

-- ДЗ №3

-- В качестве ДЗ делам прогноз Т0 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.

```
select
    o_date,
    round(sum(price)) * 30 as money
from
    (
        select
            tb_3.user_id,
            price,
            o_date
        from orders_20190822 o
        join
            (
                select user_id
                from
                    (
                        select *,
                            count(*) as count_o,
                            max(o_date) as max_o_date,
                            sum(price) as sum_price
                        from
                            (
                                select *
                                from orders_20190822
                                where o_date < '2017-05-01'
                            ) as tb_1
                        group by user_id
                    ) as tb_2
                where count_o >= 3
                    and max_o_date > '2017-04-01'
            ) as tb_3
        on o.user_id = tb_3.user_id
    ) as tb_4
where o_date < '2017-06-01'
group by o_date;
```

	o_date	money
1	2016-01-01	3 192 930
2	2016-01-02	6 110 400
3	2016-01-03	7 033 080
4	2016-01-04	10 756 770
5	2016-01-05	9 585 330
6	2016-01-06	9 229 020
7	2016-01-07	7 114 200
8	2016-01-08	10 596 450
9	2016-01-09	9 012 810
10	2016-01-10	8 188 320

-- Добавим в результирующую таблицу дни недели и рассчитаем сумму денег на каждый день:

Названия строк	Сумма по полю money
1	1 542 952 800
2	2 212 620 180
3	2 936 783 880
4	2 586 078 780
5	2 344 184 760
6	1 983 467 160
7	1 583 741 340
Общий итог	15 189 828 900

-- Посчитаем во сколько раз увеличились продажи в 2017 году по сравнению с 2016

Названия строк	Сумма по полю money
2016	8 783 900 130
2017	6 390 740 610
Общий итог	15 174 640 740

-- Т.к. в анализе участвует 12 месяцев 2016г и 4 месяцев 2017, то полученные результаты разделим на 12 и 4 соответственно, а затем и друг на друга

$$k = (8\,783\,900\,130 / 12) / (6\,390\,740\,610 / 4) = 2.183$$

-- Допустим, что в следующем месяце эта группа клиентов тратит денег столько же, сколько и в прошлом году умноженные на коэффициент увеличения продаж

-- для первой строки:

$$26\,464\,050 * 2.183 = 11\,835\,773$$

-- просуммируем значения за месяц и сравним с истинными значениями:

$$\text{Ошибка} = (1\,090\,092\,806 / 1\,240\,991\,190) - 1 = -12\%$$

-- Результаты вычислений сведем в таблицу:

O_date	money	ДН	o_date	money	ДН	Предсказанные значения
01.05.2017	26 464 050	1	01.05.2016	5 422 650	7	11 835 773
02.05.2017	47 129 670	2	02.05.2016	5 609 610	1	12 243 842
03.05.2017	46 857 990	3	03.05.2016	7 676 550	2	16 755 259
04.05.2017	47 514 090	4	04.05.2016	12 607 800	3	27 518 475
05.05.2017	41 573 130	5	05.05.2016	9 728 760	4	21 234 525
06.05.2017	31 626 060	6	06.05.2016	9 447 360	5	20 620 326
07.05.2017	30 655 200	7	07.05.2016	8 090 160	6	17 658 027
08.05.2017	28 526 100	1	08.05.2016	5 634 120	7	12 297 339
09.05.2017	29 188 350	2	09.05.2016	7 243 500	1	15 810 060
10.05.2017	44 185 110	3	10.05.2016	12 084 360	2	26 375 986
11.05.2017	48 813 990	4	11.05.2016	22 758 060	3	49 672 989
12.05.2017	43 736 430	5	12.05.2016	18 517 920	4	40 418 227
13.05.2017	31 703 100	6	13.05.2016	18 128 490	5	39 568 236
14.05.2017	33 167 970	7	14.05.2016	15 348 030	6	33 499 451
15.05.2017	34 911 030	1	15.05.2016	14 731 110	7	32 152 928
16.05.2017	53 088 390	2	16.05.2016	18 105 060	1	39 517 096
17.05.2017	57 505 350	3	17.05.2016	20 316 420	2	44 343 732
18.05.2017	45 882 000	4	18.05.2016	21 322 440	3	46 539 526
19.05.2017	39 077 760	5	19.05.2016	20 051 790	4	43 766 136
20.05.2017	31 610 760	6	20.05.2016	19 666 170	5	42 924 460
21.05.2017	37 555 110	7	21.05.2016	14 317 140	6	31 249 374
22.05.2017	35 497 200	1	22.05.2016	15 395 250	7	33 602 516
23.05.2017	76 395 030	2	23.05.2016	22 817 970	1	49 803 752
24.05.2017	53 507 340	3	24.05.2016	50 549 550	2	110 332 218
25.05.2017	43 308 480	4	25.05.2016	29 142 930	3	63 608 956
26.05.2017	35 664 540	5	26.05.2016	18 947 790	4	41 356 485
27.05.2017	25 372 740	6	27.05.2016	16 347 120	5	35 680 120
28.05.2017	28 423 740	7	28.05.2016	12 332 130	6	26 916 783
29.05.2017	31 849 650	1	29.05.2016	10 116 420	7	22 080 653
30.05.2017	39 161 670	2	30.05.2016	16 528 050	1	36 075 028
31.05.2017	41 039 160	3	31.05.2016	20 449 650	2	44 634 527
1 240 991 190						1 090 092 806
						-12%

-- 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)

```
select
    o_date,
    round(sum(price)) * 30 as money
from
    (
        select
            tb_3.user_id,
            price,
            o_date
        from orders_20190822 o
        join
            (
                select user_id
                from
                    (
                        select *,
                            count(*) as count_o,
                            max(o_date) as max_o_date,
                            sum(price) as sum_price
                        from
                            (
                                select *
                                from orders_20190822
                                where o_date < '2017-05-01'
                            ) as tb_1
                        group by user_id
                    ) as tb_2
                where count_o >= 3
                    and max_o_date < '2017-04-01'
            ) as tb_3
        on o.user_id = tb_3.user_id
    ) as tb_4
where o_date <= '2017-06-01'
group by o_date;
```

	o_date	money
1	2016-01-01	15 565 800
2	2016-01-02	20 222 610
3	2016-01-03	25 589 010
4	2016-01-04	28 354 350
5	2016-01-05	28 811 880
6	2016-01-06	28 858 680
7	2016-01-07	22 096 320
8	2016-01-08	24 992 610
9	2016-01-09	24 305 250
10	2016-01-10	28 724 400

-- Добавим в результирующую таблицу дни недели и рассчитаем сумму денег на каждый день:

Названия строк	Сумма по полю money
1	3 019 465 650
2	4 221 782 490
3	3 725 395 620
4	3 457 710 240
5	2 774 687 790
6	2 320 903 380
7	2 354 058 090
Общий итог	21 874 003 260

-- Посчитаем во сколько раз увеличились продажи в 2017 году по сравнению с 2016г

Названия строк	Сумма по полю money
2016	17 974 674 420
2017	3 899 328 840
Общий итог	21 874 003 260

-- Т.к. в анализе участвует 12 месяцев 2016г и 4 месяцев 2017, то полученные результаты разделим на 12 и 4 соответственно, а затем и друг на друга

$$k = (17\,974\,674\,420 / 12) / (3\,899\,328\,840 / 4) = 0.651$$

-- Допустим, что в следующем месяце эта группа клиентов тратит денег столько же, сколько и в прошлом году умноженные на коэффициент увеличения продаж

-- для первой строки:

$$17\,914\,530 * 0.651 = 11\,658\,844$$

-- просуммируем значения за месяц и сравним с истинными значениями:

$$\text{Ошибка} = (816\,084\,836 / 663\,367\,860) - 1 = 23\%$$

-- Результаты вычислений сведем в таблицу:

o_date	money	ДН	o_date	money	ДН	Предсказанные значения
01.05.2017	17 196 900	1	01.05.2016	17 914 530	7	11 658 844
02.05.2017	22 963 560	2	02.05.2016	23 484 120	1	15 283 555
03.05.2017	27 637 800	3	03.05.2016	23 813 040	2	15 497 617
04.05.2017	19 915 590	4	04.05.2016	30 593 580	3	19 910 418
05.05.2017	22 740 150	5	05.05.2016	29 445 330	4	19 163 133
06.05.2017	19 015 830	6	06.05.2016	24 389 130	5	15 872 539
07.05.2017	15 173 880	7	07.05.2016	24 329 700	6	15 833 861
08.05.2017	14 334 420	1	08.05.2016	18 943 230	7	12 328 326
09.05.2017	17 395 740	2	09.05.2016	22 900 530	1	14 903 752
10.05.2017	24 051 540	3	10.05.2016	31 171 800	2	20 286 726
11.05.2017	27 456 540	4	11.05.2016	50 573 160	3	32 913 205
12.05.2017	19 808 430	5	12.05.2016	38 037 570	4	24 754 995
13.05.2017	14 524 320	6	13.05.2016	39 727 500	5	25 854 808
14.05.2017	18 816 390	7	14.05.2016	36 372 390	6	23 671 290
15.05.2017	25 075 170	1	15.05.2016	36 035 820	7	23 452 249
16.05.2017	27 884 250	2	16.05.2016	53 900 520	1	35 078 664
17.05.2017	30 398 280	3	17.05.2016	52 598 250	2	34 231 141
18.05.2017	27 149 460	4	18.05.2016	56 090 130	3	36 503 670
19.05.2017	29 543 190	5	19.05.2016	48 935 100	4	31 847 149
20.05.2017	16 803 870	6	20.05.2016	41 671 470	5	27 119 951
21.05.2017	18 909 570	7	21.05.2016	37 770 120	6	24 580 938
22.05.2017	19 157 970	1	22.05.2016	36 738 360	7	23 909 465
23.05.2017	37 247 670	2	23.05.2016	44 978 760	1	29 272 348
24.05.2017	27 251 130	3	24.05.2016	133 722 330	2	87 027 001
25.05.2017	25 417 800	4	25.05.2016	67 802 400	3	44 126 060
26.05.2017	13 982 310	5	26.05.2016	43 306 590	4	28 184 094
27.05.2017	13 350 330	6	27.05.2016	34 562 160	5	22 493 185
28.05.2017	13 210 170	7	28.05.2016	27 130 680	6	17 656 750
29.05.2017	18 766 620	1	29.05.2016	30 045 870	7	19 553 967
30.05.2017	16 434 900	2	30.05.2016	47 134 170	1	30 675 097
31.05.2017	21 754 080	3	31.05.2016	49 846 110	2	32 440 038
663 367 860						816 084 836
						23%

-- 3. Отдельно разобрать пользователей с 1 и 2 покупками за все время.

```
select
  o_date,
  round(sum(price)) * 30 as money
from
  (
    select
      tb_2.user_id,
      price,
      o_date
    from
      (
        select
          user_id,
          count(*) as count_o
        from
          (
            select *
            from orders_20190822
            where o_date < '2017-05-01'
          ) as tb_1
        group by user_id
      ) as tb_2
    join
      orders_20190822 o
    on
      o.user_id = tb_2.user_id
    where count_o < 3
  ) as tb_3
where o_date < '2017-06-01'
group by o_date;
```

	o_date	money
1	2016-01-01	15 972 660
2	2016-01-02	28 546 800
3	2016-01-03	38 965 140
4	2016-01-04	36 203 490
5	2016-01-05	44 987 520
6	2016-01-06	39 756 540
7	2016-01-07	33 620 580
8	2016-01-08	38 172 480
9	2016-01-09	46 219 650
10	2016-01-10	52 813 230

-- Предположим, что эта группа от года к году поведение свое не меняет.

-- просуммируем значения за месяц и сравним с истинными значениями:

Ошибка = $(1\ 495\ 610\ 850 / 903\ 286\ 050) - 1 = 66\%$

-- Предположим, что большая ошибка прогнозирования связана с тем, что данные начинаются с 2016г и пользователи, делающие покупки в мае 2016г, также делали покупки и ранее, и на самом деле не относятся к этой группе.

-- Результаты вычислений сведем в таблицу:

O_date	money	ДН	o_date	money
01.05.2017	24 282 450	1	01.05.2016	21 413 460
02.05.2017	30 728 790	2	02.05.2016	22 226 790
03.05.2017	34 568 220	3	03.05.2016	32 075 460
04.05.2017	31 845 000	4	04.05.2016	38 526 390
05.05.2017	33 747 540	5	05.05.2016	34 805 280
06.05.2017	23 483 940	6	06.05.2016	38 474 580
07.05.2017	22 647 150	7	07.05.2016	26 814 090
08.05.2017	25 133 370	1	08.05.2016	27 039 360
09.05.2017	21 854 820	2	09.05.2016	25 027 380
10.05.2017	30 640 260	3	10.05.2016	45 872 280
11.05.2017	33 592 920	4	11.05.2016	51 852 360
12.05.2017	26 044 320	5	12.05.2016	48 330 960
13.05.2017	25 944 150	6	13.05.2016	46 892 160
14.05.2017	19 850 130	7	14.05.2016	45 437 610
15.05.2017	30 048 600	1	15.05.2016	50 081 040
16.05.2017	33 607 440	2	16.05.2016	64 957 650
17.05.2017	46 590 870	3	17.05.2016	60 141 390
18.05.2017	32 936 940	4	18.05.2016	62 102 670
19.05.2017	31 069 290	5	19.05.2016	59 451 510
20.05.2017	19 802 580	6	20.05.2016	51 249 090
21.05.2017	29 083 140	7	21.05.2016	41 489 370
22.05.2017	25 508 910	1	22.05.2016	41 059 530
23.05.2017	43 666 200	2	23.05.2016	51 970 710
24.05.2017	37 220 160	3	24.05.2016	109 149 420
25.05.2017	31 542 900	4	25.05.2016	90 562 590
26.05.2017	26 179 140	5	26.05.2016	66 702 060
27.05.2017	26 866 920	6	27.05.2016	40 049 100
28.05.2017	24 499 440	7	28.05.2016	36 712 380
29.05.2017	24 464 550	1	29.05.2016	47 532 900
30.05.2017	27 195 960	2	30.05.2016	59 012 430
31.05.2017	28 639 950	3	31.05.2016	58 598 850
	903 286 050			1 495 610 850
				66%