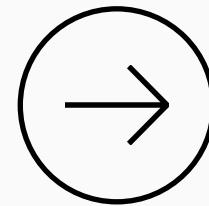


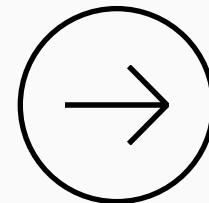
# **Итоговый проект: Прогнозирование оттока клиентов бизнес-сегмента Альфа-Банка**

Как ML-модель и анализ поведения помогают удерживать пользователей

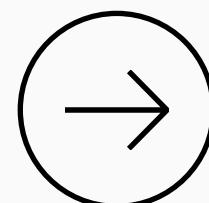
# Что имеем и что хотим



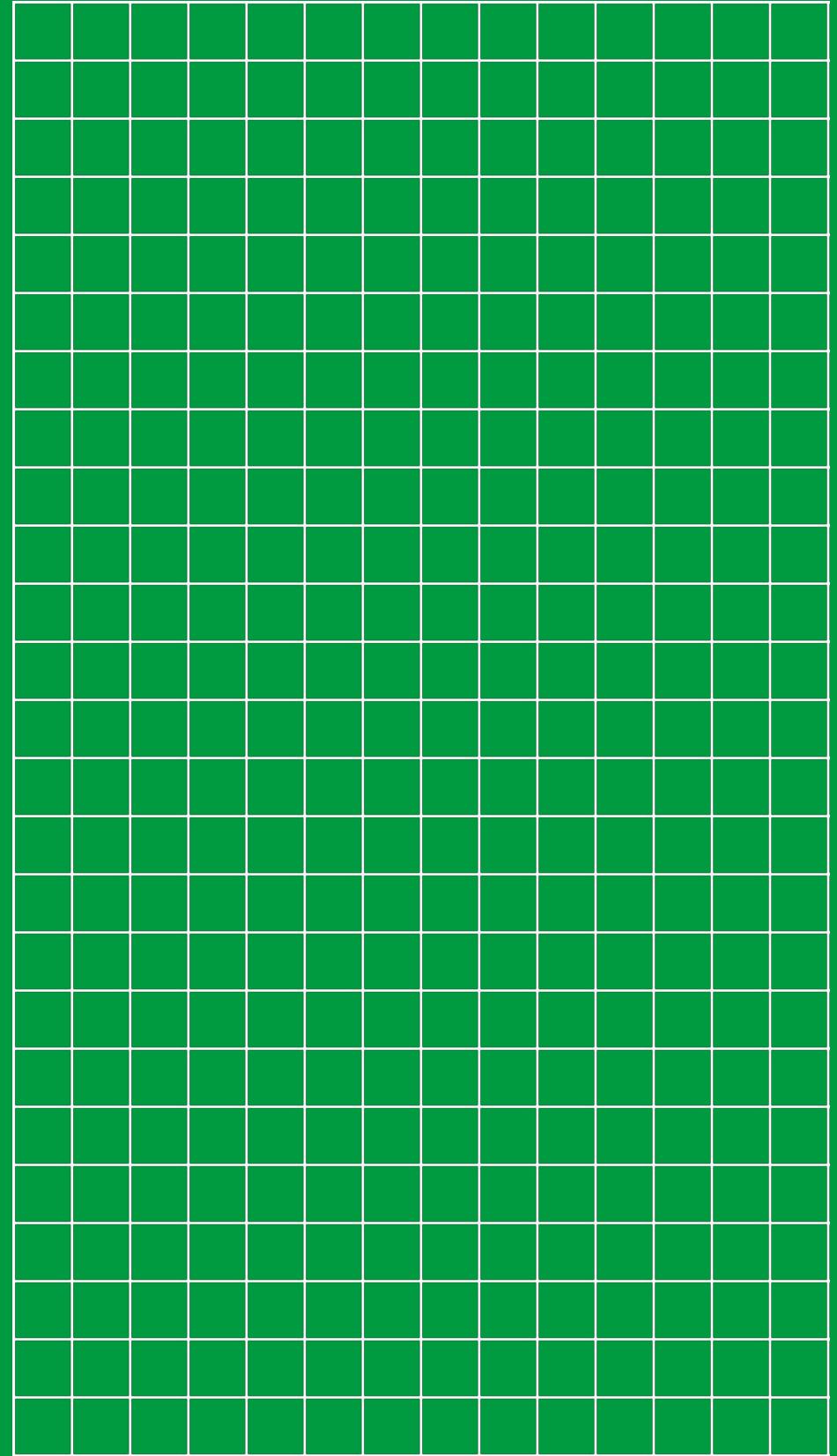
124 920 клиентов, 89 признаков.



Глобальный уровень оттока: 7.1%

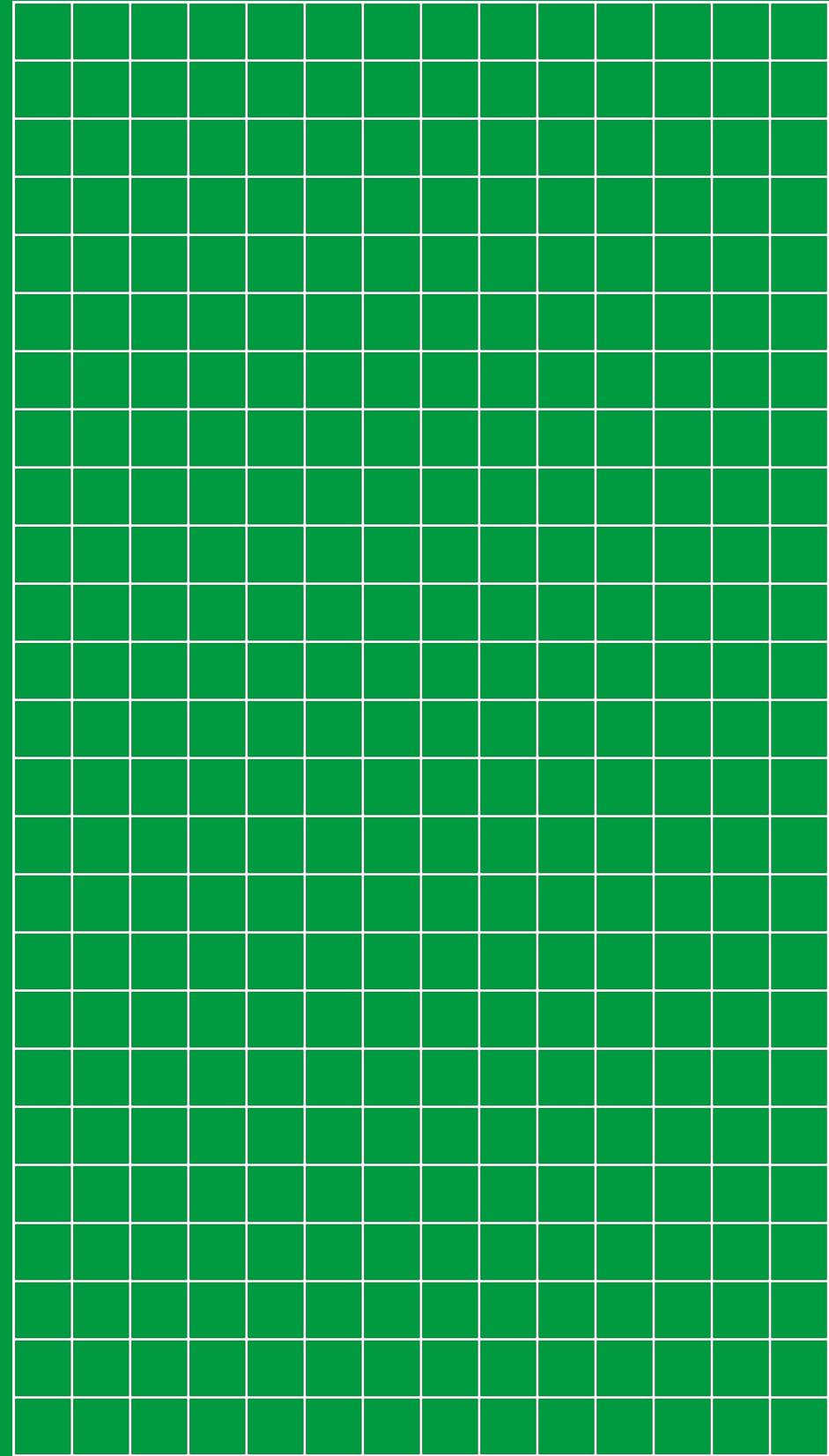


Построить ML-модель для раннего выявления клиентов с высоким риском оттока и определить ключевые факторы риска.



---

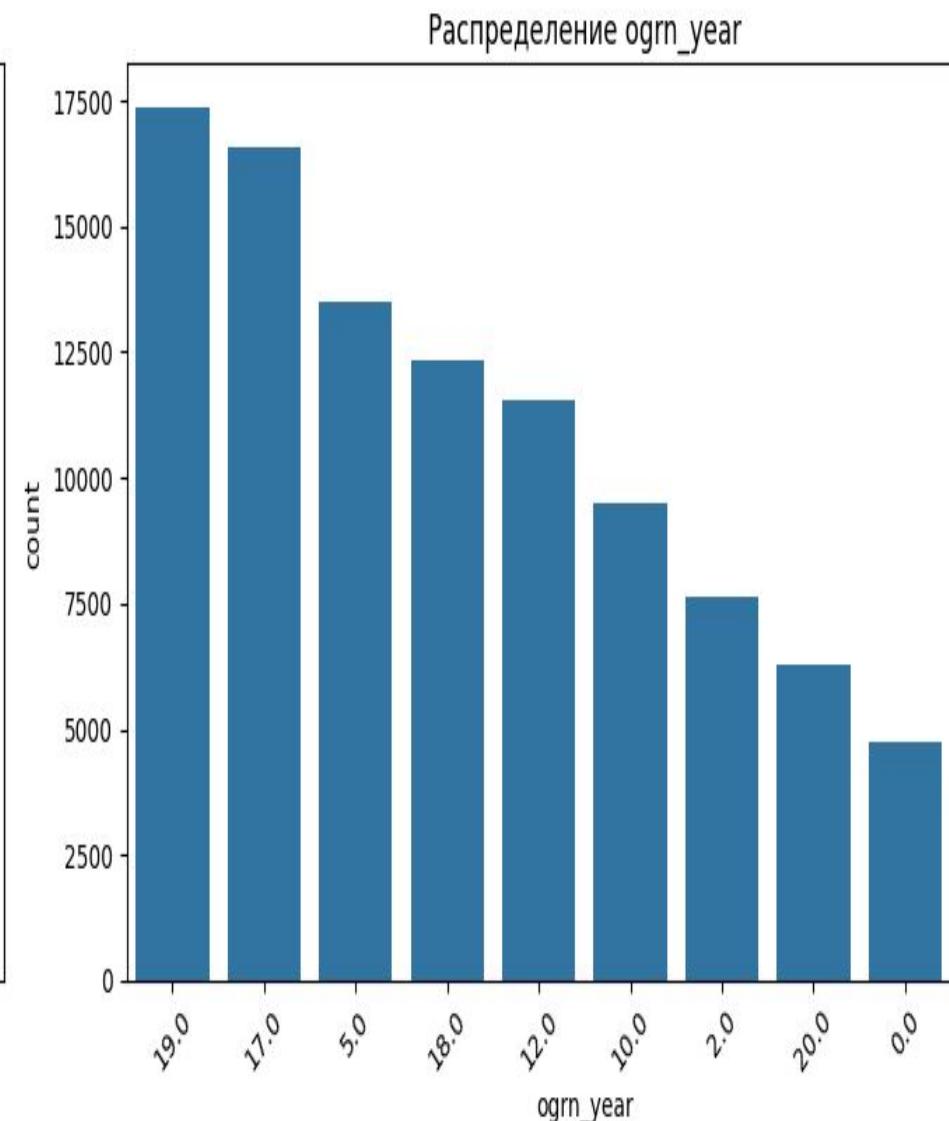
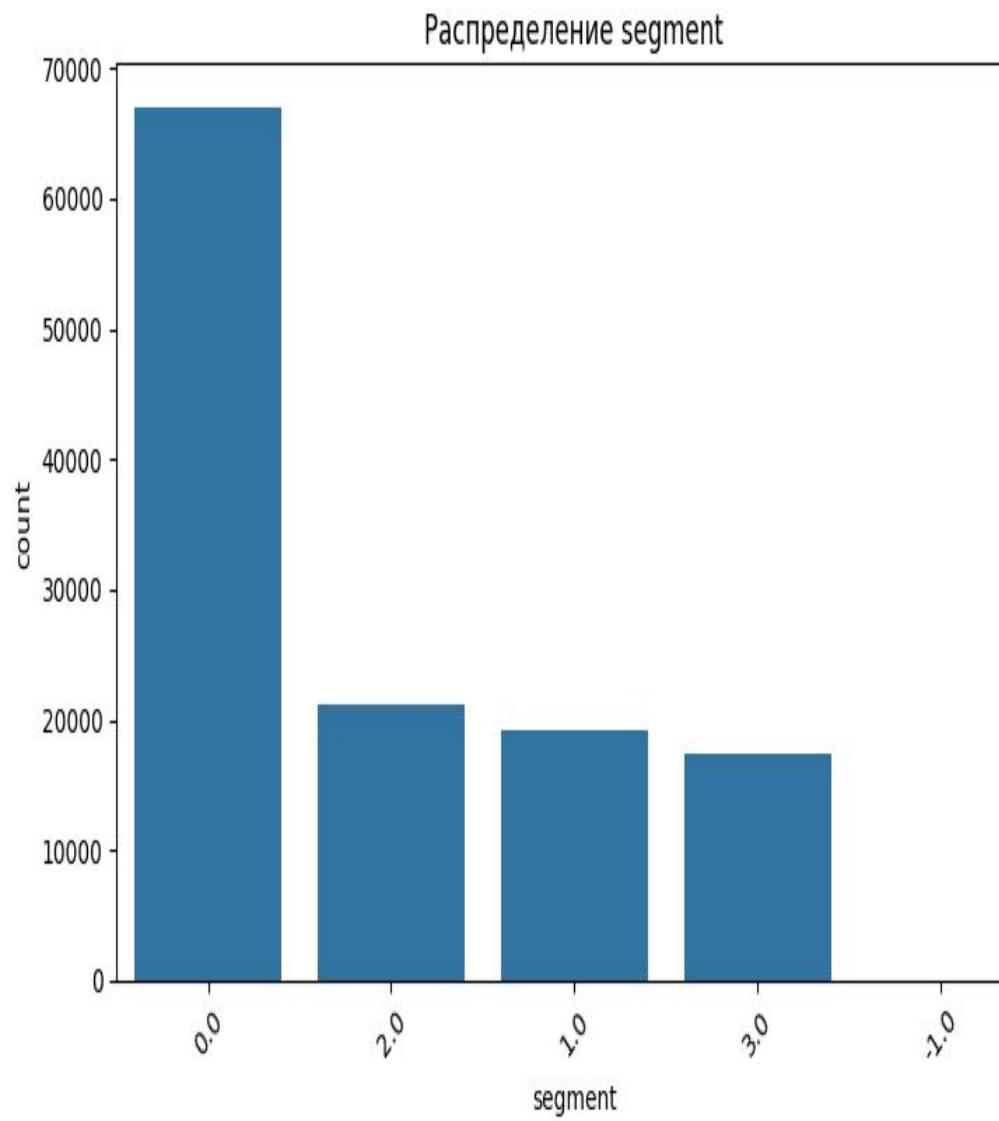
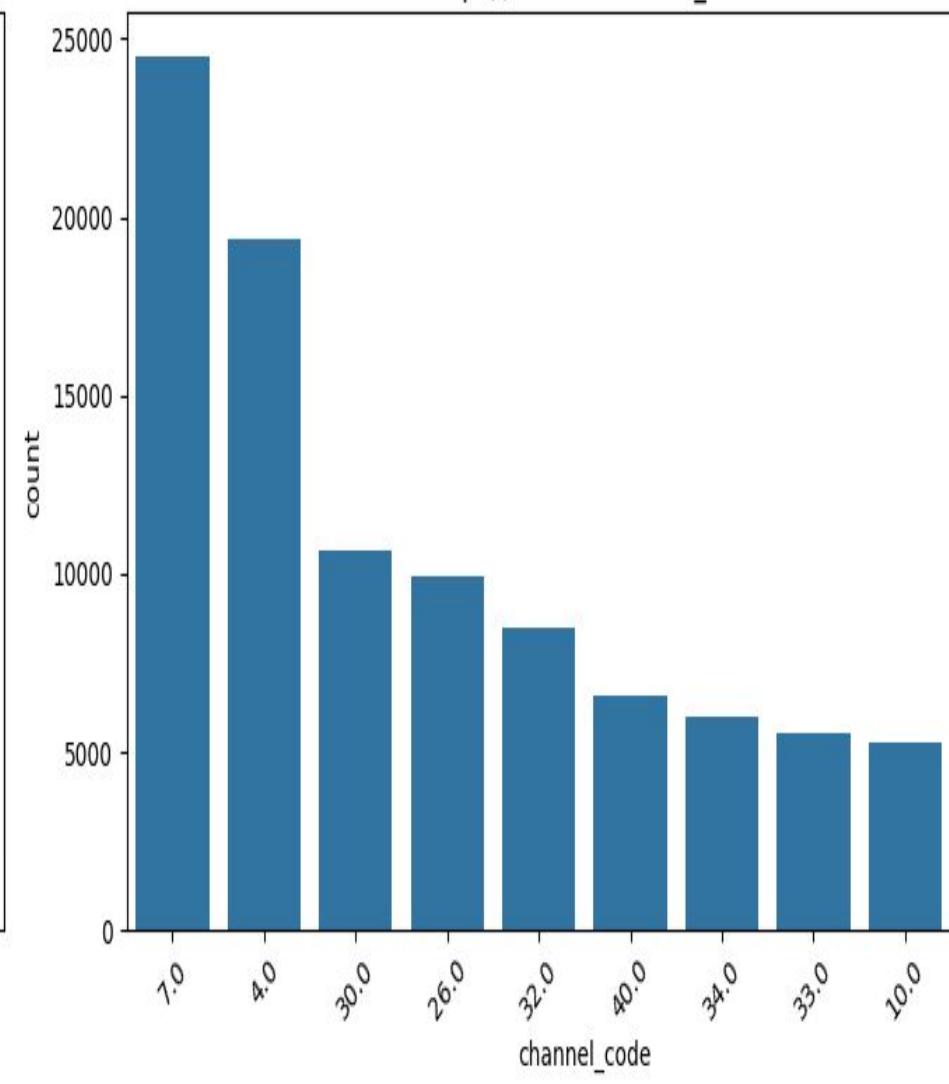
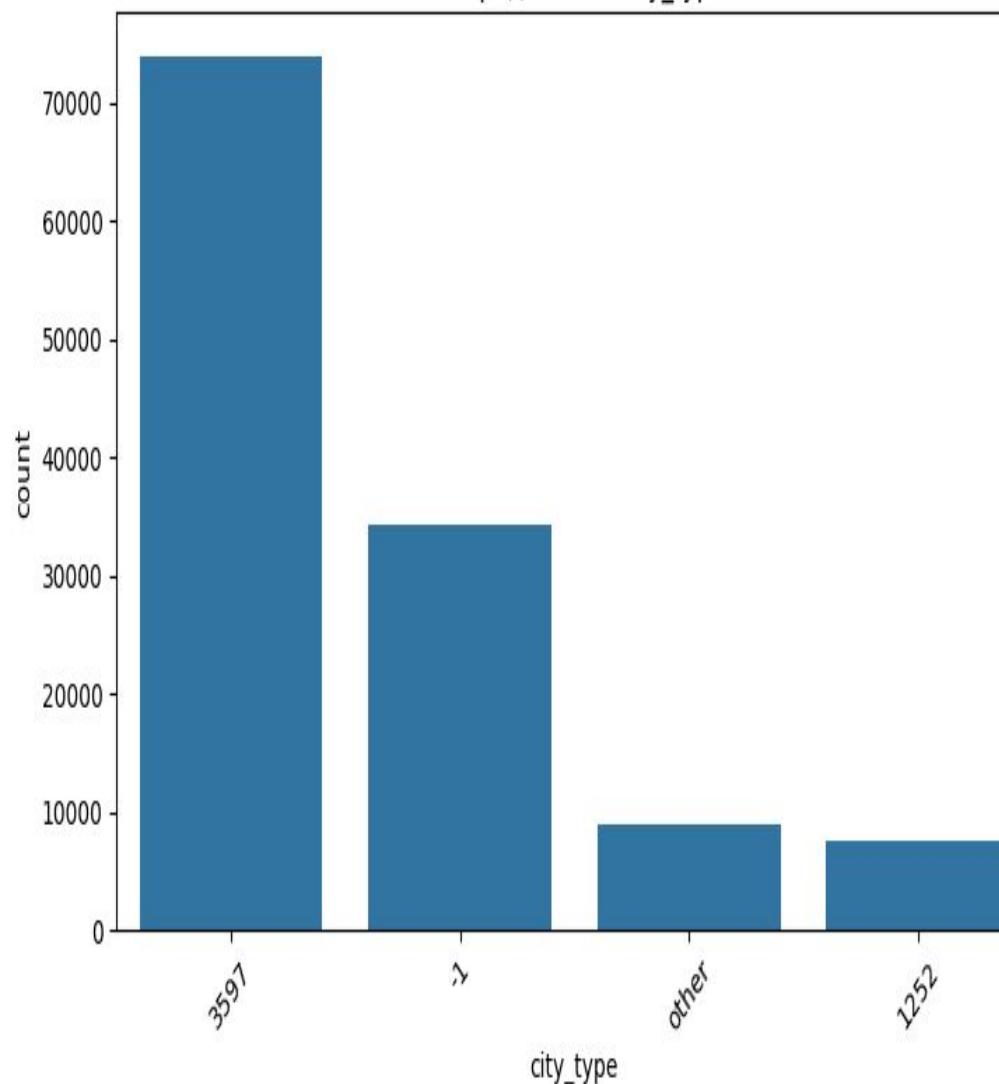
# Портрет клиента и EDA



---

## 4 важных категориальных признака

# ВОТ ОНИ



(01)

тип города

(02)

канал привлечения

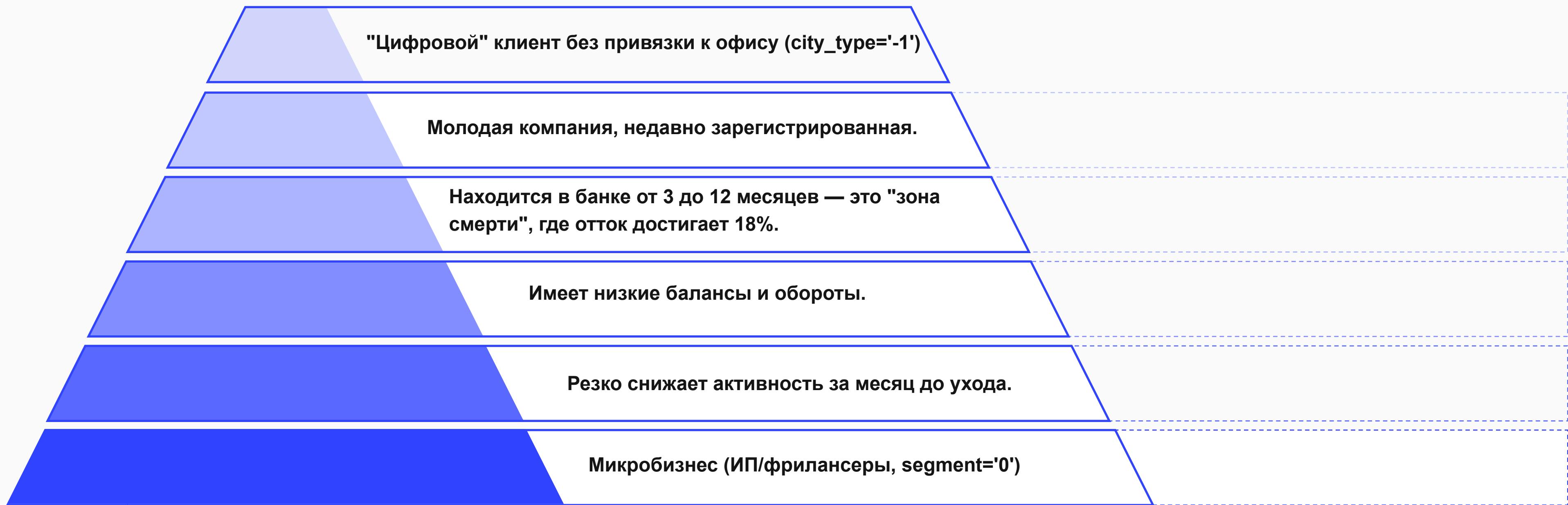
(03)

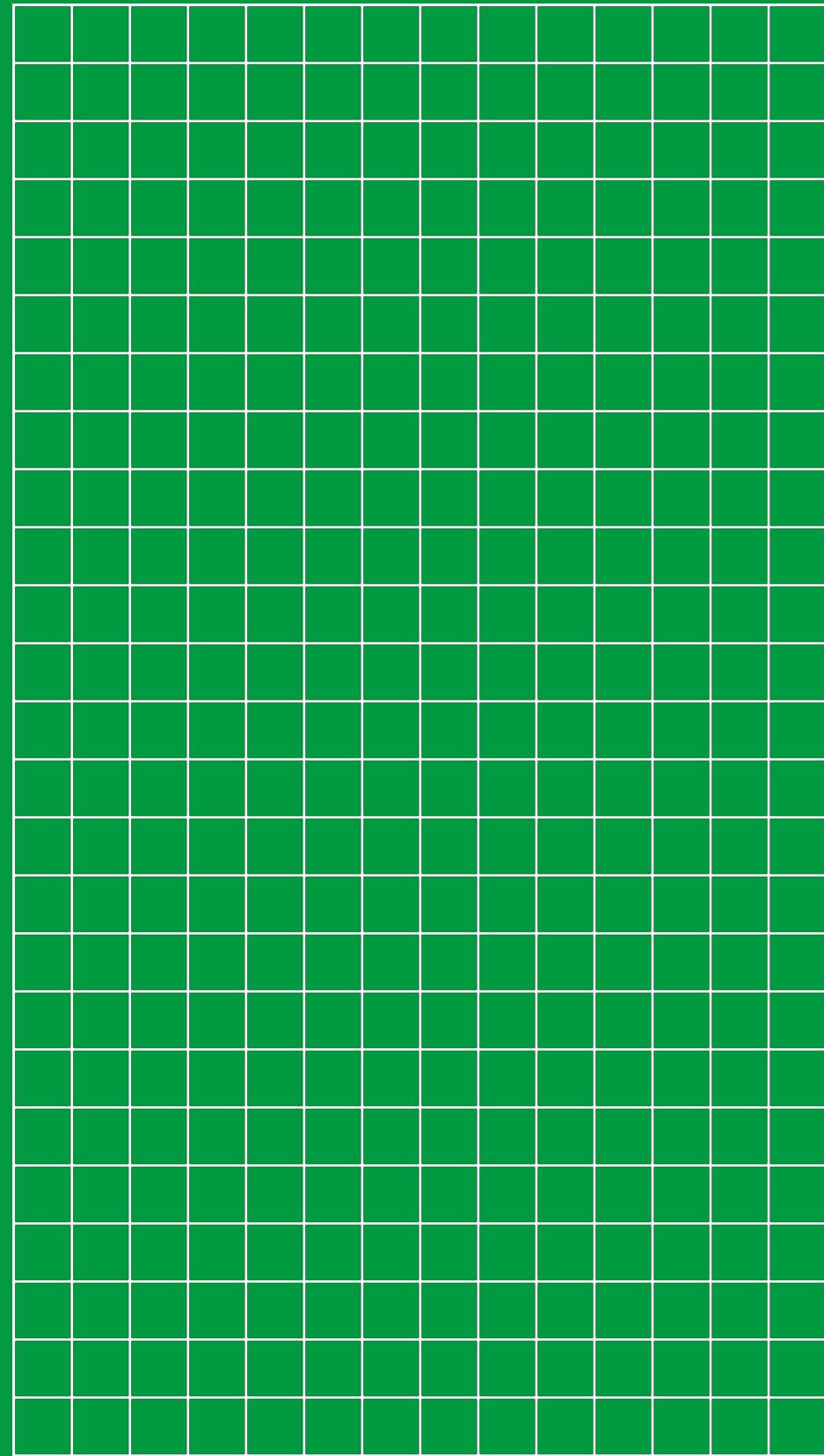
размер и статус кошелька  
клиента

(04)

дата рождения бизнеса

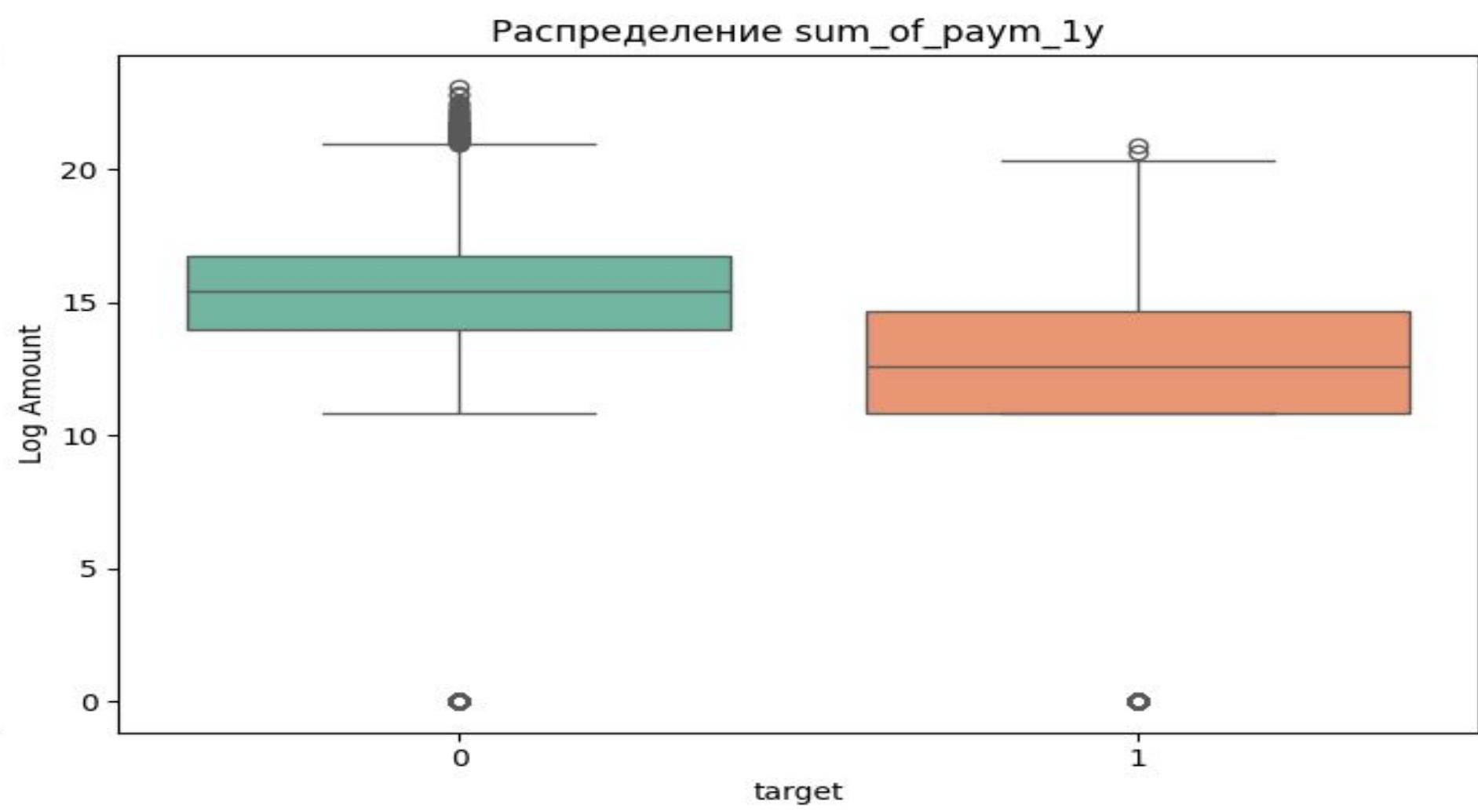
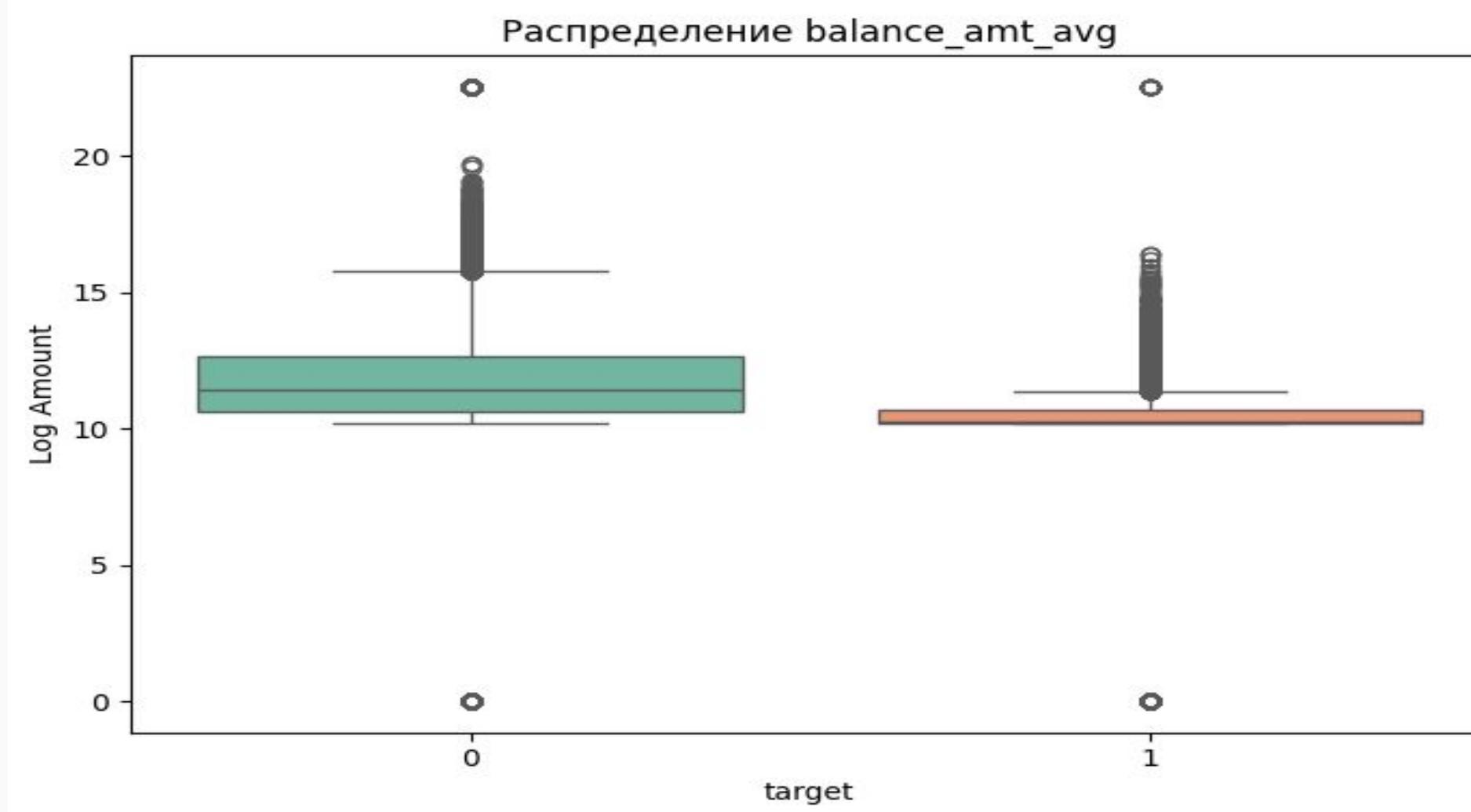
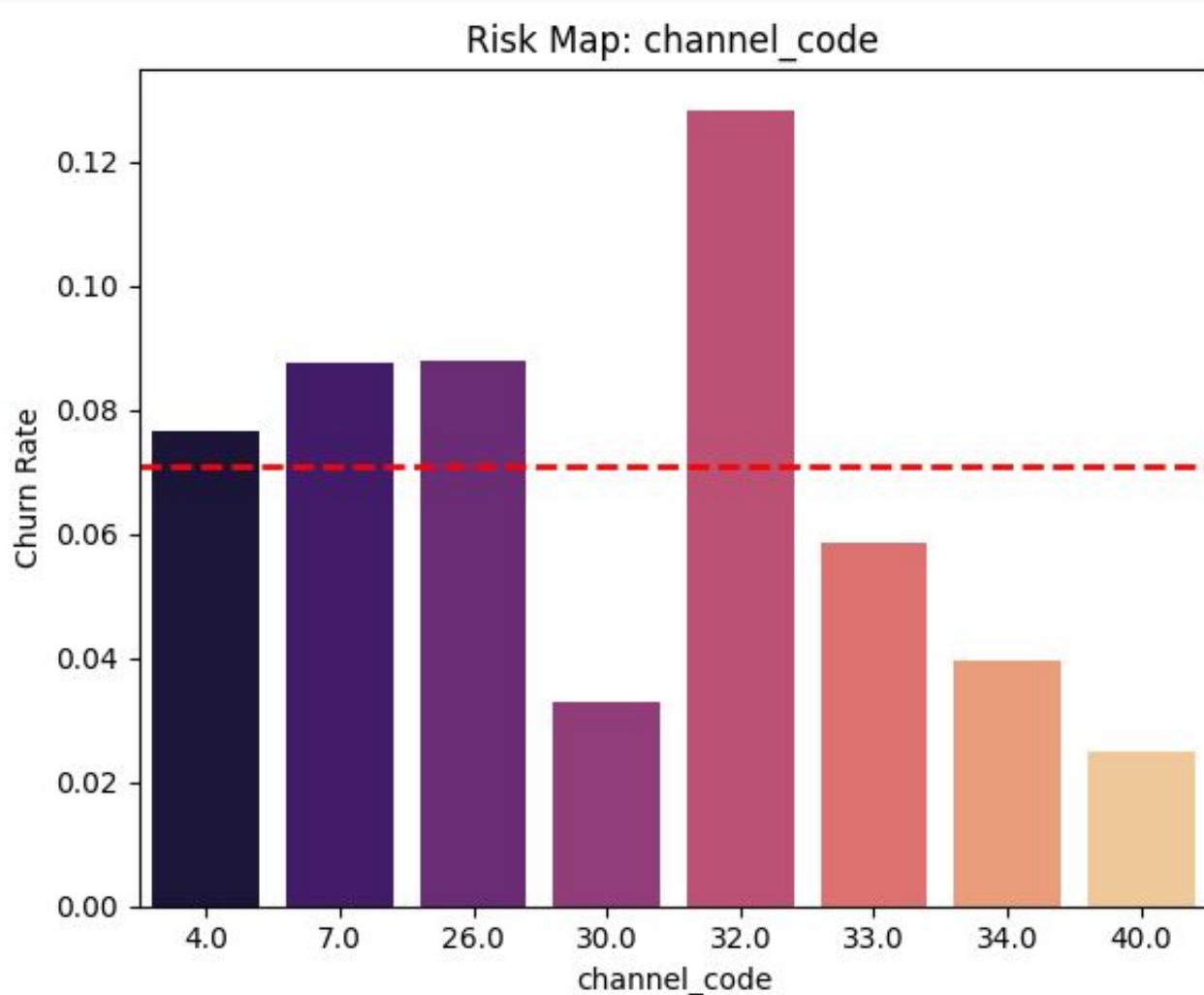
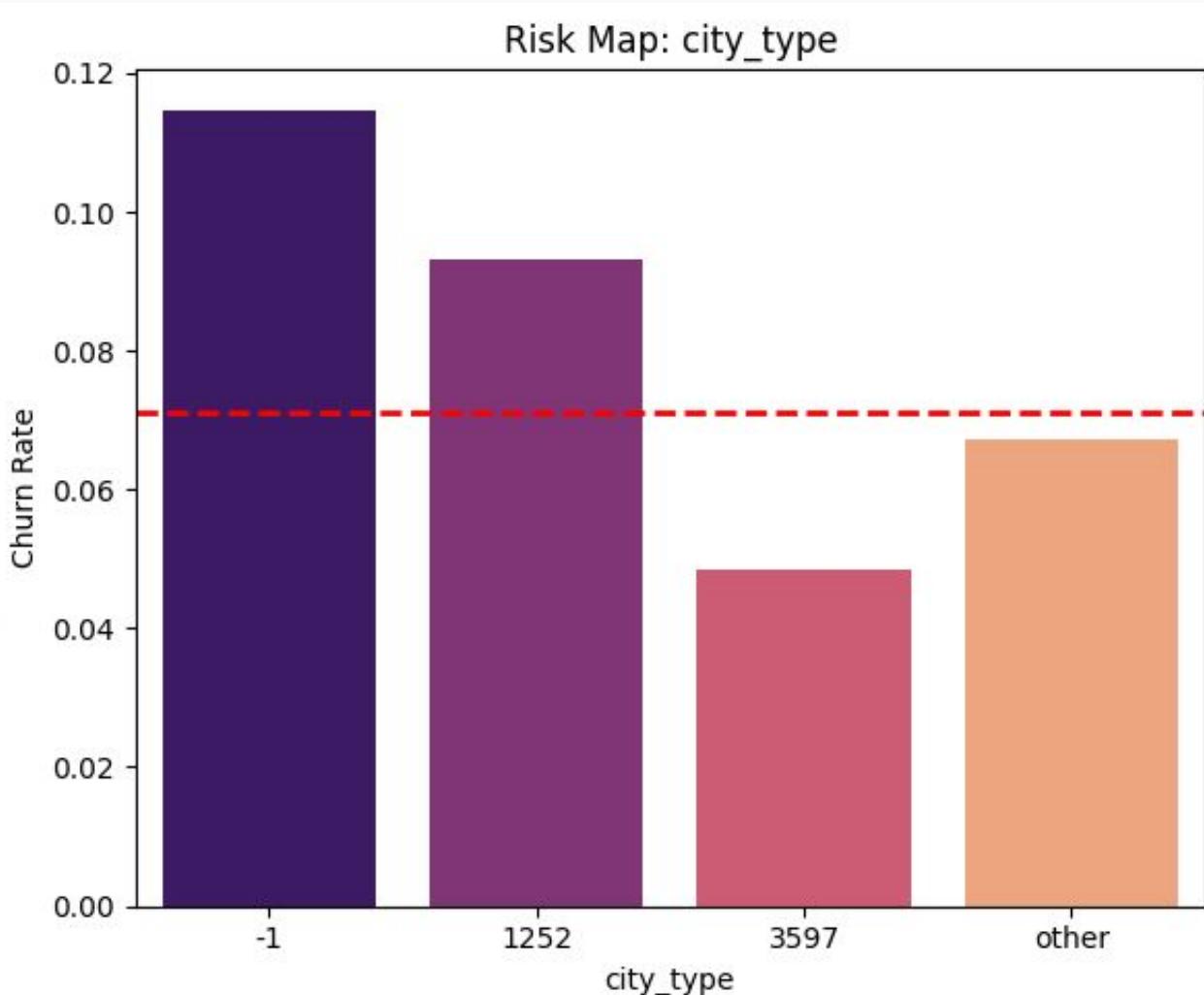
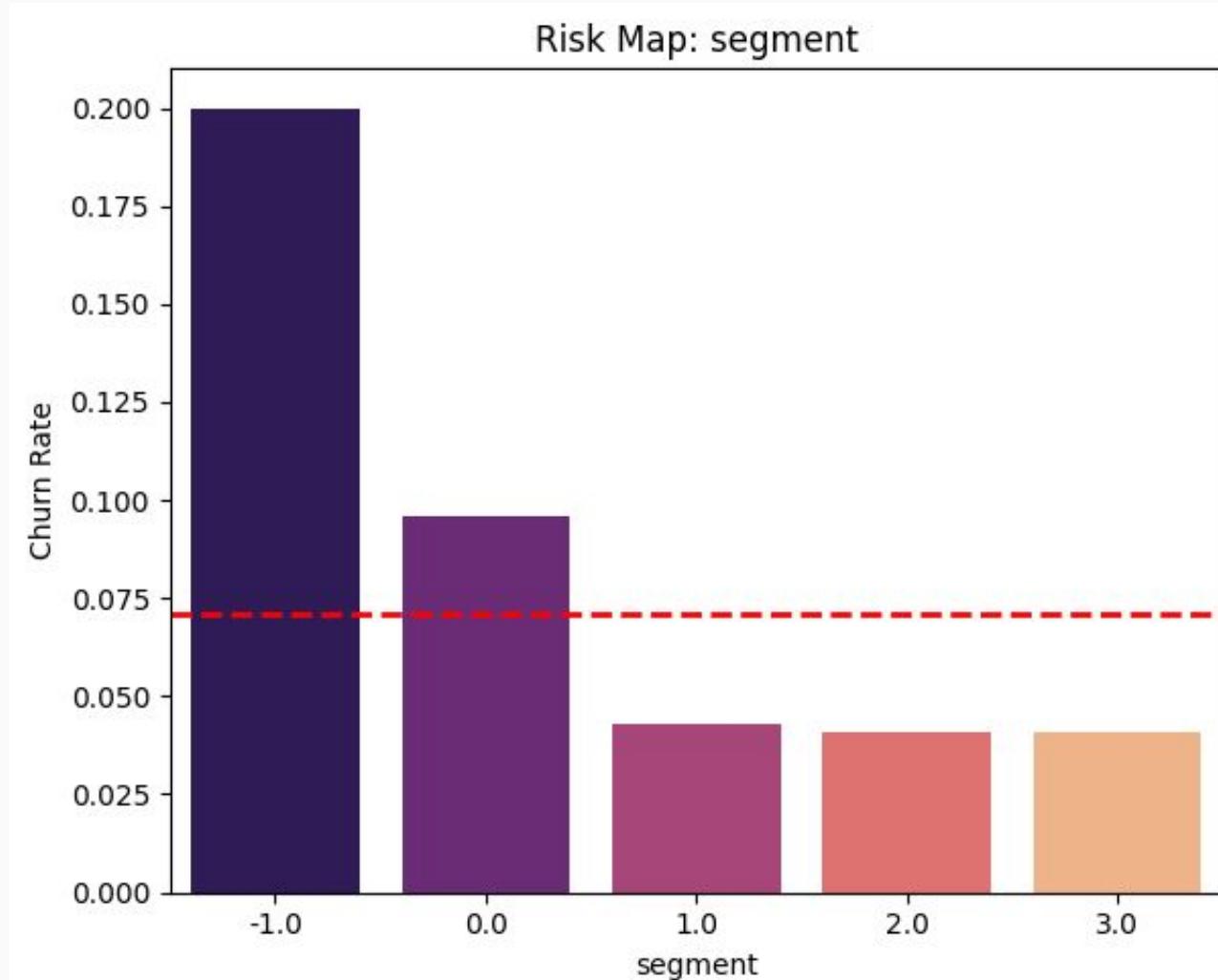
# Кто он, уходящий клиент?

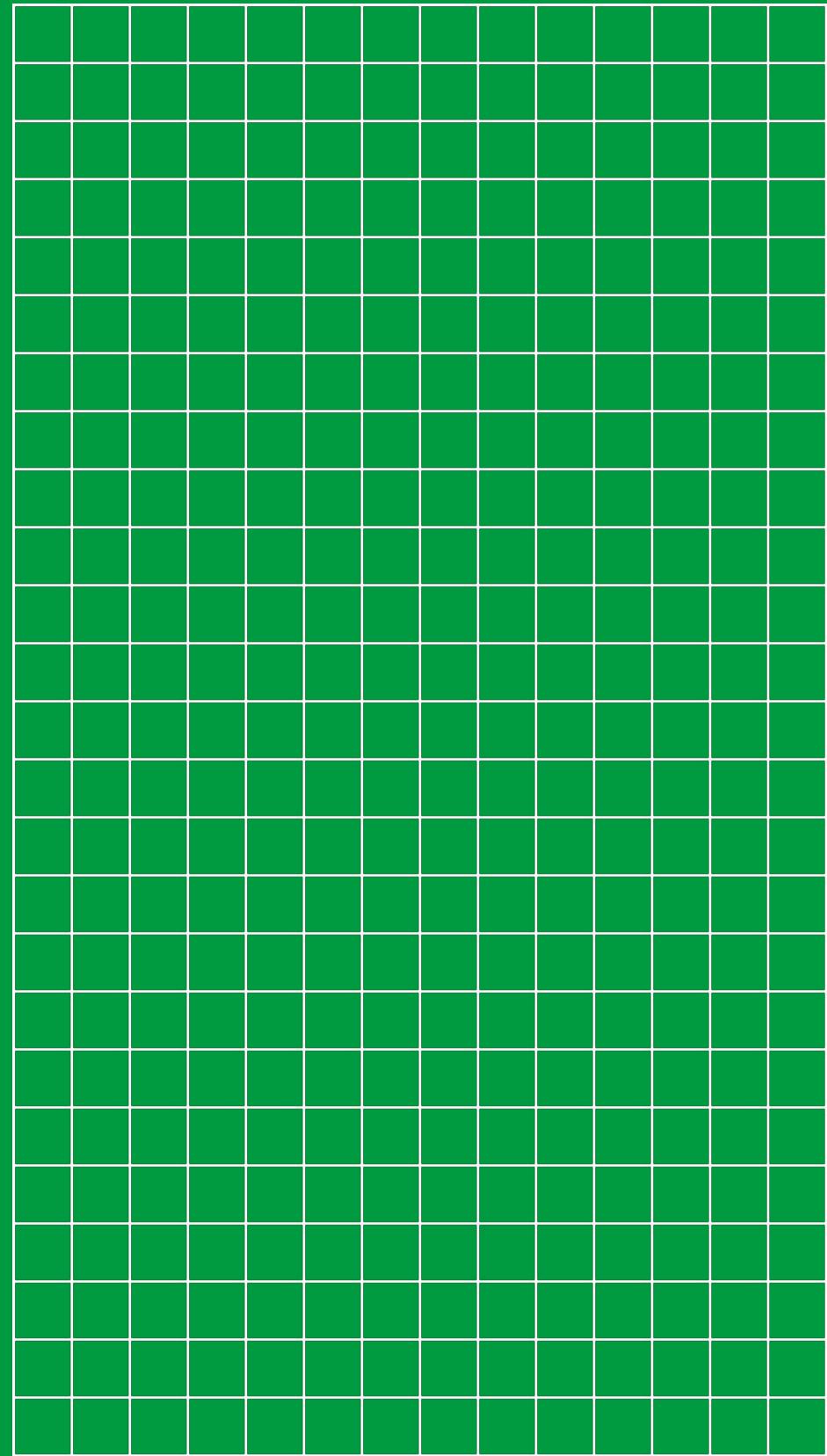




---

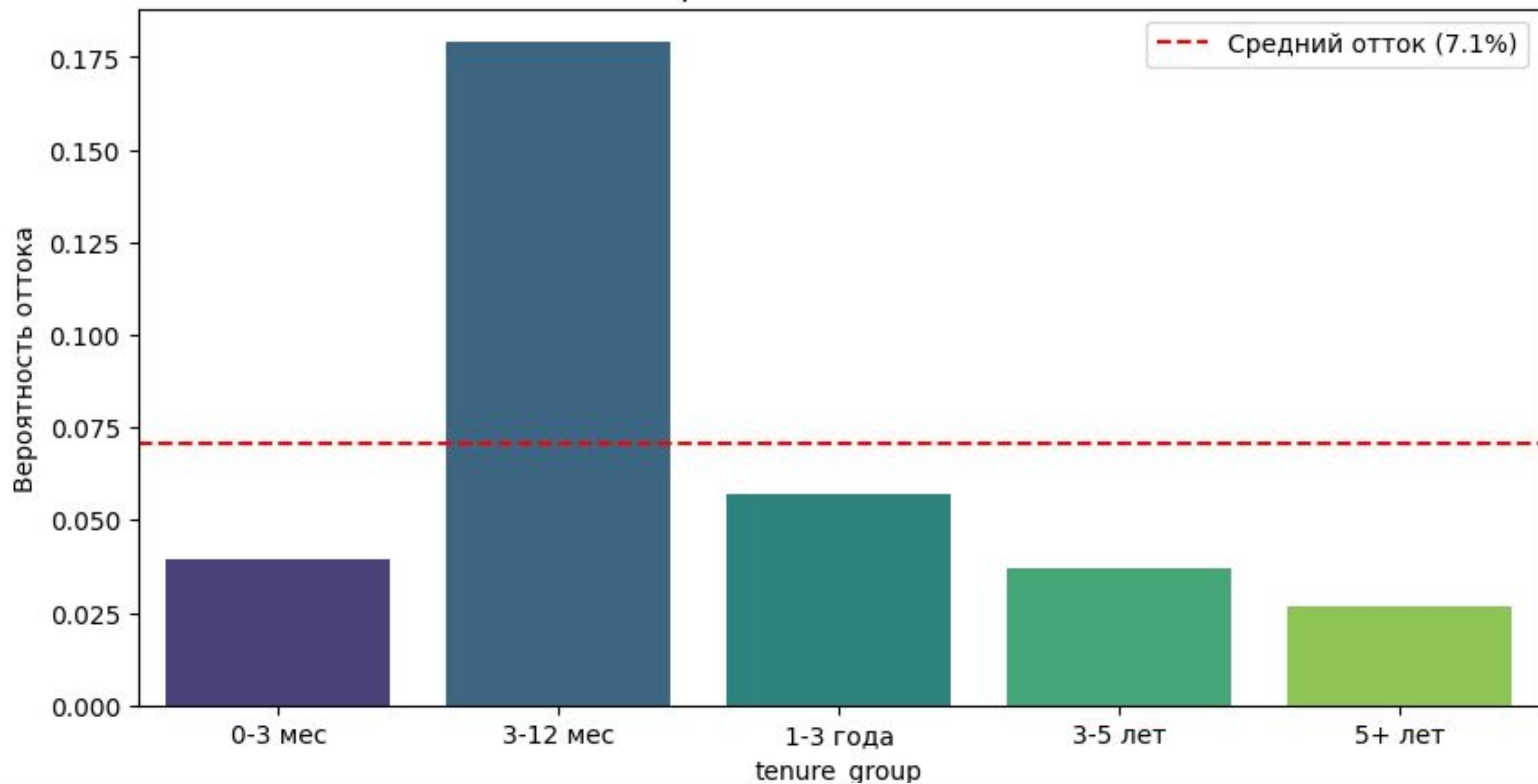
# Карта риска

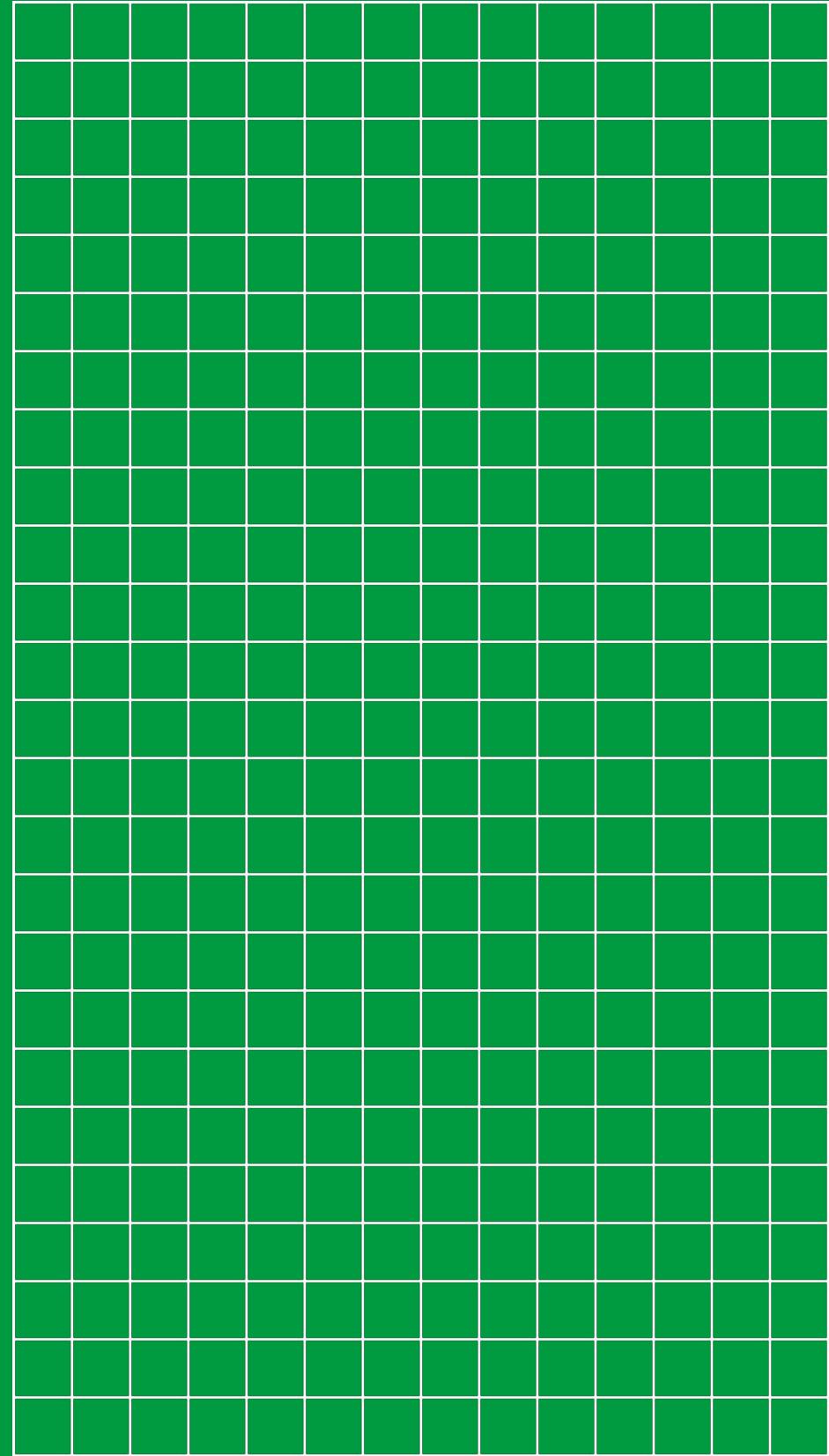




# Кри́вая вы́живаемости

## Кривая выживаемости

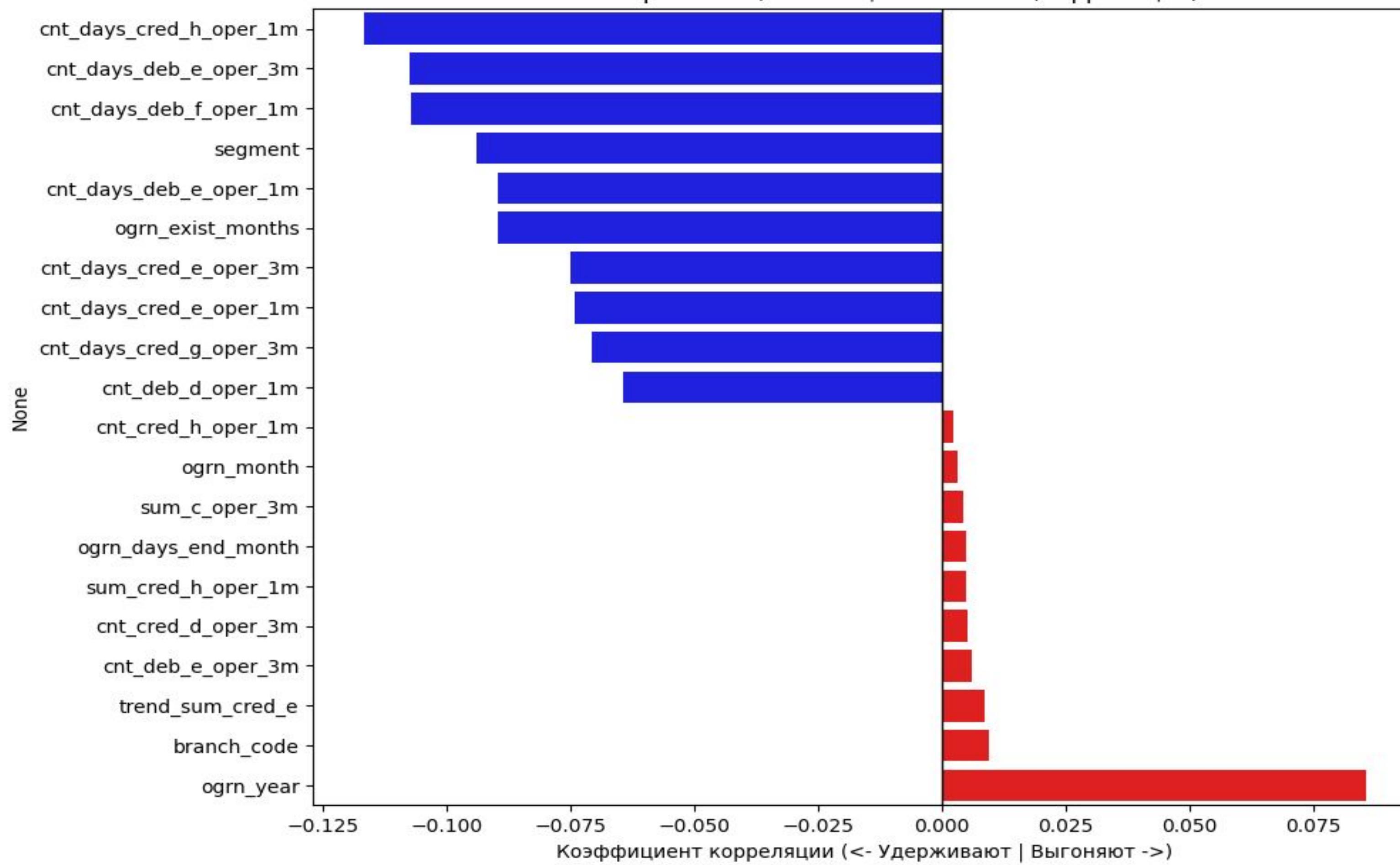




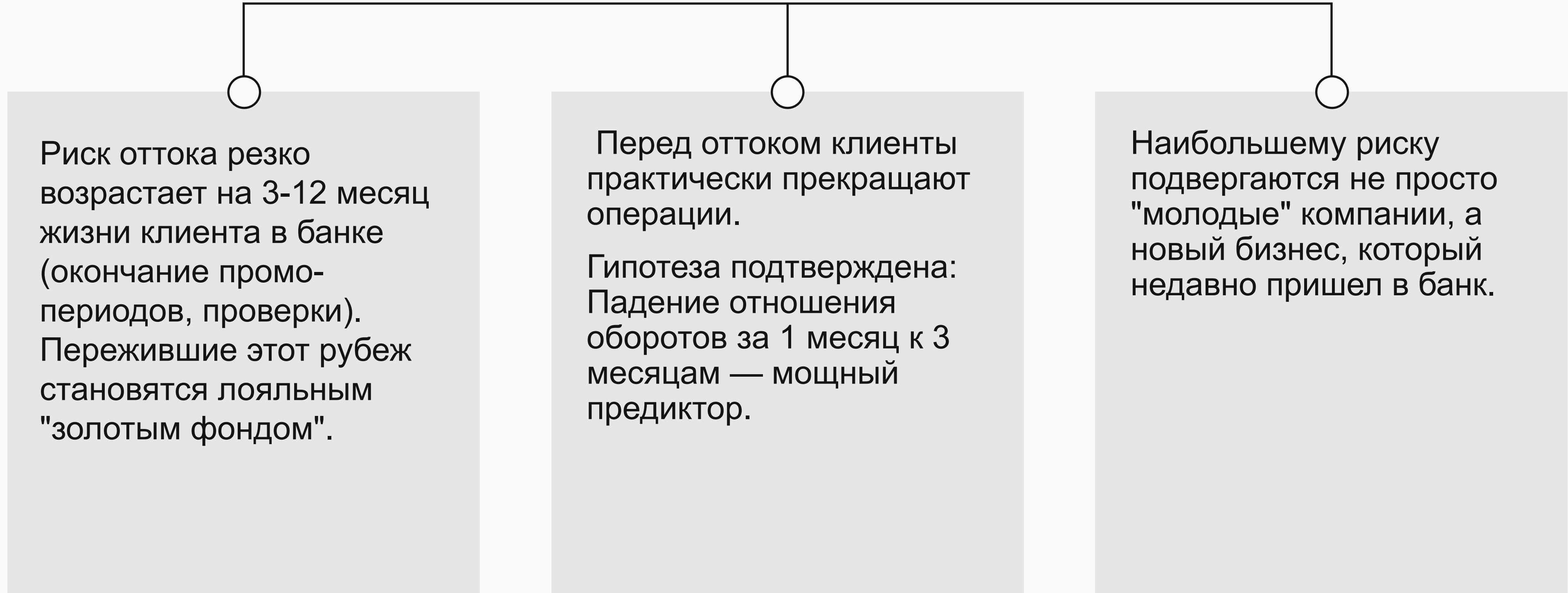
---

# Топ-20 признаков влияющих на отток

### ТОП-20 признаков, влияющих на Отток (Корреляция)

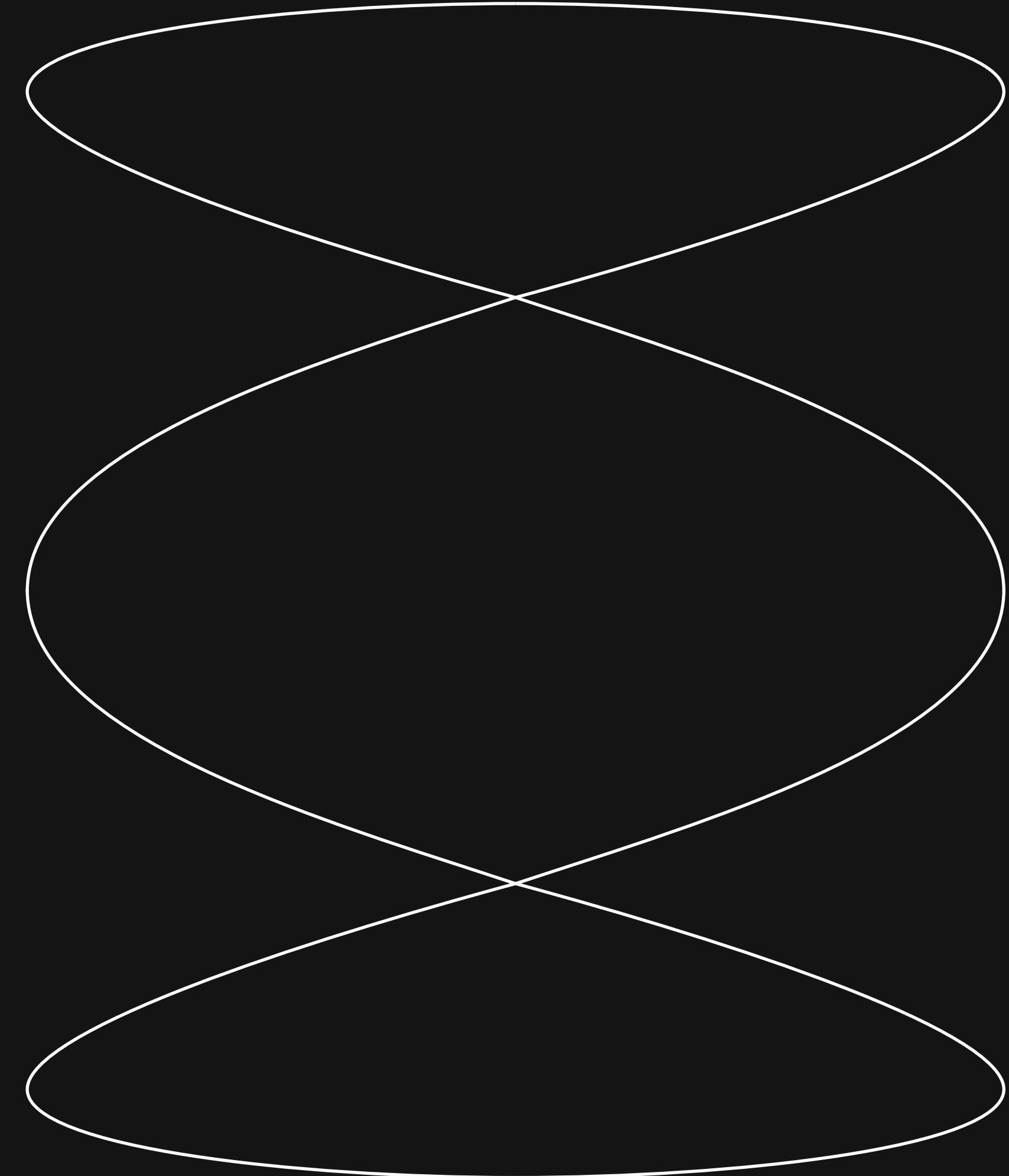


# EDA: "Зона смерти" и "Замирание"



# Аномалии

## Feature engineering



# Аномалии

Статистические методы (Z-score, IQR) подтвердили наличие экстремальных, но реальных значений.

ML-методы (Isolation Forest, LOF, SVM) позволили оценить "странные" в многомерном пространстве и создать композитный скор аномальности.

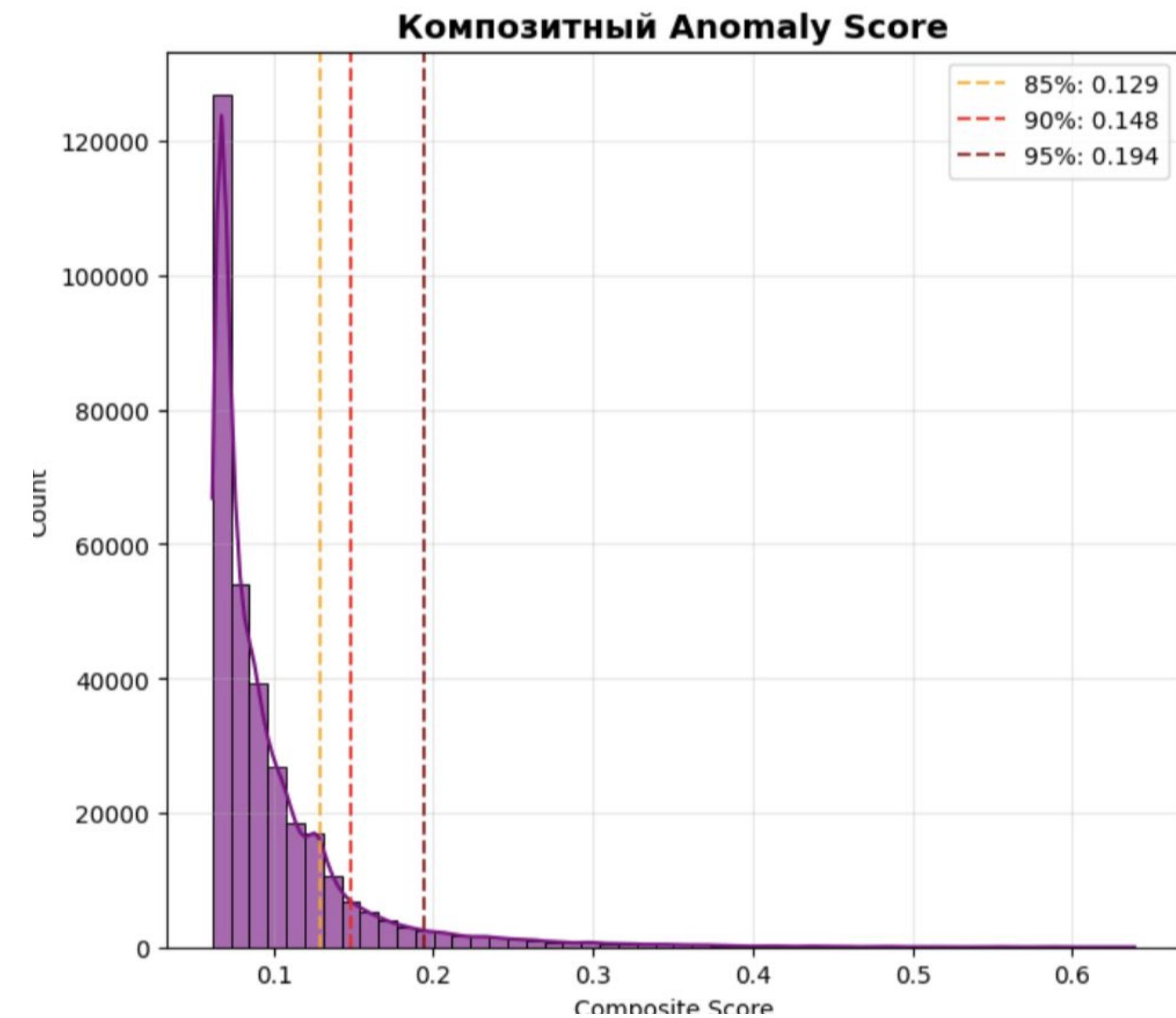
Главный инсайт: Доля оттока монотонно растет с ростом уровня аномальности. Аномальные клиенты уходят чаще.

**Для статистических тестов:**

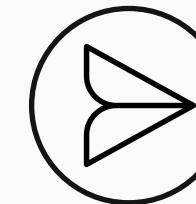
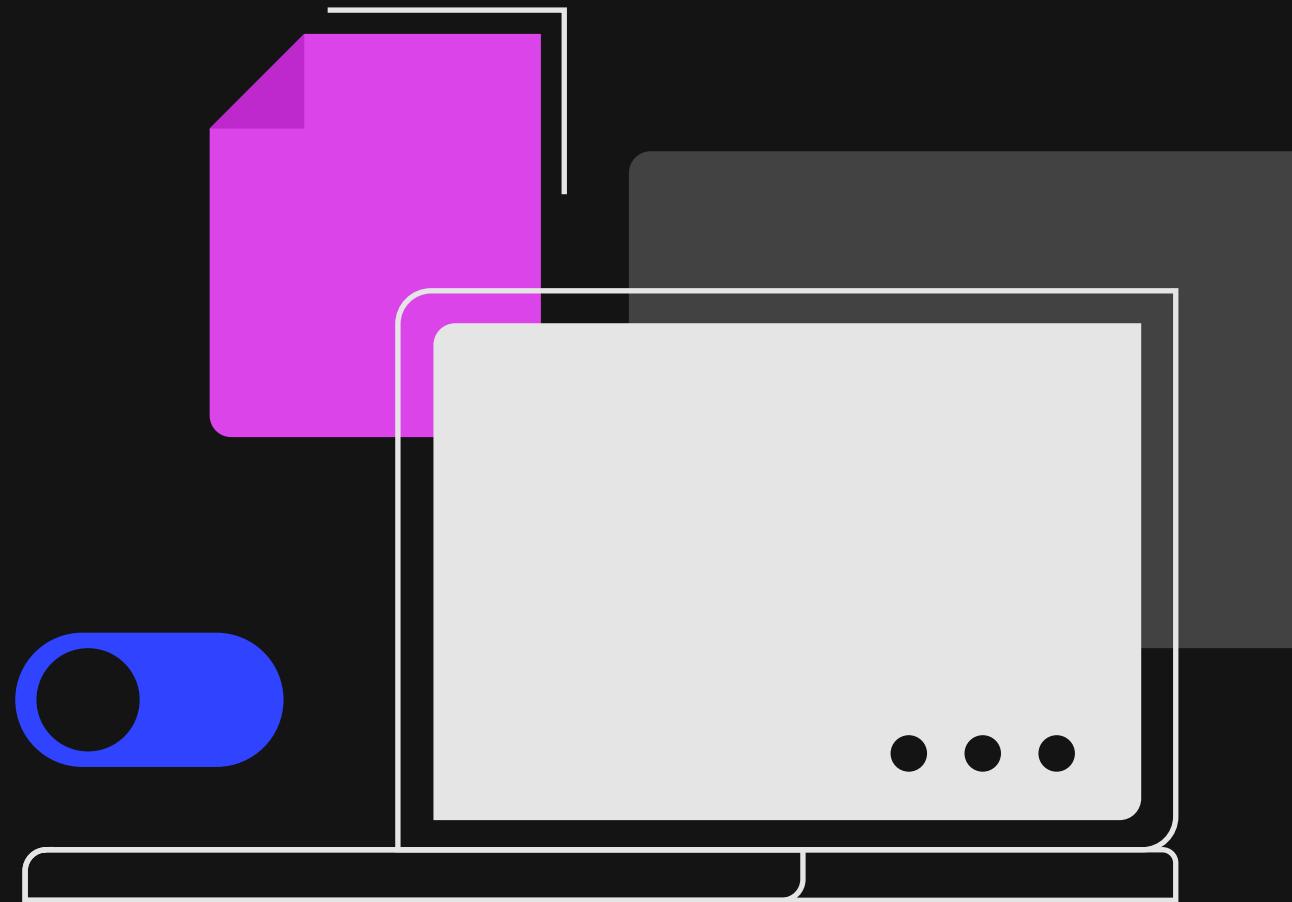
- **total\_anomaly\_flags** (общее число аномалий)
- **total\_zscore\_strength** (сила отклонений по z)
- **total\_iqr\_strength** (расстояние до границ)

**Для ML аномалий:**

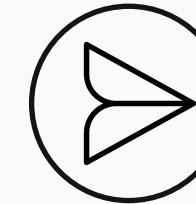
- **iso\_score** (аномальность по дереву)
- **lof\_score** (loc outlier factor)
- **svm\_score** (аномальность по svm)



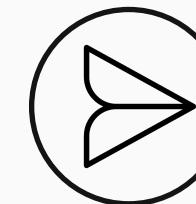
# Feature Engineering: Как мы усилили данные



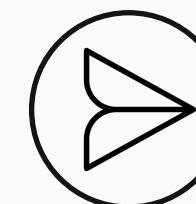
**Кодировка категорий: Target, Frequency, Label Encoding**



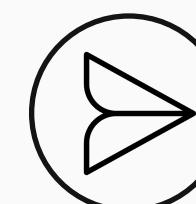
**Признаки соседей (KNN): Сравнение с "похожими" клиентами (баланс, платежи).**



**Временные признаки: Циклическое кодирование, логарифмирование, отношение срока в банке к возрасту бизнеса.**

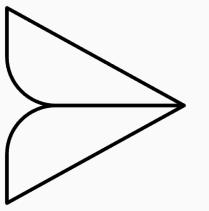


**Финансовые индикаторы: Стабильность баланса, оборачиваемость средств, регулярность активности.**

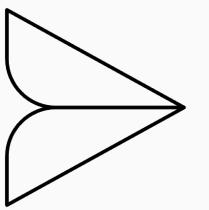


**Итого: получили много интересных и информативных признаков**

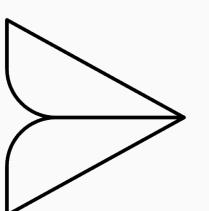
# Итого по моделям



Модель: CatBoost Classifier  
(градиентный бустинг на деревьях).



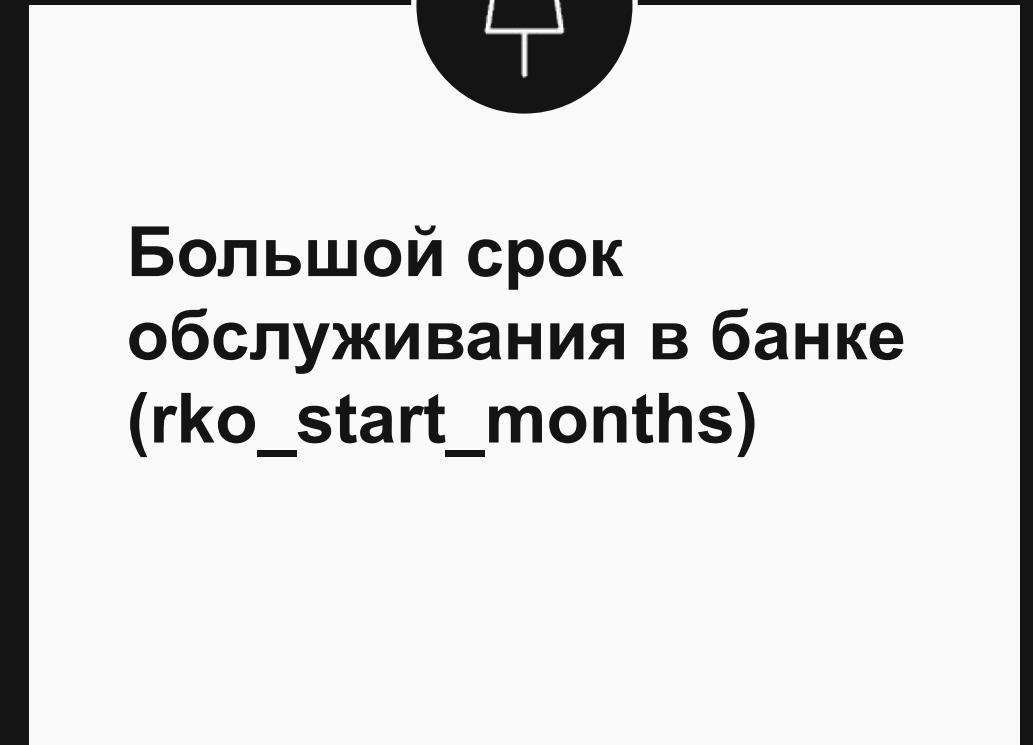
Процесс отбора признаков: Консенсус  
важности из 4 методов + удаление  
коллинеарных признаков.



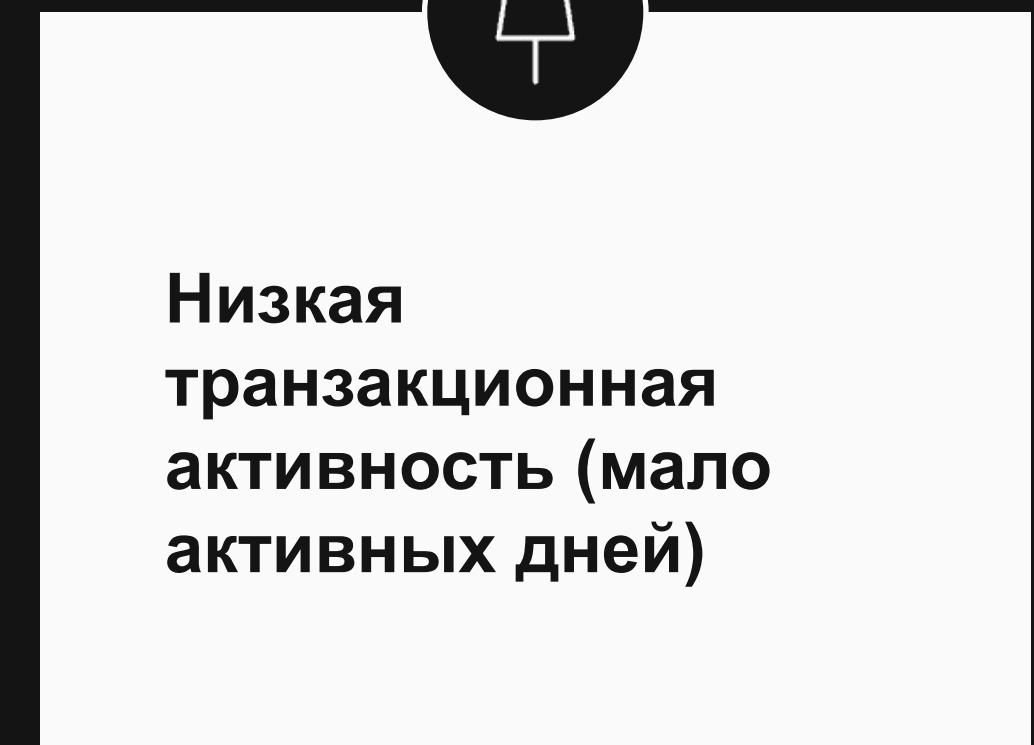
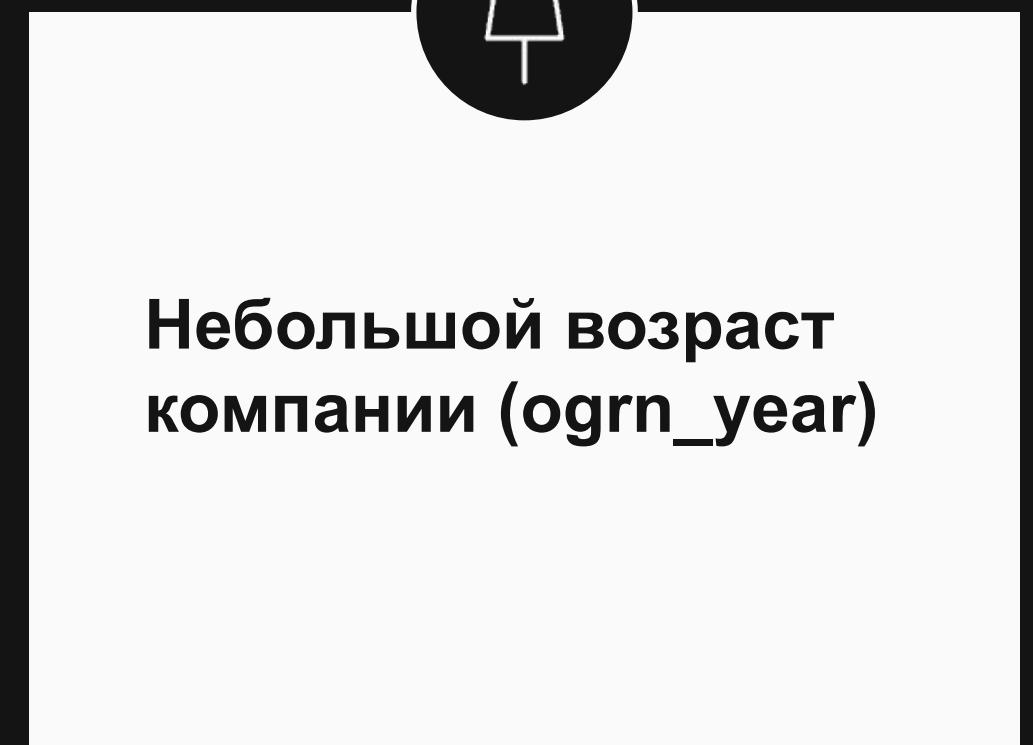
Итоговая метрика: ROC-AUC = 0.9012  
Модель демонстрирует высокую  
прогнозную способность.

# Интерпретация модели: Глобальные драйверы (SHAP)

Удерживает

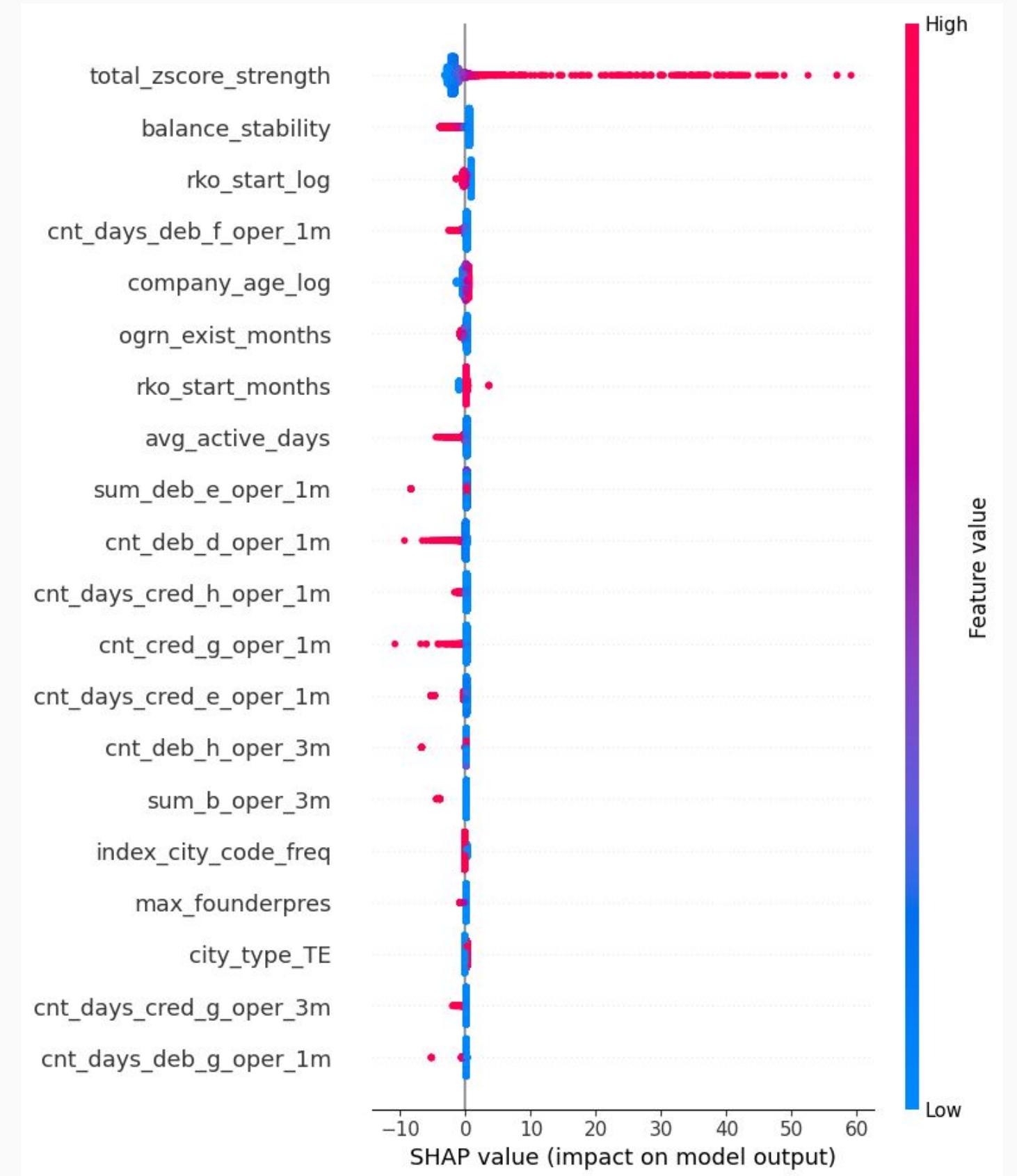


Отталкивает

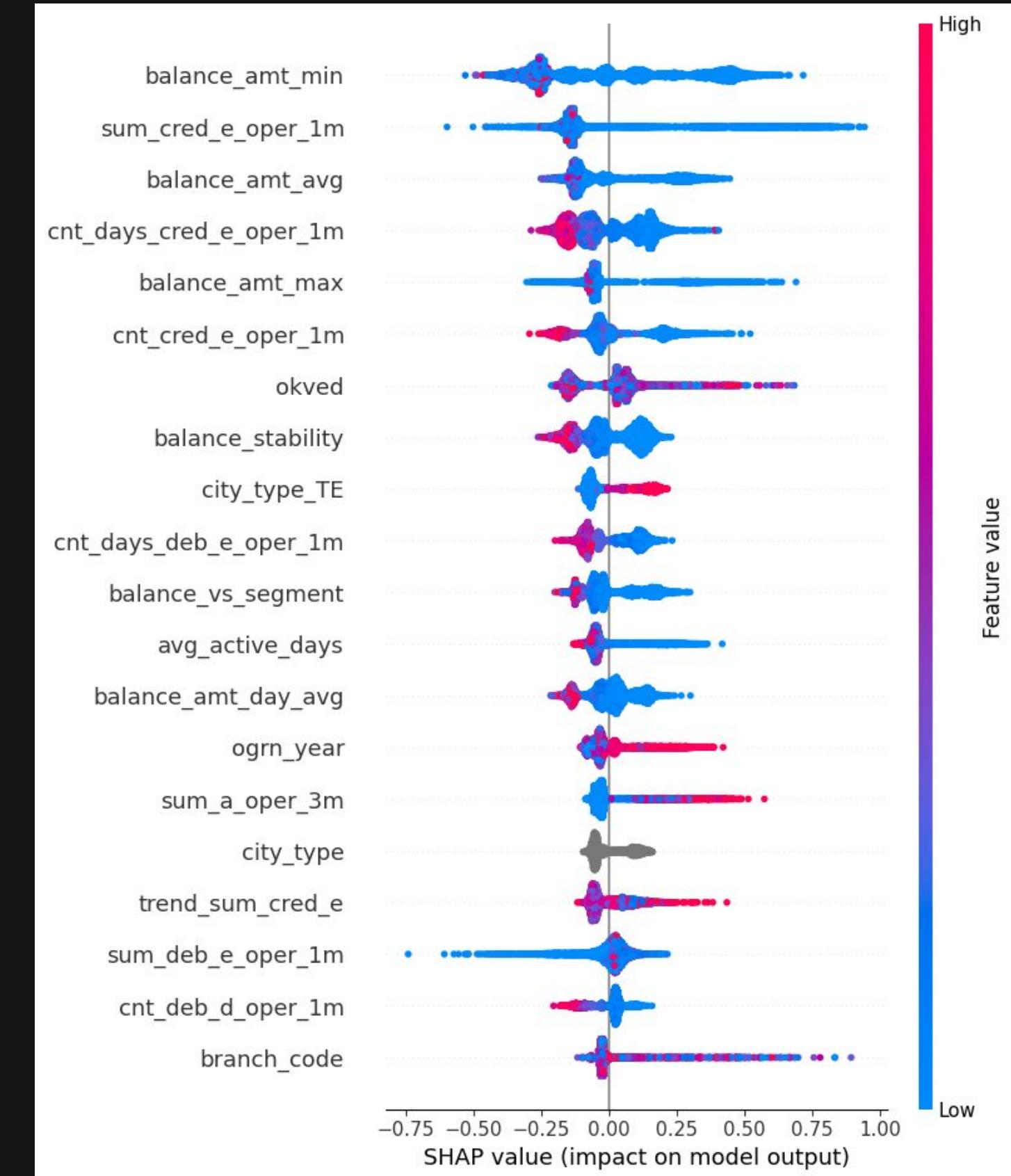


# CatBoost vs LR

Logistic Regression



CatBoost

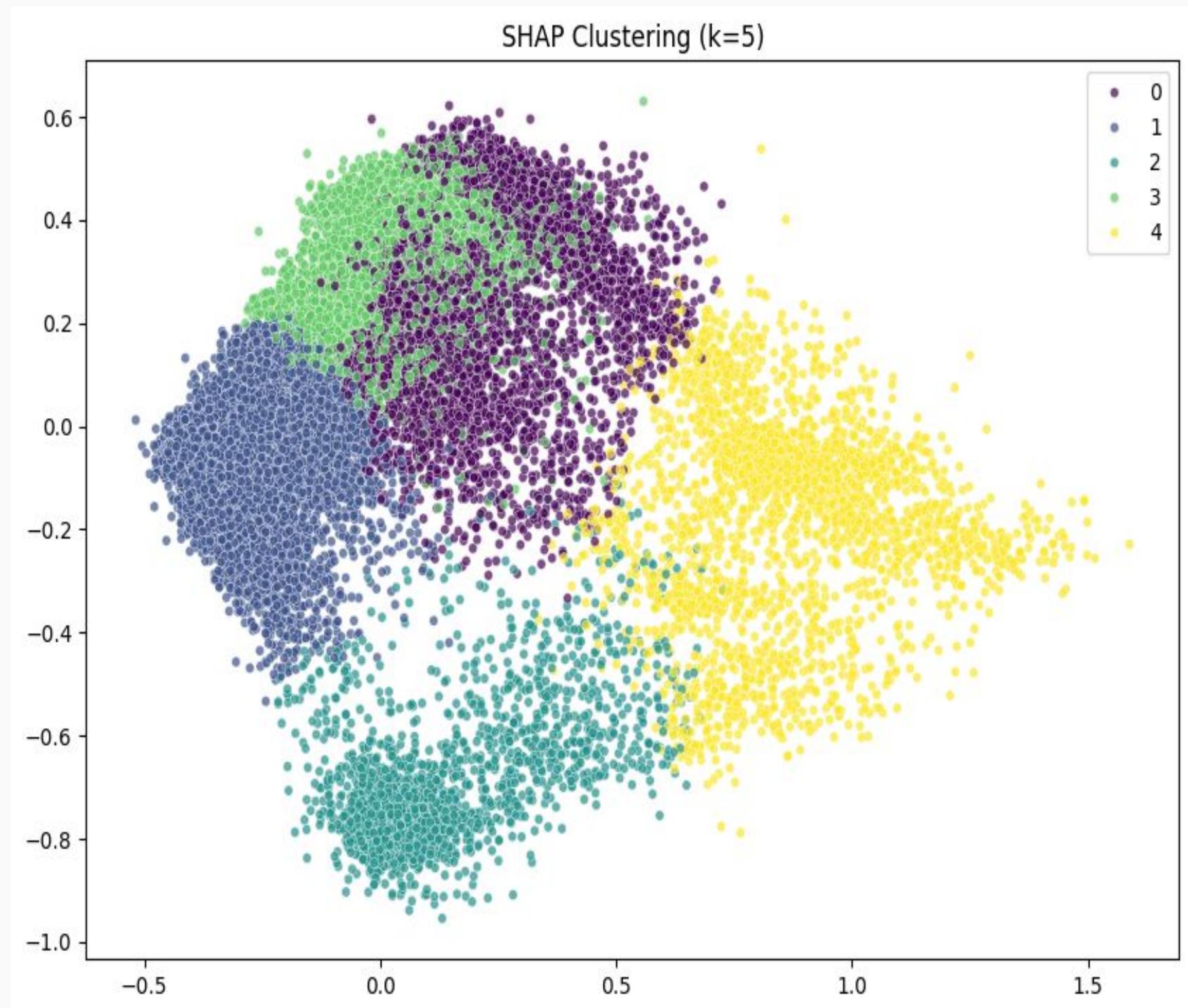


# Сегменты оттока

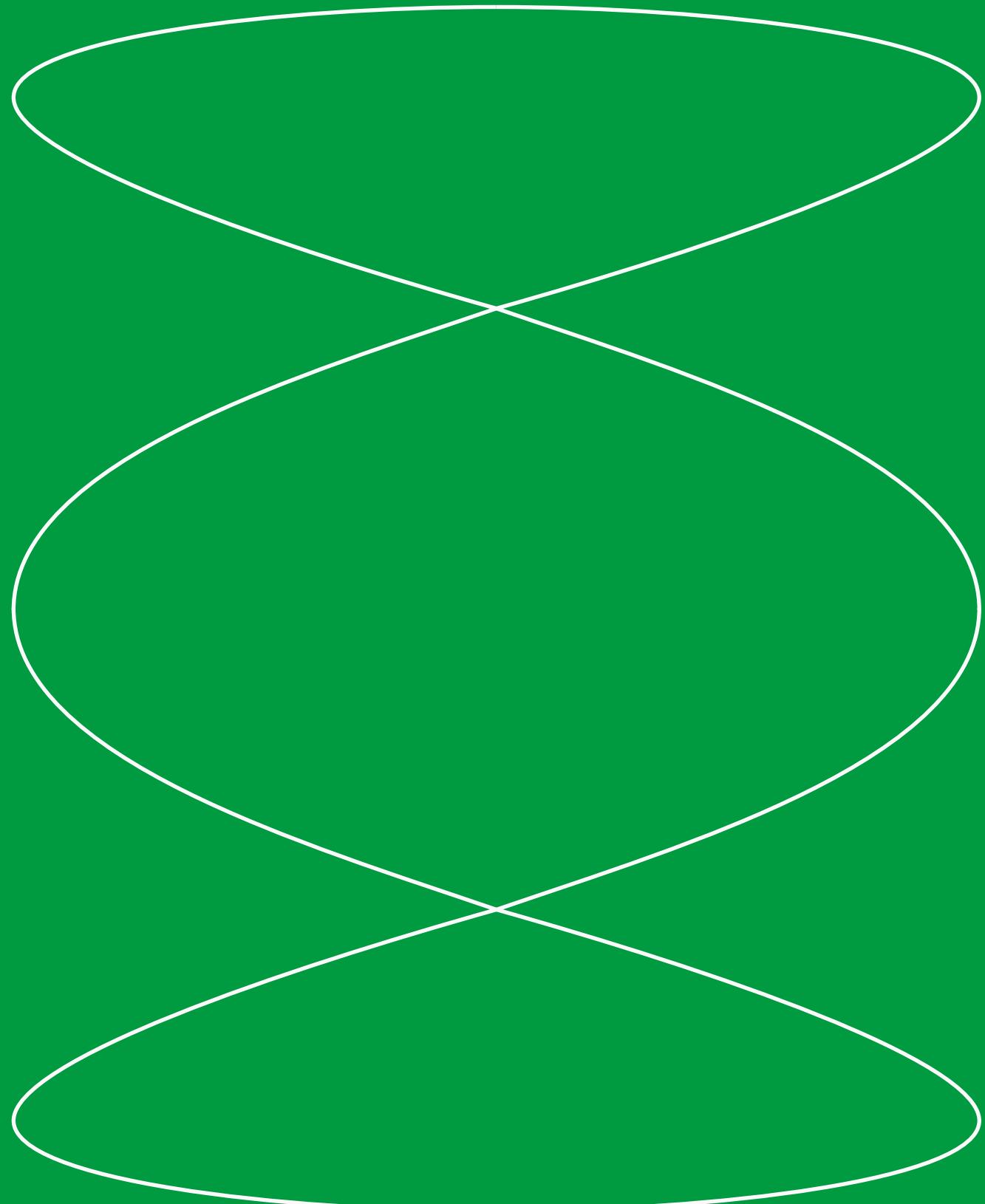
**SHAP-значения клиентов были сгруппированы в 5 кластеров.**

**Каждый кластер - это архетип клиента с уникальной комбинацией причин оттока (напр., "Новички", "Финансово истощенные").**

**Бизнес-ценность:** Позволяет перейти от общей стратегии удержания к персонализированным действиям для каждой группы риска.

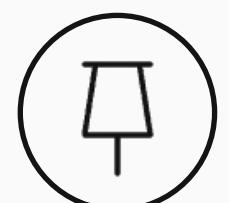


# Эксперименты и выводы по устойчивости модели



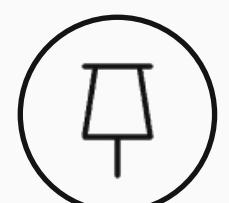
**Удаление "странных" клиентов (SHAP-аномалии) → Метрика УПАЛА.**

Крайние случаи критически важны для обучения.



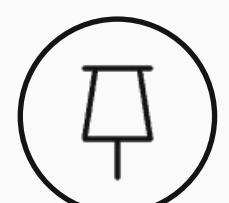
**Добавление номера SHAP-кластера как признака → Прироста НЕТ.**

CatBoost уже сам выявляет эти паттерны.



**Обучение только на SHAP-значениях → Метрика СИЛЬНО УПАЛА.**

SHAP — инструмент объяснения, а не замены feature engineering.

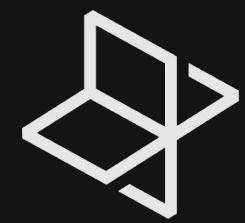


**Общий вывод: Исходный feature engineering оказался оптимальным, а модель CatBoost — робастной.**

## Что советуем бизнесу?

---

- 1. Сфокусироваться на "зоне смерти": Активная программа удержания для клиентов на 3-12 месяце жизни в банке.**
- 2. Внедрить мониторинг "замирания": Триггерные алерты при резком падении операционной активности клиента.**
- 3. Сегментировать действия: Использовать кластеры риска для точечных предложений: новичкам — поддержка, "истощенным" — финансовые консультации.**
- 4. Принять "аномалии": Не игнорировать нестандартное поведение, а исследовать его как ранний сигнал.**



ЦЕНТРАЛЬНЫЙ  
УНИВЕРСИТЕТ

В целом всё!

