**Project : Customer Segments**

So far, you've seen Supervised Learning from start to finish, twice. The workflow you practiced provides the strongest foundation for applied machine learning.

Yet here's the best part... the workflow for Unsupervised Learning is a **simpler version** of the one for Supervised Learning.

Because there's no target variable to predict, there are no widely-accepted performance metrics for evaluating models. That means we can actually **skip some steps**, such as declaring hyperparameter grids and implementing cross-validation.

Up to now, the datasets have also been in the right format for machine learning. Sure, you had to do data cleaning and feature engineering, but you didn't need to drastically restructure the data.

However, often, some of the most interesting machine learning applications require you to creatively restructure (a.k.a. **wrangle**) the data. This project is an example.

By nature, Unsupervised Learning tasks are more **open-ended**.

For the sake of efficiency, we'll drastically condense exploratory analysis, data cleaning, and feature engineering and focus on the new concepts.



*Cluster analysis is very popular and effective for e-commerce, retail, and digital marketing.*

**Project Context**

Our client is an online retailer. They sell all-occasion gifts, and many of their customers are wholesalers.

* Most of their customers are from the UK, but they have a small percent of customers from other countries.
* They want to create groups of these international customers based on their previous purchase patterns.
* Their goal is to provide more tailored services and improve the way they market to these international customers.

**Current Solution**

Currently, the retailer simply groups their international customers by country. As you'll see in the project, this is quite inefficient because:

1. There's a large number of countries (which kinda defeats the purpose of creating groups).
2. Some countries have very few customers.
3. This approach treats large and small customers the same, regardless of their purchase patterns.

**Our Role**

The retailer has hired us to help them create customer clusters, a.k.a **"customer segments,"** through a data-driven approach.

* They've provided us a dataset of past purchase data at the transaction level (you'll see why this will be important).
* Our task is to build a clustering model using that dataset.
* Our clustering model should factor in both **aggregate sales patterns** *and* **specific items purchased**.

**Problem Specifics**

It's always helpful to explicitly lay out the problem specifics before starting.

Deliverable: **Clusters for customers**

Machine learning task: **Clustering**

Target variable: **N/A (Unsupervised learning)**

Win condition: **N/A (Subjective results)**

Now that you have the context and problem specifics, let's dive right in!