

# 1 Questions

- How does neighborhood poverty level impact school enrollment rates across different regions in the United States?
- How do poverty levels around schools vary by region in the United States, and what might these differences suggest about educational inequality?

## 2 Data Sources

### 2.1 Description of Data Sources

- **Dataset 1: School Neighborhood Poverty Estimates 2020-2021**

This dataset from the National Center for Education Statistics (*NCES*) includes data on neighborhood poverty levels surrounding schools, which can serve as a proxy for educational access and community socioeconomic conditions. [1]

- **Dataset 2: Report Card Enrollment**

This dataset provides detailed school enrollment statistics disaggregated by school, district, and state for the 2022-23 school year. It includes student counts by demographics, which can help analyze how regional poverty levels might influence enrollment and highlight disparities across regions. [2]

### 2.2 Data Structure and Quality

- **School Neighborhood Poverty Estimates 2020-2021**

NAME	IPR_EST	IPR_SE	SCHOOLYEAR	LAT	LON
Filter	Filter	Filter	Filter	Filter	Filter
Albertville Middle School	259	87	2020-2021	34.2602	-86.2062
Albertville High School	261	92	2020-2021	34.2622	-86.2049
Albertville Intermediate School	139	61	2020-2021	34.2733	-86.2201
Albertville Elementary School	227	110	2020-2021	34.2527	-86.221806
Albertville Kindergarten and PreK	373	124	2020-2021	34.2898	-86.1933
Albertville Primary School	224	109	2020-2021	34.2533	-86.2218
Kate Duncan Smith DAR Middle	296	85	2020-2021	34.5337	-86.2541

Figure 1: School neighborhood poverty estimate 2020-2021 dataset

The data is organized with geospatial data (latitude and longitude) defining the locations of various education systems. It also includes the income-to-poverty ratio (IPR) which measure the socioeconomics conditions in surroundings to the schools. Additionally, It defines the standard error of the IPR for each area.

### • Report Card Enrollment

NAME	IPR_EST	IPR_SE	SCHOOLYEAR	LAT	LON
Filter	Filter	Filter	Filter	Filter	Filter
Albertville Middle School	259	87	2020-2021	34.2602	-86.2062
Albertville High School	261	92	2020-2021	34.2622	-86.2049
Albertville Intermediate School	139	61	2020-2021	34.2733	-86.2201
Albertville Elementary School	227	110	2020-2021	34.2527	-86.221806
Albertville Kindergarten and PreK	373	124	2020-2021	34.2898	-86.1933
Albertville Primary School	224	109	2020-2021	34.2533	-86.2218
Kate Duncan Smith DAR Middle	296	85	2020-2021	34.5337	-86.2541

Figure 2: Report card Enrollment

The data is structured for analyzing school-specific or regional educational trends. Mainly it contains the statistical metrics of enrollment in different Education system. Also, providing the income category by which students are belonging.

### 2.3 Licenses and Permissions

The data sources are publicly available on under open-data licenses. Detailed license information can be found at License

## 3 Data Pipeline

The data pipeline engineering is designed and developed with python has three main modules: extractor, transform, and loader, each module contains respective functionalities. Furthermore, the pipeline helper has responsibility to initialize the desire configuration for data sources.

After configuration Initialization for data sources, each module perform its responsibility. First **extract** from extractor module is used to extract the data source from url, second **delete\_columns** from transform module deletes the list of useless columns specified for every dataset. Additionally, eradicate the rows contain null or empty record, once all the transformations have been applied, dataset is then loaded to sqlite database using **load\_df\_to\_sqlite** from loader module.

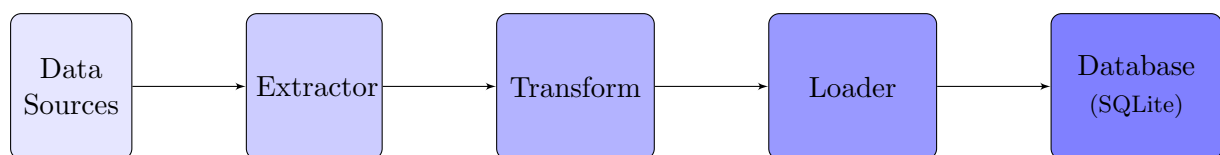


Figure 3: ETL Pipeline Flow

## 4 Result and Limitations

The pipeline stored an output in SQLite database in table format. This is one of the fastest and easier method to deal with the data. The pipeline is smart-enough to maintain the stored data quality.

- All necessary information is available to support the analysis questions.

- The data is consistent in their formats.
- The data reflects the real world and are correct indicators.

## References

- [1] Open. School neighborhood poverty estimates, 2020-21. <https://catalog.data.gov/dataset/school-neighborhood-poverty-estimates-2020-21>, 2021.
- [2] Open. Report card enrollment 2022-23 school year, year-end update. <https://catalog.data.gov/dataset/report-card-enrollment-2022-23-school-year-year-end-update>, 2023.