

Исследование видов кредитования, актуальных для клиентов

Команда #3

- Первушин Владислав Олегович
- Черненко Игорь Сергеевич
- Грищук Арсений Васильевич

Описание датасета

Датасет состоит из заявок клиентов на кредитование.

Описание основных переменных:

- monthly_income_amt – месячный доход клиента (у.е.)
- req_amt – запрошенная сумма кредитования (у.е.)
- req_term – запрошенный срок кредитования (мес.)
- gender_code – пол клиента
- car_flg – наличие машины у клиента
- car_price – цена машины клиента (у.е.)
- realty_flg – наличие недвижимости у клиента
- realty_price – цена недвижимости клиента (у.е.)
- agreement_product_code – код согласованного продукта
- success_activate_flg – успешность активации кредита

Исследовательский вопрос

Какие характеристики клиента определяют, какой вид кредитования для него наиболее актуален?

Обработка выбросов

Оставляем:

- По возрасту (18 ~ 60)
- По месячному доходу (1 ~ 43 у.е.)
- По стоимости недвижимости (< 2000 у.е.)
- По стоимости машины (< 480 у.е.)
- Только одобренные и выданные кредиты

Предварительный анализ

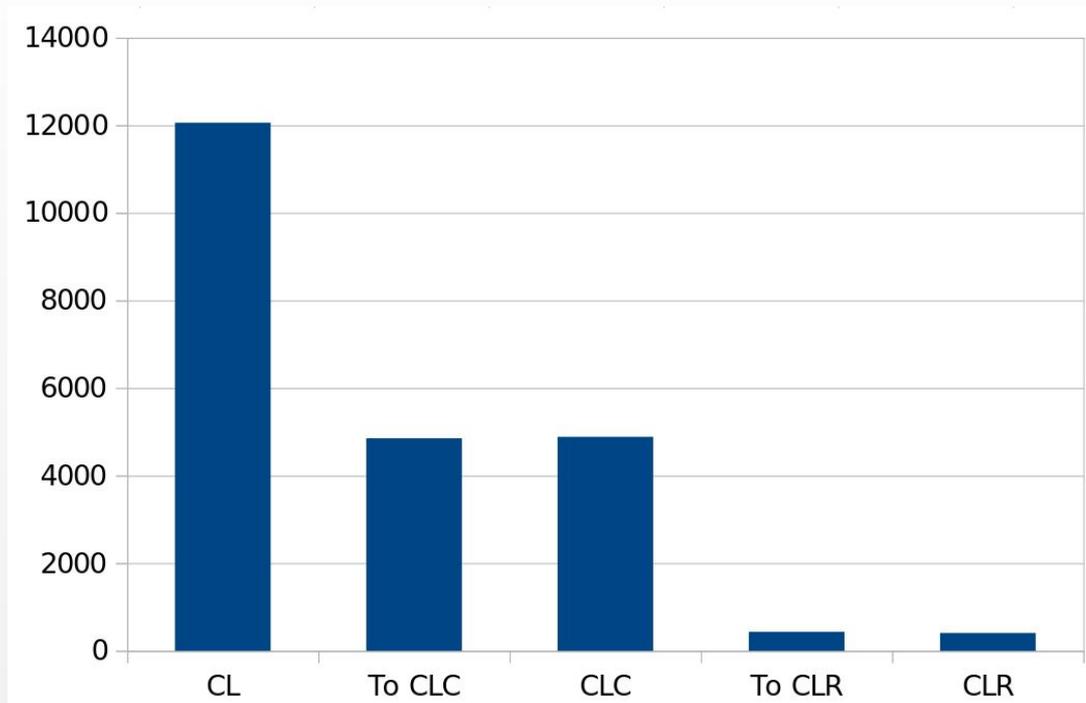


График распределения видов кредитования

CL - кредит наличными

CLR - кредит с залогом под
недвижимость

CLC - кредит с залогом под
авто

To CLR - upsale под
недвижимость

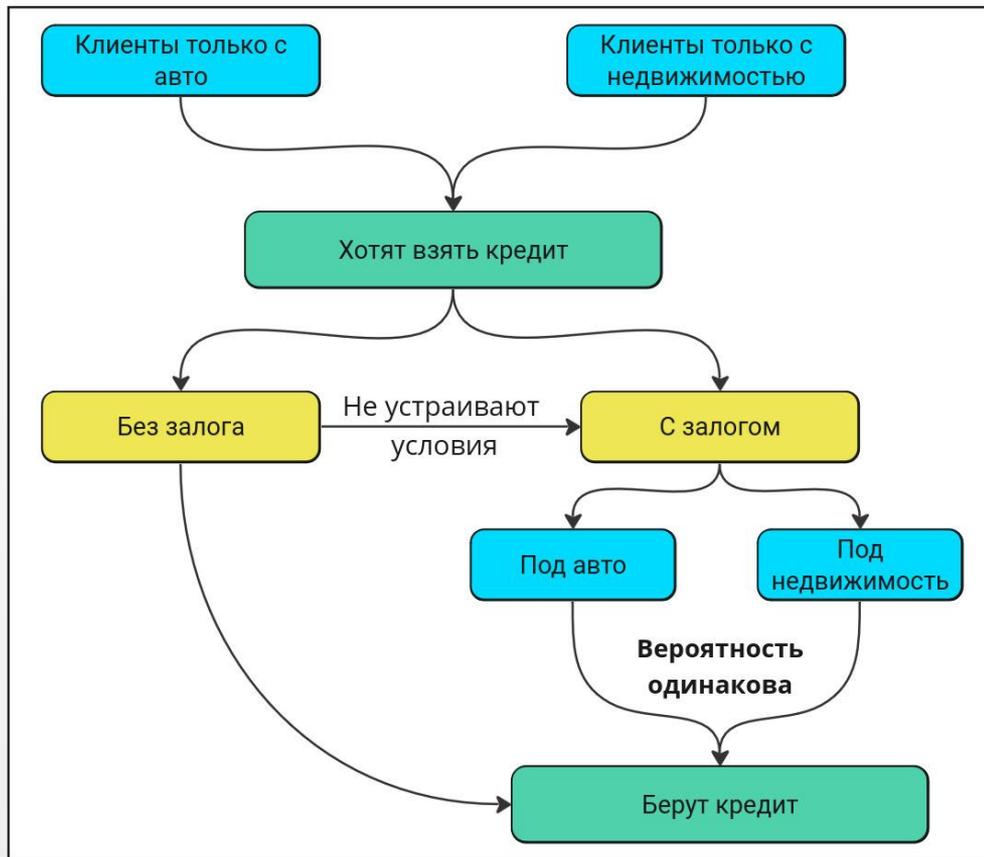
To CLC - upsale под авто

Гипотеза

Нулевая гипотеза: если у клиента есть только один вид имущества (автомобиль или недвижимость), вероятность того, что он оставит свою собственность в залог, одинакова

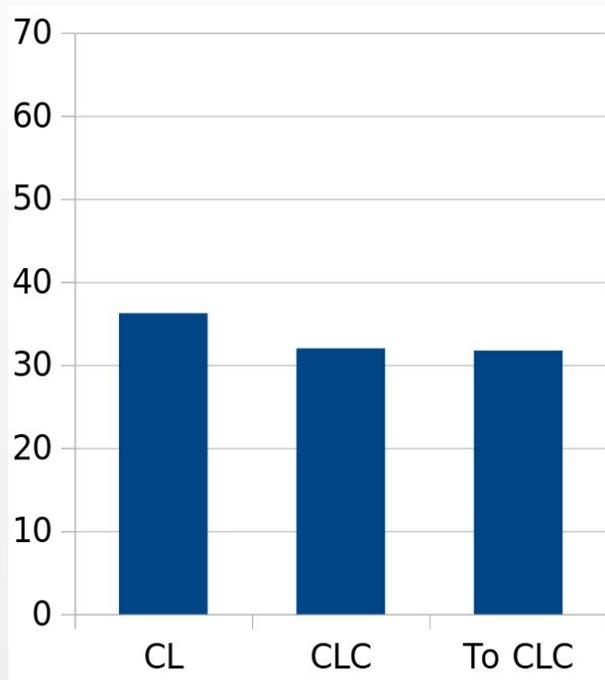
Альтернативная гипотеза: Если у клиента есть только один вид имущества, вероятность того, что он оставит свою собственность в залог, отличается.

Механизм

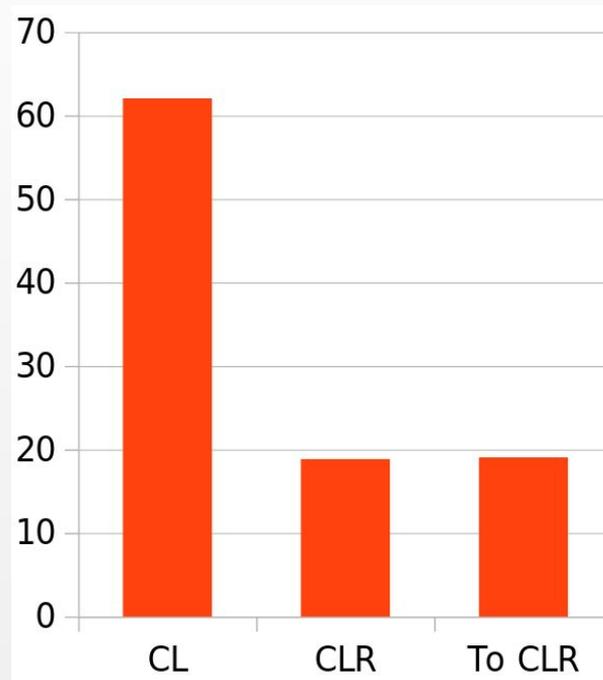


Результаты исследования

Графики распределения видов кредитования среди владельцев одного вида имущества



Владельцы авто



Владельцы недвижимости

CL - кредит наличными

CLR - кредит с залогом под недвижимость

CLC - кредит с залогом под авто

To CLR - upsale под недвижимость

To CLC - upsale под авто

Математическая модель

Цель: оценить отличие средних значений в двух группах (клиенты с недвижимостью / машинами)

Метод: *t*-тест

Сравнивали две группы наблюдений (владельцы только автомобилей, владельцы только недвижимости).

Устойчивость модели

Была проверена устойчивость модели с помощью исключения из данных клиентов женского пола.

Клиенты только с авто

CL	34%
CLC	33%
To CLC	33%

Клиенты только с недвижимостью

CL	58%
CLR	21%
To CLR	21%

Вывод

Подставив значения в математическую модель, получаем $p_value \approx 0.009$, которое в свою очередь меньше 0.05 , а также оценив график распределения, получаем, что нулевая гипотеза неверна. Соответственно, верна

альтернативная гипотеза:

Если у клиента есть только один вид имущества, вероятность того, что он оставит свою собственность в залог, отличается.

Применение результатов исследования

Для повышения прибыли компании стоит чаще предлагать владельцам автомобилей кредит с залогом под автомобиль, а владельцам недвижимости – кредит без залога.

Ограничения исследования

- Достоверность данных не гарантирована, т.к. данные заполняют сами пользователи
- Актуальность данных не гарантирована, т.к. компания не всегда может своевременно обновлять базу данных

