WORLDQUANT
PREDICTIVE

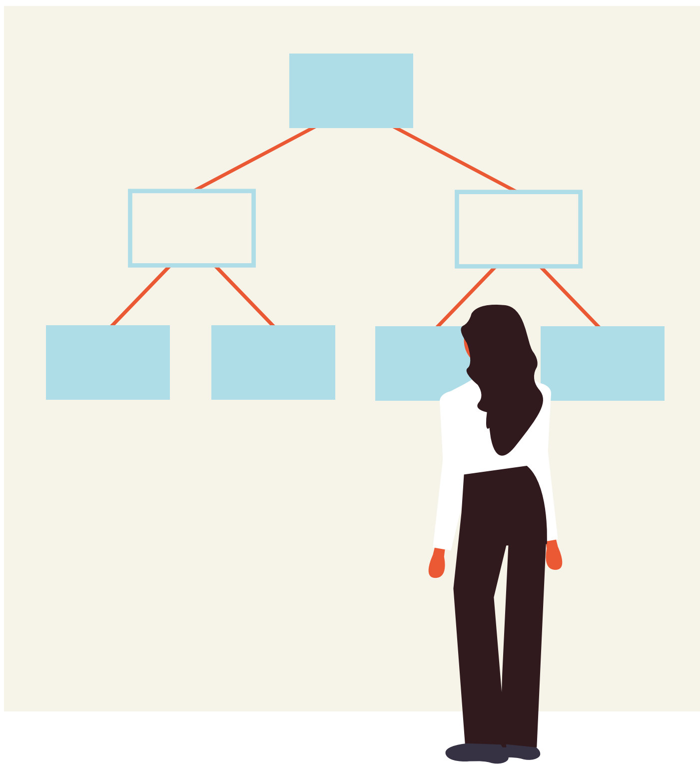
Customer Decision Trees

A Predictive Approach

Jason Yang, Ramana Reddy, Robert Molnar

Customer Decision Trees (CDTs) are powerful visual illustrations of how customers make purchasing decisions based on the product's features & characteristics

> We propose a fully automated method for creating omnichannel CDTs



575 Fifth Avenue, New York, NY 10017

Table of Contents

DEFINING CUSTOMER DECISION TREES

Customer Decision Trees (CDTs) are powerful visual illustrations of how customers make purchasing decisions within a category of products, based on the product's features and characteristics. CDTs convey information about the extent to which products are substitutable or complementary in any given category, and how shoppers make tradeoff decisions when choosing products based on their differentiating factors.

Reading CDTs

CDTs are displayed using a hierarchical diagram in which both the vertical and the horizontal dimensions carry business meaning. Each CDT is organized into a number of horizontal levels and a number of nodes (also called partitions) on each level.

Interpreting nodes and levels

- Each node in the CDT describes the products they contain with a specific set of features
- Nodes are organized in a hierarchical layout, with "child" nodes belonging to "parent" nodes
- Child nodes inherit the features of the parent and contain further differentiating features; there is a part/whole relationship between child nodes and parent nodes
- The lower the level, the more specific the features that define the products in the nodes
- Products in the same nodes compete for the same customers, as they tend to offer similar benefits
- The vertical position of the nodes describes how substitutable products are across the nodes:
    - In higher levels, benefits are less substitutable
    - In lower levels, benefits are more substitutable
- The end nodes of the CDT, known as Competitive Sets or Switching Boxes, represent the most detailed combination of features that are important to customers, beyond which they do not see or differentiate between products

Exhibit 1.1 – Example of CDT I Beer Category I Illustration, Simplified, Vertical Layout

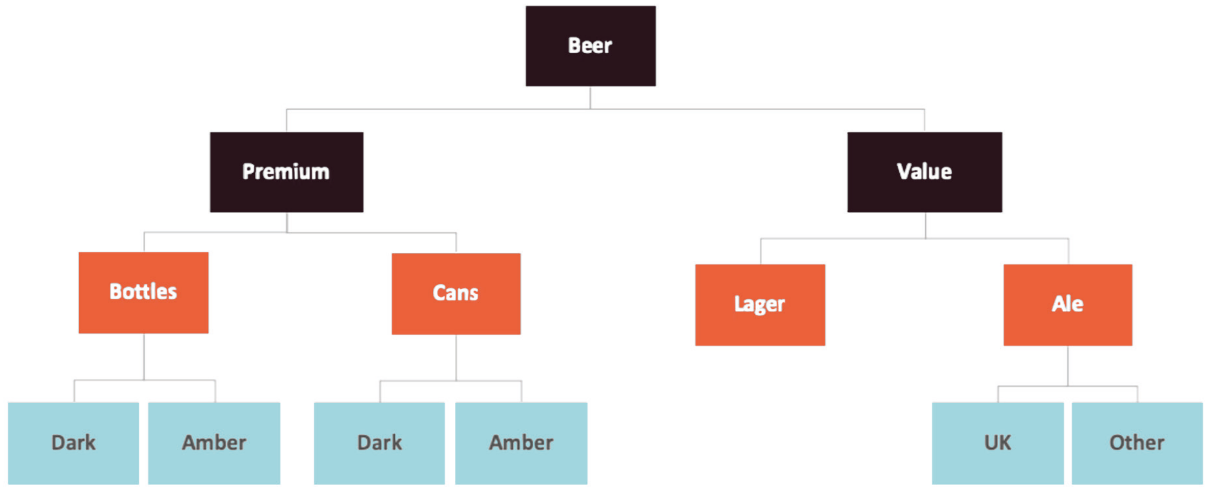
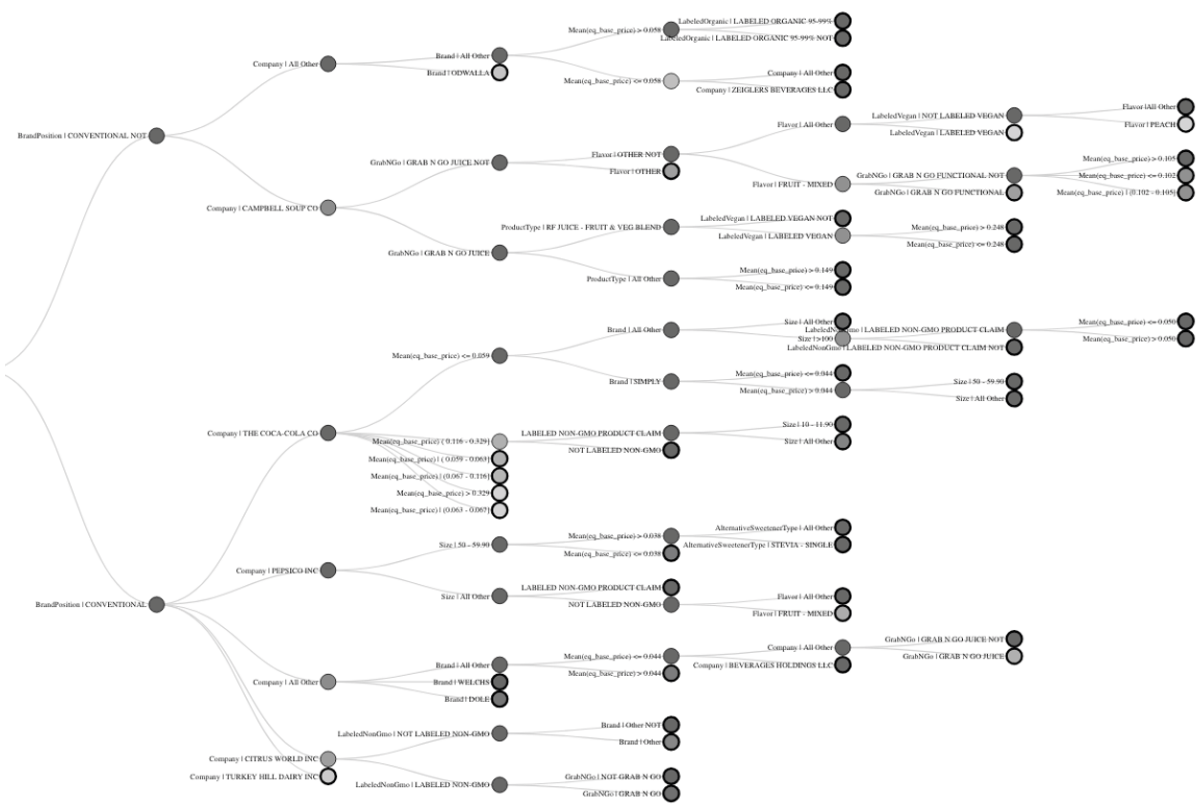


Exhibit 1.2 – Example of CDT I Beverages I Actual Results, All Nodes, Horizontal Layout



VALUE FOR BUSINESSES

CDTs offer a systematic way of segmenting product categories into subsets that reflect similar customer needs and substitutable product benefits. They help retailers and manufacturers understand what attributes are driving customer purchases, customer loyalty, and differentiation, as well as when customers are likely to switch between products.

Using CDTs, marketers have a robust framework for evaluating both gaps in the market and opportunities for innovation, as well as redundancies such as too many products offering the same benefit. CDTs also enable effective pricing, promotion, and placement decisions by identifying the direct competitor of each product in the mind of the customers.

For Retailers

CDTs Help to Identify

- Category segments important to shoppers
- Customer preferences by regions and zones
- Brands driving loyalty and repeat purchases
- Importance of price vs other product features
- Redundant items that customers view as interchangeable
- Gaps in the portfolio for new vendors to fill
- Ideal positioning for private label items
- Effective space planning by planogram
- Optimal assortment in stores and online

For Manufacturers

CDTs Help to Identify

- The importance of brand relative to other features for the customers
- The level of brand loyalty in the market vs other manufacturers
- The relevant competitive sets and the most important competitors
- Redundant (cannibalizing) items in the portfolio
- Gaps in the portfolio for innovation products
- Product differentiators to guide marketing focus
- Potential quick wins from price increases or packaging changes
- Better pricing and promotion strategies
- Optimal assortments by channel and accounts

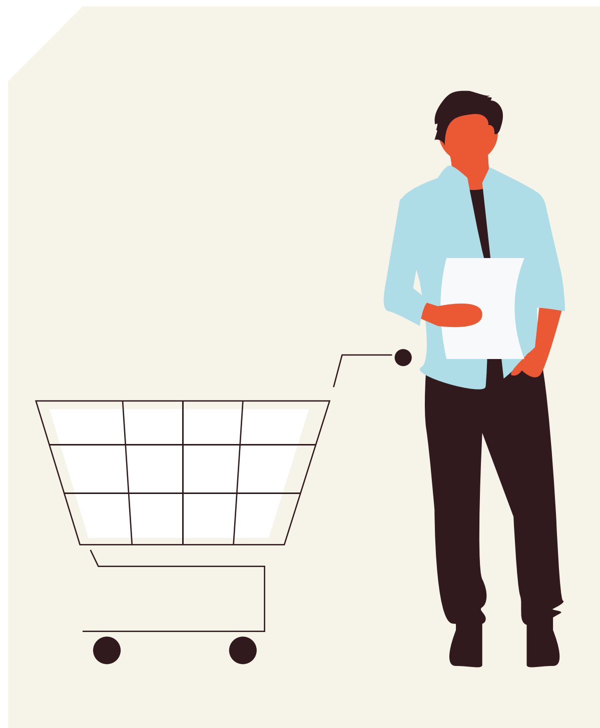
CHALLENGES WITH CDTS

CDTs are traditionally created by analyzing product associations based on consumer switching in repeat transactions. The information on switching behavior is gleaned from shopper panels, point of sale (POS) transactions, or via online browsing behavior.

There are three main types of challenges that prevent consumer-facing companies from accessing meaningful CDTs: data, readability and speed.

Data

Retailers with loyalty card programs and anonymized credit card information have the advantage of being able to identify unique repeat customers. However, this in itself is inadequate to ensure omnichannel coverage and is also insufficient for product categories that are slow-moving and have low repeat purchase rates.



Manufacturers without access to POS data at the customer level rely exclusively on third-party vendors to supply data on repeat purchases or to synthesize preferences for product attributes. Established vendors of third-party data include consumer panel companies and those that syndicate retail data sourced from multiple retailers. Each has its own challenges.

The panel data, while insightful about repeat purchase patterns, is often patchy, only covering a subset of desired product categories, locations, and segments. Syndicated data, on the other hand, can ensure high coverage of the products within most categories and can offer segmentation by market and channel. However, the challenge with syndicated data is that it needs to be purchased in an aggregated shape that can be subject to data consistency errors (e.g., multiple instances of categorical values due to misspelling, missing financial data).

Readability

Once the data is in place, CDT structures are created using specialized algorithms that search for distinct patterns in the purchase behavior of customers. Proper calibration is necessary for these algorithms so that the nodes they create adequately reflect customer behavior and lay out the associations necessary for subsequent classification and labeling.

An equally important part of the methodology is labeling the CDT nodes. Even if the right data have been used and the correct associations identified, the nodes of the CDT are not "human-readable" unless a meaningful label is assigned to them. It is therefore crucial to apply intuitive labels to nodes within a CDT, so business users are able to utilize insights appropriately and easily.

Speed

The ability to quickly create production-grade CDTs has value for both retailers and manufacturers. For retailers, CDTs can help to systematize category reviews, planogram configurations and vendor negotiations. For manufacturers, having CDTs available on short notice can help with marketing decisions, product innovation, and with price and promotion negotiations for key accounts.

Delays can occur in three stages during the CDT creation process: i) data gathering, ii) algorithm development, and iii) node labeling. Without rigorous data and analytics pipelines, each step can add weeks to the delivery time.

Most often, bottlenecks in the process are caused by the last step: the labeling of the nodes. Traditionally, nodes are labeled using an "eyeball" approach by domain experts. This process is both prone to error and time-intensive, requiring set up of templates and spreadsheets, and significant workhours to find common attributes of products within each node.

WORLDQUANT PREDICTIVE APPROACH

CDTs for Brick & Mortar

Data

We combine two types of aggregated POS data provided by established vendors in the market:

- National Shopper Card Data
- Retailer Syndicated Data

Data from both sources are based on a standardized transformation logic with the following characteristics:

- Time horizon: latest two years
- Aggregation: weekly
- Granularity: UPC
- Market: Total US

We also enhance our list of product features with data gathered from online sources (e.g., nutritional content of products). This blend of data can be further augmented with client-specific data available on a UPC level (e.g., additional product attributes, custom KPIs).

Methodology

We create a blend of the two data sources with special indicators to quantify product associations observed in historical customer purchases, then we deploy a proprietary two-step approach:

- One class of algorithms identifies product associations based on the indicators that are not explained by chance or product market share, but instead are driven by customer preferences for certain product features. The result of this process is the dendrogram.
- Another class of algorithms is applied on different vertical cutoff points (i.e., levels) of the dendrogram to classify the nodes based on available attributes. The result is the CDT with labeled nodes, ready for business use.

Creating Dendograms

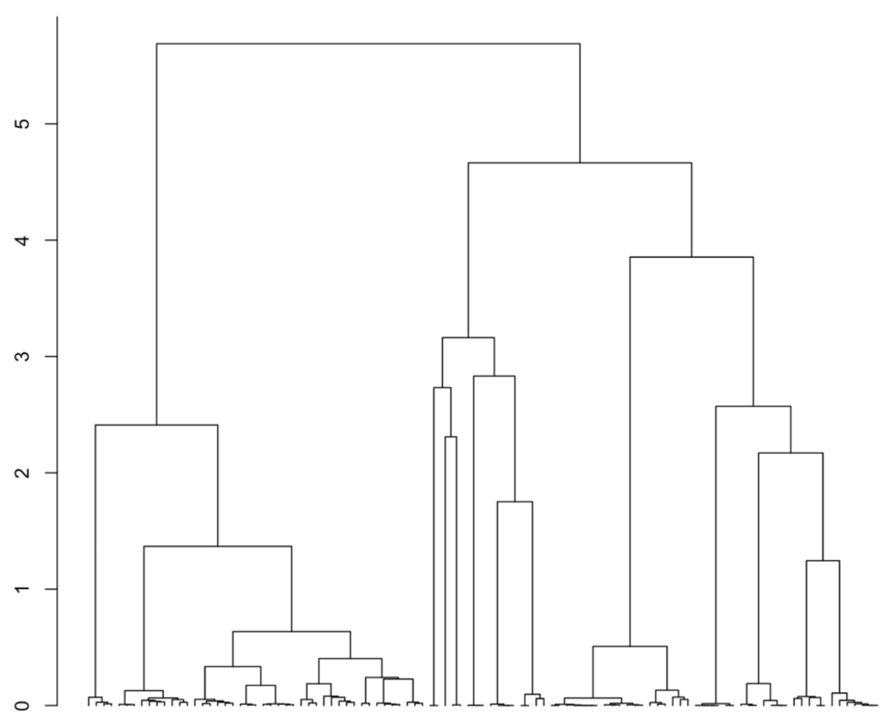
Dendrograms are the first steps in creating CDTs. From a layout perspective, a dendrogram looks similar to a CDT, but has two key differences: the nodes of the dendrograms are not labeled, and each node splits into two additional nodes (vs a CDT, which can have three or more leaf-nodes).

Technically, a dendrogram is a multilevel hierarchy where clusters (nodes) at one level are joined together to form the clusters at the next level. Each cluster of the dendrogram is a group of Unique Product Codes (UPCs), containing products that consumers think of as interchangeable or substitutable. The degree of substitutability decreases with each level of the hierarchy, from top to bottom.

We apply a two-step process to build dendrograms:

- Create measure of association for pairs of UPC, based on transactional data for a given time frame
- Use agglomerative (i.e. bottom-up) clustering technique to arrive at hierarchical clusters.

Exhibit 2.1 – Example of Dendrogram I Illustrational



Creating CDTs from Dendograms

Once the structure of the dendrogram is established, the next step is to label the nodes; this step transforms a dendrogram into a CDT.

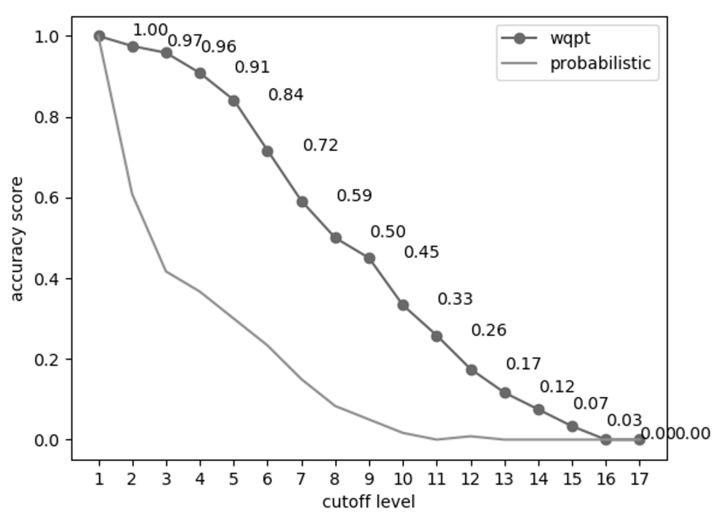
If we take the nodes on a specific level of the dendrogram and the nodes in their respective subtrees as clusters, we can use these clusters as the training data for CDT classification. If we only look at the dendrogram from the top root down to a specific level, that level becomes the “cutoff” level. Systematically experimenting with different dendrogram cutoff levels can determine which level gives the CDT the greatest classification accuracy.

This approach is a creative way to take a dendrogram obtained through unsupervised hierarchical clustering and then use it in supervised decision-tree building. In the absence of prior node labels, generating labels based on picking the optimal cut-off level of the dendrogram turns out to be a useful “bootstrap” approach to creating accurate CDTs.

Measuring Accuracy

The plot below illustrates the classification accuracies of the CDTs built using different dendrogram cutoff levels. It is noteworthy that up to cutoff level=5, the accuracy score remains above 90%. The score deteriorates steadily with the lower cutoff levels as they get closer and closer to the bottom leaf nodes and lose generalizability, thus getting closer to over-fitting the training data.

Exhibit 3.1 – Accuracy Level of Classification I Different Cutoff Levels of the Dendrogram



For comparison, there is also a simple probabilistic approach plotted (in orange) against the cutoff levels. In the probabilistic case, the classification of a sample is the result of the random assignment of a cluster according to its weight in terms of the number of cluster members.

Overall, our CDT—built from unsupervised dendrogram information—performs well and has a significant predictive edge over the probabilistic model. By tuning for hyperparameters such as the cutoff levels, we can build CDTs that not only make reliable predictions but also are conducive to human interpretation.

In summary, WQP's fully automated approach removes bias, reduces the potential for errors, and allows us to create highly accurate and easy to read CDTs significantly faster than the more traditional manual process.

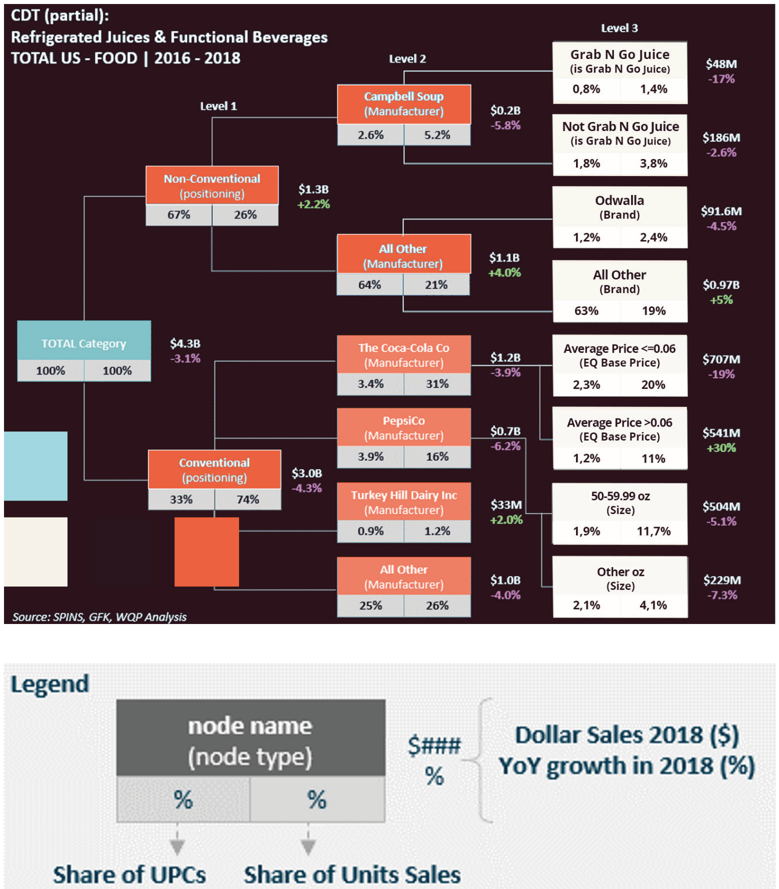
Real World Results

Below we provide an example of applying our proprietary methodology in an end-to-end process to create a CDT for the category of Refrigerated Juices and Functional Beverages in the Total U.S. market (brick and mortar channel).

Using Retailer Syndicated Data, we have inserted KPIs into each of the nodes to capture share of SKUs, share of unit, dollar sales, and year-over-year growth. These indicators provide immediate insights for sizing the subsegments and for assessing productivity.

For purposes of this example, we have structured a series of insights into two parts reflecting various partial snapshots within the CDT. The whole CDT—with 10 levels and up to five nodes per level—is too complex to meaningfully display in this paper. However, in any client project, the whole CDT is used and delivered with an exhaustive list of insights and recommendations.

Exhibit 4.1 – ACDT (3 Levels) Refrigerated Juices & Functional Beverages I Total US - Food 2016 - 2018

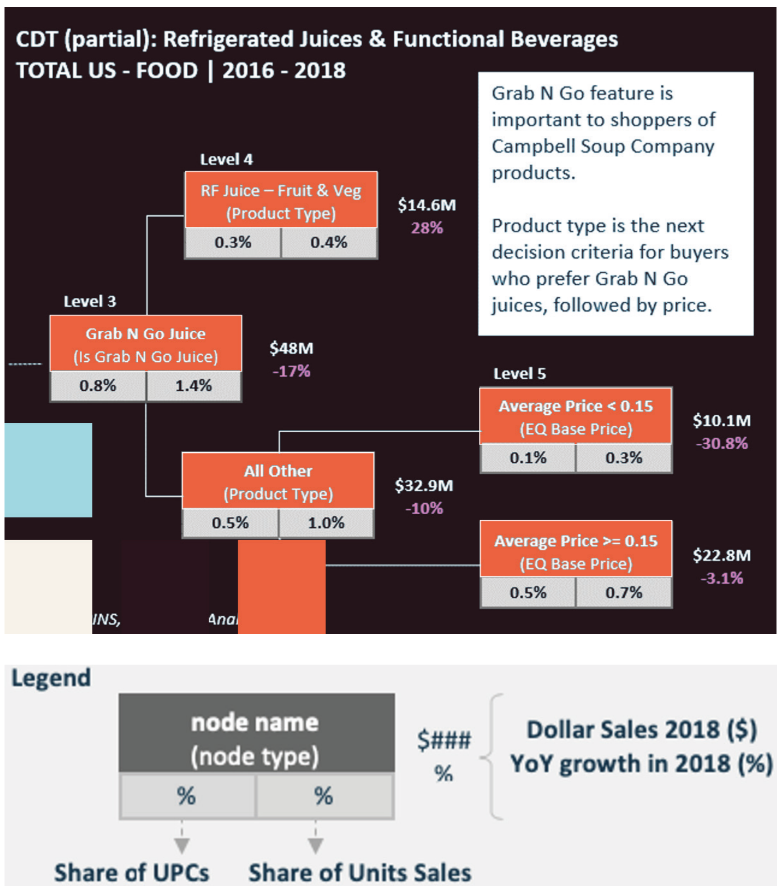


Sample Insights

- With yearly sales of $4.3 billion, the category of Refrigerated Juices and Functional Beverages has declined by 3.1% in 2018 in Total US (brick and mortar).
- The CDT is segmenting the product category by the importance of product attributes in the shopper's purchasing decisions.
- The positioning of the brands (as defined by SPINS) is the most important criteria affecting shoppers' decisions.
- The Non-Conventional group (specialty, natural and functional products) has grown by 2.2%. Although it contains 67% of products, it generated only 26% of unit sales (less than half of its fair share).

- For shoppers of both Conventional and Non-Conventional products, the second most important criteria is the manufacturer of the product (see level 2).
- In the Conventional branch, Coca-Cola, PepsiCo and Turkey Hill have the highest brand loyalty, while Campbell Soup is the most differentiated in the mind of the consumer.

Exhibit 4.2 – CDT (3-5 Levels) I Refrigerator & Functional Beverages I Total US – Food I 2016 - 2018



Sample Insights

- Grab N Go feature is important to shoppers of Campbell Soup Company products.
- Product type is the next decision criteria for buyers who prefer Grab N Go
- For shoppers of Fruit & Veg RF Juices, the vegan label is important

- Non-vegan products are outperforming vegan ones, with vegan products in significant decline (year over year) and underperforming vs their fair share
- For shoppers of all other types of juices within Grab N Go, price per ounce is the next most important criteria

CDTs for e-Channel

Data

In building CDTs for e-channels, our approach makes use of the following two data sources, which are available in most online channels:

- Fundamental information of products, e.g., product ID, price, category, manufacturer, etc.
- A list of other products, considered in relation to a given product and ranked by relatedness to that product.

There are three main types of challenges that prevent consumer-facing companies from accessing meaningful CDTs: data, readability and speed.

Methodologies

Two different approaches exist to address the challenges in developing accurate CDTs, each taking a different perspective on the data.

- The bottom-up approach first looks at individual data points and then tries to measure similarities between them. By recursively combining “most similar” points into “meta nodes,” this approach gradually builds a hierarchical tree from the bottom up.
- The top-down approach looks at the initial agglomeration of data points and tries to find reasonable ways to split them into smaller and smaller groups until a stop criterion further down can be met.

The Bottom Approach

The bottom-up approach to developing CDTs works by first measuring the similarities between any two nodes and merging the ones with the most similarities into meta-nodes. By recursively repeating the process on the meta-nodes, we build a phylogenetic tree that encompasses all nodes.

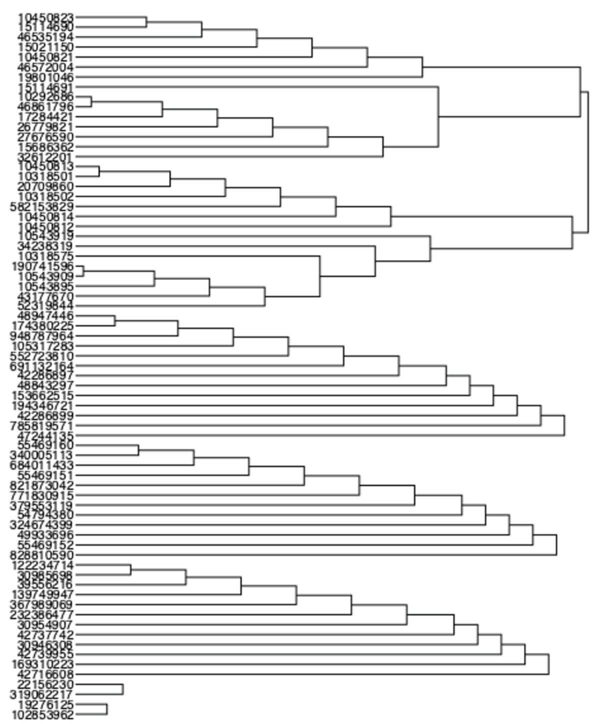
We structure the bottom-up process in two steps:

- We model the data as a graph by viewing each individual product as a single node and viewing the relationships among them, such as “also bought” or “also considered,” as edges. Depending on whether we treat the edges as directed or undirected, we could either use the Info-map algorithm (in the directed case), or the Louvain algorithm (in the undirected case), to label each product with its cluster.

- Further recursively applying the aforementioned algorithm to all the current clusters creates clusters of clusters (meta-clusters), thus gradually building a hierarchical tree from the bottom up.

The following is a hierarchical tree based on using the Louvain algorithm1:

Exhibit 5.1 – Dendrogram Using Louvain Algorithm I Leftmost Numbers Represent Product IDs



The bottom-up approach works by clustering the nodes based on certain similarity measures and recursively working on the clusters. As a first step in exploring the data set, it offers useful insights into the structure of the data and gives a rough idea of how the data should be classified.

The Top-Approach

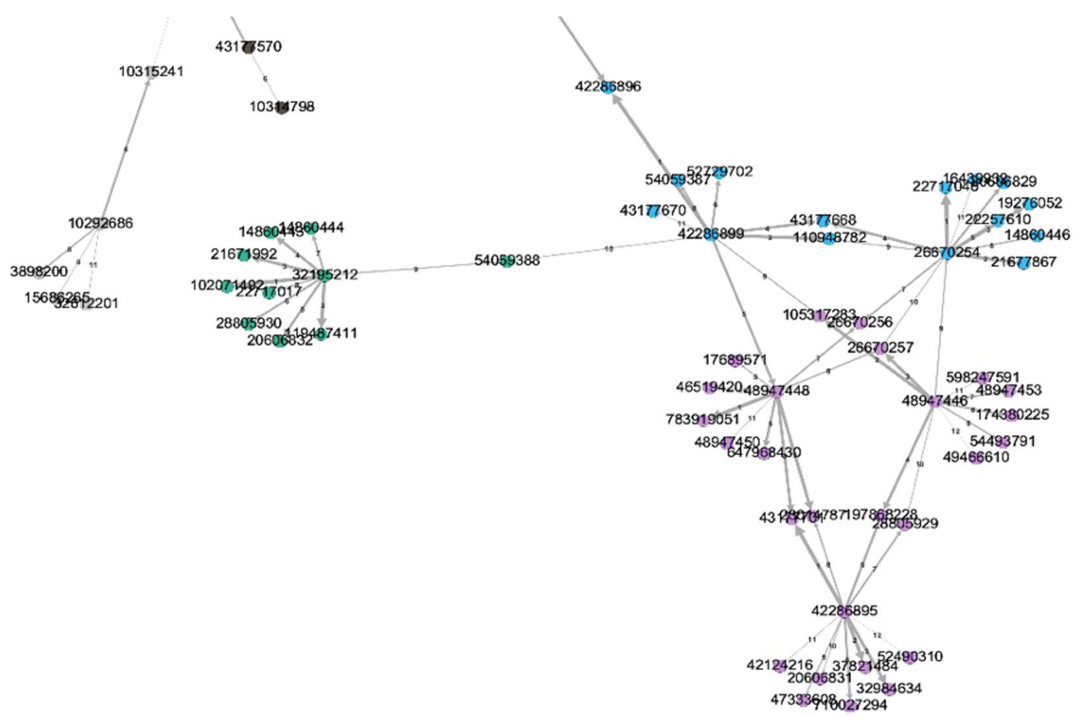
This approach uses decision tree classification, which works by gradually separating the whole dataset, according to a set criterion, into smaller and smaller bins represented by the tree leaves—hence the name, “top-down.”

The top-down approach takes the following steps:

1. Graph Creation

- Using the data of products and their “also considered” counterparts, we are able to create a directed graph, with products being the vertices and directed edges pointing from one product to its counterparts. See example below.

Exhibit 6.1 – Directed Graph



2. **Cluster Detection**
    - Next, we apply the Info-map algorithm2 to detect clusters on the directed graph. When the algorithm finishes, each product will have been labeled with its own cluster.

3. **Data Preprocessing**
    - Next, we apply preprocessing to the fundamental information of the products. We adopt a one-hot encoding scheme for all categorical variables; namely, adding a separate binary feature for each possible value of a categorical variable.

4. **Tree Building**
    - Augmenting the data set we obtained in Step 3 by matching products with clusters, we are now able to set up a classic classification problem in machine learning, with the clusters as class labels and all other variables as features. By applying the decision tree algorithm to this problem, we are able to create a hierarchical tree out of our data.

    - The decision tree algorithm works by recursively splitting the data set on some feature in the data, e.g., brand or flavor. The decision of where to make the split in the data is made based on the principle of entropy (or uncertainty in the resulting split). The algorithm makes the split based on the feature which reduces the most entropy.

    - When a data set contains products of only one cluster (equivalent to when entropy equals 0, i.e., when there is no confusion about cluster assignments, as all products belong to one single cluster), the split will come to an end. The end result would be a hierarchical tree that shows all the split criteria and intermediate data at each step.

Comparing the Two Approaches

Compared with the bottom-up approach, the top-down approach has better interpretability. First, the nodes in the tree automatically have interpretable labels thanks to the explicit split criteria used in the decision tree algorithm. In addition, inherent to the decision tree algorithm is the fact that the features used as split criteria higher up in the tree are relatively more important/informative in revealing the factors that impact purchase behaviors

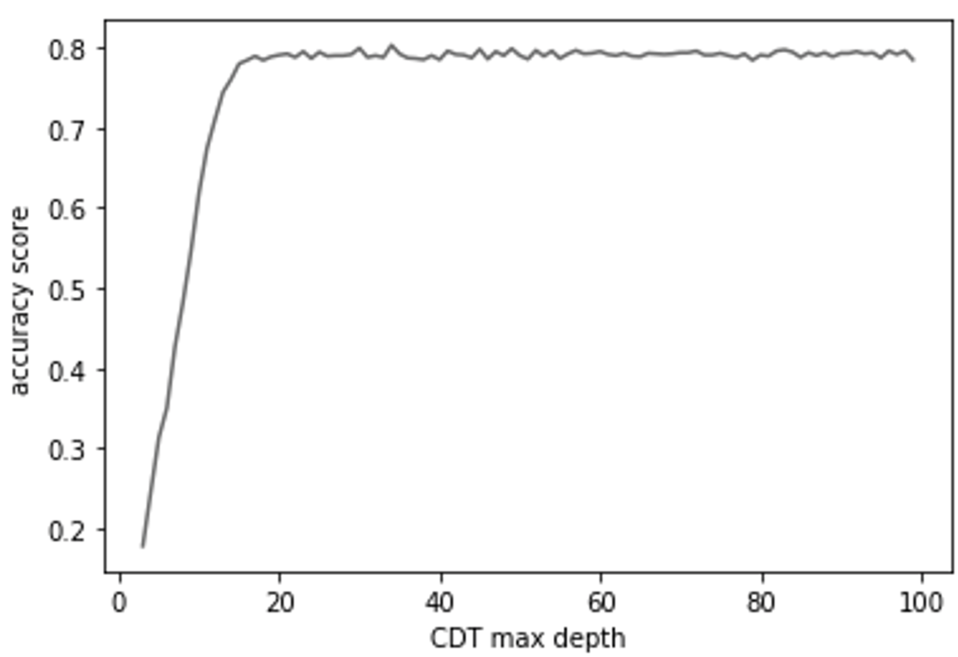
In contrast, CDTs built using a bottom-up approach would have to resort to either subject matter experts' eyeballing or convoluted algorithms to extract the label or topic/motif of a given cluster of nodes which, more often than not, contain data that are so fuzzy and unclean that the topic-extraction algorithms would have a hard time returning intuitive and clear results.

Moreover, as another edge over the bottom-up CDT approach, the top-down CDT can also be used to easily classify brand-new products into different clusters and to naturally obtain accuracy scores or probabilistic estimates for any given product belonging to different clusters/nodes.

Measuring Accuracy

We have conducted experiments by using different CDT max depths, training the CDTs on 80% of all data and measuring their performances on the remaining 20%. The following plots the accuracy scores against the varying maximum depths of the CDTs. It is noteworthy that the best performance can be achieved with no more than max depth = 20.

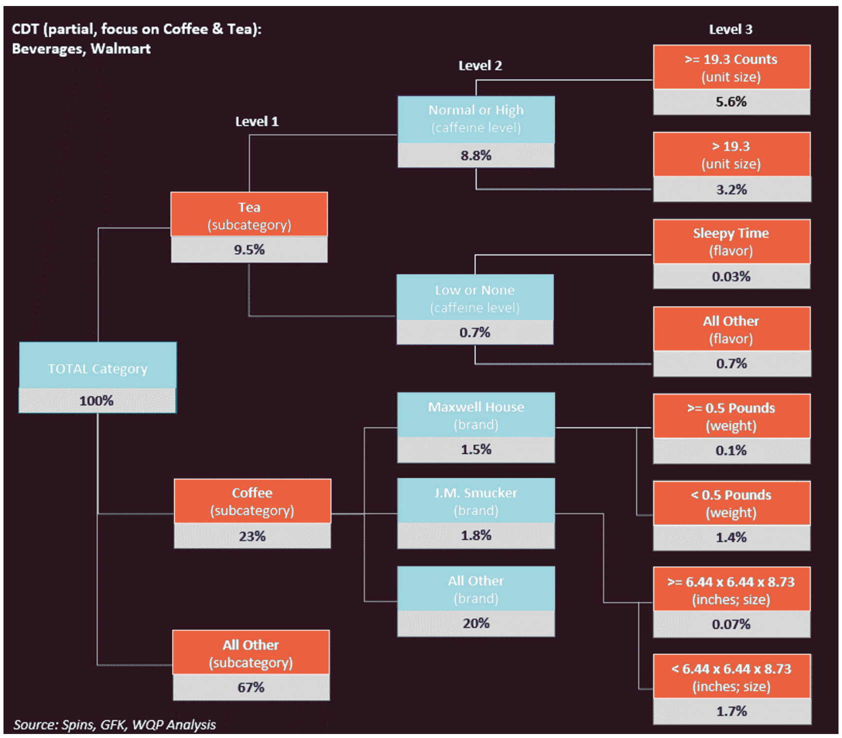
Exhibit 7.1 – Directed Graph

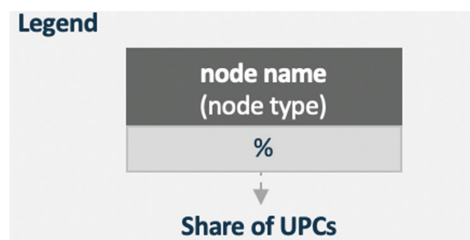


Real World Results

We have applied the top-down approach to data available on Walmart.com to create a sample CDT for the category of Beverages (non-alcoholic). The data points required were gathered from the website and engineered according to our proprietary methodology. The CDT as illustrated below is specific to the channel Walmart.com.

Exhibit 8.1 – CDT (3 Levels) I Beverages, Focus on Coffee & Tea I Walmart.com





Sample Insights

- The CDT segments the product category by the importance of product attributes in the shopper's purchasing decisions.
- The first split of the CDT groups product by Tea vs Coffee vs All Other; this means that customers shopping for Tea or Coffee are not likely to substitute outside these subsegments
- In the Tea subsegment, the caffeine level is the next decision criteria, indicating that online tea shoppers have a preference for caffeine levels that are either normal to high, or low to none

- The tea flavor “Sleepy Time” in the low-caffeine group is highly differentiated in the mind of the consumer, despite its relatively low share of product in the category
- In the Coffee segment, the brands Maxwell House and J.M. Smuckers drive the most loyalty
- For both of the loyalty coffee brands, size and weight, respectively, are the next differentiating factors for consumers

CONCLUSIONS

Accurate Customer Decision Trees (CDTs) are increasingly valuable for companies competing for well-informed customers purchasing highly differentiated products or services. Insights gleaned from CDTs enable optimal assortment, efficient investment in innovation products, better pricing, and more effective marketing tactics.

We have developed a standardized approach for creating CDTs for most product categories in relevant geographical markets, store formats and channels. We combine established data sources using a proprietary set of algorithms to rapidly create highly accurate CDTs that are ready for business use in an omnichannel context.

Our predictive methodology solves for the known issues in all stages of the CDT creation process that keep most consumer-facing companies from realizing value from the results. We aim to commoditize CDTs by making them available as a turnkey prediction product for manufacturers and retailers in both online channels and traditional brick and mortar ones.

REFERENCES

[1] Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. J Stat Mech P10008:

[2] Rosvall M, Bergstrom CT. Maps of random walks on complex networks reveal community structure. Proceedings of the National Academy of Sciences. 2008;105(4):1118–1123. 10.1073/pnas.0706851105

www.wqpredictive.com