**Rohlik Sales Forecasting Challenge**

**Overview**

The **Rohlik Sales Forecasting Challenge** is a machine learning competition aimed at predicting future sales based on historical data. Organized by **Rohlik Group**, a leading European e-grocery innovator, the challenge focuses on leveraging historical sales data to enhance supply chain processes, delivery logistics, and inventory management. Accurate forecasts contribute to reducing waste and ensuring efficient operations.

**Challenge Scope**

Participants are tasked with forecasting **sales volume** for selected inventory items over the next **14 days** using historical sales data. The dataset contains sales records from Rohlik’s **11 warehouses** operating in **Czech Republic, Germany, Austria, Hungary, and Romania**.

**Competition Timeline**

* **Start Date:** 3 months ago
* **Close Date:** 14 hours ago (Late Submission Accepted)

**Dataset Description**

The dataset consists of multiple files that provide detailed information about sales, inventory, calendar events, and warehouse operations. Some columns have been altered to maintain confidentiality.

**📂 Files in the Dataset**

* sales\_train.csv – Historical sales data for training.
* sales\_test.csv – Testing dataset for making predictions.
* inventory.csv – Metadata about inventory, including product categories.
* calendar.csv – Calendar events such as public holidays and warehouse-specific closures.
* solution.csv – Sample submission file format.
* test\_weights.csv – Weights used in the evaluation metric.

**📝 Key Columns in sales\_train.csv & sales\_test.csv**

* **unique\_id** – Unique inventory ID.
* **date** – Date of the record.
* **warehouse** – Warehouse name.
* **total\_orders** – Historical order count (available for test set too).
* **sales** – Target variable (missing in test set).
* **sell\_price\_main** – Selling price of the item.
* **availability** – Proportion of the day when inventory was available.
* **type\_0\_discount, type\_1\_discount, …** – Various promotional discount rates applied.

**🏷️ Key Columns in inventory.csv**

* **unique\_id** – Inventory ID.
* **product\_unique\_id** – Product ID across warehouses.
* **name** – Product name.
* **L1\_category\_name, L2\_category\_name, …** – Hierarchical product categorization.
* **warehouse** – Warehouse location.

**📅 Key Columns in calendar.csv**

* **warehouse** – Warehouse location.
* **date** – Date of the record.
* **holiday\_name** – Public holiday name (if applicable).
* **holiday** – Binary flag for holidays (0/1).
* **shops\_closed** – Indicator for whether shops were closed.
* **winter\_school\_holidays** – Binary flag for winter school holidays.
* **school\_holidays** – Binary flag for school holidays.

**⚖️ Key Columns in test\_weights.csv**

* **unique\_id** – Inventory ID.
* **weight** – Weight used for computing the final evaluation metric.

**📊 Evaluation Metric**

Submissions are evaluated using the **Weighted Mean Absolute Error (WMAE)** between the predicted and actual sales values. Weights for test evaluation are provided in the test\_weights.csv file.

**📌 Submission Format**

Each submission file must contain **predicted sales values** for each id in the test set. The format should be as follows:

id,sales\_hat

840\_2024-06-10,12.01

2317\_2024-06-15,13.32

738\_2024-06-10,14.12

3894\_2024-06-11,3.03

3393\_2024-06-08,53.03

**🚀 Why This Challenge Matters?**

Accurate sales forecasting is crucial for optimizing **planning, logistics, and inventory management** in the e-grocery industry. This competition provides an opportunity to apply machine learning techniques to real-world problems, contributing to more sustainable and efficient supply chain operations.

**🌍 Your Impact**

By participating in this challenge, you directly support **Rohlik Group’s mission** to enhance customer service, improve sustainability, and optimize e-grocery operations.