



Кредиты и где они обитают

Команда ENTROPIX

Михаил Фадин, Антон Ершов, Тимур Инамов,
Аким Качалиев, Злата Кобалиа, Таисия Квон



19

Количественные
переменные

4

Бинарные
переменные

6

Категориальные
переменные

Структура данных

**Тинькофф
Банк**

Источник
данных

Клиент

Единица
наблюдения



Анализ качества данных

Анализ заполнения

≈ 88%

ячеек заполнено

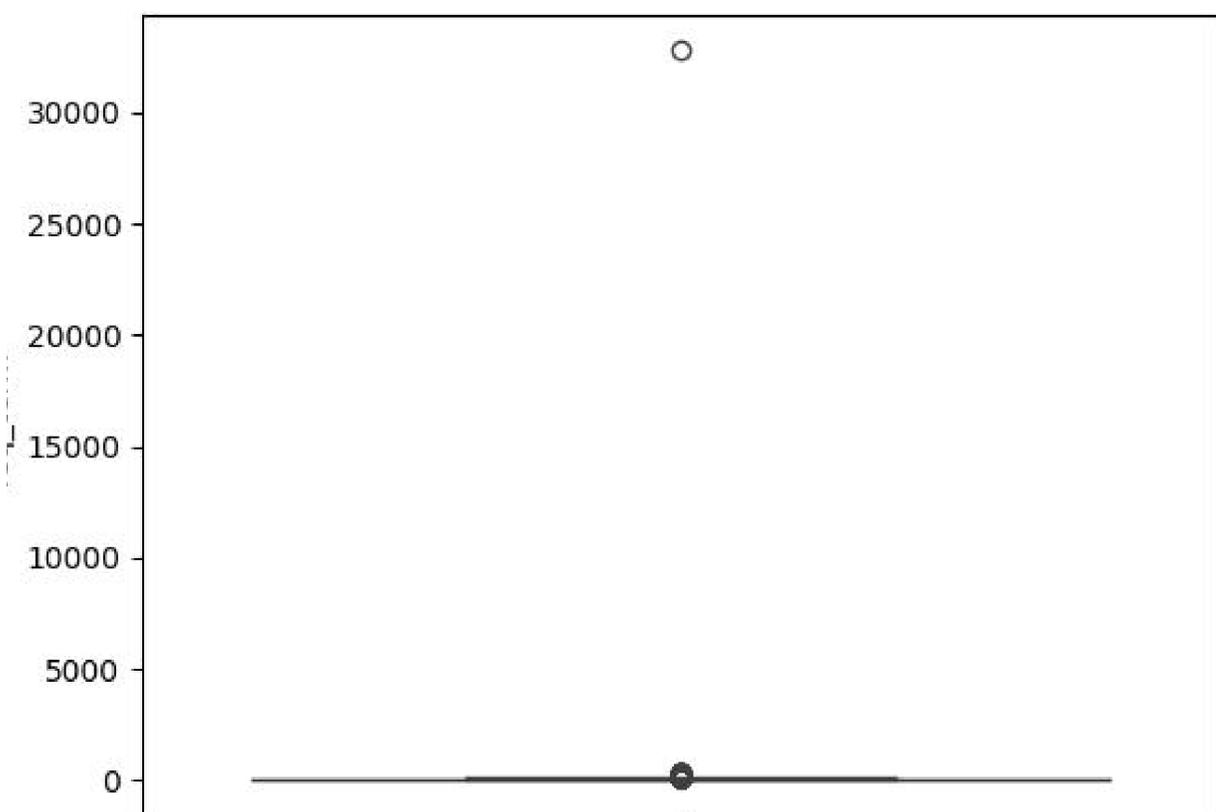
При проверке
распределений
переменных
выяснилось,
что все
переменные
распределены
ненормально

Анализ дубликатов

Данные
разбиты по
заявкам, а не
по клиентам

Очистка выбросов

Было

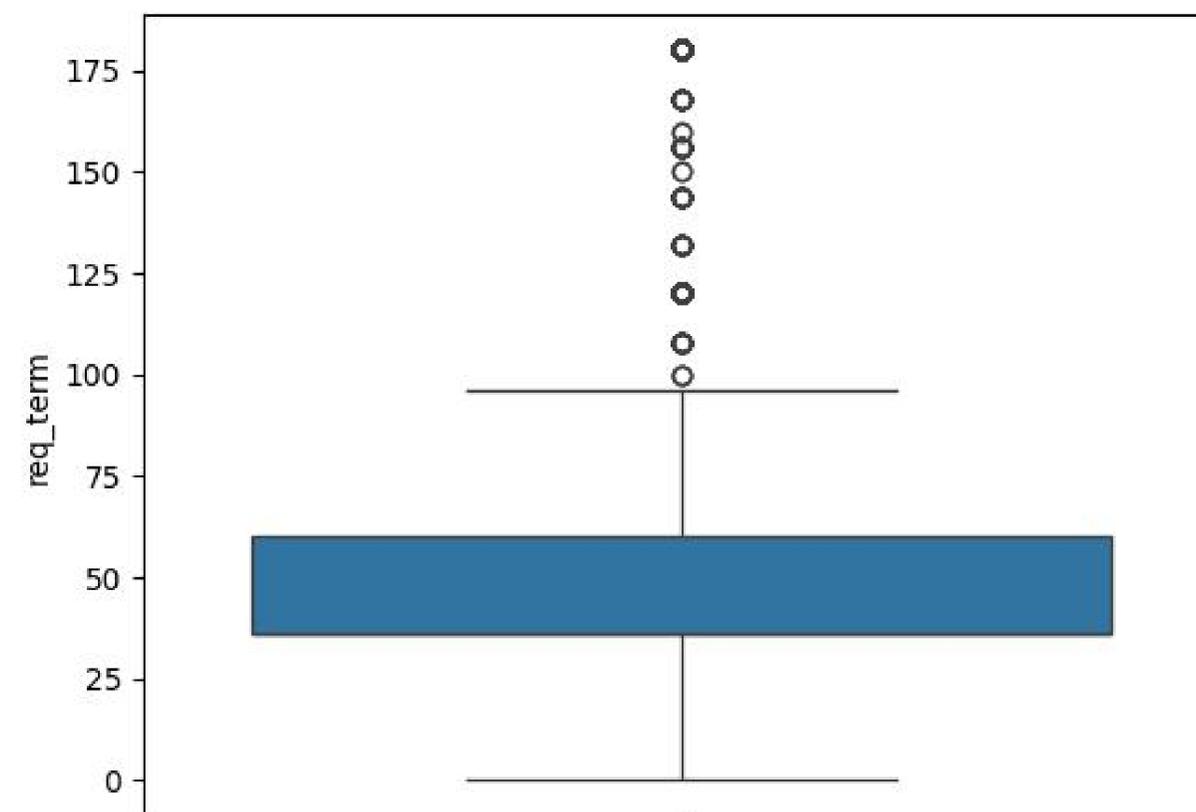


Записей: 225.954

Удалено
25.924 записи



Стало

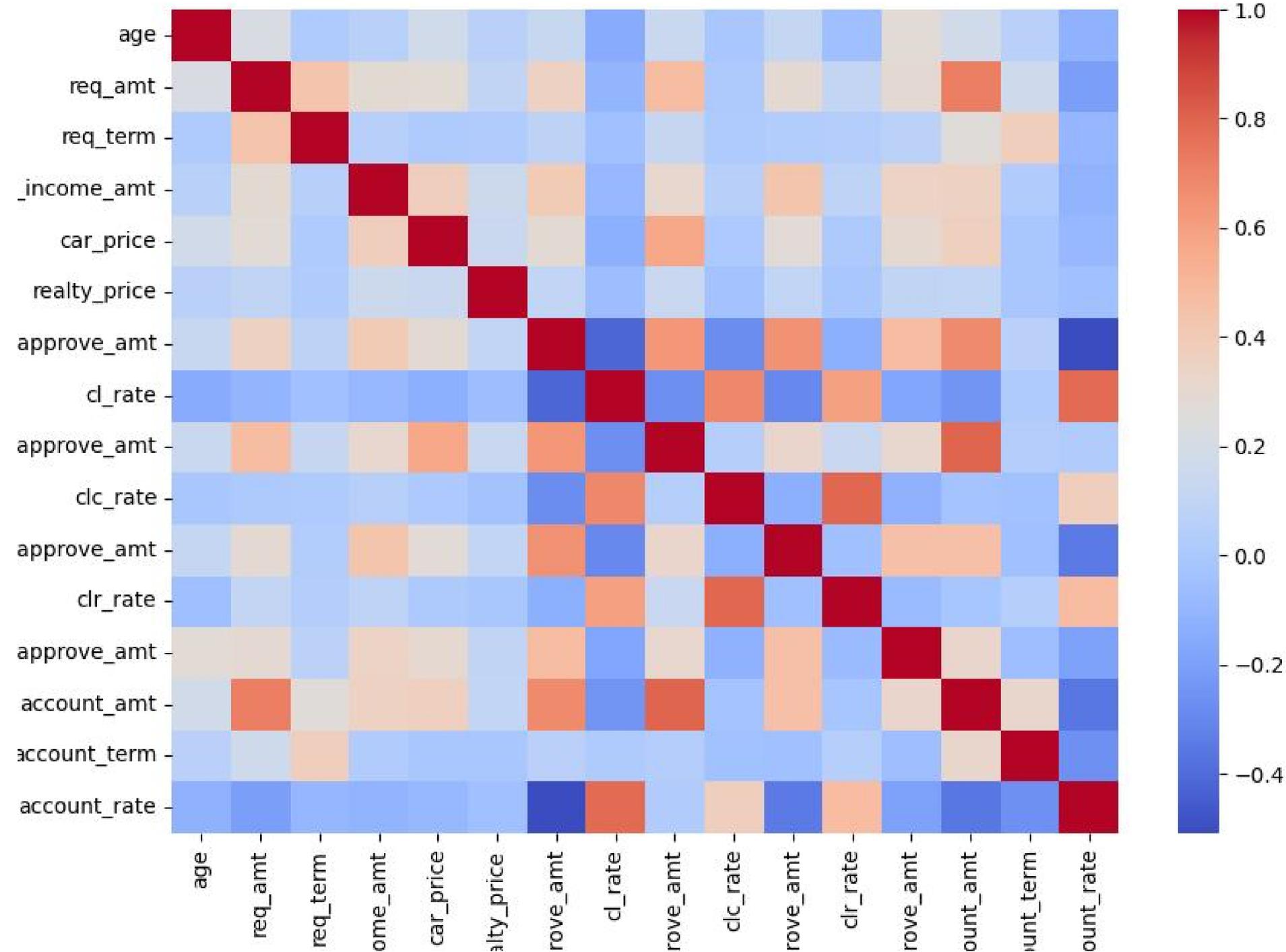


Записей: 200.030

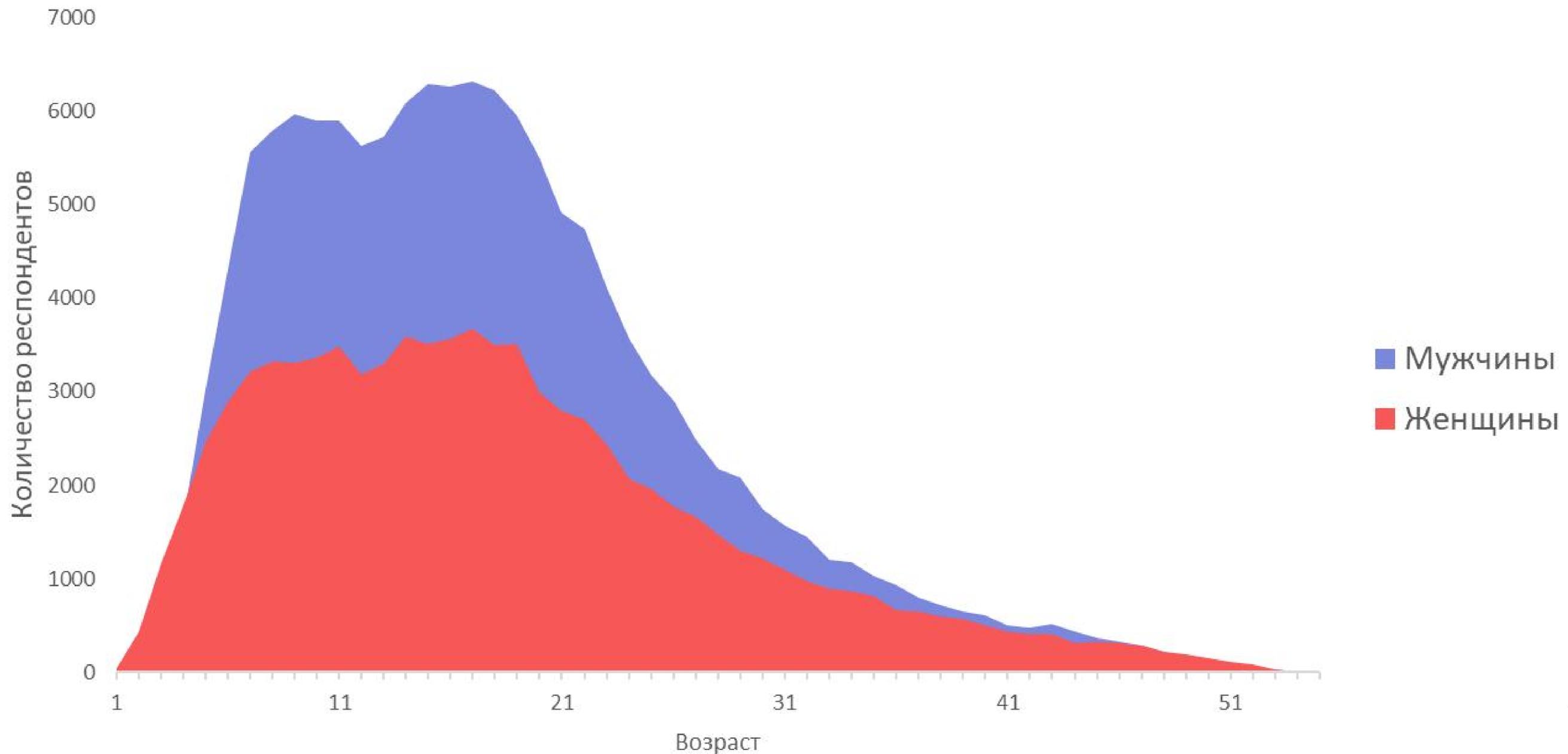
**Мы провели
анализ
распределений с
Shapiro test p-
value < 0.05**

**По итогу все
переменные
распределены не
нормально.**

Матрица корреляций количественных данных

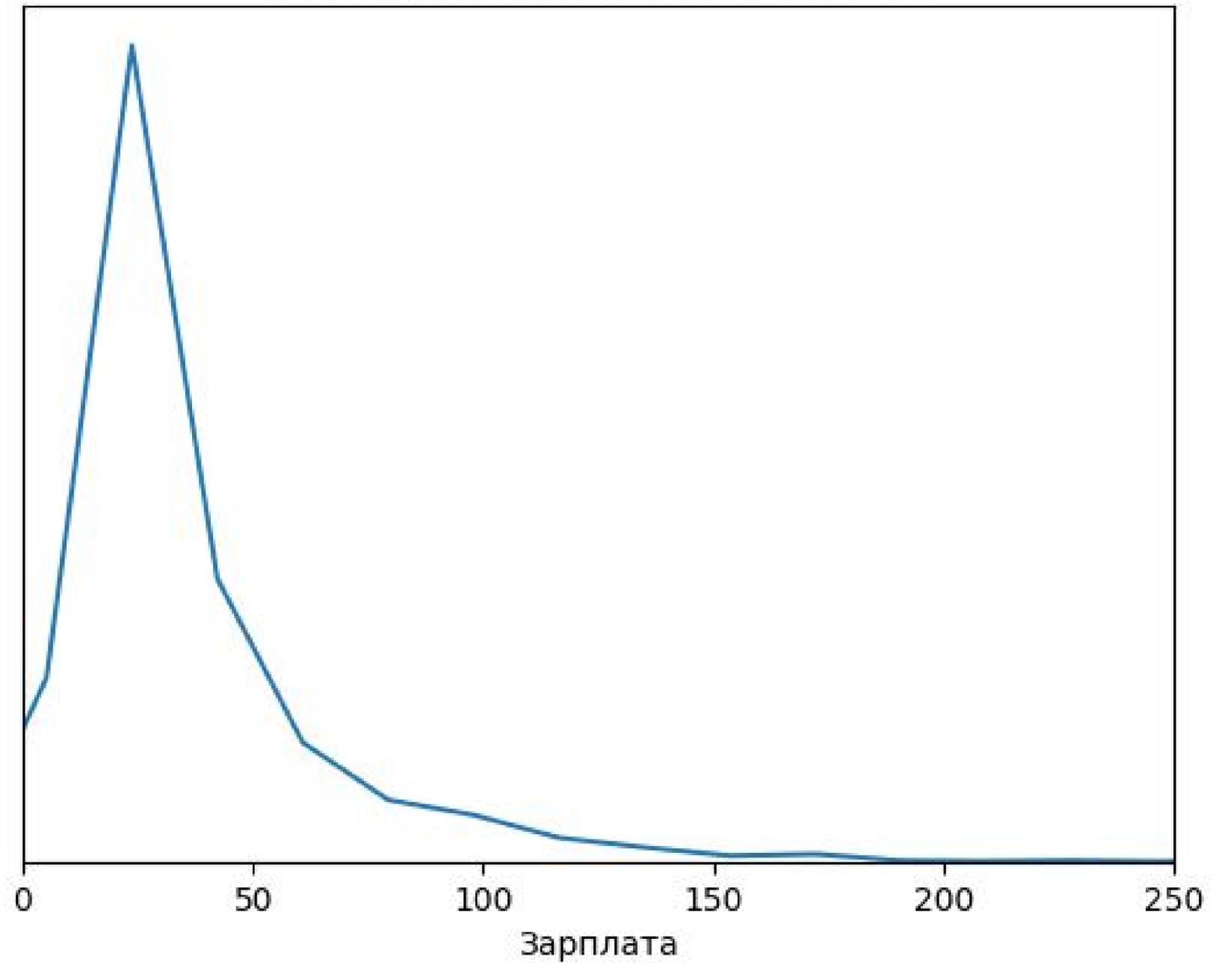


Поло-возрастное распределение



Распределение зарплат

График распределения зарплат



Исследовательский вопрос:

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

Мотивация

Повышение доли клиентов, которые соглашаются с условиям кредитования на лучших для банка условиях → увеличение прибыли



Гипотеза

**Наличие кредитной карты
влияет на успешный
UpSell**



Преобразование датасета

**Объединили заявки с
одинаковыми ID в строку**

**Добавили
столбец result**

**Sell
UpSell
Unsuccess**

**Добавили
столбец has_to**

**True
False**



Механизм ГИПОТЕЗЫ

Пользование кредитной картой



Человек кредитно грамотен



Сможет закрыть кредит в срок



Альтернативный механизм гипотезы

Люди привыкли жить в
кредит с кредитной картой

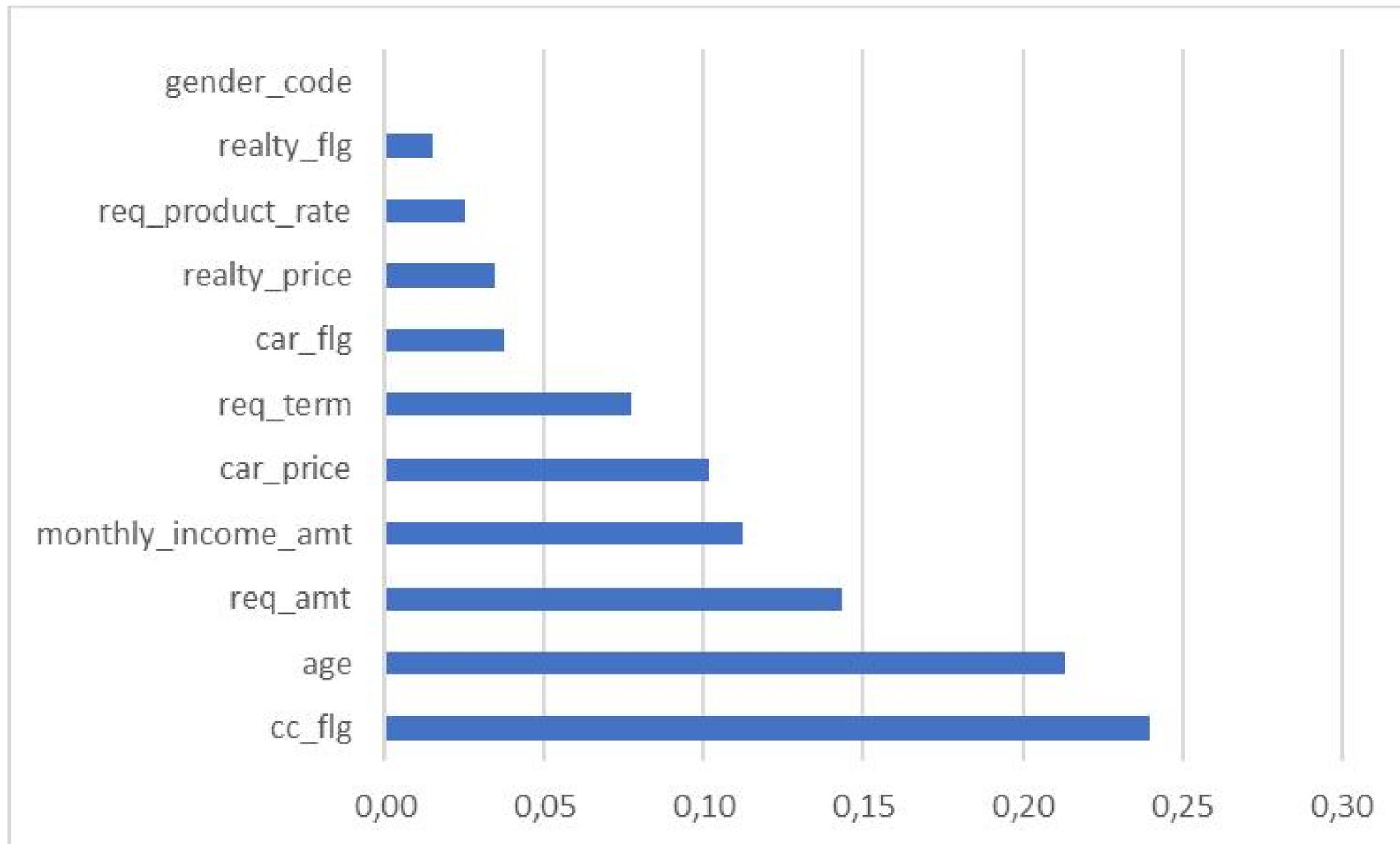


Недооценивают риск
потерять залог



Охотнее соглашаются
оставить залог

FOREST



Средний F1-score кросс-валидации: 0.61
F1-score на тестовой выборке: 0.62



Выводы

01

Гипотеза №1 подтвердилась.

02

Проверка на устойчивость

Интерпретация результатов

Основываясь на данном исследовании, можно сделать вывод, что наличие кредитной карты является основным критерием, который влияет на шанс успешного UpSell.

Ограничения

01

Утилизация последнего этапа воронки

В нашем датасете данные взяты только по РФ, поэтому наши выводы в других странах могут не работать.

02

Нет других Банков

В нашем датасете данные взяты только у банка Тинькофф, поэтому наши выводы могут не работать в других банках.

03

Только клиенты

В нашем датасете не включены те люди, которые ушли без продукта.

04

Фокусировка на UpSell

В нашем исследовании мы сфокусировались на возможности продажи UpSell продуктов.

Policy implication

01

Целевая аудитория

На основе датасета мы можем сделать выводы о текущей целевой аудитории. Банку следует сохранять упор на владельцах КК среднего возраста

02

Студенческие кредиты

В процессе предварительного анализа была замечена низкая активация одобренных кредитов среди людей до 25 лет ($\leq 50\%$)



Вопросы

Контакты

ENTROPIX