

Содержание

1	Проблемы и задачи	2
2	Описание датасета	3
3	Поиск причин неоплаченных заявок	5
4	Поиск лишних операций	7
5	Поиск повторных проверок.	9

1 Проблемы и задачи

- Не все счета успешно доходят до оплаты. Необходимо выяснить, где они останавливаются и почему.
- Определить лишние операции в процессе, определить топ сотрудников, их совершающих и время, затраченное на эти действия.
- В процессе производятся излишние (повторные проверки). Определить причины повторов и излишние затраты времени на них.

Дедлайн: 24 октября, 9:30.

2 Описание датасета

Собраны данные по оплатах примерно за два года.

Счета бывают двух типов: оплата возмещения сотруднику и счёт перед юр. лицом.

- **'Экземпляр'** - id счёта
- **'Операция'** - тип операции, проведённой над счётом. Всего 15 типов.
- **'Дата начала'** - дата начала операции. Самая ранняя: 2021-01-07 02:49:12', поздняя: '2022-12-06 23:42:17'
- **'Тип операции'** - тип операции: ['Получение', 'Согласование', 'Платеж', 'Сбор данных']
- **'Стоимость активности'** - стоимость операции
- **'Сумма'** - сумма счёта
- **'Автоматизация'** - тип операции: ['Автоматизированная', 'Ручная']
- **'Ответственный'** - ФИО сотрудника
- **'Статус счета'** - статус: ['Закрит', 'В процессе']
- **'Тип счета'** - тип счёта по размеру и категории поставщика: ['Счет на малую сумму', 'Счет на среднюю сумму', 'Доверенный поставщик', 'Авансовый отчет', 'Партнер', 'Счет на большую сумму']
- **'Регион'** - регион принятия решения по данному типу конкретной операции
- **'Клиент'** - ФИО клиента
- **'Тип клиента'** - тип клиента: ['Обычный клиент', 'Импульсивный клиент', 'Клиент со скидкой', 'Внутренний клиент', 'Партнер']
- **'Поставщик'** - юр. имя поставщика
- **'Тип поставщика'** - тип поставщика (в основном 'Производственное оборудование'): ['Производственное оборудование', 'Инженерное оборудование', 'Сотрудники', 'Офисные принадлежности']
- **'Отдел'** - отдел принятия решения по данному типу конкретной операции, в основном 'Внутренний контроль'
- **'Пользователь'** - тот, кто принимал решение на данном этапе оплаты счёта
- **'Путь'** - 357 уникальных путей. Это тип цепочки событий процесса.
- **'Тип пути'** - алгоритм выплаты счёта

Экземпляр	Операция	Дата начала	Тип операции	...
IF-1506417	A. Получение счета	2021-06-27 10:00:23	Получение	...
IF-1506417	C. Проверка счета	2021-06-27 10:00:25	Получение	...
IF-1506417	G. Финальная проверка счета	2021-06-27 10:41:50	Согласование	...
IF-1506417	H. Согласование	2021-06-28 10:10:55	Согласование	...
IF-1506417	N. Оплата счета	2021-07-01 17:01:47	Платеж	...

Таблица 1: Пример лога для чека

По обработка заявки заняла 4 дня и 7 часов. В обработке участвовало 3 сотрудника.

Обработка одного платежа имеет множество этапов, проходит согласование у разных сотрудников на разных уровнях и в разных регионах. В данной ситуации различных комбинаций путей заявки слишком много, так что задачу оптимизации не решить методом "пристального взгляда" на гистограмму.

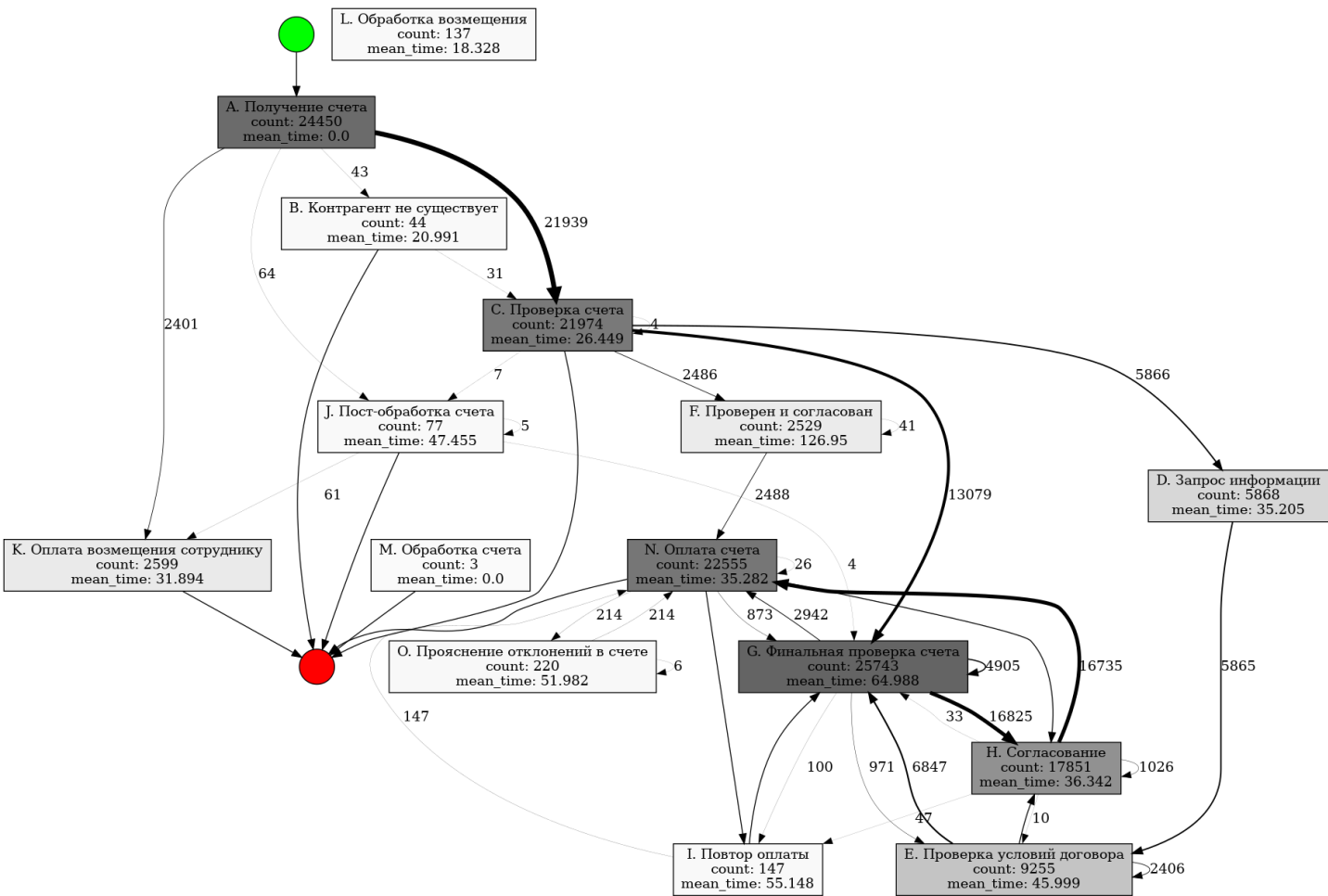


Рис. 1: Граф популярных путей обработки чека.

3 Поиск причин неоплаченных заявок

Для нахождения слабого звена в процессе оплаты счетов использовалась библиотека SberPM.

Рассмотрим цепочки, которые не завершились оплатой.

	index	count	ids	trace_length	unique_activities_num
0	(А. Получение счета, С. Проверка счета)	529	{IF-1507819, IF-1512735, IF-1504335, IF-160086...	2	2
1	(А. Получение счета, J. Пост-обработка счета)	3	{IF-1609640, IF-1604375, IF-1506373}	2	2
2	(А. Получение счета, М. Обработка счета)	3	{IF-1503173, IF-1504556, IF-1503174}	2	2
3	(А. Получение счета, С. Проверка счета, J. Пос...	2	{IF-1609219, IF-1609220}	4	3
4	(А. Получение счета, С. Проверка счета, J. Пос...	1	{IF-1610820}	5	3

Рис. 2: Цепочки, не дошедшие до оплаты.

Видим повторяющиеся фамилии. На данном этапе есть две гипотезы:

- сотрудники систематически не закрывают оплату счёта
- через этих сотрудников проходит значительный поток платежей, в том числе и не оплаченных

Для дальнейшего расследования рассмотрим самый популярный сценарий обработки заявки, не дошедший до платежа.: А. Получение счета → С. Проверка счета.

На всех заявках не был назначен ответственный. Заявки останавливались на данных сотрудниках:

Пользователь	Количество
Система	529
Кондрат Григорьевич Ковалев	525
Макаров Измаил Брониславович	2
Эмилия Константиновна Суворова	1
Нинель Антоновна Кудряшова	1

Разберёмся, как в принципе проходили платежи через сотрудника Кондрат Григорьевич Ковалев. Для этого соберём все счета, проходившие через него и визуализируем на графе со временными и количественными метриками.

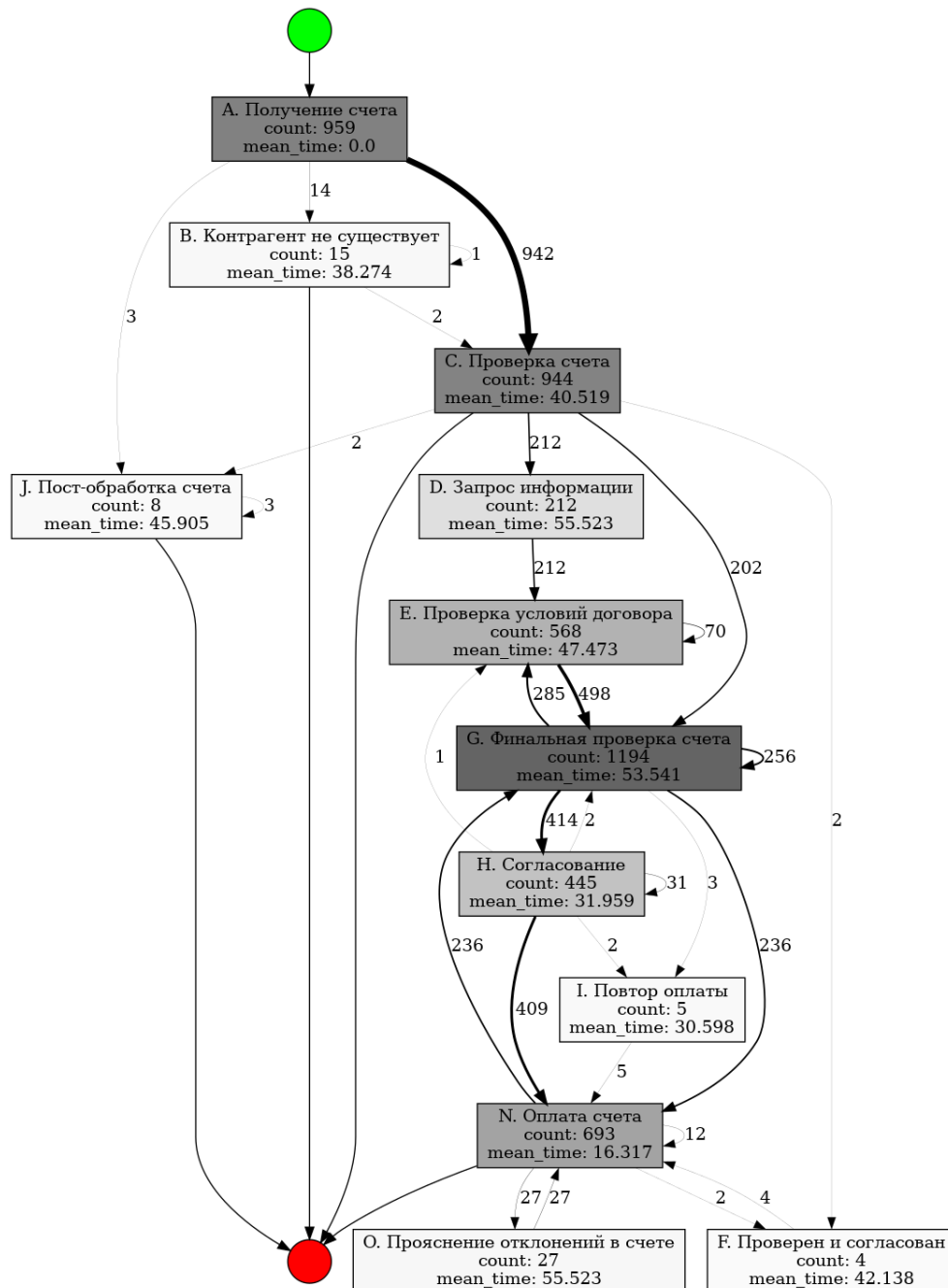


Рис. 3: Работа подозрительного сотрудника.

Выводы

По какой-то причине руководитель Кондрат Григорьевич Ковалев не назначал ответственного за оплату и счета останавливались на этапе проверки или финального согласования.

4 Поиск лишних операций

Для выявления лишних операций и сотрудников, которые их выполняют предлагается выполнить следующие действия:

1. Выделить признаки, характеризующие каждый чек (объект) на всех этапах его обработки в общем. Важно не использовать признаки, отвечающие за то, какие этапы и через кого проходил объект.
2. Сопоставить каждому объекту время, затраченное на его обработку.
3. Векторизовать выбранные признаки. В данном случае достаточно закодировать через ONE.
4. Понизить размерность и кластеризовать объекты (решил выделить 5 кластеров).
5. В каждом кластере выбрать самые долгие по времени обработки объекты, сравнить их обработку с другими объектами.
6. Выделить сотрудников, выполнявших эти лишние действия, и посчитать затраченное время.

Таким образом можно выделить похожие платежи и выделить лишние действия.

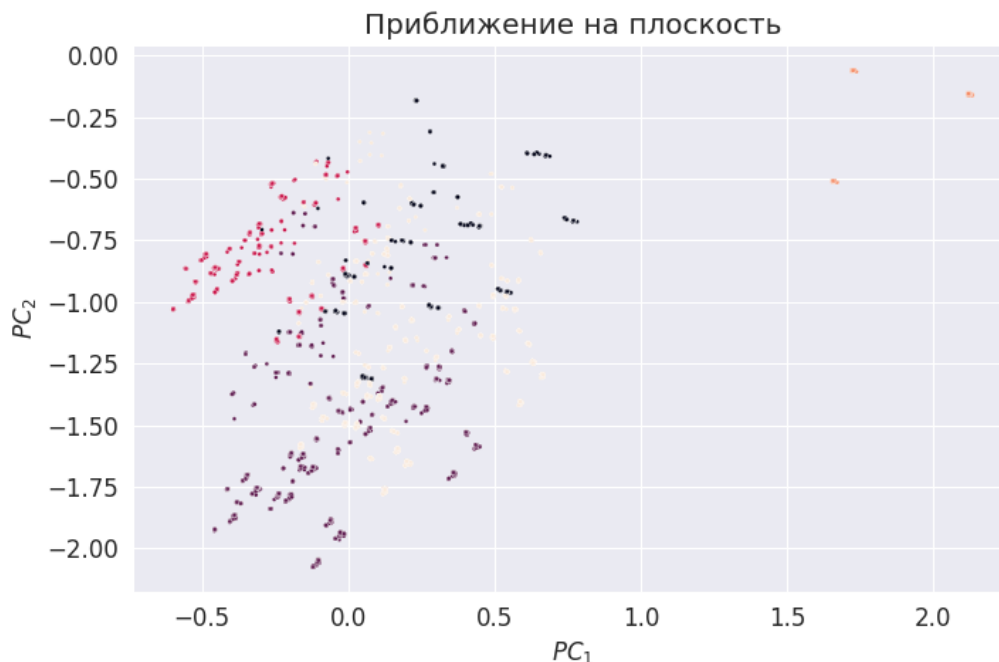


Рис. 4: Распределение чеков.

Лишние действия были найдены по алгоритму:

- Проходим по всем кластерам, рассматриваем плохие объекты (выходящие за 3σ).
- По хорошим объектам ищем счастливый путь как самую частую последовательность.
- Смотрим, как путь плохих объектов отличается от счастливого и записываем лишние действия (с сотрудниками и затраченным временем).

Получил топ сотрудников, выполнявших лишние действия.

excessive_users	excessive_time_mean	excessive_stages_count
Ковалева Надежда Оскаровна	1062.090247	9
Белякова Антонина Юльевна	1055.748839	14
Зиновьева Майя Макаровна	887.343207	11
Ипатий Адрианович Ефимов	476.295139	1
Глеб Валерианович Журавлев	325.766429	2
...

5 Поиск повторных проверок.

Рассмотрим причины заикливания на самых частых путях с $loop_percent > 25$. Визуализируем самые частые циклы.

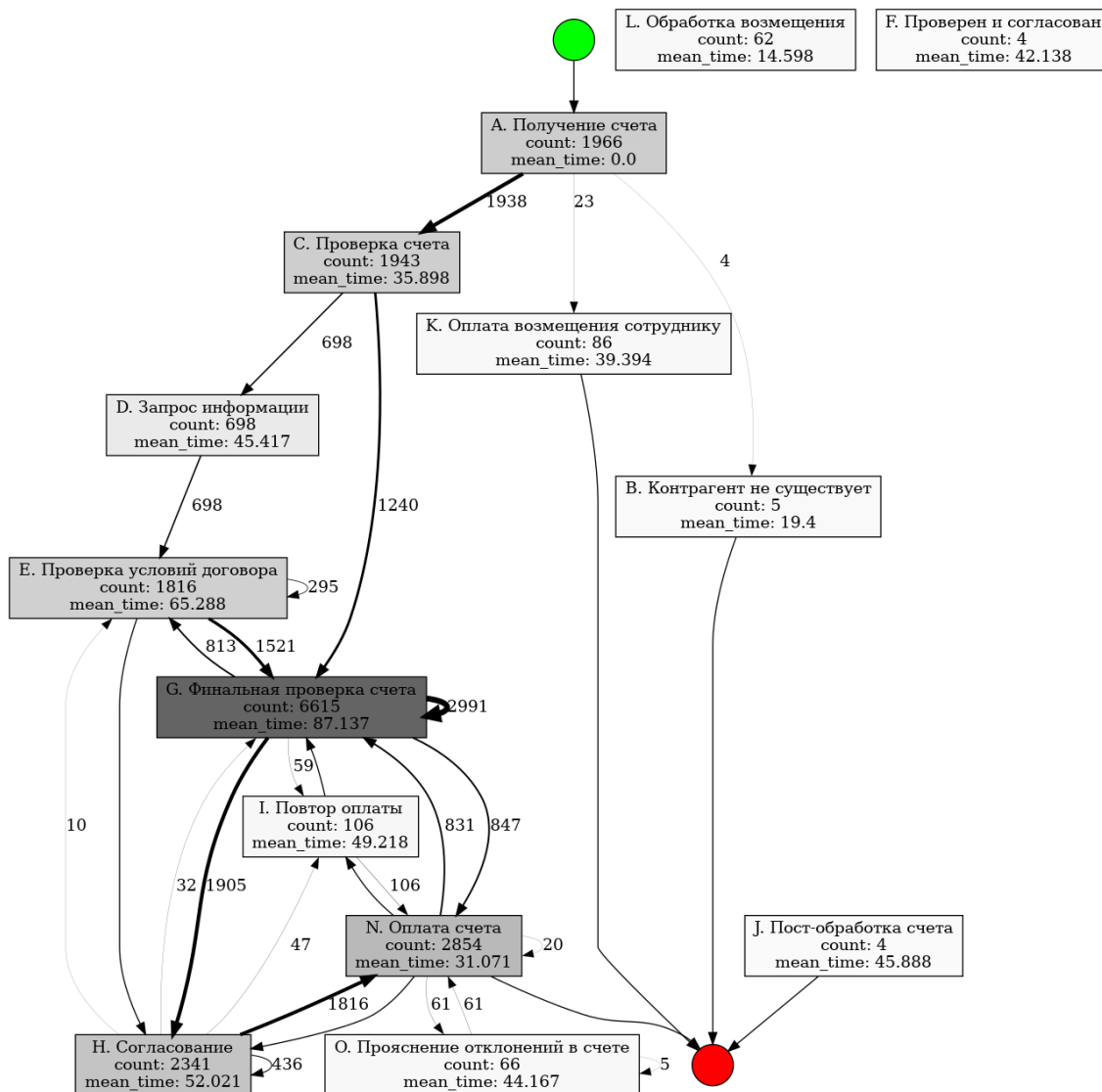


Рис. 5: Схема популярных циклов.

Большинство заикливаний происходит на этапе финальной проверки счёта. Счета отправляются на повторной проверке условий, либо передаются другому пользователю для финальной проверки.

Для оценки времени, затраченного на циклы я просто умножу коэффициент заикленности на среднее время выполнения цепочки операций, то есть по формуле:

$$count \times loop_percent \times mean_duration$$

Затраченное на циклы время: 24960 дней;

Затраченное время на обработку чеков в общем: 163007 дней;

Доля времени, потраченного на циклы: 15%.