Data Table Documentation

Introduction

This document provides an overview of the data tables used in this project. The tables are organized into two main folders: `kernel` and `extended`.

The only table that is strictly required is `receivals`, located in the `kernel` folder. However, it is highly recommended to also use `purchase_orders` for improved context and analysis. Other tables are optional and can provide additional insight or metadata.

Receivals

File: `/data/kernel/receivals.csv`

`rm_id`
Unique identifier for the raw material.

`product_id`
Identifier for each specific product received. Each rm_id can have multiple product_id's.

`purchase_order_id`
Links the receival to a corresponding purchase order.

`purchase_order_item_no`
Links the receival to a corresponding purchase order item.

`batch_id`
Each purchase can be split into multiple batches, each batch as a unique id.

`receival_item_no`
Each purchase order item can be split into several receivals. receival_item_no identifies each single receival for the item.

`batch_id`
Identifier for the "batch" this raw material delivery becomes a part of.

`date_arrival`
UTC timestamp of material arrival. This is the timestamp we use to decide which date something is received.

receival_status
Current status in text, such as "Completed". All statuses are valid and counts towards the target variable.

net_weight
Weight of the product alone, excluding packaging or containers. This is the basis for the target variable.

supplier_id
An id for the supplier for each purchase.

Purchase Orders

File: /data/kernel/purchase_orders.csv

purchase_order_id
Order identifier, joinable with receivals and transportation.

purchase_order_item_no
Order item identifier, joinable with receivals and transportation.

quantity
Amount of product ordered.

delivery_date
Expected delivery date. Sometimes placed at the end of the month or year.

product_id
Product identifier. Hydro orders a specific product_id, and later distinguish which rm_id the receival has.

product_version
Version or subtype of the product.

created_date_time
Timestamp when the record was created.

modified_date_time
Timestamp for last edit.

unit_id
ID of the unit used (e.g., kg).

unit
Name of the unit, typically "kg".

status_id
Identifier for status.

status
Status text, e.g., "Closed".

Materials

File: /data/extended/materials.csv

rm_id
Unique identifier for the raw material.

product_id
Identifier for the product.

product_version
Product version identifier.

raw_material_alloy
Name for the raw material.

raw_material_format_type
Physical form of raw material (e.g., block, powder).

stock_location
Storage or warehouse position.

Transportation

File: `/data/extended/transportation.csv`

rm_id
Unique identifier for the raw material.

product_id
Identifier for the product.

purchase_order_id
Linked order for transport.

purchase_order_item_no
Linked order item for transport.

receival_item_no
Identifies each single receival for the production order item.

batch_id
Identifier for the "batch" this raw material delivery becomes a part of.

transporter_name
Anonymized transporter's name (e.g., `transporter_0`, `transporter_1`).

vehicle_no
Anonymized vehicle identifier (e.g., `vehicle_0`, `vehicle_1`).

unit_status
Status such as loaded or in-transit.

vehicle_start_weight
Vehicle weight before loading.

vehicle_end_weight
Vehicle weight after unloading.

gross_weight
Total transported weight including packaging.

tare_weight
Empty vehicle weight.

net_weight
Net weight of material alone.

wood, ironbands, plastic, water, ice, other, chips, packaging, cardboard
Weight of non-material elements used for impurity deduction.