
Задача 1

Раз в 10 дней компания отправляет всем действующим клиентам предложение закрыть текущий займ досрочно и получить новый на лучших условиях. Как можно оценить эффективность такой рассылки?

1. Самое важное - это конечно же прибыльность данного мероприятия, иначе в чем смысл, верно?
 - а. Выгодно ли компании делать предложение на лучших условиях в общем случае, и как часто люди принимают выгодное для компании решение?
 - б. Если каналов, с помощью которых отправляют предложение - несколько, покрывают ли выгодные решения содержание этих каналов?
 - в. Действительно ли это выгодное предложение, или это может впоследствии вести к репутационным издержкам?
 - г. Часто ли люди досрочно закрывают займ, но не соглашаются на новый после? Выгодно ли это?
 - д. Приоритетнее удержать клиента или выдать новый займ?
2. Отследить, доставляются и просматриваются ли отправленные предложения, или основная часть просто игнорируется или помечается как спам
3. Стоит протестировать, лучше ли всего работает "раз в 10 дней"? Возможно лучше раз в 15 дней (зп/аванс), или ближе к платежу по займу (обычно раз в 30 дней)?
4. Пересекаются ли между собой разные способы отправки предложения, или для всех он один (един)?
5. Оценить вероятность просмотра и выгодного решения в зависимости от вида рассылки:
смс / пуш / email / звонок
6. Как близко к погашению займа люди соглашаются на предложенные условия? (возможно стоит делать предложение, когда он погашен хотя бы наполовину)

Задача 2

Вам дана sql таблица с действиями клиентов на сайте:

client_id - int	created_at - timestamp	action_type - int
1	2021-01-01 12:00:00	1
1	2021-01-01 12:01:00	1
1	2021-01-01 12:02:00	2

Подозрительными считаются действия разного типа (action_type) - совершенные с интервалом менее 10 секунд одним пользователем. напишите sql запрос который выведет все такие действия.

(client_id | created_at | prev_created_at* | diff_sec* | action_type | prev_action_type | suspicious*)

```
SELECT
client_id,
created_at,
prev_created_at, --можно удалить
diff_sec, --можно удалить
action_type,
prev_action_type,
act_is_diff is true AND diff_sec <10 as suspicious --можно удалить
FROM(
    SELECT
        client_id,
        created_at,
        action_type,

        LAG(created_at) OVER(PARTITION BY client_id
                            ORDER BY created_at) AS prev_created_at, --можно удалить

        EXTRACT(EPOCH FROM created_at) - EXTRACT(EPOCH FROM (LAG(created_at)
OVER(PARTITION BY client_id
        ORDER BY created_at))) AS diff_sec,

        LAG(action_type) OVER(PARTITION BY client_id
                            ORDER BY created_at) AS prev_action_type,

        action_type != LAG(action_type) OVER(PARTITION BY client_id
        ORDER BY
created_at) AS act_is_diff

    FROM test_task
) AS tmp

WHERE act_is_diff is true AND diff_sec <10
ORDER BY client_id, created_at
```

Задача 3

Проанализировать данные (data.csv) и поискать в них интересные инсайты, построить графики.

Инструменты: jupyter notebook или excel + отчет в pdf

Общая картина по средней / медиане среди подавших заявление на кредит по наиболее важным, с точки зрения весов обученной catboost модели, параметрам

mean / median	Денежная сумма всех активных кредитов <code>all_active_creds_sum_all</code>	Сумма закрытых кредитов <code>all_closed_creds_sum_all</code>	Средняя сумма МФО кредитов, выданных за последние 3м <code>mfo_cred_mean_sum_3lm</code>	Сумма закрытых кредитов за последний год <code>all_closed_creds_sum_ly</code>	Количество запросов в МФО <code>mfo_inqs_count_month</code>	Количество кредитов, взятых за последний месяц <code>all_creds_count_lm</code>	Сумма лимитов кредитных карт, оформленных за <code>cred_sum_cc_ly</code>
approved = 1	230000 / 57300	270000 / 92000	8000 / 6500	72000 / 25000	2 / 1	2.4 / 2	24000 / 0
approved = 0	82000 / 7000	117000 / 15000	2500 / 0	10750 / 0	0.9 / 0	0.5 / 0	4200 / 0
diff	147000 / 50000	155000 / 77000	5500 / 6500	61000 / 25000	1 / 1	2 / 2	20500 / 0
bad = 0	240000 / 57350	288000 / 96500	8000 / 6500	75000 / 27000	2 / 1	2.4 / 2	24250 / 0
bad = 1	205000 / 57350	230000 / 82000	7900 / 6500	60000 / 20000	2.5 / 1	2.5 / 2	26000 / 0
diff	32500 / 0	59000 / 14500	100 / 0	16500 / 7000	-0.5 / 0	-0.1 / 0	-1750 / 0

Распределение по профессии

	1	2	3	4	5
count	2084	4881	4834	3570	7747
mean income	46000	43000	30000	42000	40000
approved	0.418	0.39	0.22	0.33	0.28
bad	0.203	0.26	0.253	0.252	0.32

...	a	b	c	d
12	32.677573	10.473591	cred_day_overdue_all_sum_all	Суммарное количество дней просрочки текущих ак...
20	0.019572	9.504745	region	Регион
5	5.852429	8.062407	all_active_creds_sum_all	Денежная сумма всех активных кредитов
2	4.490736	7.685468	all_closed_creds_sum_all	Сумма закрытых кредитов
11	12.000302	7.297244	all_creds_count_all	Общее количество кредитов
10	8.498030	6.268948	delay_more_sum_all	Количество просрочек более чем на 90 дней по всем
3	3.565815	5.840570	bank_inqs_count_quarter	Количество запросов на кредиты в банки
0	2.584674	5.780241	cred_sum_cc_all	Сумма кредитов по кредитным картам
18	0.076591	5.440926	work_code	Код профессии
8	8.432062	4.643692	all_closed_creds_sum_ly	Сумма закрытых кредитов за последний год
9	2.154291	4.218801	mfo_cred_mean_sum_3lm	Средняя сумма МФО кредитов, выданных за 3м
19	0.835774	3.517918	month_income	Месячный доход
4	1.387907	3.152882	cred_max_overdue_max_ly	Максимальная просрочка за год
14	2.467738	2.955134	mfo_closed_count_ly	Количество закрытых МФО кредитов, взятых за по...
17	3.086431	2.846169	all_creds_count_lm	Количество кредитов, взятых за последний месяц
15	4.203589	2.510604	cred_sum_overdue_cc_all	Сумма просрочек по кредитным картам
7	3.253168	2.480508	cred_sum_cc_ly	Сумма лимитов кредитных карт, оформленных за п...
1	0.010422	2.352062	mfo_inqs_count_month	Количество запросов на кредиты в другие в МФО
6	0.220919	2.122665	mfo_last_days_all	Количество дней с последнего займа в МФО
16	3.704051	1.961531	count_overdue_all_3lm	Количество кредитов на просрочке, взятых за la...
13	0.477925	0.883893	cred_max_overdue_max_3lm	Максимальная сумма просроченной задолженности,...

a - коэфы при фичах у нормально обученного классификатора
 b - коэфы у намеренно переобученного классификатора
 Обучены на классификацию approved

...	a	b	c	d
20	7.917794	12.963582	region	Регион
5	5.523752	9.145701	all_active_creds_sum_all	Денежная сумма всех активных кредитов
2	4.242894	8.326008	all_closed_creds_sum_all	Сумма закрытых кредитов
18	11.555119	8.057935	work_code	Код профессии
9	4.628952	7.577755	mfo_cred_mean_sum_3lm	Средняя сумма МФО кредитов, выданных за 3м
3	2.461938	7.226434	bank_inqs_count_quarter	Количество запросов на кредиты в банки
0	2.257415	6.579274	cred_sum_cc_all	Сумма кредитов по кредитным картам
8	9.616167	6.082713	all_closed_creds_sum_ly	Сумма закрытых кредитов за последний год
11	2.622224	5.222219	all_creds_count_all	Общее количество кредитов
1	11.330668	4.310161	mfo_inqs_count_month	Количество запросов на кредиты в другие в МФО
14	5.991318	4.250507	mfo_closed_count_ly	Количество закрытых МФО кредитов, взятых за по...
7	5.431810	3.857108	cred_sum_cc_ly	Сумма лимитов кредитных карт, оформленных за п...
4	5.418327	3.641775	cred_max_overdue_max_ly	Максимальная просрочка за год
17	2.305597	3.532451	all_creds_count_lm	Количество кредитов, взятых за последний месяц
19	2.800698	3.495925	month_income	Месячный доход
10	1.141860	1.909360	delay_more_sum_all	Количество просрочек более чем на 90 дней по всем
13	5.229233	1.343399	cred_max_overdue_max_3lm	Максимальная сумма просроченной задолженности,...
6	1.846753	1.020242	mfo_last_days_all	Количество дней с последнего займа в МФО
16	5.505044	0.662985	count_overdue_all_3lm	Количество кредитов на просрочке, взятых за la...
15	1.229385	0.587205	cred_sum_overdue_cc_all	Сумма просрочек по кредитным картам
12	0.943053	0.207261	cred_day_overdue_all_sum_all	Суммарное количество дней просрочки текущих ак...

a - коэфф при фицах у нормально обученного классификатора
b - коэфф у намеренно переобученного классификатора
Обучены на классификацию bad

Интересные наблюдения

Неочевидные моменты с матрицы корреляции

Козф корреляции Коррелирующие параметры

0.4

1. Сумма закрытых кредитов за последний год + Количество кредитов, взятых за последний месяц

0.5

1. Сумма закрытых кредитов + Сумма задолженности по всем кредитам

2. Общее количество кредитов + Количество кредитов, взятых за последний месяц

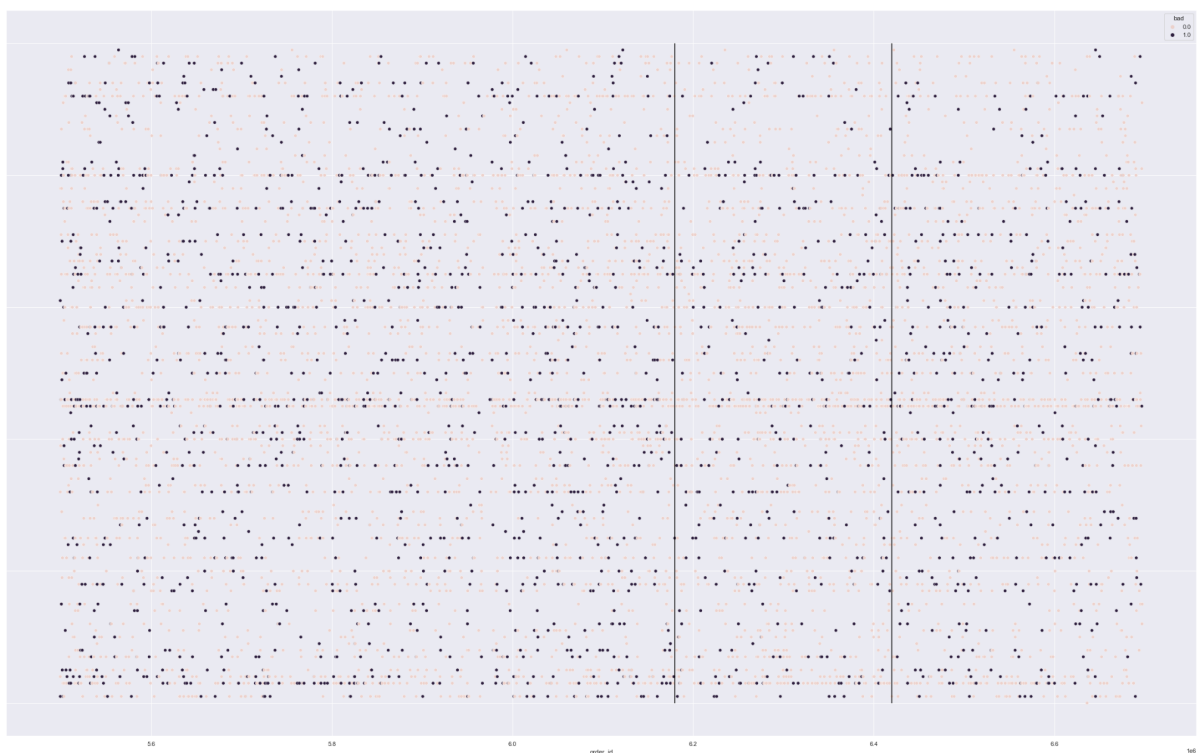
0.7

1. количество запросов на кредиты в другие в МФО + Количество кредитов, взятых за последний месяц

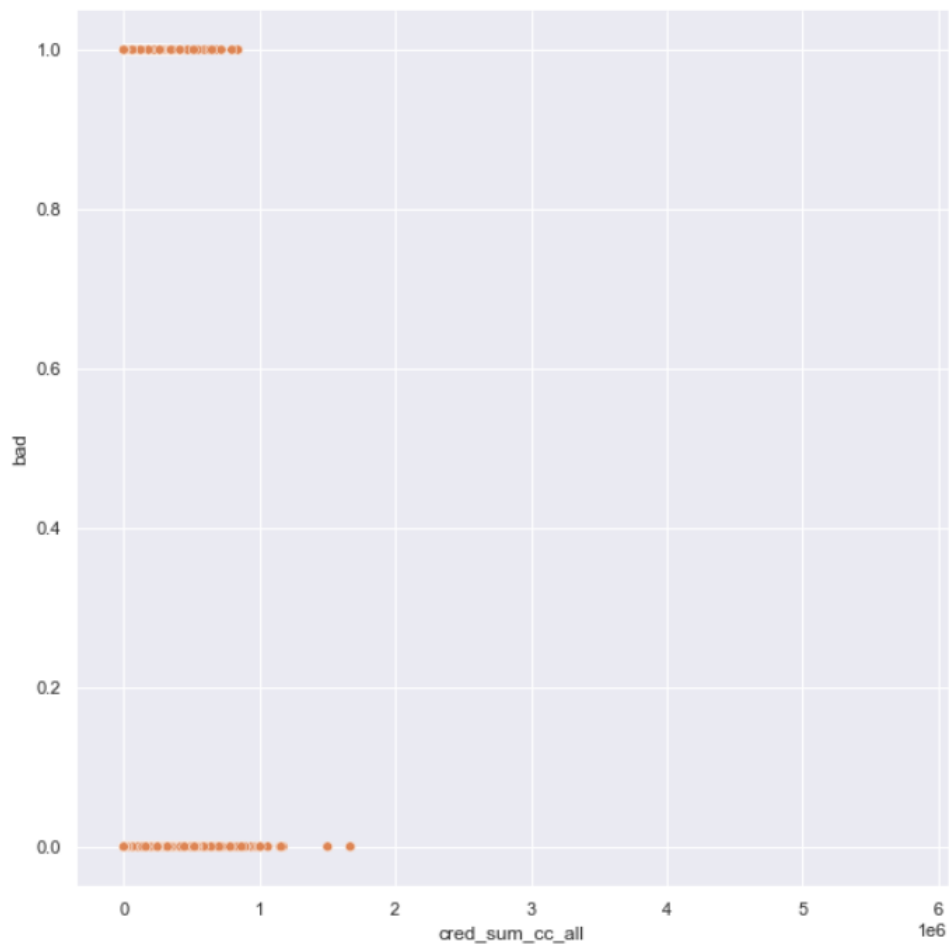
0.9

1. Общее количество кредитов + Количество закрытых МФО кредитов, взятых за последний год

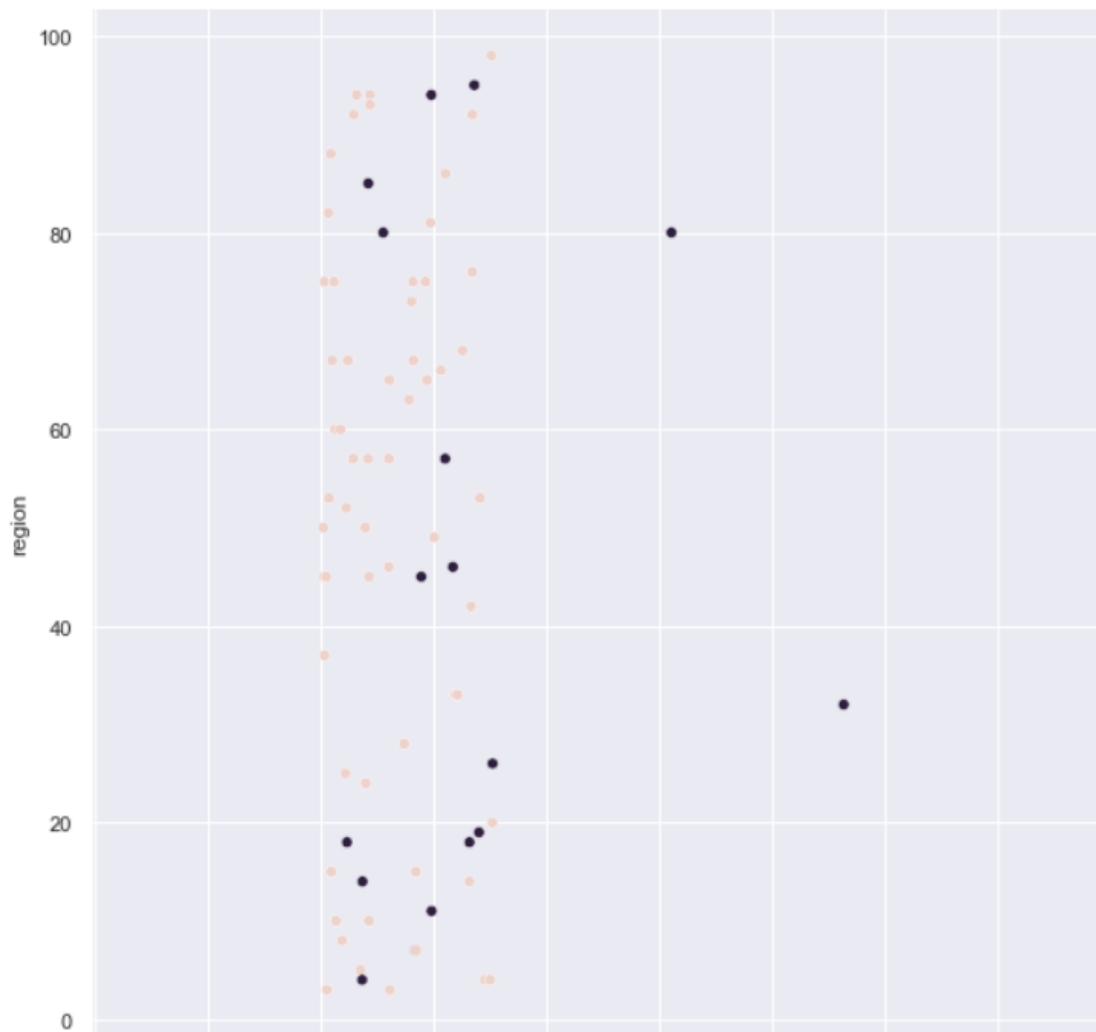
1. Большое количество одобрений в районе id 618**** - 642****, при этом это не повлияло на возвраты и визуально количество просрочек ничем не выделяется



2. У тех, у кого сумма кредитов по кредитным картам свыше ~900т. без просрочек



3. Доход меньше 10к у id 550**** - 568****
work_code 3 (один с 2 кодом)
month income 0 (один с 2к и один с 8к)
среди них отличился 60-79 регионы с минимумом бэдов



76/322 approved (0.23, стандартно)
 17/76 bad (0,22/0,25 обычного, даже лучше справились)

при $\text{count_overdue_all_3lm} > 1$ (при наличии какой-либо просрочки) – 0,4
 вероятность просрочить (и/или этот) кредит

$\text{cred_sum_overdue_cc_all}$ Сумма просрочек по кредитным картам При
 больше 0 вероятность получить одобрение падает с 0.337804 до 0.147443,
 при этом вероятность отдать кредит не меняется 0.269742 к 0.27512 (кто-то
 умудрился просрочить 1р, забавно)

$\text{mfo_cred_mean_sum_3lm} == 0$ – 0,12 вероятность получить кредит против
 $\text{mfo_cred_mean_sum_3lm} > 0$ – 0,5 вероятности
 При этом 0.25 и 0.27 не вернуть соответственно

delay_more_sum_all (Количество просрочек более чем на 90 дней по всем кредитам) и all_creds_count_all (Общее количество кредитов) **поглядеть**

Есть 173 человека с закрытыми 1483р за год, из них одобрено 42, 11 бэдов (стандартные вероятности)