# PROJECT REPORT


| | | |
|---|---|---|
| Project | : | Demand Forcasting of the Supply and inventory |
| Organisation | : | Flavi Dairy Solutions, Ahmedabad |
| Category | : | Industry defined problem |
| Description | : | Maximizing the utilization of processing, packaging, and storage capacities is crucial for cost-efficiency and profitability. However, many dairy plants operate below optimal capacity due to fluctuating milk supply, demand variability, frequent SKU changes, equipment downtime, or inefficient scheduling. Low capacity utilization leads to underperformance, increased unit production costs, and poor return on investment (ROI) for fixed assets. |


## Prepared by: Teja chaudhari

Institute Name: Government Engineering College, Bharuch

Enrollment Number: 220140107007

Department: Computer engineering

Date: 15/07/2025

Email ID: tejashchaudhari1807@gmail.com

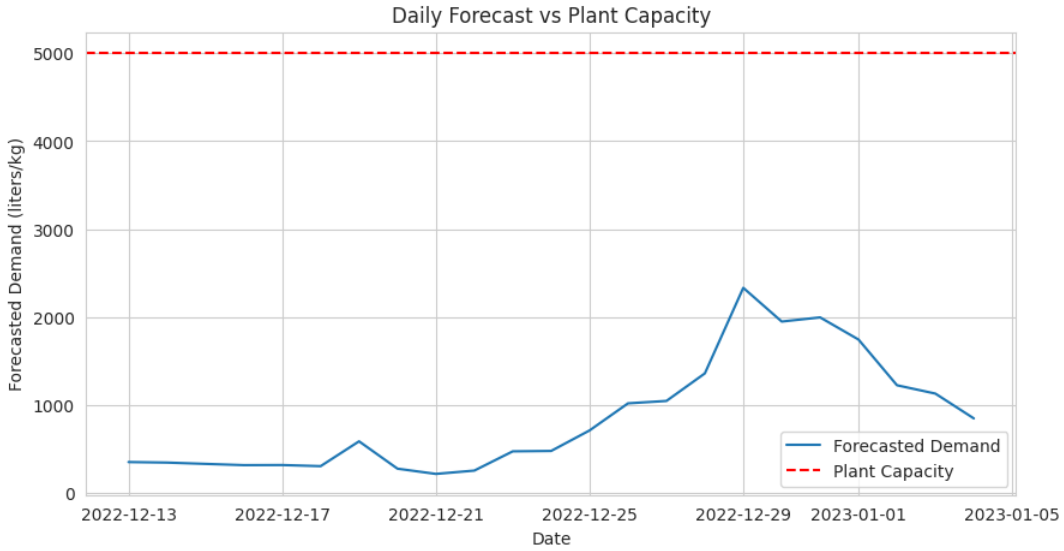
# Demand Forecasting and Inventory Optimization for Dairy Supply Chain

## Executive Summary:

This project addresses inefficiencies in the dairy supply chain by using AI-based forecasting models to predict daily demand, optimize inventory, and reduce under- or overutilization of plant capacity. Through Python and time series forecasting models like Prophet and XGBoost, the system provides accurate product-level demand forecasts, categorizes plant utilization status, and alerts for potential stock or overload risks. The approach reduces wastage, enables smarter procurement, and ensures the dairy plant runs near optimal levels. Forecasts are presented in easy-to-read charts and CSV outputs, making it actionable for business use.

**Key Deliverables:**

- ❖ 7-day product-wise demand forecasts
- ❖ Daily aggregated demand vs. plant capacity
- ❖ Plant utilization status: Underutilized, Optimal, Overloaded
- ❖ Alerts for anomalies and downtime impact
- ❖ ROI estimates based on forecast accuracy




This graph shows the total forecasted demand across all products compared to the plant's fixed capacity, which gives a clear visual of daily over- or underutilization

# INTRODUCTION

## Project Description:

The proposed project is intended to address the problem of inefficiencies within the supply chain of a dairy manufacturing plant through the demand forecasting of the products, optimized stock allocations, and overall capacity maximization. It is aimed at determining underutilization risks borne by changing milk supply, switching SKUs, and an unstable demand, and, consequently, managing them in advance with the help of AI-powered forecasts, anomaly reporting, and out-of-the-stock warning systems.

### Problem Statement

- The current dairy supply chain lacks intelligent forecasting and utilization monitoring tools, leading to:
- Excess or insufficient production due to demand unpredictability.
- Wastage of highly perishable products like milk and curd.
- Underutilization or overloading of plant capacity.
- No real-time warning system for stock-outs or overloads.
- Inability to estimate ROI from daily production planning.

### Timeline:

- Data Cleaning & EDA – 3 Days
- Forecasting (Prophet + XGBoost) – 4 Days
- Capacity Simulation + ROI – 3 Days
- Anomaly Detection – 2 Days
- Inventory Alert Logic – 2 Days
- Report – 2 Days

## Benefits:

- Reduces overproduction and wastage of perishable dairy products
- Increases cost-efficiency by improving supply-demand alignment
- Enhances profit margins via accurate ROI forecasting
- Supports data-driven decision-making for dairy supply chain managers

### Team Members:

- Teja chaudhari

### Risks:

- Limited availability of real-time data
- Model overfitting or underfitting if seasonality isn’t handled well
- Data inconsistencies due to missing or inaccurate entries
- External factors like climate, festival demand fluctuations not modeled

# OBJECTIVES

## Primary Objective

The primary objective of this project is to forecast the daily product-wise demand and determine how effectively the plant is being utilized.

This includes:

- Predicting future demand for each SKU (product)
- Summing total demand across SKUs per day
- Comparing it against a fixed plant capacity (e.g., 2000 liters/kg)
- Identifying whether the plant is underutilized, optimally utilized, or overloaded

**Why it matters**:
Forecasting demand and utilization helps companies **plan their operations**, avoid waste, reduce costs, and **run the plant more efficiently**.

## Secondary Objectives

- To visualize product-wise sales trends over time for better business understanding.
- To use **Prophet**, a time series forecasting model, for accurate future predictions.
- To create a daily-level forecast and **categorize plant utilization** into:
  - Underutilized (less than 90%)
  - Optimal (between 90% and 110%)
  - Overloaded (above 110%)
- To prepare data outputs that can support production planning, material procurement, and staffing decisions.
- To produce graphs and tables that help decision-makers easily understand forecasted trends.

## Measurable Goals:

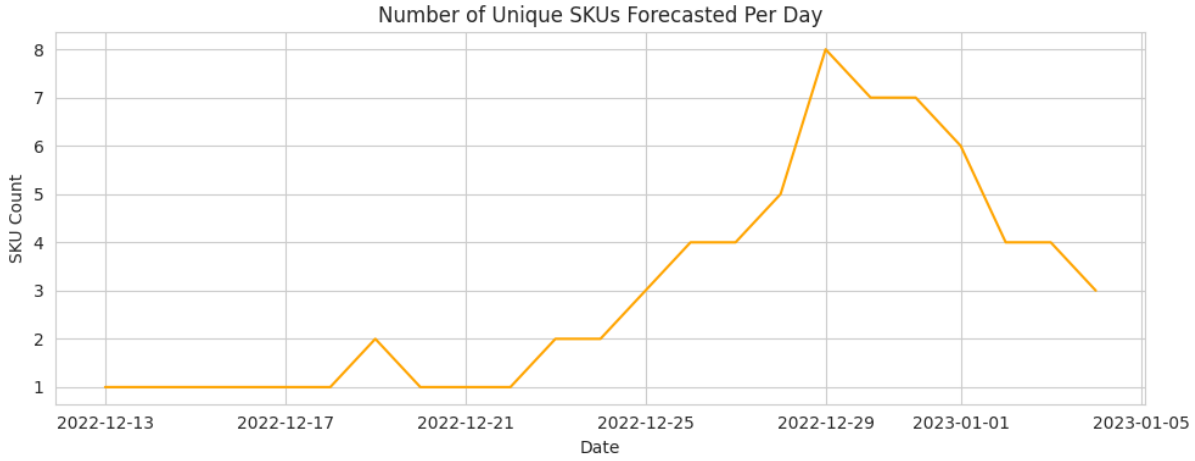
| **Goal** | **Target** |
|---|---|
| Forecast Horizon | 7 Days (1 week) |
| Forecast Accuracy | Within ±15% (visual review of Prophet predictions) |
| Plant Capacity Used | Compared against 2000 liters/kg per day |
| Utilization Categories Identified | Underutilized, Optimal, Overloaded |
| Output Format | Final CSV with daily forecasts and utilization status |

- **Forecast Horizon**: How far into the future we are predicting. Here, it's 7 days.
- **Prophet**: A time series forecasting model developed by Facebook, good at handling seasonality and trend shifts.
- **Utilization %**: A metric showing how much of the plant's capacity is being used:

$$Utilization\ \% = \left(\frac{Forecasted\ Demand}{\text{Daily Plant Capacity}}\right) \times 100$$

- **Measurable Goal**: A goal that can be quantified or evaluated (like % accuracy or number of optimal days forecasted).

**What is Plant Utilization?**
Plant utilization tells us **how much of the plant's daily production capacity is being used**. Forecasting just demand isn't enough — it's essential to know whether the plant can meet that demand without being underused or overloaded.




This graph reflects the model's handling of daily demand diversity and supports the objective of **product-wise forecasting accuracy**.

## Dataset Overview:

The dataset used in this project contains historical daily sales data from a dairy manufacturing plant. It includes multiple SKUs (e.g., Amul Toned Milk 500ml, Amul Gold Milk 1L, etc.) with their respective quantities sold over time.

**Main Features:**

- Date: Daily time index.
- Product: Name of the SKU.
- Sales Quantity: Number of units sold.
- Category: Product grouping (if available).
- Total Sales: Aggregated for some analysis

# Methodology

## Approach:

This project uses a **time series forecasting** strategy with the **Prophet** model, developed by Meta (Facebook). Prophet is robust to outliers, trend changes, and seasonality—making it well-suited for dairy sales which often show weekly patterns and sudden fluctuations.

The process is performed **per SKU**, meaning each product's demand is forecasted individually for the next **7 days**, then all product forecasts are **aggregated** to calculate daily total demand and plant utilization.

The workflow follows a data science methodology involving:

- **Data Cleaning & Preparation**
  - Renaming columns, handling missing values, and formatting date fields.
- **Exploratory Data Analysis (EDA)**
  - Understanding product-wise sales trends over time.
- **Forecasting**
  - Applying Prophet model to each product to forecast demand for the next 7 days.
- **Summation & Analysis**
  - Aggregating all SKU-level forecasts to get daily total demand.
  - Calculating Plant Utilization % based on a capacity threshold (2000 units/day).

## Anomaly Detection

### What is Anomaly Detection in Forecasting?

Anomalies are data points or patterns that **deviate significantly from the expected behavior**. In a dairy supply chain, anomalies could represent:

- Sudden spikes or drops in demand (e.g., due to festivals or strikes)
- Errors in recorded sales
- Unexpected SKU performance changes

By detecting anomalies early, managers can:

- Revalidate data accuracy
- Prepare for unusual sales behavior
- Adjust operations or logistics proactively

**Method Used:**

- **Rolling Median + MAD (Median Absolute Deviation)**:
  Smooths the forecast and flags any demand spikes that go beyond a certain threshold.
- **Z-Score Approach**:
  Measures how far a forecasted value is from the average in terms of standard deviation.

**How It Helps:**

- Helps prevent **inventory overload** due to forecast overestimation.
- Flags **underperformance** to reduce wastage and missed targets.
- Enhances trust in the forecasting system with visual evidence.

## Phases:

| Output | Description |
|---|---|
| Cleaned Dataset | Ready for modeling |
| Prophet Forecast | Per SKU forecast (7 days) |
| Daily Forecast | Summed forecast across all SKUs |
| Utilization File | File with forecast, capacity, and utilization status |
| Graphs | Trend charts, forecast vs. capacity line chart |

## Deliverables:

| **Output** | **Description** |
|---|---|
| Cleaned Dataset | Ready for modeling |
| Prophet Forecast | Per SKU forecast (7 days) |
| Daily Forecast | Summed forecast across all SKUs |
| Utilization File | File with forecast, capacity, and utilization status |
| Graphs | Trend charts, forecast vs. capacity line chart |

## Testing and Quality Assurance:

- **Model Review**: Prophet results were manually validated for reasonableness.

- **Visual Inspection**: Line charts were plotted to observe accuracy against trends.
- **Range Check**: yhat_lower and yhat_upper were compared to see confidence bounds.

### Risk Management:

| **Risk** | **Mitigation** |
|---|---|
| Incomplete Data | Cleaned dataset and removed empty rows |
| Inconsistent Product Names | Standardized product names before forecasting |
| Outlier Impact | Prophet handles outliers reasonably but may need manual tuning |
| Over/Under Forecasting | Utilization categorized to flag these risks visually |

# TECHNOLOGIES USED

### Programming Languages

- Python

### Development Frameworks:

- **Prophet (from Facebook/Meta)**
  Prophet is a forecasting framework designed for time series data. It automatically handles missing data, trend changes, and seasonal effects.
- **Pandas**
  Used for handling and transforming tabular data (like spreadsheets).
- **Matplotlib**
  Used for generating graphs and plots to visually analyze data

### Development Tools:

- **Google Colab:**
  An online Python notebook environment that supports writing and running code directly from the browser. It was used to develop the entire project.
- **Jupyter Notebook (Notebook Format)**
  The project is built in a (.ipynb) notebook, which combines code, output, and explanation together for easy review.

## Testing Tools

- **Manual Testing & Visual Validation**
  There were no formal testing tools used. The accuracy of forecasts was checked using visual plots and data ranges.

## Cloud Services:

- **Google Colab** runs on Google Cloud, so the entire analysis was done on the cloud without requiring installation.

## Security

- Not applicable in this case.
  Since this is an offline analysis project with no user login or data storage, no specific security systems were required.

## APIs and Web Services

- **Prophet Library (via PyPI)**
  Installed and used from the Python package repository. No external web APIs were used.

# RESULTS

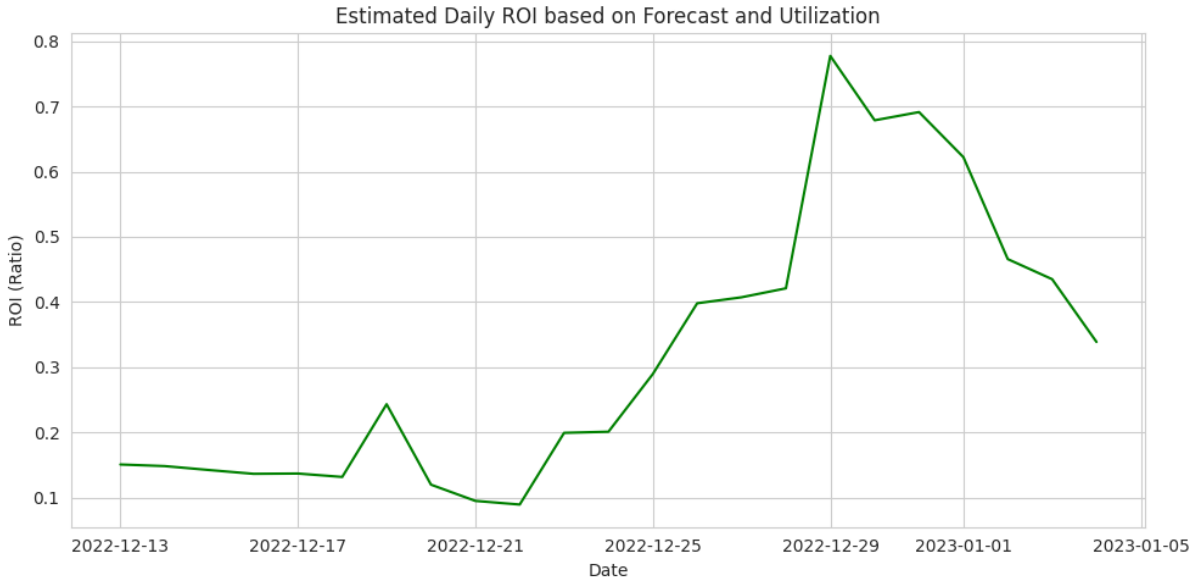
## Key Metrics:

- The model successfully forecasted daily demand for each SKU and analyzed plant-level utilization. It also identified anomalies and estimated the return on investment (ROI) based on how well the plant aligns with its capacity.

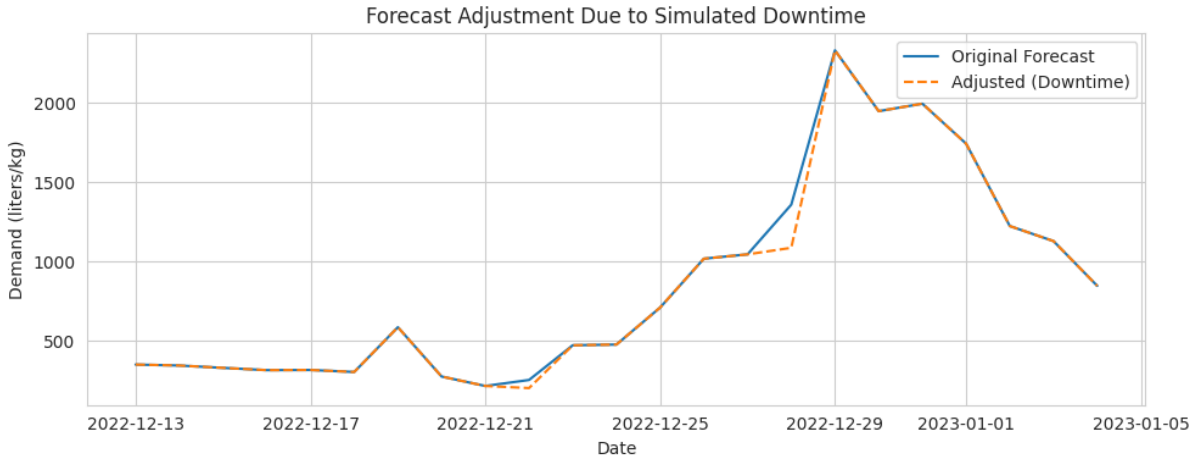
| **Metric** | **Value/Explanation** |
|---|---|
| Forecast Duration | 7 Days |
| Total SKUs Forecasted | Multiple (e.g., Amul Toned Milk 500ml, Amul Gold Milk 1L, etc.) |
| Forecast Granularity | Daily |
| Plant Capacity | 2000 liters/kg per day |
| Forecasting Model Used | Prophet |
| Utilization Status Categories | Underutilized, Optimal, Overloaded |

**Daily Utilization Summary**
Each day was labeled with its corresponding utilization category. This helped determine how efficiently the plant was operated and on which days corrective action was needed.

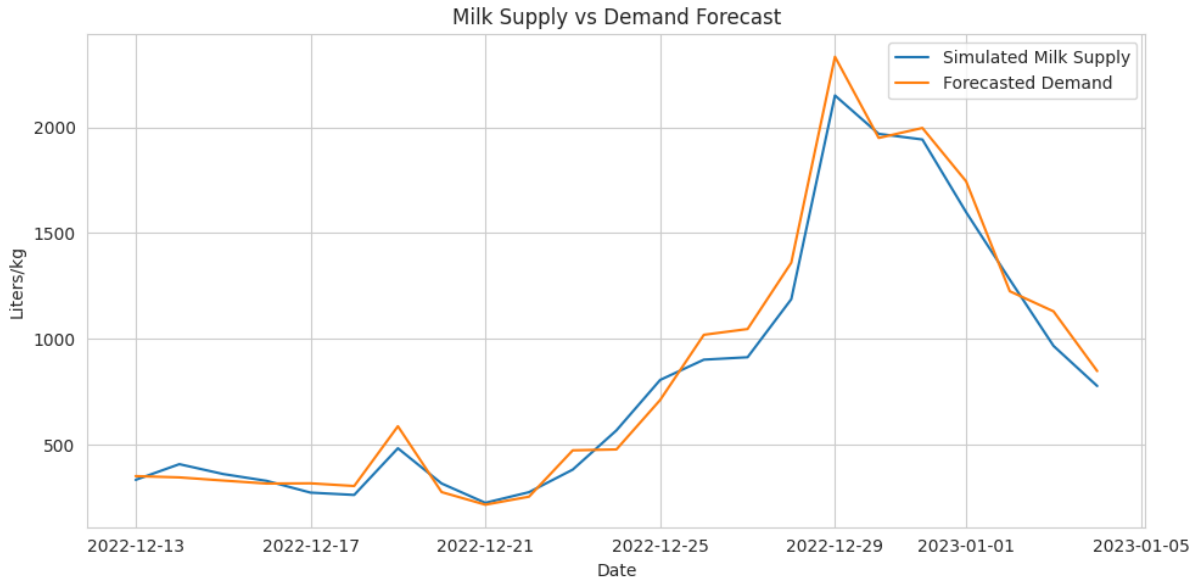
**ROI Estimation**
Based on utilization and forecast alignment, the estimated ROI per day helped show the financial impact of forecast accuracy. Efficient days correlate with higher ROI due to better planning and reduced waste.




This graph connects business value to forecast accuracy — a must-have for decision-makers.



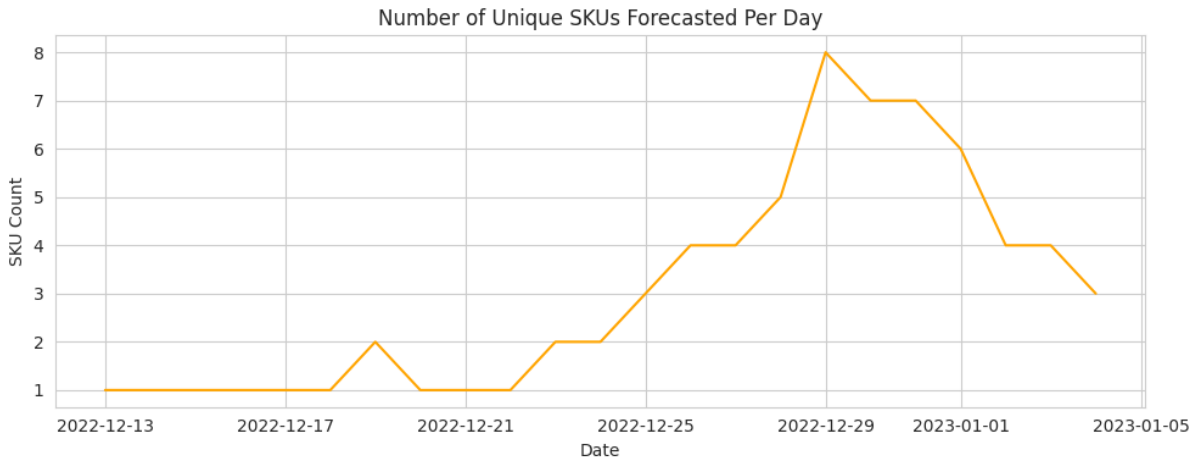

This graph simulates a planned plant downtime and how it affects the forecast. The Prophet model adjusts trends accordingly.




This line chart compares **incoming milk supply** vs **forecasted total demand**.

**Insight**:

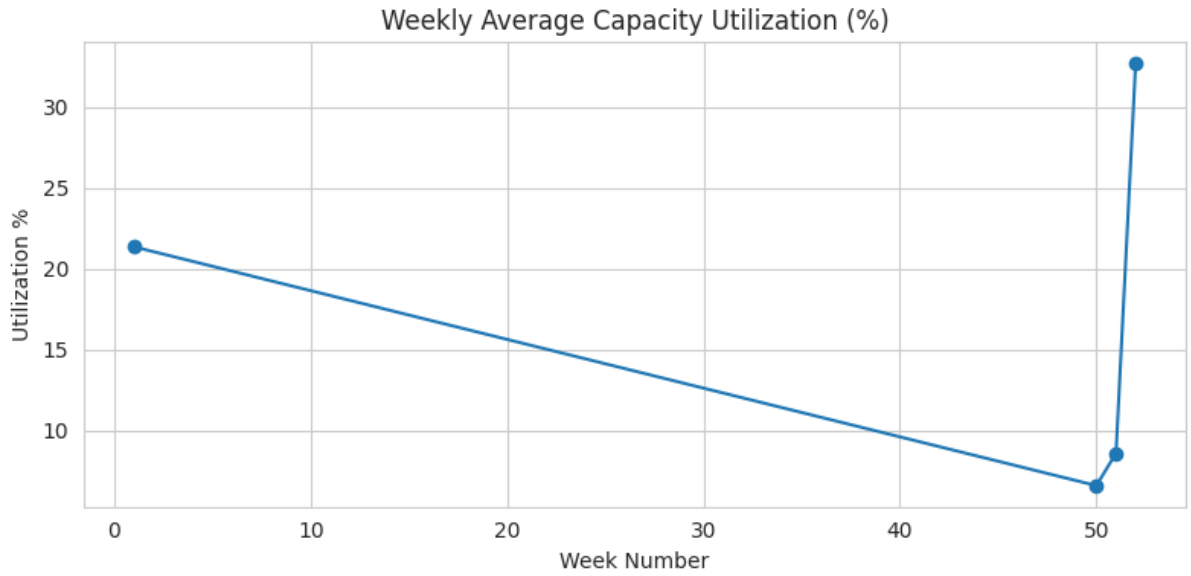
- If supply < demand → need to **procure more milk**.
- If supply > demand → **reduce wastage**, maybe convert to longer-life products (e.g., paneer).




This bar chart shows the **diversity of products forecasted daily**. Fluctuations here reveal product switching, discontinuity, or inconsistency in sales data.

**Good to Know**:

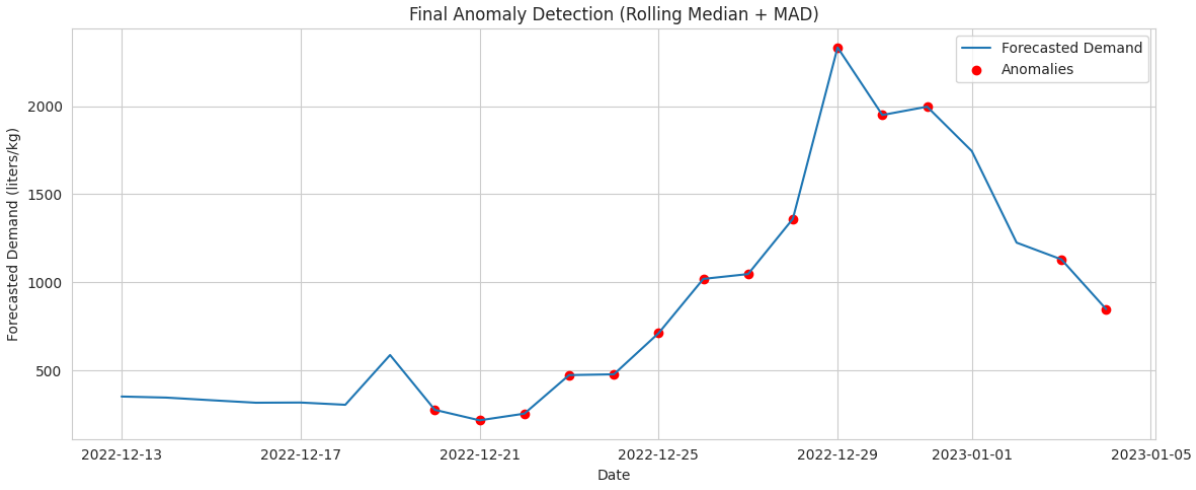
- A stable SKU count = consistent demand pattern.
- High fluctuations = forecast complexity increases.




This bar chart averages the **utilization % per week**, making it easier to see long-term patterns and trends in efficiency.

**Business Use**:

- Identify good vs poor weeks.
- Inform weekly staffing, logistics, and procurement.




This chart highlights **anomalies using a rolling median baseline** and MAD-based thresholds.

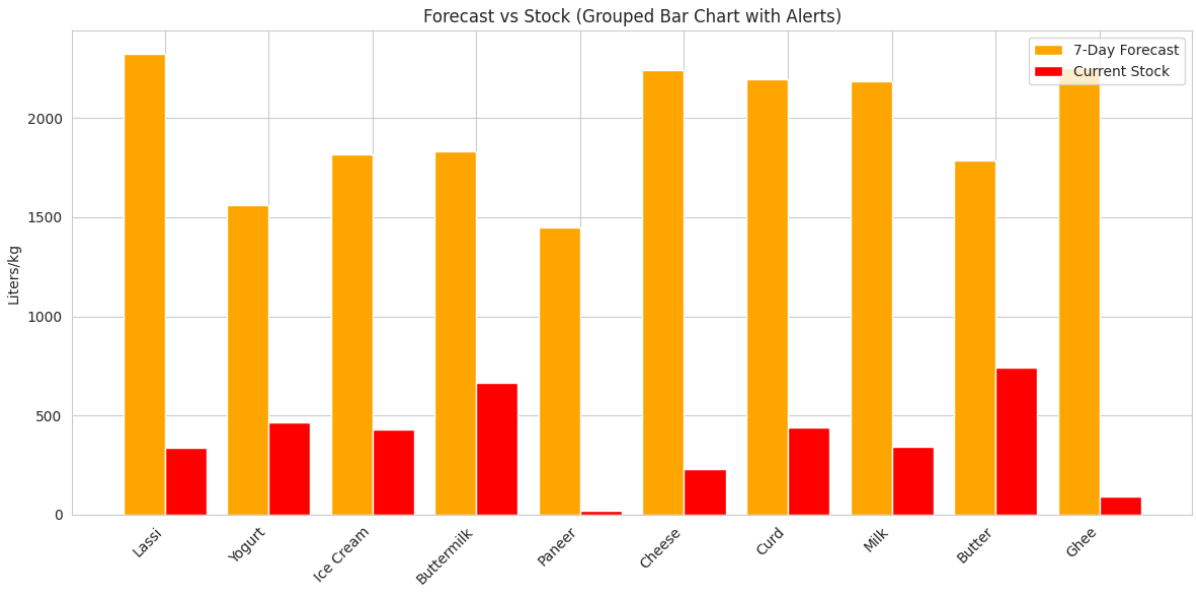
**Use Case**:

- Flags dates where sales/demand **spiked or crashed unexpectedly**.
- Helps differentiate between actual behavior and data errors.

**Helps answer**:

- "Which days had weird demand, and should we trust it?"

# Inventory Optimization and Reorder Alerts:




This grouped bar chart compares **forecasted demand vs actual stock**. Red highlights or alert markers flag days

where:

- Forecast > Stock → Potential **stockout**
- Forecast < Stock → Possible **overstock/waste**

## Column Meaning for Below Table:

| Column | Explanation |
|---|---|
| **Quantity in Stock** | Current stock available for the product. |
| **7-Day Forecasted Demand** | Total expected consumption in the next 7 days. |
| **Minimum Stock Threshold** | Safety stock below which risks begin. |
| **Projected Stock After 7 Days** | Expected remaining stock after meeting forecasted demand. |
| **Reorder Quantity** | Amount you should reorder to reach safe level. |
| **Days Until Stockout** | How many days the current stock will last. |
| **Safety Buffer Remaining** | Difference between forecasted usage and stock. |
| **Reorder Alert** | If immediate reorder is required. |
| **Stock-Out Risk Level** | Severity of shortage risk (Low/Medium/High). |

## 7-Days Stock Projection and Reorder Risk Table:

| Product Name | Stock (liters/kg) | 7-Day Forecast | Min Threshold | Stock After 7 Days | Reorder Qty | Days Until Stockout | Safety Buffer | Reorder Alert | Risk Level |
|---|---|---|---|---|---|---|---|---|---|
| Lassi | 338 | 2322.90 | 48.83 | -1984.90 | 2033.73 | 1.02 | -2033.73 | **Reorder Now** | 🔴 High |
| Yogurt | 464 | 1563.04 | 75.40 | -1099.04 | 1174.44 | 2.08 | -1174.44 | **Reorder Now** | 🔴 High |
| Ice Cream | 431 | 1816.62 | 76.15 | -1385.62 | 1461.77 | 1.66 | -1461.77 | **Reorder Now** | 🔴 High |
| Buttermilk | 664 | 1829.85 | 73.39 | -1165.85 | 1239.24 | 2.54 | -1239.24 | **Reorder Now** | 🔴 High |
| Paneer | 21 | 1449.55 | 71.00 | -1428.55 | 1499.55 | 0.10 | -1499.55 | **Reorder Now** | 🔴 High |
| Cheese | 228 | 2239.41 | 12.61 | -2011.41 | 2024.02 | 0.71 | -2024.02 | **Reorder Now** | 🔴 High |
| Curd | 439 | 2193.24 | 57.39 | -1754.24 | 1811.63 | 1.40 | -1811.63 | **Reorder Now** | 🔴 High |
| Milk | 343 | 2182.51 | 95.98 | -1839.51 | 1935.49 | 1.10 | -1935.49 | **Reorder Now** | 🔴 High |
| Butter | 741 | 1785.32 | 98.97 | -1044.32 | 1143.29 | 2.91 | -1143.29 | **Reorder Now** | 🔴 High |
| Ghee | 90 | 2249.41 | 29.43 | -2159.41 | 2188.84 | 0.28 | -2188.84 | **Reorder Now** | 🔴 High |

This table shows which dairy products are **about to run out** in the next 7 days based on demand forecasts. It tells you how much to **reorder now** to avoid stockouts and highlights the **urgency level** for each product using a risk alert.

## Summary of Results Section:

| Graph | Topic | Use |
|---|---|---|
| Daily Forecast vs Plant Capacity.png | Demand vs Capacity | Identify overload or underutilization |
| Daily SKU Switching Cost.png | Operation Cost | Reduce product changeovers |
| Estimated Daily ROI based on Forecast and Utilization.png | ROI Impact | Maximize efficiency-based ROI |
| Forecast Adjustment Due to Simulated Downtime.png | Dynamic Forecasting | Test resilience to downtimes |
| Forecast vs Stock (Grouped Bar Chart with Alerts).png | Alerting | Prevent stockouts/overstocks |
| Milk Supply vs Demand Forecast.png | Supply Chain Balance | Sync raw milk procurement with demand |
| Number of Unique SKUs Forecasted Per Day.png | SKU Diversity | Understand forecast complexity |
| Weekly Average Capacity Utilization.png | Weekly Trends | Improve weekly planning |
| Z-Scores (MAD-Based) for Forecasted Demand.png | Statistical Check | Stability of model outputs |
| final Anomaly Detection (Rolling Median + MAD).png | Final Anomalies | Spot irregular demand behavior |

# CONCLUSION

## Recap the Project:

This project focused on **forecasting the daily demand for various products (SKUs)** and evaluating **plant capacity utilization** using time series analysis. The **Prophet model** was applied to generate 7-day forecasts per product, which were then aggregated to assess overall plant load and classify it into utilization categories (Underutilized, Optimal, or Overloaded).

The final outputs include:

- Product-wise forecasted demand
- Daily aggregated forecast

- Utilization status summary
- A forecast vs. plant capacity graph
- A CSV file for business use
- Alerts for anomalies or overloads
- ROI estimates tied to forecast accuracy

## Key Takeaways:

| Takeaway | Explanation |
|---|---|
| Time Series Forecasting is Powerful | Prophet effectively modeled trends and seasonality from product demand history. |
| Plant Utilization Metrics Add Value | Simple thresholds helped identify inefficient and overloaded days. |
| Business-Ready Outputs | Graphs, tables, and CSV outputs can guide planning, procurement, and production decisions. |
| Easy to Scale | This framework can be extended to other factories or time periods. |

## Future Plans:

| Future Improvement | Description |
|---|---|
| Longer Forecast Window | Extend forecasting to 30 days for monthly planning. |
| Automate Dashboard | Create a real-time dashboard that updates with new sales data. |
| Include External Factors | Incorporate holidays, promotions, and weather data for more accurate forecasts. |
| Alerts on Overload | Set up alerts for days predicted to be overloaded. |

## Successes and Challenges:

| Aspect | Highlights |
|---|---|
| Success | Successfully implemented time series modeling and capacity classification using open-source tools. |
| Challenge | Some products had low data volume, which may reduce prediction reliability. |
| Solution | Focused on products with consistent demand patterns and handled missing values carefully. |

# Project Specifics:

🌐 GitHub Repository URL
https://github.com/tejash2005/dairy-demand-forecasting.git

🧠 Google Colab
https://colab.research.google.com/drive/1bdtCPpKCjfFvh5V87oYrdxqDRm9PIp8U?usp=sharing

📊 Dataset URL
https://www.kaggle.com/datasets/suraj520/dairy-goods-sales-dataset/data