

## **Medicine Matching Code Documentation:**

1. Nomenclature of medicine - Brand | Strength | Medicine Type | Number, For Example  
- ADPANTA 40MG TABLET 10S
2. The structured format of the distributor data is - ['item\_code', 'brand', 'pack',  
'manufacturer', 'catg', 'subcatg', 'mrp']
3. We apply synonyms to both distributor and Master datasets as the this sheet -  
[Synonyms](#)
4. We will be concatenating Pack info with brand in distributor in the following way - if  
the pack is empty we will avoid concatenation, if the pack is non-empty but already  
present in the brand name then we will skip concatenation, and if the pack is  
non-empty and not already present in the brand name then we will concat the brand  
with pack.
5. Now we will be lower casing to both distributor brand and master brand.
6. Now we will apply some standard replacements to the brand like - '-' -> '' etc
7. Now we will be classifying the SKU's into 4 different categories - tablet, capsule,  
injection, syrup, others based on the brand
8. We will be handling each of the categories differently (Syrup and others category are  
handled together)
9. Firstly, we match the exact medicine name (cosine similarity = 1) and would want  
max cases to exactly match after applying predefined replacements
10. Secondly, for the remaining medicines we will be calculating expected SKU brand  
and match with the master table SKU brand to reduce the search space

For Example - Brand - ADVEN PILCARE DROP 30ML

Exp\_sku\_brand - ADVEN PILCARE DROP

Exp\_sku\_brand2 - ADVEN PILCARE

Exp\_sku\_brand3 - ADVEN

11. We will be getting  $\geq 1$  matches for the remaining medicines ( $>90\%$  of max cosine match) which have exact SKU match
12. We will be modifying the cosine value for cases where (Syrup & Suspension), (Suspension and oral suspension), (Powder and dusting powder) or (Tablet and Capsule) in brand name in master and distributor data by replacing them with blank space and applying cosine
13. The remaining medicines will not have any matching in the master table