Intel and Cloudera Improve Company's Targeted Marketing With Consumer Segmentation

Intel developed consumer segmentation for purchasers of cold cut products, increasing the effectiveness of marketing and sales.

Why Intel and Cloudera

Intel and Cloudera take the guesswork out of Hadoop. Using a unique collaborative approach, we delivery excellent performance, security, and quality distribution, built on open standards. Working with more vendors across the ecosystem, only a solution built on CDH can ensure freedom from lock-in, enabling you to build a robust big data solution to meet the needs of your business today and into the future.

- Uniquely aligned product roadmaps for software and hardware to drive innovation faster, providing more industry firsts than any other Hadoop alternative.
- Deep partnerships with virtually every provider in the data center, streamlining the process for building Big Data solutions.
- Proven track records of identifying the driving industry standards, so you don't run the risk of stranding yourself on an island.

An international food processing company greatly improves the accuracy and effectiveness of its targeted marketing campaigns by using Cloudera distribution of Apache Hadoop (CDH) to develop in-depth segmentation of consumers who purchase cold cut products.

The customer segmentation is built from various data sources: consumer demographics, responses to surveys, previous purchase history, and more. The company uses its increased marketability from the consumer segmentation to more accurately market cold cut products to targeted groups, predict customer purchasing trends, and increase profits.

Results

- Provides demographical, purchase, and preference behavior patterns of cold cut consumers.
- Uses generated patterns to provide four segments of consumers across various geographical locations for better targeted marketing for cold cut products.
- Improves the outcomes of consumer segmentation based on demographic information and survey responses only.

Business drivers

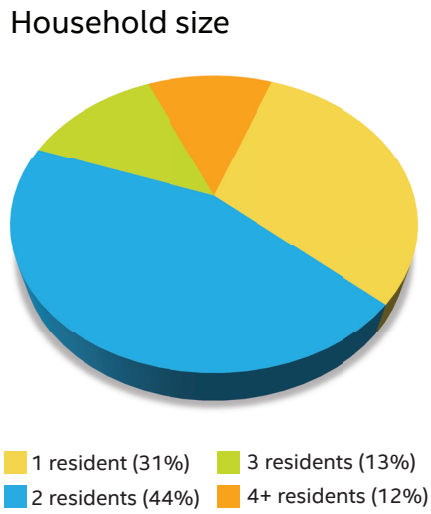
The Company was experiencing difficulty understanding consumer behavior in certain segments, in part due to their limited ability to merge various sources of consumer data—such as purchases, demographics, and survey responses—and use these in their predictive analytics. The ever-changing environment of consumer behavior was restricting the Company's targeted marketing efforts, especially with regard to customers who had purchased cold cuts in the past.

To better understand their consumers, the Company wanted to recognize purchase behavior patterns across various dimensions, segment them accordingly, and use the data for better and more personalized marketing.

Solution details

The objective of consumer segmentation was to group cold cut consumers into four segments—based on purchase, demographic, and survey response behaviors—with the purpose of recognizing behavior patterns, and using those insights to perform better and more profitable targeted marketing.

Using CDH running on Intel® Xeon® processor E5-based servers, Intel data scientists helped design and build the analytics engine to meet the Company's consumer segmentation requirements.



Data is obtained from four sources:

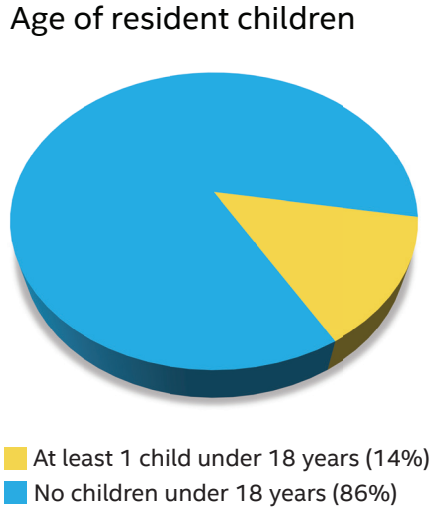
- Purchase history.
- Product information.
- Survey responses.
- Consumer demographics.

It was important to group the questions and responses in a meaningful and insightful way, so the survey responses needed to cover various aspects of consumers' preferences. Consumers were asked questions about the following:

- **Meal habits.** Whether they preferred to prepare meals at home, purchase prepared meals at a retail store, order takeout food online, stick to healthier food products, or compare prices and deals before making decisions.
- **Quick-service restaurants.** The consumers' preference of QSR establishments, such as burger joints, pizza joints, sub shops, coffee shops, or other fast food restaurants.
- **Cold cut preferences.** Any preferences regarding cold cuts. For example, attentiveness to cold cut details, cold cut brand loyalties, noticeable indifference to cold cut products, etc.
- **Food lifestyle habits.** Attitudes towards cooking meals at home and eating healthy foods versus junk foods.
- **Store brand preferences.** The consumers' name brand preferences of store brands.

Bringing spending and purchase behavior into consideration was insightful for segmentation purposes. For consumers who had purchased cold cut products, the Company discovered the total amount spent in other categories was highly indicative of spending intentions in cross-categories so they created a grouping of amounts spent for different categories.

The Company derived grouping variables from trips, channels, and total amounts spent to better understand a given consumer's affinity towards various products and brands. They generated a set of



derived variables from this information and implemented principal component analysis (PCA) to identify the following variables:

- **Active variables.** Amounts spent, store brands, food habits, etc.
- **Spend variables.** Amounts spent by consumers in various food categories, such as dairy, bakery, and others.
- **Trip variables.** Trips by channels, average spent per trip, attitudes about lifestyles, cooking, eating, and convenience.
- **Demographic variables.** Household size, income bracket, number and ages of children, education, occupation, and employment.

With the derived variables and insights, the Company executed clustering processes to identify four distinct segments of consumers (*Table 1*). The segments derived from clustering clearly showed distinction of consumer behavior in terms of purchase history, intention, habits, amount spent, and demographics.

Cloudera Enterprise

Given the need for flexibility, large dataset storage, and security, the Company saw a schema-on-read architecture for data ingestion as the only way to support the objectives, and thus eliminated traditional databases as a solution.

The Company's evaluation of open source options led to Hadoop, which addresses a number of business requirements. Beyond the flexibility of schema-on-read and the ability to handle a variety of data types, its standard set of capabilities resolve a number of data management needs. Additionally, due to the broad community and commercial support for Hadoop, it has matured to a readiness of mainstream adoption in the enterprise.

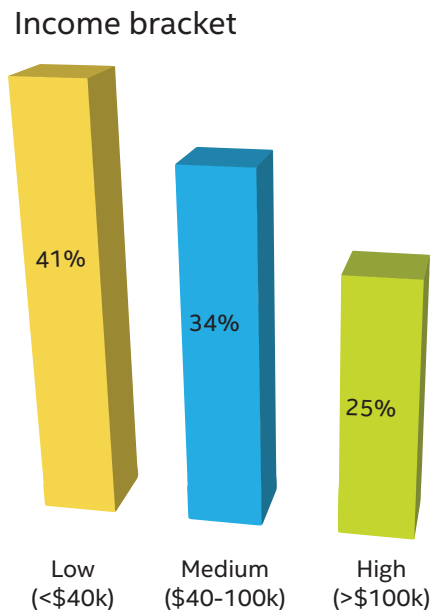


Table 1 **Customer segmentation.** The Intel/Cloudera solution allowed the Company to identify cold cut (CC) consumers and market to them appropriately based on one of four segments they fall into.

Segment 1	Higher attention to CC detail and CC brand loyal.
	Know to cook/connoisseurs.
	Higher spends in sandwich/coffee/chicken stores.
	Low % of CC spends on deals.
Segment 2	Skew to low income bracket.
	Primarily HH with a single member.
	High spends in CC.
	Low spends in *Entrées*, *Snacks/Meat*, *Brown Sugar*, *Sea Food*, and *Olive Oil*.
	Indifferent to CC brands and moderate attention to CC details.
	Infrequent visits to coffee shops.
	High % of CC spends on.
Segment 3	Skew to high income bracket and HH with 3 or more members.
	Lower propensity to consume store brands.
	High spends on most of the grocery F&B items.
	Low QSR trips and low on unhealthy/junk food.
Segment 4	Skew to HH with single member.
	Low spends on CC but moderate spend on organic CC.
	High spend on meal combos and precut salads.
	Low spends on most of the other grocery F&B items.
	Low/no attention to CC details.

From among the Hadoop distribution vendors, the Company selected Cloudera Enterprise because it offers a secure enterprise-ready Hadoop distribution. Of particular interest to the Company was the fact that Cloudera allows them to secure personally identifiable information (PII) by encrypting the data on the disk and restricting access to it within the Hadoop ecosystem. Furthermore, the Hadoop ecosystem offers tools to ingest data (using Sqoop and Flume), cleanse and prep data (with tools like Pig), and analyze data (with analytical libraries like Mahout and R).

Summary

With the new segmentation for cold cut consumers, the Company can now perform better personalized and targeted marketing. The customer segmentation they derive from the analysis factors in the unique behavior of cold cut consumers and their purchasing behaviors across various other categories and brands.

Intel was able to help the Company better understand consumer behavior in the cold cut segment, information they can use in their predictive analytics and apply to other product categories.

Let us help your business too.

Spotlight on Cloudera

Cloudera is revolutionizing enterprise data management by offering a unified platform for Big Data, an enterprise data hub built on Apache Hadoop*. Cloudera offers enterprises one place to store, access, process, secure, and analyze all their data, empowering them to extend the value of existing investments while enabling fundamental new ways to derive value from their data.

Cloudera's open source Big Data platform is the most widely adopted in the world, and Cloudera is the most prolific contributor to the open source Hadoop ecosystem. As the leading educator of Hadoop professionals, Cloudera has trained over 40,000 individuals worldwide. Over 1,800 partners and a seasoned professional services team help deliver greater time to value. Finally, only Cloudera provides proactive and predictive support to run an enterprise data hub with confidence. Leading organizations in every industry plus top public sector organizations globally run Cloudera in production.

For more information, visit **www.cloudera.com**.

cloudera®

Meeting your needs

We look forward to meeting with you to define your requirements and meet your objectives.

- **Accelerate time to value:** Achieve real-time cost savings, respond to market trends, and drive innovation.
- **Secure Big Data:** Deploy a sustainable Big Data program that doesn't put your organization, or you, at risk.
- **Maintain control:** Work with a partner who educates your team so you become self-sufficient.
- **Increase business potential**: Create and execute a plan that helps you adapt now, and in the future.

Table 2**Hadoop sizing guide.**

		Cluster size		
		Small	**Medium**	**Large**
CPU		**Intel® Xeon® processor E5**		
Storage (TB)		<72 TB	72 to 570 TB	>570 TB
Node count	**Master**	2 to 3	4 to 7	≥8
	Slaves	<12	12 to 95	≥ 96
Memory (GB)	**Master**	64 GB	128 GB	≥256 GB
	Slaves	48 GB	96 GB	≥128 GB
Network		1 Gbps	10 Gbps	10 Gbps

Hardware configuration is highly dependent on workload. A high storage density cluster may be configured with a 4 TB JBOD hard disk, while a compute intensive cluster may be configured with a higher memory configuration.

Contact us

Contact your sales rep or e-mail us.

Intel.com/bigdata/services

333169-001