Consumer Segmentation Analytics

# Abstract

AXANTEUS is a market research agency specializing in tracking consumer purchase behavior in both durable and nondurable consumer goods. In one major research project, AXANTEUS tracks numerous consumer product categories (e.g., detergents), and within each category, potentially dozens of brands.

To monitor purchase behavior, AXANTEUS has established consumer panels in over 100 cities and towns across Thailand, covering most of the Thailand urban market. The consumers were carefully selected using stratified sampling to ensure a representative sample. A subset of 600 records is analyzed here. The strata were defined based on socioeconomic status and market location (a collection of cities).

AXANTEUS maintains both transaction data (each row represents a transaction) and consumer data (each row represents a consumer). The consumer data includes the following information:

* **Demographics of the consumers** (updated annually)
* **Possession of durable goods** (e.g., car, washing machine; updated annually; an “affluence index” is computed from this information)
* **Purchase data of product categories and brands** (updated monthly)

AXANTEUS has two categories of clients:

1. Advertising agencies that subscribe to the database services, obtain updated data every month, and use the data to advise their clients on advertising and promotion strategies
2. Consumer goods manufacturers, which monitor their market share using the AXANTEUS database.

# Major Challenges

AXANTEUS has traditionally segmented markets on the basis of purchaser demographics. They would first like to segment the market based on two primary sets of variables more directly related to the purchase process and to brand loyalty:

1. Purchase behavior (volume, frequency, susceptibility to discounts, and brand loyalty)
2. Basis of purchase (price, selling proposition)

Doing so would allow AXANTEUS to gain information about what demographic attributes are associated with different purchase behaviors and degrees of brand loyalty, and thus deploy promotion budgets more effectively. More effective market segmentation would enable AXANTEUS’s clients (in this case, a firm called XRB) to design more cost-effective promotions targeted at appropriate segments. Thus, multiple promotions could be launched, each targeted at different market segments at different times of the year. This would result in a more cost-effective allocation of the promotion budget to different market segments. It would also enable XRB to design more effective consumer reward systems and thereby increase brand loyalty.

# Project Requirements

1. Start with required descriptive statistics then, data cleaning, transformation, and variable reduction.
2. Perform data segmentation by focusing on following areas:

* Data that describes the purchase behavior (including brand loyalty)
* Data that describes the basis for purchase
* Data that describes both the basis for purchase and the purchase behavior

1. Use segmentation results to develop a predictive model to classify clients as valued conscious or not.

* Value consciousness means that consumers pay more attention to deals and special offers and are prepared to buy in bulk to secure discounts.

1. Use segmentation results to develop a predictive model to predict Brand Runs in order to help potential loyal customers based on their demands.

# Dataset

The data in the following table profile each consumer, each row containing the data for one consumer

| Group Name | Attribute | Definition |
| --- | --- | --- |
| Demographic | Member id | Consumer id number |
|  | SEC | Socioeconomic Class (1 = High, 5 = Low) |
|  | FEH | Eating Habit (1 = Vegetarian, 2 = Vegetarian but eggs, 3 = Non vegetarian, 0 = Not Specified) |
|  | MT | Native Language (0~19 language code) |
|  | SEX | Gender of consumer (1 = Male, 2 = Female) |
|  | AGE | Age of Consumer |
|  | EDU | Consumer Education (1 = Minimum, 9 = Maximum) |
|  | HS | Number of member of Consumer household |
|  | CHILD | Number of consumer children |
|  | CS | Television Availability (1 = Available, 2 = Not Available |
|  | Affluence Index | Weighted value of durable possessed |
| Purchase Summary Over the Period | No. of Brands | Number of Brand Purchased |
|  | Brand Runs | Number of instances of consecutive purchase of brands |
|  | Total Volume | Sum of Volume |
|  | No. of Trans | Number of Purchase transactions (multiple brands purchased in a month are counted as separate transactions) |
|  | Value | Sum of values |
|  | Trans / Brand Runs | Average transaction per brand run |
|  | Vol/Tran | Average volume per transaction |
|  | Avg. Price | Average price of purchase |
| Purchase within Promotion | Pur Vol No Promo - % | Percent of value under no promotion |
|  | Pur Vol Promo 6 % | Percent of value under promotion code 6 |
|  | Pur Vol Other Promo % | Percent of value under order promotion |
| Brand wise Purchase | Br. Cd. 57, 144 | Percent of volume purchased under brand code |
|  | Br. Cd. 55 |
|  | Br. Cd. 272 |
|  | Br. Cd. 286 |
|  | Br. Cd. 24 |
|  | Br. Cd. 481 |
|  | Br. Cd. 352 |
|  | Br. Cd. 5 |
|  | 999 (Others) |
| Price Category wise Purchase | Pr Cat 1 | Percent of volume purchased under price category (1:4) |
|  | Pr Cat 2 |
|  | Pr Cat 3 |
|  | Pr Cat 4 |
| Selling Proportion wise Purchase | PropCat 5 | Percent of volume purchased under the product proposition category (5:15) |
|  | PropCat 6 |
|  | PropCat 7 |
|  | PropCat 8 |
|  | PropCat 9 |
|  | PropCat 10 |
|  | PropCat 11 |
|  | PropCat 12 |
|  | PropCat 13 |
|  | PropCat 14 |
|  | PropCat 15 |

Consumer Segmentation Analytics Requirements

# Requirements

Project Name: Consumer Segmentation Analytics

Dataset: Consumer.csv

Method: Data analytics modeling

Models: Classification, prediction, and clustering

Approach:

1. Project introduction
2. Business and analytics goals
3. Data preprocessing (such as attributes definition, data exploration, checking missing value, checking zero, and more)
4. Predictor analysis and relevancy
5. Dimension reduction (if needed)
6. Data transformation (if needed)
7. Data partitioning methods
8. Model selection
9. Model fitting, validation accuracy and test accuracy
10. Report models performance
11. Model evaluation (of the selected models)
12. Observation and conclusion

Deadline: Week 13

Delivery:

* Week 10: part 1, 2, and 3
* Week 11: part 4, 5, 6, 7, and 8
* Week 12: part 9, 10, and 11
* Week 13: complete work submission (documentation and presentation)

# Deliverable Policy

Demonstrate your analytics ability as an entry level data analytics professional. Your work should be delivered on time, late work will not be accepted. Do not share your work with other students. Avoid machine learning blunders and fundamental mistakes.