# Data Cleaning Policy Document

## Introduction

This document outlines the data cleaning and transformation policy for the agricultural and environmental data within our database. The purpose of this policy is to ensure data accuracy, completeness, and consistency by defining procedures for handling missing values, inaccuracies, and data transformation.

## Data Tables and Treatment

### Crop Table

#### Column: Treatment

Crop Type: Remove rows with missing data.

Crop Yield: Replace missing data with the average yield per crop type.

Growth Stage: Replace missing data with "seedlings."

Pest Issues: Replace missing data with "no pest issues."

### Location Table

Remove rows with null data entirely. The whole row is removed when any data is missing.

### Pest Table

#### Column: Treatment

Pest Type: No missing data.

Pest Description: Select the most occurring description within a subset of pest type.

Pest Severity: Select the most occurring severity within a pest type.

### Sensor Table

Calculate the mean (average) value for all numeric columns. All numeric columns are transformed by replacing missing data with their respective column averages.

### Soil Table

Calculate the mean (average) value for all columns. All columns are transformed by replacing missing data with their respective column averages.

### Weather Table

Fill missing numeric values with the average value within their respective weather condition column.