who scored a 4 in math
101. Grade 5 Math 4s - Economically Disadvantaged - -- No. of students in 5th grade with this particular background who scored a 4 in math
102. Grade 6 ELA - All Students Tested - -- No. of students tested for ELA in 6th grade
103. Grade 6 ELA 4s - All Students - -- No. of students in 6th grade who scored a 4 in ELA
104. Grade 6 ELA 4s - American Indian or Alaska Native - -- No. of students in 6th grade with this particular background who scored a 4 in ELA
105. Grade 6 ELA 4s - Black or African American - -- No. of students in 6th grade with this particular background who scored a 4 in ELA
106. Grade 6 ELA 4s - Hispanic or Latino - -- No. of students in 6th grade with this particular background who scored a 4 in ELA
107. Grade 6 ELA 4s - Asian or Pacific Islander - -- No. of students in 6th grade with this particular background who scored a 4 in ELA
108. Grade 6 ELA 4s - White - -- No. of students in 6th grade with this particular background who scored a 4 in ELA
109. Grade 6 ELA 4s - Multiracial - -- No. of students in 6th grade with this particular background who scored a 4 in ELA
110. Grade 6 ELA 4s - Limited English Proficient - -- No. of students in 6th grade with this particular background who scored a 4 in ELA
111. Grade 6 ELA 4s - Economically Disadvantaged - -- No. of students in 6th grade with this particular background who scored a 4 in ELA
112. Grade 6 Math - All Students Tested - -- No. of students in 6th grade tested for Math
113. Grade 6 Math 4s - All Students - -- No. of students in 6th grade who scored a 4 in math

114. Grade 6 Math 4s - American Indian or Alaska Native - -- No. of students in 6th grade with this particular background who scored a 4 in math
115. Grade 6 Math 4s - Black or African American - -- No. of students in 6th grade with this particular background who scored a 4 in math
116. Grade 6 Math 4s - Hispanic or Latino - -- No. of students in 6th grade with this particular background who scored a 4 in math
117. Grade 6 Math 4s - Asian or Pacific Islander - -- No. of students in 6th grade with this particular background who scored a 4 in math
118. Grade 6 Math 4s - White - -- No. of students in 6th grade with this particular background who scored a 4 in math
119. Grade 6 Math 4s - Multiracial - -- No. of students in 6th grade with this particular background who scored a 4 in math
120. Grade 6 Math 4s - Limited English Proficient - -- No. of students in 6th grade with this particular background who scored a 4 in math
121. Grade 6 Math 4s - Economically Disadvantaged - -- No. of students in 6th grade with this particular background who scored a 4 in math
122. Grade 7 ELA - All Students Tested - -- No. of students tested for ELA in 7th grade
123. Grade 7 ELA 4s - All Students - -- No. of students in 7th grade who scored a 4 in ELA
124. Grade 7 ELA 4s - American Indian or Alaska Native - -- No. of students in 7th grade with this particular background who scored a 4 in ELA
125. Grade 7 ELA 4s - Black or African American - -- No. of students in 7th grade with this particular background who scored a 4 in ELA
126. Grade 7 ELA 4s - Hispanic or Latino - -- No. of students in 7th grade with this particular background who scored a 4 in ELA
127. Grade 7 ELA 4s - Asian or Pacific Islander - -- No. of students in 7th grade with this particular background who scored a 4 in ELA
128. Grade 7 ELA 4s - White - -- No. of students in 7th grade with this particular background who scored a 4 in ELA
129. Grade 7 ELA 4s - Multiracial - -- No. of students in 7th grade with this particular background who scored a 4 in ELA
130. Grade 7 ELA 4s - Limited English Proficient - -- No. of students in 7th grade with this particular background who scored a 4 in ELA
131. Grade 7 ELA 4s - Economically Disadvantaged - -- No. of students in 7th grade with this particular background who scored a 4 in ELA

132. Grade 7 Math - All Students Tested - -- No. of students in 7th grade tested for Math
133. Grade 7 Math 4s - All Students - -- No. of students in 7th grade who scored a 4 in math
134. Grade 7 Math 4s - American Indian or Alaska Native - -- No. of students in 7th grade with this particular background who scored a 4 in math
135. Grade 7 Math 4s - Black or African American - -- No. of students in 7th grade with this particular background who scored a 4 in math
136. Grade 7 Math 4s - Hispanic or Latino - -- No. of students in 7th grade with this particular background who scored a 4 in math
137. Grade 7 Math 4s - Asian or Pacific Islander - -- No. of students in 7th grade with this particular background who scored a 4 in math
138. Grade 7 Math 4s - White - -- No. of students in 7th grade with this particular background who scored a 4 in math
139. Grade 7 Math 4s - Multiracial - -- No. of students in 7th grade with this particular background who scored a 4 in math
140. Grade 7 Math 4s - Limited English Proficient - -- No. of students in 7th grade with this particular background who scored a 4 in math
141. Grade 7 Math 4s - Economically Disadvantaged - -- No. of students in 7th grade with this particular background who scored a 4 in math
142. Grade 8 ELA - All Students Tested - -- No. of students tested for ELA in 8th grade
143. Grade 8 ELA 4s - All Students - -- No. of students in 8th grade who scored a 4 in ELA
144. Grade 8 ELA 4s - American Indian or Alaska Native - -- No. of students in 8th grade with this particular background who scored a 4 in ELA
145. Grade 8 ELA 4s - Black or African American - -- No. of students in 8th grade with this particular background who scored a 4 in ELA
146. Grade 8 ELA 4s - Hispanic or Latino - -- No. of students in 8th grade with this particular background who scored a 4 in ELA
147. Grade 8 ELA 4s - Asian or Pacific Islander - -- No. of students in 8th grade with this particular background who scored a 4 in ELA
148. Grade 8 ELA 4s - White - -- No. of students in 8th grade with this particular background who scored a 4 in ELA
149. Grade 8 ELA 4s - Multiracial - -- No. of students in 8th grade with this particular background who scored a 4 in ELA

150. Grade 8 ELA 4s - Limited English Proficient - -- No. of students in 8th grade with this particular background who scored a 4 in ELA
151. Grade 8 ELA 4s - Economically Disadvantaged - -- No. of students in 8th grade with this particular background who scored a 4 in ELA
152. Grade 8 Math - All Students Tested - -- No. of students in 8th grade tested for Math
153. Grade 8 Math 4s - All Students - -- No. of students in 8th grade who scored a 4 in math
154. Grade 8 Math 4s - American Indian or Alaska Native - -- No. of students in 8th grade with this particular background who scored a 4 in math
155. Grade 8 Math 4s - Black or African American - -- No. of students in 8th grade with this particular background who scored a 4 in math
156. Grade 8 Math 4s - Hispanic or Latino - -- No. of students in 8th grade with this particular background who scored a 4 in math
157. Grade 8 Math 4s - Asian or Pacific Islander - -- No. of students in 8th grade with this particular background who scored a 4 in math
158. Grade 8 Math 4s - White - -- No. of students in 8th grade with this particular background who scored a 4 in math
159. Grade 8 Math 4s - Multiracial - -- No. of students in 8th grade with this particular background who scored a 4 in math
160. Grade 8 Math 4s - Limited English Proficient - -- No. of students in 8th grade with this particular background who scored a 4 in math
161. Grade 8 Math 4s - Economically Disadvantaged - -- No. of students in 8th grade with this particular background who scored a 4 in math

ABSTRACT

Brief introduction about the ASSIGNMENT Purpose, Data Set, Kind of analysis performed, The kind of questions you could ask/ could answer using your data, conclusion reached. (200 words max - must be only a paragraph)

The assignment aims to derive insights from the dataset ,answer the questions that might arise and finally asking the right questions.

The possible tracks for insights include:
1.Relationship between geographical location ,economic need index, and the distribution of the later through out the country.

2.Relationship between Race(White, Hispanic, Asian, Black, Multiracial) and geographical location.

3.Distribution of diversity according to race in the various schools

4.Realtionship between economic need index and racial distribution

5. Which demographic makes up schools with higher highest grades and lower lowest grades?

6. Which demographic makes up schools with higher attendance rate?

And many more

EXPLORATORY ANALYSIS

Data Cleaning:
1.The dataset we received was cleaned as specified in the guidelines with all the NAN's for categorical columns to be replaced with its previous row values and all the NAN's for numeric columns to be replaced with average of the column .This was carried out using the **fillna** function defined in the **numpy** library along with the attribute "**ffil**" forward fill for categorical columns to be replaced with its previous row values

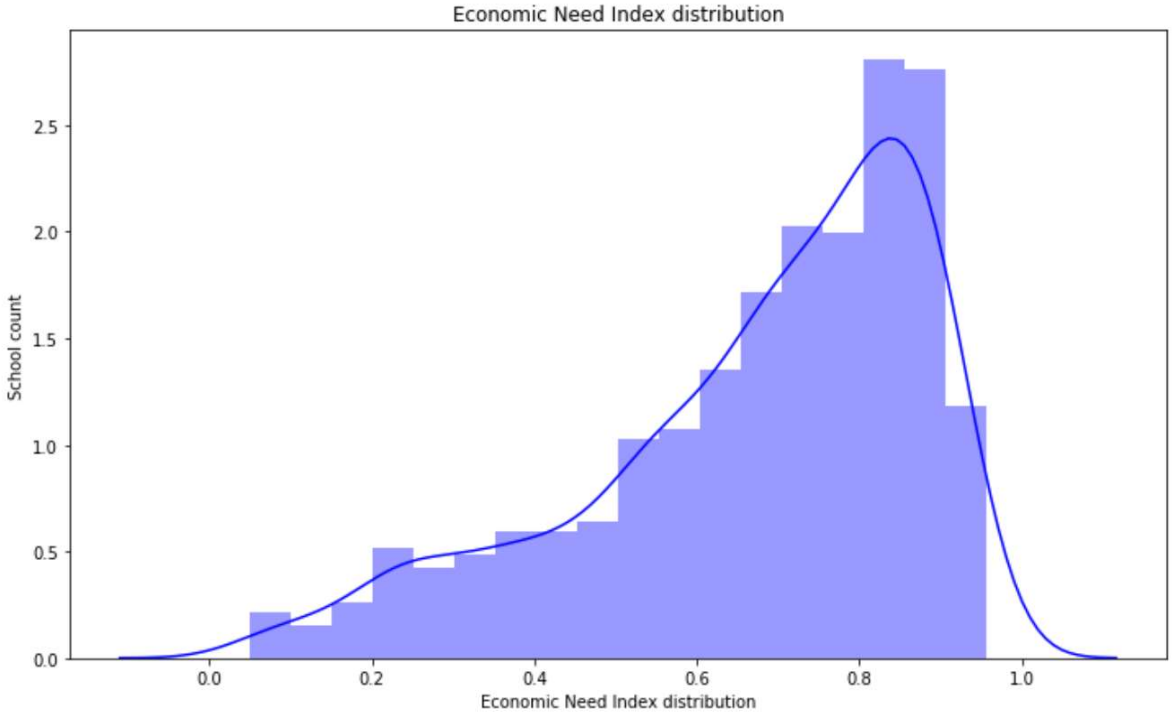
2.Then it was required to convert the columns with **percentage value**s to float .
The **column.astype(str).apply(p2f)** method was used for this purpose.

3.The columns regarding income had to be stripped off their unwanted characters('$' ,'\' etc).
The **column .str.replace('$', '')** method was used for this purpose.

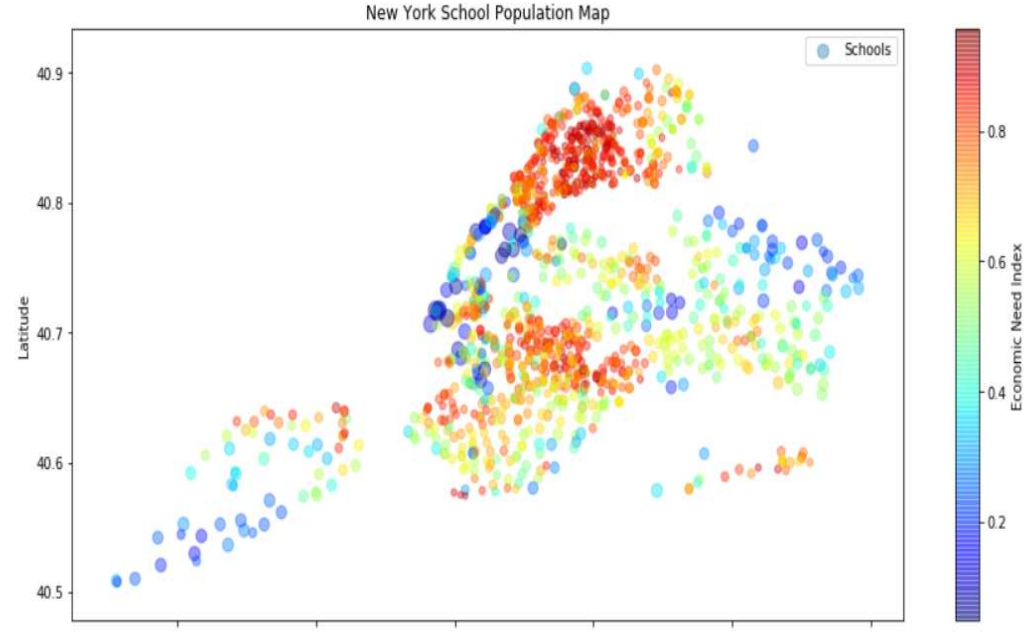
4. The columns Rigorous Instruction Rating, Collaborative Teachers Rating. Supportive Environment Rating, Effective School Leadership Rating, Strong Family-Community Ties Rating, Trust Rating, Student Achievement Rating' have been dropped due to large size of columns(161)
The function used here is **dataframe.drop** from the **pandas package**

Descriptive Analysis:

1.**The left skewed nature of distribution of economic need index distribution is observable in the following plot.**



2. The Economic need index of all the schools mentioned ,has been plotted with latitude and Longitude as x and y axis so it is a close representation to a map of USA with all the states and Color intensities specifying the economic need.
The following scatter plot explains it:



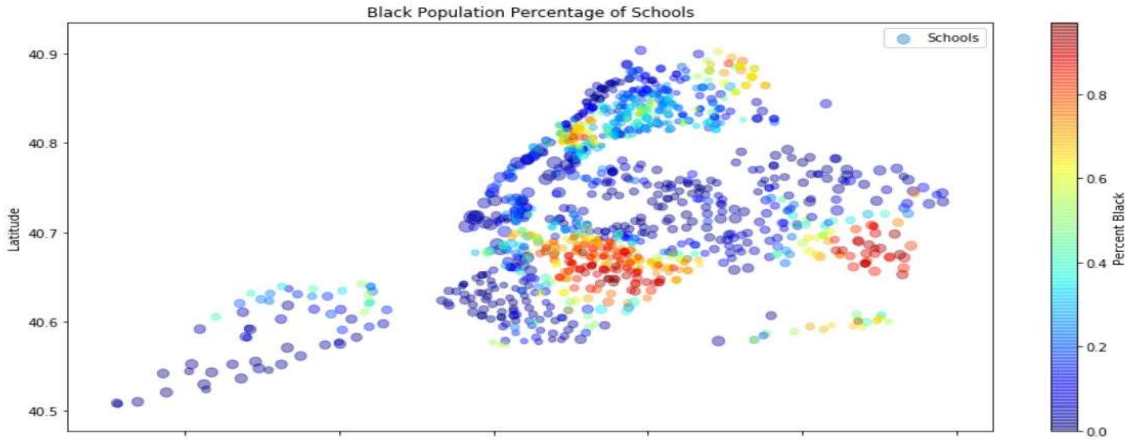
We can see that the school when ploted according to the location from latitude and longitude , the schools with higher economic needs are found in the regions of Central New York according to the School Income Estimate.

The inference drawn here is that schools when ploted according to the location from latitude and longitude , the schools with higher economic needs are found in the regions of Central New York according to the School Income Estimate.

3. Distribution of Races across New York :

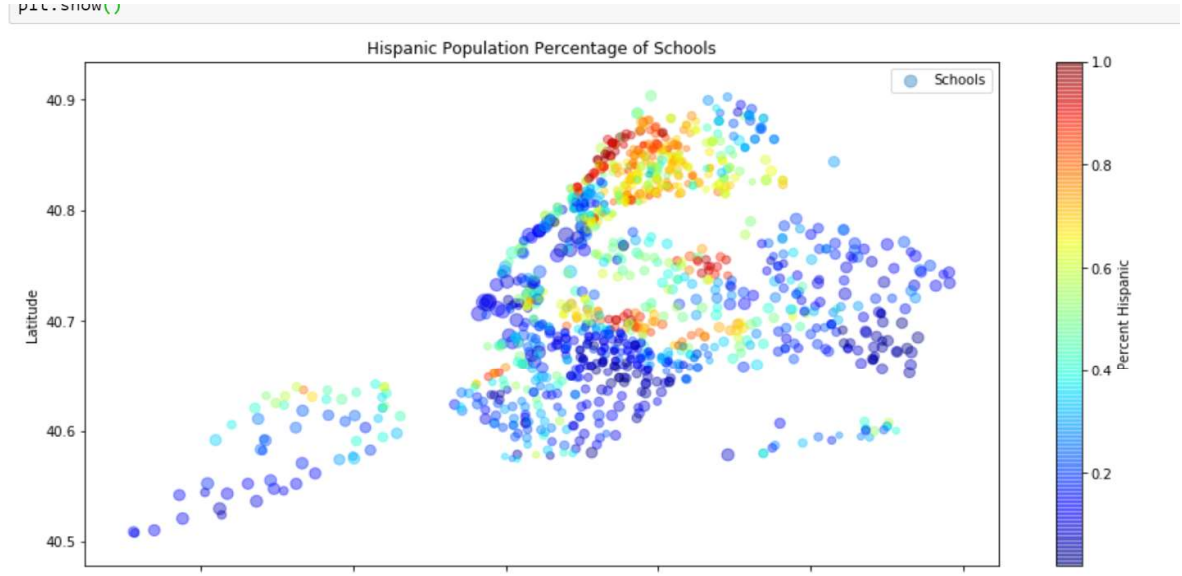
Central New York contains most of the Schools with Blacks: 1) As we can see here that the Majority of the black population are living in Central New York . 2)If you compare this to the Hispanic Population,

they are totally isolated from each other.

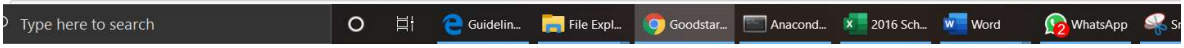


Central New York contains most of the Schools with Blacks: 1) As we can see here that the Majority of the black population are living in Central New York . 2) you compare this to the Hispanic Population, they are totally isolated from each other.(Red)

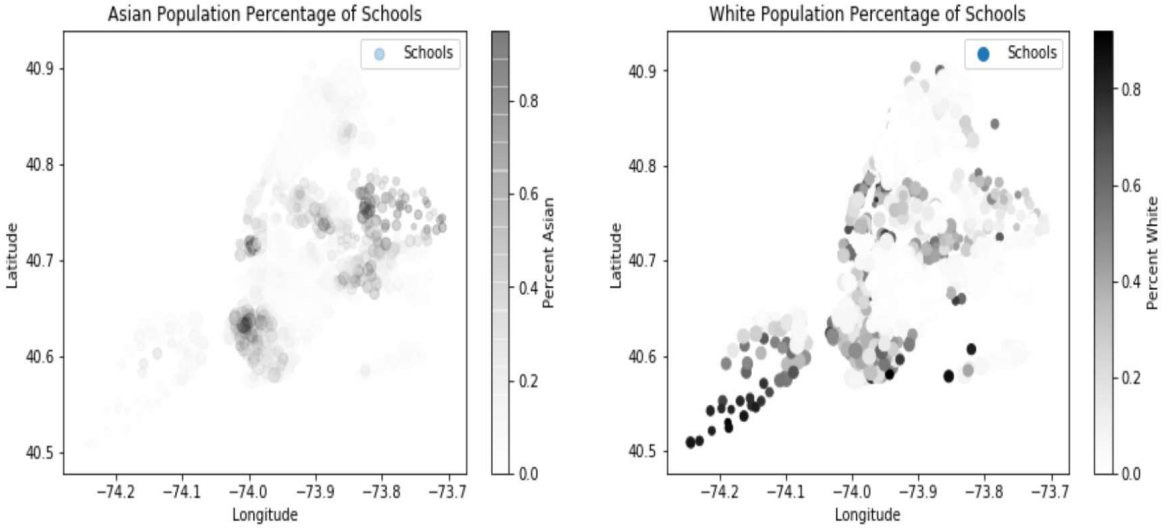
4.



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 t of Blacks in Central New York.



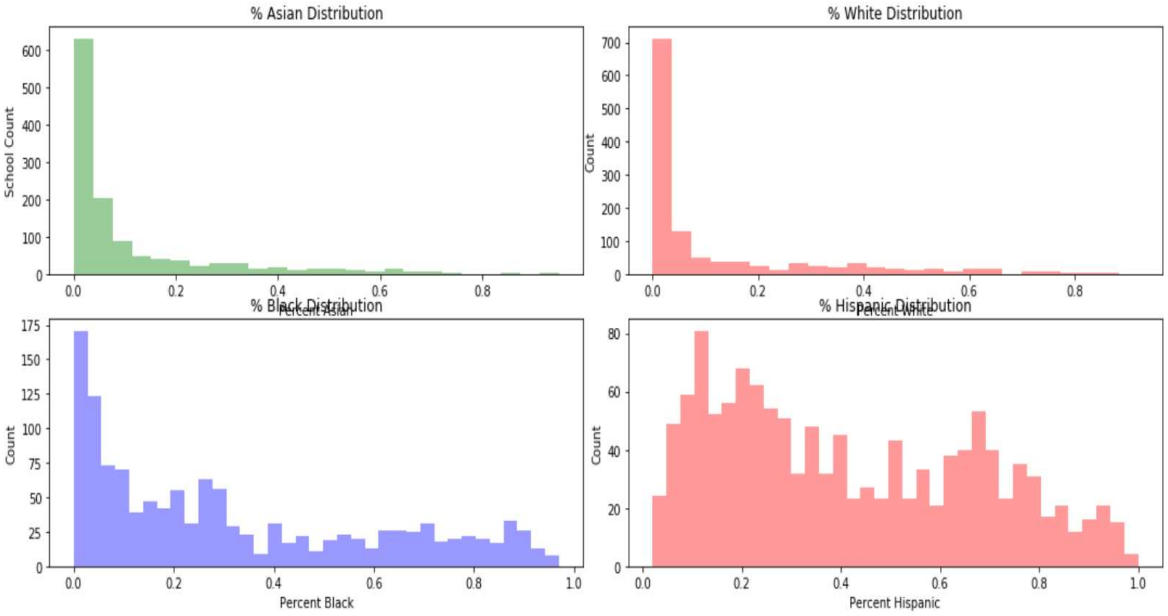
Majority of the hispanic population are living in Upper New York. Here we notice that there are a few Hispanics in Central New York while there are a majority of Blacks in Central New York.



The Asian population are mostly living in Central New York.Noticeable from the black color.

The White Population has a very noticeable cluster near the edge and lower parts of New York.

5. The Asian population are mostly living in Central New York .Noticeable from the black color. The White Population has a very noticeable cluster near the edge and lower parts of New York.



6**. Now we see the percent distribution of the Different Races**.
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

7. **Hypothesis testing**

NULL Hypothesis : Schools with a Higher Hispanic and Black Population tend to have a lower Economic Need Index Score than schools with a Higher White and Asian
Alternate Hypothesis: Schools with a Higher Hispanic and Black Population tend to have a higher Economic Need Index Score than schools with a Higher White and Asian

in other words NULL Hypothesis:

mean_black_and_hispanic_Economic_Need_Index <= mean_white_asian_Economic_Need_Index
m2 - m1 <= 0

Alternate Hypothesis:

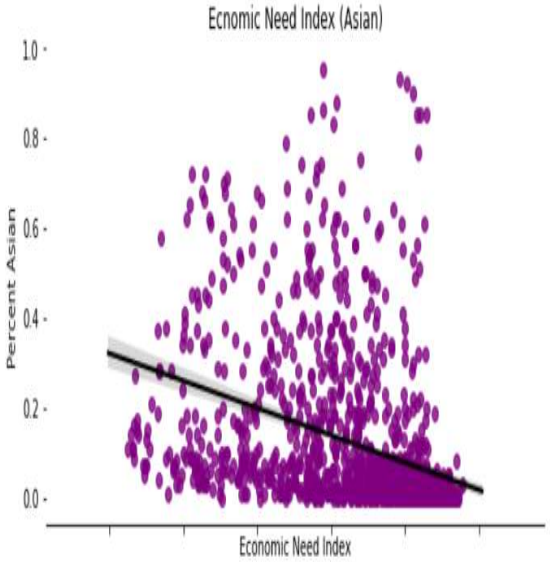
mean_black_and_hispanic_Economic_Need_Index > mean_white_asian_Economic_Need_Index

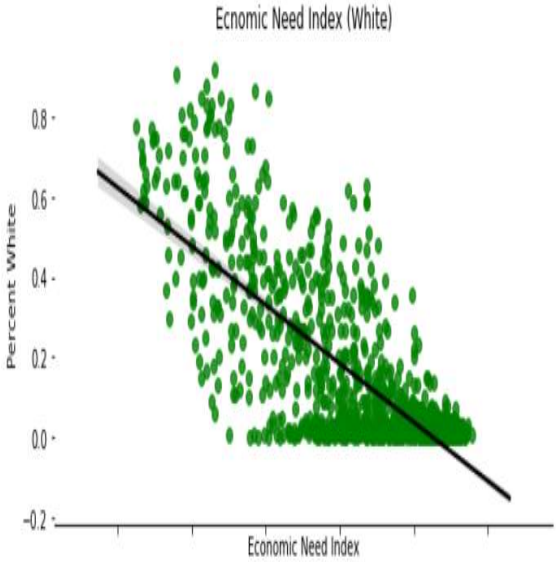
m2 -m1 > 0

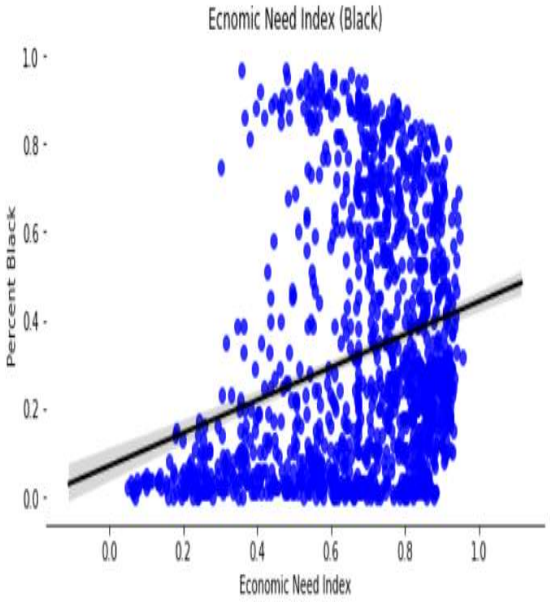
Now computing the Z score: z = (m2-m1)- 0/math.sqrt(s1^2/n1 + s2^2/n2)
= 16.09252112179623

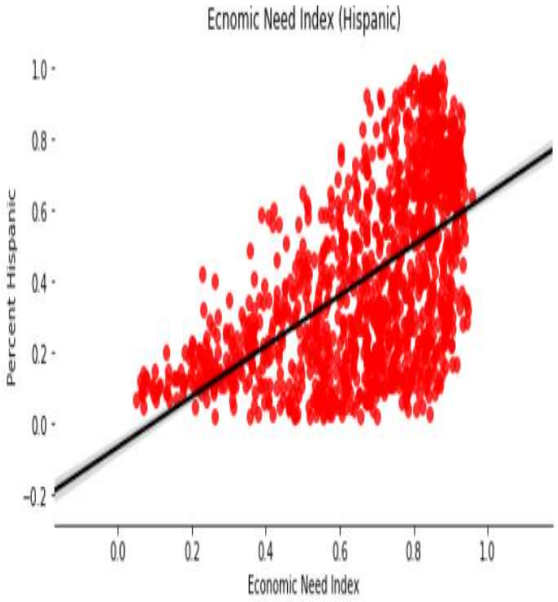
Since the value 16.0952112179623 ~= 0 P < 0.05 We can reject the NULL Hypothesis

We have graphical Evidence to justify the Hypothesis testing .sns.replot is used to visualize a linear relationship as determined through regression.The function draws a scatterplot of two variables, x and y, and then fit the regression model y ~ x and plot the resulting regression line and a 95% confidence interval for that regression:





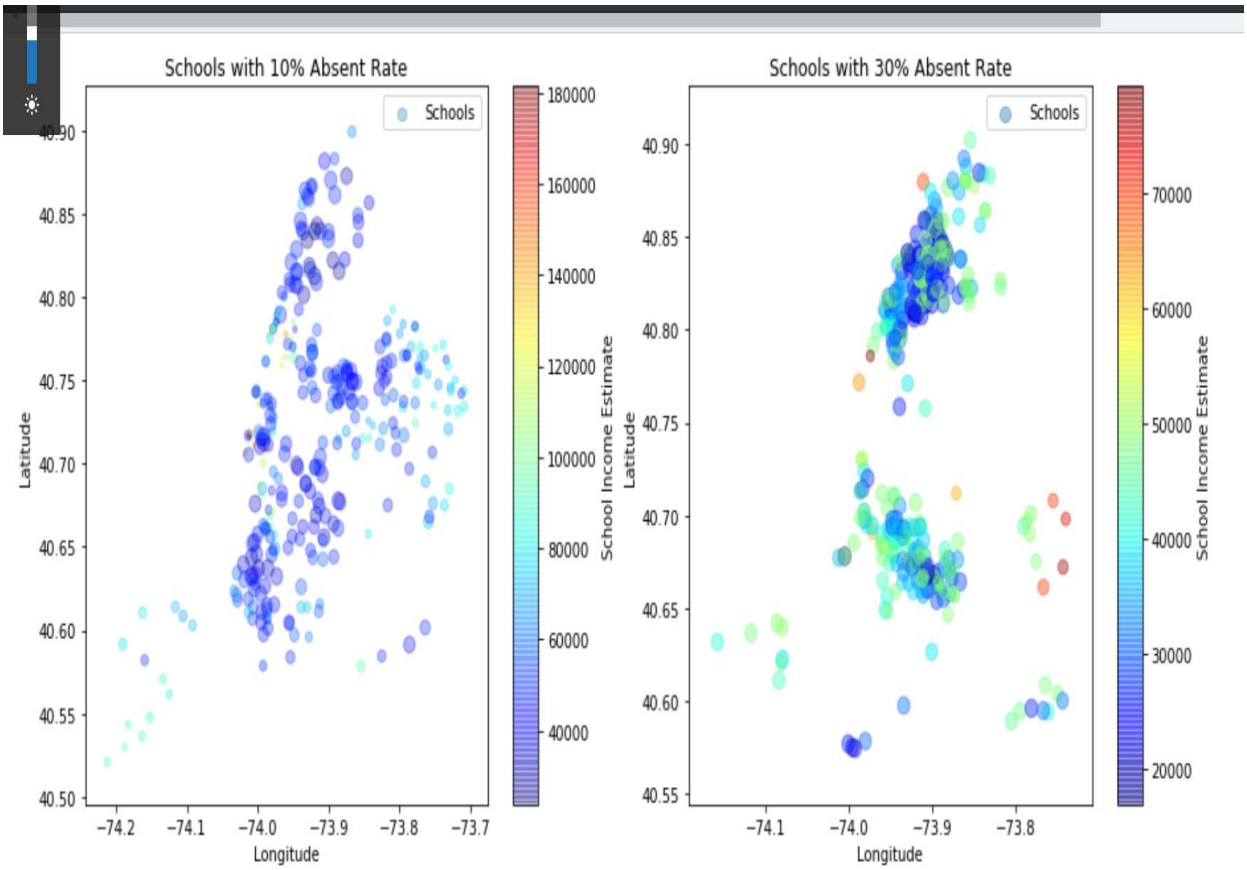




Schools with a Higher White and Asian Population tend to have a Lower Economic Need Index Score.
Schools with a Higher Hispanic and Black Population tend to have a Higher Economic Need Index Score.

Here we have used **Regression plot** to plot the Economic need index relative to the different races.
As we can see:
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.

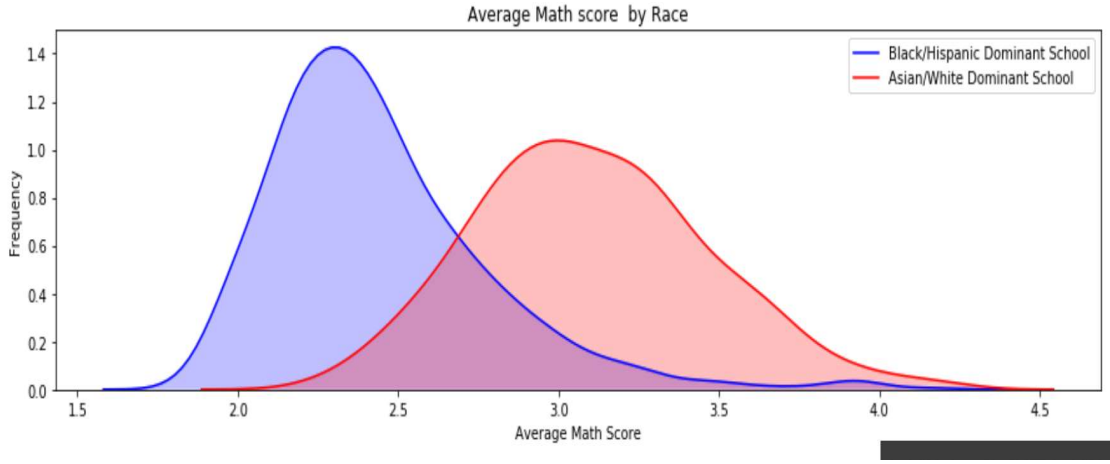
8.

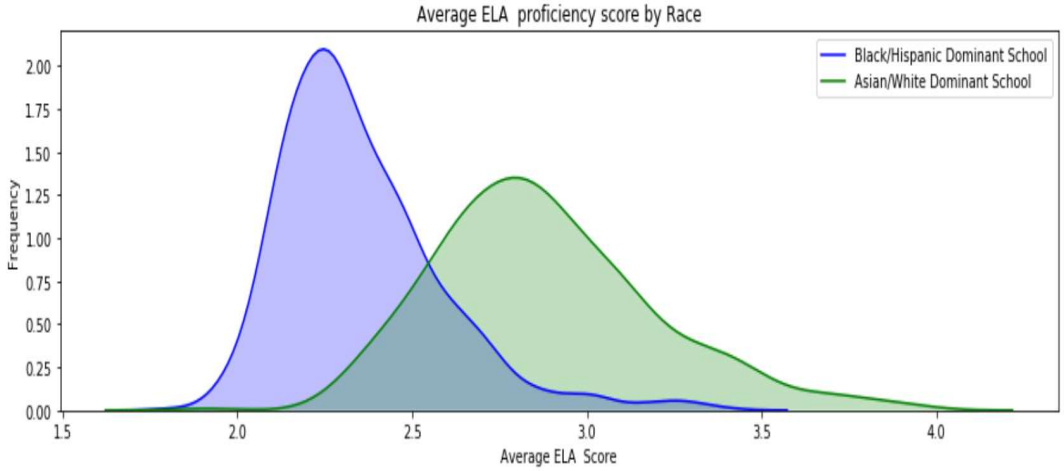


Absent Rates and Schools Income Estimate distribution Realtion:
The plot below shows the Schools Income Estimate distribution for 10% and 30% attendence rate.
As we can see the 30% attendance rate has School Income distributed approximately into two classes. The intenstites of color are lower than 10% although there seems to be concentration

However the 10% attendance has Schools Income Estimate distribution scattered through out. Schools Income Estimate distribution here is higher than the 10% ,denoted by the color intensities.

9.





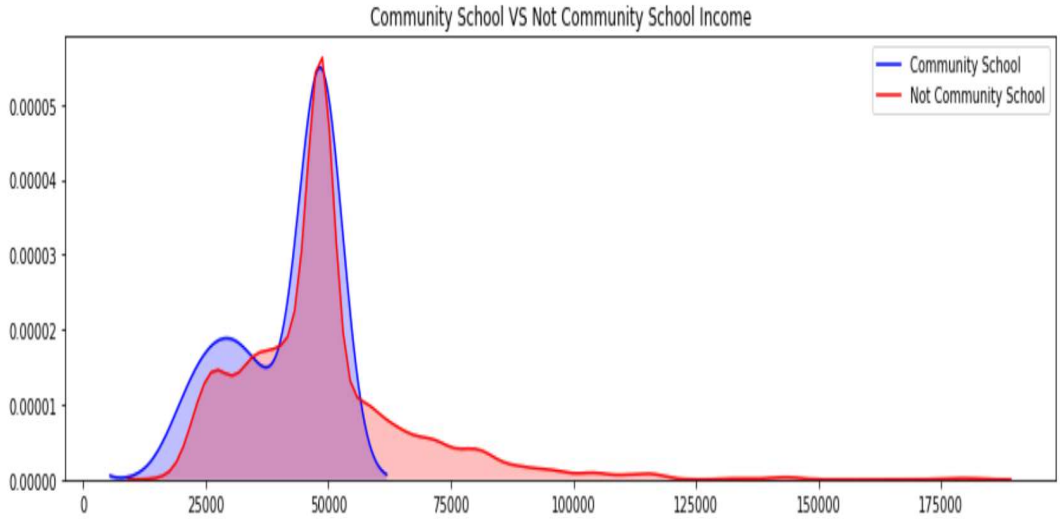
Relation between Race and scores:

```
For Black/Hispanic: Average ELA Proficiency     2.352136
Average Math Proficiency    2.439736

For White/Asian: Average ELA Proficiency     3.046919
Average Math Proficiency    3.326919
```

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

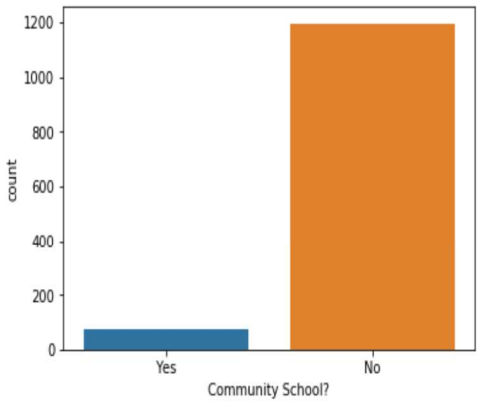
10.



Here we can see that a greater proportion of the community schools have a greater Income Estimate than non Community schools. Also here non community schools are more in number.

School Income comparison between community and non-community:
Here we can see that a greater proportion of the community schools have a greater Income Estimate than non Community schools.

11.

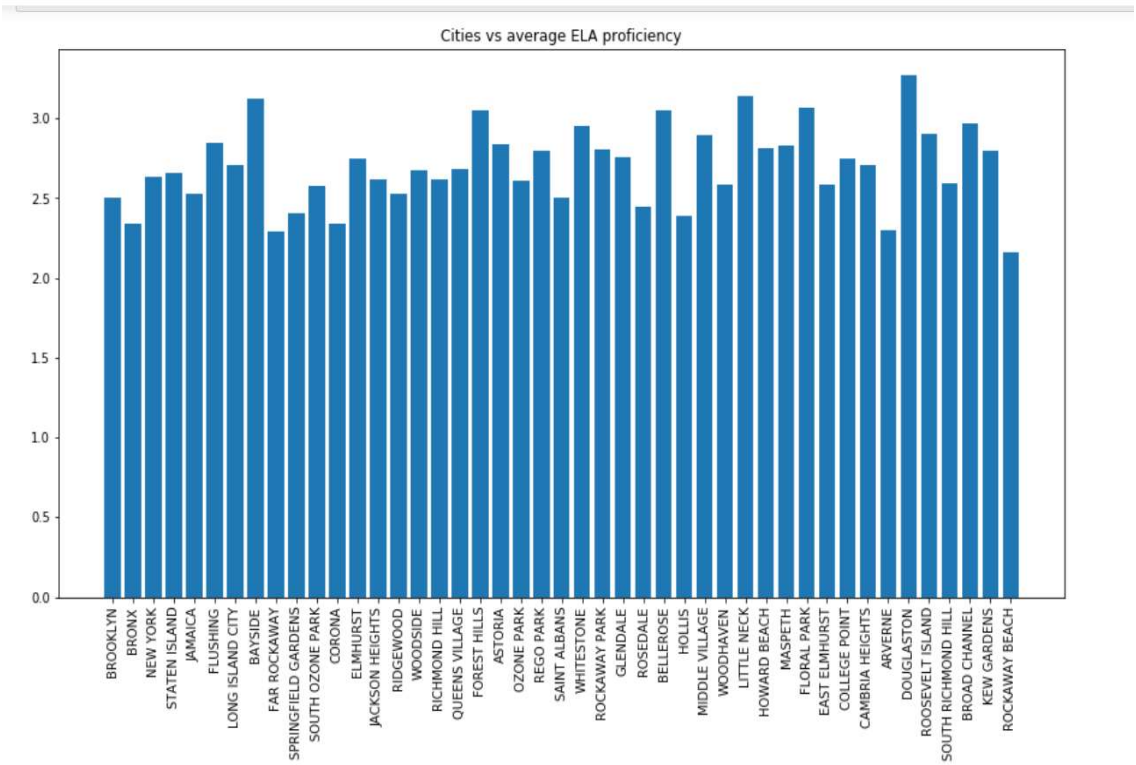


As we can see that there are more number of Non Community schoools.Also community schools have greater income estimate.

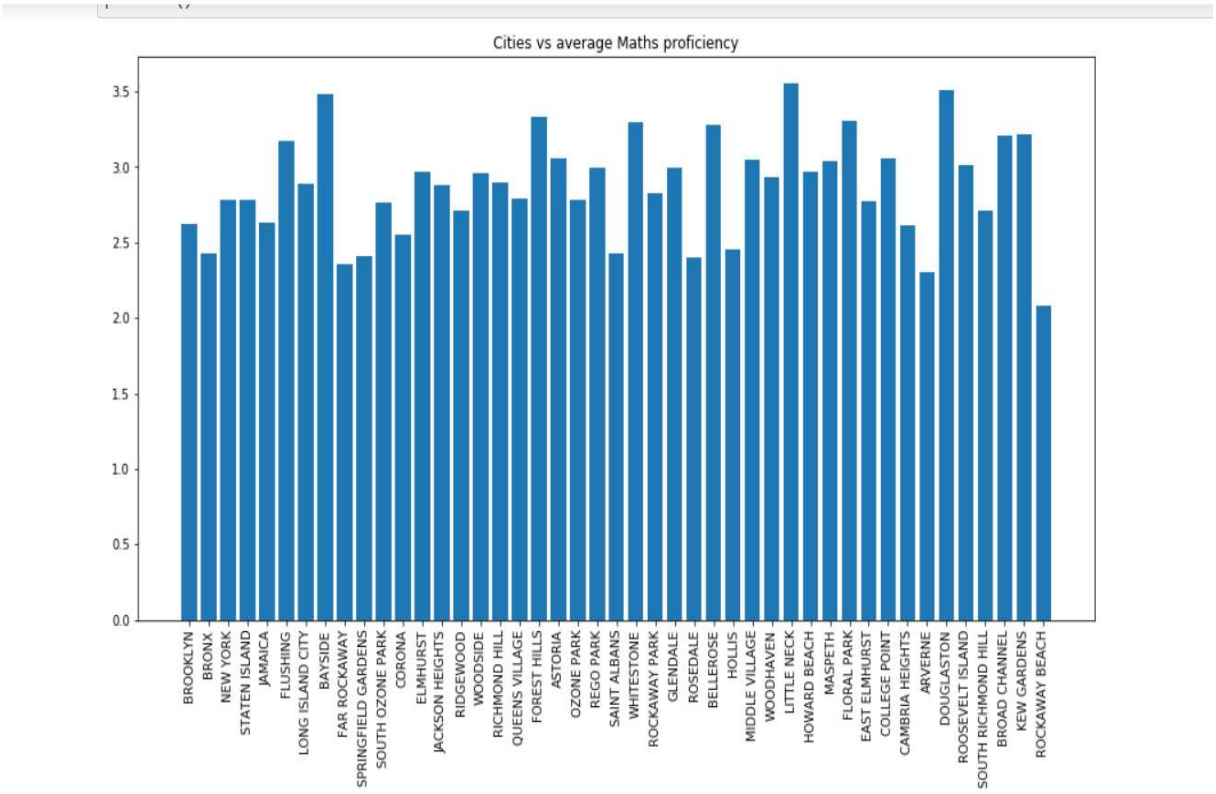
Categorization of schools as community and non-community:
As we can see that there are more number of Non Community schoools.Also community schools have greater income estimate.

12. **Now Comparing the scores for ELA and MAth proficiency among the various schools**

Cities vs Average ELA proficiency:

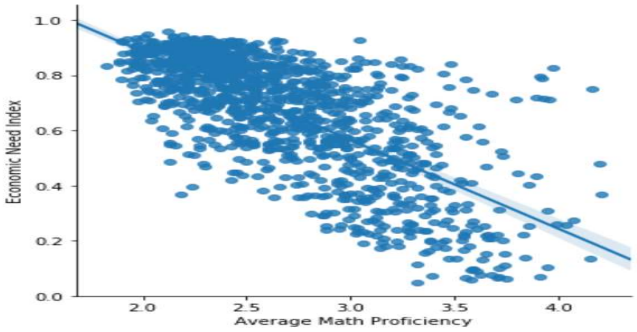


Cities vs Average Maths proficiency



13.**Now we see the relation between the grades and the Economis need Index** . Firstly we compare the Average Math Proficiency against economic need index

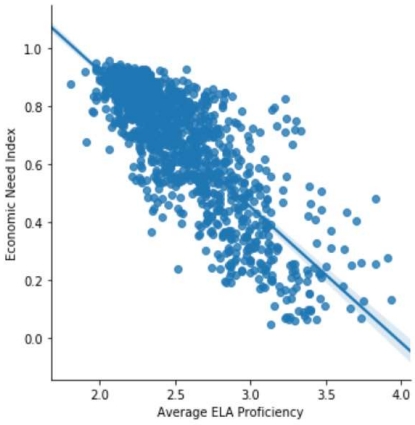
```
In [36]: sns.lmplot(x='Average Math Proficiency', y='Economic Need Index',
                    fit_reg=True)

Out[36]: <seaborn.axisgrid.FacetGrid at 0x21915e174a8>
```



Here we see that an inverse relation seems to exist between them

Here from the regression plot from the regression line we see that an inverse relation seems to exist between them.

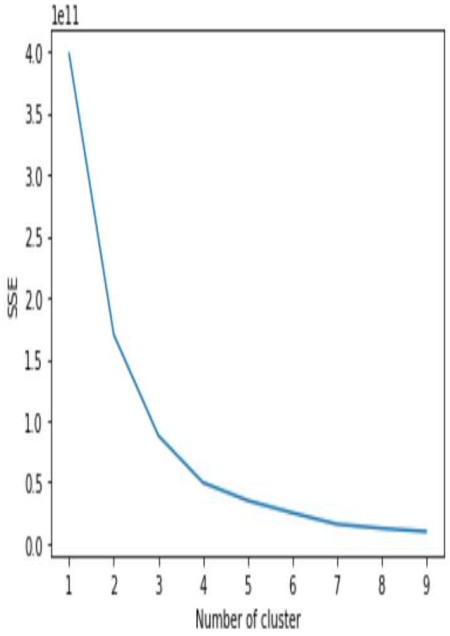
```
In [37]: sns.lmplot(x='Average ELA Proficiency', y='Economic Need Index', data=df,
                    fit_reg=True)

Out[37]: <seaborn.axisgrid.FacetGrid at 0x21916096f28>
```



Next we compare the Average ELA Proficiency against economic need index Here also we see that an inverse relation seems to exist between them

Here from the regression plot from the regression line we see that an inverse relation seems to exist between them.

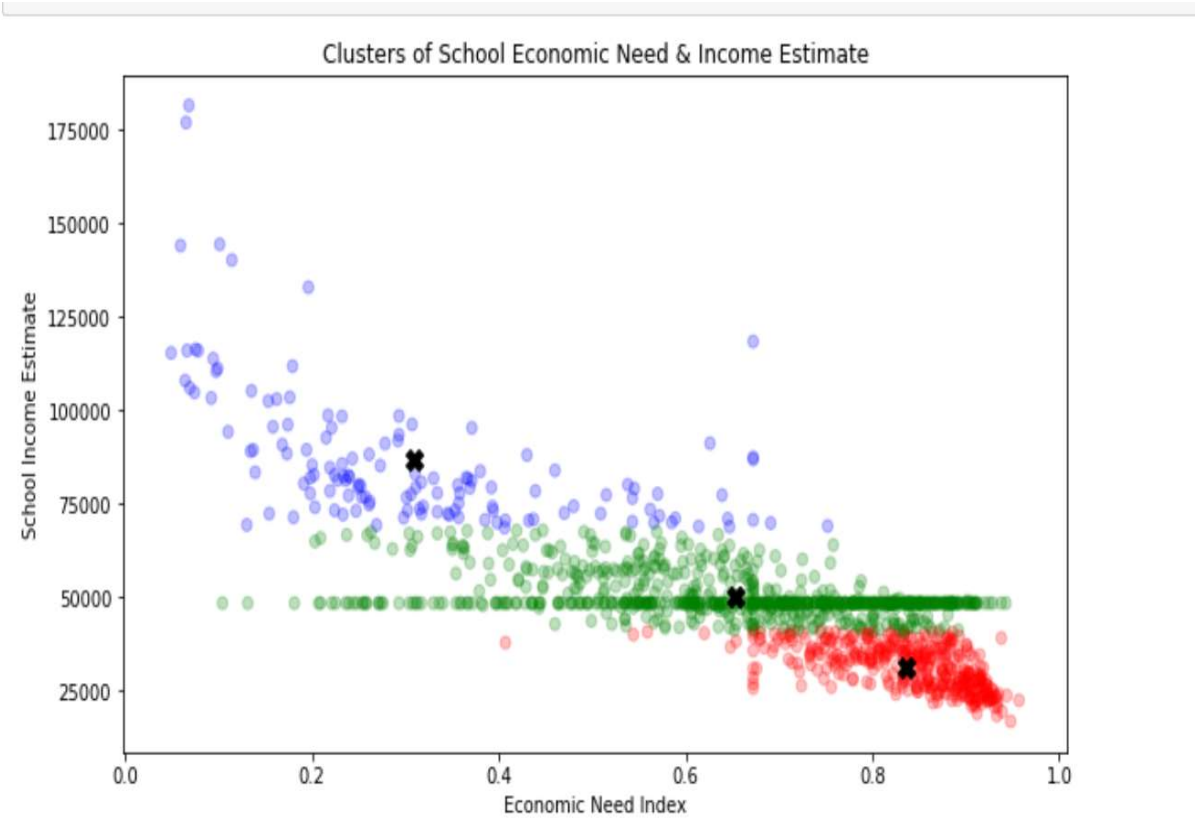
14. **K-Means Clustering:** School Economic Need and Income Estimate



This a plot of Sum of distances of samples to their closest cluster center(SSE)ERROR to the number of clusters considered. Here we consider the x co-oridinate where there is a change in slope of the error ,to be the ideal k (or number of classes) that should be considered for clustering .Here we can see that its 3(k=3)

Calculating k value:

This a plot of Sum of distances of samples to their closest cluster center(SSE)ERROR to the number of clusters considered. Here we consider the x co-oridinate where there is a change in slope of the error ,to be the ideal k (or number of classes) that should be considered for clustering .Here we can see that its 3(k=3).



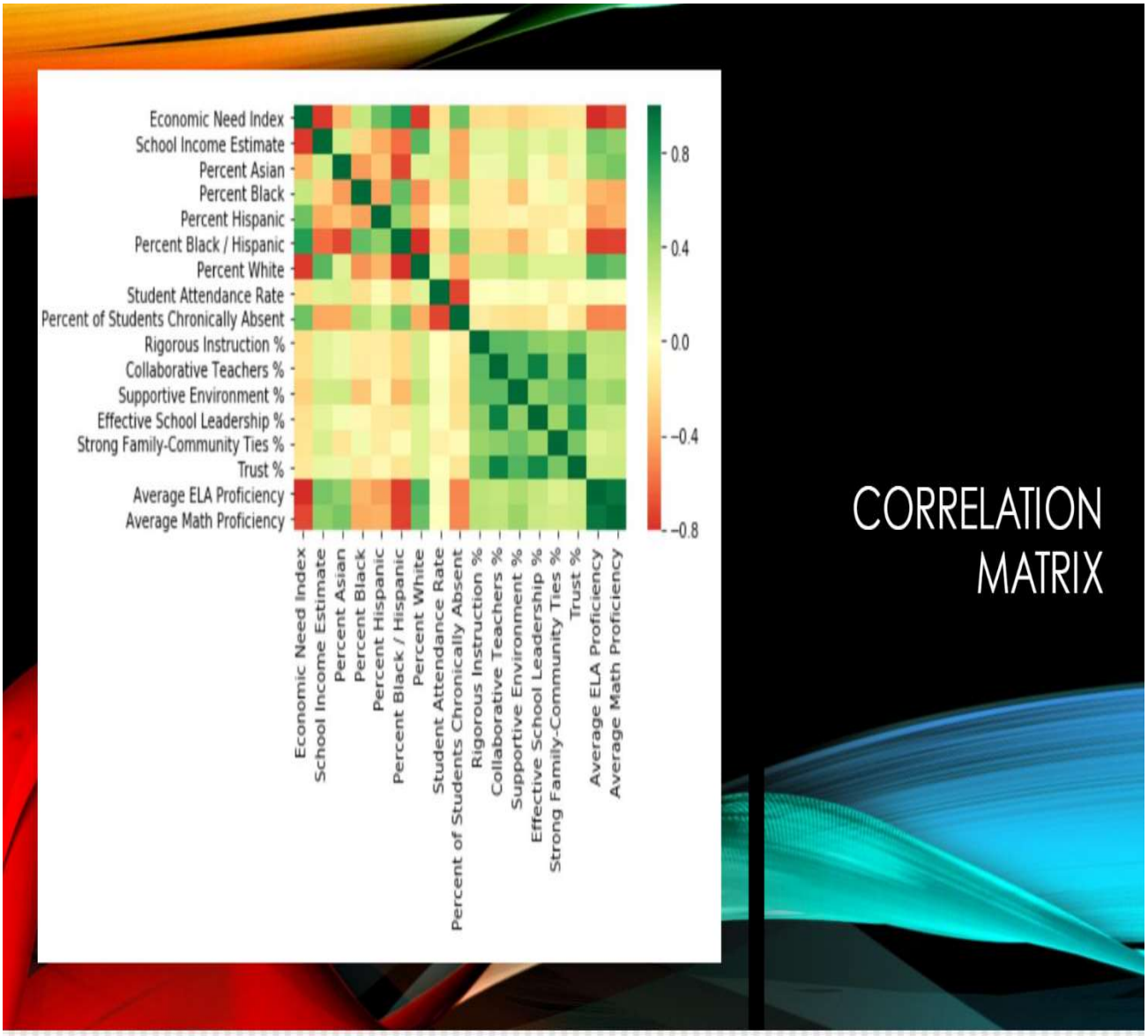
We can see three clusters here:

1:the red represents the schools 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)

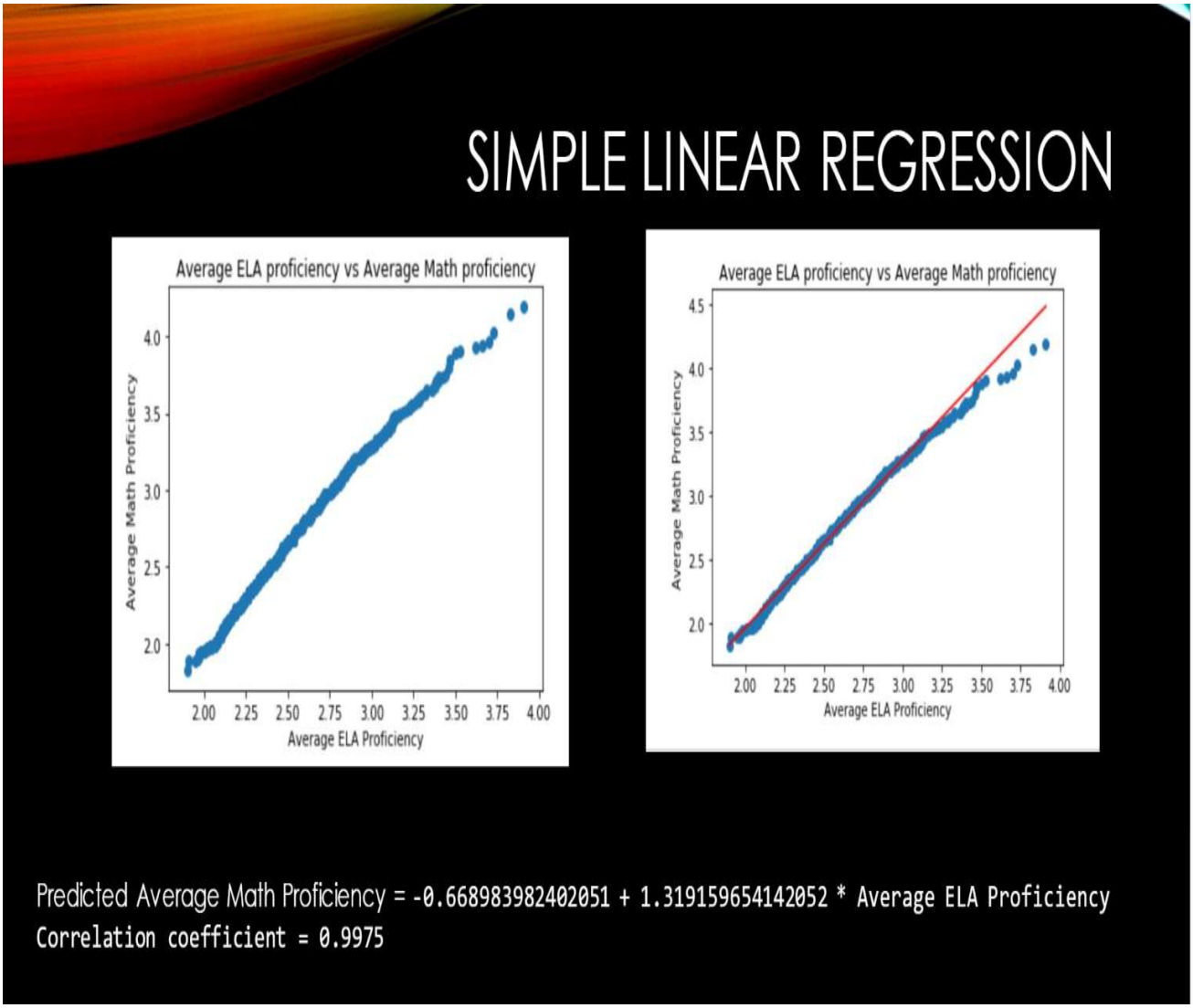
3:The blue represents the schools with low economic need ,but have surplus income.Hence are in no need of any inremented income

14. **A plot correlation matrix to show the relationship between various columns**.



The above heat map shows the relationship between various columns which aids in the selection of linear regression done later.

15. Lastly we have done a Linear Regression to predict Average Math Proficiency



Correlation coefficient : 0.9975
Predicted Average Math Proficiency: - 0.669 + 1.32 * Average ELA Proficiency

CONCLUSIONS:

The left skewed nature of distribution of economic need index distribution is observable in the following plot.

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

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 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

Schools with a Higher Hispanic and Black Population tend to have a higher Economic Need Index Score than schools with a Higher White and Asian

As we can see the 30% attendance rate has School Income distributed approximately into two classes. The intenstites of color are lower than 10% although there seems to be concentration

However the 10% attendance has Schools Income Estimate distribution scattered through out. Schools Income Estimate distribution here is higher than the 10% ,denoted by the color intensities.

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

Here we can see that a greater proportion of the community schools have a greater Income Estimate than non Community schools

From kmeans clustering we can see three clusters:
1:the red represents the schools 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 inremented income

There is very high correlation between Average ELA proficiency and Average Math Proficiency

Linear Regression to predict Average Math Proficiency yields: Predicted Average Math Proficiency: -0.669 + 1.32 * Average ELA Proficiency