The goal of this report is to explain the method and steps used to study socio-economic and spatial data of U.S. counties. The study looked for patterns and information about jobs, education, and income in different counties by using geospatial maps. I collected data on education, income, and economic characteristics (employment or unemployment) for the years 2023, 2022, 2021, and 2019. However, I could not integrate the shapefile, so I was unable to organize the data properly.

After this point, I described what I would have done if I had been able to complete the process.

1-Load the shapefile using **GeoPandas**.
2- Load employment, education, and income data in **CSV format** using **Pandas**.
3- Merge datasets based on a common column (**county FIPS code**).

4- Handle missing values in the dataset.
5- Ensure that **shapefile geometries** are valid.
6- Normalize indicators like **income** and **employment rate**.

7- Use **GeoPandas** to visualize the geographical distribution of counties.
8- Use **Matplotlib** to create basic graphs.
9- Use **Seaborn** for more readable charts.

10- Use .corr() method to calculate the correlation between **education level and income**.
11- Use **sm.OLS()** to estimate how income changes based on **education level and employment rate**.
12- Use **sm.add_constant(X)** to add an intercept to the regression model.
13- Create an **OLS (Ordinary Least Squares) regression model** with **sm.OLS(y, X).fit()**.
14- Print the statistical summary of the model using **model.summary()**.

If education level and employment rate **significantly affect income**, the model coefficients will be **meaningful ($p < 0.05$)**, and the **R^2 value** will be high.

- **Positive Coefficients**: If education level and employment rate have **positive coefficients**, this means that as these values increase, **income also increases**.
- **High p-value ($p > 0.05$)**: If the **p-value** is high, education and employment rate **do not have a significant effect** on income.