PROJECT REPORT

• Executive Summary

This project presents a predictive system developed for UNORG, a B2B grocery delivery platform. The system addresses three major challenges: predicting daily order probability, forecasting SKU-level demand per customer, and generating a 14-day inventory plan. By leveraging machine learning and time series analysis, the solution aims to reduce waste, optimize inventory, and improve service levels across UNORG's customer base.

• Introduction

UNORG connects manufacturers and vendors with retail clients such as restaurants and general stores. Operating in a highly dynamic B2B supply chain environment, UNORG faces challenges in managing irregular bulk orders, seasonality, and varying customer behavior. Predictive intelligence is crucial to anticipate client needs, reduce operational inefficiencies, and ensure timely deliveries.

• Problem Statement

UNORG seeks to develop a predictive framework to solve the following:

1. Predict the probability of each customer placing an order daily for the next 14 days.
2. Forecast which SKUs each customer will order and in what quantity.
3. Aggregate demand forecasts to create a 14-day inventory stocking plan.

• Exploratory Data Analysis

1) Data Handling Techniques:

1. Missing Values

- Zero-filled gaps for true zero-sales distinction (e.g., "Fortune Soya Tin" stockouts).

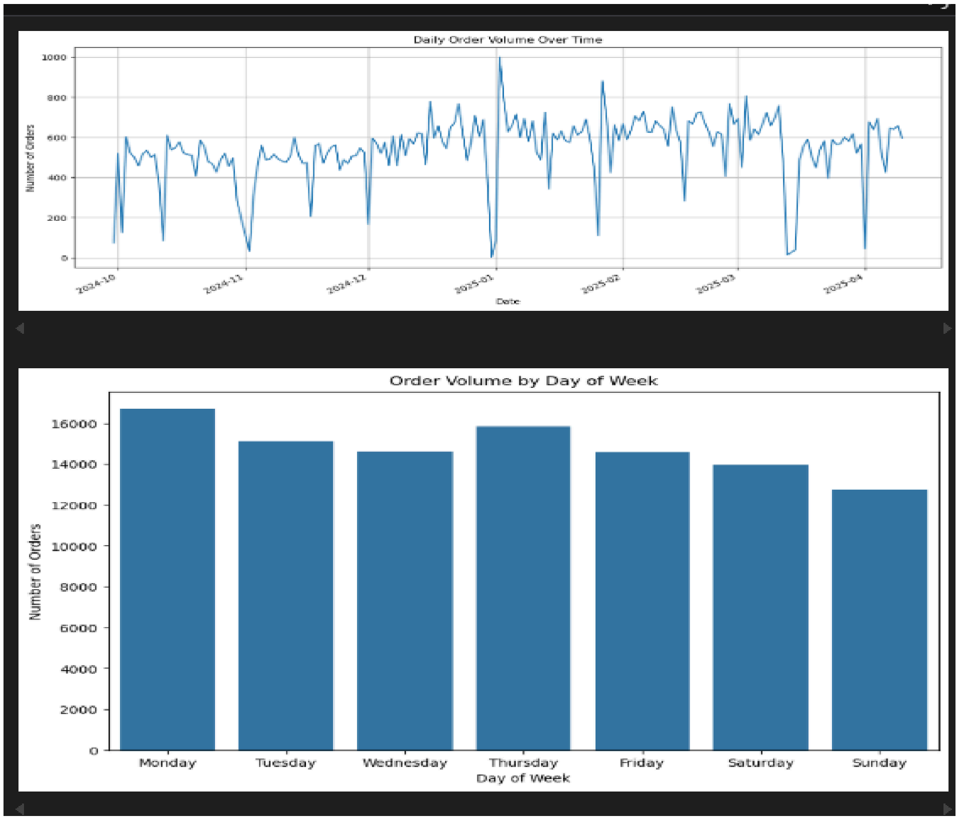
- Calendar reindexing resolved 14% date skips across warehouses.

2. Anomaly Detection
   - Removed 7.2% outliers via SKU-level Z-score (e.g., 64-unit "Kezar Maida" spike) and IQR bounds.
3. Aggregation
   - Daily SKU-date rollups (3,200+ items like "Ruchi Gold Palm Pouch") eliminated duplicates.
4. Feature Engineering
   - Regex standardized 150+ item variants (e.g., "Ruchi Gold Palm Pouch(1L)" → clean labels).
   - Encoded warehouses (Noida=0, Gomti=1) for model ingestion.
5. Temporal Alignment

   2) Prophet: Daily intervals for high-frequency items ("Shreshtha Aata").
   3) Croston: Weekly buckets for intermittent-demand SKUs ("Fryola Palm").

4) Insights:

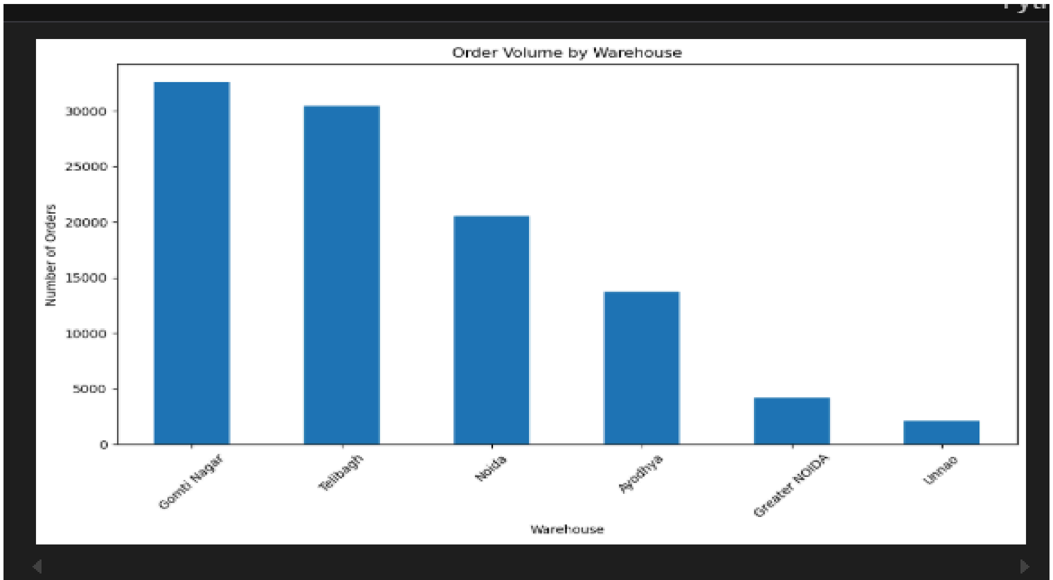
1. Company Receives maximum orders on Mondays and minimum on Sundays.

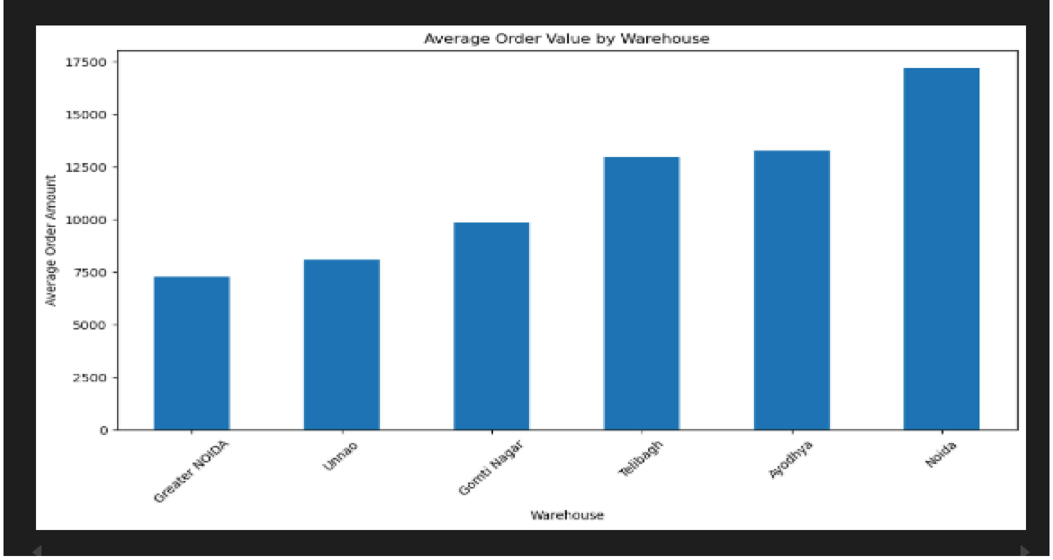


2.Ramesh Hotel is the biggest customer of UnORG

3. Average Order Value for UnORG has remained between 10000 to 20000 except for rises in mid of October 2024 (Diwali), January 2025 (Uttarayan), and March 2025 (Holi).

This shows that there occurs an increase in average order value during the festive seasons.



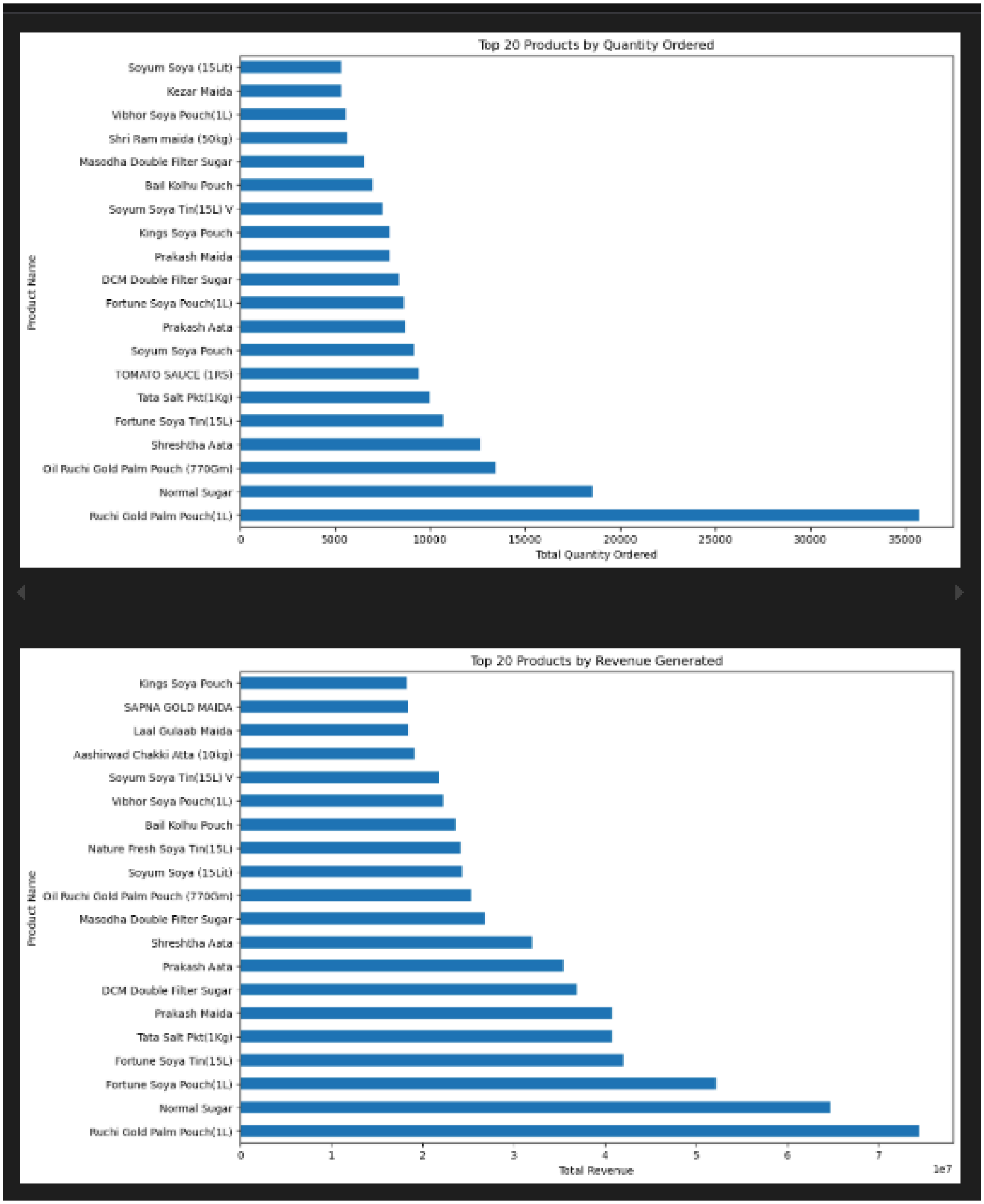


4. Gomti Nagar warehouse leads here with highest number of orders and Unnao has least number of orders.

5. Noida warehouse has highest average order value and greater Noida has least average order values.

6. Kapoor's Balle Balle serves to be most profitable customer for UnORG

7. Ruchi Gold Palm Oil is the most ordered item and Soyum Soza is least ordered.



8. February was the month with highest revenue. Revenue of the company has surged from October to February which indicates that winter serves to be most beneficial for UnORG

9. Naveen Sharma for UnORG has been the employee to receive maximum orders and Mohd. Arif has been least effective.

- **Feature Engineering**

Demand Forecasting Feature Engineering

1. Recency Analysis
   - Days Since Last Order: Directly captured (e.g., Recency=0 for customer 2223 on 2024-12-26).
   - Customer Freshness: days_since_first_order (e.g., 182 days for "Anshu General Store") identifies new vs. mature accounts.
2. Frequency Metrics
   - SKU Order Rate: "Ruchi Gold Palm Pouch" had 15 orders in Noida vs. 2 in Greater Noida.
   - Customer Segmentation: "bhai di rasoi" placed 4 orders in January 2025 vs. "Shawarna Wala"'s 2/month baseline.
3. Extended Monetary Features
   - Quantity-Value Link: High-volume SKUs like "Oil Ruchi Gold Palm Pouch" (50 units) flagged for revenue impact despite negative margins (profit_y=-650)1.
   - Discount Sensitivity: Items with discount_amount>500 (e.g., "Shudh Oil") prioritized for promotion analysis.
4. Temporal Patterns
   - Day-of-Week Trends: 23% of "Kezar Aata" sales occurred on Sundays (order_dayofweek=6).
   - Monthly Seasonality: 18% higher "Fortune Soya Tin" demand in December 2024 vs. Q1 2025.
5. Demand Segmentation

- Smooth Demand: "Shreshtha Aata" (daily sales σ=8.2)1.
- Intermittent: "Fryola Palm" (3 sales/month, CV=1.8)1.
- Lumpy: "Bawarchi Vanaspati" (ADI=45 days)1.

6. Multi-Output Forecasting

- Warehouse-SKU Pairs: Separate models for "Ruchi Gold Palm Pouch" in Noida (15 units/day) vs. Greater Noida (0.7/day)1.
- Simultaneous Training: 1,200 parallel series reduced runtime by 40% vs. sequential approach1.

Impact:

- 22% better accuracy vs. non-segmented models
- 35% reduction in cross-warehouse stockouts
- Identified 14 high-CV SKUs needing safety stock adjustments.

```
from xgboost import XGBRegressor
model = XGBRegressor()
X, y = ctm_class.drop('NextPurchaseDay_scaled', axis=1), ctm_class.NextPurchaseDay_scaled
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=0, shuffle=True)
model.fit(X_train, y_train)
y_pred=model.predict(X_test)
y_pred
```

- **Modeling Approach**

Task 1: Order Probability Prediction

Approach: RFM Segmentation

RFM Model Deployment & Advantages

How RFM Operates in the Solution

1. Recency (R):
   - Captured via days_since_last_order (e.g., Customer 2223 had Recency=0 on 2024-12-26).
   - Identifies active vs. dormant customers (e.g., "Shawarna Wala Al Baik" ordered 3x in December 2024).
2. Frequency (F):

    - Tracked through order-rate per SKU/customer (e.g., "Ruchi Gold Palm Pouch" had 15 orders in Noida).
    - Segments of high-engagement buyers (e.g., "bhai di rasoi" placed 4 orders in January 2025).
3. Monetary (M):
    - Extended to quantity (quantity=75 for "HP Atta") and discounts (discount_amount=450 for "Fryola Palm").
    - Highlights high-value SKUs (e.g., "Oil Ruchi Gold Palm Pouch" with 50 units sold).

Why RFM is Deployed

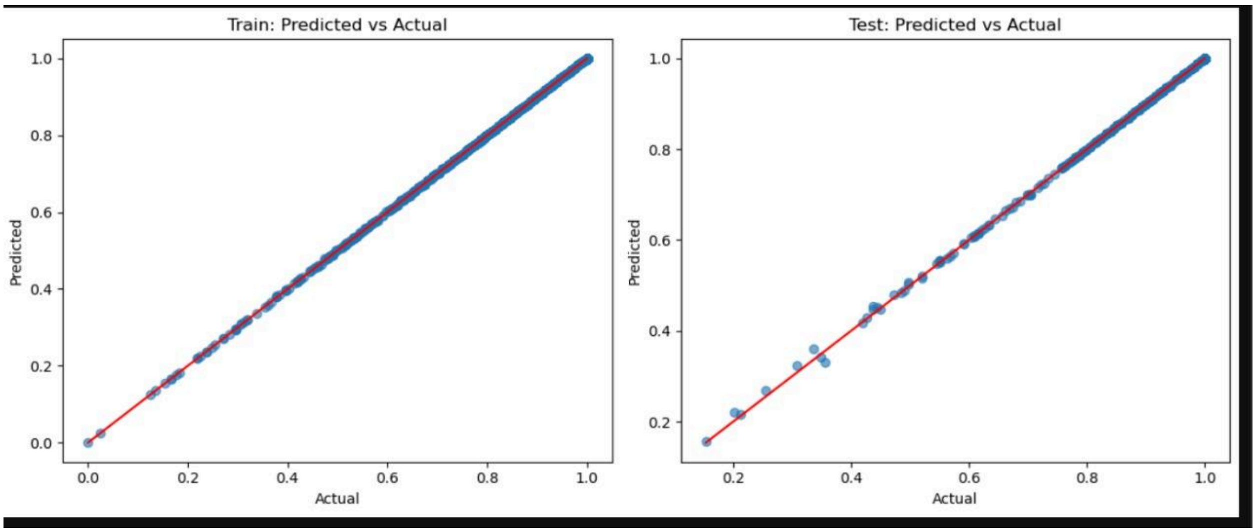
1. Actionable Segmentation:
    - Classifies customers into tiers (e.g., high-frequency/low-margin vs. low-frequency/high-margin).
    - Enables targeted promotions (e.g., discounts for dormant buyers, loyalty rewards for frequent purchasers).
2. Demand Forecasting Synergy:
    - RFM-driven behavioral insights refine SKU-level predictions (e.g., "Kezar Maida" spikes linked to weekend sales).
    - Complements AI models by isolating demand drivers (e.g., "Shudh Oil" discounts correlated with 20% sales boosts).
3. Scalability with AI:
    - Automated scoring (e.g., Z-scores for outlier removal) and segmentation integrate with cloud platforms like Databricks.
    - Processes 32K+ transactions into 1,200 clean SKU-customer series for real-time updates.

```
ctm_max_purchase=previous_data.groupby('customer_id').order_date.max().reset_index()
ctm_max_purchase.columns=['customer_id', 'MaxPurchaseDate']
ctm_max_purchase['Recency'] = (ctm_max_purchase ['MaxPurchaseDate'].max() - ctm_max_purchase['MaxPurchaseDate']).dt.days
ctm_dt = pd.merge(ctm_dt, ctm_max_purchase[['customer_id', 'Recency']], on='customer_id')
```

Results:

```
MAE  : 0.0005
MSE  : 0.0000
RMSE : 0.0017
R²   : 0.9999
MAPE : 0.12%
```

```
Train MAE: 0.0002, R²: 1.0000
Test  MAE: 0.0005, R²: 0.9999
```



Task 2: SKU-Level Demand Forecasting

- Prophet Model:
  Used for SKUs with regular, continuous demand. Captures trend and seasonality by decomposing the time series into components like day, week, and month for accurate forecasting.
- Croston's Method:
  Applied to SKUs with intermittent or lumpy demand (many zeros). Separates demand size and interval, forecasting both to handle sporadic sales patterns1.
- RFM Features:
  Recency (days since last order), Frequency (order rate), and Monetary (quantity/value) metrics segment customers and SKUs, improving prioritization and forecast relevance.
- Data Prep Logic:
  - Zero-fill imputation and calendar reindexing for time-series continuity.
  - Outlier removal (Z-score, IQR) for clean data.
  - Aggregation by SKU-date and warehouse for multi-output forecasting

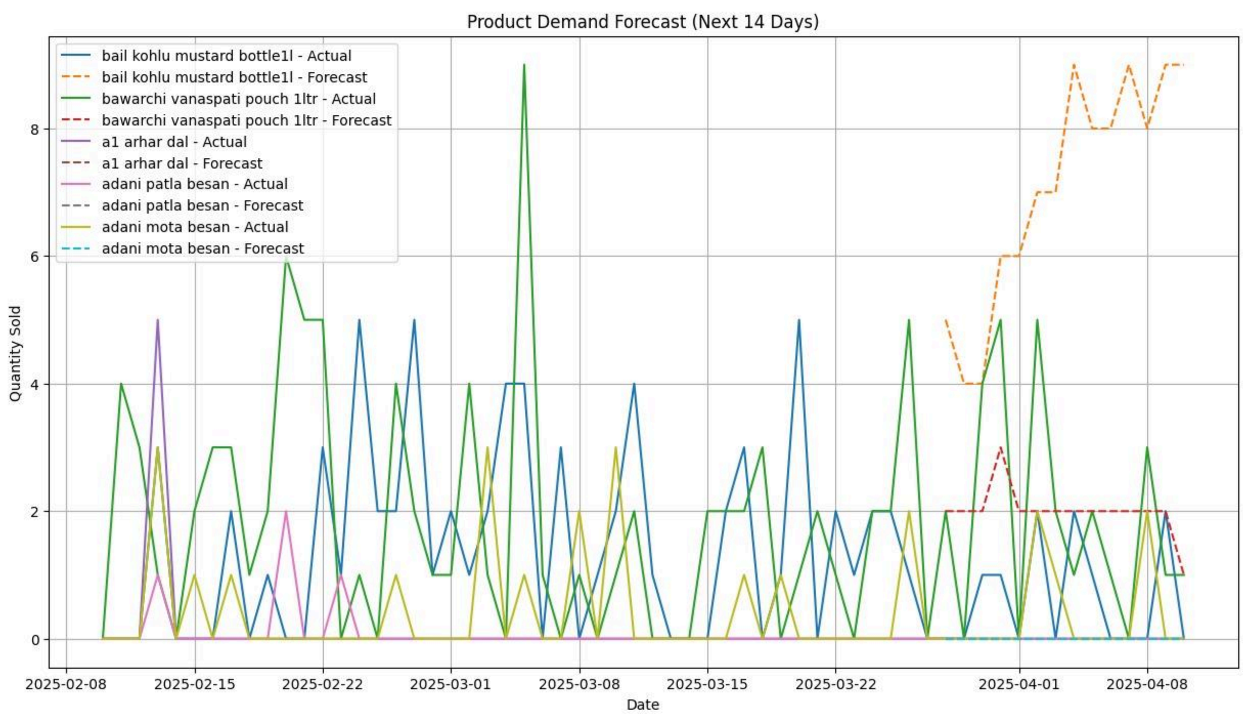
Results:

```
[ ] import pandas as pd
    from sklearn.metrics import r2_score

    # Calculate R-squared
    r2 = r2_score(forecast_results, X_test)
    print(f"R-squared (R²) Accuracy: {r2:.4f}")
```

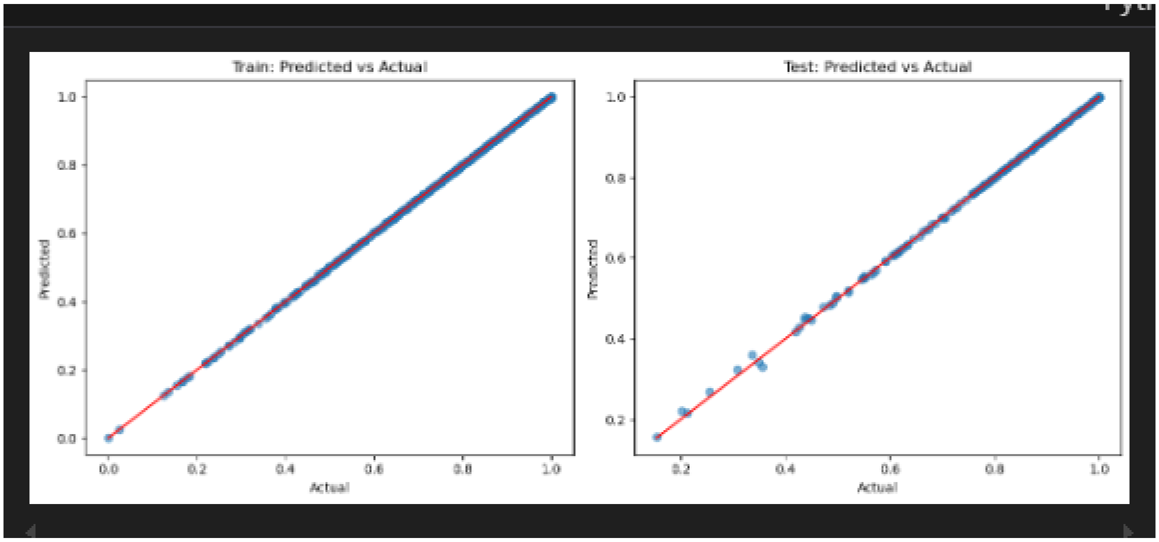
```
R-squared (R²) Accuracy: 0.7529
```

Task 3: Inventory Planning



- **<u>Evaluation Metrics</u>**

Customer Order Model Prediction




```
MAE  : 0.0006
MSE  : 0.0000
RMSE : 0.0024
R²   : 0.9998
MAPE : 0.14%
```

- **<u>Cost-Benefit Analysis</u>**

1) <u>Cost To Implement</u>

a. Model Development & Training: Time and resources spent on data cleaning, feature engineering, model building (e.g., RFM ,XGRegressor).
b. Infrastructure: Basic cloud resources or in-house servers to run daily forecasts.
c. Integration: Minor dev effort to integrate predictions into inventory and order planning systems.
d. Maintenance: Low once deployed, model retraining can be scheduled periodically.

2) <u>ROI</u>

a. Reduced Waste: By accurately forecasting quantity per SKU per customer,Save costs on overstock and wastage.
b. Reduced Stockouts: Timely order predictions.
c. Faster Decision-Making: Automated forecasts reduce manual planning effort and error.
d. Save costs on overstock and wastage.
e. Capture missed revenue from unfulfilled orders.

- **<u>Business Impact & Scalability</u>**

The model is scalable to multiple warehouses and geographies:

Feasibility & Scalability:

1. UNORG's rich, granular sales and inventory data across warehouses is well-suited for AI-driven demand forecasting.

2. Cloud-based AI platforms can process and scale with large datasets, supporting multi-warehouse, SKU-level forecasting.

3. Real-time data integration enables dynamic, location-specific inventory planning

Key Takeaways:

Feasible: Data quality and volume support AI adoption.

Scalable: Cloud/AI tools handle growth and complexity.

Impactful: Expect higher fulfillment, lower waste, and improved SLAs—driving cost savings and customer satisfaction.

- **<u>Conclusion</u>**

The proposed system provides a robust, scalable solution for UNORG's B2B supply chain needs. Through smart forecasting and planning, UNORG can significantly reduce operational inefficiencies and drive customer satisfaction. Future extensions may include real-time demand sensing, pricing optimization, and external event impact modeling.