Dairy Goods Sales Revenue Driver Analysis

Wen Yang

Springboard Capstone 2

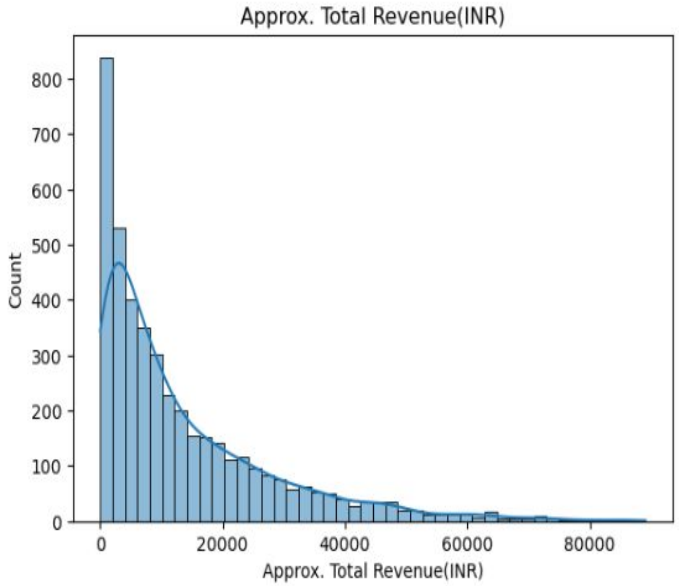
Project Overview

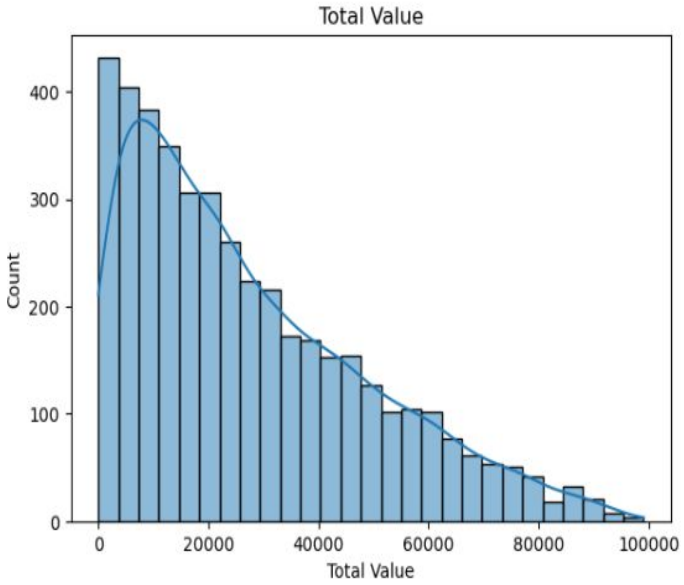
- Goal: Identify key operational and commercial drivers of dairy revenue.
- Focus: Diagnostic insights, not forecasting.
- Dataset: 4,325 records (2019–2022).

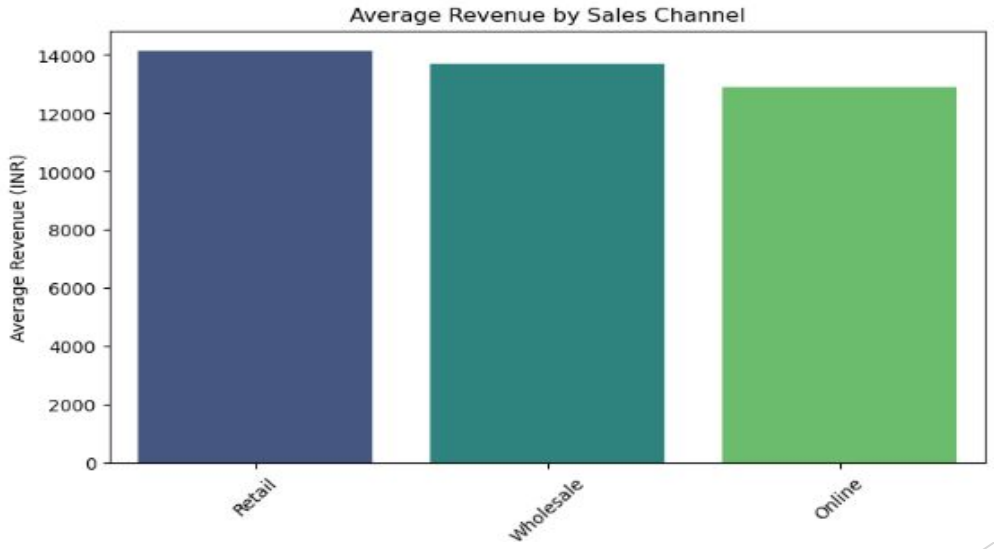
Dataset Description

- Features:
  Product attributes
  shelf life, pricing
  production volumes
  reorder behavior
  stock levels
  farm characteristics
- Target Variable: Approx. Total Revenue (INR).

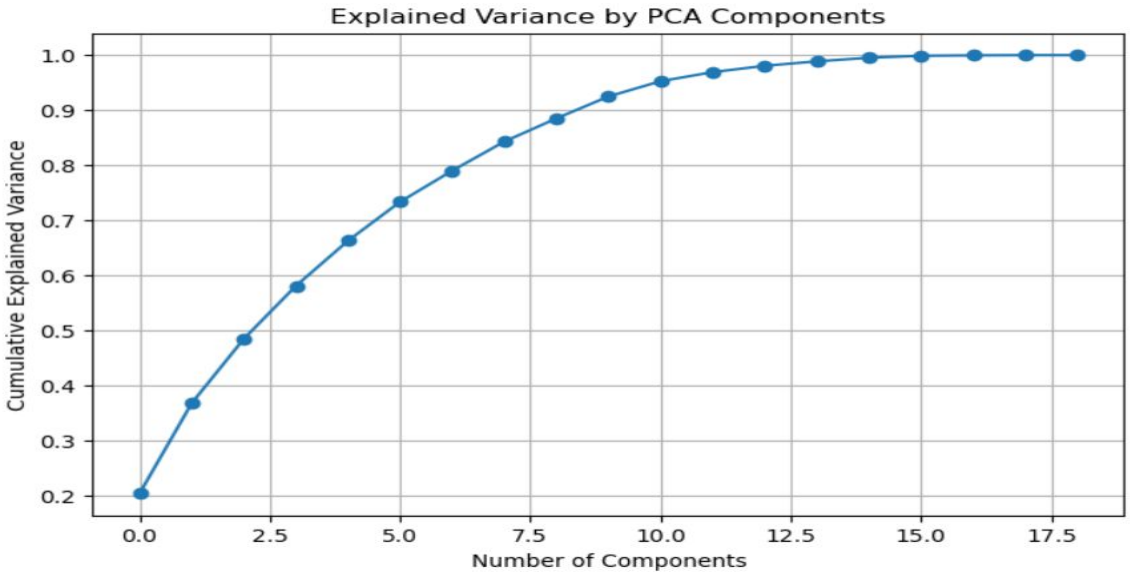
EDA Highlights (1/2)

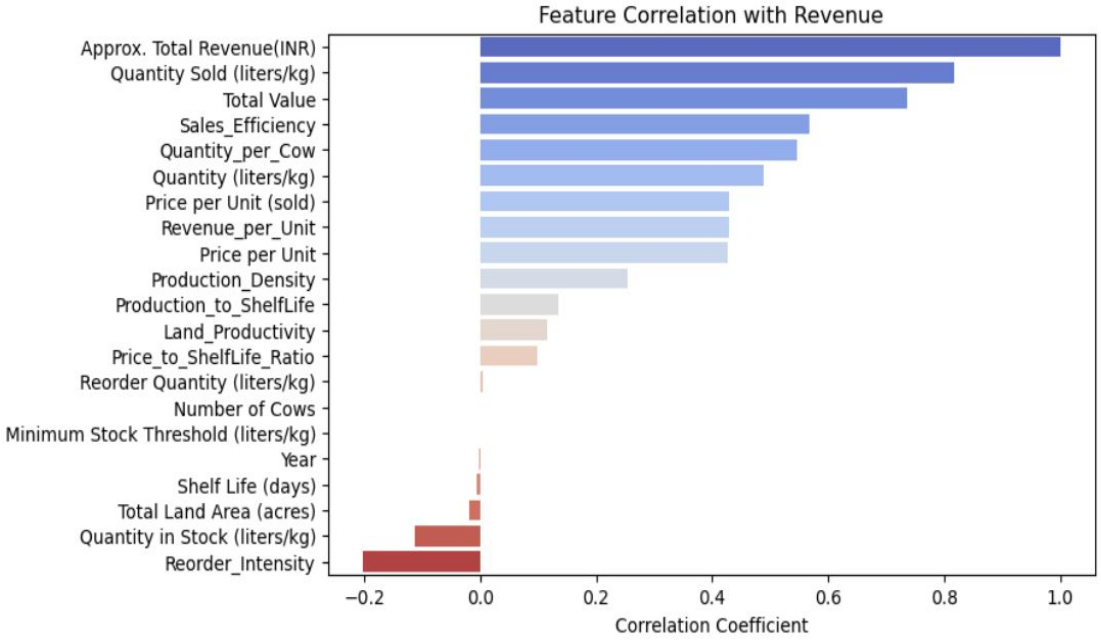






EDA Highlights (2/2)





Feature Engineering

Key engineered metrics:

Productivity & Operational Efficiency:

- Quantity_per_Cow
- Production_Density
- Sales_Efficiency

Demand Strength & Stock Movement:

- Reorder_Intensity
- Production_to_ShelfLife

Pricing Dynamics:

- Price_to_ShelfLife_Ratio

Preprocessing Steps

- One-hot encoding for categorical features
- StandardScaler for numeric variables
- Train/Test split (80/20)
- Removed leakage features:

  Quantity Sold,
  Price per Unit,
  Revenue_per_Unit,
  Total Value,
  Quantity

Modeling Approach

Models built:

- Linear Regression
- Random Forest Regressor
- Gradient Boosting Regressor (Final Model)

Model Performance

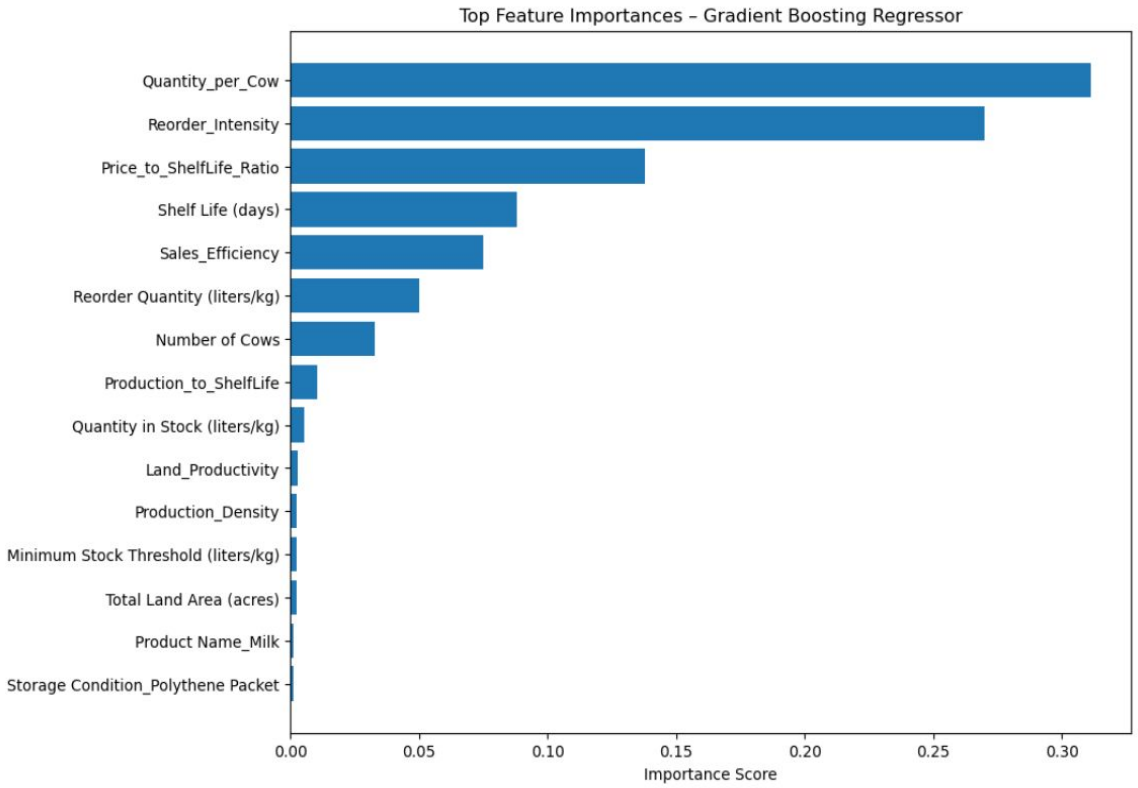
- Linear Regression: CV MAE ~7610, R^2 = 0.527
- Random Forest: CV MAE ~3632, R^2 = 0.866
- Gradient Boosting: CV MAE ~2589, R^2 = 0.924

- Final Model: Gradient Boosting Regressor--
  the strongest ability to capture nonlinear relationships

Key Drivers

Top Drivers:

- 1. Quantity_per_Cow
- 2. Reorder_Intensity
- 3. Price_to_ShelfLife_Ratio
- Other: Sales_Efficiency, Shelf Life, Land Productivity



Business Recommendations

- 1. Improve per-cow productivity.
- 2. Prioritize high-demand (reorder-intensity) products.
- 3. Align pricing strategy with shelf-life behavior.

Limitations

- Missing external factors (seasonality, competition, logistics).
- Revenue ≠ profit (lack of cost data).
- Some engineered features remain correlated.

Future Work

- Incorporate cost for profit optimization.
- Use SHAP for deeper interpretability.
- Add time-series or forecasting.
- Integrate external economic/environmental data.