Exploratory Data Analysis of Indian Agricultural Costs and Yields

This report presents a comprehensive Exploratory Data Analysis (EDA) of the cost of cultivation in India, based on the provided dataset. The following sections visualize and explain key findings related to costs, yields, profitability, and regional variations.

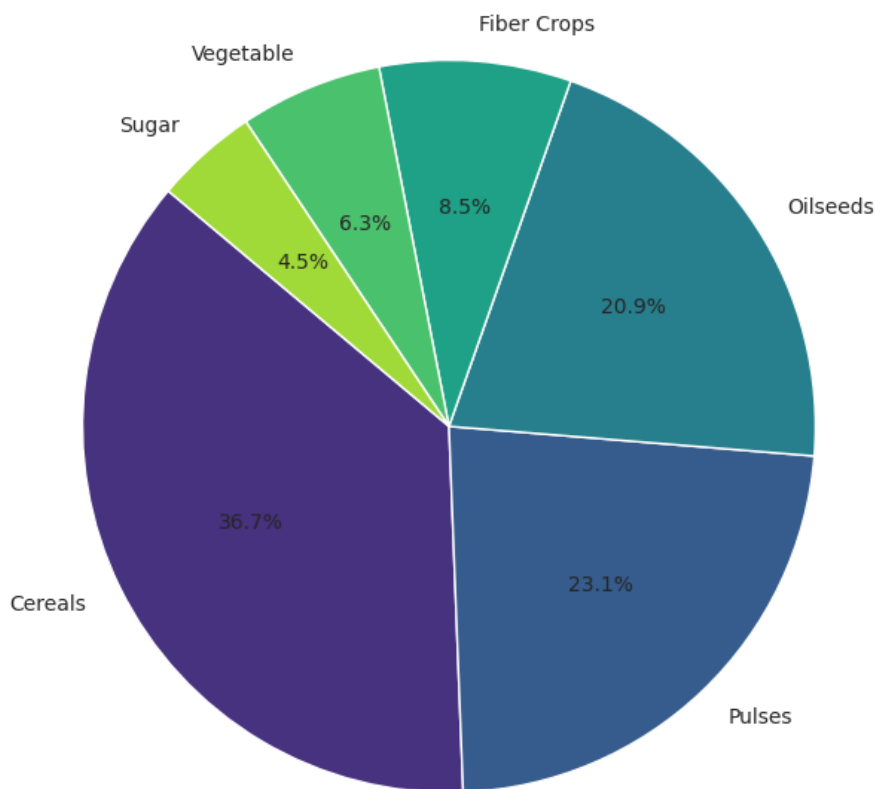## Glossary: Understanding Agricultural Cost Concepts

The dataset uses standard cost concepts defined by India's Commission for Agricultural Costs and Prices (CACP). Understanding these is key to the analysis. The primary metric used in this report is Cost C2, as it is the most comprehensive measure of a farmer's expenses.

- **Cost A1**: All actual paid-out expenses in cash and kind. This includes hired labor, seeds, fertilizers, irrigation, etc.

- **Cost A2**: Cost A1 + rent paid for any land leased by the farmer.

- **Cost A2+FL**: Cost A2 + the imputed value of unpaid labor contributed by the farmer and their family.

- **Cost B1**: Cost A1 + interest on the value of owned fixed capital assets (e.g., machinery), excluding land.

- **Cost B2**: Cost B1 + the rental value of the farmer's own land and rent paid for leased land.

- **Cost C1**: Cost B1 + imputed value of family labor.

- **Cost C2 (Comprehensive Cost)**: Cost B2 + imputed value of family labor. This is the most comprehensive cost, accounting for all paid and imputed expenses. It is the 'Total Cultivation Cost' used throughout this report.

- **Cost C3**: Cost C2 + 10% of Cost C2 to account for the farmer's managerial functions.
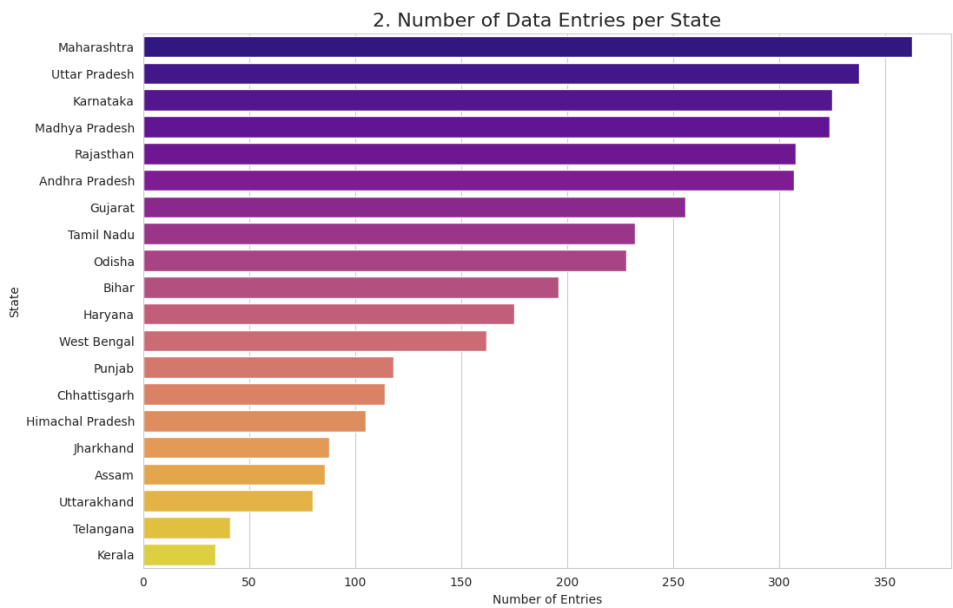
Distribution by Crop Type

This pie chart shows the proportion of data entries for each major crop category. Cereals constitute the largest portion, highlighting their significance in Indian agriculture as recorded in this dataset.
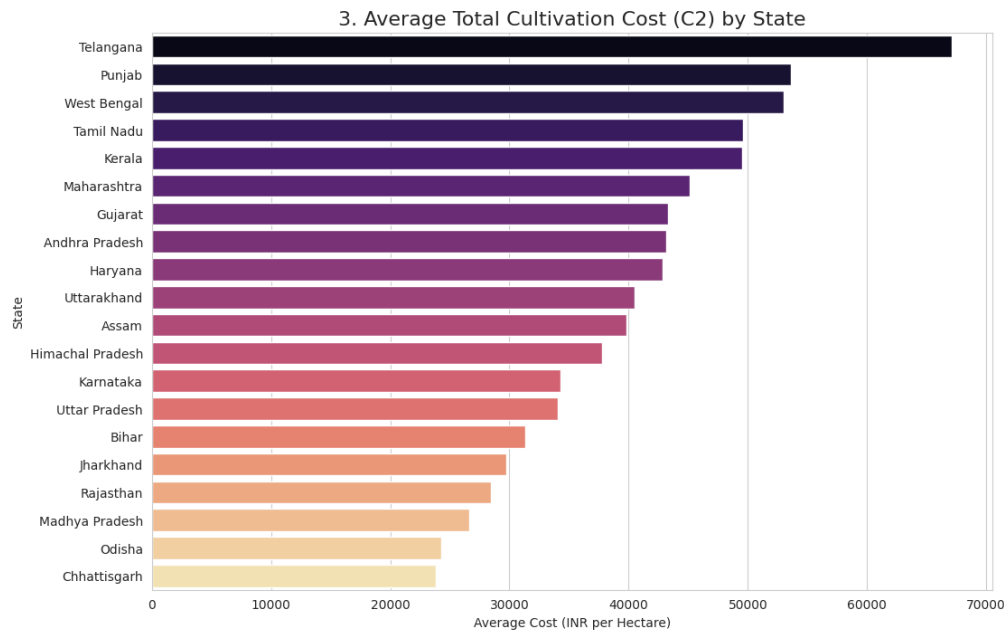
1. Distribution of Data by Crop Type

Data Entries per State

This bar chart displays the geographic distribution of the dataset, showing the number of agricultural data entries from each Indian state. States like Uttar Pradesh and Madhya Pradesh have the highest representation.
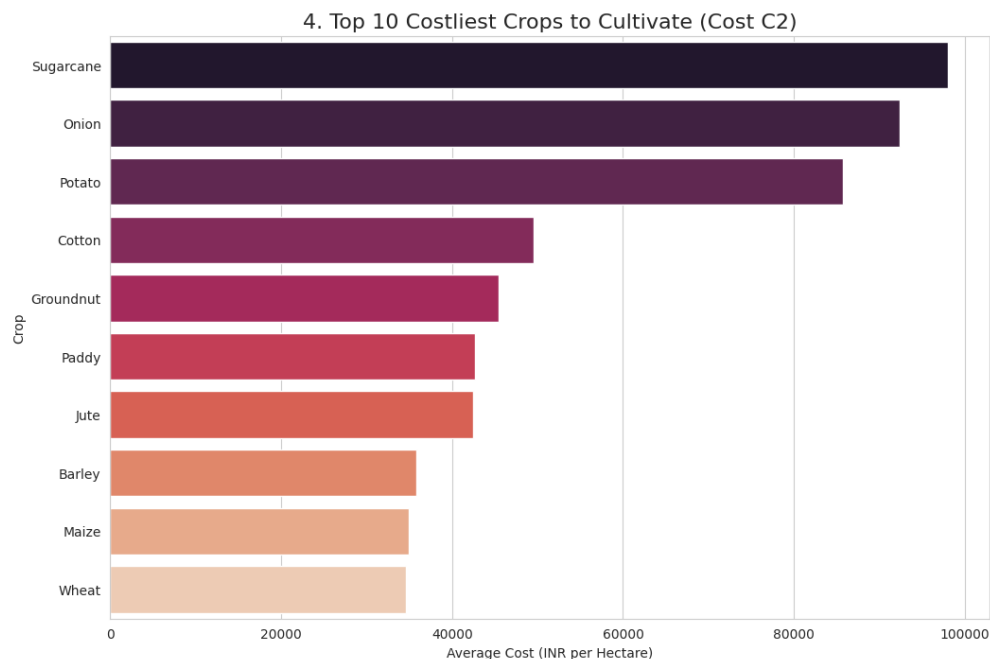
Average Cultivation Cost by State

This chart ranks states by their average total cost of cultivation (Cost C2). It reveals significant regional disparities in farming expenses, with states like Andhra Pradesh showing the highest costs.
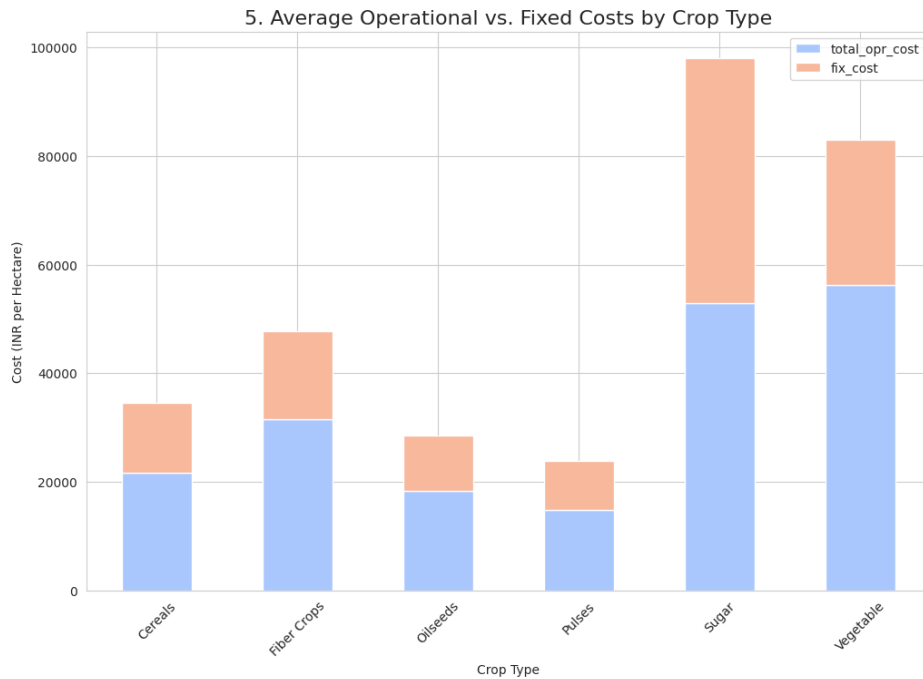

3. Average Total Cultivation Cost (C2) by State

Top 10 Costliest Crops

This bar chart identifies the ten crops with the highest average cultivation cost. High-value crops like Sugarcane and Potato dominate this list, indicating they are capital-intensive to grow.

4. Top 10 Costliest Crops to Cultivate (Cost C2)
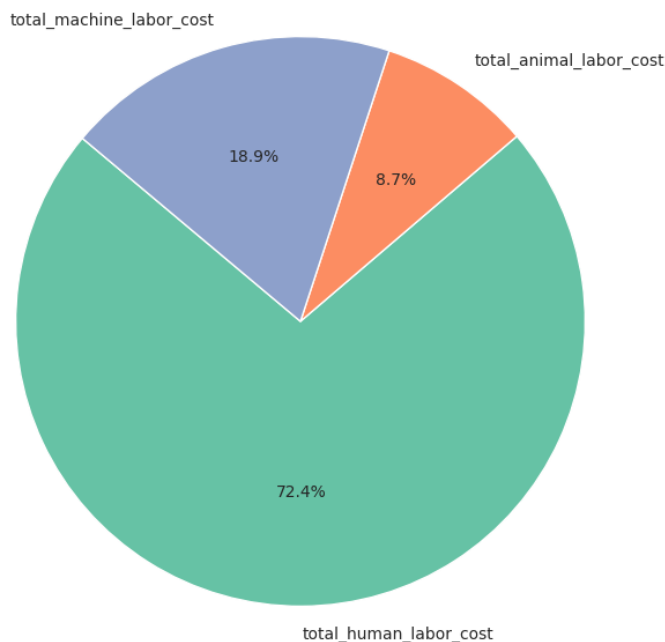
Operational vs. Fixed Costs

This stacked bar chart breaks down the total cultivation cost into operational (variable) and fixed components for each crop type. It clearly shows that operational costs form the bulk of expenses across all categories.

5. Average Operational vs. Fixed Costs by Crop Type
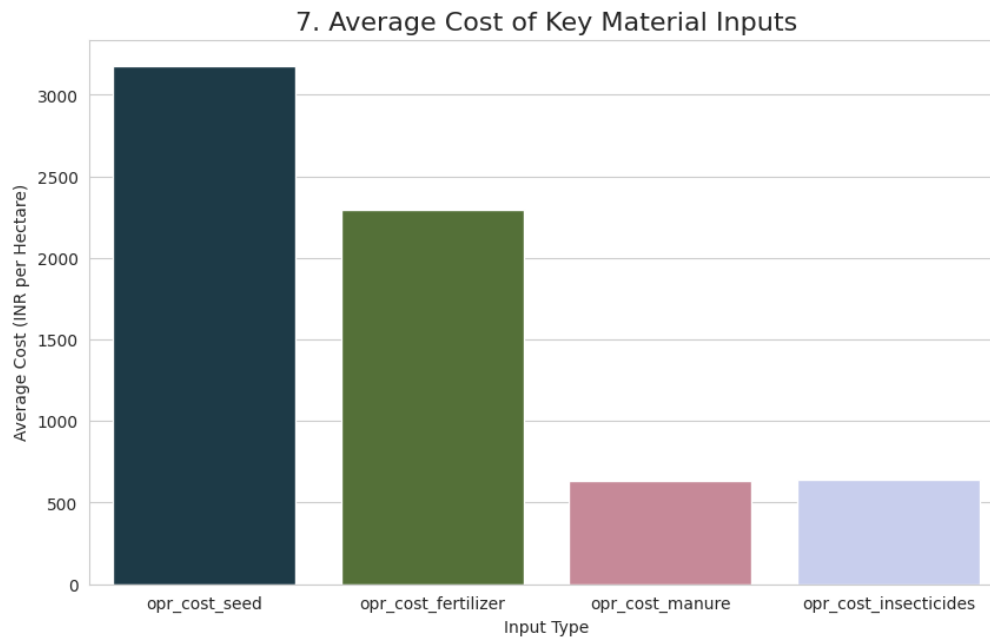
Labor & Machinery Cost Breakdown

This pie chart illustrates the average share of human, animal, and machine labor in the total labor and machinery expenditure. Human labor is the most significant component, followed closely by machinery, signifying a shift towards mechanization.
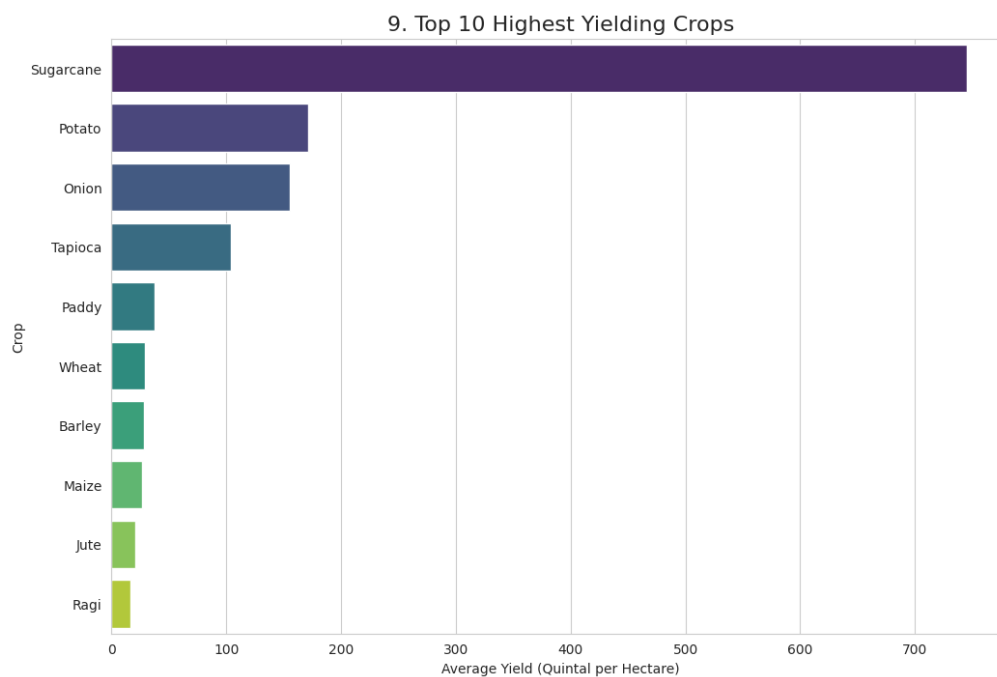

6. Average Composition of Labor & Machinery Costs

Key Input Costs

This bar chart compares the average per-hectare expenditure on four critical inputs: seeds, fertilizer, manure, and insecticides. Fertilizer and seeds typically represent the highest material input costs for farmers.
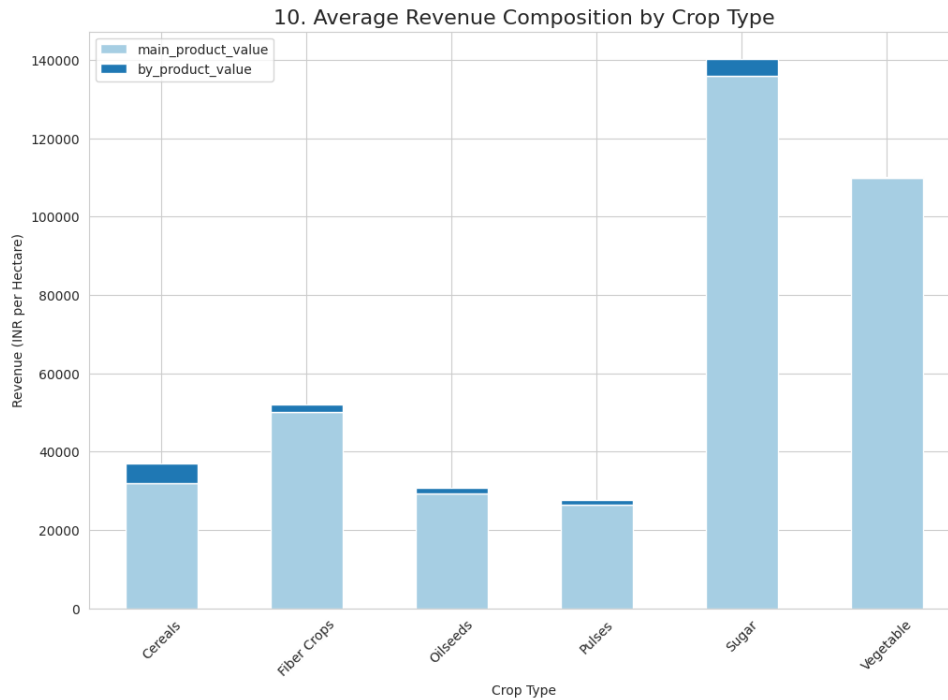

7. Average Cost of Key Material Inputs

Top 10 Highest Yielding Crops

This chart ranks the top ten crops by their average yield. Unsurprisingly, crops like Sugarcane and Potato, which are bulky, are at the top of this list.

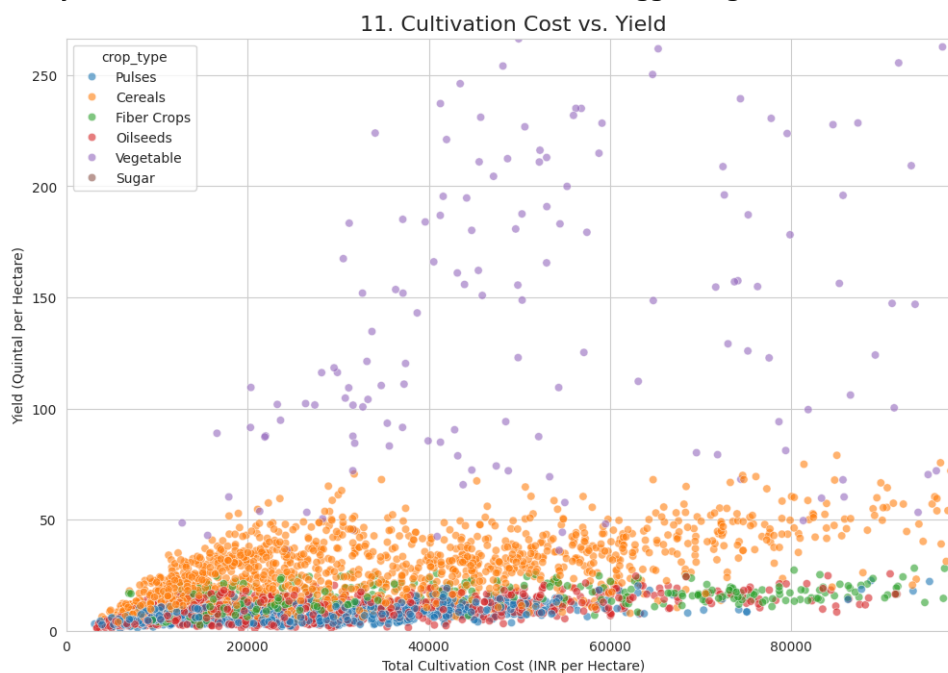9. Top 10 Highest Yielding Crops

Revenue Composition

This stacked bar chart illustrates the average revenue generated from the main product versus the by-product for each crop type. For most crops, the main product is the primary revenue source, but for some, like certain cereals, the by-product (fodder) is also significant.
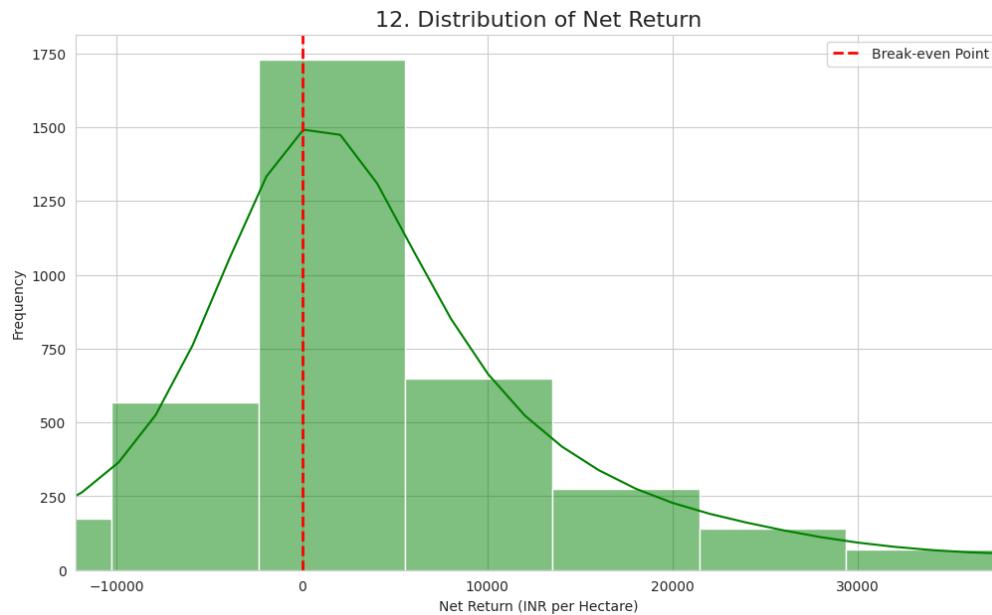

Cost vs. Yield

This scatter plot explores the relationship between total cultivation cost and yield, with points colored by crop type. Generally, higher investment leads to higher yield, but the relationship is not always linear and shows considerable variation, suggesting differences in efficiency.
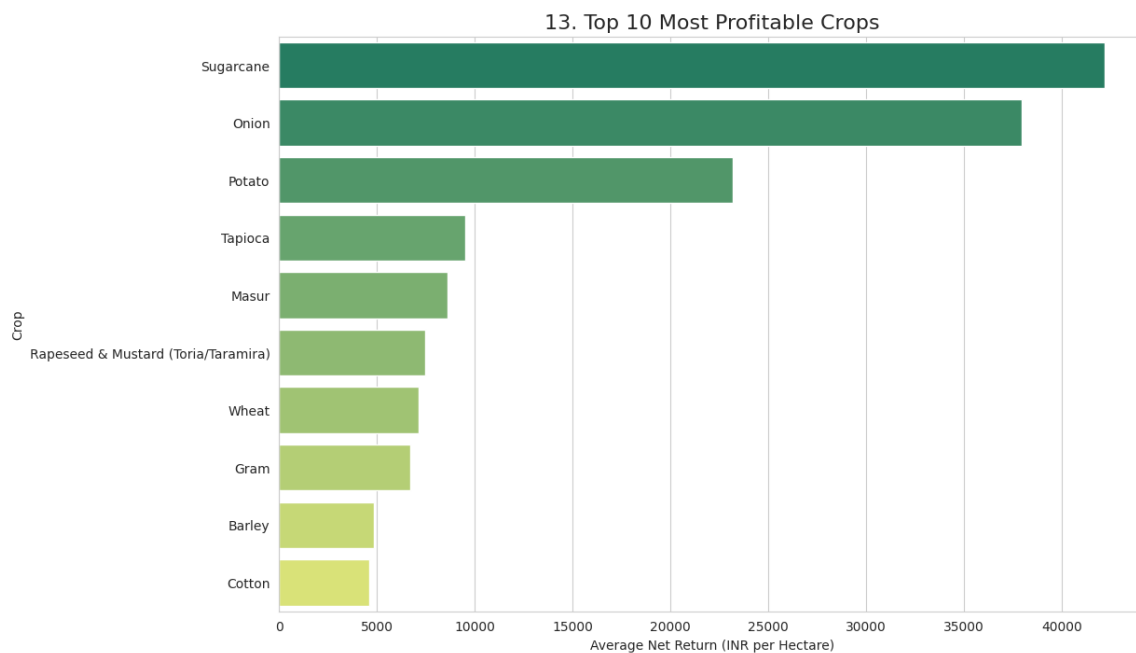
Net Return Distribution

This histogram shows the distribution of net returns (profit/loss). The red dashed line at zero represents the break-even point. A significant portion of the distribution falls to the left of this line, indicating that a substantial number of farming operations recorded in the dataset resulted in a financial loss.
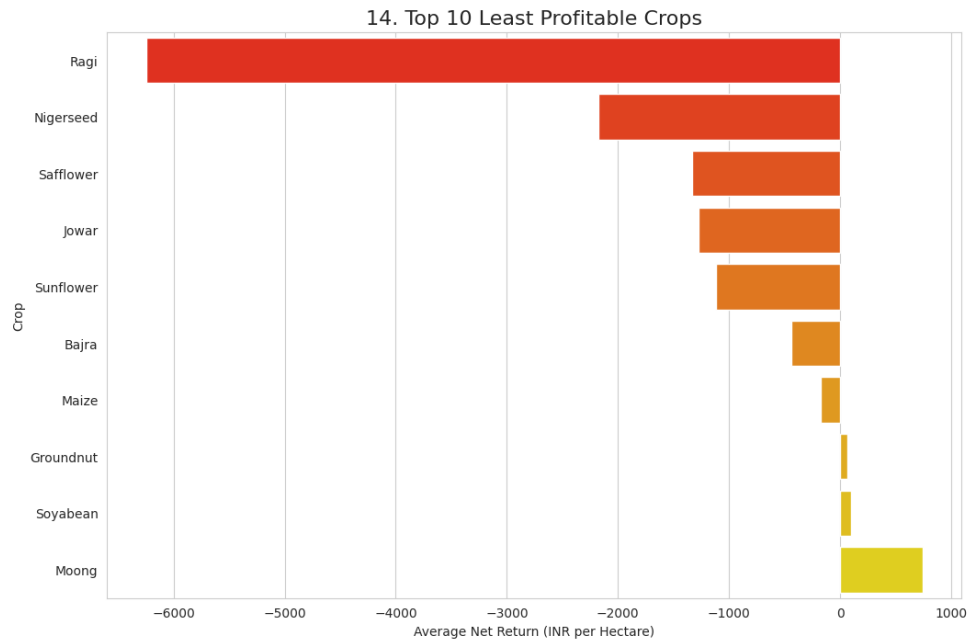

Top 10 Most Profitable Crops

This chart highlights the ten crops with the highest average net return. Commercial crops like Sugarcane and Potato are among the most profitable, assuming favorable conditions.
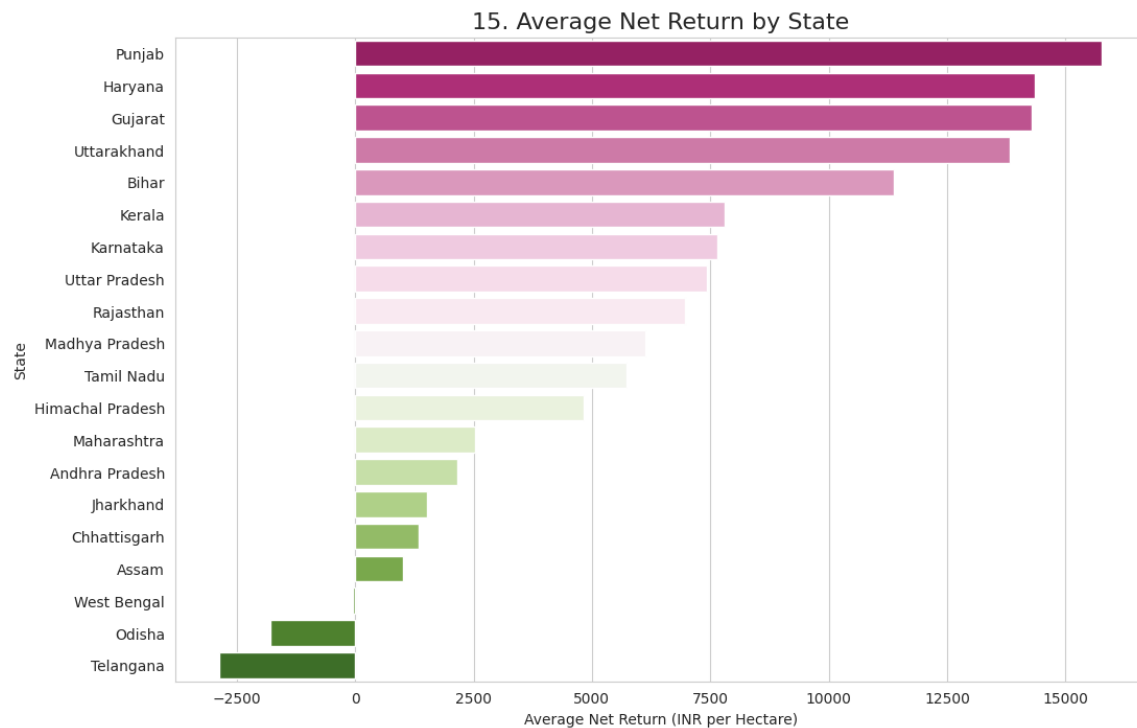
Top 10 Least Profitable Crops

This bar chart shows the ten crops that, on average, resulted in the largest financial losses. It underscores the high financial risk associated with certain crops, where costs frequently exceed revenues.

14. Top 10 Least Profitable Crops
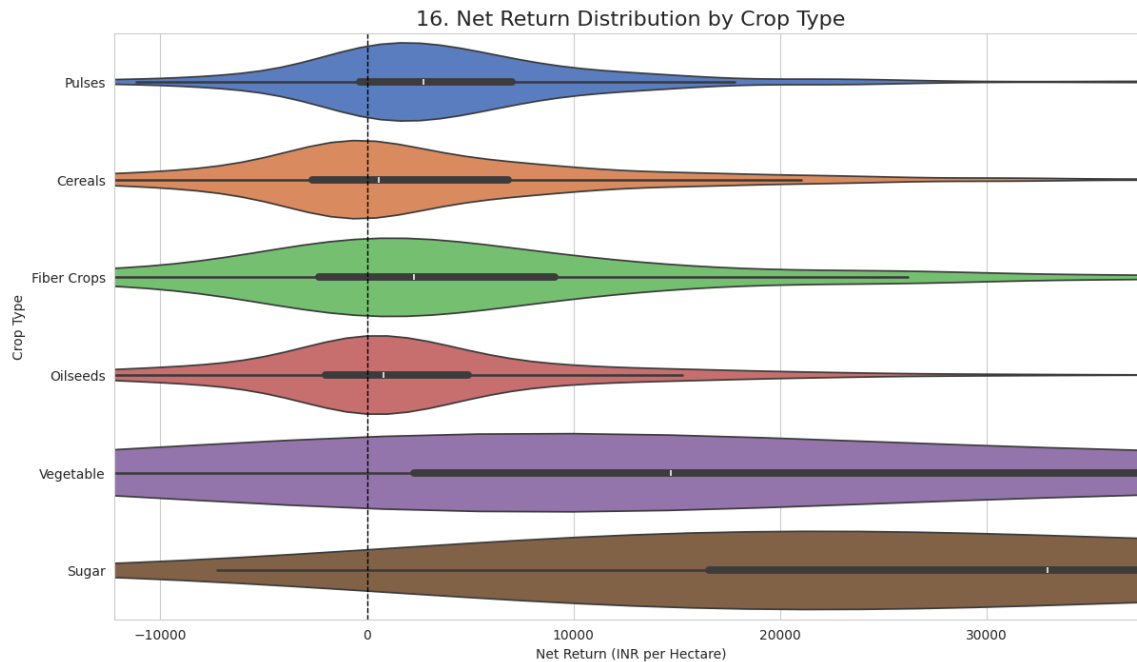
Average Net Return by State

This chart ranks states by their average net return from agriculture. It reveals which states have, on average, more profitable farming sectors. States like Haryana and Punjab often show higher profitability, while others face systemic challenges.
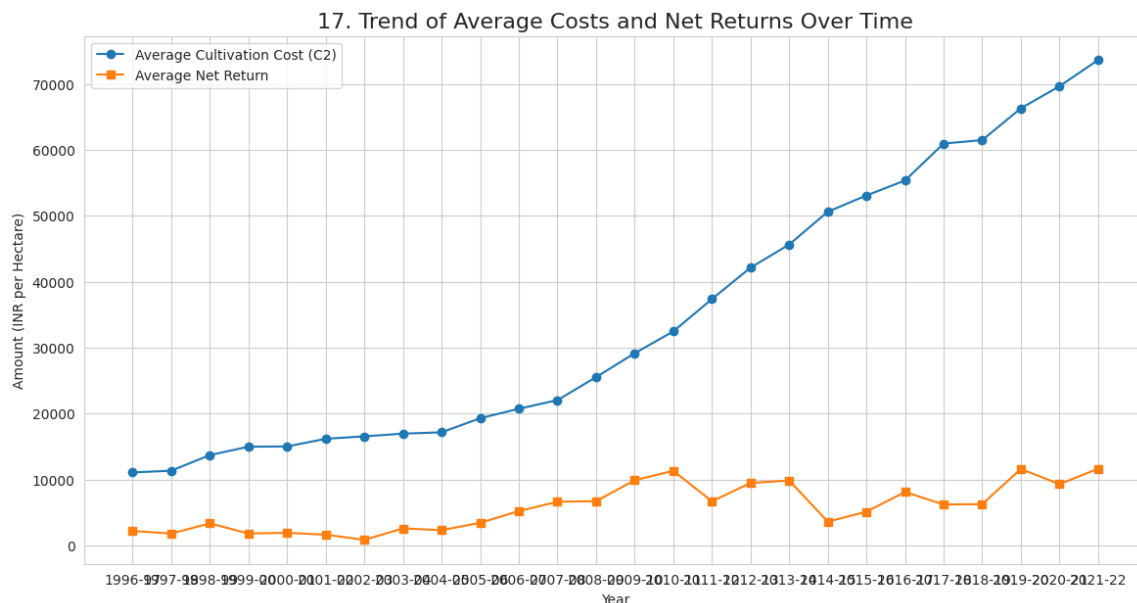

15. Average Net Return by State

Net Return Distribution by Crop Type

This violin plot provides a deeper look at profitability by showing the full distribution of net returns for each crop type. The width of the violin indicates the frequency of data points at that profit level. It effectively visualizes both the profitability and the financial risk (spread) associated with different crop categories.

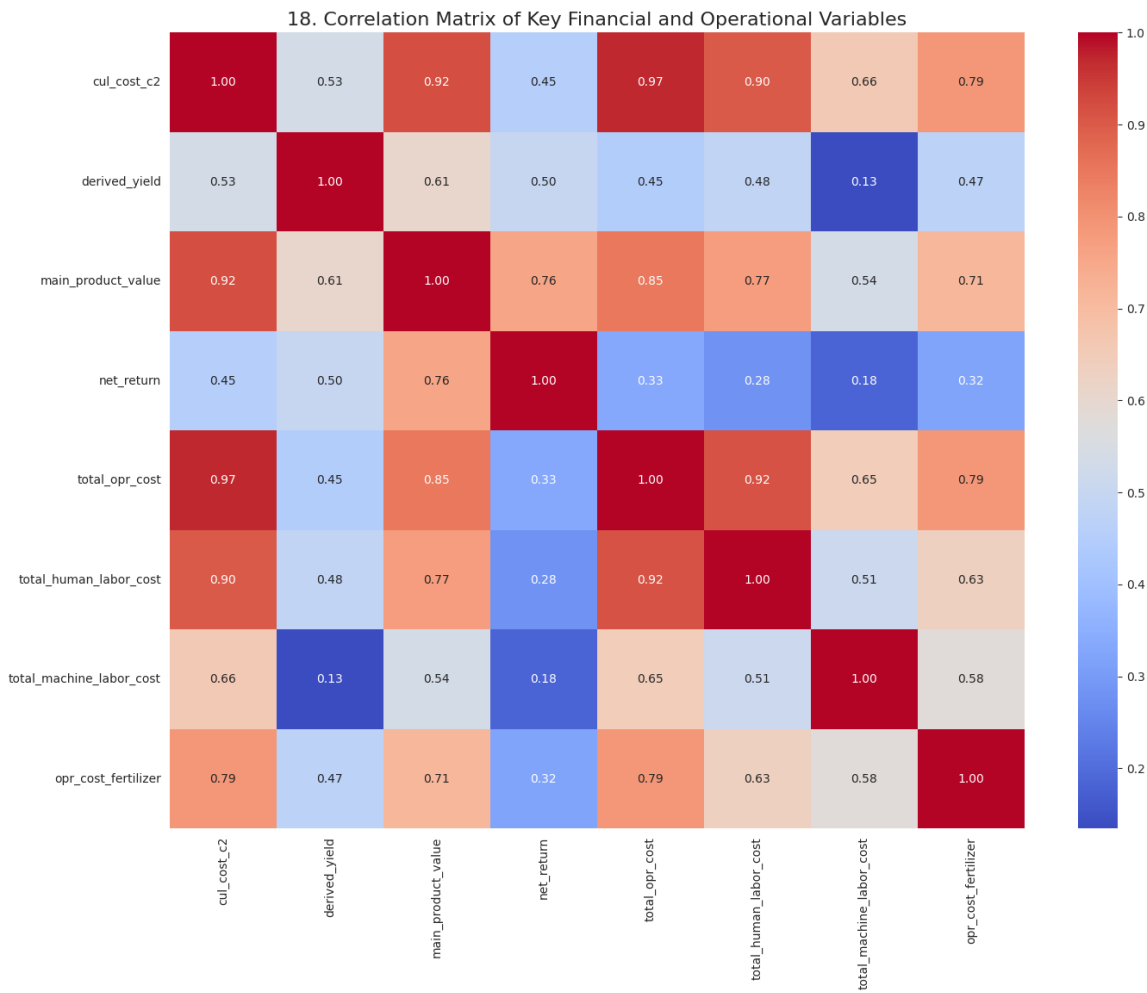16. Net Return Distribution by Crop Type

Trends of Costs and Returns

This line chart tracks the average total cultivation cost and average net return from 1999 to 2022. It illustrates the 'cost-price squeeze,' showing that while costs have steadily risen, net returns have been much more volatile and have not kept pace.

17. Trend of Average Costs and Net Returns Over Time
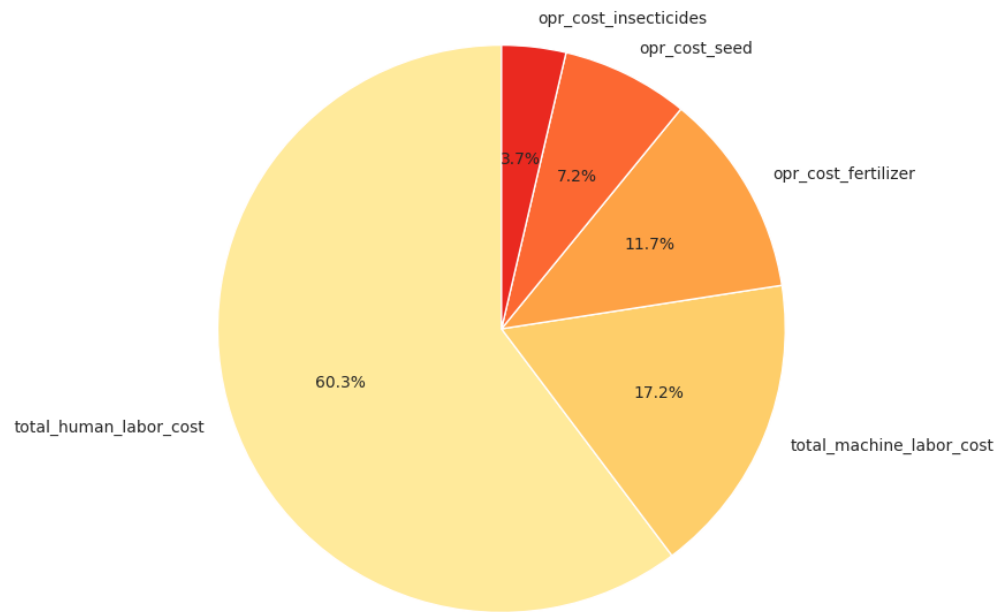
Correlation Heatmap

This heatmap shows the correlation coefficients between key variables. Red indicates a positive correlation, while blue indicates a negative one. It helps to statistically identify relationships, such as the strong positive link between operational costs and total cultivation costs, and the relationship between yield and revenue.

18. Correlation Matrix of Key Financial and Operational Variables

Cost Structure for Paddy

This pie chart breaks down the major operational costs for Paddy. It highlights the primary drivers of expense for cultivating this specific crop.
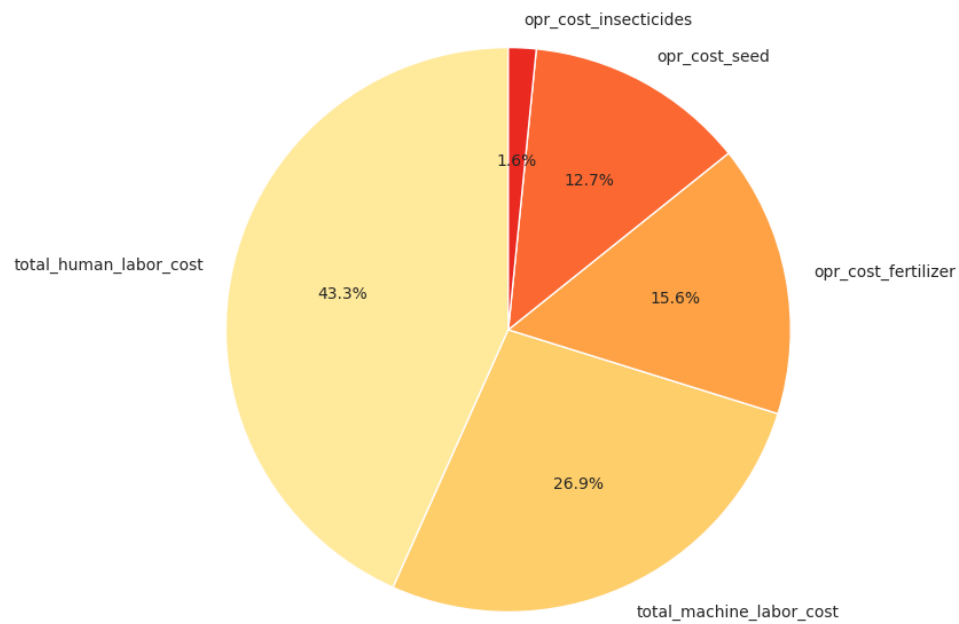
19. Operational Cost Structure for Paddy

Cost Structure for Wheat

This pie chart breaks down the major operational costs for Wheat. It highlights the primary drivers of expense for cultivating this specific crop.
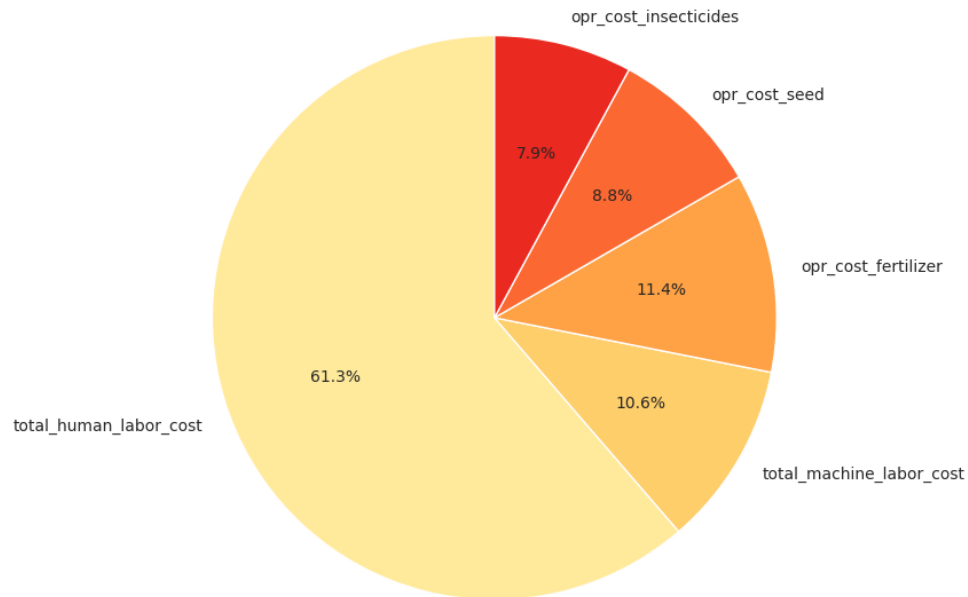
20. Operational Cost Structure for Wheat

Cost Structure for Cotton

This pie chart breaks down the major operational costs for Cotton. It highlights the primary drivers of expense for cultivating this specific crop.
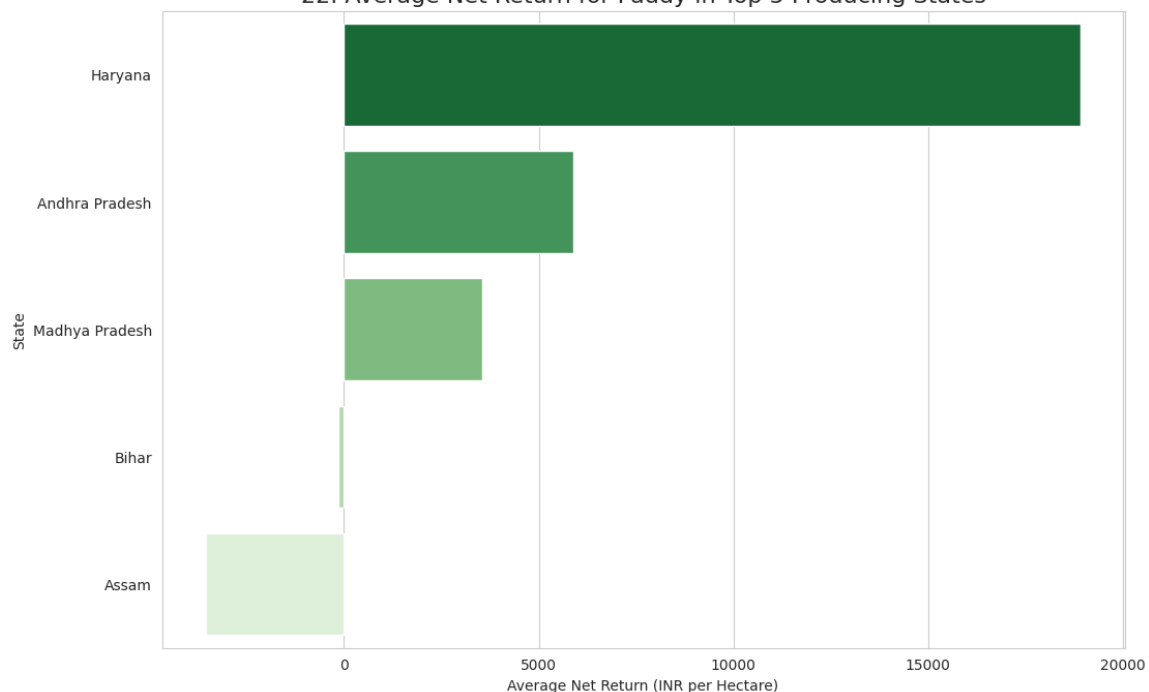

21. Operational Cost Structure for Cotton

Paddy Profitability by State

This chart compares the average profitability of Paddy cultivation across the top 5 producing states (by data volume). It shows that even for the same crop, profitability varies dramatically by region.

22. Average Net Return for Paddy in Top 5 Producing States

Wheat: Fertilizer vs. Yield

This scatter plot specifically examines the relationship between money spent on fertilizer and the resulting yield for Wheat. It helps analyze the efficiency of input use for a key crop.

23. Fertilizer Cost vs. Yield for Wheat