**В качестве ДЗ делам прогноз ТО на 05.2017. В качестве метода прогноза - считаем сколько денег тратят группы клиентов в день:**

**1. Группа часто покупающих (3 и более покупок) и которые последний раз покупали не так давно.**

-- Считаем сколько денег оформленного заказа приходится на 1 день. Умножаем на 30.

**CREATE** **TEMPORARY** **TABLE** **IF** **NOT** **EXISTS** forecast\_2017\_05 **AS**

(

**SELECT** user\_id, **sum**(price)/(timestampdiff(**DAY**, **min**(new\_o\_date), **date**('2017-05-01'))) \* 31 **AS** forecast

**FROM** orders\_20190822

**WHERE** **date**(new\_o\_date) < **date**('2017-05-01')

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

**HAVING** **count**(id\_o) > 2 -- пользователи, у которых 3+ покупок,

**AND** (timestampdiff(**DAY**, **max**(new\_o\_date), **date**('2017-05-01'))) <= 60 -- последний заказ не более 60 дней назад

)

-- ---------------------

-- прогноз на май 2017

-- -- итог 111453993.3725687

**SELECT** **sum**(forecast) **AS** forecast\_05\_2017

**FROM** forecast\_2017\_05



-- факт оборот за май 2017 по тем пользователям, которые попали в прогнозную выборку по условиям задания

-- -- итог 50641804.4999995

**SELECT** **sum**(price) **AS** fact\_05\_2017

**FROM** orders\_20190822 o

**JOIN** (**SELECT** user\_id **FROM** forecast\_2017\_05) **AS** t

**ON** o.user\_id = t.user\_id

**WHERE** **month**(new\_o\_date) = 5

**AND** **year**(new\_o\_date) = 2017 -- т.к. в таблице два года



***Расхождение очень большое! На уроке мы обсуждали, что есть несколько способов откалибровать такое расхождение, используя данные 2016 года. На уроке обсуждали соотношение между фактом и прогнозом для разных месяцев, вычленили кэф 3.***

***Для интереса в домашке использовала другой способ***

-- найти соотношение между маем и апрелем 2016 года для данной группы пользователей и применить данный кэф для 2017 года

-- -- выборка id пользователей категории 3+ за апрель 2016

**CREATE** **TEMPORARY** **TABLE** **IF** **NOT** **EXISTS** users\_2016\_04 **AS**

(

**SELECT** user\_id

**FROM** orders\_20190822

**WHERE** **date**(new\_o\_date) < **date**('2016-05-01')

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

**HAVING** **count**(id\_o) > 2

**AND** (timestampdiff(**DAY**, **max**(new\_o\_date), **date**('2016-05-01'))) <= 60

);

-- -- факт сумма по пользователям категории 3+ за май 2016

-- -- итог 24409660.80000082

**SELECT** **sum**(price) **AS** fact\_05\_2016

**FROM** orders\_20190822 o

**JOIN** (**SELECT** user\_id **FROM** users\_2016\_04) **AS** t

**ON** o.user\_id = t.user\_id

**WHERE** **month**(new\_o\_date) = 5

**AND** **year**(new\_o\_date) = 2016



-- -- факт сумма по пользователям категории 3+ за апрель 2016

-- -- апрель 2016 50427776.69999895

**SELECT** **sum**(price) **AS** fact\_04\_2016

**FROM** orders\_20190822 o

**JOIN** (**SELECT** user\_id **FROM** users\_2016\_04) **AS** t

**ON** o.user\_id = t.user\_id

**WHERE** **month**(new\_o\_date) = 4

**AND** **year**(new\_o\_date) = 2016



***Таким образом на примере 2016 года видно, что заказы по данной группе пользователей сократились почти в два раза в мае по сравнению с апрелем. Теперь найдем ТО для данной группы пользователей в апреле 2017 года и применим полученный кэф***

-- -- выборка факт суммы по пользователям категории 3+ за апрель 2017

-- апрель 2017 87836087.49999124

**SELECT** **sum**(price) **AS** fact\_04\_2017

**FROM** orders\_20190822 o

**JOIN** (**SELECT** user\_id **FROM** forecast\_2017\_05) **AS** t

**ON** o.user\_id = t.user\_id

**WHERE** **month**(new\_o\_date) = 4

**AND** **year**(new\_o\_date) = 2017 -- т.к. в таблице два года



***Таким образом с учетом кэфа прогнозные продажи по данной группе пользователей на май 2017 года составят порядка 44 000 000, что не далеко от факта (ближе, чем прошлый прогноз с анализом денег, которые клиент приносит за день)***



**2. Группа часто покупающих, но которые не покупали уже значительное время. Так же можем сделать вывод, из такой группы за след месяц сколько купят и на какую сумму. (постараться продумать логику)**

***Способ анализа соотношения месяц к месяцу в 2016 году и применение кэфа к 2017 году оказался неэффективным для этой группы. Скорее всего это связано с тем, что с начала 2016 года слишком маленький промежуток времени для анализа.***

***Другой метод оказался более эффективным: я взяла пользователей, которые соответствуют условию на конец февраля 2017 и проверила на какую сумму они сделали заказы в марте 2017***

-- -- выборка id по пользователям, которые соответствуют условию на конец февраля 2017

**DROP** **TABLE** **IF** **EXISTS** users\_2017\_02;

**CREATE** **TEMPORARY** **TABLE** **IF** **NOT** **EXISTS** users\_2017\_02 **AS**

(

**SELECT** user\_id

**FROM** orders\_20190822

**WHERE** **date**(new\_o\_date) < **date**('2017-03-01')

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

**HAVING** **count**(id\_o) > 2 -- пользователи, у которых 3+ покупок,

**AND** (timestampdiff(**DAY**, **max**(new\_o\_date), **date**('2017-03-01'))) > 60 -- последний заказ более 60 дней назад

)

-- -- на какую сумму они сделали заказы в марте 2017

-- факт март 2017 - 14258064.100000124

**SELECT** **sum**(price) **AS** fact\_03\_2017

**FROM** orders\_20190822 o

**JOIN** (**SELECT** user\_id **FROM** users\_2017\_02) **AS** t

**ON** o.user\_id = t.user\_id

**WHERE** **month**(new\_o\_date) = 3

**AND** **year**(new\_o\_date) = 2017 -- т.к. в таблице два года



***Таким же образом нашла id пользователей из нужной группы в марте 2017 и отследила их заказы в апреле 2017***



***Получается, что такие пользователи делали покупки в среднем на сумму 13 млн в предыдущие два месяца. Этот показатель соответствует фактической ситуации за май 2017 для пользователей данной категории:***



**3. Отдельно разобрать пользователей с 1 и 2 покупками за все время, прогнозируем их.**

***Попыталась сделать прогноз, высчитав сколько денег за день приносит пользователь с двумя покупками, который делал заказы не более 60 дней назад, и затем спроецировав эту сумму на май 2017. Расхождение колоссальное.***



***Делаю прогноз по другой схеме – высчитаю отношение мая к апрелю 2016 года для данной группы пользователей и применю этот кэф для 2017 года***

-- найти соотношение между маем и апрелем 2016 года для данной группы пользователей и применить данный кэф для 2017 года

-- -- выборка id пользователей категории 2 и менее заказов на конец апреля 2016

**DROP** **TEMPORARY** **TABLE** **IF** **EXISTS** users\_2016\_04;

**CREATE** **TEMPORARY** **TABLE** **IF** **NOT** **EXISTS** users\_2016\_04 **AS**

(

**SELECT** user\_id

**FROM** orders\_20190822

**WHERE** **date**(new\_o\_date) < **date**('2016-05-01')

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

**HAVING** **count**(id\_o) <= 2 -- пользователи, у которых 2 и менее заказов покупок,

**AND** (timestampdiff(**DAY**, **max**(new\_o\_date), **date**('2016-05-01'))) <= 60 -- последний заказ не более 60 дней назад

);

-- -- выборка факт суммы по пользователям категории 2 и менее заказов за апрель 2016

-- апрель 2016 - 88823746.19999178

**SELECT** **sum**(price) **AS** fact\_04\_2016

**FROM** orders\_20190822 o

**JOIN** (**SELECT** user\_id **FROM** users\_2016\_04) **AS** t

**ON** o.user\_id = t.user\_id

**WHERE** **month**(new\_o\_date) = 4

**AND** **year**(new\_o\_date) = 2016



-- -- выборка факт суммы по пользователям категории 2 и менее заказов за май 2016

-- май 2016 - 13580379.400000185

**SELECT** **sum**(price) **AS** fact\_05\_2016

**FROM** orders\_20190822 o

**JOIN** (**SELECT** user\_id **FROM** users\_2016\_04) **AS** t

**ON** o.user\_id = t.user\_id

**WHERE** **month**(new\_o\_date) = 5

**AND** **year**(new\_o\_date) = 2016



***Кэф май к апрелю – 0,15. Найдем фактические продажи для данной группы пользователей в апреле 2017 года и с помощью кэфа сделаем прогноз на май 2017***

-- -- выборка id пользователей категории 2 и менее заказов на конец апреля 2017

**DROP** **TEMPORARY** **TABLE** **IF** **EXISTS** users\_2017\_04;

**CREATE** **TEMPORARY** **TABLE** **IF** **NOT** **EXISTS** users\_2017\_04 **AS**

(

**SELECT** user\_id

**FROM** orders\_20190822

**WHERE** **date**(new\_o\_date) < **date**('2017-05-01')

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

**HAVING** **count**(id\_o) <= 2 -- пользователи, у которых 2 и менее заказов покупок,

**AND** (timestampdiff(**DAY**, **max**(new\_o\_date), **date**('2017-05-01'))) <= 60 -- последний заказ не более 60 дней назад

);

-- -- выборка факт суммы по пользователям категории 2 и менее заказов за апрель 2017

-- апрель 2017 109790618.69997722

**SELECT** **sum**(price) **AS** fact\_04\_2017

**FROM** orders\_20190822 o

**JOIN** (**SELECT** user\_id **FROM** users\_2017\_04) **AS** t

**ON** o.user\_id = t.user\_id

**WHERE** **month**(new\_o\_date) = 4

**AND** **year**(new\_o\_date) = 2017 -- т.к. в таблице два года



***Применяем кэф и прогнозная сумма по данной группе пользователей – 16 786 032,12, что не так уж далеко от факта***



***Для пользователей с двумя покупками, которые делали заказы более 60 дней назад не оптимально использовать кэфы 2016 года, т.к. в 2016 году слишком маленький промежуток времени для анализа.***

***Как и для пользователей с 3+ покупками я взяла пользователей, которые соответствуют условию на конец февраля 2017 и проверила на какую сумму они сделали заказы в марте 2017***



***Таким же образом нашла id пользователей из нужной группы в марте 2017 и отследила их заказы в апреле 2017***



***Получается, что такие пользователи делали покупки в среднем на сумму 15 млн в предыдущие два месяца. Этот показатель близок к фактической ситуации за май 2017 для пользователей данной категории:***



***Для пользователей с 1 покупкой находим кэф соотношения между пользователями с 1 покупкой в апреле 2016 года с их фактическими покупками в мае 2016 года. Затем находим аналогичных пользователей в апреле 2017 года и с помощью кэфа делаем прогноз на май 2017 года.***

-- -- Находим id пользователей с 1 заказом в апреле 2016 года

**DROP** **TEMPORARY** **TABLE** **IF** **EXISTS** users\_2016\_04;

**CREATE** **TEMPORARY** **TABLE** **IF** **NOT** **EXISTS** users\_2016\_04 **AS**

(

**SELECT** user\_id

**FROM** orders\_20190822

**WHERE** **date**(new\_o\_date) < **date**('2016-05-01')

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

**HAVING** **count**(id\_o) = 1 -- пользователи, у которых 1 заказ

);

-- -- выборка факт суммы по пользователям с 1 заказом за апрель 2016

-- апрель 2016 65885910.29999638

**SELECT** **sum**(price) **AS** fact\_05\_2016

**FROM** orders\_20190822 o

**JOIN** (**SELECT** user\_id **FROM** users\_2016\_04) **AS** t

**ON** o.user\_id = t.user\_id

**WHERE** **month**(new\_o\_date) = 4

**AND** **year**(new\_o\_date) = 2016



-- -- выборка факт суммы по пользователям с 1 заказом за май 2016

-- май 2016 11173024.800000062

**SELECT** **sum**(price) **AS** fact\_05\_2016

**FROM** orders\_20190822 o

**JOIN** (**SELECT** user\_id **FROM** users\_2016\_04) **AS** t

**ON** o.user\_id = t.user\_id

**WHERE** **month**(new\_o\_date) = 5

**AND** **year**(new\_o\_date) = 2016



***Получается, что коэф соотношения мая к апрелю 2016 для пользователей с 1 заказом порядка 0,17***

***Аналогичным образом находим id пользователей с 1м заказом в апреле 2017 года и вычисляем ТО по ним за апрель 2017***



***Применяем кэф 2016 года и делаем прогноз для этих пользователей на май 2017 – 14 338 460,42***

***Близко, однако с фактом есть существенное расхождение:***



***Осталось сделать прогноз по пользователям, которые зарегистрируются и сделают первый заказ в мае 2017 года.***

-- -- пользователи, которые новые в мае - сравнить долю таких заказов к общей сумме в 2016 и применить коэф

-- -- -- это общий товарооборот по месяцам

**SELECT** (**concat**(**month**(new\_o\_date), '-', **year**(new\_o\_date))), **sum**(price)

**FROM** orders\_20190822

**GROUP** **BY** (**concat**(**month**(new\_o\_date), '-', **year**(new\_o\_date)));

-- -- -- это ТО по месяцам только по первым заказам

**SELECT** (**concat**(**month**(new\_o\_date), '-', **year**(new\_o\_date))), **sum**(price)

**FROM** orders\_20190822 o2

**WHERE** id\_o **IN**

(**SELECT** **min**(id\_o)

**FROM** orders\_20190822 o3

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

)

**GROUP** **BY** (**concat**(**month**(new\_o\_date), '-', **year**(new\_o\_date)));

***Получив данные, я проанализировала их в Excel – доля первых заказов по отношению к общему ТО в мае 2016 года равна 53%***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **ОБЩИЙ ТО 2016** | | **ТО ПЕРВЫЕ ЗАКАЗЫ 2016** | | **ДОЛЯ 1 ЗАКАЗОВ** |
| янв.16 | 96 759 906,95 | янв.16 | 75 854 606,10 | 78% |
| фев.16 | 95 393 477,90 | фев.16 | 61 399 819,60 | 64% |
| мар.16 | 115 204 009,90 | мар.16 | 69 783 732,90 | 61% |
| апр.16 | 139 251 522,90 | апр.16 | 77 388 738,00 | 56% |
| май.16 | 108 522 070,30 | май.16 | 57 294 786,50 | 53% |
| июн.16 | 116 599 742,70 | июн.16 | 59 274 145,00 | 51% |
| июл.16 | 113 116 567,20 | июл.16 | 58 658 016,20 | 52% |
| авг.16 | 139 492 714,20 | авг.16 | 66 065 085,10 | 47% |
| сен.16 | 152 859 092,40 | сен.16 | 77 574 611,10 | 51% |
| окт.16 | 211 599 255,00 | окт.16 | 110 357 001,30 | 52% |
| ноя.16 | 256 630 130,40 | ноя.16 | 141 289 666,70 | 55% |
| дек.16 | 258 901 675,20 | дек.16 | 146 588 827,00 | 57% |
|  | 1 804 330 165,05 |  | 1 001 529 035,50 | 56% |

***Сделанные ранее прогнозы:***

|  |  |
| --- | --- |
| **Группа покупателей** | **Прогноз ТО на май 2017** |
| 3+ <=60 | 44 000 000 |
| 3+ >60 | 13 000 000 |
| 2 <=60 | 16 786 032 |
| 2 >60 | 15 000 000 |
| 1 | 14 338 460 |
|  | **103 124 493** |

***103 124 493 – это прогноз на май 2017 по всем заказам, кроме первых, и это предполагаемо составляет 47% общего ТО, т.к. ранее нашли кэф 53% для первых заказов в мае***

***Таким образом полный проноз на май 2017 года:***

|  |  |  |
| --- | --- | --- |
|  | **103 124 493** | 47% |
| NEW | **116 289 321** | 53% |
| общий ТО | **219 413 814** | 100% |

***Сравним с фактом:***

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Прогноз** | **Факт** | **Дельта %** |
|  | **103 124 493** | 108 048 458,70 | -4,6 |
| NEW | **116 289 321** | 108 048 458,70 | 7,6 |
| общий ТО | **219 413 814** | 217 040 888,40 | 1,1 |

***Я считаю, что расхождения минимальные***