

## **DAIRY PRODUCTS SALES PERFORMANCE ANALYSIS**

## ABSTRACT

The dairy business plays a vital role in the food industry by providing essential dairy products while generating revenue for companies. Managing operations efficiently in this sector is inherently complex due to the diversity of products, brands, farm sizes, and regional markets. Dairy companies face the constant challenge of balancing production levels, inventory management, and sales strategies to meet market demand without incurring wastage. Understanding which products and brands sell the most, recognizing regional sales variations, and identifying the most profitable sales channels are critical steps in enhancing operational efficiency and profitability. Efficient resource management in the dairy sector involves evaluating farm capacities, monitoring production outputs, and maintaining optimal stock levels. By examining the number of cows on farms of varying sizes across different locations, companies can assess their production potential and align it with demand forecasts. Inventory levels for individual products are carefully monitored to ensure adequate supply while avoiding overstocking, which can lead to spoilage or increased storage costs. This approach enables businesses to maintain a steady supply chain and make timely adjustments to production schedules according to market needs.

Analyzing regional performance is an important aspect of effective dairy management. By examining how products and brands perform across different locations, companies can identify which products are most successful and which markets have the highest potential. Seasonal demand trends and periods of peak consumption are also studied to optimize production schedules, allocate resources efficiently, and ensure that popular products are available when needed most. Additionally, sales channels are evaluated to determine how efficiently products are distributed and how much revenue each channel generates. By understanding which products perform best through retail, distributors, or online platforms, companies can focus on the channels that maximize revenue. Combining data on brands, products, and locations provides a comprehensive view of overall performance and helps identify the customer segments and regions that contribute most significantly to profitability. The insights gained from such an integrated analysis enable dairy companies to make data driven decisions that improve production efficiency, optimize inventory management, and enhance overall operational performance. By focusing on high-performing products, brands, and regions, companies can not only increase revenue but also ensure sustainable growth in a competitive market.

## 1. INTRODUCTION

The dairy industry is a major segment of the food sector, providing essential dairy products such as milk, cheese, butter, yogurt, and other processed items to consumers while generating significant revenue for companies. Efficient management of production, inventory, and sales is essential, as dairy businesses operate across multiple product lines, brands, farm sizes, and regions. Managing these aspects efficiently ensures profitability, reduces wastage, and supports the sustainability of operations. However, many dairy companies face challenges in identifying top-performing products and brands, understanding regional sales variations, and evaluating the contribution of different sales channels to overall revenue. In addition, factors such as farm capacity, production levels, stock availability, and market demand are closely interconnected. Mismanagement in any of these areas can lead to revenue loss, inventory shortages, product wastage, and suboptimal utilization of resources.

Focuses on analyzing dairy production and sales data to provide insights that help improve operational efficiency and strategic decision-making. The dataset includes critical information such as farm size, number of cows, types of dairy products, brand details, production quantities, sales volumes, stock levels, customer locations, and sales channels. By studying these variables collectively, the project aims to uncover patterns and relationships that influence production performance, inventory management, sales trends, and revenue generation across multiple regions and product lines. Understanding these patterns enables dairy businesses to make data-driven decisions for optimizing operations and enhancing profitability.

The analysis begins with a structured data cleaning and preprocessing phase to ensure accuracy and consistency. This phase involves handling missing values, standardizing categorical variables such as farm size and product type, and preparing the data for visualization and statistical examination. Following preprocessing, exploratory data analysis (EDA) techniques are employed to identify trends, distributions, and relationships among key variables. Visual tools such as bar charts, stacked bar charts, tables, and heatmaps are used to present insights in an intuitive and interpretable manner. These visualizations highlight important aspects, including production quantities across different farm sizes, top-selling products and brands, regional sales distribution, stock availability, and revenue contributions from various sales channels. Special attention is given to regional and farm-level differences, which helps identify high-performing areas and potential bottlenecks in production or sales.

A major focus of this is to examine production and sales distribution across products and brands. This analysis identifies high-demand items, top-performing brands, and regions with the highest sales, providing a clear understanding of market trends. In addition, sales channel analysis is conducted to determine which channels, such as direct retail, wholesale, or online platforms, contribute most significantly to overall revenue. By assessing the performance of different channels, the project highlights opportunities to optimize distribution, improve customer reach, and increase profitability. Inventory levels are also analyzed relative to sales to detect potential overstock or understock situations, ensuring that production and stock management align with market demand.

Further investigates farm resources and production capacity by comparing small, medium, and large farms. This includes an assessment of the number of cows, production output, and sales generated from farms of different sizes. Evaluating farm-level efficiency provides insights into resource utilization and helps businesses identify which farm categories contribute most effectively to overall operations. Revenue contribution by product, brand, and location is analyzed to identify profitable items, high-value customers, and regions generating maximum income. These insights support targeted business strategies, such as focusing on high-demand products, allocating production resources efficiently, and planning marketing campaigns for specific regions segments.

To analyzes farm-level data to understand how the number of cows and farm size affect production capacity. It evaluates sales performance of dairy products and brands across regions to identify top-selling items and profitable markets. Stock levels are monitored to prevent overstocking and shortages, while the contribution of different sales channels to revenue is assessed to optimize distribution. Trends in product demand are studied to highlight seasonal and regional variations. Revenue contributions by products and brands are prioritized to focus on high-performing items. Geographical comparisons help allocate resources effectively, ensuring stock meets local demand. Inventory management is reviewed to reduce waste and improve efficiency. The work also examines how production, sales, and stock interact to influence overall profitability. Insights from the analysis enable better planning of marketing and distribution strategies. By combining farm, product, and sales data, the project offers a comprehensive view of operational performance. These findings assist dairy businesses in planning production, managing inventory, and enhancing profitability, providing practical guidance for improving efficiency and supporting growth in the dairy sector.

## **1.1 PROBLEM STATEMENT**

The dairy industry faces significant challenges in efficiently managing production, inventory, and sales due to the diversity of products, brands, farm sizes, and operational regions. Many dairy businesses struggle to identify which products and brands perform best, how sales vary across different locations, and which sales channels contribute most to overall revenue. Ineffective management of farm capacity, production levels, stock availability, and market demand can lead to operational inefficiencies, revenue loss, product wastage, and suboptimal utilization of resources.

Dairy production and sales analysis focuses on evaluating farm-level and sales data to uncover meaningful insights that can drive operational improvements and strategic decision making. By analyzing trends in farm resources, product types, regional sales, stock availability, and sales channel contributions, the study aims to highlight key factors affecting overall business performance. Additionally, it investigates demand fluctuations, identifies top performing products and brands, and evaluates the efficiency of different sales channels. The goal is to support dairy companies in building a data-driven, efficient, and profitable operational framework. This approach helps minimize wastage and optimize inventory management. It also ensures that production and sales remain aligned with customer needs across regions. Ultimately, it contributes to sustainable growth and long-term competitiveness in the dairy sector.

- Dairy businesses often lack clarity on which products or brands contribute most to overall revenue.
- Sales performance varies widely across regions, making it difficult to plan production and allocate resources effectively.
- Stock levels are frequently misaligned with actual demand, leading to shortages, overstock, wastage, and financial losses.
- The contribution of different sales channels to revenue is not analyzed thoroughly, limiting the optimization of distribution strategies.
- The relationship between farm capacity, production levels, and sales performance is not well understood, affecting operational efficiency.
- Seasonal and regional variations in product demand are often overlooked, resulting in poor inventory management and weak marketing strategies.

## 1.2 TOOL DESCRIPTION - JUPYTER NOTEBOOK

Jupyter Notebook is an open-source, web-based interactive computing environment that allows users to create and share documents containing live code, equations, visualizations, and narrative text. It supports various programming languages, with Python being the most commonly used in data science. Jupyter Notebook served as the primary development platform due to its ability to combine code execution, data analysis, and visualization in a single, organized document. It helped streamline the workflow by enabling step-by-step execution, immediate output display, and easy debugging. The notebook format also allowed combining charts, tables, and explanatory notes in a readable and organized way, which helped in better understanding the data and presenting the findings clearly.

## PACKAGES USED

To perform the analysis on dairy production and sales data, the following Python packages were used:

- **Pandas:** Used for loading and managing the dataset. It helped in cleaning the data, handling missing values, filtering records, grouping data by farm size, product, brand, and region, and performing operations such as sorting and summarizing production and sales information.
- **Matplotlib:** A basic plotting library used to create bar charts, line graphs, and pie charts for visualizing production levels, sales performance, stock availability, and regional comparisons in the dairy sector.
- **Seaborn:** Built on top of Matplotlib, it made it easier to create attractive and informative visualizations such as heatmaps, boxplots, and trend comparisons. It enhanced the clarity of visual outputs and was particularly useful for highlighting product demand patterns and regional sales variations.
- **Plotly Express:** This library was used for creating interactive visualizations such as bar plots, scatter plots, and regional comparisons. It allowed dynamic exploration of the data with hover insights and filter options, making the analysis more engaging and insightful.
- **Plotly Graph Objects:** Used for creating advanced and customized interactive charts like stacked bar charts and detailed breakdowns of sales and stock data. These visualizations helped present complex relationships between production, inventory, and sales in a clear and interactive way.

## **2. DOMAIN - DAIRY PRODUCTION AND SALES ANALYTICS**

The domain Dairy Production and Sales Analytics, which plays an important role in the agricultural and food industry. Dairy products such as milk, curd, butter, ghee, and cheese are consumed daily by people of all age groups, making this sector a key contributor to food security and economic growth. With demand steadily increasing, companies face challenges in balancing production, inventory, and sales across different regions and channels. Seasonal variations, regional consumption patterns, and changing customer preferences add complexity, and analyzing production and sales data becomes essential to improve efficiency, minimize wastage, and enhance profitability.

This domain focuses on analyzing both farm-level production data and market-level sales performance. Factors such as farm size, number of cows, production capacity, product variety, sales volumes, regional demand, and revenue contributions are examined to identify productive farms, top-performing products, and high-demand regions. By using data effectively, dairy companies can align production with market demand and make informed decisions on distribution, inventory, and resource allocation.

The analysis is carried out using exploratory data analysis (EDA), descriptive statistics, and data visualization techniques. Sales by region can reveal strong markets, product comparisons can show high-revenue items, and seasonal demand analysis helps prepare for peak periods like festivals or summer months. These insights support forecasting, production planning, and timely product availability for consumers.

The significance of this domain lies in its practical business impact. By identifying patterns in sales and production, companies can reduce wastage, improve profitability by focusing on high-demand products, and design targeted marketing strategies. Understanding sales channels such as wholesale, retail, and direct distribution also enables better logistics and revenue optimization. Linking farm capacity with sales ensures efficient use of cattle, feed, and manpower, leading to stronger operations.

Dairy production and sales analytics provide a data-driven approach that improves decision-making, resource utilization, and customer satisfaction. By connecting farm operations with market data, businesses gain a clear view of performance and strengthen their competitiveness in an expanding industry.

## **2.1 CASE STUDY - DAIRY SALES AND PRODUCTION ANALYSIS**

Dairy company faced challenges in managing production, inventory, and sales across multiple farms, product lines, brands, and regions. The company needed insights into the number of farms by location, herd sizes by farm category (small, medium, and large), production volumes, stock availability, and sales performance across regions and channels. Additionally, it was important to track top-selling products and brands, monthly sales trends, products with the longest shelf life, and revenue contributions for better business monitoring. To address these challenges, the company analyzed historical production and sales data using Python with Pandas, Matplotlib, Seaborn, and Plotly. Key variables included number of farms by location, cows by farm size, production and stock levels, quantity sold by product and brand, sales channel distribution, revenue contributions, monthly sales trends, and top performing products, brands, and customer locations. This analysis revealed patterns in production, inventory, and sales, helping the company focus on high-performing products and regions.

### **OBJECTIVES**

The main objective of this work to analyze dairy production and sales data using Python to uncover key trends and patterns through data visualization.

- Determine the number of farms by location and analyze herd distribution by farm size to understand supply at the farm level.
- Analyze production, sales, and stock levels by product to monitor inventory and ensure alignment with market demand.
- Examine regional sales performance by product and brand to identify high-demand areas and guide distribution strategies.
- Assess the revenue contributions of products and brands to prioritize profitable items and optimize business decisions.
- Study overall sales channel distribution and product-wise performance to enhance sales strategies and improve channel efficiency.
- Analyze monthly dairy sales trends and identify top-performing products, brands, and customer locations to support operational planning and strategic decision-making.

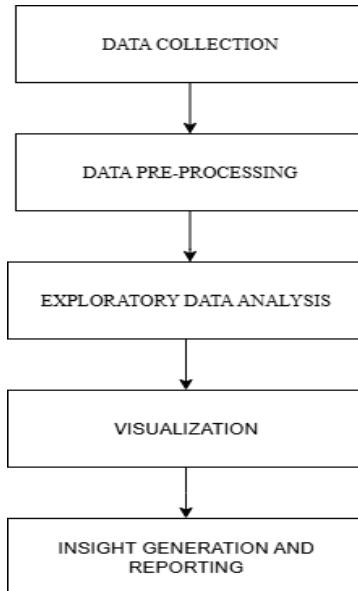
This analysis helps the dairy company make data-driven decisions, ensuring efficient production, optimal inventory management, and increased profitability while aligning supply with market demand.

### 3. DATA MODELLING

The dairy farm dataset contains farm information, herd size, product details, sales, stock levels, and customer information. These features are used to analyze production efficiency, sales performance, inventory management, and regional demand patterns. The system visualizes key insights such as top-performing products and brands, regional sales trends, stock shortages, and high-revenue customer locations, enabling dairy businesses to optimize production, distribution, and strategic planning.

#### 3.1 PROCESS FLOW

The process flow diagram shows how raw dairy farm data is transformed into actionable insights. The process involves data collection, cleaning, and encoding, followed by exploratory data analysis (EDA) and visualizations. Key factors such as farm location, herd size, product type, sales quantity, revenue, shelf life, and stock levels are analyzed, leading to informed decisions on inventory management, sales strategy, and business planning.



**Figure 3.1 – Process Flow**

Figure 3.1 represents the process flow, starting with data collection and preprocessing to prepare the dairy dataset, including farm, production, sales, and stock records. It then moves to exploratory data analysis to identify patterns in farm sizes, product demand, regional sales, and channels. The process concludes with generating business insights to support better planning, inventory control, and sales growth in the dairy sector.

### 3.2 DATA COLLECTION

The dataset was sourced from dairy farm operations, production records, and sales systems. It contains 4,325 records used to study production, sales, inventory, and regional performance. Key attributes include Location, Farm Size, Number of Cows, Product Name, Brand, Production Date, Expiration Date, Shelf Life, Quantity Sold, Revenue, Customer Location, and Sales Channel. The data was cleaned and organized to ensure accuracy and consistency.

### 3.3 SAMPLE DATASET

The dataset contains important details about dairy farms, products, and sales. It includes farm related fields such as Location, Farm Size, and Number of Cows, along with product information like Product ID, Product Name, Brand, Shelf Life, and Storage Condition. Sales details such as Quantity Sold, Price per Unit, Revenue, and Customer Location capture business performance. Inventory attributes like Quantity in Stock, Minimum Stock Threshold, and Reorder Quantity help track stock availability. Overall, the dataset provides a complete view of production, sales, and inventory in the dairy business.

| A  | B              | C                       | D              | E         | F          | G          | H            | I                    | J                    | K              | L           | M                 | N                 |                 |
|----|----------------|-------------------------|----------------|-----------|------------|------------|--------------|----------------------|----------------------|----------------|-------------|-------------------|-------------------|-----------------|
| 1  | Location       | Total Land Area (acres) | Number of Cows | Farm Size | Date       | Product ID | Product Name | Brand                | Quantity (liters/kg) | Price per Unit | Total Value | Shelf Life (days) | Storage Condition | Production Date |
| 2  | Telangana      | 311                     | 96             | Medium    | 17-02-2022 | 5          | Ice Cream    | Dodla Dairy          | 222                  | .86            | 19064       | 25                | Frozen            | 27-12-2021      |
| 3  | Uttar Pradesh  | 19                      | 44             | Large     | 01-12-2021 | 1          | Milk         | Amul                 | 687                  | .43            | 29294       | 22                | Tetra Pack        | 03-10-2021      |
| 4  | Tamil Nadu     | 582                     | 24             | Medium    | 28-02-2022 | 4          | Yogurt       | Dodla Dairy          | 503                  | .36            | 18377       | 30                | Refrigerated      | 14-01-2022      |
| 5  | Telangana      | 908                     | 89             | Small     | 09-06-2019 | 3          | Cheese       | Britannia Industries | 823                  | .27            | 21836       | 72                | Frozen            | 15-05-2019      |
| 6  | Maharashtra    | 862                     | 21             | Medium    | 14-12-2020 | 8          | Buttermilk   | Mother Dairy         | 148                  | .84            | 12391       | 11                | Refrigerated      | 17-10-2020      |
| 7  | Telangana      | 88                      | 51             | Medium    | 07-01-2019 | 6          | Curd         | Raj                  | 594                  | .86            | 50804       | 5                 | Refrigerated      | 03-01-2019      |
| 8  | Karnataka      | 73                      | 74             | Small     | 05-08-2022 | 3          | Cheese       | Dynamix Dairies      | 757                  | .37            | 27885       | 83                | Frozen            | 14-06-2022      |
| 9  | Bihar          | 68                      | 77             | Medium    | 14-02-2019 | 9          | Paneer       | Mother Dairy         | 203                  | .29            | 5914        | 14                | Refrigerated      | 06-01-2019      |
| 10 | West Bengal    | 72                      | 76             | Medium    | 02-08-2020 | 6          | Curd         | Raj                  | 949                  | .23            | 21630       | 6                 | Refrigerated      | 06-06-2020      |
| 11 | Telangana      | 413                     | 36             | Large     | 04-03-2022 | 8          | Buttermilk   | Mother Dairy         | 386                  | .42            | 16228       | 9                 | Refrigerated      | 03-02-2022      |
| 12 | Madhya Pradesh | 868                     | 82             | Small     | 22-12-2021 | 4          | Yogurt       | Dodla Dairy          | 53                   | .67            | 3538        | 24                | Refrigerated      | 02-11-2021      |
| 13 | Chandigarh     | 649                     | 99             | Small     | 14-10-2020 | 7          | Lassi        | Sudha                | 653                  | .85            | 55247       | 14                | Refrigerated      | 22-08-2020      |
| 14 | Delhi          | 237                     | 81             | Large     | 18-02-2020 | 3          | Cheese       | Passion Cheese       | 12                   | .58            | 722         | 69                | Refrigerated      | 18-01-2020      |
| 15 | Telangana      | 856                     | 39             | Large     | 26-12-2019 | 5          | Ice Cream    | Mother Dairy         | 612                  | .94            | 57806       | 25                | Frozen            | 15-12-2019      |
| 16 | Gujarat        | 182                     | 70             | Medium    | 10-04-2022 | 5          | Ice Cream    | Dodla Dairy          | 798                  | .89            | 70723       | 28                | Frozen            | 26-03-2022      |
| 17 | Kerala         | 15                      | 71             | Small     | 22-02-2020 | 1          | Milk         | Mother Dairy         | 822                  | .12            | 9513        | 1                 | Polythene Packet  | 31-01-2020      |
| 18 | Madhya Pradesh | 234                     | 95             | Large     | 17-12-2022 | 10         | Ghee         | Mother Dairy         | 663                  | .49            | 32391       | 138               | Ambient           | 26-11-2022      |
| 19 | Kerala         | 413                     | 88             | Medium    | 24-06-2020 | 6          | Curd         | Amul                 | 385                  | .26            | 10090       | 6                 | Refrigerated      | 08-05-2020      |
| 20 | Madhya Pradesh | 421                     | 25             | Large     | 19-01-2020 | 8          | Buttermilk   | Amul                 | 367                  | .41            | 15071       | 10                | Refrigerated      | 01-12-2019      |
| 21 | Tamil Nadu     | 534                     | 52             | Medium    | 15-06-2021 | 1          | Milk         | Amul                 | 897                  | .94            | 84057       | 28                | Tetra Pack        | 22-05-2021      |
| 22 | Jharkhand      | 705                     | 24             | Large     | 19-05-2020 | 9          | Paneer       | Raj                  | 89                   | .88            | 7763        | 10                | Refrigerated      | 04-04-2020      |
| 23 | Bihar          | 338                     | 85             | Small     | 08-08-2022 | 3          | Cheese       | Britannia Industries | 220                  | .14            | 3067        | 43                | Frozen            | 01-07-2022      |
| 24 | Gujarat        | 31                      | 41             | Large     | 16-04-2020 | 4          | Yogurt       | Amul                 | 525                  | .84            | 44316       | 28                | Frozen            | 04-04-2020      |
| 25 | Karnataka      | 396                     | 79             | Small     | 03-03-2022 | 3          | Cheese       | Britannia Industries | 879                  | .90            | 79208       | 36                | Frozen            | 04-02-2022      |
| 26 | Maharashtra    | 276                     | 80             | Small     | 24-10-2021 | 6          | Curd         | Sudha                | 213                  | .81            | 17283       | 5                 | Refrigerated      | 26-08-2021      |
| 27 | Rajasthan      | 552                     | 12             | Large     | 17-11-2021 | 9          | Paneer       | Sudha                | 886                  | .62            | 55140       | 8                 | Refrigerated      | 23-10-2021      |
| 28 | Madhya Pradesh | 110                     | 27             | Small     | 26-07-2022 | 2          | Butter       | Warana               | 430                  | .90            | 38586       | 37                | Frozen            | 05-07-2022      |
| 29 | Jharkhand      | 16                      | 90             | Small     | 25-09-2022 | 2          | Butter       | Warana               | 440                  | .77            | 34115       | 36                | Frozen            | 20-08-2022      |
| 30 | Chandigarh     | 456                     | 45             | Medium    | 24-06-2020 | 6          | Curd         | Mother Dairy         | 565                  | .86            | 48830       | 6                 | Refrigerated      | 21-05-2020      |
| 31 | Kerala         | 98                      | 87             | Small     | 13-01-2021 | 7          | Lassi        | Sudha                | 102                  | .19            | 1929        | 14                | Refrigerated      | 07-01-2021      |

Figure 3.3 - Sample Dataset

Figure 3.3 presents a sample of the dairy production and sales dataset. Key attributes include Location, Farm Size, Number of Cows, Product Details, Quantity, Price, Shelf Life, Stock Levels, and Sales Information. These features provide insights into production, inventory management, and sales performance, helping analyze product trends, farm productivity, and revenue across regions and brands.

### 3.4 DATASET DESCRIPTION

**Table 3.4 -Dataset Description**

| Attribute                           | Description   |
|-------------------------------------|---|
| Location                            | The geographical location of the dairy farm.  |
| Total Land Area (acres)             | The total land area occupied by the dairy farm.   |
| Number of Cows                      | The number of cows present in the dairy farm.   |
| Farm Size                           | The size of the dairy farm (in sq.km).  |
| Date                                | The date of data recording.   |
| Product ID                          | The unique identifier for each dairy product.   |
| Product Name                        | The name of the dairy product.  |
| Brand                               | The brand associated with the dairy product.  |
| Quantity (liters/kg)                | The quantity of the dairy product available.  |
| Price per Unit                      | The price per unit of the dairy product.  |
| Total Value                         | The total value of the available quantity of the dairy product.                                 |
| Shelf Life (days)                   | The shelf life of the dairy product in days.  |
| Storage Condition                   | The recommended storage condition for the dairy product   |
| Production Date                     | The date of expiration for the dairy product.   |
| Expiration Date                     | The date of expiration for the dairy product sold.  |
| Quantity Sold (liters/kg)           | The quantity of the dairy product sold  |
| Price per Unit (sold)               | The price per unit at which the dairy product was sold.   |
| Approx. Total Revenue (INR)         | Total Revenue (INR) The approximate total revenue generated from the sale of the dairy product. |
| Customer Location                   | The location of the customer who purchased the dairy product.                                   |
| Sales Channel                       | The channel through which the dairy product remaining in stock.                                 |
| Quantity in Stock (liters/kg)       | The quantity of the dairy product remaining in stock.   |
| Minimum Stock Threshold (liters/kg) | The minimum stock threshold for the dairy product.  |
| Reorder Quantity (liters/kg)        | The recommended quantity to reorder for the dairy product.                                      |

Table 3.4 shows that the dataset contains production, inventory, and sales attributes of dairy farms, including details like Location, Total Land Area, Number of Cows, Product Name, Brand, Quantity Produced, Price, Stock Levels, Shelf Life, and Sales Channel. These variables are used to analyze farm productivity, product performance, stock management, and revenue trends.

### 3.5 DATA PREPROCESSING

Data preprocessing was carried out to clean and prepare the dairy production and sales dataset for analysis. Missing values were handled, and categorical columns such as Product Name, Brand, Location, and Sales Channel were standardized for consistency. Dates like Production Date and Expiration Date were converted to datetime format to extract features such as production month and expiry month. Duplicate or irrelevant records were removed to ensure the dataset was ready for accurate visualizations and analysis.

**Checking for Missing Values:** The dataset was thoroughly checked for null and inconsistent values across all fields. Since no significant missing data was detected in critical columns, the dataset was retained without the need for imputation or removal of rows. This helped preserve the completeness and quality of the dataset.

**Data Type Conversion:** The Date column was converted into a proper datetime format. This conversion enabled the extraction of temporal features such as Year, Month, and Year-Month, which were later used for analysing monthly sales trends, seasonal demand variations, and shelf-life patterns.

**Categorical Variables:** Categorical fields such as Product Name, Brand, Location, and Sales Channel were standardized and cleaned for uniformity. This step ensured that categories with slight variations or inconsistencies were unified, preventing duplication and making cross-comparisons more accurate during grouped visualizations.

**Grouping for Analysis:** Certain numerical and categorical attributes were grouped into meaningful ranges to simplify interpretation. For example, Farm Size was categorized into Small, Medium, and Large, while Shelf Life was grouped into short, medium, and long durations. Similarly, stock levels were summarized into ranges to highlight low, medium, and high availability. These groupings made the analysis more interpretable and provided clearer insights into product performance, farm capacity, and inventory management.

## 4. ANALYSIS AND REPORT

The dairy production and sales dataset was carefully analysed to understand production efficiency, inventory management, and sales performance. The analysis focused on key areas such as farm location, farm size, number of cows, product details, quantity produced, sales channels, stock levels, and revenue generation. This helped identify trends in product performance, stock shortages, and high-performing sales channels, enabling better farm management and strategic business decisions.

### FARM COUNT BY LOCATION

Farm count by location visualization shows shows the number of dairy farms across various regions. It provides an overview of how farms are distributed geographically. Understanding these patterns highlights areas with high or low farm density. Such insights indicate potential production capacity in different locations.



**Figure 4.1 Farm Count by Location**

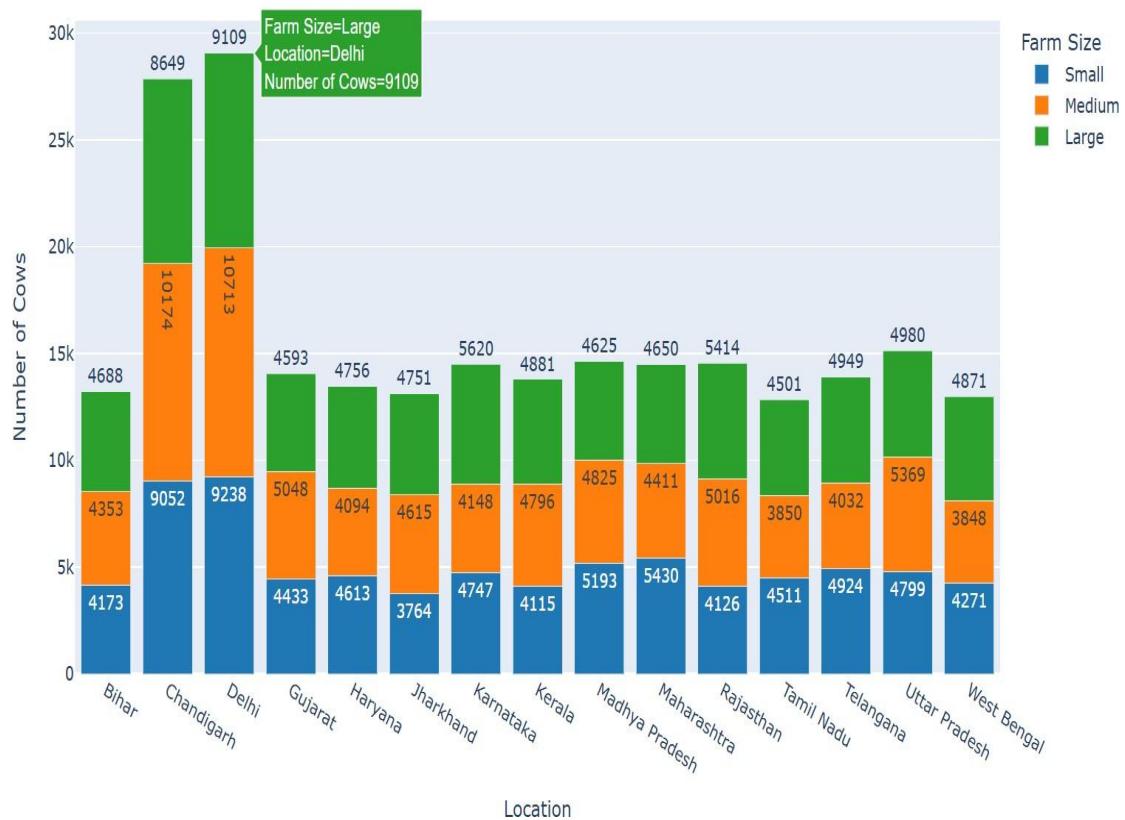
Figure 4.1 visually represents the number of dairy farms across different region showing that Delhi (525 farms) and Chandigarh (519 farms) have the highest farm counts, while other states have comparatively fewer farms. This indicates higher production capacity in regions with greater farm density.

### NUMBER OF COWS BY FARM SIZE ACROSS LOCATIONS

The distribution of cows across farms of different sizes small, medium, and large across various locations. The visualization highlights the total number of cows in each farm size

category for every region, showing which locations have the largest herds and how production capacity is distributed geographically.

Number of Cows by Farm Size across Locations



**Figure 4.2 Number of Cows by Farm Size Across Locations**

Figure 4.2 visually represent the distribution of cows across small, medium, and large farms in different locations. The chart highlights that certain regions have a higher number of cows in large farms, while other regions are dominated by small and medium farms. This indicates how herd sizes contribute to overall dairy production in each location.

## PRODUCTION, SALES, AND STOCK BY PRODUCT

Production, Sales, and Stock by Product refers to tracking the quantity of each product that is manufactured, sold, and remaining in inventory. Production shows how much of a product is made, sales indicate how much is sold to customers, and stock reflects the leftover quantity available. This helps businesses monitor performance, manage inventory, and plan future production effectively.

### Production, Sales, and Stock by Product

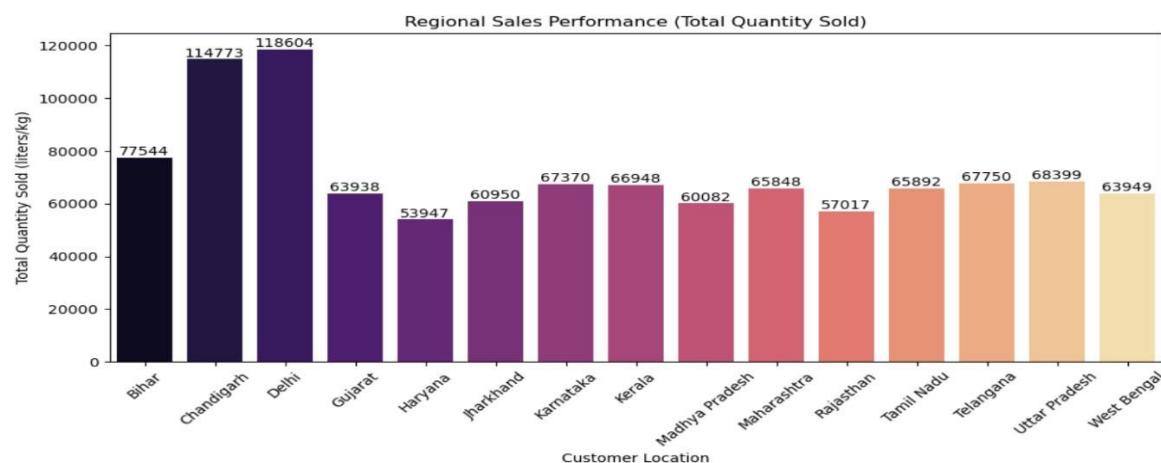
| Product Name | Quantity Produced | Quantity Sold | Quantity in Stock |
|--------------|-------------------|---------------|-------------------|
| Butter       | 225565            | 109578        | 115784            |
| Buttermilk   | 211411            | 100402        | 110789            |
| Cheese       | 206369            | 104065        | 102112            |
| Curd         | 254171            | 120415        | 133525            |
| Ghee         | 196566            | 98513         | 97858             |
| Ice Cream    | 208617            | 104780        | 103629            |
| Lassi        | 226342            | 112731        | 113385            |
| Milk         | 215246            | 109852        | 105187            |
| Paneer       | 210587            | 107931        | 102438            |
| Yogurt       | 210444            | 104744        | 105490            |

**Figure 4.3 Production, Sales, and Stock by Product**

Figure 4.3 represents the total production, sales, and remaining stock for each dairy product. It provides a comparative view of product-wise performance, showing how much was produced, the quantity sold, and what remains in inventory. This helps in understanding demand patterns and inventory management efficiency.

### REGIONAL SALES PERFORMANCE QUANTITY SOLD BY LOCATION

Regional Sales Performance many dairy products are sold in each region. It helps compare sales performance across locations to identify which areas are doing well and which may need improvement through better marketing or distribution.

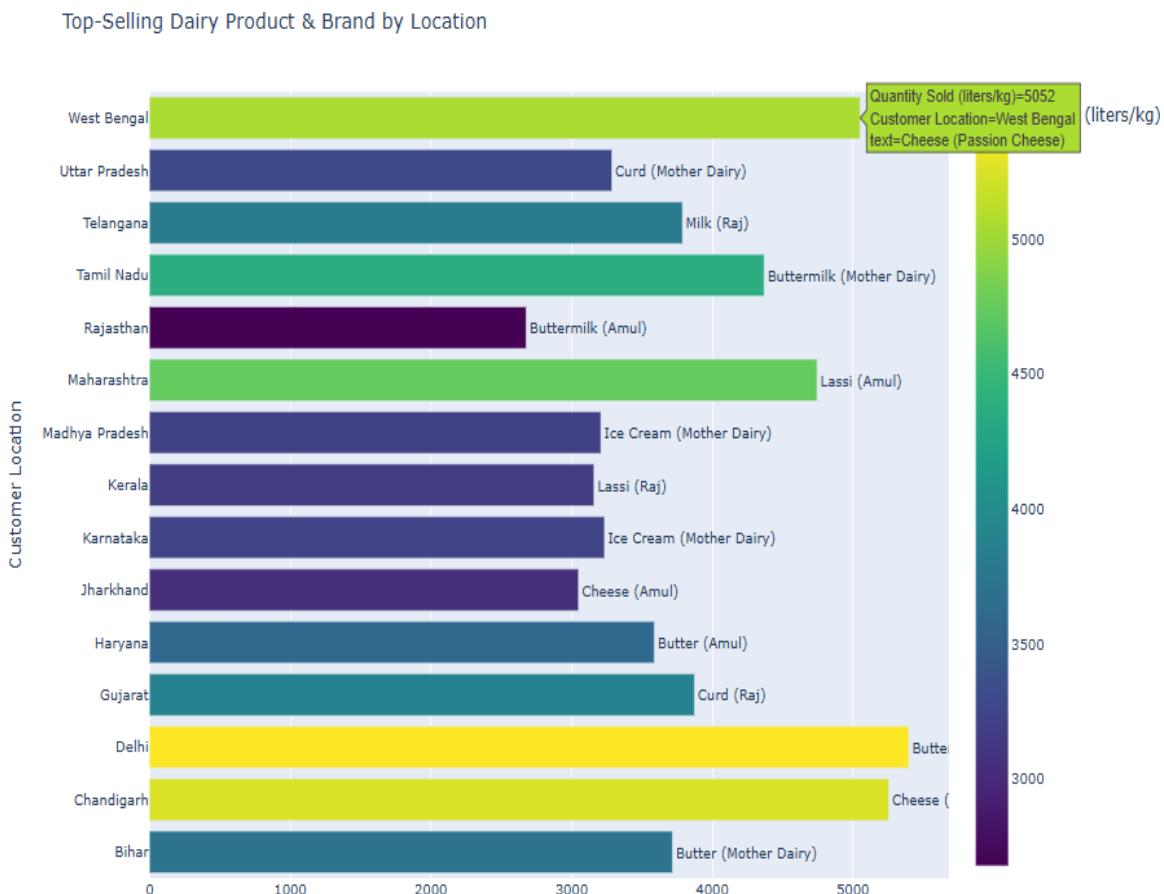


**Figure 4.4- Regional Sales Performance Quantity Sold by Location**

Figure 4.4 represents the bar chart showing the quantity of dairy products sold across different locations. The chart allows comparison of sales volumes among regions, making it easier to identify high-performing areas with strong consumer demand and underperforming regions where sales improvement strategies may be required.

## TOP SELLING PRODUCT AND BRAND BY LOCATION

Top Selling Product and Brand by Location visualization highlights the best-performing dairy products and brands in each location. It shows which items are most preferred by customers across different regions. By comparing locations, it helps identify regional favourites and key sales drivers.

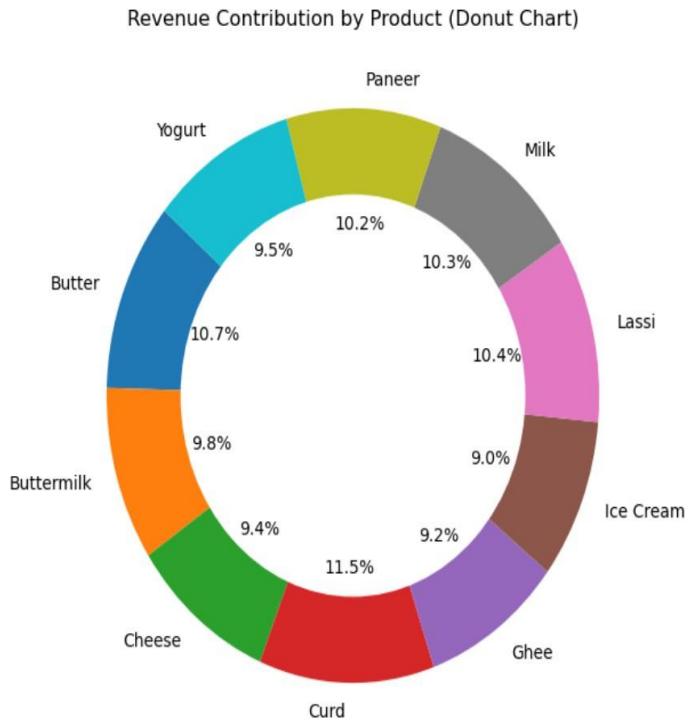


**Figure 4.5 Top Selling Product and Brand by Location**

Figure 4.5 visually represent top-selling dairy products and brands across locations. It highlights regional preferences and leading brands, helping in product planning, distribution, and marketing strategies.

## REVENUE CONTRIBUTION BY PRODUCT

Revenue contribution by product analysis illustrates the revenue generated by each dairy product, highlighting the most profitable items and supporting strategic product and sales decisions.



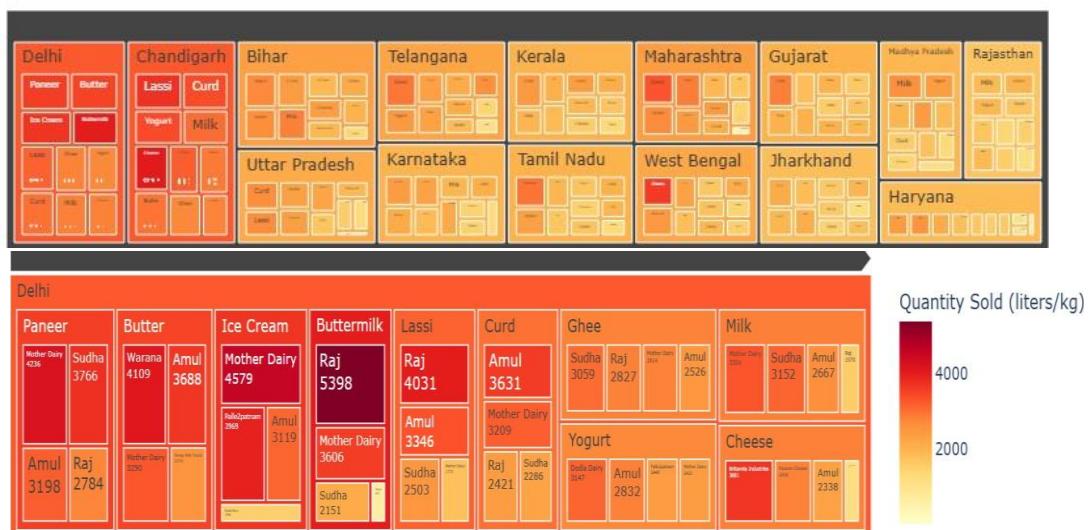
**Figure 4.6 Revenue Contribution by Product**

Figure 4.6 displays the revenue contribution of the top-selling product, Curd, which alone accounts for 11.5% of the total revenue. This significant share emphasizes the crucial role of Curd in driving overall business performance. The product's strong contribution highlights its consistent customer demand, positioning it as a key revenue generator within the product portfolio.

## REGIONAL SALES PERFORMANCE BY BRAND AND PRODUCT

The sales performance of each dairy brand and product across various locations provides a clear picture of regional demand and consumption patterns. It shows which brands and products are performing best in specific regions, highlighting regional preferences and market strengths. By comparing brand and product combinations, it helps identify areas with high sales potential, opportunities for expansion, and regions that may require targeted marketing strategies to improve performance. It supports better resource allocation by focusing on high-performing regions. It also assists in optimizing distribution channels to ensure availability of popular products. Ultimately, such insights contribute to building stronger brand positioning and improving overall market competitiveness.

Regional Sales Performance by Product and Brand (Quantity Sold)



**Figure 4.7 Regional Sales Performance by Brand and Product**

Figure 4.7 represents the sales data for each brand and product across different locations. It provides a comparative view of how products and brands perform regionally, helping to identify top-performing combinations and assess regional competitiveness in the dairy market.

## SALES CHANNEL DISTRIBUTION

This sales channel distribution analysis highlights the contribution of each sales channel to overall sales. Retail is the highest contributor, followed closely by wholesale and online channels, indicating a relatively balanced distribution across channels.

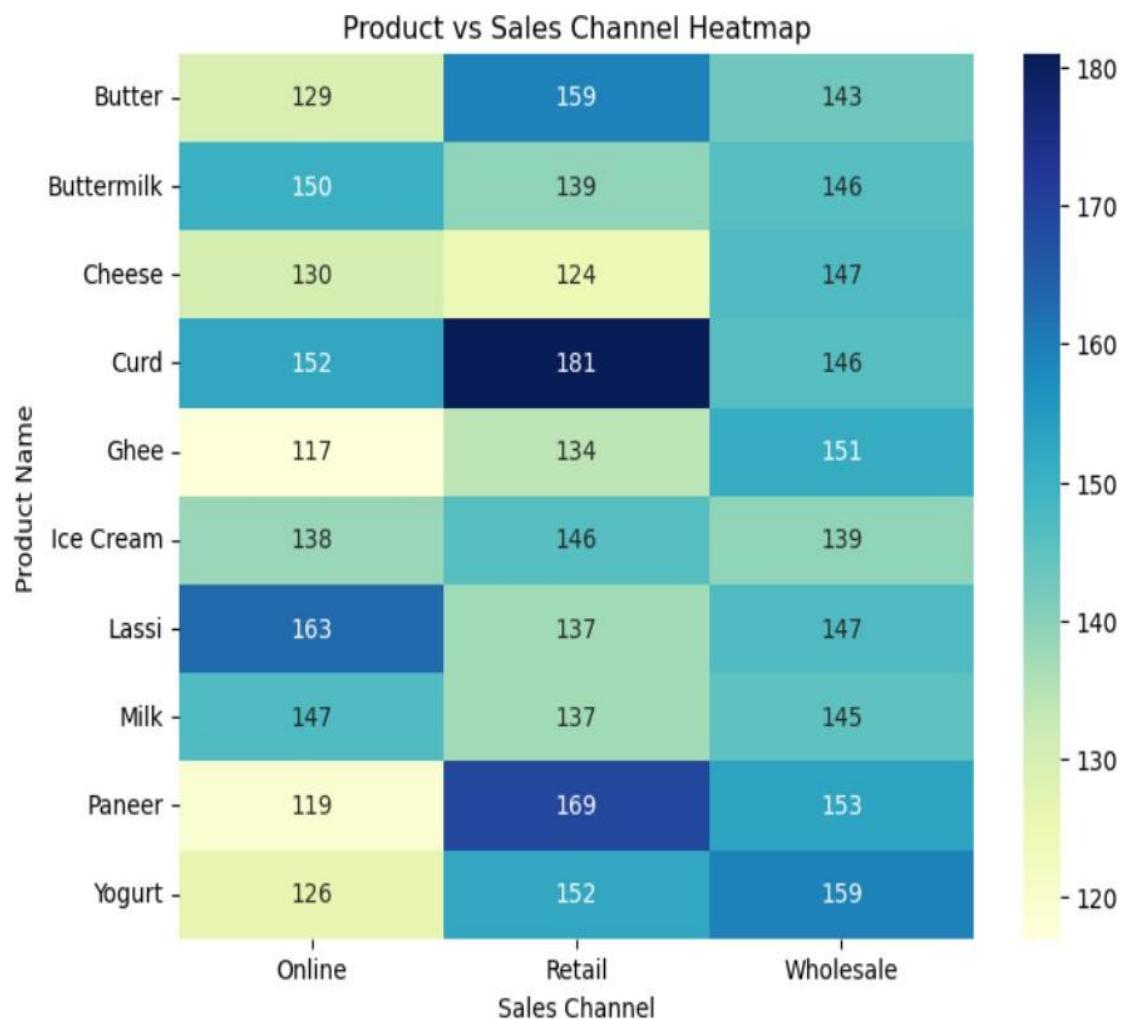


**Figure 4.8 Sales Channel Distribution**

Figure 4.8 represents the total number of dairy product sales through different channels. Retail accounts for 1,478 sales, wholesale accounts for 1,446 sales, and online sales account for 1,371 sales. This visualization highlights the contribution of each sales channel to overall dairy product distribution.

## SALES CHANNEL DISTRIBUTION BY PRODUCT

This Sales Channel Distribution by Product helps identify which products perform best in specific channels. For example, some products may sell more through retail, while others have stronger online or wholesale sales, enabling targeted channel strategies for each product.

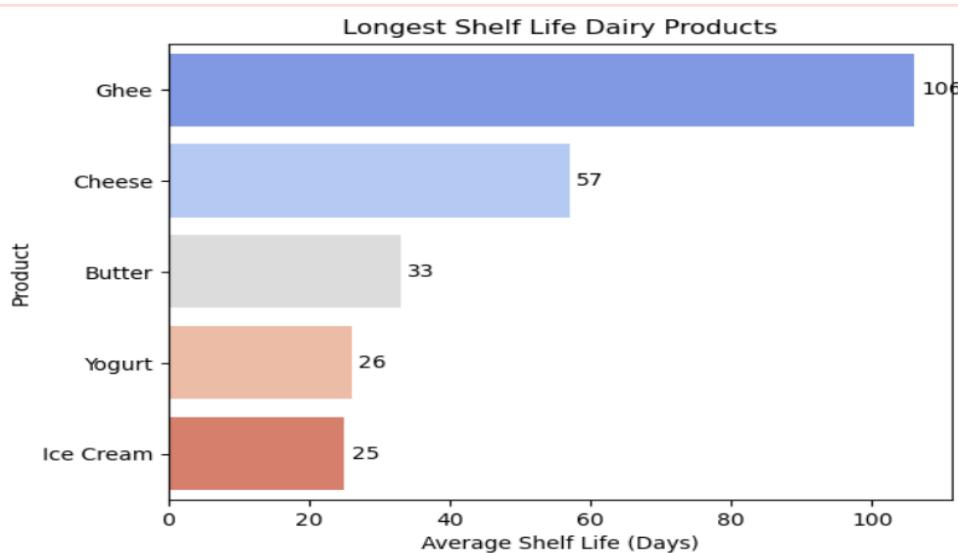


**Figure 4.9- Sales Channel Distribution by Product**

Figure 4.9 represents a heat map showing how individual dairy products are sold through retail, wholesale, and online channels. This visualization helps identify which products perform best in each channel and supports targeted channel-specific strategies to optimize overall sales performance.

## SHELF-LIFE ANALYSIS OF DAIRY PRODUCTS

Dairy products with the longest shelf life helps identify items that stay fresh longer, enabling better inventory management and reducing waste. This insight supports strategic decisions for storage, production planning, and sales prioritization.

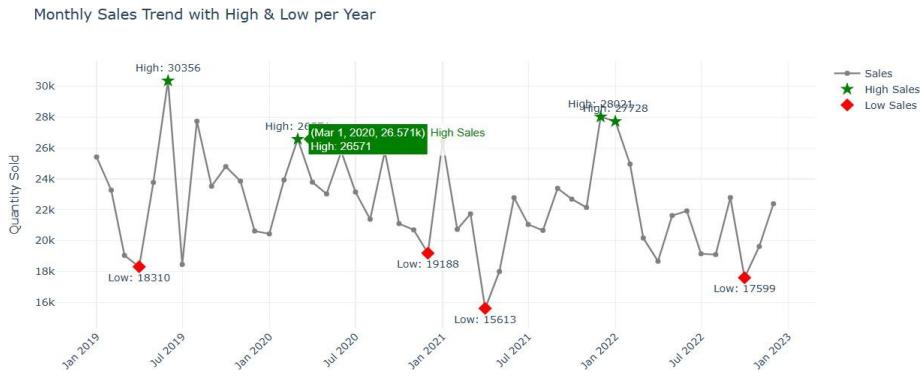


**Figure 4.10 - Shelf-Life Analysis of Dairy Products**

Figure 4.10 visually represents the top 5 dairy products with the longest shelf life. Taller bars indicate longer-lasting products, while shorter bars represent items with lower shelf longevity. This visualization highlights products that remain fresh longer, supporting inventory management, production planning, and strategies to reduce waste.

## MONTHLY DAIRY SALES TRENDS

Monthly sales trends enable better production planning, inventory management, and targeted marketing efforts. Seasonal highs guide scaling of production, while lows indicate periods that may benefit from promotions or focused sales strategies.

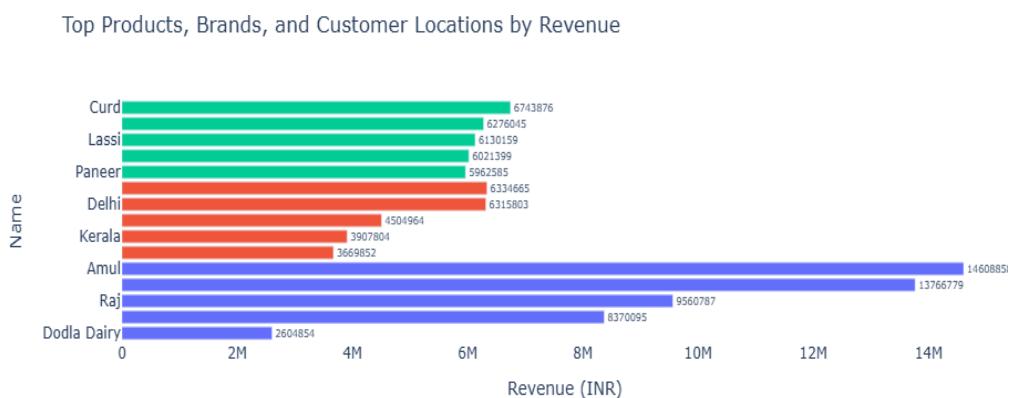


**Figure 4.11- MONTHLY DAIRY SALES TRENDS**

Figure 4.11 displays the monthly sales volumes of dairy products. Green indicates higher sales while red indicates lower sales. In the line chart, higher points represent higher sales and lower points represent lower sales. This visualization highlights seasonal peaks and fluctuations in demand, supporting informed decisions for production planning, inventory management, and targeted marketing strategies.

## TOP PRODUCTS, BRANDS, AND CUSTOMERS LOCATION BY REVENUE

Customer analysis identifies the products, brands, and customer segments that generate the most revenue. It helps the business recognize top performers, focus on high-value areas, and make informed decisions for growth and strategic planning.



**Figure 4.12 Top Products, Brands, and Customers by Revenue**

Figure 4.12 shows the top-performing dairy products, brands, and customer locations by revenue. Curd, Lassi, and Paneer lead in product sales, while Amul and Raj are the highest grossing brands. Among customer locations, Delhi and Kerala generate the most revenue. This visualization helps identify key revenue drivers and focus areas for strategic growth.

## **5. CONCLUSION**

The analysis of the dairy business provides a comprehensive view of supply, production, sales, and revenue performance. Farm and herd distribution shows that regions such as Delhi and Chandigarh have higher concentrations of farms and larger herd sizes, indicating strong production capacity. Understanding the distribution of small, medium, and large farms supports effective resource allocation, logistics planning, and potential expansion strategies.

Production and stock assessment highlights how products align with market demand. By examining production volumes, sales, and remaining stock, fast-moving products and slow-moving inventory are identified, enabling better inventory management and minimizing the risk of overproduction. Stock-to-sales ratios further help optimize inventory levels and ensure timely availability of high-demand products.

Regional sales performance and product and brand analysis provide insights into market dynamics. High-performing regions, products, and brands are identified, revealing regional preferences and competitive strengths. Cross-analysis of brand and product sales across locations informs targeted marketing strategies and supports regional growth initiatives.

Sales channel insights reveal how products perform across retail, wholesale, and online channels, guiding channel-specific strategies. Temporal trends highlight seasonal and monthly fluctuations in sales, allowing proactive planning for production, inventory, and promotional activities.

Finally, the top performing products, brands, and customer segments provides key insights into the areas driving the most revenue and business impact. By integrating supply, production, sales, and revenue data, it offers a holistic view of operational performance, regional trends, and channel effectiveness. These insights enable management to focus on high value products, optimize resource allocation, enhance inventory management, and make informed strategic decisions that improve profitability, strengthen market position, and support sustainable long-term growth.

## REFERENCES

1. Hitesh S.M. and Yukthi A., Sales Prediction Using Machine Learning Techniques, International Journal of Novel Research and Development (IJNRD), Volume 9, Issue 1, January 2023.
2. Kumar P. and Singh R., Machine Learning Applications in Retail and Dairy Sales Forecasting, ResearchGate, August 2022.
3. Pandas Documentation, Python Data Analysis Library, available at <https://pandas.pydata.org/>, accessed July 2025.
4. Hunter J. D., Matplotlib: A 2D Graphics Environment, Computing in Science & Engineering, Volume 9, Issue 3, May 2007.
5. Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., et al., Scikit-learn: Machine Learning in Python, Journal of Machine Learning Research, Volume 12, 2011.
6. Rasheed A., Data-Driven Approaches for Dairy Production and Sales Analytics, International Journal of Agriculture and Food Science, Volume 8, Issue 2, March 2022.
7. Giordani M., Polese M., Roy A., Castor D., and Zorzi M., AI and IoT Applications in Supply Chain Monitoring, IEEE Communications Magazine, Volume 58, Issue 3, March 2020.
8. Sharma R. and Mehta V., Real-Time Data Analytics in Food Supply Chains, International Journal of Computer Applications (IJCA), Volume 184, Issue 42, April 2022.
9. Xu J., Fang J., and Lin W., AI-Driven Monitoring and Forecasting in Agriculture and Dairy Systems, National Center for Biotechnology Information (NCBI), February 2021.
10. Kaggle Datasets, Dairy Production and Sales Data, available at <https://www.kaggle.com/>, accessed July 2025.

## APPENDIX

```
# Import Libraries

!pip install squarify

import pandas as pd, seaborn as sns, matplotlib.pyplot as plt, squarify

import plotly.express as px, plotly.graph_objects as go

from plotly.offline import init_notebook_mode

init_notebook_mode(connected=True)

# Load Data

df = pd.read_csv(r"C:\Users\vivek\Downloads\dairy_dataset (2).csv")

df['Date'] = pd.to_datetime(df['Date'])

df['YearMonth'] = df['Date'].dt.to_period('M').astype(str)

# 1. Farms by Location

loc_counts = df['Location'].value_counts().reset_index()

loc_counts.columns = ["Location", "Count"]

fig = px.bar(loc_counts, x="Count", y="Location", orientation="h", text="Count",

title="Number of Farms by Location", color="Location")

fig.update_traces(textposition="outside"); fig.show()

# 2. Cows by Farm Size & Location

farm_loc = df.groupby(["Location", "Farm Size"])["Number of Cows"].sum().reset_index()

fig = px.bar(farm_loc, x="Location", y="Number of Cows", color="Farm Size",

barmode="stack", text="Number of Cows",

title="Cows by Farm Size across Locations")

fig.update_traces(textposition="outside"); fig.show()
```

```
# 3. Dairy Sales by Region
```

```
region_sales=df.groupby("CustomerLocation")["QuantitySold  
(liters/kg)"].sum().reset_index()  
  
sns.barplot(x="Customer Location", y="Quantity Sold (liters/kg)", data=region_sales,  
palette="magma")  
  
plt.title("Regional Sales Performance"); plt.xticks(rotation=45); plt.show()
```

```
# 4. Quantity Produced, Sold, Stock (Table)
```

```
qty = df.groupby("Product Name")[["Quantity (liters/kg)",  
"Quantity Sold (liters/kg)","Quantity in Stock (liters/kg)"]].sum().reset_index()  
  
fig = go.Figure(data=[go.Table(  
  
header=dict(values=list(qty.columns), fill_color="blue", font=dict(color="white")),  
  
cells=dict(values=[qty[c] for c in qty.columns]))])  
  
fig.update_layout(title="Production, Sales & Stock by Product"); fig.show()
```

```
# 5. Top-Selling Product & Brand per Location
```

```
top_selling = (df.groupby(["Customer Location","Product Name","Brand"])["Quantity Sold  
(liters/kg)"]  
  
.sum().reset_index().sort_values(["CustomerLocation","QuantitySold(liters/kg)"],ascending=  
[True,False]).groupby("Customer Location").first().reset_index())  
  
fig = px.bar(top_selling, x="Quantity Sold (liters/kg)", y="Customer Location",  
orientation="h",text=top_selling["Product Name"]+" ("+top_selling["Brand"]+")",  
color="Quantity Sold (liters/kg)", color_continuous_scale="Blues",  
title="Top-Selling Product & Brand by Location")  
  
fig.update_traces(textposition="outside");  
  
fig.show()
```

```
# 6. Revenue Contribution by Product
```

```
prod_rev = df.groupby("Product Name")["Approx. Total Revenue(INR)"].sum().reset_index()
```

```

fig = px.pie(prod_rev, names="Product Name", values="Approx. Total Revenue(INR)",
             title="Revenue Contribution by Product")
fig.show()

# 7. Regional Sales by Product & Brand (Treemap)

loc_sales = df.groupby(["Customer Location","Product Name","Brand"])["Quantity Sold
(liters/kg)"].sum().reset_index()

fig = px.treemap(loc_sales, path=["Customer Location","Product Name","Brand"],
                  values="Quantity Sold (liters/kg)", color="Quantity Sold (liters/kg)",
                  color_continuous_scale="YlOrRd",
                  title="Regional Sales by Product & Brand")
fig.show()

# 8. Sales Channel Distribution

channel_counts = df['Sales Channel'].value_counts()
squarify.plot(sizes=channel_counts.values,
              label=[f'{c}\n{v}' for c,v in zip(channel_counts.index, channel_counts.values)],
              alpha=.8); plt.title("Sales Channels Distribution"); plt.axis("off"); plt.show()

# 9. Product vs Sales Channel (Heatmap)

cross = pd.crosstab(df["Product Name"], df["Sales Channel"])

sns.heatmap(cross, annot=True, cmap="YlGnBu", fmt="d")
plt.title("Product vs Sales Channel"); plt.show()

# 10. Monthly Sales Trend (High & Low)

monthly = df.groupby("YearMonth")["Quantity Sold (liters/kg)"].sum().reset_index()

high, low = monthly["Quantity Sold (liters/kg)"].max(), monthly["Quantity Sold
(liters/kg)"].min()

fig = go.Figure()

```

```

fig.add_trace(go.Scatter(x=monthly["YearMonth"], y=monthly["Quantity Sold (liters/kg)"],
mode="lines+markers", name="Sales"))

fig.add_trace(go.Scatter(x=monthly[monthly["QuantitySold(liters/kg)"]==high]["YearMont",
y=[high], mode="markers+text", marker=dict(color="green", size=12),
text=[f"High: {high}"], textposition="top center"))

fig.add_trace(go.Scatter(x=monthly[monthly["QuantitySold(liters/kg)"]==low]["YearMonth"],
y=[low], mode="markers+text", marker=dict(color="red", size=12),
text=[f"Low: {low}"], textposition="bottom center"))

fig.update_layout(title="Monthly Dairy Sales Trend", xaxis_tickangle=-45);
fig.show()

```

#### # 11. Top Products, Brands & Locations by Revenue

```

top_prod=df.groupby("ProductName")["Approx.Total
Revenue(INR)"].sum().nlargest(5).reset_index()

top_brand=df.groupby("Brand")["Approx.TotalRevenue(INR)"].sum().nlargest(5).reset_inx()
top_loc=df.groupby("CustomerLocation")["Approx.Total
Revenue(INR)"].sum().nlargest(5).reset_index()

for data,title in [(top_prod,"Top Products"),(top_brand,"TopBrands"),(top_loc,"Top
Locations")]:


sns.barplot(x="Approx. Total Revenue(INR)", y=data.columns[0], data=data,
palette="viridis")plt.title(title+" by Revenue");

plt.show()

```

#### # 12. Longest Shelf Life Products

```

shelf = df.groupby("Product Name")["Shelf Life (days)"].mean().nlargest(5).astype(int)

sns.barplot(y=shelf.index, x=shelf.values, palette="coolwarm")

plt.title("Longest Shelf Life Dairy Products"); plt.show()

```

