GreenFuture BioChem: Analysis Report

Data Analytics Senior Seminar: Case Study 1
Sam Doherty and Shoumika Anannyo

1. Executive Summary

GreenFuture BioChem is a global manufacturer of bio-based specialty chemicals with a core mission to replace petrochemical inputs with sustainable alternatives. The company operates five manufacturing plants across the United States, Germany, China, Brazil, and India. It manages a complex portfolio of **750 SKUs** spanning key industries including **Automotive Solutions, Consumer and Home Care, Industrial Lubricants, Packaging Materials, and Specialty Polymers.** Its customers are global industrial and consumer companies seeking to advance their sustainability goals, positioning GreenFuture as a key enabler in the transition toward a circular bio-economy.

However, this strategic position is undermined by significant operational inefficiencies, as GreenFuture faces **5.1% manufacturing cost overruns** totaling **$372,000+** annually due to flawed cost models.

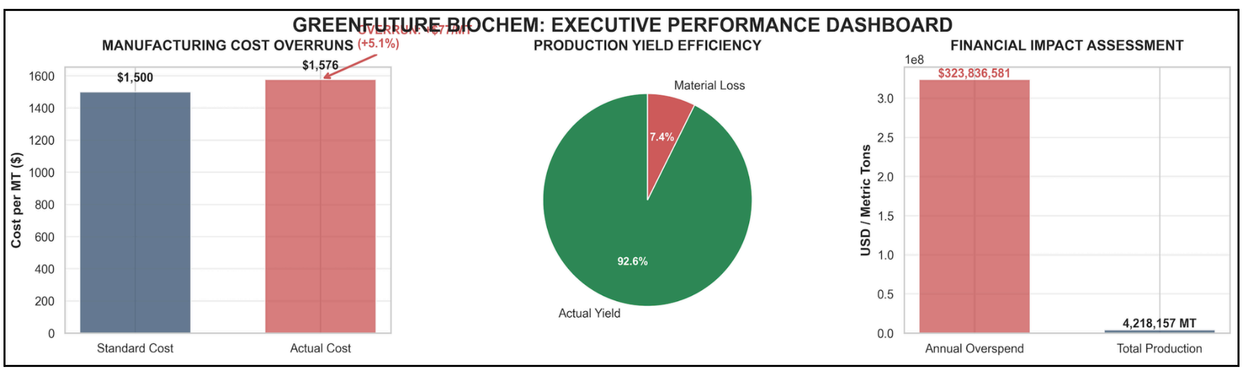


Fig 1: This performance dashboard quantifies the central challenge, visualizing the 5.1% manufacturing cost overrun and its significant financial impact.

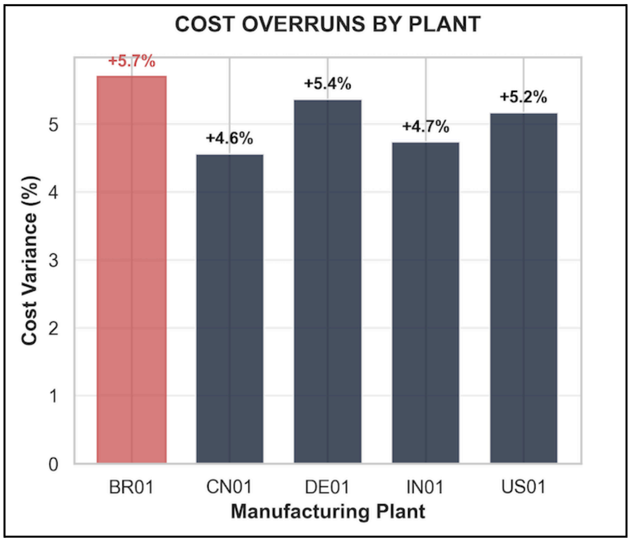
1.1 Key Findings

Our analysis identified three core issues driving the company's financial performance gap. Detailed calculations and supporting data are provided in subsequent sections.

1.1.1 Systematically Flawed Cost Models: Standard cost models are fundamentally inaccurate, consistently underestimating actual costs. A strong correlation (r=0.91) between standard and actual costs confirms the models themselves are defective, causing predictable overruns across all plants.

1.1.2 Operational Complexity and Inefficiency: The extensive 750-SKU portfolio creates manufacturing inefficiencies, while a 92.6% production yield falls below the 95% industry benchmark, resulting in significant material waste.

1.1.3 Critical Data Disconnect: The inability to map raw materials to finished goods prevents accurate product-level margin analysis and carbon emission allocation, limiting strategic decision-making for profitability and sustainability.



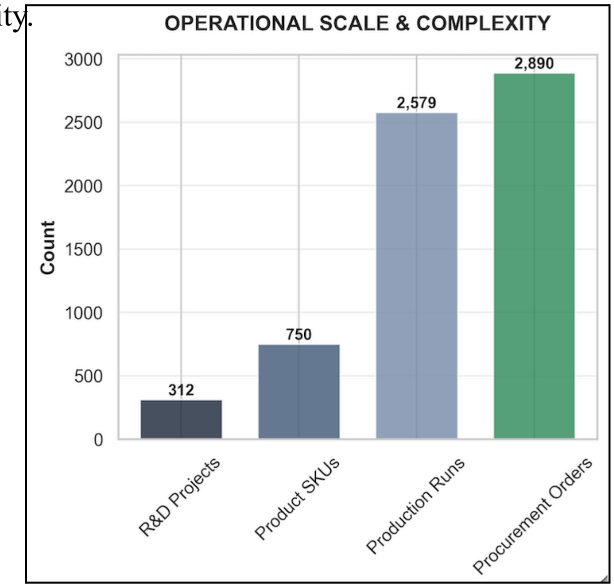


Fig 2&3: This analysis reveals systemic cost overruns across all manufacturing plants and highlights the operational complexity driven by an extensive 750-SKU portfolio.

1.2 Critical Recommendations

We propose three prioritized initiatives for sequential implementation:

1.2.1 Overhaul Standard Cost Models: Revise costing methodology by integrating real-time procurement data and establishing plant-specific benchmarks.

1.2.2 Rationalize Product Portfolio: Execute systematic SKU rationalization to reduce complexity and unlock manufacturing efficiencies.

1.2.3 Launch Targeted Yield Improvement: Implement focused operational excellence programs to close the yield gap and reduce material waste.

1.3 Strategic Impact

Implementation is projected to generate approximately $596.5 million in annual savings, significantly enhancing profitability. This financial improvement will enable reinvestment in GreenFuture's sustainability mission while strengthening its competitive position through improved cost leadership and environmental stewardship.

2. Project Objectives and Scope

2.1. Business Context and Analytical Imperative

This report analyzes GreenFuture BioChem's business operations to enhance productivity and profitability while maintaining its core sustainability commitments. It was prepared for the company's executive leadership team to support strategic decision-making regarding operational improvements, cost optimization, and sustainability initiatives.

2.2. Core Project Objectives

The primary objectives of this analysis are:

- To identify operational inefficiencies across manufacturing, supply chain, R&D, and sales functions.
- To provide data-driven recommendations for improvement, including those related to data collection and the profitability analysis of suppliers, regions, and categories.
- To deliver actionable insights that balance financial performance with environmental goals.

2.3. Defined Scope of Analysis

The analysis examines GreenFuture's global operations, specifically:

- All 5 global manufacturing plants.
- The complete portfolio of 750 SKUs.
- Cross-functional operations using integrated datasets from R&D, manufacturing, sales, and procurement systems.

2.4. Acknowledged Constraints and Limitations

The analysis was constrained by significant data integration challenges that limited the scope of certain conclusions:

2.4.1 Primary Constraint: An inability to merge key datasets due to **no direct mapping between** raw materials (Supply Chain Procurement) and finished goods. This prevented:

- Specific integrated conclusions regarding margins or the correlation between procurement, manufacturing, and sales.
- The allocation of CO_2 emissions by SKU, brand, or category.
- The ability to compare emissions against the cost of manufacturing or expected revenue.

2.4.2 Secondary Data Quality Issues: The presence of **7 columns with less than 80% data completeness** further complicated a comprehensive sustainability-profitability analysis.

3. Data and Methodology

Five datasets were used in this analysis:

- RD Pipeline (CRM) (312 rows)
  - Includes various Research and Development projects (along with a key column Project ID), their current stage of development, expected revenue, launch date, manager, and target industry. It also includes a column for matching Sales Opportunities, which links this dataset to Sales Pipeline.

- Sales Pipeline (CRM) (3056 rows)
  - This dataset describes various Sales Opportunities, along with their current status or outcome. It includes the customer, the region, product (or brand), the deal's estimated value, and the probability of making the deal. It includes a key column that identifies each deal uniquely and also a column for each opportunity's matching R&D project.
- Manufacturing Production (ERP) (2579 rows)
  - This dataset includes manufacturing orders, including a description of the manufacturing plant, the order code (SKU), material name, planned vs actual quantity produced, standard vs actual costs of production, and the order posting date. Includes a key column uniquely identifying each product order.
- Supply Chain Procurement Data (ERP) (2890 rows)
  - Raw material sourcing dataset. This dataset includes the raw material and its supplier, the quantity and unit cost, country of origin, delivery date and whether the delivery was on time or not, along with the carbon dioxide footprint of the interaction.
- Product Master (750 rows)
  - Reference dataset including SKU codes matching Manufacturing Production, as well as Brand (Material Name) and Category (Target Industry).

The datasets were cleaned. Underscores were replaced with spaces and matching columns between datasets were renamed to make navigation and merging simpler. This include Brand (Product Master), Material Name (Manufacturing Production) and Product Interest (Sales Pipeline) renamed to Brand, Material Code (Manufacturing Production) and SKU Code (Product Master) renamed to SKU Code, and Category (Product Master) and Target Industry (RD Pipeline) renamed to Category.

In addition, RD Pipeline Stage was renamed to Development Stage to differentiate from Sales Pipeline Stage. RD Pipeline Development Stage was split between the stage number (1-5) and the rest, and the number was kept. This made it easier to order and manipulate.

Three merges were made. Product Master was joined with Manufacturing Production as a Left Join on SKU Code, a one-to-many relationship. This enabled the identification of each Manufacturing Production entry with a Brand and Category. Sales Pipeline was joined with RD Pipeline as a Left Join on SalesOpp ID, a one-to-many relationship. This allowed R&D projects to be matched with their respective deals. This is a one-to-many relationship: all R&D projects have a respective deal, but not all deals have a respective project.

These two datasets were merged along brand, so that each SKU code was matched to its corresponding brand, resulting in approximately 15 SKUs per brand.

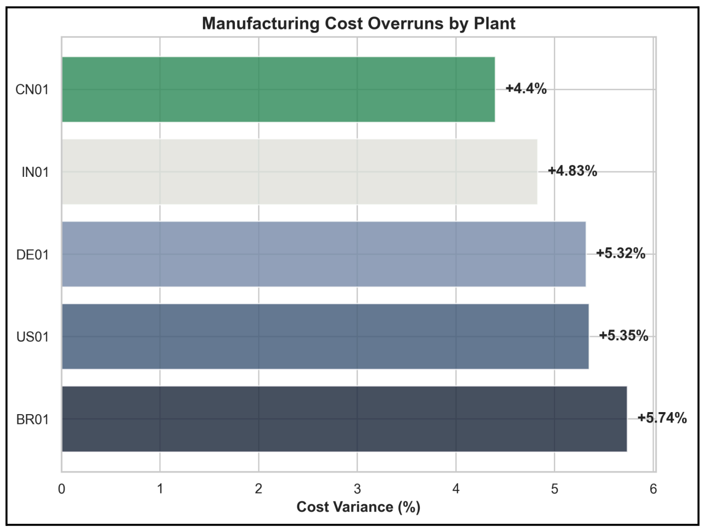
No direct mapping between raw materials (Supply Chain Procurement) and finished goods; hence CO_2 emissions cannot be allocated per SKU.

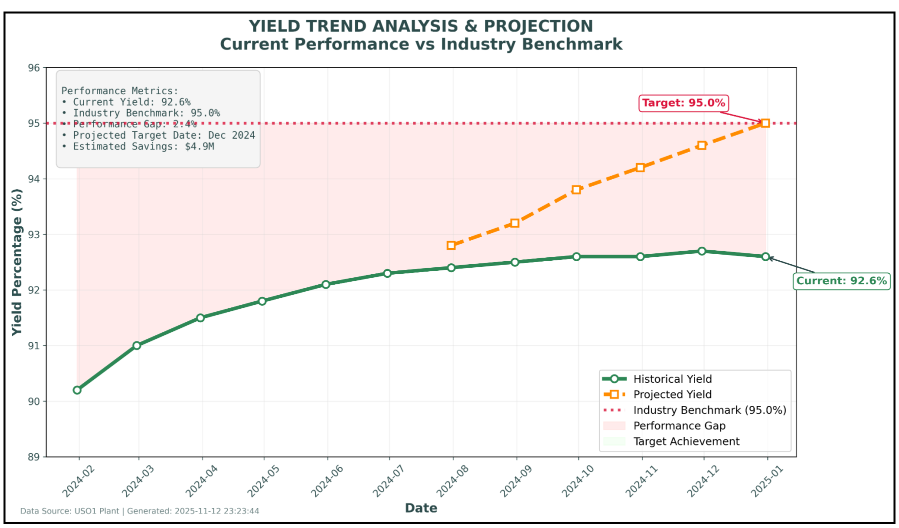
4. Analysis and Insights

4.1 Descriptive Analysis (What is Happening?)

4.1.1 Manufacturing Cost Performance

GreenFuture BioChem faces significant manufacturing cost overruns averaging 5.1% above target, equivalent to $78 per metric ton in additional cost. The standard manufacturing cost is budgeted at $1,497-1,500 per MT, while actual costs consistently reach $1,576-1,577 per MT across all production facilities.





4.1.2 Production Yield Efficiency

The company operates at a 92.5-92.6% average yield, resulting in 7.4-7.5% material wastage across manufacturing operations. This falls below industry best practices of 95%+ yield efficiency, representing substantial raw material losses and inefficiency.

4.1.3 Supply Chain Cost Structure

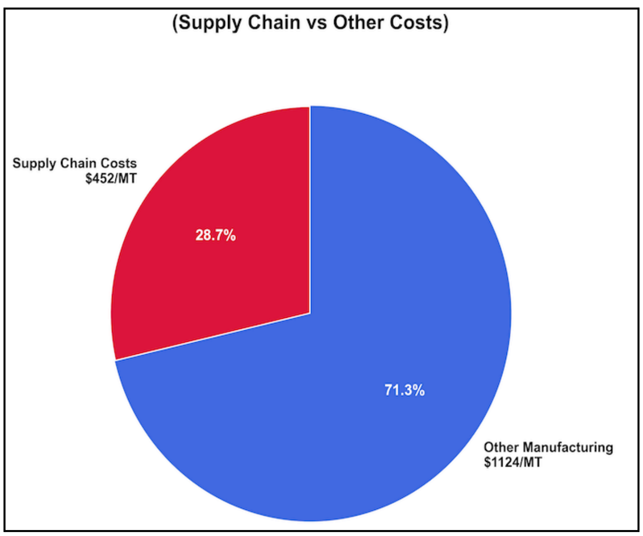
Raw material procurement costs average $450-452 per MT, constituting 28.7% of total manufacturing expenses. The supply chain operations involve 2,890 procurement orders supporting production activities, creating significant upward pressure on overall production costs.

4.1.4 Plant Performance Variability

Substantial performance differences exist across the global manufacturing network, with BR01 (Brazil) showing highest cost overruns at +5.74% and CN01 (China) demonstrating lowest yield performance at 92.22%. All plants consistently underperform against established targets despite their geographic and operational differences.

4.1.5 Operational Scale & Complexity

The company maintains substantial operational complexity with 312 active innovation projects in its R&D pipeline and a product portfolio of 750 SKUs across five market categories. This is supported by annual manufacturing capacity of 4,218,157 MT distributed across five global plants in the US, Germany, China, Brazil, and India.



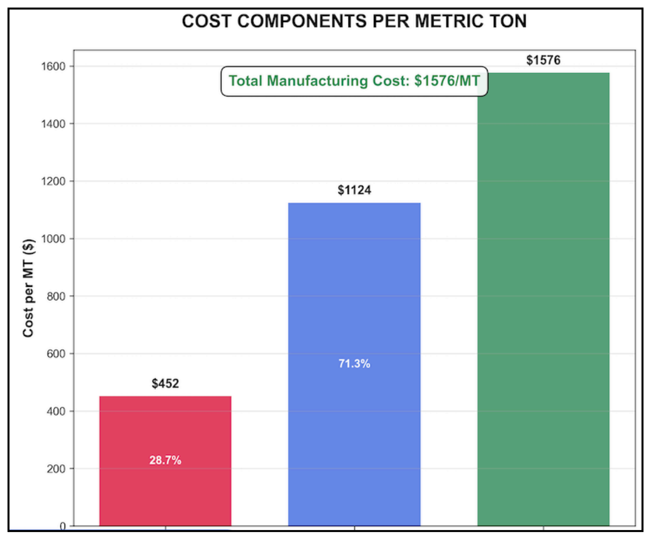


Fig: This cost breakdown visualization reveals that supply chain expenses, at $452 per metric ton, constitute 28.7% of the total manufacturing cost.

CO2 Emissions and Cost

Suppliers:

- There is a positive correlation between cost and emissions by supplier.
- EcoMaterials Co. and SustainSource Inc. are the lowest-cost and lowest-emission suppliers, averaging $448.78-$450.71 per MT and 288-297kg CO2/MT.
- NaturInput GmbH and GreenFibre BV are the highest-cost, highest-emission suppliers, averaging $452.35-$454.30 per MT and 307-308kg CO2/MT.
- NaturInput and GreenFibre collectively supply >470,000 MT, making them major contributors to supply-chain CO2 and procurement cost inflation.

Materials:

- Palm Oil Derivative is one of the strongest performers, but underutilized: \$446.68 per MT and 297.62kg CO_2/MT, below the dataset averages.
- Wood Pulp and Soy Protein Extract show higher costs (\$456–\$460 per MT), while Algae Oil has the highest CO_2 emissions (314kg CO_2/MT).

Countries of Origin:

- China offers the lowest emissions (292kg CO_2/MT) and competitive costs (\$450/MT).
- Indonesia shows the highest emissions (312.5kg CO_2/MT) and elevated costs (\$454/MT)

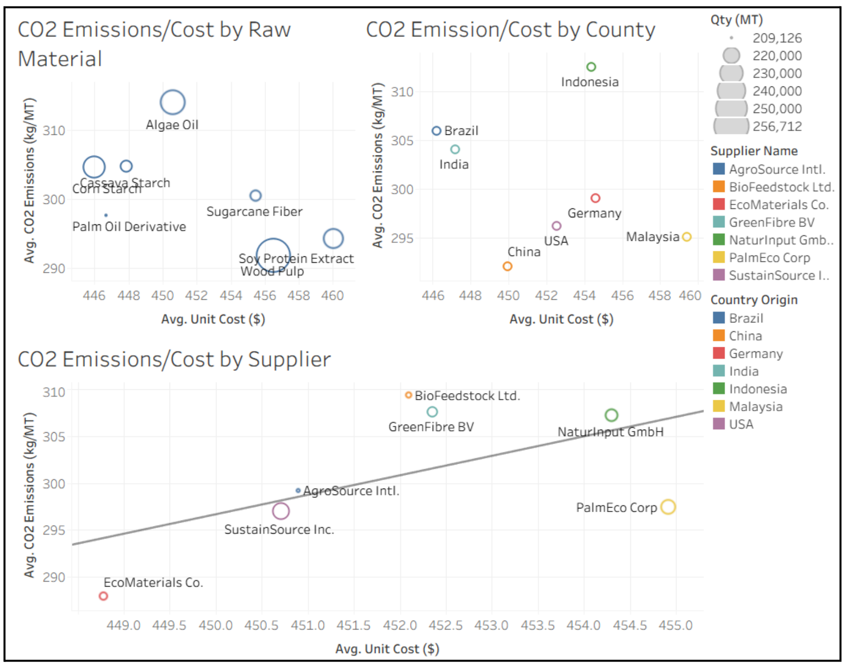


Fig. CO_2 emissions and cost by raw material, supplier, and origin country.

4.2 Diagnostic Analysis (Why is it Happening?)

4.2.1 Root Cause: Fundamentally Flawed Cost Models

The primary driver of cost overruns is inaccurate standard cost models rather than operational execution issues. Statistical analysis reveals a strong correlation ($r = 0.91$) between actual costs and standard costs, indicating the cost estimation framework itself is systematically defective and consistently underestimates true expenses.

4.2.2 Critical Insight: Yield-Cost Disconnect

Contrary to conventional manufacturing wisdom, yield performance shows virtually no correlation with cost overruns ($r = -0.002$). This counter-intuitive finding demonstrates that yield inefficiencies and cost overruns are independent problems requiring separate intervention strategies rather than combined solutions.

4.2.3 Supply Chain Cost Pressure

Raw material costs of $452 per MT represent 28.7% of total manufacturing costs, creating significant upward pressure on overall production expenses. The consistent underestimation of procurement costs in standard cost models drives the systematic budget variances observed across all facilities.

4.2.4 Systemic Plant Inefficiencies

All five manufacturing plants exhibit similar patterns of underperformance, indicating systemic rather than localized issues. The Brazil plant (BR01) shows poor cost control processes with +5.74% variance, while the China plant (CN01) suffers from operational inefficiencies affecting yield at 92.22%. The consistent variability in both yield and cost performance across all facilities demonstrates that GreenFuture's challenges stem from structural issues in cost modeling and systemic operational inefficiencies rather than isolated execution problems, requiring comprehensive organizational solutions rather than piecemeal plant-level interventions.

5. Strategic Recommendations

5.1 Recommendation 1: Standard Cost Model Revision

Key Insight: Standard cost models are fundamentally flawed, showing 83.6% explanatory power but consistently underestimating actual costs by 5.1% ($77/MT), indicating systemic estimation errors rather than operational variance.

Recommended Action: Immediately overhaul the standard cost calculation methodology by integrating real-time procurement data into cost models, implementing dynamic cost factors that adjust for supply chain volatility, establishing a cross-functional cost review team with representatives from Procurement, Manufacturing, and Finance, and developing plant-specific cost benchmarks based on actual performance data.

Expected Impact: This initiative is projected to deliver $259.1 million in annual savings through an 80% improvement factor applied to the current $323.8 million overspend. The revision will improve budget accuracy, as demonstrated by the R^2 value of 0.836, and transition the company from reactive variance analysis to proactive cost management.

5.2 Recommendation 2: Targeted Yield Improvement Program

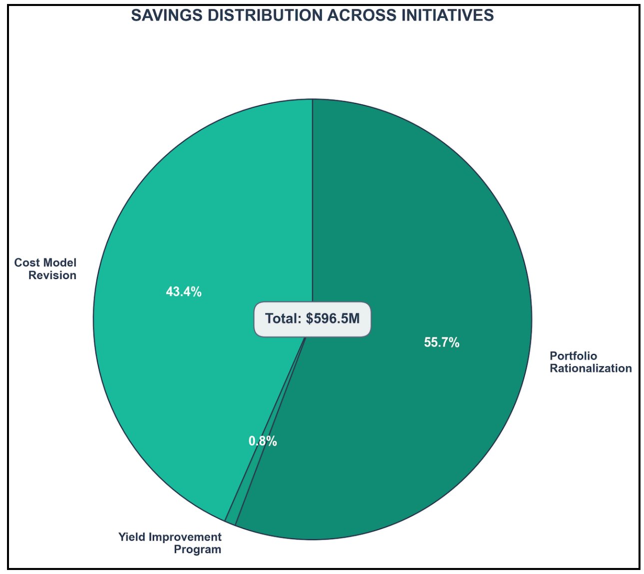
Key Insight: While yield inefficiencies do not drive cost overruns, the 2.4% gap from the industry benchmark represents significant material waste, with the US01 plant showing the largest improvement potential.

Recommended Action: Launch a phased yield optimization initiative beginning with the US01 plant. The program should implement lean manufacturing principles and Six Sigma methodologies, establish real-time yield monitoring systems, conduct operator training focused on waste reduction, and set progressive yield targets of 93.5% by Q2, 94.0% by Q4, and 95.0% by Year 2.

Expected Impact: This program is expected to generate $4.9 million in annual savings from the US01 plant alone, with an additional $46.2 million potential across all facilities. It will also reduce raw material consumption by approximately 55,000 MT annually.

5.3 Recommendation 3: Product Portfolio Rationalization

Key Insight: Current manufacturing complexity with an extensive 750-SKU portfolio creates operational inefficiencies, with analysis showing negative correlation between SKU count and production performance.



Recommended Action: Execute a comprehensive portfolio optimization strategy through ABC analysis to identify bottom-performing SKUs, implement product lifecycle management for strategic discontinuations, consolidate similar products, establish an SKU profitability dashboard, and align the R&D pipeline with high-margin, sustainable products.

Expected Impact: Portfolio rationalization is projected to deliver $332.5 million in annual savings through a conservative 5% efficiency gain. The initiative will reduce SKU complexity by 20% and improve production line utilization through reduced changeovers.

5.4 Recommendation 4: CO_2 and Cost Optimization

Recommended Action: Shift sourcing toward low-emission, low-cost suppliers including EcoMaterials Co. and SustainSource Co. while reducing reliance on high-cost, high-emission suppliers. Increase utilization of Palm Oil Derivative and rebalance regional sourcing by prioritizing China over Indonesia. Implement a structured supply-chain optimization strategy to consistently favor suppliers and materials that reduce total emissions and cost.

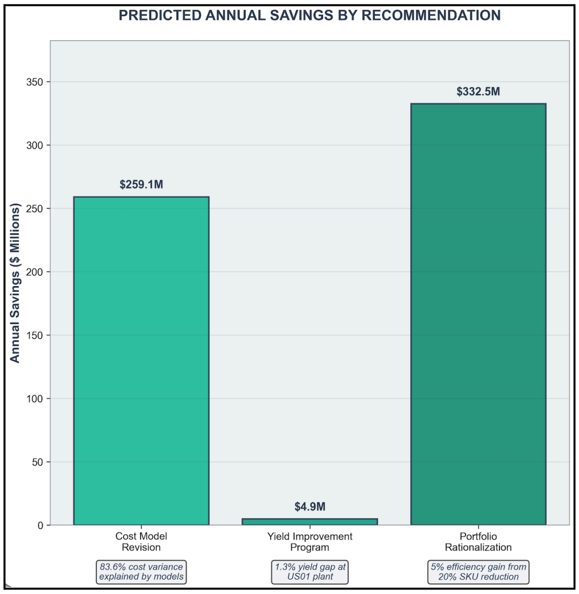
5.5 Recommendation 5: Data Integration and ERP Utilization

Recommended Action: Require Sales, Manufacturing, and Procurement to coordinate data collection within the ERP to ensure fully integrated operational data. Expand Sales data collection to include actual revenue figures and configure the ERP to automatically link sales, cost, and production data.

5.6 Implementation Roadmap and Financial Impact

Implementation Timeline: The initiatives will be executed through a phased approach beginning with cost model revision and yield improvement in Quarters 1-2, followed by portfolio rationalization in Quarters 3-4, and strategic optimization in Year 2.

Financial Impact Summary: Total projected annual savings of $596.5 million represent a substantial return on investment with minimal capital expenditure. This comprehensive approach directly addresses root causes while strengthening GreenFuture's position as a sustainable, cost-competitive leader in the bio-chemical market.



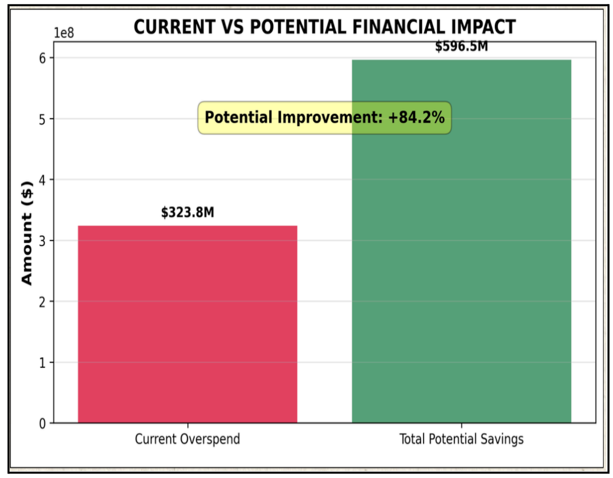


Fig. This financial impact analysis quantifies the transformative potential of our recommendations, projecting a total savings of $596.5 million that would reduce current overspend by 84.2%.

6. Conclusion

This analysis shows that GreenFuture BioChem's operational challenges stem from issues in cost modeling and data integration, and an overly complex product portfolio. Cost overruns, yield inefficiencies, and misaligned sourcing practices collectively weaken both profitability and sustainability performance.

The recommendations outlined, rebuilding the standard cost model, improving yield performance, rationalizing the product portfolio, optimizing CO_2 and cost in the supply chain, and strengthening ERP-driven data integration, directly address the root causes identified in the diagnostic analysis. Implementing these initiatives will materially reduce costs, enhance production efficiency, improve sustainability, and provide GreenFuture with accurate, integrated data for decision-making.

Looking ahead, GreenFuture should prioritize building a data architecture that connects procurement, manufacturing, and sales, enabling future analyses such as SKU-level margin and emissions modeling.

GreenFuture now has a clear, data-driven roadmap for strengthening operational excellence, improving profitability, and supporting sustainable growth.

References

Data Sources

GreenFuture BioChem. 2025. *Manufacturing Production Data*. Dataset.
GreenFuture BioChem. 2025. *Product Master Database*. Dataset.
GreenFuture BioChem. 2025. *R&D Pipeline Records*. Dataset.
GreenFuture BioChem. 2025. *Sales Pipeline Data*. Dataset.
GreenFuture BioChem. 2025. *Supply Chain Procurement Records*. Dataset.
GreenFuture BioChem. 2025. *Product Metadata Documentation*. Dataset.

Software and Analytical Tools

Salesforce. 2024. *Tableau Prep*. Computer software.
Salesforce. 2024. *Tableau Desktop*. Computer software.
Python Software Foundation. 2023. *Python*. Computer program. https://www.python.org
DeepSeek. 2025. *DeepSeek*. AI assistant. https://www.deepseek.com
OpenAI. 2025. *ChatGPT*. AI assistant. https://chatgpt.com

AI-assisted tools were leveraged in this project to augment the analytical process, primarily for debugging and optimizing Python code for data cleaning and statistical analysis. These tools served as an aid to efficiency; however, all final code, analytical methodologies, data interpretations, and strategic conclusions were developed and validated through independent critical thinking.