Documentation for Indian state wise analysis.

To perform a state-wise analysis, I followed the following steps:

Data Preparation:

Initially, I utilized the pandas library to split the "State & District" column into separate components, namely "State," "District_name," and "District_code." Consequently, I removed the original "State & District" column.

State & District
district_code:1, state_name:JAMMU AND KASHMIR, district_name:Kupwara
district_code:2, state_name:JAMMU AND KASHMIR, district_name:Badgam

District_code	State	District_name
1	JAMMU AND KASHMIR	Kupwara
2	JAMMU AND KASHMIR,	Badgam

Data Optimization:

To streamline the dataset and enhance efficiency, I removed several redundant columns, as outlined below:

Population Column: The "Population" column was eliminated since the sum of "Male" and "Female" populations yields the total population, improving data efficiency.

Literate Column: To reduce redundancy, I removed the "Literate" column, as it can be calculated by summing the "Male_Literate" and "Female_Literate" columns.

SC Column: The "SC" column was also deleted, as it provides the same information as the combined "Male_SC" and "Female_SC" columns.

ST Column: Similarly, the "ST" column was removed, as it duplicates the information found in the "Male_ST" and "Female_ST" columns.

Workers Column: To simplify the dataset and reduce redundancy, I eliminated the "Workers" column, as its data can be derived by aggregating "Main_Workers," "Marginal_Workers," "Non_Worker," "Cultivator_Workers," "Agricultural_Workers," "Household_Workers," and "Other_Workers."

Total_Education Column: I also removed the "Total_Education" column since it duplicates information already present in the "Literate_Education" and "Illiterate_Education" columns.

Household Column: The "Household" column was deemed redundant as it provides the same data as the combined "Rural_Household" and "Urban_Household" columns.