**Geospatial Analysis Assignment: Exploring Climatic Health Variables**

**Objective:**The objective of this assignment is to perform a comprehensive geospatial analysis using climate and health variables to identify potential correlations or patterns in a given region.

**Dataset Description:** You are provided with a village level dataset containing the following variables:

*[Note: The dataset is anonymized with dummy values]*

* 'rajas\_id': Unique identification code for the region
* 'block\_code\_census': Census block code for the region
* 'rain': Rainfall data
* 'population\_density': Population density in the region
* 'rural\_facilities': Number of healthcare facilities in rural areas
* 'land\_surface\_temperature\_during\_day': Daytime land surface temperature
* 'land\_surface\_temperature\_during\_night': Nighttime land surface temperature
* 'ultraviolet\_index': Ultraviolet index
* 'pm\_2.5': Particulate Matter (PM 2.5) concentration
* 'geometry': Geometric data representing the region
* 'number\_of\_tb\_cases': Number of tuberculosis cases
* 'number\_of\_heart\_disease\_cases': Number of heart disease cases
* 'number\_of\_cancer\_cases': Number of cancer cases
* 'number\_of\_child\_malnutrition\_cases': Number of child malnutrition cases
* 'number\_of\_maternal\_anemia\_cases': Number of maternal anemia cases
* 'number\_of\_smoker\_cases': Number of cases related to smoking
* 'Number\_of\_lung\_problem\_cases': Number of cases related to lung problem

**Filename:** climate\_health\_anonymized\_dataset.geojson

**Tasks:**

1. **Data Exploration (10 points):**
   * Load the dataset and perform initial exploratory data analysis (EDA) to understand the structure, summary statistics, and distribution of variables.
2. **Geospatial Visualization (20 points):**
   * Visualize the geographic distribution of variables such as rainfall, population density, healthcare facilities, etc., on a map using appropriate visualization libraries
3. **Correlation Analysis (10 points):**
   * Analyze the correlation between climatic variables (rainfall, temperature, UV index, etc.) and health-related variables (number of TB cases, heart disease cases, etc.). Visualize these correlations using suitable techniques.
4. **Spatial Analysis (30 points):**
   * Conduct spatial analysis to identify spatial patterns or clusters of health-related cases across different regions. Use spatial clustering techniques or mapping tools to visualize these patterns.
5. **Hypothesis Testing (10 points):**
   * Formulate a hypothesis (Does higher PM 2.5 concentration correlate with a higher number of lung diseases?), perform statistical tests, and present findings with appropriate visualizations.
6. **Insights and Recommendations (20 points):**
   * Based on your analysis, provide insights or recommendations for policymakers or healthcare professionals. Identify areas of concern or potential interventions based on the observed patterns or correlations.

**Submission Guidelines:**

* Software to be used: R / Python / ArcGIS / QGIS
* Present your analysis in a Jupyter Notebook or on Google Colab or in a presentation format.
* Include detailed explanations, code, visualizations, and interpretations for each step.
* Provide actionable insights and recommendations based on your analysis.