Schools Join Methodology

I joined the LCGMS schools dataset (updated July 15 each year) with the School Locations dataset (updated November 26, 2024) using the ATS/ATS System Code as the common field.

- Dataset sizes and accuracy:
    - The LCGMS dataset contains 1,900 rows and provides the most up-to-date and accurate representation of schools. This figure closely matches the DOE's published total for the 2023–2024 school year: 1,870 schools (1,596 public and 274 charter).
    - The School Locations dataset contains 1,950 records.
- Join results:
    - The join successfully matched all but 68 records (or 16 from the LCGMS table, accounting for discrepancies between the two datasets).
    - Most unmatched entries are not traditional schools: they include home schools, administrative offices, Alternative Learning Centers (ALCs), or charter schools.
- Notable errors:
  Several public schools failed to match due to what I assume are errors in the ATS code, including, P.S. 076 A. Philip Randolph, M.S. 250 West Side Collaborative Middle School, The Children's School, and P.S. 025 Eubie Blake School. I retained all unmatched schools in the dataset and also exported a GeoJSON file of unmatched records.
- Duplicate locations:
  I also exported a file called "SchoolsPoints_Count", which includes an ICOUNT field showing how many schools share the same lat/long. This file contains 1,381 unique points, meaning that over ~500 schools share a building or lot with another school.
    - In most cases, these are multiple schools within the same building, confirmed by having the same street address. Because the analysis is concerned with school buildings rather than individual schools, I don't think it's necessary to adjust the overlapping points.
    - In some cases, schools have different street addresses but were geolocated to the same point because they fall within the same parcel and therefore share the same BBL (Borough Block Lot). These points could be adjusted in the future to match their physical locations.
- A next step could be to match school BBLs with the building/energy data.