Judging from the plot we just created, it doesn't appear that there's an extremely strong correlation between sat\_score and total\_enrollment. If there was a very strong correlation, we'd expect all of the points to line up. Instead, there's a large cluster of schools, and then a few others going off in three different directions.

However, there's an interesting cluster of points at the bottom left where total\_enrollment and sat\_score are both low. This cluster may be what's making the r value so high. It's worth extracting the names of the schools in this cluster so we can research them further.

Instructions

* Filter the combined dataframe to keep only those rows where total\_enrollment is under 1000 and sat\_score is under 1000. Assign the result to low\_enrollment.
* Display all of the items in the School Name column of low\_enrollment.
* Use [Wikipedia](https://www.wikipedia.org/) and [Google](https://www.google.com/) to research the names of the schools.
* Can you discover anything interesting about them?

Answer

# we brake the filter in steps first we filter the combined df and assign it to a variable

low\_enrollment = combined[combined["total\_enrollment"] < 1000]

# then we use this variable with low enrlolment to filter the low score also

low\_enrollment = low\_enrollment[low\_enrollment["sat\_score"] < 1000]

# then we use this dataframe to filter the school name

print (low\_enrollment["School Name"])

Our research on the last screen revealed that most of the high schools with low total enrollment and low SAT scores have high percentages of English language learners. This indicates that it's actually ell\_percent that correlates strongly with sat\_score, rather than total\_enrollment. To explore this relationship further, let's plot out ell\_percent vs sat\_score.