Peter Lovberg

PyCity Schools Analysis, Conclusions, and Attributions

The data provided from the csv files schools\_complete.csv, and students\_complete.csv provided data to create a profile of this particular high school district’s trends regarding students passing or not passing math and reading, and their overall trends for passing their grades based on passing math and reading during the same school year. We used data including type of school, amount of funding resources spent on each student, and school size to determine if there were relationships between these categories of data, and the propensity to receive an overall passing grade per school year. Here were the main conclusions I arrived at.

1. There was no trend describing passing either math, reading or overall, broken out by grade level. The results were inconclusive; sometimes the freshman (9th graders) had lower or higher reading, math, or overall passing rates than the seniors (12th graders) in the same school. Sometimes the reverse was the case; as one would expect, the seniors had higher passing rates in math, reading or overall than freshmen in the same schools.
2. The bottom performing schools were all of the District type
3. The top performing schools were all of the Charter type
4. There was an INVERSE relationship between passing rates, and the budget$ spent per student. So in summary, throwing more money at a school will not necessarily result in higher passing rates, as many teachers unions would suggest. (this also assumes that Charter Schools within the district receive the same per student budget allocations as those in the District Schools)
5. There was an INVERSE relationship between school Size (by enrollment) and passing rates. When Large schools were compared against Small and Medium schools. The results were significant, indicating a 58.2% passing rate overall across large schools, vs. an 89.88 and 90.62 passing rate overall in Small, and Medium sized schools, respectively.

During the course of the Challenge exercise, I utilized course materials from Module 4 Activities and Solutions files to reference coding conventions within Pandas. I also utilized W3Schools Pandas Tutorial found at <https://www.w3schools.com/python/pandas/default.asp> to workout some issues.

I utilized the help of a tutor-Hassan out of Toronto, Canada, to help me preview some binning issues that I might encounter while completing the Challenge Activity.

Finally, I used Chat GPT as a last resort to assist in debugging data frames 43, 47, and 50, where the starter code provided did not reference columns as indicated in the outputs provided. This occurred in a column referenced in the starter code titled [“ % Overall Passing”], in all three of those data frames, which did not appear in the output provided. After of hours of attempting to debug by myself, I utilized ChatGpt to help debug this. Incidentally, Chat GPT was not great at debugging, I actually stumbled upon the discrepancy myself, while Chat sent me down rabbit hole after rabbit hole.