**АНАЛИЗ ПОСЛЕДОВАТЕЛЬНОСТЕЙ**

1. **Описание задачи**

Во многих магазинах накапливается информация о продажах товаров с течением времени. При этом в первую очередь регистрируются такие данные как дата продажи, наименование товара и идентификатор клиента (если он постоянный).

Эти данные можно использовать для выявления наиболее часто встречающихся последовательностей покупок.

Например, если человек покупает стационарный компьютер, то затем с высокой вероятностью он купит принтер, сканер и постоянно начнет покупать расходные материалы для принтера.

Необходимо оценить какой товар с максимальной вероятностью купит каждый постоянный клиент при очередном заходе в магазин.

Кроме получения указанной оценки можно попытаться также оценить среднюю периодичность покупок и их усредненную стоимость.

В общем случае задача очень сложна, поскольку необходимо определить вероятность появления всех возможных перестановок номеров товаров, встретившихся в данных о продажах.

При этом, при большом ассортименте товаров количество перестановок становится настолько велико, что превзойдет количество продаж, совершенных даже в самых крупных магазинах за несколько лет.

Поэтому подсчитывают частоту только тех перестановок, которые встречаются в имеющейся базе данных.

Возможен упрощенный вариант поиска – определяется частота только двухэлементных перестановок

Полученные таким образом данные используются затем для планирования объемов продаж.

1. **Пример**

На сетевом диске имеется файл "Последовательности.xls". В нем на листе “Продажи” приводятся следующие сведения о продажах:

* День продажи (общее количество дней 500);
* № продажи (на один день работы магазина приходится по 20 продаж);
* Идентификатор покупателя (общее количество постоянных клиентов равно 300);
* Идентификатор товара (общее количество товаров равно 14). Наименования товаров и их цены приведены на втором листе.

По этим данным можно определить множество параметров влиящих на успешность работы магазина.

Основным таким параметром является оценка вероятного объема продаж в очередной день работы магазина. Для его вычисления необходимо выполнить следующую цепочку вычислений

1. ***Определяются вероятности появления каждой последовательности продаж.***

Для определения этого показателя используется макрос MakeSequence. Результат его работы формируется на листе “Последовательности” и представлен в табл. 1.

Таблица 1

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
| 1 | 0 | 29 | 28 | 3 | 57 | 0 | 44 | 13 | 11 | 24 | 15 | 13 | 17 | 6 |
| 2 | 0 | 0 | 85 | 5 | 1 | 326 | 2 | 5 | 7 | 7 | 4 | 6 | 4 | 51 |
| 3 | 15 | 225 | 12 | 14 | 13 | 13 | 8 | 13 | 12 | 13 | 10 | 12 | 13 | 11 |
| 4 | 10 | 2 | 2 | 1 | 0 | 1 | 118 | 1 | 2 | 0 | 34 | 15 | 74 | 0 |
| 5 | 24 | 21 | 22 | 29 | 0 | 20 | 22 | 26 | 18 | 22 | 21 | 20 | 16 | 21 |
| 6 | 33 | 40 | 48 | 36 | 32 | 370 | 45 | 45 | 42 | 41 | 40 | 35 | 46 | 44 |
| 7 | 6 | 10 | 6 | 9 | 3 | 8 | 274 | 9 | 4 | 11 | 3 | 10 | 257 | 96 |
| 8 | 7 | 6 | 4 | 9 | 6 | 6 | 8 | 153 | 6 | 7 | 8 | 280 | 7 | 9 |
| 9 | 7 | 7 | 9 | 5 | 11 | 5 | 6 | 71 | 55 | 6 | 4 | 121 | 13 | 3 |
| 10 | 8 | 8 | 4 | 7 | 14 | 9 | 7 | 8 | 13 | 342 | 76 | 214 | 7 | 11 |
| 11 | 2 | 6 | 6 | 9 | 6 | 10 | 6 | 11 | 6 | 92 | 335 | 9 | 11 | 198 |
| 12 | 62 | 53 | 50 | 50 | 62 | 57 | 69 | 66 | 69 | 71 | 62 | 634 | 63 | 57 |
| 13 | 21 | 23 | 29 | 26 | 28 | 22 | 28 | 24 | 30 | 33 | 28 | 30 | 269 | 259 |
| 14 | 53 | 62 | 54 | 56 | 42 | 50 | 69 | 59 | 42 | 58 | 66 | 55 | 57 | 1136 |

Данные таблицы интерпретируются следующим образом:

* каждая строка соответствует первому элементу последовательности;
* столбцы соответствуют второму элементу последовательности;
* в ячейках таблицы находятся число, которое указывает сколько раз встречается данная последовательность в таблице продаж.

Например, ни разу не встречаются последовательности продаж 1 – 1, 1 – 6, 2 – 1, 2 – 2, 4 – 5, 4 – 14, 5 - 5. Наиболее часто встречающейся последовательностью является 14 – 14 (встречается 1136 раз).

1. ***Определяется последний товар, который купил каждый клиент***

Для этого используется макрос LastPurchase. Результат его работы выводится на лист “Последняя покупка”. Часть полученной таблицы показана в табл. 2.

Таблица 2

|  |  |
| --- | --- |
| клиент | товар |
| 1 | 7 |
| 2 | 12 |
| 3 | 13 |
| 4 | 11 |
| 5 | 12 |
| 6 | 5 |
| 7 | 5 |
| 8 | 7 |
| 9 | 6 |
| 10 | 12 |
| 11 | 2 |
| 12 | 6 |
| 13 | 2 |
| 14 | 6 |
| 15 | 13 |
| 16 | 6 |
| 17 | 10 |
| 18 | 7 |
| 19 | 11 |
| 20 | 14 |

1. ***Определяется вероятность появления в магазине каждого покупателя***

Расчет основан на количестве покупок каждого покупателя незавимо от того, какой товар он купил. Эта величина рассчитывается с помощью макроса CountPurchase. Результат его работы выводится на лист “Вероятность появления”. Часть полученной таблицы показана в табл. 3.

Таблица 3

|  |  |
| --- | --- |
| клиент | количество покупок |
| 1 | 37 |
| 2 | 36 |
| 3 | 23 |
| 4 | 25 |
| 5 | 34 |
| 6 | 35 |
| 7 | 30 |
| 8 | 32 |
| 9 | 28 |
| 10 | 37 |
| 11 | 33 |
| 12 | 36 |
| 13 | 30 |
| 14 | 28 |
| 15 | 23 |
| 16 | 41 |
| 17 | 31 |
| 18 | 31 |
| 19 | 33 |
| 20 | 33 |
| 21 | 27 |
| 22 | 39 |
| 23 | 33 |

Поскольку в день ожидается по 20 покупок (приходов покупателей), полученный выше список сортируется по убыванию по колонке «Количество покупок»

Результат сортировки показан в табл. 4.

Таблица 4

|  |  |
| --- | --- |
| клиент | количество покупок |
| 124 | 51 |
| 164 | 48 |
| 282 | 48 |
| 84 | 47 |
| 213 | 47 |
| 216 | 47 |
| 91 | 45 |
| 105 | 45 |
| 196 | 45 |
| 39 | 44 |
| 56 | 44 |
| 262 | 44 |
| 271 | 44 |
| 185 | 43 |
| 95 | 42 |
| 139 | 42 |
| 156 | 42 |
| 182 | 42 |
| 202 | 42 |
| 237 | 42 |
| 278 | 42 |
| 280 | 42 |
| 286 | 42 |
| 290 | 42 |

Здесь выделены первые 20 покупателей.

1. ***Определение общей суммы выручки***

Для определения этого показателя необходимо определить какие товары будут куплены и подсчитать их суммы. Для этого используются данные из табл. 1, 2, 4.

Например, согласно табл. 4 в магазине появится клиент под номером 124. Согласно данным листа “Последняя покупка” этот клиент посдний раз купил товар № 14. А согласно табл. 1 наиболее вероятной покупкой вслед за товаром 14 является также товар 14. Поэтому сумма первой покупки будет равна стоимости товара 14 (берется с листа “Товары”).

Подсчет производится с помощью макроса SumPurchase.

Результаты расчетов выводятся на лист «Вероятная стоимость» и показаны в табл. 5.

Таблица 5

|  |  |
| --- | --- |
| Товар | Цена |
| 14 | 450 |
| 12 | 950 |
| 10 | 2400 |
| 6 | 300 |
| 14 | 450 |
| 2 | 12000 |
| 5 | 1200 |
| 7 | 550 |
| 14 | 450 |
| 10 | 2400 |
| 6 | 300 |
| 10 | 2400 |
| 11 | 6700 |
| 13 | 250 |
| 6 | 300 |
| 6 | 300 |
| 5 | 1200 |
| 10 | 2400 |
| 14 | 450 |
| 14 | 450 |
| Всего | 35900 |

1. **Задания**

Оценить наиболее вероятную сумму продаж на 501 день работы магазина, удалив при этом из БД продаж часть данных согласно таблице:

|  |  |
| --- | --- |
| № варианта | Удаляемые данные |
| 1 | 1..50 день |
| 2 | 26..75 день |
| 3 | 51..100 день |
| 4 | 76..125 день |
| 5 | 101..150 день |
| 6 | 126..175 день |
| 7 | 151..200 день |
| 8 | 176..225 день |
| 9 | 201..250 день |
| 10 | 226..275 день |
| 11 | 251..300 день |
| 12 | 276..325 день |
| 13 | 301..350 день |
| 14 | 326..375 день |
| 15 | 351..400 день |
| 16 | 376..425 день |
| 17 | 401..450 день |
| 18 | 426..475 день |

**Примечание**

Приведенные в примере расчеты опираются на результаты работы соответствующих макросов. В свою очередь макросы используют именованный диапазон “Продажи”. Убедитесь в том, что в результате удаления части данных адресация указанного диапазона соответствущим образом откорректировалась. В противном случае макросы будут работать не корректно.