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** I will add details here.

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: First 7 rows of school neighborhood poverty estimate 2020-2021 dataset.

- **Report Card Enrollment** I will add details here.

GradeLevel	All Students	Female	Male	Highly Capable	Homeless	Low-Income	Mobile	Non-Highly Capable	Non-Homeless	Non-Low Income	Non Mobile
Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter
8th Grade	169	84	85	4	0	35	4	165	169	134	165
All Grades	521	264	257	18	0	125	5	503	521	396	516
4th Grade	155	79	76	11	0	31	1	144	155	124	154
5th Grade	155	70	85	10	1	33	3	145	154	122	152
All Grades	473	216	257	21	2	93	5	452	471	380	468
Kindergarten	123	55	68	0	1	21	0	123	122	102	123

Figure 2: First 6 rows of report card enrollment

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.

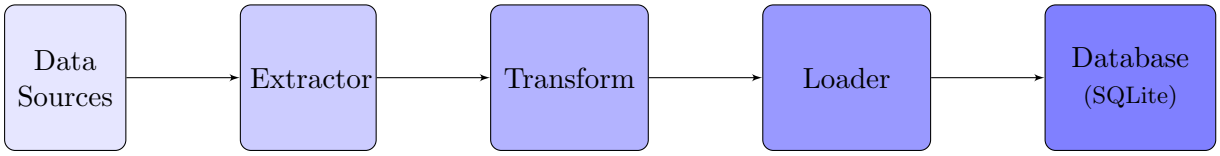

Figure 3: ETL Pipeline Flow

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 word 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.