

Data Mining for Business Intelligence

Assignment

Year 2021

Notes and Instructions

- Use Bathsoap.xls data sheet
 - Use Weka software tool, if necessary
 - Submission deadline : January, 12 2021
 - Marks of this assignment will be counted for the final examination
-

IMRD: Segmenting Consumers of Bath Soap

Business Situation

The Indian Market Research Bureau (IMRB) is a leading market research agency that specializes in tracking consumer purchase behavior in consumer goods (both durable and non-durable).

IMRB tracks about 30 product categories (e.g. detergents, etc.) and within each category, about 60 – 70 brands. To track purchase behavior, IMRB has constituted about 50,000 household panels in 105 cities and towns in India, covering about 80% of the Indian urban market. (In addition to this, there are 25,000 sample households selected in rural areas; however, we are working with only urban market data). The households are carefully selected using stratified sampling. The strata are defined on the basis of socio-economic status, and the market (a collection of cities).

IMRB has both transaction data (each row is a transaction) and household data (each row is a household), and, for the household data, maintains the following information:

- Demographics of the households (updated annually)
- Possession of durable goods (car, washing machine, etc.; updated annually) and a computed "affluence index" on this basis
- Purchase data of product categories and brands (updated monthly).

IMRB has two categories of clients: (1) Advertising agencies who subscribe to the database services. They obtain updated data every month and use it to advise their clients on advertising and promotion strategies. (2) Consumer goods manufacturers who monitor their market share using the IMRB database.

Key Problems

IMRB has traditionally segmented markets on the basis of purchaser demographics. They would like now to segment the market based on two key 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), and
2. Basis of purchase (price, selling proposition)

Doing so would allow IMRB to gain information about what demographic attributes are associated with different purchase behaviors and degrees of brand loyalty, and more effectively deploy promotion budgets.

The better and more effective market segmentation would enable IMRB's clients 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 a year. This would result in a more cost-effective allocation of the promotion budget to different market-segments. It would also IMRB to design more effective customer reward systems and thereby increase brand loyalty.

Data

File: Bathsoap.xls

Sheet: DM_Data

The data in this sheet profile each household – each row contains the data for one household.

Member Identification	Member id		Unique identifier for each household
Demographics	SEC	1 – 5 categories	Socio Economic Class (1=high, 5=low)
	FEH	1 – 3 categories	Food eating habits (1=vegetarian, 2=veg. but eat eggs, 3=non veg., 0=not specified)
	MT		Native language (see table in worksheet)
	SEX	1: male 2: Female	Sex of homemaker
	AGE		Age of homemaker
	EDU	1 – 9 categories	Education of homemaker (1=minimum, 9 = maximum)

Demographics	HS	1 - 9	Number of members in the household
	CHILD	1 – 4 categories	Presence of children in the household
	CS	1 - 2	Television available. 1: Available 2: Not Available
	Affluence Index		Weighted value of durables possessed

Summarized Purchase Data

Purchase summary of the household over the period	No. of Brands	Number of brands 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 value
	Trans / Brand Runs	Avg. transactions per brand run
	Vol/Tran	Avg. volume per transaction
	Avg. Price	Avg. price of purchase

Purchase within Promotion	Pur Vol No Promo - %	Percent of volume purchased under no-promotion
	Pur Vol Promo 6 %	Percent of volume purchased under Promotion Code 6
	Pur Vol Other Promo %	Percent of volume purchased under other promotions

Brand wise purchase	Br. Cd. (57, 144), 55, 272, 286, 24, 481, 352, 5 and 999 (others)	Percent of volume purchased of the brand
Price category wise purchase	Price Cat 1 to 4	Per cent of volume purchased under the price category

Selling proposition wise purchase	Proposition Cat 5 to 15	Percent of volume purchased under the product proposition category
-----------------------------------	--------------------------------	--

Questions: Make a report to answer these questions

1. Use k-means clustering to identify clusters of households based on
 - a. The variables that describe demographic factors
 - b. The variables that describe purchase behavior
 - c. The variables that describe both demographic factors and purchase behavior.
 2. How should k be chosen?
 3. Select what you think is the best segmentation and comment on the characteristics of these clusters. (This information would be used to guide the development of advertising and promotional campaigns.)
-