

## **ИСПОЛЬЗОВАНИЕ МЕТОДОВ DATA MINING В ПРАКТИКЕ МАРКЕТИНГОВЫХ ИССЛЕДОВАНИЙ (НА ПРИМЕРЕ РЫНКА СОКОВ)**

**Ю. В. Сарайкин**, канд. экон. наук, доцент кафедры статистики ГОУ ВПО «Мордовский государственный университет имени Н.П. Огарева»

**Е. И. Кручинина, В. А. Курганова, О. А. Мельникова, М. М. Спирина**, студенты II курса специальности «Маркетинг» ГОУ ВПО «Мордовский государственный университет имени Н.П. Огарева»

*В статье рассмотрен пример использования подхода data mining для исследования рынка соков, обосновывается применение данного метода в практике маркетинговых исследований.*

**Ключевые слова:** DATA MINING, маркетинговые исследования, анализ данных, статистика, рынок соков, кластерный анализ, непараметрическая статистика

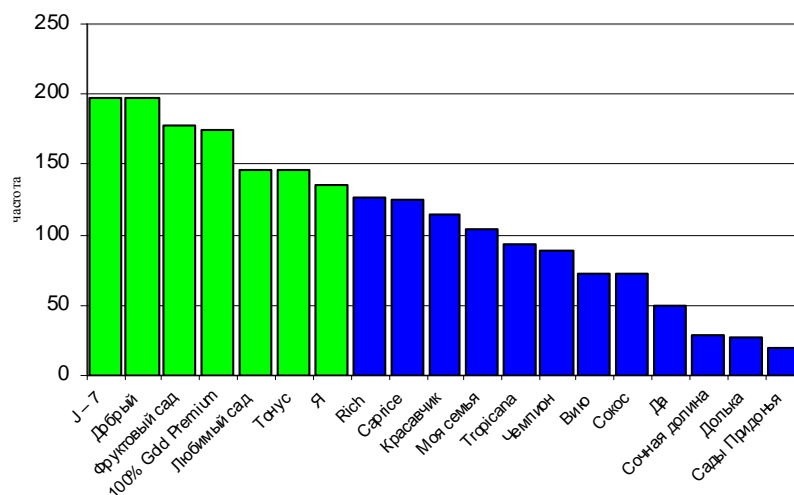
Маркетинговые исследования характеризуются большим количеством первичных необработанных данных. В связи с этим возникает необходимость использования специальных методов их обработки. Одним из подходов к анализу данных является data mining. Этот термин введен Григорием Пятецким-Шапиро в 1989 году и не имеет однозначного перевода на русский язык (добыча данных, вскрытие данных, информационная проходка, извлечение данных/информации) поэтому в большинстве случаев используется в оригинале. Наиболее удачным непрямым переводом считается термин «интеллектуальный анализ данных» (ИАД). Data Mining (Интеллектуальный анализ данных) – выявление скрытых закономерностей или взаимосвязей между переменными в больших массивах необработанных данных.

Задачи, решаемые data mining:

- классификация – отнесение входного вектора (объекта, события, наблюдения) к одному из заранее известных классов;
- кластеризация – разделение множества входных векторов на группы (кластеры) по степени «похожести» друг на друга;
- сокращение описания – для визуализации данных, лаконизма моделей, упрощения счета и интерпретации, сжатия объемов собираемой и хранимой информации;
- ассоциация – поиск повторяющихся образцов. Например, поиск «устойчивых связей в корзине покупателя» (англ. market basket analysis);
- прогнозирование;
- анализ отклонений – например, выявление нетипичной сетевой активности позволяет обнаружить вредоносные программы.

– визуализация.

Для исследования рынка соков г. Саранска в марте 2009г. было проведено обследование 30 магазинов разных типов полевым методом. Основными задачами данного мероприятия являлись изучение представленности брендов (марок) сока, их ценовых позиций и ценовой политики торговых точек разных типов в зависимости от размера торговой точки.



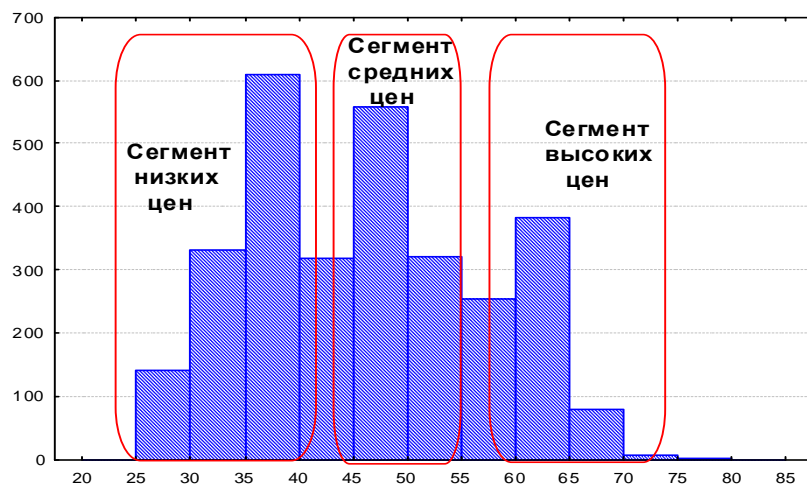
Р и с у н о к 1 Представленность брендов соков в магазинах г. Саранска

Анализ выявил, что наиболее представленными марками соков являются: J-7, 100% Gold Premium, Добрый, Фруктовый сад, Любимый сад, Тонус, Я.

Данные исследования магазинов позволили также выявить три ценовых сегмента:

- сегмент низких цен;
- сегмент средних цен, обладающий наименьшим «разбросом» цен;
- сегмент высоких цен.

Следует отметить, что сегменты низких и высоких цен имеют самый широкий «разброс» по ценам (рис. 2).



Р и с у н о к 2 Ценовые диапазоны соков магазинов г. Саранска (март 2009 г.).

Для выявления зависимости между ценой сока и такими факторами как «тип магазина» и «бренд» был использован дисперсионный анализ (табл.1).

Т а б л и ц а 1

**Дисперсионный анализ зависимости цены сока от типа магазина и бренда**

	Дисперсия	Степень свободы	Дисперсия в расчете на одну степень свободы	F-расчетный	p-уровень
«Бренд»	176092	13	13546	1278,0	0,000
«Тип магазина»	1036	2	518	48,9	0,000
«Бренд» & «тип магазина»	2552	26	98	9,3	0,000
Ошибка	26720	2521	11		
R	0,953	41		606,3	0,000
R <sup>2</sup>	0,908				

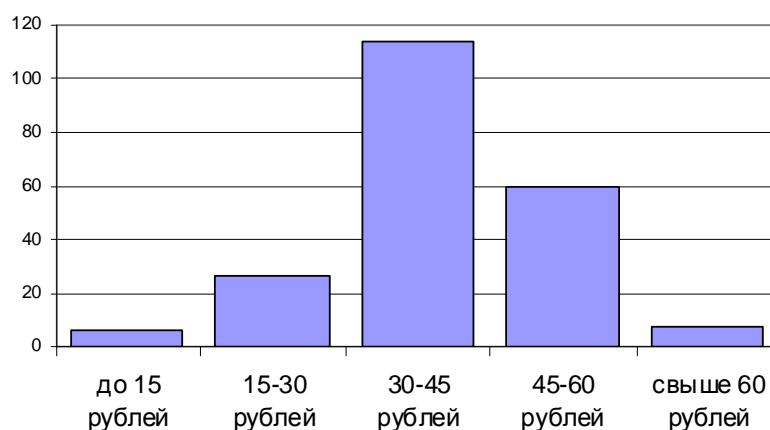
Расчеты подтвердили статистически значимую зависимость формирования ценовой политики торговой точки в отношении соков от факторов ее размера и бренда сока. Коэффициент детерминации равен  $R^2=0,908$ , т.е. 90,8% вариации цен зависит от выбранных факторов.

Также было проведено исследование потребительских предпочтений соков среди молодежи с использованием подхода data mining. Для сбора информации была разработана специальная анкета, вопросы которой касались предпочитаемых вкусов и брендов соков, частоты потребления сока, ценового диапазона покупки соков, а также мнения о степени разрекламированности того или иного бренда сока.

По результатам опроса были получены следующие результаты:

- от 84,3 до 98,7% студентов употребляют соки различных вкусов и марок регулярно;
- ежемесячные затраты на сок для молодежи составляют от 245 до 287 рублей на человека;
- оценка погрешности составляет 7,2%.

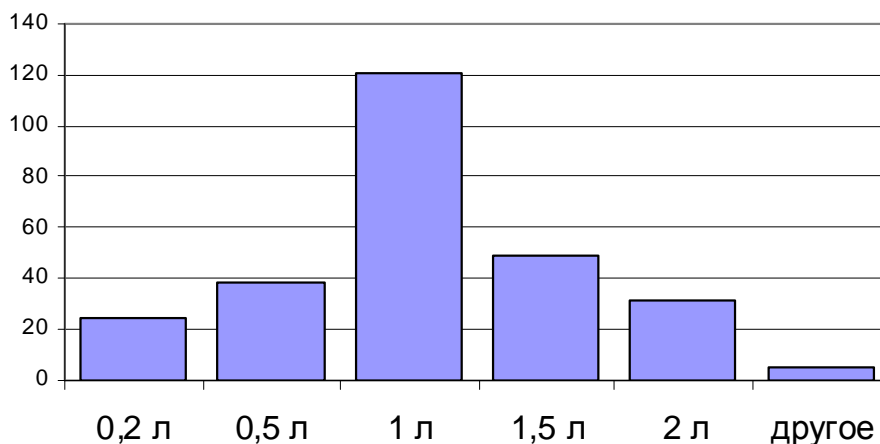
В ценовых предпочтениях студентов лидирует диапазон 30-45 рублей (рис.3).



**Р и с у н о к 3 Распределение ценовых диапазонов покупки сока среди студентов**

Можно сделать однозначный вывод, что студенты являются потребителями соков из сегмента низких цен, который является наиболее массовым.

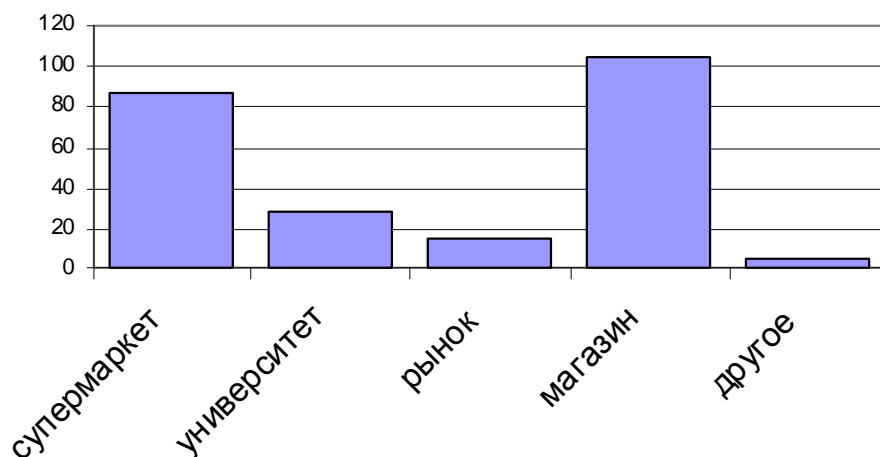
Наибольшей популярностью пользуется упаковка, объемом 1 литр, что обусловлено, прежде всего, удобством именно этого объема.



Р и с у н о к 4 Распределение объема упаковки соков при покупке среди студентов

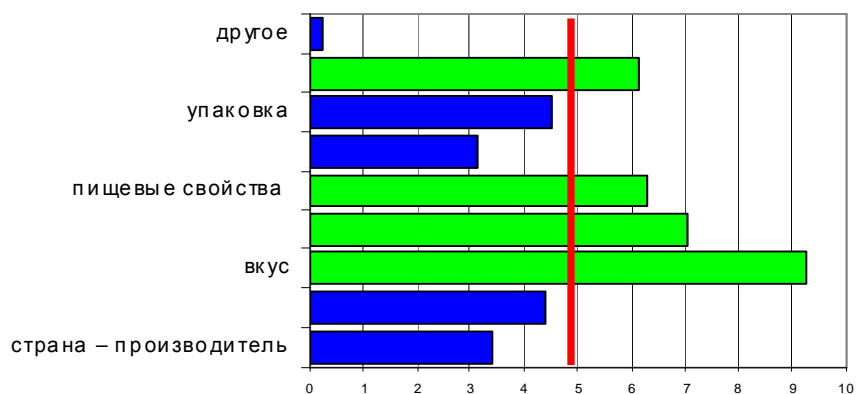
Характеристика потребительских предпочтений соков по объему представлена на рисунке 4.

Также было выявлено, что наиболее популярными местами покупки являются магазин и супермаркет (рис.5).



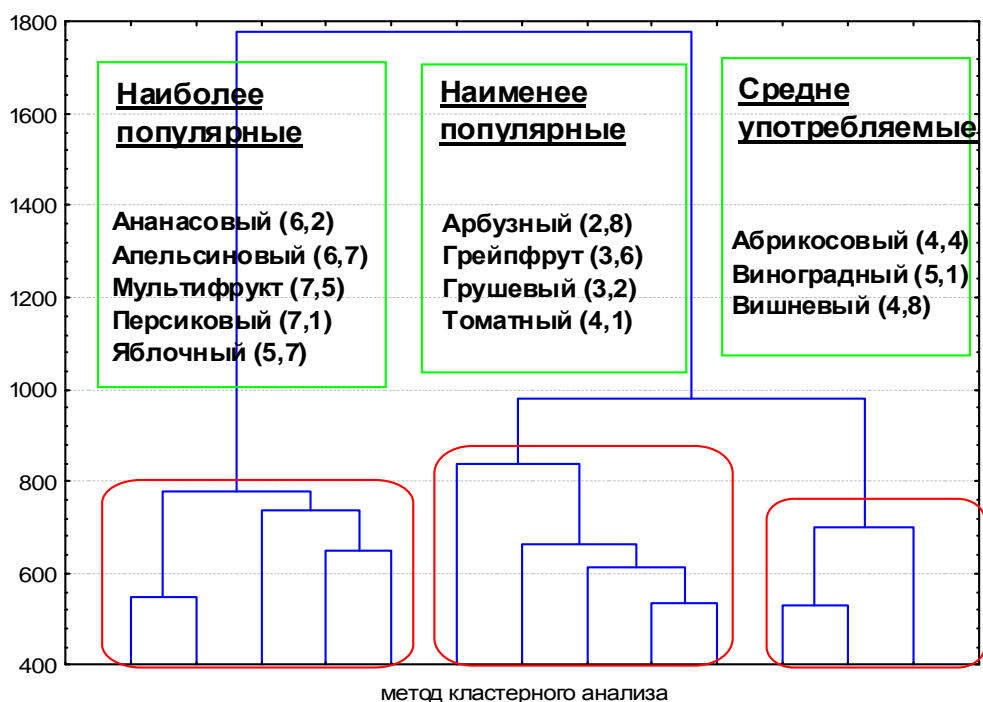
Р и с у н о к 5 Распределение мест покупки сока среди студентов

При проведении анализа было выявлено, что наиболее значимыми факторами при выборе сока являются вкус, цена, пищевые свойства и торговая марка (рис.6).



Р и с у н о к 6 Рейтинг факторов, влияющих на потребительский выбор

Для решения задачи сегментации рынка был использован иерархический агломеративный кластерный анализ. Он подразумевает группировку многомерных объектов, основан на представлении результатов точками подходящего геометрического пространства с последующим выделением групп как «сгустков» этих точек.



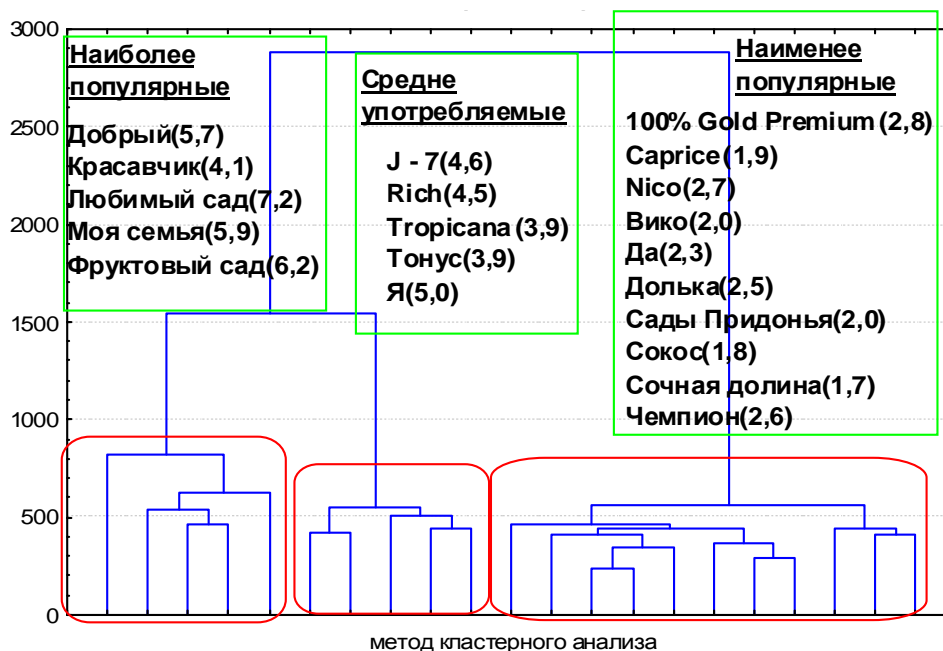
Р и с у н о к 7 Сегментация потребительских предпочтений вкусов сока

С помощью данного метода были выделены три сегмента вкусов соков (рис. 7):

- сегмент наиболее популярных вкусов: апельсиновый, мультифруктовый, персиковый;
- сегмент менее употребляемых вкусов: абрикосовый, виноградный, вишневый;
- сегмент наименее популярных вкусов: грушевый, томатный и др.

Следует отметить, что данная сегментация подтверждается рейтингом предпочтений по вкусам (его значения указаны в скобках).

При кластерном анализе потребительских предпочтений по брендам были выявлены следующие сегменты (рис.8).



Р и с у н о к 8 Сегментация потребительских предпочтений по брендам (маркам) соков

– сегмент самых популярных брендов: Добрый, Любимый сад, Моя семья, Фруктовый сад, т.е. недорогие разрекламированные соки;

– сегмент соков средней популярности: J – 7, Rich, Tropicana, Тонус, Я, т.е. разрекламированные соки премиум сегмента;

– «аутсайдеры»: такие бренды как Сады Придонья, Сокок, Сочная долина, Чемпион не являются часто выбираемыми при покупке.

Данные рейтинга также указаны в скобках.

При анализе потребительских предпочтений брендов соков нельзя не учесть влияние на выбор сока степени его разрекламированности, в таблице 3 представлены данные корреляционного анализа Кендалла. С его помощью было установлено, что степень разрекламированности бренда сока влияет на выбор сока.

Т а б л и ц а 2

**Непараметрический корреляционный анализ зависимости частоты покупки сока и его разрекламированности**

Наименование бренда сока	$\tau$ -Kendall	p-уровень
1	2	3
1. 100% Gold Premium	0,391	0,000
2. Caprice	0,346	0,000
3. Сочная долина	0,341	0,000
4. Да	0,314	0,000
5. Сокок	0,305	0,000

Наименование бренда сока	$\tau$ -Kendall	p-уровень
1	2	3
6. Любимый сад	0,304	0,000
7. Долька	0,302	0,000
8. Сады Придонья	0,300	0,000
9. J – 7	0,289	0,000
10. Чемпион	0,288	0,000
11. Вико	0,279	0,000
12. Тropicana	0,267	0,000
13. Фруктовый сад	0,263	0,000
14. Я	0,208	0,000
15. Niso	0,205	0,000
16. Моя семья	0,201	0,000
17. Rich	0,200	0,000
18. Тонус	0,199	0,000
19. Красавчик	0,168	0,000
20. Добрый	0,137	0,005

Таким образом, необходимо выделить следующие моменты:

– безусловными лидерами рынка соков для студентов являются Добрый, Любимый сад, Фруктовый сад, Моя семья, J – 7, Rich;

– представленность соков в магазинах достаточно хорошо отражает потребительские предпочтения;

– наиболее значимыми факторами при выборе сока являются вкус, цена, торговая марка.

Подводя итоги исследованию, можно констатировать, что использование data mining способствует статистической поддержке управленческих решений в сфере маркетинга.