Finding Nemo - Active Work

1. **Business Case: Attribute Normalization for flavor, scent, and color**

   Normalization means identifying a limited list of attributes that we want to use for matching each department.

   Normalized - condensed into specific groups of attributes that are most prevalent across the category.

1. **Challenges:**

   The attribute value and attribute code from PIM table would be used to determine the predicted attribute ' for GTIN_NO.   Below are challenges:

   - The GTIN_NO might contain multiple predicted attribute values
   - There is similarity among attributes.
   - The predicted attribute values do not show the relevance.

1. **Assumptions:**

   - Used predicted attribute values/completeness to determine the relevance of attribute values which would be used as normalized attribute.
   - The attribute normalization is product categories related.
   - The assumed target areas within customer facing taxonomy for Flavor are Beverage, Pantry and Candy
   - The assumed target areas within customer facing taxonomy for Scent are Beauty, Personal Care and Cleaning & Household.
   - The assumed target areas within customer facing taxonomy for Color are Beauty, Kitchen & Dining and Home Décor

1. **Methodology:**

   - Match process

     Utilized ML algorithm rapidfuzz.process.extract to match the attribute value and attribute code from PIM table 'CURIDQ.PIMMART. DIM_ATB_INF_LKP' with the DPT_DSC from PIM table 'CURIDQ.PIMMART.ITEM_VIEW' to create the predicted attribute 'attrib_field'.

   - Normalization
     - Created 'PERCENT_CUMULATIVE_COUNT' by cumulating percent prevalence of single attributes from 'attrib_field ' and ranked from top. (Excluded attribute value = 'OTHER').
     - Picked up top n from the 'PERCENT_CUMULATIVE_COUNT' to create "normalized_list".
     - Consulted with SME to adjust the "normalized_list" and add 'OTHER','MULTI xxxx'].
     - Utilized "normalized_list" to against value in 'attrib_field ' to get the 'NORMALIZED_xxxx' .
     - If value in 'attrib_field ' was also in "normalized_list", assigned the value in the "normalized_list" to create 'NORMALIZED_xxxx'.
     - The xxxx might be flavor, or scent, or color.