**Understanding Rice Cultivation Practices and Yield Patterns among Smallholder Farmers in India: An Exploratory Data Analysis**

**Introduction**

Agricultural productivity among smallholder farmers in India remains variable due to diverse agroecological conditions and management practices. Understanding these variations is essential to identify bottlenecks and effective strategies for improving crop outcomes.

This project uses detailed farm-level survey data collected by Digital Green in collaboration with JEEViKA and FAIR Forward. The dataset captures information on rice cultivation practices, land characteristics, input use, pest management, and yields for over 5,000 farmers in India.

The project aims to explore the data to describe current farming practices, identify trends, and detect relationships between agronomic decisions and yield outcomes.

**Project Objectives**

1. Descriptive profiling of rice farming practices across surveyed farmers.
2. Identify patterns and trends in yield outcomes based on farm size, soil type, fertilizer use, irrigation frequency, pest management, and other variables.
3. Compare yield and input-use characteristics across subgroups (e.g. districts, landholding size categories, number of irrigation events).
4. Document challenges and data gaps that may inform future data collection and program design.
5. Generate evidence-based recommendations for farmer-facing extension services.

**Research Questions**

* What are the most common agronomic practices (fertilizer use, irrigation, transplanting timing, pest control) among rice farmers in the dataset?
* How does rice yield vary by farm size, input application, irrigation frequency, and other land or management factors?
* Are there observable trends in yield associated with specific practices (e.g., use of basal fertilizer, organic manure, early transplanting)?
* How do farming practices and outcomes differ by region, soil type, or irrigation access?
* What data quality or consistency issues exist that might affect program evaluation or policy recommendations?

**Data Description**

The dataset comprises approximately 5,000 farm-level observations on rice cultivation practices and yields, with over 40 variables including:

* **Farm and land information**: land size, soil type, tillage type, drainage conditions.
* **Input use**: basal and top-dress fertilizer application (quantities by type: Urea, DAP, NPK, SSP, Zinc), organic inputs.
* **Crop management**: transplanting date, nursery practices, irrigation count, pest/disease management type.
* **Harvest data**: harvest date, method, post-harvest practices.
* **Exposure to extension services**: digital scorecards, video viewing, geofencing activities.
* **Yield outcomes**: crop yield (kg) per main plot area.

**Expected Deliverables**

1. Cleaned and structured rice cultivation dataset.
2. EDA report with:
   * Descriptive statistics.
   * Visualizations of input use and yield patterns.
   * Subgroup comparisons and trend analyses.
   * Data quality assessment.
3. Summary of key agronomic trends and yield variations.
4. Recommendations for extension agents and program planners.
5. Dashboards for field-level use.

**Timeline**

| **Phase** | **Duration** |
| --- | --- |
| Data Cleaning & Prep |  |
| Exploratory Data Analysis |  |
| Subgroup & Trend Analysis |  |
| Data Quality & Insight Report |  |
| Recommendations & Reporting |  |

**Potential Impact**

* Provides a clear, data-driven snapshot of rice cultivation practices among smallholders.
* Highlights areas for targeted intervention (e.g., increasing irrigation access, promoting balanced fertilization).
* Equips extension workers with evidence to guide farmers toward practices associated with better outcomes.
* Informs future data collection priorities and survey tool improvements.
* Aligns with broader sustainable agriculture and food security goals in Bihar/India.

**References**

* Digital Green, & GiZ. (n.d.). *FRAME templates* [Data set]. GitHub. <https://github.com/digitalgreenorg/frame-templates>