ETL – Project for Week 13

Team: Eve Barkley, John McIntyre, Rupesh Patel

Our team decides to look at the National School Lunch Program. We had an unrealistic assumption that the reporting of student participation was managed and stored at the National Level because the program ties into the SNAP program, which uses US Census records and the USDA program for child nutrition. The School Lunch Program was in signed into law in 1946, by Harry Truman. It was originally established as a way to prop up food prices by absorbing farm surpluses, that would provide food to school age children who qualified to get free or reduced lunches at school. The plan was started as subsidy program that gave most of it’s support as cash reimbursement for meals served. In addition, schools and school districts were strongly encouraged to receive and purchase surpluses of agricultural commodities in their locales. Up until the 1960’s counts of participating students and cash payments were managed at a National level.

In 1962 Congress modified the NSLP to adjust fund distribution based on the budget allocated for each state to determine actual need for assistance. Now state boards of education certify the enrollment numbers and the numbers of students who are identified for NLSP and the numbers of students receiving NLSP support.

To access the best possible raw data, our team found two sources:

The USDA Food and Nutrition Service NLSP for annual State totals: <https://www.fns.usda.gov/nslp>

* This data was collected from Excel spreadsheets

National Center for Education Statistics for annual National totals for enrollment, testing scores and demographic information on students <https://nces.ed.gov/>

* This data was collected from Excel spreadsheets

USDA – for Direct Certification of each States’ NLSP over a two-year period <http://www.fns.usda.gov/ops/reseach-and-analysis>

* This data was collected from a report pdf

The data collected had limits that eventually defined the scope of what information over time we could collect. We settled on collecting Annual information by State participation from 2016 to 2019.

We collected National data from 2016 to 2019 as it was available to match to the State collections of data and for student test scores, we were able to find complete datasets for 4th, 8th and 12th grade students. We chose the 8th grade test score set because middle school scores are found to be good indicators of future student performance.

Transform – goes here

The final database design was created for upload to Postgres. The datasets were not granular enough to create a normalized database architecture [Boyce Codd Normal Form]. We found we had two relatable structures using a star pattern. This design still allows for fast queries and will show activity and enrollment numbers in a year or over 5 years. The two star designs are maintained within the same database to allow some cross-relationships to show student test performance and student poverty across states.

Diagrams – Given

State:

National: