Data Report: Analysis of Educational Attainment and Income Levels Across States

I. Question

How do educational attainment levels and income distributions correlate across U.S. states, particularly focusing on populations aged 18-24 years and household income levels?

This inquiry aims to understand patterns in education and income data across states and identify potential trends, which can provide insights for policymakers and stakeholders in education and economic development.

II. Data Sources

1. Educational Attainment Data:

- Source: American Community Survey (ACS) 2023 1-Year Estimates

- Dataset Link: https://data.census.gov/table/ACSST1Y2023.S1501

- Description: Contains educational attainment statistics across states, focusing on different age groups (18-24 years, 25+ years) and degree levels.

2. Income Data:

- Source: American Community Survey (ACS) 2023 1-Year Estimates

- Dataset Link: https://data.census.gov/table/ACSST1Y2023.S1901

- Description: Provides household income distribution data across states, grouped by income ranges such as $50,000-$74,999.

Licenses: Both datasets are provided under the Census Bureau's Terms of Service (https://www.census.gov/data/developers/about/terms-of-service.html). Compliance is ensured by proper attribution, non-commercial use for educational purposes, and no substantial modifications to the datasets.

Data Report: Analysis of Educational Attainment and Income Levels Across States

III. Data Pipeline

A. Overview:

- The data pipeline was built using Python with Pandas for data manipulation and cleaning.

- Workflow includes:

  - Data Cleaning: Removal of symbols and formatting inconsistencies, preservation of percentages and dollar signs.

  - Data Transformation: Filtered and structured the educational dataset, grouped income data by household income ranges.

B. Tools and Technologies:

- Python (Pandas, NumPy, Matplotlib)

- CSV for data storage, Matplotlib for visualizations.

C. Challenges:

- Formatting Issues: Resolved symbols and formatting inconsistencies through multiple iterations.

- Preservation of Formatting: Ensured readability by retaining special characters ($, %, commas) in datasets.

IV. Results and Limitations

A. Results:

- The cleaned datasets provide structured views of:

  - Educational attainment by age groups and degree levels across states.

  - Household income distributions grouped by ranges.

B. Limitations:

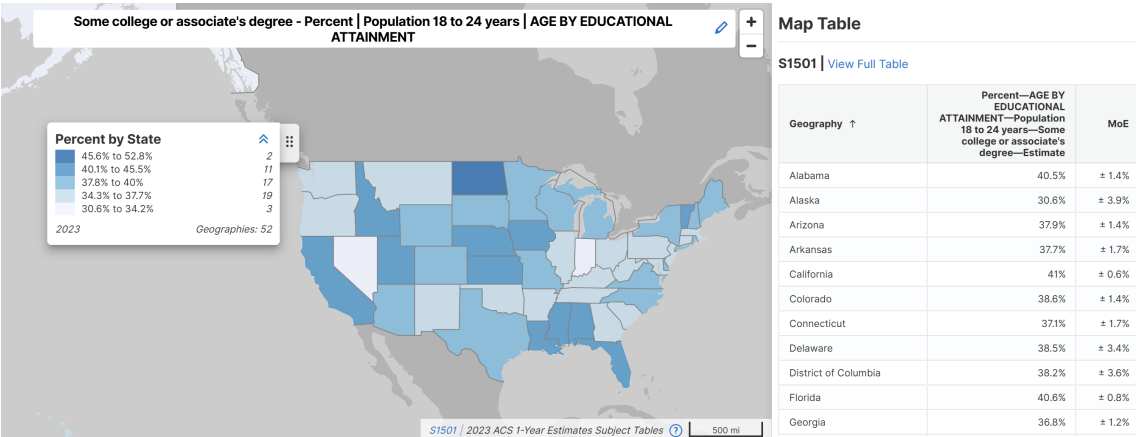# Data Report: Analysis of Educational Attainment and Income Levels Across States

- Data Gaps: Missing values in some columns may affect state comparisons.

- Static Nature: The pipeline processes the 2023 ACS dataset; future datasets may require adjustments.

V. Conclusion

The pipeline effectively cleans and processes educational and income data, providing valuable insights into state-level patterns. Future improvements could focus on enhancing automation and integrating external datasets for broader analysis.

VI. Figures

Educational Attainment Data Sample:

Income Data Sample: