**Steps Explained**

**1. Load and Merge Datasets**

* **What we did**: Loaded product, order, and metadata files and merged them to create a comprehensive dataset for analysis.
* **Why**: This provides a consolidated view of products, their orders, and associated metadata for further processing.

**2. Add Current Inventory Column**

* **What we did**: Calculated the total historical orders for each product and used this as a proxy for the current inventory.
* **Why**: This step assumes that past order counts approximate the stock currently available.
* **Visualization**:
  + **Histogram of Current Inventory**: Displays how the inventory is distributed across products.
  + **Insight**: Identifies products with very low or very high inventory levels, highlighting potential stockout or overstock risks.

**3. Analyze Repeat Orders**

* **What we did**:
  + Identified repeat orders for each product (instances where a product was reordered).
  + Computed the repeat\_order\_count and repeat\_order\_ratio (percentage of repeat orders relative to inventory).
* **Why**: Repeat order behavior is critical for understanding customer demand patterns.
* **Visualization**:
  + **Histogram of Repeat Order Ratios**: Shows how frequently products are reordered compared to their inventory.
  + **Insight**: Identifies high-demand products likely to require restocking more frequently.

**4. Prepare Dataset for Modeling**

* **What we did**:
  + Aggregated total orders per product (total\_orders) as the target variable.
  + Used current\_inventory, repeat\_order\_count, and repeat\_order\_ratio as features.
* **Why**: These features represent key indicators for predicting future inventory needs.

**5. Train a Random Forest Regressor**

* **What we did**:
  + Split the dataset into training and testing sets.
  + Trained a **Random Forest Regressor** to predict total\_orders based on the features.
* **Why**: Random Forest models are robust, can handle nonlinear relationships, and provide feature importance metrics.

**6. Evaluate the Model**

* **What we did**: Computed the Root Mean Squared Error (RMSE) to evaluate model accuracy.
* **Why**: RMSE quantifies prediction errors; a lower value indicates better accuracy.
* **Insight**: RMSE provides confidence in the model's ability to predict inventory needs.

**7. Visualize Feature Importance**

* **What we did**:
  + Visualized the importance of each feature used in the model.
  + current\_inventory, repeat\_order\_count, and repeat\_order\_ratio were ranked.
* **Why**: Identifies which features most influence inventory predictions.
* **Visualization**:
  + **Feature Importance Bar Plot**: Highlights the key drivers of inventory predictions.
  + **Insight**: Helps prioritize data collection or focus on key aspects influencing demand.

**8. Predict Future Inventory Needs**

* **What we did**:
  + Predicted future inventory (predicted\_inventory) for all products.
  + Added this as a column in the dataset for actionable insights.
* **Why**: This enables proactive inventory planning to minimize stockouts and overstocking.
* **Visualization**:
  + **Scatter Plot of Actual vs. Predicted Inventory**: Compares current inventory to predicted needs.
  + **Insight**: Products above the diagonal may have sufficient stock, while those below it may need restocking.

**Key Takeaways from the Visualizations**

1. **Inventory Distribution**:
   * Understand the current stock levels and identify outliers (e.g., products with excessive or insufficient stock).
2. **Repeat Order Ratios**:
   * Helps focus on products with high repeat orders, which are likely to drive future demand.
3. **Feature Importance**:
   * Identifies the most critical features influencing predictions, guiding better decision-making.
4. **Actual vs. Predicted Inventory**:
   * Visual validation of how well the model's predictions align with reality.