General Project Instructions

Objective:

Develop a project that leverages geospatial data, machine learning techniques, and visualization tools to solve a real-world problem. Use **Geopandas** for geospatial data processing and **Folium** for creating interactive visualizations.

1. Project Scope

- Choose a topic that involves geospatial data and machine learning. Examples include:
    - Predicting flood-prone areas
    - Analyzing traffic congestion patterns
    - Identifying optimal locations for renewable energy sites
    - Mapping crime hotspots
    - Analyzing urban growth or deforestation
    - Assessing the impact of climate change on a specific region
- Clearly define the problem you want to address.

2. Tools and Technologies

- **Programming Language:** Python
- **Libraries:**
    - Geopandas: Geospatial data handling and analysis
    - Folium: Interactive maps and visualizations
    - Scikit-learn: Machine learning models
    - Pandas & Numpy: Data analysis and manipulation
    - Matplotlib & Seaborn: Data visualization
- **Optional Tools:**
    - Google Earth Engine or QGIS for additional geospatial data.

3. Steps to Follow

Step 1: Topic Selection and Research

- Research potential topics and select a specific problem to solve.
- Identify key geospatial datasets relevant to your topic. For example:
    - Environmental data (e.g., terrain, vegetation, rainfall)
    - Socioeconomic data (e.g., population, infrastructure)

    - Satellite imagery or shapefiles

Deliverable: A 1-paragraph problem statement explaining the selected topic.

Step 2: Data Collection

- Search for reliable geospatial datasets from sources like:
    - **OpenStreetMap**
    - **NASA EarthData**
    - **World Bank Data**
    - Local or regional government open-data portals
- Save datasets.

Deliverable: A list of datasets used with brief descriptions and links.

Step 3: Data Preprocessing

- Load the datasets using Geopandas and Pandas.
- Perform cleaning steps:
    - Handle missing values.
    - Standardize coordinate reference systems (CRS).
    - Remove irrelevant features or outliers.
- Merge or join datasets as needed.

Deliverable: Cleaned and merged dataset ready for analysis.

Step 4: Exploratory Data Analysis (EDA)

- Use visualizations to understand the data:
    - Plot geospatial features using Geopandas.
    - Create heatmaps, choropleth maps, or other relevant plots.
- Identify patterns, trends, or correlations.

Deliverable: Graphs, charts, and maps summarizing your findings.

Step 5: Feature Engineering

- Generate features from geospatial data, such as:
    - Distances to specific landmarks (e.g., roads, rivers, schools).
    - Area sizes, population density, or terrain attributes.
- Use these features for machine learning.

Deliverable: Final dataset with engineered features.

Step 6: Machine Learning Model

- Choose a suitable machine learning approach:
- Split the data into training and testing sets.
- Train the model and evaluate its performance using appropriate metrics (e.g., accuracy, RMSE, etc.).

Deliverable: Trained machine learning model and evaluation results.

Step 7: Visualization Using Folium

- Visualize your results on an interactive map:
    - Use Folium to plot predictions or clusters.
    - Add relevant layers (e.g., roads, rivers, buildings).

Deliverable: An interactive map in HTML format.

Step 8: Documentation and Presentation

- Prepare a final report that includes:
    - Problem statement
    - Data sources and preprocessing steps
    - Machine learning methodology and results
    - Visualizations and insights
- Present your findings using slides and your interactive map.

Deliverable: Final report and presentation slides.

4. Deliverables

1. **Problem Statement:** A clear explanation of the chosen topic.
2. **Processed Data:** Cleaned and ready-to-use dataset.
3. **Machine Learning Model:** Trained model with performance evaluation.
4. **Interactive Map:** A Folium map showing key insights.
5. **Report and Presentation:** A detailed summary of the project.

5. Evaluation Criteria

1. **Relevance and Originality of Topic:** Is the topic well-defined and impactful?
2. **Data Handling:** Quality of data collection, preprocessing, and feature engineering.
3. **Model Implementation:** Accuracy and appropriateness of the machine learning model.
4. **Visualization:** Effectiveness and clarity of the interactive map.
5. **Documentation and Presentation:** Clarity and depth of the report and slides.