Improving Profitability for Sham's Small-Scale Wheat-Paddy Farm in Haryana

A Mid-Term report for the BDM capstone Project

Submitted by

Name: Litesh

Roll number: 24f2003468

IITM Online BS Degree Program,

Indian Institute of Technology, Madras, Chennai

Tamil Nadu, India, 600036

Contents

1. Executive summary

This report examines a farm in rural India, which primarily grows wheat and paddy. The farm faces several pressing challenges: high irrigation expenses due to reliance on diesel pumps and rented tube wells; rising fertilizer and pesticide costs that strain the budget; increasing labor costs and seasonal shortages; and a persistent gap between MSP and rising input costs. These issues collectively threaten the farm's profitability and long-term sustainability.

To explore these challenges, we collected detailed primary data through direct field visits, structured interviews, and observation. Key variables collected include irrigation costs, fertilizer and pesticide usage, seed expenses, labor wages and days, crop yield, and minimum support price (MSP). All information was initially recorded in a notebook and later digitized using Google Sheets. The dataset covers three crop cycles and provides a detailed breakdown of both cost and production variables, enabling comprehensive descriptive statistical analysis.

Using this dataset, the analysis began with the systematic data cleaning, standardization, and the calculation of descriptive statistics such as mean, median, mode, standard deviation, minimum, and maximum for all major variables. Visualizations, including pie and bar charts, were used to illustrate cost structures and compare input expenses across crop cycles. The findings reveal that labor, irrigation, and fertilizer are the largest and most variable costs, with paddy cultivation being the most input-intensive but also the most profitable. These insights highlight the main cost drivers and profit patterns.

2. Proof of originality

To establish the authenticity and originality of the data and analysis, the following evidence is provided:

- **Video Interaction with the Farm Owner:**
  A video interview was conducted with the farm owner, covering the farm's background, key challenges, and expectations from this project.
- **Images of the Farm:**
  Photographs documenting the farm premises, equipment, and the data collection process are included as visual proof of fieldwork.
- **Data Collection Notebook:**
  All primary data was initially recorded by hand in a dedicated notebook during field

visits. Scanned pages or photographs of the original notebook entries are provided as evidence of firsthand data collection.

- **Letter from the Farm Owner:**
  A signed letter from the farm owner, confirming their participation and consent for data collection, is attached**.**

All supporting materials, including **the video, images, notebook scans, and dataset,** are available in the shared Google Drive folder Proof of Originality Link (click me).

3. Metadata

- The project data was collected through structured interviews with the farm owner conducted during **February–April 2025.** All data represent consolidated expenses for the entire **2-acre farm**, as the operator does not maintain separate records for individual acres.
- Data format: Excel sheets (.xlsx) organized by crop cycle and expense category.
- Temporal Coverage :
  - Wheat Cycle (November 2023 to April 2024)
  - Paddy Cycle (june 2024 to October 2024)
  - Wheat Cycle (November 2024 to April 2025)
- Farm Size: 2 acres (total consolidated operation)
- Collection Method: 5 structured interviews (approximately 30 min each) with verification through local market price surveys

Information about the Wheat-Paddy Operations Dataset (2023-2025):

Link : Wheat-Paddy Operations Dataset (2023-2025)

About the Google Sheet Dataset

The primary dataset for this project is maintained in a Google Sheet that is organized into several distinct sheets, each serving a specific purpose:

- Wheat-2024, Wheat-2025, Paddy-2024:
  Each of these sheets contains detailed input cost data for a specific crop cycle, including categories such as irrigation, fertilizers, seeds, and labor, with itemized costs, quantities, and total expenses.
- Financial Overview Sheets:
  For each crop cycle, a separate financial overview sheet summarizes total production, quantity sold, MSP, total revenue, total cost, and total profit.
- Descriptive Statistics:
  This sheet compiles summary statistics (mean, median, mode, standard deviation, min, max, and range) for all major variables, along with category-wise cost breakdowns and calculated percentages.

Table 1 : Wheat-2024, Wheat-2025, and Paddy-2024 sheet columns

Key (Column Name)	Description	Data Type	Format/Unit
Category	The broad classification of the expense or input (e.g., Irrigation, Fertilizers, Seeds, Labour)	Text	String
Items	The specific item or activity (e.g., diesel, DAP, wheat seeds, harvesting, etc.)	Text	String
Units	Measurement units (see detailed unit analysis below in **Table 3**)	Text	String
Cost per Unit (₹)	The price for each unit of the item, in Indian Rupees	Numeric	Currency (₹), Decimal
Quantity	The number of units used or purchased	Numeric	Integer/Decimal
Total Cost (₹)	The total expenditure for the item (Cost per Unit × Quantity)	Numeric	Currency (₹), Decimal

Table 2 : Financial Overview Sheet Columns

Key (Column Name)	Description	Data Type	Format/Unit
Total Production (Quintal)	Total crop yield for the season, in quintals (1 quintal = 100 kg)	Numeric	Decimal, Quintal
Quantity Stored at Home	Amount of crop kept for home use, in quintals	Numeric	Decimal, Quintal
Total Quantity Remaining	Crop quantity available for sale after storage deduction, in quintals	Numeric	Decimal, Quintal
MSP	Minimum Support Price per quintal, as notified by the government	Numeric	Currency (₹)

Table 3 : Units Glossary

Unit	Description
Liters (L)	Volume measurement for diesel or liquid agrochemicals.
Hours	Time measurement for equipment or tube well operation.
Service	It’s used for the maintenance purpose.
50 kg	Standard fertilizer bag size for DAP/urea.
135 kg, 4 kg, 3 kg, 90kg	Weight of input used (often for micronutrients or when not a full bag). Specify product in data dictionary.
80 g	Small-quantity agrochemical or micronutrient, typically for specialized applications.
1 packet/Packet	One retail packet.
1 Bottle	One retail bottle
8 Tanks	Eight full spray tanks used for 1 acre of land
Acres	Land area measurement for operations.
Trolleys	Local unit for chaff (bhoosa): one trolley ≈ 25 quintals (2,500 kg).
Quintal	100 kilograms; standard unit for crop yield and sales.
Gram (g)	Used for very small quantities of chemicals or micronutrients.

Data Quality Note:

All data reflect consolidated expenses for the full 2-acre operation. Most values are based on farmer recall and market price verification; minor expenses may be underreported. Units like "trolley" and "tank" are based on local practice and defined in the Units Glossary above.

4. Descriptive Statistics

This section presents quantitative summaries of the farm’s key input costs, production outputs, and financial outcomes across three crop cycles (Wheat 2024, Wheat 2025, and Paddy 2024)

Summary Statistics for Input Costs

Table 4 below presents the mean, median, mode, standard deviation, minimum, and maximum for each major cost category across the three crop cycles.

Table 4: Summary Statistics for Input Costs

Category	Mean (₹)	Median (₹)	Std. Dev. (₹)	Min (₹)	Max (₹)	Range (₹)
Irrigation	10,880.93	9,766.80	3,771.51	7,792	15,084	7,292
Fertilizers	9,793.33	7,494	3,987.76	7,488	14,398	6,910
Seeds	3,353.33	3,330	40.41	3,330	3,400	70
Labor	16,906.67	15,340	2,853.23	15,180	20,200	5,020
Total Cost	40,934.27	35,930.80	10,574.56	33,790	53,082	19,292

Table 4 shows that labor is the largest and most variable cost category, with a mean of ₹16,907 and a standard deviation of ₹2,853. Irrigation and fertilizer costs are also substantial and show considerable variability across seasons. Seed costs are the lowest and most stable, with minimal variation. The total cost per cycle ranges from ₹33,790 to ₹53,082, highlighting significant fluctuations in overall farm expenditure.

Table 5 : Summary Statistics for Production and Revenue

Variable	Mean	Median	Std. Dev.	**Min**	Max	Range
Yield (Quintals)	41.33	32.00	19.73	28.00	64.00	36.00

MSP (₹/Quintal)	2,194.33	2,183.00	81.77	2,125	2,275	150
Total Revenue (₹)	93,824.00	72,800.00	40,756.96	67,872	140,800	72,928
Total Profit (₹)	52,889.73	36,869.20	30,368.54	34,082	87,718	53,636

Table 5 shows that Paddy 2024 achieved the highest yield and revenue, resulting in the greatest profit among all cycles. Wheat cycles had lower yields and profits, with Wheat 2024 being the least profitable. The standard deviations for yield, revenue, and profit indicate substantial variability across crop cycles, reflecting the influence of both market prices and input costs on farm income.

Table 6 : Cost-to-Output Ratios and Profit Margins

Crop/Year	Cost per Quintal (₹)	Revenue per Quintal (₹)	Profit Margin(%)
Wheat 2024	1,206	2,121	50.2
Wheat 2025	1,198	2,275	50.6
Paddy 2024	830	2,200	62.3

Table 6 shows that Paddy 2024's profit margin (62.3%) significantly outperforms wheat cycles (~50%), suggesting better financial resilience despite higher input costs.

To understand what drives these costs, the following pie charts illustrate the cost-structure breakdown for each crop cycle, showing the share of irrigation, fertilizers, seeds, and labor in total expenses.

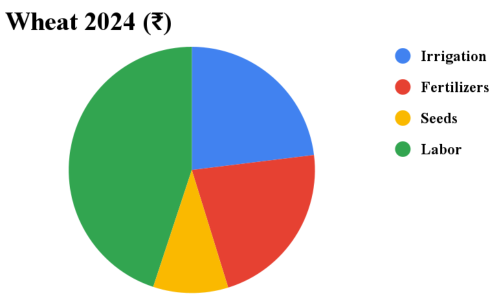

Figure 1: Cost Structure for Wheat 2024

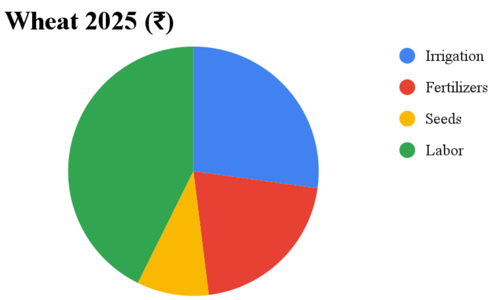

Figure 2: Cost Structure for Wheat 2025

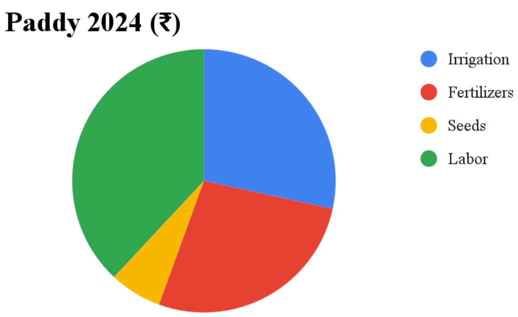

Figure 3: Cost Structure for Paddy 2024

As shown in Figures 1–3, labor consistently makes up nearly half of total costs across Wheat 2024, Wheat 2025, and Paddy 2024, with irrigation and fertilizers as secondary expenses and seeds contributing the least.

These cost patterns suggest that initiatives targeting labor efficiency and irrigation optimization would have the greatest impact on overall profitability

5. Analysis Process and Methods

The data preprocessing and cleaning began by transferring manually recorded data into Google Sheets, where a systematic cleaning process was applied. This involved

- Checking for Errors: All entries were reviewed for typographical mistakes, unrealistic values, or missing information.
- Standardizing Units: Measurements were standardized (e.g., all fertilizer quantities in 50 kg bags, all costs in Indian rupees) to maintain uniformity across the dataset.
- Handling Missing Data: Any missing or unclear values were clarified through follow-up discussions with the farm owner. If a value could not be confirmed, it was marked as "not available" and excluded from calculations.
- Removing Duplicates: Duplicate or repeated entries were identified and removed to prevent double-counting.

After cleaning the data, I calculated standard descriptive statistics (mean, median, mode, standard deviation, and range) using built-in Google Sheets functions.

Cost-to-Output Ratios

To evaluate economic efficiency across crop cycles, I computed two key performance ratios:

- Cost per Quintal = Total Cost ÷ Yield
- Profit Margin (%) = (Total Profit ÷ Total Revenue) × 100

These ratios directly address the gap between MSP and rising input costs by quantifying profitability differences between wheat and paddy cultivation.

These statistical calculations provided the quantitative foundation for the analysis. To translate these numbers into more intuitive insights, visualization methods were employed.

I generated pie charts in Google Sheets to represent the cost structure for each crop cycle. This made it easier to identify which categories (labor, irrigation, fertilizers, and seeds) contributed most to total costs.

While visualizations made the data accessible, the choice of descriptive statistics as the primary analytical approach requires explanation. It is the most suitable method for this stage because they provide a clear, quantitative summary of the main features of the dataset, especially when working with a small sample size and primary data. This approach allows for easy comparison across crop cycles and cost categories and helps identify patterns or outliers without overcomplicating the analysis.

This analytical approach was intentionally chosen to address the farm's specific operational challenges. By summarizing averages, variability, and cost breakdowns, this step directly addresses the project's problem statements-such as identifying which costs are the largest or most unpredictable and comparing profitability across crops. The insights gained here form the foundation for deeper analysis and future recommendations.

The methodological choices behind this analysis deserve further explanation. Descriptive statistical analysis was chosen because it is the most effective and transparent method for summarizing and understanding small, primary datasets like those collected from Sham's farm.

This approach allows for clear identification of central tendencies (such as average costs and profits), variability (such as which costs fluctuate most), and cost structure (such as which input dominates expenses).

Unlike more complex statistical models (e.g., regression analysis or machine learning), descriptive statistics do not require large sample sizes or advanced technical resources, making them ideal for farm-level decision-making and reporting.

Beyond basic statistics, additional analytical techniques enhanced the practical value of the findings. The use of ratios and visualizations further enhances interpretability, allowing both the researcher and the farmer to quickly grasp which areas require intervention.

By directly linking these descriptive findings to the problem statements (high irrigation and labor costs, rising fertilizer expenses, and squeezed profit margins), this method ensures that the analysis is relevant, actionable, and tailored to the real challenges faced by the farm.

6. Results and Findings

This section presents the main results from the analysis of the farm data

Input Cost Patterns

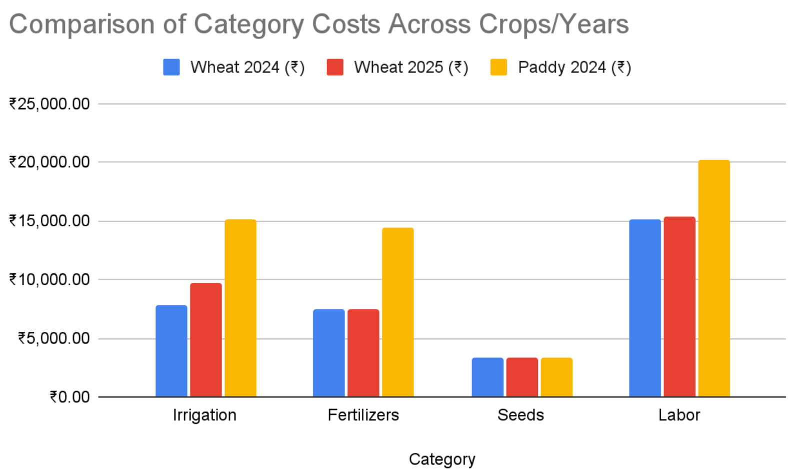

Figure 4. Comparison of Category Costs Across Crops/Years

Figure 4 compares the major input costs-irrigation, fertilizers, seeds, and labor-for Wheat 2024, Wheat 2025, and Paddy 2024. It shows that

- Irrigation and fertilizer costs are much higher for paddy in 2024 than for wheat, reflecting the water and nutrient intensity of paddy cultivation. Specifically, paddy requires approximately 40% more irrigation (₹15,000 vs. ₹10,000) due to its submerged growing conditions and greater nutrient demands.
- Labor costs are the largest single expense in all crop cycles, with the highest value for Paddy 2024 (approximately ₹20,000). This directly supports Sham's concern about labor availability and rising wages, which now constitute nearly 50% of total production costs.
- Seed costs are relatively stable and much lower compared to other categories, averaging only ₹3,000 per crop cycle and showing minimal variation.
- These findings quantitatively confirm that irrigation, fertilizers, and labor are the main drivers of total cost, especially for paddy, directly supporting the first three problem statements. Labor alone accounts for nearly half of all expenses, making it the most critical cost component to optimize.

Revenue and Profit Trends

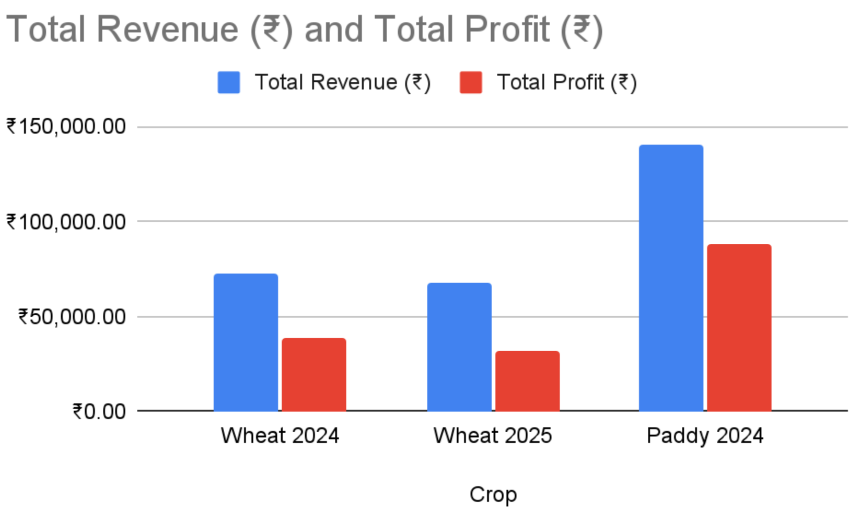

Figure 5. Total Revenue and Profit by Crop Cycle

Figure 5 illustrates the total revenue and total profit for each crop cycle. The chart shows that

- Paddy 2024 generated the highest revenue (approximately ₹145,000) and profit (about ₹90,000) among all cycles, despite also having the highest input costs. This represents a nearly 50% increase in revenue compared to either wheat cycle.
- Both wheat cycles had similar revenues (approximately ₹75,000) and profits (about ₹38,000), which were noticeably lower than those of paddy. This consistency across wheat cycles suggests stable yet modest returns.
- The profit margin for paddy is significantly higher (62.3%), indicating that, although it requires greater investment, paddy cultivation can yield better financial returns when managed efficiently. This directly addresses Sham's concern about maintaining profitability despite rising input costs.

This pattern highlights the importance of crop selection and input management. While paddy incurs higher costs, it also offers the potential for greater profitability, provided input costs are kept in check and yields remain strong.

Trends in Cost, Revenue, and Profit Over Crop Cycles

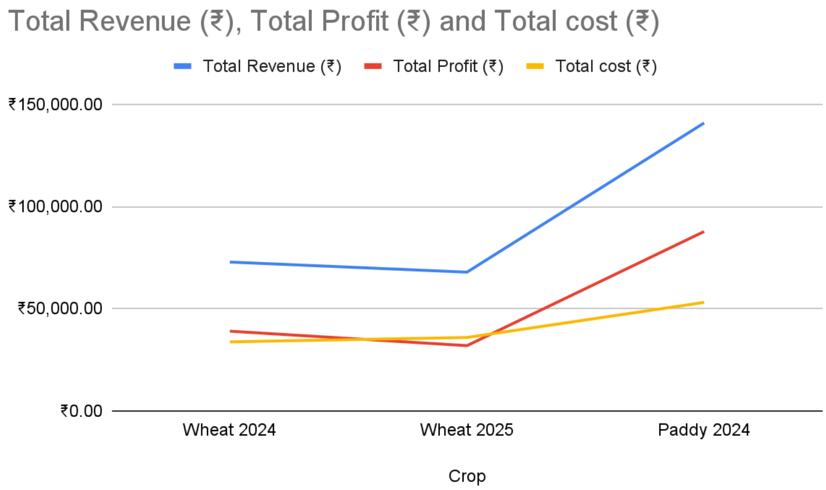

Figure 6. Trends in Total Cost, Revenue, and Profit Across Crop Cycles

Figure 6 shows the changes in total cost, total revenue, and total profit across the three crop cycles. The chart reveals that

- All three metrics (cost, revenue, and profit) increase sharply for Paddy 2024 compared to both wheat cycles, with revenue rising more steeply than costs (by approximately ₹70,000 vs. ₹20,000).
- The gap between revenue and cost is widest for Paddy 2024, indicating improved profitability despite higher expenses. This gap is nearly twice as large for paddy (₹90,000) as for either wheat cycle (₹38,000).
- Wheat cycles show relatively stable costs and profits, with only a slight increase from 2024 to 2025. This stability contrasts with the rising input costs Sham reports, suggesting that wheat profit margins are gradually being squeezed.

This trend demonstrates that although paddy cultivation is more expensive (requiring approximately ₹52,000 in total inputs compared to ₹36,000 for wheat), it can also be more profitable when input costs are managed and yields are high. It also confirms that rising input costs are a key concern, as they have a direct impact on profitability.

Collectively, these findings directly address the farm's core challenges: the dominance of labor costs (47-50% of total expenses) validates his labor shortage concerns; the significant irrigation expenses (₹15,000 for paddy) confirm his issues with diesel pump costs; and paddy's superior profit margin (62.3% versus wheat's 50.6%) provides a potential solution to his concern about the gap between MSP and rising input costs.
The results suggest that optimizing labor efficiency and irrigation costs-particularly for paddy cultivation-would have the greatest impact on overall farm profitability. While paddy requires a higher initial investment, its superior yield and profit margins offer a more sustainable path forward if the primary input constraints can be effectively managed.