

Урок 8. Типовые структуры БД. Изменение групп пользователей при RFM анализе.

ДЗ делаем по бд orders В качестве ДЗ сделаем карту поведения пользователей. Мы обсуждали, что всех пользователей можно разделить, к примеру, на New (совершили только 1 покупку), Regular (совершили 2 или более на сумму не более столько-то), Vip (совершили дорогие покупки и достаточно часто), Lost (раньше покупали хотя бы раз и с даты последней покупки прошло больше 3 месяцев). Вся база должна войти в эти группы (т.е. каждый пользователь должен попадать только в одну из этих групп).

Задача:

1. Уточнить критерии групп New,Regular,Vip,Lost

Lost - с момента последней покупки прошло более 60 дней; New - с момента первой покупки прошло менее 60 дней (включительно) и была только одна покупка; Regular - сумма покупок менее 5000, было более 1 покупки и клиент не попал в группу Lost; Vip - сумма покупок более 5000 (включительно), было более 1 покупки и клиент не попал в группу Lost.

2. По состоянию на 1.01.2017 понимаем, кто попадает в какую группу, подсчитываем кол-во пользователей в каждой.

Используем переменную для задания даты анализа

```
set @analys_date = date('2017-01-01');
```

В основном запросе воспользуемся тем что конструкция case when then возвращает первое успешное совпадение. Сперва отсеим потерянных клиентов, из оставшихся выберем новых (с одной покупкой), из оставшихся выберем регулярных (с суммарной покупкой менее 5000p), остальные попадут в Vip.

```
select
  category,
  count(user_id) as users_count,
  sum(total_revenue) as total_revenue,
  sum(orders_count) as orders_count
from (
  select
    user_id,
```

```

count(id_o) as orders_count,
sum(price) as total_revenue,
case
  when TIMESTAMPDIFF(DAY, max(o_date), @analys_date) > 60 then "Lost"
  when count(id_o) = 1 then "New"
  when sum(price) < 5000 then "Regular"
  else "Vip"
end as category
from orders
where o_date < @analys_date
group by user_id
) group_by_user_id
group by category;

```

category	users_count	total_revenue	orders_count
Lost	291973	902733049.400	414086
New	98243	205359002.800	98243
Regular	22151	62450202.500	62752
Vip	32725	634480397.950	286265

Для проверки, посчитаем число пользователей в выборке

```
select count(distinct user_id) from orders where o_date < @analys_date;
```

count(distinct user_id)
445092

Число пользователей сходится, значит никого не потеряли.

3. По состоянию на 1.02.2017 понимаем, кто вышел из каждой из групп, а кто вошел.

```
set @analys_date = date('2017-02-01');
```

category	users_count	total_revenue	orders_count
Lost	347612	1112856133.200	515021

category	users_count	total_revenue	orders_count
New	91484	200968670.700	91484
Regular	19807	56078969.100	54900
Vip	30708	631172272.850	283628

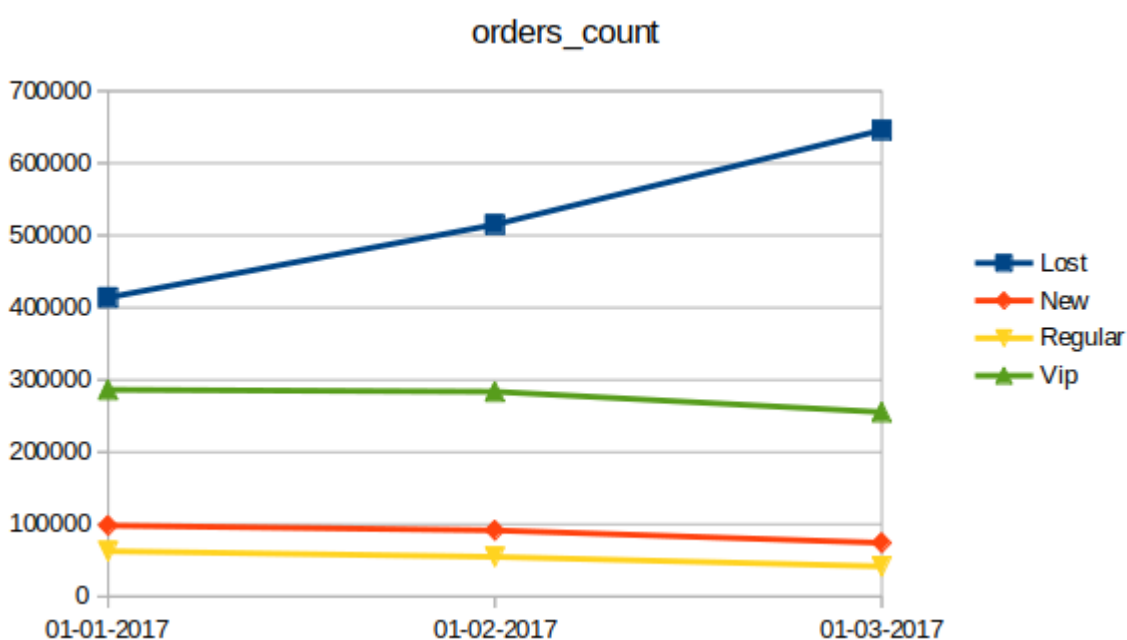
4. Аналогично смотрим состояние на 1.03.2017, понимаем кто вышел из каждой из групп, а кто вошел.

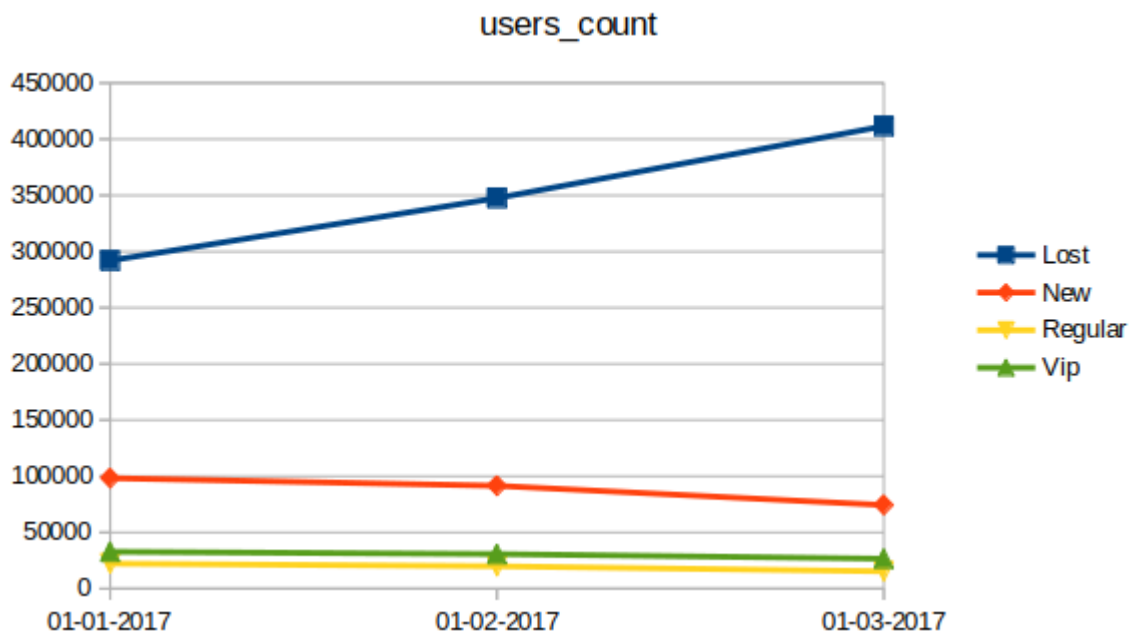
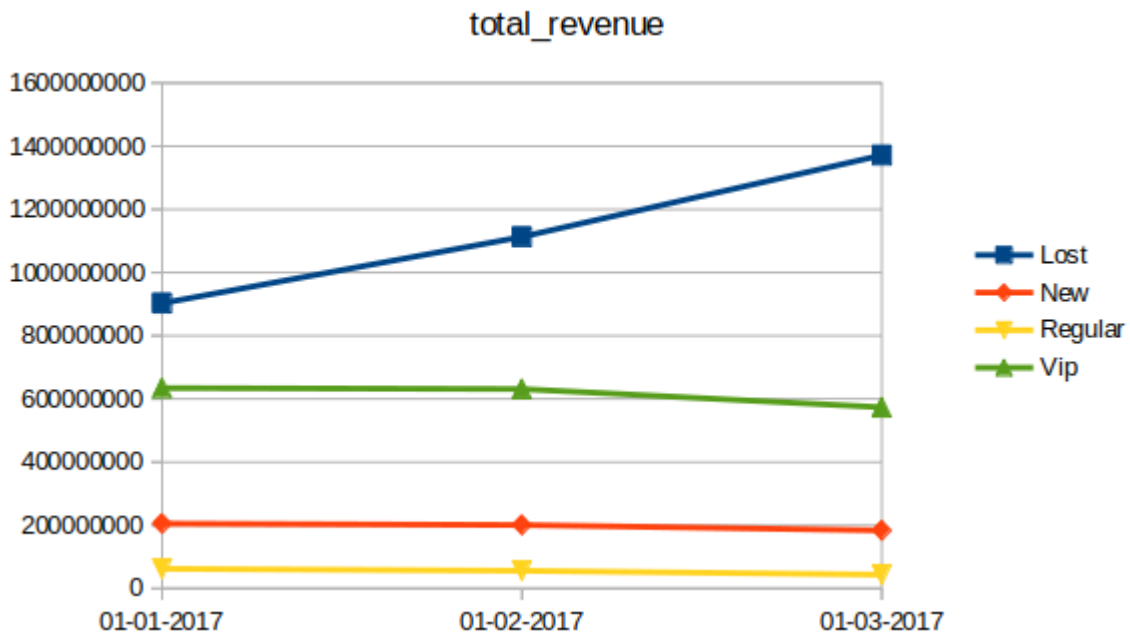
```
set @analys_date = date('2017-03-01');
```

category	users_count	total_revenue	orders_count
Lost	411553	1371155888.900	645737
New	74336	183864223.900	74336
Regular	15316	43815328.200	41719
Vip	26571	573547544.150	255316

5. В итоге делаем вывод, какая группа уменьшается, какая увеличивается и продумываем, в чем может быть причина.

Отобразим динамику на графиках:





Видно что число потерянных пользователей растёт. Это результат естественного вымирания пользовательской аудитории.

Так же виден сильный спад числа новых, регулярных и вип пользователей на дату анализа 1 марта 2017. Это может быть связано с тем что для этих групп временной диапазон даты последней покупки ограничен двумя месяцами до даты анализа. То есть многие пользователи сделали покупки в декабре и перестали их делать в январе. Если бы мы считали потерянными клиентов, сделавших последнюю покупку более месяца назад, то этот спад пришелся бы на 1 февраля 2017.