

## **СБОР И АНАЛИЗ ДАННЫХ ДЛЯ МОНИТОРИНГА АУДИТОРИИ САЙТОВ**

*Г.В. Сорокин*

Для обеспечения эффективности интернет-рекламы современные компании используют механизмы, которые позволяют выделить из всей имеющейся аудитории только ту часть, которая удовлетворяет заданным критериям, и показать рекламу именно ей. В статье приведены методы, которые могут использоваться при сборе и анализе данных для мониторинга аудиторий, а также реализация этих методов в Microsoft Excel с помощью VBA.

Ключевые слова: куки, эффективность рекламы, интернет-ресурсы, VBA.

HTTP Cookie или куки представляют собой небольшой набор данных, который сервер посещенного сайта записывает в браузер клиента. Они применяются для аутентификации пользователя, сохранения пользовательских настроек для сайта и для ведения статистики рекламными компаниями [1]. В последнем случае с их помощью отслеживаются действия пользователей на сайте. Далее из этих данных рекламные компании формируют

анонимные профили пользователей для более точного нацеливания рекламы [2]. Например, когда пользователь посещает сайт учебного заведения, рекламная компания в профиле отмечает для себя, что пользователю интересна информация, связанная с образованием, и в дальнейшем предоставляет ему подобные ресурсы в виде рекламы в интернете.

Пользователи (куки) формируются в группы по общим признакам: возраст, пол, интересы, доходы и т.д. Пользователю с предположительным возрастом до 18 лет будет предлагаться реклама, отличная от тех, кому за 40. В более узком смысле можно формировать группу кук, которые оформляли покупки в конкретном интернет-магазине или просто посещали его. В случае падения аудитории сайта – привлекать новых пользователей, близких по параметрам к тем, кого этот сайт уже заинтересовал.

Также стоит отметить, что кука не долговечная и в рекламных целях не имеет смысл хранить ее более 30 дней.

Рассмотрим следующую задачу: рекламная компания отслеживает изменение в аудитории пользователей определенных магазинов. В нашем доступе имеются следующие данные (рис. 1):

- 1) количество пользователей;
- 2) количество новых пользователей за расчетный день;
- 3) некоторый показатель ведения рекламной кампании.

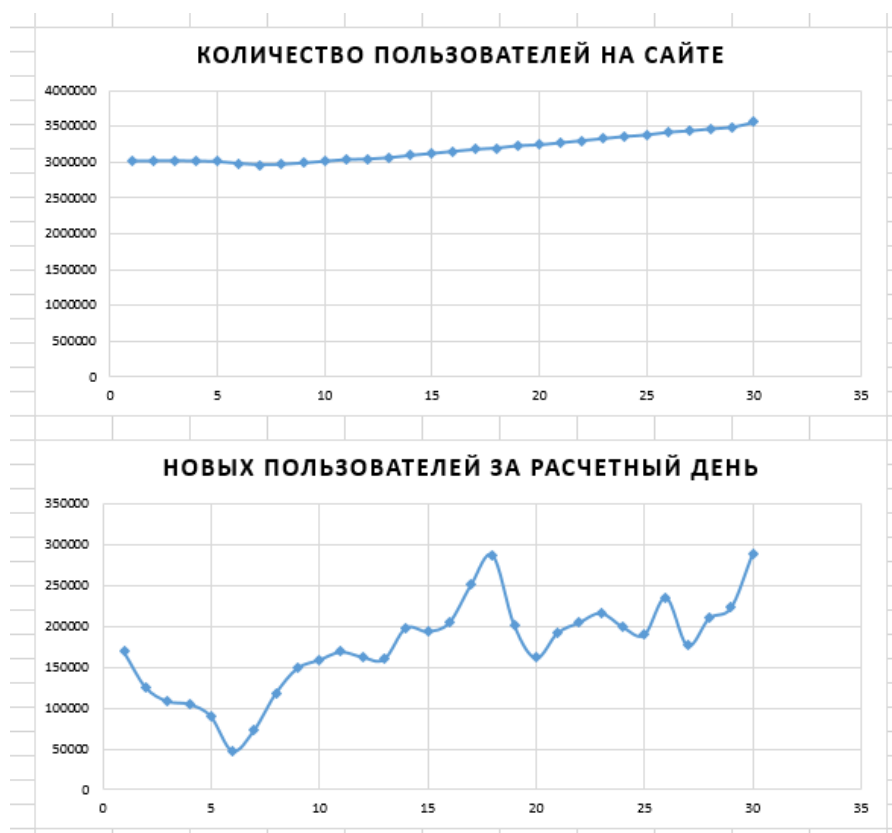


Рис. 1. Начальные данные, представленные в виде графиков

Необходимо отслеживать, чтобы аудитория в группе не сокращалась в объеме и имела стабильный ежесуточный прирост. Для отслеживания используется alert-система, которая оповещает ответственных сотрудников, если количество пользователей будет меньше минимального допустимого количества – показателя  $\text{minAuditory}$ , а минимальное суточное обновление аудитории меньше показателя  $\text{minNewAuditory}$ .

Данные показатели должны быть подобраны таким образом, чтобы небольшие колебания не провоцировали alert-события, а динамика понижения показателей день ото дня наоборот была бы донесена до сведения специалистов.

Возьмем за правило, что для определения нормативного показателя в графиках с устойчивым ростом, устойчивым падением или волнообразностью нужно высчитать среднесуточное значение метрики, а для определения нормативного показателя в равномерных графиках нужно высчитать среднесуточное значение метрики и уменьшить его на 15 %. Также если при обновлении контрольных цифр было выявлено значительное плановое (30 % и более) либо резкое сокращение значения одного из показателей, необходимо уведомить специалиста независимо от alert-системы.

Если в показателе ведения рекламной кампании за последние 7 дней нет данных – уведомление о падении не требуется (скорее всего кампания отключена).

Таким образом, для определения нормативного показателя необходимы параметры роста (устойчивое падение, рост) и отклонения от среднего значения (волнообразность).

Проверяем показатели за последний месяц. В качестве роста мы используем величину Рост. Она показывает, на сколько процентов увеличивается показатель ежедневно.

$$\text{Рост} = \frac{y_n - y_{n-1}}{y_{\text{ср}}} 100\%$$

Поскольку показания  $Y$  имеют дискретные значения, мы будем использовать линейную интерполяцию. Коэффициенты находим по формулам:

$$\text{Коэф. } A = \frac{N \sum xy - \sum x \sum y}{N \sum x^2 - (\sum x)^2}$$

$$\text{Коэф. } B = \frac{\sum y - a \sum x}{N}$$

Полученные данные можно классифицировать семантически, воспользовавшись нечеткими множествами [3] (рис. 2).

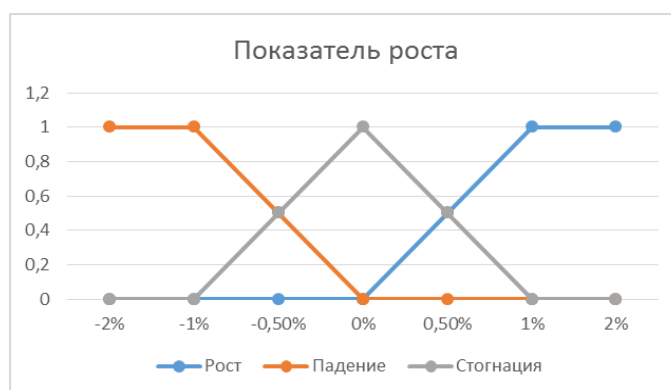


Рис. 2. Значения параметра роста

Так при значениях показателя роста меньше -1 % мы можем говорить о явном падении аудитории, тогда как в промежутке от -0,5 % до 0,5 % показатель будет больше стабильным, чем изменяющимся.

На данном этапе мы можем задать контрольную цифру, равную среднесуточному значению динамики с явным ростом или явным падением. В качестве отклонения мы используем формулу стандартного отклонения:

$$\text{Отклонение} = \sqrt{\frac{\sum(y - y_{\text{ср}})^2}{N}}$$

Случай «скорее стабильный, чем изменяющийся» будет проверяться на «волнообразность» – отклонение больше или меньше 10 %. Неявный рост и падение мы также оценим на «волнообразность» увеличив нормированное значение до 12 %.

Если график не имеет большого отклонения от среднего значения, делается контрольная проверка по Росту. На вход подаются показатели уже за три последних месяца, показатели роста сравниваются со значениями – 0,75 % и +0,75 %.

Данный алгоритм проверки был реализован на MS Excel средствами VBA. Написанный сценарий импортирует таблицы с данными аудитории из закрытых web-ресурсов рекламной компании, используя коллекцию QueryTables. Полученные из алгоритма контрольные значения округляются с необходимой точностью и собираются в отдельный CSV-файл, который передается в alert-систему (табл.).

Таблица

Правила округления контрольных значений

10	с точностью до 5
100	две значащих цифры
1 000	две значащих цифры
10 000	с точностью до 500
100 000	с точностью до 5000
1 000 000	три значащих цифры
10 000 000	три значащих цифры

Также в сценарии отслеживаются аудитории с устойчивым падением. Под критерий попадают случаи, когда:

- линия тренда нисходящая и конечное ее значение на 30 и более процентов меньше начального;
- линия тренда возрастает/не изменяется в течение периода и текущее значение метрики на 30 и более процентов меньше конечного значения линии тренда.

#### Библиографический список

1. Бабаев, А. Контекстная реклама. Учебник / А. Бабаев, Н. Евдокимов, А. Иванов. – СПб.: Питер, 2011. – 304 с.
2. Геддс, Б. Google AdWords. Исчерпывающее руководство / Б. Геддс. – М.: «Манн, Иванов и Фербер», 2014. – 624 с.
3. Кофман, А. Введение в теорию нечетких множеств / А. Кофман. – М.: Радио и связь, 1982. – 432 с.

[К содержанию](#)