CS 411 – Project Task 1 (PT1) Task 1

Title	Description
Team Name	SelectStar
Project Title	Data-Driven Forecasting and Visualization of Fresh Food Demand in a Retail Store
Team Members	Mihir Shah (mihirss3@illinois.edu) Mohit Badve (mbadve2@illinois.edu) Prajakta Pikale (ppikale2@illinois.edu) Riya Tendulkar (rtend@illinois.edu)

Table of Contents

1. Project Title

The topic that we have decided for our Project Track 1 is "*Data-Driven Forecasting and Visualization of Fresh Food Demand in a Retail Store".* Further details about the project description and problem statement are mentioned below.

2. Project Summary

The "Data-Driven Forecasting and Visualization of Fresh Food Demand in a Retail Store" project is all about changing how inventory is managed by using historical sales data to predict what fresh food items will be in demand. By using data-driven approach through machine learning and advanced data visualization, the system provides retail stores with accurate insights into buying patterns and what they need to stock up on. It's a centralized platform where analysts can keep an eye on inventory levels, track how products are selling, and spot trends as they happen, helping stores make smarter decisions.

This platform will help stores keep the right number of products in stock, reducing waste and supporting sustainability. By offering real-time demand forecasting and interactive visual dashboards, it aims to make operations smoother, increase profits, and ensure customers always find fresh products on the shelves. The easy-to-use interface, along with these data-driven insights, will allow stores to adjust their inventory and marketing strategies proactively, boosting efficiency and customer satisfaction.

3. Problem Statement

Food waste is a significant global issue, particularly in the fresh food sector, where a large amount of produce is discarded due to inefficiencies in supply chain management, inventory control, and consumer habits. According to the Food and Agriculture Organization (FAO), approximately one-third of all food produced for human consumption is lost or wasted globally, amounting to about 1.3 billion tons annually. In the United States alone, fresh food such as fruits, vegetables, dairy, and meat products make up nearly 40% of all food waste, with grocery stores contributing significantly to this problem due to overstocking, improper storage, and unpredictable demand. This not only leads to economic losses but also contributes to environmental harm, as food waste in landfills generates harmful greenhouse gases. Overstocking leads to waste and financial loss, while understocking results in missed sales and unhappy customers. Retail stores often struggle to find the right balance.

Our goal is to solve this problem by creating a real-time data-driven forecasting and visualization system which will provide predictions of what and how much will be needed, helping retailers optimize stock levels, cut down on waste, and ensure optimal product availability. The solution empowers stores to make well-informed decisions, improving efficiency and promoting a more sustainable way of doing business.

Currently, many retail stores rely on traditional methods or basic software tools for inventory management, which often lack the sophistication to provide accurate, real-time demand forecasting. Without a system that can dynamically analyze purchasing patterns and predict future needs, stores are left making reactive decisions rather than proactive ones. This gap in effective inventory management not only impacts the store's profitability and sustainability efforts but also affects the customer experience by limiting product availability. It's evident that a more advanced, data-driven solution is needed—one that can offer actionable insights and make the entire inventory management process more efficient.

4. Usefulness and Benefits

In the competitive grocery retail sector, effective inventory management and demand forecasting is essential for ensuring optimal stock levels, reducing waste, improving sales and enhancing customer satisfaction. All of these factors can be catered to in developing a real-time data-driven forecasting and visualization application that would help retail grocery stores streamline their operations, get a better understanding of their inventory, and enhance decision making. The centralized visualization platform would help business and operations analysts to track inventory, monitor sales performance, identify trends and buying patterns all at one place. Additionally, using demand forecasting, stores can ensure that there is sufficient stock to meet the demands and also identify if over stocking is taking place which would cause loss for the money spent on purchasing and storing the excess products. Using the dashboard, the stores can then take steps to mitigate under/over stocking of products. Detailed advantages of utilizing the dashboard application would be as follows –

- **Enhanced Inventory Management**
    a. With visualizations on the unified dashboard, the analysts of the company can easily track and monitor the current grocery stock across all stores. This would also enable them to gain insights into the sales demographics of each product and make real time informed decisions.
    b. The dashboard will aid in understanding the customer buying patterns and emerging trends in grocery purchases allowing for strategic adjustments to inventory and marketing strategies, ultimately leading to increased efficiency and profitability.

- **Reduced waste and enhanced sustainability**
    a. By leveraging forecasting to predict grocery sales and determine the required inventory, the company can ensure stores receive only the required amount of inventory, minimizing wastage of food and resources. This would also ensure that the grocery items maintain their quality by not storing them for longer periods of time and fresh inventory is available to customers at all times.
    b. Targeted inventory management helps in conserving the company's resources and financial savings by avoiding wastage of surplus grocery and the costs associated with excess stocking.

- **Ensure availability is always optimum to avoid Out of Stock situations**
    a. During "Out of Stock" Situations, sales opportunities are lost due to unavailability or lack of inventory.
    b. By implementing an application that accurately predicts demand and optimizes stock levels, the company can ensure that high-demand products are always available on the shelves.

- **Environmental Impact and Corporate Responsibility**
    a. Adopting a data-driven approach would help in supporting the company's commitment to environmental responsibilities.
    b. By reducing food waste, the company could lower its carbon footprint and contribute to more sustainable practices.
    c. This proactive stance on sustainability would enhance the company's reputation as a responsible corporate citizen and align with growing consumer expectations for environmentally-conscious business practices.

To ensure that the above-mentioned advantages are integrated into the application, the analysts will have the following functionality in the dashboard –

1. Analysts will be able to login into the application.
2. They can view the dashboards containing the graphs and charts to understand inventory stock.
3. They will also have the functionality to apply filters on products, market, time frame amongst others to view the required data.
4. Option to generate required reports to view the past, current or predicted future trends will be provided.
5. Analysts will also be able to view the forecasted required inventory and compare it to the current stock which is present.
6. They will be able to make lists of required produce and share it with stores and manufacturers.

5. Users and Functionalities

The primary users of our applications would be Business Analysts, Operations Analysts and Managers of the retail store. There will be role-based access-control implemented in the application to ensure minimum necessary privileged access. The functionalities provided for these users would be as follows –

1) Login & Password Recovery

- Description:
    - The users of the application will have to login to view the accessible dashboards. Since the dashboard will be displaying sensitive sales and inventory data, it is essential that only authorized users will be able to access the system. Role-based access control will be implemented in the application to ensure the minimum necessary privileged access.

        - It is implied that the users will be registered by an administrator, and they should not have the option to sign up themselves. They will get an account setup link from the administrator to set their passwords and will have the option to recover/change the password later if forgotten.
    - Actions Performed on the Database:
        - A dedicated table for managing authentication and authorization - UserAuth. CRUD operations will be performed on this table.

2) **Current Inventory Analysis**
    - Description:
        - The authorized users will be able to view the dashboards detailing the current inventory stock and apply filters and selections for fields such as product, category, and time amongst others as per their needs.
        - The dashboards and analysis will be getting updated real-time as the retail stores usually take orders digitally. For the scope of this project, we are simulating this through manual entry or bulk entry via file upload of orders through an administrator.
    - Actions Performed on the Database:
        - A dedicated table for managing orders - Orders. CRUD operations will be performed on this table.
        - A dedicated table for managing products - Products. CRUD operations will be performed on this table.
        - A dedicated table for managing the daily inventory - Inventory. CRUD operations will be performed on this table.

3) **Inventory Expiry Tracking**
    - Description:
        - The system will track the expiry dates of perishable items and provide alerts for products that are nearing their expiration date. This will help stores reduce waste by identifying products that need to be sold quickly.
    - Actions Performed on the Database:
        - The Inventory table will include an expiry_date field.
        - Automated checks will trigger alerts based on expiry dates and store these in the Alerts table.

4) **Demand Forecast**
    - Description:
        - It will provide a detailed unified view of the historical fresh food consumption. It will have an option to filter the data by the product, category, and time amongst others as per their needs.
        - The dashboards will have top N products that are being consumed at a higher rate indicating the increase in the demand and hence the supply.

        - It will have a section to display the predicted demand of all the products calculated by applying prediction algorithms on historical data.
        - There will be a trend line to better understand the trend of demands.
    - Actions Performed on the Database:
        - Usually, the historical data is stored in cold line archive storage to reduce the cost involved by trading off the retrieval speed. But, for the scope of this project, we will use databases and optimized queries for faster retrieval.
        - One good simulation that we will do is to set a trigger after working hours to move the data from indexed tables which are used for daily transactional purposes and load it into the dedicated denormalized wide table for running machine learning algorithms. This will have zero impact on the daily transactional data management and will also ease out the retrieval by emptying the unnecessary records.

5) **Check Required Future Stock Demand**
    - Description:
        - Based on the above two views, the application would aim to provide a comprehensive data on which products are currently present in the inventory and as per forecasted demand, what would be the quantity that would be required to restock.
        - Simliarly, based on the expected forecast it would also aim to provide mitigation strategies for excess stock (like allocating the excess stock to some other market). For example, if there are excess sugar packets for sale, they could be allocated to the "Bakery" department to be utilized for generating the bakery items.
    - Actions Performed on the Database
        - Read operation on the denormalized wide table for analytics.

6) **View & Download Report**
    - Description:
        - Apart from seeing the visualization of the dashboard, users can generate reports to check the historic trends, sales, etc. and can also have reports to check the amount of profit generated, saved food wastage due to demand forecasting.
    - Actions Performed on the Database:
        - A dedicated table for managing audit history - AuditHistory. CRUD operations will be performed on this table. This will keep track of all the user activity.

6. Realness

As we discussed the usefulness of the application for real retail stores, we naturally want to keep the realness of the application intact.

The Data sources selected for the project are as follows:

1. For fresh food products,
    a. https://www.kaggle.com/datasets/willianoliveiragibin/retail-analytics-trends?select=All_Data_ASDA.csv
    b. https://www.kaggle.com/datasets/willianoliveiragibin/retail-analytics-trends?select=All_Data_Aldi.csv
    c. https://www.kaggle.com/datasets/willianoliveiragibin/retail-analytics-trends?select=All_Data_Morrisons.csv
    d. https://www.kaggle.com/datasets/willianoliveiragibin/retail-analytics-trends?select=All_Data_Sains.csv
    e. https://www.kaggle.com/datasets/willianoliveiragibin/retail-analytics-trends?select=All_Data_Tesco.csv
2. For customers and orders,
    f. https://www.kaggle.com/datasets/aungpyaeap/supermarket-sales
    g. https://data.world/ahmedmnif150/walmart-retail-dataset/workspace/file?filename=Walmart-Retail-Dataset.csv

As we are trying to build a real application, as data cleaning is involved in any data-drive system, we will apply data cleaning techniques to extract out the data we need for our solution.

We will also generate mocked data for testing purposes, but we are looking for current datasets as well, if possible, from one of the retail stores with all customer details masked. We will connect with the professor and TAs to help us out here.

7. Integration of Creative Components

To make the application interactive and intuitive, we will be using dashboards having multiple graphs, heat maps, trend lines, filter and selection options which would help the analysts grasp patterns easily. This would significantly enhance the interface of the application having intuitive layouts, customizable views, and adaptive displays that change based on user roles or preferences. The application would also have detailed filters and selection criteria to modify the dashboard's graphs according to the use case ensuring that the analyst can see relevant and required data at a time. Components like drag-and-drop widgets, reports, and alert systems that respond to user interactions would help in making complex data accessible and actionable.

Beyond visualizations, we would be leveraging Machine Learning algorithms for demand forecasting to predict the future required inventory or stock. By analysing

historical sales data, seasonal trends, market conditions, and external factors, our application can provide precise, data-driven insights. This advanced approach enables proactive inventory management, helping to optimize stock levels, minimize waste, and ensure that products are available to meet customer demand, ultimately driving increased efficiency and profitability for the business. Common Machine Learning Algorithms like Regression Models, Gradient Boosting or Ensemble Models would help us with implementing the forecasting.

Here are some key database concepts and how they will be applied:

1. Complex SQL Queries:

These queries will be used to retrieve and analyze historical sales data, forecast demand, and provide real-time inventory insights.

Usage:

- Multi-table joins to fetch data from related tables like Orders, Products, and Inventory.
- Subqueries for filtering and identifying top-selling products or products nearing expiry.
- Aggregate functions (e.g., SUM, COUNT, AVG) to generate sales trends and compute the total quantity sold for demand forecasting.

Example: A query to fetch the top N products with the highest demand over a specific period, helping analysts predict future demand more accurately.

2. Triggers:

To automate tasks based on events such as product expiry, stock level changes, or new sales data.

Usage:

- Expiry Tracking: A trigger will automatically alert when a product's ExpiryDate is approaching, notifying the system to take action (e.g., discounting the product).
- Stock Level Alerts: A trigger will ensure that when inventory levels fall below a certain threshold, the system will notify the store to restock.

Example: A trigger can update an Alerts table whenever stock levels fall below the reorder point for certain products, ensuring that high-demand items are always in stock.

3. Constraints:

To maintain data integrity and consistency across the system.

Usage:

- Primary and Foreign Key Constraints to enforce relationships between tables (e.g., between Products, Suppliers, and Orders).

- Check Constraints to validate specific conditions like ensuring the UnitPrice in the Inventory table is always positive, or that expiration dates are valid.
- Unique Constraints to ensure critical fields like ProductName or Email in the User table are unique.

Example: A check constraint ensuring that ExpiryDate is always later than ManufactureDate in the Inventory table.

4. Stored Procedures:

To encapsulate frequently used queries and improve performance by reducing the need to repeat complex query logic.

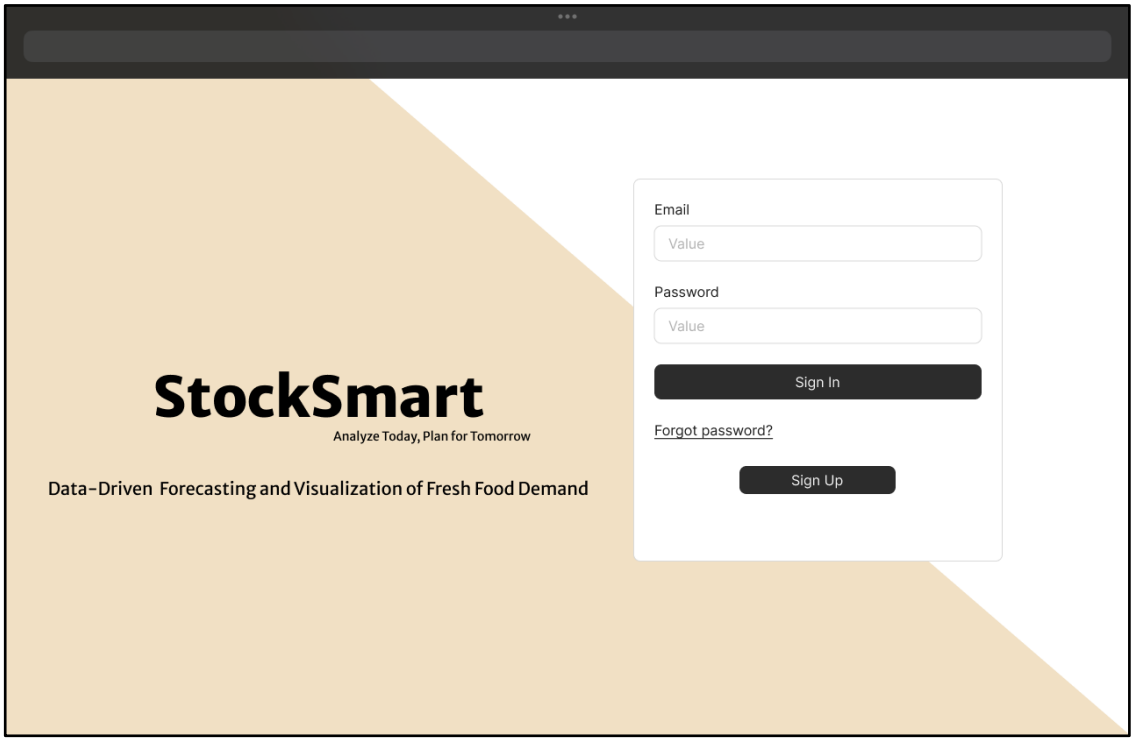
Usage:

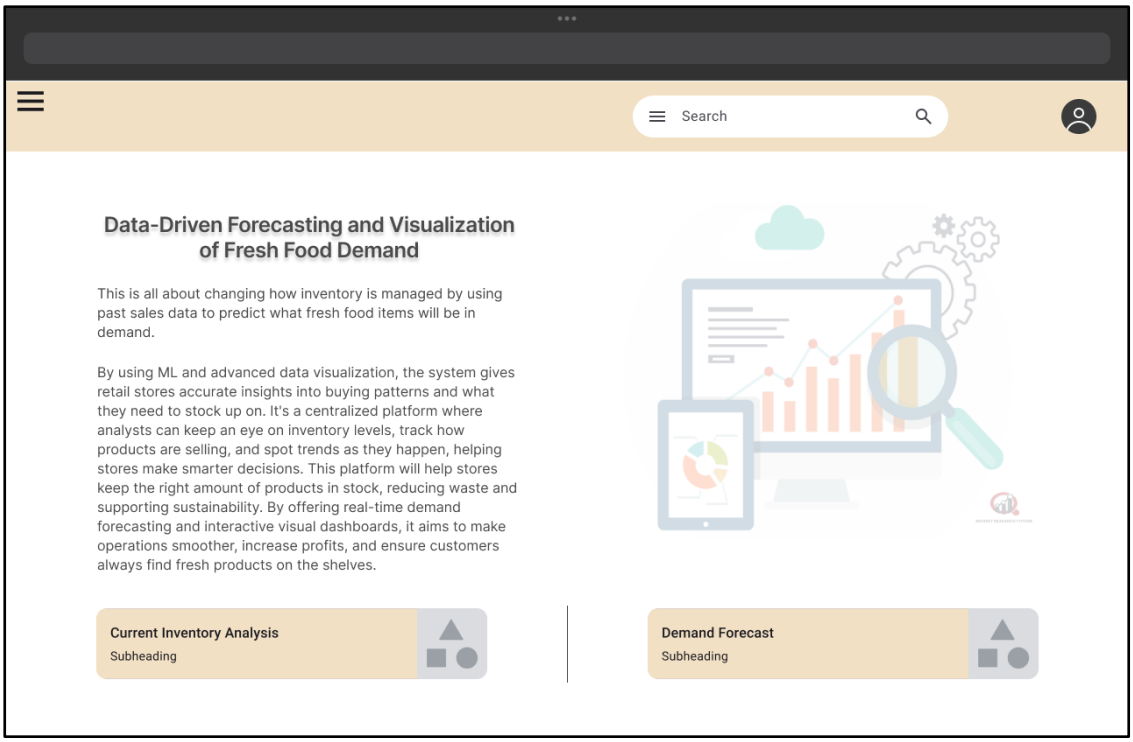
- Stored procedures will be used for calculating forecasts based on past sales data and updating forecast tables at the end of each day.
- Can be used to generate reports on demand with predefined filters (e.g., daily, weekly, or monthly sales).

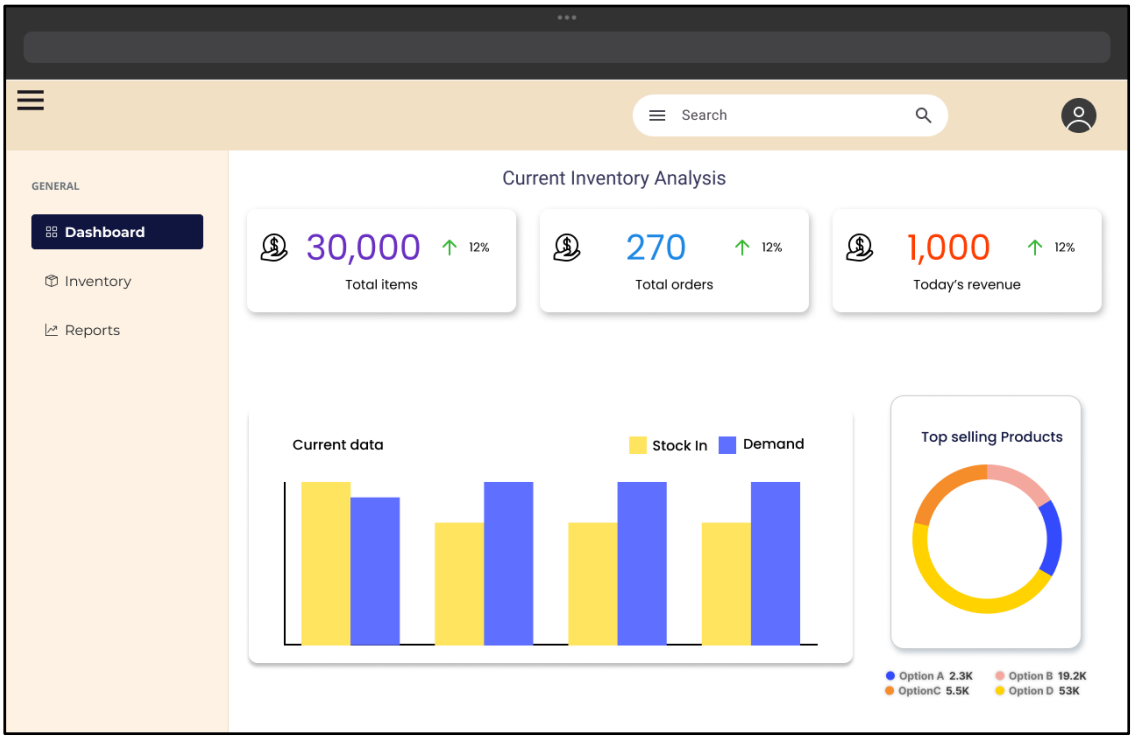
Example: A stored procedure that is triggered at the end of each business day to aggregate the day's sales, update forecasted demand, and store it in the Forecast table.

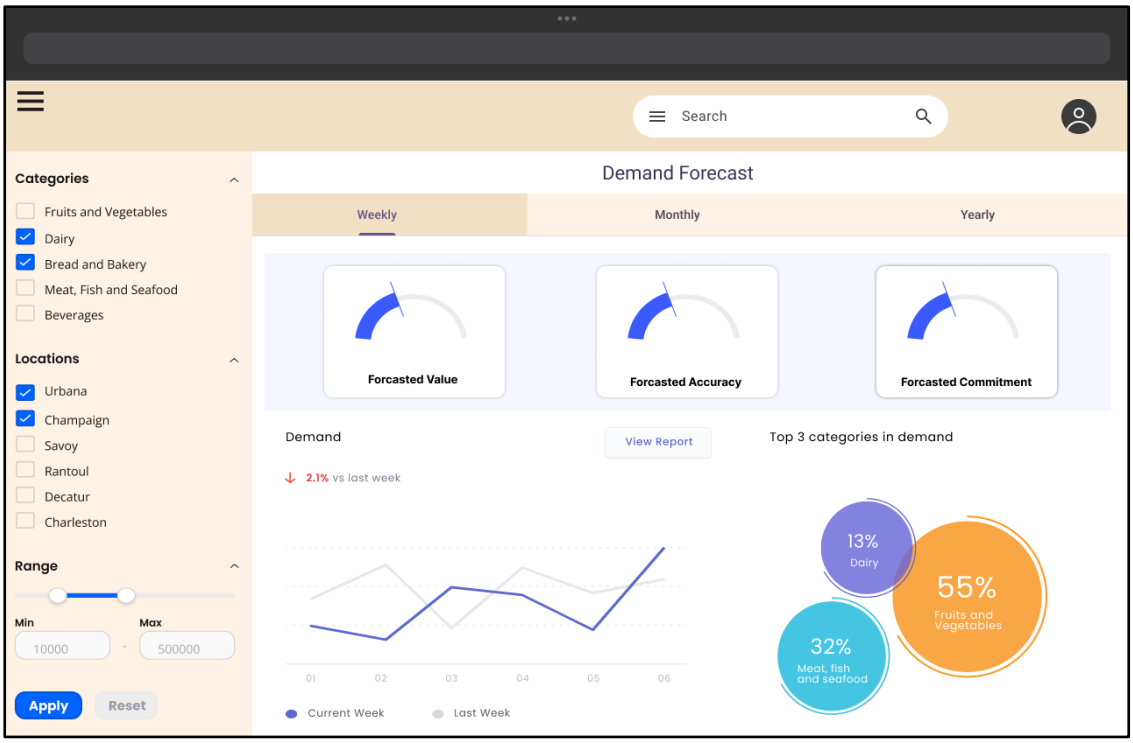
8. Detailed Description of Delivery of Web Application

8.1. UI Mock-up




Search
Data-Driven Forecasting and Visualization of Fresh Food Demand
This is all about changing how inventory is managed by using past sales data to predict what fresh food items will be in demand.
By using ML and advanced data visualization, the system gives retail stores accurate insights into buying patterns and what they need to stock up on. It's a centralized platform where analysts can keep an eye on inventory levels, track how products are selling, and spot trends as they happen, helping stores make smarter decisions. This platform will help stores keep the right amount of products in stock, reducing waste and supporting sustainability. By offering real-time demand forecasting and interactive visual dashboards, it aims to make operations smoother, increase profits, and ensure customers always find fresh products on the shelves.
Current Inventory Analysis
Subheading
Demand Forecast
Subheading


Search
GENERAL
Dashboard
Inventory
Reports
Current Inventory Analysis
30,000
12%
Total items
270
12%
Total orders
1,000
12%
Today's revenue
Current data
Stock In
Demand
Top selling Products
Option A 2.3K
Option B 19.2K
OptionC 5.5K
Option D 53K



8.2. Project Work Distribution

Task	Team Member
Creation of Database Schema	All Members
Data Cleaning and Loading	All Members
Login and Password Recovery	Riya, Mihir
Current Inventory Analysis	Prajakta, Riya
Demand Forecasting	Mohit, Mihir
Create and Download Report	Mohit, Riya
Add/Delete/Update Functionalities	Mohit, Prajakta
Creation of frontend views	Prajakta, Mihir
Integration of backend with frontend	Mohit, Riya
Prediction Model Development for Forecasting	All Members