

Problem Statement Worksheet (Hypothesis Formation)

How can dairy producers leverage historical sales data to identify the key factors influencing total revenue, and develop a predictive model that optimizes pricing and operational decisions to maximize overall sales performance?

H

1 Context

The dairy industry faces constant challenges balancing product freshness, pricing, and inventory. The Dairy Goods Sales dataset (2019–2022) contains detailed information on farm operations, product characteristics, and sales transactions — including farm size, cow population, storage conditions, shelf life, quantity sold, and total revenue. Despite having access to this data, producers often lack a systematic understanding of which factors most influence revenue performance. A data-driven approach is needed to identify the variables that drive revenue and to optimize pricing, production, and sales decisions accordingly.

2 Criteria for success

- Build a model that accurately predicts Approx. Total Revenue (INR).
- Provide insights that can guide pricing adjustments, shelf-life management, and storage strategies.
- Present results in a clear and interpretable format for business stakeholders.

3 Scope of solution space

- Analyze the Dairy Goods Sales dataset (2019–2022) to explore relationships between pricing, shelf life, storage conditions, and product attributes.
- Limit scope to sales and operational factors; exclude external economic or climatic variables not included in the dataset.

4 Constraints

- Dataset may contain missing, inconsistent, or outlier values that require cleaning.
- Certain qualitative factors (e.g., marketing campaigns or local economic trends) are not represented.
- Model performance depends on feature completeness and accuracy of sales data.

5 Stakeholders to provide key insight

- Farm Owners / Dairy Producers: Seek to maximize profitability through improved pricing and production planning.
- Sales and Distribution Managers: Use revenue insights to optimize sales channels and inventory allocation.

6 Key data sources

- Dairy Goods Sales Dataset (2019–2022): Includes farm-level, product-level, and sales-level details such as price, quantity, shelf life, and revenue.

H

D

E

I

P