

BestHack 2021.Data

Команда RedLink



Михайлов
Баир



Сурков
Вячеслав



Гапонов
Максим



Тищенко
Евгений



Задание



ДАНО:

Aggregates – месячные отчеты о клиенте

Transaction – месячные отчеты о тратах клиента по категориям

ЦЕЛЬ:

- Провести разведочный анализ
- Придумать метрику качества
- Построить модель склонности к потребительскому кредиту

Склонность



ИДЕЯ

Для каждого клиента выбрать отчеты за последовательные 12 месяцев и разбить на 2 периода по 6.

Прошлое и настоящее

Первый период. В нем обобщаем информацию о клиенте, считаем её за известное.

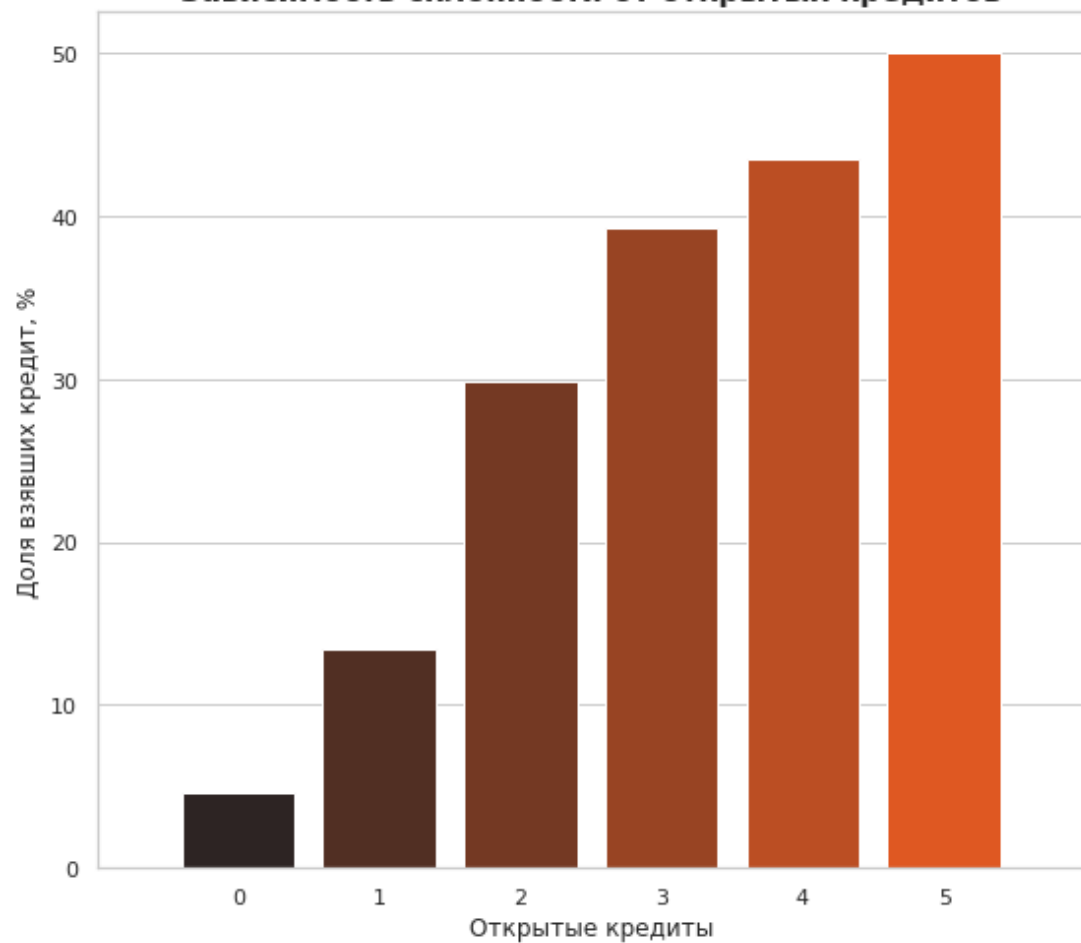
Будущее

Второй период. По нему строится целевая переменная – взял ли клиент кредит. Если взял, то высокая склонность.

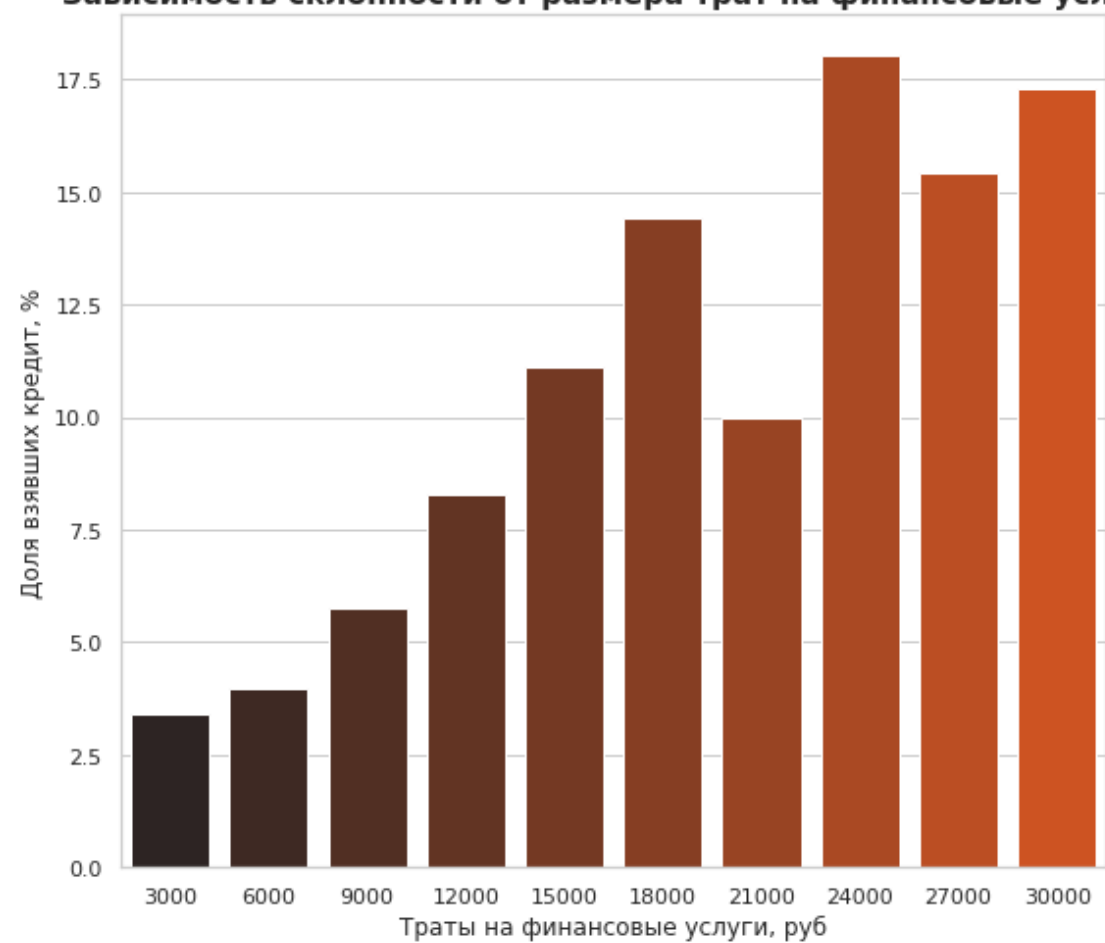
Разведочный анализ



Зависимость склонности от открытых кредитов



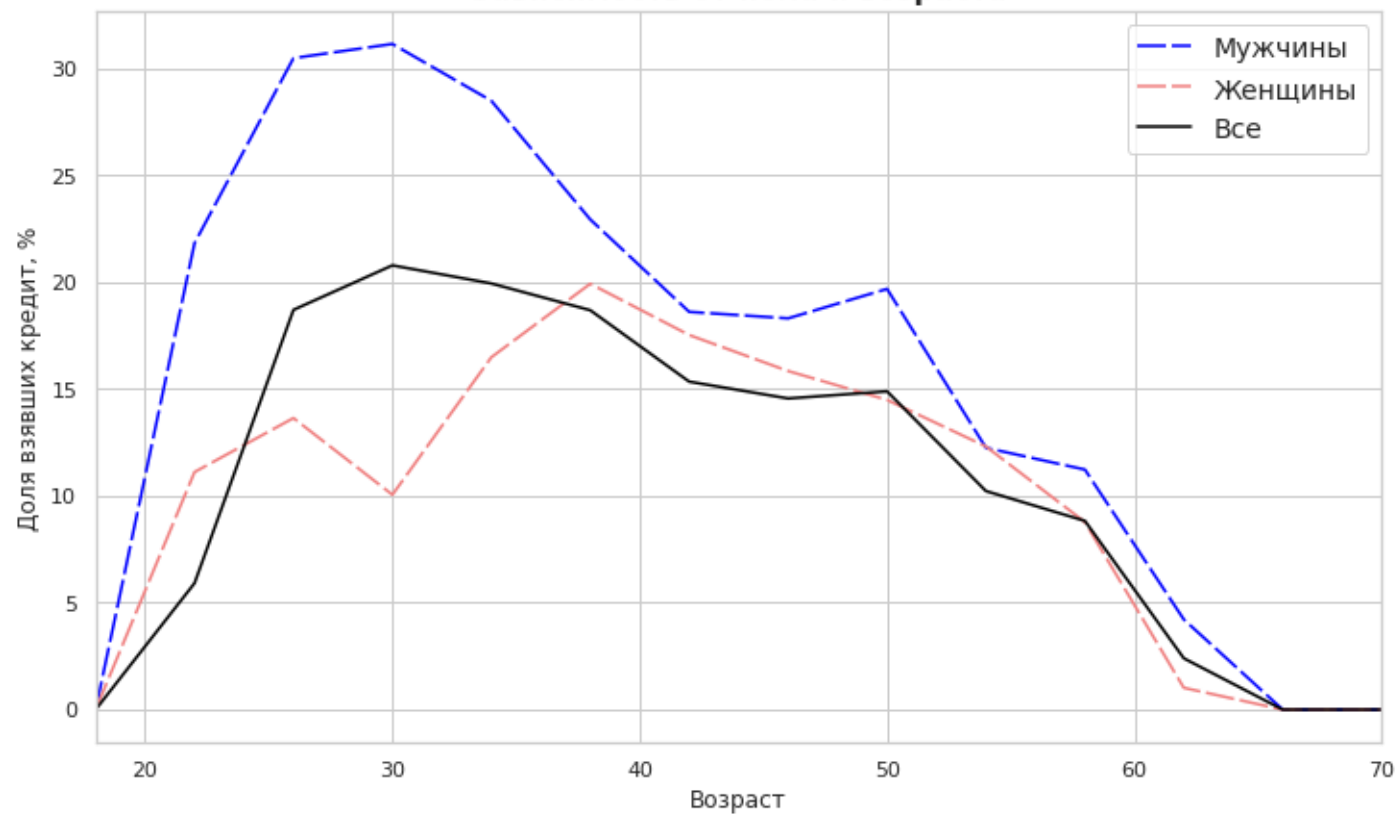
Зависимость склонности от размера трат на финансовые услуги



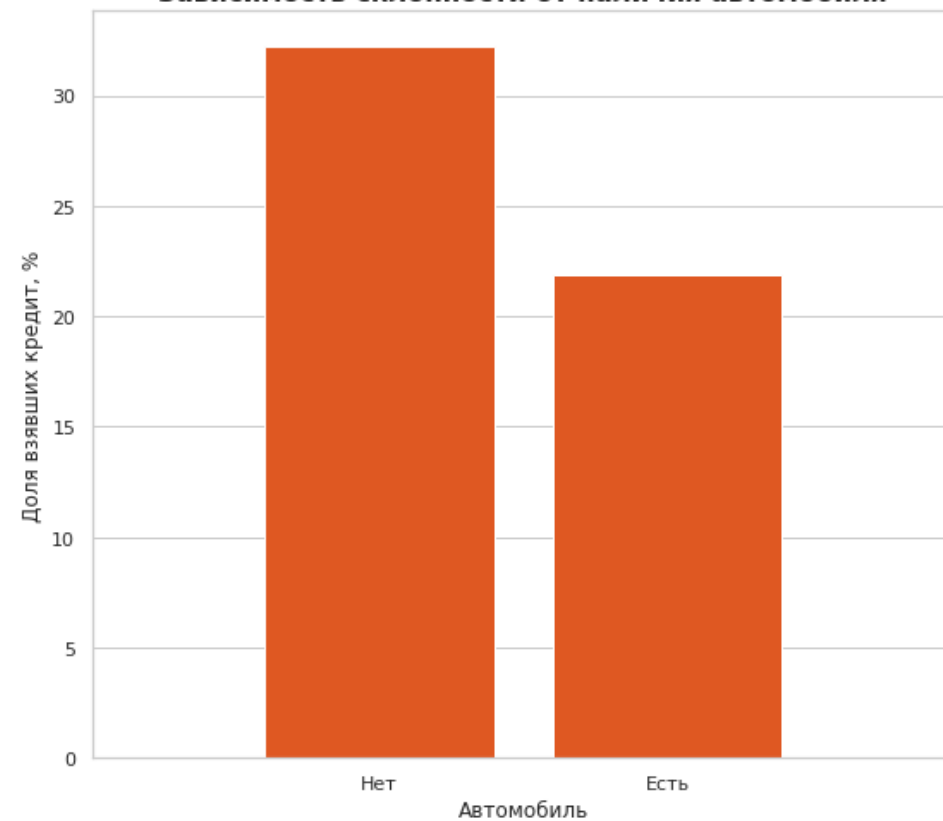
Разведочный анализ



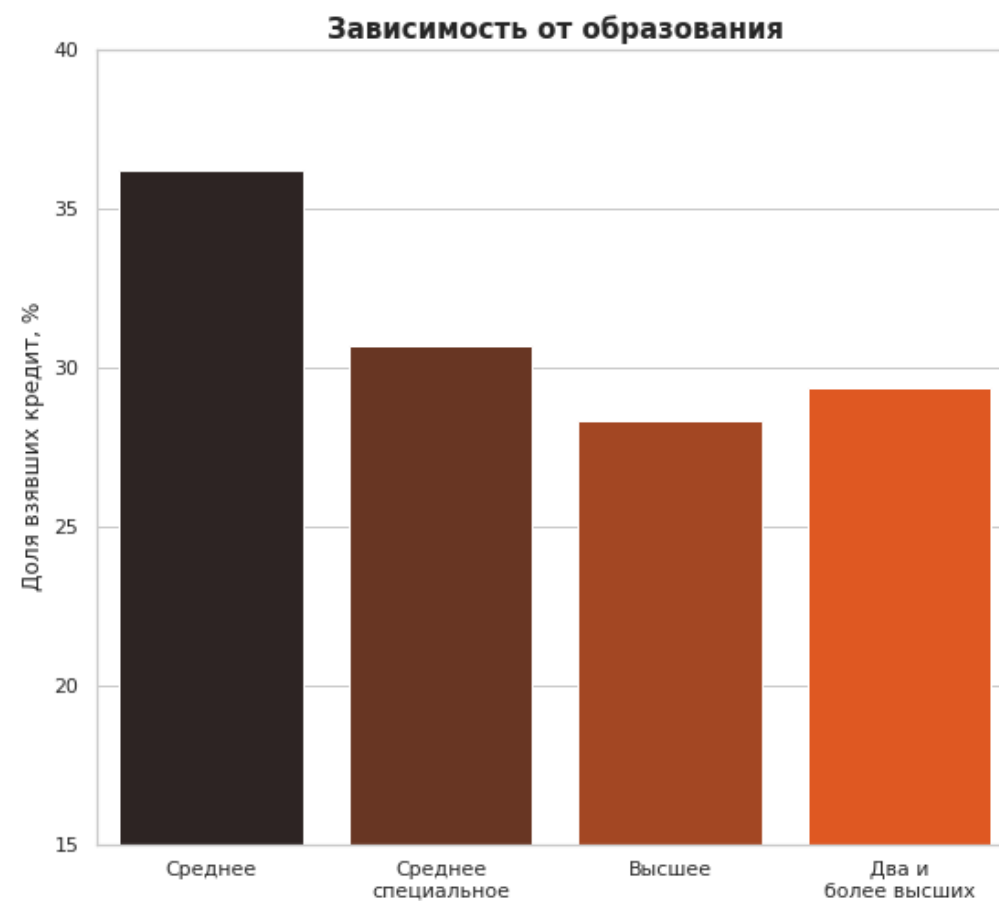
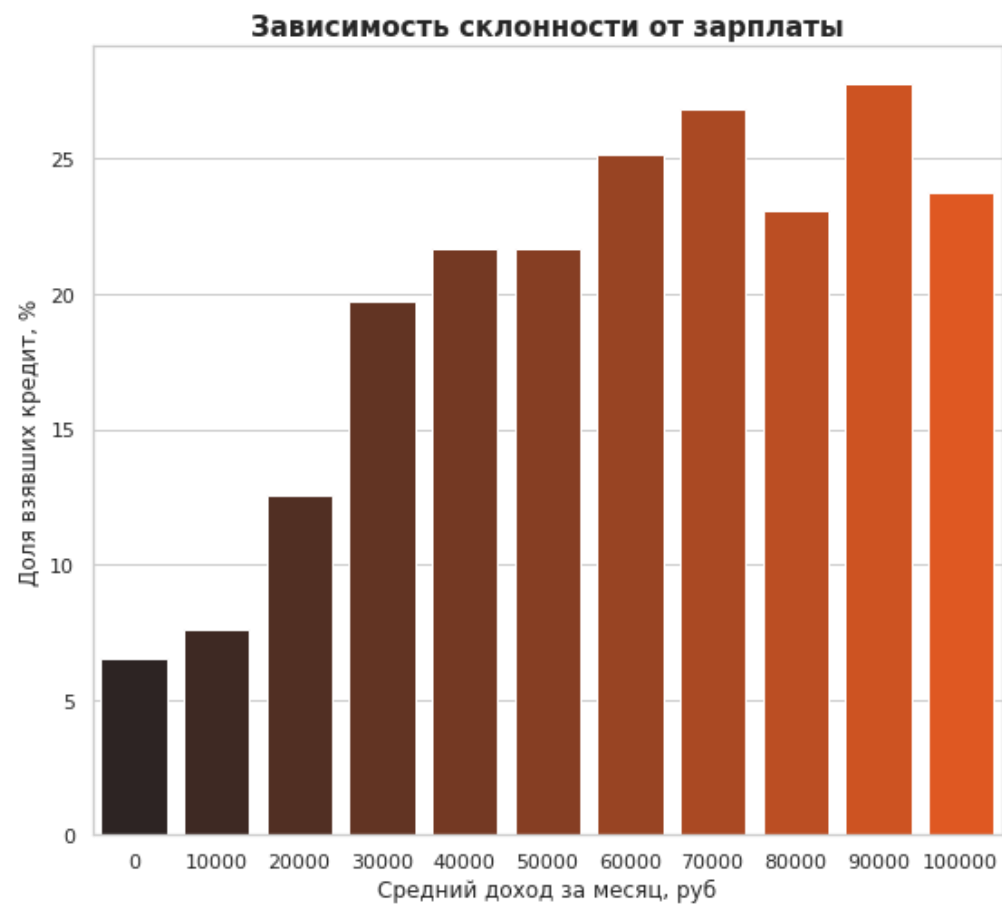
Зависимость от пола и возраста



Зависимость склонности от наличия автомобиля



Разведочный анализ



Метрика



Модель возвращает склонность как число от 0 до 1.
Как оценить ошибку?



Важно не упустить потенциального клиента, а показать рекламу незаинтересованному не настолько критично!



Модель предсказала склонность y_{pred} , а правильный ответ – y_{true} .
Метрика сильнее штрафует за $|y_{true} - y_{pred}|$, если $y_{test} = 1$, и слабее, если $y_{true} = 0$.

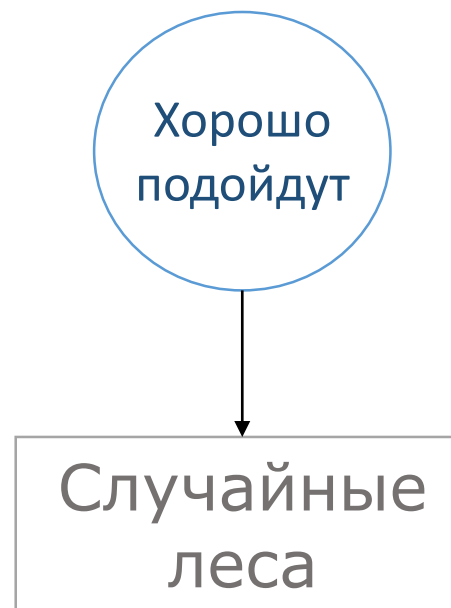


Модель



Особенности данных:

- Клиентов немного (~10000 человек)
- Много пропусков
- Большое количество атрибутов
- Данные разнородны
- Зависимости не линейны



Модель



Модель	Ошибка по метрике
Вывести 0	0.39
Вывести 1	0.33
Вывести случайно 0/1	0.36
Случайный лес (RandomForestRegressor)	0.14