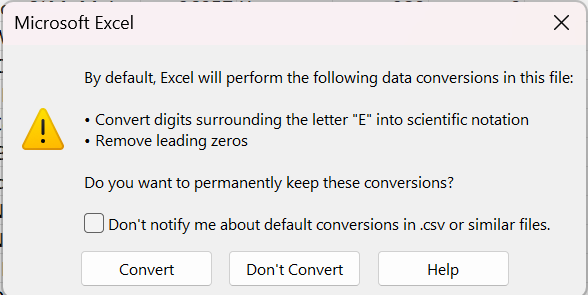
“Local Education Agencies

The number of public school systems shown in this volume generally corresponds to counts of ‘‘local education agencies’’ in reports by state offices of education and presented in published reports by the National Center for Education Statistics, with the exception of certain administrative and other education activities that do not meet Census Bureau classification criteria for school district governments or dependent public school systems.”

**Question One: How many school districts were there in total in 2017-18 using Excel?**

Count of LEA IDs including nulls for which there is data in other columns (COUNT when converting data): **17152**

****

Count of LEA IDs not including nulls in this column (COUNT when not converting data): **14259**

Count of rows with data in them (COUNTA): **17604**

452 rows have LEA IDs that also have numbers in them (CC for California, SOP for juvenile justice centers in multiple states)

**Question Two: How many school districts were there in 2017-18 using Python?**

Using Python, there were 17,616 school districts in 2017-2018.

**Question Three: Why would you need encoding=“cp1252” for US data?**

We would need this encoding for U.S. data because data from the United States is reported in English, or using abbreviations/acronyms with the Latin alphabet. This encoding allows the system to read data and convert the single-byte characters to readable text.

**Question 4: This number does not match the CRDB database, but both databases claim to have the list of school districts in the US in the period 2017-18. Explain why.**

SGLA reports 18,843 LEAs when we count in Python.

CRDC:

* “The initial target population for the 2017-18 data collection included 17,744 LEAs and 99,089 schools. Discovered during data collection, some LEAs merged, closed, or opened. Thus, they were added or removed from the initial target population. These adjustments resulted in a revised final target population of 17,637 LEAs and 97,632 schools. Puerto Rico was also added to the collection for the 2017-18 collection.”
* “Data collection began for state educational agencies (SEAs) submitting data on behalf of their LEAs on January 23rd, 2019 and LEAs on February 4th, 2019. The data collection closed to all data submitters on June 21st, 2019.”

LEAs could self report or SEAs could report on their behalf

SLGA: “This data set contains address geocodes (estimated latitude/latitude values) and other geographic indicators to public schools, public local education agencies, private schools, and postsecondary schools.”

* <https://nces.ed.gov/programs/edge/geographic/schoollocations>
  + “Geocodes for public schools and school district administrative offices are based on data reported in the NCES Common Core of Data (CCD), an annual collection of administrative data about enrollment, staffing, and program participation for schools, local education agencies (LEAs), and state education agencies (SEAs).”
  + SEAs reported on behalf of LEAs

The discrepancy in the count of LEA (Local Education Agency) IDs between Python and Excel can be attributed to several factors. Differences in data cleaning and formatting methods in Excel and Python may cause variations in the count. Excel often applies automatic data formatting, while Python provides more manual control over data cleaning, potentially leading to differences. Additionally, Python can be more sensitive to whitespace, special characters, and case sensitivity, so ensuring consistent handling of these aspects is crucial. Furthermore, the data type treatment, such as text or numeric interpretation, should match in both tools to avoid inconsistencies. Missing data or empty cells may also affect the count, and encoding disparities can lead to variations in special character handling. Finally, Excel functions like COUNTA and COUNTIF may count cells differently than a custom Python script, so using similar counting methods in both tools is important for accurate results. To resolve the issue, it's essential to thoroughly review the data, ensure uniform data sources and formats, employ similar counting techniques, and handle data cleaning consistently in both Python and Excel.

**Question 5: At this point, imagine that you are given the task of finding out the number of school districts in the nation. Which file do you trust more? Both are federal databases. Two different analysts might end up with different results using different datasets, or, worse, using the same database, as we showed with the exploration of the CRDC file. Reflect on this.**

Algebra I\*

Corporal Punishment\*

Expulsions\* - “extra data after last expected column”

Harassment and Bullying\*\*\*

Justice Facilities\*\*\*

Suspensions\*\*\*

**Issue with null row:**

Harassment and bullying

Justice facilities

Suspensions

Single-sex classes (null for LEA state name on row 97634)

School Support

* ERROR: missing data for column "lea\_state\_name" CONTEXT: COPY School\_Support, line 97634: ""

School Expenditures same issue

Referrals and Arrests - need to delete null row

Restraints and Exclusions - need to delete null row

School Characteristics - need to delete null row