

7. Post processing

Along the adjustment steps described in the previous page, we have other post-processing strategies to tune the results based on the needs of the different product families.

Hard boundaries

A hard-boundary is a feature for which if the value **differs** between a target CPC and a comparable, we **filter out** the comparable from the set.

For example, if region is a hard-boundary, we only compare CPCs having the same region as the target.

The hard boundaries are decided based on business intuition and analysis of the business teams. They are defined at a product family level.

- If hard boundaries drive a big price impact, they should normally already be captured by the model.
- Note: **We should aim to add as few hard boundaries as possible.** Adding too many hard boundaries is not recommended because it impacts the number of CPCs in the comparable set as well as the coverage.

Inverse boundaries :

Inverse boundaries are the **opposite** as hard boundaries.

We currently use it on shipto_code and soldto_code to not compare with CPC on different products of the same customer.

Volume boundaries :

A **threshold** can be set at a product family level to limit the comparison of CPC having a **too big difference** in volumes.

This is an **additional security** when the volume adjustment function is not enough to cover high variations in volume.

If a threshold of 10 is used, it means that a CPC will only be compared with others in a range of 10x its volume, be it lesser or greater. (e.g. a CPC with a volume of 1 000 can only be compared to CPCs with a volume between 100 and 10 000).

Similarity threshold

To define a similarity threshold, we set a matching percentage for certain features that must be identical between comparables.

```
"match_percentage_similarity_threshold": 0.4,
"match_percentage_cols": {
  "shared": [
    "country_shipto",
    "end_use",
    "gbu_customer_seg",
    "product_group"
  ]
}
```

=> the threshold is currently set to 0

For each target CPC, all comparable CPCs are ranked according to similarity distance, then we select the set of comparable CPCs as the minimum between the top 10 and the threshold for final price calculation.