*Advanced Data Analytics – Final Project Preliminary Report*

*Name:*

**Data:** The dataset is a metadata file from the *Wind for Schools* project, which includes key details about wind energy installations at various educational and community facilities. The fields in this dataset are:

* **Turbine ID**: Unique identifier for each wind turbine.
* **Project Name**: The designated name for each wind project.
* **Facility**: The educational or community facility where the turbine is installed.
* **State**: The U.S. state where the facility is located.
* **Coordinates**: Geographic coordinates (latitude and longitude) for each wind installation.
* **Address**: Physical address for certain facilities, where available.
* **Installed Capacity**: The rated power generation capacity of the turbine in megawatts (MW).
* **Number of Units**: The number of turbines installed per project.

**Analysis:**

To conduct a thorough analysis of wind energy adoption patterns across facilities, this project will follow a structured approach, broken into the following steps:

**1. Define Objectives and Key Questions:**

* **Goal**: Understand how wind energy is distributed across the U.S., which facility types tend to adopt wind energy, and identify regions with notable wind capacity.
* **Questions**:
  + What are the most common states and facility types for wind installations?
  + What is the range of installed capacities, and are there any notable trends in size or scale?
  + How do installations differ geographically or by facility type?

**2. Data Collection and Understanding:** This dataset includes essential project details, so the initial step is to familiarize ourselves with each field’s purpose and limitations. A descriptive analysis will provide a basic overview of the distributions and identify any anomalies.

**3. Data Cleaning and Preparation:**

* **Handling Missing Data**: Address missing values in fields like *Coordinates* and *Installed Capacity* by either imputing them or excluding incomplete rows where appropriate.
* **Type Corrections and Standardization**: Ensure numerical values are accurately typed for computation (e.g., converting *Installed Capacity* to a numerical format).
* **Coordinate Processing**: Separate latitude and longitude for effective mapping and spatial analysis.

**4. Exploratory Data Analysis (EDA):**

* **Statistical Summary**: Calculate key statistics, such as mean, median, and range, particularly for *Installed Capacity* and *Number of Units*.
* **Visualization**:
  + **Geospatial Mapping**: Use mapping techniques to visualize the geographic spread of wind installations and highlight areas with high densities.
  + **Capacity Analysis**: Create visualizations such as histograms or box plots to explore installed capacity distribution, identifying average and outlier capacities.
  + **Cluster Analysis**: Group projects by state or facility type to reveal clusters, showing any notable patterns or concentrations in wind energy adoption.

**5. In-Depth Analysis and Modeling:** Based on EDA findings, a more focused analysis will address the following:

* **Regional and State-Level Analysis**: Group projects by state to identify trends in wind energy adoption by region.
* **Comparative Facility Analysis**: Compare characteristics of installations across different facility types, such as K-12 schools vs. universities, to detect any preferred installation practices or capacity variances.
* **Capacity-Units Relationship**: Investigate any correlation between the number of turbines (units) and the installed capacity per project to explore the impact of project scale on output.

**6. Interpret and Summarize Findings:** Summarize findings with a combination of visuals and insights. Key points may include:

* Identification of states or regions with a high number of installations and their corresponding capacities.
* Trends in capacity distribution, highlighting facility types that frequently adopt wind energy.
* Insights into potential areas for new wind projects based on patterns observed.

**7. Present and Document Results:** Develop a report and presentation that effectively communicates the project’s findings. Use visuals, including charts and geographic maps, to highlight significant patterns and support your insights. Focus on answering the core questions, explaining the implications of the findings, and suggesting how they could inform policy or energy project decisions.