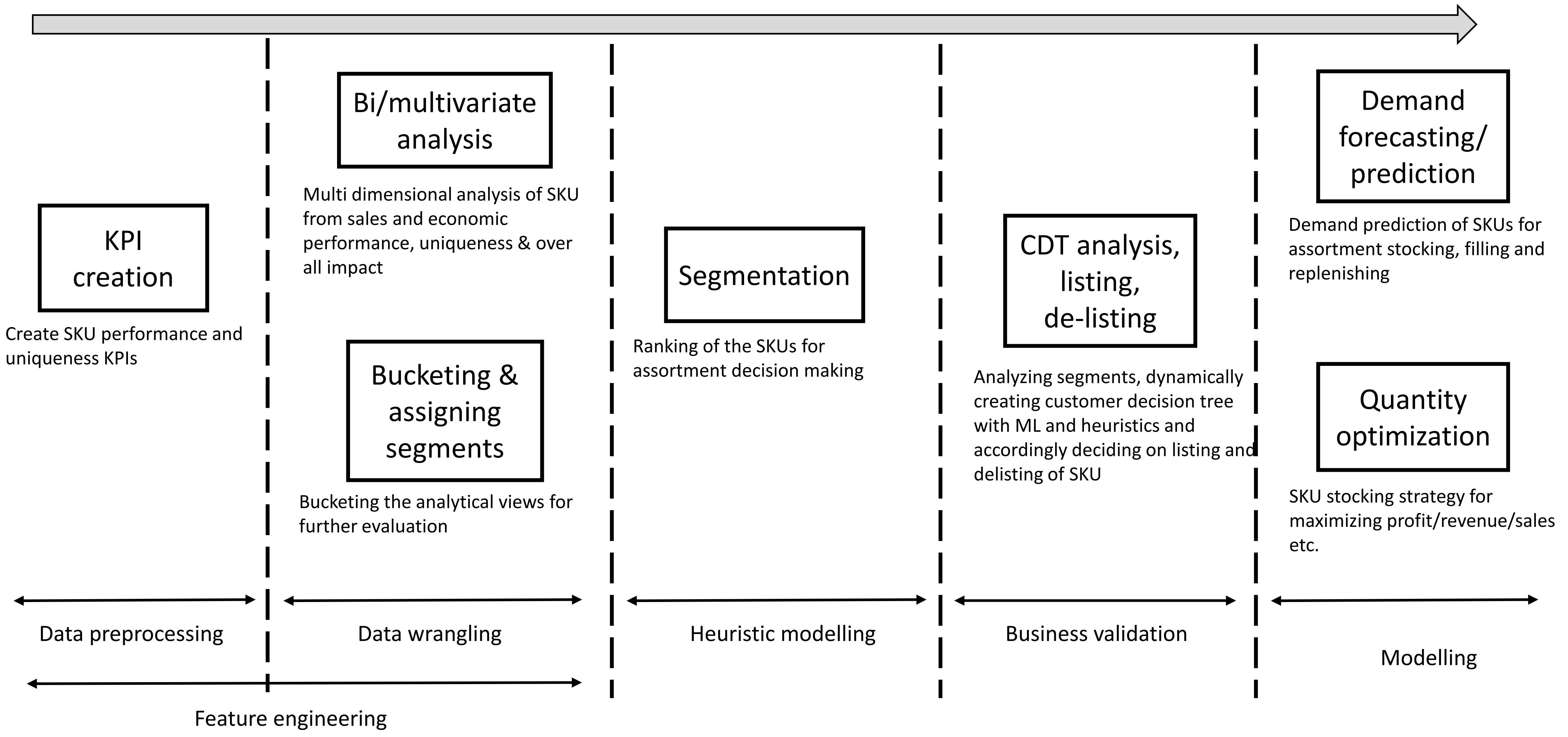
Assortment planning (data science & analytical) model pipeline high level overview



If the assortment planning is seasonal, this approach (in a pipeline) can be easily deployed/is dynamic and scalable enough to run and to handle all angles and exceptions.

Financial, operational and strategic elements are the fundamentals for taking assortment decisions in a business. With the evolving customer base and behavior, assortment optimization is a continuous cycle, in which ***de-listing and listing of SKUs*** is done by assessing SKU performance from multiple dimensions and need identification.

SKU performance is measured by calculating/deriving multiple KPIs like:

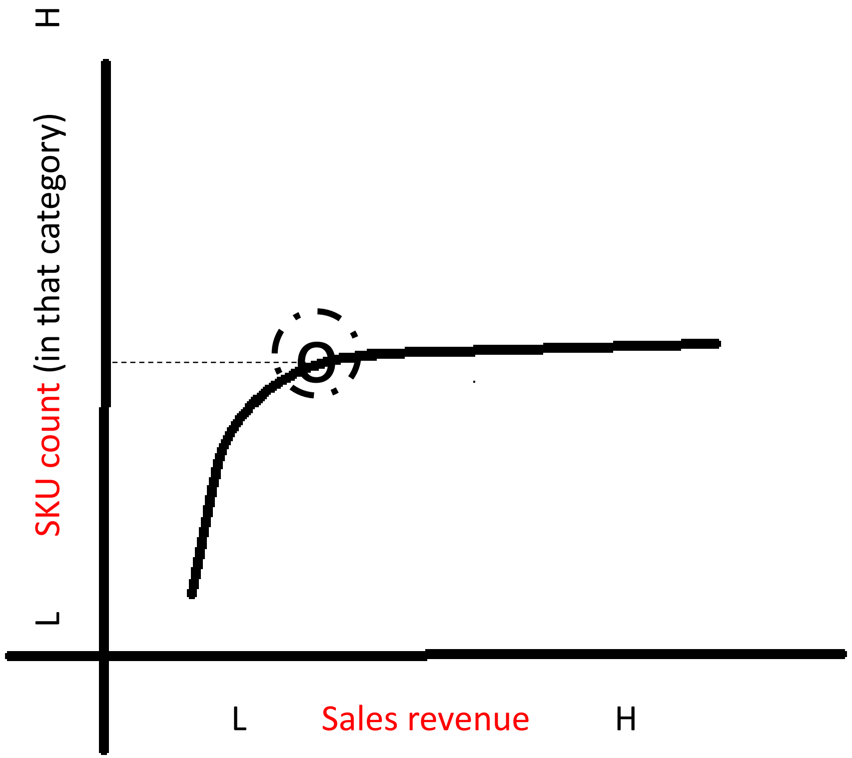
Financial performance KPIs	Sales Analysis KPIs	Cost analysis KPIs	Supply chain KPIs
• **Gross profit margin** = [(revenue – COGs)/revenue] • **ROI** = [net profit/COGs] • **Contribution margin** = [(revenue – variable costs)/revenue] • **Sales revenue** = total revenue generated in FY or a time frame	• **Sell through rate** = [Total units sold/(total units sold + EOP inventory)] • **Sales per week per store** = [total sales of SKU/No. of stores] • **Volume of sales** = Avg. demand • **Demand variability** = [standard deviation/avg. sales] • **Full price realization** = [full price net unit sold/total net unit sold]	• **Cost to serve** = [total cost to serve/Number of units sold] • **Inventory holding cost** = [avg. inventory value of SKU/total inventory value] • **% contribution of sales in its category** = [sales of SKU/total sales of category]	• **Inventory turns** = [total sales/EOP inventory] • **Days of supply(cover)** = [EOP inventory/avg. daily sales] • **Inventory turnover** = [COGs/avg. inventory]
COGs = cost of goods sold; EOP = End of period inventory			

Exhibit 1

Need identification of SKUs is measured by:

- **Uniqueness** of SKU [SKU **substitutability**]
- **Customer decision tree** creation as per product lines

Choice count analysis (category wise)



This point shows the optimal number of SKUs that should be planned (under that category). Increasing the no. of SKUs won't optimize the assortment and might eat up more space, also leaving the customer confused.

SKU performance analysis

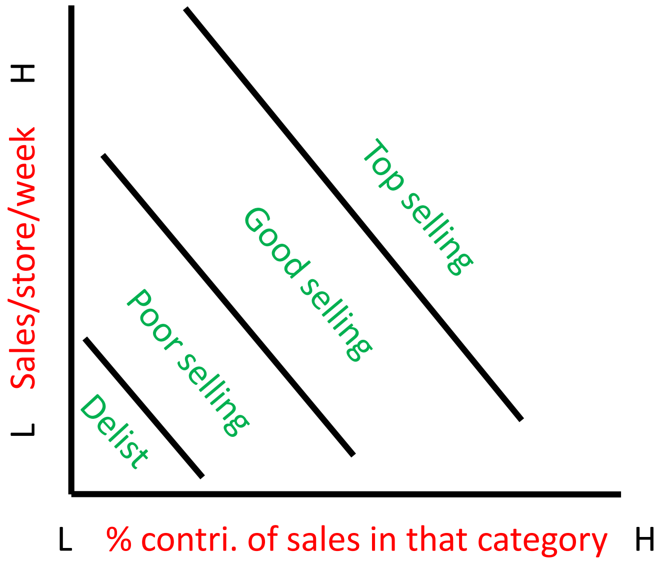


Exhibit 2

Calculated KPIs

Grouping/assigning names as per buckets and analysis

Diversified individual SKU performance

Multidimensional SKU evaluation

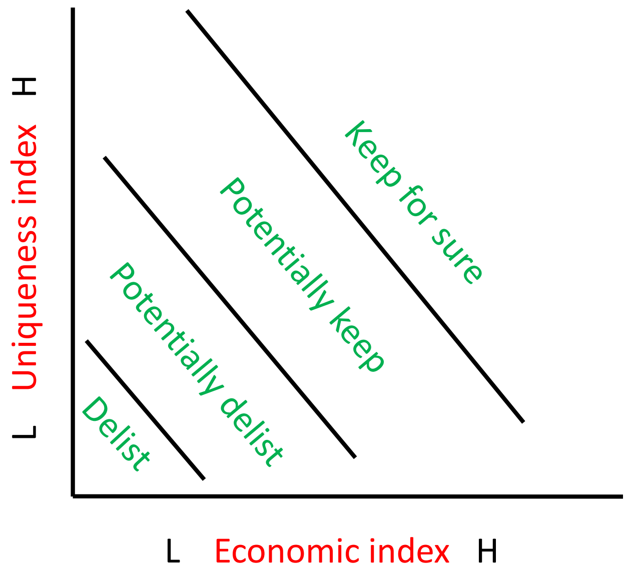


Exhibit 3

Uniqueness, can be a measurement of need identification of the SKU. It will help category planner to understand the product diversity better. Later while substituting, similar characteristics will be considered.

SKU uniqueness index calculation (all SKUs under one category)

- **Jaccard similarity coeff.:**

  Data structure – SKU: {bought by customer set}
  SKU1: {customer1, customer2, customer3}
  SKU2: {customer2, customer3, customer4}
  SKU3: {customer1, customer4, customer5}
  ***Jaccard similarity between 2 sets (of SKUs) =** Intersection of sets / Union of sets*
  Or, use sklearn.metrics jaccard_score function

- **Euclidian distance:**

  *Data structure – SKU: [weight, shelf life, price, purchase frequency, **Market basket support & confidence score**]*
  SKU1: [1.2, 30, 36, 2.5, 0.3, 0.5]
  SKU2: [0.8, 25, 45, 5.0, 0.6, 0.2]
  SKU3: [1.0, 18, 30, 3, 0.6, 0.9]
  ***Euclidian distance =** np.linalg.norm (vector1 - vector2)*

Note: Take the avg. Jaccard similarity score and avg. Euclidian distance of every SKU

Uniqueness index per SKU = w1*Jaccard score + w2*Euclidian score

w1 & w2 are weight and can be tweaked as per business objectives as Jaccard score is based on customer purchase pattern and Euclidian score is as per product description

SKU economic performance index calculation

Economic index = w1*sales revenue + w2*profit margin + w3*ROI + w4*Inventory turnover

Normalize the values to bring it under same scale; w1,w2,w3 & w4 are weights are can be tweaked as per business needs i.e. e.g. is goal is more profit then assign more weights to w2 & w3, but if goal is more revenue then high weights to w1 & w4

Areas to grow or optimize

Overall Impact of SKU in that category assortment

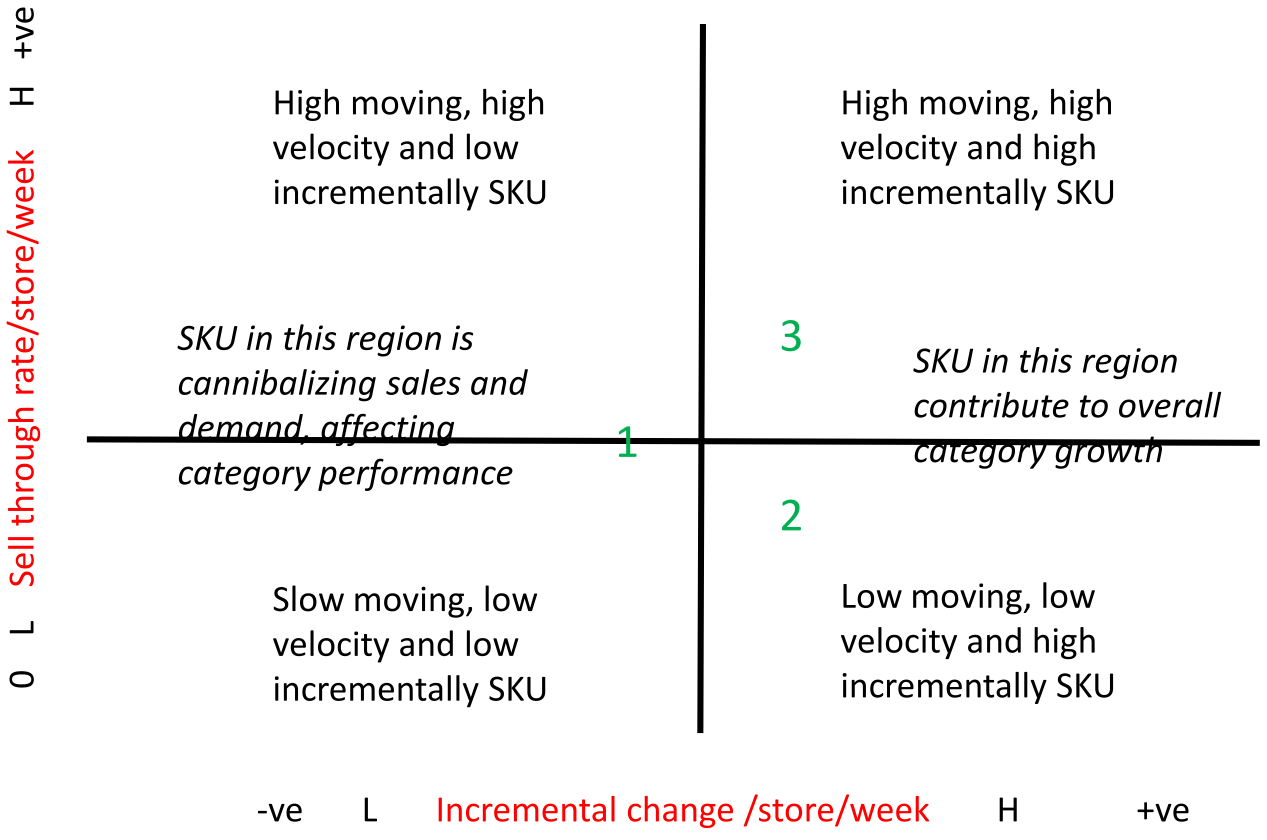


Exhibit 4

1: Cannibalizes sales and category performance
2: Nurture/substitute
3: High performing

Incremental value of SKU calculation within a category

By new SKU introduction or within existing SKU base

Baseline revenue = No. of existing SKUs * average weekly revenue per SKU
Total impact = New total revenue
Incremental value = mod (Total impact – baseline revenue)
Incremental value per store per week* = *Incremental value/(no. of store * no. of weeks)

Reduction rate in sales per SKU per week = Average ([previous sales – new sales]/previous sales)

{ Note: applicable only if +ve i.e. prev. sales > new sales }
{ If –ve i.e. prev. sales < new sales, then, 1 + above equation value = increased rate }

Cannibalization or Increased = ***Avg. Reduction in sales per SKU or Increased rate * Avg. price per unit***

Adjusted Incremental value = *Incremental value – cannibalization or Increased*

Adjusted Incremental value per store per week = Adjusted Incremental value/(no. of store * no. of weeks)

Incremental change = Adjusted Incremental value per store per week - Incremental value per store per week
Can also be decremented change if cannibalization happens.

Mining behavioral & attitudinal insights from customer transactions

Replicating customer thought process while buying a product

The decision path taken by a customer to purchase a product (SKU) under one single category depends on multiple factors like :

- Major Features of every SKU (under one category) - [Exhibit 5] *Note: groupings defined as per business understanding and creation of CDT*
  - **Budget**: *price, promotion effective*
  - **Popularity**: *brand reputation, product reviews*
  - **Longevity**: *availability, shelf life*
  - **Specialty**: *product category, specific description (benefits, additional advantages etc.)*
  - **Presentation**: *packaging, size*

- Other features of other products/categories bought with it in that transaction -
  - *Customer demographics*, *basket information* (with what all products was it bought, how many products had similar characteristics even if bought from a different category or product family)

The dataset is **x-variables**: *{all features of the SKU that was purchased, other features of the products/categories and customers demographics – customer X transaction level data}* and **Y-variable** is the *SKU {SKU1, SKU2, SKU3… etc. of one single product category}*
Fit a decision tree classifier ML algorithm that predicts the purchase of a SKU as per the transaction patterns of different customers.

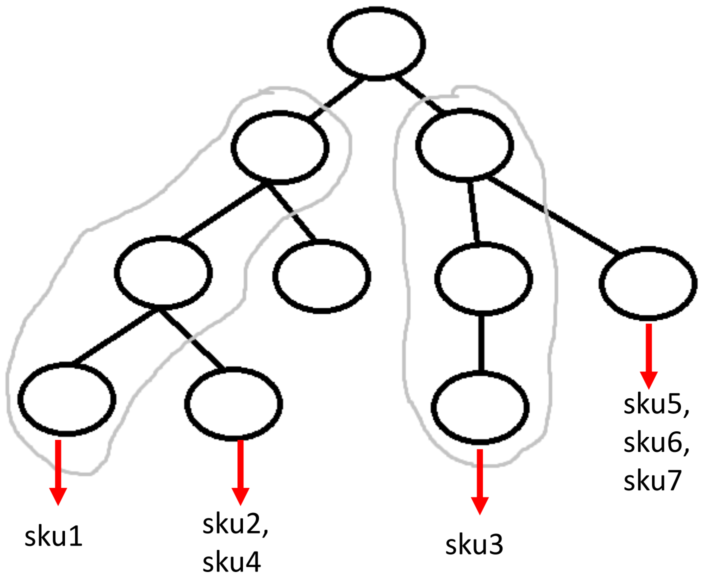
In a decision tree, the splits on features are prioritized on the information gain. The combination of splitting decisions, when the decision tree is created/trained to it's complete depth, we get the maximum information gain, that helps us to reach to that particular SKU (under that category) being bought by a customer. These priority wise splitting decision can me mapped to a customer buying pattern behavior for purchasing the product.

Understanding the tree to create CDT :

- Running the decision algorithm for individual category to predict the purchase of SKUs (under that category)
- Visualize the tree and understand the splitting process that get to the SKU
  - Different categories will have different combinations in which feature splitting process is prioritized to get the maximized information gain

<u>On the basis of the major features group of every SKU (under one category) (Exhibit 5), create the Customer Decision Tree (CDT) as per the priority wise splits that the DT model suggested to reach that product, thus replicating a customer buying decision behavior/pattern.</u>

Understanding the purchase behavior of customer through product dimensions (uniqueness)

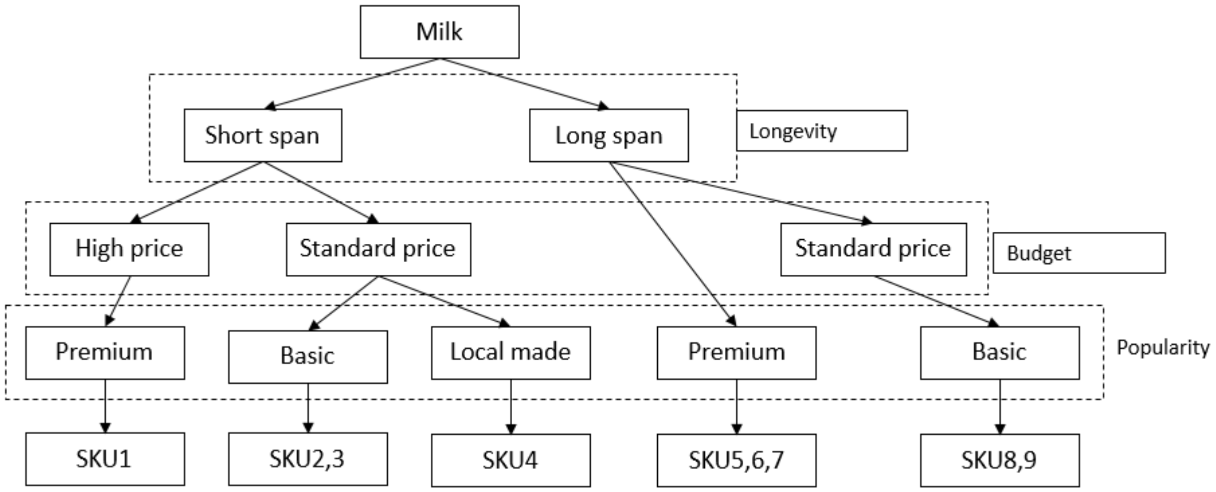


To buy **sku1** & **sku3**, customers drill down by taking a **specific decision path**, thus making the SKUs **unique**. These customers are very ***<u>less likely to change to a different SKU</u>*** as they are looking for a specific class of product under that category. Thus planning these SKUs in the assortment would ***improve customer satisfaction and retention***.

Sku2, **sku4** has a litter lesser likelihood of customers switching to a different product, as for purchasing these, customers too take a long decision path. So if these products aren't performing up-to expectations *(can be accessed with Exhibit 2, Exhibit 3 and Exhibit 4)* category managers ***needs to either nurture them with proper marketing strategies or substitute them for generating overall category growth***.

Sku5, sku6 & sku7 has higher likelihood of customers switching from one product to another as they can be similar, mainstream and commonly purchased products under that category.

<u>For example</u>, running the DT model on milk category products (which has multiple milk SKUs in it), the model created the decision tree, and from a certain decision node split, ***the priority of split was according to longevity, budget and popularity.*** So while creating the CDT for milk, create it as below.



<u>Another example</u>, running the DT model on detergent category products (which has multiple detergent SKUs in it), the model created the decision tree, and from a certain decision node split, ***the priority of split was according to budget, popularity and specialty.*** So while creating the CDT for detergent, follow this decision path to reach the SKU.

Note: This process is dynamic and is easily scalable for any category

Segmentation/ranking of SKUs for assortment decision making

	SKU evaluation KPIs	Derived KPIs using math & data science techniques	Exhibit 2 segments	Exhibit 3 segments	Exhibit 4 segments	Final SKU segment/ Rank of SKU
SKU1	Financial analysis KPIs, Sales analysis KPIs, Cost analysis KPIs, Supply chain analysis KPIs	Jaccard similarity coeff, Euclidian distance, Uniqueness index, Economic index, Incremental change (cannibalization)	Top selling, Good selling, Poor selling, Delist	Keep for sure, Potentially keep, Potentially delist, Delist	High performing, Nurture/ Substitute, Cannibalizes sales and category performance	A, B, C

Segmentation model :

Segment A : ***(Top selling || good selling) && (Keep for sure) && (High performing)***
These are the champion SKUs for the business and should be kept/prioritized in the assortment.

Segment B : ***(poor selling) && (potentially keep || potentially delist) && (Nurture/Substitute)***
These are the list of SKUs for which category planners should deep dive into leveraging the customer decision tree analysis to understand the buy pattern or behavior of customers. Accordingly take marketing decisions; replace/substitute with new SKUs of similar characteristics as per CDT so that we don't loose a niche customer base; or do a stock clearance sale and finally delist by moving it to segment C if performance still falls.

Segment C : ***(Delist) && (Delist) && (Cannibalizes sales and category performance)***
These are the poorest performing set of SKUs, that shouldn't be planned in the assortment

These would be the major combinations of segments that will be created.
Similarly for a few other combination of segments, analyze and accordingly assign them the respective groups.

Calculating the right quantity of every SKU

<u>**Demand prediction model :**</u> ***Calculation of SKU demand*** *= [Moving average demand + (Z * Standard deviation of demand)] * (Day wise factor) * (Event and Promo factor) * (holiday factor) * (trend factor for seasonality)*

- *Moving average demand:* Can be last 7 days demand past 4 weeks demand [time range as per business needs]
- *(Z * Standard deviation of demand):* Error factor as per the demand volatilities, where Z is the serviceability and std. of demand can be calc. at daily level or week level
- *Day wise factor:* If clac. Demand at a daily level then use this [calculated by taking past 3 or 6 months of daily per SKU level data, then avg. of all the records of the % of the sales for days of the week]
- *Event and Promo factor:* (promo day sales/Avg. sales of past 15 days), different promos will have different factors
- *Holiday factor:* (holiday sales/Avg. sales of past 15 days), different holiday dates will have different factors
- *Trend factor:* (last year same date forward looking avg demand/last year same date backward looking avg demand)

Demand calculation at SKU level can be done at a daily, weekly or 15 days level, depends on the replenishment cycle that the business decides to do.

<u>**SKU stocking strategy model :**</u>

Lets take an example of how to ***plan the number of each products quantities as per the shelf space allocated*** for *generating maximum revenue and high gross profit margins*

Milk = 100 units can be stocked at once in the shelf

Sku1 demand = 100 units; sku2 demand = 60 units; sku3 demand = 40 units

<u>***Mathematical/statistical model :***</u> *Calculate ratio = sku1:sku2:sku3 = 5:3:2*

So, sku1 = 50 units, sku2 = 30 units, sku3 = 20 units and once it gets finished it can again be restocked as per the same distribution

If the ***objective is to maximize profit***, by placing the right quantities of SKUs as per the demand and the space allocation, create a <u>***Mixed Integer Linear Program model which generates SKU wise quantities and maximizes net profit***</u>.

Milk = 100 units can be stocked at once in the shelf

Sku1 demand = 100 units; sku2 demand = 60 units; sku3 demand = 40 units

Profit margin : sku1 = 3; sku2 = 5; sku3 = 4

MiLp equation = MAX { summation (i =1 to 3) X(i) * profit(i) } - this is give quantity of every product to be kept

Constraint – X(i) (i = 1 to 3) <= shelf capacity

We definitely need to *add multiple constraints, handle multiple edge cases by further adding more components to the equations* and refining it