Data Center
Environmental Siting Optimizer

Cloud Chasers

Team Member	School	Major	Year
Sahana Krishnaraj	College of Arts and Sciences	Applied Statistics	2026
Divya Ramakrishnan	College of Arts and Sciences	Applied Statistics	2026

How can data center operators balance environmental sustainability with operational reliability when selecting sites for new facilities?

Data centers are energy-intensive infrastructure critical to modern digital economy, yet their environmental impact, including energy consumption, wildfire risk, drought vulnerability, and reliance on fossil fuels, is rarely considered in siting decisions.

Our product demonstrates that environmental responsibility and operational performance can coexist in data center siting. Without transparent tools such as our product to evaluate these trade-offs, companies risk building in high-risk zones, exacerbating local environmental stress, and missing renewable energy opportunities.

Who's Affected:

- Data center operators balancing growth and sustainable practices
- Policymakers managing environmental protection
- Local communities currently facing economic and environmental strain

Our Solution: An interactive web app that quantifies environmental trade-offs in data center siting.

- Demonstrates that sustainable and profitable siting are compatible goals.
- Allows companies to identify low-risk, higher-sustainable locations without sacrificing operation.
- Data center developers gain a competitive advantage through more environmental and therefore more socially-acceptable siting.

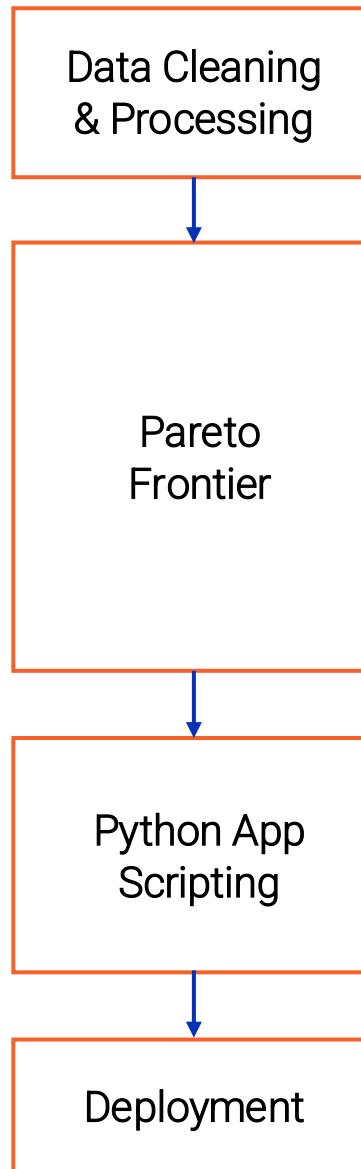## Data Center Environmental Siting Optimizer Web App

Data Sources

- US Wind Power Plant Capacity
- US Oil and Gas Production Values by County
- US Wildfire Hazard Potential by County
- US Drought Risk by County
- US Census State and County-Level Geodata

Considerations and Limitations

- Some counties do not have drought-risk data.
- Not all counties have information on oil and gas production.
  - This is due partly to counties not producing any energy and partly to the production data missing in the data source.
- Although recent, this data is not updated real-time.

Project Workflow

Data Cleaning & Processing

- Cleaned raw data for each factor and normalized all quantities into scores ranging from 0-1.
- Merged data with shapefiles to obtain locations to display on map.

Pareto Frontier

Pareto Frontier: In an optimization problem with many distinct factors, the Pareto front represents the set of solutions that are equally good. It allows us to restrict attention to the set of efficient choices and make trade-offs with this set.

What and Why?

- Selects counties that are better than the other countries in all other factors and are strictly better on at least one factor.
- Originally an economics principle, but it was applicable to our project because it helps accurately assess tradeoffs.

Steps:

1. Iterate through the dataset, comparing each county to every other county in the dataset. Nested loops used for this approach.
2. Check for dominance. A county is dominated if another county has a lower Drought Risk, Wildfire Hazard Risk, Oil production, Gas production and higher Wind Power capacity.
3. Check to see if the other county is strictly better in at least one of the factors. If so, the previous county is not considered Pareto efficient.
4. Run this loop through the entire dataset and subset only Pareto efficient counties. From this set, 10 counties were selected.

Python App Scripting

- Developed functions that computed the environmental suitability score, which is a weighted average of the user's desired weight for each metric. These were computed on the 10 Pareto efficient counties.
- Normalized the suitability score from 0 to 100 for scalability and comparability purposes.
- Sorted the scores in ascending order. Displayed user-specified number of counties.
    - The most suitable county will always have a suitability score of 100%. The least suitable county will always have a score of 0%.
- Streamlit components were integrated (slider, text, table), and the Folium package was used to display the data on a map.

Deployment

- Deploy the application to Streamlit.

Web App Overview

Risk Factors
These sliders allow the user to choose how much they prioritize Wildfire and Drought risk when building a data center. The slider ranges from 0 to 1. A higher weight closers decreases the environmental compatibility score.

Energy Sources
These sliders allow the user to choose how much they prioritize the energy sources. Wind energy ranges from 0 to 1. Oil and Gas production ranges from -1 to 1. A higher value indicative that the user prefers higher oil/gas production, which decreases the environmental compatibility score. However, a higher value for Wind energy increases the environmental score.

Location
This slider allows the user to choose number of counties to be displayed.

Data Center Siting Optimizer Prototype

Recommends optimal locations to build data centers based on environmental risk factors.

Slider Options

Wildfire Weight — Low Wildfire Risk — 0.46 — High Wildfire Risk (0.00 – 1.00)

Drought Weight — Low Drought Risk — 0.54 — High Drought Risk

Wind Energy Weight — Low Priority — 0.58 — High Priority

Oil Production Weight — Low Preference — -0.66 — High Preference

Gas Production Weight — Low Preference — -0.69 — High Preference

Number of Locations — 3

Top 3 Recommended Counties:

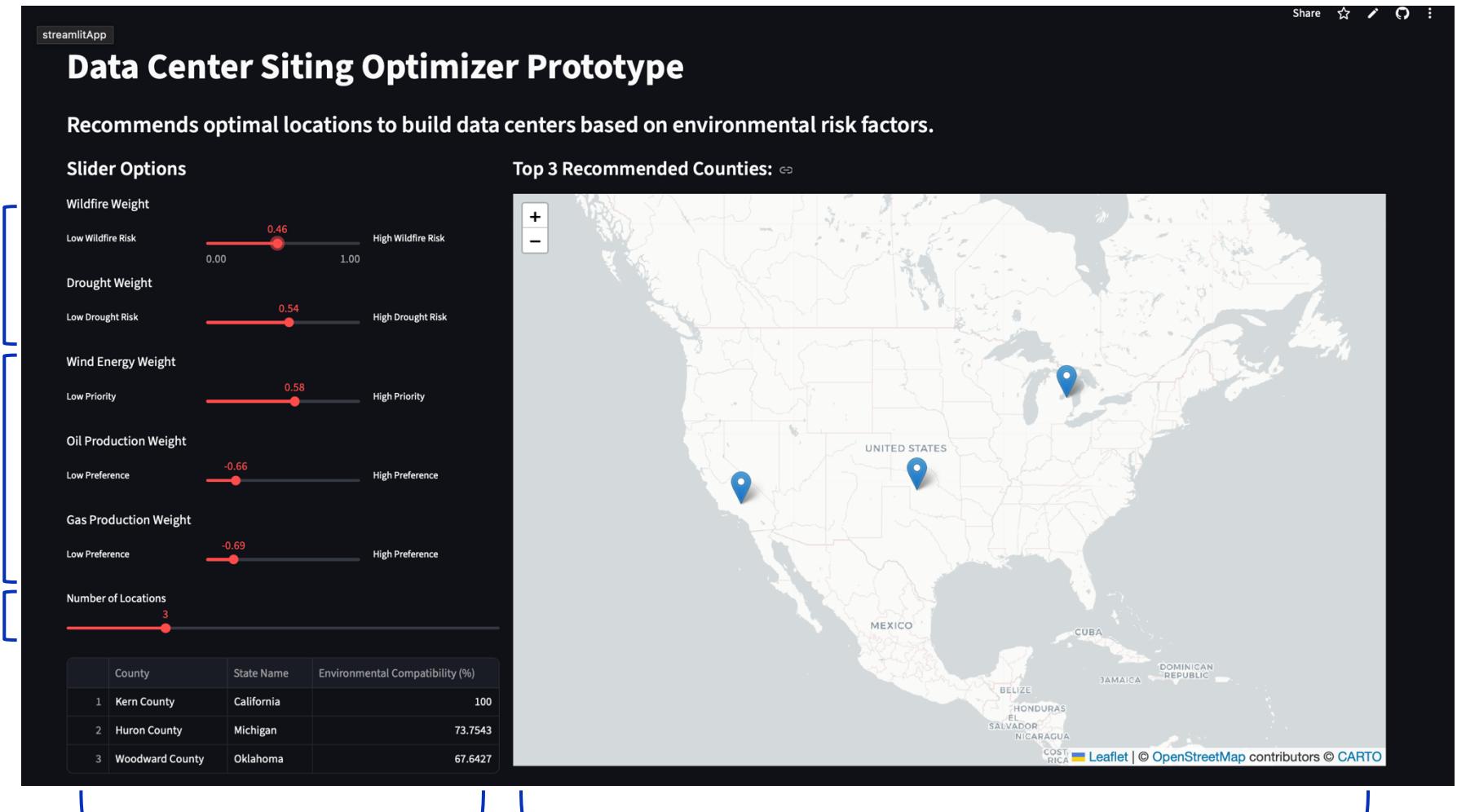|	County	State Name	Environmental Compatibility (%)
1	Kern County	California	100
2	Huron County	Michigan	73.7543
3	Woodward County	Oklahoma	67.6427

Table of Recommended Counties
A table with the county/state and a respective environmental compatibility score ranked from highest to lowest. This score is computed using user weights and is represented as a percentage. A higher score indicates better suitability to build a data center based solely on environmental factors.

Map Locations
This displays the selected counties on the U.S Map, making it easier to visualize.

Key Insight #1: 10 Pareto Efficient Counties Identified

These counties represent optimal trade-offs across sustainability metrics (wildfire risk, drought risk, renewable energy availability). No other U.S. counties offer superior environmental performance across all dimensions.

Key Insight #2: Minimal Overlap with Current Data Center Hotspots

Many environmentally-optimal counties do not overlap with existing data center hotspots, revealing an untapped potential for new, sustainable hotspot growth.

These insights demonstrate that environmentally-conscious siting is not only viable but in abundance. They challenge the assumption that operators must choose between being environmentally conscious and being profitable.

These surprising results go to show that there is a true economic niche in building data centers at sustainable locations, one that policymakers and operators can push for.

Potential Risks, Biases, & Unintended Consequences

- **Overreliance on Product:** Users may treat suggested counties as a definitive decision.

- **Incomplete Data Coverage:** Gas and Oil production data has geographic gaps that may skew results.

- **Narrow Scope:** The current model prioritizes environmental metrics but excludes economic, social, and operational factors critical to real-world feasibility

Community & Stakeholder Impact

- **Local Communities**: Counties may lack infrastructure, economic capacity, or political will to support data center development, including equity concerns.

- **Energy Producers**: Oil, gas, and wind energy production plants may be opposed to supporting the construction and operation of a data center. Our product only showcases potential resources yet does not promise their use.

Safeguards

- **Expand Metrics**: integrate county-level characteristic data (such as population level, economic readiness, social opinion, and legislation).

- **Policy Context:** include context on county-wide policies regarding data centers.

- **Clear Disclaimers:** Remind users that our product is decision-support, not prescriptive. The responsibility of due diligence lies with the user.

Data Completeness

- **Fill Geographic Gaps:** Obtain comprehensive data for all current metrics, especially oil, gas, and wind power.

- **Expand Renewable Energy Coverage:** Add metrics for solar, hydro, and geothermal energy.

- **Land Availability:** Integrate data on suitability of land, zoning, and environment constraints.

Broaden Criteria

- **Operational Sustainability:** Water and air availability for cooling. Consideration of air quality and climate change.

- **Economic Viability:** Proximity to major cities and demand markets, labor markets, and construction costs.

- **Social Consideration:** Community attitudes, employment impact, noise pollution, and environmental justice efforts.

- **Legislative Consideration:** State and county permitting requirements and local renewable and fossil fuel energy mandates.

Additional Data Needed

- **Comprehensive Data:** Obtain permission to collect and use data for each metric from every US county from sources such as the Department of Energy, US Geological Survey, Census Bureau, and more.

- **Web-Scraped Data:** Social and legislative consideration data must be web-scraped from social media platforms and official policy documentation.

- **Real-Time Data:** Connect to live energy production, risk, social media platforms, and legislative platforms for relevant data updates.

GitHub Repository

Data Sources:
- Wind Power Plant Data
- Gas and Oil Production Data
- Drought Risk Data
- Wildfire Hazard Potential Data
- US States Shapefile
- US Counties Shapefile
- US Census State & County Shapefile

Inspiration:
- NREL Siting Optimizer
- Electricity Grid Mix

Research:
- Data Center Metrics
- Environmental Cost of Data Centers
- DOE Request for Information on AI Infrastructure
- Pareto Frontier