#\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

#Создаём таблицу с пользователями с 3-мя и более покупками

**DROP** **TABLE** **IF** **EXISTS** active\_users;

**CREATE** **TABLE** active\_users(user\_id **BIGINT**);

# И заполняем её

**INSERT** **INTO** active\_users

**SELECT** user\_id

**FROM** orders

**WHERE** **datediff**(**DATE\_FORMAT**("2017.04.30", "%Y.%m.%d"), o\_date) > 0

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

**HAVING** **COUNT**(**DISTINCT** id\_o) > 3 **AND** **COUNT**(**DISTINCT** id\_o) < 100

**ORDER** **BY** user\_id

;

#\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

#Собираем информацию на тех, кто активен и покупал недавно

**DROP** **TABLE** **IF** **EXISTS** active\_users\_info;

**CREATE** **TABLE** active\_users\_info

(user\_id **BIGINT**,

first\_buy **DATE**,

last\_buy **DATE**,

total\_sum **BIGINT**,

pay\_by\_month **BIGINT**);

**INSERT** **INTO** active\_users\_info (user\_id, first\_buy, last\_buy, total\_sum)

**SELECT** **DISTINCT** a\_u.user\_id, t3.o\_date, t2.o\_date, t4.total\_sum

**FROM** active\_users a\_u

**JOIN**

(**SELECT** o1.user\_id, o1.o\_date

**FROM** orders o1

**LEFT** **JOIN** orders o2

**ON** o1.user\_id = o2.user\_id

**AND** o1.o\_date < o2.o\_date

**WHERE** o2.o\_date **IS** **NULL**

**ORDER** **BY** o1.user\_id) t2

**ON** (a\_u.user\_id = t2.user\_id)

**JOIN**

(**SELECT** o1.user\_id, o1.o\_date

**FROM** orders o1

**LEFT** **JOIN** orders o2

**ON** o1.user\_id = o2.user\_id

**AND** o1.o\_date > o2.o\_date

**WHERE** o2.o\_date **IS** **NULL**

**ORDER** **BY** o1.user\_id) t3

**ON** (a\_u.user\_id = t3.user\_id)

**JOIN**

(**SELECT** user\_id, **SUM**(price) **AS** total\_sum

**FROM** orders

**WHERE** **datediff**(**DATE\_FORMAT**("2017.04.30", "%Y.%m.%d"), o\_date) > 0

**GROUP** **BY** user\_id) t4

**ON** (a\_u.user\_id = t4.user\_id)

**WHERE** **DATEDIFF**(**DATE\_FORMAT**("2017.04.30", "%Y.%m.%d"), t2.o\_date) < 30

;

#\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

#Считаем среднюю сумму для активных и недавних

**DROP** **TABLE** **IF** **EXISTS** active\_users\_sum;

**CREATE** **TABLE** active\_users\_sum (

user\_id **BIGINT**,

pay\_by\_month **BIGINT**

);

**INSERT** **INTO** active\_users\_sum

**SELECT** user\_id, ((total\_sum / **datediff**(last\_buy, first\_buy))\*30) **AS** pay\_per\_month **FROM** active\_users\_info;

#\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

#Собираем информацию на тех, кто активен и покупал давно

**DROP** **TABLE** **IF** **EXISTS** active\_passive\_users\_info;

**CREATE** **TABLE** active\_passive\_users\_info

(user\_id **BIGINT**,

first\_buy **DATE**,

last\_buy **DATE**,

total\_sum **BIGINT**,

pay\_by\_month **BIGINT**,

buy\_count **BIGINT**

);

**INSERT** **INTO** active\_passive\_users\_info (user\_id, first\_buy, last\_buy, total\_sum, buy\_count)

**SELECT** **DISTINCT** a\_u.user\_id, t3.o\_date, t2.o\_date, t4.total\_sum, t5.buy\_count

**FROM** active\_users a\_u

**JOIN**

(**SELECT** o1.user\_id, o1.o\_date

**FROM** orders o1

**LEFT** **JOIN** orders o2

**ON** o1.user\_id = o2.user\_id

**AND** o1.o\_date < o2.o\_date

**WHERE** o2.o\_date **IS** **NULL**

**ORDER** **BY** o1.user\_id) t2

**ON** (a\_u.user\_id = t2.user\_id)

**JOIN**

(**SELECT** o1.user\_id, o1.o\_date

**FROM** orders o1

**LEFT** **JOIN** orders o2

**ON** o1.user\_id = o2.user\_id

**AND** o1.o\_date > o2.o\_date

**WHERE** o2.o\_date **IS** **NULL**

**ORDER** **BY** o1.user\_id) t3

**ON** (a\_u.user\_id = t3.user\_id)

**JOIN**

(**SELECT** user\_id, **SUM**(price) **AS** total\_sum

**FROM** orders

**WHERE** **datediff**(**DATE\_FORMAT**("2017.04.30", "%Y.%m.%d"), o\_date) > 0

**GROUP** **BY** user\_id) t4

**ON** (a\_u.user\_id = t4.user\_id)

**JOIN**

(**SELECT** user\_id, **COUNT**(**DISTINCT** id\_o) **AS** buy\_count

**FROM** orders

**WHERE** **datediff**(**DATE\_FORMAT**("2017.04.30", "%Y.%m.%d"), o\_date) > 0

**GROUP** **BY** user\_id) t5

**ON** (a\_u.user\_id = t5.user\_id)

**WHERE** **DATEDIFF**(**DATE\_FORMAT**("2017.04.30", "%Y.%m.%d"), t2.o\_date) > 30

;

#

#Считаем среднюю сумму для активных и давних

**DROP** **TABLE** **IF** **EXISTS** active\_passive\_users\_sum;

**CREATE** **TABLE** active\_passive\_users\_sum (

user\_id **BIGINT**,

pay\_by\_month **BIGINT**

);

**INSERT** **INTO** active\_passive\_users\_sum

**SELECT** user\_id, ((total\_sum / **datediff**(last\_buy, first\_buy))\*30) **AS** pay\_per\_month **FROM** active\_passive\_users\_info;

#\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**SELECT** **avg**(**DATEDIFF**(**DATE\_FORMAT**("2017.04.30", "%Y.%m.%d"), last\_buy)) **FROM** active\_passive\_users\_info;

**SELECT** **avg**(**DATEDIFF**(last\_buy, first\_buy)/ buy\_count) **FROM** active\_passive\_users\_info;

**SELECT** **COUNT**(user\_id) **FROM** active\_passive\_users\_info **WHERE** **DATEDIFF**(**DATE\_FORMAT**("2017.04.30", "%Y.%m.%d"), last\_buy) > 150 **AND** **DATEDIFF**(**DATE\_FORMAT**("2017.04.30", "%Y.%m.%d"), last\_buy) < 180;

# Примерно таким образом я посмотрел, сколько пользователей из общей массы активно-неактивных покупали последние полгода и нашёл процентное соотношение от общей массы.

# Получилось примерно 7%, это число и возьмём для расчёта

**SELECT** **COUNT**(user\_id) **FROM** active\_passive\_users\_info

**UNION**

**SELECT** **COUNT**(user\_id) **FROM** active\_users\_info

**UNION**

**SELECT** **COUNT**(user\_id) **FROM** active\_users;

#\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

#Создаём таблицу с пользователями с менее чем 3-мя покупками

**DROP** **TABLE** **IF** **EXISTS** inactive\_users;

**CREATE** **TABLE** inactive\_users(user\_id **BIGINT**);

# И заполняем её

**INSERT** **INTO** inactive\_users

**SELECT** user\_id

**FROM** orders

**WHERE** **datediff**(**DATE\_FORMAT**("2017.04.30", "%Y.%m.%d"), o\_date) > 0

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

**HAVING** **COUNT**(**DISTINCT** id\_o) < 3

**ORDER** **BY** user\_id

;

#\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**SELECT** **SUM**(t2.price), **MONTH**(t2.o\_date), **YEAR**(t2.o\_date)

**FROM** inactive\_users t1

**JOIN** orders t2 **ON** (t1.user\_id = t2.user\_id)

**GROUP** **BY** **MONTH**(t2.o\_date), **YEAR**(t2.o\_date)

**ORDER** **BY** **YEAR**(t2.o\_date), **MONTH**(t2.o\_date);

**SELECT** **SUM**(price), **MONTH**(o\_date), **YEAR**(o\_date)

**FROM** orders

**GROUP** **BY** **MONTH**(o\_date), **YEAR**(o\_date);

# В Excel сравниваем результаты и приходим к выводу, что доля таких покупателей занимает примерно 50% всей выручки за месяц

**SELECT** **AVG**(pay\_by\_month) \* **COUNT**(user\_id) **FROM** active\_users\_sum; #=54827421.9142

**SELECT** **AVG**(pay\_by\_month) \* **COUNT**(user\_id) \* 0.07 **FROM** active\_passive\_users\_sum; #=10683544.260549

#Сумму предыдущих шагов умножаем на 1.50, получаем 98 266 449.26

**SELECT** **SUM**(price) **FROM** orders **WHERE** **YEAR**(o\_date) = 2017 **AND** **MONTH**(o\_date) = 5; # 217075552.50557518

#Получилось на -118 809 103.2 меньше, ошиблись на 54%, ужас)

#При этом если банально посчитать, во сколько раз больше выручка для того же месяца через год по предыдущим месяцам, выходит, что в среднем на 70%

#Тогда для прогноз на май выходит 184 755 325, что на 14.9% меньше реального значения, уже не 50%, но всё равно плохо

#Буду признателен за комментарии