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

***1. Определяем критерии для каждой буквы R, F, M (т.е. к примеру, R – 3 для клиентов, которые покупали <= 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.)***

-- R – время с последнего действия

-- -- 3. 0-30 дней с даты покупки – высокий шанс покупки

-- -- 2. 30-60 – высокий риск ухода

-- -- 1. больше 60 - Чем больше промежуток, тем меньше шанс повторного заказа

-- F – количество действий за все время

-- -- 3. 3+ - отлично

-- -- 2. 2-3 – норм

-- -- 1. Меньше 2 - такое

-- M – сколько всего выручки с пользователя

-- -- 3. 15 тыс + - хорошо

-- -- 2. 7-15 тыс + - норм

-- -- 1. Меньше 7 тыс – такое

**DROP** **TEMPORARY** **TABLE** **IF** **EXISTS** rfm\_users;

**CREATE** **TEMPORARY** **TABLE** **IF** **NOT** **EXISTS** rfm\_users **AS**

**SELECT** user\_id,

**concat**(

(**CASE**

**WHEN** timestampdiff(**DAY**, **max**(new\_o\_date), **date**('2018-01-01')) > 0

**AND** timestampdiff(**DAY**, **max**(new\_o\_date), **date**('2018-01-01')) <= 30

**THEN** '3'

**WHEN** timestampdiff(**DAY**, **max**(new\_o\_date), **date**('2018-01-01')) > 30

**AND** timestampdiff(**DAY**, **max**(new\_o\_date), **date**('2018-01-01')) <= 60

**THEN** '2'

**ELSE** '1'

**END**), -- recency

'',

(**CASE**

**WHEN** **count**(id\_o) > 3

**THEN** '3'

**WHEN** **count**(id\_o) > 1

**AND** **count**(id\_o) <= 3

**THEN** '2'

**ELSE** '1'

**END**), -- frequency

'',

(**CASE**

**WHEN** **sum**(price) >= 15000

**THEN** '3'

**WHEN** **sum**(price) > 7000

**AND** **sum**(price) < 15000

**THEN** '2'

**ELSE** '1'

**END**) -- monetary

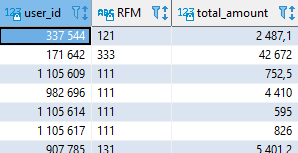
) **AS** RFM,

**sum**(price) **AS** total\_amount

**FROM** orders\_20190822

**GROUP** **BY** user\_id

***Получаем таблицу такого плана по базе за два года:***



***2. Для каждого пользователя получаем набор из 3 цифр (от 111 до 333, где 333 – самые классные пользователи)***

-- делаем сводную из полученной таблицы:

**DROP** **TEMPORARY** **TABLE** **IF** **EXISTS** RFM\_cross;

**CREATE** **TEMPORARY** **TABLE** **IF** **NOT** **EXISTS** RFM\_cross **AS**

**SELECT** RFM, **count**(user\_id) **AS** qty\_users, **sum**(total\_amount) **AS** total\_volume

**FROM** rfm\_users

**GROUP** **BY** RFM;

***Получаем выгрузку следующего вида:***

|  |  |  |
| --- | --- | --- |
| **RFM** | **Q-ty users** | **Total volume** |
| 111 | 590801 | 1 026 637 318,70 |
| 112 | 24616 | 240 696 484,00 |
| 113 | 8097 | 182 339 849,30 |
| 121 | 85269 | 266 389 528,30 |
| 122 | 17413 | 170 533 746,60 |
| 123 | 6893 | 176 863 976,10 |
| 131 | 16766 | 76 233 098,20 |
| 132 | 19374 | 202 544 690,60 |
| 133 | 18295 | 644 345 111,55 |
| 211 | 57102 | 105 382 365,20 |
| 212 | 2312 | 22 290 636,90 |
| 213 | 756 | 16 800 219,10 |
| 221 | 11358 | 38 996 552,00 |
| 222 | 2872 | 27 884 402,70 |
| 223 | 1144 | 29 150 123,80 |
| 231 | 2824 | 13 705 388,20 |
| 232 | 4621 | 49 232 178,10 |
| 233 | 6309 | 227 220 627,20 |
| 311 | 85583 | 153 637 796,20 |
| 312 | 2438 | 22 948 364,60 |
| 313 | 735 | 16 730 837,20 |
| 321 | 17162 | 58 023 861,70 |
| 322 | 3718 | 35 781 831,40 |
| 323 | 1243 | 31 387 033,30 |
| 331 | 4450 | 21 692 500,20 |
| 332 | 7631 | 80 869 999,00 |
| 333 | 11990 | 603 417 110,80 |
| **ИТОГО** | **1 011 772** | **4 541 735 630,95** |

***3. Вводим группировку, к примеру, 333 и 233 – это Vip, 1XX – это Lost, остальные Regular (можете ввести боле глубокую сегментацию)***

|  |  |  |  |
| --- | --- | --- | --- |
| **R** | **F** | **M** | **status** |
| 1 | 1 | \* | lost\_new |
| 1 | 2 | \* | lost\_reg |
| 1 | 3 | 1 | lost\_reg |
| 1 | 3 | 2 | lost\_reg |
| 1 | 3 | 3 | lost\_vip |
| 2 | 1 | \* | new |
| 3 | 1 | \* | new |
| 3 | 3 | 3 | vip |
| 2 | 3 | 3 | vip |
| 2 | 2 | \* | regular |
| 2 | 3 | 1 | regular |
| 2 | 3 | 2 | regular |
| 3 | 2 | \* | regular |
| 3 | 3 | 1 | regular |
| 3 | 3 | 2 | regular |

***4. Для каждой группы из п. 3 находим кол-во пользователей, кот. попали в них и % товарооборота, которое они сделали на эти 2 года.***

-- из сводной таблицы делаем еще одну с разбивкой по группам, которые спланировали ранее:

**SELECT**

**CASE**

**WHEN** RFM = 111 **OR** RFM = 112 **OR** RFM = 113

**THEN** 'lost\_new'

**WHEN** RFM = 121 **OR** RFM = 122 **OR** RFM = 123 **OR** RFM = 131 **OR** RFM = 132

**THEN** 'lost\_regular'

**WHEN** RFM = 133

**THEN** 'lost\_vip'

**WHEN** RFM = 211 **OR** RFM = 212 **OR** RFM = 213 **OR** RFM = 311 **OR** RFM = 312 **OR** RFM = 313

**THEN** 'new'

**WHEN** RFM = 233 **OR** RFM = 333

**THEN** 'vip'

**ELSE** 'regular'

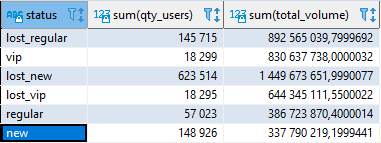
**END** **AS** status,

**sum**(qty\_users),

**sum**(total\_volume)

**FROM** RFM\_cross

**GROUP** **BY** status;



***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 собьются цифры). То же самое делаем и по деньгам.***

|  |  |  |  |
| --- | --- | --- | --- |
| **status** | **qty users** | **total\_volume** | **доля ТО** |
| lost\_regular | 145 715 | 892 565 040 | **20%** |
| vip | 18 299 | 830 637 738 | **18%** |
| lost\_new | 623 514 | 1 449 673 652 | **32%** |
| lost\_vip | 18 295 | 644 345 112 | **14%** |
| regular | 57 023 | 386 723 870 | **9%** |
| new | 148 926 | 337 790 219 | **7%** |
|  | **1 011 772** | **4 541 735 631** |  |

***Итоговое количество юзеров и сумма соответствует изначальной выгрузке, которую приводила ранее*** [***здесь.***](#ИТОГО)