**AgriSmart AI: Precision Agriculture Data Cleaning & Analysis Report**

**1. Introduction**

As a data scientist hired by AgriSmart AI, I was tasked with preparing and analyzing climate-action-related agricultural data to assist in precision agriculture operations. The dataset, **climate\_action\_data.csv**, contains information from farm sensors and drones measuring environmental factors crucial for optimizing irrigation and fertilization.

**2. Dataset Overview**

* **Initial records**: 913
* **Variables (Columns)**: 10
* **Issues identified:**
  + Missing values in several columns (e.g., 'Date')
  + Erroneous entries labeled as 'error'
  + Incorrect data types (dates as strings)
  + Duplicate rows

**3. Data Cleaning Steps**

1. **Removed Duplicate Records:**

python

CopyEdit

df = df.drop\_duplicates()

1. **Replaced Erroneous Values:**
   * Values labeled as 'error' replaced with np.nan
2. **Data Type Corrections:**
   * Converted 'Date' column to datetime format.
3. **Handled Missing Data:**
   * Dropped rows where critical variables were missing (e.g., 'Soil\_Moisture(%)', 'Crop\_Type')
   * Missing non-critical numeric values were filled using column means/medians.
4. **Standardized Column Names:**

python

CopyEdit

df.columns = df.columns.str.strip()

1. **Final Dataset Shape:** 799 records × 10 variables
2. **Exported Clean Dataset:** cleaned\_precision\_agriculture\_data.csv

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

**4.1 Descriptive Statistics**

* **Soil Moisture (%):** Mean ~45%, range ~20% - 80%
* **Soil pH:** Normally distributed around 6.5
* **Temperature (°C):** Mean ~28°C, but with values exceeding 30°C in hot periods

**4.2 Visualizations**

* **Histograms** plotted for key numeric variables to inspect distributions
* **Correlation Heatmap** identified:
  + Strong positive correlation between **Humidity(%)** and **Soil Moisture(%)**
  + Fertilizer recommendations showed moderate correlation with **Soil pH** and **Soil Moisture(%)**

**5. Key Insights & Recommendations**

**5.1 Variables Influencing Fertilizer Recommendations**

* **Soil pH** and **Soil Moisture(%)** have the strongest correlation with **Fertilizer\_Recommended(kg/ha)**.
* Slight influence from **Temperature(C)** as well.

**5.2 Crop Type with Highest Average Soil Moisture**

python

CopyEdit

df.groupby('Crop\_Type')['Soil\_Moisture(%)'].mean().sort\_values(ascending=False).head(1)

**Wheat:** **47.36%** average soil moisture

**5.3 Irrigation Adjustments for Crops with Average Temperatures > 30°C**

python

CopyEdit

hot\_crops = df[df['Temperature(C)'] > 30]

hot\_crops.groupby('Crop\_Type')['Soil\_Moisture(%)'].mean()

* Crops experiencing **>30°C** include Beans, Lettuce, Maize, Tomatoes, and Wheat.
* **Tomatoes** had the highest soil moisture in hot conditions (**48.65%**), suggesting they retain moisture better under heat.
* **Recommendation:** Increase irrigation frequency **for Lettuce and Beans**, which show relatively lower average moisture during high temperatures.

**6. Exported Clean Dataset**

climate\_cleaned\_data.csv

**7. Conclusion**

Cleaning and analyzing the climate-action dataset revealed key relationships between environmental conditions and agricultural needs. These insights can directly support **precision irrigation schedules** and **fertilizer planning**, improving yield efficiency at AgriSmart AI