**Goal**

Buying parts can become very messy very quickly in the whole supply chain cycle!

Data science can be extremely useful to automatically organize the items in categories so

that they can be easily found by the engineers for the product that they are building.

The goal of this challenge is to look at product purchase history and create categories of items that

are likely to be bought together and, therefore, should belong to the same section.

**Challenge Description**

In the current version of the supply chain, we have manually grouped the items into a few categories intuitively.

However, they now have a lot of data about product purchase history. Therefore, they would like to

put the data into use!

This is what they asked you to do:

* The company founder wants to meet with some of the best products to go through a

focused demand management. You are asked to send the ID of the following products to the

founder:

* Top 5 percentile of the products that bought the most items overall in their lifetime
  + - For each item, the product that bought that item the most
* Which are the top 3 most expensive products to manufacture?
  + Find out the correlation between Lead Time and Cost. Do items with higher Lead Time cost more than the items with Low Lead Time?
  + If we want to focus on reducing Lead Time of items by placing them on certain specialty programs. Which parts should we focus on?

* Cluster items based on products co-purchase history. That is, create clusters of items that have the highest probability of being bought together. The goal of this is to replace the old/manually created categories with these new ones. Each item can belong to just one cluster.

**Data Definition**

We have 2 table -

The 2 tables are:

"item\_to\_id" - for each item, it gives the corresponding id

Columns:

* Item\_name : the name of the item
* Item\_id : the id of the item. Can be joined to the id in the other table. It is unique by item
* Item\_cost : cost of the item in USD
* Item\_lt : Lead time of the item in days

"purchase\_history" - for each product purchase, the items bought

Columns:

* product\_id : the id of the product.
* id : comma-separated list of items bought together in that transaction.