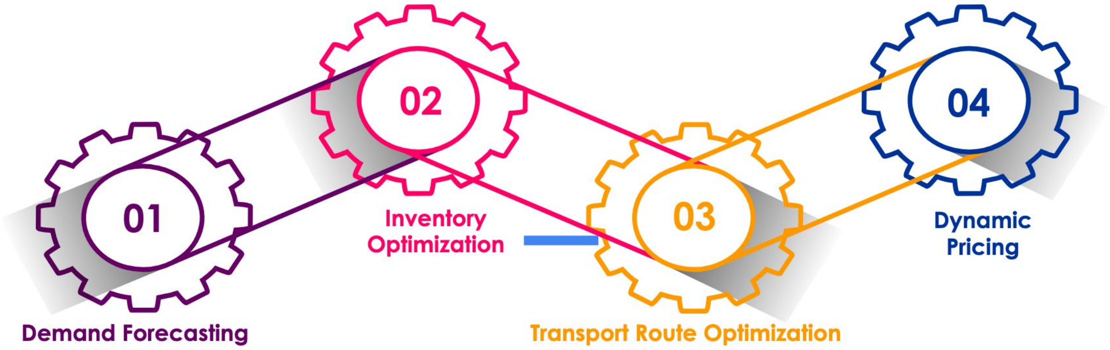
Industry 4.0 Project
Team : 4.0 Pioneers

21f1001107	Utkarsh Sahu
21f3001154	Saarthak Saran
21f1002100	Ashirbad Arun Dash
21f1003267	Abhijeet Kumar Garg
21f1006777	Kartikeya Tiwari

Problem Statement

Booming Personal Care Industry: India’s personal care industry is growing rapidly due to rising incomes, evolving self-care trends, and social media influence.

- Supply Chain Challenges:
  - Inefficient inventory management leading to stockouts or overstocking.
  - High transportation costs impacting profitability.
  - Inconsistent supplier performance causing delays and disruptions.
  - Production bottlenecks affecting timely delivery of products.
- Impact on Brands:
  - Lakmé: Stockouts during peak seasons, leading to lost sales opportunities.
  - Mamaearth: Struggles with managing complex product variations, affecting operational efficiency.
- **Research Objective:** Explore data based solutions to improve the overall supply chain with special focus on Demand Forecasting, Inventory Optimization, Dynamic Pricing, Transport Route Optimization, Defect Prediction and Supplier Performance Evaluation

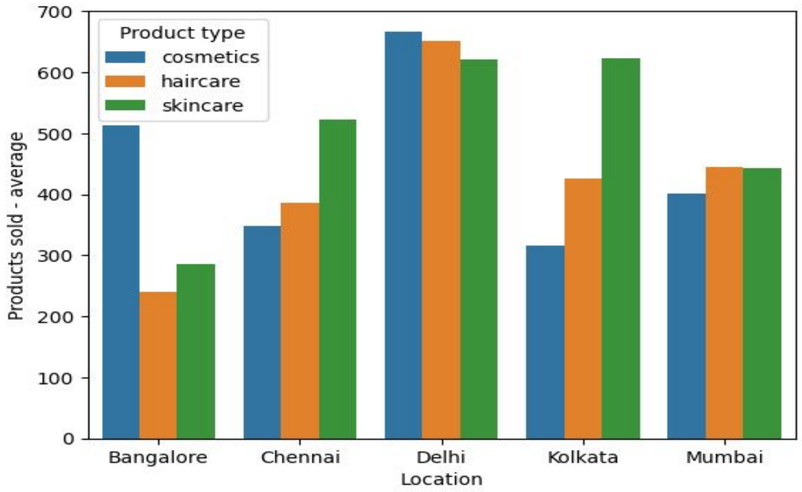


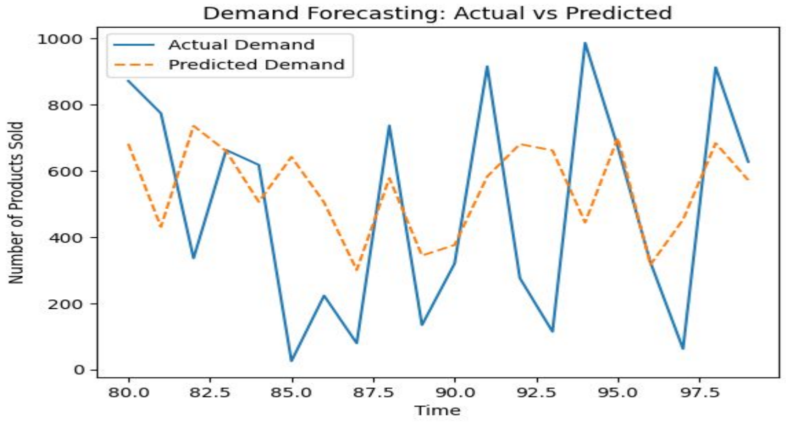
Objective and Constraints

Demand Forecasting	Forecast product demand to optimize inventory and prevent stock-outs or overstocking.	• Using different regression models (like, Linear Regression, XGBRegressor, etc. with features like price, availability, stock levels, and shipping times to predict the number of products sold.
Inventory Optimization	Minimize the total order quantity needed to meet demand.	• Constraints: ○ Order quantity should meet or exceed the forecasted demand (taking into account lead time)
Transport Route Optimization	Minimize total shipping costs while meeting delivery time constraints and incorporating route efficiency.	• Constraints: ○ Each route must use exactly one transportation mode. ○ Delivery time per route ≤ specified threshold (here 7 days).
Dynamic Pricing	Adjust prices to maximize revenue by modeling price elasticity and customer behavior.	• Price Elasticity Modeling: Training a machine learning model to predict revenue based on price, customer demographics, and availability.

Overview of the Relevant Data

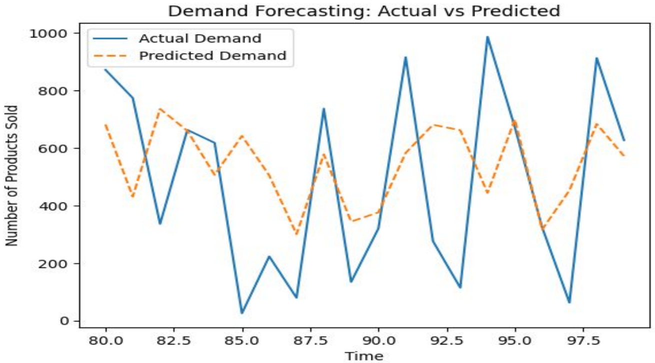
- Dataset contains 100 entries with 24 columns.
- Product Information: Product type (cosmetics, skincare, haircare), SKU (Stock Keeping Unit identifier), Price per product, Availability (units available).
- Sales Metrics: Number of products sold, Revenue generated per product.
- Inventory Data: Stock levels (units available), Order quantities.
- Lead Times: Lead times for shipping and manufacturing processes.
- Production Data: Production volumes, Manufacturing lead time.
- Cost Metrics: Manufacturing costs per product, Shipping costs per route and transport mode, Overall operational costs.
- Quality Control Data: Inspection results (Pass, Fail, Pending), Defect rates per product category.
- Transportation and Logistics Data: Transportation modes used (Road, Rail, Air, Sea), Routes used (Route A, Route B, Route C), Shipping times, Costs associated with each route and mode.



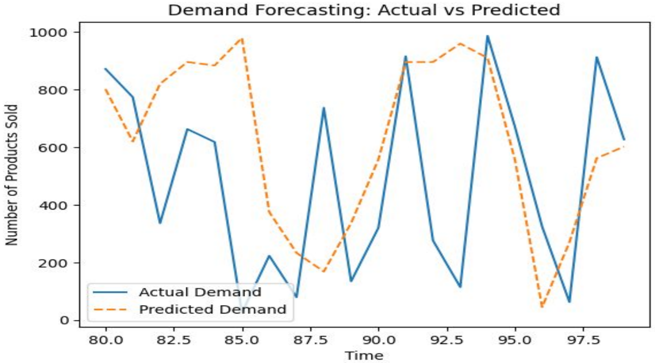


Methodology

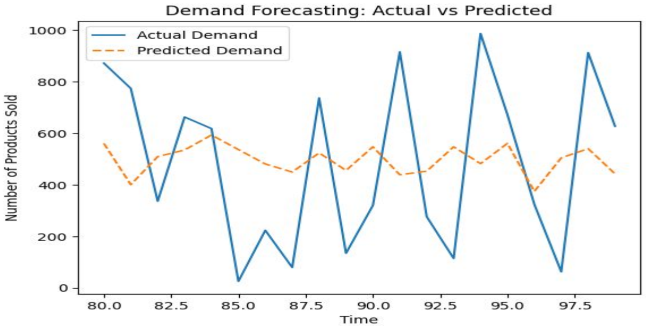
- **Data Acquisition and Preparation**: The team imported necessary libraries (pandas, numpy, seaborn, plotly) and loaded supply chain data from Kaggle.
- **Exploratory Data Analysis (EDA)**: The code shows descriptive statistics generation to understand numerical features (mean, standard deviation, min/max values) of price, availability, stock levels, and other metrics, revealing insights like average stock levels of 47.77 units and average lead times of 15.96 days.
- **Product Performance Analysis**: The team aggregated data by product type to compare sales volumes across categories (skincare: 20,731 units, haircare: 13,611 units, cosmetics: 11,757 units) and analyzed operational metrics like average lead times by product category.
- **Statistical Analysis and Correlation**: The code includes statistical testing (chi2_contingency) and correlation analysis to identify relationships between variables, with imported libraries suggesting the relationships between inventory levels, sales, and other supply chain metrics.
- **Optimization Modeling**: The team used PuLP (linear programming library) for optimization problems, likely for transportation mode selection, inventory management, and route planning as indicated by the imported library and the analysis results showing route efficiency comparisons.
- **Machine Learning Implementation**: The code imports LinearRegression, StandardScaler, and metrics like mean_squared_error and r2_score from scikit-learn, indicating predictive modeling was used for demand forecasting and potentially for dynamic pricing optimization as mentioned in the analysis results.



XG Boost

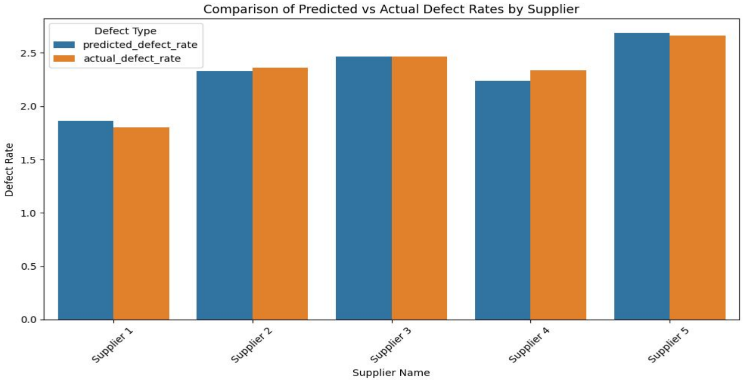


Decision Tree

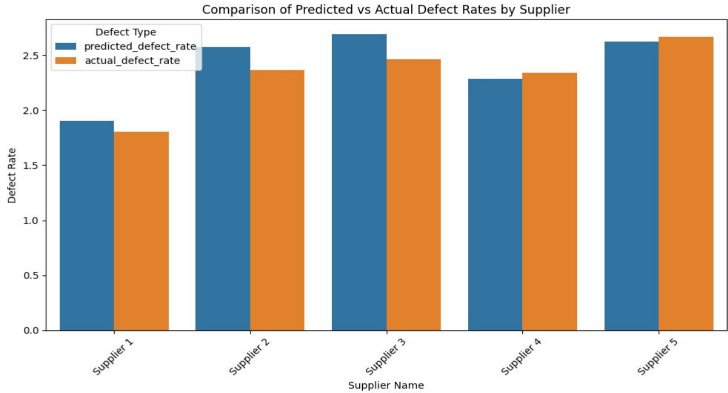


Random Forest

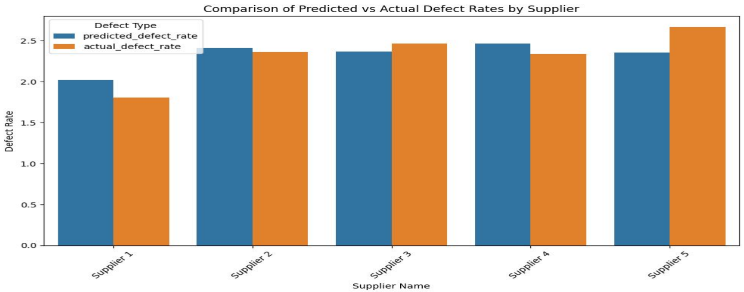
XG Boost



Gradient Boost Regressor



Linear Regressor



Results and Implications

Product Performance:
Skincare products account for the **highest sales volume (40%).** Average defect rates vary by product type, with **skincare having the lowest defect rates.** Revenue distribution varies significantly across product types, with **skincare generating the most revenue**

Transportation & Logistics:
Road transportation accounts for the **highest sales volume (29%). Air** transportation shows the **lowest defect rates**, while Sea has the highest. **Route A is the most efficient route** with an efficiency score of 0.08857. Optimal transportation mode selection can reduce shipping costs while maintaining efficiency.

Dynamic Pricing:

The model identified significant differences in optimal pricing **based on product availability using RandomForestRegressor.**

With **high availability,** the **optimal price was $2.68,** whereas with **low availability, the optimal price shifted to $88.34,** indicating that **scarcity justifies a higher price to maximize revenue** while balancing reduced supply.

The observed variation highlights the model's sensitivity to **key supply chain factors such as availability and customer demographics.**

Geographic Insights:
Kolkata shows the highest total number of products sold, followed by **Mumbai** Revenue generation varies by location, with **Mumbai generating the highest revenue.** Different locations show varying preferences for product types. **Bangalore having lowest selling of products.** Companies needs to be understand why this city having less selling even though it's major city in India.

Supplier Analysis:
Supplier 1 demonstrates the lowest defect rates (1.80%), while Supplier 5 has the highest (2.67%).
XGBRegressor with objective as reg:squarederror, n_estimators=4000 and max_depth=10 is the most optimised model for defect prediction.
Supplier 1 and Supplier 2 were identified as optimal suppliers based on cost, defect rate, and lead time