

Extract

I extracted my data from Kaggle.com. Kaggle.com got their information from the following sources:

Enrollment

<https://nces.ed.gov/ccd/stnfis.asp>

Financials

<https://www.census.gov/programs-surveys/school-finances/data/tables.html>

Academic Achievement

<https://www.nationsreportcard.gov/ndecore/xplore/NDE>

The data that I gathered was in the format of csv files.

Transform

The data was cleaned in the pandas and in the csv file. I encountered dates that had superscripts which became problematic to clean in pandas. Therefore, I dropped those dates in the csv file. The data a reformatted to fit PGAdmin requirements with all columns switched to lower case letters. I also added another column to the math test score data to create a foreign key to the revenue data.

Load

The data was then loaded into PGAdmin. I created two tables to house the data from my pandas dataframes. I chose PGAdmin because I felt that my data fit its rigid structure and I have the best understanding of PGAdmin over the other choices at this time.