

Финансовый анализ и выявление фродовых транзакций

([github pet-project fraud-analysis-financial-transactions](#))



Аналитик Киселев Е.Ю.

Июнь 2025 г.

Исходные данные

Работа проводилась над **синтетическим** датасетом (созданным благодаря коду ИИ)

Названия колонок и их роль

600 наблюдений и 12 колонок

+/- реалистичные бизнес-данные

Несколько пропущенных значений (NaN) в разных столбцах

Чуть больше 35+ фрод-транзакций с подозрительными признаками для инсайтов

Вид структуры датасета

	Transaction_ID	Date	Transaction_Type	Department	Vendor	Payment_Method	Category	Approved_by	Recurring	Project_Code	Amount	Is_Fraudulent
211	0e806352-928a-498b-be5c-390011c39e60	2022-04-01	income	Sales	Gonzalez Group	card	taxes	CFO	False	PRJ_019	9823.71	False
246	c1c2d6f5-d57d-46c7-a62f-91e97817bbbc	2024-01-03	expense	Marketing	Alexander, White and Banks	bank_transfer	salary	CEO	False	PRJ_003	-8265.76	False
423	a75c371e-d974-4d22-8ed9-17d7884a01fc	2023-05-21	income	Sales	Valdez, Robinson and Fernandez	crypto	training	HR_Manager	True	PRJ_016	9613.30	False
511	8f263a7e-21cd-46c9-a7ca-670c1026e45d	2024-01-28	income	HR	Hays-Thomas	card	operational	CFO	True	PRJ_007	8858.78	False
135	84ce6228-390e-4b7c-a5c4-0c4b123b524	2024-03-17	expense	Finance	Lewis, Carroll and Walker	card	taxes	CTO	False	PRJ_003	-7818.21	False
226	4ffe8a41-b572-4bec-90a2-20edb512143	2021-11-17	expense	Sales	Patterson, Harding and Phillips	bank_transfer	taxes	COO	True	PRJ_020	-7122.59	False
116	a7a93b49-9e5a-48c2-b68e-58fe5576afd7	2021-11-01	expense	Finance	Henry, Hill and Johnson	card	travel	HR_Manager	True	PRJ_007	-3763.78	False

Рисунок 1 - сэмплинг семи случайных строк dataframe

Фокус работы

Все вопросы EDA-анализа

Цель: изучить финансовые данные с целью получения инсайтов для долгосрочного развития организации, а также понять природу фрод-транзакций

Задачи:

1. Выполнить предобработку данных: очистить от NaN, упростить регистр названия колонок, оценить распределение
2. Провести EDA-анализ, ответив на 15 исследовательских вопросов (в т.ч. про фрод транзакции) с использованием статистических тестов
3. Интерпретировать получившиеся результаты анализа

1. Как распределены транзакции по типам (transaction_type)?
2. Какое общее соотношение доходов и расходов по сумме (amount)?
3. Какие департаменты совершают больше всего транзакций и на какие суммы?
4. Какое распределение способов оплаты (payment_method) по типам транзакций?
5. Какие категории (category) являются самыми затратными?
6. Как распределены суммы (amount) в зависимости от одобряющего лица (approved_by)?
7. Какие проекты (project_code) потребляют наибольшие бюджеты?
8. Как меняется активность транзакций по месяцам и годам?
9. Есть ли сезонность в доходах и расходах?
10. Какие признаки чаще встречаются в фродовых транзакциях (is_fraudulent=True)?
11. Какое распределение сумм у фродовых vs. нормальных транзакций?
12. Какие поставщики (vendor) чаще фигурируют во фродовых операциях?
13. Есть ли связь между фродом и регулярностью (recurring)?
14. Есть ли департаменты, которые чаще попадают в фрод?
15. Какие наиважные признаки у фродовых и нормальных транзакций?

Проблема №1

Преобладание количества транзакций расходов над доходами

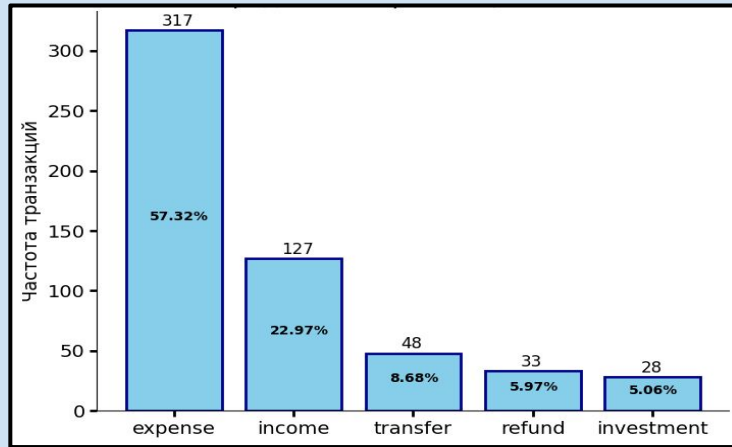


Рисунок 2 - распределение транзакций по их типам

Каждая вторая транзакция является расходом

На каждую транзакцию income приходится примерно по **2,5** транзакции expense

Возможно, это из-за сезонности дохода с несколько разными объемами сумм, нежели расходов

Другая гипотеза - большой портфель проектов, которые мало окупают себя

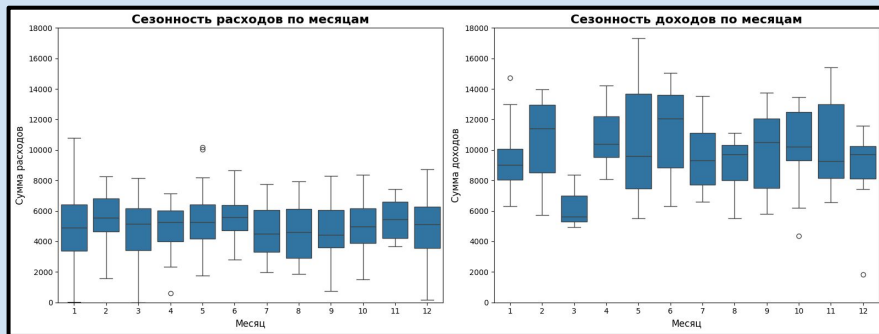


Рисунок 3 - сезонности расходов и доходов

Проблема №2

Потенциальный убыток в соотношении expenses/income за месяца



Рисунок 4 - тест Манн Кендалла по количеству, среднему и объему за месяц income-транзакций

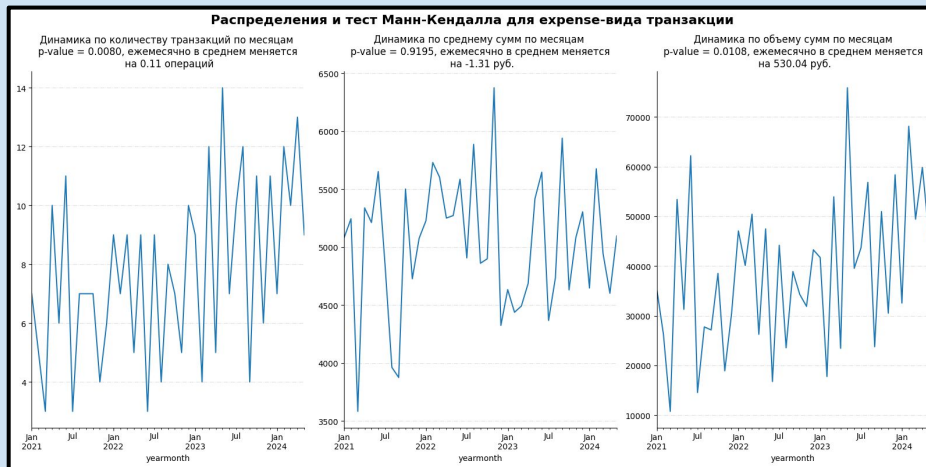


Рисунок 5 - тест Манн Кендалла по количеству, среднему и объему за месяц expense-транзакций

Expenses имеют **статистический** тренд к росту как по количеству, так и сумме за месяца

Income по суммам в месяц опережают expenses всего на 1,84 руб.



Сковывание финансовых ресурсов для развития и...

в недалекой перспективе объем затрат по тренду превысит объем дохода, вызвав убыток

Проблема №3

Уменьшение ресурсов из-за влияния фрод-транзакций

Доля фрод транзакций = **6,69%**

Медиана фрод БОЛЬШЕ не фрод транзакций на 600 руб.

Статистически, нет связи не по среднему распределения, не по медиане

Однако разница в 600 руб. в масштабах бизнеса может быть серьезной даже так. И потому волнует внимания!

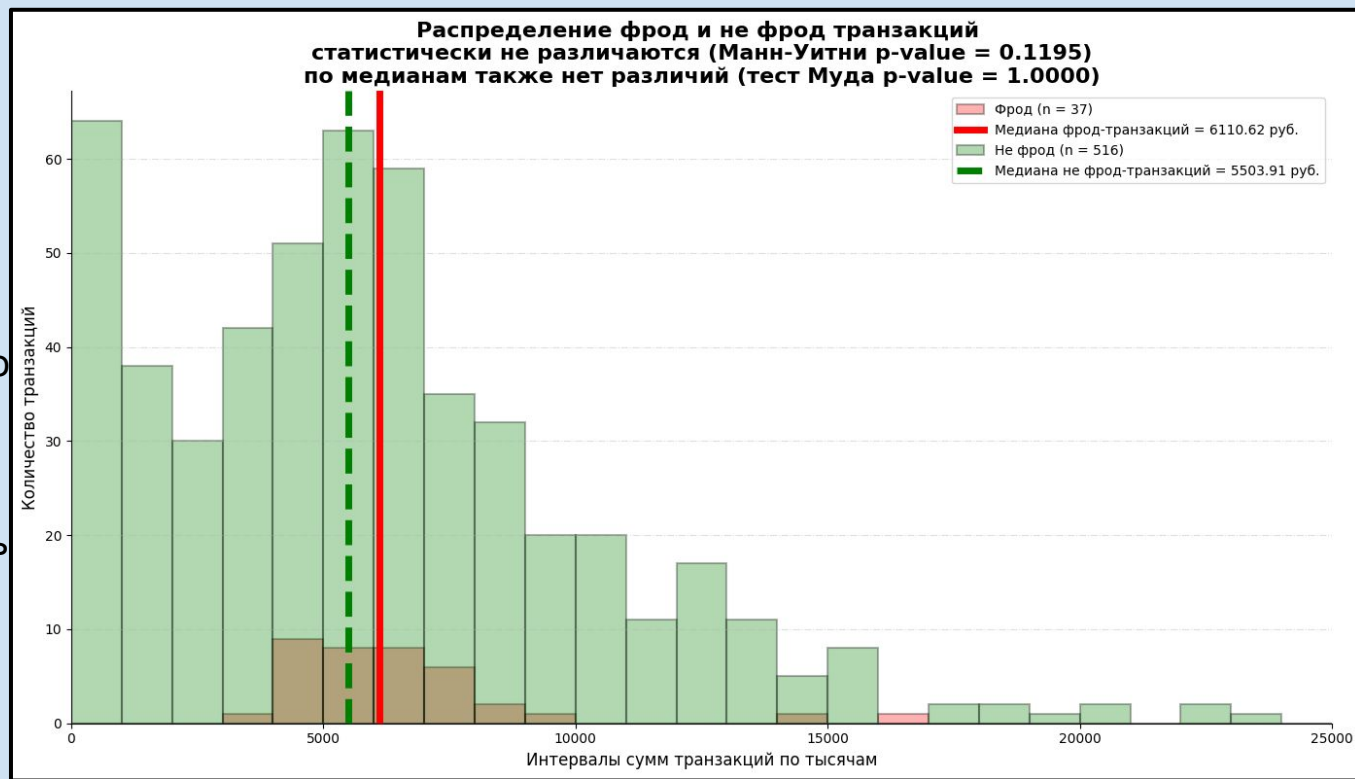


Рисунок 6 - распределение сумм фрод и не фрод транзакций со стат.тестами на медиану и на среднее

Решение проблемы №1 и №2

Уменьшение доли расходов

Распределение категорий по затратности одной транзакции (в среднем)

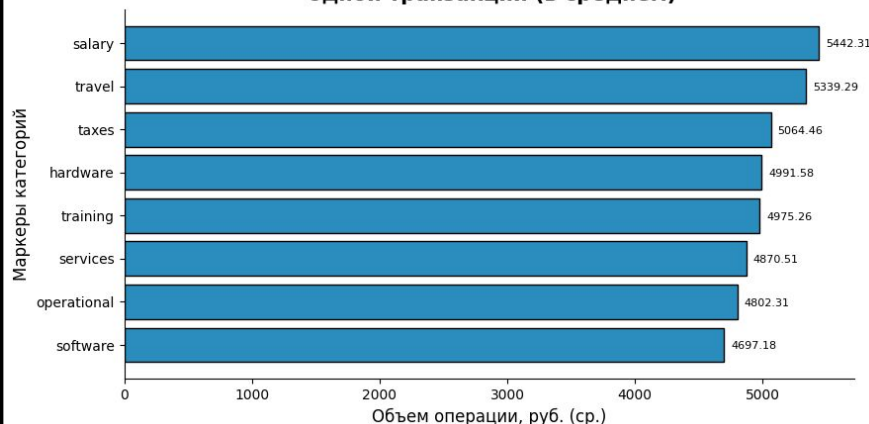


Рисунок 7 - распределение средней суммы транзакции-расхода по категориям

Наиболее затратные категории - это зарплата и путешествия

так почему бы их и не минимизировать - отдавая предпочтения приоритетным действительно полезным направлениям?

Распределение проектов по наибольшим затратам

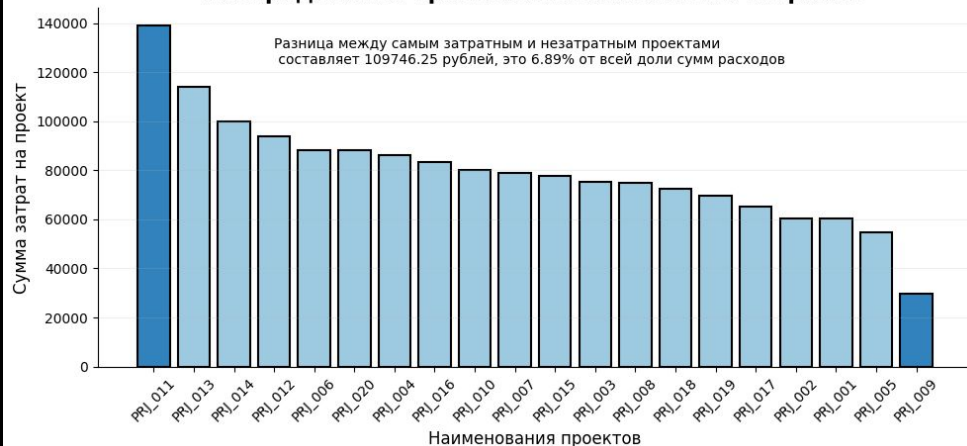


Рисунок 8 - распределение проектов по общей сумме расходов транзакций на них

Компания имеет целых **20** активных проектов, пожирающих ресурсы. И далеко не каждый из них выгодный.

Возможно, пора отбросить те, что не окупают себя и не окупят в ближайшем будущем?

Топ 10 признаков по важности

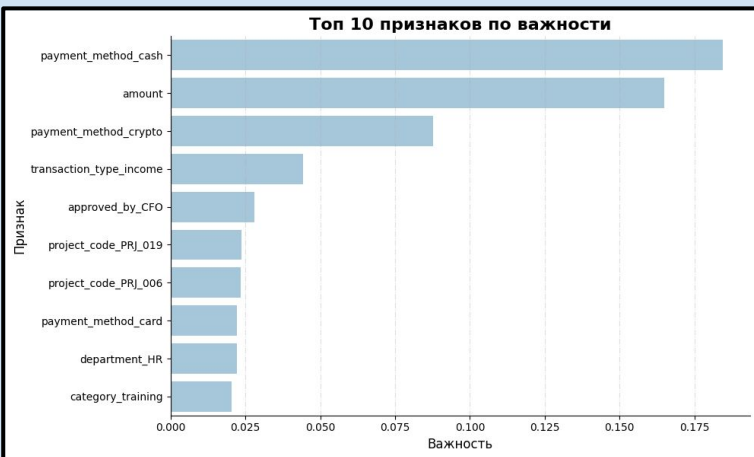


рисунок 9 - топ признаков фрод/не фрод randomforestclassifier

Решение проблемы №3

Комплексная оценка транзакций на фрод

Хоть в датасете всего 7% фрод-транзакций - этого вполне достаточно для понимания общих признаков их.
(напр. оплата наличными или криптой).

Однако следует трактовать признаки осторожно, поскольку их большая часть может принадлежать и обычным транзакциям

Распределение поставщиков по типам транзакций

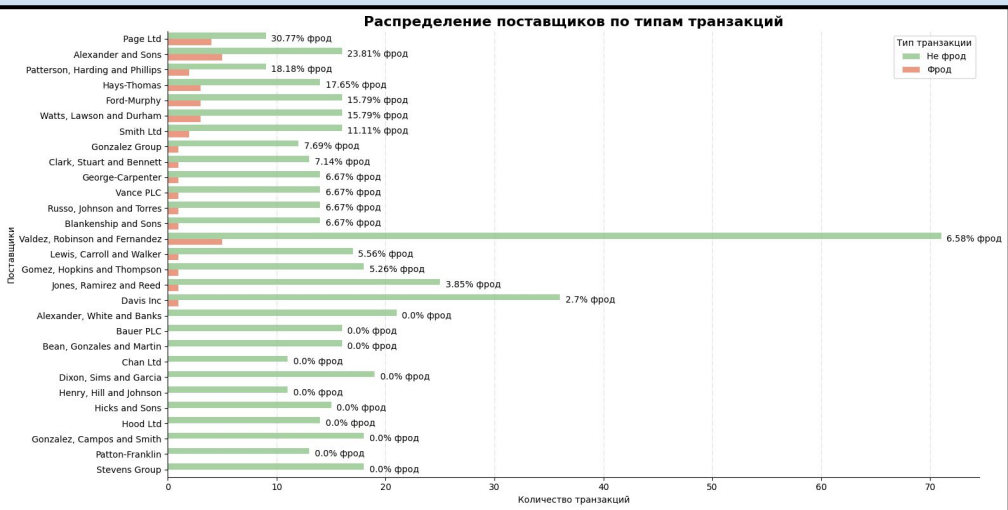


рисунок 10 - распределение поставщиков/контрагентов по фрод/не фрод транзакциям

В свете этого я предлагаю комплексный подход с присуждением баллов за каждый подозрительный признак в транзакции.

- транзакция с поставщиком, у которого 30% транзакций - фрод? +1 балл к фроду
- транзакция категории training? +1 балл к фроду

Набирается 3-5 очков фрода - и операция отклоняется.
Это должно минимизировать лжесрабатывания и, в целом, сами фрод транзакции