Fulfilment Remote Exercise - Data Science

In this task you will be looking at a particular dataset the Data Science team has extracted. The dataset involves sets of products that are picked up from a store and packed into containers to be delivered to customers.

The exercise is open ended and the expectation is that you will perform some exploratory analysis, visualisation and perhaps some modelling and/or discussion regarding avenues to explore. The task has no predetermined time limit and *what we ask as a deliverable is a .zip file with your work and documentation and/or a report describing your arguments, findings and proposed ideas.*

Dataset

Products denoted by their SKU are picked from shelves within bays in a store and packed in containers. The following fields are available:

- *containerId*: Unique identifier for a container.
- *containerRegime*: Categorical describing regime in which the container is placed.
- *orderId*: Unique identifier for an order to which the products belong to.
- *skuId*: Unique identifier for the product sku.
- *quantityToPick*: Number of Items for a given sku that need to be picked to fulfil the order.
- *locationId*: Unique identifier for the position where the product is located. It is composed by the bay identifier and the shelf number (number after last hyphen).
- *completionTime*: Timestamp indicating the completion of the process where the product is picked and placed in the container.
- *container_pt: Partition timestamp.*
- *width:* Width dimension of the product.
- *height*: Height dimension of the product.
- *weight*: Weight of the product.
- *depth*: Depth dimension of the product.
- *seq*: Sequence number that specifies where a product is located in the store using a single number. Close numbers for two products means they are closely located in the store; -1 value means this product has an unknown sequence number

Additionally the dimensions of all containers are:

CONTAINER_WIDTH_CM = 65.0
CONTAINER_HEIGHT_CM = 20.0
CONTAINER_DEPTH_CM = 40.0
CONTAINER_MAX_WEIGHT_KG = 20.0

Instructions

We would like you to use Python for this exercise, but use whatever other tool(s)/libraries you would like, as long as you explicitly call out their usage. The exercise will be evaluated on the basis of the following criteria:

1. **Exploratory analysis**. We will look at what you decided to explore, visualisations and insights. We are keen on all types of observations you think we might find interesting.
2. **Reasoning**: Given your findings what motivated your next steps.
3. **Data Science Outputs**: Findings, models, predictions, etc.
4. **Communication**: Ability to explain and justify process and outputs.

The work will be reviewed by data scientists and product managers - they will receive the assets you send back- so make sure you are able to convey what you did to this audience.