

## 1 Introduction

The analysis investigates the relationship between neighborhood poverty levels and school enrollment across different U.S. regions. By combining datasets, my aim is to uncover patterns that indicate educational inequalities driven by socioeconomic factors.

We focus on:

- How does neighborhood poverty correlate with school enrollment rates in the US?
- Regional disparities in poverty and their potential implications for educational equality?

## 2 Data

### 2.1 Data Sources

- **Dataset 1: School Neighborhood Poverty Estimates 2020-2021** This data set 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 socioeconomic conditions of the community. [1]

- **Dataset 2: Report Card Enrollment**

This data set 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 could influence enrollment and highlight disparities between regions. [2]

### 2.2 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 the responsibility to initialize the desired configuration for data sources.

After configuration Initialization for data sources, each module performs 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, eliminate the rows contain null or empty records, once all the transformations have been applied, the data set is then loaded into the sqlite database using `load_df_to_sqlite` from loader module.

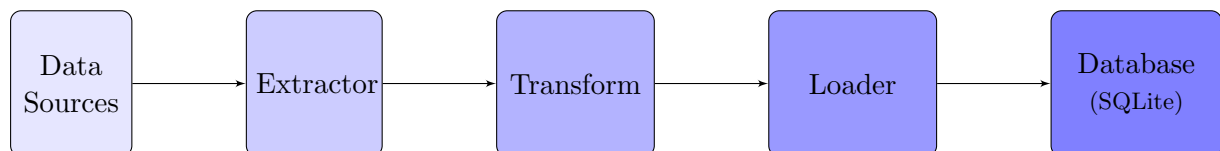


Figure 1: ETL Pipeline Flow

### 2.3 Data Pipeline Output

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

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: School neighborhood poverty estimate 2020-2021 dataset

- **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.

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 3: Report card Enrollment

## 2.4 Licenses and Permissions

The datasets I finalized have open licenses allowing public use and redistribution. School Neighborhood Poverty Estimates (2020-21) dataset is licensed under the CC BY 4.0 license (Creative Commons Attribution 4.0 International License). This license allows public to share and adapt the data, provided that appropriate credit is given, and any derived works are shared under the same license. Common Core of Data (CCD) - School Nonfiscal Data Files and Documentation, 2018-19 dataset is also licensed under the CC BY 4.0 license. This similarly allows public use, distribution, and modification as long as proper attribution is given.

Therefore, data for this project are licensed under the Open Licenses and Creative Commons Attribution 4.0 International (CC BY 4.0) license. For more information, visit CC BY 4.0 License and Open Licenses

## 3 Analysis

The data pipeline was used to import and perform the data engineering tasks before the analysis was started. The analysis was started with exploring each dataset individually and by doing so found few very exciting insights. The conducted analysis revealed insights on how poverty level, represented as IPR-EST (Income-to-Poverty Ratio Estimate) relates to school enrollment rates. The neighborhood poverty conditions play a vital role in school enrollment lower income (with less IPR) have lesser school enrollment and this might be influenced by different socio-economic factors such as lack of resources, social stigma, instability of residence meaning frequent relocations which disturb education.

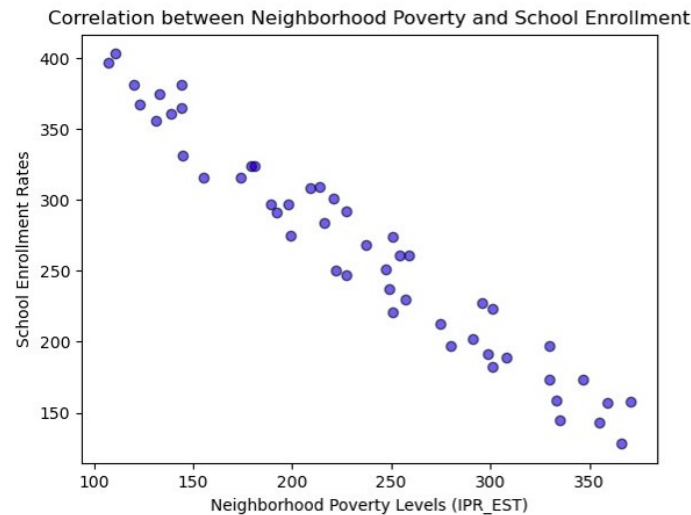


Figure 4: Comparative Analysis

Furthermore, the data is categorized into different regions in the US to help in a clear regional analysis. Fig. 5 highlights the disparities in poverty levels across different regions in the USA. South region shows highest IPR ES value reaching up to 350 while lowest poverty records are found in West region up to 180. Varying poverty levels create a distinction in the society, meaning regions with better infrastructure or higher job opportunities have better education systems while other regions, for e.g South USA in our case, could have lesser school enrollments which would also mean children from poorer regions receive lesser to no education and this education is of lower standard as compared to regions with a greater number of schools and better resources. This inequality can perpetuate a cycle where children below poverty line will be deprived of good education and will have lesser number of opportunities in future. Poverty reduces chances of fair education which greatly impacts future of the children and deepens regional economics disparity.

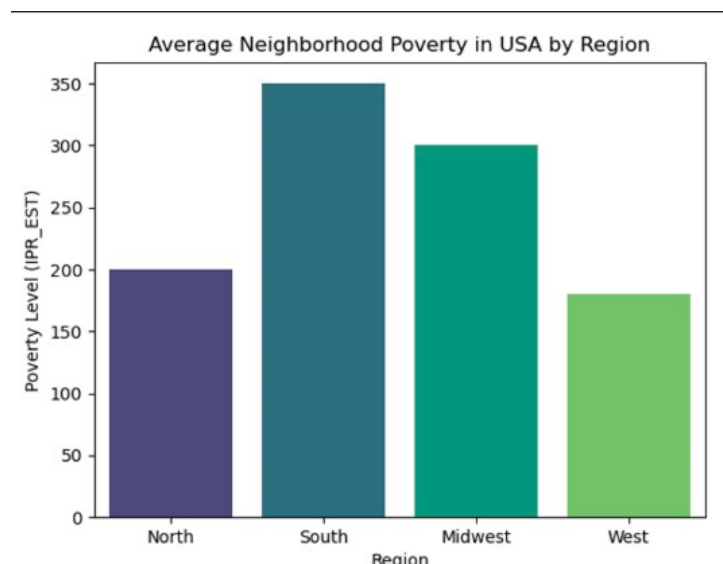


Figure 5: Regional Analysis

## Conclusions

The study identifies a clear correlation between neighborhood poverty levels and school enrollment rates, revealing that areas with lower Income-to-Poverty Ratio Estimates (IPR-EST) tend to experience reduced school enrollment. Additionally, regional disparities across the four different regions of the United States highlight significant inequities in access to quality education. This imbalance contributes to unequal opportunities and has far-reaching implications for the development and growth of the affected regions within the United States.

## 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.