# Objective

The objective is to identify potential customers who are highly likely to purchase a particular category (**cl, kids, npr, jj, sun, powered sun and non powered sun**) based on their similarity to existing buyers. The model aims to assign a **similarity score** to customers who have never purchased a particular category from us by comparing their characteristics with those of customers who have previously bought that category.

By leveraging this similarity score, we can effectively **target and cross-sell** the desired category to high-potential customers.

# Methodology

For my analysis only numerical variables were being used to generate the similarity scores for each category.

The process to generate similarity scores for customers who have not purchased the particular category involves the following steps:

1. **Segmentation of Customers:**
   * The dataset is divided into two groups within each customer category (P1 to P6):
     + Customers who have purchased the category.
     + Customers who have not purchased the category.
2. **Feature Weight Calculation Using JS Divergence:**
   * To quantify the importance of numerical features, Jensen-Shannon (JS) divergence is used to measure the difference between the distributions of the two groups.
   * This helps in determining how distinct the feature distributions are between buyers and non-buyers.
   * The obtained JS divergence values are then used to derive **relative feature weights**, reflecting the importance of each numerical variable.
3. **Creating the Representative Target Vector:**
   * The representative target vector is calculated by multiplying the computed feature weights with the mean values of the numerical variables for the buyer group.
   * This weighted center serves as a reference profile representing typical characteristics of customers who have purchased the category.
4. **Similarity Score Calculation:**
   * Cosine similarity is used to measure the closeness between the representative target vector and each customer in the non-buyer group.
   * This similarity score indicates how closely a non-buyer customer resembles the typical buyer profile.

Cosine Similarity magnitude will lie between 0 and 1. The larger the value, the closer a non-buyer is to the buyer profile.