

### Задание для Урока 3.

Главная задача: сделать RFM-анализ на основе данных по продажам за 2 года (из предыдущего дз).

Что делаем:

1. Определяем критерии для каждой буквы R, F, M (т.е. к примеру, R – 3 для клиентов, которые покупали  $\leq 30$  дней от последней даты в базе, R – 2 для клиентов, которые покупали  $> 30$  и менее 60 дней от последней даты в базе и т.д.)
2. Для каждого пользователя получаем набор из 3 цифр (от 111 до 333, где 333 – самые классные пользователи)
3. Вводим группировку, к примеру, 333 и 233 – это Vip, 1XX – это Lost, остальные Regular ( можете ввести более глубокую сегментацию)
4. Для каждой группы из п. 3 находим кол-во пользователей, кот. попали в них и % товарооборота, которое они сделали на эти 2 года.
5. Проверяем, что общее кол-во пользователей бьется с суммой кол-во пользователей по группам из п. 3 (если у вас есть логические ошибки в создании групп, у вас не сойдутся цифры). То же самое делаем и по деньгам.

!!! Результаты присылаем скриншотом или в pdf (где можно увидеть название группы, правила попадания в группу, кол-во пользователей, их товарооборот, а так же процент по кол-ву пользователей и по деньгам)

### Решение

1) – 3) пункты:

```
-- все пользователи Vip
SELECT COUNT(user_id), SUM(s)
FROM
(
SELECT user_id, r, f, m, s,
CASE WHEN (r = 3 AND f = 3 AND m = 3) OR (r = 2 AND f = 3 AND m = 3) THEN "Vip"
      WHEN r = 1 THEN "Lost"
      ELSE "Regular" END AS user_group
FROM
(
SELECT user_id, SUM(price) s,
CASE WHEN TIMESTAMPDIFF(DAY, MAX(o_date), date('2017-12-31')) > 60 THEN "1"
      WHEN TIMESTAMPDIFF(DAY, MAX(o_date), date('2017-12-31')) <= 60 AND TIMESTAMPDIFF(DAY, o_date, date('2017-12-31')) > 30 THEN "2"
      ELSE "3" END AS r,
CASE WHEN COUNT(*) <= 1 THEN "1"
      WHEN COUNT(*) > 1 AND COUNT(*) <= 4 THEN "2"
      ELSE "3" END AS f,
CASE WHEN SUM(price) < 5000 THEN "1"
      WHEN SUM(price) >= 5000 AND SUM(price) < 15000 THEN "2"
      ELSE "3" END AS m
FROM orders.orders_
GROUP BY user_id
) t_rfm
GROUP BY user_id) t_groups
WHERE user_group = "Vip"
```

Сначала создаются столбцы r,f,m: R1 days > 60, R2 30 < days <= 60, R3 days <= 30;  
F1 orders <= 1, F2 1 < orders <= 4, F3 > 4; M1 price < 5000, 5000 <= M2 < 15000; M3 >= 15000.  
Далее создаются столбцы users\_group согласно пункту 3 задания. После чего по каждому значению users\_group подсчитывается количество user\_id и общая сумма покупок.

4) — 5) пункты:

Экспортирую полученные значения в Excel, после чего проверяю сходятся ли суммы пользователей и покупок (строка Сумма). После чего считаю процент группы пользователей от всех и процент общей суммы чека группы от всей суммы чека.

	<u>Sum users</u>	<u>Sum price</u>	<u>% users</u>	<u>% price</u>
<u>All</u>	1015119	4542687441	100,00%	100,00%
<u>Vip</u>	17251	805627749	1,70%	17,73%
<u>Lost</u>	788069	2972151619	77,63%	65,43%
<u>Regular</u>	209799	764908073	20,67%	16,84%
Сумма	1015119	4542687441		