Objective

The goal of this project is to explore how time-series environmental data can be used to build a machine learning model that identifies correlations or predictive relationships between grow conditions and plant growth.

Key Questions to Explore

- How can multi-variable time-series data (temperature, humidity, nutrients, etc.) be structured and synchronized with periodic plant weight measurements?

- What types of ML models are appropriate for this problem (e.g., regression, time-series forecasting, sequence models)?

- Can the model identify which environmental factors most strongly influence lettuce growth over time?

- Is it possible to predict future plant weight or growth rate based on recent environmental conditions?

- What insights or actionable takeaways can be extracted to improve growing conditions?

Scope and Deliverables

- Data preprocessing and feature engineering (e.g., aggregations, rolling averages, lag features).

- Selection and implementation of one or more ML approaches.

- Model evaluation and interpretation (correlations, feature importance, or sensitivity analysis).

- A short report or presentation summarizing findings and recommendations for optimizing the grow tent setup.

Constraints / Notes

- Environmental data is high-frequency (continuous or near-continuous).

- Plant weight measurements are sparse (every 4 days).

- The focus is on learning and insight generation rather than production-level deployment.

If you'd like, I can also:

- Simplify this for a **junior student**

- Make it more **research-oriented**

- Add **example datasets or starter ideas**

- Turn it into a **one-page project brief** or **intern-style assignment**